THE UNIVERSITY *of* EDINBURGH

Developing novel optogenetic tools in *Caenorhabditis elegans*

Kieran Garry Baxter

Doctor of Philosophy in Neuroscience

The University of Edinburgh

2022

Declaration

I declare that this thesis has been composed solely by myself and that all work is my own, except where explicitly stated otherwise. This work has not been submitted, in whole or in part, for any other degree or professional qualification.

Kieran Garry Baxter
August 2022

Acknowledgements

First and foremost, thank you to my supervisor, Sebastian Greiss, for giving me the opportunity to pursue this PhD. His guidance, patience, and support, have been invaluable for my MSc project, my PhD, and my scientific career.

Thank you also to the other worm PIs, Maria Doitsidou and Emanuel Busch, and all the members of the worm labs. Particularly, thank you to Daniel-Cosmin Marcu for his statistics wizardry, Lloyd Davis, Andromachi Xypnitou, and Zhiyan Xi, for showing me the ropes when I first came to the Greiss lab and for helping me throughout, and to Cristina Sampedro Torres-Quevedo, Jack O'Shea, Samanta Paz Recio, Feng Xue, Angel Goutou, Lourdes Riquelme Domínguez, Viktoria Bajuszova, Eugenia Goya, Gintarė Bagačionkaitė, Ailish Tynan, Jose Vázquez Rodríguez, and Deep Prakash, for making coming to the lab a joy and helping me keep my sanity through long nights and failed experiments. Thank you also to Pasin Chatchalermwit and Jinkai Zhang, for putting up with my attempts to supervise and train.

The HRB technical team also deserve my full gratitude for keeping the place functional despite our best attempts: Carol Wollaston, James Griffiths, Andrew Harrower, Alastair Couttie, Isobel Couttie, Krisztina Vinko, Michael Molinek, and Yvonne Mushet.

Thank you to our collaborators, Henrik Bringmann and Inka Busack, for agreeing to take on and test the tools we have developed within their own work. It is always a good point to know that what we developed works and is useful to other researchers.

Finally, thank you to all the friends I've made throughout my years in Edinburgh, and to my parents for supporting me over the past nine years.

Abstract

Proteins are biopolymers constructed from 20 canonical amino acids which, while limited in number, work together to carry out an extensive variety of functions essential to life. Genetic code expansion allows for the site-specific incorporation of non-canonical amino acids with novel functions not found in nature, creating proteins with unique properties that can be applied to tasks that are otherwise unachievable.

Photocaged amino acids contain a bulky 'caging' group conjugated to the side chain of a canonical amino acid. This caging group can render the protein inactive by blocking its active site, but can be rapidly removed by illumination with 365 nm light, restoring the canonical amino acid and allowing for the photoactivation of the protein. Photocaged amino acids have been previously used in *Caenorhabditis elegans* to develop tools for controlling gene expression, apoptosis, and protein-protein interactions. In this thesis, I give a general introduction on genetic code expansion and the use of photocaged amino acids, as well as the use of *C. elegans* as a model organism. I then explore the use of photocaged amino acids in developing optogenetic tools in *C. elegans*.

The main focus of this thesis is the development of a photocaged FLP recombinase which can drive gene expression with high spatiotemporal control. I show that this can be done in two ways, by photocaging either the catalytic lysine residue, or the catalytic tyrosine residue. In both cases, expression of a target gene is inhibited by the use of an FRT-flanked transcriptional terminator, which can be excised by the photoactivated FLP to drive gene expression with single-cell resolution. This system provides a valuable tool for the study of functions of cells that could not be targeted by other methods. We demonstrate this by using photocaged FLP to drive expression of a channelrhodopsin in the PVC neurons, which do not have their own cell-specific promoter, allowing us to study the contributions of the PVCs in locomotion and sleep.

I also discuss the use of photocaged amino acids in the generation of gene-targeted random mutagenesis tools, an area of genetic screening that is so far underutilised in *C. elegans*. These tools are designed to be targeted to a specific gene of interest and generate a random set of mutations in the area, which would facilitate the study of that gene in reverse genetics experiments.

Lay Summary

Proteins are molecules which contribute to almost every possible function in a cell, from maintaining the structure of the cell, to replicating DNA, to catalysing chemical reactions. Proteins are made from a combination of 20 amino acids which each have unique chemical properties. The blueprint for any specific protein is contained within our DNA, which determines which amino acids make up the protein, as well as where and when in the body the protein is made.

Recently, researchers have developed ways of including new amino acids into proteins, beyond the 20 natural amino acids. These new amino acids can have chemical properties that are not found in the natural amino acids, allowing researchers to make designer proteins with novel functions. For example, photocaged amino acids can block a protein's function until illumination with a specific wavelength of light, which converts the photocaged amino acid into a normal amino acid. In this way, a protein can be switched on by activation with light, which is a powerful tool for researchers to use. In this thesis, I investigate the use of photocaged amino acids in developing new tools for researchers to use in *Caenorhabditis elegans*, a small nematode worm commonly used for studying the nervous system, disease, and aging.

The first tool I developed is a tool which can be used to activate the expression of any other protein. With this tool being controllable by light, researchers can ensure that any desired protein is present in specific cells at specific times, which can be used to aid in the study of that protein or the cell. As an example, we focus on using our tool to express a protein in neurons which can be used to activate the neuron. In this way, we can use our tool to express the protein only in a single neuron, which would allow us to study the role of that neuron.

The second set of tools I investigate with photocaged amino acids are those that can cause changes in a specific region of DNA. Changes in the DNA can cause proteins to be made with different amino acids in them, which can disrupt the function of the protein and is often a cause of disease. By forcing changes in a specific protein's DNA code, we can study the function of that protein. These proteins in the worm often have equivalent proteins in other animals, including humans, and studying them in the worm provides an easier way of figuring out how and why diseases are caused in humans.

Abbreviations

A site	Aminoacyl Site
aa-AMP	Aminoacyl-Adenylate
aaRS	Aminoacyl-tRNA Synthetase
aa-tRNA	Aminoacyl-tRNA
AID	Activation-Induced Cytidine Deaminase
AMP	Adenosine Monophosphate
AsLOV2	LOV domain from *Avena sativa* phototropin 1
ATP	Adenosine Triphosphate
ATR	All-Trans Retinal
att$B$	Attachment site Bacteria
att$P$	Attachment site Phage
att$L$	Attachment site Left
att$R$	Attachment site Right
Boc lysine	N^{ε}-tert-butyloxycarbonyl-L-lysine
BONCAT	Bioorthogonal Non-Canonical Amino Acid Tagging
Cas9	CRISPR-Associated Protein 9
CDS	Coding Sequence
CFP	Cyan Fluorescent Protein
ChR	Channelrhodopsin
ChR2	Channelrhodopsin 2
Cre	Cyclisation Recombinase
CRISPR	Clustered Regularly Interspersed Short Palindromic Repeats
DanAla	2-amino-3-(5-(dimethylamino) naphthalene-1-sulfonamido) propanoic acid
dCas9	Dead Cas9
DNA	Deoxyribonucleic Acid
E site	Exit Site
EMS	Ethyl Methanesulphonate
ENU	N-ethyl-N-nitrosourea
ePDZ	Engineered PDZ (PSD-95, Dlg1, ZO-1) domain
F2A	2A self-cleaving peptide from Foot-and-mouth Disease Virus
FlEx	Flip-Excision

FLP	Flippase
FMN	Flavin Mononucleotide
FRT	FLP Recognition Target
GCE	Genetic Code Expansion
GFP	Green Fluorescent Protein
GOI	Gene of Interest
gRNA	Guide RNA
HA	Haemagglutinin
HCK	Hydroxycoumarin Lysine
HL	Happy Linker
HPLC	High Performance Liquid Chromatography
Indel	Insertion or Deletion Mutation
K-NPY	Lysine-Nitropiperonyl Tyrosine (Dipeptide)
L1	Larval Stage 1
L2	Larval Stage 2
L3	Larval Stage 3
L4	Larval Stage 4
LANS	Light Activated Nuclear Shuttle
LB	Lysogeny Broth
LOV	Light-Oxygen-Voltage sensing domain
LUCA	Last Universal Common Ancestor
Mb	*Methanosarcina barkeri*
MiniSOG	Mini Singlet Oxygen Generator
Mj	*Methanocaldococcus jannaschii*
Mm	*Methanosarcina mazei*
MNPY	Methylnitropiperonyl Tyrosine
MosSCI	Mos-1-Mediated Single Copy Insertion
mRNA	Messenger RNA
ncAA	Non-canonical Amino Acid
nCas9	Nicking Cas9
NES	Nuclear Export Sequence
NGM	Nematode Growth Medium
NLS	Nuclear Localisation Sequence
NMD	Nonsense Mediated Decay

NPY	Nitropiperonyl Tyrosine
NPYRS	Nitropiperonyl Tyrosyl-tRNA Synthetase
O-ribosome	Orthogonal Ribosome
ONBY	Orthonitrobenzyl Tyrosine
P site	Peptidyl Site
PATC	Periodic A_n/T_n Cluster
pcAA	Photocaged Amino Acid
PCC	Photocaged Cysteine
PCK	Photocaged Lysine; Methyl-*o*-nitropiperonyl Lysine
PCKRS	Photocaged Lysyl-tRNA Synthetase
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
piRNA	Piwi-Interacting RNA
PolI	DNA Polymeras I
Pre-mRNA	Pre-messenger RNA
PYLIS	Pyrrolysine Insertion Sequence
PylRS	Pyrrolysyl-tRNA Synthetase
RFP	Red Fluorescent Protein
RNA	Ribonucleic Acid
RNAi	RNA Interference
rRNA	Ribosomal RNA
scFV	Single Chain Fragment Variable Antibody
SECIS	Selenocysteine Insertion Sequence
SL	Spliced Leader
SNP	Single Nucleotide Polymorphism
SSR	Site-Specific Recombinase
TALEN	Transcription Activator-Like Effector Nuclease
TB	Terrific Broth
TBD	Thioredoxin-Binding Domain
tRNA	Transfer RNA
UTR	Untranslated Region
UV	Ultraviolet
WT	Wild-type
YFP	Yellow Fluorescent Protein

ZFN	Zinc-Finger Nuclease

Contents

Chapter 1

Introduction

1.1 The Central Dogma of Molecular Biology

Proteins are fundamental biomolecules required for all known forms of life. As an integral part of cellular machinery, proteins are involved in a diverse range of functions, including enzymatic processes, signalling between and within cells, and maintaining the structure of cells. Proteins are polymers made up of amino acids, with the specific sequence determining the final 3D structure of the protein, as well as the functions it performs. In turn, the sequence of the protein is determined by the sequence of DNA, which is processed through an RNA intermediate to give rise to the protein. This transfer of information from DNA to RNA to protein is referred to as the 'Central Dogma of Molecular Biology'[1,2] (Fig 1). This dictates that DNA can be replicated to give rise to duplicate copies of DNA, or it can be transcribed to give rise to RNA, specifically a messenger RNA (mRNA), which is then translated into a protein. While information can flow backwards from mRNA to DNA via the process of reverse transcription, it cannot flow back from protein to mRNA.

Figure 1: The Central Dogma of Molecular Biology The flow of biological information generally proceeds from DNA to RNA to Protein. Reverse transcription can occur to reverse the flow from RNA to DNA, while DNA can be replicated.

The transfer of information between DNA and RNA, either by transcription or reverse transcription, is relatively simple. Both DNA and RNA are biopolymers, like proteins, which are made up of four bases: adenine (A), thymine (T), guanine

(G), and cytosine (C) in the case of DNA; or adenine, uracil (U), guanine, and cytosine in the case of RNA. These sets of bases are complementary to each other, with adenine naturally aligning and interacting with thymine or uracil, and guanine interacting with cytosine. As a result, the process of transcription is relatively simple, with an RNA polymerase able to carry out the majority of the steps. During transcription, the RNA polymerase binds the DNA double helix and separates the two strands, then simply reads along the DNA and generates an RNA strand by adding each RNA base complementary to the DNA base. Reverse transcription occurs by much the same mechanism, though in this case using a Reverse Transcriptase to generate complementary DNA from an RNA template. Translation, on the other hand, is a much more involved and complex process. This is primarily because of two reasons. Firstly, unlike DNA/RNA, there are no innate complementary interactions involved between RNA and protein, meaning that the cell must use additional machinery to assign each amino acid to an RNA sequence. Secondly, while RNA only uses four unique bases, proteins are generally made up of 20 unique amino acids, meaning a single base cannot possibly code for a single amino acid. Instead, amino acids are encoded by three bases, referred to as a triplet codon. With four unique bases, this gives rise to 64 possible triplet codons that must each be handled by the cell machinery. Of these, three are designated as stop codons, which act as signals to terminate translation, while the remaining 61 are assigned as 'sense' codons, mapping to the 20 individual amino acids. Translation is the process by which the triplet codons on the mRNA are read by the ribosome and converted into an amino acid chain. In this process, each amino acid has a cognate aminoacyl-tRNA synthetase (aaRS), which loads the amino acids onto a cognate transfer RNA (tRNA). The aminoacylated tRNA is recruited to the ribosome, where each tRNA binds to a specific codon of the mRNA and its amino acid is added to the growing polypeptide chain (Fig 2). As there are 20 amino acids, this process requires 20 different aminoacyl-tRNA synthetases, while the 61 sense codons require 61 unique tRNAs. In reality, not all of the 61 codons have a corresponding tRNA, as wobble base pairing at the third position of each codon means that a minimum of only 32 tRNAs are required for full decoding[3,4].

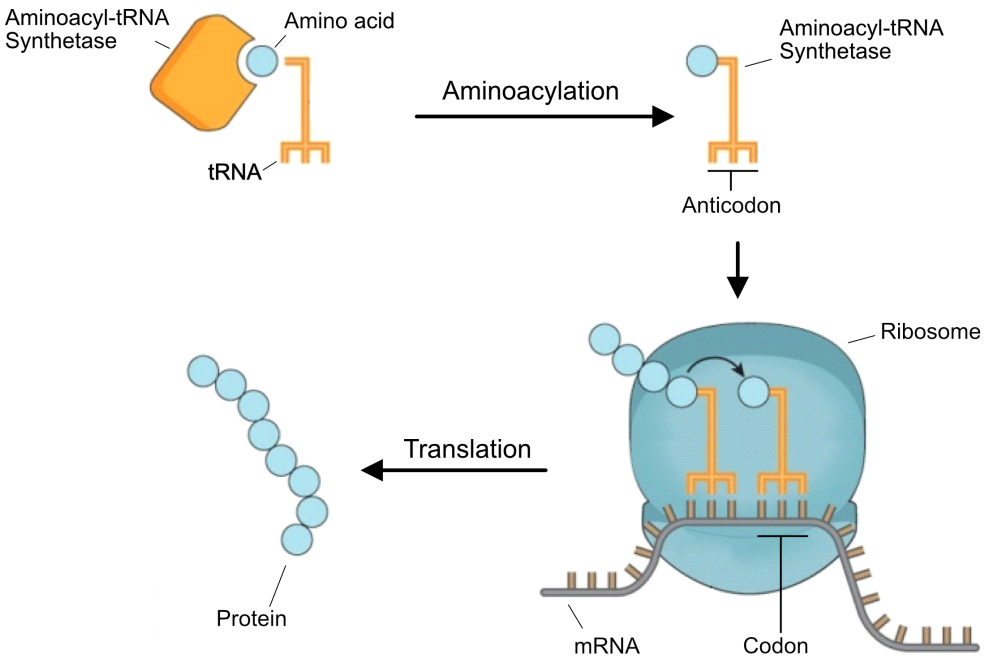

Figure 2: Protein Translation A tRNA is aminoacylated with its cognate amino acid by an aminoacyl-tRNA Synthetase. The aminoacyl-tRNA moves to the ribosome, where its anticodon interacts with a complementary codon. The ribosome attaches the amino acid to a growing polypeptide chain to produce a protein. Adapted from Chin, 2017[5].

1.1.1 Protein Translation

1.1.1.1 Aminoacyl-tRNA Synthetases

As a first step in translation, each amino acid must be attached to an appropriate tRNA. This step is catalysed by a group of enzymes called the aminoacyl-tRNA synthetases, with at least one aaRS for each amino acid[6]. This esterification of the tRNA and the amino acid produces an aminoacyl-tRNA (aa-tRNA), which is then delivered to the ribosome by elongation factors for translation[7].

The esterification reaction driven by the aaRS occurs in a two-step mecahnism[7]. Firstly, amino acid activation occurs, where the aaRS recognises and binds to the amino acid and activates it with adenosine triphosphate (ATP). In the second step, the tRNA attacks the amino acid to generate the aminoacyl-tRNA.

Due to the broad similarities in amino acid structures, aaRSs use a range of methods to determine specificity and bind the correct amino acid, including size exclusion, using charged binding pockets, or using metal ions to bind specific chemical groups[7]. For example, glycine, the smallest amino acid, is bound by its cognate aaRS, GlyRS, using a small, highly negatively-charged binding pocket, which prevents binding of other amino acids[8]. The ThrRS distinguishes between the similarly-structured threonine and valine by using a zinc ion in the binding pocket, which is stabilised by the amino and hydroxyl groups present on threonine[9]. Because valine contains a methyl group instead of the hydroxyl group, it is unable to be bound and activated by ThrRS. Similar mechanisms of zinc ion coordination are used by CysRS and SerRS to prevent binding of noncognate amino acids[7]. In addition to binding specificity, some aaRSs also contain proofreading capabilities, where a noncognate amino acid which has bound the binding pocket is detected and expelled from the enzyme[7].

Upon binding of the relevant amino acid to the aaRS, an ATP also binds within the active site, positioned where the α-carboxylate oxygen of the amino acid can perform a nucleophilic attack on the α-phosphate of the ATP, forming aminoacyl-adenylate (aa-AMP) (Fig 3A). Next, the hydroxyl group from the terminal base in the tRNA attacks the α-carbon in the aa-AMP to transfer the amino acid to the tRNA, generating an aminoacyl-tRNA and releasing the AMP (Fig 3B).

1.1.1.2 Transfer RNAs

The tRNAs are responsible for decoding the mRNA sequence and supplying the correct amino acid for each codon during protein translation. The primary and secondary structures of the tRNAs were first described for the $tRNA^{Ala}$ from yeast[10], and have since been found to be conserved across almost all tRNAs. tRNAs generally consist of between 76 and 90 bases, approximately 12% of which are modified variants of canonical bases, including Ribothymidine, Pseudouridine, Dihydrouridine, and Inosine[3]. One of the most highly conserved features of the primary structure of tRNAs is the CCA sequence contained at the 3' terminus, which is the site of aminoacylation. Another feature within the primary structure of tRNA, though variable amongst different tRNA species, are the determinant and anti-determinant identity elements, which are nucleotides which regulate the binding of the cognate aaRS, or block the binding of the noncognate aaRS,

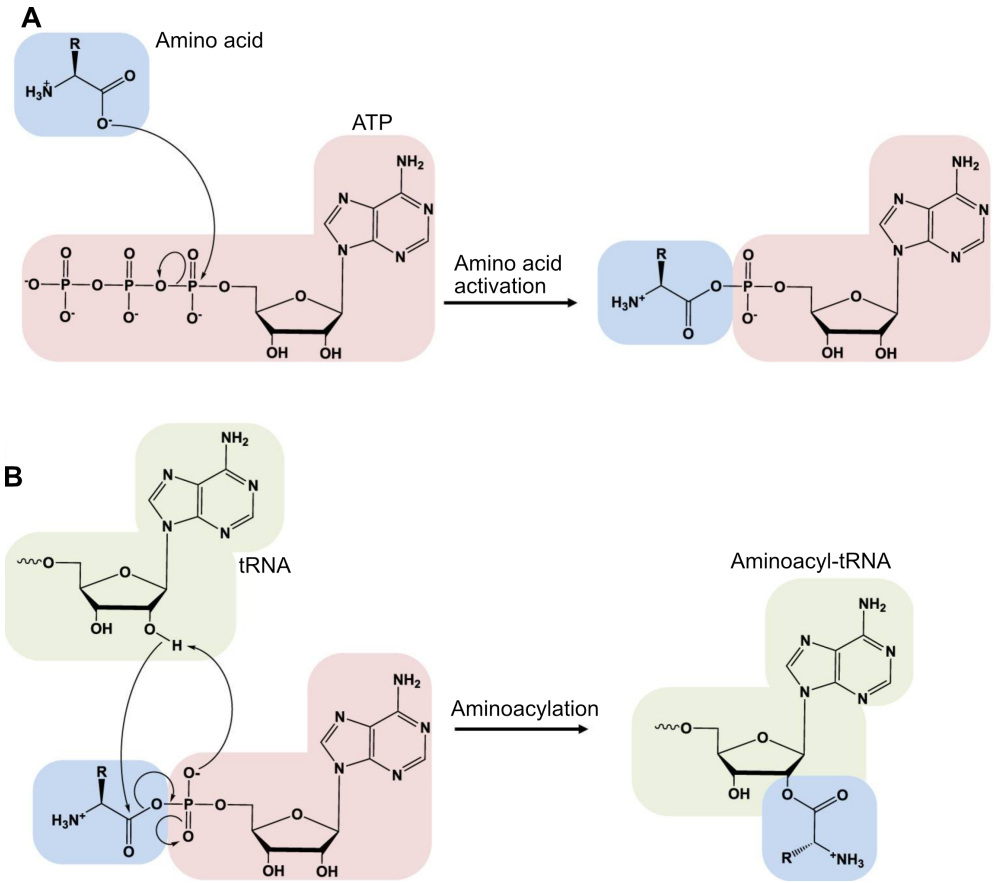

Figure 3: tRNA Aminoacylation Mechanism **(A)** The amino acid is activated by ATP. **(B)** The activated amino acid becomes covalently bound to the tRNA. Adapted from Rubio Gomez and Ibba, 2020[7].

respectively[11].

Through internal base pairing, tRNAs form a stable 'cloverleaf' structure, consisting of five individual domains (Fig 4A). Firstly, the acceptor stem is formed by base pairing of the 5' terminal and 3' terminal nucleotides, just upstream from the CCA sequence. Moving upstream from the 3' end, the next structure is the TΨC stem-loop, named for the Thymine-Pseudouridine-Cytosine (TΨC) sequence found in the loop, which is involved in binding to the ribosome[12]. Next is the variable arm, which varies in size and is thought to increase the specificity of the tRNA for the cognate aaRS[13]. The anticodon loop is directly opposite the acceptor stem and contains three bases which are complementary to the mRNA codon, and is therefore responsible for decoding the mRNA. Along with the acceptor stem-loop, the anticodon stem-loop contains the majority of the determinant and anti-determinant identity elements for aaRS recognition.

Finally, the D stem-loop, named for the Dihydrouridine (D) present within the loop, is involved in stabilising the tertiary structure of the tRNA[3]. In the tertiary structure of tRNAs, an L-shape is formed with one end containing the acceptor stem and the TΨC stem-loop, while the other end contains the D stem-loop, the variable arm, and the anticodon stem-loop (Fig 4B)[14].

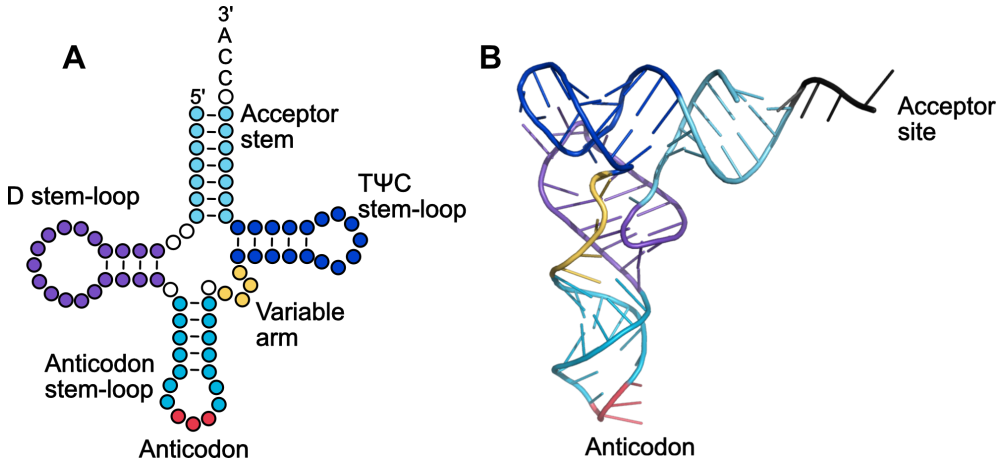

Figure 4: Structure of a tRNA **(A)** Secondary structure of a tRNA. tRNA adopts a cloverleaf structure with several distinct domains. **(B)** Tertiary structure of a tRNA. tRNA adopts an L-shape with one end containing the acceptor stem to bind the amino acid, and the other end containing the anticodon stem to interact with the mRNA. This structure shows the yeast $tRNA^{Phe}$, pdb=4TNA[15], with domains coloured to match those in **(A)**.

Once the tRNA has bound to its amino acid, the aminoacyl-tRNA must then be delivered to the ribosome to provide the amino acid for translation. This is performed in prokaryotes by the elongation factor EF-Tu, or in eukaryotes and archaea by the homologous EF-1A[16], which binds to the aminoacyl-tRNA transports it into the A-site of the ribosome to initiate translation.

1.1.1.3 Ribosomes and Translation

The ribosome is a large complex consisting of a small and large protein subunit, along with ribosomal RNAs (rRNA), though the exact composition, size, and numbers of proteins and RNAs contained in each subunit differ across domains. The ribosome components are typically identified by their sedimentation coefficients, given in units of S (Svedberg), which indicates the speed at which the protein or RNA travels through a solution during

centrifugation. In prokaryotes and archaea, the ribosome is 70S and consists of a 30S small subunit and a 50S large subunit[17]. The 30S small subunit consists of 21 ribosomal proteins and a 16S rRNA, while the larger 50S subunit consists of 33 ribosomal proteins, a 23S rRNA, and a 5S rRNA[18]. The eukaryotic ribosome is slightly larger at 80S, containing a 40S small subunit and a 60S large subunit. The 40S small subunit contains 33 ribosomal proteins and an 18S rRNA, while the 60S subunit contains 50 proteins and a 5S rRNA, a 5.8S rRNA, and a 28S rRNA[18].

The small subunit is primarily involved in binding to the mRNA and the anticodon loop of the tRNA, while the large subunit binds to the tRNA at the acceptor stem and catalyses the peptidyl transferase reaction to form the growing polypeptide chain. Both subunits form three tRNA-binding sites, designated the A (aminoacyl), P (peptidyl), and E (exit) sites[19], with translation occurring in steps involving each tRNA-binding site in turn. Firstly, translation initiation occurs where the ribosome binds to the mRNA and positions itself at the start of the reading frame. Next, elongation of the polypeptide occurs across the reading frame. Finally, termination of translation occurs and the mRNA and newly-formed protein are released from the ribosome.

Initiation of translation differs between prokaryotic and eukaryotic ribosomes, though both begin translation at an initial methionine residue. Archaeal translation initiation is not well understood, though likely is most similar to prokaryotic initiation with some other elements shared with eukaryotic mechanisms. In prokaryotes, the initiator tRNA, $tRNA^{fMet}$ is aminoacylated with methionine, which is then formylated to generate fMet-$tRNA^{fMet}$, which is more stable than methionine. An elongation factor, IF2, binds to fMet-$tRNA^{fMet}$, which together are then bound by the small subunit. After initiator tRNA binding, the mRNA then binds to the small subunit by recognition of the ribosome-binding Shine-Dalgarno sequence, which is complementary to the 16S rRNA. This binding places the start codon, AUG, in the P site of the ribosome. In eukaryotes, the initiator tRNA is $tRNA_i^{Met}$, which is distinct from the $tRNA^{Met}$ used during elongation. $tRNA_i^{Met}$ binds methionine to generate Met-$tRNA_i^{Met}$, which is then bound by eIF2 and transported to the small subunit. The mRNA is then recognised and bound by the ribosome by its 5'-cap, with the ribosome then moving along the mRNA until it encounters the AUG start codon. In both prokaryotes and eukaryotes, binding of the initiator tRNA and mRNA to

the small subunit then promotes binding of the large subunit to form the full ribosome complex[17], which can then proceed with the elongation and terminator steps, which are broadly similar between prokaryotes and eukaryotes.

During elongation, the growing polypeptide chain, or the initial methionine if no elongation has occurred yet, is kept in the P site of the ribosome as a peptidyl-tRNA (Fig 5A). An aminoacylated tRNA is transported to the ribosome by elongation factors EF-Tu or EF-1A, which deliver the aa-tRNA to the A-site of the ribosome (Fig 5B). If the new aa-tRNA in the A-site contains an anticodon loop complementary to the mRNA codon, then a peptidyl transferase reaction occurs to transfer the polypeptide chain from the P-site to the A-site and form a new peptide bond between the growing polypeptide chain and the new amino acid (Fig 5C). The now deacylated tRNA in the P-site is transferred to the E-site, while the peptidyl-tRNA is transferred to the P-site, freeing up the A-site for the next aa-tRNA to enter[17]. The E-site has low affinity for the tRNA, with the tRNA quickly dissociating from the site or being displaced during the next round of elongation (Fig 5D).

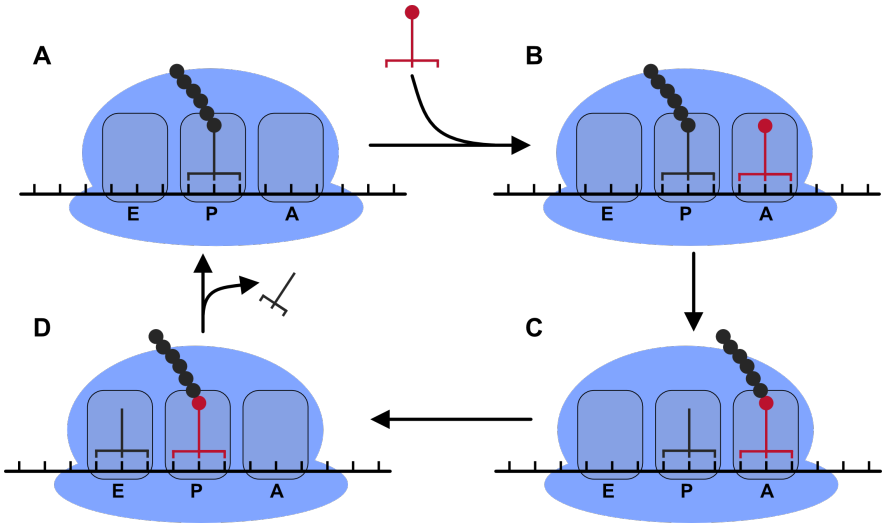

Figure 5: Elongation in the Ribosome **(A)** At the start of each step, the polypeptide chain is attached to a tRNA in the P (Peptidyl) site. **(B)** A new aminoacyl-tRNA enters at the A (Aminoacyl) site. **(C)** A peptidyl transferase transfers the polypeptide chain to the aminoacyl-tRNA in the A site. **(D)** Peptidyl-tRNA is transferred to P site and deacylated tRNA is transferred to E (Exit) site, where it is removed from the ribosome.

Translational termination occurs when the ribosome reaches an in-frame stop

codon, UAA (ochre), UAG (amber), or UGA (opal). These are recognised by release factors, eRF1 in eukaryotes or RF1 and RF2 in prokaryotes. In eukaryotes, eRF1 recognises all three stop codons, while prokaryotic RF1 recognises UAA and UAG, and RF2 recognises UAA and UGA[17]. When a stop codon enters the A-site, the release factors bind to the codon, which triggers the hydrolysis and release of the polypeptide chain from the P-site tRNA[19]. The release factors also recruit ribosome recycling factors, which release the mRNA from the ribosome and cause dissociation of the ribosomal subunits in order to begin translation of the next mRNA transcript[19].

1.1.1.4 Exceptions to the Genetic Code

For efficient and consistent protein translation to occur, a strict set of rules governing the codon-amino acid relationship is required. This is referred to as the genetic code, and was first partially deciphered in *E. coli* in the early 1960s[20], when 50 of the possible 64 triplet codons were assigned to amino acids. Since then, all 61 sense codons and 3 nonsense codons have been determined and, surprisingly, found to be highly conserved across all known forms of life, with only minor changes in a small number of organisms. This universality of the genetic code goes back at least as far as the Last Universal Common Ancestor (LUCA), though beyond that it is unknown if pre-LUCA life contained differing genetic codes[21], and since LUCA only a small number of changes have been made to the genetic code.

Despite the apparent universality, a study by Ambrogelly *et al*[22] noted a total of 16 changes found in nuclear and mitochondrial genetic codes (Fig 6), with the majority of changes being found in the mitochondria. This is thought to be due to minimisation of the mitochondrial genome, which has led to the loss of tRNA genes, leaving only 22 tRNAs in the mammalian mitochondria and leading to compensatory codon reassignment. A relatively common occurrence of tRNA loss is the loss of a $tRNA^{Ile}$ and subsequent reassignment of the AUA codon to Met. In the canonical genetic code, AUU, AUC, and AUA all code for Ile, while AUG codes for Met. This is unusual, as the majority of tRNAs are not required to differentiate between an A or a G in position 3, and some species utilise post-transcriptional modifications to properly decode and differentiate AUA from AUG[23]. In some other species, the $tRNA^{Ile}$ for AUA has been lost completely, with the AUA and AUG codons being decoded by $tRNA^{Met}$, while Ile is encoded

only by AUU and AUC[24]. This change is found in a number of species, including several fungal and metazoan species, as well as mitochondria. In other cases, codon reassignments have been traced to mutations in the tRNA. For example, in the mitochondria of several yeast species, the $tRNA_{UAG}$, which normally decodes leucine, contains mutations in the anticodon stem-loop which cause it to instead decode threonine[25].

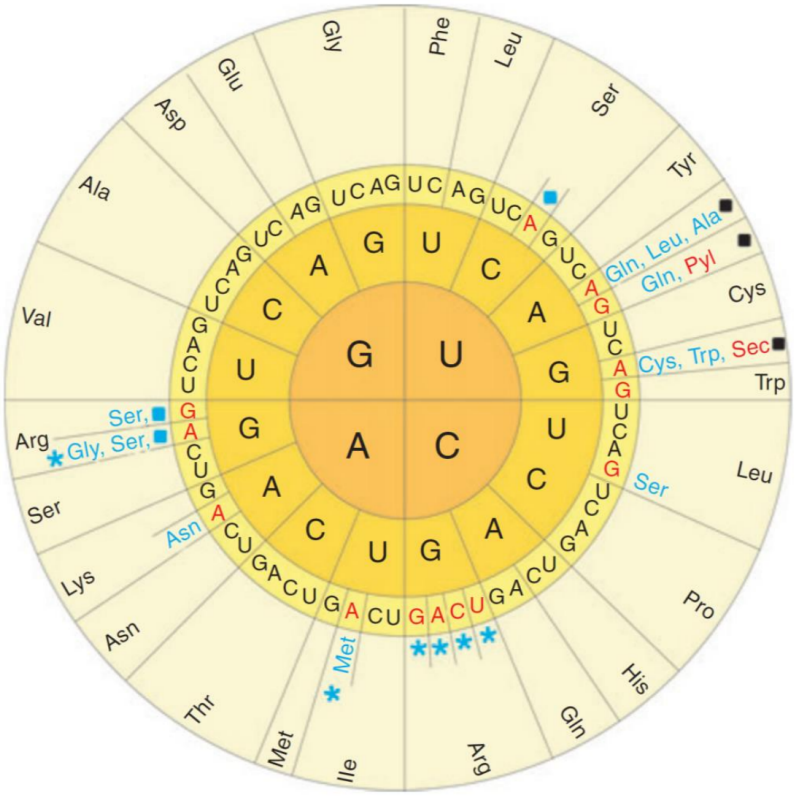

Figure 6: Exceptions to the Natural Genetic Code Codons are represented in the wheel with the inner circle, second circle, and third circle, representing the first, second, and third base in the triplet, respectively. If the final base is shown in red, this codon has been found to encode multiple amino acids. Amino acids in black are those that are encoded by the canonical genetic code, while those in blue show deviations. Red amino acids are the two rarer amino acids beyond the standard 20. Black squares represent stop codons, blue asterisks represent codons that may be unassigned in some organisms. Taken from Ambrogelly *et al*, 2007[22].

While the majority of genetic code changes are found as codon reassignments for the existing amino acids, some organisms have also been found to encode one

of two extra amino acids in addition to the standard 20 amino acids. These are selenocysteine (Sec) and pyrrolysine (Pyl) (Fig 7), often referred to as the 21st and 22nd amino acids, which are encoded by the opal UGA and amber UAG stop codons, respectively. In both cases, the stop codons are not completely reassigned, and so these amino acids require mechanisms to override and suppress the stop codon when being incorporated into the appropriate proteins.

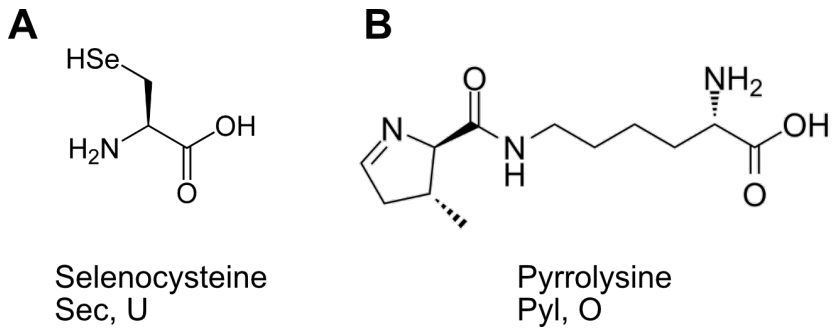

Figure 7: 21st and 22nd Amino Acids Structures of **(A)** Selenocysteine and **(B)** Pyrrolysine.

Selenocysteine is structurally similar to cysteine, but with a selenium atom in place of the sulphur. Unique among amino acids, selenocysteine requires specialised elongation factors, SelB in prokaryotes or EFsec in eukaryotes, to transport the Sec-tRNA^{Sec} to the ribosome, as well as a unique element in the mRNA, named the Selenocysteine Insertion Sequence (SECIS), downstream of the UGA stop codon[26]. SECIS forms a hairpin loop structure, which is recognised and bound by SelB or EFsec. The elongation factor bound to SECIS then localises to the ribosome in high concentrations, preferentially binding the ribosome over the release factor and resulting in incorporation of the Sec amino acid. The formation of Sec-tRNA also follows an unusual pathway, with tRNA^{Sec} initially being charged with a serine residue by SerRS, which is then phosphorylated to produce phosphoseryl-tRNA^{Sec}. The phosphate is then replaced with selenium to produce Sec-tRNA^{Sec}, which can incorporate Sec during translation[26]. Although not found in all organisms, Sec and selenoproteins are present in all domains of life, with humans containing approximately 25 separate selenoproteins[27].

Pyrrolysine is a much less common amino acid, found only in a group of methane-producing archaea, as well as some methane-producing prokaryotes. First discovered in the methanogen methyltransferase in *Methanosarcina barkeri*[28], pyrrolysine is encoded by the UAG codon and utilises its own unique

tRNA synthetase, pyrrolysyl-tRNA synthetase (PylRS), and tRNA, $tRNA^{Pyl}$ pair[29,30]. It was shown that expression of PylRS and $tRNA^{Pyl}$ in *E. coli* grown in the presence of pyrrolysine could lead to incorporation of pyrrolysine at endogenous UAG stop codons[31]. Furthermore, PylRS and $tRNA^{Pyl}$ were found to be contained within a genetic cassette alongside three genes with then-unknown functions, PylB, PylC, and PylD. Expression of the entire cassette in *E. coli* led to incorporation of pyrrolysine at the UAG stop codon without the addition of pyrrolysine by researchers[32], indicating a role for PylBCD in the biosynthesis of pyrrolysine from endogenous components. Indeed, these enzymes were found to generate pyrrolysine by linking a methylated pyrroline carboxylate group to the ε-amino group of L-lysine[33–35]. In this reaction, PylB first converts lysine to 3R-methyl-D-ornithine, which is then ligated to the ε-amino group of a second L-lysine molecule by PylC to generate L-lysine-N^{ε}-3R-methyl-D-ornithine. PylD then oxidises this molecule, resulting in the formation of pyrrolysine (Fig 8)[34].

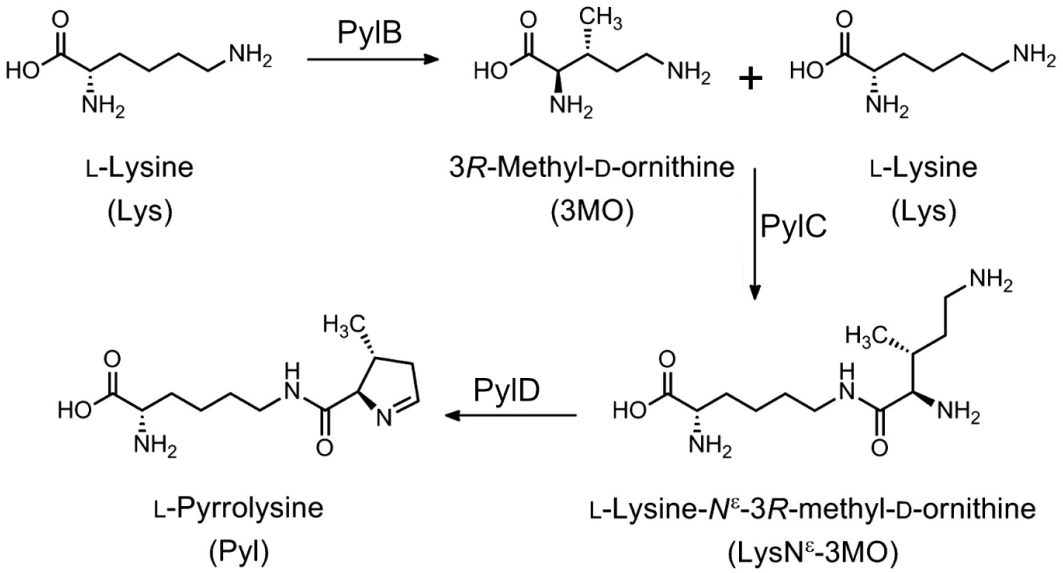

Figure 8: Biosynthesis of Pyrrolysine Pyrrolysine is biosynthesised from two molecules of L-Lysine and proceeds through three catalytic steps involving the enzymes PylB, PylC, and PylD. Adapted from Quitterer *et al*, 2012[34]

Pyrrolysine is encoded by the UAG amber stop codon and, as mentioned, is incorporated using its own PylRS and $tRNA^{Pyl}$ pair. However, it was

unclear how the UAG codon is reassigned from 'stop' to Pyl when required. A sequence similar to the SECIS was found in pyrrolysine-containing genes, named the Pyrrolysine Insertion Sequence (PYLIS), and it was initially assumed that designation of a pyrrolysine-incorporating amber codon followed the same pathway as selenocysteine. However, not all Pyl-incorporating genes contained the PYLIS sequence, and further studies found that PYLIS had no effect on pyrrolysine incorporation[36]. Additionally, incorporation of Pyl in *E. coli* indicated that no specialised elongation factors are required for Pyl-mediated UAG suppression as is the case for Sec incorporation[31,37]. It was further theorised that Pyl-containing species had completely reassigned UAG to Pyl, negating the need for release factors to act on UAG. However, although used much more rarely than the UAA and UGA stop codons, Pyl-containing species such as *Methanosarcina* were found to still utilise UAG as a terminal stop codon in some instances, while the release factors were found to still act on all three stop codons[38]. Instead, incorporation of Pyl at the UAG stop codon is likely a result of direct competition between Pyl-tRNA^{Pyl} and the release factor. In this case, some internal Pyl-incorporating UAG codons would be treated as a stop codon and truncated product would be degraded, while the remainder of the transcripts successfully incorporate Pyl to generate full-length proteins. At terminal stop codons, some readthrough of UAG stop codons by incorporation of Pyl would be expected. However, it was found that UAA or UGA stop codons are often placed immediately downstream of terminal UAG stop codons, providing an extra measure to prevent detrimental read-through of terminal UAG stop codons[39], which may indicate that these Pyl-containing archaea are currently in a state of ongoing, but incomplete, reassignment of the UAG codon.

1.2 Genetic Code Expansion

1.2.1 Early Genetic Code Expansion

Although the 22 canonical amino acids, along with post-translational modifications, can perform an already wide range of functions in proteins, for the past few decades researchers have attempted to synthesise and incorporate novel, non-canonical amino acids (ncAA) to further expand the functions of proteins beyond what can be achieved in nature. For example, ncAAs with

chemically-reactive groups could be used to modify or tag proteins, photo-crosslinking amino acids can be used to covalently link a protein to another biomolecule, photocaged ncAAs can be used to control protein function, or ncAAs that resemble post-translationally modified amino acids can be used to impart permanent modification to a protein[40]. The incorporation of ncAAs is commonly referred to as Genetic Code Expansion (GCE).

The earliest attempts at GCE relied on using a cell's endogenous synthetases to incorporate ncAAs that were structurally very similar to its cognate canonical amino acid. This method was very restrictive, and typically could only be done using close derivatives of the canonical amino acid. For example, TrpRS was used to incorporate 5-hydroxytryptophan in place of tryptophan in *E. coli*, which could be used to aid in the study of the structure and dynamics of the 5-hydroxytryptophan-containing protein[41] due to the enhanced spectral readout of 5-hydroxytryptophan compared to tryptophan. In another case, AlaRS was used to incorporate 3-fluorophenylalanine, which allowed proteins to be studied using ^{19}F nuclear magnetic resonance spectroscopy[42]. In both of these cases, as well as all other cases using this method of GCE, the primary requirement is for the ncAA to be effectively indistinguishable from a canonical amino acid, as perceived by the aaRS. This is severely limiting, as only derivatives of canonical amino acids with relatively minor modifications can be used. Additionally, because an endogenous aaRS is used for incorporation, the ncAA will be inserted at every codon that the aaRS decodes. This further restricts the ncAA to those that are required across an entire protein, rather than at a specific site, as well as restricting it to those that are not detrimental to a protein's function when incorporated in such high amounts.

By mutating the endogenous aaRS proteins, particularly in the proofreading domains, some researchers were able to use the endogenous aaRS to incorporate an even wider range of ncAAs. For example, mutations in the proofreading and editing domain of LeuRS allowed researchers to incorporate variants of valine as well as variants of leucine[43], though it should be noted that valine and leucine structures are already similar enough to require the proofreading function of LeuRS in the first place. While this technique allowed a wider range of ncAAs to be incorporated, these were still fairly limited in structure and the issue of proteome-wide incorporation remained.

1.2.2 Site-Specific Incorporation of Non-Canonical Amino Acids

One of the first methods for site-specific GCE was developed by the Schultz lab, and it allowed for the incorporation of almost any ncAA into any desired protein at the UAG stop codon[44]. Of the 64 codons, the 3 nonsense codons (UAG, UAA, and UAG) are particularly attractive for incorporation as these sites would not break the existing proteome. Of these, the amber codon, UAG, is most suitable for use in GCE as it is the most rarely-used nonsense codon in the majority of organisms, including *E. coli*, yeast, mammals, and other commonly-used model organisms, such as *C. elegans*,[45]. A suppressor tRNA which is directed to incorporate at the UAG stop codon is first chemically aminoacylated *in vitro* with the desired ncAA. The ncAA-$tRNA_{CUA}$ is then used in a cell-free protein synthesis system to generate full length protein with the ncAA incorporated. By replacing any codon with UAG, the ncAA could be site-specifically incorporated at any desired site in the protein. However, this is labour-intensive and difficult to achieve, due to the requirement to chemically aminoacylate the tRNA *in vitro*.

A better method of site-specific GCE would be to use an aaRS to aminoacylate the tRNA *in vivo* (Fig 9). However, this cannot be done using a cell's endogenous aaRSs without incorporating the ncAA at multiple sites and affecting the entire proteome. Therefore, an additional, non-endogenous aaRS/tRNA pair would be required to site-specifically incorporate ncAAs *in vivo*. This aaRS/tRNA pair would have three main requirements: firstly that it can incorporate at a codon that is not already assigned to another canonical amino acid; secondly, that it can be evolved to recognise and charge multiple different ncAAs; and thirdly, that it is entirely orthogonal to a cell's endogenous machinery, so that it cannot react with endogenous aaRS or tRNAs. The TyrRS/$tRNA^{Tyr}$ pair from *Methanocaldococcus jannaschii* was found to be orthogonal in *E. coli*, amenable to mutation of the tRNA's anticodon, and lacks a proofreading domain[46], and so provided a perfect platform for site-specific GCE. The *Mj*$tRNA^{Tyr}$'s anticodon loop was mutated so that it would recognise the UAG stop codon (*Mj*$tRNA^{Tyr}_{CUA}$), and a mutant library of this was generated to identify variants that did not cross-react with endogenous *E. coli* machinery. This library was passed through a negative selection, where a toxic barnase gene was used with an internal

UAG stop codon. If the *Mj*tRNA$^{Tyr}_{CUA}$ was acylated by endogenous *E. coli* machinery, the resultant suppression of the UAG codon would lead to toxic barnase being produced to kill the cells. Next, a positive selection was used by expressing a β-lactamase gene with an internal UAG codon in the presence of *Mj*TyrRS and *Mj*tRNA$^{Tyr}_{CUA}$ to ensure that the machinery could incorporate at the UAG codon[46]. A library of *Mj*TyrRS variants was then generated by mutating the tyrosine-binding pocket to change the specificity of *Mj*TyrRS from Tyrosine to any other desired ncAA. This was used to successfully incorporate O-methyl-L-tyrosine[46], while further studies were able to evolve *Mj*TyrRS for incorporation of L-3-(2-naphthyl)alanine[47], p-azido-L-phenylalanine[48]. and p-benzoyl-L-phenylalanine[49].

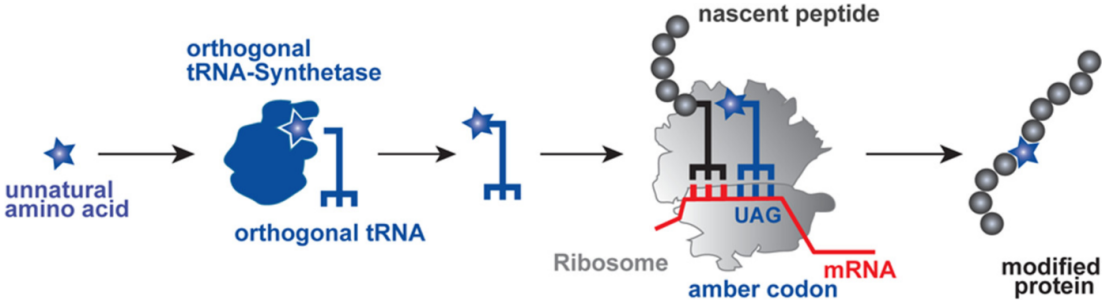

Figure 9: *In Vivo* Site-specific Genetic Code Expansion An orthogonal aminoacyl-tRNA synthetase / tRNA pair are used to incorporate a non-canonical amino acid at a UAG amber stop codon. Adapted from Lang and Chin, 2014[50].

One major limitation of the *Mj*TyrRS/tRNA$^{Tyr}_{CUA}$ system was a lack of orthogonality in mammalian cells, though this was later solved through altering both the *Mj*TyrRS and the *Mj*tRNA$^{Tyr}_{CUA}$ to remove elements which were also found in mammalian endogenous systems[51]. However, before this study, the *M. jannaschii* system could not be used in mammalian cells, and therefore an alternative system was sought. A breakthrough came with the discovery of the PylRS/tRNA^{Pyl} system in *Methanosarcina barkeri*[29]. The natural specificity of this system for the UAG stop codon, as well as the absence of Pyl and its incorporation machinery in *E. coli* as well as yeast and other eukaryotes, made the PylRS/tRNA^{Pyl} an attractive system for GCE. Technically, because *E. coli* does not naturally utilise Pyl, the first use of the PylRS/tRNA^{Pyl} system for genetic code expansion came when it was used to drive incorporation of Pyl in *E. coli* in the early 2000s[31,32].

PylRS/tRNA^{Pyl} was first used for incorporation of ncAAs in 2008, only a few years after it was discovered, by Neumann *et al*[52]. In this study, a library of *Mb*PylRS variants were generated by mutagenesis of six binding-pocket residues. Mutants that could recognise the ncAA N$^{\varepsilon}$-acetyl-lysine were selected for through both positive and negative selection methods, which resulted in a fully orthogonal N$^{\varepsilon}$-acetyl-LysRS/tRNA$_{CUA}$ pair which could site-specifically incorporate N$^{\varepsilon}$-acetyl-lysine into *E. coli* proteins *in vivo*. This N$^{\varepsilon}$-acetylation is a post-translational modification found commonly in histone proteins with roles in regulating transcription, DNA repair and replication, and chromatin condensation[53]. Since this study, the PylRS/tRNA^{Pyl} system has been evolved and used for incorporation of a wide range of post-translationally modified amino acids, including phosphorylation, sulphation, nitration, acylation, and methylation[54], which have greatly aided in the study of post-translational modifications.

Around the same time as the PylRS/tRNA^{Pyl} was developed for ncAA incorporation in *E. coli*, another group applied the system to mammalian GCE. Here, as a proof of principle, the PylRS/tRNA^{Pyl} system from *Methanosarcina mazei* was used unmodified to incorporate N$^{\varepsilon}$-tert-butyloxycarbonyl-L-lysine (Boc-lysine) into proteins in Chinese hamster ovary cells and human HEK293 cells[55]. Next, the group evolved *Mm*PylRS to recognise a larger lysine variant, N$^{\varepsilon}$-benzyloxycarbonyl-L-lysine, which could be done with only five total mutations in PylRS, with only two mutations in the binding pocket. Finally, the group transplanted the mutations in the *Mb*PylRS that were used to generate N$^{\varepsilon}$-acetyl-LysRS[52] into the *Mm*PylRS, and successfully used this to incorporate N$^{\varepsilon}$-acetyl-lysine into mammalian cell lines[55].

Since these studies, *Mm*PylRS/tRNA^{Pyl} and *Mb*PylRS/tRNA^{Pyl} have become the platforms of choice for genetic code expansion, as the orthogonality in both prokaryotes and eukaryotes allows for the evolution of PylRS in bacteria, which can then be transplanted into eukaryotes. More recently, the use of PylRS/tRNA^{Pyl} has been expanded from single cellular bacteria, yeast, and mammalian cell cultures, to multicellular animals. The first animal to have its genetic code expanded in this way was *Caenorhabditis elegans* in 2011[56], followed a year later by *Drosophila melanogaster*[57], *Mus musculus* in 2016[58], and *Danio rerio* in 2017[59].

1.2.3 Improving Genetic Code Expansion

Although GCE has been used successfully in a large number of cases, problems still remain in the efficiency of incorporation of the aaRS/tRNA pairs used. Because of this, a large amount of effort has been devoted to further improving the existing systems to improve the rate of incorporation while preventing cross-reactivity with endogenous aaRS/tRNA pairs. Additionally, GCE has so far been capable of only incorporating a single ncAA at a time, and so work has been undertaken in expanding this to be capable of incorporation of multiple ncAAs.

1.2.3.1 Improving ncAA Incorporation Efficiency

Incorporation of ncAAs using the evolved PylRS/$tRNA^{Pyl}$ remained relatively low, with some evolved PylRS variants catalysing their aminoacylation reactions at less than 0.1% of the rate of wild-type PylRS[60]. The low catalytic activity of PylRS variants, along with competition for UAG recognition between $tRNA^{Pyl}$ and endogenous release factors[61], leads to relatively low levels of incorporation of the ncAA. Although some studies have reassigned UAG completely, replacing all terminal UAG stop codons with UAA or UGA and removing the UAG-specific release factor in *E. coli*[61], this is very impractical in higher eukaryotes. Therefore, effort has been focused towards evolving the PylRS and $tRNA^{Pyl}$ for higher efficiency of aminoacylation and incorporation of the ncAA.

One of the simplest methods of improving incorporation of ncAA is to increase the availability of ncAA, PylRS, or $tRNA^{Pyl}$, to overcome the inefficiency. For example, expression of PylRS or $tRNA^{Pyl}$ from a multi-copy plasmid was found to improve incorporation and yield of ncAA-containing protein by up to 20 times as compared to genomically integrated single-copy expression[62]. Increasing the availability of ncAA also aids in ncAA incorporation rates, either by simply increasing the concentration of ncAA supplied, or improving the uptake of the ncAA by the cells. For example, a study found that p-benzoyl-L-phenylalanine was not incorporated well at low concentrations. However, increasing the concentration of p-benzoyl-L-phenylalanine, or supplying it as a dipeptide to enhance uptake, could significantly improve incorporation efficiency[63]. However, these methods are merely compensatory for the low catalytic rate of evolved PylRS/$tRNA^{Pyl}$ and do not actually solve the problems.

To improve catalytic efficiency, researchers must look at the PylRS variants produced during directed evolution. It should be noted that the low efficiency is not always a direct consequence of a change in binding specificity, and it is often possible to generate PylRS variants specific to certain ncAAs while retaining wild-type levels of incorporation efficiency. However, due to the sheer number of mutants generated during library genesis and the subsequent amount of screening required, these efficient variants may not always be found. However, improved mutagenesis and screening protocols could help to generate PylRS variants that are specific to ncAAs without a loss in activity. For example, one group used the Phage-Assisted Continuous Evolution technique[64] to generate variants of PylRS with a higher efficiency than wild-type PylRS[65], with the mutations then being transplanted into evolved PylRS to improve incorporation of ncAAs.

Other more rational and directed approaches have been taken to improve PylRS efficiency. For example, the N-terminal domain of PylRS is known to be involved in the interaction with its $tRNA^{Pyl}$ partner[66,67], but the highly hydrophobic properties of the N-terminal domain leads to aggregation of PylRS and a resultant low cytosolic concentration[67]. Because of this, overexpression of PylRS may not improve incorporation as much as desired, as a higher concentration would only lead to further aggregation. Therefore, it was desirable to evolve the N-terminal domain to increase the solubility of PylRS. This was performed by Sharma *et al*[68], who performed error-prone PCR on the N-terminal domain and screened subsequent mutants for incorporation efficiency. This allowed for generation of mutant PylRS variants with higher incorporation efficiency than wild-type PylRS, without lowering the orthogonality of the system, though it was unclear if this improvement is due to a higher solubility or improved tRNA-binding of the PylRS N-terminal domain[68]. Another study solved the same solubility problem with an alternative approach by attaching SmbP, a protein tag that increases solubility[69], to the PylRS N-terminal domain, which improved solubility of PylRS and increased incorporation efficiency[70]. Another potentially detrimental feature of PylRS was found during analysis of the *Mm*PylRS sequence, which found a cryptic nuclear localisation sequence (NLS) and a nucleolar localisation sequence near the N-terminus end of the protein[71], with immunofluorescence staining confirming the nuclear localisation of PylRS. Since aminoacylation primarily occurs in the cytosol, this NLS could lower overall incorporation efficiencies. To counteract the NLS, a strong nuclear export signal

(NES) sequence was added to the N-terminal end of PylRS. This was found to increase incorporation efficiency to rates up to 15-fold greater than wild-type PylRS[71].

In addition to altering PylRS, it is also possible to modify $tRNA^{Pyl}$ for improved incorporation of ncAAs. One approach to achieve this in *E. coli* was to modify residues in the acceptor stem and TΨC stem-loop of $tRNA^{Pyl}$, which are both involved in interacting with the elongation factor EF-Tu[72]. This was found to improve incorporation efficiency in *E. coli*[73], though it is unclear if the same mutations, or at least a similar approach, could apply to $tRNA^{Pyl}$'s interaction with EF-1A in eukaryotes. In another approach, rational engineering of $tRNA^{Pyl}$ was performed to improve compatibility of this tRNA with endogenous mammalian host machinery[74]. It was noted that the $tRNA^{Pyl}$ secondary structure is divergent from other canonical tRNAs, typically featuring a longer anticodon stem-loop and missing nucleotides in the D-stem-loop and variable arm[75], which may cause lower compatibility with endogenous machinery. While differing from the vast majority of known tRNA structures, it was noted that $tRNA^{Pyl}$ shows remarkable similarity with the mitochondrial $tRNA^{Ser}$, which showed better incorporation efficiencies[76]. Therefore, chimera tRNA variants were generated by introducing PylRS identity elements from $tRNA^{Pyl}$ into the mitochondrial $tRNA^{Ser}$, with the best of these variants, $tRNA^{C15}$, increasing ncAA incorporation more than 2-fold compared to $tRNA^{Pyl}$ whilst retaining orthogonality[74]. In the same study, a second set of $tRNA^{Pyl}$ variants were generated as chimeras of $tRNA^{Pyl}$ and human tRNAs. The best performing tRNA from this set, $tRNA^{M15}$, was found to improve incorporation at similar rates to the $tRNA^{C15}$ variant[74]. The secondary structures of $tRNA^{C15}$ and $tRNA^{M15}$, as well as the structures of mitochondrial $tRNA^{Ser}$ and a canonical human tRNA used to generate them, are shown in Figure 10.

In our lab, a combination of PylRS modifications and $tRNA^{Pyl}$ modifications are used in *C. elegans* to greatly increase incorporation efficiency over the wild-type PylRS/$tRNA^{Pyl}$ pair. Both the $tRNA^{C15}$ and $tRNA^{M15}$ variants showed much higher levels of incorporation in our hands[77,78]. Additionally, we used the NES method developed by Nikić *et al*[71] to further improve incorporation. For this, a number of NES sequences were tested, including p120cts, Smad4, and PKI-α, which were all derived from human proteins[79] and S-NES, which is the same NES used in the original Nikić *et al* study in mammalian cell culture[71].

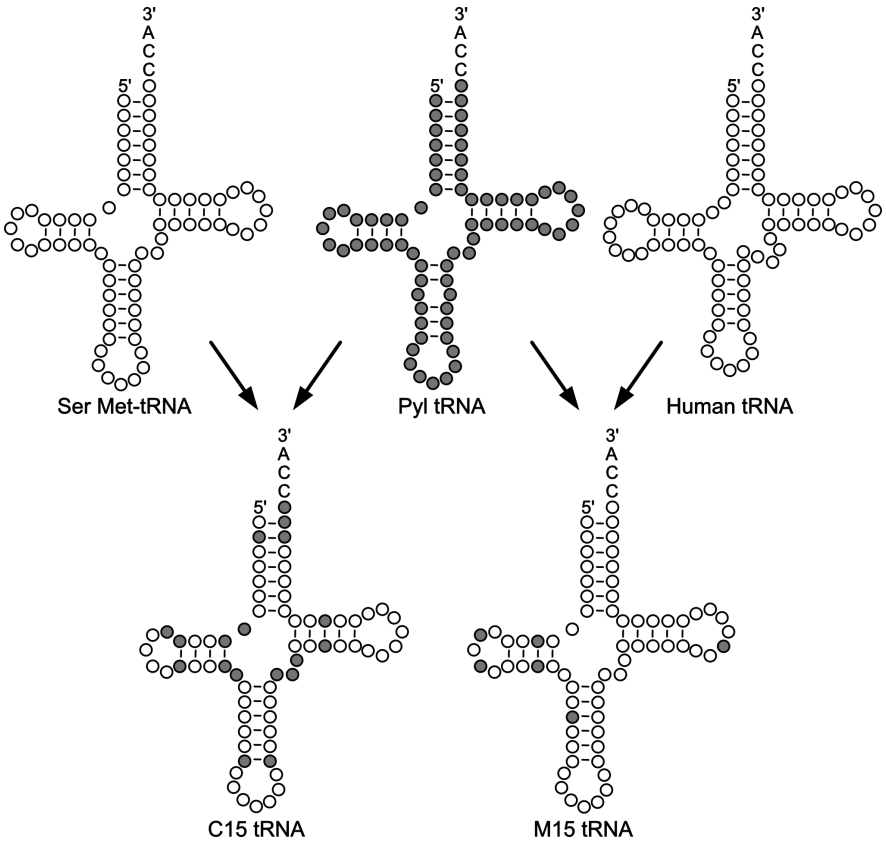

Figure 10: Generation of tRNA^{C15} and tRNA^{M15} Variants. tRNA structures for a mitochondrial tRNA^{Ser}, tRNA^{Pyl}, a canonical human tRNA, tRNA^{C15}, and tRNA^{M15}. tRNA^{C15} was generated by inserting specific tRNA^{Pyl} modifications (Grey circles) onto a mitochondrial tRNA^{Ser} backbone (White circles). tRNA^{M15} was generated by inserting specific tRNA^{Pyl} modifications (Grey circles) onto a canonical human tRNA backbone (White circles)[74].

Although S-NES and p120cts were found to have very little effect on incorporation in *C. elegans*, both PKI-α and Smad4 NES showed a greatly increased level of incorporation. Together, the NES-PylRS fusion and the tRNA^{C15} or tRNA^{M15} increased efficiency of incorporation by more than 50-fold compared to unmodified PylRS/tRNA^{Pyl}.

1.2.3.2 Incorporation of Multiple ncAAs

The techniques for GCE described so far typically utilise the UAG stop codon and an orthogonal aaRS/tRNA pair. Because of this, it is typically only possible to incorporate one ncAA in any one strain, as a second ncAA would require the use

of another stop codon as well as another mutually orthogonal aaRS/tRNA pair. However, the incorporation of more than one unique ncAA would be useful as it would open up further possibilities for studying protein functions and creating proteins with more advanced and novel properties. Work to incorporate multiple unique ncAAs has therefore been ongoing, and primarily focuses on generating new codons or reassigning existing codons, as well as identifying or evolving new orthogonal aaRS/tRNA pairs.

One possibility to increase the number of codons available for GCE is to remove codons from the endogenous code. Because 61 sense codons encode 20 canonical amino acids, most amino acids are encoded by more than one codon, with most being encoded by 2 or 4 codons. Although codon choice can influence gene expression and cell fitness[80], many groups decided to try cutting down this redundancy by rewriting entire genomes to replace certain codons with a synonymous codon, thereby freeing up that codon[81–85]. In one study carried out by the Chin lab[81], an *E. coli* genome was completely redesigned and synthesised to replace serine TCG and TCA codons, as well as the TAG stop codon, with synonymous AGC, AGT, and TAA codons, respectively. This synthetic *E. coli*, named Syn61, grew at a much reduced rate compared to wild-type *E. coli*, but only used 61 codons instead of 64. Syn61 was capable of incorporating non-canonical amino acids at the TCG codon without negative effects[81], and in future studies could be used to incorporate a unique ncAA at each of the three free codons.

In another approach, the Chin lab has developed orthogonal ribosomes (O-ribosome) by altering the ribosomal RNAs to direct the O-ribosome to an orthogonal mRNA[86,87], which contained different recognition sequences from endogenous mRNAs. The most recent orthogonal ribosome, Ribo-Q, is optimised for quadruplet decoding, rather than triplet coding, and was capable of incorporating up to four unique ncAAs in a single protein[87].

Although the previously discussed study used orthogonal ribosomes optimised to decode quadruplet codons, other studies have shown it to be possible to use quadruplet codons with endogenous translation machinery[88–90]. As quadruplet codons consist of four bases instead of three, there are a total of 256 possible quadruplet codons, vastly increasing the coding space for including ncAAs. Our lab has recently developed a system for quadruplet decoding in *C. elegans* based off the PylRS/tRNA^{Pyl} pair[78]. This required optimisation of the tRNA

to contain quadruplet-decoding anticodon loops, and resulted in incorporation rates approximately 40% that of triplet-decoding systems. Although this lower rate could limit the number of ncAAs incorporated, further improvements and optimisation of the quadruplet PylRS/tRNA^{Pyl} pair could lead to quadruplet codons being a viable choice for incorporation of multiple ncAAs. In addition to this, quadruplet codons could reduce incorporation of ncAAs at terminal UAG stop codons, which may reduce the overall impact on the endogenous proteome caused by triplet decoding[78].

While all the described methods free up or otherwise expand the coding space available for incorporation of ncAAs, there is still the requirement of additional orthogonal aaRS/tRNA pairs to facilitate the incorporation. Although the two most common PylRS/tRNA^{Pyl} pairs, *Mm*PylRS/tRNA^{Pyl} and *Mb*PylRS/tRNA^{Pyl}, are not mutually orthogonal, they have been found to be mutually orthogonal with the *Mj*PylRS/tRNA^{Pyl} and so can be used in the same strain to incorporate two unique ncAAs[91]. The PylRS/tRNA^{Pyl} from another methanogenic archaeon species, designated as *ISO4-G1* has also been found to be mutually orthogonal with *Mm*PylRS/tRNA^{Pyl} and can therefore be used together[91]. Another methanogenic archaeon, *Methanomethylophilus alvus* was found to have a PylRS/tRNA^{Pyl} pair that is not mutually orthogonal to the *Mm*PylRS/tRNA^{Pyl} pair. However, directed evolution of *Ma*PylRS/tRNA^{Pyl} generated a variant that was found to be mutually orthogonal with *Mm*PylRS/tRNA^{Pyl}, allowing them to then be used together[91]. Work is currently ongoing in our lab to apply *Ma*PylRS/tRNA^{Pyl} and *Mm*PylRS/tRNA^{Pyl} pairs to decoding quadruplet codons and incorporating two ncAAs in a single animal. Further exploration of PylRS/tRNA^{Pyl} pairs from other species, or evolution of these pairs as with the *Ma*PylRS/tRNA^{Pyl} pair, may yet yield additional mutually orthogonal pairs.

1.2.4 Photocaged Amino Acids

One particularly useful group of ncAAs are the photocaged amino acids (pcAA), which contain photolabile protective groups, referred to as a caging group, which block the activity of the underlying amino acid[92]. Upon illumination with a specific wavelength of light, the caging group is removed and a canonical amino acid structure is revealed, which is fully functional. To date, photocaged

variants of aspartic acid[93], serine[94], glycine[95], tyrosine[96], cysteine[97], and lysine[98] have been generated, as well as photocaged variants of phosphoserine and phosphotyrosine[99].

These pcAAs can be used to control the activity of proteins by replacing catalytic residues with their photocaged counterparts to inactive the protein, then activate them by uncaging the pcAA[100]. Using GCE, this has been applied to a wide range of proteins. For example, a photocaged lysine (PCK, Fig 11) has been used to control the activation of the bacteriophage T7 RNA polymerase, which allowed for spatiotemporal control of gene expression from the T7 promoter[101]. In another example, PCK was placed into the nuclear localisation sequence (NLS) of a transcription factor, which could then be used to block the nuclear import of the transcription factor until illumination[102]. Photocaged tyrosines[103,104] have also been applied to controlling protein activity, including the activity of a DNA recombinase[105], a STAT1 transcription factor[106], and a zinc-finger nuclease[107], while photocaged cysteine has been used to photocage the activity of the TEV protease[108] and inteins[109].

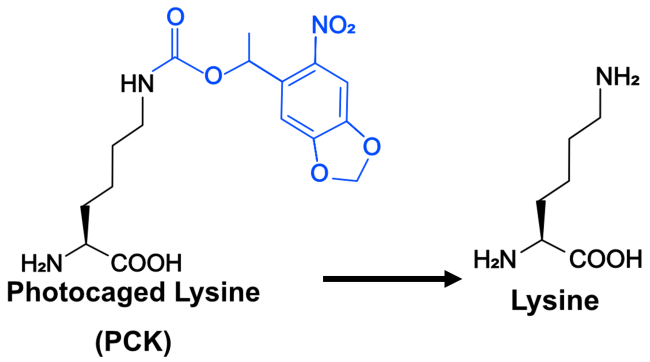

Figure 11: Uncaging of Photocaged Lysine. Methyl-*o*-nitropiperonyl Lysine (Photocaged lysine, PCK) contains a 'caging' group (blue) which blocks activity of the underlying lysine. Upon stimulation with 365 nm light, the cage group is removed to restore a canonical lysine.

Although multiple different photolabile groups exist, one of the most common groups used for photocaged amino acids is the ortho-nitrobenzyl group, or its derivatives[100], which can undergo Norrish reactions, a carbon-carbon cleavage which occurs during photo-irradiation[110]. In this thesis, I use several different photolabile groups, including orthonitrobenzyl (Fig 12A), nitropiperonyl (Fig

12B), and methylnitropiperonyl (Fig. 12C), which each have differing uncaging kinetics[104,111]. These groups all uncage with illumination by 365 nm wavelength ultraviolet (UV) light.

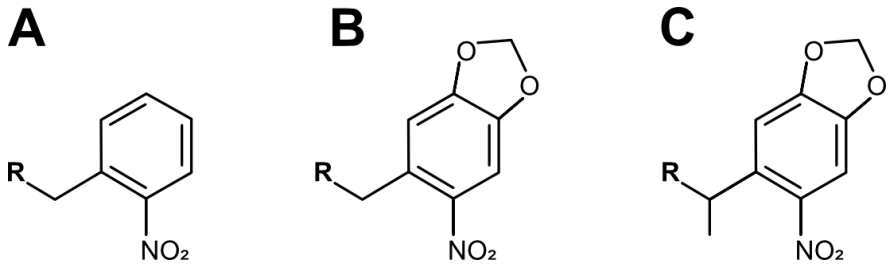

Figure 12: Various Photocaging Moieties **(A)** orthonitrolbenzyl, **(B)** nitropiperonyl, **(C)** methylnitropiperonyl.

1.3 *Caenorhabditis elegans*

1.3.1 *Caenorhabditis elegans* as a Model Organism

Caenorhabditis elegans is a free-living nematode worm discovered in 1900 by Émile Maupus, who first isolated and characterised it from soil samples in Algeria[112], though it has since been found to be widespread and living in almost all parts of the world[113]. After its discovery in 1900, *C. elegans* remained relatively obscure until the late 1940s, when it was proposed as a model for genetic study by Ellsworth Dougherty and Hermione Calhoun[114]. However, by far the most significant player in establishing *C. elegans* as a model organism was Sydney Brenner who, after working to elucidate the nature of the genetic code with Crick[115], decided to turn his attention towards applying molecular biology to the fields of developmental biology and neuroscience. In searching for a model animal which was relatively simple, genetically tractable, and could be handled in large numbers, Brenner settled on *Caenorhabditis*, obtaining an isolate of *C. elegans* from Dougherty in 1963 and beginning his work in characterising the nervous system of the animal[116]. The strain Brenner was given is referred to as the Bristol strain, which was isolated by Nicholas *et al* in 1959 from mushroom compost in Bristol, England[117], though at the time it was erroneously identified as an isolate of *Caenorhabditis briggsae*. From this Bristol strain, Brenner isolated a single hermaphrodite and allowed it to propagate by self-fertilisation to give rise

to a new strain, named 'N2', which remains the most commonly used strain in *C. elegans* research to this day[118].

C. elegans are only about 1 mm in length as adults and exist as either self-fertilising hermaphrodites or males, with the vast majority of animals (>99%[119]) being hermaphrodite. This makes them ideal for genetics studies, as hermaphrodites and males can be used in genetic crosses, with the hermaphrodites then isolated and self-fertilising to become homozygous for a genotype. In addition, a single hermaphrodite can produce upwards of 200 progeny[120] during its life, allowing for strains to be established relatively quickly and large-scale experiments performed. In the lab, *C. elegans* can be kept on petri dishes filled with nematode growth medium (NGM) agar and fed on simple monocultures of *E. coli*, which makes them very simple and easy to maintain.

The development and life cycle of *C. elegans* is very quick, taking only around three days from fertilisation to adulthood. After egg laying and hatching, a young *C. elegans* animal goes through four larval stages, named L1 to L4, during which its cells continue to divide and differentiate to give rise to an adult with approximately 1000 somatic cells[121]. The lineages of each of these cells has been fully mapped throughout the entire life cycle of the worm[122], and is found to be invariant. If conditions are unfavourable for worm development, the L1 larva will instead enter a developmentally arrested stage, named 'dauer', which is unable to feed and is more resistant to external stresses, allowing the worm to survive for several months[120]. When encountering favourable conditions, the dauer stage develops into an L4 larva and continues development to adulthood. The life cycle is shown in Figure 13.

Many other features contribute to *C. elegans*'s utility as a model organism. The animal is completely transparent, allowing for visualisation of internal cells and tissues. The adult has an invariant number of neurons, 302, with all connections between these neurons having been mapped[123], and a large number of human genes, including 40% of disease-causing genes[124], have orthologs in the *C. elegans* genome, making studies of *C. elegans* relevant to the study of human function and disease. Indeed, a number of *C. elegans* models of human diseases have been generated, including those of Alzheimer's disease[125], Parkinson's disease[126], muscular dystrophy diseases[127–129], and cancer[130].

Throughout its many years as a model organism, *C. elegans* has contributed to the discovery of a number of other vital discoveries. Programmed cell

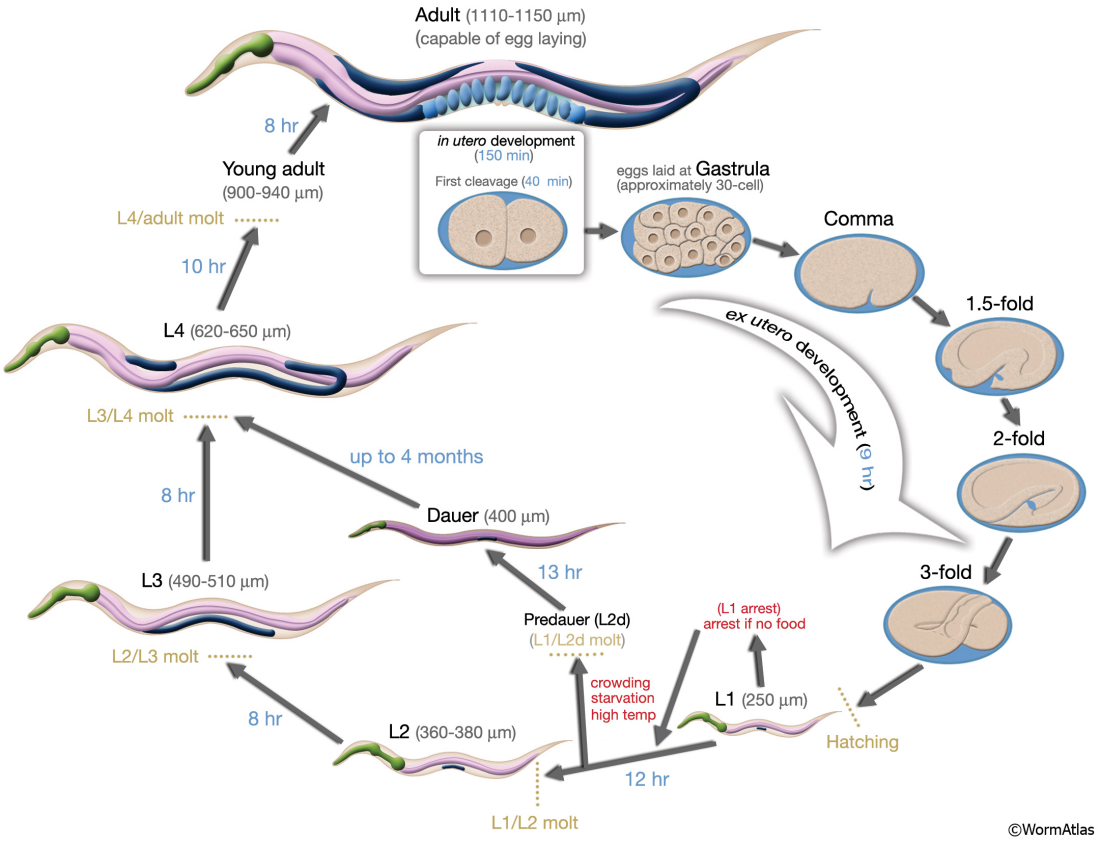

Figure 13: The Life Cycle of *Caenorhabditis elegans* The *C. elegans* egg is partially developed *in utero*, with development finishing *ex utero*. After hatching, *C. elegans* development proceeds through four larval stages before becoming adult. Under unfavourable conditions, the larval stage L1 can develop into the dauer stage. On return of favourable conditions, dauers re-enter the life cycle at the L4 stage. Times of development at each stage are approximate for *C. elegans* hermaphrodites grown at 22℃. Taken from Corsi, 2006[137].

death, an important part of normal animal development, was first studied in *C. elegans*[131,132] and earned Nobel prizes for Sydney Brenner, John Sulston, and Robert Horvitz, in 2002. Further Nobel prizes came in 2006 for Andrew Fire and Craig Mello for their discovery of RNA interference (RNAi)[133], a host immune response which causes downregulation of gene expression by complementary binding of mRNA. RNAi is now a widely used tool for gene downregulation in *C. elegans*[134], as well as in other model organisms and is now being explored as a therapeutic option against several diseases[135]. *C. elegans* also played a role in the development of green fluorescent protein (GFP) as a reporter for gene

expression and protein localisation. After testing GFP in *E. coli*, Martin Chalfie then applied GFP to *C. elegans* and was able to use it to visualise individual neurons within the worm[136]. This earned Chalfie a Nobel prize in 2008, and GFP, along with its differently-coloured derivatives, have since become some of the most commonly used biological tools across multiple model organisms.

With the overwhelming amount of data available on *C. elegans*, a number of tools and databases have been set up to aid in research. One of the most extensive databases is the WormBase[138], which contains fully annotated genomes, descriptions of all known coding and non-coding genes, protein interactome data, cell lineage data, and more. WormAtlas[139] provides detailed anatomies and functions of every cell in the *C. elegans* body, including every neuron. The Caenorhabditis Genetic Centre (CGC) maintains and distributes thousands of different *C. elegans* strains with well characterised mutations for use in research.

1.3.2 Optogenetics in *Caenorhabditis elegans*

The transparency of the *C. elegans* body makes it particularly amenable to light-based experiments, broadly referred to as 'optogenetics'. The dictionary definition of optogenetics is a combination of optical methods and genetic engineering, and in its broadest sense can include any method of controlling or monitoring biological function through the use of light[140], including methods as simple as visualisation of fluorescently-labeled cells, though the term is most commonly applied to the activation and inhibition of neurons through the use of opsins[141].

Opsins are transmembrane photoreceptors found in animals, bacteria, archaea, and algae, and include a number of uses in optogenetics. Channelrhodopsins (ChR) are used by green algae to sense light and drive plasma membrane depolarisation, increasing the cytosolic concentration of positive ions and triggering signaling cascades which mediate the algae's flagellar movement to drive the algae towards the light[142]. ChRs have since become a common tool in optogenetics, as the cation currents generated during light stimulus can activate action potentials within neurons, allowing for the spatiotemporal control of neuron activation. Other opsins, such as the bacteriorhodopsins and halorhodopsins from archaea, pump positive ions out of the cell or pump negative ions into the cell upon light stimulus which can be used in optogenetics

to hyperpolarise and inhibit neuron function[143,144]. *C. elegans* was the first organism to be behaviourally-manipulated using Channelrhodopsin 2 (ChR2) from *Chlamydomonas reinhardtii* to excite muscle cells and mechanosensory neurons in 2005[145], and since then a large number of studies have been performed to study neural function in *C. elegans* using these channels[146].

Another important area of optogenetics is the control of protein-protein interactions and protein localisation. This can been done using AsLOV2, a LOV (Light-Oxygen-Voltage sensing) domain from *Avena sativa*'s phototropin 1, which can interact with an engineered PDZ domain (ePDZ, named after the three proteins in which the domain was first discovered: PSD-95, Dlg1, and ZO-1) after exposure to blue light[147]. By fusing proteins to both the AsLov2 and ePDZ, protein-protein interactions between them can be induced by stimulation with blue light. In *C. elegans* this system has also been used to drive organelle transport by using an organelle membrane-localised AsLOV2 to recruit an ePDZ-fused motor protein, which led to trafficking of the organelle[148,149].

A major limitation of optogenetic techniques is the level of spatiotemporal control involved. Although many are completely inactive without stimulus, some, such as the AsLOV2/ePDZ system[150], show significant background activity even in the dark. It can also be difficult to achieve spatial precision in certain experiments. For example, few cell-specific promoters exist in *C. elegans* so the optogenetic tools often must be expressed in a number of cells at one time. However, it can often be difficult to activate the tool in a single cell in a free-moving animal, making the study of individual cells difficult. One emerging area of optogenetics in *C. elegans* which could solve these problems is the use of photocaged amino acids to control protein function.

1.3.3 Genetic Code Expansion in *Caenorhabditis elegans*

In 2011, *C. elegans* became the first multicellular organism to have its genetic code expanded[56]. In the initial study, Greiss and Chin[56] showed that *Mm*PylRS/tRNA^{Pyl} could be used to incorporate Boc-lysine into an amber codon placed in a linker between GFP and mCherry fluorescent proteins, which would result in mCherry production only after incorporation of Boc-lysine. However, two main problems were encountered during the study. Firstly, very low transmission of the transgenes was seen, resulting in populations where

very few animals expressed the transgenes. This was solved by injecting a Hygromycin B phosphotransferase gene into the worms and growing them in the presence of Hygromycin B, which resulted in transmission rates of 100%[56], and was later developed to include a selectable Hygromycin B resistance cassette on the transgene expression constructs themselves[151], which additionally allowed for selection of transgenic animals after transgenesis. The second issue encountered was that worms that did express the transgenes often showed very low expression of GFP due to degradation of the mRNA by nonsense mediated decay (NMD). However, crossing strains expressing the transgenes into the *smg-2(e2008)* strain, which is deficient in NMD, results in a significant increase in GFP signal. Although incorporation rates were low, with only 5% of worms showing incorporation and readthrough of mCherry, this study marked the first use of GCE in a multicellular system, and also showed that a PylRS/$tRNA^{Pyl}$ pair evolved in *E. coli* could be used for ncAA incorporation in multicellular eukaryotes[56]. Additionally, there were no observable effects on the health of the animals. This is likely because endogenous 3' UTR sequences encode protein sequences that tag a read-through protein for degradation[152], which would prevent accumulation of proteins that have incorporated the ncAA at the endogenous amber stop codons.

A later study demonstrated the use of the tyrosyl-tRNA synthetase/$tRNA^{Tyr}$ and leucyl-tRNA synthetase/$tRNA^{Leu}$ pairs from *E. coli* for incorporation of ncAAs in *C. elegans*[153], which was used to incorporate the ncAAs *o*-methyl-L-tyrosine and 2-amino-3-(5-(dimethylamino) naphthalene-1-sulfonamido) propanoic acid (DanAla). Noteably, although *o*-methyl-L-tyrosine is structurally similar to tyrosine, DanAla is structurally distinct from all canonical amino acids. This study also further demonstrated the use of dipeptides to increase incorporation efficiency, with the DanAla monopeptide being sequestered into intestinal cells, while an Ala-DanAla dipeptide was successfully transported into other tissues[153].

Although these studies showed that GCE was possible in *C. elegans* and could be used to study a range of biological problems, few studies since have utilised the method, with one of the only uses being in the process of bioorthogonal noncanonical amino acid tagging (BONCAT). A typical BONCAT experiment involves the use of azidohomoalanine, an ortholog of methionine, which is incorporated into proteins in place of methionine and can be used to selectively

tag and identify newly synthesised proteins in cells, allowing for temporal studies of the proteome[154]. A study by Yuet *et al*[155] developed BONCAT further in *C. elegans* by using genetic code expansion to selectively incorporate the non-canonical amino acid tags in specific cells. In this study, a modified phenylalanyl-tRNA synthetase was used to incorporate *p*-azido-L-phenylalanine, which could be used to tag and identify proteins similarly to azidohomoalanine. By expressing the modified PheRS using cell-specific promoters, the method could be used for the temporal study of the proteome of a specific subset of cells[155].

Following these studies, our lab has continued to develop, optimise, and apply GCE to *C. elegans*, primarily with the use of the PylRS/tRNA^{Pyl} pair to incorporate pcAAs. Due to its transparent body, *C. elegans* is a particularly useful platform for light-based ncAAs, such as photocaged amino acids, photo-crosslinking amino acids, and fluorescent amino acids. By inserting a photocaged lysine into the active site of the Cre recombinase, Davis *et al*[77] were able to generate photo-activatable Cre which could be used to drive gene expression. By laser-targeting, the photo-activatable Cre was used to drive expression of channelrhodopsins in individual neurons, which allowed them to be illuminated in free-moving animals to activate the neuron. In this study, the modifications and optimisations to PylRS/tRNA^{Pyl} described above allowed us to reach much higher levels of incorporation, with over 95% of animals incorporating pcAA into Cre. In a following study, Xi *et al*[78] used photocaged cysteine (PCC) to control the activity of Caspase-3, which could be used to optically ablate cells. This study also applied quadruplet codons to drive ncAA incorporation for the first time in an animal, which was optimised to incorporate at close to the levels reached by triplet decoding. Finally, a study by O'Shea *et al*[156] photocaged protein-protein interactions by using photocaged tyrosine to control the binding of a nanobody to its antigen.

1.3.4 Thesis Objectives

In this thesis, I aim to expand the toolkit available to *C. elegans* researchers by developing novel optogenetics tools. To this end, I have developed a new optogenetic tool for targeted gene expression and attempted to develop a set of optogenetic tools for gene-targeted random mutagenesis. These tools utilise photocaged amino acids incorporated by genetic code expansion to confer a high

level of spatiotemporal control not found in other tools.

In Chapter 2, I detail the development of a photocaged FLP recombinase. This tool can be used to drive gene expression, and I show that a laser can be used to target individual cells to drive gene expression with extremely high spatiotemporal precision. This chapter focuses primarily on the development and optimisation of photocaged FLP recombinase, and the characterisation of its activity before and after activation, with some brief work at the end of the chapter where we use photocaged FLP to study neuronal functions.

In Chapter 3, I attempt to develop multiple tools for gene-targeted random mutagenesis in *C. elegans*, which could potentially be used for reverse genetics studies, as well as for directed evolution in *C. elegans*, which could be useful in further improving the efficiency of incorporation of ncAAs. Although the development of these tools was unsuccessful, I propose future work to solve the problems encountered during the course of this thesis and to successfully develop these optogenetic mutagenesis tools.

In Chapter 4, I will give an overview of the materials and methods used to achieve the objectives of this thesis. In Chapter 5, I will give an overview of the tools developed and introduce potential future directions in the continued development and applications of these optogenetic tools.

Chapter 2

Photocaged FLP Recombinase for Spatiotemporal Gene Expression

2.1 Introduction

2.1.1 Site-Specific Recombinases

The site-specific recombinases (SSR) are a class of enzymes which recognise short DNA sequences and catalyse the recombination between two of these sequences, which allows them to perform four basic functionalities depending on the orientations of their recognition sites: excision, insertion, inversion, and translocation of DNA (Fig. 14). Thousands of SSRs have been identified in a wide range of organisms, including viruses, bacteria, archaea, and single-cell eukaryotes[157], carrying out roles such as the integration of viral DNA in host genomes, the resolution of plasmid and chromosome multimerisation during replication, the modification of gene expression, and the increasing of genetic diversity by mediating gene transfer[158,159].

SSRs generally contain at least two domains, the catalytic domain and the DNA-binding domain, though some SSRs have lost the DNA-binding domain and others have other further domains which act in regulating the SSR activity[158]. Although thousands of SSRs have been identified, they can be categorised into two families: the Serine-SSRs and the Tyrosine-SSRs, depending on which amino acid within the catalytic domain is used as a nucleophile to attack the DNA backbone during recombination. Although members of both families function in all of the roles described above, the two families are unrelated and their mechanisms differ[159].

Serine-SSRs are best characterised by γδ resolvase, the prototypical and most well-studied Serine-SSR, which is naturally found as part of a transposable element in the *E. coli* genome[160]. γδ resolvase, as well as other Serine-SSRs,

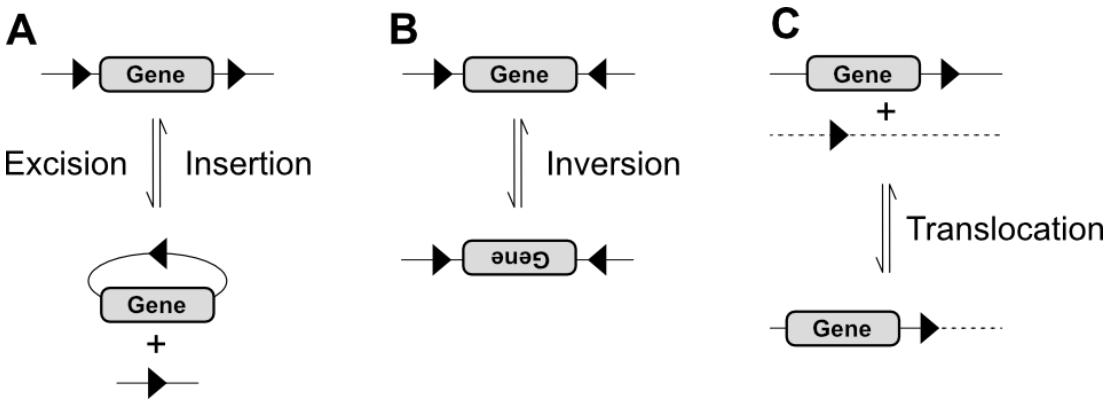

Figure 14: Possible Outcomes of Site-Specific Recombination. The products of the recombination reaction depend on the orientation of the recombinase recognition sites (black triangles), and can give rise to **(A)** excision or insertion, **(B)** inversion, or **(C)** translocation of the genetic elements.

naturally exist as dimers, with each dimer binding one recognition site and then organising with a second dimer to form a synaptic complex, which consists of a core of the Serine-SSR tetramer with two duplex DNA molecules on the outside. All four DNA strands are cleaved, which is caused by nucleophilic attack of the phosphodiester bond by the catalytic serine residues, linking the Serine-SSRs covalently to the 5' end of the DNA and leaving a free 3' hydroxyl end. One half of the synaptic complex then rotates around 180° to bring the cleaved DNA strands into a recombined configuration, where the free 3' hydroxyl ends then attack the 5' phosphoserine linkages to reform the DNA duplexes (Fig. 15A)[159].

Tyrosine-SSRs share a similar domain structure, though with the key difference of using a tyrosine in the catalytic domain. Interestingly, the catalytic domain of Tyrosine-SSRs share high sequence, structural, and mechanistic similarity with type IB topoisomerases, indicating a shared ancestor[159], whereas the Serine-SSRs do not seem to share ancestry with either the Tyrosine-SSRs or the type IB topoisomerases. Tyrosine-SSRs also differ when it comes to the formation of the synaptic complex, where the two DNA molecules are contained within the tetramer, allowing them much closer access to each other than in the Serine-SSR conformation. The cleavage reaction also differs, with Tyrosine-SSRs performing the reaction in two steps instead of one. Firstly, only two of the four Tyrosine-SSR proteins are active, and the nucleophilic tyrosine of these active subunits cleave one strand of each of the DNA duplexes, linking the protein

covalently to the 3' end of the DNA and leaving the 5' hydroxyl free. The DNA strands exchange, with the free 5' hydroxyls attacking the 3' phosphotyrosine linkages of the opposing strands. The two previously-inactive protomers then become active, repeating the exact same process to recombine the remaining two DNA strands to fully recombine the DNA[159] (Fig. 15B).

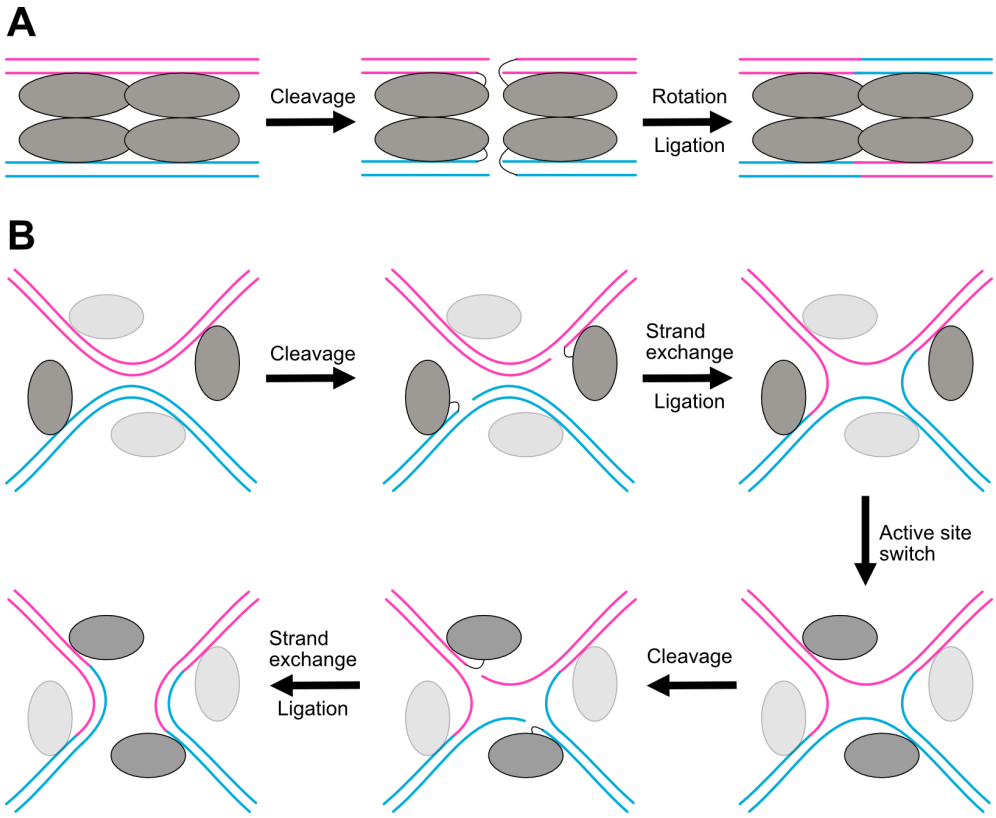

Figure 15: Site-Specific Recombination (SSR) Mechanisms. (A) Serine-SSR. Four protomers (grey ovals) of the Serine-SSR cleave a DNA strand each (red and blue lines), forming phosphoserine bonds (black lines). Two protomers within the complex rotate and DNA is re-ligated in a recombined configuration. **(B)** Tyrosine-SSR. Two active protomers (dark grey oval) cleave their DNA strands and form phosphotyrosine bonds (black lines). Strands are exchanged and ligated, forming a Holliday junction. Two previously inactive protomers (light grey) become active, repeating the reaction to resolve the Holliday junction.

2.1.2 Applications of SSRs

Of the thousands of SSRs that exist in nature, several have been repurposed as molecular biology tools. These include the Tyrosine-SSRs Cre (Cyclisation recombination) from the P1 bacteriophage, FLP (Flippase) from the 2 μm plasmid from *Saccharomyces cerevisiae* and λ integrase from the λ phage, and the Serine-SSR ΦC31 from the ΦC31 phage[157]. These have been applied to a number of roles, including plasmid construction, transgene integration and control of gene expression.

One common cloning technique, the Gateway cloning system developed by Invitrogen, uses the λ integrase to transfer genetic elements between plasmids[161]. λ integrases normally facilitate the integration of bacteriophage λ's circular genome into the genome of its host, *E. coli*, by recombining between the host *att*B (Bacterial attachment) site and its own *att*P (Phage attachment) site. This gives rise to a prophage flanked by two newly generated sites, *att*L (Left attachment) and *att*R (Right attachment), which can then recombine to excise the prophage and restore the original *att*B and *att*P sites[162]. In the development of Gateway cloning, mutations of these four sites gave rise to a number of variations, which could recombine with specific sites but not others. For example, *att*B1 and *att*P1 can recombine, but *att*B1 and *att*P2 cannot. Using these sites, two cloning techniques were developed - the BP reaction and the LR reaction (Fig. 16). In the BP reaction, a PCR product flanked by two different *att*B sites is integrated into a 'donor' vector with two equivalent *att*P sites, which generates a single 'entry' vector with a transgene flanked by two *att*L or *att*R sites, depending on the orientation of the initial *att*B and *att*P sites. In the LR reaction, one or more 'entry' vectors are recombined with a 'destination' vector, which contains equivalent *att*R sites, producing a single expression plasmid from multiple genetic elements.

Another common use of SSRs is to integrate transgenes into the genome, with the first studies utilising the λ integrase to integrate transgenes into the *E. coli* genome[163]. Since then, multiple other SSRs have been used and in multiple model organisms, including the use of FLP and Cre recombinases to integrate transgenes into mammalian genomes[164,165]. The use of SSRs for this purpose in *C. elegans* has been fairly limited, with the majority of genomic integrations being done by UV irradiation, CRISPR/Cas9 editing, or transposon-based integration[166].

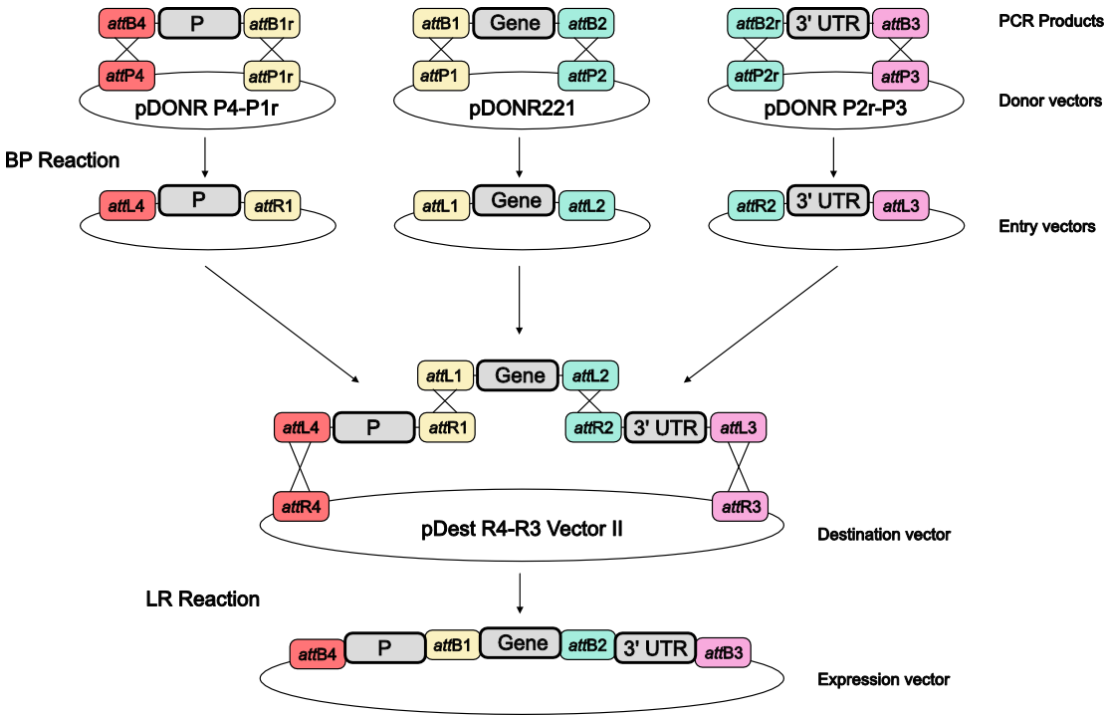

Figure 16: Gateway Cloning Schematic. λ integrases are utilised in the Gateway cloning method (Invitrogen). *att*B, *att*P, *att*L, and *att*R sites are used in series of recombination events to transfer genetic elements from a PCR product to a donor vector (BP reaction), and then from the donor vector into a destination vector (LR reaction). Multiple entry and destination vectors exist, shown here is an example using the Gateway vectors used in this thesis. In this example, a promoter (P) can be placed in position 1, a gene in position 2, and a 3' untranslated region (UTR) in position 3, allowing for easy switching of components to generate different expression constructs.

Recently, a system was developed in *C. elegans* where FLP is used to integrate a transgene along with a detectable marker, and Cre was then used to excise the marker after confirmation of integration, leaving behind only the transgene in the genome[167].

In *C. elegans*, one of the most common uses for SSRs, particularly FLP, is the control of gene expression. This can be done in many different ways to achieve both gene knockout and gene activation. In one method, which achieves activation of expression, a transcriptional terminator is placed between the promoter and the coding sequence (CDS) of the transgene, with the terminator flanked by FLP recognition sites. Upon addition of FLP to the system, the

transcriptional terminator is excised from the cassette, allowing expression of the transgene to proceed[168] (Fig. 17A). Conversely, using FLP to excise the gene from an expression construct[169] can be used to inhibit transgene expression (Fig. 17B). Another method for control of gene expression is to place the transgene in an inverted orientation between the recombination recognition sites, which prevents functional expression until the SSR is introduced to recombine and invert the sequence[170]. Although inversion between identical recombination sites is reversible and unbiased, with inversion occurring for as long as the SSR is present, asymmetric recombination sites exist which give rise to two different asymmetric sites when inverted. If the second pair of recombinase recognition sites is much less efficient than the first pair, then this leads to inversions that are very biased in one direction[171] (Fig. 17C). A more advanced method, the Flip-Excision (FlEx) switch, utilises two pairs of recombinase recognition sites, with one pair being a mutant pair that does not recombine with the other wild-type pair. By combining inverted and non-inverted transgenes with these pairs of recombination sites, both inversion and excision can be used to switch the expression between the two transgenes[172,173] (Fig. 17D).

2.1.3 Limitations of SSR-mediated Gene Expression

Although powerful in its approach, the use of SSR to activate or repress gene expression is not without limitations. One of the main issues is the lack of spatial control of the gene expression. In multicellular organisms, including *C. elegans*, promoters generally drive expression of their genes in groups of cells, often across entire tissues or subsections of tissues[174]. However, researchers often need to express transgenes in specific cells, especially in neuroscience where each individual neuron may have a distinct function. For the vast majority of cases, a cell-specific promoter does not exist, which hinders the study of these cells. Additionally, the SSRs lack temporal control, which may limit the expression of toxic transgenes or transgenes which may otherwise interfere with the normal development of the animal. For example, the tetanus neurotoxin is often used in the study of neuron functions, but expression of tetanus in developing worms can arrest development or lead to other neurons compensating for the loss of the neuron[168,175], hindering the study of the neuron of interest.

Several methods have been developed to introduce spatiotemporal control

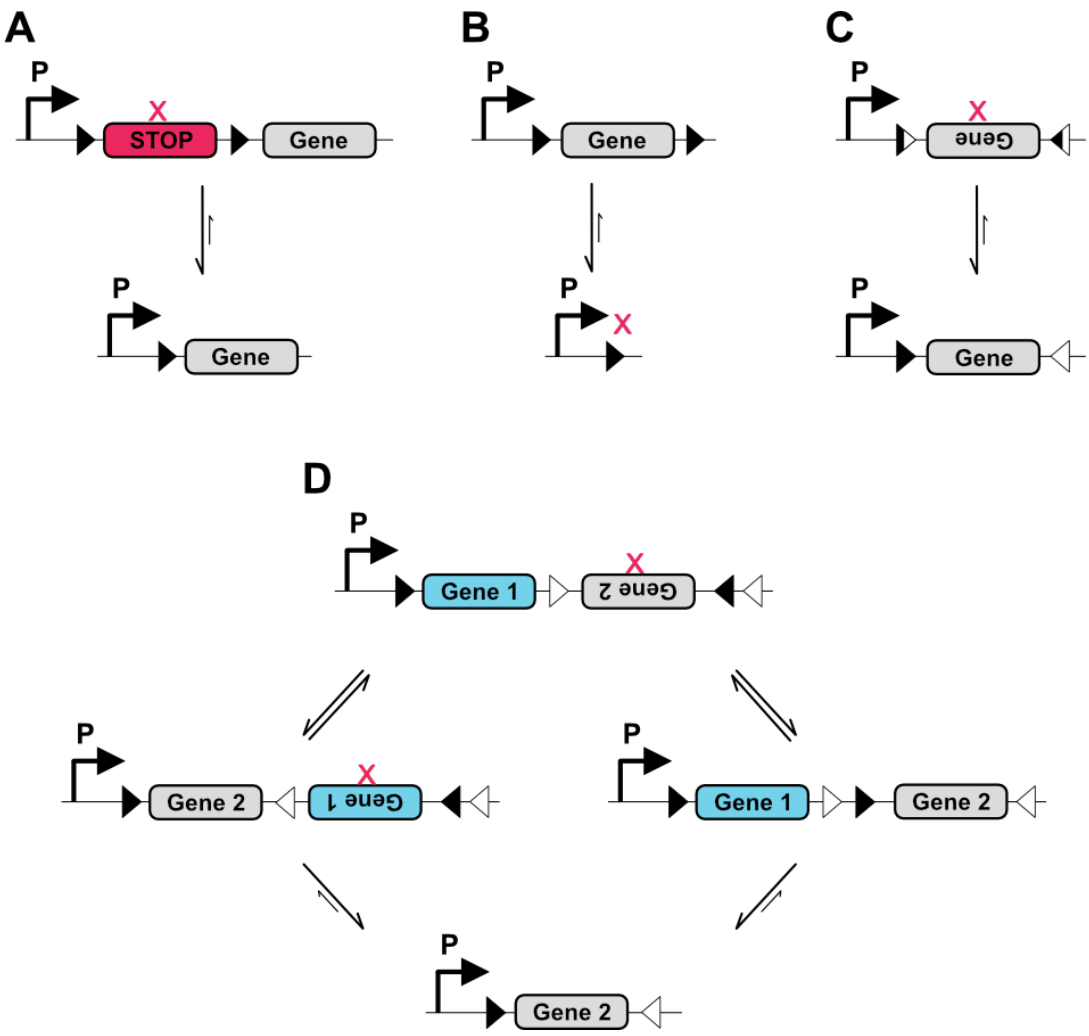

Figure 17: Gene Expression Using Site-Specific Recombination. Red 'X' represents lack of expression of gene product. **(A)** Activation of gene expression. STOP indicates a transcriptional terminator, which is excised by the site-specific recombinase (SSR) to allow expression of Gene. **(B)** Gene knockout using SSR. **(C)** Gene activation by SSR inversion using mutant recombination sites. White regions of recombination sites indicate mutated versions, black regions indicate wild-type sequences. When recombined, fully wild-type or fully mutant sites are generated, which recombine at much lower rates. **(D)** FlEx switch. Using two pairs of recombinase sites (white or black), the activation and repression of gene expression in a two-gene system can be switched. Two pathways are shown, depending on which sites recombine first, but always result in the same end state.

into SSR-mediated gene activation systems. A common method is to use two overlapping promoters to express the transgene and the SSR[168], which can introduce either higher spatial control or temporal control into the system. For higher spatial control, this method uses two overlapping spatial promoters, with one expressing the transgene behind a transcriptional terminator (Fig. 17A), and the other expressing FLP. In this setup, only the cells where both promoters are active will express the transgene. In many cases, it is possible to find overlapping promoters which allow for the expression of the transgene in single cells. However, this is not always possible, and the technique is still limited to using very specific promoters, which may further limit the research if these promoters are only active at certain developmental stages, or have a very low level of expression. Furthermore, this technique still does not allow for expression of toxic transgenes, as expression will occur whenever both spatial promoters are active. For temporal control, the transgene can be expressed using the spatial promoter as before, but the FLP can be expressed using a temporal promoter. In *C. elegans*, the only temporal promoters are the ubiquitous heat shock promoters, including *hsp16.41p* which was used in the Davis *et al.* study[168], though some other heat shock promoters exist with differing levels of activity in different tissues[176]. Using this promoter, FLP will not be expressed under standard conditions and the transcriptional terminator will not be excised. Upon heat shock, the heat shock promoter activates, driving expression of FLP and activating expression of the transgene. While this method allows the expression of toxic genes at specific developmental stages, it is very spatially limited and is unlikely to be fit for use in single cells, except in those rare cases where a single-cell promoter exists. Additionally, the activation of the heat shock response has fundamental changes on the animal, including effects on transcription and translation of other proteins and even altering the lifespan of the animal[177], which may cause unintended behavioural effects which could interfere with the study of the cells of interest.

A further development of this system includes both FLP and Cre recombinases, which allows for a greater level of control[178]. In this system, two terminator sequences are placed before the transgene, with one flanked by FLP recognition sites, and one flanked by Cre recognition sites, thereby requiring the presence of both FLP and Cre to activate expression of the transgene. With one of the recombinases expressed behind an overlapping spatial promoter and one expressed behind the temporal heat shock promoter, this system combines

the separate spatial and temporal controls possible from the system developed by Davis *et al*[168]. With this, transgenes can be expressed when desired in a very small subset of cells, or even in single cells, facilitating the study of previously inaccessible cells. However, some limitations as discussed earlier still apply, as not all individual cells, especially in the nervous system, may have promoter pairs that would express solely in that neuron. This is especially true of neuron pairs, where the neurons exist as a bilateral symmetric pair[179], which may be genetically indistinguishable but functionally different. For example, the neurons AVJ, CEPD, URYV, RIM, and PVC, all exist as neuron pairs, but only one of each pair is presynaptic to RIS, an interneuron involved in sleep[180], indicating asymmetrical roles within each of these neuron pairs despite them being genetically identical. As two-thirds of *C. elegans* neurons exist as neuron pairs[179], tools capable of dissecting the pairs would be very beneficial in increasing the understanding of the neural circuits and functions.

A method was developed using the heat shock promoter to drive FLP expression in single cells by Churgin *et al*[181,182]. This method uses a pulsed laser to heat the cell of interest up to 50 ℃ to induce heat shock in only that cell, activating expression of FLP and subsequently the transgene. While achieving much greater spatiotemporal control than the other methods described and without the need to identify suitable promoter pairs, this still requires inducing heat shock in the cells being studied, which could have potentially detrimental effects. In addition, it was found that heating of over 30 ℃ was experienced as far as 6 μm away from the focal point. With the average neuron size being between 5-10 μm[183], with some neurons being as small as 3 μm[181], this has a high potential for off-target activation. Indeed, in attempts to activate expression in one of the smallest neurons in the worm, ADL, the authors found that only 30% of worms targeted expressed the transgene only in the target cell, with a further 10% expressing in off-target cells. Testing the system with a second neuron, AWB, gave similar results, with only 26% of targeted worms expressing the transgene in the target cell only, and a further 11% expressing in off-target cells as well[181]. Although better results were seen in larger cells, such as intestinal cells, this relatively low activation rate and high off-target activation rate is potentially inhibitory for study of cells that are small and tightly packed, including neurons.

Though as yet unused in *C. elegans*, light-dependent domains can also be used to reconstitute split versions of FLP[184] or Cre[185]. In these systems, the SSR is

split into two inactive halves, with each fused to one photoswitch domain, such as those found in the Magnet system[186]. Upon exposure to light, the photoswitch domains assemble together, reconstituting and activating the full SSR protein. Although the use of low-level light is much better for cells than high temperature, this system is found to have a relatively high level of background, with one experiment finding approximately 10% of cells expressing the transgene under dark conditions[184]. The use of wavelengths of light in the visible range also necessitates keeping the cells expressing the system completely in the dark, which makes working with them difficult in transparent animals, such as *C. elegans* or zebrafish embryos. These systems are also reconstituted by blue light, which can interfere with optogenetic and behavioural studies, which often use blue light-activated channelrhodopsins, such as ChR2.

It would therefore be useful for further development of SSRs with high spatiotemporal control. An ideal system would allow for the activation of transgene expression down to single-cell resolution, with no off-target activation in neighbouring cells and without relying on very specific overlapping promoters. Instead, the ideal system would allow researchers to express their transgene from any desired promoter which best suits their needs. The system would be inducible, with no activity before induction and high activity after. Ideally, the method of induction would allow a researcher to manually select individual cells for expression, allowing them to select single cells or any desired subset of cells. The temporal control would best be provided by stimuli that are non-damaging and non-invasive, such as low-level light, but that are also not susceptible to activation by light produced in a typical lab setting, either by the overhead lighting or by light- and fluorescent-stereomicroscopes.

2.1.4 Photocaged SSRs

One method which fits all criteria of an ideal spatiotemporal control system is the use of photocaged amino acids to confer photocontrol onto the SSR, by using the caging group to block activity of a catalytic residue. This was first performed for the Cre recombinase in mammalian cells by Edwards *et al*[187], where the catalytic tyrosine residue was replaced with a photocaged tyrosine, *o*-nitrobenzyl tyrosine (ONBY), which loses its cage group and reverts to a canonical tyrosine upon exposure to 365 nm light. The Cre-ONBY was found to be completely

inactive, and uncaging restored activity to approximately the level of wild-type Cre, allowing Cre-ONBY to be used for spatiotemporal gene control. This system was expanded upon by Luo *et al*[188], who found that a conserved lysine could also be photocaged to install photocontrol into the Cre. The photocaged lysine (PCK), methyl-*o*-nitropiperonyl Lysine, shows higher uncaging kinetics, higher solubility, and higher biocompatibility of the caging group compared to ONBY, and so is better suited to use in biological systems[188].

Photocaged Cre was developed further by Brown *et al*[189], with the photocaged Cre being used in zebrafish embryos. While still caging the critical lysine residue, this study used hydroxycoumarin lysine (HCK) instead of PCK, which had a decaging wavelength of 405 nm. Although single-cell activation was not performed, this study successfully used the photocaged Cre to perform lineage tracking of the heart, telencephalon, and tail, by illuminating cell populations in an early embryonic stage. A similar variant of photocaged lysine, bromo-hydroxycoumarin lysine, can additionally be uncaged by two-photon excitation[190], which increases the depth within tissues that the amino acid can be uncaged.

We have since applied photocaged Cre to *C. elegans*[77], which we used to drive expression of an optogenetic channel in single neurons of a neuron pair. This photocaged Cre had its critical lysine, K201, replaced with PCK, which showed no expression of the optogenetic channel before uncaging. Global uncaging of the photocaged Cre expressed in glutamatergic neurons showed expression of the optogenetic channel and subsequent behavioural responses in over 95% of animals. This system was then used to express the optogenetic channel in individual neurons by targeting with a 365 nm laser, which allowed us to study the individual contributions of neurons within a neuron pair. Although this was used to control transgene expression in *C. elegans*, it could also apply to controlling the expression of endogenous genes by inserting recombination target sites into the *C. elegans* genome, as well as controlling expression of transgenes and endogenous genes in other model organisms.

To date, all photocaged SSRs developed have been Cre recombinases[77,187–189]. It would be useful to expand this toolkit to include other photocaged SSRs, such as photocaged FLP, which may have advantages over Cre in certain cases. One consideration is the optimum temperatures for recombinase activity. As Cre is derived from bacteriophages targeting *E. coli*, its optimum temperature is 37

℃, while the optimum temperature for the yeast-derived FLP recombinase is lower, at 30 ℃[191]. Since *C. elegans* are grown between 15-25 ℃, it is likely that FLP would be much more active in *C. elegans* than Cre, though both have been used successfully in a large number of studies in *C. elegans*[192]. Another factor is the observations of DNA-damaging effects of Cre recombinase in mammalian cells[193], which can lead to general toxicity, but has not been observed for FLP recombinase[194].

Aside from specific advantages of FLP over Cre, the development of a photocaged FLP could complement the use of photocaged Cre. One study found that FLP and Cre seemed to have different efficiencies when expressed under different promoters, with some promoters leading to FLP being more active, while some promoters led to Cre being more active[195]. Because of this, having both photocaged FLP and photocaged Cre systems available to a researcher would expand the potential applications of the systems. Additionally, once a photocaged FLP has been developed, it could be used in conjunction with the photocaged Cre to drive expression of two transgenes in two cells or two distinct groups of cells in a single animal. Therefore, I set out to develop a photocaged FLP recombinase for use as a tool for transgene expression in *C. elegans*.

2.1.5 FLP Recombinase

FLP is a member of the tyrosine-SSR family, along with Cre and λ integrase, and is found as part of the 2 μm plasmid, a selfish genetic element which resides as multiple copies in the nucleus of *Saccharomyces cerevisiae*[196]. The biological function of FLP is to prevent decreases in the numbers of 2 μm plasmids in the yeast cell, by inverting regions of the plasmid to change from a bi-directional replication system to a uni-directional rolling circle system[197]. While bi-directional replication results in only a single replicated plasmid being generated, as the meeting of the two replication forks prevents any further replication from occurring, uni-directional replication will continue unhindered until FLP activity re-inverts the sequence to restore the bi-directional replication system.

A FLP dimer recognises and binds the FLP Recognition Target (FRT) sequence, a site consisting of three 13 bp repeats flanking an asymmetrical 8 bp spacer[198], though recombination can still occur between sites with a 7 or

9 bp spacer[199], or with only two repeat sequences rather than three[200]. Once dimers have bound two FRT sites, they tetramerise to form the synaptic complex consisting of four FLP protomers and two DNA duplexes. At this point, two of the protomers become active, while two remain active, a state referred to as half-of-the-sites activity[201]. Unusually for tyrosine-SSRs, which normally use active sites consisting of residues all from a single protomer, FLP assembles its catalytic sites *in trans*, with residues from two protomers forming the active site. While most active site residues are provided by a single protomer, the catalytic tyrosine, Y343, is provided by an adjacent protomer. The remainder of the active site is made up of highly conserved R191, K223, H305, R308, and W330 residues (Fig. 18A,B). The two arginines interact with and position the scissile phosphate group of the substrate DNA[201]. The tyrosine acts as the nucleophile which attacks the substrate, with the histidine likely serving as the general base which activates the tyrosine. Although unconfirmed in FLP, the catalytic lysine in Cre has been shown to act as the general acid, and so is also likely to act as the general acid in FLP. The tryptophan seems to help position Y343 in its active orientation and may also help in stabilising the leaving 5' hydroxyl group during cleavage[202,203].

Recombination by FLP occurs in a two-step mechanism, with 2 protomers active at one time (Fig. 15B). This activity can be seen by the positioning of Y343, which is well ordered and in close proximity to the phosphate when contributing to an active catalytic site (Fig. 18A), and disordered and more distant from its target when contributing to the inactive sites[201] (Fig. 18B). The active Y343 is deprotonated by H305 and is then able to attack the phosphodiester bond, generating a covalent linkage to the 3' end of the DNA, while K223 donates a proton to the free 5' hydroxyl end, breaking the DNA strand. Strand exchange then occurs by the 5' hydroxyl group attacking the phosphotyrosine linkage of the opposing DNA strand, which displaces the tyrosine and ligates the strands in a recombined position (Fig. 18C), which forms a Holliday junction. The half-of-the-sites state then switches, with the two previously inactive protomers becoming active. These sites repeat identical cleavage-exchange steps, but on the two non-recombined strands, completing the recombination reaction[159,202].

With photocaged variants of both tyrosine[104] and lysine[204] available, FLP can be photocaged by caging either the K223 or the Y343 residues (Fig. 19A), with the equivalent residues in Cre, K201[77,188,189] and Y324[187] having previously been photocaged successfully. The critical roles of K223 and Y343 in the recombination

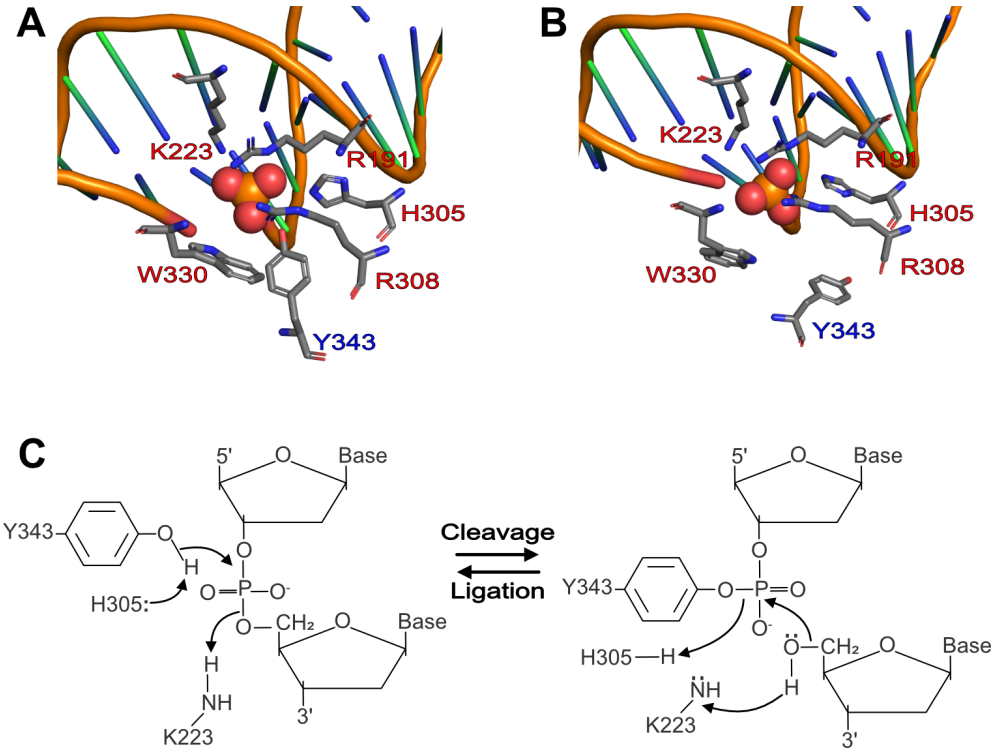

Figure 18: FLP Active Site and Mechanism of Action. **(A)** Active catalytic site of FLP with catalytic residues highlighted. R191, K223, H305, R308, and W330 residues are provided by a single protomer. Y343 is donated by an adjacent protomer. **(B)** Inactive catalytic site of FLP. Y343 is moved out of position and is unable to perform a nucleophilic attack on phosphate. **(A,B)** PDB: 1FLO[201], showing the Holliday junction state of FLP. It is noted that the Holliday junction was captured with the phosphotyrosine bonds, which are subject to slow hydrolysis, which explains why inactive site **(B)** shows cleaved DNA. **(C)** Mechanism of cleavage and ligation reactions performed by FLP. After cleavage, strands are exchanged and ligated to the opposing strand.

reaction, along with the evidence from photocaged Cre, would indicate that caging either of the residues should prevent any activity of FLP before uncaging. By uncaging and activating FLP, gene expression can be controlled with high spatiotemporal precision (Fig. 19B). A photocaged FLP would expand the toolkit available to researchers, both within the field of *C. elegans* research and in those of other model organisms, and both photocaged FLP and photocaged Cre could eventually be combined into a single system, further increasing the control of gene expression.

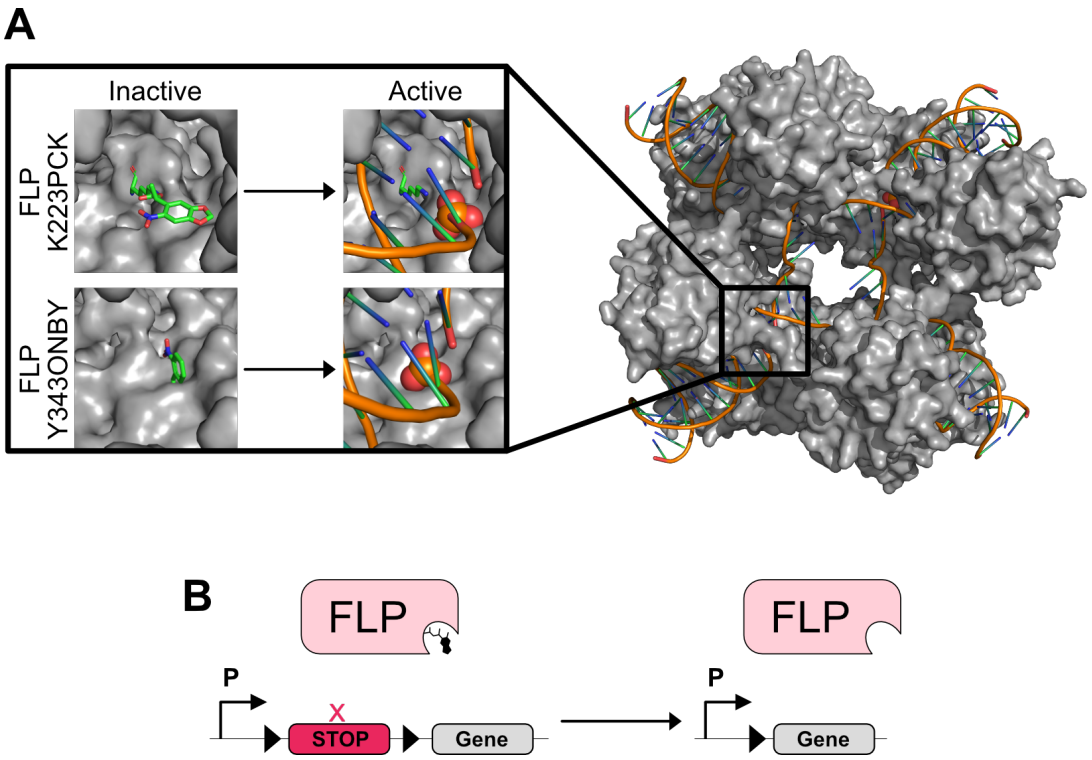

Figure 19: Photocaging of FLP Recombinase. **(A)** Structure of FLP recombinase with either K223 or Y343 replaced with equivalent photocaging groups (PCK and ONBY). pdb = 1FLO[201]. PCK and ONBY structures are custom drawn in Pymol. **(B)** Schematic for gene expression using photocaged FLP recombinase.

2.2 Results

2.2.1 FLP-K223PCK Can Drive Gene Expression With Light-Mediated Control

Due to the recent success in photocaging the Cre recombinase in *C. elegans* by replacing the catalytic lysine, K201, with PCK, we first decided to test photocaged FLP with PCK at position K223 (FLP-K223PCK). Four genetic constructs were designed to facilitate testing of the FLP-K223PCK, with the first construct containing FLP with an amber codon (TAG) at the codon position of K223 to direct incorporation of PCK at this site. The G5D variant of FLP was used, which has the original glycine at position 5 replaced with an aspartic acid residue and was previously shown to increase the recombination efficiency of FLP[169]. For initial testing of whether FLP could be photocaged, constructs were expressed under the control of the *myo-3p* promoter, which drives expression in the body wall muscle, a group of 95 muscle cells extending across the entire length of the worm body. The FLP construct itself also contained a Haemagglutinin (HA) tag to facilitate potential future characterisation of the expression by Western Blot, as well as an *egl-13* nuclear localisation sequence (NLS) to bring FLP into the nucleus. As the NLS is attached to the C-terminal end, truncated FLP (terminating at K223TAG) in the absence of PCK would not translate the NLS, and therefore would not be imported into the nucleus where it may cause damage[77]. The construct containing FLP was expressed in an artificial operon with GFP for visualisation by placing an SL2-accepting splice site (SL2) between FLP and GFP. The second construct contained an mKate2 reporter preceded by an FRT-flanked *let-858* 3' UTR terminator sequence, which acts as a 'STOP' cassette and has previously been used in FLP systems in *C. elegans*[168] to control gene expression. This reporter is driven by the ubiquitous *eft-3p* promoter, which would allow us to test FLP-K223PCK in multiple tissues without changing this construct. On the third construct, an aminoacyl-tRNA synthetase (aaRS) evolved to load PCK (PCKRS), derived from the directed evolution of the pyrrolysyl-tRNA synthetase (PylRS) from *Methanosarcina mazei*[204], is placed behind the *myo-3p* promoter, while the fourth construct uses *prp-1p* to drive expression of the tRNA which is charged with PCK by PCKRS and attaches PCK to the growing polypeptide chain at the UAG codon during translation.

The PCKRS contains a human Smad4 nuclear export sequence fused to the N-terminus which was found to improve incorporation by directing PCKRS to the cytoplasm[71,77], while the tRNA used, $tRNA^{C15}_{CUA}$ is derived from the bovine mitochondrial $tRNA^{Ser}_{CUA}$ modified to contain elements from $tRNA^{Pyl}$, which improved incorporation compared to unmodified $tRNA^{Pyl74}$. All constructs are shown in Fig. 20A.

C. elegans strains expressing these four constructs can therefore be detected by presence of GFP, but if maintained off PCK would not be expected to express mKate2, as only truncated FLP would be produced and the STOP cassette would not be excised. Animals can then be fed on PCK, which will allow for the translation of full-length FLP-K223PCK to proceed. This would be expected to be inactive due to the caging of the catalytic lysine, and no mKate2 will be expressed. Upon illumination with 365 nm light, the caging group of PCK should be removed to give rise to a fully active and effectively wild-type FLP recombinase, which would excise the STOP cassette and drive expression of mKate2 in cells which have been exposed.

The four constructs were introduced into the wild-type N2 strain by biolistic bombardment and screened for presence of the GFP reporter. Strains were generated which expressed FLP consistently across the entire body wall muscle. Although clearly visible under a fluorescent stereomicroscope and under higher magnification (Fig. 20B, bottom row), this signal was too dim to see clearly under 10X magnification (Fig. 20B, top and middle rows). These worms did not show any visible mKate2 signal in the absence of PCK, nor in the presence of PCK but without uncaging (+PCK, -UV), as expected. Upon uncaging by exposure to 365 nm light, 5 mW/cm^2, for 5 minutes (+PCK, +UV), bright mKate2 signal could be seen along the entire body wall muscle (Fig. 20B, middle row). Single-cell uncaging could also be performed by using a 365 nm laser to target only a single cell, which resulted in GFP and mKate2 signals being seen only in the targeted cell, while other cells expressed only GFP (Fig. 20B, bottom row). This demonstrates that caging the K223 residue of FLP is enough to break recombination activity, while uncaging can restore the activity and can be done down to single-cell resolution.

We next attempted to apply the FLP-K223PCK to drive expression of a channelrhodopsin in neurons, which we could use to study the function of the neuron by illuminating the channel to optogenetically activate the neuron. For

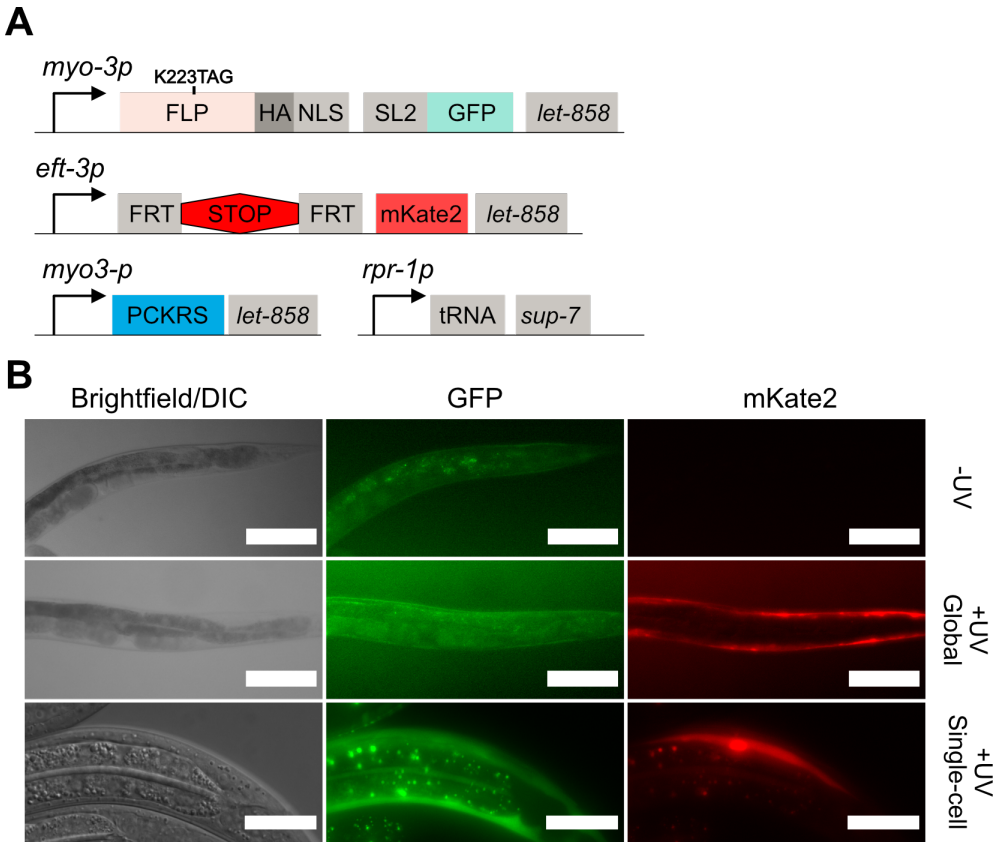

Figure 20: FLP Can Be Made photo-activatable by Photocaging K223. **(A)** Genetic constructs used for testing FLP-K223PCK in body wall muscle cells. **(B)** Expression of the constructs in worms, in the presence of PCK and presence or absence of UV light. (Top row) Before uncaging, worms show GFP signal but no mKate2. (Middle row) After global illumination with 365 nm light, strong mKate2 signal is seen throughout the body wall muscles. (Bottom row) Single-cell uncaging can be performed using a 365 nm laser. Scale bars for top two rows are 100 μm. Scale bars for bottom row are 25 μm. All fluorescent images are max intensity projections, brightfield/DIC images are single slices.

this, similar constructs (Fig. 21A) were used as before, but after the STOP cassette we used the Chrimson channel protein fused to the mKate2. Chrimson is a channelrhodopsin from *Chlamydomonas noctigama* which is red-shifted relative to other common channelrhodopsins[205,206] and so is particularly useful for *C. elegans* optogenetics since blue light is potentially damaging to the worm and they exhibit avoidance behaviour responses upon blue-light stimulation[207]. The STOP cassette was also changed, with the *let-858* 3' UTR being replaced by

the 3' UTR from *lacZ*[208]. Although commonly used without issue, one study with a particularly sensitive assay found low-level expression of the transgene downstream of the *let-858* 3' UTR[209], indicating some level of leakiness or readthrough of the STOP cassette. This may be due to the relatively short size of the *let-858* 3' UTR, at less than 500 bp, while the *lacZ* 3' UTR is just over 3.5 kb and so is less likely to have this issue. The final change to the constructs was to the promoters. For testing the expression of Chrimson, we expressed FLP behind the *mec-7p* promoter, which drives expression in the mechanosensory neurons, a set of 6 neurons (AVM, ALML, ALMR, PVM, PLML, PLMR) which sense touch and transmit signals to drive locomotion in an avoidance behaviour[210]. The Chrimson::mKate2 fusion expression was driven by *maco-1p*, a pan-neuronal promoter with a high level of expression in young adult worms. As expected, strains expressing the FLP-K223PCK showed no visual expression of Chrimson::mKate2 before exposure to 365 nm light. After uncaging with global illumination with 365 nm light, Chrimson::mKate2 signal could be seen localised to the edge of the neurons, as expected for a reporter fused to a channel protein (Fig. 21B).

To test the activity of Chrimson, the transgenic strains were grown on PCK for two days, globally uncaged and transferred to plates supplemented with all-trans retinal (ATR), the cofactor required for channelrhodopsin activity. Animals were then left for 24 hours to express the Chrimson::mKate2 and uptake ATR, then transferred to behaviour plates and stimulated with red light for 5 second pulses to activate Chrimson. As an output of worms responding to stimulus, we chose to measure the speed of the worms during stimulus rather than the velocity. Although measuring speed only gave us the absolute change in movement, rather than showing whether the worms accelerated forwards or reversed as velocity would, we chose not to measure velocity because of a high level of mosaicism in the expression of the transgenes, with the majority of worms only showing expression of the transgenes in one or two neurons. Since some of the mechanosensory neurons mediate forward movement and some mediate backwards movement, this mosaicism resulted in stimulated worms of the same strain carrying out a mix of forward and backward accelerations, depending on which neurons were expressing the transgenes in that particular worm, which would average out to a low velocity across multiple worms. Worms expressing FLP-K223PCK did not show any change in speed between stimulus and non-stimulus time points without

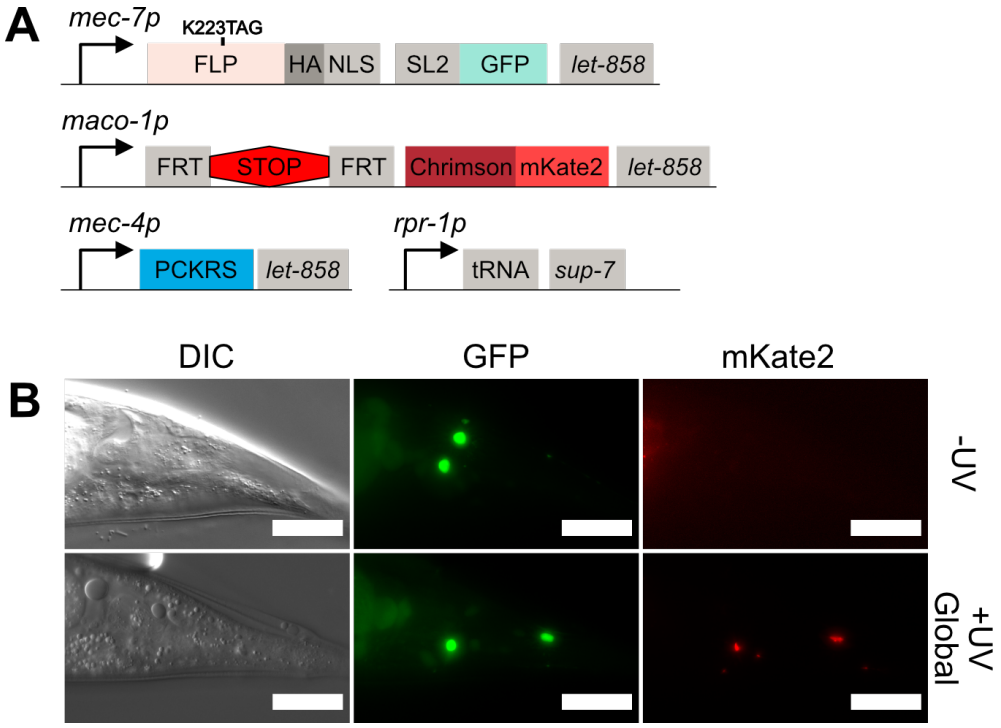

Figure 21: Driving Expression of Chrimson::mKate2 Using FLP-K223PCK (A) Genetic constructs used for testing FLP-K223PCK in mechanosensory neurons. **(B)** Expression of the constructs in worm tail (PLM neurons), in the presence of PCK and presence or absence of UV light. (Top row) Before uncaging, worms show GFP signal but no mKate2. (Bottom row) After global illumination with 365 nm light, strong mKate2 signal is seen in the mechanosensory neurons. Scale bars are 25 μm. All fluorescent images are max intensity projections, DIC images are single slices.

activation of FLP. Once uncaged to activate FLP, worms showed a sharp increase in the speed during red-light stimulus compared to unstimulated conditions (Fig. 22), showing that Chrimson had been expressed at a functional level and that its expression was entirely dependent upon uncaging and activation of FLP-K223PCK. However, it should be noted that the data shown in Fig. 22 represents a small sample size and this experiment was not replicated due to difficulties in visualisation of the Chrimson::mKate2 construct.

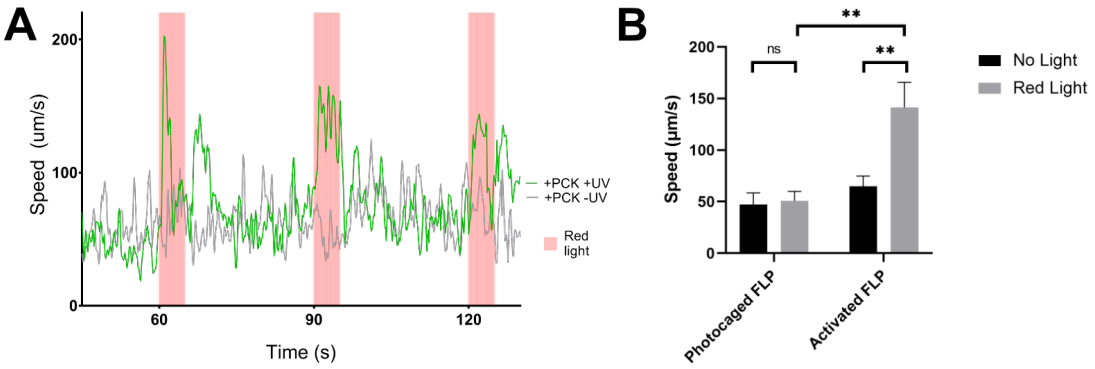

Figure 22: FLP-K223PCK Can Drive Chrimson Expression for Behavioural Analysis. Behavioural analysis is performed with worms using FLP-K223PCK to express Chrimson in mechanosensory neurons. **(A)** Speed of worms during stimulus with red light (red bar) stimuli. **(B)** Average speed of worms in the 5 s before and 5 s during red-light stimulus. Mean±SD. **=p≤0.01, ns=p>0.05; Mixed-effects analysis with post-hoc Sidak's multiple comparisons. The data represents the average of 8 animals in a single non-replicated experiment.

2.2.2 Improving Visibility of Chrimson-Fluorescent Reporter Transgene

Although the photocaged FLP-K223PCK system has been proven to work, the visibility of the mKate2 was very poor when fused to the Chrimson, with visualisation of Chrimson::mKate2 only being possible using objective lenses more powerful than a standard fluorescent stereomicroscope uses, which may prove to limit its wider use as a tool in other labs. We therefore sought to improve the visibility of the system. To do this, we expressed wild-type FLP under a pan-neuronal promoter, *snb-1p*, and the Chrimson construct with the STOP cassette behind the pan-neuronal *maco-1p* promoter. We first swapped around the fluorophores, with FLP being expressed in an operon with mKate2 and Chrimson directly fused to GFP. This was done because the RFP channel on our fluorescent stereomicroscope showed high background fluorescence, which could have prevented the visualisation of dim mKate2 signal, whereas the GFP channel had much less background fluorescence. However, although the mKate2 in the operon with FLP could now be easily seen, the Chrimson::GFP fusion was not visible (Data not shown). We then decided to switch the GFP for

a YFP, Citrine2, for similar reasoning, as the YFP channel on our fluorescent stereomicroscope showed even less background fluorescence than the GFP or RFP channels. However, the Chrimson::Citrine2 fusion again proved very dim (Fig. 23A) and was difficult to detect using a fluorescent stereomicroscope.

Because fusions of Chrimson to the fluorescent protein proved to be too dim to visualise, while the fluorescent proteins expressed on their own were easily visible, we reasoned that the direct fusion may be the cause. This may be due to misfolding of the fluorescent reporter when in such close proximity to the Chrimson[211]. To attempt to create better folding conditions for the reporter, we inserted a 24 amino acid linker between the Chrimson and the Citrine2, named the Happy Linker (HL), in order to spatially distance the two proteins. However, this did not seem to increase brightness of the Citrine2 signal (Fig. 23B), and remained too dim to detect using the fluorescent stereomicroscope.

Next, we decided to separate Chrimson from the fluorescent reporter, so that any potential folding issues caused by the fusion of the two proteins could be avoided completely. To do this we expressed Citrine2 in an operon with Chrimson, by placing the intergenic region from the *gdp-2/3* operon between them, named 'SL2', which contains an SL2-accepting splice site. The mechanisms of these operons are discussed further in Section 2.2.2.1. Briefly, all genes in the operon are transcribed from a single promoter to generate a polycistronic pre-messenger RNA (pre-mRNA), which is then further processed to generate multiple individual mRNAs corresponding to each gene from within the operon[212]. Because this processing occurs after transcription, we reasoned that expression of both genes should be subject to the transcriptional terminator in the STOP cassette, preventing expression until excision by FLP. We initially looked at the embryos of these strains immediately after bombardment and found bright mKate2 and Citrine2 signals throughout the egg (Data not shown), which we initially believed to be the expected pan-neuronal signal. We therefore generated strains with FLP-K223PCK and the Chrimson::SL2::Citrine2. However, we found that the embryos of these animals also showed Citrine2 signal, despite not being fed on PCK or uncaged. When observed at a later life stage, we found that expression of the Citrine2 was seen in a ubiquitous pattern (Fig. 23C), despite the FLP-K223PCK not being activated, and no ubiquitous promoter being present on any construct. The mKate2 in the operon with FLP was expressed pan-neuronally, as expected. This is briefly explored further in Sections 2.2.2.2 and 2.2.2.3.

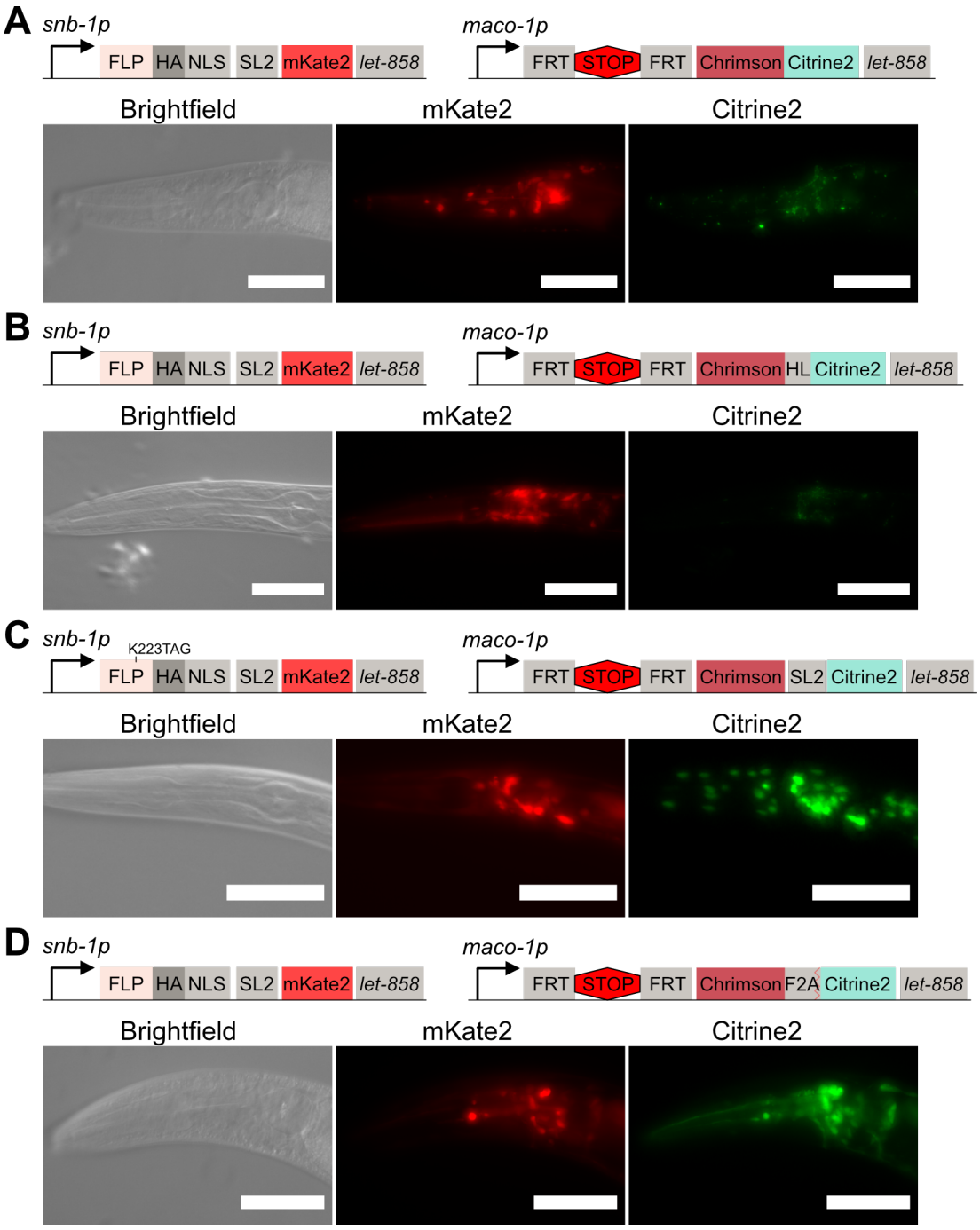

Figure 23: Various Methods to Improve Visualisation of Chrimson::FP Fusion. **(A)** Direct fusion of Chrimson::Citrine2. **(B)** Increasing the length of the linker (Happy Linker - HL) between Chrimson and Citrine2. **(C)** expressing Citrine2 in an operon with Chrimson. **(D)** Placing a self-cleaving peptide (F2A) between Chrimson and Citrine2. All scale bars are 25 μm. All brightfield images are single slices, fluorescent images are max intensity projections.

Finally, we decided to use the F2A self-cleaving peptide from the foot-and-mouth disease virus[213] to separate Chrimson and Citrine2. The F2A peptide, along with other peptides of the 2A family[214], are short amino acid sequences which cleave themselves during translation. By placing these sequences between two protein-coding genes, translation of both proteins can be performed in one process, but the cleaving of the F2A results in both proteins being present after translation as separate units, and in this way any folding issues resulting from the direct fusion between Chrimson and Citrine2 should be avoided. Indeed, expression of this construct showed clearly visible Citrine2 signal in all neurons (Fig. 23D). Although we also found that adding an NLS to the Citrine2 increased visibility even further (Data not shown), we decided to continue without the NLS, as this was visible enough to work with without risking increasing the load of proteins in the nucleus.

As the Chrimson::F2A::Citrine2 proved to be very easily visible, we next tested these constructs with the photocaged K223 (Fig. 24A). For these, we decided to use the *glr-p* promoter, which expresses in the glutamatergic neurons. We chose this promoter because we found it gave much more stable expression than *mec-7p* and with much less mosaicism, and it offered a number of suitable and interesting candidates to test behaviourally with our system, as discussed later. As expected, no Citrine2 signal is seen before uncaging of FLP-K223PCK. After global uncaging, Citrine2 signal could be seen strongly in all glutamatergic neurons (Fig. 24B). Therefore, all future experiments with photocaged FLP expressed FLP in an operon with mKate2. Behavioural experiments expressed Chrimson::F2A::Citrine2 behind the STOP cassette, while non-behavioural experiments expressed only Citrine2 behind the STOP cassette.

2.2.2.1 Background on Operons in *C. elegans*

Operons are genetic units which express at least two genes from a single promoter, and are often found to express genes involved in the same functional pathways[215]. Although most commonly associated with prokaryotic gene regulation, such as in the case of the *lac* operon used by *E. coli* to express genes involved in lactose metabolism[215], operons are also found in some eukaryotes, though this is much less common than in prokaryotes[216]. One eukaryote where operons are relatively common is *C. elegans*, with 15% of genes organised into over 1000 operons[217,218].

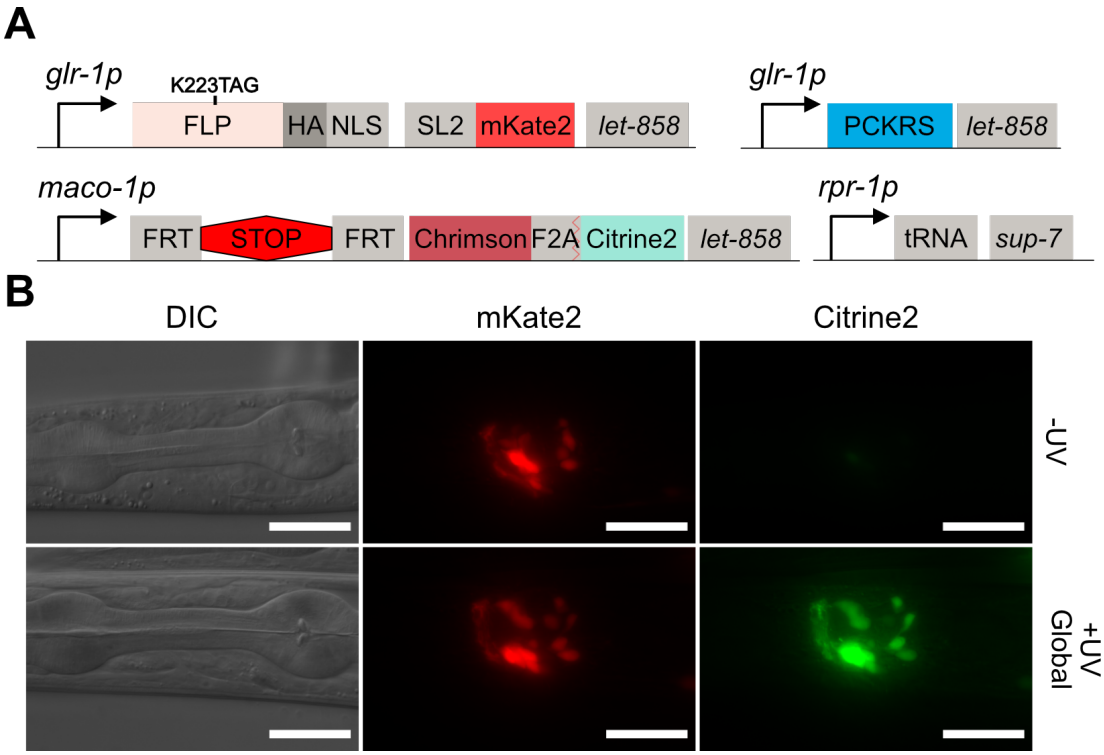

Figure 24: FLP-K223PCK Driving Expression of Chrimson::F2A::Citrine2 in Glutamatergic Neurons. **(A)** Genetic constructs for expression of photocaged FLP-K223PCK in glutamatergic neurons, to drive expression of Chrimson::F2A::Citrine2. **(B)** Expression of constructs in worms grown on PCK, with or without activation of FLP. Scale bars are 25 μm. All DIC images are single slices, fluorescent images are max intensity projections.

In *C. elegans*, operons generally consist of between 2-8 separate genes[218], which are transcribed as a single pre-mRNA. The polycistronic pre-mRNA then undergoes a trans-splicing event, which generates multiple monocistronic mRNAs[219]. During trans-splicing, a cleavage and polyadenylation sequence at the end of each gene results in the generation of monocistronic pre-mRNAs. In most *C. elegans* operons, the 5' end of the first gene's pre-mRNA is then spliced to the Spliced Leader 1 (SL1), an RNA sequence which provides a cap for the pre-mRNA and generates a fully matured mRNA[220]. Downstream genes, on the other hand, are generally spliced to the Spliced Leader 2 (SL2) site to generate the mature mRNAs. There seems to be some flexibility, with some upstream genes in operons being spliced to SL2 and downstream genes spliced to SL1, though

this is much less common[220]. These SL sites provide caps to the mRNAs which are required for the stability of the mRNA, protecting them from exonuclease activity, as well as controlling the recruitment of translation machinery[221].

Although it has not been definitively proven why some genes are clustered into operons, the most common line of reasoning is that operons generally express genes that are involved in similar functions, which allows them to be regulated together. This is the case in prokaryotic operons[217], and has been found to be true in many cases for the operons of *C. elegans*. For example, there exists operons which express multiple subunits of a single protein complex, or operons that express a protein along with a second protein that modifies the activity of the first, or operons that express multiple related mitochondrial proteins[222]. Genes contained within operons that were previously thought to be unrelated or have unknown functions have often subsequently been found to be involved in related functions[219,223]. For example, the gene encoding the ICln ion channel was found contained in an operon with two other genes, though the functions of these other genes was unknown. However, based on their presence in the operon, it was speculated and then experimentally confirmed that these genes encoded proteins that were involved in regulating the ICln channel[224]. However, not all genes found in operons together are functionally related, nor do they always seem to require the same level of regulation[219], and so it is not always clear why certain operons exist.

Although genes in an operon are co-expressed from a single pre-mRNA, their expression levels are not equal, with transcripts of the first gene in an operon expressing at levels approximately twice as high as the downstream genes[225]. Since downstream genes show no further decrease in transcript levels, it is likely a difference between the SL1 and SL2 sites which confers differing levels of stability on the transcripts, with SL2 being seemingly less stable than SL1. Additionally, in endogenous operons, a correlation is found between codon usage and gene position in operons, with upstream genes generally having much more optimised codon usage than downstream genes[225], likely increasing the translation efficiency of the upstream gene and leading to higher expression levels of the protein.

Operon structures are used widely when expressing transgenes in *C. elegans* due to the possibility of co-expressing two transgenes without having to physically link the proteins together, which can in some cases cause problems for the folding and function of the proteins. For example, a transgene can be co-expressed

with a fluorophore, which would allow for visualisation of the expression of the transgene without disrupting the function of the first transgene. This is most often done by using the intergenic region from between the *gpd-2* and *gpd-3* genes[166], which contains the cleavage site, the polyadenylation site, and the SL2 accepting site. *gpd-2* and *gpd-3* are contained within the operon CEOPX036, which contains *mai-1* as the upstream gene, and *gpd-2*, *gpd-3*, and *gpcp-1* as downstream genes[138]. By placing the *gdp-2/3* intergenic region between two transgenes, the transgenes are then treated as if they were contained within the operon and become co-expressed. For brevity, the use of the *gpd-2/3* intergenic region in a genetic construct will be referred to simply as "SL2".

In our hands, the use of SL2 has consistently and robustly allowed for the co-expression of two transgenes under the same promoter. However, here we found that expressing Chrimson::SL2::Citrine2 behind a transcriptional terminator resulted in expression of the Citrine2 protein in an apparently ubiquitous and constitutive pattern. However, we had expected the transcriptional terminator to prevent any production of the polycistronic pre-mRNA containing both Chrimson and Citrine2, thereby preventing expression of both proteins. Additionally, the promoter upstream of the transcriptional terminator was a pan-neuronal promoter, while the expression pattern seen for Citrine2 was ubiquitous. Therefore, it seemed that it was not simply a case of the SL2 element somehow bypassing the terminator. Although the reasons for this are still unknown, in the following subsections, we briefly explore the use of SL2 to co-express transgenes behind a transcriptional terminator and discuss possible reasons for the abnormal expression patterns seen.

2.2.2.2 A Transcriptional Terminator Preceding Chrimson::SL2::Citrine2 Results in Ubiquitous Expression of Citrine2

During testing of a Chrimson::Citrine2 fusion protein, we found that Citrine2 signal was too dim to be easily visualised. Due to the possibility of the direct fusion causing folding issues for Citrine2, we attempted to physically separate the two proteins to allow for better visualisation. To do this, we co-expressed Chrimson and Citrine2 in an operon by placing the SL2 sequence between them. These constructs were placed downstream from the pan-neuronal *maco-1p* promoter, with a transcriptional terminator between the promoter and the Chrimson::SL2::Citrine2. This was expressed alongside a photocaged FLP

recombinase, which was expressed by the pan-neuronal *snb-1p* promoter and co-expressed with an mKate2 reporter in an operon (Fig 25A). In theory, we would expect to see no Citrine2 signal before activation of the photocaged FLP, with activation then leading to expression of Citrine2, as well as Chrimson, only in the neurons. However, we found expression of Citrine2 in an apparent ubiquitous pattern, even before feeding the worms on photocaged amino acid and activating the FLP recombinase (Fig 25B,C). This signal was strongest where it overlaps with FLP expression, but is also clearly visible throughout the body.

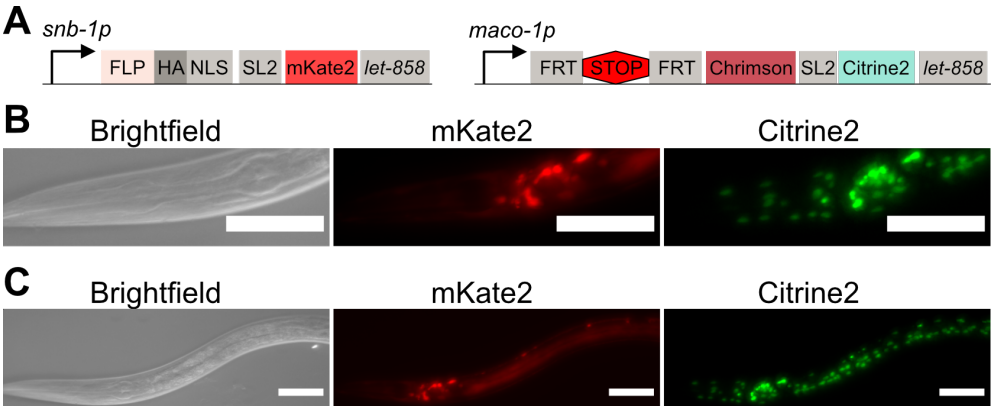

Figure 25: Expression of Chrimson::SL2::Citrine2 Behind a STOP Cassette Driven by a Pan-neuronal Promoter. **(A)** Genetic constructs. **(B)** Expression of mKate2 and Citrine2 in the head. **(C)** Expression of mKate2 and Citrine2 across the entire body. **(B)** and **(C)** show the same worm and is the same as that shown in Fig 23C. Worm was not fed on photocaged amino acid and not subjected to UV illumination before imaging. Scale bars are 50 μm. Fluorescent images are maximum intensity projections, brightfield images are single slices.

2.2.2.3 Only Transgene Downstream of SL2 is Expressed Behind a Transcriptional Terminator

Although the ubiquitous and constitutive expression of Citrine2 was unexpected based on our knowledge of SL2 operon elements, it is unclear to what extent the introduction of the transcriptional terminator broke the expected expression pattern. In particular, we did not know if the transgene before the SL2 acceptor site, in our case Chrimson, was also expressed abnormally. Therefore, we designed constructs with an mKate2 fluorescent reporter before the SL2 acceptor

site, and a Citrine2 reporter after the SL2 acceptor site, which would allow us to visualise expression of both transgenes within the operon. These were driven by the *glr-1p* promoter, which was chosen both to rule out any potential effects of the promoter, and because we found our *snb-1p* promoter to express very inconsistently, whereas *glr-1p* resulted in much more consistent and clean expression of the transgenes. The *lacZ* 3' UTR was placed between the promoter and the first fluorescent reporter as a transcriptional terminator, with flanking FRT sites. This construct was either introduced into worms alone, or alongside a *glr-1p*-driven wild-type FLP recombinase, co-expressed with a CFP fluorescent reporter. We also generated worms with the mKate2::SL2::Citrine2 construct without a preceding transcriptional terminator. All constructs are shown in Figure 26.

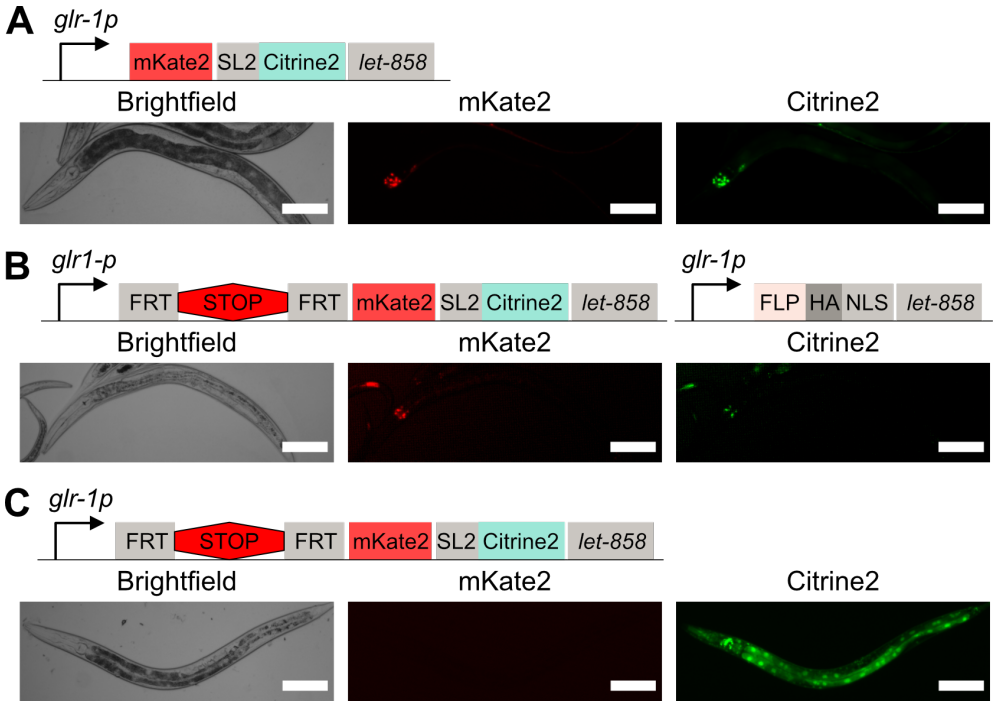

Figure 26: Expression of mKate2::SL2::Citrine2 With or Without a STOP Cassette and With or Without FLP Recombinase. **(A)** Expression of mKate2::SL2::Citrine2 without a STOP cassette. **(B)** Expression of mKate2::SL2::Citrine2 with a STOP cassette in the presence of FLP recombinase. **(C)** Expression of mKate2::SL2::Citrine2 with a STOP cassette in the absence of FLP recombinase. Scale bars are 50 μm. Fluorescent images are maximum intensity projections, brightfield images are single slices.

As expected, worms expressing mKate2::SL2::Citrine2 without the transcriptional terminator showed signal for both fluorescent reporters only in the glutamatergic neurons (Fig 26A). Likewise, worms expressing mKate2::SL2:Citrine2 with the transcriptional terminator in the presence of wild-type FLP recombinase showed expression of both fluorescent reporters only in the glutamatergic neurons (Fig 26B). No worms were found on these plates that showed a ubiquitous expression pattern of Citrine2. When the transcriptional terminator was placed before mKate2::SL2::Citrine2 and no FLP was included to excise it, we again see a ubiquitous expression pattern of the Citrine2 (Fig 26C). As would be expected, we see no expression of mKate2. It therefore seems that only the transgene downstream of the SL2 acceptor site is abnormally expressed, while the transgene upstream of the SL2 acceptor site is expressed as expected of a gene placed behind a transcriptional terminator. The abnormal expression pattern seems to be caused by addition of a transcriptional terminator upstream of both transgenes, and this effect is removed with removal of the terminator.

It should be noted that, for unknown reasons, no CFP signal was seen co-expressed with FLP recombinase. This has also been seen by others in our lab using the same entry plasmid containing the CFP, and therefore may be a fault with the plasmid itself. For this reason, the CFP channel was not imaged for these experiments.

2.2.3 FLP-K223PCK Can Drive Gene Expression in Multiple Different Tissues

Although our primary focus with the photocaged FLP was as a tool for neuroscience research, it also has potential for use in other areas of research on *C. elegans*. Therefore, we also tested whether FLP-K223PCK could be used to drive expression in a range of different tissues. For this, we chose two other tissues to test the system in: the body wall muscle and the gut. For the body wall muscle test, FLP-K223PCK expression was driven by the *myo-3p* promoter as before. For testing gut expression, the *elt-2p* was used, which primarily expresses in the intestine. For both body wall muscle and gut expression, the ubiquitous *eft-3p* promoter was used to drive expression of Citrine2 after STOP cassette excision. The promoter driving expression of the PCKRS was the same as that driving expression of the FLP, while the tRNA expression was driven by *rpr-1p*.

As can be seen in Fig. 27, FLP-K223PCK was successfully expressed in both tissues, as indicated by expression of the mKate2 co-marker. Citrine2 expression was seen in both tissues only after uncaging and activation of FLP-K223PCK to excise the STOP cassette. In addition to using the system in neurons, we have shown that our system can be used to successfully drive gene expression in a range of different tissues with high temporal control. It should be noted that the number of worms expressing Citrine2 was very high in the body wall muscle, but very low in the gut. Although the data was not recorded for these strains, visually at least 90% of animals expressing FLP-K223PCK in the body wall muscle showed expression of Citrine2 after uncaging. However, less than 10% of animals expressing FLP-K223PCK in the gut showed expression of Citrine2 after uncaging. This may be due to the higher penetrance the UV light requires to reach the gut cells, which are at the very centre of the worm body, as opposed to the body wall muscle cells which are very close to the edge of the worm's body. Additionally, a high PCK concentration was used to feed these worms, which we later found to slightly decrease uncaging efficiency (Section 2.2.4). This could be due to the excess PCK absorbing the light used for uncaging, preventing as much light from reaching the PCK within the FLP proteins. It may be possible therefore to increase uncaging efficiency in the gut, as well as in other tissues, by decreasing the PCK concentration, removing worms from the PCK source for some time before uncaging to allow for the flushing out of excess PCK, or by increasing the power of the light used to uncage.

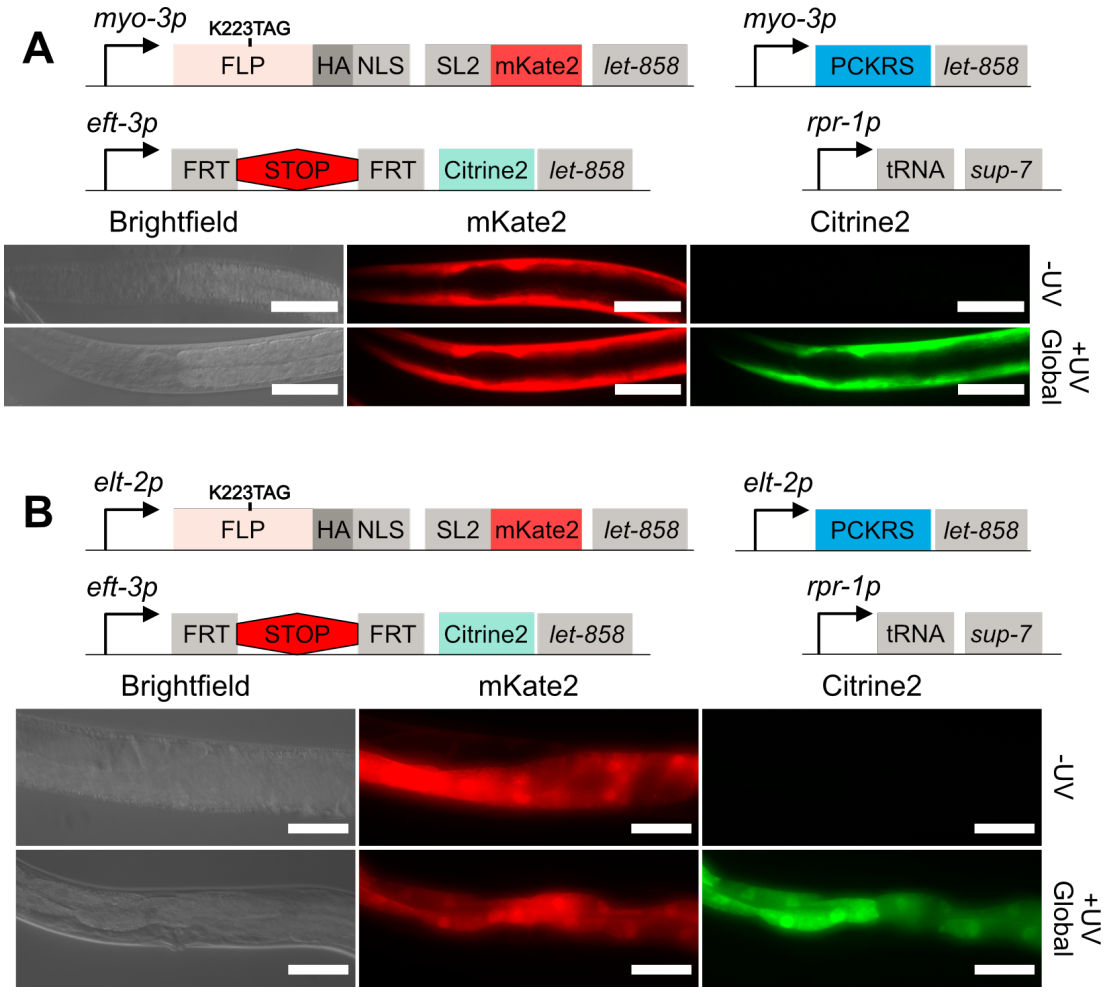

Figure 27: FLP-K223PCK Can Drive Gene Expression in Multiple Tissues. FLP-K223PCK-mediated expression of Citrine2 in the **(A)** body wall muscles and **(B)** gut. Scale bars are 25 µ. Brightfield images are single slices. Fluorescent images in **A** are single slices, in **B** are max intensity projections.

2.2.4 Minimum Conditions for Use of FLP-K223PCK

With a working and visible FLP-K223PCK system developed, we next sought to characterise the minimum conditions for activation of FLP-K223PCK with regards to how long it takes to uptake PCK to produce enough FLP-K223PCK for recombination activity, and how little PCK is required. Determining the minimum time for PCK uptake determines the minimum life stage that can be studied with this tool, though it should be noted that, while untested in this study, worms that hatch from eggs laid by adults grown on PCK could also be viable. The concentration of PCK required is also important for general use of the tool, as PCK is currently expensive and may be prohibitive for many labs if a high concentration is required.

To test these conditions, we generated strains containing photocaged FLP-K223PCK driven by the *glr-1p* promoter, along with a Citrine2 driven by *maco-1p* with a STOP cassette, as well as the necessary incorporation machinery (Fig. 28A). We also generated equivalent strains containing Cre-K201PCK in place of FLP-K223PCK, as described by Davis *et al*[77] (Fig. 28B), which allowed us to compare both tools.

Worms synchronised at the L1 stage were plated on NGM plates supplemented with varying concentrations of PCK. After set times, worms were uncaged by global illumination with 365 nm light and returned to NGM plates without PCK to recover and begin expression of the Citrine2 protein. All worms were imaged at the adult stage, approximately 48 hours after the final uncaging time point. Average fluorescence values for both mKate2 and Citrine2 were calculated for the area of the head between the two pharyngeal bulbs, which contained the majority of *glr-1*-expressing neurons (Fig. 28C). The raw results for this experiment are shown in Fig. 29 for FLP-K223PCK and wild-type FLP, and Fig. 30 for Cre-K223PCK. It should be noted that, although the Citrine2 fluorescence in the uncaged FLP-K223PCK strains did not reach nearly as high as in the wild-type FLP strains (Fig. 29), both were very visible even under a standard fluorescent stereomicroscope. Fluorescence values for Cre-K201PCK strains were similar to those for FLP-K223PCK strains.

We next determined how many worms in each population had uncaged and activated expression of Citrine2. For both FLP-K223PCK and Cre-K223PCK, we defined 'uncaged' worms as those that had an average Citrine2 fluorescence above

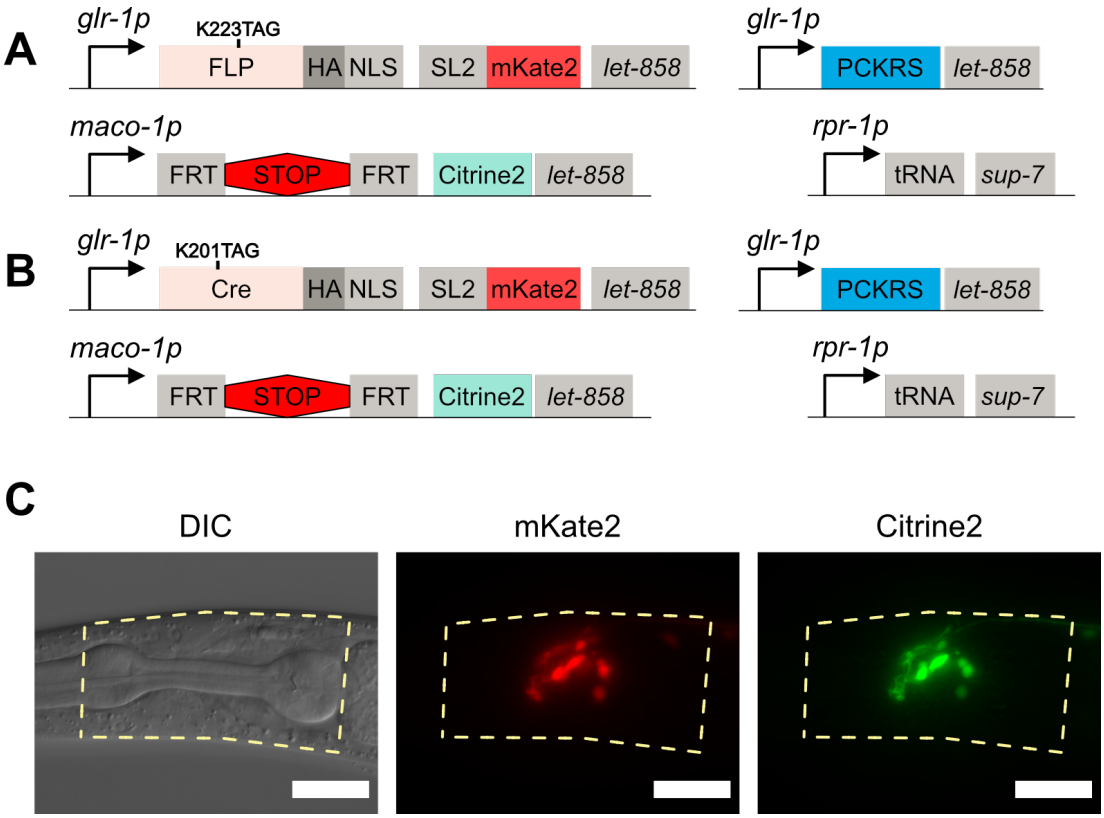

Figure 28: Strains Used for Characterisation of Minimum Conditions for FLP-K223PCK and Cre-K201PCK. **(A)** Genetic constructs for expression of FLP-K223PCK. **(B)** Genetic constructs for expression of Cre-K201PCK. **(C)** Yellow mask shows the area of the worm head used for counting fluorescence of mKate2 and Citrine2.

a baseline. This baseline was set to be slightly higher than the highest Citrine2 fluorescence value found in uncaged worms (-UV), not including any obvious outliers. This data is shown in Fig. 31 and Fig. 32 for FLP-K223PCK and Cre-K223PCK, respectively. For both cases, we found some worms expressing very low levels of mKate2, likely due to the presence of the transgenes on an extrachromosomal array giving rise to some mosaicism in the expression. To account for these, we also separately considered only the top 75% of worms as ranked by mKate2 fluorescence (Fig. 31B and Fig. 32B), which removed the majority of animals that expressed no or very low levels of mKate2.

As seen, no or almost no worms show Citrine2 fluorescence levels corresponding to active FLP or Cre without uncaging (-UV), even when fed

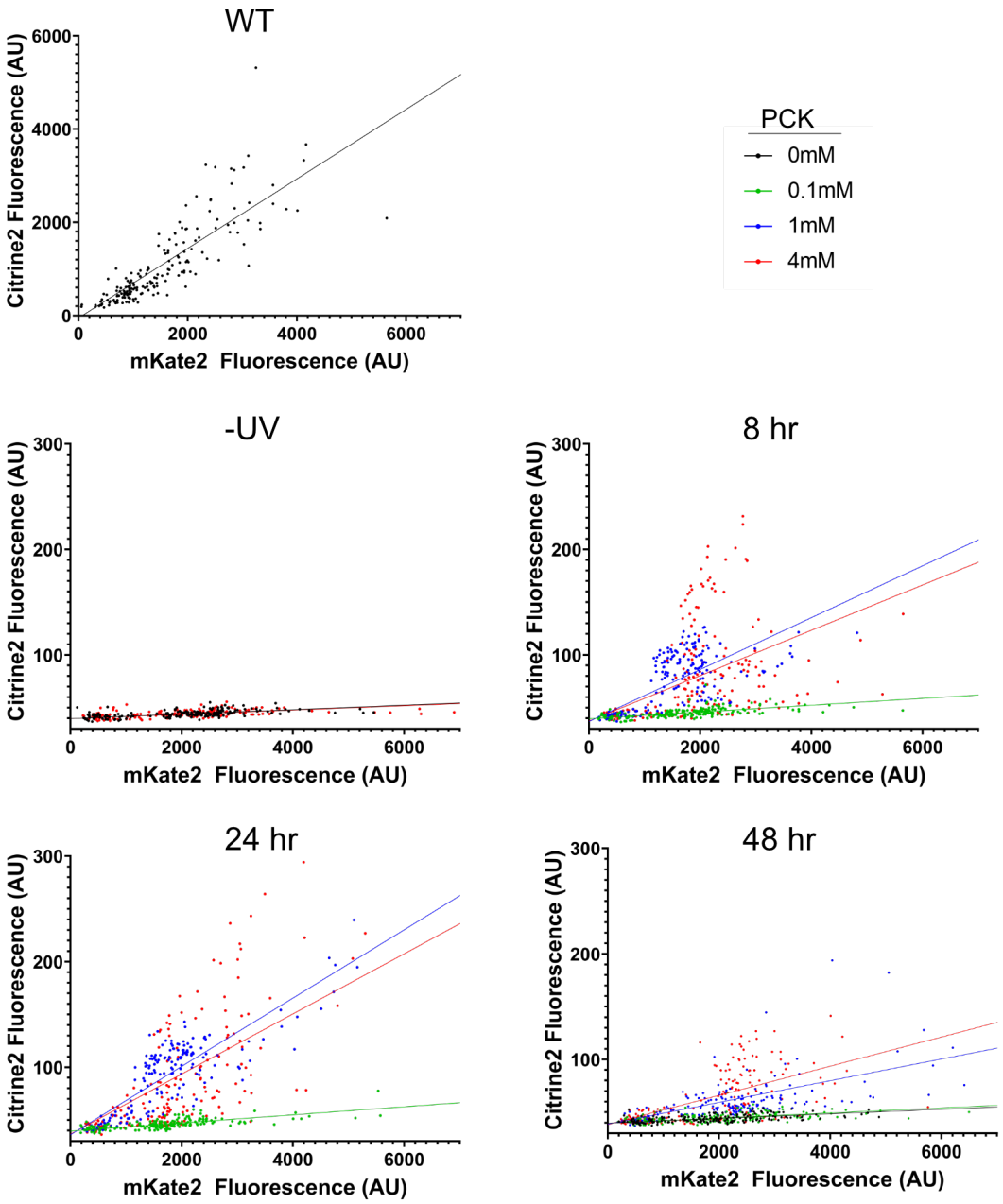

Figure 29: Characterisation of FLP-K223PCK and Wild-type FLP. Raw fluorescent data from FLP characterisations. Graph titles refer to the time on PCK before uncaging, except for the top right graph which shows worms that were not uncaged. Coloured dots show the fluorescence values for each individual worm counted. Different colours are used to refer to the concentration of PCK that worms were fed on, as detailed in the legend. Trend lines are simple linear regressions.

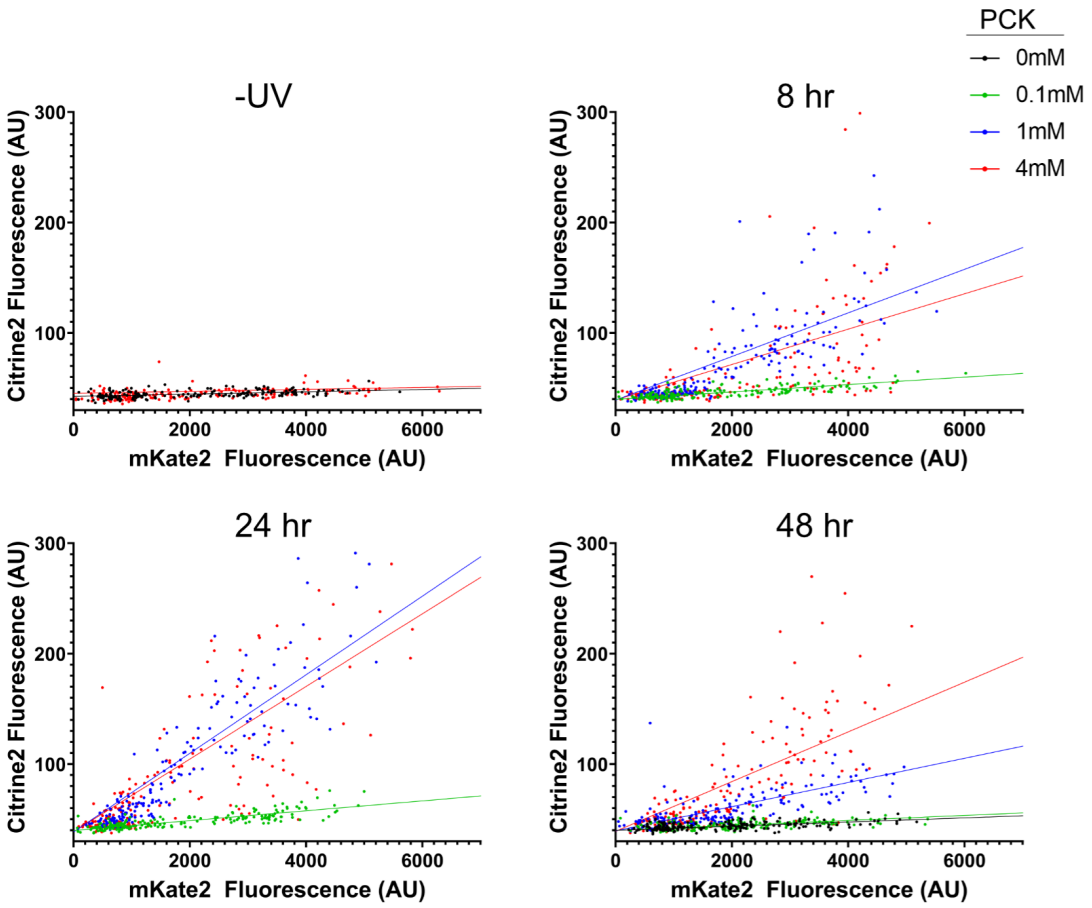

Figure 30: Characterisation of Cre-K201PCK. Raw fluorescent data from Cre-K201PCK characterisations. Graph titles refer to the time on PCK before uncaging, except for the top right graph which shows worms that were not uncaged. Coloured dots show the fluorescence values for each individual worm counted. Different colours are used to refer to the concentration of PCK that worms were fed on, as detailed in the legend. Trend lines are simple linear regressions.

on 4 mM PCK. Worms also do not show Citrine2 expression after illumination with 365 nm light, but without having been fed on PCK (48 hr, 0 mM PCK). Generally, 0.1 mM PCK was not enough to produce enough FLP-K223PCK or Cre-K201PCK to drive gene expression after activation. Both 1 mM PCK and 4 mM PCK are efficient to produce active recombinases, though no statistically significant difference was found in transgene expression in worms fed on 1 mM PCK or 4 mM PCK. However, with both 8 hr and 24 hr of feeding for both FLP-K223PCK and Cre-K201PCK, the number of animals uncaging after feeding on 1 mM PCK seemed slightly higher than those fed on 4 mM PCK, and so

under more sensitive conditions this may prove to make a statistically significant difference. For example, when expressing in the gut as discussed previously (Fig. 27), reduction of PCK concentration to 1 mM might make enough of a difference to noticeably increase the efficiency of uncaging compared to 4 mM PCK. Feeding worms on PCK for 24 hours before uncaging appeared to give the best uncaging results, with 95±3% and 85±7% of worms showing Citrine2 expression after feeding on 1 mM PCK for 24 hours for FLP-K223PCK and Cre-K201PCK strains, respectively. A full list of pairwise comparisons of the data from Fig. 31 and Fig. 32 is shown in Table 2. We found no statistical significance between FLP-K223PCK and Cre-K201PCK (Fig. 33).

The slight reduction in the number of animals with activated FLP recombinase after 48 hours on PCK is likely an artefact of the promoter used in this study. As worms were plated at the L1 stage and kept at 20 ℃ throughout growth, this would put them at approximately the L4 or young adult stage after 48 hours[226], with variation depending on the exact timing of the egg hatching between bleaching and plating the worms. The *glr-1p* promoter which drives FLP-K223PCK and Cre-K201PCK is most active in the late embryo and earlier larval stages, but its expression drops considerably in the young adult stage (Data obtained from WormBase entry for *glr-1*[138]). As our worms are likely to have reached the young adult stage by the time we uncage them for the 48 hour time point, it is likely that by this time, no or very lower levels of new FLP-K223PCK or Cre-K201PCK are being synthesised. If the proteins have a high turnover rate, then there may be very little protein remaining during the uncaging event, leading to low expression levels of the Citrine2 reporter.

It should also be noted that we could not record data for how many cells in each individual animal had uncaged, but instead only recorded whether an animal had an average fluorescence above a threshold. This was because images were taken with low magnification, and the lack of an NLS on the Citrine2 protein often made it difficult to tell individual neurons apart when in larger clusters. It may be possible that in an individual worm uncaged under optimal conditions, all glutamatergic neurons express Citrine2, whereas in a worm uncaged under sub-optimal conditions, only a few cells express Citrine2. Under both of these conditions, so long as the average Citrine2 fluorescence was still above the threshold, the worm would be considered a positive result and any differences in efficiency would not be noted. It would therefore be worth repeating the

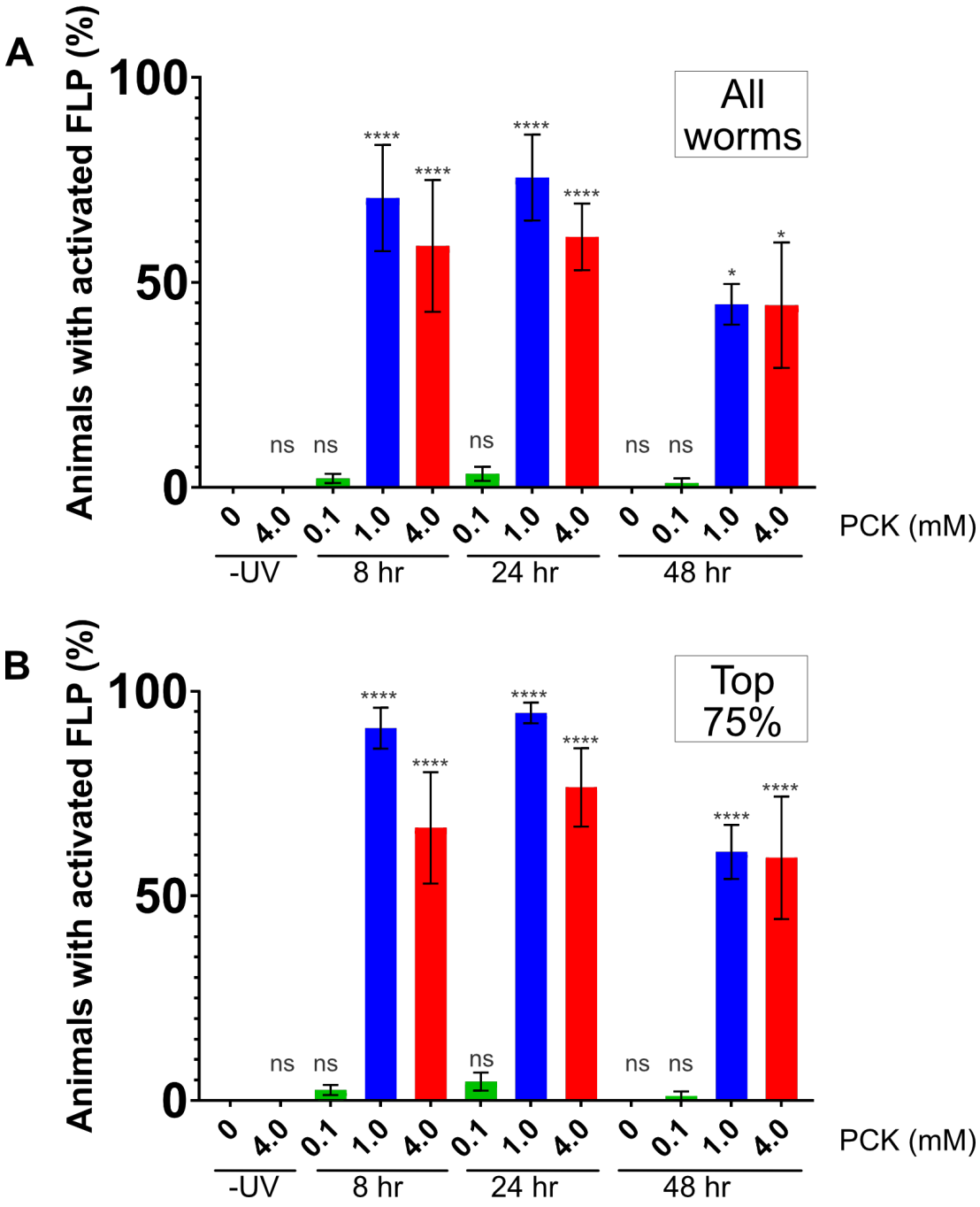

Figure 31: Percentage of Worms Uncaging in FLP-K223PCK Strains. Percentage of worms with a Citrine2 fluorescence level above that of uncaged (-UV) worms. **(A)** All worms analysed. **(B)** The top 75% of worms by mKate2 fluorescence. Mean ± SEM. All statistical significance values relative to 0 mM PCK -UV control. ns = $p>0.05$, * = $p\leq0.05$, *** = $p\leq0.001$, **** = $p\leq0.0001$; One-way ANOVA with post-hoc Sidak's multiple comparisons.

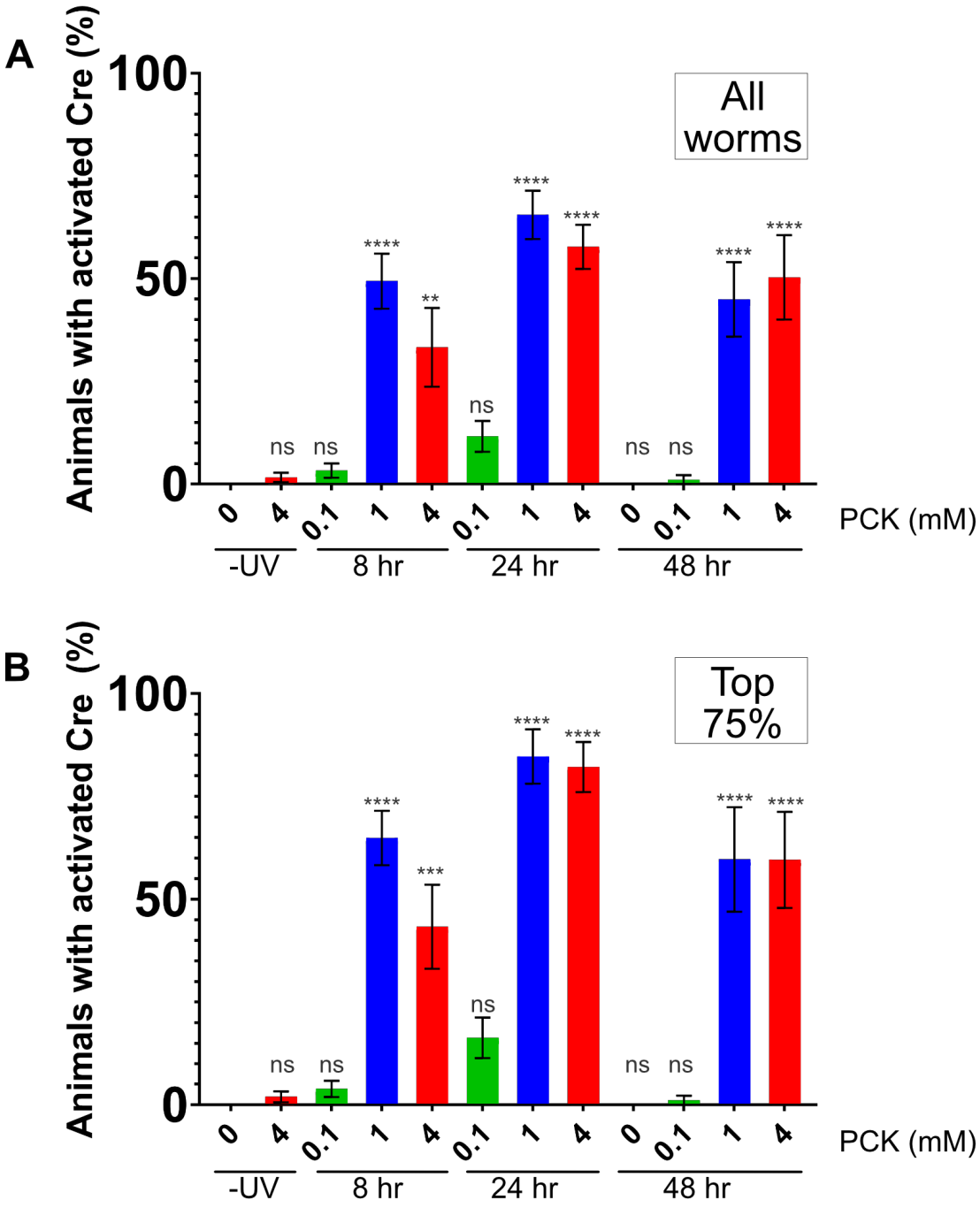

Figure 32: Percentage of Worms Uncaging in Cre-K201PCK Strains. Percentage of worms with a Citrine2 fluorescence level above that of uncaged (-UV) worms. **(A)** All worms analysed. **(B)** The top 75% of worms by mKate2 fluorescence. All statistical significance values relative to 0 mM PCK -UV control. Mean ± SEM. ns = $p>0.05$, * = $p\leq0.05$, *** = $p\leq0.001$, **** = $p\leq0.0001$; One-way ANOVA with post-hoc Sidak's multiple comparisons.

experiment with an NLS on the Citrine2 and with higher magnification images to be able to detect these differences in recombinase activation in individual cells.

Table 2: Pairwise Comparisons for FLP-K223PCK and Cre-K201PCK. Data shown in full in Fig. 31 and Fig. 32. Conditions refer to time on PCK and concentration of PCK.. ns = $p>0.05$, * = $p\leq0.05$; Sidak's multiple comparisons.

FLP-K223PCK			Cre-K201PCK		
Comparison		Significance	Comparison		Significance
8 hr, 1 mM	8hr, 4 mM	ns	8 hr, 1 mM	8hr, 4 mM	ns
24 hr, 1 mM	24hr, 4 mM	ns	24 hr, 1 mM	24hr, 4 mM	ns
48 hr, 1 mM	48hr, 4 mM	ns	48 hr, 1 mM	48hr, 4 mM	ns
8 hr, 1 mM	24 hr, 1 mM	ns	8 hr, 1 mM	24 hr, 1 mM	ns
8 hr, 1 mM	48 hr, 1 mM	ns	8 hr, 1 mM	48 hr, 1 mM	ns
24 hr, 1 mM	48 hr, 1 mM	*	24 hr, 1 mM	48 hr, 1 mM	ns
8 hr, 4 mM	24 hr, 4 mM	ns	8 hr, 4 mM	24 hr, 4 mM	ns
8 hr, 4 mM	48 hr, 4 mM	ns	8 hr, 4 mM	48 hr, 4 mM	ns
24 hr, 4 mM	48 hr, 4 mM	ns	24 hr, 4 mM	48 hr, 4 mM	ns

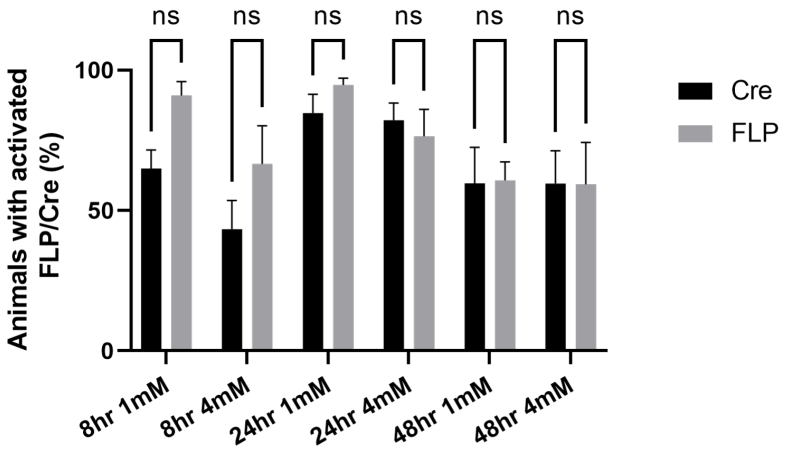

Figure 33: Comparison of FLP-K223PCK and Cre-K223PCK. Percentage of worms with a Citrine2 fluorescence level above that of uncaged worms. Data is the same as that found in Fig. 31 and Fig. 32, represented here to show comparisons between FLP-K223PCK and Cre-K201PCK. Mean ± SEM. ns = $p>0.05$; One-way ANOVA with post-hoc Sidak's multiple comparisons.

2.2.5 FLP-K223PCK and Cre-K201PCK Begin Excising Within 5 Minutes of Activation

We next sought to determine how quickly the photocaged recombinases become active after illumination with 365 nm light. This has previously been

determined *in vitro*, with excision by FLP recombinase beginning within 5 minutes, and the majority of excision events completed within 30 minutes at 23 ℃, though the actual rate of excision would depend on the relative concentrations of FLP and the substrate. On the other hand, Cre seemed very inactive at 23 ℃, with only a low level of excision occurring even after 1 hour[191], though as noted earlier, Cre was much more active at higher temperatures, with the highest level of activity seen at 37–42 ℃. Although FLP recombinase activity has also been measured *in vivo* in mammalian cells[191], this used a system where FLP was fused to the ligand binding domain of the androgen receptor, which inactivated FLP in the absence of the androgen[227]. However, activation of FLP and excision of the DNA was very slow with this setup, with a high level of excision seen 24–48 hours after addition of the androgen, compared to $<$ 1 hour *in vitro* activity[191]. This is likely due to the slow uptake of the androgen into cells, and a potentially lower activity of FLP when fused to another protein domain. With photocaged FLP, we would expect the rate of recombination to more closely match that found in *in vitro* assays, as uncaging occurs on the second or microsecond timescale[111]. Additionally, all FLP-K223PCK is produced prior to the uncaging event and does not depend on expression of the protein after induction, allowing for nearly instantaneous activation of all photocaged FLP-K223PCK proteins.

To test this, we designed PCR primers to amplify the STOP cassette. Nine sets of primers were designed and tested with a range of annealing temperatures, to find the best condition for this experiment (Fig. 34). The primers and primer pairs used are found in Appendix Tables A7 and A8. In Fig. 34A, extension times were used corresponding to the size of the excised band, which ranged from 229 bp (Primer pair **9**) to 653 bp (Primer pair **2**). However, we found that using extension times corresponding to the non-excised band (4 kb) allowed visualisation of both the excised and non-excised bands, despite the large difference in size between the two fragments, and therefore we used this for all other PCRs. It should be noted that this was performed in worms expressing the recombinase from the *glr-1p* promoter, and therefore in the majority of cells the FLP would not be expressed. Due to this, non-excised DNA fragments will always be seen in high abundance, no matter how active the FLP recombinase was. We found primer pair **5** with an annealing temperature of 60 ℃ to be the best condition for detection of excision, and this was used for all following PCRs. As these primers bound in the upstream promoter and the downstream Citrine2 transgene, we used the same primer pair

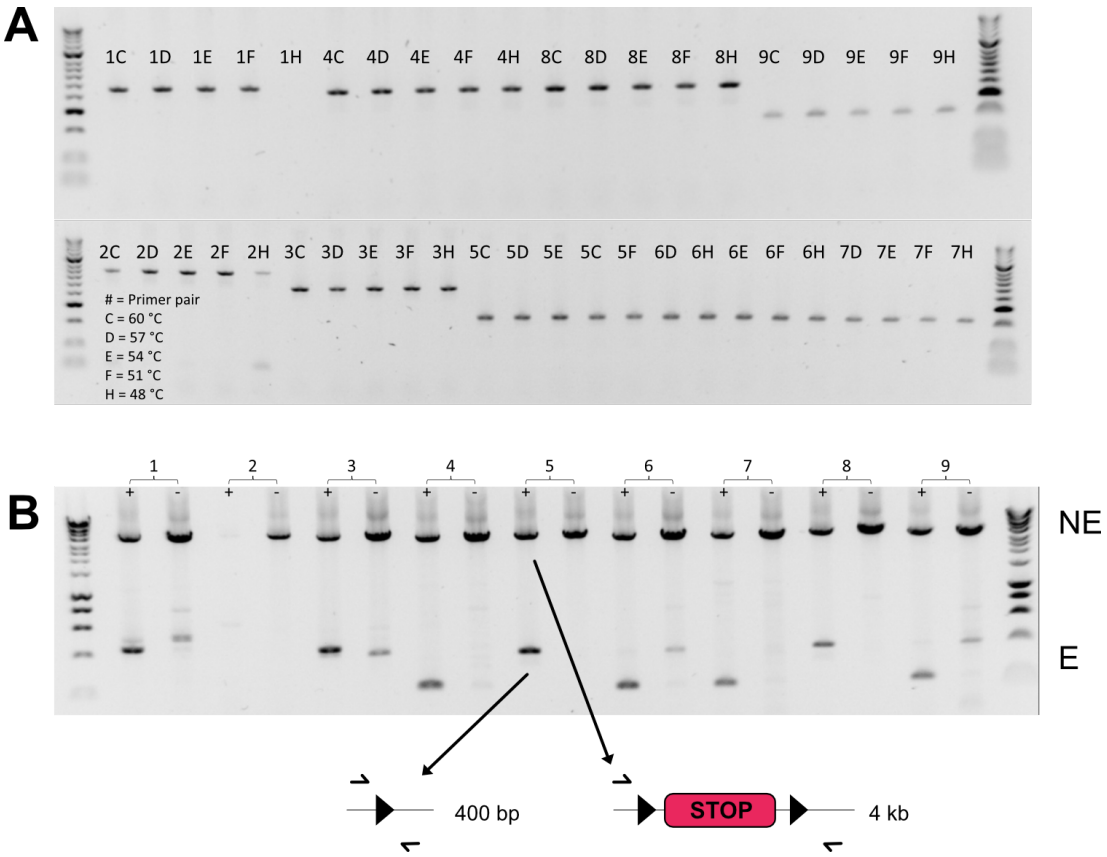

Figure 34: Testing of PCR Conditions for FLP Excision. **(A)** PCR performed with various primer pairs and annealing temperatures on lysates of uncaged worms. Ladder is Hyperladder 50 bp. **(B)** PCR performed with various primer pairs, with 60 ℃ annealing temperature, on lysates of caged and uncaged worms. Ladder is Hyperladder 1kb. +/- = lysates from uncaged/caged worms. NE/E = non-excised/excised DNA. Schematic shows the PCR of the excised and non-excised fragments, with approximate expected sizes.

to test both FLP-K223PCK and Cre-K201PCK activities.

Next, we determined the minimum exposure time to UV to give detectable recombinase activity (Fig. 35), which would allow us to more accurately determine how quickly recombinase activity occurred after UV exposure. Worms were fed on 4 mM PCK and uncaged after 1 day with the indicated exposure time of UV light. Worms were then allowed to recover for 4 hours to allow enough time for the recombinases to excise the majority of the substrate. For both FLP-K223PCK and Cre-K201PCK, we saw little activity after only 30 s of UV

stimulation. For FLP-K223PCK, little difference was seen in the excision above 1 minute, while Cre-K201PCK showed some difference between 1 minute and 2 minutes of UV, but little difference above this. We therefore, chose to illuminate worms for 2 minutes with UV light in all future PCR experiments. However, we note that these PCRs are not quantitative, and therefore any values obtained from these experiments are rough estimates. Additionally, since transgenes are present at varying copy numbers, comparisons between strains expressing FLP-K223PCK and strains expressing Cre-K201PCK cannot be definitively made.

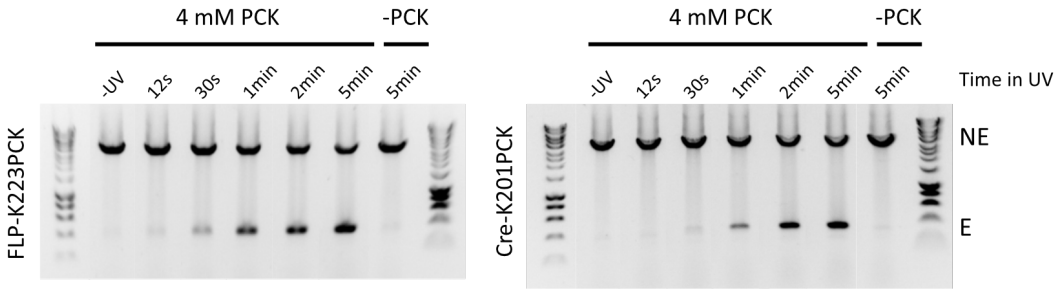

Figure 35: Minimum Time Exposed to UV for FLP-K223PCk and Cre-K201PCK Activity. Worms grown on PCK were exposed to UV light for the indicated times and lysed 4 hours later. PCR was performed to determine how much of the substrate had been excised, and therefore how active the recombinase was. NE/E = non-excised/excised DNA.

With optimal conditions established for uncaging of the recombinases and detection of the excision events, we next sought to determine the time it takes for FLP-K223PCK and Cre-K201PCK to activate and excise their target substrates. This would provide one of the first proper *in vivo* characterisations of FLP and Cre kinetics, and would give insight to researchers wishing to use these tools with regards to the duration of experiments. Worms were grown on 4 mM PCK for 24 hours before being uncaged for 2 minutes in the UV Crosslinker. At certain intervals, lysates were frozen by placing them in PCR tubes on a metal tube rack on dry ice, which fully froze the samples within approximately 5-10 seconds. For FLP-K223PCK, we observed recombination within 5 minutes of UV activation and continuing up to 1 hour after UV activation, with little more recombination occurring after this point (Fig. 36), which aligns relatively closely with *in vitro* findings[191]. As expected, Cre-K201PCK activity appeared relatively low at 20 ℃, though unlike the *in vivo* experiment[191], we do see obvious recombination

occurring earlier than 1 hour after uncaging. In all cases, a very dim band could be seen at the same size as the excised DNA band would be expected. However, running the PCR on the STOP cassette-containing plasmid also revealed this band. It is therefore likely to be an artefact of the PCR itself, rather than a sign of background activity, and may arise from short incomplete PCR fragments annealing together at the FRT sites. Finally, we also tested the stability of the caging group under standard lab lighting, which used fluorescent ceiling lamps. For both recombinases, we saw no uncaging or activity of the recombinase in worms exposed to this lighting for 4 hours.

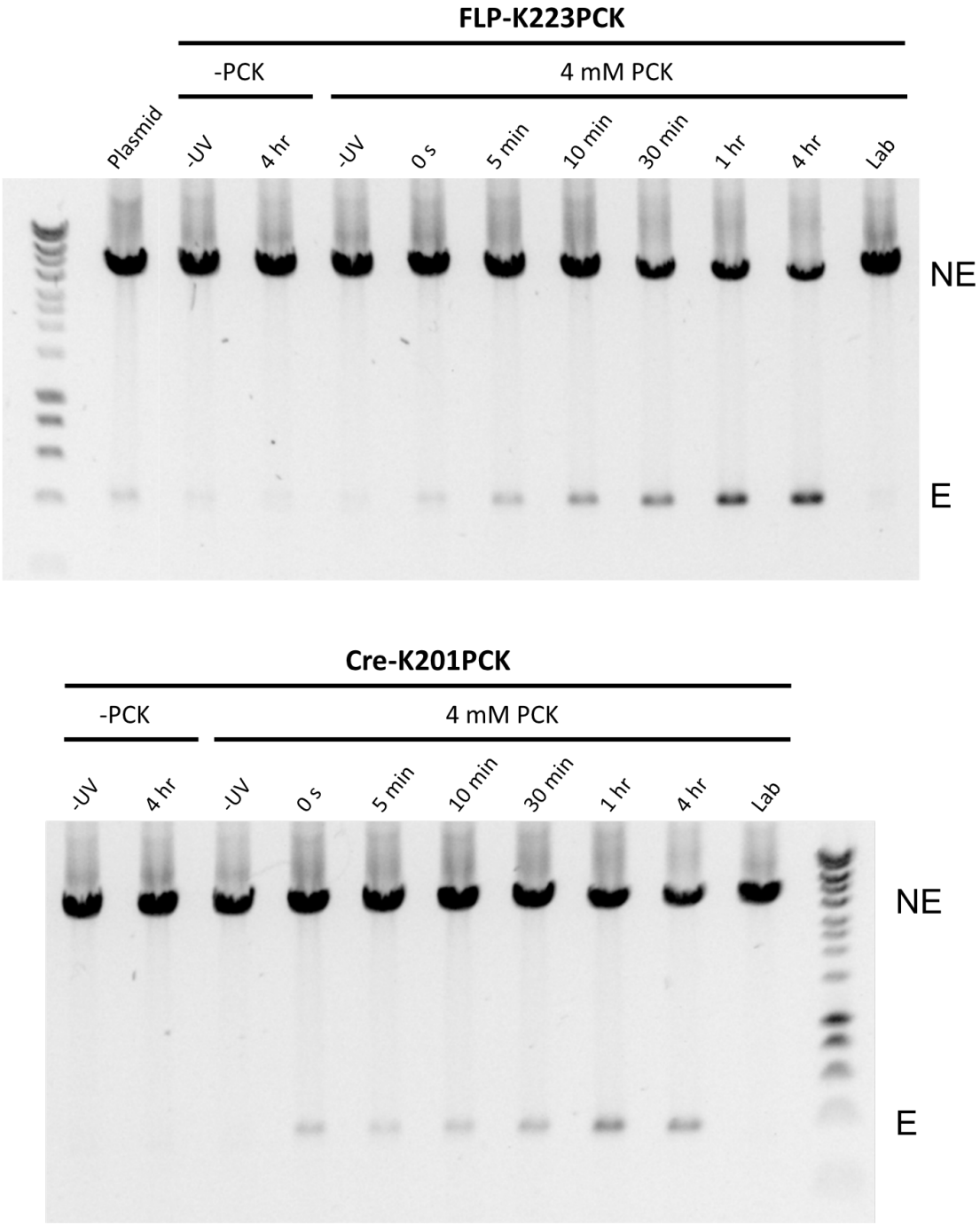

Figure 36: ***In Vivo*** **FLP-K223PCK and Cre-K201PCK Activity.** All worms uncaged for 2 minutes after feeding on 0 mM or 4 mM PCK. Time indicates how long after uncaging worms were frozen to stop reaction. 'Plasmid' is control PCR of only the plasmid used to generate transgenic animals. '-UV' samples were frozen after 4 hr in dark. 'Lab' samples were left on the benchtop exposed to room lights for 4 hr. NE/E = non-excised/excised DNA.

2.2.6 FLP Can Be Photocaged at the Catalytic Tyrosine

With it being possible to photocage FLP activity by replacement of the critical K223 residue with a photocaged lysine, we next wanted to determine if the tyrosine can also be photocaged. This would particularly be useful in future applications where a photocaged Cre and a photocaged FLP could be used at one time in the same system, with one being caged at the lysine residue and one being caged at the tyrosine residue. Additionally, some photocaged variants of tyrosine, notably ortho-nitrobenzyl tyrosine (ONBY), are considerably cheaper than PCK, likely due to the simpler route of synthesis involved[204,228]. Although more recent studies with Cre recombinase have focused on photocaging the lysine[77,188,189], the first study to photocage Cre successfully photocaged the tyrosine[187]. Given that FLP and Cre have nearly identical mechanisms and the critical tyrosine and lysine residues are conserved between them, it should also be possible to photocage FLP at the critical Y343 residue.

Constructs were generated as before, but with the codon at position 343 replaced with the amber stop codon to facilitate incorporation of the photocaged tyrosine. Instead of the PCKRS, we used an aaRS which recognises nitropiperonyl tyrosine (NPY, NPYRS), which has been evolved from the same *M. mazei* PylRS as the PCKRS. All other components, including the plasmid containing the tRNA and the plasmid containing the STOP cassette and citrine2 transgene, were kept identical to those used for testing FLP-K223PCK. Multiple variants of photocaged tyrosine exist, with slight differences in the caging group which change the uncaging dynamics[104]. Although specific aaRS have been developed for some of these photocaged tyrosines with improved efficiencies[229], we find that the NPYRS can be used sufficiently for all variants, removing the requirement to generate new strains for each photocaged tyrosine variant tested. The three variants we test here are nitropiperonyl tyrosine (NPY), ortho-nitrobenzyl tyrosine (ONBY), and methylnitropiperonyl tyrosine (MNPY). The structures of these ncAAs are shown in Fig. 37. Additionally, we also test a dipeptide of lysine and NPY (K-NPY), as dipeptides containing the ncAA have previously been found to improve incorporation[153]. The dipeptide exploits dipeptide transporters to improve uptake, though after uptake the dipeptide would be cleaved to give an NPY residue. Therefore, worms feeding on NPY and worms feeding on K-NPY will produce identical FLP-Y343NPY proteins, and any differences should only be

due to incorporation of K-NPY as compared to NPY, and subsequent processing of K-NPY to give the NPY monopeptide.

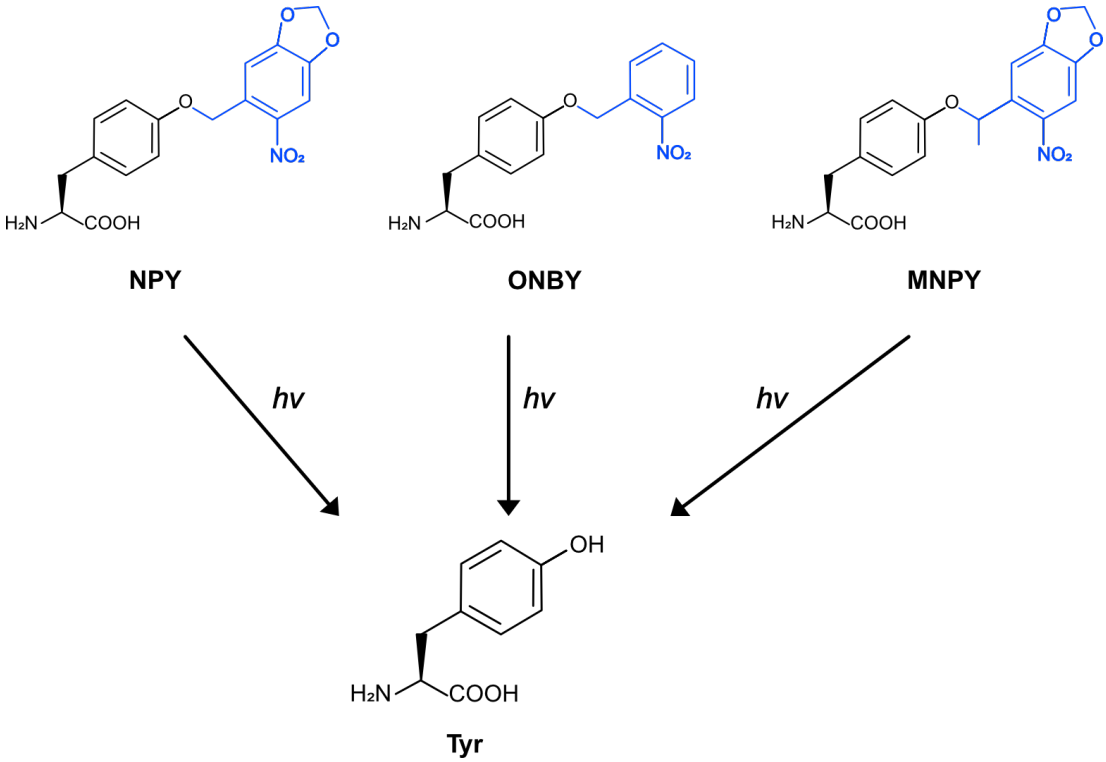

Figure 37: Structures of Photocaged Tyrosines. Nitropiperonyl tyrosine (NPY), Ortho-nitrobenzyl tyrosine (ONBY), and Methylnitropiperonyl tyrosine (MNPY) structures. Caging group is shown in blue. All three can be converted to tyrosine (Tyr) upon illumination with 365 nm light.

With worms expressing these constructs, we performed the same assay as before to detect FLP-mediated excision by PCR. However, due to the low solubility of all photocaged tyrosine variants, other than K-NPY, we could only produce NGM plates containing 0.5 mM ncAA. For K-NPY, we also produced NGM plates containing 4 mM K-NPY, as a more direct comparison with the previous PCK results. All experiments were performed as in Fig. 35, with worms fed on the ncAA for 24 hours before being uncaged for the indicated time, frozen 4 hours later, lysed, and the STOP cassette analysed by PCR. We first compared PCK and K-NPY (Fig. 38). Although we should again note that these are not directly comparable, as the analysis is done in different strains and the PCR is not quantitative, we find that K-NPY seems to work very effectively in photocaging

FLP. Photocaged FLP can be used with as little as 0.5 mM of PCK or K-NPY, and neither strain showed background uncaging even with higher concentrations of ncAA. As with PCK, K-NPY also showed high stability under normal lab lighting, with no uncaging seen. With K-NPY, photocaged FLP showed very high levels of recombination with very little UV, with near maximal uncaging seeming to be reached at 1 minute of UV, with high recombination still seen with less time under UV.

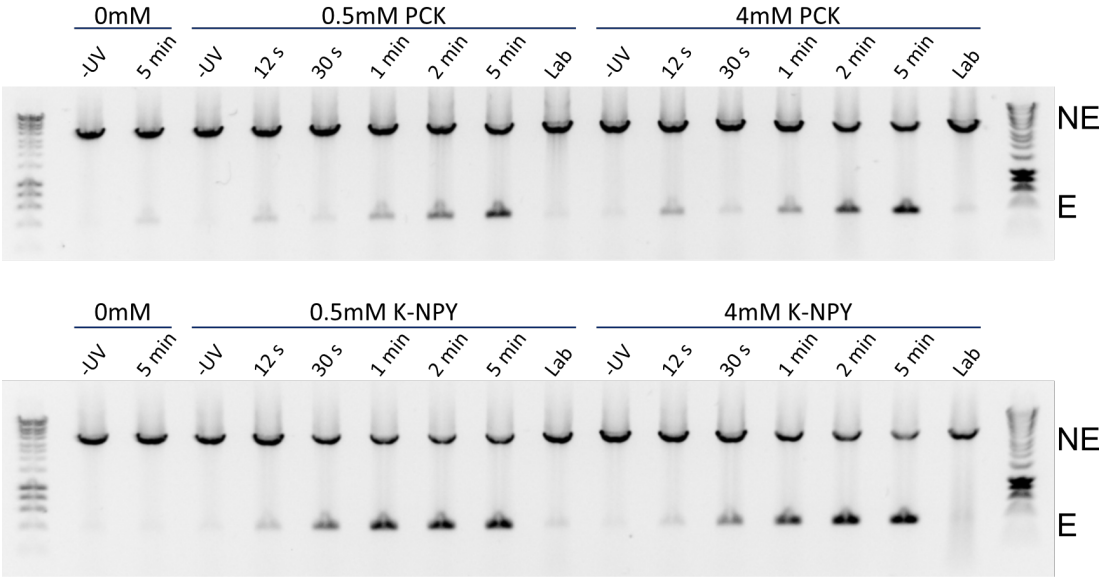

Figure 38: Activity of FLP Photocaged with PCK or K-NPY After Uncaging. Worms grown on 0.5 mM or 4 mM of PCK or K-NPY were uncaged for the times indicated. After four hours, worms were lysed and the excision analysed by PCR. NE/E = non-excised/excised DNA.

We next tested the other three photocaged tyrosine variants, NPY, MNPY, and ONBY (Fig. 39). These are much less soluble, and therefore only the 0.5 mM condition could be tested. We found MNPY and NPY groups to be very sensitive to the 365 nm light, with strong recombination seen after only 30 s of UV exposure. On the other hand, ONBY was much less reactive, with no visible recombination occurring with less than 1 minute of UV. None of the three photocaged tyrosine variants uncaged under normal lab light, and no signal was seen in the absence of the ncAA. There is little obvious difference observed between NPY and K-NPY (Fig. 38 and Fig. 39).

Worms expressing photocaged FLP-K223 and FLP-Y343 and uncaged under

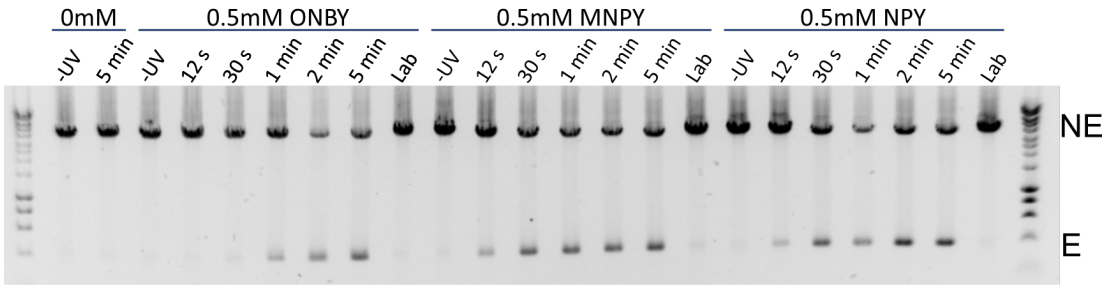

Figure 39: Activity of FLP-Y343NPY, FLP-Y343MNPY, and FLP-Y343ONBY. Worms grown on 0.5 mM of ncAA were uncaged for the times indicated. After four hours, worms were lysed and the excision analysed by PCR. NE/E = non-excised/excised DNA.

the conditions described above were also checked for expression of Citrine2 after 24 hours (Figs 40-46). In all cases, no expression of the Citrine2 transgene is seen without feeding on the photocaged amino acid or without UV illumination. We also see no uncaging in any case when illuminated with only 12 s of UV. All amino acids also seem to be stable under standard lab lighting, with no transgene expression seen. For all photocaged amino acids, bright signal is seen after 5 minutes of UV. For K-NPY, and MNPY, and NPY, 1 minute of UV results in a comparable level of transgene expression (Figs 42, 43, 45, 46). However, with PCK and ONBY, 1 minute of UV uncaging results in slightly dimmer expression of the transgene (Figs 40, 41, 44). These results correlate closely with the level of uncaging observed by excision of the STOP cassette in Figures 38 and 39.

Although we have shown that all three photocaged tyrosine variants can be used to successfully photocage FLP, further characterisation would need to be performed before they are used for practical purposes. Although ONBY appeared to be the least efficient residue, it is roughly similar to PCK in its uncaging efficiency and its low cost makes it an enticing choice for future work. As with PCK, all photocaged tyrosine variants appear to completely block FLP activity until uncaging.

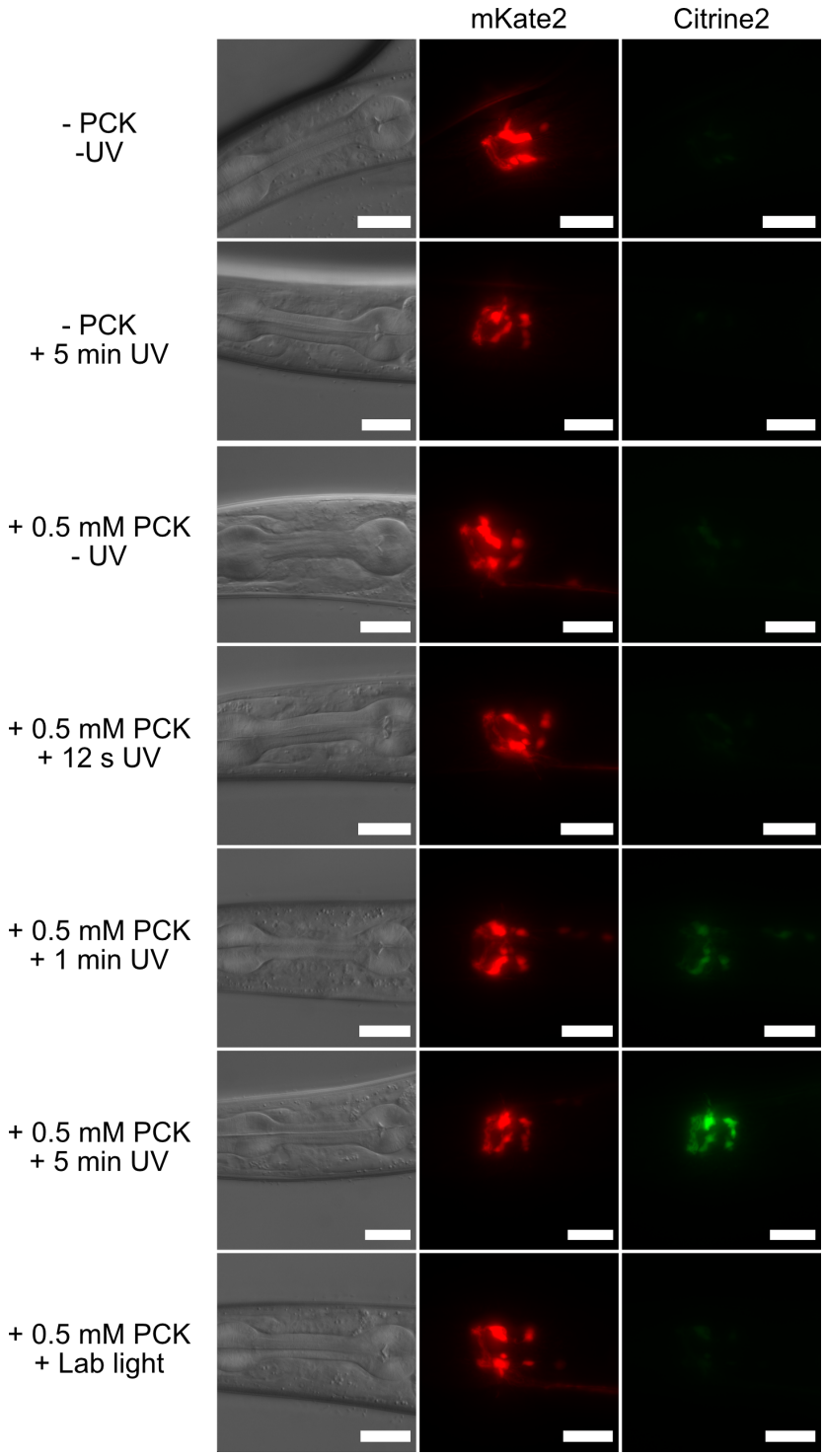

Figure 40: FLP-K223PCK With 0.5 mM PCK. Worms grown on 0.5 mM PCK and uncaged by illumination with UV for indicated time. DIC images are single slices, fluorescent images are maximum intensity projections. Scale bars are 25 μm.

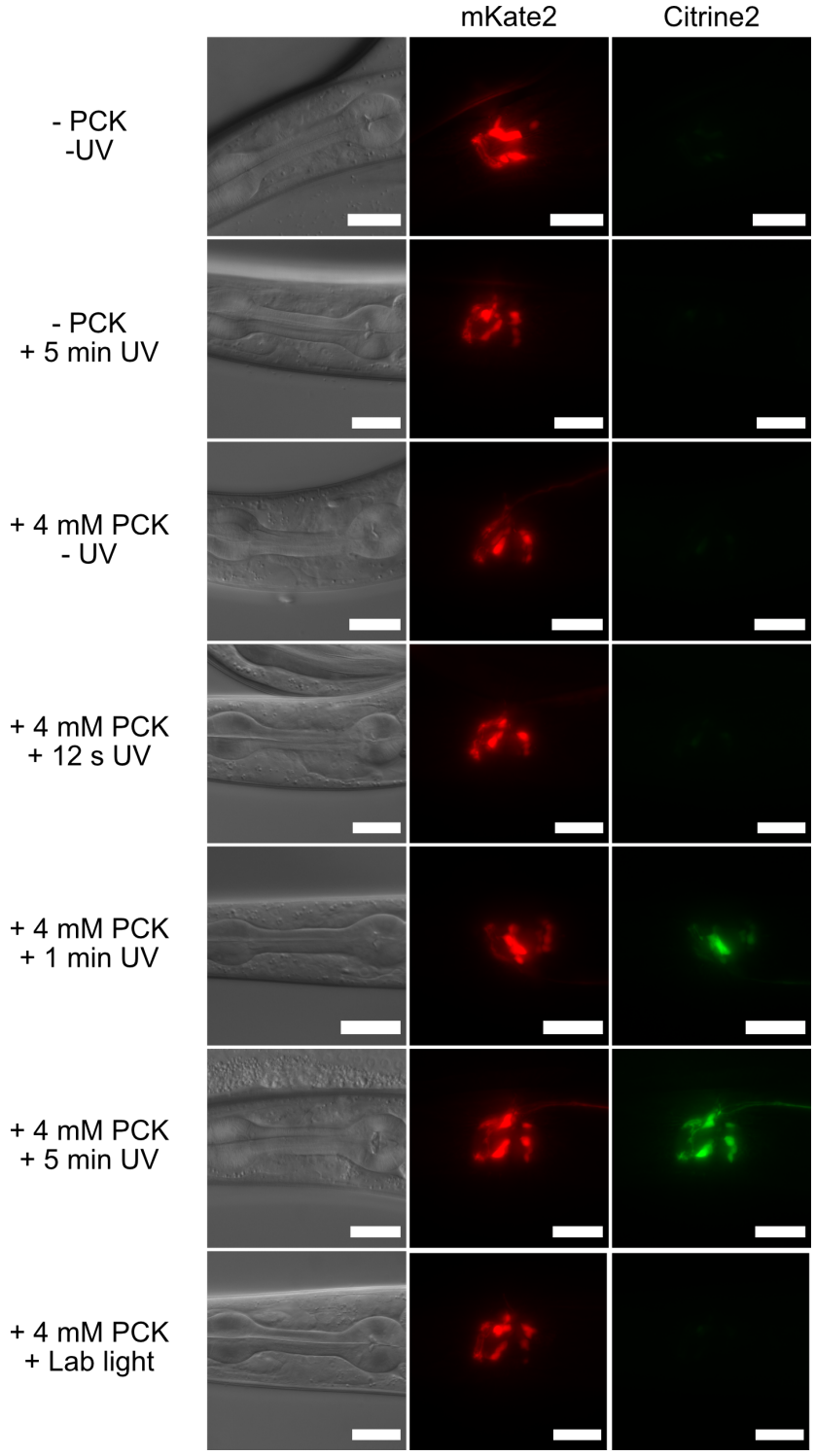

Figure 41: FLP-K223PCK With 4 mM PCK. Worms grown on 4 mM PCK and uncaged by illumination with UV for indicated time. DIC images are single slices, fluorescent images are maximum intensity projections. Scale bars are 25 μm.

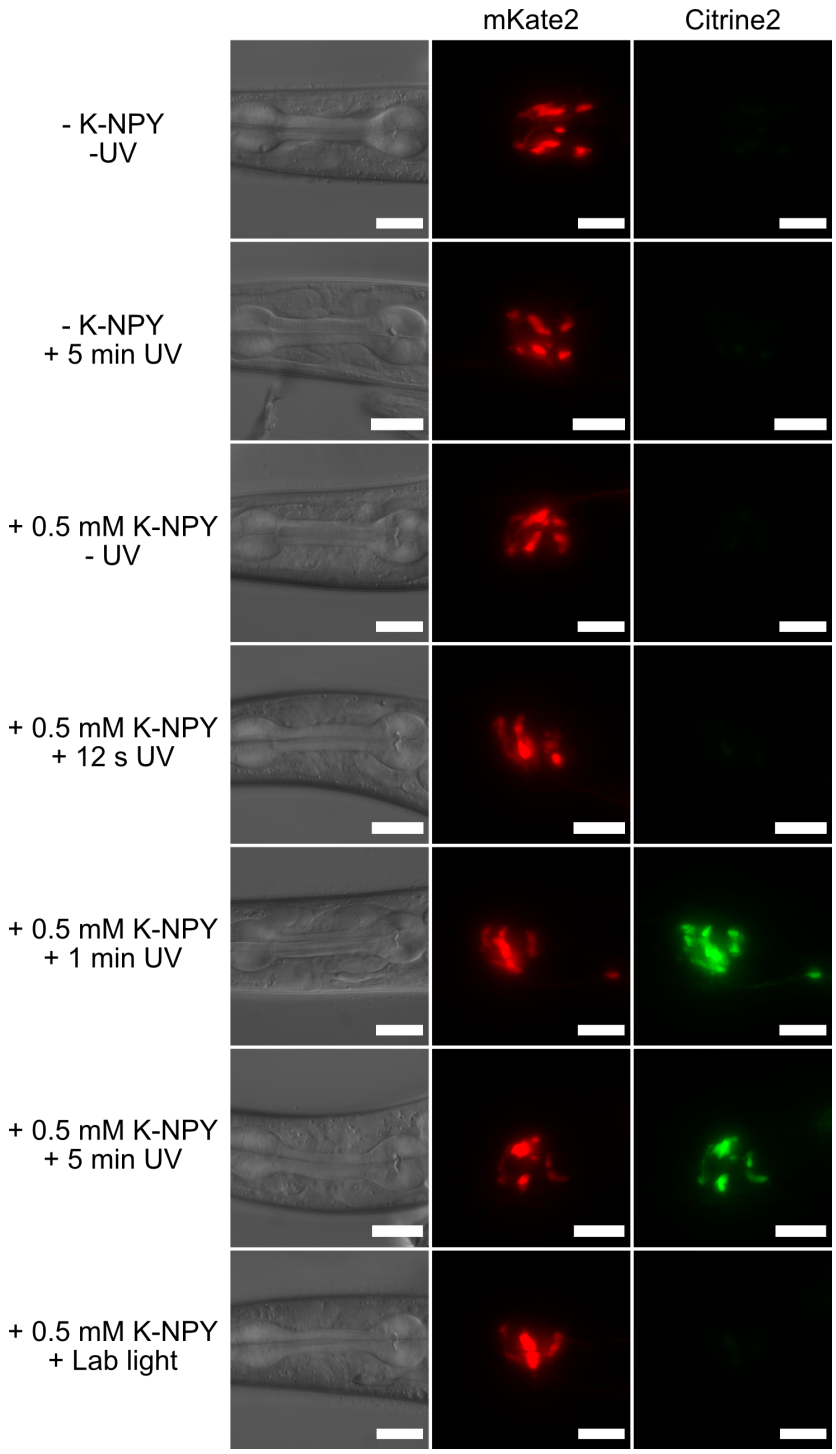

Figure 42: FLP-Y343NPY With 0.5 mM K-NPY. Worms grown on 0.5 mM K-NPY and uncaged by illumination with UV for indicated time. DIC images are single slices, fluorescent images are maximum intensity projections. Scale bars are 25 μm.

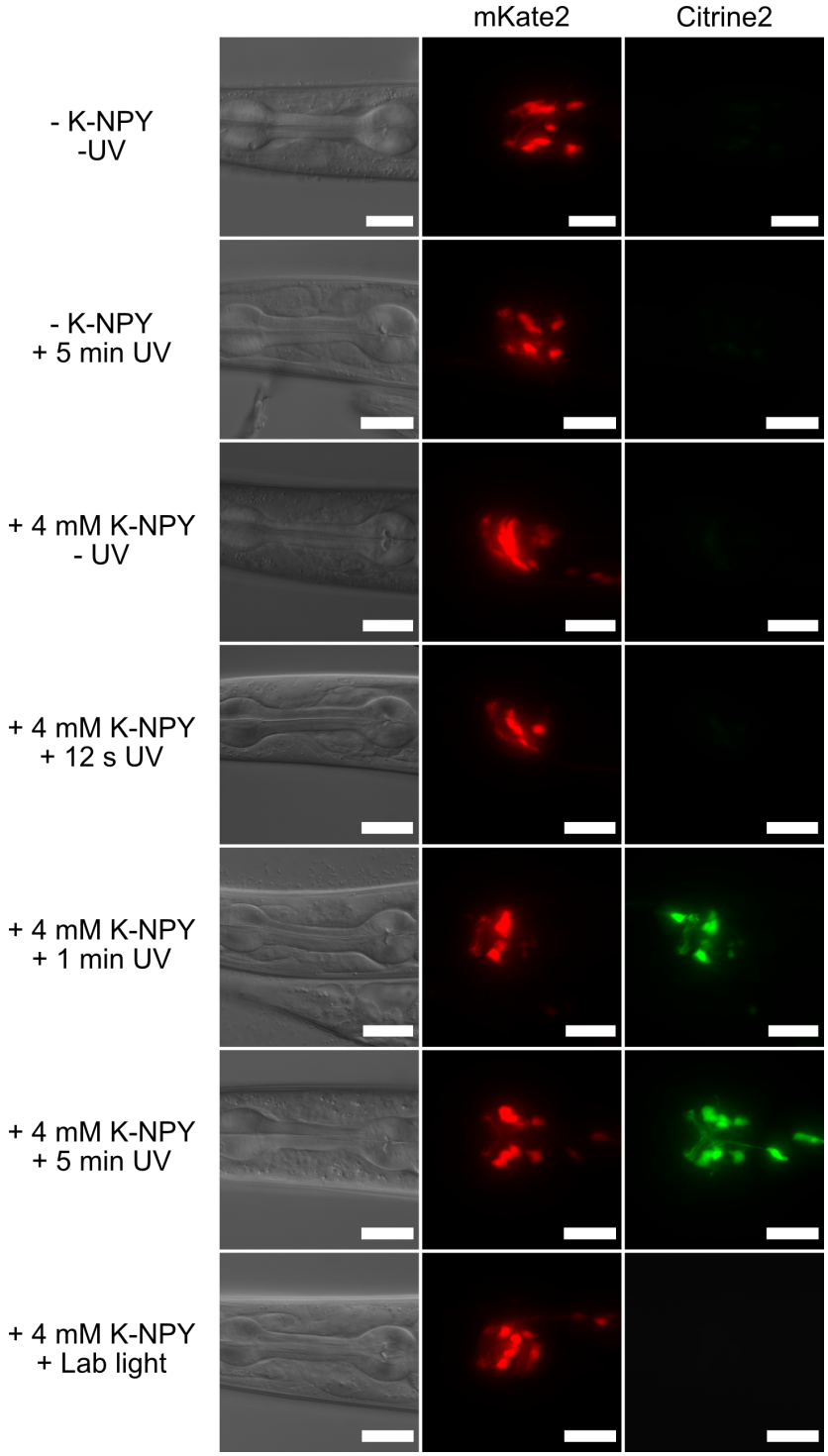

Figure 43: FLP-Y343NPY With 4 mM K-NPY. Worms grown on 4 mM K-NPY and uncaged by illumination with UV for indicated time. DIC images are single slices, fluorescent images are maximum intensity projections. Scale bars are 25 μm.

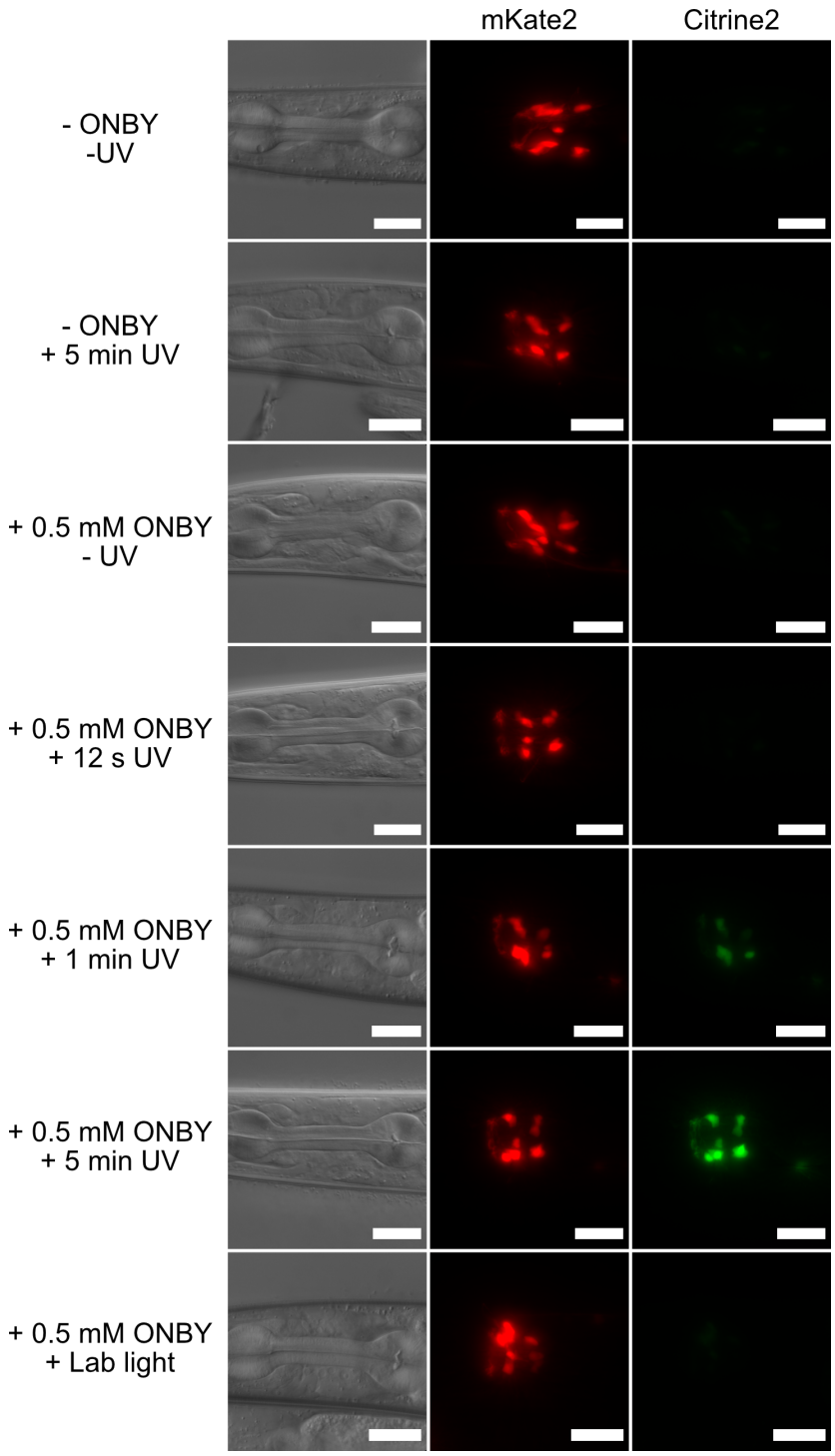

Figure 44: FLP-Y343ONBY With 0.5 mM ONBY. Worms grown on 0.5 mM ONBY and uncaged by illumination with UV for indicated time. DIC images are single slices, fluorescent images are maximum intensity projections. Scale bars are 25 μm.

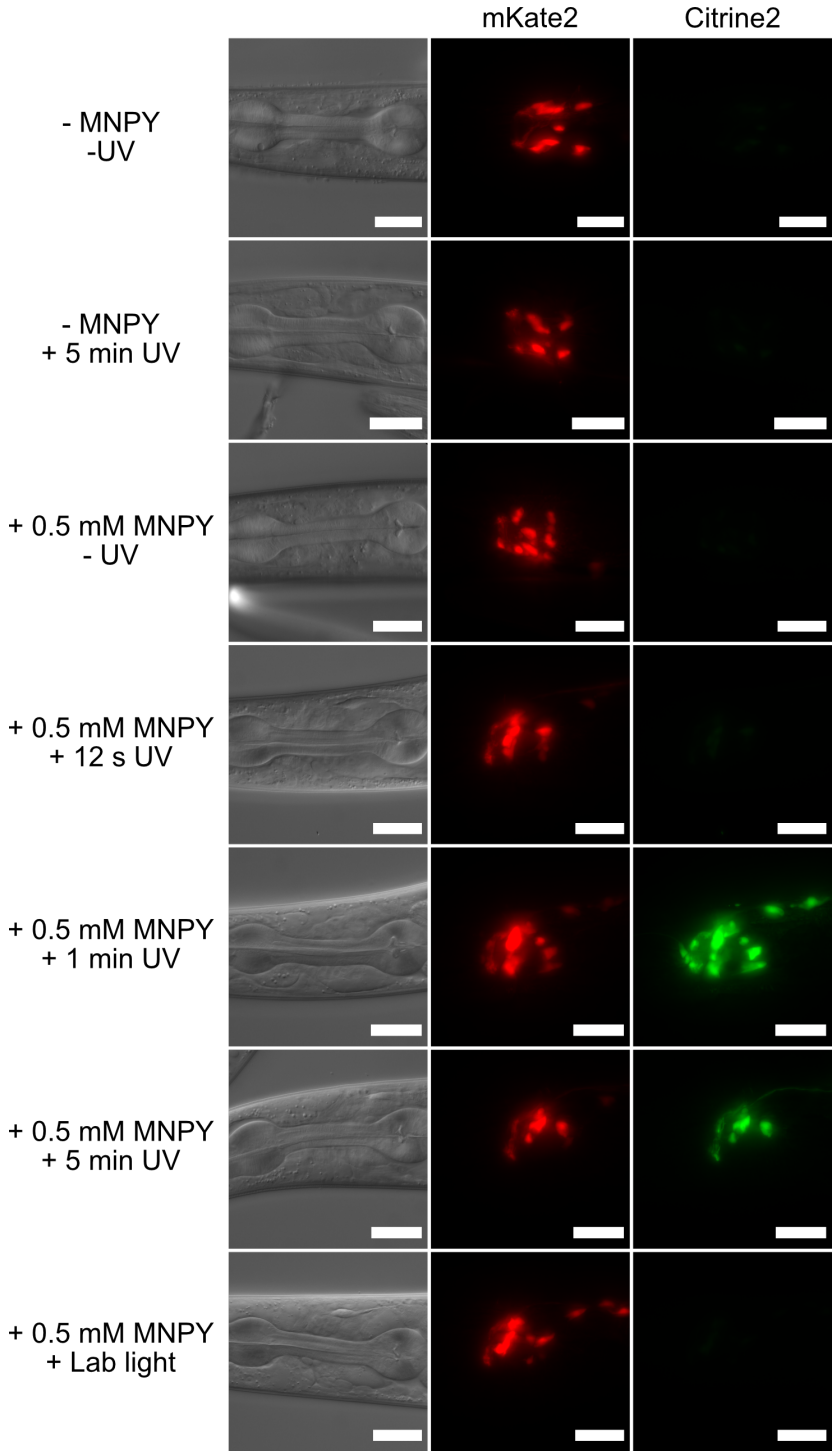

Figure 45: FLP-Y343MNPY With 0.5 mM MNPY. Worms grown on 0.5 mM MNPY and uncaged by illumination with UV for indicated time. DIC images are single slices, fluorescent images are maximum intensity projections. Scale bars are 25 μm.

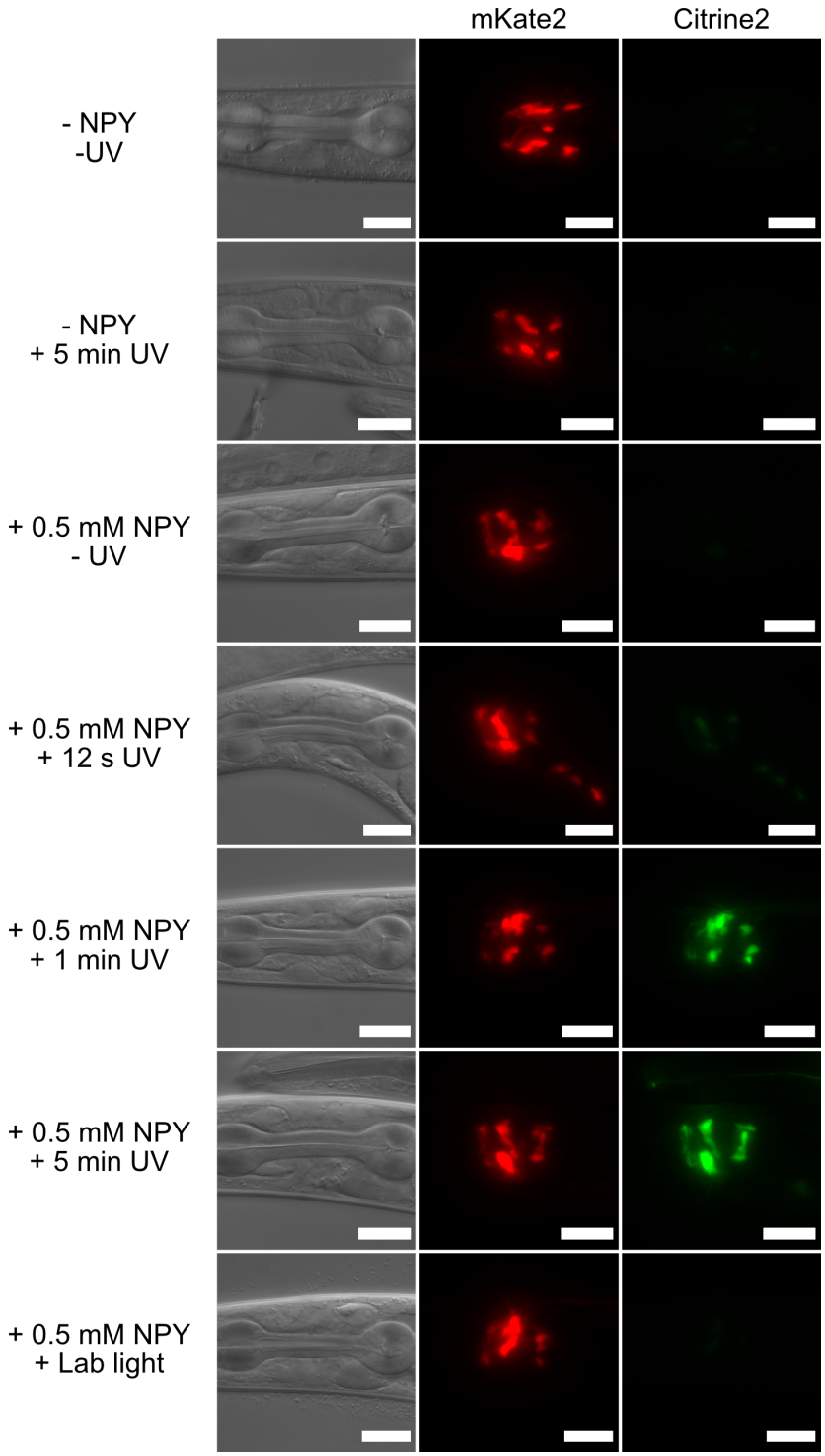

Figure 46: FLP-Y343NPY With 0.5 mM NPY. Worms grown on 0.5 mM NPY and uncaged by illumination with UV for indicated time. DIC images are single slices, fluorescent images are maximum intensity projections. Scale bars are 25 μm.

2.2.7 FLP-K223PCK Used to Study Neuronal Function

A particularly useful aspect of our photocaged FLP system is the ability to use a laser to activate FLP and drive gene expression in single cells (Fig. 20). For neuroscientists in particular, the expression of a protein such as a channelrhodopsin, which can optogenetically activate the neuron, in single neurons would significantly aid in the study of neuron functions. To facilitate this we use used the set-up as described in Section 2.2.2 to express a Chrimson channelrhodopsin, with a Citrine2 co-marker by expressing them together with an F2A self-cleaving peptide. This is placed downstream of an FRT-flanked STOP cassette, and driven by the *maco-1p* promoter, which has a consistently high expression across all life stages, with an increase at young adult, allowing us to perform behavioural analysis at this point in the life cycle. FLP-K223PCK expression was driven by the *glr-1p* promoter (Fig. 47A), which drove its expression in approximately 16 glutamatergic neurons[230]. GLR-1 is primarily expressed in the locomotory interneurons, including AVB and PVC neurons, which drive forward locomotion, and AVA, AVD, and AVE neurons, which drive backwards locomotion[231]. Ablating these neurons can lead to a loss of reversals in the worms, as well as uncoordinated movement when moving forwards[232]. Collectively, stimulation of these neurons drives a backwards movement[233], which provides an easy to score phenotype to test the activity of Chrimson as expressed by our FLP-K223PCK system. All of these interneurons are present as bilateral pairs[123], though are unable to be genetically dissected.

Animals expressing these constructs were synchronised by bleaching and grown on 4 mM PCK for 48 hours before being globally uncaged by exposure to 365 nm light for 5 minutes. After uncaging, worms were transferred to NGM plates supplemented with all ATR. After 24 hours, worms with visible Citrine2 signal were picked onto behaviour plates, and their behavioural responses to red light stimulus recorded (Fig. 47B). Worms that had an average speed in the first 2 seconds of stimulus at least 2 standard deviations above their speed in the 2 seconds before stimulus were considered to be responders (Fig. 47C). Under these criteria, approximately 70% of worms responded after feeding on PCK and uncaging, compared to approximately 5% or less of worms that had not been fed on PCK or uncaged. It should be noted however that these worms that have not had their FLP proteins activated may not be responding to the

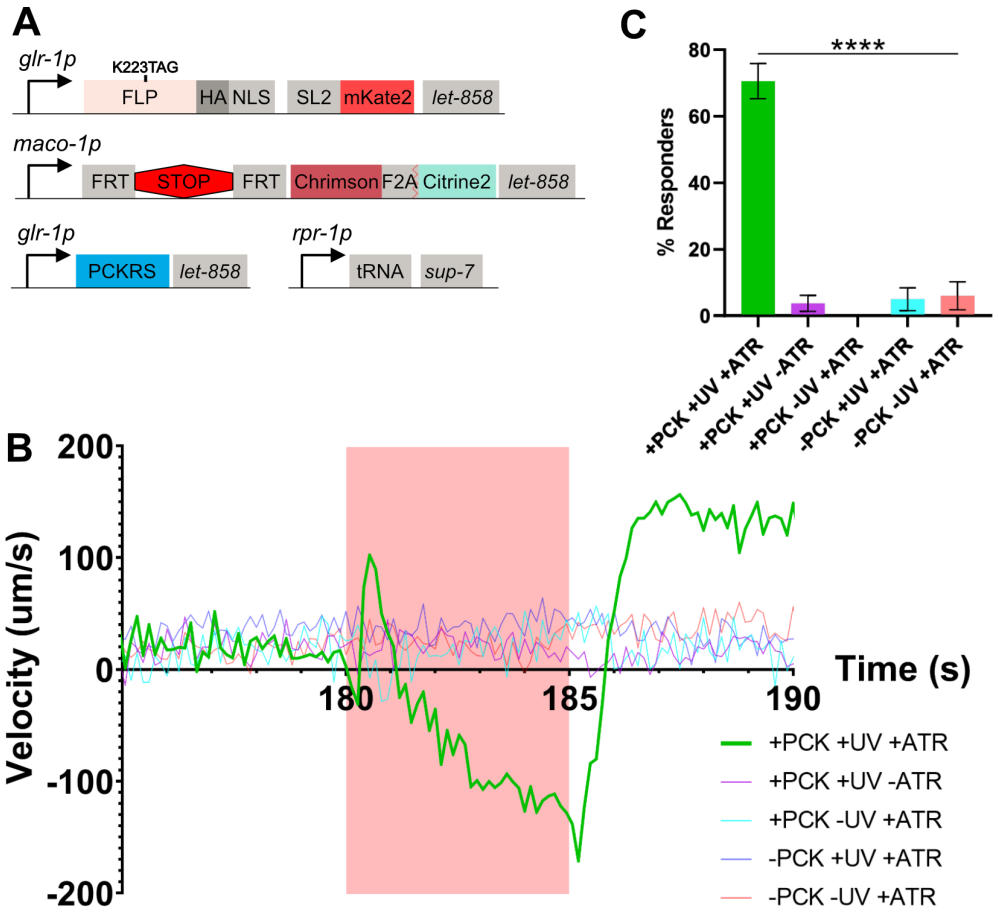

Figure 47: *glr-1p*::FLP-K223PCK Global Response to Stimulus. A Average velocity of worms during 5 s red light stimulus.**(B)** Percentage of worms responding in each condition. Responders are considered to be those worms whose average speed in the first 2 seconds of stimulus was at least 2 standard deviations higher than the average speed in the 2 seconds before stimulus. Mean ± SEM. **** = p< 0.0001, One-way ANOVA with post-hoc Dunnett's multiple comparisons to compare all conditions to +PCK +UV +ATR condition.

stimulus, but may for example be stationary for the majority of the time and a small movement coincidentally during the stimulus may lead to the worm being scored as a responder. We also find that individual worms show slightly different behaviours, with a small number of worms moving forwards for part or all of the stimulus time, though many of these then change direction and reverse for the remainder of the stimulus. This is likely due to mosaicism in the expression of the transgenes, and would likely be solved by integrating the array. Interestingly,

although there is no further stimulus, worms expressing Chrimson show a marked increase in velocity after the end of the 5 s stimulus compared to worms without active Chrimson. While we did not further test this, we speculate that this may be due to activation of the RID neuron, which forms chemical and electrical synapses with PVC and AVB, respectively, and has been found to sustain long periods of forward locomotion[234].

We next sought to use FLP-K223PCK to drive Chrimson in single neurons. We chose to study PVC, a neuron pair located in the tail which has so far never been genetically targeted due to a lack of PVC-specific promoters. Although one study has attempted to express an optogenetic channel in the PVC neurons using the overlapping promoter system with Cre recombinase[195], this proved unsuccessful as expression was found in a number of cells other than PVC. Another study managed to target PVC by illuminating the entire tail of the worm[233], however this does not have high enough spatial precision to target individually PVCL or PVCR, or individual neurons that are closely clustered with other neurons. As mentioned previously, PVC is an interneuron within the locomotory circuit which mediates forward movement. It is also involved in the posterior harsh touch response[235] and in developmental sleep, through its activation of the sleep-active neuron RIS[180]. In addition to making possible the expression of a transgene in the PVC neuron pair only, FLP-K223PCK could also be used to study the individual PVCL and PVCR neurons that make up the pair. This is particularly interesting in the sleep response, as it is known that only PVCL is presynaptic to RIS, while PVCR is not[123,180,236], indicating differing roles for the two PVC neurons. Additionally, only PVCL forms gap junctions with PLML, while only PVCR forms gap junctions with PLMR[236]. As we found differences in the PLM neurons during development and testing of the photocaged Cre system[77], this also points towards differing contributions of PVCL and PVCR in locomotion. Furthermore, both PVCL and PVCR form gap junctions with the locomotory interneurons AVAL and AVAR, which mediate backwards locomotion. However, while both PVCL and PVCR form two gap junctions with AVAL, PVCL forms six gap junctions with AVAR while PVCR forms 10 gap junctions with AVAR[236]. Finally, PVCR has been found to be presynaptic to both AVAL and AVAR, while PVCL is presynaptic to neither neuron[236], further indicating differing roles for PVCL and PVCR in the locomotory circuit.

To study the PVCs, we sent the strain expressing FLP-K223PCK from

the *glr-1p* promoter to the lab of Dr Henrik Bringmann at the Technical University Dresden. **All following data was generated by Dr Inka Busack, Bringmann Lab, Technical University Dresden.**

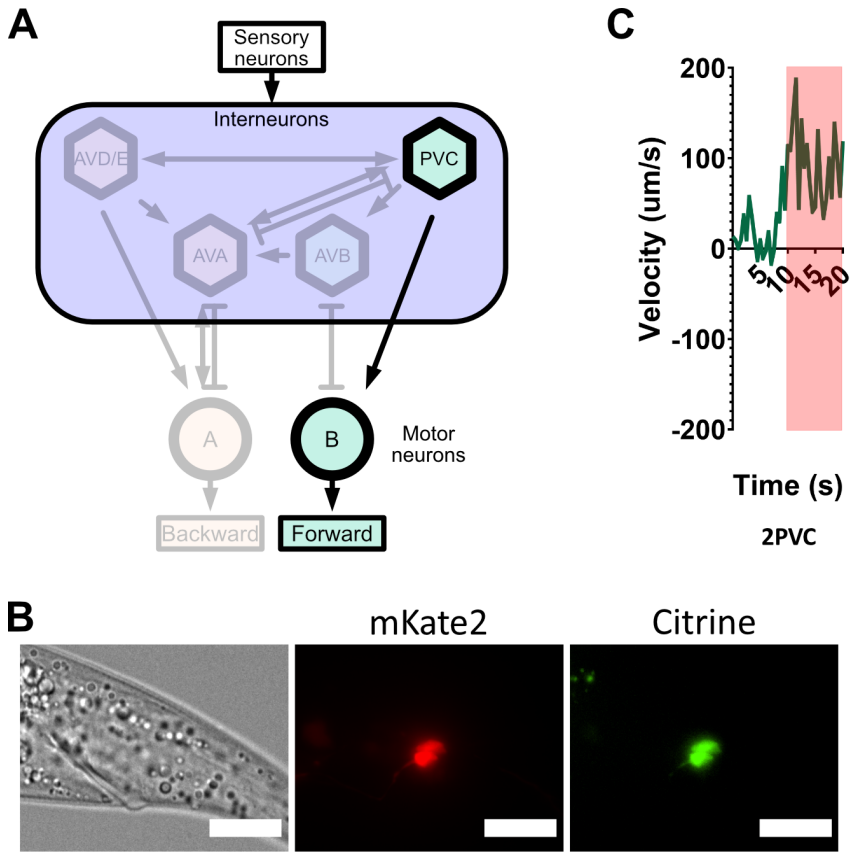

Figure 48: Activation of PVC Neurons. **A** Circuit diagram of the locomotory interneurons. Only PVC is activated, which drives forward locomotion. **B** PVC activation in the worm. Brightfield image is a single slice, fluorescent images are maximum intensity projections. Scale bars are 25 μm. **C** Average velocity of worms during red light stimulus. All data generated by Dr Inka Busack, Bringmann lab, TU Dresden.

These strains were fed on 4 mM PCK for 24 hours before being mounted on an agar pad on a microscope slide. The PVCs were identified and shot with a 355 nm laser to uncage FLP (Fig 48B). After recovering for 24 hours, the behaviour of the worms was analysed by stimulation with red light. As the ultimate goal for the Bringmann lab is to study the PVCs' role in sleep induction, a long (1 minute) red light stimulus was used. While this was normally done using the ReaChR channelrhodopsin[237,238], it seemed that Chrimson desensitised much

more quickly, with the response dropping during the 1 minute stimulus (data not shown). Regardless, as expected from only the PVCs being activated, we see a sharp forward acceleration upon stimulation of Chrimson (Fig 48C). Thus, our tool makes it possible to genetically target neurons which have not previously been targetable. Work is now ongoing to determine the optimum uncaging conditions to target PVCL or PVCR independently.

2.3 Discussion

2.3.1 Photocaged FLP Recombinase

Site-specific recombinases are powerful tools for controlling gene expression *in vivo*, but lack innate spatiotemporal control mechanisms. Although several methods have been developed to introduce spatiotemporal control[181,184,185,239], none have sufficiently reached the level of spatial control required to target individual cells without high levels of off-target expression while also employing a method for non-damaging temporal control. The development of a photocaged Cre[187,188] was one of the first methods that introduced this high level of spatiotemporal control. The use of photocaged Cre in higher organisms, including *C. elegans*[77], provided the first way to reliably drive gene expression in single cells that could not otherwise be genetically targeted, including individual neurons within a bilateral pair. Here, we have expanded this toolkit to include a photocaged FLP recombinase, which we foresee being particularly useful in neuroscience research in *C. elegans*, which is often hampered by a lack of cell-specific promoters. We show this by using photocaged FLP to drive expression of a channelrhodopsin in the PVC neurons, which do not have their own cell-specific promoter and thus cannot be genetically targeted. Although initially encountering issues with expression and visualisation of the channelrhodopsin with a fused fluorescent reporter, we find that including a self-cleaving peptide to separate the channelrhodopsin from the fluorescent reporter vastly improves the visible signal seen, making this tool more broadly accessible for use in other labs.

Photocaged FLP was generated by replacing a catalytic lysine residue, K223, with a photocaged lysine, PCK. The K223 residue acts as the general acid during the recombination reaction carried out by FLP and, like its counterpart K201 in Cre recombinase, we find that photocaging the residue completely abolishes FLP

activity. Uncaging the residue is a very quick reaction, requiring short exposures to 365 nm light and uncaging within seconds or minutes, depending on the source of light used, which restores the catalytic lysine residue and leaves a functional FLP recombinase. If a STOP cassette is used to prevent gene expression, this cassette can be excised in a relatively short time to allow for gene expression to begin, though the rate of gene expression itself would depend on the transgene being expressed, the promoter, and the age of the animal.

We go on to characterise FLP-K223PCK and determine the minimal conditions required to use the system. Here, we find that as little as 8 hours of feeding on PCK may be enough to generate enough photocaged FLP to drive target gene expression after activation in almost over 90% of animals, with little difference after feeding for a full 24 hours. Although further characterisation would need to be done to determine if all cells can be targeted with such short a time feeding on PCK, this result is promising as it facilitates the study of animals at points very early in their life cycles. As little difference is seen between 8 hours and 24 hours of feeding on PCK, it may even be possible to further reduce this time to target even younger animals. Alternatively, unpublished observations have shown that a gravid adult feeding on PCK can supply the egg with the PCK, which would even allow us to drive gene expression in the embryo. We have also determined the minimum PCK concentration required for uncaging, as the high cost of PCK will often be the limiting factor in performing experiments with the FLP-K223PCK system, and a low concentration would allow a large number of experiments to be performed for relatively little cost. Here, although we commonly use 4 mM PCK for our experiments, we find that 1 mM PCK results in similar, or even slightly better, levels of activity. Furthermore, although not as well characterised, we find that using 0.5 mM PCK can also lead to sufficient FLP activity after uncaging, which would drive costs down even more and make the tool more accessible.

The dynamics of FLP-K223PCK activity after uncaging closely follow those found in *in vitro* studies of wild-type FLP recombinase[191], and provide one of the first looks at the dynamics of FLP-mediated recombination *in vivo*. Although *in vivo* studies have been done previously[191], these relied on slow induction methods, which took several hours to express FLP in sufficient quantities, limiting the analysis of the activity of FLP itself. With our system, FLP-K223PCK is produced in a primed state, with a brief exposure to 365 nm light causing uncaging

and activation of FLP within seconds, which provides very accurate insight into the dynamics of FLP-mediated recombination.

Furthermore, we have found that we can also photocage the catalytic tyrosine of FLP, the key residue which characterises all Tyrosine-SSRs. While we did not characterise photocaged FLP-Y343 strains further, they have the potential to make the use of this system even more accessible, with ONBY being cheap and commercially available. Other photocaged tyrosine variants, MNPY and NPY, showed even greater uncaging dynamics than ONBY, with FLP becoming active after much shorter exposure times to 365 nm light, decreasing the risk of damaging cells when uncaging. However, neither MNPY nor NPY are currently commercially available and are required to be custom synthesised, which can be prohibitively expensive, though this could change as photocaged amino acids become more mainstream and more photocaged amino acids become commercially available, and synthesis routes are improved.

For single-cell gene expression, photocaged FLP shows a lot of promise. While the work is ongoing, we have shown that we can activate gene expression in the PVC neurons which do not have their own cell-specific promoter, allowing us to study the role of PVC. Furthermore, the PVC neuron exists as a bilateral neuron pair consisting of PVCL and PVCR, and work is now ongoing to determine the optimum targeting conditions to drive expression in each of these individually using the FLP-K223PCK, which would allow us to study the individual contributions of each neuron to the worm's behaviour for the first time. Two-thirds of *C. elegans* neurons exist as bilateral pairs and the functions of the vast majority of these pairs have never before been dissected. However, the individual neurons within each pair often make different connections within the *C. elegans* connectome, which would imply differing contributions of each individual neuron to the overall functions of the pair. We showed this when we developed photocaged Cre[77], where we investigated the neuron pair PLM and found that PLML and PLMR make slightly different contributions to the tail touch response, with PLMR activation showing a much stronger behavioural response than PLML activation. Through a collaboration with Dr Henrik Bringmann, we are now studying the individual roles of PVCL and PVCR in both locomotion and developmental sleep in *C. elegans*.

We have focused primarily on using FLP-K223PCK to express transgenes in neurons. However, we have also briefly shown that the tool can be applied to

other tissues, including body wall muscle and gut tissues, though in some cases this requires further optimisation to improve the efficiency of the system. Further, strains could be generated with FRT flanked endogenous genes, which would allow our system to be used to switch off endogenous gene expression in specific cells. Finally, with both photocaged FLP and photocaged Cre now developed, it could be possible to combine both systems in a single animal to provide even greater transgene control. For example, each photocaged SSR could be used to drive two different optogenetic channels in two different neurons, allowing for the activation of each neuron independently in a single animal. Alternatively, one SSR could be used to express a neuron-activating channelrhodopsin, while the other is used in a different neuron to express a neuron-inhibiting halorhodopsin.

The application of photocaged amino acids in Cre and FLP may be more broadly applicable to other recombinases, especially those of the Tyrosine-SSR class. As all recombinases in this class are defined by a catalytic tyrosine, the photocaging of this residue is likely to have the same effect as the photocaging of the Cre[187] and FLP catalytic tyrosines. With only a few exceptions, the majority of Tyrosine-SSRs also use a lysine residue as the general acid[240], which we have also found to be a good candidate for photocaging in both Cre[77] and FLP. Serine-SSRs may take a bit more work to photocage, as they share neither the nucleophilic tyrosine of the Tyrosine-SSRs, nor do they tend to use lysine as the general acid in the reaction. Instead, Serine-SSRs utilise serine as the nucleophile for attacking DNA, and often use arginine as a general acid[241]. Although no photocaged arginine has been reported in the literature so far, a photocaged serine has been reported[242] which could potentially be used to photocage recombinases from the Serine-SSR class, further expanding the toolkit available to researchers.

2.3.2 SL2 and Operons

In Section 2.2.2, we attempted to place two transgenes in an operon behind a transcriptional terminator in order to co-express them after FLP-based excision of the terminator sequence. However, we found that expression of the second transgene bypassed the transcriptional terminator and became expressed in a ubiquitous pattern, despite the upstream promoter expressing only in neurons. We have explored this further to determine the full extent of this abnormal expression. We find that placing mKate2::SL2::Citrine2 downstream of the

transcriptional terminator and expressing this in worms in the presence of wild-type FLP recombinase to excise the transcriptional terminator resulted in the expected expression pattern of both transgenes. In the absence of FLP, we would expect not to see either transgene expressed due to the transcriptional terminator. However, while we see no expression of the mKate2 protein, we see ubiquitous expression of Citrine2. No ubiquitous promoter was present on our transgene constructs, with all promoters used being the glutamatergic neuron-specific *glr-1p* promoter. The reasons for the ubiquitous expression of Citrine2 despite the transcriptional terminator are unknown.

One possibility could be cryptic promoters, either in the transcriptional terminator sequence or the SL2 acceptor site. These promoters may be weak enough that expression from the actual promoter on the construct may override expression from the cryptic promoter, with the transcriptional terminator downstream from the actual promoter biasing expression towards the cryptic promoter. We reason that the cryptic promoter is unlikely to be in the transcriptional terminator, however, as if this were the case we may expect to see expression of mKate2 as well as Citrine2, and in Section 2.2.2 we would expect no difference in Citrine2 expression patterns between Chrimson::Citrine2, Chrimson::HL::Citrine2, Chrimson::SL2::Citrine2, and Chrimson::F2A::Citrine2, as the cryptic promoter would precede all of these. However, a role for the transcriptional terminator cannot be ruled out and it may be useful to test other terminator sequences in this setup to determine if it is only the *lacZ* 3'UTR that causes abnormal expression.

Therefore, if a cryptic promoter exists, it is most likely in the SL2 acceptor site. Indeed, promoters have been found within operons, which allows for differential expression of the genes within an operon[220]. From the operon we use to co-express transgenes, it has been noted that expression of the downstream genes, *gpd-2* and *gpd-3*, is much higher than that of the upstream gene *mai-1*[243], contrary to the typical expression pattern of operons where the downstream genes are expressed at much lower levels than the upstream gene[225], which indicates that an internal promoter is likely involved. However, the same study noted that this internal promoter is likely between *mai-1* and *gpd-2*, as the *gpd-2/3* intergenic region alone does not allow for expression of a transgene, while the intergenic region between *mai-1* and *gpd-2* could drive gene expression alone. Since the intergenic region we use is that from between *gpd-2* and *gpd-3*, the idea of an internal

promoter being present is unlikely, though it is still a possibility that one could exist but is too weak to drive expression on its own. However, if an internal promoter were to be present, we would expect to see some expression from this promoter wherever we used this SL2 acceptor site. Since we use this on almost all of our genetic constructs, but only see the ubiquitous expression when placed behind the transcriptional terminator, it is unlikely to be the case that an internal promoter is causing the abnormal expression. It may be interesting to replace our SL2 acceptor site with the SL2 acceptor sites from other operons, both those with known internal promoters and those without, to determine if the effect is specific to the *gpd-2/3* intergenic region.

One interesting observation is seen when expressing FLP to remove the transcriptional terminator. Without FLP, we see Citrine2 expression ubiquitously. However, when FLP is present to excise the transcriptional terminator, expression of mKate2 and Citrine2 is then restricted to only the cells that are expressing FLP. Although this expression pattern is entirely expected and would otherwise not be noteworthy, it raises interesting questions when considered alongside the observation that, without FLP excision of the terminator, expression of Citrine2 is seen ubiquitously, while expressing FLP in only a subset of cells changes this ubiquitous expression to a localisation expression. Why then does expression of Citrine2 in all of the other cells change, when FLP is not expressed in these cells and the transcriptional terminator is not being excised? This points towards an epigenetic regulation of the gene expression, with the epigenetic effects carried through from the earliest cell divisions to the adult worm. One interesting experiment to test this would be to use photocaged FLP recombinase to excise the transcriptional terminator in a subset of cells only in the adult worm. If the ubiquitous expression of Citrine2 is really due to an epigenetic factor carried through from an early life stage, we would expect excision of the transcriptional terminator in adult worms not to have an effect on Citrine2 expression in cells where this excision does not occur.

It should be possible to cross our ubiquitous mKate2::SL2::Citrine2 strains with strains that have had certain chromatin-modifying proteins mutated, and this may result in a normal expression of Citrine2 if epigenetic factors are involved in the abnormal expression pattern. A scenario where downregulation of chromatin-modifying proteins can cause abnormal expression patterns has already been shown in the case of multivulva mutant worms[244]. During normal

development of the vulva, the RTK/Ras/Map signalling pathway promotes vulval cell specification, while synMuvs, a set of genes which recruit and activate chromatin-modifying genes, represses vulval cell specification. In multivulva worms, mutations in the synMuv genes or the other chromatin-modifying genes cause misexpression of germline-specific genes in somatic cells, leading to the multivulva phenotype[245] and indicating a role for epigenetics in the expression pattern of vulval fate-determining genes. Epigenetic modifications have been found to be involved in determining and maintaining cell fate in all *C. elegans* tissues, and likely lead to differential expression of genes in these cells[246]. In a similar manner, if the SL2 acceptor site is in some way interfering with chromatin-modifying proteins during development, it could lead to abnormal expression patterns of the transgene which could explain the ubiquitous expression we see in our strains. By crossing these with strains defective for chromatin modulation, we may see expression of our transgene become restricted to the pattern expected from the promoters we use.

Chapter 3

Tools for *In Vivo* Gene-Targeted Random Mutagenesis

3.1 Introduction

3.1.1 Mutagenesis and Genetic Screening

The introduction of mutations into an organism's genome is an important process for both forward genetics, where genes involved in particular biological processes can be identified, as well as reverse genetics, where the specific functions of genes can be studied[247]. With its short generation time and hermaphroditic lifecycle, *C. elegans* is particularly well-suited to genetic screens and as such has been used for decades, with a wide range of methods developed for both forward and reverse genetic screens.

Forward genetic screening is often the first step in identifying and studying a biological pathway and typically involves genome-wide mutagenesis, followed by screening for a desired phenotype and identification of the mutations. In *C. elegans*, the mutagenesis step is typically done either by chemical mutagenesis[116,248,249], radiation[250,251], or random insertion of transposons[252]. While differing in the specific types of mutations generated, these techniques allow for the generation of mutations across the entire genome, which facilitates identifying previously unknown genes, regardless of their position in the genome. By screening for animals with altered phenotypes after the mutagenesis steps, such as defective behavioural patterns, altered susceptibility to drugs, or abnormal development and lifespan, researchers can then use genetic mapping or whole-genome sequencing to identify the causative mutations and therefore the genes involved in the biological pathways involved in the observed phenotypes.

Once a forward genetic screen has identified a particular gene of interest (GOI), reverse genetics techniques can then be used as a more focused approach

to further study the functions of that gene. While both forward and reverse genetics involved the generation of mutations, reverse genetics involved generating mutations only in the GOI. Therefore, the specific techniques used differs between forward and reverse genetic screens. The mutagenesis step of reverse genetic screens falls into two main categories, target-selected mutagenesis and gene-targeted mutagenesis, each with their own advantages and disadvantages.

Target-selected mutagenesis utilises random genome-wide mutagenesis similar to forward genetic screens, though with the additional step of screening for animals where the mutations occurred in the GOI[247]. This screening is most commonly done by Polymerase Chain Reaction (PCR), whereby an insertion or deletion (indel) of a large region in the GOI can be identified[253]. Animals that give a positive hit by PCR can then be isolated and the effect of the indel can then be studied. While relatively simple in its approach, this technique is fairly limited as it can only reliably detect large indels, but is unsuited to detecting single-nucleotide polymorphisms (SNP) or even small indels that would result in a change in band size that is too small to detect by PCR.

Some studies have used whole genome sequencing to detect SNPs in libraries of mutant strains. For example, the Million Mutant Project was a project that generated and fully sequenced 2,000 mutagenised *C. elegans* strains with over 800,000 total SNPs representing changes in all of the approximately 20,000 genes[254], which could be used to isolate strains with SNPs in a GOI. However, although this represents a colossal library to study, each gene is only represented by an average of eight non-synonymous SNPs[254], which may not be enough coverage to fully study a gene's function. Additionally, the generation of this library is a result of a collaboration between multiple labs and is both labour-intensive and costly to reproduce, especially for individual labs that may only desire to study mutations in a single GOI. Even if SNPs were easily detectable, whole-genome mutagenesis is very inefficient for single-gene studies due to the low likelihood of mutations being generated in the GOI. To put it into perspective, the size of the *C. elegans* genome is approximately 100 Mb (100,000,000 bases)[120], whereas the median gene size is less than 2,000 bases[255]. Assuming an equal distribution of mutations across the entire genome, this gives only a 0.002% chance of each individual mutation hitting a typical GOI. In a typical mutagenesis study, 2-300 mutations per genome are generated, giving a 5% chance of a single study generating mutations in a single typical GOI. Some genes, as well as

other genomic regions such as those encoding RNAs, are even smaller, with an even smaller chance of being hit during whole-genome mutagenesis. Additionally, this approach requires extensive backcrossing to remove mutations occurring in other parts of the genome, as otherwise these would interfere with analysis of the mutant.

An alternative approach for reverse genetics studies is to use gene-targeted mutagenesis, which gives rise to mutations only in the GOI[247]. This is useful to researchers interested in particular genes or pathways, as there is no requirement to sift through a large number of irrelevant mutant strains. These techniques primarily include the functionally-similar Zinc-Finger Nuclease (ZFN) and Transcription Activator-Like Effector Nuclease (TALEN) systems[256], as well as the Clustered Regularly Interspersed Palindromic Repeat (CRISPR) systems[257,258]. All three of these systems can be targeted to specific DNA sequences, where they cleave the DNA. The DNA is then repaired by the cell's endogenous non-homologous end joining pathway, which is fairly error-prone and can introduce small indels at the repair site[247]. Alternatively, homology-directed repair can be promoted by providing a template for repair, consisting of a DNA fragment containing a desired mutation flanked by short DNA sequences with homology to the cut site[259,260]. This allows the introduction of very specific mutations, including SNPs or the insertion or deletion of several base pairs at the target site.

While a range of methods have been developed for mutagenesis in *C. elegans*, limitations still remain. As can be seen, most methods in use can either randomly mutagenise a whole genome, or they can introduce specific mutations at specific sites in a target gene. However, it is not currently possible to introduce random mutations at only a specific GOI, without also affecting the rest of the genome and without having to decide which specific mutations to generate. Therefore, one additional category of tools that would be useful to develop for genetic screening would be those that could perform gene-targeted random mutagenesis. This would be useful for researchers studying a GOI that is very poorly characterised, as they would not have to decide on specific, known mutations as they would be using the described gene-targeted approaches, but can instead utilise the randomness of the mutagenesis to determine which specific amino acids or regions of the gene are important for its function. Additionally, compared to genome-wide mutagenesis methods, a gene-targeted random mutagenesis method would

be much more efficient, as far fewer animals would have to be screened since all, or nearly all, mutations would be in the GOI.

Furthermore, gene-targeted random mutagenesis tools could be used for directed evolution in *C. elegans*. In particular, these tools could be used to improve the efficiency of genetic code expansion in *C. elegans*. As discussed in Chapter 1.2.3, a large number of improvements have been made to the incorporation efficiency of ncAAs by generating mutations in both the tRNA and the synthetase sequences. Additionally, mutagenesis is required to generate synthetases that can incorporate novel ncAAs. However, the mutagenesis and screening processes involved are generally done in *E. coli*, before being effectively transplanted into *C. elegans*. While these developments have so far proven effective in increasing incorporation efficiency in *C. elegans*, there may be lost potential due to the differing translation mechanisms between the prokaryotic *E. coli* and eukaryotic *C. elegans*. Instead, performing the mutagenesis and screening directly in *C. elegans* may prove to yield new tRNA and synthetase variants with much higher incorporation efficiencies than current methods.

To develop a gene-targeted random mutagenesis tool, I attempted to adapt three other tools which have been developed in other systems for similar purposes. The first of these tools is a Histone-bound MiniSOG[261], which has been used for whole-genome mutagenesis in *C. elegans*. I planned to further develop this tool for site-specific mutagenesis by using a Cas9 protein, which could target it to a desired genetic sequence. The next tool is named EvolvR[262], which uses Cas9 to guide an error-prone polymerase and has been used successfully in bacteria for continuous mutagenesis at a specific target site. Finally, the CRISPR-X[263] system uses Cas9 to guide a cytidine deaminase, which has been used for directed evolution in mammalian cells.

3.1.2 His-MiniSOG

The mini Singlet Oxygen Generator (MiniSOG) is derived from Phototropin 2, a blue light photoreceptor from Arabidopsis, and was developed for use as a tag for light and electron microscopy[264]. This photoreceptor contains a Light-Oxygen-Voltage-sensing (LOV) domain, which under non-stimulated conditions is non-covalently bound to a flavin mononucleotide (FMN) cofactor. Upon absorption of photons, the FMN enters an excited state and subsequently forms a covalent

bond with a cysteine residue of the LOV domain. This causes a conformational change in the protein, leading to activation of the photoreceptor kinase activity and initiation of signalling cascades[265]. In MiniSOG, this cysteine has been mutated to a glycine, which prevents the covalent bond from being formed and results in the energy from the excited FMN being transferred to ground state oxygen molecules ($^3O_2$) to generate singlet oxygen ($^1O_2$)[264]. This singlet oxygen is highly reactive and can be used to oxidise diaminobenzidine to cause it to form high-contrast precipitates, which are more easily imaged by electron microscopy.

The high reactivity of $^1O_2$ has also been found to cause irreversible damage to biomolecules including proteins and DNA, where it can cause strand breaks and convert guanine into 8-oxo-7,8-dihydroguanine, which primarily results in G:C to T:A transversions[266] when resolved. This damaging feature of $^1O_2$ has led to MiniSOG being used for cell ablation studies in *C. elegans*. By targeting MiniSOG to the mitochondria[267] or the cell membrane[268], a high concentration of singlet oxygen can be generated by directing blue light at the cell, which leads to cell degeneration and eventually death.

The DNA-damaging effect of $^1O_2$ has led to MiniSOG being used as a tool for mutagenesis. By attaching MiniSOG to a histone protein, HIS-72, to bring it in close proximity to genomic DNA, Noma *et al* found that the singlet oxygen species produced by light-activated MiniSOG could induce heritable mutations in *C. elegans*[261]. This approach was found to induce deletions, rearrangements, and SNPs, which differed from those produced by chemical mutagenesis. For example, ethyl methanesulfonate (EMS), one of the most common chemical mutagens, is heavily biased towards G:C to A:T transitions[269]. Because of this bias, some mutations are very unlikely to occur in EMS-based screens. For example, alanine is encoded by GCN, where N denotes any base, and therefore it is unlikely that EMS would cause another codon to be mutated to give rise to alanine. Other mutations are highly likely to occur with EMS, such as nonsense mutations resulting in AT-rich stop codons[248], which are useful for forward genetic screens but not very useful when studying a GOI's functions. Another common mutagen, N-ethyl-N-nitrosourea (ENU), is less biased than EMS, but still biases towards G:C to A:T and A:T to G:C transitions, with relatively little transversions occurring[270]. His-MiniSOG, on the other hand, was found to induce almost no G:C to A:T transitions. Instead, His-MiniSOG was found to bias more towards G:C to T:A transversions, as expected from its use of singlet oxygen species,

as well as G:C to C:G transversions[261]. This gives rise to a range of possible mutations not found using EMS or ENU, and so provides a good complement to chemical mutagen-based mutagenesis. Using alanine again as an example, the codon GGN, which encodes glycine, can be mutated to alanine, GCN, by the His-MiniSOG, though it is very unlikely to occur using EMS or ENU.

As a development on His-MiniSOG mutagenesis, we could instead target MiniSOG to specific GOIs, or other regions of DNA, by use of CRISPR-associated Protein 9 (Cas9). Cas9 is a nuclease which functions as an adaptive immune system in bacteria, integrating guide RNAs (gRNA) derived from viral genetic material during exposure to infection and using these as a guide during future infections to target the nuclease to invading viral particles, cleaving their DNA or RNA and neutralising them[271]. The ability for a researcher to provide custom gRNA to the system has led to Cas9 being rapidly adopted as a tool for genome editing, with thousands of papers published every year using the tool[272]. One development of Cas9 is dead Cas9 (dCas9), which retains its DNA-binding capability, but has had its nuclease activity removed by mutating the catalytic residues in its two nuclease domains[273]. By fusing MiniSOG to this dCas9, it should be possible to direct MiniSOG only to a desired region of DNA, including to a GOI. There, it can be activated by blue light to generate a local population of reactive single oxygen, leading to mutagenesis only at that locus (Fig 49).

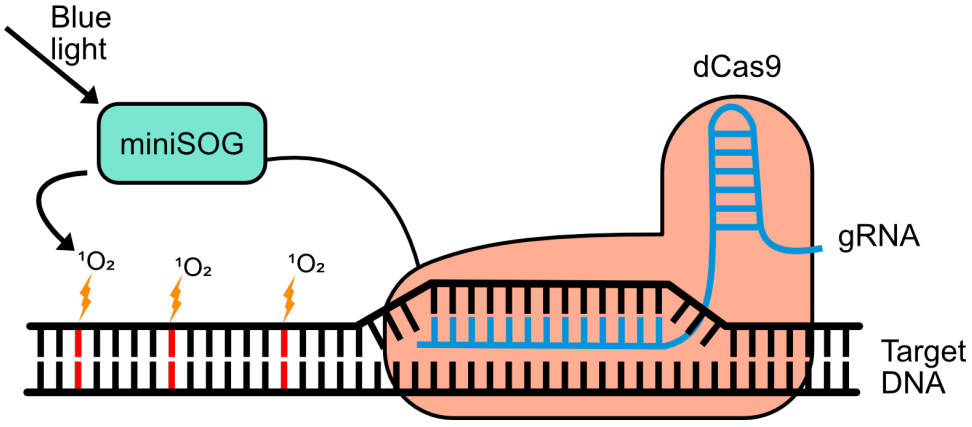

Figure 49: dCas9-MiniSOG. Dead Cas9 (dCas9) can be used to target MiniSOG to a specified DNA (black) locus using a guide RNA (gRNA, blue). Upon stimulation with blue light, MiniSOG generates a local population of reactive singlet oxygen species ($^1O_2$). $^1O_2$ attacks guanine residues in the target DNA to mutagenise them to a new base (red).

3.1.3 EvolvR

The second tool I aimed to adapt to *C. elegans* is EvolvR, a Cas9-guided error-prone DNA polymerase, which was developed and used successfully as a gene-targeted random mutagenesis tool in *E. coli*[262]. This method depends on DNA Polymerase I (PolI), a *E. coli* polymerase which plays a role in DNA repair and lagging-strand replication of chromosomal DNA by initiating replication at sites with a single-stranded break, or nick[274]. A highly error-prone version of PolI has been generated by including the D424A, I709N, and A759R point mutations (PolI3M). The I709N mutation is thought to increase the misincorporation of nucleotides occurring in the substrate-binding pocket of the polymerase[275], while D424A results in a loss of exonuclease proofreading activity[276] and A759R lowers the fidelity of PolI[277]. Together, these 3 mutations result in an average rate of mutagenesis 200,000-fold greater than wild-type PolI[277].

While the initial study that generated PolI3M was used for mutagenesis of a PolI-dependent plasmid[277], Halperin *et al*[262] adapted the system for use on defined regions of chromosomal DNA by the use of Cas9. Because PolI3M is recruited to nicked DNA, a nicking version of Cas9, nCas9, was used. nCas9 has only one of two nuclease domains inactivated, allowing it to bind to its target DNA and cleave only a single strand[278]. After nicking the target site, the PolI3M would then bind and re-synthesise a stretch of the DNA at the GOI, introducing random mutations in the process (Fig 50). The mutation types using this tool are all substitutions, predominantly G:C to A:T and A:T to G:C transitions, though with lower rates (approximately 20% of mutations) of transversions[277].

As PolI3M does not require prokaryote-specific cofactors[279], and is instead recruited to the DNA by the nCas9 and the DNA nick, it should be possible to use the system directly in *C. elegans*. Similarly to the dCas9-MiniSOG system, we can provide custom gRNAs to guide nCas9-PolI3M to a GOI to be mutated. One issue with this system is that it would be constitutively active, which would prevent proper screening of mutants. A way around this would be to use a photo-activatable nCas9 to prevent targeting of PolI3M during non-stimulus conditions (Discussed further in Chapter 3.1.5).

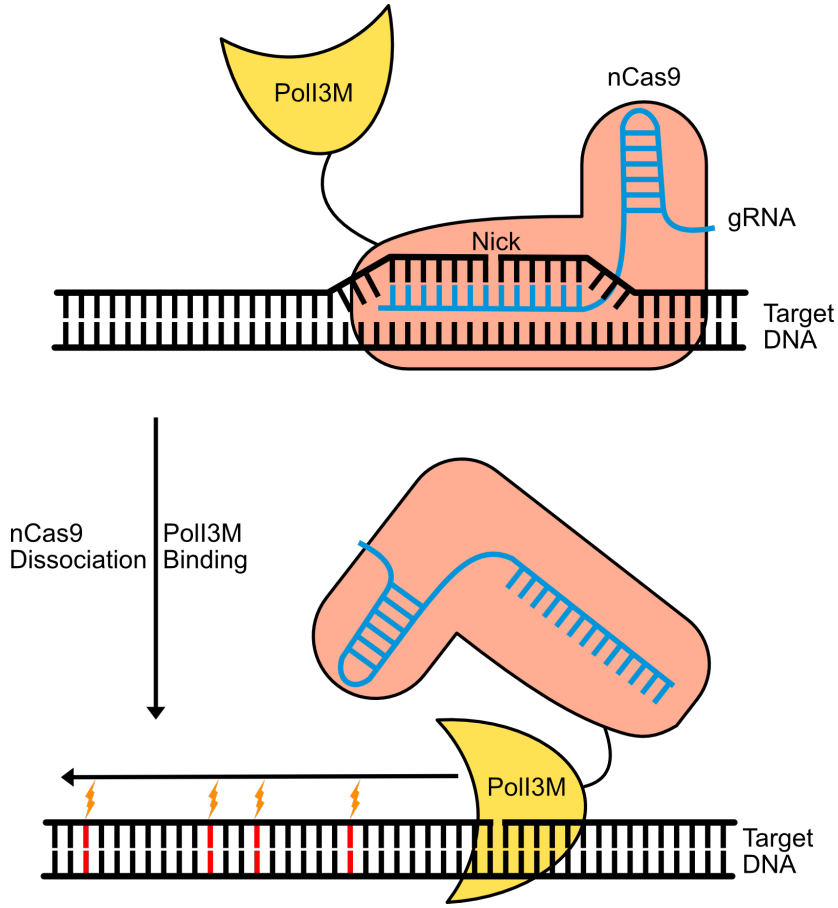

Figure 50: EvolvR. Nicking Cas9 (nCas9) is fused to the error-prone DNA Polymerase I (PolI3M). nCas9 is guided to the target locus (black) and induces a single-stranded break at the target site. After dissociating, PolI3M recognises and binds the DNA nick and initiates polymerase activity, introducing random mutations (red).

3.1.4 CRISPR-X

The final tool I attempted to apply to *C. elegans* is the CRISPR-X system, which was developed in mammalian cells[263]. This exploits the somatic hypermutation process, which uses activation-induced cytidine deaminase (AID) to drive mutations in antibodies during adaptation of the immune system to foreign bodies[280]. AID deaminates the cytosine in a C:G pair to uracil, resulting in a U:G basepair which can then be replicated as either T:A or C:G. Alternatively, the base excision repair pathway can excise uracil and replace it with any random base, or the mismatch repair pathway can recruit an error-prone polymerase at the site to introduce further mutations[281]. dCas9 was used to

guide AID to a target, giving rise to mutation rates approximately 1,000,000-fold higher than wild-type at the target locus, representing all possible transitions and transversions from cytosine and guanine, as well as a low level of mutagenesis from adenine and thymine. Like EvolvR, this system is constitutively active and would also require photocaging of the dCas9 to prevent high background mutagenesis.

3.1.5 Photocaged Cas9

As described, all three tools depend on Cas9 to target the mutagenesis to a specific DNA locus. Both MiniSOG and CRISPR-X tools would use dead Cas9 (dCas9), a Cas9 variant that has both nuclease domains inactivated to remove any cleavage activity. For EvolvR, a nicking Cas9 (nCas9) is used, a variant which has only one nuclease domain inactivate so that nCas9 can only cleave, or nick, a single strand. In all three cases, introducing photo-control into the dCas9 or nCas9 protein would be useful in controlling when mutagenesis can occur and preventing background mutagenesis. For this, it should be possible to photocage dCas9 or nCas9 with the use of photocaged amino acids, which can be used to prevent activity of a critical residue until the photocage group is removed by UV illumination.

A photocaged Cas9 has previously been developed with a critical lysine, K866, photocaged to prevent DNA cleavage[282]. K866 is present in the HNH domain, one of the 2 nuclease domains of Cas9, indicating a role in cleavage. Structural studies of Cas9 support this, as binding of the gRNA to Cas9 induces a conformational change in K866, which seems to convert the HNH domain to an active state[283]. Therefore, Cas9-K866PCK was found to be completely inactive and unable to cleave its DNA substrate, while uncaging of the K866PCK residue restored catalytic activity and allowed Cas9 to cleave the DNA. Since this study photocaged the DNA cleavage activity of Cas9, it should also be possible to use this principle to confer photo-control to nCas9, as we could photocage the HNH nuclease domain using PCK at position K866, while the second nuclease domain, RuvC, is completely deactivated. In this way, the photocaged protein should be completely inactive and unable to cleave the DNA until the photocage group is removed, which would then activate and restore only one nuclease domain, allowing the protein to nick its substrate but not fully cleave it. By using this photocaged nCas9 to control PolI3M localisation, we could prevent nicking and

subsequent PolI3M activity until photo-activation, conferring temporal control to the mutagenesis. Additionally, as translation of the full-length nCas9 product would depend upon incorporation of PCK, a truncated version of nCas9 would be produced during standard maintenance conditions and would be unable to bind to the target DNA. This would introduce two layers of protection of DNA during non-stimulus conditions, preventing background mutagenesis from occurring.

Unlike nCas9, dCas9 would not be able to be photocaged in this way. Since K866 is involved in cleavage, it is likely that dCas9-K866PCK would still retain DNA-binding capabilities, and therefore photocaging K866 would have no effect. Although it would still have the benefit of preventing full-length translation of the dCas9, it would still be preferable to also be able to block DNA binding. For this, I have identified a number of residues as potential candidates. Firstly, K163 has already been identified as a residue which could be photocaged to prevent activity of Cas9[282], though no further study of this residue was carried out. K163 is present in the gRNA-binding REC I domain and is highly conserved, indicating a likely role for it in gRNA-binding[284]. A number of residues have also been identified which extend into the gRNA channel or make direct contact with the gRNA, including K742, Y1013, K1097, K1123, and Y1356[284]. These residues could all be tested by using dCas9's DNA-binding abilities to repress gene activation, as described by Bikard *et al*[273]. This involves targeting photocaged dCas9 to the promoter region of a gene, such as GFP, which blocks access by RNA polymerase and prevents or lowers expression of the gene. Successful candidates can then be applied to the CRISPR-X system.

3.2 Results

3.2.1 dCas9-MiniSOG to Target *dpy-5*

I first set out to test only the MiniSOG system, as this is the only tool that has already been shown to work in *C. elegans* and, since MiniSOG itself is photo-inducible and therefore we did not need to use a photocaged dCas9, it required the least amount of extra variables to be introduced. We designed expression plasmids (Fig. 51A) containing NLS::dCas9::HL::mKate2::HL::MiniSOG::NLS, driven by the germline-specific *mex-5p* promoter. The *tbb-2* 3' UTR was used, as this has been found to aid in germline expression[285]. The nuclear localisation sequences

(NLS) were used to ensure strong localisation of the construct to the nucleus, where mutagenesis can occur. 'HL' denotes Happy Linker, a 24-amino acid linker used to spatially separate each component. Although MiniSOG itself displays some fluorescent signal, this signal is dim[264] and observation of MiniSOG's signal would activate the protein and mutagenise the animal, so mKate2 was added to facilitate visualisation of transgenic animals.

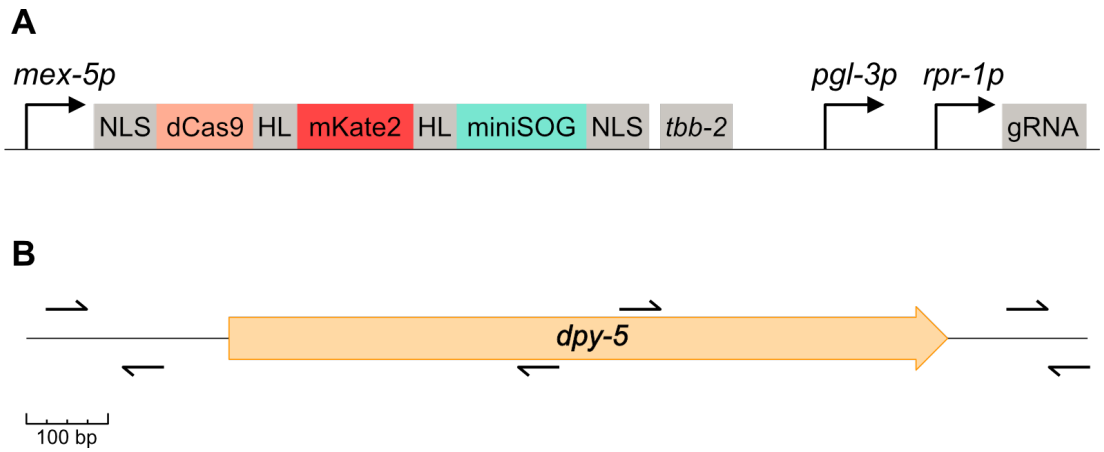

Figure 51: Plasmid Design and Targets for dCas9-MiniSOG. (A) MiniSOG was fused to dCas9 with an mKate2 in between. All 3 components were linked together by Happy Linker (HL). Two nuclear localisation sequences (NLS) were placed at the N' and C' termini to direct the protein into the nucleus. Germline promoter *mex-5p* was used to drive expression. Expression of the gRNA was driven by a dual *pgl-3p* and *rpr-1p* promoter system. **(B)** gRNAs were targeted to *dpy-5*, either to the coding or non-coding strand and targeting either the coding sequence, or the 5' or 3' flanking regions.

Also present on the plasmid is a 96 bp gRNA specific to the *dpy-5* gene. This is driven by a dual promoter system consisting firstly of the germline-specific *pgl-3p* promoter followed by the RNA promoter *rpr-1p*. The use of the first promoter is based on work by Sagi *et al*[286], which found that using the promoter from a protein-coding gene can support the proper expression of a tRNA. *dpy-5* was chosen as the target as mutations in this gene are semi-dominant and give rise to the very obvious 'dumpy' phenotype, characterised by a short and stout body[287]. Several gRNAs were designed, targeting either the promoter, the gene, or the 3' UTR, and targeting either the coding strand or the non-coding strand (Fig. 51B).

While transgenic *C. elegans* strains were generated with these expression constructs, no mKate2 signal could be detected. Additionally, irradiating with blue light gave rise to no dumpy offspring (data not shown). We reasoned that this may be due to expressing from an extrachromosomal array, which often leads to silencing of the transgenes in germline cells[288], likely due to protective pathways that attempt to limit foreign DNA expression in the germline to prevent heritable defects. One method that has been found to help in germline expression is to integrate low copy numbers of the transgene into the *C. elegans* genome[289]. We generated similar expression plasmids as before, but this time in a backbone compatible with the Mos-1-mediated Single Copy Insertion (MosSCI) integration system[290], which uses the Mos1 transposase to integrate a single copy of the expression plasmid into the genome. These strains were generated by SunyBiotech, but still did not give visible signal (data not shown).

3.2.2 Germline Expression of Transgenes

To improve germline expression, constructs were redesigned according to principles described by Al Johani *et al*[291], which facilitate the expression of extrachromosomal arrays in the germline. Firstly, transgenes are codon optimised to more closely match the codon usage of *C. elegans*. Secondly, piwi-interacting RNA (piRNA)-targeting sites are identified and removed using the pirScan tool[292]. piRNAs are small non-coding RNAs which target mRNAs for silencing to prevent transposons from acting within the genome, but their target sequences can often be found in transgenes. Finally, an intron from a ribosomal protein is placed within the first 100 bp of the transgene, and 2-3 Periodic A_n/T_n Cluster (PATC)-rich introns were added throughout the remainder of the transgene. PATCs are characterised by short (2-5 bp) stretches of adenine and thymine with a periodicity of 10 bp, which is approximately the length of 1 helical turn of DNA, resulting in the A/T clusters aligning along a single face of the helix, which causes the helix to bend[293]. These regions make up 6% of the *C. elegans* genome, with the majority of PATCs contained within the introns and 5' and 3' UTRs of coding sequences[294]. A correlation was found between highly PATC-rich genes and the expression of those genes in the germline, and it is thought that the bending of the DNA helix by PATCs prevents nucleosome positioning[294], which is known to be involved in mediating germline expression[295]. Further studies

found that inserting PATC-rich introns into transgenes could prevent germline silencing and drastically increase expression levels[293].

As well as modifying the expression plasmids for better germline expression, constructs were also redesigned to change the way in which MiniSOG or AID were recruited to the target. Instead of having the protein fused directly to dCas9 or nCas9, the SunTag system[296] was used to recruit a larger number of copies of the proteins to the target. In this setup, the dCas9 or nCas9 is fused to a SunTag scaffold, which consists of a repeat of 10 GCN4 epitopes. These are recognised and bound by a single chain fragment variable (scFV) antibody, which is fused to the MiniSOG or the AID protein. This allows a single dCas9 or nCas9 to recruit up to 10 MiniSOG or AID proteins, respectively, increasing the range and extent of mutagenesis that can occur. Additionally, the SunTag scaffold can be extended to contain between 10 and 24 GCN4 epitopes, which may allow for the tuning of the tools to induce more or fewer mutations.

Expression plasmids for MiniSOG using the SunTag system and the modifications for germline expression were generated. Expression plasmids were generated for the EvolvR system with germline expression modifications, and the K866 position of the nCas9 replaced with an amber codon to facilitate incorporation of PCK. Gateway entry plasmids for CRISPR-X were generated with the SunTag system and germline expression modifications, however expression plasmids were not generated as it is still unknown if dCas9 can be photocaged. All expression plasmids are shown in Fig. 52.

Expression plasmids for a GFP-mCherry reporter were also generated, which would allow us to quickly test germline expression and incorporation of PCK (Fig 53A). This reporter consists of a GFP and an mCherry gene separated by an amber stop codon. This would be co-expressed with a tRNA/aaRS synthetase pair to incorporate PCK.

A GFP-mCherry reporter, along with MiniSOG constructs with gRNAs targeting *dpy-5* were bombarded into the N2 strain. However, due to external circumstances, these bombardments could not be screened for several months and had to be kept at room temperature (18-20 °C), rather than the preferred 25 °C for germline expression[291]. Although survivors were screened and transferred to 25 °C for 1-2 generations, these did not prove to express in the germline, instead expressing only in the intestine (Fig 53B).

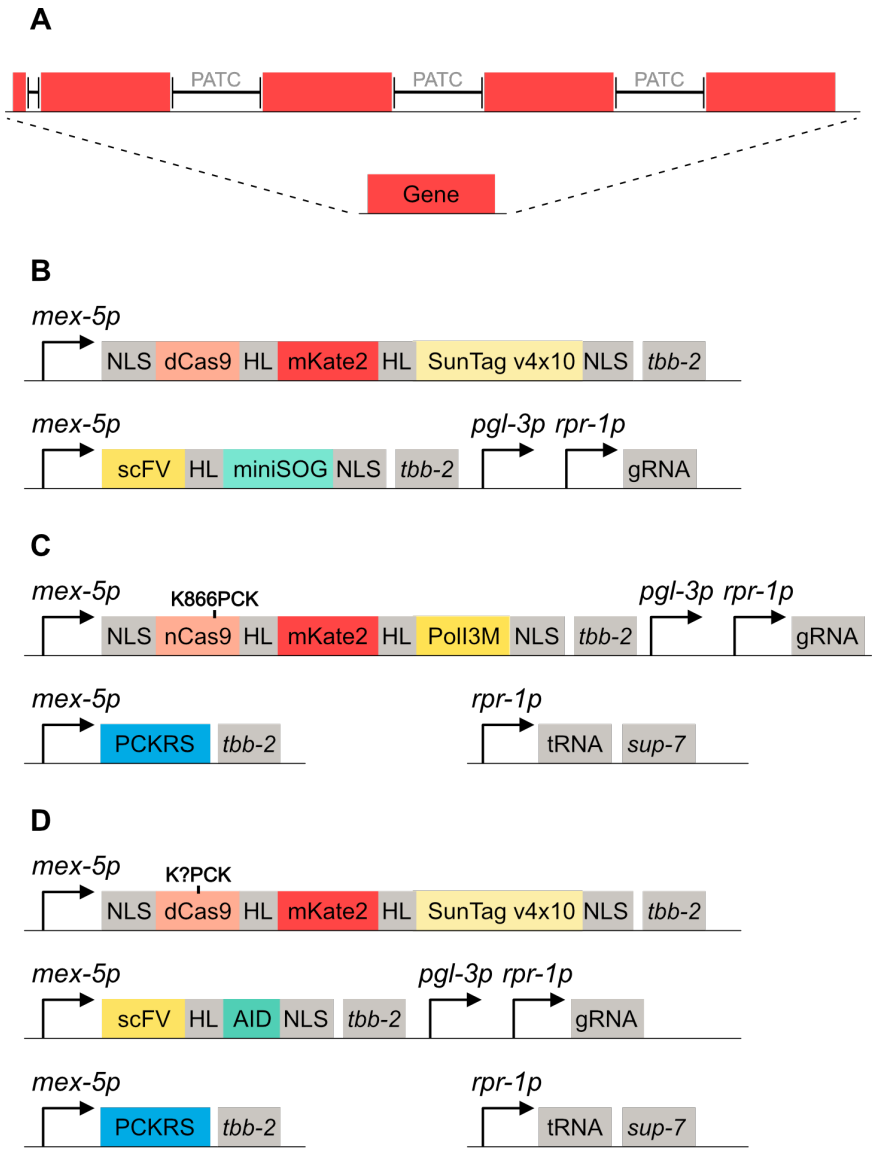

Figure 52: Redesigned Constructs for Germline Expression and Improved Recruitment. **(A)** Insertion of introns for germline expression is shown with an example gene. Within the first 100 bp, an intron from a ribosomal protein is inserted, then three PATC-rich introns spaced throughout the remainder of the gene. For brevity, introns are not shown in **(B-D)**. **(B)** Constructs for expression of dCas9-MiniSOG using the SunTag recruitment system with 10 epitope sites fused to the dCas9 and each MiniSOG protein fused to an scFV antibody. **(C)** Constructs for expression of nCas9-PolI3M, with nCas9 photocaged at the K866 position to incorporate photocaged lysine (PCK). A PCK synthetase (PCKRS) and corresponding tRNA are included. **(D)** Potential constructs for photocaged CRISPR-X system.

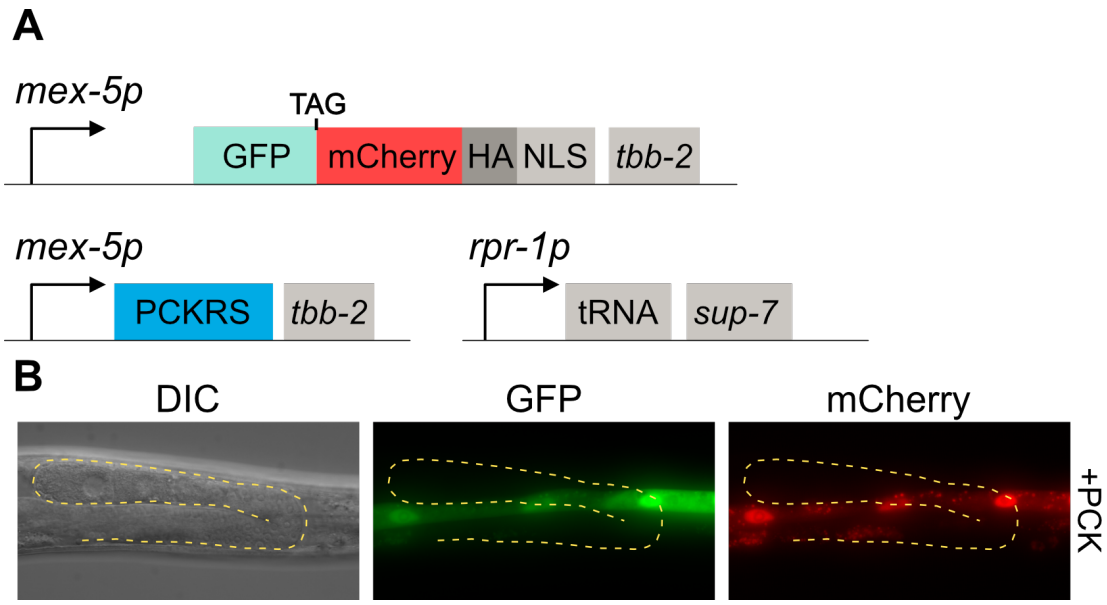

Figure 53: Germline GFP-mCherry Expression. **(A)** Expression constructs used to express GFP-mCherry in the germline. All protein-coding plasmids contain the PATC-rich introns as described previously. **(B)** Expression of *mex5-p* GFP::mCherry. Yellow dotted outline shows the germline of the worm. GFP and mCherry signal are seen in the intestine. Worm fed with photocaged lysine (PCK) prior to imaging. Scale bars are 25 μm.

3.3 Discussion

The aim of this work was to develop tools that could be used for gene-targeted random mutagenesis in *C. elegans*. To achieve this, I attempted to adapt three tools that have previously been used for genome-wide mutagenesis in *C. elegans*, or gene-targeted random mutagenesis in other systems. However, I could not achieve germline expression of any transgene, and because of this the goals of this project could not be achieved. This is not unexpected, as germline expression is especially difficult to achieve in *C. elegans* due to the silencing of both extrachromosomal transgenes and single-copy integrated transgenes, likely used by *C. elegans* as a defence mechanism against foreign DNA[293]. Although measures were introduced to enhance germline expression, these could not be fully tested due to time limitations introduced by the COVID-19 lockdowns. In this case, *C. elegans* strains with germline-optimised constructs could not be screened for a long period of time after the transgenesis event and could not be kept at

permissible temperatures, and therefore did not express properly in the germline. While strains were eventually transferred to 25 °C, germline silencing can be inherited due to epigenetic modification of the chromatin[297] and expression was still not seen. A repeat of these experiments under proper conditions could still yield strains expressing our constructs in the germline and be taken forward for testing. Other measures could also be taken to further optimise the expression of these constructs, including integrating single copies of the germline-optimised constructs into the genome via MosSCI[290].

Once germline expression can be achieved, the 3 tools would require extensive testing and characterisation. Since MiniSOG is the only system used so far in *C. elegans*, and is the only 1 of the 3 that wouldn't require the involvement of genetic code expansion, this is likely to be the simplest to achieve. Strains expressing the dCas9-MiniSOG system can be easily tested by targeting to *dpy-5*, illuminating L4 or young adult worms with blue light, and monitoring offspring for dumpy mutants. Sequencing of the region around *dpy-5* would be done to determine the range of mutagenesis from the target site and the types of mutants generated, though this should be similar to the types found by Noma *et al*[295], and whole-genome sequencing used to determine the off-target mutagenesis rate. The level of mutagenesis should be easily tuneable by adjusting both the intensity of the blue light, as well as the time of illumination. Additionally, the range of mutagenesis could be increased or decreased by changing the number of GCN4 epitopes that make up the SunTag scaffold. Another consideration which will need to be tested is the inclusion of an NLS on the MiniSOG protein. While this may lead to higher mutation rates, as MiniSOG would be better positioned for recruitment to the SunTag scaffold, it may also increase off-target mutation rates if MiniSOG has any level of background activity. One further possibility would be to use the Light Activated Nuclear Shuttle (LANS) system, which uses the blue light-sensitive LOV domain with an embedded NLS and exposed NES[79]. In its unstimulated state, the complex, along with any fused protein, will be present in the cytoplasm by the NES. Upon blue light-stimulated conformational changes, the NLS is exposed and the complex is pulled into the nucleus. Should background activity of MiniSOG prove to be an issue, this system would prevent it from mutagenising the genome until activated by blue light, which would cause nuclear localisation and activation of MiniSOG within a single step.

EvolvR can also be tested immediately, though with some differences from

MiniSOG. Firstly, it would need to be determined if the system can work at all in eukaryotes, and for this it would need to be introduced with the nCas9 unmodified for photocaging. In this way, assuming the system works in *C. elegans*, mutagenesis would be continuous and dumpy worms would arise within the first generations. After proving that EvolvR can work, photocaged nCas9 (pc-nCas9), with K866 replaced with a PCK, would then be used to determine if EvolvR can be made photoinducible. Strains with these constructs would be expected to give rise to no dumpy mutants under standard maintenance conditions (-PCK, -UV) as well as when fed with PCK but not uncaged (+PCK, -UV). Assuming these criteria are met, the pc-nCas9 can then be uncaged (+PCK, +UV), which should then activate nCas9, nicking the target DNA and beginning mutagenesis by PolI3M. The mutants would then need to be characterised to determine the extent of mutagenesis. Since uncaging of a photocaged amino acid is irreversible, this system would not be able to be turned off once started, but instead would depend on the turnover and depletion of PolI3M to stop mutagenesis. However, the extent of mutagenesis should still be somewhat controllable, as the level of pc-nCas9 can be limited by lowering the concentration of PCK used to feed the worms, while the level of uncaging can be controlled by varying the intensity and time of UV illumination. Other tuning mechanisms are possible, though would require generation of new strains. These include lowering the mutation rate by using PolI1M (D424A only) or PolI2M (D424A, I709N), or increasing the mutation rate by using an enhanced nCas9, whereby three mutations (K848A, K1003A, R1060A) lower the binding affinity of Cas9, preventing it from remaining bound after nicking and blocking access by PolI3M[262]. Additionally, the 'window' in which mutations can be generated can also be increased. While the original nCas9-PolI3M generated mutations predominantly within 20 bp of the nick site, the introduction of a bacteriophage thioredoxin-binding domain (TBD) into the polymerase (nCas9-PolI3M-TBD) resulted in a higher mutation rate as far as 56 bp from the nick site[262].

Development of the CRISPR-X system requires more extensive work, as this also requires the development of photocaged dCas9 (pc-dCas9). As discussed previously, multiple candidates exist that could potentially inhibit gRNA or target DNA binding by dCas9. While these residues may not be catalytic like K866, their replacement with PCK may be enough to sterically hinder access to the gRNA or DNA within the binding channels. The ability of pc-dCas9 to bind to target DNA

can be investigated through the use of dCas9-mediated transcriptional activation or repression in *E. coli*[273]. Briefly, the repression method involves targeting the dCas9 to the promoter of a fluorescent reporter, which blocks access by RNA polymerase and inhibits the expression of the fluorescent reporter. The activation method involves fusing dCas9 to the ω subunit of RNA polymerase, expressing this in an *E. coli* strain which lacks the ω subunit, and testing for expression of a reporter. Successful candidates can then be applied to the CRISPR-X system in *C. elegans*. Assuming it is not possible to photocaged dCas9, it could instead be possible to use the LANS system to sequester AID in the cytoplasm until required.

The development of a pc-dCas9 would have wider benefits beyond use for these mutagenesis tools. The method of using dCas9 for transcriptional activation and suppression in *E. coli* has also been applied to *C. elegans*, in this case by fusing either an activating VP16 domain or a repressing Krüppel-associated box domain to the dCas9, respectively[298]. The use of pc-dCas9 would provide a temporal control to the gene upregulation or downregulation, which is particularly useful for the study of toxic or essential genes. dCas9 is also used to direct other tools to target DNA regions, such as enzymes which can apply epigenetic modifications[299,300], which may benefit from greater temporal control. In all cases, pc-dCas9 could be used with higher spatial control, as uncaging could be done using a UV laser to activate the process in single cells.

Each of the 3 tools would have its own advantages and disadvantages, but would also complement each other. For example, MiniSOG is likely to be the most accessible tool, as it does not require photocaged amino acids to function. It is also likely to be the most easily tuneable, requiring only a dimmable light source to achieve varying rates of mutagenesis. On the other hand, MiniSOG and CRISPR-X are likely to exhibit higher off-target mutagenesis rates than EvolvR, as they could affect any nearby base, including those that may be genetically distant but in close spatial proximity. Since EvolvR depends on the nick introduced by nCas9, it should not have this issue and mutations generated by EvolvR should be much more focused at the target site. In either case, backcrossing would still need to be done to clean the genetic background before study of new mutant strains, though the extent of backcrossing required would be less than if using whole-genome or target-select mutagenesis methods. An advantage of CRISPR-X and EvolvR over MiniSOG is the wider range of mutations they can induce, though all 3 tools

could be used to complement each other to get the broadest coverage of mutations possible.

Once developed, these tools have great potential for use in gene-targeted random mutagenesis screens, a currently unutilised area in *C. elegans* genetics[247]. This process would allow for more efficient study of gene function by allowing researchers to generate mutations only in their GOI, perturbing the function of that gene without the influence of the whole genome.

Chapter 4

Materials and Methods

4.1 Plasmid Generation

All plasmids used in this thesis are described in Appendix Tables A2–A6. Transgenes were codon-optimised for expression in *C. elegans* using an online codon optimisation tool[301]. For expression in the germline, methods developed by Al Johani *et al*[291] were followed.

Genetic fragments were either obtained by restriction digest or amplification by PCR from genomic DNA or from existing plasmids, or were purchased as gBlocks (IDT) or GeneArt fragments (Thermo Fisher Scientific). All primers were purchased from IDT. PCRs were carried out using Q5 HotStart High-Fidelity 2X MasterMix (New England Biolabs) or KAPA HiFi HotStart ReadyMix PCR Kit (Roche). PCRs were performed according to manufacturer instructions for each enzyme mix, with the annealing temperatures for each primer pair determined using the online NEB Tm Calculator. PCR fragments were resolved by gel electrophoresis and recovered using the Zymogen Gel Recovery Kit (Zymo Research). Restriction digests were performed using enzymes and buffers from New England Biolabs. Ligation of restriction enzyme digest products was performed using Rapid T4 Ligase (Roche) and Antarctic Phosphatase (New England Biolabs). Assembly of PCR fragments was performed using NEBuilder 2X Master Mix (New England Biolabs).

Expression plasmids were assembled via Gateway Cloning (Thermo Fisher Scientific). Briefly, linear DNA fragments were recombined into entry vectors (Either pDONR P4-P1r, pDONR221, or pDONR P2r-P3) using BP Clonase to generate donor vectors. Three donor vectors were recombined with a destination vector (pDEST R4-R3 Vector II, or modifications of this vector) to assemble an expression plasmid using LR Clonase II Plus. Donor and expression plasmids were transformed into NEB-5α cells (New England Biolabs) by chemical transformation in accordance with manufacturer's specifications. Destination and entry vectors were grown in One Shot *ccdB* Survival cells (Thermo Fisher

Scientific). Transformants were grown overnight on Terrific Broth (TB) agar supplemented with the appropriate selection antibiotic. Individual colonies were picked and grown overnight in TB broth supplemented with the appropriate antibiotic, and plasmids were recovered using the QIAprep Spin Miniprep kit (Qiagen).

Plasmids were verified by colony PCR using OneTaq HotStart 2X Master Mix (New England Biolabs), by restriction digest, and by Sanger sequencing (Source BioScience).

4.2 *C. elegans* Maintenance

The Bristol N2 strain was used as the wild-type *C. elegans* strain for all experiments in this thesis[116]. All transgenic strains were created in the N2 background.

C. elegans strains were maintained according to standard procedures[302]. Strains were cultured on 1X nematode growth medium (NGM) agar seeded with the *E. coli* OP50 strain. For seeding, OP50 was streaked onto TB or Lysogeny Broth (LB) agar and grown at 37 ℃ overnight. A single colony was picked and used to inoculate liquid TB. The OP50 culture was grown for approximately 24 hours, and 300 µL of the culture was used to seed each 6 cm NGM plate. For transgenic strains, Hygromycin B (Formedium) was added to a final concentration of 0.3 mg/ml for selection of the extrachromosomal arrays. For both transgenic strains and wild-type N2, Gentamicin (final concentration 200 µg/ml, Formedium) and Nystatin (final concentration 50 µg/ml, Sigma Aldrich) were added to inhibit bacterial and fungal contamination. Unless otherwise stated, worms were maintained at 20 ℃.

4.3 Bleaching of *C. elegans*

To clean and synchronise worms, bleaching was performed. Gravid adults were washed off of plates with M9 and transferred into Falcon tubes. The volume was made up to 15 mL and the tubes centrifuged at 4400 rpm for 1 minute. Supernatant was removed using a vacuum pump. The pellet containing the gravid adults was resuspended in 5 mL of a bleaching solution (0.6 mL dH_2O, 1.9 mL sodium hypochlorite (5%), 2.5 mL NaOH (1 M)) and shaken continuously for

up to 5 min until adult cuticles were fully dissolved and only eggs remained in solution. Three washes with M9 were performed to remove the bleach and the tube was gently shaken overnight at room temperature to allow for eggs to hatch into M9 buffer.

In cases that only required the cleaning of strains, a spot bleaching method was used. Here, a drop of bleaching solution was deposited on an unseeded region of an NGM plate. Several gravid adults were picked into the drop where they would be dissolved. After eggs had hatched, L1s were allowed to crawl away from the bleach area before being picked onto a fresh NGM plate.

4.4 Freezing *C. elegans* for Long Term Storage

Worms were frozen at -80 ℃ for long term storage. Two 6 cm NGM plates were used to grow worms until a freshly starved population of L1s was achieved. Worms were washed from these plates with 1 mL S Buffer and transferred to cryovials (Thermo Fisher Scientific). 1 mL S Buffer supplemented with 30% glycerol was added to each cryovial and mixed. 600 µL from each cryovial was then combined in a third cryovial, and all three were placed in a polystyrene box at -80 ℃. The next day, the combined vial was thawed and worms plated on NGM to test for survival, while the other two were stored at -80 ℃.

4.5 Lysis of *C. elegans* for Genomic and Extrachromosomal DNA Extraction

For PCR of genomic DNA or of extrachromosomal arrays, worms were first lysed. 50 mL of lysis buffer (10 mL KCl (1 M), 20 mL Tris (0.1 M, pH 8.3), 500 µL $MgCl_2$ (1 M), 900 µL IGEPAL, 900 µL Tween-20, and 1 mL Gelatin (2%), made up with dH_2O) was made and stored at 4 ℃. Just before lysis was performed, 1 µL of Proteinase K (2 mg/ml, Life Technologies) was added to 100 µL lysis buffer.

For lysis, individual worms were picked into 10 µL lysis solution, or 10 µL of worms in M9 were mixed with 10 µL lysis solution. Samples were frozen at -80 ℃ for 30 minutes, then incubated at 65 ℃ for 90 minutes, followed by incubation at 95 ℃ for 15 minutes for heat-inactivation. After being cooled to room temperature, samples were used directly for PCR or stored at -20 ℃.

4.6 Growing Concentrated HB101 for Feeding

E. coli HB101 strain was used to feed liquid cultures. The HB101 strain was transformed with pUC18 to confer ampicillin resistance.

1 L TB was made up in each of 2 L baffled flasks and autoclaved immediately. *E. coli* HB101 was streaked on TB agar supplemented with carbenicillin (200 μg/ml) and a colony used to inoculate the 1 L liquid TB supplemented with carbenicillin (200 μg/ml). This culture was grown at 37 ℃ with shaking (220 rpm) for 24 hours. Cultures were transferred to 1 L centrifuge buckets and centrifuged at 4000 g for 30 minutes at 4 ℃. Supernatant was removed, and the pellet resuspended in 15 mL M9 supplemented with Kanamycin (100 μL of 100 mg/ml stock) and Gentamicin (100 μL of 100 mg/ml stock). Resuspended pellets were transferred to Falcon tubes and stored at 4 ℃ for up to 2 weeks.

4.7 Generation of Transgenic *C. elegans* Strains

4.7.1 Preparation of DNA for Bombardment

For each bombardment, typically three or four plasmids would be included, with each bombardment containing a total of 10-15 μg of DNA. Each plasmid is linearised overnight by restriction digest. Enzymes are chosen that digest only in the backbone of the plasmid, and that can be heat inactivated. After overnight linearisation, all plasmid and enzyme mixes are heat inactivated for 20 minutes at 80 ℃. A small amount of the DNA is run on an agarose gel to verify linearisation before all DNA for a single bombardment is mixed.

For each bombardment, we use 6 mg of 0.3–3 μm diameter gold beads (ChemPur). Gold for all bombardments is weighed out and cleaned by vortexing in 1 mL 70% HPLC-grade ethanol (Sigma-Aldrich) for 30 minutes. Gold is then pelleted in a desktop mini centrifuge and washed three times with sterile water, with each wash involving 1 minute of vortexing before pelleting the gold and removing the supernatant. 6 mg of cleaned gold is distributed to a DNA LoBind tube (Eppendorf), pelleted, and supernatant discarded. The gold is then resuspended in 50 μL of linearised DNA solution. While vortexing, 50 μL of $CaCl_2$ (2.5 M) is added, then 20 μL of Spermidine (100 mM, Merck Life Sciences) added dropwise to the solution. Mixtures were then vortexed for 30 minutes to

allow DNA to precipitate and coat onto the gold. DNA was then pelleted and supernatant discarded, then washed once with 300 µL 70% HPLC-grade ethanol and once with 1 mL 100% HPLC-grade ethanol. Coated gold beats were then resuspended in 170 µL of 100% HPLC-grade ethanol and stored at -20 ℃ until use.

4.7.2 Preparation of *C. elegans*

N2 worms were grown on 9 cm 3X NGM plated seeded with HB101. 3X NGM refers to NGM with 3 times the standard concentration of peptone, which allows for thicker bacterial lawns to be grown. Worms were grown on these plates until they starved, resulting in a dense population of L1s. These worms were washed off the plates with S Medium and passed through a 5 µm polypropylene filter bag (Spectrum), which removed most worms larger than L1. Approximately 500 mL of L1s in S Medium was transferred to an autoclaved baffled 1 L flask. 5 mL of 100X Penicillin/Streptomycin/Amphotericin B mixture (Gibco) and 5 mL of Nystatin solution (10,000 units/mL, Sigma-Aldrich) was added to limit contamination. Flasks were stored at 15 ℃ rotating at 140 rpm for up to 2 weeks prior to bombardment.

Prior to bombardment, the liquid culture of synchronised L1s is fed with freeze-dried OP50 (LabTIE) or concentrated HB101. Worms were kept rotating at 15 ℃ until they reached the young adult stage, where they were then cleaned by sucrose flotation. Worms were first pelleted in 4 50 mL flasks by centrifuging at 1000 G for 1 minute and as much supernatant as possible removed. This was repeated in the same 50 mL flasks until all worms had been pelleted. Final pellets were resuspended in 20 mL of 0.1 M NaCl and cooled on ice. 20 ml of 60% pre-chilled sucrose was added to each worm solution and mixed, then 5 mL of pre-chilled 0.1 M NaCl was layered on top. These were centrifuged at 1200 G for 5 minutes at 4 ℃, with the deceleration speed set to 6/9. After centrifugation, the worms were present in a layer at the bottom of the NaCl layer, above the sucrose. These were recovered using a glass Pasteur pipette and transferred to a fresh Falcon tube with chilled 0.1 M NaCl. Worms were left to settle by gravity, then the supernatant removed and fresh 0.1 M NaCl added. This was repeated until the supernatant was completely clear.

After cleaning, 0.5–1 mL of pelleted adult worms were transferred to an

unseeded, pre-chilled 0X NGM plate using a glass Pasteur pipette. This plate contains no peptone, to limit bacterial and fungal contamination, and is supplemented with 200 µg/ml Kanamycin, 400 µg/ml Carbenicillin, and 200 µg/ml Gentamicin. Worms are spread as evenly as possible across the plate, and the plate returned to 4 ℃ to limit worm activity and movement until the plates had dried, at which point they could be bombarded.

4.7.3 Biolistic Bombardment and Screening

Biolistic bombardment was performed used the Biolistic PDS-1000/He Particle Delivery System with the Hepta Adaptor attachment (Bio-Rad). For each bombardment one rupture disk (900 psi, Bio-Rad) and seven macrocarriers (Inbio Gold) were washed in 100% isopropanol and allowed to dry. After drying, macrocarriers were placed into each of the seven slots in the lower section of the Hepta Adaptor. DNA-coated gold was then spread evenly across each macrocarrier, and this section of the Hepta Adaptor placed into the vacuum chamber. The chamber was depressurised until the ethanol was entirely evaporated. The upper section of the Hepta Adaptor was then attached to the apparatus with the rupture disk inside and secured using a torque wrench. A stopping screen is placed in the lower section of the Hepta Adaptor, and this section is placed into the chamber with the gold facing down. A dried 9 cm plate containing the adult worms is placed without a lid underneath the Hepta Adaptor in the vacuum chamber.

The vacuum chamber was depressurised until the vacuum reached 27.5 inches of mercury, then helium fired into the top chamber until the rupture disk burst and the DNA-coated gold was fired into the worms. The plate containing the worms was then removed and worms allowed to recover on the plate for at least 30 minutes, after which the plate was chunked equally onto six 9 cm 3X NGM plates seeded with HB101. After two days, once most of the eggs had hatched, Hygromycin B was added to a concentration of 0.3 mg/ml, along with Gentamicin (final concentration 200 µg/ml) and Nystatin (final concentration 50 µg/ml). One day later, 50 µL of freeze-dried OP50 or concentrated HB101 was added to each selection plate. Plates were left for two weeks for strains to establish before screening.

After two weeks, the bombardments were screened. All screening and

subsequent work with fluorescent animals was performed using a Leica M165FC fluorescent stereomicroscope. Each of the six selection plates was chunked to a 9 cm 0X NGM plate supplemented with 50 µL of freeze-dried OP50 or concentrated HB101. After allowing 1–2 days for worms to crawl out from the selection plate, these worms were washed off the plate using M9 supplemented with 0.001% Triton X-100 and transferred to ncAA plates.

If the transgenes allowed for visualisation of incorporation alone (For example, in the GFP-TAG-mCherry strains), then screening was performed 1–2 days later from the ncAA plates. Worms that showed the correct expression pattern of the appropriate fluorescent proteins were singled onto a seeded 6 cm 1X NGM and maintained. If the transgene only allowed for visualisation of uncaging (For example, with the FLP and Cre strains), then worms were left on the ncAA plate for 2 days, before being globally uncaged. 1–2 days later, worms showing expression of the appropriate fluorescent proteins were singled onto 6 cm 1X NGM plates.

4.8 ncAA Feeding

All ncAAs were provided to worms by dissolving into the NGM. Plates containing photocaged amino acids were stored in the dark and only used in dark rooms under red LED light.

Methyl-*o*-nitropiperonyl Lysine (PCK, M_w = 384.14 g/mol) was synthesised by ChiroBlock GmbH following a synthesis protocol described by Gautier *et al*[204]. PCK was dissolved in 0.5 M HCl before being added to molten NGM. NGM was neutralised with equimolar NaOH.

Orthonitrobenzyl Tyrosine (ONBY, M_w = 352.77 g/mol) is commercially available and was purchased from Fluorochem. ONBY was dissolved in 5 M NaOH before being added to molten NGM. NGM was neutralised with equimolar HCl.

Methylnitropiperonyl Tyrosine (MNPY, M_w = 374.35 g/mol) was synthesised by NewChem Technologies as a salt with Trifluoroacetate (Final M_w = 488.37 g/mol). MNPY was dissolved in 5 M HCl before being added to molten NGM. NGM was neutralised with equimolar NaOH.

Nitropiperonyl Tyrosine (NPY, M_w = 360.32 g/mol) was synthesised by NewChem Technologies. NPY was dissolved in 5 M NaOH before being added

to molten NGM. NGM was neutralised with equimolar HCl.

The dipeptide of lysine and NPY (K-NPY, M_w = 488.54 g/mol) was synthesised by NewChem Technologies as a salt with di-Trifluoroacetate (Final M_w = 716.54 g/mol). K-NPY was dissolved in water before being added to molten NGM.

Worms were either grown on seeded 6cm NGM plates and washed onto ncAA plates, or bleached and the L1s transferred directly to the ncAA plate. 50 µL of freeze-dried OP50 (LabTie, dissolved in water as per manufacturer's instructions) was added to the ncAA plate as food. Worms grown on ncAA plates were kept at 20 ℃ before uncaging.

4.9 Uncaging

4.9.1 Global Uncaging

Global uncaging was performed using a CL-1000L UV Crosslinker (UVP). Worms would be washed from the ncAA plate and the drop placed on the inside of an empty 35 mm plate without the lid. Worms were illuminated in the crosslinker for 5 minutes at 5 mW/cm^2. After uncaging, worms were transferred to a fresh 6 cm seeded NGM plate.

4.9.2 Single-cell Uncaging

Single-cell uncaging was performed using a Zeiss M2 Imager with a Micropoint Galvo module fitted with a 365 nm dye cell (Andor Technology/Oxford Instruments). Worms were mounted on a 2.5% agar pad and immobilised in 25 mM NaN_3 in M9 buffer. Neurons were identified by the fluorescent reporter co-expressed with photocaged FLP. The nucleus of each targeted neuron was targeted with the laser, with each region being targeted three times with 10-repeat firing at a power setting of 34 (Chosen so that partial bleaching could be observed[77]). After uncaging, worms were washed off the pad with M9 supplemented with 0.001% Triton X-100 and transferred to a seeded 6 cm NGM plate.

Single-cell uncaging by Dr Inka Busack (Section 2.2.7) was performed using a Nikon TiE inverted microscope with a Rapp OptoElectronic DPSL 255/14 at

1% power. To uncage both PVC neurons, a 10 s illumination time was used.

4.10 Imaging of *C. elegans*

Worms were imaged using a Zeiss M2 Imager. Worms were mounted on a 2.5% agar pad and immobilised in 25 mM NaN_3. Fluorescent images were processed and analysed using ImageJ software.

For characterisation of FLP-K223PCK and Cre-K201PCK experiments, 50 µM Floxuridine was added to the plates before imaging to prevent new eggs from hatching. To determine fluorescent intensity of mKate2 and Citrine2, a region was drawn in the head of each worm from the front of the first pharyngeal bulb to the back of the second pharyngeal bulb. The average fluorescence across the entire region was taken, and the minimum value of fluorescence in the same region subtracted to remove background.

4.11 Behavioural Assays

30 µL of 5 mM ATR was added to the lawn of a seeded 6cm NGM plate. Immediately after uncaging, worms were transferred to these plates and left to recover for 24 hours before behavioural assays were performed.

Behaviour plates were made at least 30 minutes before behaviour assays were performed. A 17 mm copper ring was cleaned in ethanol, heated in a flame, and deposited onto the surface of an unseeded 6 cm NGM plate. 20 µL of freeze-dried OP50 (LabTie, prepared according to manufacturers instructions then diluted 1/40 in M9) was placed in the centre of the ring and allowed to dry. Once dried, worms were picked into the centre of the ring and left for at least 15 minutes to acclimatise. A maximum of 10 worms was picked onto each behaviour plate.

Tracking and stimulation of the worms was carried out using the Wormlab system (MBF Bioscience). Chrimson was activated using the integrated 617 nm LED at full power (74 mW/cm^2). Wormlab software (MBF Bioscience) was then used to track each worm. Where multiple tracks were generated for a single worm, these tracks were combined into a single track. Where tracks could not be clearly identified, for example when worms overlapped each other, tracks were discarded.

Behavioural analysis performed by Dr Inka Busack (Section 2.2.7) was done by mounting worms in a microchamber as described by Turek *et al*[303]. Worms

were imaged using a Nikon TiE inverted scope and Chrimson activated using a Texas red filter. For each frame, the nose and the grinder positions were extracted and used to calculate the velocity of the worm.

4.12 PCR Detection of FLP Excision Events

To detect FLP-mediated excision of the STOP cassette, mixed populations of worms were grown on the relevant ncAA for 24 hours before being uncaged for the stated times. Approximately the same number of worms were placed in each drop for uncaging. After uncaging, 10 µL of worms in M9 were added to PCR tubes containing 10 µL lysis solution. These PCR tubes were pre-chilled on a metal heat block placed in dry ice, which froze the solutions in approximate 5-10 s, before proceeding with lysis protocol.

After lysis, 0.5 µL of the worm lysate was used as a template for PCR with OneTaq HotStart 2X Master Mix (New England Biolabs). A number of PCR primer pairs were tested, with the final pair used being primers K321 and K322. All primers tested are listed in Appendix Tables A7 and A8. All PCRs were performed following the manufacturer's protocol, with an annealing temperature of 60 ℃ and an extension time of 4 minutes, which corresponds to the length of the non-excised cassette. PCR products were resolved by agarose gel electrophoresis.

4.13 Statistical Analysis

All analysis of data was performed using GraphPad Prism 9. Statistical tests used are stated in the figure legends. Unless otherwise stated, two independent repeats of each strain were used, and each experiment on each strain was repeated three times, to give an n = 6.

Chapter 5

Conclusions and Future Work

This thesis described the development of optogenetic tools in *C. elegans* based on the use of photocaged amino acids introduced by genetic code expansion. Here, I briefly summarise each tool and discuss potential future developments and applications of the tools.

In Chapter 2, I developed a photo-activatable FLP recombinase by incorporating photocaged amino acids into the active site. This tool was then used to drive gene expression with an extremely high level of spatiotemporal control and with an activation efficiency unattainable by other tools previously developed for the purpose. Using the photo-activatable FLP, we have shown that individual cells can be targeted using a laser, and it would be possible to selectively activate individual cells or any desired combination of cells which cannot be genetically targeted by other methods. Although this can be used to drive the expression of any desired transgene, we focus on using photo-activatable FLP to drive expression of a channelrhodopsin, which allows us to then study the behaviour and function of specific neurons. We show this by expressing the Chrimson channelrhodopsin in the PVC neurons, a neuron pair which does not have its own cell-specific promoter, allowing us to study the function of PVC without interference from other neurons. Work is now ongoing in a collaboration with the Bringmann lab at TU Dresden to target the individual PVCL and PVCR neurons, which would allow us to determine the individual contributions of each of these neurons within the neural circuit.

In this thesis, we focused on using photo-activatable FLP recombinase to drive transgene expression from an extrachromosomal array introduced by biolistic bombardment. However, it may also be possible to use our tool to knockout endogenous gene expression. By introducing FRT sites flanking an endogenous gene, it would be possible to target FLP to remove that gene from the genome in a spatiotemporal manner, allowing for the study of the gene's function in particular cells and at particular timepoints.

While the majority of the work focused on the use of photocaged lysine, we

briefly showed that photocaged tyrosine can also be used to cage FLP activity. Because variants of photocaged tyrosine, particularly ortho-nitrobenzyltyrosine, are much cheaper than photocaged lysine, it would be useful to further characterise the use of photocaged tyrosine to determine its effectiveness in caging FLP activity.

Additionally, the use of photocaged tyrosine to cage FLP would allow us to use both photocaged FLP and photocaged Cre[77] in a single system, by using photocaged tyrosine for one and photocaged lysine for the other. For this, two different codons would be required for incorporation, as well as two mutually orthogonal aaRS/tRNA pairs. Work published by Xi *et al*[78] showed that quadruplet codons can be used to incorporate non-canonical amino acids in *C. elegans*, which drastically increases the coding space available for genetic code expansion, allowing us to incorporate multiple amino acids at one time. Based on work by Willis and Chin[91], work is now ongoing in our lab to develop a mutually orthogonal aaRS/tRNA pair, with promising results already seen using the PylRS/tRNA$^{\mathrm{Pyl}}$ pairs from *Methanosarcina mazei* and *Methanomethylophilus alvus* in conjunction, or those from *Methanosarcina mazei* and the methanogenic archaeon *ISO4-G1* (Greiss lab, unpublished data). These pairs, along with the quadruplet codons, should make it possible to incorporate two unique non-canonical amino acids at two unique sites in a single animal. Using this, photocaged FLP and photocaged Cre could be caged separately and used to drive expression of two different transgenes in two different cells or groups of cells. For example, expression of a neuron-inactivating halorhodopsin and a neuron-activating channelrhodopsin could be driven in two neurons in a circuit at the same time, allowing for greater control in the activation and inhibition of the neural circuit and facilitating deeper dissection of neural circuits than previously possible.

The development of photo-activatable FLP recombinase raised further questions in other areas, particularly in the mechanisms of the SL2 acceptor site in operons. As previously discussed, the most likely cause is an epigenetic factor in the regulation of the SL2 acceptor site. Therefore, it would be useful to cross the strains generated here with strains that are defective in a number of chromatin-modifying genes, which may help to answer if an epigenetic factor is causing the abnormal expression, and determine which pathways are involved.

In Chapter 3, I attempted to develop optogenetic tools for gene-targeted

random mutagenesis. These tools, if fully developed, could be used for reverse genetics screens and directed evolution experiments in a more efficient way than current tools allow. However, as noted, I was unable to drive expression of our transgenes in the germline without silencing. While measures[291] were taken to negate germline silencing, these proved unsuccessful. However, this is most likely due to disruptions caused by the COVID-19 lockdowns, which prevented us from adequately screening our strains after transgenesis. Therefore, repeating the transgenesis under proper conditions and with timely screening may yield animals expressing the transgenes in the germline. Alternatively, combining the germline expression measures described in the chapter with single-copy genome insertions may increase the likelihood of successful germline expression. Once developed, these tools would be useful in the study of a gene of interest, as multiple mutations in the gene could be generated to study which parts of the gene are important for particular functions. Additionally, it could be possible to use these tools for directed evolution. For example, I discussed in Chapter 1 the importance of PylRS and $tRNA^{Pyl}$ evolution in improving incorporation rates and orthogonality for genetic code expansion experiments. However, for *C. elegans* genetic code expansion, evolution of PylRS/$tRNA^{Pyl}$ is done in bacteria, then transplanted into *C. elegans*. While this is often enough to improve PylRS/$tRNA^{Pyl}$ in *C. elegans*, different mechanisms between eukaryotes and prokaryotes may mean that it is not finding as many beneficial modifications as possible. With these optogenetic gene-targeted mutagenesis tools, PylRS or $tRNA^{Pyl}$ could be targeted and evolved directly in *C. elegans*, with strains generated after mutagenesis being tested for improved efficiencies. Although this would still be slower than directed evolution in bacteria, it may prove to increase the effectiveness of PylRS and $tRNA^{Pyl}$ in eukaryotes in a more efficient way.

Development of a new photocaged Cas9 would be useful for these mutagenesis tools, but would also have wider uses. Although a photocaged Cas9 has previously been developed[282], this involved caging the cleavage activity of Cas9. Therefore, this method would not work for caging dead Cas9 (dCas9), as no cleavage activity is involved. It would therefore be useful to photocage the gRNA- or DNA-binding activity of Cas9, which would then be capable of photocaging dCas9. As described in the chapter, dCas9 is used to guide a variety of effector proteins to a specified genomic locus, and the caging of dCas9 could be useful to impart spatiotemporal control on this activity.

This thesis has focused primarily on applying genetic code expansion to incorporate photocaged amino acids into FLP recombinase and into tools for mutagenesis in *C. elegans*. However, genetic code expansion is a rapidly evolving field, and, especially in *C. elegans*, a relatively underutilised method which could be applied to solving a wide range of problems. Much of the work in *C. elegans* has focused on using photocaged amino acids to confer photocontrol onto proteins[77,78,156], and many more possibilities lie in this area. However, a large number of other types of non-canonical amino acid exist, which have so far not been applied to *C. elegans* research. For example, those non-canonical amino acids that introduce permanent or cage-able post-translational modifications can be used to study the use of the modifications in *C. elegans* proteins, while others, such as photocrosslinking amino acids, could be used to study protein-protein and protein-nucleic acid interactions[40]. The current capabilities of genetic code expansion are extensive, and the further development of novel non-canonical amino acids could lead to even more functionalities imparted into proteins, which would help to advance our understanding and control of biological functions.

Appendix

In Table A1, all transgenic *C. elegans* strains generated during this study are listed. Strains generated by me are given the designation 'KB' followed by a number. All KB strains used here are generated by biologistic bombardment into the N2 wild-type strain. All strains are listed with a description of all genetic constructs introduced, followed by the designations of each specific plasmid introduced, as well as the amount in µg of the plasmid linearised and coated onto gold beads prior to bombardment. SG07 was generated by SunyBiotech using the MosSCI integration method.

In Table A2, all expression constructs are listed. All constructs were generated by Multisate Gateway cloning with three entry vectors and one destination vector. All plasmids named 'KB' were generated by myself, 'AT' by Ailish Tynan, 'IR' by Inja Radman, 'SE' or 'SG' by Sebastian Greiss, 'LD' by Lloyd Davis, 'ZX' by Zhiyan Xi, 'JOS' by Jack O'Shea, and 'AX' by Andromachi Xypnitou. pCFJ150 was generated by Frøkjær-Jensen[304].

All destination and entry vectors are listed in Tables A3, A4, A5, and A6. If generated by myself, a brief description on how each construct was generated is also included. pDEST pDONR R4-R3 Vector II, pDONR P4-P1r, pDONR221, and pDONR P2r-P3 are standard destination and entry vectors sold by Invitrogen.

Table A7 lists all primers used for cloning and for the FLP excision experiment

Table A8 lists the primer pairs used for the FLP excision experiment, as well as the expected excised DNA size.

Table A1: List of all transgenic *C. elegans* strains used in this study.

Strain	Allele	Plasmid 1		Plasmid 2		Plasmid 3		Plasmid 4	
		Name	µg	Name	µg	Name	µg	Name	µg
SGR100	*greSi12[mex-5p::sv40NLS::Cas9::HL::mKate2::HL::MiniSOG::egl-13NLS; pgl-3p::rpr-1p::gRNA(dpy-5centre)]* Note: Generated by MosSCI Insertion by SunyBiotech	KB175	n/a						
SGR101	*greEx86[mex-5p::Smad4::MmPCKRS (With PATC-introns); sur-5p::hygR; mex-5p::GFP_(TAG)_mCherry (With PATC-introns); rpr-1p::tRNA^{C15}]*	KB223	3.5	KB222	1.5	SE150	5		
SGR102	*greEx87[myo-3p::FLP_G5D_K223TAG::HA::egl-13NLS::SL2::GFP; eft-3p::FRT_let-858 3'UTR_FRT::mCherry::HA::egl-13NLS; sur-5p::hygR myo-3p::MmPCKRS; rpr-1p::tRNA^{C15}]*	KB319	6	KB301	2.5	KB320	1.5	SE150	4.5
SGR103	*greEx88[mec-7p::FLP_G5D_K223TAG::HA::egl-13NLS::SL2::GFP; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::mKate2::HA; sur-5p::hygR; mec-4p::PKIα::MmPCKRS; rpr-1p::tRNA^{C15}]*	KB313	6	KB332	2.5	SE333	1.5	SE150	4.5

Table A1: List of all transgenic *C. elegans* strains used in this study.

Strain	Allele	Plasmid 1		Plasmid 2		Plasmid 3		Plasmid 4	
		Name	µg	Name	µg	Name	µg	Name	µg
SGR104	*greEx89[snb-1p::FLP_G5D_K223TAG::HA::egl-13NLS::SL2::mKate2; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::SL2::Citrine2; rps-0p::hygR; snb-1::Smad4::PCKRS; rpr-1p::tRNA^{C15}]*	KB364	6	KB413	2.5	SE247	1.5	SE150	4.5
SGR105	*greEx90[snb-1p::FLP_G5D_WT::HA::egl-13NLS::SL2::mKate2; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::Citrine2; rps-0p::hygR; snb-1::Smad4::PCKRS; rpr-1::tRNA^{C15}]*	KB392	6	KB437	2.5	SE247	1.5	SE150	4.5
SGR106	*greEx91[snb-1p::FLP_G5D_WT::HA::egl-13NLS::SL2::mKate2; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::F2A::Citrine2; rps-0p::hygR; snb-1::Smad4::PCKRS; rpr-1p::tRNA^{C15}]*	KB392	6	KB443	2.5	SE247	1.5	SE150	4.5
SGR107	*greEx92[glr-1p::FLP_G5D_K223TAG::HA::egl-13NLS::SL2::mKate2; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::F2A::Citrine2; rps-0p::hygR; glr-1::Smad4::PCKRS; rpr-1p::tRNA^{C15}]*	KB438	6	KB443	2.5	SE174	1.5	SE150	4.5

Table A1: List of all transgenic *C. elegans* strains used in this study.

Strain	Allele	Plasmid 1		Plasmid 2		Plasmid 3		Plasmid 4	
		Name	µg	Name	µg	Name	µg	Name	µg
SGR108	*greEx93[myo-3p::FLP_G5D_K223TAG::HA::egl-13NLS::SL2::mKate2; eft-3p::FRT_lacZ 3'UTR_FRT::Citrine2; sur-5p::hygR; myo-3p::Smad4::PCKRS; rpr-1p::tRNA^{C15}]*	KB379	6	KB450	2.5	KB320	1.5	SE150	4.5
SGR109	*greEx94[elt-2p::FLP_G5D_K223TAG::HA::egl-13NLS::SL2::mKate2; eft-3p::FRT_lacZ 3'UTR_FRT::Citrine2; sur-5p::hygR; elt-2p::Smad4::PCKRS; rpr-1p::tRNA^{C15}]*	KB382	6	KB450	2.5	KB381	1.5	SE150	4.5
SGR110	*greEx95[glr-1p::Cre_K201TAG::HA::egl-13NLS::SL2::mKate2; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::F2A::Citrine2; rps-0p::hygR; glr-1::Smad4::PCKRS; rpr-1p::tRNA^{C15}]*	KB389	6	KB462	2.5	SE174	1.5	SE150	4.5
SGR111	*greEx96[snb-1p::FLP_G5D_WT::HA::egl-13NLS::SL2::mKate2; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::HL::Citrine2; rps-0p::hygR; snb-1::Smad4::PCKRS; rpr-1p::tRNA^{C15}]*	KB392	6	KB459	2.5	SE247	1.5	SE150	4.5

Table A1: List of all transgenic *C. elegans* strains used in this study.

Strain	Allele	Plasmid 1		Plasmid 2		Plasmid 3		Plasmid 4	
		Name	µg	Name	µg	Name	µg	Name	µg
SGR112	*greEx97[glr-1p::FLP_G5D_Y343TAG::HA::egl-13NLS::SL2::mKate2; maco-1p::FRT_lacZ 3'UTR_FRT::Chrimson::F2A::Citrine2; sur-5p::hygR; glr-1::Smad4::NPYRS; rpr-1p::tRNA^{C15}]*	KB463	6	KB454	2.5	AT11	1.5	SE150	4.5
SGR113	*greEx98[glr-1p::FRT_LacZ 3'UTR_FRT::mKate2::SL2::Citrine2; rps-0p::hygR]*	KB474	8	IR87	2				
SGR114	*greEx99[glr-1p::FRT_LacZ 3'UTR_FRT::mKate2::SL2::Citrine2; glr-1p::FLP_G5D_WT::HA::egl-13NLS::SL2::Citrine2; rps-0p::hygR]*	KB474	8	KB475	2				
SGR115	*greEx100[glr-1p::mKate2::SL2::Citrine2; rps-0p::hygR]*	KB476	8	IR87	2				

Table A2: Expression plasmids generated by Gateway cloning.

Name	Destination vector	Entry vectors		
		pDONR P4-P1r	pDONR221	pDONR P2r-P3
AT11	SE80	IR157	KB126	SG606
IR87	*Not known.*			
KB175	pCFJ150	KB121	SE106	AX35
KB222	pDEST R4-R3 Vector II	KB207	KB204	KB205
KB223	SE80	KB207	KB203	KB193
KB301	SE114	AX67	SG743	KB299
KB313	pDEST R4-R3 Vector II	ZX204	KB257	SE306
KB319	pDEST R4-R3 Vector II	ZX58	KB257	SE306
KB320	SE80	ZX58	SE153	SG606
KB332	pDEST R4-R3 Vector II	SE77	LD222	LD287
KB364	pDEST R4-R3 Vector II	IR91	KB257	IR322
KB379	pDEST R4-R3 Vector II	ZX58	KB257	IR322
KB381	SE80	SG639	SE154	SG606
KB382	pDEST R4-R3 Vector II	SG639	KB257	IR322
KB389	pDEST R4-R3 Vector II	IR157	SE17	IR322
KB392	pDEST R4-R3 Vector II	IR91	KB258	IR322
KB413	KB408	IR91	LD222	KB399
KB437	pDEST R4-R3 Vector II	SE77	LD222	KB402
KB438	pDEST R4-R3 Vector II	IR157	KB257	IR322
KB443	pDEST R4-R3 Vector II	SE77	LD222	KB431
KB450	pDEST R4-R3 Vector II	AX67	LD222	KB442
KB454	pDEST R4-R3 Vector II	SE77	LD222	KB442
KB459	pDEST R4-R3 Vector II	SE77	LD222	KB449
KB462	pDEST R4-R3 Vector II	SE77	SG365	KB442
KB463	pDEST R4-R3 Vector II	IR157	KB321	IR322
KB474	KB407	IR157	LD222	KB469
KB475	IR98	IR157	KB258	ZX114
KB476	pDEST R4-R3 Vector II	IR157	KB468	SE239
SE174	IR98	IR157	SE154	SG606
SE247	IR98	IR91	SE154	SG606
SE333	SE80	LD400	SE153	SG606

Table A3: Destination plasmids used to generated expression constructs.

Name	Description	Source
pCFJ150	ttTi5605 left recomb; Unc-119p::CB-Unc-119(+); ttTi5605 right recomb	Frøkjær-Jensen et al[304]
IR98	*rps-0::hygR* in pDEST R4-R3 Vector II backbone	Radman et al[151]
KB407	*SL2::mCitrine* in pDEST R4-R3 Vector II backbone	Made from JOS13 (*bglII::SL2::GFP::NLS*) and SE239 (*mCitrine*). JOS13 was opened up by PCR with K286 and K289. *mCitrine2* was isolated from SE239 by PCR with K285 and K288. Fragments were assembled by NEBuilder
KB408	*SL2::Citrine2* in pDEST R4-R3 Vector backbone	Made from JOS13 (*bglII::SL2::GFP::NLS*) and SE308 (*Citrine2*).JOS13 was opened up by PCR with K286 and K289. *Citrine2* was isolated from SE239 by PCR with K290 and K291. Fragments were assembled by NEBuilder
pDEST R4-R3 Vector II	Standard pDEST backbone	Invitrogen
SE80	*sur-5p::hygR* in pDEST R4-R3 Vector backbone	Generated by Sebastian Greiss
SE114	*let-858 3'UTR* in pDEST R4-R3 Vector backbone	Generated by Sebastian Greiss

Table A4: Entry plasmids in the pDONR P4-P1r (Position 1) backbone.

Name	Description	Source
AX67	*eft-3p*	Generated by Andromachi Xypnitou
IR91	*snb-1p*	Generated by Inja Radman
IR157	*glr-1p*	Generated by Inja Radman
KB121	*mex-5p::sv40NLS::dCas9::HL*	Made from KB119 and AX15. KB119 was opened up by restriction digest with AscI. *mex-5p* was isolated from AX15 by PCR with K52 and K53. Fragments were assembled by NEBuilder.
KB207	*mex-5p*	*mex-5p* was isolated from AX15 by PCR with primers K84 and K85. Fragment was inserted into pDONR P4-P1r by BP reaction
LD400	*mec-4p*	Generated by Lloyd Davis
SE77	*maco-1p*	Generated by Sebastian Greiss
SG639	*elt-2p*	Generated by Sebastian Greiss
ZX204	*mec-7p*	Generated by Zhiyan Xi
ZX58	*myo-3p*	Generated by Zhiyan Xi

Table A5: Entry plasmids in the pDONR221 (Position 2) backbone.

Name	Description	Source
KB126	*Smad4::MmNPYRS*	Made by inserting *Smad4* NES into LD184. LD184 was opened up by PCR with primers S430 and S431. *Smad4* was inserted by NEBuilder of this fragment with primer S434
KB203	*Smad4::MmPCKRS* (With PATC-introns)	Made by Golden Gate cloning between KB194, KB188, KB189, and KB190, digested with BsaI. KB194 is *Smad4::MmPCKRS* with Golden Gate acceptor sites at each intron position. KB188, KB189, and KB190 each contain a PATC-rich intron (Sequence provided by Christian Frøkjær-Jensen[291])
KB204	*GFP::TAG* (With PATC-introns)	Made by Golden Gate cloning between KB195 and KB188, digested with BsaI. KB195 is *GFP::TAG* with Golden Gate acceptor sites at the intron position. KB188 contains a PATC-rich intron (Sequence provided by Christian Frøkjær-Jensen[291])
KB257	*FLP_G5D_K223TAG::HA::egl-13NLS*	Made from SE13 by opening plasmid with primers S674 and S675 and re-ligating the plasmid using NEBuilder to insert the G5D mutation
KB258	*FLP_G5D_WT::HA::egl-13NLS*	Made from SE14 by opening plasmid with primers S674 and S675 and re-ligating the plasmid using NEBuilder to insert the G5D mutation
KB321	*FLP_G5D_Y343TAg::HA::egl-13NLS*	Made from KB258 by opening plasmid with primers S667 and S668 and re-ligating the plasmid using NEBuilder to insert the Y343TAG mutation
KB468	*mKate2::HA*	Made by amplifying mKate2 from KB264 with primers S58 and K307
LD222	*FRT_LacZ 3'UTR_FRT*	Generated by Lloyd Davis
SE17	*Cre_K201TAG::HA::egl-13NLS*	Generated by Sebastian Greiss
SE106	*mKate2::HL::MiniSOG::egl-13NLS*	Generated by Sebastian Greiss
SE153	*PKIa::MmPCKRS*	Generated by Sebastian Greiss
SE154	*Smad4::MmPCKRS*	Generated by Sebastian Greiss
SG365	*LoxP_LacZ 3'UTR_LoxP*	Generated by Sebastian Greiss
SG743	*FRT_let-858 3'UTR_FRT*	Generated by Sebastian Greiss

Table A6: Entry plasmids in the pDONR2r-P3 (Position 3) backbone.

Name	Description	Source
AX35	*pgl-3p::rpr-1p::gRNA(dpy-5 centre)*	Generated by Andromachi Xypnitou
IR322	*SL2::mKate2*	Generated by Inja Radman
KB193	*tbb-2 3'UTR*	Made by isolating *tbb-2* terminator from SE106 by PCR with primers K110 and K111 and inserting fragment into pDONR P2r-P3 by BP reaction
KB205	*mCherry* (With PATC-introns)	Made by Golden Gate cloning between KB191, KB189, and KB190, digested with BsaI. KB191 is *mCherry* with Golden Gate acceptor sites at the intron position. KB189 and KB190 each contain a PATC-rich intron (Sequence provided by Christian Frøkjær-Jensen[291])
KB299	*mCherry:HA::egl-13NLS*	Made by inserting ATG start sequence into SG79
KB399	*Chrimson*	Made by isolating *Chrimson* from KB362 with primers S691 and K269 and inserting it into pDONR P2r-P3 by BP reaction
KB402	*Chrimson::Citrine2::HA::egl-13NLS*	Made from KB362 and SE308. KB362 was opened by PCR with primers K242 and K243. *Citrine2* was isolated from SE308 by PCR with primers K244 and K245. Fragments were assembled by NEBuilder
KB431	*Chrimson::F2A::Citrine2::HA*	KB402 was opened up with primers K304 and K305. *F2A* was inserted as primer S420 by NEBuilder
KB442	*Citrine2::HA*	*Citrine2* was isolated from KB402 by PCR with primers K306 and L145. Fragment was inserted into pDONR P2r-P3 by BP reaction
KB449	*Chrimson::HL::Citrine2::HA*	KB402 was opened up with primers K309 and K310 and *HL* inserted as primer K308 By NEBuilder
KB469	*mKate2::HA*	*mKate2* isolated from KB264 with primers S114 and K314 and fragment inserted into pDONR P2r-P3 by BP reaction
LD287	*Chrimson::mKate2::HA*	Generated by Lloyd Davis
SE150	*rpr-1p::tRNA*^{C15}	Generated by Sebastian Greiss
SE239	*SL2::mCitrine*	Generated by Sebastian Greiss
SE306	*SL2::mEGFP*	Generated by Sebastian Greiss
SG606	*let-858 3'UTR*	Generated by Sebastian Greiss
ZX114	*SL2::CFP*	Generated by Zhiyan Xi

Table A7: Primers used for cloning and excision experiments

Name	Description	Sequence
K52	Pmex-5.FOR	CTTTGTATAGAAAAGTTGTAatatcagtttttaaaaaattaaaccataa aacaaataatataaccca
K53	Pmex-5.REV	gctgtatttttgtctaccttacgcttcttctttggcatttttctctgt
K84	Pmex-5 attB.FOR	GGGGACAACTTTGTATAGAAAAGTTGTAatatcagtttttaaaaaa ttaaaccataaaacaaataatataaccca
K85	Pmex-5 attB.REV	GGGGACTGCTTTTTTGTACAAACTTGTtctctgtctgaaacattca attgattatcgataaaattaattaagag
K110	tbb-2 pos3.FOR	GGGGACAGCTTTCTTGTACAAAGTGGTAgataaatgcaagatcctt tcaagcattcc
K111	tbb-2 pos2.FOR	GGGGACAACTTTGTATAATAAAGTTGTgacttttttcttggcggca caataaagtttcctct
K242	Chrimson.REV	TTTCGATACCATTCCTCCTGGCGCGCCTCCTCCGGC
K243	Chrimson.FOR	AGCTCTACAAGTACCCATATGATGTCCCAGACTAC
K244	Citrine2.FOR	AGGCGCGCCAGGAGGAATGGTATCGAAAGGaGAGGA aTTATTTAC
K245	Citrine2.REV	TATGGGTACTTGTAGAGCTCtcCCATTCCGAGGG
K269	Chrimson(stop).R	GGGGACAACTTTGTATAATAAAGTTGTTTAGACTGTA TCCTCATCTTCCTCTTCCAC
K285	SE305_239-JOS13.R	TTTCGTCTACGGCTCATCTTGTAGAGCTCGTCCATTCC
K286	JOS13-SE305.R	ACCTTTAGAAACCATtttttctaccggtacagcagtttccc
K288	SE239-JOS13.F	gtagaaaaaATGGTATCGAAAGGCGAGGAG
K289	JOS13-SE239_308.R	CCTTTcGAtACCATtttttctaccggtacagcagtttccc
K290	SE308_JOS13.R	TCGTCTACGGCTCATCTTGTAGAGCTCTCCCATTCC
K291	SE308_JOS13.F	accggtagaaaaaATGGTATCGAAAGGAGAGG
K304	Chrimson_F2A.REV	TGCTTGACTCCGGATCCTGGGACTGTATCCTCATC TTCCTC
K305	Chrimson_F2A.FOR	GTCGAGTCCAACCCAGGACCAATGGTATCGAAAGG aGAGG
K306	attB3_Citrine2.F	GGGGACAGCTTTCTTGTACAAAGTGGTAAAAAAATG GTATCGAAAGGAGAGGAATTATTTACTG
K307	attB1_mKate.F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTaaaaaA TGTCCGAGCTCATCAAGGAGAACATG
K308	Happy Linker	ATCTTGGGGGCTCCATCTGGTGGCGGCGCTACGGCC GGTGCTGGGGGAGCCGGAGGTCCAGCCGGTTTGATT
K309	HLtoKB402.R	ACCAGATGGAGCCCCCAAGATGGCTGGGTCGACTGT ATCCTC
K310	HLtoKB402.F	GCCGGAGGTCCAGCCGGTTTGATTATGGTATCGAAA GGaGAGGAaTTATTTACTGG

Table A7: Primers used for cloning and excision experiments

Name	Description	Sequence
K314	attR1_mKate.F	GGGGACAGCTTTCTTGTACAAAGTGGTAaaaaaATG TCCGAGCTCATCAAGGAG
K319	Excision.FOR1	agctcaagttgatgtgaccaga
K320	Excision.REV1	ACGAGGATTGGAACGACACC
K321	Excision.FOR2	tgtttccgattccttttaacgtct
K322	Excision.REV2	CATGTGGTCTGGGTAACGGG
K323	Excision.FOR3	tccgattccttttaacgtcttct
K324	Excision.REV3	GAGGATTGGAACGACACCAGT
L145	attB3 3' GFP let858 R	GGGGACAACTTTGTATAATAAAGTTGatacggattcgca tttgccaag
S58	HA attB2R	GGGGACCACTTTGTACAAGAAAGCTGGGTTTAAGCGTAG TCTGGGACATCATATG
S114	HA attB3R	GGGGACAACTTTGTATAATAAAGTTGTTTAAGCGTAGTC TGGGACATCATATGG
S420	F2A attB2F	GGGGACAGCTTTCTTGTACAAAGTGGTACCAGGATCCGG AGTCAAGCAAAC
S430	MmPylRS NEB F	ATGGACAAGAAGCCACTCAACAC
S431	MmPylRS NEB R	CATTTTTGCAGCCTGCTTTTTTGTACA
S434	Smad4 MmPylRS	AAAAGCAGGCTGCAAAAATGGGAATCGACCTCTCCGGA CTCACCCTCCAAATGGACAAGAAGCCACTCAA
S667	FLP Y343 TAG R	gatggcggtgatttggtgggtCTAggtggtacgggcgacggcggaggcacgcttg
S668	FLP Y343 TAG F	gtaccaccTAGacccaccaaatcaccgccatcccagac
S674	FLP G5D F	acaattcgACatcctctgcaagaccccaccaaag
S675	FLP G5D R	cttgcagaggatGTcgaattgtggcatttttagcctg
S691	Chrimson attB2F	GGGGACAGCTTTCTTGTACAAAGTGGTAAAAAATGGC CGAACTCATCTCATCG

Table A8: Primer pairs used for excision experiments

Pair	Primer 1	Primer 2	Excised length (bp)
1	K319	K320	466
2	K319	K322	653
3	K319	K324	464
4	K321	K320	235
5	**K321**	**K322**	**422**
6	K321	K324	233
7	K323	K320	231
8	K323	K322	418
9	K323	K324	229

Bibliography

1. Crick, F. H. On protein synthesis. eng. *Symposia of the Society for Experimental Biology* **12,** 138–163. ISSN: 0081-1386 (1958).
2. Crick, F. Central Dogma of Molecular Biology. en. *Nature* **227.** Number: 5258 Publisher: Nature Publishing Group, 561–563. ISSN: 1476-4687. doi:10.1038/227561a0 (Aug. 1970).
3. Berg, M. D. & Brandl, C. J. Transfer RNAs: diversity in form and function. eng. *RNA biology* **18,** 316–339. ISSN: 1555-8584. doi:10.1080/15476286.2020.1809197 (Mar. 2021).
4. Crick, F. H. C. Codon—anticodon pairing: The wobble hypothesis. en. *Journal of Molecular Biology* **19,** 548–555. ISSN: 0022-2836. doi:10.1016/S0022-2836(66)80022-0 (Aug. 1966).
5. Chin, J. W. Expanding and reprogramming the genetic code. en. *Nature* **550.** Number: 7674 Publisher: Nature Publishing Group, 53–60. ISSN: 1476-4687. doi:10.1038/nature24031 (Oct. 2017).
6. Ibba, M. & Söll, D. Aminoacyl-tRNAs: setting the limits of the genetic code. en. *Genes & Development* **18.** Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, 731–738. ISSN: 0890-9369, 1549-5477. doi:10.1101/gad.1187404 (Apr. 2004).
7. Rubio Gomez, M. A. & Ibba, M. Aminoacyl-tRNA synthetases. *RNA* **26,** 910–936. ISSN: 1355-8382. doi:10.1261/rna.071720.119 (Aug. 2020).
8. Valencia-Sánchez, M. I. *et al.* Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases*. en. *Journal of Biological Chemistry* **291,** 14430–14446. ISSN: 0021-9258. doi:10.1074/jbc.M116.730382 (July 2016).
9. Sankaranarayanan, R. *et al.* The structure of threonyl-tRNA synthetase-tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. eng. *Cell* **97,** 371–381. ISSN: 0092-8674. doi:10.1016/s0092-8674(00)80746-1 (Apr. 1999).

10. Holley, R. W. *et al.* Structure of a Ribonucleic Acid. *Science* **147.** Publisher: American Association for the Advancement of Science, 1462–1465. doi:10.1126/science.147.3664.1462 (Mar. 1965).

11. Giegé, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research* **26,** 5017–5035. ISSN: 0305-1048. doi:10.1093/nar/26.22.5017 (Nov. 1998).

12. Nissen, P., Kjeldgaard, M., Thirup, S., Clark, B. F. C. & Nyborg, J. The ternary complex of aminoacylated tRNA and EF-Tu-GTP. Recognition of a bond and a fold. en. *Biochimie* **78,** 921–933. ISSN: 0300-9084. doi:10.1016/S0300-9084(97)86714-4 (Jan. 1996).

13. Tukalo, M., Yaremchuk, G., Kovalenko, O. P., Kriklivyi, I. & Gudzera, O. I. Recognition of tRNAs with a long variable arm by aminoacyl-tRNA synthetases. doi:10.7124/BC.000825 (2013).

14. Mukai, T. *et al.* Transfer RNAs with novel cloverleaf structures. *Nucleic Acids Research* **45,** 2776–2785. ISSN: 0305-1048. doi:10.1093/nar/gkw898 (Mar. 2017).

15. Hingerty, B., Brown, R. S. & Jack, A. Further refinement of the structure of yeast tRNAPhe. en. *Journal of Molecular Biology* **124,** 523–534. ISSN: 0022-2836. doi:10.1016/0022-2836(78)90185-7 (Sept. 1978).

16. Xu, B., Liu, L. & Song, G. Functions and Regulation of Translation Elongation Factors. *Frontiers in Molecular Biosciences* **8.** ISSN: 2296-889X (2022).

17. Marintchev, A. & Wagner, G. Translation initiation: structures, mechanisms and evolution. en. *Quarterly Reviews of Biophysics* **37.** Publisher: Cambridge University Press, 197–284. ISSN: 1469-8994, 0033-5835. doi:10.1017/S0033583505004026 (Nov. 2004).

18. Bhavsar, R. B., Makley, L. N. & Tsonis, P. A. The other lives of ribosomal proteins. *Human Genomics* **4,** 327–344. ISSN: 1473-9542. doi:10.1186/1479-7364-4-5-327 (June 2010).

19. Ramakrishnan, V. Ribosome Structure and the Mechanism of Translation. English. *Cell* **108.** Publisher: Elsevier, 557–572. ISSN: 0092-8674, 1097-4172. doi:10.1016/S0092-8674(02)00619-0 (Feb. 2002).

20. Nirenberg, M. W. *et al.* On the Coding of Genetic Information. en. *Cold Spring Harbor Symposia on Quantitative Biology* **28.** Publisher: Cold Spring Harbor Laboratory Press, 549–557. ISSN: 0091-7451, 1943-4456. doi:10.1101/SQB.1963.028.01.074 (Jan. 1963).

21. Koonin, E. V. & Novozhilov, A. S. Origin and evolution of the genetic code: the universal enigma. *Iubmb Life* **61,** 99–111. ISSN: 1521-6543. doi:10.1002/iub.146 (Feb. 2009).

22. Ambrogelly, A., Palioura, S. & Söll, D. Natural expansion of the genetic code. en. *Nature Chemical Biology* **3.** Number: 1 Publisher: Nature Publishing Group, 29–35. ISSN: 1552-4469. doi:10.1038/nchembio847 (Jan. 2007).

23. Muramatsu, T. *et al.* A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from Escherichia coli. en. *Journal of Biological Chemistry* **263,** 9261–9267. ISSN: 0021-9258. doi:10.1016/S0021-9258(19)76533-8 (July 1988).

24. Sengupta, S., Yang, X. & Higgs, P. G. The Mechanisms of Codon Reassignments in Mitochondrial Genetic Codes. *Journal of Molecular Evolution* **64,** 662–688. ISSN: 0022-2844. doi:10.1007/s00239-006-0284-7 (June 2007).

25. Miranda, I., Silva, R. & Santos, M. A. S. Evolution of the genetic code in yeasts. eng. *Yeast (Chichester, England)* **23,** 203–213. ISSN: 0749-503X. doi:10.1002/yea.1350 (Feb. 2006).

26. Schmidt, R. L. & Simonović, M. Synthesis and decoding of selenocysteine and human health. *Croatian Medical Journal* **53,** 535–550. ISSN: 0353-9504. doi:10.3325/cmj.2012.53.535 (Dec. 2012).

27. Gonzalez-Flores, J., Shetty, S. P., Dubey, A. & Copeland, P. R. The Molecular Biology of Selenocysteine. *Biomolecular concepts* **4,** 349–365. ISSN: 1868-5021. doi:10.1515/bmc-2013-0007 (Aug. 2013).

28. Hao, B. *et al.* A new UAG-encoded residue in the structure of a methanogen methyltransferase. eng. *Science (New York, N.Y.)* **296,** 1462–1466. ISSN: 1095-9203. doi:10.1126/science.1069556 (May 2002).

29. Polycarpo, C. *et al.* An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. eng. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 12450–12454. ISSN: 0027-8424. doi:10.1073/pnas.0405362101 (Aug. 2004).

30. Srinivasan, G., James, C. M. & Krzycki, J. A. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. eng. *Science (New York, N.Y.)* **296,** 1459–1462. ISSN: 1095-9203. doi:10.1126/science.1069588 (May 2002).

31. Blight, S. K. *et al.* Direct charging of tRNACUA with pyrrolysine in vitro and in vivo. en. *Nature* **431.** Number: 7006 Publisher: Nature Publishing Group, 333–335. ISSN: 1476-4687. doi:10.1038/nature02895 (Sept. 2004).

32. Longstaff, D. G. *et al.* A natural genetic code expansion cassette enables transmissible biosynthesis and genetic encoding of pyrrolysine. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 1021–1026. ISSN: 0027-8424. doi:10.1073/pnas.0610294104 (Jan. 2007).

33. Quitterer, F., List, A., Eisenreich, W., Bacher, A. & Groll, M. Crystal Structure of Methylornithine Synthase (PylB): Insights into the Pyrrolysine Biosynthesis. *Angewandte Chemie International Edition* **51.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201106765, 1339–1342. ISSN: 1521-3773. doi:10.1002/anie.201106765 (2012).

34. Quitterer, F., List, A., Beck, P., Bacher, A. & Groll, M. Biosynthesis of the 22nd Genetically Encoded Amino Acid Pyrrolysine: Structure and Reaction Mechanism of PylC at 1.5Å Resolution. en. *Journal of Molecular Biology* **424,** 270–282. ISSN: 0022-2836. doi:10.1016/j.jmb.2012.09.007 (Dec. 2012).

35. Cellitti, S. E. *et al.* D-Ornithine coopts pyrrolysine biosynthesis to make and insert pyrroline-carboxy-lysine. en. *Nature Chemical Biology* **7.** Number: 8 Publisher: Nature Publishing Group, 528–530. ISSN: 1552-4469. doi:10.1038/nchembio.586 (Aug. 2011).

36. Namy, O. *et al.* Adding pyrrolysine to the Escherichia coli genetic code. en. *FEBS Letters* **581.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1016/j.febslet.200 5282–5288. ISSN: 1873-3468. doi:10.1016/j.febslet.2007.10.022 (2007).

37. Théobald-Dietrich, A., Frugier, M., Giegé, R. & Rudinger-Thirion, J. Atypical archaeal tRNA pyrrolysine transcript behaves towards EF-Tu as a typical elongator tRNA. *Nucleic Acids Research* **32,** 1091–1096. ISSN: 0305-1048. doi:10.1093/nar/gkh266 (2004).

38. Alkalaeva, E. *et al.* Translation termination in pyrrolysine-utilizing archaea. *FEBS letters* **583,** 3455–3460. ISSN: 0014-5793. doi:10.1016/j.febslet.2009.09.044 (Nov. 2009).

39. Zhang, Y., Baranov, P. V., Atkins, J. F. & Gladyshev, V. N. Pyrrolysine and Selenocysteine Use Dissimilar Decoding Strategies*. en. *Journal of Biological Chemistry* **280,** 20740–20751. ISSN: 0021-9258. doi:10.1074/jbc.M501458200 (May 2005).

40. Wang, L., Xie, J. & Schultz, P. G. Expanding the genetic code. eng. *Annual Review of Biophysics and Biomolecular Structure* **35,** 225–249. ISSN: 1056-8700. doi:10.1146/annurev.biophys.35.101105.121507 (2006).

41. Ross, J. B. *et al.* Spectral enhancement of proteins: biological incorporation and fluorescence characterization of 5-hydroxytryptophan in bacteriophage lambda cI repressor. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 12023–12027. ISSN: 0027-8424 (Dec. 1992).

42. Hinds, M. G., King, R. W. & Feeney, J. 19F n.m.r. studies of conformational changes accompanying cyclic AMP binding to 3-fluorophenylalanine-containing cyclic AMP receptor protein from Escherichia coli. *Biochemical Journal* **287,** 627–632. ISSN: 0264-6021 (Oct. 1992).

43. Tang, Y. & Tirrell, D. A. Attenuation of the editing activity of the Escherichia coli leucyl-tRNA synthetase allows incorporation of novel amino acids into proteins in vivo. eng. *Biochemistry* **41,** 10635–10645. ISSN: 0006-2960. doi:10.1021/bi026130x (Aug. 2002).

44. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. eng. *Science (New York, N.Y.)* **244,** 182–188. ISSN: 0036-8075. doi:10.1126/science.2649980 (Apr. 1989).

45. Nakamura, Y., Gojobori, T. & Ikemura, T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. eng. *Nucleic Acids Research* **28,** 292. ISSN: 0305-1048. doi:10.1093/nar/28.1.292 (Jan. 2000).

46. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science* **292.** Publisher: American Association for the Advancement of Science, 498–500. doi:10.1126/science.1060077 (Apr. 2001).

47. Wang, L., Brock, A. & Schultz, P. G. Adding L-3-(2-Naphthyl)alanine to the genetic code of E. coli. eng. *Journal of the American Chemical Society* **124,** 1836–1837. ISSN: 0002-7863. doi:10.1021/ja012307j (Mar. 2002).

48. Chin, J. W. *et al.* Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli. eng. *Journal of the American Chemical Society* **124,** 9026–9027. ISSN: 0002-7863. doi:10.1021/ja027007w (Aug. 2002).

49. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **99,** 11020–11024. ISSN: 0027-8424. doi:10.1073/pnas.172226299 (Aug. 2002).

50. Lang, K. & Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chemical Reviews* **114.** Publisher: American Chemical Society, 4764–4806. ISSN: 0009-2665. doi:10.1021/cr400355w (May 2014).

51. Thibodeaux, G. N. *et al.* Transforming a Pair of Orthogonal tRNA-aminoacyl-tRNA Synthetase from Archaea to Function in Mammalian Cells. *PLoS ONE* **5,** e11263. ISSN: 1932-6203. doi:10.1371/journal.pone.0011263 (June 2010).

52. Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N-acetyllysine in recombinant proteins. *Nature Chemical Biology* **4,** 232–234. doi:10.1038/nchembio.73 (Apr. 2008).

53. Kouzarides, T. Chromatin Modifications and Their Function. en. *Cell* **128,** 693–705. ISSN: 0092-8674. doi:10.1016/j.cell.2007.02.005 (Feb. 2007).

54. Chen, J. & Tsai, Y.-H. Applications of Genetic Code Expansion in Studying Protein Post-translational Modification. en. *Journal of Molecular Biology. Recent Advances in Genetic Code Expansion: from Cell Engineering to Protein Design* **434,** 167424. ISSN: 0022-2836. doi:10.1016/j.jmb.2021.167424 (Apr. 2022).

55. Mukai, T. *et al.* Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. eng. *Biochemical and Biophysical Research Communications* **371,** 818–822. ISSN: 1090-2104. doi:10.1016/j.bbrc.2008.04.164 (July 2008).

56. Greiss, S. & Chin, J. W. Expanding the Genetic Code of an Animal. *Journal of the American Chemical Society* **133,** 14196–14199. ISSN: 0002-7863. doi:10.1021/ja2054034 (Sept. 2011).

57. Bianco, A., Townsley, F. M., Greiss, S., Lang, K. & Chin, J. W. Expanding the genetic code of Drosophila melanogaster. en. *Nature Chemical Biology* **8.** Number: 9 Publisher: Nature Publishing Group, 748–750. ISSN: 1552-4469. doi:10.1038/nchembio.1043 (Sept. 2012).

58. Ernst, R. J. *et al.* Genetic code expansion in the mouse brain. en. *Nature Chemical Biology* **12.** Number: 10 Publisher: Nature Publishing Group, 776–778. ISSN: 1552-4469. doi:10.1038/nchembio.2160 (Oct. 2016).

59. Liu, J., Hemphill, J., Samanta, S., Tsang, M. & Deiters, A. Genetic Code Expansion in Zebrafish Embryos and Its Application to Optical Control of Cell Signaling. *Journal of the American Chemical Society* **139.** Publisher: American Chemical Society, 9100–9103. ISSN: 0002-7863. doi:10.1021/jacs.7b02145 (July 2017).

60. Guo, L.-T. *et al.* Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. *Proceedings of the National Academy of Sciences of the United States of America* **111,** 16724–16729. ISSN: 0027-8424. doi:10.1073/pnas.1419737111 (Nov. 2014).

61. Lajoie, M. J. *et al.* Genomically Recoded Organisms Expand Biological Functions. *Science (New York, N.Y.)* **342,** 357–360. ISSN: 0036-8075. doi:10.1126/science.1241459 (Oct. 2013).

62. Amiram, M. *et al.* Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. eng. *Nature Biotechnology* **33,** 1272–1279. ISSN: 1546-1696. doi:10.1038/nbt.3372 (Dec. 2015).

63. Huang, L.-Y. *et al.* Unnatural Amino Acid Replacement in a Yeast G Protein-Coupled Receptor in Its Native Environment. *Biochemistry* **47.** Publisher: American Chemical Society, 5638–5648. ISSN: 0006-2960. doi:10.1021/bi701866e (May 2008).

64. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. en. *Nature* **472.** Number: 7344 Publisher: Nature Publishing Group, 499–503. ISSN: 1476-4687. doi:10.1038/nature09929 (Apr. 2011).

65. Bryson, D. I. *et al.* Continuous directed evolution of aminoacyl-tRNA synthetases. en. *Nature Chemical Biology* **13.** Number: 12 Publisher: Nature Publishing Group, 1253–1260. ISSN: 1552-4469. doi:10.1038/nchembio.2474 (Dec. 2017).

66. Herring, S. *et al.* The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity. *FEBS letters* **581,** 3197–3203. ISSN: 0014-5793. doi:10.1016/j.febslet.2007.06.004 (July 2007).

67. Jiang, R. & Krzycki, J. A. PylSn and the Homologous N-terminal Domain of Pyrrolysyl-tRNA Synthetase Bind the tRNA That Is Essential for the Genetic Encoding of Pyrrolysine. *The Journal of Biological Chemistry* **287,** 32738–32746. ISSN: 0021-9258. doi:10.1074/jbc.M112.396754 (Sept. 2012).

68. Sharma, V. *et al.* Evolving the N-terminal Domain of Pyrrolysyl-tRNA Synthetase for Improved Incorporation of Noncanonical Amino Acids. *Chembiochem : a European journal of chemical biology* **19,** 26–30. ISSN: 1439-4227. doi:10.1002/cbic.201700268 (Jan. 2018).

69. Perez-Perez, D. A. *et al.* Engineered small metal-binding protein tag improves the production of recombinant human growth hormone in the periplasm of Escherichia coli. eng. *FEBS open bio* **10,** 546–551. ISSN: 2211-5463. doi:10.1002/2211-5463.12808 (Apr. 2020).

70. Koch, N. G., Baumann, T. & Budisa, N. Efficient Unnatural Protein Production by Pyrrolysyl-tRNA Synthetase With Genetically Fused Solubility Tags. *Frontiers in Bioengineering and Biotechnology* **9.** ISSN: 2296-4185 (2021).

71. Nikić, I. *et al.* Debugging Eukaryotic Genetic Code Expansion for Site-Specific Click-PAINT Super-Resolution Microscopy. *Angewandte Chemie (International Ed. in English)* **55,** 16172–16176. ISSN: 1433-7851. doi:10.1002/anie.201608284 (Dec. 2016).

72. Eargle, J., Black, A. A., Sethi, A., Trabuco, L. G. & Luthey-Schulten, Z. Dynamics of Recognition between tRNA and Elongation Factor Tu. *Journal of molecular biology* **377,** 1382–1405. ISSN: 0022-2836. doi:10.1016/j.jmb.2008.01.073 (Apr. 2008).

73. Fan, C., Xiong, H., Reynolds, N. M. & Söll, D. Rationally evolving tRNAPyl for efficient incorporation of noncanonical amino acids. eng. *Nucleic Acids Research* **43,** e156. ISSN: 1362-4962. doi:10.1093/nar/gkv800 (Dec. 2015).

74. Serfling, R. *et al.* Designer tRNAs for efficient incorporation of non-canonical amino acids by the pyrrolysine system in mammalian cells. *Nucleic Acids Research* **46,** 1–10. ISSN: 0305-1048. doi:10.1093/nar/gkx1156 (Jan. 2018).

75. Nozawa, K. *et al.* Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. eng. *Nature* **457,** 1163–1167. ISSN: 1476-4687. doi:10.1038/nature07611 (Feb. 2009).

76. Watanabe, Y. *et al.* Higher-order structure of bovine mitochondrial tRNA(SerUGA): chemical modification and computer modeling. *Nucleic Acids Research* **22,** 5378–5384. ISSN: 0305-1048 (Dec. 1994).

77. Davis, L. *et al.* Precise optical control of gene expression in C elegans using improved genetic code expansion and Cre recombinase. *eLife* **10** (eds Hobert, O., Sengupta, P. & Bringmann, H.) Publisher: eLife Sciences Publications, Ltd, e67075. ISSN: 2050-084X. doi:10.7554/eLife.67075 (Aug. 2021).

78. Xi, Z. *et al.* Using a quadruplet codon to expand the genetic code of an animal. *Nucleic Acids Research* **50,** 4801–4812. ISSN: 0305-1048. doi:10.1093/nar/gkab1168 (May 2022).

79. Yumerefendi, H. *et al.* Control of Protein Activity and Cell Fate Specification via Light-Mediated Nuclear Translocation. *PLoS ONE* **10.** ISSN: 1932-6203. doi:10.1371/journal.pone.0128443 (June 2015).

80. Plotkin, J. B. & Kudla, G. Synonymous but not the same: the causes and consequences of codon bias. eng. *Nature Reviews. Genetics* **12,** 32–42. ISSN: 1471-0064. doi:10.1038/nrg2899 (Jan. 2011).

81. Fredens, J. *et al.* Total synthesis of Escherichia coli with a recoded genome. en. *Nature* **569.** Number: 7757 Publisher: Nature Publishing Group, 514–518. ISSN: 1476-4687. doi:10.1038/s41586-019-1192-5 (May 2019).

82. Isaacs, F. J. *et al.* Precise Manipulation of Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **333.** Publisher: American Association for the Advancement of Science, 348–353. doi:10.1126/science.1205822 (July 2011).

83. Ostrov, N. *et al.* Design, synthesis, and testing toward a 57-codon genome. eng. *Science (New York, N.Y.)* **353,** 819–822. ISSN: 1095-9203. doi:10.1126/science.aaf3639 (Aug. 2016).

84. Robertson, W. E. *et al.* Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science* **372.** Publisher: American Association for the Advancement of Science, 1057–1062. doi:10.1126/science.abg3029 (June 2021).

85. Venetz, J. E. *et al.* Chemical synthesis rewriting of a bacterial genome to achieve design flexibility and biological functionality. eng. *Proceedings of the National Academy of Sciences of the United States of America* **116,** 8070–8079. ISSN: 1091-6490. doi:10.1073/pnas.1818259116 (Apr. 2019).

86. Wang, K., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. en. *Nature Biotechnology* **25.** Number: 7 Publisher: Nature Publishing Group, 770–777. ISSN: 1546-1696. doi:10.1038/nbt1314 (July 2007).

87. Dunkelmann, D. L., Oehm, S. B., Beattie, A. T. & Chin, J. W. A 68-codon genetic code to incorporate four distinct non-canonical amino acids enabled by automated orthogonal mRNA design. *Nature chemistry* **13,** 1110–1117. ISSN: 1755-4330. doi:10.1038/s41557-021-00764-5 (Nov. 2021).

88. Anderson, J. C. *et al.* An expanded genetic code with a functional quadruplet codon. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 7566–7571. ISSN: 0027-8424. doi:10.1073/pnas.0401517101 (May 2004).

89. Chatterjee, A., Lajoie, M. J., Xiao, H., Church, G. M. & Schultz, P. G. A bacterial strain with a unique quadruplet codon specifying non-native amino acids. eng. *Chembiochem: A European Journal of Chemical Biology* **15,** 1782–1786. ISSN: 1439-7633. doi:10.1002/cbic.201402104 (Aug. 2014).

90. Wang, N., Shang, X., Cerny, R., Niu, W. & Guo, J. Systematic Evolution and Study of UAGN Decoding tRNAs in a Genomically Recoded Bacteria. eng. *Scientific Reports* **6,** 21898. ISSN: 2045-2322. doi:10.1038/srep21898 (Feb. 2016).

91. Willis, J. C. W. & Chin, J. W. Mutually orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs. en. *Nature Chemistry* **10.** Number: 8 Publisher: Nature Publishing Group, 831–837. ISSN: 1755-4349. doi:10.1038/s41557-018-0052-5 (Aug. 2018).

92. Slanina, T. *et al.* In Search of the Perfect Photocage: Structure–Reactivity Relationships in meso-Methyl BODIPY Photoremovable Protecting Groups. *Journal of the American Chemical Society* **139.** Publisher: American Chemical Society, 15168–15175. ISSN: 0002-7863. doi:10.1021/jacs.7b08532 (Oct. 2017).

93. Mendel, D., Ellman, J. A. & Schultz, P. G. Construction of a light-activated protein by unnatural amino acid mutagenesis. *Journal of the American Chemical Society* **113.** Publisher: American Chemical Society, 2758–2760. ISSN: 0002-7863. doi:10.1021/ja00007a063 (Mar. 1991).

94. Cook, S. N. *et al.* Photochemically Initiated Protein Splicing. *Angewandte Chemie International Edition in English* **34.**

_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.199516291, 1629–1630. ISSN: 1521-3773. doi:10.1002/anie.199516291 (1995).

95. England, P. M., Lester, H. A., Davidson, N. & Dougherty, D. A. Site-specific, photochemical proteolysis applied to ion channels invivo. *Proceedings of the National Academy of Sciences* **94.** Publisher: Proceedings of the National Academy of Sciences, 11025–11030. doi:10.1073/pnas.94.20.11025 (Sept. 1997).

96. Miller, J. C., Silverman, S. K., England, P. M., Dougherty, D. A. & Lester, H. A. Flash Decaging of Tyrosine Sidechains in an Ion Channel. en. *Neuron* **20,** 619–624. ISSN: 0896-6273. doi:10.1016/S0896-6273(00)81001-6 (Apr. 1998).

97. Wu, N., Deiters, A., Cropp, T. A., King, D. & Schultz, P. G. A Genetically Encoded Photocaged Amino Acid. *Journal of the American Chemical Society* **126.** Publisher: American Chemical Society, 14306–14307. ISSN: 0002-7863. doi:10.1021/ja040175z (Nov. 2004).

98. Endo, M., Nakayama, K. & Majima, T. Design and Synthesis of Photochemically Controllable Restriction Endonuclease BamHI by Manipulating the Salt-Bridge Network in the Dimer Interface. *The Journal of Organic Chemistry* **69.** Publisher: American Chemical Society, 4292–4298. ISSN: 0022-3263. doi:10.1021/jo035774n (June 2004).

99. Rothman, D. M. *et al.* Caged Phosphoproteins. *Journal of the American Chemical Society* **127.** Publisher: American Chemical Society, 846–847. ISSN: 0002-7863. doi:10.1021/ja043875c (Jan. 2005).

100. Courtney, T. & Deiters*, A. Recent Advances in the Optical Control of Protein Function through Genetic Code Expansion. *Current opinion in chemical biology* **46,** 99–107. ISSN: 1367-5931. doi:10.1016/j.cbpa.2018.07.011 (Oct. 2018).

101. Hemphill, J., Chou, C., Chin, J. W. & Deiters, A. Genetically Encoded Light-Activated Transcription for Spatiotemporal Control of Gene Expression and Gene Silencing in Mammalian Cells. *Journal of the American Chemical Society* **135.** Publisher: American Chemical Society, 13433–13439. ISSN: 0002-7863. doi:10.1021/ja4051026 (Sept. 2013).

102. Engelke, H., Chou, C., Uprety, R., Jess, P. & Deiters, A. Control of protein function through optochemical translocation. eng. *ACS synthetic biology* **3,** 731–736. ISSN: 2161-5063. doi:10.1021/sb400192a (Oct. 2014).

103. Deiters, A., Groff, D., Ryu, Y., Xie, J. & Schultz, P. G. A Genetically Encoded Photocaged Tyrosine. *Angewandte Chemie International Edition* **45.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.200600264, 2728–2731. ISSN: 1521-3773. doi:10.1002/anie.200600264 (2006).

104. Luo, J., Torres-Kolbus, J., Liu, J. & Deiters, A. Genetic Encoding of Photocaged Tyrosines with Improved Light-Activation Properties for the Optical Control of Protease Function. eng. *Chembiochem: A European Journal of Chemical Biology* **18,** 1442–1447. ISSN: 1439-7633. doi:10.1002/cbic.201700147 (July 2017).

105. Chou, C., Young, D. D. & Deiters, A. A light-activated DNA polymerase. eng. *Angewandte Chemie (International Ed. in English)* **48,** 5950–5953. ISSN: 1521-3773. doi:10.1002/anie.200901115 (2009).

106. Arbely, E., Torres-Kolbus, J., Deiters, A. & Chin, J. W. Photocontrol of Tyrosine Phosphorylation in Mammalian Cells via Genetic Encoding of Photocaged Tyrosine. *Journal of the American Chemical Society* **134.** Publisher: American Chemical Society, 11912–11915. ISSN: 0002-7863. doi:10.1021/ja3046958 (July 2012).

107. Chou, C. & Deiters, A. Light-Activated Gene Editing with a Photocaged Zinc-Finger Nuclease. *Angewandte Chemie International Edition* **50.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201101157, 6839–6842. ISSN: 1521-3773. doi:10.1002/anie.201101157 (2011).

108. Nguyen, D. P. *et al.* Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. eng. *Journal of the American Chemical Society* **136,** 2240–2243. ISSN: 1520-5126. doi:10.1021/ja412191m (Feb. 2014).

109. Ren, W., Ji, A. & Ai, H.-w. Light Activation of Protein Splicing with a Photocaged Fast Intein. *Journal of the American Chemical Society* **137.** Publisher: American Chemical Society, 2155–2158. ISSN: 0002-7863. doi:10.1021/ja508597d (Feb. 2015).

110. Bochet, C. G. Photolabile protecting groups and linkers. en. *Journal of the Chemical Society, Perkin Transactions 1.* Publisher: The Royal Society of Chemistry, 125–142. ISSN: 1364-5463. doi:10.1039/B009522M (Jan. 2002).

111. Klán, P. *et al.* Photoremovable Protecting Groups in Chemistry and Biology: Reaction Mechanisms and Efficacy. *Chemical Reviews* **113.** Publisher: American Chemical Society, 119–191. ISSN: 0009-2665. doi:10.1021/cr300177k (Jan. 2013).

112. Maupas, É. Modes et formes de reproduction des nématodes. *Archives de zoologie expérimentale et générale* **8,** 463–624 (1900).

113. Schulenburg, H. & Félix, M.-A. The Natural Biotic Environment of Caenorhabditis elegans. *Genetics* **206,** 55–86. ISSN: 0016-6731. doi:10.1534/genetics.116.195511 (May 2017).

114. Dougherty, E. C. & Calhoun, H. G. Possible Significance of Free-living Nematodes in Genetic Research. en. *Nature* **161.** Number: 4079 Publisher: Nature Publishing Group, 29–29. ISSN: 1476-4687. doi:10.1038/161029a0 (Jan. 1948).

115. Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. General Nature of the Genetic Code for Proteins. en. *Nature* **192.** Number: 4809 Publisher: Nature Publishing Group, 1227–1232. ISSN: 1476-4687. doi:10.1038/1921227a0 (Dec. 1961).

116. Brenner, S. The genetics of Caenorhabditis elegans. eng. *Genetics* **77,** 71–94. ISSN: 0016-6731. doi:10.1093/genetics/77.1.71 (May 1974).

117. Nicholas, W. L., Dougherty, E. C. & Hansen, E. L. Axenic Cultivation of Caenorharditis Briggsae (nematoda: Rhabditidae) with Chemically Undefined Supplements; Comparative Studies with Related Nematodes*. en. *Annals of the New York Academy of Sciences* **77.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1959.tb36902.x, 218–236. ISSN: 1749-6632. doi:10.1111/j.1749-6632.1959.tb36902.x (1959).

118. Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R. *Origins of the Model* en. Publication Title: C. elegans II. 2nd edition (Cold Spring Harbor Laboratory Press, 1997).

119. Hodgkin, J., Horvitz, H. R. & Brenner, S. Nondisjunction Mutants of the Nematode CAENORHABDITIS ELEGANS. eng. *Genetics* **91,** 67–94. ISSN: 0016-6731. doi:10.1093/genetics/91.1.67 (Jan. 1979).

120. Corsi, A. K., Wightman, B. & Chalfie, M. A Transparent Window into Biology: A Primer on Caenorhabditis elegans. *Genetics* **200,** 387–407. ISSN: 0016-6731. doi:10.1534/genetics.115.176099 (June 2015).

121. Alberts, B. *et al.* Caenorhabditis Elegans: Development from the Perspective of the Individual Cell. en. *Molecular Biology of the Cell. 4th edition.* Publisher: Garland Science (2002).

122. Sulston, J. E. & Horvitz, H. R. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. en. *Developmental Biology* **56,** 110–156. ISSN: 0012-1606. doi:10.1016/0012-1606(77)90158-0 (Mar. 1977).

123. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. eng. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **314,** 1–340. ISSN: 0962-8436. doi:10.1098/rstb.1986.0056 (Nov. 1986).

124. Culetto, E. & Sattelle, D. B. A role for Caenorhabditis elegans in understanding the function and interactions of human disease genes. eng. *Human Molecular Genetics* **9,** 869–877. ISSN: 0964-6906. doi:10.1093/hmg/9.6.869 (Apr. 2000).

125. Alvarez, J., Alvarez-Illera, P., Santo-Domingo, J., Fonteriz, R. I. & Montero, M. Modeling Alzheimer's Disease in Caenorhabditis elegans. eng. *Biomedicines* **10,** 288. ISSN: 2227-9059. doi:10.3390/biomedicines10020288 (Jan. 2022).

126. Cooper, J. F. & Van Raamsdonk, J. M. Modeling Parkinson's Disease in C. elegans. *Journal of Parkinson's Disease* **8,** 17–32. ISSN: 1877-7171. doi:10.3233/JPD-171258 (2018).

127. Ellwood, R. A., Piasecki, M. & Szewczyk, N. J. Caenorhabditis elegans as a Model System for Duchenne Muscular Dystrophy. *International Journal of Molecular Sciences* **22,** 4891. ISSN: 1422-0067. doi:10.3390/ijms22094891 (May 2021).

128. Hewitt, J. E. *et al.* Muscle strength deficiency and mitochondrial dysfunction in a muscular dystrophy model of Caenorhabditis elegans and its functional response to drugs. *Disease Models & Mechanisms* **11,** dmm036137. ISSN: 1754-8403. doi:10.1242/dmm.036137 (Dec. 2018).

129. Miguel-Aliaga, I. *et al.* The Caenorhabditis elegans orthologue of the human gene responsible for spinal muscular atrophy is a maternal product critical for germline maturation and embryonic viability. eng. *Human Molecular Genetics* **8,** 2133–2143. ISSN: 0964-6906. doi:10.1093/hmg/8.12.2133 (Nov. 1999).

130. Kyriakakis, E., Markaki, M. & Tavernarakis, N. Caenorhabditis elegans as a model for cancer research. *Molecular & Cellular Oncology* **2,** e975027. ISSN: 2372-3556. doi:10.4161/23723556.2014.975027 (Dec. 2014).

131. Sulston, J. E. Post-embryonic development in the ventral cord of Caenorhabditis elegans. eng. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* **275,** 287–297. ISSN: 0962-8436. doi:10.1098/rstb.1976.0084 (Aug. 1976).

132. Ellis, H. M. & Horvitz, H. R. Genetic control of programmed cell death in the nematode C. elegans. en. *Cell* **44,** 817–829. ISSN: 0092-8674. doi:10.1016/0092-8674(86)90004-8 (Mar. 1986).

133. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. en. *Nature* **391.** Number: 6669 Publisher: Nature Publishing Group, 806–811. ISSN: 1476-4687. doi:10.1038/35888 (Feb. 1998).

134. Conte, D., MacNeil, L. T., Walhout, A. J. M., Mello, C. C. & Mello, C. C. RNA Interference in Caenorhabditis elegans. *Current protocols in molecular biology* **109,** 26.3.1–30. doi:10.1002/0471142727.mb2603s109 (Jan. 2015).

135. Xu, W., Jiang, X. & Huang, L. RNA Interference Technology. *Comprehensive Biotechnology,* 560–575. doi:10.1016/B978-0-444-64046-8.00282-2 (2019).

136. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W. & Prasher, D. C. Green Fluorescent Protein as a Marker for Gene Expression. *Science* **263.** Publisher: American Association for the Advancement of Science, 802–805. doi:10.1126/science.8303295 (Feb. 1994).

137. Corsi, A. K. A Biochemist's Guide to C. elegans. *Analytical biochemistry* **359,** 1–17. ISSN: 0003-2697. doi:10.1016/j.ab.2006.07.033 (Dec. 2006).

138. Davis, P. *et al.* WormBase in 2022-data, processes, and tools for analyzing Caenorhabditis elegans. eng. *Genetics* **220,** iyac003. ISSN: 1943-2631. doi:10.1093/genetics/iyac003 (Apr. 2022).

139. Altun, Z. & Hall, D. Handbook of C. elegans anatomy. *WormAtlas* (2021).

140. Joshi, J., Rubart, M. & Zhu, W. Optogenetics: Background, Methodological Advances and Potential Applications for Cardiovascular Research and Medicine. *Frontiers in Bioengineering and Biotechnology* **7.** ISSN: 2296-4185 (2020).

141. Fenno, L., Yizhar, O. & Deisseroth, K. The Development and Application of Optogenetics. *Annual Review of Neuroscience* **34.** _eprint: https://doi.org/10.1146/annurev-neuro-061010-113817, 389–412. doi:10.1146/annurev-neuro-061010-113817 (2011).

142. Schneider, F., Grimm, C. & Hegemann, P. Biophysics of Channelrhodopsin. *Annual Review of Biophysics* **44.** _eprint: https://doi.org/10.1146/annurev-biophys-060414-034014, 167–186. doi:10.1146/annurev-biophys-060414-034014 (2015).

143. Zhang, F. *et al.* Multimodal fast optical interrogation of neural circuitry. en. *Nature* **446.** Number: 7136 Publisher: Nature Publishing Group, 633–639. ISSN: 1476-4687. doi:10.1038/nature05744 (Apr. 2007).

144. Deisseroth, K. Optogenetics: 10 years of microbial opsins in neuroscience. *Nature neuroscience* **18,** 1213–1225. ISSN: 1097-6256. doi:10.1038/nn.4091 (Sept. 2015).

145. Nagel, G. *et al.* Light Activation of Channelrhodopsin-2 in Excitable Cells of Caenorhabditis elegans Triggers Rapid Behavioral Responses. en. *Current Biology* **15,** 2279–2284. ISSN: 0960-9822. doi:10.1016/j.cub.2005.11.032 (Dec. 2005).

146. Koopman, M., Janssen, L. & Nollen, E. A. A. An economical and highly adaptable optogenetics system for individual and population-level manipulation of Caenorhabditis elegans. *BMC Biology* **19,** 170. ISSN: 1741-7007. doi:10.1186/s12915-021-01085-2 (Aug. 2021).

147. Strickland, D. *et al.* TULIPs: Tunable, light-controlled interacting protein tags for cell biology. *Nature methods* **9,** 379–384. ISSN: 1548-7091. doi:10.1038/nmeth.1904 (Mar. 2012).

148. Harterink, M. *et al.* Light-controlled intracellular transport in Caenorhabditis elegans. en. *Current Biology* **26,** R153–R154. ISSN: 0960-9822. doi:10.1016/j.cub.2015.12.016 (Feb. 2016).

149. De Henau, S., Pagès-Gallego, M., Pannekoek, W.-J. & Dansen, T. B. Mitochondria-Derived H2O2 Promotes Symmetry Breaking of the C. elegans Zygote. en. *Developmental Cell* **53,** 263–271.e6. ISSN: 1534-5807. doi:10.1016/j.devcel.2020.03.008 (May 2020).

150. Lungu, O. I. *et al.* Designing photoswitchable peptides using the AsLOV2 domain. *Chemistry & biology* **19,** 507–517. ISSN: 1074-5521. doi:10.1016/j.chembiol.2012.02.006 (Apr. 2012).

151. Radman, I., Greiss, S. & Chin, J. W. Efficient and Rapid C. elegans Transgenesis by Bombardment and Hygromycin B Selection. en. *PLOS ONE* **8.** Publisher: Public Library of Science, e76019. ISSN: 1932-6203. doi:10.1371/journal.pone.0076019 (Oct. 2013).

152. Arribere, J. A. *et al.* Translation Readthrough Mitigation. *Nature* **534,** 719–723. ISSN: 0028-0836. doi:10.1038/nature18308 (June 2016).

153. Parrish, A. R. *et al.* Expanding the genetic code of Caenorhabditis elegans using bacterial aminoacyl-tRNA synthetase/tRNA pairs. eng. *ACS chemical biology* **7,** 1292–1302. ISSN: 1554-8937. doi:10.1021/cb200542j (July 2012).

154. Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (BONCAT). *Proceedings of the National Academy of Sciences of the United States of America* **103,** 9482–9487. ISSN: 0027-8424. doi:10.1073/pnas.0601637103 (June 2006).

155. Yuet, K. P. *et al.* Cell-specific proteomic analysis in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences of the United States of America* **112,** 2705–2710. ISSN: 0027-8424. doi:10.1073/pnas.1421567112 (Mar. 2015).

156. O'Shea, J. M. *et al.* Generation of Photocaged Nanobodies for Intracellular Applications in an Animal Using Genetic Code Expansion and Computationally Guided Protein Engineering**. en. *ChemBioChem* **n/a.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cbic.202200321, e202200321. ISSN: 1439-7633. doi:10.1002/cbic.202200321 (2022).

157. Olorunniji, F. J., Rosser, S. J. & Stark, W. M. Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochemical Journal* **473,** 673–684. ISSN: 0264-6021. doi:10.1042/BJ20151112 (Mar. 2016).

158. Prorocic, M. M. & Stark, W. M. en. in *Brenner's Encyclopedia of Genetics (Second Edition)* (eds Maloy, S. & Hughes, K.) 459–462 (Academic Press, San Diego, Jan. 2013). ISBN: 978-0-08-096156-9. doi:10.1016/B978-0-12-374984-0.01435-2.

159. Grindley, N. D. F., Whiteson, K. L. & Rice, P. A. Mechanisms of site-specific recombination. eng. *Annual Review of Biochemistry* **75,** 567–605. ISSN: 0066-4154. doi:10.1146/annurev.biochem.73.011303.073908 (2006).

160. Hatfull, G. F., Noble, S. M. & Grindley, N. D. F. The resolvase induces an unusual DNA structure at the recombinational crossover point. English. *Cell* **49.** Publisher: Elsevier, 103–110. ISSN: 0092-8674, 1097-4172. doi:10.1016/0092-8674(87)90760-4 (Apr. 1987).

161. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA Cloning Using In Vitro Site-Specific Recombination. *Genome Research* **10,** 1788–1795. ISSN: 1088-9051 (Nov. 2000).

162. Landy, A. & Ross, W. Viral Integration and Excision: Structure of the Lambda att Sites. *Science (New York, N.Y.)* **197,** 1147–1160. ISSN: 0036-8075 (Sept. 1977).

163. Atlung, T., Nielsen, A., Rasmussen, L. J., Nellemann, L. J. & Holm, F. A versatile method for integration of genes and gene fusions into the attachment site of Escherichia coli. en. *Gene* **107,** 11–17. ISSN: 0378-1119. doi:10.1016/0378-1119(91)90291-I (Oct. 1991).

164. O'Gorman, S., Fox, D. T. & Wahl, G. M. Recombinase-mediated gene activation and site-specific integration in mammalian cells. eng. *Science (New York, N.Y.)* **251,** 1351–1355. ISSN: 0036-8075. doi:10.1126/science.1900642 (Mar. 1991).

165. Fukushige, S. & Sauer, B. Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. eng. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 7905–7909. ISSN: 0027-8424. doi:10.1073/pnas.89.17.7905 (Sept. 1992).

166. Nance, J. & Frøkjær-Jensen, C. The Caenorhabditis elegans Transgenic Toolbox. *Genetics* **212,** 959–990. ISSN: 0016-6731. doi:10.1534/genetics.119.301506 (Aug. 2019).

167. Nonet, M. L. Efficient Transgenesis in Caenorhabditis elegans Using Flp Recombinase-Mediated Cassette Exchange. *Genetics* **215,** 903–921. ISSN: 0016-6731. doi:10.1534/genetics.120.303388 (Aug. 2020).

168. Davis, M. W., Morton, J. J., Carroll, D. & Jorgensen, E. M. Gene Activation Using FLP Recombinase in C. elegans. *PLoS Genetics* **4,** e1000028. ISSN: 1553-7390. doi:10.1371/journal.pgen.1000028 (Mar. 2008).

169. Muñoz-Jiménez, C. *et al.* An Efficient FLP-Based Toolkit for Spatiotemporal Control of Gene Expression in Caenorhabditis elegans. *Genetics* **206,** 1763–1778. ISSN: 0016-6731. doi:10.1534/genetics.117.201012 (Aug. 2017).

170. Flavell, S. W. *et al.* Serotonin and the Neuropeptide PDF Initiate and Extend Opposing Behavioral States in C. elegans. *Cell* **154,** 1023–1035. ISSN: 0092-8674. doi:10.1016/j.cell.2013.08.001 (Aug. 2013).

171. Oberdoerffer, P., Otipoby, K. L., Maruyama, M. & Rajewsky, K. Unidirectional Cre-mediated genetic inversion in mice using the mutant loxP pair lox66/lox71. *Nucleic Acids Research* **31,** e140. ISSN: 0305-1048. doi:10.1093/nar/gng140 (Nov. 2003).

172. Chakravarti, D., Caraballo, L. D., Weinberg, B. H. & Wong, W. W. Inducible Gene Switches with Memory in Human T Cells for Cellular Immunotherapy. *ACS Synthetic Biology* **8.** Publisher: American Chemical Society, 1744–1754. doi:10.1021/acssynbio.8b00512 (Aug. 2019).

173. Boniface, E. J., Lu, J., Vicotroff, T., Zhu, M. & Chen, W. FlEx-based transgenic reporter lines for visualization of Cre and Flp activity in live zebrafish. *Genesis (New York, N.Y. : 2000)* **47,** 10.1002/dvg.20526. ISSN: 1526-954X. doi:10.1002/dvg.20526 (July 2009).

174. Dupuy, D. *et al.* A First Version of the Caenorhabditis elegans Promoterome. en. *Genome Research* **14.** Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab, 2169–2175. ISSN: 1088-9051, 1549-5469. doi:10.1101/gr.2497604 (Oct. 2004).

175. White, J. Q. *et al.* The Sensory Circuitry for Sexual Attraction in C. elegans Males. en. *Current Biology* **17,** 1847–1857. ISSN: 0960-9822. doi:10.1016/j.cub.2007.09.011 (Nov. 2007).

176. Bacaj, T. & Shaham, S. Temporal Control of Cell-Specific Transgene Expression in Caenorhabditis elegans. *Genetics* **176,** 2651–2655. ISSN: 0016-6731. doi:10.1534/genetics.107.074369 (Aug. 2007).

177. Zevian, S. C. & Yanowitz, J. L. Methodological Considerations for Heat Shock of the Nematode Caenorhabditis elegans. *Methods (San Diego, Calif.)* **68,** 450–457. ISSN: 1046-2023. doi:10.1016/j.ymeth.2014.04.015 (Aug. 2014).

178. Ge, M.-H. *et al.* Dual Recombining-out System for Spatiotemporal Gene Expression in C. elegans. en. *iScience* **23,** 101567. ISSN: 2589-0042. doi:10.1016/j.isci.2020.101567 (Oct. 2020).

179. Hobert, O., Johnston, R. J. & Chang, S. Left–right asymmetry in the nervous system: the Caenorhabditis elegans model. en. *Nature Reviews Neuroscience* **3.** Number: 8 Publisher: Nature Publishing Group, 629–640. ISSN: 1471-0048. doi:10.1038/nrn897 (Aug. 2002).

180. Maluck, E. *et al.* A wake-active locomotion circuit depolarizes a sleep-active neuron to switch on sleep. *PLoS Biology* **18,** e3000361. ISSN: 1544-9173. doi:10.1371/journal.pbio.3000361 (Feb. 2020).

181. Churgin, M. A., He, L., Murray, J. I. & Fang-Yen, C. Efficient Single-Cell Transgene Induction in Caenorhabditis elegans Using a Pulsed Infrared Laser. *G3: Genes—Genomes—Genetics* **3,** 1827–1832. ISSN: 2160-1836. doi:10.1534/g3.113.007682 (Oct. 2013).

182. Churgin, M. A., He, L., Murray, J. I. & Fang-Yen, C. Construction of a system for single-cell transgene induction in Caenorhabditis elegans using a pulsed infrared laser. *Methods (San Diego, Calif.)* **68,** 431–436. ISSN: 1046-2023. doi:10.1016/j.ymeth.2014.05.001 (Aug. 2014).

183. Chen, C.-H., Chen, Y.-C., Jiang, H.-C., Chen, C.-K. & Pan, C.-L. Neuronal aging: learning from C. elegans. *Journal of Molecular Signaling* **8,** 14. ISSN: 1750-2187. doi:10.1186/1750-2187-8-14 (Dec. 2013).

184. Jung, H. *et al.* Noninvasive optical activation of Flp recombinase for genetic manipulation in deep mouse brain regions. en. *Nature Communications* **10.** Number: 1 Publisher: Nature Publishing Group, 314. ISSN: 2041-1723. doi:10.1038/s41467-018-08282-8 (Jan. 2019).

185. Kawano, F., Okazaki, R., Yazawa, M. & Sato, M. A photoactivatable Cre–loxP recombination system for optogenetic genome engineering. en. *Nature Chemical Biology* **12.** Number: 12 Publisher: Nature Publishing Group, 1059–1064. ISSN: 1552-4469. doi:10.1038/nchembio.2205 (Dec. 2016).

186. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. en. *Nature Communications* **6.** Number: 1 Publisher: Nature Publishing Group, 6256. ISSN: 2041-1723. doi:10.1038/ncomms7256 (Feb. 2015).

187. Edwards, W. F., Young, D. D. & Deiters, A. Light-Activated Cre Recombinase as a Tool for the Spatial and Temporal Control of Gene Function in Mammalian Cells. *ACS Chemical Biology* **4.** Publisher: American Chemical Society, 441–445. ISSN: 1554-8929. doi:10.1021/cb900041s (June 2009).

188. Luo, J. *et al.* Genetically encoded optical activation of DNA recombination in human cells †Electronic supplementary information (ESI) available: Experimental protocols. See DOI: 10.1039/c6cc03934k Click here for additional data file. *Chemical Communications (Cambridge, England)* **52,** 8529–8532. ISSN: 1359-7345. doi:10.1039/c6cc03934k (July 2016).

189. Brown, W., Liu, J., Tsang, M. & Deiters, A. Cell-Lineage Tracing in Zebrafish Embryos with an Expanded Genetic Code. en. *ChemBioChem* **19.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cbic.201800040, 1244–1249. ISSN: 1439-7633. doi:10.1002/cbic.201800040 (2018).

190. Luo, J. *et al.* Genetically Encoded Optochemical Probes for Simultaneous Fluorescence Reporting and Light Activation of Protein Function with Two-Photon Excitation. *Journal of the American Chemical Society* **136,** 15551–15558. ISSN: 0002-7863. doi:10.1021/ja5055862 (Nov. 2014).

191. Buchholz, F., Ringrose, L., Angrand, P.-O., Rossi, F. & Stewart, A. F. Different Thermostabilities of FLP and Cre Recombinases: Implications for Applied Site-Specific Recombination. *Nucleic Acids Research* **24,** 4256–4262. ISSN: 0305-1048. doi:10.1093/nar/24.21.4256 (Nov. 1996).

192. Hubbard, E. J. A. FLP/FRT and Cre/lox recombination technology in C. elegans. eng. *Methods (San Diego, Calif.)* **68,** 417–424. ISSN: 1095-9130. doi:10.1016/j.ymeth.2014.05.007 (Aug. 2014).

193. Schmidt-Supprian, M. & Rajewsky, K. Vagaries of conditional gene targeting. en. *Nature Immunology* **8.** Number: 7 Publisher: Nature Publishing Group, 665–668. ISSN: 1529-2916. doi:10.1038/ni0707-665 (July 2007).

194. Nern, A., Pfeiffer, B. D., Svoboda, K. & Rubin, G. M. Multiple new site-specific recombinases for use in manipulating animal genomes. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 14198–14203. ISSN: 0027-8424. doi:10.1073/pnas.1111704108 (Aug. 2011).

195. Schmitt, C., Schultheis, C., Husson, S. J., Liewald, J. F. & Gottschalk, A. Specific Expression of Channelrhodopsin-2 in Single Neurons of Caenorhabditis elegans. *PLoS ONE* **7,** e43164. ISSN: 1932-6203. doi:10.1371/journal.pone.0043164 (Aug. 2012).

196. Strope, P. K. *et al.* 2µ plasmid in Saccharomyces species and in Saccharomyces cerevisiae. *FEMS Yeast Research* **15,** fov090. ISSN: 1567-1356. doi:10.1093/femsyr/fov090 (Dec. 2015).

197. Chan, K.-M., Liu, Y.-T., Ma, C.-H., Jayaram, M. & Sau, S. The 2 micron plasmid of Saccharomyces cerevisiae: A miniaturized selfish genome with optimized functional competence. en. *Plasmid. Special Issue based on the International Society for Plasmid Biology Meeting: Santander 2012* **70,** 2–17. ISSN: 0147-619X. doi:10.1016/j.plasmid.2013.03.001 (July 2013).

198. Zhu, X.-D. & Sadowski, P. Cleavage-dependent ligation by the FLP recombinase: Characterization of a mutant FLP protein with an alteration in a catalytic amino acid. *The Journal of biological chemistry* **270,** 23044–54. doi:10.1074/jbc.270.39.23044 (Oct. 1995).

199. Senecoff, J. F. & Cox, M. M. Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. en. *Journal of Biological Chemistry* **261,** 7380–7386. ISSN: 0021-9258. doi:10.1016/S0021-9258(17)38402-8 (June 1986).

200. Vetcher, A. *et al.* DNA Topology and Geometry in Flp and Cre Recombination. *Journal of molecular biology* **357,** 1089–104. doi:10.1016/j.jmb.2006.01.037 (May 2006).

201. Chen, Y., Narendra, U., Iype, L. E., Cox, M. M. & Rice, P. A. Crystal Structure of a Flp Recombinase–Holliday Junction Complex: Assembly of an Active Oligomer by Helix Swapping. en. *Molecular Cell* **6,** 885–897. ISSN: 1097-2765. doi:10.1016/S1097-2765(05)00088-2 (Oct. 2000).

202. Jayaram, M. *et al.* An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective. *Microbiology Spectrum* **3.** Publisher: American Society for Microbiology, 3.4.12. doi:10.1128/microbiolspec.MDNA3-0021-2014 (July 2015).

203. Chen, Y. & Rice, P. A. The Role of the Conserved Trp330 in Flp-mediated Recombination: FUNCTIONAL AND STRUCTURAL ANALYSIS*. en. *Journal of Biological Chemistry* **278,** 24800–24807. ISSN: 0021-9258. doi:10.1074/jbc.M300853200 (July 2003).

204. Gautier, A. *et al.* Genetically Encoded Photocontrol of Protein Localization in Mammalian Cells. *Journal of the American Chemical Society* **132.** Publisher: American Chemical Society, 4086–4088. ISSN: 0002-7863. doi:10.1021/ja910688s (Mar. 2010).

205. Klapoetke, N. C. *et al.* Independent Optical Excitation of Distinct Neural Populations. *Nature methods* **11,** 338–346. ISSN: 1548-7091. doi:10.1038/nmeth.2836 (Mar. 2014).

206. Schild, L. C. & Glauser, D. A. Dual Color Neural Activation and Behavior Control with Chrimson and CoChR in Caenorhabditis elegans. *Genetics* **200,** 1029–1034. ISSN: 0016-6731. doi:10.1534/genetics.115.177956 (Aug. 2015).

207. Lee, K. H. & Aschner, M. A Simple Light Stimulation of Caenorhabditis elegans. *Current protocols in toxicology / editorial board, Mahin D. Maines (editor-in-chief) ... [et al.]* **67,** 11.21.1–11.21.5. ISSN: 1934-9254. doi:10.1002/0471140856.tx1121s67 (Feb. 2016).

208. Macosko, E. Z. *et al.* A hub-and-spoke circuit drives pheromone attraction and social behaviour in C. elegans. en. *Nature* **458.** Number: 7242 Publisher: Nature Publishing Group, 1171–1175. ISSN: 1476-4687. doi:10.1038/nature07886 (Apr. 2009).

209. Ruaud, A.-F., Katic, I. & Bessereau, J.-L. Insulin/Insulin-like growth factor signaling controls non-Dauer developmental speed in the nematode Caenorhabditis elegans. eng. *Genetics* **187,** 337–343. ISSN: 1943-2631. doi:10.1534/genetics.110.123323 (Jan. 2011).

210. Goodman, M. Mechanosensation. *WormBook.* ISSN: 15518507. doi:10.1895/wormbook.1.62.1 (2006).

211. Snapp, E. Design and Use of Fluorescent Fusion Proteins in Cell Biology. *Current protocols in cell biology / editorial board, Juan S. Bonifacino ... [et al.]* **CHAPTER,** Unit–21.4. ISSN: 1934-2500. doi:10.1002/0471143030.cb2104s27 (July 2005).

212. Blumenthal, T. *Trans-splicing and operons in C. elegans* en. Publication Title: WormBook: The Online Review of C. elegans Biology [Internet] (WormBook, 2018).

213. Ryan, M. D., King, A. M. & Thomas, G. P. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. eng. *The Journal of General Virology* **72 (Pt 11),** 2727–2732. ISSN: 0022-1317. doi:10.1099/0022-1317-72-11-2727 (Nov. 1991).

214. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. en. *Scientific Reports* **7.** Number: 1 Publisher: Nature Publishing Group, 2193. ISSN: 2045-2322. doi:10.1038/s41598-017-02460-2 (May 2017).

215. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. en. *Journal of Molecular Biology* **3,** 318–356. ISSN: 0022-2836. doi:10.1016/S0022-2836(61)80072-7 (June 1961).

216. Hurst, L. D., Pál, C. & Lercher, M. J. The evolutionary dynamics of eukaryotic gene order. en. *Nature Reviews Genetics* **5.** Number: 4 Publisher: Nature Publishing Group, 299–310. ISSN: 1471-0064. doi:10.1038/nrg1319 (Apr. 2004).

217. Osbourn, A. E. & Field, B. Operons. *Cellular and Molecular Life Sciences* **66,** 3755–3775. ISSN: 1420-682X. doi:10.1007/s00018-009-0114-3 (Dec. 2009).

218. Blumenthal, T. *et al.* A global analysis of Caenorhabditis elegans operons. en. *Nature* **417.** Number: 6891 Publisher: Nature Publishing Group, 851–854. ISSN: 1476-4687. doi:10.1038/nature00831 (June 2002).

219. Blumenthal, T. & Gleason, K. S. Caenorhabditis elegans operons: form and function. en. *Nature Reviews Genetics* **4.** Number: 2 Publisher: Nature Publishing Group, 110–118. ISSN: 1471-0064. doi:10.1038/nrg995 (Feb. 2003).

220. Blumenthal, T., Davis, P. & Garrido-Lecca, A. *Operon and non-operon gene clusters in the C. elegans genome* en. Publication Title: WormBook: The Online Review of C. elegans Biology [Internet] (WormBook, 2018).

221. Bitar, M., Boroni, M., Macedo, A., Machado, C. R. & Franco, G. The spliced leader trans-splicing mechanism in different organisms: molecular details and possible biological roles. *Frontiers in Genetics* **4.** ISSN: 1664-8021 (2013).

222. Blumenthal, T. Operons in eukaryotes. *Briefings in Functional Genomics* **3,** 199–211. ISSN: 2041-2649. doi:10.1093/bfgp/3.3.199 (Nov. 2004).

223. Dossena, S. *et al.* Use of the Operon Structure of the C. elegans Genome as a Tool to Identify Functionally Related Proteins. *Cellular Physiology and Biochemistry* **32.** Publisher: Karger Publishers, 41–56. ISSN: 1015-8987, 1421-9778. doi:10.1159/000356623 (2013).

224. Fürst, J. *et al.* ICln Ion Channel Splice Variants in Caenorhabditis elegans: VOLTAGE DEPENDENCE AND INTERACTION WITH AN OPERON PARTNER PROTEIN*. en. *Journal of Biological Chemistry* **277,** 4435–4445. ISSN: 0021-9258. doi:10.1074/jbc.M107372200 (Feb. 2002).

225. Cutter, A. D., Dey, A. & Murray, R. L. Evolution of the Caenorhabditis elegans Genome. *Molecular Biology and Evolution* **26,** 1199–1234. ISSN: 0737-4038. doi:10.1093/molbev/msp048 (June 2009).

226. Byerly, L., Cassada, R. C. & Russell, R. L. The life cycle of the nematode Caenorhabditis elegans. I. Wild-type growth and reproduction. eng. *Developmental Biology* **51,** 23–33. ISSN: 0012-1606. doi:10.1016/0012-1606(76)90119-6 (July 1976).

227. Logie, C. & Stewart, A. F. Ligand-regulated site-specific recombination. eng. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 5940–5944. ISSN: 0027-8424. doi:10.1073/pnas.92.13.5940 (June 1995).

228. Guo, X., Zhang, B., He, Y., Liu, Y. & Tian, C. Improved chemical synthesis of o-nirtrobenzyl-tyrosine for concise site-specific 15N-tyrosine NMR analysis demonstrated by plant ABA receptor PYL10. en. *Tetrahedron Letters* **58,** 3764–3767. ISSN: 0040-4039. doi:10.1016/j.tetlet.2017.08.034 (Sept. 2017).

229. Baumann, T. *et al.* Computational Aminoacyl-tRNA Synthetase Library Design for Photocaged Tyrosine. *International Journal of Molecular Sciences* **20,** 2343. ISSN: 1422-0067. doi:10.3390/ijms20092343 (May 2019).

230. Brockie, P. J., Madsen, D. M., Zheng, Y., Mellem, J. & Maricq, A. V. Differential Expression of Glutamate Receptor Subunits in the Nervous System of Caenorhabditis elegans and Their Regulation by the Homeodomain Protein UNC-42. *The Journal of Neuroscience* **21,** 1510–1522. ISSN: 0270-6474. doi:10.1523/JNEUROSCI.21-05-01510.2001 (Mar. 2001).

231. Gjorgjieva, J., Biron, D. & Haspel, G. Neurobiology of Caenorhabditis elegans Locomotion: Where Do We Stand? *Bioscience* **64,** 476–486. ISSN: 0006-3568. doi:10.1093/biosci/biu058 (June 2014).

232. Chalfie, M. *et al.* The neural circuit for touch sensitivity in Caenorhabditis elegans. *The Journal of Neuroscience* **5,** 956–964. ISSN: 0270-6474. doi:10.1523/JNEUROSCI.05-04-00956.1985 (Apr. 1985).

233. Stirman, J. N. *et al.* Real-time multimodal optical control of neurons and muscles in freely-behaving Caenorhabditis elegans. *Nature methods* **8,** 153–158. ISSN: 1548-7091. doi:10.1038/nmeth.1555 (Feb. 2011).

234. Lim, M. A. *et al.* Neuroendocrine modulation sustains the C. elegans forward motor state. *eLife* **5,** e19887. ISSN: 2050-084X. doi:10.7554/eLife.19887 (2016).

235. Li, W., Kang, L., Piggott, B. J., Feng, Z. & Shawn Xu, X. Z. The neural circuits and sensory channels mediating harsh touch sensation in C. elegans. *Nature communications* **2,** 315. ISSN: 2041-1723. doi:10.1038/ncomms1308 (May 2011).

236. Cook, S. J. *et al.* Whole-animal connectomes of both Caenorhabditis elegans sexes. *Nature* **571,** 63–71. ISSN: 0028-0836. doi:10.1038/s41586-019-1352-7 (July 2019).

237. Urmersbach, B., Besseling, J., Spies, J.-P. & Bringmann, H. Automated analysis of sleep control via a single neuron active at sleep onset in C. elegans. eng. *Genesis (New York, N.Y.: 2000)* **54,** 212–219. ISSN: 1526-968X. doi:10.1002/dvg.22924 (Apr. 2016).

238. Busack, I., Jordan, F., Sapir, P. & Bringmann, H. The OptoGenBox – a device for long-term optogenetics in C. elegans. *Journal of Neurogenetics* **34.** Publisher: Taylor & Francis _eprint:

https://doi.org/10.1080/01677063.2020.1776709, 466–474. ISSN: 0167-7063. doi:10.1080/01677063.2020.1776709 (Oct. 2020).

239. Davis, L. & Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. en. *Nature Reviews Molecular Cell Biology* **13.** Number: 3 Publisher: Nature Publishing Group, 168–182. ISSN: 1471-0080. doi:10.1038/nrm3286 (Mar. 2012).

240. Wang, J. *et al.* A novel family of tyrosine integrases encoded by the temperate pleolipovirus SNJ2. *Nucleic Acids Research* **46,** 2521–2536. ISSN: 0305-1048. doi:10.1093/nar/gky005 (Mar. 2018).

241. Keenholtz, R. A. *et al.* Arginine as a General Acid Catalyst in Serine Recombinase-mediated DNA Cleavage. *The Journal of Biological Chemistry* **288,** 29206–29214. ISSN: 0021-9258. doi:10.1074/jbc.M113.508028 (Oct. 2013).

242. Lemke, E. A., Summerer, D., Geierstanger, B. H., Brittain, S. M. & Schultz, P. G. Control of protein phosphorylation with a genetically encoded photocaged amino acid. en. *Nature Chemical Biology* **3.** Number: 12 Publisher: Nature Publishing Group, 769–772. ISSN: 1552-4469. doi:10.1038/nchembio.2007.44 (Dec. 2007).

243. Spieth, J., Brooke, G., Kuersten, S., Lea, K. & Blumenthal, T. Operons in C. elegans: Polycistronic mRNA precursors are processed by trans-splicing of SL2 to downstream coding regions. en. *Cell* **73,** 521–532. ISSN: 0092-8674. doi:10.1016/0092-8674(93)90139-H (May 1993).

244. González-Aguilera, C., Palladino, F. & Askjaer, P. C. elegans epigenetic regulation in development and aging. *Briefings in Functional Genomics* **13,** 223–234. ISSN: 2041-2649. doi:10.1093/bfgp/elt048 (May 2014).

245. Unhavaithaya, Y. *et al.* MEP-1 and a Homolog of the NURD Complex Component Mi-2 Act Together to Maintain Germline-Soma Distinctions in C. elegans. en. *Cell* **111,** 991–1002. ISSN: 0092-8674. doi:10.1016/S0092-8674(02)01202-3 (Dec. 2002).

246. DasGupta, A., Lee, T. L., Li, C. & Saltzman, A. L. Emerging Roles for Chromo Domain Proteins in Genome Organization and Cell Fate in C. elegans. *Frontiers in Cell and Developmental Biology* **8.** ISSN: 2296-634X (2020).

247. Kutscher, L. M. & Shaham, S. Forward and reverse mutagenesis in C. elegans. *WormBook,* 1–26. doi:10.1895/wormbook.1.167.1 (Jan. 2014).

248. De Stasio, E. A. & Dorman, S. Optimization of ENU mutagenesis of Caenorhabditis elegans. en. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **495,** 81–88. ISSN: 1383-5718. doi:10.1016/S1383-5718(01)00198-X (Aug. 2001).

249. Barstead, R. J. & Moerman, D. G. en. in *C. elegans: Methods and Applications* (ed Strange, K.) 51–58 (Humana Press, Totowa, NJ, 2006). ISBN: 978-1-59745-151-2. doi:10.1385/1-59745-151-7:51.

250. Stewart, H. I., Rosenbluth, R. E. & Baillie, D. L. Most ultraviolet irradiation induced mutations in the nematode Caenorhabditis elegans are chromosomal rearrangements. en. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **249,** 37–54. ISSN: 0027-5107. doi:10.1016/0027-5107(91)90131-7 (July 1991).

251. Babu, P. & Brenner, S. Spectrum of 32P-induced mutants of Caenorbabditis elegans. en. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **82,** 269–273. ISSN: 0027-5107. doi:10.1016/0027-5107(81)90156-1 (July 1981).

252. Boulin, T. & Bessereau, J.-L. Mos1-mediated insertional mutagenesis in Caenorhabditis elegans. en. *Nature Protocols* **2.** Number: 5 Publisher: Nature Publishing Group, 1276–1287. ISSN: 1750-2799. doi:10.1038/nprot.2007.192 (May 2007).

253. Lesa, G. M. Isolation of Caenorhabditis elegans gene knockouts by PCR screening of chemically mutagenized libraries. en. *Nature Protocols* **1.** Number: 5 Publisher: Nature Publishing Group, 2231–2240. ISSN: 1750-2799. doi:10.1038/nprot.2006.345 (Dec. 2006).

254. Thompson, O. *et al.* The million mutation project: A new approach to genetics in Caenorhabditis elegans. *Genome Research* **23,** 1749–1762. ISSN: 1088-9051. doi:10.1101/gr.157651.113 (Oct. 2013).

255. Spieth, J. Overview of gene structure. *WormBook.* ISSN: 15518507. doi:10.1895/wormbook.1.65.1 (2006).

256. Wood, A. J. *et al.* Targeted Genome Editing Across Species Using ZFNs and TALENs. *Science* **333.** Publisher: American Association for the Advancement of Science, 307–307. doi:10.1126/science.1207773 (July 2011).

257. Frøkjær-Jensen, C. Exciting Prospects for Precise Engineering of Caenorhabditis elegans Genomes with CRISPR/Cas9. *Genetics* **195,** 635–642. ISSN: 0016-6731. doi:10.1534/genetics.113.156521 (Nov. 2013).

258. Waaijers, S. *et al.* CRISPR/Cas9-Targeted Mutagenesis in Caenorhabditis elegans. *Genetics* **195,** 1187–1191. ISSN: 0016-6731. doi:10.1534/genetics.113.156299 (Nov. 2013).

259. Paix, A., Folkmann, A., Rasoloson, D. & Seydoux, G. High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes. *Genetics* **201,** 47–54. ISSN: 0016-6731. doi:10.1534/genetics.115.179382 (Sept. 2015).

260. Dickinson, D. J., Ward, J. D., Reiner, D. J. & Goldstein, B. Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination. eng. *Nature Methods* **10,** 1028–1034. ISSN: 1548-7105. doi:10.1038/nmeth.2641 (Oct. 2013).

261. Noma, K. & Jin, Y. Optogenetic mutagenesis in Caenorhabditis elegans. *Nature Communications* **6,** 8868–8868. doi:10.1038/ncomms9868 (Dec. 2015).

262. Halperin, S. O. *et al.* CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature* **560,** 248–252. doi:10.1038/s41586-018-0384-8 (Aug. 2018).

263. Hess, G. T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature methods* **13,** 1036–1036. doi:10.1038/NMETH.4038 (2016).

264. Shu, X. *et al.* A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS biology* **9,** e1001041–e1001041. doi:10.1371/journal.pbio.1001041 (Apr. 2011).

265. Takemiya, A. & Shimazaki, K.-i. Arabidopsis phot1 and phot2 phosphorylate BLUS1 kinase with different efficiencies in stomatal opening. en. *Journal of Plant Research* **129,** 167–174. ISSN: 1618-0860. doi:10.1007/s10265-015-0780-1 (Mar. 2016).

266. Sies, H. & Menck, C. F. M. Singlet oxygen induced DNA damage. en. *Mutation Research/DNAging* **275,** 367–375. ISSN: 0921-8734. doi:10.1016/0921-8734(92)90039-R (Sept. 1992).

267. Qi, Y. B., Garren, E. J., Shu, X., Tsien, R. Y. & Jin, Y. Photo-inducible cell ablation in Caenorhabditis elegans using the genetically encoded singlet oxygen generating protein miniSOG. eng. *Proceedings of the National Academy of Sciences of the United States of America* **109,** 7499–7504. ISSN: 1091-6490. doi:10.1073/pnas.1204096109 (May 2012).

268. Xu, S. & Chisholm, A. D. Highly efficient optogenetic cell ablation in C. elegans using membrane-targeted miniSOG. en. *Scientific Reports* **6.** Number: 1 Publisher: Nature Publishing Group, 21271. ISSN: 2045-2322. doi:10.1038/srep21271 (Feb. 2016).

269. Flibotte, S. *et al.* Whole-Genome Profiling of Mutagenesis in Caenorhabditis elegans. *Genetics* **185,** 431–441. ISSN: 0016-6731. doi:10.1534/genetics.110.116616 (June 2010).

270. De Stasio, E. *et al.* Characterization of Revertants of Unc-93(e1500) in Caenorhabditis Elegans Induced by N-Ethyl-N-Nitrosourea. en. *Genetics* **147.** Publisher: Oxford University Press, 597. doi:10.1093/genetics/147.2.597 (Oct. 1997).

271. Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: Biology, mechanisms and applications. en. *Biochimie. Special Issue: Regulatory RNAs* **117,** 119–128. ISSN: 0300-9084. doi:10.1016/j.biochi.2015.03.025 (Oct. 2015).

272. Adli, M. The CRISPR tool kit for genome editing and beyond. en. *Nature Communications* **9.** Number: 1 Publisher: Nature Publishing Group, 1911. ISSN: 2041-1723. doi:10.1038/s41467-018-04252-2 (May 2018).

273. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* **41,** 7429–37. doi:10.1093/nar/gkt520 (Aug. 2013).

274. Kelly, R. B., Cozzarelli, N. R., Deutscher, M. P., Lehman, I. R. & Kornberg, A. Enzymatic Synthesis of Deoxyribonucleic Acid: XXXII. REPLICATION OF DUPLEX DEOXYRIBONUCLEIC ACID BY POLYMERASE AT A SINGLE STRAND BREAK. en. *Journal of Biological Chemistry* **245,** 39–45. ISSN: 0021-9258. doi:10.1016/S0021-9258(18)63419-2 (Jan. 1970).

275. Patel, P. H., Kawate, H., Adman, E., Ashbach, M. & Loeb, L. A. A Single Highly Mutable Catalytic Site Amino Acid Is Critical for DNA Polymerase Fidelity*. en. *Journal of Biological Chemistry* **276,** 5044–5051. ISSN: 0021-9258. doi:10.1074/jbc.M008701200 (Feb. 2001).

276. Derbyshire, V. *et al.* Genetic and crystallographic studies of the 3',5'-exonucleolytic site of DNA polymerase I. eng. *Science (New York, N.Y.)* **240,** 199–201. ISSN: 0036-8075. doi:10.1126/science.2832946 (Apr. 1988).

277. Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **100,** 9727–9732. ISSN: 0027-8424. doi:10.1073/pnas.1333928100 (Aug. 2003).

278. Smith, C. J. *et al.* Enabling large-scale genome editing at repetitive elements by reducing DNA nicking. *Nucleic Acids Research* **48,** 5183–5195. ISSN: 0305-1048. doi:10.1093/nar/gkaa239 (May 2020).

279. Patel, P. H., Suzuki, M., Adman, E., Shinkai, A. & Loeb, L. A. Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection11Edited by J. H. Miller. en. *Journal of Molecular Biology* **308,** 823–837. ISSN: 0022-2836. doi:10.1006/jmbi.2001.4619 (May 2001).

280. Papavasiliou, F. N. & Schatz, D. G. Somatic Hypermutation of Immunoglobulin Genes: Merging Mechanisms for Genetic Diversity. English. *Cell* **109.** Publisher: Elsevier, S35–S44. ISSN: 0092-8674, 1097-4172. doi:10.1016/S0092-8674(02)00706-7 (Apr. 2002).

281. Lans, H. & Vermeulen, W. Nucleotide Excision Repair in Caenorhabditis elegans. *Molecular Biology International* **2011.** ISSN: 2090-2182. doi:10.4061/2011/542795 (2011).

282. Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A. & Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *Journal of the American Chemical Society* **137,** 5642–5645. ISSN: 0002-7863. doi:10.1021/ja512664v (May 2015).

283. Zhu, X. *et al.* Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. eng. *Nature Structural & Molecular Biology* **26,** 679–685. ISSN: 1545-9985. doi:10.1038/s41594-019-0258-2 (2019).

284. Nishimasu, H. *et al.* Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. English. *Cell* **156.** Publisher: Elsevier, 935–949. ISSN: 0092-8674, 1097-4172. doi:10.1016/j.cell.2014.02.001 (Feb. 2014).

285. Merritt, C., Rasoloson, D., Ko, D. & Seydoux, G. 3' UTRs are the primary regulators of gene expression in the C. elegans germline. *Current biology : CB* **18,** 1476–1482. ISSN: 0960-9822. doi:10.1016/j.cub.2008.08.013 (Oct. 2008).

286. Sagi, D. *et al.* Tissue- and Time-Specific Expression of Otherwise Identical tRNA Genes. *PLoS Genetics* **12,** e1006264. ISSN: 1553-7390. doi:10.1371/journal.pgen.1006264 (Aug. 2016).

287. Liu, P. *et al.* Heritable/conditional genome editing in C. elegans using a CRISPR-Cas9 feeding system. *Cell research* **24,** 886–9. doi:10.1038/cr.2014.73 (July 2014).

288. Kelly, W. G., Xu, S., Montgomery, M. K. & Fire, A. Distinct Requirements for Somatic and Germline Expression of a Generally Expressed Caernorhabditis Elegans Gene. *Genetics* **146,** 227–238. ISSN: 0016-6731 (May 1997).

289. Praitis, V., Casey, E., Collar, D. & Austin, J. Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. *Genetics* **157,** 1217–1226. ISSN: 0016-6731 (Mar. 2001).

290. Zeiser, E., Frøkjær-Jensen, C., Jorgensen, E. & Ahringer, J. MosSCI and Gateway Compatible Plasmid Toolkit for Constitutive and Inducible Expression of Transgenes in the C. elegans Germline. en. *PLOS ONE* **6.** Publisher: Public Library of Science, e20082. ISSN: 1932-6203. doi:10.1371/journal.pone.0020082 (May 2011).

291. Aljohani, M. D., El Mouridi, S., Priyadarshini, M., Vargas-Velazquez, A. M. & Frøkjær-Jensen, C. Engineering rules that minimize germline silencing of transgenes in simple extrachromosomal arrays in C. elegans. en. *Nature Communications* **11.** Number: 1 Publisher: Nature Publishing Group, 6300. ISSN: 2041-1723. doi:10.1038/s41467-020-19898-0 (Dec. 2020).

292. Wu, W.-S. *et al.* pirScan: a webserver to predict piRNA targeting sites and to avoid transgene silencing in C. elegans. eng. *Nucleic Acids Research* **46,** W43–W48. ISSN: 1362-4962. doi:10.1093/nar/gky277 (2018).

293. Frøkjær-Jensen, C. *et al.* An abundant class of non-coding DNA can prevent stochastic gene silencing in the C. elegans germline. *Cell* **166,** 343–357. ISSN: 0092-8674. doi:10.1016/j.cell.2016.05.072 (July 2016).

294. Fire, A., Alcazar, R. & Tan, F. Unusual DNA Structures Associated With Germline Genetic Activity in Caenorhabditis elegans. *Genetics* **173,** 1259–1273. ISSN: 0016-6731. doi:10.1534/genetics.106.057364 (July 2006).

295. Shin, T. H. & Mello, C. C. Chromatin regulation during C. elegans germline development. en. *Current Opinion in Genetics & Development* **13,** 455–462. ISSN: 0959-437X. doi:10.1016/S0959-437X(03)00109-6 (Oct. 2003).

296. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. eng. *Cell* **159,** 635–646. ISSN: 1097-4172. doi:10.1016/j.cell.2014.09.039 (Oct. 2014).

297. Grishok, A., Sinskey, J. L. & Sharp, P. A. Transcriptional silencing of a transgene by RNAi in the soma of C. elegans. *Genes & Development* **19,** 683–696. ISSN: 0890-9369. doi:10.1101/gad.1247705 (Mar. 2005).

298. Long, L. *et al.* Regulation of transcriptionally active genes via the catalytically inactive Cas9 in C. elegans and D. rerio. *Cell Research* **25,** 638–641. ISSN: 1001-0602. doi:10.1038/cr.2015.35 (May 2015).

299. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. eng. *Nature Biotechnology* **33,** 510–517. ISSN: 1546-1696. doi:10.1038/nbt.3199 (May 2015).

300. Xiong, T. *et al.* Targeted DNA methylation in human cells using engineered dCas9-methyltransferases. en. *Scientific Reports* **7,** 1–14. ISSN: 2045-2322. doi:10.1038/s41598-017-06757-0 (July 2017).

301. Redemann, S. *et al.* Codon adaptation–based control of protein expression in C. elegans. en. *Nature Methods* **8,** 250–252. ISSN: 1548-7105. doi:10.1038/nmeth.1565 (Mar. 2011).

302. Stiernagle, T. Maintenance of C. elegans. *WormBook.* ISSN: 15518507. doi:10.1895/wormbook.1.101.1 (2006).

303. Turek, M., Besseling, J. & Bringmann, H. Agarose Microchambers for Long-term Calcium Imaging of Caenorhabditis elegans. eng. *Journal of Visualized Experiments: JoVE,* e52742. ISSN: 1940-087X. doi:10.3791/52742 (June 2015).

304. Frøkjær-Jensen, C. *et al.* Single-copy insertion of transgenes in Caenorhabditis elegans. en. *Nature Genetics* **40.** Number: 11 Publisher: Nature Publishing Group, 1375–1383. ISSN: 1546-1718. doi:10.1038/ng.248 (Nov. 2008).