Application of fluorescent markers onto CRISPR-associated proteins

Name: Lochlan Parr

Supervisors: Dr. Ed Bolt - University of Nottingham

Dr. Stephen Harding - University of Nottingham

Company: University of Nottingham

Key Words: CRISPR, Cas1, Cas2, Fluorescent Protein, PCR, SDM, Mutagenesis.

Report/Research paper submitted to The University of Nottingham in partial

fulfilment of the requirements for the degree of Master of Research in Biomolecular Technology.

Total length of abstract: 200 words

Total length of report: 3390 words

Contents

Abstract

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems mediate prokaryotic adaptive immunity against mobile genetic elements (MGEs), where MGE sequences are captured and stored as spacers in CRISPR loci. CRISPR-associated (Cas) proteins Cas1 and Cas2 form a complex that integrates these spacers into CRISPR arrays, generating adaptive immunity through CRISPR interference. While the mechanisms of the Cas1-Cas2 complex are understood, less is known about its regulatory pathways, particularly because it has been shown to interact with proteins far outside its genetic locality, sometimes from different chromosomes. Fluorescent proteins (FPs) have successfully been used to tag proteins for over a century, enhancing our understanding of their intermolecular interactions. Site directed mutagenesis (SDM) can be used to substitute all Cas1 cysteine residues for alternative inert amino acids, and implementation of another cysteine at a more accessible site would allow the binding of fluorescent maleimide proteins, accommodating advanced microscopy techniques to investigate molecular interactions *in vivo*, thereby improving our understanding of Cas1 regulatory interactions. The methods to produce the target mutant Cas1 protein are discussed in this review, including techniques such as Sanger sequencing, pilot protein overexpression and spacer acquisition assays, which are used to verify the quality and viability of the product.

1. Introduction

The clustered regularly interspaced short palindromic repeat (CRISPR) is a segment of DNA which, alongside a host of CRISPR-associated (Cas) proteins, is involved in RNA-guided prokaryotic defence against invading foreign nucleic acids or mobile genetic elements (MGEs) from plasmids and phages. The array is made up of identical short (~20-50bp) repeat sequences

interspaced by similarly short variable spacer sequences, derived by acquisition and integration of foreign DNA fragments.

New spacers are captured from invader DNA by the highly conserved Cas1-Cas2 complex, which then inserts it into the CRISPR locus. CRISPR locus transcription and processing produces short mature CRISPR RNAs (crRNAs) which form Cas protein-crRNA complexes. These complexes recognize and precisely degrade invading nucleic acids complementary to the crRNA in interference reactions, creating adaptive immunity. The molecular mechanisms of these steps are now well characterised, both in molecular and functional terms, however Cas1-Cas2 acquisition is poorly understood. Researchers await detailed analysis, particularly regarding allosteric regulation. The complex has been shown to interact with a variety of proteins with loci far from Cas1, including different chromosomes, which adds challenge to the investigation. One way that the complex can be studied is with the use of a fluorescent marker to image its movement around the cell under different conditions.

2. CRISPR-Cas

In humans, pathogens and infected cells are destroyed by an incredibly complex multicellular immune system, where adaptive immunity is mounted against foreign proteins and other structures recognised by B and T cells. The demonstration of adaptive immunity by prokaryotes was therefore surprising (Barrangou et al., 2007), since it must necessarily be unicellular. Prokaryotes utilise a standalone system based on a DNA region called Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), which protects against viruses and other mobile genetic elements (MGEs) by targeting DNA and RNA.

First observed in *Escherichia coli*, CRISPR has been found in both the chromosomal and plasmid DNA in around 84% of archaea and 45% of bacteria (Ishino et al., 1987; Alberts et al., 2002; Pourcel et al., 2020) – it's worth noting that this difference in prevalence is likely affected by sampling bias, as almost twenty times more bacteria have been analysed than archaea.

CRISPR activity is accomplished through CRISPR-associated (*Cas*) proteins, with genes usually found adjacent to the CRISPR (Ishino et al., 1987). Descendants benefit from spacer acquisition since the chromosomal genome is modified, creating a chromosomal record of viruses the cell and its ancestors have encountered and become resistant against.

CRISPR-Cas mediated immunity can be divided into three stages; adaptation, expression, and interference (Jackson et al., 2017; Pickar-Oliver & Gersbach, 2019).

The first stage, adaptation, leads to insertion of new spacers in the CRISPR locus. In the second stage, expression, the system gets ready for action by expressing the *cas* genes and transcribing the CRISPR into a long precursor CRISPR RNA (pre-crRNA). The pre-crRNA is subsequently processed into mature crRNA by Cas proteins and accessory factors. In the third and last stage, interference, target nucleic acid is recognized and destroyed by the combined action of crRNA and Cas proteins.

3. Fluorescent Markers

The desire to improve our understanding of biological systems has motivated the development of methods to detect and identify biomolecules. The advent of fluorescent protein technology has paved the way for the study of molecular structure and interactions with high spatial and temporal resolution. Previously, low-resolution, unstable and less easily detectable radioisotopes were used to detect and identify molecular compounds (Maloth et al., 2014).

Biomolecules are now more commonly labelled by the use of safer fluorescent dyes or proteins as tags or 'probes.'

3.1 A Brief History of Fluorescent Markers

In 1962 green fluorescent protein (GFP) was discovered in *Aequorea* jellyfish by Shimomura et al while investigating the chemiluminescent protein aequorin. Just two months later the team noted that its emission spectrum peak of 508nm was close to the peak luminescence wavelength of living *Aequorea* tissue, while pure chemiluminescent aequorin peaked at 470nm, a blue colour close to GFP's excitation peak (Johnson et al, 1962). After witnessing the same colour shift in related coelenterates in 1971, Morin and Hastings were the first to suggest in vivo GFP excitation by radiationless energy transfer, which Morise et al verified three years later (Morise et al., 1974). Critically, Chalfie et al and Inouye & Tsuji showed that when the gene cloned by Prasher et al in 1992 was expressed in other organisms, fluorescence was still achieved (Prasher et al., 1992).

3.2 Thiol-Reactive Probe Labelling

Biothiols are biological compounds containing a sulfhydryl group (-SH) and can be categorized into two classes; protein thiols and non-protein thiols (Little & O'Brien, 1968). Thiol-reactive probes are chemical reagents that have been specifically designed to bind to thiol groups in molecules, making them valuable tools in molecular biology and biochemistry for labelling, detecting, and studying biomolecules. Thiol-reactive reagents in the probes react with thiol groups on proteins to form thioether-coupled products. These reagents react rapidly around

pH7 and can be coupled with thiol groups selectively in the presence of amine groups (Peng et al., 2012).

Functional groups contained in the probes bond covalently to thiol-containing reactive groups such as maleimide, iodoacetamide and alkyl halides, involving the formation of a covalent linkage through a variety of possible mechanisms including nucleophilic substitution, cyclization, redox reactions, Michael addition and metal complexes coordination (Wang, Peng and Wang, 2014; Qu et al., 2014). The probes are conjugated with fluorescent dyes or other labels such as biotin, allowing labelled molecules to be detected and studied using fluorescence microscopy visualisation. Streptavidin-based assays are used to detect biotinylated probes (Green, 1975; Lakshmipriya, Gopinath & Tang, 2016). Special probes capable of differentiating between cysteine, homocysteine, and glutathione were limited, but in the last decade several groups of thiol-reactive probes have been designed to selectively detect specific amino acids (Yin et al., 2017). For determination of specific thiols, chromatography and capillary electrophoresis can be used to separation, and chemically derived thiols can also be selectively detected using mass spectrometry (Wang, Peng and Wang, 2014).

Maleimide-based probes are widely used due to their specificity and stability. Based on the 1,8-naphthalic anhydride fluorophore, these probes have an attached maleimide unit which binds to thiols by Michael addition, forming a stable diether bond. Cell experiments have shown good cell-membrane permeability of the probe, thus it can be used to mark thiol within living cells. (Qu et al., 2014). Research for the use of maleimide probes to specifically label cysteine is ongoing and not without significant analytical challenge, but since cysteine is the only naturally occurring thiol peptide, they can be used to label proteins (Kim et al., 2008; McConnell, Smythers and Hicks, 2020).

Chemoproteomic platforms such as isoTOP-ABPP (Weerapana et al., 2011) apply a cysteine-reactive iodoacetamide-alkyne probe which, coupled with quantitative mass spectrometry, can monitor cysteine-reactivity changes. Techniques such as these have allowed relative reactivity quantification across hundreds of cysteines in live cells and lysates (Abo, Bak & Weerapana, 2017; Abo & Weerapana, 2015). Applications include the monitoring of post-translational modifications of cysteine including oxidation, nitrosation and metal chelation (Abo, Li and Weerapana, 2017).

Alkyl halides are less commonly used than maleimide and iodoacetamide but can be useful in certain applications. Since the structures are undisclosed, most are designed for *in vitro* assays instead of live cells (Kar et al., 2019). Alkyl halide probes utilize the same principles of high reactivity to -SH as maleimides, and the application is mostly overlapped (Wang et al., 2022).

4. Principles of Methods

Cas1 must be modified to become suitable for FP labelling. Wild type *E. coli* Cas1 has 4 cysteine residues (Uniprot, 2024) which must be removed for the application of maleimide labelling (Dillard et al., 2019). None of the existing cysteine residues are easily accessible to FPs, so the best course would be to introduce a new cysteine residue at a better position – see figure 1. S257S substitution could be a good candidate for its proximity to DNA, unlikelihood of Cas1-Cas2 complex formation interference, and very accessible protruding side group.

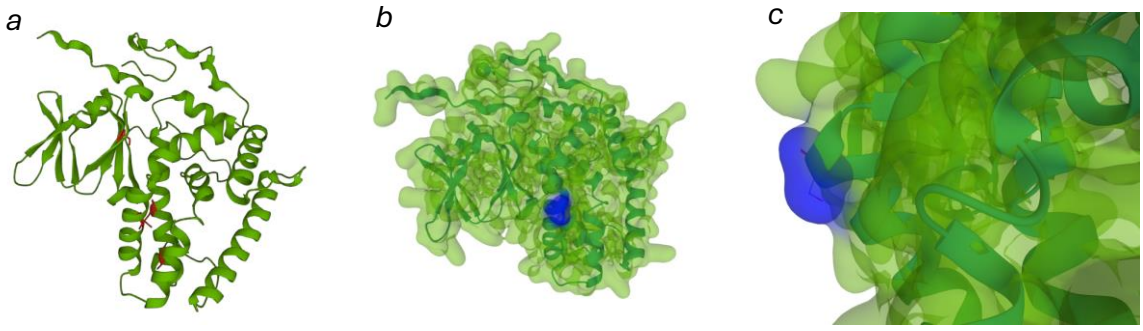

Figure 1: (a) Backbone model of *E. coli* Cas1 with cysteine residues highlighted in red; *(b)* Spacefill model of proposed *E. coli* Cas1 mutation site S257S highlighted in blue; *(c)* Close-up of the site shown in *b*.

Cysteine substitution can be achieved using site directed mutagenesis (SDM); polymerase chain reaction (PCR) success can be checked using agarose gel electrophoresis. To reach high enough yield for sanger sequencing, PCR product containing Cas1 can be amplified up to 30X by transformation into bacteria, after circularisation into a plasmid (Gutterson & Koshland, 1983). Cysteine substitution could cause a change to Cas1 expression (Akabas, 2015) so it must be checked by a pilot protein overexpression experiment. Cysteine substitution can also lead to protein misfolding (Mason, 1994; Vollbrandt et al., 2004), so mutant Cas1 DNA binding function should be assessed by an electrophoretic mobility shift assay (EMSA), fluorescence anisotropy assay, or spacer acquisition assay. Once mutant Cas1 is confirmed to function as well as wild type, maleimide labelling can be attempted.

4.1 Site Directed Mutagenesis (SDM)

Site-directed mutagenesis is a technique used to make intentional, targeted mutations to DNA sequences by the insertion, deletion, or substitution of specified nucleotides within a gene. A

pair of synthetic DNA oligonucleotides complementary to the target DNA sequence but containing the desired mutation are first designed. These 'primers' bind to template DNA during a polymerase chain reaction (PCR), allowing the enzyme DNA polymerase to synthesize new strands of DNA incorporating the mutations specified by the primers (Zhu et al., 2020). The PCR product is circularized and transformed into bacteria for exponential amplification, producing a high DNA yield that can be extracted and sequenced to confirm successful SDM (Gutterson & Koshland, 1983).

PCR is a hugely impactful method of DNA amplification used in a number of laboratory and clinical techniques including genetic diagnosis, DNA fingerprinting, and detection of disease (Jeffreys et al., 1988; Zhu et al., 2020). The first successful specific enzymatic amplification of DNA *in vitro* was performed by Mullis et al. in 1986 by denaturing double stranded DNA (dsDNA) into single stranded DNA (ssDNA) in a buffered solution with forward and reverse primers to define total amplicon length, and then adding *Escherichia coli* DNA polymerase I alongside deoxynucleotide triphosphates (Mullis et al., 1986). The polymerase enzyme had to be replaced after each cycle, which made the entire process incredibly tedious, until Randall et al. then improved the method in 1988 by instead using a thermostable DNA polymerase extracted from the thermophilic bacteria *Thermus aquaticus* known as Taq polymerase. This enzyme automates the repetitive amplification step and does not need to be replenished thanks to its stability up to 94°C (Randall et al., 1988).

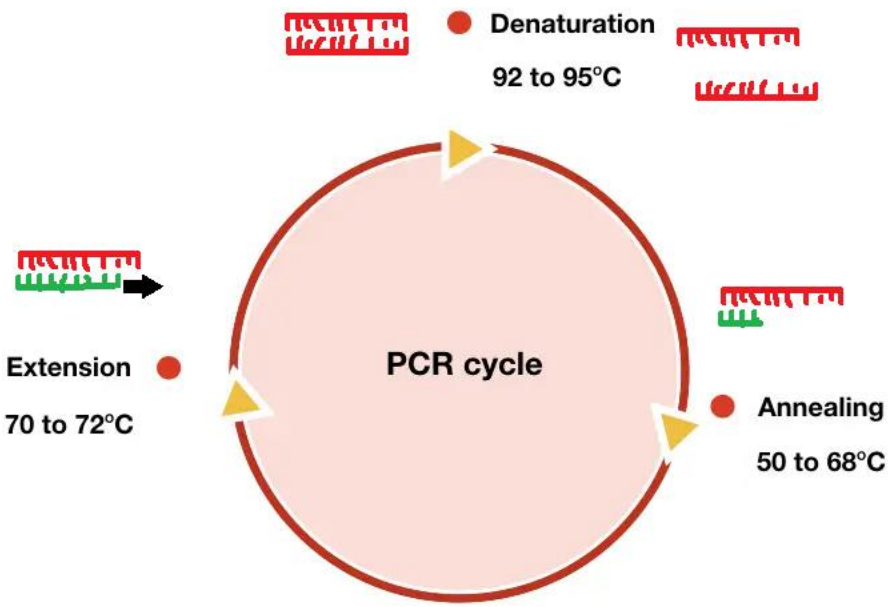

Figure 2: The steps of the polymerase chain reaction (PCR) cycle. Adapted from: Geneticeducation.co.in.

A number of challenges still exist with the application of PCR techniques. The success of its practical applications relies heavily on the quality of template DNA (He et al., 2011; Silva Jr. et al., 2000). High GC-content template sequence amplification also present challenge, due to the formation of secondary structures that hinder denaturation and primer annealing. This can sometimes be overcome by the use of specially designed polymerases and shorter annealing times (PCRBio, 2024; Mammedov et al., 2009).

4.2 Gel Electrophoresis

Developed in 1971 by Chrambach and Rodbard, polyacrylamide gel electrophoresis (PAGE) is a method of separating macromolecules such as DNA, RNA, and proteins based on their chain

length for analysis (Chrambach & Rodbard, 1971). Polyacrylamide is a synthetic polymer used to form pores that can be varied by adjusting its concentration.

To begin, a sample containing the molecules of interest is loaded into wells at one end of the gel. An electric current is then applied across the gel. Because the gel matrix is porous, smaller molecules move through it more quickly than larger ones. The electric field causes the molecules to migrate through the gel, with their movement being influenced by their size and charge. DNA is a negatively charged molecule, so will move towards the cathode (Al-Tubuly, 2000). Unlike DNA, proteins have varying mass/charge ratios depending on the bases in their sequence, which means they will not be separated by length alone (Lee et al., 2012). To counteract this, proteins are coated in sodium dodecyl-sulfate (SDS), which denatures them and creates a uniform negative charge across the molecule (Nowakowski, Wobig & Petering, 2015).

The gel will contain a stain that is added before setting, or in the case of SDS-PAGE, after the run is complete. This dye binds to the molecules of interest, allowing them to be visualized. For DNA or RNA, common stains include ethidium bromide or SYBR Green (Dragan et al., 2012; Franklin & Locker, 1981; Voytas, 2001.), which fluoresce under ultraviolet light. For proteins, stains like Coomassie Brilliant Blue or silver staining are used (Gallagher & Sasse, 1998). The separated molecules form distinct bands within the gel, which can be analyzed to determine their size, quantity, and sometimes their specific characteristics by comparison to a molecular ladder of known weights. This technique is widely used in molecular biology for applications such as DNA fingerprinting, gene analysis, and protein characterization (Johansson, 1972; Lee et al., 2012; Schaberg, Tompkins & Falkow, 1981).

4.3 Bacterial Transformation

PCR product is linear and should be circularized before transformation for maximum efficiency (Le et al., 2013). Product is added to a mixture of kinase, restriction endonuclease and ligase at room temperature for 5-60 minutes (New England Biolabs, 2024). Kinase phosphorylates both the 5' and 3' ends of the PCR product (Wang, Lima and Shuman, 2002), preparing it for DNA ligase, which circularizes (ligates) the DNA (Rossi et al., 1997). Template DNA is destroyed by a restriction endonuclease such as DpnI so that it cannot be transformed into bacteria (Mierzejewska et al., 2014).

The newly circularized recombinant plasmid is introduced into the bacteria on ice, before the bacterial membrane is made permeable to DNA by heat shock, electroporation, or chemical treatments like calcium chloride.

After overnight incubation, the transformed bacteria are sub-cultured by spreading onto agar plates that contain selective antibiotics. The plasmid includes an antibiotic resistance gene, so only the bacteria that have successfully taken up the recombinant plasmid will survive and grow on the selective medium. Any bacteria picked after another overnight incubation should contain high yields of the desired recombinant plasmid, which can then be extracted by cell lysis and ultracentrifugation filtering (Moore et al., 2004).

4.4 Sanger Sequencing

Sanger sequencing is a method used to determine the nucleotide sequence of DNA. The process involves several key steps that enable the precise reading of DNA sequences:

First, the DNA sample to be sequenced is denatured to separate the two strands. A short piece of DNA called a primer is then annealed to the template strand. This primer is designed to bind to a specific region of the DNA and provides a starting point for DNA synthesis.

Next, a DNA polymerase enzyme is used to extend the primer by adding nucleotides complementary to the template strand. This process is carried out in a reaction mixture that contains normal deoxynucleotide triphosphates (dNTPs) as well as modified, fluorescently labelled dideoxynucleotide triphosphates (ddNTPs). These ddNTPs are analogous to normal nucleotides but lack a 3'-OH group, which is necessary for forming a phosphodiester bond between nucleotides. As a result, their incorporation into the growing DNA strand causes the chain to terminate because it cannot establish a bond with the 5' phosphate of the following dNTP (Eren, Taktakoğlu & Pirim, 2022; Slatko, 2001). When the DNA polymerase encounters a ddNTP during synthesis, it incorporates the ddNTP at a specific position, terminating the chain at that point. Since each ddNTP is labelled with a different colour, the terminated fragments can be distinguished based on a fluorescent signal of changing wavelengths.

After the extension reaction, the mixture contains a collection of DNA fragments of varying lengths, each ending at a ddNTP. These fragments are then separated by size using capillary electrophoresis. The electrophoresis process involves passing the fragments through a capillary tube, where smaller fragments move faster than larger ones. As the fragments exit the capillary tube, they pass through a detector that reads the fluorescent labels of the ddNTPs. The sequence of fluorescent signals corresponds to the sequence of nucleotides in the original DNA strand. By analysing these signals, the complete nucleotide sequence of the DNA can be reconstructed (Eren, Taktakoğlu & Pirim, 2022).

Though Sanger sequencing is known for its high accuracy and is widely accepted as the current gold standard for sequencing, new techniques are rapidly overcoming its limitations. Next

generation sequencing techniques boast greater sensitivity, parallel samples analysis, lower error rate and decreased cost (Fernandes & Zhang, 2014; Hu et al., 2021; Slatko, Gardner & Ausubel, 2018).

4.5 Pilot Protein Overexpression

Obtaining substantial amounts of pure protein is essential in innumerable biological studies and indispensable to the biochemical characterization of proteins. The ease of growth, well-characterized genetics, and the large number of tools for gene expression have long made *Escherichia coli* the organism of choice for protein overproduction (James et al., 2021; Molina et al., 2009). Cysteine mutation can cause changes in protein expression (Akabas, 2015), leading to false results in future experiments. To confirm that mutant Cas1 expresses normally, a pilot protein overexpression experiment can be run. Recombinant plasmids containing the protein of interest are designed such that expression can be tightly regulated and the protein expressed at extreme levels. One such commercially available strain of *E. coli* possessing a T7-based protein expression system, BL21AI, is widely used for recombinant protein production because of its engineered capacity to produce T7 polymerase and its deficiency in Lon and OmpT proteases (Ratelade et al., 2009; Studier and Moffatt, 1986).

4.6 Spacer Acquisition Assay

The spacer acquisition assay is a technique developed to assess the impact of chaperone protein DnaK on Cas1-Cas2 mediated DNA integration into the chromosomal CRISPR locus (Killelea et al., 2023). It can be used to evaluate the impact of cysteine mutations on Cas1-Cas2 function.

Cas1-Cas2 is inducibly over-expressed from a plasmid as described in section 4.5 to overcome chromosomal Cas1 repression by H-NS (Mitić et al., 2020). Cas1-Cas2 then self-targets chromosomal DNA, which is captured and integrated as spacers into the CRISPR locus, causing the chromosomal CRISPR-1 site to expand (Wimmer & Beisel, 2020). PCR of the CRISPR locus produces DNA fragments which can be separated by agarose gel electrophoresis to compare CRISPR-1 site expansion by mutant Cas1 and wild type Cas1.

5. Conclusion

Various techniques have been developed to investigate the interactions of macromolecules *in vivo*. By systematically mutating Cas1 to substitute cysteine residues for others, such as alanine, it can be tagged by a fluorescent marker protein and its regulatory pathways can be better understood. It is important to check that the mutations have not impacted the DNA binding and normal function of Cas1 beforehand, using novel techniques such as spacer acquisition assays.

References

- Abo, M., Bak, D. and Weerapana, E., 2017. Optimization of Caged Electrophiles for Improved Monitoring of Cysteine Reactivity in Living Cells. *ChemBioChem*., 18(1), 81–84. DOI: 10.1002/cbic.201600524.

- Abo, M., Li, C. and Weerapana, E., 2017. Isotopically-Labeled Iodoacetamide-Alkyne Probes for Quantitative Cysteine-Reactivity Profiling. *Molecular Pharmaceutics*, 15 (3), 743-749. DOI: 10.1021/acs.molpharmaceut.7b00832.

- Abo, M. and Weerapana, E., 2015. A Caged Electrophilic Probe for Global Analysis of Cysteine Reactivity in Living Cells. *J. Am. Chem. Soc.*, 137(1), 7087– 7090. DOI: 10.1021/jacs.5b04350.

- Alberts, B. et al., 2002. The Adaptive Immune System. *Molecular Biology of the Cell,* 4(1), chapter 24. ISBN-10: 0-8153-3218-1.

- Al-Tubuly, A., 2000. SDS-PAGE and Western Blotting. In: George, A. and Urch, C. (eds) Diagnostic and Therapeutic Antibodies. *Methods in Molecular Medicine*, 40(1). DOI: 10.1385/1-59259-076-4:391.

- Akabas, M., 2015. Cysteine Modification: Probing Channel Structure, Function and Conformational Change. *Novel Chemical Tools to Study Ion Channel Biology,* 869(1), 25-54. ISBN : 978-1-4939-2844-6.

- Barrangou, R., et al., 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, 315(5819), 1709-1712. DOI: 10.1126/science.1138140.

- Bunyat-zada, A., and Ross, A., 2023. Highlights of Bioinformatic Tools and Methods for Validating Bioinformatics Derived Hypotheses for Microbial Natural Products Research. *Current Opinion in Chemical Biology,* 76. DOI: 10.1016/j.cbpa.2023.102367.

- Chrambach, A., and Rodbard, D., 1971. Polyacrylamide Gel Electrophoresis. *Science*, 172, 440-451. DOI:10.1126/science.172.3982.440.

- Dillard, K. et al., 2019. Sortase-Mediated Fluorescent Labeling of CRISPR Complexes. *Methods in Enzymology,* 616(1), 43-59. DOI: 10.1016/bs.mie.2018.10.031.

- Dragan, A. et al., 2012. SYBR Green I: Fluorescence Properties and Interaction with DNA. *J Fluoresc,* 22(1), 1189–1199. DOI: 10.1007/s10895-012-1059-8.

- Eren, K., Taktakoğlu, N. and Pirim, I., 2022. DNA Sequencing Methods: From Past to Present. *Eurasian J Med,* 54(1), 47-56. DOI: 10.5152/eurasianjmed.2022.22280. PMID: 36655445; PMCID: PMC11163357.

- Fernandes, H. and Zhang, P., 2014. Overview of Molecular Diagnostics in Clinical Pathology. *Pathobiology of Human Disease,* 3287-3303. DOI: 10.1016/B978-0-12-386456-7.06306-1.

- Franklin, W. and Locker, J., 1981. Ethidium Bromide: a Nucleic Acid Stain for Tissue Section. *Journal of Histochemistry & Cytochemistry*, 29(4) 572-576. DOI: 10.1177/29.4.6166660.

- Gallagher, S. and Sasse, J., 1998. Protein Analysis by SDS-PAGE and Detection by Coomassie Blue or Silver Staining. *Current Protocols in Pharmacology*, 2(3). DOI: 10.1002/0471141755.pha03bs02.

- Geneticeducation.co.in, 2024. Optimizing PCR Cycles to get Good Results [online]. Available at: https://geneticeducation.co.in/optimizing-pcr-cycles-to-get-good-results/?utm_content=cmp-true [Accessed 08 September 2024].

- Green, N., 1975. Avidin. *Advances in Protein Chemistry,* 29(1), 85-133. DOI: 10.1016/S0065-3233(08)60411-8.

- Gutterson, N. and Koshland D., 1983. Replacement and Amplification of Bacterial Genes with Sequences Altered In Vitro. *Proceedings of the National Academy of Sciences,* 80(16), 4894-4898. DOI: 10.1073/pnas.80.16.4894.

- He, G. et al., 2011. Evaluating the Reliability of Microsatellite Genotyping from Low-Quality DNA Templates with a Polynomial Distribution Model. *Chin. Sci. Bull,* 56(1), 2523–2530. DOI: 10.1007/s11434-011-4634-5.

- Hu, T., Chitnis, N., Monos, D. and Dinh, A., 2021. Next Generation Sequencing Technologies: an Overview. *Human Immunology,* 82(11), 801-811. DOI: 10.1016/j.humimm.2021.02.012.

- Ishino, Y., et al., 1987. Nucleotide Sequence of the iap Gene, Responsible for Alkaline Phosphatase Isozyme Conversion in Escherichia coli, and Identification of the Gene Product. *J Bacteriol,* 169(12). DOI: 10.1128/jb.169.12.5429-5433.1987.

- Jackson, S. et al., 2017. CRISPR-Cas: Adapting to change. *Science,* 356(1). DOI: 10.1126/science.aal5056.

- James, J. et al., 2021. Protein Over-Expression in Escherichia coli Triggers Adaptation Analogous to Antimicrobial Resistance. *Microb Cell Fact,* 20(13). DOI: 10.1186/s12934-020-01462-6.

- Jeffreys, A., Wilson, V., Neumann, R. and Keyte, J., 1988. Amplification of Human Minisatellites by the Polymerase Chain Reaction: Towards DNA Fingerprinting of Single Cells, *Nucleic Acids Research*, 16(23), 10953–10971. DOI: 10.1093/nar/16.23.10953.

- Johansson, B., 1972. Agarose Gel Electrophoresis. *Scandinavian Journal of Clinical and Laboratory Investigation*, 29(124), 7–19. DOI: 10.3109/00365517209102747.

- Johnson, F., Shimomura, O., Saiga, Y., Gershman, L., Reynolds, G. and Waters, J., 1962. Quantum Efficiency of *Cypridina* Luminescence, With a Note on That of *Aequorea*. *J. Cell. Comp. Physiol.*, 60(1), 85-103. DOI: 10.1002/jcp.1030600111.

- Kar, S. et al., 2019. Hydrogen Sulfide Ameliorates Homocysteine-Induced Cardiac Remodeling and Dysfunction. *Front Physiol*., 10(1), 598. DOI: 10.3389/fphys.2019.00598.

- Killelea, T. et al., 2023. Cas1-Cas2 Physically and Functionally Interacts with DnaK to Modulate CRISPR Adaptation. *Nucleic Acids Res.,* 51(13), 6914-6926. DOI: 10.1093/nar/gkad473.

- Kim, Y. et al., 2008. Efficient Site-Specific Labeling of Proteins via Cysteines. *Bioconjugate Chemistry*, 19 (3), 786-791. DOI: 10.1021/bc7002499.

- Lakshmipriya, T., Gopinath, S. and Tang, T., 2016. Biotin-Streptavidin Competition Mediates Sensitive Detection of Biomolecules in Enzyme Linked Immunosorbent Assay. *PLOS ONE,* 11(3), 151-153. DOI: 10.1371/journal.pone.0151153.

- Little, C. and O'Brien, P., 1968. The Effectiveness of a Lipid Peroxide in Oxidizing Protein and Non-Protein Thiols. *Biochem J*, 106 (2), 419–423. DOI: 10.1042/bj1060419.

- Le, Y., et al., 2013. Thermostable DNA Ligase-Mediated PCR Production of Circular Plasmid (PPCP) and its Application in Directed Evolution via In Situ Error-Prone PCR. *DNA Res.* 20(4), 375-82.

- Lee, P., Costumbrado, J., Hsu, C. and Kim, Y., 2012. Agarose Gel Electrophoresis for the Separation of DNA Fragments. *J. Vis. Exp.* 1(62), 3923. DOI: 10.3791/3923.

- Maloth, K. et al., 2014. Radioisotopes: An Overview. *International Journal of Case Reports and Images*, 5(9), 604. DOI: 10.5348/ijcri-201457-RA-10012.

- Mamedov, T., et al., 2008. A Fundamental Study of the PCR Amplification of GC-rich DNA Templates. *Comput. Biol. Chem.,*32(6), 452-7. DOI: 10.1016/j.compbiolchem.2008.07.021.

- Mason, A., 1994. Functional Analysis of the Cysteine Residues of Activin A. *Molecular Endorcrinology,* 8(3) 325-332. DOI: 10.1210/mend.8.3.8015550.

- Mierzejewska, K. et al., 2014. Structural Basis of the Methylation Specificity of R.DpnI. *Nucleic Acids Research*, 42(13), 8745–8754. DOI: 10.1093/nar/gku546.

- Mitić, D., Radovčić, M., Markulin, D. and Ivančić-Baće, I., 2020. StpA represses CRISPR-Cas Immunity in H-NS Deficient Escherichia coli. *Biochimie,* 174(1), 136-143. DOI: 10.1016/j.biochi.2020.04.020.

- Molina, D. et al., 2008. Engineering Membrane Protein Overproduction *in Escherichia coli. Protein Science,* 17(1), 673-680. DOI: 10.1110/ps.073242508.

- Moore, E. et al., 2004. Simplified Protocols for the Preparation of Genomic DNA from Bacterial Cultures. *Molecular Microbial Ecology Manual,* 2(1), 3-18. DOI:10.1007/978-1-4020-2177-0_101.

- Morin, J. and Hastings, J., 1971. Energy Transfer in a Bioluminescent System. *J. Cell. Physiol.,* 77, 313-318. DOI: 10.1002/jcp.1040770305.

- Morise, H., Shimomura, O., Johnson, F., and Winant, J., 1974. Intermolecular Energy Transfer in the Bioluminescent System of *Aequorea*. *Biochemistry*, 13 (12), 2656–2662. DOI: 10.1021/bi00709a028.

- Mullis, K. et al., 1986. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harb Symp Quant Biol.,* 51, 263-273. DOI: 10.1101/SQB.1986.051.01.032.

- New England Biolabs, 2024. KLD Enzyme Mix [online]. Available at: https://www.neb.com/en-gb/products/m0554-kld-enzyme-mix [Accessed 10 September 2024].

- Nowakowski, A., Wobig, W. and Petering, D., 2015. Native SDS-PAGE: High Resolution Electrophoretic Separation of Proteins with Retention of Native Properties Including Bound Metal Ion. *Metallomics*, 6(5), 1068-1078. DOI: 10.1039/c4mt00033a.

- PCR Biosystems, 2024. GC-Rich PCR [online]. Available at: Polymerases for GC-Rich PCR | PCR Biosystems [Accessed 08 September 2024].

- Peng, H. et al., 2012. Thiol Reactive Probes and Chemosensors. *Sensors*, 12(11):15907-15946. DOI: 10.3390/s121115907.

- Pickar-Oliver, A., Gersbach, C., 2019. The Next Generation of CRISPR–Cas Technologies and Applications. *Nat Rev Mol Cell Biol,* 20(1), 490–507. DOI: 10.1038/s41580-019-0131-5.

- Pourcel, C. et al., 2020. CRISPRvCasdb A Successor of CRISPRdb Containing CRISPR Arrays and Cas Genes from Complete Genome Sequences, and Tools to Download and Query Lists of Repeats and Spacers. *Nucleic Acids Res.*,48(D1), D535-D544. DOI: 10.1093/nar/gkz915.

- Prasher, D., et al 1992. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111(2), 229-233. DOI: 10.1016/0378-1119(92)90691-H.

- Qu, L. et al., 2014. A Maleimide-Based Thiol Fluorescent Probe and its Application for Bioimaging. *Sensors and Actuators B: Chemical,* 195(1), 246-251. DOI: 10.1016/j.snb.2014.01.026.

- Ratelade, J. et al, 2009. Production of Recombinant Proteins in the lon-Deficient BL21(DE3) Strain of Escherichia coli in the Absence of the DnaK Chaperone. *Appl Environ Microbiol.* 75(11), 3803-7. DOI: 10.1128/AEM.00255-09.

- Rossi, R., Montecucco, A., Ciarrocchi, G. and Biamonti, G., 1997. Functional Characterization of the T4 DNA Ligase: a New Insight into the Mechanism of Action. *Nucleic Acids Research*, 25(11) 2106–2113. DOI: 10.1093/nar/25.11.2106.

- Saiki, R., et al., 1988. Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. *Science*, 239, 487-491. DOI:10.1126/science.2448875.

- Schaberg, D., Tompkins, L., Falkow, S., 1981. Use of Agarose Gel Electrophoresis of Plasmid Deoxyribonucleic Acid to Fingerprint Gram-Negative Bacilli. *J Clin Microbiol* 13(1). DOI: 10.1128/jcm.13.6.1105-1108.1981.

- Shimomura, O., Johnson, F.H. and Saiga, Y., 1962. Extraction, Purification and Properties of Aequorin, a Bioluminescent Protein from the Luminous Hydromedusan, Aequorea. *J. Cell. Comp. Physiol.,* 59(3), 223-239. DOI: 10.1002/jcp.1030590302.

- Silva, W. et al., 2001. PCR Template Preparation for Capillary DNA Sequencing. *BioTechniques*, 30(3), 537–542. DOI: 10.2144/01303st05.

- Slatko, B., Albright, L., Tabor, S. and Ju, J., 1999. DNA Sequencing by the Dideoxy Method. *Current Protocols in Molecular Biology*, 47(7). DOI: 10.1002/0471142727.mb0704as47.

- Studier, F. and Moffatt, B., 1986. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J Mol Biol,* 189(1), 113-30. DOI: 10.1016/0022-2836(86)90385-2.

- Slatko, B., Gardner, A. and Ausubel, F., 2018. Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology,* 122(59). DOI: 10.1002/cpmb.59.

- Uniprot, 2024. ygbt – CRISPR-Associated Endonuclease Cas1 [online]. Available at: https://www.uniprot.org/uniprotkb/Q46896/entry [Accessed 10 September 2024].

- Vollbrandt, T. et al., 2004. Consequences of Cysteine Mutations in Calcium-binding Epidermal Growth Factor Modules of Fibrillin-1. *Glycobiology and Extracellular Matrices*, 279(31), 32924-32931.

- Voytas, D., 2000. Agarose Gel Electrophoresis. *Current Protocols in Molecular Biology*, 51(2). DOI: 10.1002/0471142727.mb0205as51.

- Wang, L. et al., 2022. Fluorescent Probes and Mass Spectrometry-Based Methods to Quantify Thiols in Biological Systems. *Antioxid Redox Signal.,* 36(6), 354-365. DOI: 10.1089/ars.2021.0204.

- Wang, L., Lima, C. and Shuman, S., 2002. Structure and Mechanism of T4 Polynucleotide Kinase: an RNA Repair Enzyme. *The Embo Journal,* 21(1), 3873-3880. DOI: 10.1093/emboj/cdf397.

- Wang, K., Peng, H. and Wang, B., 2014. Recent Advances in Thiol and Sulfide Reactive Probes. *Journal of Cellular Biochemistry,* 11(6), 1007-1022. DOI: 10.1002/jcb.24762.

- Weerapana, E. et al., 2011. Quantitative Reactivity Profiling Predicts Functional Cysteines in Proteomes. *Nature,* 468(7325) 790-795. DOI: 10.1038/nature09472.

- Wimmer, F. and Beisel, C., 2020. CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Front. Microbiol.,* 10(1), 3078. DOI: 10.3389/fmicb.2019.03078.

- Yin, C. et al., 2017. Fluorescent Probes with Multiple Binding Sites for the Discrimination of Cys, Hcy, and GSH. *Angew Chem. Int. Ed. Engl.,* 56(43), 13188-13198. DOI: 10.1002/anie.201704084.

- Zhu, H. et al., 2020. PCR Past, Present and Future. *BioTechniques*, 69(4), 317–325. DOI: 10.2144/btn-2020-0057.

Substitution of Cysteine Residues in CRISPR-Associated Protein Cas1

Name: Lochlan Parr

Supervisors: Prof. Stephen Harding - University of Nottingham

Prof. Ed Bolt - University of Nottingham

Company: University of Nottingham

Key Words: CRISPR, Cas1, Cas2, Site Directed Mutagenesis, mStayGold.

Report/Research paper submitted to The University of Nottingham in partial fulfilment of the requirements for the degree of Master of Research in Biomolecular Technology.

Total length of abstract: 260 words

Total length of report: 3995 words

Table of contents

1. Abstract

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems mediate prokaryotic adaptive immunity against mobile genetic elements (MGEs), where MGE sequences are captured and stored as spacers in CRISPR loci. CRISPR-associated (Cas) proteins Cas1 and Cas2 form a complex that integrates these spacers into CRISPR arrays, generating adaptive immunity through CRISPR interference. While the mechanisms of the Cas1-Cas2 complex are understood, less is known about regulatory pathways, particularly because it has been shown to interact with proteins far outside its genetic locality, sometimes from different chromosomes. Fluorescent proteins (FPs) have successfully been used to tag Cas1 to study its interaction with other proteins. mStayGold, a novel highly photostable FP, shows promise of enhancing the observation of Cas1 interactions in live cells. In contrast to challenges with photobleaching in conventional FPs, mStayGold offers tenfold superior photostability at no cost to brightness, which is helpful for advanced microscopy techniques. Wildtype Cas1 presents difficulties with mStayGold fusion due to the presence of multiple cysteine residues which can destabilize the tertiary structure of FPs. To address this, site-directed mutagenesis (SDM) was employed to substitute cysteine for serine residues in Cas1, aiming to create a mutant Cas1 with a single, strategically placed cysteine residue that optimally binds mStayGold. There were significant setbacks caused by low PCR yields and undesirable mutations, but the first successful mutant, Cas1 C51S does not exhibit impaired protein expression. There is no apparent reason the remaining cysteine residues cannot continue to be substituted, allowing advanced microscopy techniques to investigate molecular interactions *in vivo*, aiming to improve our understanding of Cas1 regulatory interactions.

2. Introduction

Prokaryote adaptive immunity against mobile genetic elements (MGEs) is provided by specialised chromosomal sites called CRISPRS (Clustered Regularly Interspaced Short Palindromic Repeats). CRISPR-associated (Cas) proteins Cas1 and Cas2 form a complex capable of capturing MGE fragments as seen in figure 1, and integrating them into a CRISPR as 'spacers', generating immunity by CRISPR 'adaptation' that is delivered by CRISPR 'interference' reactions. Pre-spacer DNA is defined by length and Protospacer-Adjacent Motif (PAM) end sequences, though these PAMs are not integrated into DNA as part of the spacer.

Figure 1: Model of Cas1-Cas2 complex bound to DNA containing PAM. Four Cas1 subunits (Teal, orange, pink and blue) and two Cas2 subunits (brown and grey) combine to form a complex capable of binding pre-spacer DNA in two Cas1 active sites. Cas2 dimer acts as scaffolding. Adapted from Nuñez et al., 2015.

RNA transcribed from CRISPR is cleaved within repeat regions into single-spacer crRNAs, which interference complexes (Cascade-Cas3 in *Escherichia coli*, a class 2 CRISPR system) can bind to. These interference complexes 'lock' into DNA R-loops where a PAM is at an MGE sequence complementary to the crRNA, which triggers MGE nuclease destruction.

Fluorescent protein (FP) photobleaching is a major limitation to their use in advanced microscopy, but since understanding of its molecular mechanism is limited, improving photostability remains challenging, (Ren et al., 2016). mStayGold is a novel FP with remarkable resistance to photobleaching, (Ivorra-Molla et al., 2023), making it ideal for live cell imaging tagged proteins. Cas1 has previously been tagged with the *Aequorea* mutant eYFP as fluorescence resonance energy transfer (FRET) donors and acceptors, (Killelea et al., 2023), revealing interaction with chaperone protein DnaK. mStayGold is over one order of magnitude more photostable than any other available FP and just as bright, (Yonemaru et. al., 2024), enhancing the outcome of time-lapse fluorescence microscopy. By fusing Cas1 with mStayGold in a similar manner as was done with eYFP, novel molecular interactions could be more easily discovered.

Wildtype Cas1 would not be a suitable candidate for fusion with mStayGold due to the number and position of cysteine residues in its polypeptide chain. Cysteine is an important amino acid in FPs for a number of reasons relating to structure. It is easily oxidised into cystine due to the presence of a sulfhydryl group that forms a disulphide bridge to link the dimer. These disulfide bridges play a crucial role in intramolecular cross-linking, usually maintaining a protein's secondary structure, but in the case of a fusion protein, multiple cysteine residues on the protein of interest could destroy the FP's fluorophore (Isarankura-Na-Ayudhya et al., 2009). Removal or substitution of all cysteines, however, decreases a FP's photostability.

Cas1 cysteine residues can be sequentially substituted for other inert amino acids using site directed mutagenesis (SDM) polymerase chain reaction (PCR), resulting in mutant Cas1 that has only one cysteine residue remaining for optimal mStayGold binding. As shown in figure 2, none of these cysteine residues are adequately accessible, so the best course would be to instead substitute them all and introduce one at a better location to maintain fluorophore stability. S257 was chosen to be substituted for cysteine, for its accessibility and distance from active site, seen in figure 3.

Figure 2: Model of Cas1 with cysteine residues highlighted in red. From highest to lowest: C51, C186, C251, and C179. They are not very easily accessible by mStayGold, but may still bind and disrupt its structure during protein unfolding and / or binding of the active site to DNA.

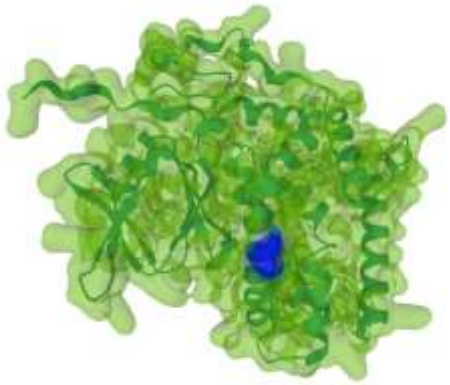

Figure 3: Model of Cas1 with proposed serine to cysteine substitution site S257 highlighted in blue. It is easily accessible by mStayGold on the outside of the protein.

3. Materials and Methods

3.1 Strains, plasmids and media

E. coli strain DH5-α are competent cells designed to maximise transformation efficiency due to a deoR mutation, (Yale, 2024). *E. coli* BL21-AI is a commercially available arabinose-inducible strain possessing a phage T7-based protein-expression system, used for recombinant protein overexpression (Bhawsinghka, Glenn and Schaaper, 2020.) *E. coli* EB377 approximates wild type *E. coli* while still being low-risk. Cells were grown at 37 C in LB broth (10 g / l bacto-tryptone, 5 g / l yeast extract, 10 g / l NaCl) and on LB agar plates (supplemented with 15 g of agar / L for solid media) unless otherwise stated. pEB505 is a wild type Cas1 expression plasmid under control of T7 promoter; *E. coli* ygbT cloned into pET14-b (Novagen). pEB628 is a wild type Cas1-Cas2 expression plasmid under control of arabinose inducible araBAD promoter; *E. coli* ygbT and ygbF cloned into pBadHisA (Invitrogen).

3.2 Site Directed Mutagenesis

Mutations were produced at target sites using polymerase chain reaction (PCR) site directed mutagenesis (SDM). PCR was performed using the Routine Vent PCR protocol (New England Biolabs, 2024) with an Applied Biosystems™ Veriti™ 96 Well Thermal Cycler. Standard cycle conditions are shown in table 1, with annealing step temperature being the main variable, provided in table 3. Oligonucleotide primers described in table 2 were designed using NEBaseChanger (New England Biolabs, 2024). This software calculates primer-DNA annealing temperature (T_a) with the assumption that Q5 polymerase is used, but these experiments were instead using Vent polymerase. Therefore, estimated annealing temperatures were adjusted by including the difference between calculated T_a using Q5 and Vent polymerase by NEB Tm calculator (New England Biolabs, 2024), which does not consider base mismatch. In the circumstance that SDM PCR produces no usable product, annealing temperature is the first variable to be adjusted, in 3°C increments.

Table 1: Standard PCR cycler conditions. Adapted from Protocol For a Routine Vent PCR (New England Biolabs, 2024).

	Step	Temperature (°C)	Time (Minutes)
	Initial denaturation	95	5
25 Cycles	Denaturation	95	0.5
	Annealing	Variable	0.5
	Extension	72	5
	Final extension	72	5
	Hold	10	

Table 2: Description of designed SDM PCR primers. *There was an error in this primer's design, see section 4.1.

Name	Intended Mutation	Sequence
oTK172 F*	**C51S** (TGC → TCC)*	5' TGGCTCGGTTCCCTGCATCAT
oTK172 R	none	5' ACAGGAATATGAGTGCGG
oTK173 F	**C179S** (TGT → TCT)	5' GCAACTTCCTCTTTATACGGCG
oTK173 R	**C186S** (TGC → TCC)	5' AGCGCTAATGGATTGGTTG
oTK174 F	**C251S** (TGC → TCC)	5' CGTTTGGCGTCCAGGGATATTTTTCG
oTK174 R	none	5' GACTTCCCGGTCCGGCTC
oTK175 F	**S257C** (AGT → TGT)	5' TATTTTTCGCTGTAGTAAAACATTAGCCAAATTGATTCC
oTK175 R	none	5' TCCCTGCACGCCAAACGG
oTK219 F	**C51S** (TGC → TCC)	5' CTCGGTTGCCAGCATCATGCTGG
oTK219 R	none	5' CCAACAGGAATATGAGTGCCG

3.3 Agarose Gel Electrophoresis

Using the procedure described by Lee et al., 2012, PCR products were assessed for confirmation of successful mutation. Comparison against a 1kb ladder allows identification of bands containing DNA of length ~5kb, indicating the desired complete mutant PCR product is present. In the case of multiple visible bands in one PCR product, it is necessary to extract the 5kb band from the gel. In this circumstance the gel was run again but with all of the PCR product, and then the DNA with the desired mutation was excised. Since UV light is highly mutagenic, the gel was visualized and cut under a blue light in a darkroom. An example of such a gel prepared is shown in figure 4. The sliced band was then dissolved, eluted and purified.

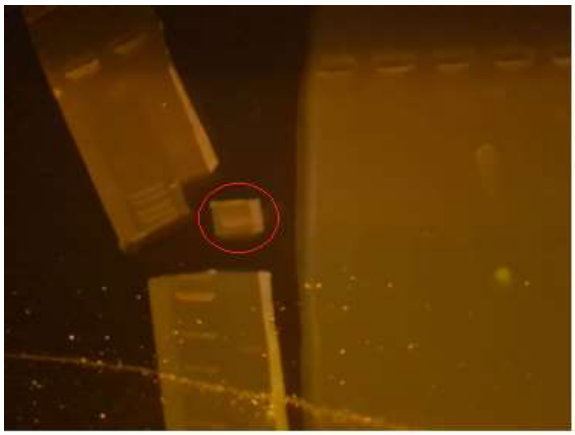

Figure 4: An excised band of PCR product from an agarose gel. 1kb protein ladder is visible on the left, and the desired PCR product is highlighted in red.

3.4 DNA Extraction and Purification

DNA was extracted from bacteria and purified using the Promega Wizard® Plus SV Minipreps DNA Purification System. DNA was extracted and purified from agarose gels using the Promega Wizard® SV Gel and PCR Clean-Up System. Purified DNA concentration was analysed using a DeNovix® DS-11 Series Spectrophotometer.

3.5 KLD Reaction

Template plasmids were destroyed and linear PCR product was circularized before insertion into bacteria by a 10µl KLD reaction. In a buffered solution, Dpn1, T4 PNK and T4 ligase were added to PCR product and incubated overnight at 18°C to cleave template plasmids, phosphorylate linear PCR product, and circularise phosphorylated PCR product respectively.

3.6 DNA amplification

SDM PCR DNA product was low in yield, so was amplified exponentially by insertion into a safe strain of competent *E. coli* called DH5-α using a standard transformation technique. 1 microlitre of purified plasmid or a full 10µl KLD reaction was added to 100µl of DH5-α cells and incubated for 30 minutes on ice, before being heat shocked at 42°C for 90 seconds to create pores in the bacterial plasma membrane for plasmid DNA to pass through into the cell, and then replaced on ice for a further 5 minutes. Liquid broth (LB) is then added and the mixture is shaken at 37°C for 60 minutes before being centrifuged at 13,000 rpm for 1 minute. The pellet is resuspended in sterile dilute water (SDW) and plated onto LB-agar containing 0.1mg/ml ampicillin, and incubated at 37°C for 16 hours. Individual colonies can then be sub-cultured to ensure clonality by transfer to a fresh nutritive medium of LB containing 0.1mg/ml ampicillin, which is shaken for 16 hours at 37°C. DNA is extracted as described in section 3.5.

3.7 Sequencing

Plasmid DNA sample mutations were confirmed by GeneWiz® sanger sequencing service. Samples were sent premixed with appropriate primers: pET Upstream in the case of pEB505-based plasmids, and pBAD F in the case of pEB628-based plasmids. The result is compared against the wildtype sequence using online Multalin software (Corpet, 1988) and any unexpected mutations were double-checked against the trace. GeneWiz® provided QS and CRL values for the sample to verify trace quality.

3.8 Pilot Protein Overexpression

A pilot protein overexpression method was used to determine the effect of mutations on protein expression. pEB505, LP003 and pET-14b were transformed into DH5-α cells and sub-cultured as described in section 3.7. Cultures were grown to OD600 = 0.6, determined using a DeNovix® DS-11 Series Spectrophotometer. Cas1 expression was induced by the addition of sterile L-Arabinose and IPTG to final concentrations of 0.2% and 1mM respectively, before a further 3 hours of incubation at 37°C with shaking. 4X SDS loading buffer and DL-Dithiothreitol (DTT) stabilising agent were added to a pelleted sample, which was then heated for 10 minutes at 95°C. These samples were used in a standard sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE) as described by Sigma Aldritch, 2024, alongside Blue Protein Standard.

3.9 Naïve Adaptation Assays

Naïve adaptation assays were based on the procedure described by Killelea et al., 2023. *E. coli* EB377 cells transformed with plasmid vector lacking Cas1–Cas2, pEB628 or pLP002 were inoculated into LB containing 0.2% L-arabinose and aerated at 37°C for 16 hours, and then sub-cultured by diluting 1:300 into fresh LB again supplemented with 0.2% L-arabinose and aerated at 37°C for a further 8

hours. Cells were harvested at identical time points and genomic DNA extracted using Wizard® Genomic DNA Purification Kit (Promega™). Spacer acquisition was monitored by PCR, utilizing 10 ng of genomic DNA and primers SW1 and SW2 (Killilea et al., 2023), with products separated using a 1.25% agarose gel stained with ethidium bromide and imaged using a U:Genius3 (Syngen Biotech).

4. Results

4.1 SDM PCR

In total, of the 30 PCR experiments, only 15 displayed bands in the agarose gel, of greatly varying visibility. Experiment success is shown in table 3. Examples of band strength can be seen in figure 5. Most SDM PCR experiments were unsuccessful, and success appears to be highly dependent on using a suitable annealing temperature for the annealing step of PCR. Unfortunately, matching the variable annealing step temperature to estimated T_a rarely results in successful PCR, as shown in table 3. Much of the successful PCR product was produced by guesswork.

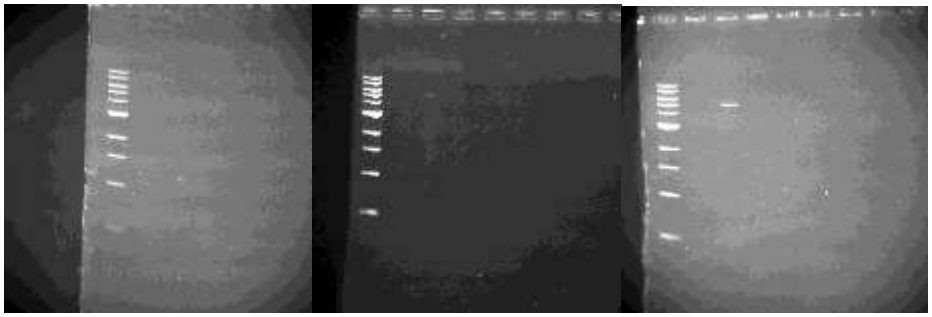

Figure 5: Three agarose gels viewed using UV imager showing: failed PCR reaction (left), PCR reaction with low yield (middle), successful PCR reaction with good yield (right). In all gels, column 1 contains 1kb ladder, column 2 contains negative control, and column 3 contains PCR product.

Table 3: PCR experiment results determined by gel electrophoresis and sanger sequencing.

Run / Product Plasmid Name	Template Plasmid	SDM Primers	Predicted T_a	Annealing Step Temperature	Apparent PCR yield	Sequencing Result
1	pEB505	oTK173	64°C	64°C	-	-
2	pEB505	oTK173	64°C	61°C	-	-
3	pEB505	oTK173	64°C	59°C	-	-
4	pEB505	oTK173	64°C	56°C	-	-
5	pEB505	oTK173	64°C	51°C	-	-
6	pEB505	oTK174	68°C	68°C	-	-
7	pEB505	oTK172	63°C	63°C	+	Yield too low for sequencing
8 / pLP001	pEB505	oTK172	63°C	63°C	+++	Undesired substitution A50P (GCC→CCC)
9 / pEB628A	pEB628	oTK172	63°C	63°C	+	Undesired substitution A50P (GCC→CCC)
10	pLP001	oTK173	64°C	64°C	-	-
11	pLP001	oTK173	64°C	64°C	-	-
12	pLP001	oTK173	64°C	57°C	-	-
13	pLP001	oTK173	64°C	57°C	++	-
14 / pEB505E	pLP001	oTK173	64°C	54°C	+++	Desired substitutions C179S and C186S. Undesired substitution Q24stop (CAG→TAG) and deletion at A182 (GCT→GC-)
15 / pEB628F	pEB628	oTK173	64°C	54°C	++	Desired substitutions C179S and C186S. Undesired deletion at A183 (GCA→-CA)
16	pEB505	oTK219	54°C	54°C	-	-
17	pEB628	oTK219	54°C	54°C	+	Yield too low for sequencing
18	pEB505	oTK219	54°C	57°C	+	Yield too low for sequencing
19	pEB628	oTK219	54°C	57°C	-	-
20	pEB505	oTK219	54°C	52°C	+	Yield too low for sequencing
21 / pLP002	pEB628	oTK219	54°C	52°C	+++	Desired substitution C51S. Undesired substitution T38A (ACA→GCA)
22 / pLP003	pEB505	oTK219	54°C	55°C	+	Desired

						substitution C51S.
23	pEB628	oTK219	54°C	55°C	-	-
24	pLP003	oTK174	61°C	61°C	++	Yield too low for sequencing
25 / pEB505AA	pLP003	oTK174	61°C	61°C	+++	No mutation
26 / pEB505AB	pLP003	oTK174	61°C	61°C	+++	Undesired substitution P91L (CCT→CTT)
27	pLP003	oTK173	64°C	61°C	-	-
28	pLP003	oTK173	64°C	54°C	-	-
29	pLP003	oTK173	64°C	63°C	-	-
30	pLP003	oTK173	64°C	58°C	+++	-

During the project it was found that run 8 in fact did not have the intended C51S mutation. An error was made during primer design, causing oTK172 to in fact direct mutagenesis to a A50P (GCC→CCC) point mutation, which was not useful. All runs subsequent runs which used pEB505A as a template were destroyed and a new primer was ordered with the correct sequence, oTK219.

pLP003 showed a mutation at E171 (GAA→GAG) but upon inspection of the trace shown below this was assumed to be false. This assumption was further supported when subsequent plasmids based on pLP003 were sequenced and showed no mutation at this site.

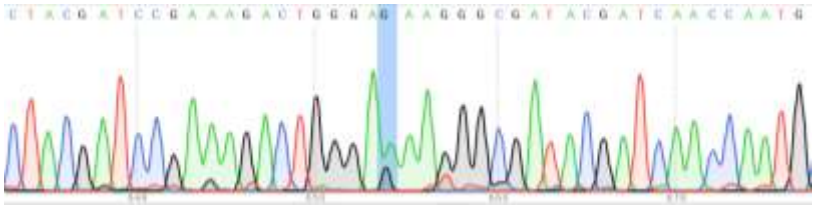

Figure 6: A section of the sanger sequencing trace for pLP003. The highlighted base shows an A→G substitution, but inspection of the trace suggests an error in reading. The peak for G is not central between its neighbouring bases, and the expected A peak behind is still stronger.

4.2 Plasmid Yield and Purity

Plasmids pEB505 and pEB628 shown in figures 7 and 8 respectively were not particularly pure. Later produced plasmids such as pLP003 (figure 9) had improved purity, though decreased yield. Yield was so low that in some instances DNA was eluted into half as much water as directed in protocol, so that concentration was high enough for further experiments.

Figure 7: Screenshot of pEB505 yield and purity from DeNovix® DS-11 Series Spectrophotometer. 74ng/µl yield is high enough for use as a template plasmid in PCR. Purity is low, particularly in regards to protein contamination.

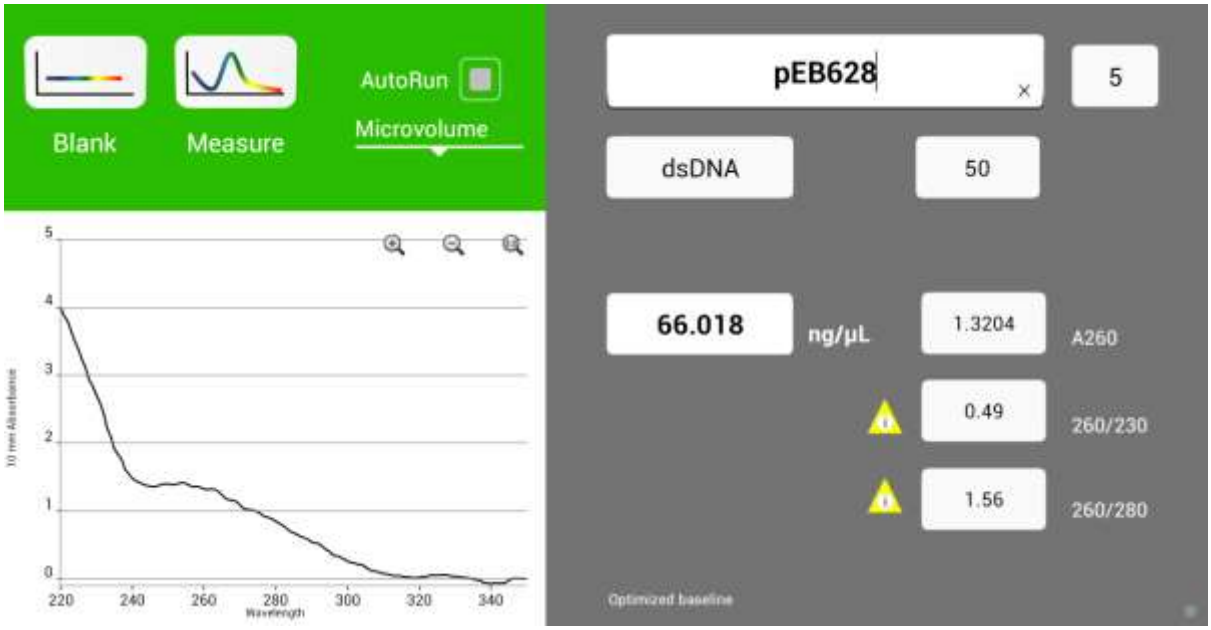

Figure 8: Screenshot of pEB628 yield and purity from DeNovix® DS-11 Series Spectrophotometer. 66ng/µl yield is high enough for use as a template plasmid in PCR. Purity is low, particularly in regards to protein contamination

Figure 9: Screenshot of pLP003 yield and purity from DeNovix® DS-11 Series Spectrophotometer. 48ng/µl yield is above the 30ng/µl required by Genewiz® for Sanger Sequencing, so the sample can be sent. Purity is satisfactory, with a 260/280nm ratio of 1.80 generally accepted as pure, with low protein contamination (source). 260/230nm ratio is expected to be above 2.0, so the result of 1.47 indicates some contamination by carbohydrates, however the sample is still usable.

4.3 Pilot Protein Overexpression

The C51S substitution in LP003 appears not to diminish Cas1 expression, as shown in figure 10, so the plasmid can be used in subsequent mutations.

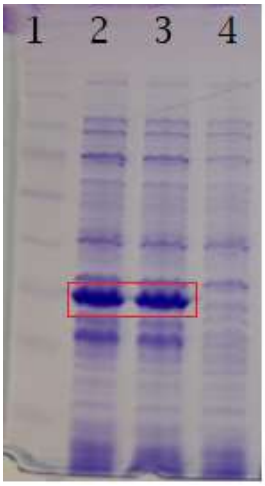

Figure 10: Completed SDS-PAGE of cell contents from pilot overexpression with columns labelled 1-4. Column 1 contains Invitrogen™ SeeBlue™ pre-stained protein ladder. Column 2 contains pEB505. Column 3 contains LP003. Column 4 contains pET14b. The two strong bands highlighted in red are Cas1. Column 1 Cas1 is wildtype, column 2 Cas1 has C51S substitution. Column 3 contains no Cas1 and is a negative control. C51S mutation has made no substantial difference to Cas1 expression.

4.4 Naïve Adaptation Assays

No bands were seen on the gel, the experiment was deemed to be a failure since the positive control was not visible.

5. Discussion

Disappointingly, many experiments over the course of this project had to be repeated multiple times, and since mutations must be performed sequentially, this meant progress was excruciatingly slow. Delays were caused mainly by failure to produce high yield PCR product.

The low success rate of SDM PCR was unexpected, and there are a multitude of possible causes. Most likely was the poor purity of pEB505 and pEB628 template plasmids. Purity plays a large role in PCR efficacy (Lorenz, 2012). Had these plasmids been purified further, and if necessary, produced from a fresh transformation, then it's possible PCR reliability would have dramatically increased, saving days or even weeks. Another factor which likely dramatically decreased PCR yield was the choice of DNA polymerase enzyme. Vent polymerase should have been a good choice with fidelity much greater than alternatives like taq polymerase due to exonuclease activity (Cline, Braman & Hogrefe, 1996). Unfortunately, this exonuclease activity may have been responsible for deletions between primers oTK173 F and oTK173 R. A different polymerase enzyme, such as Q5 polymerase, could have been used in this specific scenario. The plasmids used in these experiments approximate 5kb in length, and Vent polymerase should not lose fidelity until plasmid length approaches 15kb, (New England Biolabs, 1994).

The method used to estimate optimal annealing temperature was rudimentary and certainly could have been improved by, rather than relying on NEBaseChanger calculated T_m, using the formula $T_m = 81.5+ 0.41(\%GC)-675/N - \%$ mismatch, where N is the primer length in bases (Liu et al., 2013). The trial-and-error approach used was incredibly time consuming and rather wasteful. With unlimited resources, many PCR experiments could have been run at once, with varying annealing temperatures, but an accurate method for calculating T_m would still be preferable.

One significant and easily avoidable setback was the mistaken mutation at A50. oTK172 sequence should have been checked before use, and the mutation should have been more thoroughly confirmed when sequencing was compared to wildtype. Figure 11 shows that the intended point

mutation TGC → TCC did in fact occur, but out of phase at a nearby duplicate triplet just 5 base pairs downstream, which is why the error was not immediately spotted. A more experienced researcher might have noticed, so sequencing results should have been checked by multiple people.

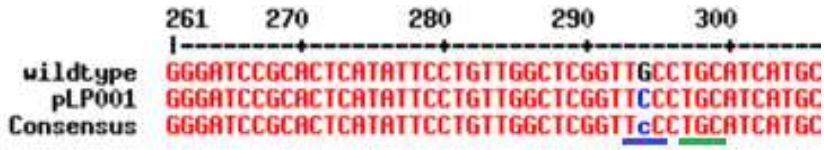

Figure 11: Multalin software showing the G to C substitution mutation in pLP001. Expected mutation was TGC → TCC, underlined in green. The actual TGC → TCC mutation underlined in blue was incorrectly identified as the desired mutation.

New England Biolabs suggest that KLD reactions typically take 5-60 minutes at room temperature, but in these experiments KLD at those conditions rarely led to successful transformation. Actual ambient temperature in the lab often measured above 30°C, so the reaction mixture was placed in a fridge. This proved successful, however, enzymatic activity is decreased at lower temperatures, so the reaction must be incubated for extended periods, further slowing rate of work. Some loss of DNA is expected following KLD, so it is important to begin with a potent PCR product shown by a strong band in the agarose gel. Unfortunately, even successful PCR often produces poor yield and must be retried, slowing workflow. KLD with low yield PCR product was attempted and never successful. KLD was performed overnight, and its outcome unknown until after another overnight transformation, thus a lot of time was lost over a typically simple process. The cause for poor KLD reaction success is unclear, but low DNA purity and degraded enzymes could be responsible.

Gel extraction is also notoriously inefficient (Carson et al., 2019, Fürst et al., 2018) and was avoided when possible; only in one instance of great PCR product was there enough DNA for successful

transformation after dissolution. It is suspected that most of the DNA is lost during washing. This was evident in PCR runs 21 and 24 seen in table 3, where despite strong bands in the agarose gel, DNA yield was too low for sequencing following gel extraction. Successful gel extraction is possible as demonstrated in run 25, but only with exceptional PCR yield.

Efforts were made to increase work rate by performing PCR reactions while waiting for sequencing results to be returned, however this was often little more than a waste of materials when the template plasmid was shown to contain undesired mutations.

Pilot protein overexpression results for pLP003 are pleasing. The fact it did not impact expression holds hope that future substitutions are likely to be as unimpactful on expression.

A successful spacer acquisition assay would provided valuable information regarding the impact of the C51S substitution on Cas1 function, especially because this cysteine was deemed the most likely to impact protein folding, so a result would have offered insight into the likelihood of success in other substitutions.

6. Further Work

The remaining substitutions can be continued to be made using the methods described above. Once accomplished, the objective mutant Cas1 can be overexpressed, purified, and assessed for DNA binding by a spacer acquisition assay and verified using an electrophoretic mobility shift assay (EMSA) or fluorescence anisotropy assay. If the modified Cas1 does bind to DNA as normal, then mStayGold fusion via GGS-linker can be attempted with work culminating in the use of TIRF and fluorescence resonance energy transfer (FRET) microscopy techniques to monitor Cas1 interactions *in vivo*.

7. References

- Bhawsinghka, N., Glenn, K. and Schaaper, R., 2020. Complete Genome Sequence of Escherichia coli BL21-AI. *Microbiol Resour Announc*., 9(10), 9-20. DOI: 10.1128/MRA.00009-20.
- Carson, S., Miller, H., Witherow, S. and Srougi, M., 2019. Lab Session 25 - Advanced CRISPR: Part I. *Molecular Biology Technique*, 4(1), 181-190, DOI: 10.1016/B978-0-12-815774-9.00025-3.
- Cline, J., Braman, C. and Hogrefe H., 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.,* 24(18), 3546-51. DOI: 10.1093/nar/24.18.3546.
- Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.*, 16 (22), 10881-10890.
- Fürst, M., Martin, C., Lončar, N. and Fraaije, M., 2018. Chapter Seven - Experimental Protocols for Generating Focused Mutant Libraries and Screening for Thermostable Proteins. *Methods in Enzymology*, 608(1), 151-187, DOI: 10.1016/bs.mie.2018.04.007.
- Isarankura-Na-Ayudhya, C., Nantasenamat, C., Tantimongcolwat, T. et al., 2009. Roles of Cysteine Residue on Chimeric Green Fluorescent Protein: Implications on Protein Solubilization and Fluorescent Property. *European Journal of Scientific Research,* 26(1), 577-593.
- Ivorra-Molla, E., Akhuli, D., McAndrew, M. et al., 2023. A monomeric StayGold fluorescent protein. *Nat Biotechnol.*, DOI: 10.1038/s41587-023-02018-w
- Killelea, T., Dimude, J., He, L. et al., 2023. Cas1-Cas2 physically and functionally interacts with DnaK to modulate CRISPR Adaptation. *Nucleic Acids Res.,* 51(13), 6914-6926, DOI: 10.1093/nar/gkad473.
- Lee, P., Costumbrado, J., Hsu, C. and Kim, Y., 2012. Agarose gel electrophoresis for the separation of DNA fragments. *J Vis Exp*, 1(62), 3923, DOI: 10.3791/3923.

- Liu, Y., Wu, T., Song, J. et al., 2013. A mutant screening method by critical annealing temperature-PCR for site-directed mutagenesis. *BMC Biotechnol.*, 13(1), 21, DOI: 10.1186/1472-6750-13-21.
- Lorenz, T., 2012. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J Vis Exp., 1*(63), 3998, DOI: 10.3791/3998.
- New England Biolabs, 1994. Vent DNA Polymerase [online]. Available at: https://www.med.upenn.edu/robertsonlab/assets/user-content/documents/vent-dna-polymerase-technical-data-sheet.pdf [Accessed 05 September 2024]
- New England Biolabs, 2024. NEBaseChanger [online]. Available at: https://nebasechanger.neb.com/ [Accessed 05 September 2024].
- New England Biolabs, 2024. Protocol for a Routine Vent PCR [online]. Available at: https://www.neb.com/en-gb/protocols/2012/09/06/protocol-for-a-routine-vent-pcr-reaction [Accessed 05 September 2024].
- New England Biolabs, 2024. Tm Calculator [online]. Available at: https://tmcalculator.neb.com/#!/main [Accessed 05 September 2024].
- Nuñez, J., Harrington, L., Kranzusch, P. et al., 2015. Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature*, 527(1), 535–538, DOI: 10.1038/nature15760
- Ren, H., Yang, B., Ma, C. et al., 2016. Cysteine Sulfoxidation Increases the Photostability of Red Fluorescent Proteins. *ACS Chemical Biology*, 11(10), 2679-2684, DOI: 10.1021/acschembio.6b00579
- Sigma Aldritch, 2024. Introduction to SDS-PAGE - Separation of Proteins Based on Size [online]. Available at: https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/protein-biology/gel-electrophoresis/sds-page [Accessed 05 September 2024]
- Yale University, 2024. The Coli Genetic Stock Center [online]. Available at: https://cgsc2.biology.yale.edu/Strain.php?ID=150015 [Accessed 05 September 2024].

- Hirano, M., Yonemaru, Y., Shimozono, S. et al., 2024. StayGold photostability under different illumination modes. *Sci Rep.,* 14(1), 5541, DOI: 10.1038/s41598-024-55213-3.