Genetic circuits for directed evolution of bioluminescent proteins

Macià Sureda Vives

Medical Research Council London Institute of Medical Sciences,
Institute of Clinical Sciences, Faculty of Medicine,
Imperial College London

Submitted for the degree of
Doctor of Philosophy

Supervised by Dr Karen Sarkisyan

April 2023

Statement of Originality

I hereby declare that this thesis and its research contents are the product of my own original work, conducted for my Doctor of Philosophy degree at Imperial College London. Any other work, figures, or intellectual contribution has been appropriately referenced or acknowledged in accordance with the standard referencing practises of the discipline.

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Abstract

Bioluminescent technologies are amongst the most commonly used tools for quantifying and visualising biological processes, and novel functions are being engineered for their application inside and outside the lab. Current methods to engineer bioluminescence-based tools rely on random or semi-random mutagenesis approaches which require intensive genotype screening, and are typically performed on expensive robotic workstations. In this project, we aimed to develop an *in vivo* platform to automate the screening of bioluminescent protein libraries using engineered bacteria. We repurposed a light-sensing circuit that allows individual bacteria to transduce the bioluminescence emission of an intracellular luciferase variant into expression of antibiotic resistance genes. This synthetic ability should allow bacteria expressing brightest luciferase variants to outcompete the rest of the population and become automatically selected by simply growing them under the appropriate antibiotic pressure. The current version of the genetic circuit confers modest but detectable growth and survival advantages to glowing cells; however, it requires further optimisation for robust performance.

Acknowledgements

I would like to start by thanking my supervisor, Karen Sarkisyan, for his mentorship and patience during these last four years, as I learned to endure and appreciate the bittersweet reality of scientific research. You have provided me with the creative freedom and independence to become a confident scientist, even if sometimes I complained about the autonomy that I now cherish. This thesis is our work, and I am grateful that you entrusted it to me. I feel prepared now to continue on this journey –wherever it might take me– in part because of you, and I will always remember this.

To the members of our group. Aubin, Ellie, and Feng, with whom I shared pipettes, frustrations, ideas, cheap meals, late nights and early mornings in the laboratory. Thank you for working in the dark to protect my experiments and listening to my peculiar music playlists always with a smile.

To Louisa González Somermeyer for helping me any time I consulted her, about the pCORE vectors she assembled, the genome integration protocol, and sorting bacterial cells.

To Bhavik Patel for training me in flow cytometry, and both designing and performing the double-sorting strategy for fluorescent bacteria presented herein.

To Remy Chait for pointing to me the importance of bacterial growth on circuit dynamics and familiarising me with some of the most relevant publications in the field.

To Carolina, Maryam, and Ksenia for becoming my temporary research family during the first lockdown, and with whom I shared one of the loveliest and most formative scientific experiences to date. It was the mindset and ethics that I saw in you while developing that antibody test that accompanied me throughout the engineering process of this other tool.

This thesis would not have been possible without plenty of other contributions, tangible or not, but it is those that were responsible for my happiness, attitude, and development that I am indebted the most. I hope that I am not missing anyone:

To Leo, Dani, Jennifer, Cassandra, and Tanara. You showed me the ropes at the institute when I first arrived, read my manuscripts and provided feedback, listened to my problems, trusted me, and I am not sure what else you did, but somehow, I always felt confident about myself after speaking to you.

To my London family: Ricardo, Stuart, and Tamara. What can I say here that I haven't said yet? I love you and I would not be or want to be here without you.

To Leire and Giada. Few people reached my heart as you did, especially at a moment when I needed friends the most. The pandemic ended when I met you girls, and since then your company has meant smiling and laughing (probably at something silly), and not feeling the burden of time.

To Doh, for reminding me that true friendship can flourish even in hopeless places, and helping become a more authentic version of myself.

To Jure, because embarking me into the world of synthetic biology is the least important thing you ever did for me, but the rest does not belong to these pages. To your unfailing love, *brate*.

To Jesús and Michalis with whom I share the pleasures of music, dancing, and literature, and most importantly my *confidants*.

To Ofri, for showing me that trust is the foundation of love. I began this journey by your side, and you blessed me with our memories. Paris will always be yours, just like a part of me.

To Agus and Toni Bauzà, for never aiming an inch lower than expected.

A n'es meus amics de s'infància: Jaume Mascaró, Albert, Maria, Tòfol, Inma, Jaume Duran, Ángel, i Nena. Gràcies per a fer-me sentir com si no hagués deixat mai sa vida de Mallorca enrere, ni jutjar sa meva decisió de partir, ni insistir en sa meva tornada. Per sa vostra insistència incansable de sempre voler veure'm. És sa vostra amistat sa que m'arrela aquí.

To Daniel Capó, my mentor and fatherly figure.

A tu Mamà, que m'ho has donat tot i segueixes donant-m'ho malgrat a vegades no sé si ho mereixo. Tot lo que he aconseguit és gràcies a tu, i a n'es padrí i sa padrina. Tot això ho has fet possible tu, que tingueres es coratge de tenir-me, que sempre m'has estimat en es sentit complet de sa paraula, i que mai no m'abandones ni deixes que m'abandoni. És meu cor sempre estarà a Porto Cristo amb voltros tres, i jo sempre estaré amb tu malgrat sa distància.

Manacor, 2023

Table of contents

List of figures

List of tables

Abbreviations

AmpR	Ampicillin Resistance
BRET	Bioluminescence Resonance Energy Transfer
CDS	Coding Sequence
CcaR	Complementary chromatic acclimation Regulator
CcaS	Complementary chromatic acclimation Sensor
CeNL	Cerulean enhanced Nano-Lantern
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
EmR	Erythromycin Resistance
FACS	Fluorescence-activated Cell Sorting
FAD	Flavin Adenine Dinucleotide
FMN	Flavin Mononucleotide
GAF	cGMP phosphodiesterase/adenylyl cyclase/FhlA
GeNL	Green enhanced Nano-Lantern
HK	Histidine Kinase
IPTG	Isopropyl-ß-D-thiogalactopyranoside
KanR	Kanamycin Resistance
LAA	Leucine-Alanine-Alanine
LB	Luria-Bertani
LED	Light-Emitting Diode
LOV	Light, Oxygen, or Voltage
LPA	Light Plate Apparatus
MIC	Minimal Inhibitory Concentration
MoClo	Modular Cloning
nnLuz	*Neonothopanus nambi* Luz
OD	Optical Density
OLuc	*Ophlophorus gracilirostris* Luciferase
Opto-T7RNA	Optogenetic T7 RNA Polymerase
PAGE	Polyacrylamide Gel Electrophoresis
PAS	Per-ARNT-Sim
paT7P	Photoactivatable T7 RNA Polymerase

PBS	Phosphate Buffered Saline
PCB	Phycocyanobilin
RBS	Ribosome Binding Site
RLuc	*Renilla reniformis* Luciferase
sfGFP	Superfolder Green Fluorescent Protein
T7RNAP	T7 RNA Polymerase
Tm	Melting Temperature
TcR	Tetracycline Resistance
TmpR	Trimethoprim Resistance
VVD	Vivid
YF1	YtvA-FixL hybrid protein 1
YNL	Yellow Nano-Lantern
YeNL	Yellow enhanced Nano-Lantern
ZeoR	Zeocin Resistance

1. Accelerating bioluminescence engineering with synthetic biology: a proposal

---

Bioluminescent technologies are useful for scientific discovery and slowly making their way for commercialisation outside the laboratory. Despite a renewed interest in this natural phenomenon and its technological applications, current methodologies for bioluminescence engineering are slow, labour-intensive or inaccessible to most laboratories. Here, a biological platform based on engineered light-sensing bacteria is proposed to automate these protocols with the intention to accelerate technological development in bioluminescence, and in other research areas by extension.

1.1 Background

1.1.1 Bioluminescence: a tool for scientific discovery

Since the emergence of life on Earth, the struggle for existence has driven a constant development and refinement of molecular structures and mechanisms to efficiently perform a plethora of functions. This pool of natural technologies, shaped for almost 4 billions years, and their complex interactions are progressively being unveiled by human curiosity and ingenuity. A distinctive trait of our species is that the discovery of knowledge is closely intertwined with its utilisation to drive innovation and solve problems, which in turn promotes further discovery. Therefore, it is not surprising that the natural world has served as an inextinguishable source of ideas and solutions from times immemorial that continues to propel human progress.

A phenomenon that has captivated humankind for millennia is the production of light in living organisms[1]. This process known as bioluminescence relies on small molecules called luciferins that can release energy in the form of light upon oxidation[2]. These molecules contain delocalised electron structures that reach excited states upon absorbing the energy freed during oxidation, which is then dissipated in a photon emission as the electron decays to the ground state (**Figure *1*A**). Numerous enzymes called luciferases have independently evolved across the tree of life to catalyse the oxidation of individual luciferins, conferring over 10,000 species with the ability to glow[3]. Thus, a vast array of light-emitting reactions exists in nature with distinct wavelengths, cofactor dependencies, and particular properties, depending on the ecological contexts in which they evolved.

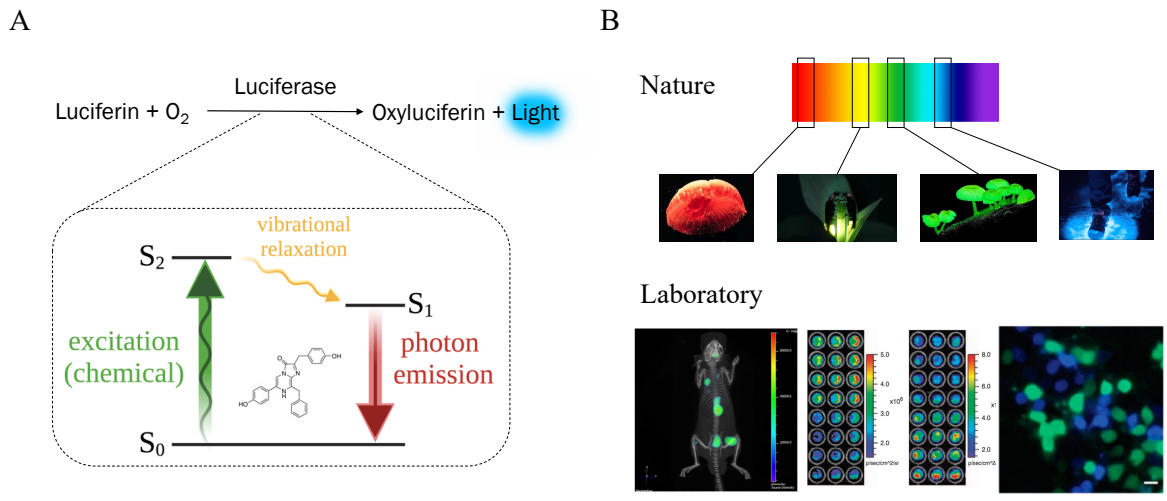


Figure 1 | A large diversity of bioluminescent reactions can be repurposed for technological innovation.
A Bioluminescent reactions only share in common their dependency on oxygen for light production, as they derive the chemical energy released during luciferin oxidation to reach light-emitting electronic states. **B** Natural bioluminescent elements need to be adapted to laboratory conditions and transformed into useful technologies for specific research contexts.

The utility of genetic elements with the intrinsic ability to produce light is conspicuous, as they can be coupled to other biological processes or entities to enable their visualisation and quantification. In fact, photon detectors are sensitive enough to detect even residual luciferase levels, resulting in methods with exceptional sensitivity that are compatible for high-throughput screening assays with miniaturised reaction volumes. Additionally, bioluminescent reactions can be modulated over a wide dynamic range, usually across 6 to 8 logs of luciferin concentration, resulting in high-resolution measurements. Finally, luciferases do not require the introduction of external light into samples for quantification, unlike fluorescent proteins, which avoids methodological disadvantages, such as sample heating, probe photobleaching, and background noise[4].

As a result, it was sufficient to isolate a few luciferase-luciferin pairs from nature and adapt them to laboratory conditions for bioluminescence to rapidly become a valuable reporter technology (**Figure *1*B**). For instance, it is now routinely used in gene expression assays[5], cell physiology[6], immunoassays[7], food analysis[8], drug screenings[9], and environmental monitoring[10]. Other more sophisticated bioluminescent technologies have allowed tracking epigenetic changes across multiple generations and quantifying molecular interactions inside living cells[11,12]. More recently, the development of sensitive digital cameras and bioluminescent reactions that can emit light through tissues of intact animals have also turned bioluminescence into a powerful *in vivo* bioimaging tool[13,14].

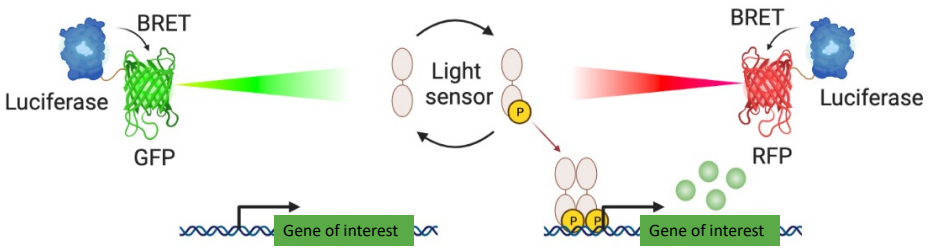
Collectively, bioluminescent reporters and imaging probes have contributed to the greater understanding of living systems that has been consolidated over the last few decades. This knowledge is enabling the engineering of biology towards tackling progressively more complex problems, which holds great promise for addressing some of the world's most pressing challenges and perhaps profoundly transforming society[15]. Consequently, it is not surprising that the potential of bioluminescence in engineering applications is presently being revisited for purposes other than measuring and observing life processes.

Inside the laboratory, luciferases are already being used as biological light sources to control physiological events and cell behaviour through light-sensitive proteins (**Figure 2A**). This approach has proven especially helpful to influence specific cell types located in inaccessible tissue regions, because luciferins are generally innocuous and can be easily distributed within the body, unlike other chemical inducers. In this context, bioluminescence has been used to induce apoptosis in cancer cells using a phototoxic fluorescent protein[16], and to repair neural networks by chronically stimulating engineered neural precursors in distinct

disease mice models, including Parkinson's[17], brain ischemia[18], and severe spinal cord injury[19]. Moreover, several proof-of-concept studies have programmed intracellular light-based molecular interactions to engineer complexity in synthetic gene circuits, as they are decoupled from the host interactome, and even for cell-to-cell optical communication[20,21].

Simultaneously, multiple companies have recently secured considerable funding to develop bioluminescent products for commercialisation outside the laboratory. The most notable entrepreneurial efforts include: Glowee, a Parisian start-up planning to reduce light pollution and electricity consumption with a living illumination system based on encapsulated bioluminescent bacteria; Light Bio, a USA-based venture creating glowing plants that reveal their inner processes with bioluminescence for home decoration; and Nyoka Design Labs, a Canadian clean-tech company selling cell-free bioluminescent sticks for recreation, fishing, and rescue operations (**Figure 2B**).

A



B

Figure 2 | Novel bioluminescent functions are emerging for applications inside and outside the lab.
A Luciferases are being used as biological light sources to control physiological events and cell behaviour through light-sensitive proteins. **B** Main start-up companies exploring the commercial viability of bioluminescence for sustainability and aesthetic purposes. Pictures and logos have been reproduced from their respective webpages: Glowee (www.glowee.com); Light Bio (www.light.bio); and Nyoka Design Labs (www.lightbynyoka.com).

1.1.2 The hurdles of bioluminescence engineering

The scope of innovation in bioluminescent technologies is limited to the light-emitting mechanisms that have been resolved and our capacity to adapt their elements for novel synthetic functions[22]. Despite the broad diversity of existing applications, they are all based on a few early discovered reactions, because the field of bioluminescence remains largely understudied. In fact, from the forty to a hundred independently evolved bioluminescent systems estimated to exist, only the structure of a dozen luciferin molecules, several luciferase gene families, and two luciferin biosynthesis pathways have been elucidated so far[23–25]

In the last decade, bioluminescent research has experienced a renaissance, leading to the discovery of new genes, substrates, and mechanisms, which have the potential to expand the bioluminescence toolkit in promising ways. Recent findings include the complete dissection of the fungal bioluminescent system and its luciferin biosynthesis genes[26–28]; the structure elucidation and organic synthesis of the luminous earthworm *Fridericia heliota*'s luciferin[29]; the luciferase-luciferin pair of the fireworm *Odontosyllis undecimdonta*[30–32]; the initial purification of the bristle worm *Chaetopterus variopedatus'* reaction components[33]; and the determination of the bacterial fatty acid reductase multiprotein complex structure[34]. This resurging interest in bioluminescence and its biotechnological opportunities have fostered international collaborations aimed at systematically uncovering nature's palette of light-emitting reactions[35]. Consequently, a wealth of new knowledge and bioluminescent components are anticipated to become available in the coming decades.

Once isolated from nature, these bioluminescent genes cannot be converted into useful applications until they have been adapted to function outside their host organisms and optimised to meet the latest research needs. Traditionally, this has been and still is a laborious and expensive process, due to the unpredictable relationship between protein sequence and function, which reduces protein engineering to mutagenesis approaches. In this process, mutations are introduced in a naturally occurring genetic sequence and evaluated at the phenotype level to identify protein variants with desired features. This is relatively simple for luciferases because even residual bioluminescence activities can be detected in living cells with electronic light sensors[4]. Consequently, bioluminescent protein engineering initially consists of increasing the stability of wild-type luciferases to improve their heterologous expression[36,37]. Further efforts usually focus on transforming the intrinsic properties of the enzyme, such as the intensity and duration of the activity[38], the structural complexity[39], the colour of the light emission[40,41], and the substrate specificity[14,38]. Finally, improved luciferases can be fused to

other molecules[42], circularly-permuted[43], and split to create proteins with novel synthetic properties[44,45].

The first generation of luciferase improvements was advanced by rational designs when the technological capacity to create and screen mutants was considerably low[36,37]. However, available structure-function knowledge is often insufficient to effectively engineer new proteins, as mutations in closely-related luciferases are not always synonymous due to their evolutionary history. Furthermore, these methods are incompatible with accessing genotypes that differ substantially from the original protein sequence, which usually correspond to high-performing variants and new enzymatic properties. Instead, systematic approaches based on random and site-saturation mutagenesis have traded off the requirement of previous knowledge for higher number of mutants screened[46]. To maximise design variables, protein changes are usually introduced in sequential cycles of mutant library construction and screening of best performing variants. In each round, candidates are selected for further mutagenesis to progressively modify the protein function in a specific direction, imitating the manufacturing process of nature, hence called directed evolution.

Some notable examples are the red-shifted *Renilla reniformis* luciferase variants, which enabled sensitive imaging in living organisms using bioluminescence[40]; NanoLuc, a synthetic luciferase evolved from the small catalytic subunit of *Ophlophorus gracilirostris* luciferase with an improved substrate analogue, resulting in a reaction ~ 2.5 million-fold brighter and resistant to catalytic autoinhibition[38]; and the more recent Akaluc, a red-shifted firefly luciferase capable of imaging single cells in deep tissues of freely moving animals[14].

Despite the powerful engineering potential of directed evolution approaches, their capacity to explore the genotype space in search for improved protein variants is limited to the number and diversity of mutants that can be screened in each selection round. In the last example mentioned, 21 mutagenesis cycles were required to evolve a luciferase that could be detected in deep tissues of freely moving animals, which accounted for 1.3 million colonies on 3,200 plates being screened[14]. To this day, these methods have continued to rely on expensive, robotic workstations for automation or laborious protocols, limiting their widespread application and typically requiring several years for the development of improved bioluminescent proteins[47]. Therefore, there is an impending need for screening methods that are high-throughput, automated, cheap, and easily implementable to accelerate the development of novel bioluminescent technologies.

1.2 Project proposal

1.1.2 Leveraging the automation advantages of biological devices

This project intended to accelerate bioluminescence engineering by simplifying the screening process of directed evolution methodologies. As mentioned above, each directed evolution cycle involves two experimental stages: genotype diversification, typically achieved using *in vitro* mutagenesis approaches; and bioluminescence selection, which involves isolating individual genotypes, quantifying their light outputs, and rationally deciding which variants might constitute the best templates for the following evolution cycle based on that information. Importantly, bioluminescent genotypes selected in each round configure the evolutionary routes that can be explored in the following rounds.

The most labour-intensive steps in these protocols reside in the bioluminescence selection stage. Genotype isolation can be automated using colony picker robots that inoculate microplate cultures with individual colonies, allowing the preparation of glycerol stock libraries. Bioluminescence screening is more complicated, because enzymatic function is influenced by a wide range of biological variables, which can easily conceal small gains in enzymatic function. This is usually accomplished by taking preventive measures to reduce variability across conditions, such as normalising culture density before quantification, and taking replicated measurements to account for inter- and intra-experimental variability. In this case, reproducibility and high-throughput are usually achieved using liquid handling robots.

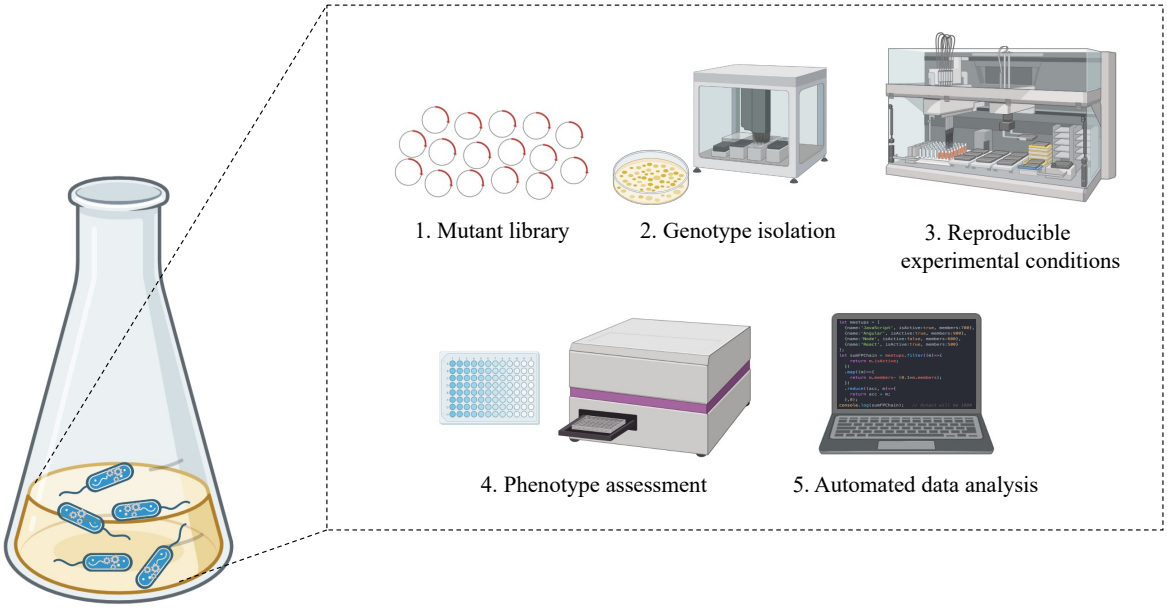


Figure 3 | Leveraging the automation advantages of biological devices with synthetic biology.
This project intended to automate, fully or partially, the experimental steps of established directed evolution protocols for bioluminescent proteins using engineered biological systems.

Finally, bioluminescence activities are typically recorded with a microplate reader and data analysis automated with a computer script (**Figure 3**). Thus, the number of genotypes that can be analysed in an evolution cycle depends on the cultures that can be grown and analysed in a reproducible manner to cross-compare results amongst plates. Evidently, this is increasingly complicated without some degree of automation.

We propose that a biological platform can be engineered to automate, fully or partially, the screening of bioluminescent mutant libraries, and that it would be more suited for the task than robotic devices and complex instrumentation. Our proposal rests on the arguments that a light-sensing circuit can be constructed to screen for user-defined bioluminescent activities, while cell compartmentalisation would allow performing this operation on each luciferase variant inside individual bacteria, simultaneously. As a result, the need for isolating genotypes would be removed and bioluminescence analysis automated with a substantial increase in throughput and an even greater reduction in experimental complexity and costs.

1.2.2 Conceptualising an *in vivo* selection platform for bioluminescence

The information processing capacity needed to screen an array of bioluminescent proteins and identify improved phenotypes is relatively simple: the light output of each protein candidate must be quantified separately and compared against a reference value for assessment. The algorithm that describes it involves two functions, a measurement and a logic statement:

(1) quantify light signal x
(2) discard if $x < y$ is True, where y is a reference value

Bacterial populations are exceptionally qualified for the first task, because they can compartmentalise each protein candidate in a cell and simultaneously run the same operation millions or billions of times in a very small culture volume. In this case, quantifying the light signal of the bioluminescent protein. Enforcing the outcome of the operation is more complicated because it involves sorting out all cells expressing a protein that performed unsatisfactorily. As hallmarked by evolution, environmental pressures can exert selective control better than internal cell regulatory mechanism, because they are more stable, especially in laboratory conditions. Therefore, as long as bioluminescent activity can be proportionally translated into fitness acquisition, an appropriate environmental challenge should be able to exclude all non-desirable phenotypes from the population. Importantly, this design enables user control of selection by simply changing the composition of the growth medium.

The proposed design relies on converting a type of information that cannot be operated by biological systems into one that can be, a light signal into a chemical signal. This process is called transduction, and reformulates the initial algorithm:

(1) convert light signal x into cell fitness y_1, where y_1 is primarily defined by a single gene

(2) discard if $y_1 < y_2$ is True, where y_2 is an environmental challenge exerted by the presence or absence of a single chemical

Evolution has already crafted a wide range of gene circuits to perform this exact operation, referred herein as light-sensing transcriptional circuits. They generically consist of a photoreceptor protein capable of regulating a series of biochemical reactions inside the cell in response to an extracellular light signal, which ultimately lead to specific gene expression. Therefore, this type of circuit could potentially be repurposed to sense light produced intracellularly by each luciferase variant, and couple that activity to the expression of an essential gene needed to grow or survive under specific culture conditions (**Figure 4A**).

In this context, bacteria expressing the brightest luciferase variants would outcompete the rest of the population and become automatically selected by simply growing a mutant cell library under the appropriate environmental pressure. This selection platform would allow users to modulate the strength of selection by tuning external parameters, such as the concentration of the environmental challenge (e.g., antibiotic concentration) or the capacity of cells to produce bioluminescence (luciferin concentration) for fitness acquisition. Furthermore, post-selection enrichment could be reported by simply adding luciferin again to the culture before plating bacteria to recover improved mutants (**Figure 4B**).

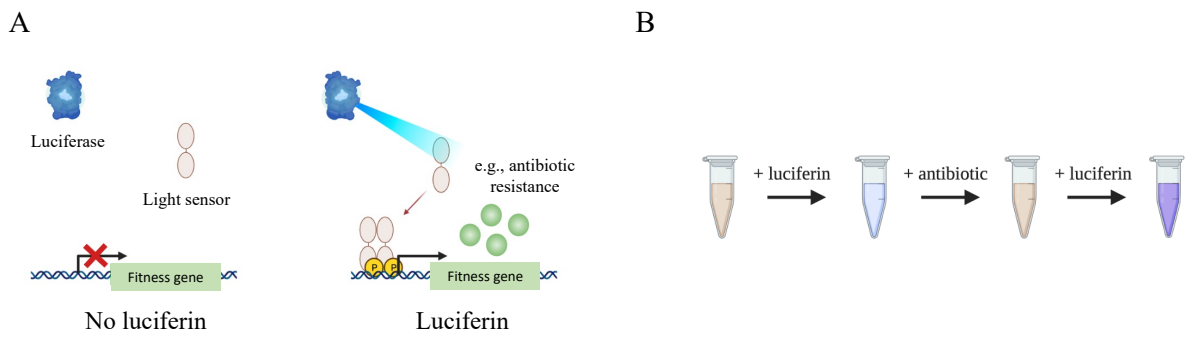


Figure 4 | Light-sensing circuit for high-throughput *in vivo* screening of bioluminescent proteins.
A Repurposed light-sensing circuit enables automatic selection of brightest luciferase variants in mutant libraries by transducing intracellular bioluminescence activity into the expression of an essential gene. **B** Two externally controlled parameters (luciferin and selective pressure) allow tunning the selection strength and reporting on post-enrichment results.

If successful, this platform would allow virtually any laboratory to develop novel bioluminescent technologies without the need of expensive automation machinery or complex protocols at a relatively fast pace. We believe that this technology could democratise protein engineering in bioluminescence at a time of increased interest in the field, and propel innovation and knowledge generation in plenty other research areas, as showcased by the past and present contribution of bioluminescent tools in scientific discovery.

1.3 Objective and milestones

The objective of this project was to automate, fully or partially, the experimental steps of established directed evolution protocols for bioluminescent proteins using engineered bacterial cells. We intended to achieve this by developing an *in vivo* bioluminescence selection platform that could couple the growth or survival of individual bacteria to their respective intracellular luciferase activities. Thus, brightest luciferase variants would become automatically selected from the rest of the population by simply growing mutant libraries in the presence of luciferin and an appropriate selective pressure.

According to our plan, the development of this *in vivo* bioluminescence selection platform would involve completing the following milestones:

I. Identifying a light-sensing transcriptional circuit with promising signalling properties for bioluminescence selection that can be reliably expressed in bacterial cells.
II. Repurposing the circuit to exclusively sense intracellular light emitted by a luciferase to guarantee that bioluminescence selection can be performed in each cell without interferences from neighbouring cells.
III. Engineering a selection strategy based on a chemical selective pressure and its counterpart selectable gene that can differentiate a practical range of bioluminescent activities for enriching luciferase mutant libraries.

Other secondary goals that were considered in this project included:

i. Collecting extensive characterisation data on circuit signalling dynamics during bacterial growth to later guide the design of selection protocols.
ii. Evaluating the potential reusability of this platform in future bioluminescence engineering projects once developed.
iii. Identifying additional selection mechanisms to suit a wider range of engineering needs.
iv. Testing alternative strategies to partially automate some of the steps of current directed evolution methods (in case the main objective of the project turned out not to be feasible).

1.4 Thesis outline

The remainder of this thesis is composed of the following chapters:

- *Chapter 2* contains a comprehensive description of the materials and methods used to engineer and evaluate the light-sensitive genetic circuits and bacterial strains used throughout this project. Gene and protein sequences corresponding to the main biological parts used to develop this technology have also been included therein.

- *Chapter 3* describes the initial exploratory phase that led to the identification of two light-controlled transcriptional circuits for later repurposing to sense intracellular light. A special emphasis was placed on characterising circuit performance during bacterial growth, which revealed novel insights on the signalling dynamics of two of the most utilised bacterial light-sensing circuits in the field of synthetic biology.

- *Chapter 4* explains the approaches that were taken to engineer a bioluminescence transduction pathway using the previously validated circuits. It also presents a photoreceptor fusion topology that functions as a modular transduction platform for bioluminescent proteins with a wide dynamic range, external tunability, and extended spectral compatibility.

- *Chapter 5* explores distinct antibiotic-based selection strategies and a FACS-based screening protocol for bioluminescent proteins using the bioluminescence transduction circuit previously engineered.

- *Chapter 6* provides a broad summary of this thesis and its major contributions, and discusses the final circuit adjustments needed to finish developing our *in vivo* bioluminescence selection platform. The closing pages also offer future prospects on some promising uses for this tool.

2. Materials and methods

A set of plasmids and biological parts were used in combination with routine microbiology and genetic engineering methods to create and manipulate the light-sensitive *E. coli* strains used in this project. Multiple protocols were created or adapted from previous articles to interrogate these strains using external light or bioluminescence pulses and track their associated gene expression responses with a microplate reader or flow cytometry analysis.

2.1 Bacterial manipulation

2.1.1 Strains and bacterial culture

A commercial *E. coli* strain called E. cloni® 10G (LGC #60107) was used for molecular cloning and in all experiments unless otherwise indicated. It is an *E. coli* K12 strain containing common genomic modifications of standard cloning strains, such as DH5α or TOP10. Experiments with the Opto-T7RNAP*(563) circuit were performed in a similar but less modified *E. coli* K12 strain, BW29655, that was unable to catalyse arabinose needed for the induction of the sensor expression[48]. Although both strains had deficient arabinose metabolism, BW29655 was more closely related to the strain used in the original Opto-T7RNAP*(563) publication and it was selected to reduce host context-dependent effects[49] (**Table 1**).

Table 1 | *E. coli* K12 strains used in this project and their genetic modifications

Strain name	Genome
E. cloni® 10G	F-, mcrA, Δ(mrr-hsdRMS-mcrBC), endA1, recA1, Φ80dlacZΔM15, ΔlacX74, araD139, Δ(ara,leu)7697, galU, galK, rpsL, nupG, λ-, tonA (StrR)
BW29655	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, Δ(envZ-ompR)520(::FRT), Δ(rhaD-rhaB)568, hsdR514

These strains were always plated on Luria-Bertani (LB) agar plates containing the necessary antibiotics for transformants selection. Single colonies, or glycerol stocks prepared from single-colony cultures, were always used to start liquid cultures. Cultures were grown overnight in LB medium with the required supplementation at 37ºC and 220 rpm in a dark incubator, unless stated otherwise. Glycerol stocks were always prepared using cultures started from single colonies by mixing 0.5 mL aliquots in equal proportions with 50% glycerol (ThermoFisher #A16205-AP), and stored at -80ºC in cryogenic tubes.

In *Chapter 3*, cultures expressing split, photoactivatable T7 RNA polymerase-based sensors were grown at some point at 28ºC to discard thermostability concerns about the folding of Vivid domains. Distinct protocol variations were tested: 22-hour overnight growth at 28ºC and light-induction at 37ºC; light-induction at 28ºC exclusively, and both overnight expansion and light-induction at 28ºC.

All cultures used to perform experiments with CcaS-CcaR and pDusk/pDawn strains were inoculated at known concentrations using seeding aliquots prepared using a previously published protocol[50]. This method has been explicitly designed to both reduce

interexperimental variability and achieve optimal results with strains expressing light-sensitive two-component systems. Essentially, the protocol consists of growing the strains in fast-dividing conditions to cause maximal intracellular dilution of active circuit components and signal output and prepare glycerol aliquots of known optical density (OD) with these washed-out cells. This was achieved by initially creating single-colony glycerol stocks that had been stopped before exponential phase exit (OD ~ 0.1-0.2) and using them to inoculate fresh cultures for seeding aliquot preparation. These second generation cultures were grown again until similar conditions were reached, and the final OD was annotated before placing them in an ice-water bath for approximately 15 minutes. Then, cultures were mixed at a 70:30 volume ratio with 50% glycerol, aliquoted into microtubes of different volumes to suit a wider variety of experiments, and stored at -80ºC. Finally, the measured OD was multiplied per 0.7 to account for the glycerol dilution and used as a reference to start experiments at desired seeding ODs. On a personal observation, CcaS-CcaR cultures retained less intracellular activation when stopped at OD values closer to 0.1 when preparing seeding aliquots and red-light illumination had no additional benefits over darkness during fast-dividing conditions.

2.1.2 Medium supplements

Several antibiotics were used to supplement agar plates and liquid cultures for plasmid selection or maintenance, as well as to challenge bacteria in fitness experiments in *Chapter 5*. They were purchased as lyophilised powder and resuspended with appropriate solvents to prepare sterile, filtered stock solutions at 1000x the recommended concentration for plasmid selection in *E. coli* (**Table 2**).

Table 2 | Antibiotics compounds used for plasmid selection and fitness experiments

Antibiotic	Commercial supplier	Solvent	Working concentration
Carbenicillin	Formedium #CAR0005	ddH_2O	100 μg/mL
Chloramphenicol	Formedium #CLA01	Ethanol	25 μg/mL
Kanamycin	ThermoFisher #11815024	ddH_2O	50 μg/mL
Spectinomycin	Sigma-Aldrich #S4014	ddH_2O	50 μg/mL
Zeocin	APExBIO #C4556	ddH_2O	20 μg/mL
Tetracycline	Sigma-Aldrich #T7660	ddH_2O	50 μg/mL
Erythromycin	Sigma-Aldrich #E5389	Ethanol	100 μg/mL
Trimethoprim	Sigma-Aldrich #T7883	DMSO	50 μg/mL

In *Chapter 3*, expression of Opto-T7RNAP*(563) sensor fragments or fused T7RNAP fragments from CAP-deficient P_{BAD} promoters, and *wild-type* T7RNA from the full-length P_{BAD} promoter, was induced by supplementing cultures at 0.1% with L-arabinose (Millipore #178680) based on the calibration data from the authors[51]. In the troubleshooting of the impaired light-regulation of Vivid-based sensors cultures were also supplemented with a gradient of flavin adenine dinucleotide (FAD) (ThermoFisher #347212500) concentrations during both overnight and light-induction experiments.

In *Chapter 5*, the failed LacI-repression of the iLux operon in both the original plasmid and the plasmid used for co-expression with CcaS-CcaR or pDawn was demonstrated by supplementing cultures with or without 1 mM Isopropyl-ß-D-thiogalactopyranoside (IPTG) (ThermoFisher #15529019). Additionally, cultures co-expressing pDawn circuit and iLux were supplemented with a gradient of D-glucose (Sigma-Aldrich #G8270) concentrations to increase the recycling of bioluminescent co-factors for enhanced photon production. Finally, these same cultures were also supplemented with 0.1% glucose and a gradient of decanal (Sigma-Aldrich #D7384) concentrations in an attempt to reduce the metabolic burden of the bioluminescent reaction.

2.1.3 *E. coli* transformation

Routine *E. coli* transformations were performed with chemically competent bacteria prepared using the Mix & Go! *E. coli* transformation kit (Zymo Research #T3001). They were prepared by diluting an overnight LB culture 1:100 in ZymoBroth™ (Zymo Research #M3015) and growing it at 37ºC, 240 rpm, and dark conditions when applicable until an OD = 0.4-0.6. Then, cells were harvested and treated with the kit buffers following the supplier instructions while strictly working on ice throughout the procedure. Chemically competent aliquots were either used fresh immediately or stored at -80ºC.

Heat-shock transformations were performed using 50 µL of chemically competent bacteria per reaction. Cells were thawed on ice for around 10 minutes and mixed with either 1-20 ng of purified DNA plasmid, 5 µL of Golden Gate or KLD reaction product, or 2 µL of Gibson assembly product, by stirring with the pipette tip. It is important that no vigorous agitation or pipetting up and down occurs, as cells membranes are relatively fragile at this point. After introducing the DNA, mixtures were incubated on ice for 10 minutes, heat-shocked for 45 seconds at 42ºC on a dry thermostat block (Biosan #TDB-120), and immediately transferred back on ice for 5 additional minutes. Then, cells were recovered for 1 hour in 950

μL of pre-heated SOC medium at 37ºC and 220 rpm in the dark. Finally, 50-100 μL were spread on LB agar plates with the required antibiotics using sterile glass plating beads.

2.2 Genetic engineering

2.2.1 Plasmids and biological parts

Plasmids used in this project were either kindly shared by the authors, purchased on Addgene, assembled from synthetic DNA fragments, or edited in the laboratory (**Table 3**).

In *Chapter 3*, a collection of cloning vectors containing distinct combinations of origins of replication and antibiotic selectable markers assembled by Dr Louisa González Somermeyer were validated for utilisation in this project. This vectors named pCORE contain the minimal genetic structure for molecular cloning using Golden Gate assembly methods, and can act both as acceptor backbones and release inserts for later assemblies (unpublished). All genetic circuits assembled from synthetic DNA fragments or requiring a plasmid backbone change for co-expression compatibility were cloned into these vectors:

i. The three circuits expressing each paT7P sensor were reproduced as accurately as possible from the sequences deposited in the GenBank database (KX980034-36).
ii. The circuit for Opto-T7RNAP*(563) sensor expression was reconstructed using the reported circuit description and the individual genetic parts published in the Supporting Information of the article[49].
iii. The bicistronic reporter used to validate these four T7RNAP-based light-sensors (P_{T7} – BBa_B0034 – CAT – x4 stop codons – BBa_B0034 – sfGFP – BBa_B0015), which substituted the originally used GFPuv reporter in paT7P circuits[52] and the mCherry gene in Opto-T7RNP*(563)[49].
iv. The reconstructed pDawn circuit was partially modified from the original design for two different reasons. First, the red fluorescent gene DsRed-Express 2 was not available due to existing conflicts of interest and was replaced by a bicistronic reporter similar to the one previously used for T7RNAP-based sensors (BBa_B0029 – mScarlet – x1 stop codon – BBa_B0034 – CAT). Second, terminator sequences were not annotated in the reported sequence and could not be identified with certainty. Hence, two strong synthetic terminators L3S1P56 and L3S2P21[53] were used to stop the transcription of the YF1-FixJ bicistronic unit and the reporter

gene, respectively; and the double terminator BBa_B0015 was placed downstream of the lambda repressor gene.

v. The iLux operon was re-localised from the original pGEX(-) vector into a backbone compatible for co-expression with CcaS-CcaR and pDusk/pDawn by amplifying the entire transcriptional unit with primers containing type II restriction site sequences with pCORE-compatible overhangs.

Two sequences were built into the MoClo level 2 acceptor vector (pAGM4673)[54] prior to the assembly and validation of the pCORE collection. These were the transcriptional unit expressing the fused T7RNAP*(563) fragments and the sequence used to create the reporter strain, which contained a T7RNAP-specific mScarlet reporter (P_{T7} – BBa_B0034 – mScarlet – BBa_B0010) and a ZeoR cassette flanked by 600 bp homology arms with a PAM mutation for CRISPR/Cas9-assisted genome integration into the *E. coli* SS9 site.

All other plasmids were generated by editing pre-existing plasmids in the laboratory through deletions, insertions, or substitutions. Except for the array of CcaS fusion protein designs tested in *Chapter 4*, the rest of modifications were done using previously characterised mutations and biological parts (**Table 4**).

Table 3 | Main plasmids used in this project

Name (link)	Description	Backbone	Source
Acceptor vectors			
pCORE_16	MoClo-like cloning vector with mScarlet selection marker	pBR322 AmpR	This group (unpublished)
pCORE_17	MoClo-like cloning vector with mScarlet selection marker	pBR322 ChlR	This group (unpublished)
pCORE_18	MoClo-like cloning vector with mScarlet selection marker	pBR322 KanR	This group (unpublished)
pCORE_19	MoClo-like cloning vector with mScarlet selection marker	pBR322 SpecR	This group (unpublished)
pCORE_20	MoClo-like cloning vector with mScarlet selection marker	pBR322 ZeoR	This group (unpublished)
pCORE_26	MoClo-like cloning vector with mScarlet selection marker	p15A AmpR	This group (unpublished)
pCORE_27	MoClo-like cloning vector with mScarlet selection marker	p15A ChlR	This group (unpublished)
pCORE_28	MoClo-like cloning vector with mScarlet selection marker	p15A KanR	This group (unpublished)
pCORE_29	MoClo-like cloning vector with mScarlet selection marker	p15A SpecR	This group (unpublished)
pCORE_30	MoClo-like cloning vector with mScarlet selection marker	p15A ZeoR	This group (unpublished)

pCORE_36	MoClo-like cloning vector with mScarlet selection marker	pSC101 AmpR	This group (unpublished)
pCORE_37	MoClo-like cloning vector with mScarlet selection marker	pSC101 ChlR	This group (unpublished)
pCORE_38	MoClo-like cloning vector with mScarlet selection marker	pSC101 KanR	This group (unpublished)
pCORE_39	MoClo-like cloning vector with mScarlet selection marker	pSC101 SpecR	This group (unpublished)
pCORE_40	MoClo-like cloning vector with mScarlet selection marker	pSC101 ZeoR	This group (unpublished)
pAGM4673	MoClo level 2 acceptor backbone	pBR322 KanR	Addgene (48014)[54]
Split, photoactivatable T7RNAP-based circuits			
pCORE_16_P_{T7}-cat-sfgfp	Bicistronic reporter expressing chloramphenicol acetyltransferase and sfGFP from P_{T7}	pBR322 AmpR	This project
pCORE_20_paT7P-2	Constitutive bicistronic expression of paT7P-2 fragments from P_{CAT}	pBR322 ZeoR	This project
pCORE_26_P_{T7}-cat-sfgfp	Bicistronic reporter expressing chloramphenicol acetyltransferase and sfGFP from P_{T7}	p15A AmpR	This project
pCORE_30_paT7P-2	Constitutive bicistronic expression of paT7P-2 fragments from P_{CAT}	p15A ZeoR	This project
pCORE_36_P_{T7}-cat-sfgfp	Bicistronic reporter expressing chloramphenicol acetyltransferase and sfGFP from P_{T7}	pSC101 AmpR	This project
pCORE_37_Opto-T7RNA	Opto-T7RNA*(563) fragments expressed from two separate CAP-deficient P_{BAD}	pSC101 ChlR	This project
pCORE_40_paT7P-2	Constitutive bicistronic expression of paT7P-2 fragments from P_{CAT}	pSC101 ZeoR	This project
pTara:500	Full-length T7RNAP expressed from P_{BAD} and constitutive AraC expressed from P_C on the complementary strand in the opposite direction	p15A ChlR	Addgene (60717)[55]
pAGM4673_T7RNAP_fusion	Fused T7RNAP*(563) with GSGSGGSG peptide linker expressed from CAP-deficient P_{BAD}	pBR322 KanR	This project
pAGM4673_P_{T7}-mScarlet_ZeoR_SS9	mScarlet expression from P_{T7} and constitutive ZeoR cassette flanked by 600 bp homology arms for *E. coli* SS9 genomic site with PAM mutation	pBR322 KanR ZeoR	This project
pSIM5	Multicistronic expression of recombineering genes λ *red* (*gam*, *bet*, and *exo*) from P_L under the control of temperature-sensitive CI857 repressor	pSC101 ($repA^{ts}$) ChlR	Court Lab[56]
pX2-Cas9	Cas9 expression from P_{BAD} and constitutive AraC expressed from P_C on the complementary strand in the opposite direction	pBBR1 KanR	Addgene (85811)[57]
pSS9_RNA	Constitutive expression of gRNA targeting the SS9 genomic site of *E. coli* K12 strains	pBR322 AmpR	Addgene (71656)[57]
CcaS-CcaR plasmids			
pSR43.6r	Constitutive expression of CcaS and metabolic enzymes Ho1 and PcyA	p15A SpecR	Addgene (63197)[58]
pNO286-3	Constitutive expression of miniCcaS#10 and metabolic enzymes Ho1 and PcyA	p15A SpecR	Addgene (107746)[59]
pSR43.6r_ΔCcaS	ΔCcaS (GG cloning site) and constitutive expression of metabolic enzymes Ho1 and PcyA	p15A SpecR	This project

pSR43.6r_ ΔN23CcaS	Constitutive expression of ΔN23-CcaS and metabolic enzymes Ho1 and PcyA	p15A SpecR	This project
pSR43.6r_ Fusion_1	nnLuz_v4-SGLRS-CcaS in pSR43.6r_ ΔCcaS	p15A SpecR	This project (confidential)
pSR43.6r_ Fusion_2	nnLuz_v1-SGLRS-GeNL-GHGTGSTGSGSS-CcaS in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_3	nnLuz_v1(N1-39)-SGLRS-GeNL-GHGTGSTGSGSS-CcaS in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_4	GeNL-SGLRS-ΔN23CcaS in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_5	GeNL-GHGTGSTGSGS-ΔN23CcaS in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_6	RLuc8.6(535)-SGLRS-ΔN23CcaS in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_7	CcaS(N1-221)-SGLRS-GeNL-SGLRS-CcaS(531-C753) in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_8	CcaS-SGLRS-GeNL in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_9	CcaS-GHGTGSTGSGSS-GeNL in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_10	miniCcaS#10-SGLRS-GeNL in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_10_ Y59G	miniCcaS#10-SGLRS-GeNL (mNeonGreen^{Y59G}) in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pSR43.6r_ Fusion_10_ R167A	miniCcaS#10-SGLRS-GeNL (NanoLuc^{R167A}) in pSR43.6r_ ΔCcaS	p15A SpecR	This project
pNO286-3_link1_CS	Constitutive expression of miniCcaS#10-SGLRS-(GG cloning site) and metabolic enzymes Ho1 and PcyA	p15A SpecR	This project
pNO286-3_link2_CS	Constitutive expression of miniCcaS#10-GHGTGSTGSGSS-(GG cloning site) and metabolic enzymes Ho1 and PcyA	p15A SpecR	This project
pNO286-3_link1_ RLuc8.6-535	RLuc8.6-535 in pNO286-3_link1_CS	p15A SpecR	This project
pNO286-3_link2_ RLuc8.6-535	RLuc8.6-535 in pNO286-3_link2_CS	p15A SpecR	This project
pSR58.6	Constitutive expression of CcaR and inducible sfGFP from $P_{cpcG2\text{-}172}$	ColE1 ChlR	Addgene (63176)[58]
pSR58.6_CS	Constitutive expression of CcaR and inducible sfGFP from $P_{cpcG2\text{-}172}$ with extra GG cloning site upstream of CcaR	ColE1 ChlR	This project
pSR58.6_ nnLuz_v3	BBa_J23105-B0032-nnLuz_v3-L3S2P21 in pSR58.6_CS	ColE1 ChlR	This project (confidential)
pSR58.6_ RLuc8.6-535	BBa_J23105-B0032-RLuc8.6-535-L3S2P21 in pSR58.6_CS	ColE1 ChlR	This project
pSR58.6_ ΔsfGFP_CS	Constitutive expression of CcaR with empty $P_{cpcG2\text{-}172}$ reporter (GG cloning site) and extra GG cloning site upstream of CcaR	ColE1 ChlR	This project

pSR58.6_ mScarlet-I_ CS	Constitutive expression of CcaR and inducible mScarlet-I from $P_{cpcG2\text{-}172}$ with extra GG cloning site upstream of CcaR	ColE1 ChlR	This project
pSR58.6_ mScarlet-I_ YNL	BBa_J23105-B0032-YNL-L3S2P21 in pSR58.6_ mScarlet-I_CS	ColE1 ChlR	This project
pSR58.6_ mScarlet-I_ GeNL	BBa_J23105-B0032-GeNL-L3S2P21 in pSR58.6_ mScarlet-I_CS	ColE1 ChlR	This project
pSR58.6_ mScarlet-I_ YeNL	BBa_J23105-B0032-YeNL-L3S2P21 in pSR58.6_ mScarlet-I_CS	ColE1 ChlR	This project
pSR58.6_ AmpR_sfGF P_CS	Bicistronic reporter expressing TEM-115 (AmpR) and sfGFP with BBa_B0034 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR AmpR	This project
pSR58.6_ KanR_sfGFP _CS	Bicistronic reporter expressing NptII (KanR) and sfGFP with BBa_B0034 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR KanR	This project
pSR58.6_ ZeoR_sfGFP _CS	Bicistronic reporter expressing *Sh* Ble (ZeoR) and sfGFP with BBa_B0034 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR ZeoR	This project
pSR58.6_ TcR_sfGFP_ CS	Bicistronic reporter expressing TetC (TcR) and sfGFP with BBa_B0034 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TcR	This project
pSR58.6_ EmR_sfGFP _CS	Bicistronic reporter expressing ErmC (EmR) and sfGFP with BBa_B0034 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR EmR	This project
pSR58.6_ TmpR_sfGF P_CS	Bicistronic reporter expressing $DHFR^{L28R}$ (TmpR) and sfGFP with BBa_B0034 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TmpR	This project
pSR58.6_ KanR_100	Single reporter NptII (KanR) with BBa_B0034 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR	This project
pSR58.6_ KanR_60	Single reporter NptII (KanR) with BBa_B0030 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR KanR	This project
pSR58.6_ KanR_30	Single reporter NptII (KanR) with BBa_B0032 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR KanR	This project
pSR58.6_ KanR_7	Single reporter NptII (KanR) with BBa_B0031 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR KanR	This project
pSR58.6_ KanR_1	Single reporter NptII (KanR) with BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR KanR	This project
pSR58.6_ AmpR_1	Single reporter TEM-115 (AmpR) with BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR AmpR	This project
pSR58.6_ ZeoR_1	Single reporter *Sh* Ble (ZeoR) with BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR ZeoR	This project

pSR58.6_TcR_1	Single reporter TetC (TcR) with BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TcR	This project
pSR58.6_EmR_1	Single reporter ErmC (EmR) with BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR EmR	This project
pSR58.6_TmpR_1	Single reporter DHFR^{L28R} (TmpR) with BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TmpR	This project
pSR58.6_KanR_1_LAA	Single reporter NptII (KanR) with LAA degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR KanR	This project
pSR58.6_AmpR_1_LAA	Single reporter TEM-115 (AmpR) with LAA degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR AmpR	This project
pSR58.6_ZeoR_1_LAA	Single reporter *Sh* Ble (ZeoR) with LAA degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR ZeoR	This project
pSR58.6_TcR_1_LAA	Single reporter TetC (TcR) with LAA degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TcR	This project
pSR58.6_EmR_1_LAA	Single reporter ErmC (EmR) with LAA degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR EmR	This project
pSR58.6_TmpR_1_LAA	Single reporter DHFR^{L28R} (TmpR) with LAA degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TmpR	This project
pSR58.6_KanR_1_LAA+4	Single reporter NptII (KanR) with LAA+4 degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR KanR	This project
pSR58.6_AmpR_1_LAA+4	Single reporter TEM-115 (AmpR) with LAA+4 degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR AmpR	This project
pSR58.6_ZeoR_1_LAA+4	Single reporter *Sh* Ble (ZeoR) with LAA+4 degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR ZeoR	This project
pSR58.6_TcR_1_LAA+4	Single reporter TetC (TcR) with LAA+4 degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TcR	This project
pSR58.6_EmR_1_LAA+4	Single reporter ErmC (EmR) with LAA+4 degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR EmR	This project
pSR58.6_TmpR_1_LAA+4	Single reporter DHFR^{L28R} (TmpR) with LAA+4 degradation tag and BBa_B0033 in pSR58.6_ ΔsfGFP_CS	ColE1 ChlR TmpR	This project
pDusk/pDawn plasmids			
pCORE_18_pDawn_mScarlet_CAT	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Light-repressible expression of cI from P_{FixK2} Bicistronic expression of mScarlet and chloramphenicol acetyltransferase from P_R	pBR322 KanR	This project

pCORE_38_pDawn_mScarlet_CAT	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Light-repressible expression of cI from P_{FixK2} Bicistronic expression of mScarlet and chloramphenicol acetyltransferase from P_R	pSC101 KanR	This project
pDawn	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Light-repressible expression of cI from P_{FixK2} Empty CDS downstream of P_R	pBR322 KanR	Addgene (43796)[60]
pDawn_CS	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Light-repressible expression of cI from P_{FixK2} GG cloning site downstream of P_R	pBR322 KanR	This project
pDawn_mScarlet-I	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Light-repressible expression of cI from P_{FixK2} mScarlet-I expression from P_R	pBR322 KanR	This project
pDawn_mScarlet-I_CS	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Light-repressible expression of cI from P_{FixK2} mScarlet-I expression from P_R with downstream GG cloning site	pBR322 KanR	This project
pDawn_mScarlet-I_RLuc8	BBa_J23105-B0032-RLuc8-L3S2P21 in pDawn_mScarlet-I_CS	pBR322 KanR	This project
pDawn_mScarlet-I_NanoLuc	BBa_J23105-B0032-NanoLuc-L3S2P21 in pDawn_mScarlet-I_CS	pBR322 KanR	This project
pDawn_mScarlet-I_CeNL	BBa_J23105-B0032-CeNL-L3S2P21 in pDawn_mScarlet-I_CS	pBR322 KanR	This project
pDusk	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Empty CDS downstream of P_{FixK2}	pBR322 KanR	Addgene (43795)[60]
pDusk_CS	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} GG cloning site downstream of P_{FixK2}	pBR322 KanR	This project
pDusk_mScarlet-I	Bicistronic expression of YF1-FixJ from P_{LacI}^{Q} Light-repressible expression of mScarlet-I from P_{FixK2}	pBR322 KanR	This project
pDusk_YF1_H22P_mScarlet-I	Bicistronic expression of $YF1^{H22P}$-FixJ from P_{LacI}^{Q} Light-inducible expression of mScarlet-I from P_{FixK2}	pBR322 KanR	This project
Other plasmids			
ilux pGEX(-)	Multicistronic expression of iLuxCDABE and Frp from P_{Tac} and constitutive expression of LacI from P_{LacI}	pBR322 AmpR	Addgene (107879)[61]
pCORE_36_iLux	Multicistronic expression of iLuxCDABE and Frp from P_{Tac}	pSC101 AmpR	This project
pCORE_36_iLux_H44A	Multicistronic expression of $iLuxCDA^{H44A}BE$ and Frp from P_{Tac}	pSC101 AmpR	This project

Table 4 | Main biological parts used in this project (excluding unpublished sequences)

Name	Description	Sequence	Source
Promoters			
BBa_K1614000	Minimal T7 promoter (P_{T7})	TAATACGACTCACTATAG	iGEM Registry

P_{araB*}	CAP-deficient arabinose-inducible promoter (P_{BAD})	TAGCATTTTTATCCATAAGATTAGCG GATCCTACCTGACGCTTTTTATCGCA ACTCTCTACTGTTTCTCCATA	CTSB Lab[49]
BBa_I14033	Constitutive promoter isolated from *cat* gene (P_{CAT})	GGCACGTAAGAGGTTCCAACTTTCAC CATAATGAAACA	iGEM Registry
BBa_J23119	Anderson parental sequence	TTGACAGCTAGCTCAGTCCTAGGTAT AATGCTAGC	iGEM Registry
BBa_J23104	Anderson mutant (72% activity)	TTGACAGCTAGCTCAGTCCTAGGTAT TGTGCTAGC	iGEM Registry
BBa_J23108	Anderson mutant (51% activity)	CTGACAGCTAGCTCAGTCCTAGGTAT AATGCTAGC	iGEM Registry
BBa_J23105	Anderson mutant (24% activity)	TTTACGGCTAGCTCAGTCCTAGGTAC TATGCTAGC	iGEM Registry
Ribosome Binding Sites (RBS)			
BBa_B0029	Community Collection RBS with 76% relative efficiency	TCTAGAGTTCACACAGGAAACCTACT AG	iGEM Registry
BBa_B0030	Community Collection RBS with 60% relative efficiency	TCTAGAGATTAAAGAGGAGAAATACT AG	iGEM Registry
BBa_B0031	Community Collection RBS with 7% relative efficiency	TCTAGAGTCACACAGGAAACCTACTA G	iGEM Registry
BBa_B0032	Community Collection RBS with 30% relative efficiency	TCTAGAGTCACACAGGAAAGTACTAG	iGEM Registry
BBa_B0033	Community Collection RBS with 1% relative efficiency	TCTAGAGTCACACAGGACTACTAG	iGEM Registry
BBa_B0034	Community Collection reference RBS (Elowitz)	TCTAGAGAAAGAGGAGAAATACTAG	iGEM Registry
Coding Sequences (CDS)			
sfGFP	Dimeric green fluorescent protein derived from *A. victoria* GFP with fast maturing dynamics	MRKGEELFTGVVPILVELDGDVNGHKF SVRGEGEGDATNGKLTLKFICTTGKLP VPWPTLVTTLTYGVQCFARYPDHMKQ HDFFKSAMPEGYVQERTISFKDDGTYK TRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNFNSHNVYITADKQKNGI KANFKIRHNVEDGSVQLADHYQQNTPI GDGPVLLPDNHYLSTQSVLSKDPNEKR DHMVLLEFVTAAGITHGMDELYK	Geoffrey Waldo Lab[62]
mScarlet	Synthetic, monomeric red fluorescent protein	MVSKGEAVIKEFMRFKVHMEGSMNGH EFEIEGEGEGRPYEGTQTAKLKVTKGG PLPFSWDILSPQFMYGSRAFTKHPADIP DYYKQSFPEGFKWERVMNFEDGGAVT VTQDTSLEDGTLIYKVKLRGTNFPPDG	Dorus Gadella Lab[63]

	with moderate acid sensitivity	PVMQKKTMGWEASTERLYPEDGVLKG DIKMALRLKDGGRYLADFKTTYKAKK PVQMPGAYNVDRKLDITSHNEDYTVV EQYERSEGRHSTGGMDELYK	
mScarlet-I	Fast-maturing mScarlet mutant (mScarlet^{T74I}) with shorter fluorescence lifetime	MVSKGEAVIKEFMRFKVHMEGSMNGH EFEIEGEGEGRPYEGTQTAKLKVTKGG PLPFSWDILSPQFMYGSRAFIKHPADIP DYYKQSFPEGFKWERVMNFEDGGAVT VTQDTSLEDGTLIYKVKLRGTNFPPDG PVMQKKTMGWEASTERLYPEDGVLKG DIKMALRLKDGGRYLADFKTTYKAKK PVQMPGAYNVDRKLDITSHNEDYTVV EQYERSEGRHSTGGMDELYK	Dorus Gadella Lab[63]
RLuc8	*R. reniformis* luciferase variant with enhanced stability and brightness	MASKVYDPEQRKRMITGPQWWARCK QMNVLDSFINYYDSEKHAENAVIFLHG NATSSYLWRHVVPHIEPVARCIIPDLIG MGKSGKSGNGSYRLLDHYKYLTAWFE LLNLPKKIIFVGHDWGAALAFHYAYEH QDRIKAIVHMESVVDVIESWDEWPDIE EDIALIKSEEGEKMVLENNFFVETVLPS KIMRKLEPEEFAAYLEPFKEKGEVRRPT LSWPREIPLVKGGKPDVVQIVRNYNAY LRASDDLPKLFIESDPGFFSNAIVEGAK KFPNTEFVKVKGLHFLQEDAPDEMGK YIKSFVERVLKNEQ	Sam Gambhir Lab[37]
RLuc8.6-535	Green-emitting *R. reniformis* luciferase variant	MASKVYDPEQRKRMITGPQWWARCK QMNVLDSFINYYDSEKHAENAVIFLHG NATSSYLWRHVVPHIEPVARCIIPDLIG MGKSGKSGNGSYRLLDHYKYLTAWFE LLNLPKKIIFVGHDWGSALAFHYAYEH QDRIKAIVHMESVVDVIESWMGWPDIE EELALIKSEEGEKMVLENNFFVETLLPS KIMRKLEPEEFAAYLEPFKEKGEVRRPT LSWPREIPLVKGGKPDVVQIVRNYNAY LRASDDLPKLFIESDPGFFSNAIVEGAK KFPNTEFVKVKGLHFLQEDAPDEMGK YIKSFVERVLKNEQ	Sam Gambhir Lab[40]
NanoLuc	Synthetic luciferase derived from the 19-kDa catalytic subunit of *O. gracilirostris* luciferase	MVFTLEDFVGDWRQTAGYNLDQVLEQ GGVSSLFQNLGVSVTPIQRIVLSGENGL KIDIHVIIPYEGLSGDQMGQIEKIFKVVY PVDDHHFKVILHYGTLVIDGVTPNMID YFGRPYEGIAVFDGKKITVTGTLWNGN KIIDERLINPDGSLLFRVTINGVTGWRL CERILA	Promega[38]
YNL	Bioluminescent BRET construct based on RLuc8^{S257G} and the yellow fluorescent protein Venus	MVSKGEELFTGVVPILVELDGDVNGHK FSVSGEGEGDATYGKLTLKLICTTGKLP VPWPTLVTTLGYGLQCFARYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYK TRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNYNSHNVYITADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTPI GDGPVLLPDNHYLSYQSKLSKDPNEKR DHMVLLEFVTAAGGTKVYDPEQRKRM ITGPQWWARCKQMNVLDSFINYYDSE KHAENAVIFLHGNATSSYLWRHVVPHI EPVARCIIPDLIGMGKSGKSGNGSYRLL DHYKYLTAWFELLNLPKKIIFVGHDWG AALAFHYAYEHQDRIKAIVHMESVVD VIESWDEWPDIEEDIALIKSEEGEKMVL	Nagai Lab[64]

		ENNFFVETVLPSKIMRKLEPEEFAAYLE PFKEKGEVRRPTLSWPREIPLVKGGKP DVVQIVRNYNAYLRASDDLPKLFIEGD PGFFSNAIVEGAKKFPNTEFVKVKGLH FLQEDAPDEMGKYIKSFVERVLKNEQ	
CeNL	Bioluminescent BRET construct based on NanoLuc and the cyan fluorescent protein mTurquoise2	MVSKGEELFTGVVPILVELDGDVNGHK FSVSGEGEGDATYGKLTLKFICTTGKLP VPWPTLVTTLSWGVQCFARYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYK TRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNYFSDNVYITADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTPI GDGPVLLPDNHYLSTQSKLSKDPNEKR DHMVLLEFVTAAGLHTLEDFVGDWRQ TAGYNLDQVLEQGGVSSLFQNLGVSV TPIQRIVLSGENGLKIDIHVIIPYEGLSGD QMGQIEKIFKVVYPVDDHHFKVILHYG TLVIDGVTPNMIDYFGRPYEGIAVFDG KKITVTGTLWNGNKIIDERLINPDGSLL FRVTINGVTGWRLCERILA	Nagai Lab[65]
GeNL	Bioluminescent BRET construct based on NanoLuc and the green fluorescent protein mNeonGreen	MVSKGEEDNMASLPATHELHIFGSING VDFDMVGQGTGNPNDGYEELNLKSTK GDLQFSPWILVPHIGYGFHQYLPYPDG MSPFQAAMVDGSGYQVHRTMQFEDG ASLTVNYRYTYEGSHIKGEAQVKGTGF PADGPVMTNSLTAADWCRSKKTYPND KTIISTFKWSYTTGNGKRYRSTARTTYT FAKPMAANYLKNQPMYVFRKTELKHS KTELNFKEWQKAFTGFEDFVGDWRQT AGYNLDQVLEQGGVSSLFQNLGVSVTP IQRIVLSGENGLKIDIHVIIPYEGLSGDQ MGQIEKIFKVVYPVDDHHFKVILHYGT LVIDGVTPNMIDYFGRPYEGIAVFDGK KITVTGTLWNGNKIIDERLINPDGSLLF RVTINGVTGWRLCERILA	Nagai Lab[65]
YeNL	Bioluminescent BRET construct based on NanoLuc and the yellow fluorescent protein Venus	MVSKGEELFTGVVPILVELDGDVNGHK FSVSGEGEGDATYGKLTLKLICTTGKLP VPWPTLVTTLGYGLQCFARYPDHMKQ HDFFKSAMPEGYVQERTIFFKDDGNYK TRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNYNSHNVYITADKQKNGI KANFKIRHNIEDGGVQLADHYQQNTPI GDGPVLLPDNHYLSYQSALSKDPNEKR DHMVLLEFVTAAMLEDFVGDWRQTA GYNLDQVLEQGGVSSLFQNLGVSVTPI QRIVLSGENGLKIDIHVIIPYEGLSGDQM GQIEKIFKVVYPVDDHHFKVILHYGTL VIDGVTPNMIDYFGRPYEGIAVFDGKKI TVTGTLWNGNKIIDERLINPDGSLLFRV TINGVTGWRLCERILA	Nagai Lab[65]
CAT	Chloramphenicol acetyltransferase gene (ChlR) isolated from the iGEM part BBa_J61000	MEKKITGYTTVDISQWHRKEHFEAFQS VAQCTYNQTVQLDITAFLKTVKKNKH KFYPAFIHILARLMNAHPEFRMAMKDG ELVIWDSVHPCYTVFHEQTETFSSLWS EYHDDFRQFLHIYSQDVACYGENLAYF PKGFIENMFFVSANPWVSFTSFDLNVA NMDNFFAPVFTMGKYYTQGDKVLMPL AIQVHHAVCDGFHVGRMLNELQQYCD EWQGGA	iGEM Registry

AadA	Aminoglycoside adenylyl transferase gene (SpecR) isolated from MoClo level 0 acceptor vector	MRSRNWSRTLTERSGGNGAVAVFMAC YDCFFGVQSMPRASKQQARYAVGRCL MLWSSNDVTQQGSRPKTKLNIMREAVI AEVSTQLSEVVGVIERHLEPTLLAVHLY GSAVDGGLKPHSDIDLLVTVTVRLDET TRRALINDLLETSASPGESEILRAVEVTI VVHDDIIPWRYPAKRELQFGEWQRNDI LAGIFEPATIDIDLAILLTKAREHSVALV GPAAEELFDPVPEQDLFEALNETLTLW NSPPDWAGDERNVVLTLSRIWYSAVT GKIAPKDVAADWAMERLPAQYQPVIL EARQAYLGQEEDRLASRADQLEEFVH YVKGEITKVVGK	Sylvestre Marillonnet Lab[54]
TEM-116	β-lactamase gene (AmpR) isolated from the iGEM part BBa_P1002	MSIQHFRVALIPFFAAFCLPVFAHPETL VKVKDAEDQLGARVGYIELDLNSGKIL ESFRPEERFPMMSTFKVLLCGAVLSRID AGQEQLGRRIHYSQNDLVEYSPVTEKH LTDGMTVRELCSAAITMSDNTAANLLL TTIGGPKELTAFLHNMGDHVTRLDRWE PELNEAIPNDERDTTMPVAMATTLRKL LTGELLTLASRQQLIDWMEADKVAGPL LRSALPAGWFIADKSGAGERGSRGIIAA LGPDGKPSRIVVIYTTGSQATMDERNR QIAEIGASLIKHW	iGEM Registry
NptII	Neomycin phosphotransfer gene (KanR) isolated from MoClo level 2 acceptor vector	MAKMRISPELKKLIEKYRCVKDTEGMS PAKVYKLVGENENLYLKMTDSRYKGT TYDVEREKDMMLWLEGKLPVPKVLHF ERHDGWSNLLMSEADGVLCSEEYEDE QSPEKIIELYAECIRLFHSIDISDCPYTNS LDSRLAELDYLLNNDLADVDCENWEE DTPFKDPRELYDFLKTEKPEEELVFSHG DLGDSNIFVKDGKVSGFIDLGRSGRAD KWYDIAFCVRSLREDIGEEQYVELFFD LLGIKPDWEKIKYYILLDELF	Sylvestre Marillonnet Lab[54]
Sh Ble	*S. hindustanus* Ble gene (ZeoR) isolated from the iGEM part BBa_K1033990	MAKLTSAVPVLTARDVAGAVEFWTDR LGFSRDFVEDDFAGVVRDDVTLFISAV QDQVVPDNTLAWVWVRGLDELYAEW SEVVSTNFRDASGPAMTEIGEQPWGRE FALRDPAGNCVHFVAEEQD	iGEM Registry
TetC	Tetracycline efflux pump *tetC* gene (TcR)	MKSNNALIVILGTVTLDAVGIGLVMPV LPGLLRDIVHSDSIASHYGVLLALYAL MQFLCAPVLGALSDRFGRRPVLLASLL GATIDYAIMATTPVLWILYAGRIVAGIT GATGAVAGAYIADITDGEDRARHFGL MSACFGVGMVAGPVAGGLLGAISLHA PFLAAAVLNGLNLLLGCFLMQESHKGE RRPMPLRAFNPVSSFRWARGMTIVAAL MTVFFIMQLVGQVPAALWVIFGEDRFR WSATMIGLSLAVFGILHALAQAFVTGP ATKRFGEKQAIIAGMAADALGYVLLAF ATRGWMAFPIMILLASGGIGMPALQA MLSRQVDDDHQGQLQGSLAALTSLTSI IGPLIVTAIYAASASTWNGLAWIVGAAL YLVCLPALRRGAWSRATST	NCBI (WP_00129 7013.1)
ErmC	*S. aureus* rRNA methylase gene (EmR) isolated from the iGEM part BBa_K2230001	MNEKNIKHSQNFITSKHNIDKIMTNIRL NEHDNIFEIGSGKGHFTLELVKRCNFVT AIEIDHKLCKTTENKLVDHDNFQVLNK DILQFKFPKNQSYKIYGNIPYNISTDIIR KIVFDSIANEIYLIVEYGFAKRLLNTKRS	iGEM Registry

		LALLLMAEVDISILSMVPREYFHPKPKV NSSLIRLSRKKSRISHKDKQKYNYFVM KWVNKEYKKIFTKNQFNNSLKHAGID DLNNISFEQFLSLFNSYKLFNK	
DHFR^{L28R}	Trimethoprim insensitive dihydrofolate reductase	MGQSSDEANAPVAGQFALPLSATFGLG DRVRKKSGAAWQGQVVGWYCTKLTP EGYAVESESHPGSVQIYPVAALERVA	UniProt (P00383)
Terminators			
BBa_B0010	Terminator 1 isolated from the *E. coli rrnB* gene	CCAGGCATCAAATAAAACGAAAGGC TCAGTCGAAAGACTGGGCCTTTCGTT TTATCTGTTGTTTGTCGGTGAACGCTC TC	iGEM Registry
BBa_B0015	Strong double terminator consisting of BBa_B0010 and BBa_B0012 (T7 terminator)	CCAGGCATCAAATAAAACGAAAGGC TCAGTCGAAAGACTGGGCCTTTCGTT TTATCTGTTGTTTGTCGGTGAACGCTC TCTACTAGAGTCACACTGGCTCACCT TCGGGTGGGCCTTTCTGCGTTTATA	iGEM Registry
L3S1P56	Strong synthetic terminator	TTTTCGAAAAAAGGCCTCCCAAATCG GGGGGCCTTTTTTATTGATAACAAAA	Voigt Lab [53]
L3S2P21	Strong synthetic terminator	CTCGGTACCAAATTCCAGAAAAGAGG CCTCCCGAAAGGGGGGCCTTTTTTCG TTTTGGTCC	Voigt Lab [53]
Other parts			
LAA	ssrA degradation tag	AANDENYALAA	Sauer Lab[66]
LAA+4	LAA tag with four additional internal residues for enhanced protein degradation	AANDENYSENYALAA	Sauer Lab[66]
5' HA* (SS9)	600 bp upstream of the SS9 genomic site of *E. coli* K12 with PAM mutation	GCTTGGTTGAGAATACGCCGAAGTTA AAATCAGCCCCTGGCTGGCAACGGCG TCGATATCTGGTGTAGGGTTACCCAC CGCCACGCCGCCACCGTTAAAACCGA CGTCCATCCAGCCCACATCGTCCAGC AAGAAAACAACCACATTCGGTTTCTT ACCGGTTTTTTTCTCAAGTTCTGCCAG CTTCTGCTGGGTTTCTTTATCCTGCGC CGGATGCTGCATTACTGGCATCATAT TGTCGGCAATAGTGGTCGCCGGTTTA ACCAGATACTGGTTTGGGTGATCGTA TCCGGCAAAGCCTTTGCGTGCGGTGG CAGTTGACGGGGTATCTGCTGCGCTG GCCATGAGAGGAAGAGCGGCGGCGA CAGCAACAACAAGACGTTTGGGTGA AAACGAAAATTCCATGCAAAATGCTC CGGTTTCATGTCATCAAAATGATGAC GTAATTAAGCATTGATAATTGAGATC CCTCTCCCTGACAGGATGATTACATA AATAATAGTGACAAAAATAAATTATT TATTTATCCAGAAAATGAATTGGAAA ATCAGGAGAGCGTTTTCAATCCTACC TCTGGCGCAGTTGATATGTCAAACAG GT	Gill Group[57]

3’ HA (SS9)	600 bp downstream of the SS9 genomic site of *E. coli* K12	TTATTATATCGCGTTGATTATTGATGC TGTTTTTAGTTTTAACGGCAATTAATA TATGTGTTATTAATTGAATGAATTTTA TCATTCATAATAAGTATGTGTAGGAT CAAGCTCAGGTTAAATATTCACTCAG GAAGTTATTACTCAGGAAGCAAAGAG GATTACAGAATTATCTCATAACAAGT GTTAAGGGATGTTATTTCCCGATTCTC TGTGGCATAATAAACGAGTAGATGCT CATTCCATCTCTTATGTTCGCCTTAGT GCCTCATAAACTCCGGAATGACGCAG AGCCGTTTACGGTGCTTATCGTCCAC TGACAGATGTCGCTTATGCCTCATCA GACACCATGGACACAACGTTGAGTGA AGCACCCACTTGTTGTCATACAGACC TGTTTTAACGCCTGCTCCGTAATAAG AGCAGGCGTTTTTTTATGTATCAGGA AGGCCCCGGAGGTGCTTGCCTCCGGG TGAGAAGGAACTACTGTGGCGGGTTA TTCTGCAACGTTAACATCAAACCGTC GCGACGCATAGCTGCAGCTTCTTCCG GCTTGTGCAGTCTGTCCAGCGCGTCG GCAAGCCATGCGTAATCGTAGGC	Gill Group[57]

2.2.2 DNA synthesis

Synthetic DNA fragments were purchased from Twist Bioscience to obtain most biological parts and circuits that were not available in the laboratory. All fragments were ordered to be compatible with Modular Cloning (MoClo) syntaxis and flanked by type II restriction enzyme sites for rapid shuttling into acceptor vectors via Golden Gate assembly[54]. Given synthesis length limitations, sequences longer than 1.8 kilobases were split into multiple fragments flanked by restriction enzyme sites with scarless overhangs to reconstitute the intended sequence upon assembly. Coding sequences were codon-optimised for *E. coli* expression using the Integrated DNA Technologies (IDT) Codon Optimization Tool (https://eu.idtdna.com/pages/tools/codon-optimization-tool) and cleared of BsaI, BpiI, BsmBI, and AarI recognition sites used in Golden Gate.

Oligonucleotide primers for amplifying, editing, or sequencing DNA samples were designed using OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html) to calculate melting temperatures (Tm) and predict DNA secondary structures and dimer primer formation. They were routinely ordered to Sigma-Aldrich as Custom DNA Oligos purified with the default desalting option and resuspended in Tris-EDTA buffer at 100 µM concentration. Primers containing long overhangs were separated using a polyacrylamide gel electrophoresis (PAGE) to exclude failed synthesis products, which become more abundant as the oligonucleotide size increases.

2.2.3 DNA amplification

Polymerase Chain Reaction (PCR) was used to amplify DNA for both genetic engineering procedures and the size validation of plasmid and genomic sequences.

PCR products needed for Golden Gate, Gibson, or KLD reactions were amplified using Q5® High-Fidelity Master Mix (NEB #M0492) to reduce the likelihood of introducing undesired changes to the DNA sequence. Reactions were prepared with 1-20 ng of template DNA in 50 µL of 1x Master Mix following the commercial supplier indications. Thermocycler conditions included a 30-second initial denaturation at 98ºC, 30 amplification cycles, and a final extension of 2 minutes at 72ºC. Annealing temperatures were calculated using the NEB Tm Calculator (https://tmcalculator.neb.com), and elongation steps lasted 30 seconds per kb, or 1 minute per kb for amplicons longer than 4 kb.

Colony PCR was performed without distinction to validate plasmid or genomic inserts. Individual colonies were picked with a small pipette tip and manually resuspended in 10 µL aliquots of Encyclo polymerase (Evrogen #PK002) Master Mix containing each primer at 5 µM final concentration. Then, pipette tips were used to inoculate 2 mL LB cultures with the appropriate antibiotic to isolate desired genotypes. Thermocycler conditions were set following the commercial supplier protocol using a 5-minute initial denaturation at 95ºC to disrupt the cells, 30 amplification cycles, and a final extension of 2 minutes at 72ºC. Annealing temperatures were calculated using Benchling default Tm estimations; and elongation steps were always set at 1 minute per kb, regardless of the amplicon length.

2.2.4 DNA separation

PCR products were generally digested with 1 µL of DpnI (NEB #R0176) in the same PCR buffer to remove the template DNA plasmid. After a 1-hour incubation at 37ºC and 20-minute enzyme inactivation at 80ºC, products were mixed with 6x Gel Loading Dye (NEB #B7025S) and separated using an agarose gel electrophoresis.

Electrophoresis gels were routinely prepared at 1% agarose in TAE buffer made by the Institute Media Kitchen Facility using TopVision Agarose Tablets (ThermoFisher #R2801) and 1 µg/mL of ethidium bromide (Biotium #40042). Tablets were allowed to dissolve first before melting in the microwave and ethidium bromide was added and mixed right before pouring into the cast. Gels were loaded with either 60 µL products for later purification or 12

μL colony PCR samples, alongside a DNA ladder with the appropriate range of molecular weights: 100 bp DNA Ladder (NEB #N3231) for products between 100-1,500 bp, or Quick-Load® Purple 1 kb Plus DNA Ladder (NEB #N0550) for products between 1-10 kb.

All gels were run at 150 Watts in 1x TAE with 1 μg/mL of ethidium bromide using a horizontal gel electrophoresis system (Cleaver Scientific #MSMAXI and # NANOPAC-300P) and revealed using a GelDoc Go Imaging System (Bio-Rad #12009077). Those products needed for further DNA manipulation were extracted from the agarose gel by excising the expected size band with a scalpel using repetitive imaging to aid the procedure.

2.2.5 DNA purification

All DNA purifications were performed using sterile, double-distilled water (ddH_2O) in the final elution step, as elution buffer reduced the efficiency of other laboratory procedures, such as electroporation and sequencing. It was pre-warmed at 70ºC to increase DNA solubility, pipetted directly on top of the column membrane and incubated for 1 minute before centrifugation.

PCR products that required additional digestions, apart from DpnI, before separation on agarose gels or previously validated amplifications with single, correct products were purified using QIAquick PCR Purification Kit (Qiagen #28104). DNA from excised agarose bands was extracted using Zymoclean Gel DNA Recovery Kit (Zymo Research #D4002).

For plasmid DNA extractions, 5 mL LB cultures with the appropriate antibiotics were inoculated in 15 mL Falcon™ conic tubes (ThermoFisher #352097) from either glycerol stocks or fresh colonies and grown overnight. Saturated cultures were spun down at 4,500 g for 10 minutes using a bench top centrifuge (Eppendorf #5920R) and bacterial pellets were frozen at -20ºC after discarding the supernatant to facilitate the disruption of cell membranes. These pellets were ready for resuspension and plasmid extraction on the same day or any other day using QIAprep Spin Miniprep Kit (Qiagen #27104).

PCR purifications were eluted with 25 μL of ddH_2O, gel extractions with 10-15 μL depending on reaction efficiency, and plasmid extractions with 40 μL. Finally, DNA products were vortexed and stored at -20ºC unless required the same day.

2.2.6 DNA assembly and editing

Genetic circuits constructed with synthetic or amplified DNA fragments were assembled using two Golden Gate-based cloning standards: MoClo[54] and pCORE (unpublished). All Golden Gate assemblies were performed using MoClo protocols and NEB enzymes (**Table 5**). Reactions were prepared in 10 μL of 1x T4 DNA Ligase Buffer (NEB #B0202S) containing equimolar concentrations of each insert and the vector backbone, and 1 μL of each enzyme including the T4 DNA ligase. After mixing the reaction components by pipetting up and down, Golden Gate reactions were briefly spun down and incubated in the thermocycler with the following programme: 45 assembly cycles (2 min at 37ºC – 5 min at 16ºC) and 1 inactivation stage (5 min at 50ºC – 10 min at 80ºC). Then, a final 15 minute incubation was performed to remove empty backbones by introducing 1 μL of the enzyme used to open the acceptor vector, followed by 20 minutes of inactivation. These reaction products were either immediately transformed or stored in the fridge/freezer for as long as necessary.

Table 5 | Cloning enzymes used for Golden Gate assemblies

Name	Recognition sequence	Optimal (inactivation) temperatures	Source
BsaI-HF® v2	5' GGTCTC $(N)_1$ 3' 3' CCAGAG $(N)_5$ 5'	37ºC (80ºC)	NEB #R3733
BbsI-HF®	5' GAAGAC $(N)_2$ 3' 3' CTTCTG $(N)_6$ 5'	37ºC (65ºC)	NEB #R3539
BsmBI-v2	5' CGTCTC $(N)_1$ 3' 3' GCAGAG $(N)_5$ 5'	55ºC (80ºC)	NEB #R0739
PaqCI® (AarI)	5' CACCTGC $(N)_4$ 3' 3' GTGGACG $(N)_8$ 5'	37ºC (65ºC)	NEB #R0745
T4 DNA Ligase	Double-stranded DNA Nicked DNA/RNA	16ºC (65º)	NEB #M0202

DNA editing was performed by amplifying the plasmid template with primers containing the desired mutations or designed to generate a truncated version of the plasmid. Then, PCR products were gel purified and circularised via Gibson assembly (NEB #E2611), or directly assembled with the KLD Enzyme Mix (NEB #M0554), which multiplexes blunt-end ligation with template plasmid digestion. In both cases, reactions were directly transformed or stored at -20ºC until needed to avoid further enzymatic activity.

CRISPR/Cas9-assisted genome integration was performed following a well-established protocol in a previously validated *E. coli* K12 genomic site using the materials deposited by the authors on Addgene[57,67]. These involve three plasmids containing an arabinose-inducible

Cas9 (pX2-Cas9), a temperature-inducible homologous recombination system (pSIM5), and a constitutive gRNA targeting the genomic sequence (pSS9_RNA). The first two plasmids are transformed into the strain to prepare electrocompetent cells with and without inducing the expression of the recombineering genes, and then electroporated with the gRNA and the dsDNA fragment for insertion. A 2-hour recovery incubation in rich medium supplemented with 0.2% L-arabinose allows sufficient time for both Cas9-gRNA to cause a double strand break in the genome and the recombineering genes to repair it using the provided insert. The insert is flanked by 600 bp homology arms with a PAM mutation to avoid further Cas9 restriction. Although only successfully engineered bacteria should survive the process, a ZeoR cassette was also included in the insert to allow for double-selection. Finally, genome insertion was validated by colony PCR and plasmids were curated with an overnight growth at 42ºC in LB containing only zeocin.

2.2.7 DNA sequencing

Modified plasmids and the insert for genome integration were initially validated in an agarose gel electrophoresis and confirmed by Sanger sequencing using GeneWiz services, now Azenta Life Sciences. DNA samples were sent pre-mixed with each primer following the company guidelines. Plasmids used in final experiments were fully-sequenced using Next-Generation Sequencing (NGS) services provided by Full Circle Labs. Sequencing results were aligned in Benchling (https://www.benchling.com) with the expected DNA sequences using the Multiple Alignment using Fast Fourier Transform (MAFFT) algorithm.

2.3 Assays

2.3.1 Plate reader assays

A Spark® multimode microplate reader (TECAN) equipped with monochromators, temperature and humidity control, and an automated lid-lifter was used throughout this project to measure bacterial growth, monitor fluorescence and bioluminescence levels over time, and characterise the emission spectra of bioluminescent proteins.

Bacterial growth assays were routinely started by inoculating LB cultures with antibiotics at an OD = 0.005 and splitting them after vortexing into 0.2 mL aliquots in 96-well, flat, transparent microplates (ThermoFisher #167008). Cultures were grown inside the plate reader at 37ºC and 220 rpm using a continuous orbital shaking mode and a humidity cassette

with sterile water to protect from desiccation. OD measurements were automatically collected at 600 nm every 5 or 15 minutes, depending on the experiment, for 24 hours.

In *Chapter 3*, OD curves were used to calculate the growth rate of multiple strains by normalising the increment in culture density at each time-point by its previous OD value and expressing the difference as a percentage. Moreover, growth rate values were plotted as a function of OD to visualise the speed of cell division across the OD curve, which allowed differentiating the distinct bacterial growth phases and designing experiments accordingly.

Bioluminescence pulses were characterised following a similar procedure using 96-well, flat bottom (chimney-well) μClear®, black microplates (ThermoFisher, # 655096) to prevent signal crosstalk between neighbouring wells. Cultures were grown until optimal induction conditions (OD = 0.1-0.2 for CcaS-CcaR; OD = 0.4 for pDawn) and treated with their respective luciferin or vehicle solutions (**Table 6**). After treatment, OD and luminescence measurements were sequentially collected at 5-minute intervals for 16 hours, using 1 second integration time without any attenuation filters to register photon emissions.

In *Chapter 4*, mScarlet-I production from pDawn in response to bioluminescence was monitored over time following an identical protocol that included an additional fluorescence measurement. mScarlet-I fluorescence was excited using a 560/20 nm monochromator and collected with top readings using a 620/20 nm monochromator, 40 μs integration time, and 20,000 μm for the z-position of the reading probe. Laser gain (81 au) was automatically calibrated using a sample containing saturating mScarlet-I expression levels as a reference for maximum signal intensity.

In *Chapter 4*, the bioluminescence emission spectra of Green enhanced Nano-Lantern and two rationally engineered mutants were determined by performing a luminescence scan from 398 to 653 nm using 15-nm reading steps and 1 second of signal integration time.

Table 6 | Luciferin substrates and vehicle solutions used in this project

Luciferin	Vehicle solution	Source
Nano-Glo® Live Cell Assay System	Heat-inactivated Nano-Glo®	Promega #N2011
ViviRen™ Live Cell Substrate	DMSO	Promega #E6491
Coelenterazine-h	Methanol	Promega #S2011
3-hydroxyhispidin	DMSO	Planta (donation)

2.3.2 Light-induction assays

External light-induction assays were performed using 0.5 mL cultures in 24-well, glass-bottom, black plates (Southern Labware, #324041) grown inside a manually-assembled Light Plate Apparatus (LPA) containing either 470 nm or 525 nm light-emitting diodes (LED) depending on the light-sensing circuit[68]. LEDs were calibrated following the article instructions and programmed using a SD card and Iris software files (http://taborlab.rice.edu/software). Except for the light-sensitivity comparison between CcaS and miniCcaS#10 photoreceptor variants in *Chapter 3*, all other LPA experiments were programmed with saturating light intensities, according to the characterisation data published for each photoreceptor. Cultures in the LPA were always grown at 37ºC and 240 rpm in a dark incubator.

In *Chapter 3*, the induction protocols published for paT7Ps[52], Opto-T7RNAP*(563)[49], CcaS-CcaR[58], and pDusk/pDawn[60] were strictly followed to implement these light-sensing circuits in the laboratory. The only protocol that could not be entirely reproduced was the Opto-T7RNAP*(563), because it required constant monitoring of culture density and manual culture dilution with fresh medium to maintain cells in exponential phase. Consequently, shorter experiments were performed to limit gene expression analyses at culture time-points within exponential phase.

Bioluminescence-induction assays with CcaS-CcaR and pDawn were always started by inoculating LB cultures with antibiotics at an OD = 0.005. Cultures were grown either in the LPA or the plate reader as specified above until optimal induction conditions (OD = 0.1-0.2 for CcaS-CcaR; OD = 0.4 for pDawn) and treated with their respective luciferin or vehicle solutions (**Table 6**). Samples were collected at different time intervals depending on the experiment and prepared for flow cytometry analysis when appropriate, as explained below.

2.3.3 Bacterial viability assay

Bacterial viability of strains carrying an inducible KanR gene under distinct kanamycin concentrations was assessed by adapting a previously published method [69]. Experiments were started by inoculating LB cultures with antibiotics at OD= 0.005 and growing them in the dark at 37ºC and 240 rpm until CcaS-CcaR optimal activation conditions (OD = 0.1-0.2). At this point, cultures were induced with either external green-light/darkness or luciferin/vehicle depending on the experiment, and challenged with different doses of kanamycin at various time-points after induction. Aliquots were collected at 2- or 4-hour intervals for up to 12 hours

depending on the experiment, and plated as droplets on LB agar after generating a dilution series across several orders of magnitude with Phosphate-Buffered Saline (PBS). Finally, plates were incubated in the dark at 37ºC for 24 hours and colonies counted next day under the microscope for each 10-fold dilution to estimate total cell viability numbers in each condition.

2.4 Flow cytometry

2.4.1 Paraformaldehyde fixation

Samples for flow cytometry analysis were collected during light-induction assays into 1.5 mL microcentrifuge tubes (ThermoFisher #3448) and protected from the light until the end of the fixation protocol. Tubes were initially placed on ice for 30 minutes to stop bacterial growth and allow fluorescent protein maturation. Next, samples were spun down at 1,400 g for 5 minutes with a MiniSpin® Plus microcentrifuge (Eppendorf #5453000060) to discard the culture medium and resuspended in 0.1 mL of ice-cold 4% paraformaldehyde (PFA) (Boster Bio #AR1068). After 10 minutes of fixation, cells were pelleted again to remove the fixative and resuspended in 1 mL PBS. Fixed samples can be stored in the fridge for months until flow cytometry analysis and are not sensitive to light anymore.

2.4.1 Flow cytometry analysis

Fixed samples were run on a BD® LSR II Flow Cytometer (BD Biosciences) using low flow rates and 10,000 events were acquired per sample. Prior to analysis, appropriate dilutions were made in PBS to obtain count rates between 500-1,000 events per second. Acquisition was performed using blue (488 nm, 20 mW, gain = 405 au) and yellow (561 nm, 50 mW, gain = 456 au) solid-state lasers and 525/50 nm and 610/20 nm band pass filters for the characterisation of sfGFP and mScarlet fluorescence, respectively. An SSC threshold was used to eliminate instrument noise events that did not correspond to cell scattering and the geometric mean of all registered events was used for data analysis. Finally, histograms were plotted using MATLAB® (MathWorks) with a script for FCS files written by Dr Jure Tica.

2.4.1 Fluorescence-activated cell sorting (FACS)

Strains used as sorting controls constitutively expressed sfGFP or RLuc8.6-535. They were inoculated in the morning by diluting overnight cultures 100-times in fresh LB with antibiotics and grown until OD ~ 0.4. Individual cultures were sequentially analysed to define

fluorescent parameters discriminating non-fluorescent bacteria and then mixed at equal proportions to validate sorting efficiency.

Strains for bioluminescent sorting expressed the luciferin-inducible CcaS-CcaR circuit with either functional or non-functional GeNL. They were inoculated the night before in 20 mL Chi.Bio turbidostats[70] (www.chi.bio) and set at an OD = 0.1 using previously calibrated values. In the morning, cultures were transferred into the LPA, individually and mixed at equal proportions, and induced for 3 hours with 1x Nano-Glo® in the dark or under saturating green light. Then, samples were collected into 1.5 mL Fisherbrand™ Black Microcentrifuge Tubes (ThermoFisher #15386548) and kept on ice while waiting to be sorted.

Bacterial samples were sorted using a BD FACSAria™ III Cell Sorter (BD Biosciences). Double-sorting was performed with low flow rates using the 4-Way Purity Precision Mode and count rates of approximately 1,000 events per second. For each sample, 100,000 events were initially sorted in PBS to enable the collection of around 7,000-10,000 events in the second round. Sorted bacteria were plated on LB agar plates with antibiotics immediately after sorting alongside their respective non-sorted controls, and grown at 37ºC overnight. Next day, fluorescent colonies were counted under a proBLUEVIEW Dual Colour transilluminator (Cleaver Scientific #PROBLUEVIEW). Instead, GeNL-expressing colonies had to be inoculated into 96-well black plates and grown for several hours before assessing their photon production inside the plate reader to distinguish bioluminescent from non-bioluminescent colonies.

3. Characterising light-controlled transcriptional circuits during bacterial growth

Three promising light-controlled transcriptional circuits from the literature were assessed in the laboratory for their project suitability. After validation, their signalling dynamics were investigated during bacterial growth to identify optimal transduction conditions for engineering bioluminescence activations. Two of them, based on bacterial two-component systems, were chosen for the next project stage: a green-light circuit named CcaS-CcaR, and two variations of a blue-light circuit called pDusk/pDawn. A set of photoactivatable, split T7 RNA polymerases could never be implemented in the laboratory despite best efforts.

3.1 Introduction

3.1.1 Selection criteria for suitable light-sensing components

As explained in *Chapter 1*, this project aims to develop a biological platform that can select improved luciferase variants from bioluminescent mutant libraries. The concept is to program bacteria with light-sensing genetic circuits that convert intracellular bioluminescent signals into an evolutionarily advantageous trait, such as a resistance to an antibiotic or an essential metabolic capacity. As a result, bacteria expressing the brightest bioluminescent phenotypes should become fitter and outcompete the rest of the population under the appropriate environmental pressure. Therefore, the properties of the light-sensing circuit used to generate fitness differences will determine how efficiently bioluminescent bacteria can be selected. These include, but are not limited to, its ability to react to the signal of interest (sensitivity) and remain inactive in the absence of it (specificity), the ratio between its upper and lower detection limits (dynamic range), the temporal dynamics of its activity (kinetics), and its regularity during bacterial growth (stability).

Two main categories of light-sensing genetic circuits are currently available to engineer bacterial behaviours with light, which target either transcription or protein activity. In both cases, control is achieved through light-sensing proteins that regulate either the expression of a gene of interest or the activity of a protein that is already available in the cell.[71] Light-controlled *transcriptional* circuits appeared more suitable for this project, because they only require changing the gene expressed from the output promoter to be repurposed. This facilitates the screening of different fitness genes that may be needed to identify suitable selection strategies. In contrast, light-controlled *protein activity* methods depend on engineering together different protein components to construct every new function, and result in different protein dynamics each time. In transcriptional circuits, every step in the gene expression process can be independently adjusted to tune the lifetime of their effects without modifying the circuit components. For instance, the production and degradation rates of the protein of interest can be respectively modified by changing the affinity of the RBS in the messenger RNA, and by adding protein degradation tags to the coding sequence[66,72]. This was an additional motivation to focus on these circuits, as it should provide higher adaptability to suit projects with different requirements.

The main criteria used to shortlist light-controlled transcriptional circuits were access to the technology, its readiness to be deployed, availability of characterisation data, and any track record of peer-reviewed applications.

3.1.2 Methodological considerations: bacterial growth and bioluminescence

In recent years, a new paradigm to engineer molecular interactions has emerged that utilises light emitted by bioluminescent proteins to regulate cell processes through light-sensitive proteins[20]. Simultaneously, other groups have shown that cell growth can be controlled with light-sensitive circuits and artificial light sources, such as LEDs[73–75]. However, the feasibility of influencing cell growth using bioluminescence remains undemonstrated. To approach this goal, it is important to consider which methodological constrains each of these processes entail to effectively couple them together.

On one side, enzymatic production of light in the laboratory is harder to control and substantially more expensive than electric light. Most bioluminescence experiments are externally induced by applying a luciferin solution to the growth medium. The main reasons being that most luciferin biosynthesis genes are still unknown, and synthetic luciferin formulations are often used to achieve better control of their light emission properties[23,24]. As a result, bioluminescent reactions are difficult to control, because they depend on the capacity of cells to both transport luciferin molecules into the cytoplasm and metabolise them, and cannot be externally inactivated. Furthermore, luciferin solutions are either chemically synthesised or extracted from natural sources, which accounts for their high commercial prices. Therefore, bioluminescence experiments need to occur in small working volumes, and experimental conditions carefully tested beforehand to achieve predictable light pulses.

On the other side, bacterial growth can affect synthetic circuit performance through two main variables: host-circuit interactions and the dilutive effect of cell division. The first one originates from the fact that genetic circuits operating within living cells depend on using their resources to function and are subjected to the internal adaptations that characterise life. For instance, it is known that bacterial growth causes global effects on gene expression due to changes in the concentration of metabolites, and transcription and translation machinery[76,77]. Specific remodelling of the host genetic programme also occurs in response to environmental stressors, such as the presence of antibiotics or starvation, which might interfere with the circuit regulation[78]. The second problem occurs when the dilutive effects of cell division are not kept constant, because it alters the steady-state concentration of circuit components, causing them

to work suboptimally[79]. Both the impact of cell division and resource availability on synthetic circuits can be mitigated by establishing continuous culture conditions in exponential growth phase. These methods accurately replace old medium for fresh medium in the culture to maintain a constant division rate and an excess of nutrients available. Turbidostats and microfluidic devices are some solutions that have been implemented to automate the process, and have both been used to reliably control bacterial growth while expressing synthetic circuits[74,80]. Most likely, these options are not going to be compatible with this project due to the high working volumes of turbidostats, and our lack of reach and expertise in microfluidic devices.

For these reasons, the characterisation of light-sensitive genetic circuits during bacterial growth could be of key importance in identifying an optimal response window to reliably induce fitness acquisition and aid in the development of a selection protocol. In line with this hypothesis, this project has focused on accurately controlling and modelling bacterial growth curves, and mapping circuit dynamics across them. In this chapter, the most promising light-controlled transcriptional circuits that could be identified within the early stages of the project (2019-2020) are presented alongside the results of their characterisation during bacterial growth.

3.2 Photoactivatable split T7 RNA polymerases

3.2.1 Background

Two collections of blue light-inducible transcriptional circuits had attracted considerable attention in the bacterial optogenetics community when this project started and seemed the most suitable option at the time[51,52]. They are all variations of a split T7 bacteriophage RNA polymerase (T7RNAP) fused to distinct versions of the small, dimerizing Vivid (VVD) photoreceptor. In these circuits, blue light enables the functional reconstitution of two T7RNAP fragments, which drives gene expression from its cognate promoter. The transcriptional autonomy conferred by T7RNAP activity and the absence of T7RNAP-specific promoter sequences in the bacterial genome make these circuits especially valuable for this project[81]. As explained in *Section 3.1.2*, decoupling transcription from the host should increase circuit stability because the availability of endogenous transcription machinery varies during growth. Nevertheless, T7RNAP expression systems are known to exhibit expression-induced toxicity due to the high transcription rate of the polymerase and often require operating under fine-tuned conditions to mitigate this problem[82,83].

The VVD sensor used to engineer them has been extensively characterised[84–87], optimised[88,89], and used in several other optogenetic applications[90–96]. It is a 21 kDa light, oxygen, or voltage (LOV) protein that uses a FAD molecule as a chromophore and dimerizes upon blue-light activation[84,87]. In the dark state, the FAD moiety is non-covalently sequestered on the protein surface, while the N-terminal dimerization region is packed against the protein core[85]. Blue-light excitation of the flavin ring catalyses the formation of a thioether bond between the chromophore and a cysteine residue that induces a conformational change in the protein. In turn, this releases the N terminus, which becomes accessible for dimerization until the bond thermally decays and the sensor reverts to its dark state[86].

The photoactivatable dimerization of VVD sensors was intended in both studies to reconstitute T7RNAP transcriptional activity by bringing the two polymerase halves in contact. Despite following similar approaches, each group took distinct considerations to maximise the utility of the sensors, resulting in a diverse group of tools. Han *et al* discovered that VVD dimerization was not required for T7RNAP reconstitution in their initial design, because it happens autonomously at the selected split position (**Figure 5A**). In fact, VVD domains enable light-regulation by blocking T7RNAP reconstitution in the dark state, either through steric hindrance or allosteric effects, and allow it after becoming displaced due to the conformational

change induced by blue light (**Figure 5B**). This persuaded the authors to further simplify their initial construct, called photoactivatable T7 RNA polymerase-1 (paT7P-1), using only one VVD domain. The resulting variants are: paT7P-2, in which only the N-terminal half of T7RNAP is fused to a VVD sensor and the C-terminal half is unmodified (**Figure 5C**); and paT7P-3, a single-chain construct that contains the VVD monomer terminally fused in between the two T7RNAP fragments (**Figure 5D**). These simpler constructs showed slower reverting kinetics because they depend on the thermal decay of the thioether bond in their single VVD monomer, whereas paT7P-1 has twice more chances of reverting to the dark state. Nonetheless, slower dark-state reversal is advantageous for this project because bioluminescence pulses are usually short lived, and luciferin solutions are expensive and sometimes toxic. Therefore, circuit activation might be difficult to sustain until bacterial selection is accomplished, which these circuits would prolongate for several hours after bioluminescent signal decay.

The structural investigation carried by Baumschlager *et al* in the selection of the T7RNAP split position led to an almost identical conclusion: Ser^{563}/Glu^{564} instead of Glu^{564}/Thr^{565}. It is quite likely, thus, that the regulatory mechanism governing the previous sensors also applies to this collection since their split positions are just one residue apart. In contrast, these authors kept the initial design, Opto-T7RNAP*(563), and focused on improving other preferred properties for synthetic dynamic regulation (**Figure 5B**). First, they used a less burdensome T7RNAP variant containing a point-mutation (R632S) at a protein region that is thought to influence polymerase processivity[97]. Second, they fused the T7RNAP halves to fast-

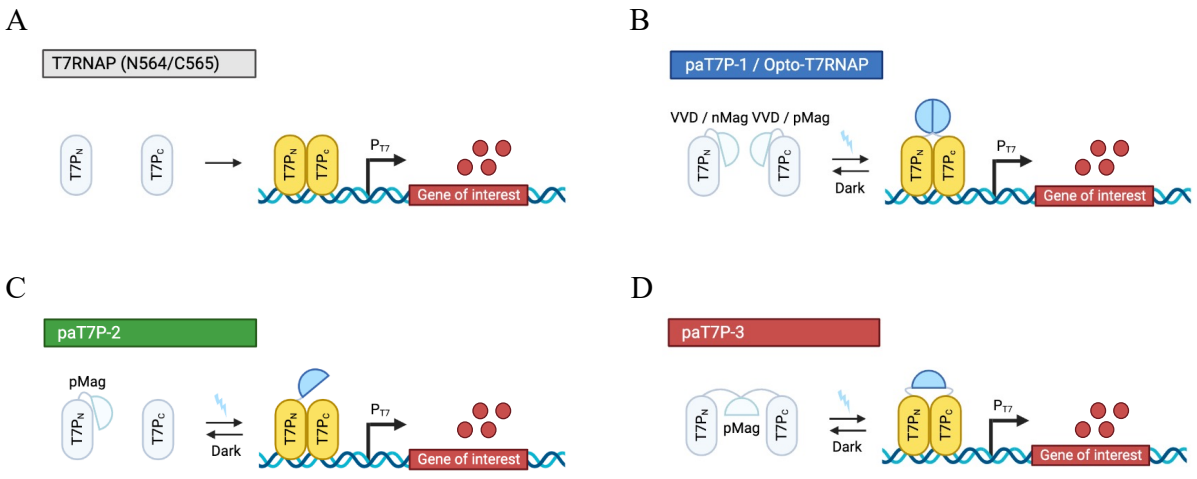


Figure 5 | Blue-light photoactivatable split T7RNAP-based circuits.
Four design variations of a light-controlled transcriptional circuit engineered by coupling the photosensitive conformational switch of VVD domains to the functional reconstitution of a split T7RNAP. **A** The T7RNAP component in all sensors was fragmented at a split position that can autonomously reconstitute inside the cell. **B** The original sensor design considered the blue light-induced dimerization of VVD domains as a mechanism for physically bringing together both T7RNAP halves and catalyse the recovery of its activity. **C, D** VVD impedes reconstitution in the dark-state conformation via steric regulation, allowing simpler designs with different ON/OFF dynamics.

reverting engineered VVD variants called Magnets[88]. Finally, they expressed each sensor half from an arabinose-inducible promoter to allow response tunability without affecting the dynamic range. To achieve precise titratable arabinose induction, these authors used an *E. coli* strain deficient in both arabinose uptake and catabolism genes, containing a mutant LacY transporter ($LacY^{A177C}$) with extended sugar specificity that allows arabinose transport in a concentration-dependent manner[98]. The availability of this strain is the main limitation for the wide-spread utilisation of Baumschlager *et al.* sensors, as it is unclear how dependent their performance might be on the expression conditions.

Despite their differences, all sensors mentioned above offered orthogonality, simple genetic structures, high dynamic ranges (62- to 300-fold induction), and high tolerance to fusion proteins; in case it was needed to fuse the luciferases to achieve bioluminescence-mediated activation. For these reasons, all three sensors from Han *et al* and the Opto-T7RNAP*(563) variant with the lowest basal activation and highest dynamic range from Baumschlager *et al* were selected for validation in the laboratory.

3.2.2 Results and discussion

The four selected circuits were assembled in the laboratory from synthetic DNA fragments based on the published sequences, as the authors could not be contacted to obtain the original materials. Although both research groups had deposited all plasmid sequences in public databases, Opto-T7RNAP*(563) entries were unavailable at the time, and that circuit was reconstructed using the individual genetic component sequences reported in the article. Adding to the difficulty of reutilising these tools, each group had used different two-plasmid co-expression systems, fluorescence reporter genes, host strains, growth conditions, and light-responsiveness assays[49,52]. Therefore, some compromises had to be made to simplify the experimental requirements to deploy these technologies in the laboratory.

3.2.2.1 Impaired dynamic range is unrelated to the circuit copy-number

In an attempt to identify the most suitable co-expression system for these circuits, a vector backbone collection previously created in the laboratory was used that contained combinations of distinct selectable markers and origins of replication (**Figure 6A**). It included three origins of replication with different copy-numbers that are independently regulated to avoid incompatibilities when transformed together: pSC101 (~ 5 copies per cell), p15A (~ 10 copies per cell), and pBR322 (~ 15-20 copies per cell). Furthermore, these vectors allow rapid

backbone exchange via a Golden Gate-type reaction to readjust the copy-number and the selectable marker of a given circuit[99].

To validate the copy-number effect on gene expression in these vectors, a constitutive transcription unit expressing the red fluorescent protein mScarlet was cloned into them. Cultures transformed with these vectors were started from identical seeding concentrations and grown for several hours before being analysed by flow cytometry. In all vectors, except for the spectinomycin-selectable subset, fluorescence signal levels were uniquely determined by the origin of replication sequence and without significant differences across antibiotic resistance genes (**Figure 6B**). Unexpectedly, p15A vectors showed lower mScarlet expression levels than pSC101, despite having a copy-number that it is normally twice higher. Still, the collection provided three distinct levels of vector replicability inside the bacterial host cell.

To determine the copy-number for each circuit component in these blue light-inducible circuits, the four sensors and a superfolder GFP (sfGFP) reporter gene under T7RNAP promoter control were systematically cloned into the ampicillin- and zeocin-selectable vector subsets. These vectors showed the most titratable effects on gene expression while retaining low variability between subsets (**Figure 6B**). Light-responsiveness assays were conducted following the Han *et al* protocol because there were no practical means of reproducing the

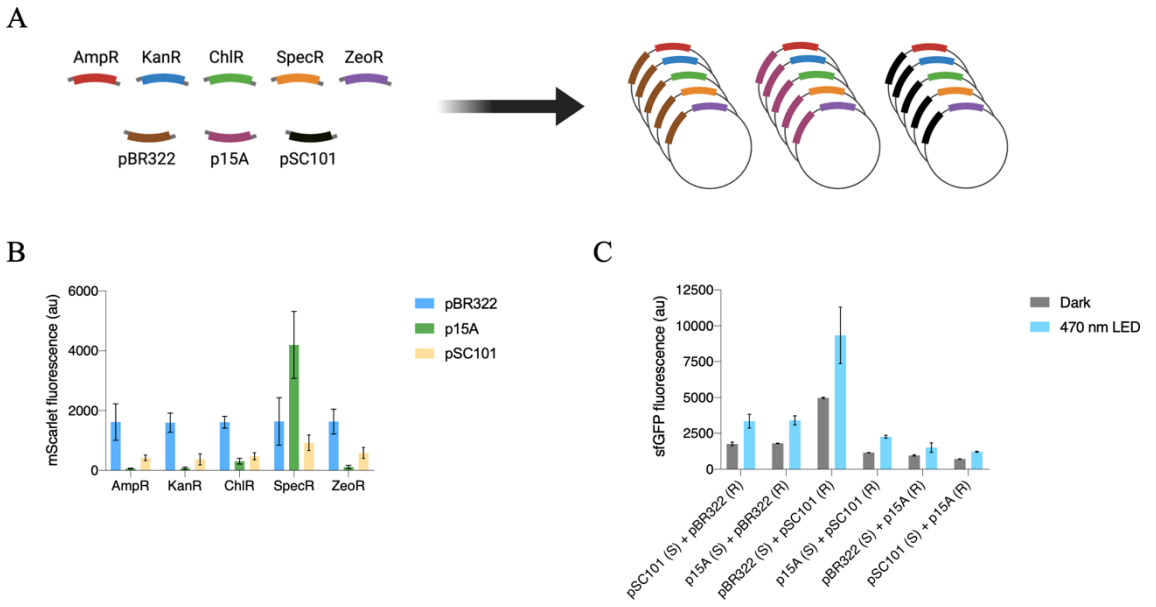


Figure 6 | Impaired light-activation of T7RNAP-based circuits is unrelated to the copy-number of their components. **A** Combinatorial collection of vector backbones based on five antibiotic selectable markers and three origin of replication sequences that allows for fast circuit relocation to adjust these variables. **B** Fluorescence characterisation of the effects of vector copy-number on constitutive gene expression in these vectors. Bars represent the average ± SD of three biological replicates analysed by flow cytometry using mean population values. **C** Fluorescence levels produced in the dark- and lit-state when co-expressing paT7P-2 sensor and the sfGFP reporter gene from distinct two-plasmid systems with different copy-number, after a 20-hour induction protocol using saturating blue light or darkness. Bars represent the average ± SD of two biological replicates analysed by flow cytometry using mean population values.

protocol from Baumschlager *et al*. All circuits elicited compromised responses after 20 hours of blue-light illumination in all co-expression systems (**Figure 6C,** data only shown for paT7P2 sensor). On one side, sensors appeared to behave similarly regardless of their copy-number, except in one strain that could not be explained (sensor in pBR322; reporter in pSC101). For instance, no significant differences were observed when expressing the reporter gene from pBR322 (highest copy-number) and comparing sensor performance from pSC101 and p15A vectors. The same was true when the reporter gene was in p15A (lowest copy-number) and sensors were expressed from either pBR322 or pSC101. On the other side, the copy-number of the reporter gene appeared to modulate subtly the transcriptional capacity of the system, as it is expected from varying the number of promoters present in the cell. In any case, it appeared clear that the short dynamic range of these circuits was a generalised problem that could not be attributed to either the circuit components or their copy-number. However, this data was insufficient to explain whether the impaired activity of these circuits resulted from high expression levels in the dark state and/or low light-inducibility.

3.2.2.2 Inefficient light-reconstitution of split T7RNAP sensors remains unsolved

The initial results obtained using two-plasmid systems showed that our versions of the four circuits failed to reproduce previously published data using the Han *et al* protocol. As it appeared that it was not a variant-specific problem, Opto-T7RNAP*(563) was selected to continue investigating these circuits. In the first place, it was noticed that high expression levels of any fitness gene in the absence of light would undermine the purpose of the project, allowing non-specific growth irrespective of the bioluminescence phenotype. To mitigate this problem, a single-copy of the T7RNAP-controlled mScarlet reporter gene was integrated into a previously validated intergenic region of the *E. coli* genome using a CRISPR-assisted homologous recombination method[57] (**Figure 7A**). The insertion site, named Safe Site 9, is located in position 3,979,535 between genes *aslA* and *glmZ*, and has been shown to provide predictable gene expression levels. This reporter strain was created in a genomic background that lacked arabinose metabolism genes to enable the modulation of Opto-T7RNAP*(563) expression levels, although it contained the native arabinose transporter rather than $LacY^{A177C}$.

To validate both the T7RNAP specificity of this promoter and the arabinose-induction compatibility of the strain, a plasmid expressing full-length T7RNAP from the arabinose-inducible promoter P_{BAD} was transformed into it. Non-transformed cultures elicited low background levels of mScarlet expression, confirming the effective decoupling of this promoter

from host transcription machinery (**Figure 7B**, black bar). The leaky expression of P_{BAD} promoter in transformed cultures was sufficient to cause a considerable activation of the promoter (13-fold), which further increased upon arabinose induction (35-fold) (**Figure 7B**, yellow and red bars). The P_{BAD} promoter is known to be considerably leaky due to its upstream regulatory CAP/CRP site[100], which Baumschlager *et al* removed from their constructs to achieve tighter control over Opto-T7RNAP*(563) expression levels. These results confirmed that mScarlet expression from this promoter is highly T7RNAP-specific and allows for a considerable fold-change response in the presence of functional T7RNAP. Therefore, any increase in mScarlet expression from this baseline in the presence of the sensors should indicate functional reconstitution of T7RNAP.

To better understand the performance of these sensors in both states, a T7RNAP construct was created to infer the highest levels of T7RNAP reconstitution that Opto-

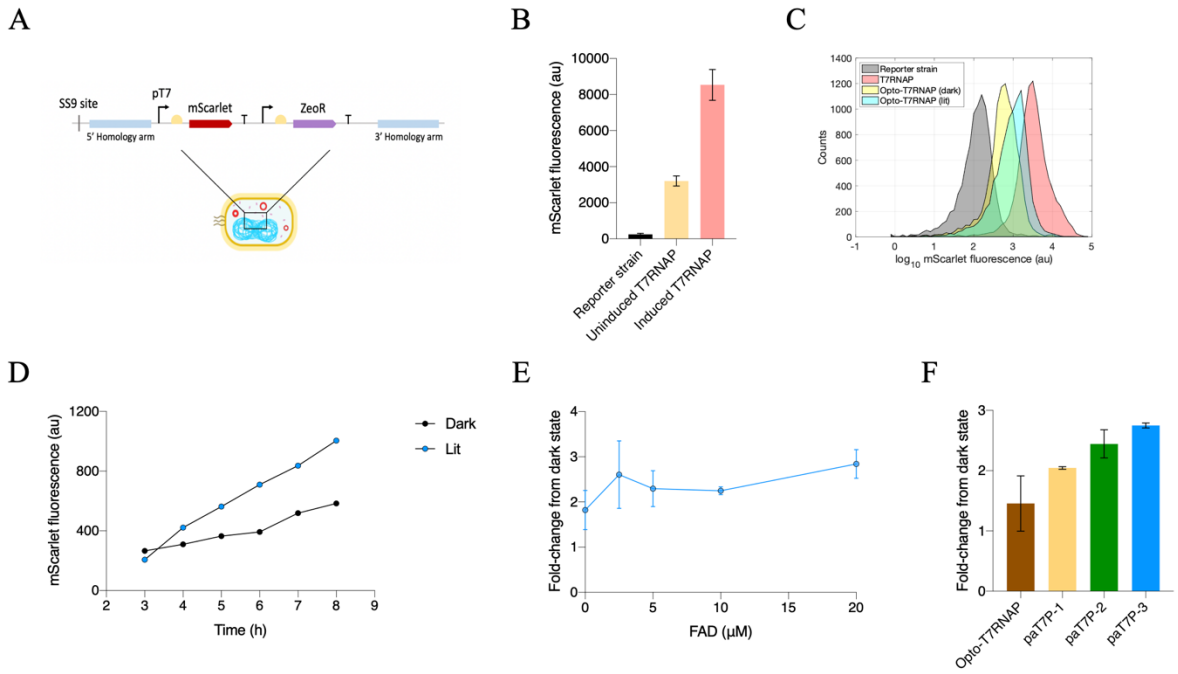


Figure 7 | Inefficient light-mediated T7RNAP reconstitution undermines the repurposing of these circuits.
A Reporter strain containing a single-copy of the T7RNAP-controlled mScarlet fluorescent gene integrated in a previously validated intergenic site of the *E. coli* genome. **B** Fluorescence validation of the reporter strain using a plasmid expressing full-length T7RNAP from the arabinose-inducible promoter P_{BAD}. Bars represent the average ± SD of three biological replicates analysed by flow cytometry using mean population values. **C** Fluorescence levels produced in the reporter strain in the absence of any T7RNAP gene (grey), Opto-T7RNAP*(563) in the dark-state (yellow), Opto-T7RNAP*(563) in the lit-state (cyan), or full-length T7RNAP (red), after a 20-hour induction protocol using either saturating blue light or darkness. Flow cytometry histograms represent 10,000 cells in each condition. **D** Temporal evaluation of Opto-T7RNAP*(563) activity in the dark or under saturating, blue-light illumination throughout the exponential growth phase. Time-points correspond to the averaged fluorescence of 10,000 cells analysed by flow cytometry. **E** Impact of FAD supplementation on Opto-T7RNAP*(563) light-responsiveness in exponential growth phase, after overnight expansion in the same supplemented conditions. Individual values represent the average ± SD of two biological replicates analysed by flow cytometry using mean population values. **F** Light-responses of each photoactivatable split T7RNAP sensor in LB medium supplemented with 2.5μM FAD and grown at 28ºC for 22 hours prior to the start of the experiment, and during the experiment. Bars represent the average ± SD of two biological replicates analysed by flow cytometry using mean population values.

T7RNAP*(563) could possibly generate. This positive control consisted of a T7RNAP construct, split in the same position as Opto-T7RNAP*(563) and expressed from the same CAP-deficient, arabinose-inducible promoter. It only differed in that both T7RNAP fragments were fused to each other to ensure reconstitution using the same linker that used to bind them to the Magnet domains in Opto-T7RNAP*(563). These two plasmids were transformed into the reporter strain and tested as previously done, under uninterrupted blue-light illumination or darkness for 20 hours, using the optimal arabinose concentration according to the original report. Flow cytometry histograms revealed that the Opto-T7RNAP*(563) circuit produced fluorescence levels that were 10-times above the reporter strain baseline, and only increased 2- to 3-times higher as an effect of blue light illumination (**Figure 7C**). This activation was just 2- to 3-times lower than the positive control, which indicated that most likely high levels of mScarlet expression in the dark condition were responsible for the reduced dynamic range of this circuit.

These results confirmed that T7RNAP reconstitution occurs in the absence of blue light in Opto-T7RNAP*(563). In contrast with the Han *et al* protocol, which induces bacterial cultures for 20 hours, Baumschlager *et al* reported to seed cultures 250-times less concentrated and followed their OD while manually adding fresh medium throughout the experiment to maintain bacteria in exponential growth. Moreover, this report always performed time-course experiments and choose time-points at which fluorescence levels stabilised during blue-light induction. This generally took between 8 to 10 hours from the moment blue light was turned on, which was at least 2.5 hours after the start of the experiment. Although it is unclear how to faithfully reproduce this methodology, it seemed that preventing bacteria from reaching stationary phase was key to achieve optimal circuit performance.

To test this hypothesis, cultures were seeded as in Baumschlager *et al* and a time-course experiment was conducted until bacteria reached stationary phase, starting blue-light illumination at 2.5 hours. Although both conditions led to mScarlet signal accumulation over time, overall fluorescence levels were much lower than in previous experiments, and cultures exposed to blue light produced fluorescence at a slightly faster pace (**Figure 7D**). Nevertheless, mScarlet production in the lit-state was not fast enough to generate a fold-change response from the dark higher than previously observed (2-fold).

Comparing this data to the 20-hour long experiment, it is clear that in stationary phase signal accumulation is much higher for all conditions, which likely result from the cessation of cell division. Since basal activation of the circuit appears to occur at a lower constant rate than

in light-induced conditions, it is possible that keeping cells in fast dividing conditions for long enough, as Baumschlager *et al* reported to do, would favour the accumulation of signal in the induced culture. Still, the light-induced signal production rate observed could not possibly generate a 332-fold response in 8-10 hours as they reported. Therefore, the light-mediated T7RNAP reconstitution of our sensor version in these conditions must have been considerably inefficient compared to the original report.

Light-reconstitution in these sensors is regulated by the VVD domains (or the Magnets versions in Opto-T7RNAP*(563)), by physically interfering with the autonomous reconstitution of T7RNAP fragments in the dark state. Upon light-induction, these photoactivatable domains experience a conformational change that displaces the steric effects and lead to T7RNAP activity. It was hypothesised that two factors could be impairing VVD function in the current conditions: either FAD availability is limiting inside the cell, thus, reducing the number of functional VVD domains; or VVD domains are folding inefficiently.

To discard the first possibility, overnight cultures were grown in FAD supplemented medium across a range of concentrations, and light-sensitivity assays were repeated using fresh supplemented medium. A slight increase in the fold-change response of this circuit was observed in all supplemented conditions compared to the non-supplemented group, which might result from higher FAD availability for incorporation into the sensors (**Figure 7E**). However, this improvement was deemed too small to explain the impaired light-reconstitution of the sensor.

The second possibility came to our attention with the publication of the "enhanced Magnets", which were engineered to fold efficiently at 37ºC and had faster dimerization and dissociation kinetics[89]. Apparently, these fungal sensors have low thermostability at 37ºC and require a preincubation at 28ºC for 18-26 hours to ensure proper folding in mammalian cells[88]. Neither of the two groups that developed the T7RNAP-based light-sensors mention any preincubation at 28ºC. To test whether a lower temperature could rescue the performance of any of the four sensors, light-induction tests were performed before and during 28ºC incubations of at least 22 hours. Still, sensor performance remained unaltered regardless of the protocol modifications introduced (**Figure 7F**).

A variety of factors might have contributed to our inability to reproduce the published data for any of these T7RNAP-based light-sensors:

After comparing our plasmids with the genetic sequences now available in online repositories, the only difference that could be identified was a 24-nucleotide cloning scar at the intergenic region between the coding sequence of the first sensor half and the RBS of the second sensor half in paT7P-1/2. Other than the vector backbones, the genetic circuits were identical for paT7P-3 and Opto-T7RNAP*(563). Therefore, it is more probable that any reproducibility problems could have originated from using host strains that differed from the original publications.

The strain used by Baumschlager *et al*. had been engineered to allow concentration-dependent transport of arabinose into the cell compared to the all-or-nothing induction response cause by the native arabinose *E. coli* transporter. While this could influence the expression conditions for Opto-T7RNAP*(563) and cause suboptimal sensor regulation dynamics, it still does not explain why Han *et al*. sensors failed to work in our hands. These sensors (paT7Ps) were constitutively expressed in the *E. coli* TOP10 strain, which according to the commercial webpage has an identical genome to the E. cloni® 10G strain used in this project.

In conclusion, protocols had been replicated and investigated to the best of our skills, expression systems and reporter genes properly validated, and genetic sequences and strains compared with the original materials. However, it was not possible to draw any clear conclusion as to what might have failed in reproducing the performance of these sensors. Hence, since the light-induced responses obtained under optimised conditions for all sensors appeared insufficient to fulfil the goal of the project, they were not used any further.

3.3 CcaS-CcaR: a cyanobacterial two-component system

3.3.1 Background

One of the earliest light-sensitive transcriptional circuits repurposed from nature to engineer synthetic behaviours was the cyanobacterial two-component system CcaS-CcaR[101]. This circuit originally evolved to adjust the light-harvesting pigments of cyanobacteria to the availability of green and red ambient light, in a process known as complementary chromatic acclimation[102,103]. Its components are hence called complementary chromatic acclimation sensor (CcaS) and regulator (CcaR). The sensor is a cyanobacteriochrome that uses a heme-derived chromophore called phycocyanobilin (PCB) to respond to green/red light, which often requires additional metabolic genes to be synthesised in heterologous hosts[104] (**Figure 8A**). PCB is covalently bound to a conserved residue in the light-sensing domain of CcaS, located near a predicted N-terminal transmembrane helix. It exists in two photoconvertible isomeric states that alter CcaS conformation when switching, and change the autokinase activity of its C-terminal transactivation domain: a green-absorbing ground state with low activity; and a red-absorbing excited state with high activity[103,105] (**Figure 8B**). In its autophosphorylated form, CcaS rapidly transfers the phosphate group to the N-terminal receiver domain of CcaR, which increases the DNA-binding affinity of its C-terminal domain (**Figure 8C**). Phosphorylated CcaR (CcaR~P) recognises a regulatory sequence upstream of its cognate promoter (P_{cpcG2}) and activates transcription (**Figure 8D**). The phosphorelay remains active until PCB relaxes to

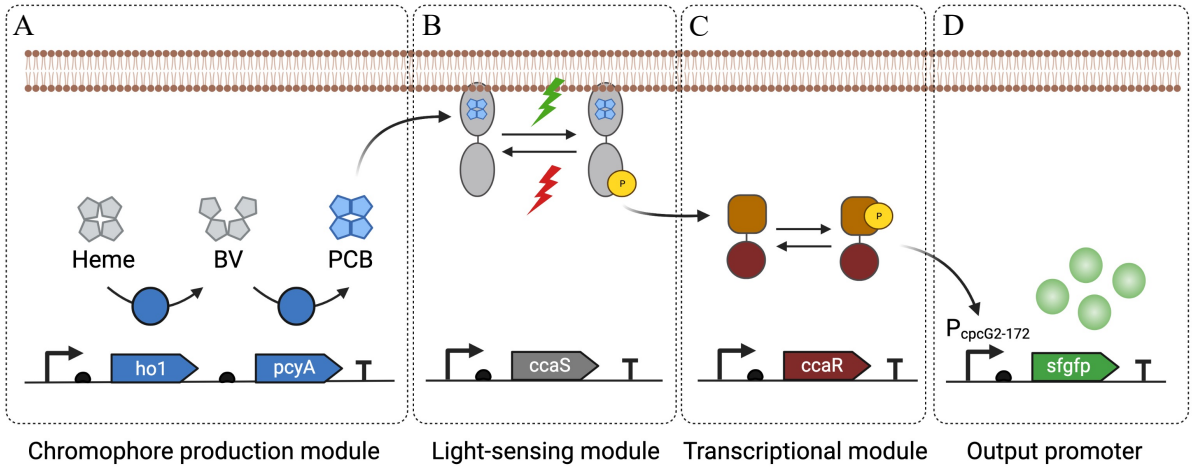


Figure 8 | CcaS-CcaR two-component phosphorelay circuit allows transcriptional control using green/red light.
A Metabolic genes *ho1* and *pcyA* are required to produce the chromophore phycocyanobilin (PCB) in *E. coli*. **B** CcaS regulates the activation of the circuit in response to green/red light which influences its autokinase activity and phosphotransfer to CcaR. **C** CcaR is susceptible to both phosphorylation and dephosphorylation by CcaS and gains DNA-binding affinity in the phosphorylated form. **D** The output promoter contains a specific regulatory sequence to which phosphorylated CcaR binds and activates transcription.

its ground state, or red light catalyses its reversal, and transcription is deactivated via a putative phosphatase CcaS activity that dephosphorylates CcaR[102].

CcaS-CcaR appears to interact minimally with endogenous signalling components and promoters in *E. coli*, which motivated its rigorous optimisation for reliably engineering light-controlled processes in bacteria[58]. First, the output promoter was truncated to remove a secondary constitutive transcriptional start site that was making it ~ 17-times leakier in the absence of green light ($P_{cpcG2\text{-}172}$). Then, PCB biosynthesis levels were optimised to maximise chromophore availability for CcaS assembly without causing heme depletion-related toxicity. It appears that insufficient PCB levels affect circuit activation because apo-CcaS has constitutive phosphatase activity, which reduces the capacity of active CcaS molecules to maintain CcaR phosphorylated upon induction. Finally, CcaS and CcaR expression levels were simultaneously tuned to identify the best ratio for optimal circuit regulation, resulting in a 20-fold improvement in the dynamic range. Later studies found that most of the internal protein sequence that separates the two functional domains in CcaS can be deleted to obtain either higher dark-state stability or inverted light-regulation[59,106]. Following its optimisation, the temporal dynamics of this CcaS-CcaR circuit were comprehensively studied in response to varying light intensities, and used to develop a predictive mathematical model that allows programming custom gene expression patterns with high reproducibility[107].

These initial efforts established a robust light-controlled gene expression platform, which have encouraged a considerable research output in less than a decade. The circuit has been multiplexed[101,108], rewired[109], and modelled with other light-sensitive genetic circuits[110]; implemented in more complex circuit architectures[111]; adapted for its deployment in other organisms[112–114]; guided the development of hardware for *in situ* characterisation and manipulation of microorganisms with light probes[68,70,115,116]; and used in several applications, including in metabolic engineering[117–119], microbiome engineering[120], and automated *in silico* feedback control of gene expression[74,80,121]. As a result, plenty of materials and insights in distinct research contexts are available to repurpose CcaS-CcaR in new directions. Most importantly, it has been used both to engineer light-controlled bacterial growth through a limiting metabolic gene[73,74], and light-controlled bacterial survival to an antibiotic using an antibiotic resistance gene[121]. Therefore, its suitability for this project is certainly justified.

3.3.2 Results and discussion

In the context of this project, CcaS-CcaR would be used to generate single-cell phenotypic differences to select specific genotypes from a bacterial population. In a hypothetical selection experiment, the basal state of CcaS-CcaR would be defined by the bioluminescence output of the template protein that wants to be improved. For instance, different variables in the system could be adjusted to set an activation threshold for this template protein, allowing only bacteria expressing protein variants with either similar or brighter light outputs to be selected. Therefore, it is essential that CcaS-CcaR remains sufficiently inactive during the experiment to prevent rescuing bacteria with poor bioluminescent phenotypes when non-favourable growth conditions are established.

Despite the current understanding of CcaS-CcaR circuit dynamics, two main limitations were faced when trying to repurpose this circuit for the project. To this date, all published insights relied on constant red-light illumination conditions as the basal circuit state to evaluate green light activation, and were strictly collected in steady-state exponential growth phase. The utilisation of red light in this project could be problematic because it desensitises the photoreceptor to green light[107]. This might complicate the bioluminescence-mediated activation of CcaS at sufficient levels to elicit practical phenotypic changes because bioluminescence outputs are expected to be dim and short-lived. Furthermore, continuous culture conditions might be prohibitively expensive for this bioluminescence experiments due to the high cost of luciferin substrates. All characterisation reports mentioned above achieved steady-state exponential growth conditions for long periods of time by starting cultures with sufficiently low concentrations of bacteria to ensure nutrient depletion occurs slower. However, seeding cultures at very low concentrations has several shortcomings: it reduces the screening throughput of bioluminescent mutants; it might cause diversity loss due to lower number of starting replicates per phenotype; and selection protocols might become too long for human implementation, unless the system is able to operate overnight unsupervised. Although the utilisation of red light, a turbidostat, and/or low seeding conditions are not yet discarded, it would be valuable to understand the circuit dynamics in the dark and outside of steady-state.

3.3.2.1 Bacterial growth dilutes CcaS-CcaR basal activation in the dark

The first step towards the implementation of CcaS-CcaR in the laboratory was to identify experimental conditions that could support steady-state exponential growth throughout a light-induction protocol. To this end, a simple mathematical approach was developed to

model bacterial growth curves that reliably predicted culture growth state from OD values (*Materials and Methods*). It consists in empirically measuring the relationship between culture OD and the incremental changes that occur between OD measurements over time. These growth curves were used to identify the turning point at which bacteria change from unrestricted growth into a progressive growth rate reduction, which characterises the exponential phase exit and transition into stationary phase. It was observed that under constant growth conditions, culture capacity to sustain exponential growth always terminated around OD_{600} ~ 0.15-0.30, irrespective of seeding conditions or the metabolic burden caused by the synthetic circuits harboured by the strain. Despite differences in strains and media, this range of values accurately matched the OD values used as a mark for exponential phase exit in previous CcaS-CcaR studies. Following this observation, the average value OD_{600} = 0.23 was used to identify seeding conditions that were able to sustain exponential growth for reasonable experimental lengths using an *E. coli* strain expressing CcaS-CcaR with sfGFP as the output gene.

To evaluate CcaS-CcaR dynamic range, cultures were grown under constant darkness or saturating green-light illumination using previously optimised growth conditions. After light-induction, flow cytometry histograms revealed two clearly differentiated fluorescent populations with a short overlap between them (**Figure 9A**). Cultures exposed to constant, saturating green light had a mean fluorescence 52-times higher than those grown in the dark. This result only differed from the originally reported 117-fold dynamic range in the fluorescence levels of the respective uninduced controls, which were twice higher in the dark than under constant, saturating red light. Therefore, it was hypothesised that CcaS might be able to switch to some extent into its active form in the absence of green light, which red light probably minimises by reverting and/or stabilising it into the ground state.

To gain further insights into the effects of bacterial growth on CcaS-CcaR performance and its basal activation in the dark, cultures were grown again in the same conditions and monitored over time until cell division ceased. The OD of each sample was measured before fluorescence analysis to map the circuit response in each condition across the bacterial growth curve. It was observed that fluorescence differences between green light-illuminated and dark cultures varied across time and were directly proportional to the division rate of the culture (**Figure 9B**) (PPC = 0.715; p-value = 0.0003). In early exponential phase, differences between conditions were very small (~ 3-fold) and became progressively larger as cells increased their division rate. They peaked around 30 minutes after reaching maximum growth rate, at the

beginning of exponential phase exit, corroborating the dynamic range detected in the previous experiment (52-fold); this delay was attributed to sfGFP maturation time (t_{90} = 39.1 ± 4.7 min)[122]. Finally, fluorescence differences between conditions waned again as bacteria transitioned into stationary phase, where they seemed to plateau (~ 8-fold). Although it is clear that the dilutive effect of cell division was responsible for the CcaS-CcaR dynamics observed, fluorescence differences between conditions should have remained constant, because both cultures divided at the same pace and were exposed to unvarying light conditions. This indicates that transcription activation in one of the two conditions is more susceptible than the other to the dilution of circuit components. Since illuminated cultures were induced with a saturating, green-light intensity, output promoter transcription must have remained near-maximal throughout the experiment in these cells. Therefore, bacterial growth appears to influence the basal activation of CcaS-CcaR in dark cultures.

In this phosphorelay circuit, CcaR~P levels control output signal transcription from $P_{cpcG2\text{-}172}$ (**Figure 8**). CcaR~P levels are determined by the net kinase activity equilibrium of active CcaS molecules with kinase activity, and inactive CcaS molecules with putative phosphatase activity. Both components in CcaS-CcaR have been shown to have unspecific activity: (a) CcaS leads to higher pathway activation in dark conditions than under constant red-light control; (b) and CcaR can start transcription in the absence of CcaS, which has been hypothesised to happen either via autophosphorylation, phosphorylation by endogenous *E. coli* histidine-kinases (HKs), or weak DNA-binding activity in the non-phosphorylated form[58].

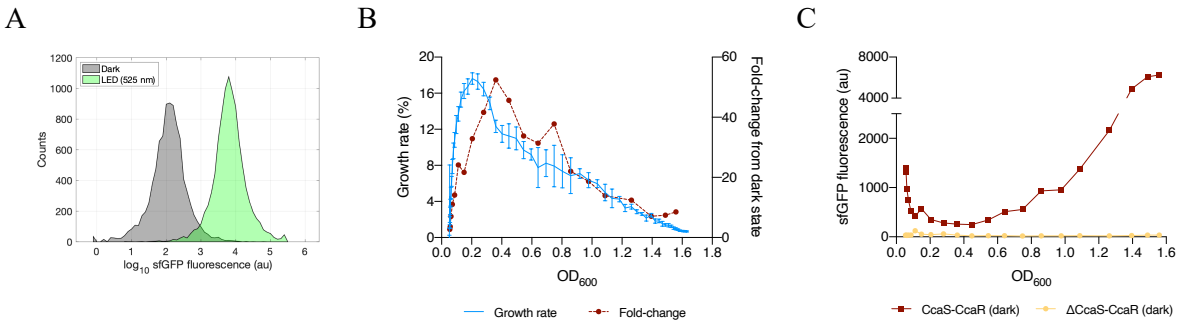


Figure 9 | CcaS unspecific activation in the dark is buffered by the dilutive effect of bacterial growth.
A Fluorescence levels produced in cultures expressing CcaS-CcaR circuit, under constant darkness (grey) or saturating green light (green), by the end of exponential growth phase. Histograms represent 10,000 events in each condition. **B** Relationship between bacterial growth rate (blue), expressed as a percentage over 15-minute intervals, and CcaS-CcaR fold-induction from the dark (red) from cultures grown under constant darkness or saturating green light until stationary phase. Fold-change measurements correspond to the light/dark ratio of the averaged fluorescence of 10,000 cells in each condition analysed by flow cytometry. Growth rate values were calculated from OD measurements collected in a microplate reader and time-points correspond to the average ± SD of twelve technical replicates. **C** Fluorescence levels produced during bacterial growth in dark cultures expressing either full CcaS-CcaR circuit (red) or a version of the circuit lacking CcaS gene (yellow). Individual values correspond to the averaged fluorescence of 10,000 cells analysed by flow cytometry. Fold-change and sfGFP fluorescence measurements in **B** and **C**, respectively, are separated by 30-minute intervals.

To assess how the dilution of each component influences signal production separately, a truncated circuit version was created by deleting CcaS gene and studied alongside the complete circuit in the dark during bacterial growth. As previously reported, CcaR was able to activate transcription in the absence of CcaS, but fluorescence levels remained relatively low and seemingly unaffected by bacterial growth phase or division rate. For fluorescence signal levels to remain constant at different division rates, CcaR unspecific activation of transcription should be higher in fast-diving cells than in low-dividing cells. The only suggested hypothesis that could explain this observation would be a higher signalling activity of cross-reactive HKs during exponential phase compared to in stationary phase. Instead, CcaS spontaneous activation resulted in varying fluorescence levels that were inversely correlated to the division rate of bacteria (**Figure 9C**, red line). They lowered as cells progressed through exponential phase, reaching a stable minimum that lasted until approximately 1 hour after exponential phase exit, and then gradually increased before suddenly saturating in non-dividing cells. Therefore, it is the dilution rate of CcaS molecules that primarily influences transcription activation in dark conditions.

Taken together, these results suggest that CcaS spontaneous activation in the dark has accumulative effects during bacterial growth that can be attenuated by fast division rates. This means that shorter cell cycles allow less time for the proportion of CcaS molecules in the cell that spontaneously activate to increase, maintaining lower levels of circuit activation. As previously observed, the relaxation of active CcaS in the dark is slow, suggesting that PCB is more stable in its red-absorbing form. Moreover, these authors suggested that dark-reversion during bacterial growth might be achieved primarily through dilution of active CcaS molecules by both cell division and *de novo* production of CcaS in its ground form[107]. This would support the hypothesis above and also be consistent with the steep increase in basal signal production observed in stationary phase (**Figure 9C**), because starvation simultaneously downregulates cell division and protein synthesis. In line with this evidence, it was concluded that bacterial growth buffers the basal activation of CcaS-CcaR in dark cultures by diluting spontaneously-activated CcaS photoreceptors, most likely, via cell division and *de novo* CcaS synthesis.

3.3.2.2 A pulse validation assay to screen for bioluminescence activations

In the previous experiments, bacterial growth was found to influence the capacity of CcaS-CcaR to generate gene expression differences in response to light by altering the basal state of the circuit. Those insights were collected using constant, fully-activating, green light

which are likely to be unsustainable conditions for a bioluminescence reaction. Instead, bioluminescence is usually produced in the form of dim light pulses that might cause transient, non-saturating activations in CcaS-CcaR. The interplay between these two factors might complicate evaluating the temporal activation of CcaS-CcaR when engineering an effective bioluminescence delivery strategy. For this reason, it appeared important to characterise the reporting capacity of CcaS-CcaR of a defined light-stimulus during bacterial growth.

To assess the pulse activation of CcaS-CcaR, distinct points around the optimal response region of the bacterial growth curve were sampled from a dark culture for induction. Samples were exposed to identical 1-hour long, saturating, green light pulses and returned to dark conditions, keeping an uninduced control in each case. Then, signal production was monitored at 30-minute intervals by flow cytometry to follow the temporal evolution of this perturbation (**Figure 10A**). As expected, CcaS-CcaR generated distinct fold-induction responses during bacterial growth in response to an identical stimulus, showing an optimal reporting capacity around near-maximal division speed (**Figure 10B**; green line). Induced gene expression differences persisted up to 1 hour after the circuit returned to the dark state, which occurred at the same time regardless of growth conditions (**Figure 10C**). These induction points were selected around late-exponential phase and early transition into stationary phase knowing that CcaS-CcaR dynamic range is larger around this part of the bacterial growth curve (**Figure 9B**). By the time these cultures have sufficiently transduced light activity into gene expression differences, the medium has become nutritionally poor. This means that the duration of these phenotypic differences is more likely to be defined by the turnover rate of the protein

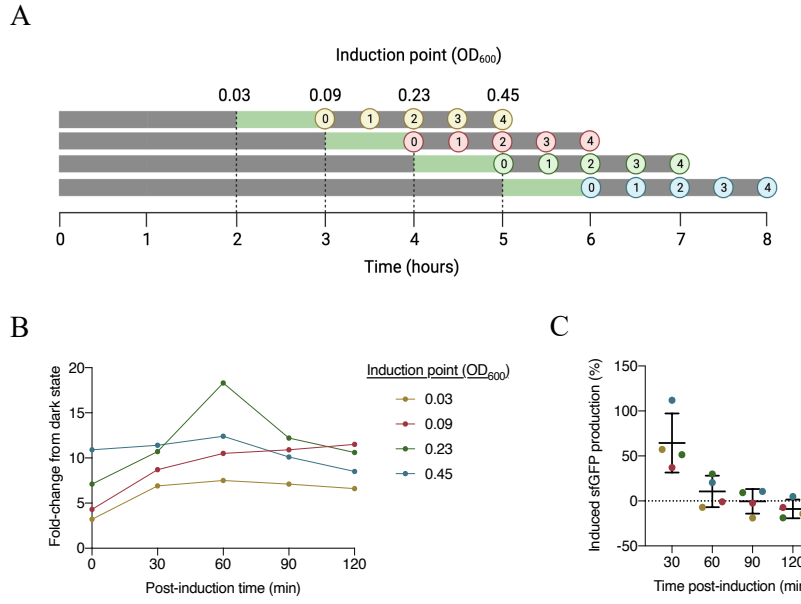


Figure 10 | CcaS-CcaR reports an identical light pulse differently during bacterial growth. **A** Distinct growth points around late-exponential phase were interrogated with a 1-hour long, saturating green light pulse, and signal production was monitored in both induced and uninduced conditions every 30 minutes. **B** Fold-inductions reported by CcaS-CcaR after pulse interrogation. Fold-change measurements correspond to the light/dark ratio of the averaged fluorescence of 10,000 cells in each condition analysed by flow cytometry. **C** Induced signal production expressed as a percentage change over 30 minutes corrected for signal production in the dark state. Bars represent average ± SD of four biological replicates.

of interest than the dilution rate of cell division. In practical terms, this suggests that survival-based selection mechanisms, such as bactericidal antibiotics and their resistance genes, might be more suitable than growth-mediated strategies to select of bioluminescent bacteria using CcaS-CcaR. However, the utilisation of stable proteins, like sfGFP in this experiment, might progressively erase these differences due to unspecific accumulation of signal, as observed in the last hour post-induction of the last induced cultures (OD600=0.23 and 0.45) (**Figure 10B**). Therefore, antibiotics targeting protein synthesis might be especially useful candidates for selection in these conditions.

3.3.2.3 Exploring CcaS sequence for downstream photoreceptor engineering

CcaS-CcaR showed a promising capacity to generate significant and lasting gene expression differences in response to temporary light emissions, but had two apparent limitations for further implementation in the project. Regarding its signalling activity, the accumulation of unspecific CcaS activity during bacterial growth could restrict both the utilisation and effectiveness of the circuit for selection purposes. In terms of compatibility with a bioluminescence-mediated activation, the chromophore-binding domain of CcaS is spatially segregated to the vicinity of the inner membrane by a putative N-terminal transmembrane helix. This complicates engineering fusion proteins at the N terminus for close-contact delivery of photons or resonance energy transfer to the chromophore; while distant delivery of bioluminescence might be dispersed and absorbed by other cellular structures before reaching the target, reducing CcaS activation efficiency.

As mentioned above, some protein regions in CcaS seem to be redundant for its activity. Its sequence comprises five recognisable domains: a predicted 23-residue long, N-terminal transmembrane (TM) helix; a cGMP phosphodiesterase/adenylyl cyclase/FhlA (GAF) domain containing the chromophore-binding site, Cys147; two internal Per-ARNT-Sim (PAS) domains of unidentified function; and a C-terminal histidine-kinase domain with an autophosphorylation site, His537 (**Figure 11A**, green). The internal PAS domains seem to constitute a vestigial protein region that used to sense other ligands[103], and their deletion results in a more stable CcaS variant when truncated at positions Gln221 and Leu513; named miniCcaS#10 [59,106] (**Figure 11A**, yellow). Furthermore, the putative N-terminal TM helix was deleted to facilitate photoreceptor purification for *in vitro* characterisation without apparent loss of light-regulation (**Figure 11A**, red). Consequently, these two truncated variants were

characterised alongside full-length CcaS to assess whether they could facilitate the following engineering stages of this project

First of all, the impact of expressing these circuit variants on bacterial growth was measured by tracking cultures harbouring each design under identical growth conditions. The growth-sensitive signalling activity of CcaS-CcaR makes it necessary that strains divide at comparable speeds and are induced at similar points in their growth curves to accurately cross-compare results. No significant differences were observed amongst the three strains with respect to exponential phase length and maximum growth rate, although miniCcaS#10 escaped lag phase with an approximate 1-hour delay (**Figure 11B**). Therefore, any strain differences in

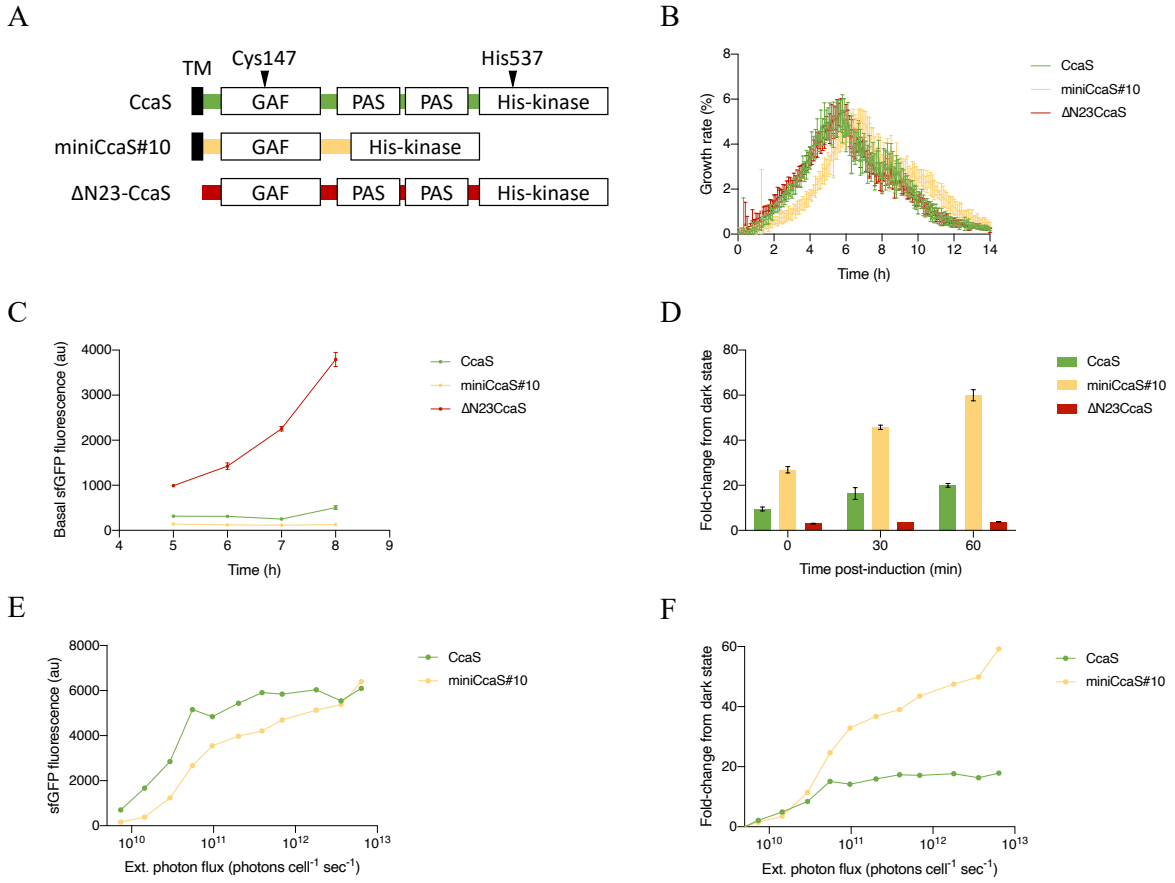


Figure 11 | Internal truncation increases CcaS ground-state stability while membrane detachment disrupts it.
A Domain architecture of full-length CcaS photoreceptor (green), miniCcaS#10 (yellow), and a soluble N-terminally truncated CcaS (red). **B** Bacterial growth over time of three strains expressing CcaS-CcaR circuit with each of the photoreceptor variants, expressed as a percentage over 5-minute intervals. Growth rate values were calculated from OD measurements collected in a microplate reader and time-points correspond to the average ± SD of twenty-four technical replicates. **C** Fluorescence levels in the dark across the optimal basal activity window of CcaS-CcaR for each photoreceptor variant. Individual values represent the average ± SD of two biological replicates analysed by flow cytometry using mean population values. **D** Fold-induction of CcaS-CcaR with each of the photoreceptor variants after a 1-hour pulse of saturating green light. Bars represent the average ± SD of two biological replicates analysed by flow cytometry using mean population values. **E** Fluorescence levels and **F** fold-inducion responses elicited by CcaS-CcaR with either the full-length photoreceptor (green) or miniCcaS#10 (yellow) across a range of 1-hour green-light pulses of varying intensities at 1-hour post-induction. Fold-change values in **D** and **F** correspond to the light/dark ratio of the averaged fluorescence of 10,000 cells in each condition analysed by flow cytometry.

fluorescence levels should reflect distinct CcaS signalling activities rather than differential accumulation of circuit components and signal output.

Secondly, the unspecific activation of these CcaS variants was compared by scanning their signal production in the dark. Previous characterisation data had shown that optimal CcaS-CcaR activity occurs at near-maximal growth rates when basal circuit activation reaches bottom levels for an approximate 2-hour window (**Figure 9C**). During this period, fluorescence levels in the dark were around 50-60% lower in miniCcaS#10 compared to full-length CcaS (**Figure 11C,** yellow line), which matched the results obtained in the original characterisation of this variant under red-light illumination[106]. This would imply that miniCcaS#10 spontaneous activation is negligible, thus rendering the catalysis of ground-state reversion by red light unnecessary. However, a second study reported 4-times higher stability in the ground state under red-light control for the same construct compared to full-length CcaS[59]. This latter characterisation was probably more accurate because it used flow cytometry instead of a microplate reader to measure fluorescence signal levels, supporting that miniCcaS#10 regulation still benefits from red-light effects despite higher ground-state stability. Although it also appeared that the optimal basal activity window of miniCcaS#10 lasted longer, the experimental length tested could not discard that this originated due to the 1-hour delay in the growth curve of this strain compared to the other two. In contrast, ΔN23-CcaS showed increasingly high fluorescence signal levels in the dark that accumulated as soon as cells exited exponential phase at an average hourly rate of ~ 56% (**Figure 11C,** yellow line); these were 3-to-9-times higher than full-length CcaS, and 7-to-29-times higher than miniCcaS#10.

Third, the transient activation of each variant was tested following the previously optimised light-pulse interrogation assay. As previously reported, miniCcaS#10 elicited slightly higher fluorescence levels in response to green light while accumulating lower basal activity in the dark compared to full-length CcaS, resulting in a ~ 3-fold improvement in its dynamic range (**Figure 11D**). Instead, the dynamic range of ΔN23-CcaS was practically impaired due to its high fluorescence levels in the dark, despite its kinase activity in the lit-state also being higher than full-length CcaS (~ 20-50%).

Finally, it was hypothesised that the lower basal activation rate of miniCcaS#10 would decrease the range of light intensities required to generate specific fold-responses compared to full-length CcaS. This is especially relevant for bioluminescence control, as the photonic rate of an enzyme is expected to be lower than that of an electronic device. To test this, both sensor variants were exposed to a titration of 1-hour pulses of green light across the range of light

intensities that a commercial 4-volt LED could deliver and fluorescence levels were analysed after allowing full dark-reversion of the circuit. In both cases, the lower detection limit of these CcaS photoreceptors could not be detected, demonstrating that they are considerably sensitive to ambient light. Cultures expressing full-length CcaS had higher fluorescence levels than those expressing miniCcaS#10 for identical light intensities at sub-saturating levels (**Figure 11E**). This is expected because green light increases the activation rate of CcaS molecules, which is higher to start with in full-length CcaS than in miniCcaS#10. Fold-induction responses are comparable between variants until full-length CcaS reaches an early dose-response saturation ($5.5x10^{10}$ photons $cell^{-1}$ sec^{-1}), supporting that their activation differences primarily originated from their distinct basal activation rates (**Figure 11F**). After this point, miniCcaS#10 continues to benefit from the catalytic effect of green light, without apparently reaching saturation by the highest light intensity tested. Therefore, the lower basal activation constant of miniCcaS#10 elongates its dose-response curve and allows it to differentiate photonic fluxes across two extra orders of magnitude and generate proportional fold-change responses. This dose-response curve is a piecewise function that has an initial linear segment of high sensitivity to light-intensity differences, similar to full-length CcaS, and a less sensitive region that responds to larger input signal differences.

In conclusion, it appears that the localisation of CcaS in the membrane has an important role is physically stabilising the photoreceptor to prevent unspecific conformation switching in the absence of green light. Since this photoreceptor is not native in *E. coli*, it is unlikely that other scaffolding or regulatory proteins are involved in this process, unless they are shared amongst bacterial two-component systems or have overlapping functions. Consequently, both solubilising the photoreceptor into the cytoplasm to facilitate distant bioluminescence activations, and fusing bioluminescent proteins to the N terminus of CcaS are unlikely to work. Furthermore, the internal truncation of CcaS in miniCcaS#10 increases its ground-state stability, which expands both its capacity to generate gene expression differences in response to light and the range of light intensities that differentially activate it. Therefore, it is likely to be a more suitable component to differentiate and select bioluminescent phenotypes, which might also allow improvement of bioluminescent proteins over a larger brightness span, and higher selection tunability when modulating luciferin concentration in the medium.

3.4 pDusk/pDawn: synthetic two-component system circuits

3.4.1 Background

One last set of light-sensitive transcriptional circuits selected for this project was the pDusk/pDawn collection based on the two-component system FixL-FixJ from the legume-root microsymbiont *Bradyrhizobium japonicum*. In the *wild-type* system, a soluble HK sensor called FixL induces the expression of nitrogen fixation genes under low oxygen conditions through the transcription regulator FixJ[123]. FixL regulates the process by switching between an oxygen-free form with autokinase activity and phosphotransfer capacity to FixJ, which activates transcription from specific promoter sequences; and an oxygen-bound form with phosphatase activity that deactivates the response[124]. This phosphorelay was reprogrammed to sense light instead of oxygen by replacing the sensory domain of FixL for the blue light-sensitive domain of *Bacillus subtilis* photoreceptor YtvA[125]. The resulting photoreceptor chimera (YF1) retained the original ligand-repressed kinase activity, phosphorylating FixJ in the dark and dephosphorylating it upon blue light-stimulation. As in other LOV proteins, blue light catalyses the formation of a covalent bond between the flavin mononucleotide (FMN) chromophore and a cysteine residue in the sensory domain, locking YF1 in its phosphatase conformation until it thermally reverts to the ground state.

The synthetic YF1-FixJ phosphorelay and a FixJ-regulated promoter (P_{FixK2}) were cloned together to engineer a light-repressed transcriptional circuit called pDusk (**Figure 12A**). However, it was found that FixJ is constitutively active in *E. coli* in the absence of YF1, which has been hypothesised to occur due to cross-phosphorylation by endogenous HKs. This implies that pDusk transcriptional control in this host is mainly achieved through the photoactivatable phosphatase activity of YF1. For practical purposes, the λ bacteriophage repressor protein cI and its target promoter P_R were placed under pDusk control to engineer a light-inducible gene expression system, resulting in the creation of pDawn[60]. In this circuit, phosphorelay activity keeps transcription tightly repressed in the dark through constant cI production, while blue light-inhibition derepresses the strong P_R promoter (**Figure 12B**). This signal-inversion cassette significantly augmented the effects of blue light by both enhancing the off-state regulation and the transcriptional capacity of the circuit upon induction, resulting in a 35-fold improvement in its dynamic range. Later studies managed to dissect the signal transduction mechanism of YF1 through crystallographic insights and extensive sequence analysis using pDusk as a functional assay[126,127]. Several YF1 mutants were identified in the process with

altered signal responses, such as inverted phenotypes with light-inducible kinase activity. These mutants can also be used to achieve light-induced gene expression in pDusk without relying on extra genetic elements.

These blue light-sensitive circuits appear to offer several practical advantages for their deployment in this project. In respect to engineering a mechanism for bioluminescence activation, YF1 is a soluble sensor that might tolerate N-terminal fusions for close-contact delivery of photons; and elicits saturating responses at low blue-light intensities in both circuit configurations. As previously argued for the photoactivatable split T7RNAP sensors, the slow-reverting dynamics of LOV domains should sustain signalling responses caused by short-lived bioluminescence pulses, which might facilitate bacterial selection. Moreover, the ubiquity of FMN synthesis should not restrict chromophore availability. In the context of bacterial growth, both pDusk and pDawn showed stable light-responses that persisted after culture saturation, with low levels of unspecific signal accumulation across 20-hour experiments. Finally, these two circuits have been combined to control bacterial growth with a toxin-antitoxin system, in which darkness causes growth defects while blue light simultaneously represses toxin expression and induces antitoxin production to restore homeostasis[75].

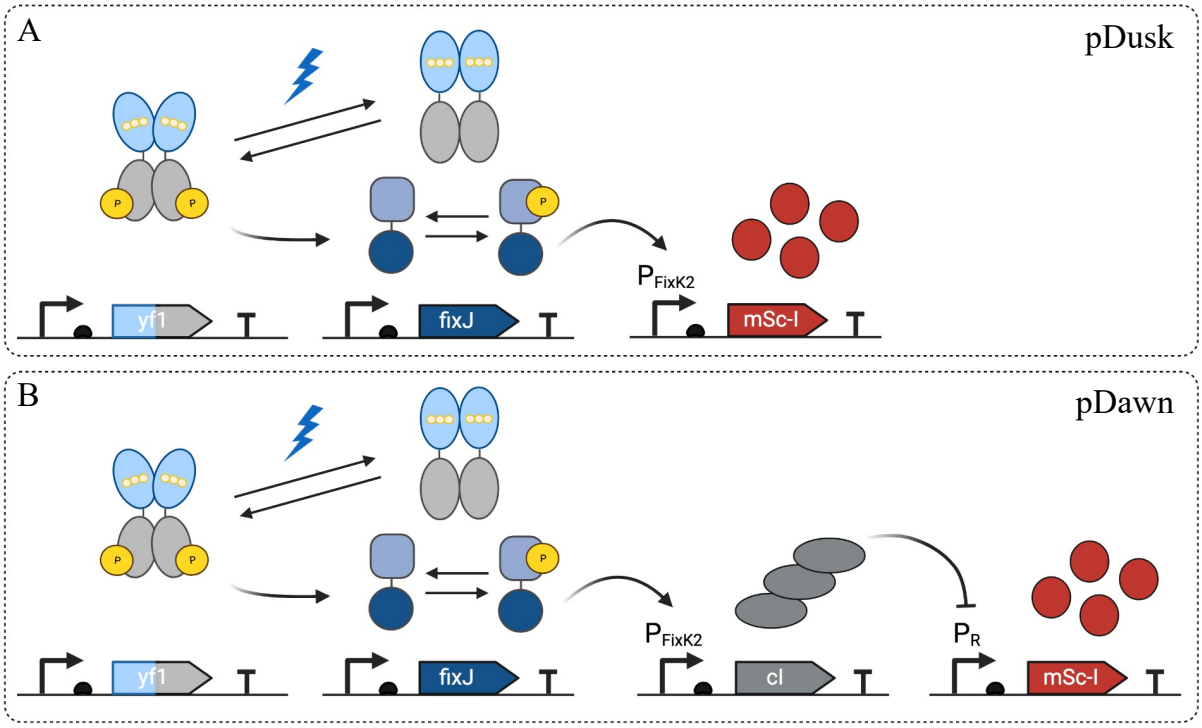


Figure 12 | Synthetic YF1-FixJ two-component system allows distinct modes of transcriptional control using blue light. **A** Light-repressed gene expression in YF1-FixJ is controlled by the blue light-inducible phosphatase activity of YF1 that deactivates FixJ-mediated transcription from FixK2 promoter (pDusk). **B** Light-inducible gene expression can also be engineered by placing the λ bacteriophage repressor protein cI and its target promoter under the control of YF1-FixJ (pDawn). In this circuit, constant production of cI in the dark keeps P_R transcription tightly-repressed and light-inhibition of the phosphorelay causes its derepression, which leads to strong signal production.

3.4.2 Results and discussion

3.4.2.1 pDawn becomes temporarily responsive during the transition into stationary phase

The light-inducible transcriptional circuit pDawn was initially selected for implementation in this project over pDusk based on its improved circuit dynamics and higher compatibility for bioluminescence-mediated fitness acquisition. The original plasmid was kindly shared by Prof Andreas Möglich and also assembled in the laboratory from synthetic DNA fragments given its relative genetic simplicity. The assembled circuit differed from the original sequence in being codon-optimised for *E. coli*, containing a bicistronic reporter co-expressing a fluorescence protein and an antibiotic resistance gene, and different terminators. In an attempt to reduce any potential transcriptional leakage that could confer unspecific fitness in the future, this version of pDawn was cloned into two different vector backbones that had either identical (pBR322) or lower copy-number (pSC101) than the original plasmid. Since the red fluorescence reporter DsRed-Express2 used in the original characterisation was unavailable due to existing conflicts of interest, the fast-maturing protein mScarlet-I was selected to assess these three pDawn circuit variants[63]. This protein had been recommended at the time as the most appropriate red fluorescent reporter for monitoring cell processes with high temporal resolution due to its monomeric nature, and *in vivo* brightness and maturation time[122].

To validate the activity of these pDawn variants and compare their performances, cultures expressing each genetic circuit were subjected to the published light-responsiveness assay. After 20 hours of growth in either darkness or constant blue-light illumination using a saturating and a near-saturating light intensity, the original pDawn circuit showed a 120-fold increase in fluorescence levels from the dark control, which is around 4-times lower than the reported dynamic range (**Figure 13A**, blue). This variation might result from using different host strains, fluorescent reporter genes, and/or flow cytometry settings. Both assembled versions of pDawn showed partially- or fully-compromised light-responsiveness, and around 5-times higher fluorescence levels in the dark compared to the original circuit regardless of their copy-number (**Figure 13A**, green and brown). The low-copy number assembled circuit elicited a lack of inducibility at both light intensities; while the unmodified copy-number version had lower sensitivity to near-saturating blue light conditions, and approximately twice lower signal production that the original pDawn under saturating illumination. These results suggest that dark-state regulation in pDawn is relatively insensitive to the number of output promoters in the cell, although some circuit components might require a certain copy-number to effectively derepress them. Furthermore, they demonstrate how non-coding DNA sequence

modifications might affect circuit performance by destabilising the expression levels of its components. For instance, it was noticed that the *wild-type* YF1-FixJ bicistronic sequence contains the RBS of FixJ embedded in the last codons of YF1, which was accidentally modified when codon-optimising the genes. This probably altered the translation rate of FixJ and, thereby, the ratio of YF1/FixJ protein levels. As shown for CcaS-CcaR, the expression ratio between HK sensor and transcription regulator can significantly influence the signalling activity of bacterial two-component systems[58]. Later in the project, it was also discovered that the bicistronic design of the dual reporter used in these assembled pDawn circuits displayed higher leakiness and impaired inducibility in CcaS-CcaR (*Chapter 5*). In light of these results, the original pDawn circuit was selected to continue with the characterisation process.

In the published characterisation of pDawn, fluorescence signal levels in the dark remained tightly regulated despite culture saturation, as corroborated above. The authors also studied the relationship between illumination length and signal production, showing that circuit responses to blue-light pulses as short as 10 minutes were detectable 20 hours later. The capacity of pDawn to respond to light across growth was not reported, but these light pulses were consistently delivered at OD ~ 0.4 without providing any specific reason. Therefore, it was hypothesised that pDawn might have an optimal response window to blue light located around this culture density.

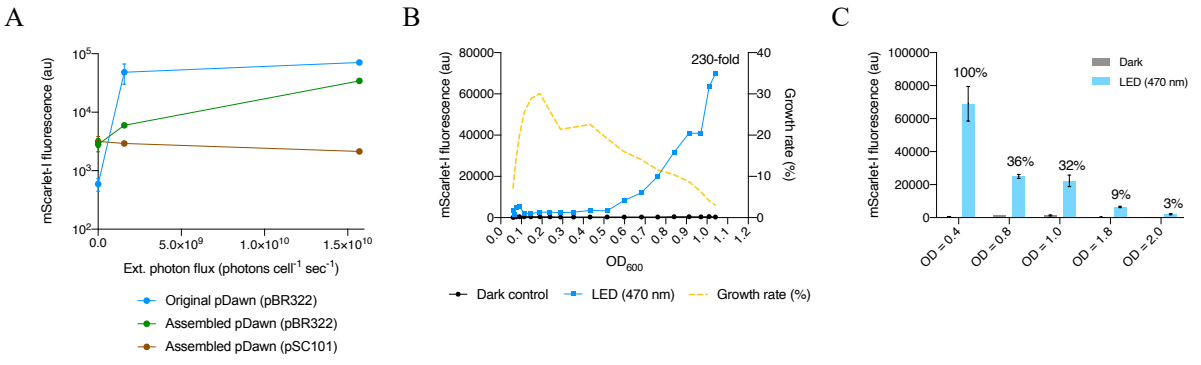


Figure 13 | pDawn becomes temporarily responsive to blue light as cultures transition into stationary phase.
A Fluorescence levels produced by the original pDawn circuit (blue), and two codon-optimised versions expressed from identical (green) or lower copy-number vector backbones (brown) after a 20-hour induction protocol using constant near-saturating or saturating blue light, or darkness. Individual values represent the average ± SD of two biological replicates analysed by flow cytometry using mean population values. **B** Fluorescence levels produced by pDawn in the dark (black) and under constant, saturating blue light (blue) during bacterial growth. Individual values were collected at 15-minute intervals and correspond to the averaged fluorescence of 10,000 cells analysed by flow cytometry. Growth rate values (yellow) were calculated from OD measurements collected in a microplate reader every 15 minutes to inform on the growth stage of cultures and represent the average of 24 technical replicates. **C** Fluorescence levels produced by pDawn in cultures grown in the dark (grey) or in the dark until a specific culture density and then induced with a 2-hour saturating pulse of blue light (blue). Bars correspond to the average ± SD of three biological replicates analysed by flow cytometry using mean population values.

To verify that these constituted the optimal pulse delivery conditions for a bioluminescence activation, fluorescence levels were initially screened in the dark and under constant saturating blue light from late-exponential phase until stationary phase while tracking culture density. Flow cytometry data revealed that pDawn remained relatively irresponsive to blue light until cultures were halfway through the transition into stationary phase (**Figure 13B**). During the first 7.5 hours of growth, fluorescence differences between dark and illuminated cultures remained small and constant (~ 8-fold), suggesting that slight P_R derepression had occurred earlier in the experiment but was practically negligible at the moment. Then, signal significantly accumulated in response to light for around 2 hours, as cells progressively ceased dividing (OD ~ 0.5-1.0). At the end of the experiment, cultures exposed to blue light showed a 230-fold induction from the dark state, which had remained fully repressed independently of time and growth. Moreover, fluorescence levels by the end of this experiment, 9.5 hours post-inoculation, were identical to the previous 20-hour induction, suggesting that most signal in response to blue light is produced during a 2-hour window right before cells reach stationary phase. It is unclear whether signal production halted due to fluorescence saturation inside cells or another unaccounted inhibition of pDawn during stationary phase. In any case, it took less than 30 minutes after induced cultures reached OD ~ 0.4 for fluorescence levels to start increasing, which is approximately the maturation half-time of mScarlet-I (25.7 ± 1.5 min)[122]. Therefore, it was concluded that pDawn is inhibited by bacterial growth and its bioluminescence activation cannot occur before this culture density.

In the previous experiment, pDawn appeared to be inhibited by some unidentified factors that disappeared as cells exited exponential phase. Although bacterial growth is not necessary for bioluminescence selection, it was hypothesised that low resource availability during stationary phase might also affect pDawn activity by either altering the expression levels of its circuit components or reducing its signal production capacity. To validate pDawn utility during this metabolically-restrictive growth phase, cultures were grown in the dark and interrogated with a 2-hour long pulse of saturating blue light at distinct densities: from the moment the circuit becomes responsive to light (OD ~ 0.4) until complete culture saturation (OD ~ 2.0). As cultures progressed into the stationary phase, pDawn signal production capacity gradually decreased in response to the same light-stimulus (**Figure 13C**). After a single cell division (from OD = 0.4 to 0.8), which took slightly over 1 hour, pDawn had lost 64% of its signal production capacity; and it further decreased by 97% as cultures totally saturated. Most likely, this results from the systemic downregulation of protein synthesis that occurs in

stationary phase to survive starvation[78]. Nevertheless, this circuit appears to be strongly influenced by the nutritional state of the host and it cannot be fully discarded that other stationary-phase genetic adaptations are interfering with pDawn activity. In conclusion, these experiments demonstrated that pDawn becomes temporarily responsive to blue light as cells transition into stationary phase, but its utility progressively decreases as cells adapt to starvation.

Failure to induce pDawn could occur for any of the following reasons or a combination of them: (1) an inability to deactivate FixJ-mediated production of cI, (2) persistence of already synthesised cI repressor, and (3) unspecific inhibition of P_R promoter by endogenous host factors. Since YF1 and FixJ are co-expressed from the same promoter and their expression can be deduced from FixJ-mediated repression of P_R, the first hypothesis would only be valid if high endogenous HKs activity during exponential phase outcompetes YF1 light-induced phosphatase activity. This would also imply that pDusk is equally difficult to repress during this growth period and constitutively produces fluorescence signal until becoming responsive to blue light. The second scenario is highly unlikely because pDawn can elicit maximal light responses in cultures that have grown in the dark until OD ~ 0.4, which means that cI production was constant until light delivery. The third explanation seems the most reasonable one since λ bacteriophage components have evolved to regulate the lysis-lysogeny cycle in response to the nutritional state of infected bacteria. This confers the virus a reproductive advantage by allowing it to replicate and lyse the cell when appropriate or integrate in the genome and wait for the right conditions[128]. In fact, multiple host components have been demonstrated to influence P_R activity, both as activators and repressors, including the secondary messenger that orchestrates the transition into stationary phase[78,129].

3.4.2.2 pDusk inducibility progressively decreases beyond exponential phase

Despite the improved circuit dynamics of pDawn, its characterisation during bacterial growth revealed that its light-responsiveness is restricted to culture conditions with low nutritional availability. This limitation seemed to arise from host-circuit interactions that repressed P_R transcriptional activity independently of light conditions. To validate this hypothesis and potentially enable the utility of this circuit before stationary phase, a light-inducible variant of pDusk, referred herein as inverted pDusk, was created for further characterisation. Inverted pDusk was engineered by introducing the point-mutation H22P in YF1 coding sequence, which had been reported as the inverted YF1 variant with highest

dynamic range (~ 14-fold)[126,127]. The original pDusk with mScarlet-I reporter gene was tested alongside inverted pDusk to assess the impact of this mutation on the signal transduction capacity of YF1. After a 20-hour induction protocol with either saturating blue light or darkness, inverted pDusk only differed from its light-repressed counterpart in slightly lower fluorescence levels in the active state, eliciting a 14-fold dynamic range as previously reported (**Figure 14A**). Both pDusk versions expressed significantly lower overall fluorescence levels than pDawn under the same flow cytometry settings, indicating that FixJ-mediated transcription in *E. coli* in the current circuit design is relatively weak. When screening fluorescence production across growth under constant light conditions, both pDusk versions displayed a tightly-regulated inactive state and a progressive accumulation of signal over time in response to either darkness (pDusk) or saturating blue light (inverted pDusk) (**Figure 14B**). As previously observed, inverted pDusk showed a slightly lower signal production capacity compared to pDusk, which indicates that FixJ phosphorylation levels are lower when YF1 is a

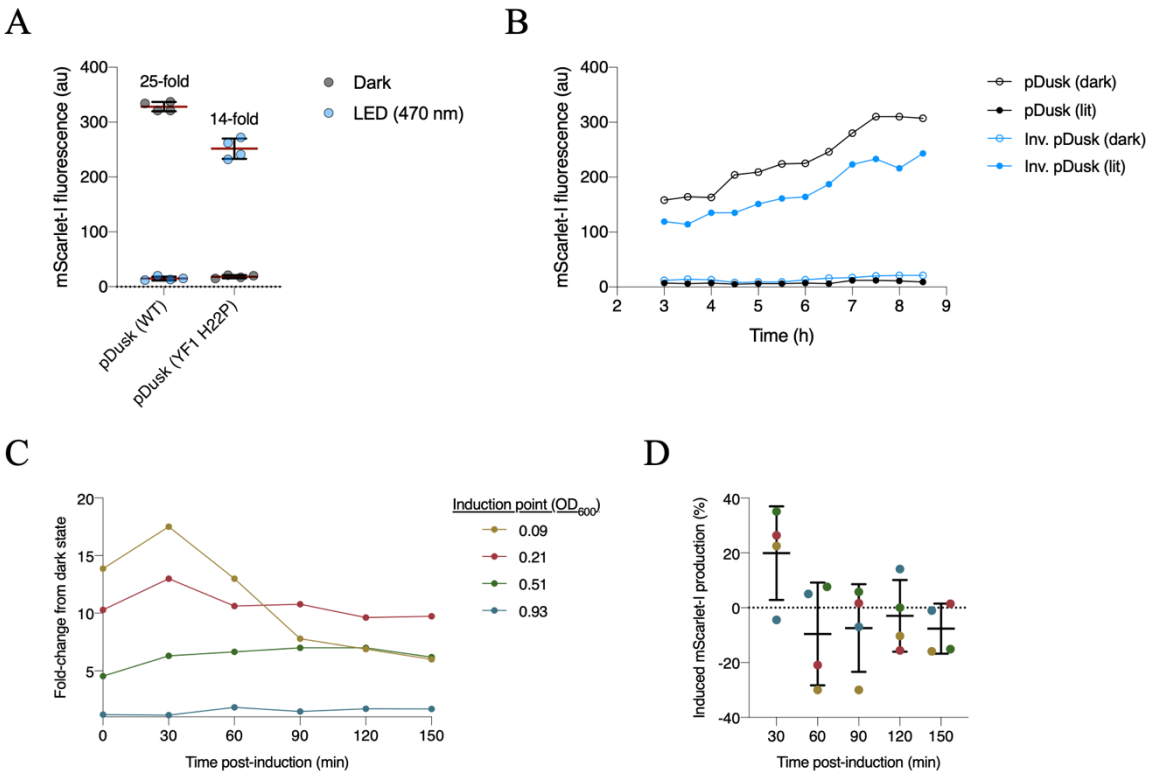


Figure 14 | pDusk induction capacity progressively decreases beyond exponential phase.
A Fluorescence levels produced by pDusk circuits in the dark (black) and under saturating blue light (blue) after a 20-hour induction protocol. Bars represent the average ± SD of four biological replicates analysed by flow cytometry using mean population values. **B** Fluorescence levels produced over time by pDusk (black) and inverted pDusk (blue) in the dark (empty) or in response to uninterrupted saturating blue light (filled). Individual values were collected every 30 minutes and correspond to the averaged fluorescence of 10,000 cells analysed by flow cytometry. **C** Fold-induction responses generated by inverted pDusk after a 1-hour pulse of saturating blue light at progressively denser culture conditions. Fold-change measurements correspond to the light/dark ratio of the averaged fluorescence of 10,000 cells in each condition analysed by flow cytometry. **D** Induced signal production expressed as a percentage change over 30 minutes corrected for signal production in the dark state. Bars represent average ± SD of four biological replicates.

phosphatase in the ground state. This suggests that signal transduction in $YF1^{H22P}$ might be less efficient and result in higher basal phosphatase levels upon induction. More importantly, this experiment confirms that the inhibition of pDawn during exponential phase is caused by the inactivity of the λ bacteriophage promoter P_R, as all other circuit components operate as expected in response to blue light. However, signal accumulation under saturating conditions was significantly slow for both pDusk circuits compared to pDawn: it took 5-6 hours depending on the version to achieve maximal differentiation from the inactive state (40-fold in pDusk; 18-fold in inverted pDusk), compared to the 230-fold increase detected in pDawn within a 2-hour window.

The slow signal production of inverted pDusk under constant, saturating light conditions was concerning for its utilisation in the bioluminescence-mediated fitness acquisition. As observed in pDawn, signal production also seemed to approximate saturation by the time cultures reached stationary phase. To assess the induction kinetics of inverted pDusk and its signal production capacity at distinct growth stages in response to a temporary light stimulus, the circuit was subjected to a pulse interrogation assay similar to the previously designed for CcaS-CcaR (*Section 3.3.2.2*). In brief, dark cultures expressing inverted pDusk were induced with identical 1-hour pulses of saturating blue light at distinct bacterial growth stages: exponential phase (OD_{600} = 0.09), exponential phase exit (OD_{600} = 0.21), transition into stationary phase (OD_{600} = 0.5), and early stationary phase (OD_{600} = 0.93). Cultures were then returned to dark conditions and signal production was simultaneously monitored in induced and uninduced controls. Fluorescence analysis showed that inverted pDusk was able to generate a maximal fold-induction in response to a 1-hour pulse of saturating blue light when illuminated during exponential phase (**Figure 14C**, yellow). However, its induction capacity progressively decreased as cultures saturated until becoming totally irresponsive to blue light by early stationary phase (**Figure 14C**, red, green, and blue). Most fluorescence production in response to light ceased within 30-60 minutes after returning to dark conditions, which indicates that phosphorelay deactivation is relatively fast (**Figure 14D**). Although the fast division rate of cultures in exponential phase diluted the fluorescence differences between conditions, cultures that had exited exponential phase confirmed that these induced differences persisted for at least 2.5 hours post-induction. Therefore, inverted pDusk transduction capacity is optimal during exponential growth, when it can quickly generate near-maximal fold-inductions in gene expression in response to a temporary blue light signal.

In conclusion, YF1-FixJ appears to be tightly-regulated in the inactive state during bacterial growth, although its overall transcriptional activity is relatively low. Previous analysis of circuit components in a similar two-component system revealed that transcriptional capacity was primarily determined by the gene expression levels of the transcription regulator[58]. In pDusk/pDawn, FixJ is expressed as the second gene of a bicistronic unit that has been kept intact from the original genetic structure. Perhaps this circuits would benefit from decoupling the expression of YF1 and FixJ, and simultaneously optimising their respective expression levels to achieve higher inducibility. Inverted YF1 was able to generate lower FixJ phosphorylated levels than the original photoreceptor, suggesting that its signal transduction capacity might be less efficient. Still, both pDusk variants were able to generate fold-induction responses that clearly differentiated from their inactive states during exponential phase, confirming that pDawn inhibition in fast dividing cultures results from the inactivity of the lysogenic λ bacteriophage promoter. Although having a fold-dynamic range that was around one order of magnitude smaller than pDawn, inverted pDusk was able to produce fast gene expression responses when induced with temporary blue light stimuli. Therefore, it might constitute a useful alternative to pDawn for bioluminescence transduction in culture conditions that support bacterial growth.

3.5 Conclusions

In the first stage of this project, three light-controlled transcriptional circuits were shortlisted from the literature for validation in the laboratory. They were selected for their apparent suitability to transduce light signals in a dynamic environment and their reliability based on published characterisation data, optimisations, and track record of applications, including the light-control of bacterial growth. Apart from one circuit that could not be implemented in the laboratory, the other two showed to be strongly influenced by bacterial growth due to the dilution of circuit components, changes in resource availability, and host-circuit interactions. For this reason, their signalling dynamics were studied in depth across distinct bacterial growth stages to identify optimal conditions for bioluminescence transduction and to aid the engineering of a fitness selection system.

The first circuit tested was a set of four photoactivatable split T7RNAP sensors that drive gene expression from a highly-orthogonal promoter. They were all variations of a similar design in which light-sensitive VVD domains fused to T7RNAP halves block the reconstitution of transcriptional activity in the dark and allow it after becoming displaced due to a conformational change induced by blue light. Despite best efforts, it was not possible to assemble functional copies of these circuits, which appeared to have an impaired dynamic range due to poor blue light-mediated reconstitution of T7RNAP activity. This failure was unrelated to sensor design, sensor/reporter gene copy-number ratio, growth and light-conditions, chromophore availability, or VVD thermostability. Most likely, the lack of reproducibility resulted from either genetic and/or host context-dependent effects, because it was not possible to obtain the original materials and they had to be reconstructed in the laboratory. As noticed with a codon-optimised version of pDawn, non-coding modifications in a circuit DNA sequence might still affect its performance by destabilising the expression levels of its components. Although certain discrepancies exist in the literature about the utilisation of these sensors, they are expected to work as reported, as they were independently engineered by two distinct research groups and recently implemented for the first time to control a synthetic bacterial behaviour using blue light[130]. In fact, this study co-regulated the growth of two different bacterial populations in the presence of a bacteriostatic antibiotic and light-controlled antibiotic resistance, supporting the utility of these sensors for this project and a potential conflict of interest at the time the authors were contacted for the materials. In any case, the available sensors were considered inappropriate for this project and discarded.

The second circuit tested was CcaS-CcaR, a cyanobacterial two-component system that regulates gene expression in response to green light. The characterisation of its signalling dynamics revealed that the photoreceptor CcaS is able to spontaneously activate in the dark and accumulate in low-dividing conditions due to a relatively slow ground-state reversion. Bacterial growth dilutes the proportion of active CcaS molecules, most likely through both cell division and *de novo* production of photoreceptors in the ground state, which buffers the basal transcription activation of the circuit. As a result, the optimal conditions for CcaS-CcaR signalling occur at near-maximal division rates with a 1-hour window before and after the end of exponential phase, which was measured to occur around OD ~ 0.23. This also indicates that probably antibiotics targeting protein synthesis might prove especially useful for select bioluminescent phenotypes using CcaS-CcaR, as they would inactivate the basal activity of the circuit in unfit mutants. Moreover, the sequence of CcaS was studied to aid the design of subsequent bioluminescence delivery strategies. On one side, it was found that CcaS cannot be released from the membrane by truncating its N-terminal transmembrane domain as it becomes highly unstable. This could have facilitated distant bioluminescence activations in the cytoplasm and N-terminal fusions with luciferases for close-contact delivery of photons to the chromophore. On the other side, internal truncation in miniCcaS#10 increased the ground stability of the photoreceptor, which expanded both its fold-induction response and the range of light intensities that differentially activate it. Therefore, this CcaS version might allow improvement of bioluminescent proteins over a larger brightness span, and higher selection tunability when modulating luciferin concentration in the medium. For these reasons, CcaS and miniCcaS#10 were selected for the next stage of the project.

The last circuit assessed in this project was pDusk/pDawn, which is based on the synthetic blue light-sensitive two-component system YF1-FixJ. Since YF1-FixJ is a light-repressed phosphorelay, two variants with inverted signalling responses were considered for engineering bioluminescence-mediated fitness acquisition: pDawn, which uses a signal-inversion cassette based on λ bacteriophage components; and inverted pDusk, which expresses a YF1 mutant with light-inducible kinase activity instead of phosphatase activity. The characterisation of pDawn showed that its transcriptional activity is suppressed during exponential phase independently of light conditions, probably due to host-circuit interactions with the λ promoter of the inversion cassette. The factors inhibiting pDawn seem to disappear as cells transition into stationary phase, around OD ~ 0.4, and it becomes temporarily responsive to blue light before its signal production capacity is compromised due to decreasing

resource availability in saturated cultures. Still, the circuit was able to induce a 230-fold change in gene expression within a 2-hour period that should be sufficient to select bioluminescent phenotypes. Instead, inverted pDusk was exempt of exponential phase inhibition, confirming the origin of the problem, although it had considerably lower overall transcription capacity. It is possible that this occurs due to low expression levels of FixJ in the current circuit design or simply lower affinity with the host transcriptional machinery. Importantly, both circuits were tightly-regulated in the dark regardless of their output promoters and growth conditions, which should prevent unspecific fitness acquisition in non-bioluminescent mutants. Therefore, these two circuit variants were also selected for further engineering.

In conclusion, two light-sensitive transcriptional pathways were successfully validated to transduce green and blue bioluminescent signals into fitness gene expression. Their optimal conditions for light transduction during bacterial growth were precisely characterised to maximise their activation with bioluminescence in the next stage of the project. Importantly, CcaS is a membrane-bound photoreceptor while YF1 is soluble, which increased the flexibility of designs available for bioluminescence delivery to their chromophores. Moreover, these photoreceptors sense light wavelengths in the region of the visible spectrum where the majority of bioluminescent proteins in nature emit[3,23]. This facilitated the identification of spectrally-compatible luciferases to engineer bioluminescence transduction strategies in *Chapter 4* and should expand the utility of the system once it is developed.

4. Engineering bioluminescence transduction in bacterial cells

A light-sensing circuit was repurposed to transduce light produced exclusively inside bacterial cells into gene expression levels. After testing distinct bioluminescence delivery strategies, it was found that close molecular distance to the photoreceptor sensory domain was necessary to transduce enzymatic light production in bacteria. Hence, the cyanobacterial light-sensor CcaS was engineered accordingly to create modular platform for bioluminescence transduction with a wide dynamic range, external tunability, and extended spectral compatibility.

4.1 Introduction

4.1.1 Bioluminescence activation of light-sensing proteins

In the previous chapter, two light-sensitive transcriptional circuits were identified that could reliably transduce light signals into gene expression levels. The next step towards developing the technology proposed in this project involved repurposing the photoreceptors controlling these circuits to sense bioluminescence emissions produced exclusively inside the cell rather than external light.

As mentioned earlier, bioluminescence had been used in multiple research contexts to activate a range of light-absorbing proteins, including rhodopsins[131,132], fluorescent proteins[64,133], LOV domains[132,134,135], plant cryptochromes[134], and sensor of blue-light using FAD (BLUF) domains[136], suggesting that it is a generalisable approach that can translate into biologically relevant functions. In fact, all reported bioluminescence activations in the literature had been comparable to external light sources despite producing lower photonic outputs, which demonstrates that intracellular light delivery is not only possible but can constitute a more efficient strategy to reach the chromophores of light-sensing proteins. Moreover, some studies had shown that it was possible to tune induction levels or timing of effects in a dose-dependent manner with existing bioluminescent materials[133–137]. This is especially important for this project because it should contribute to adjusting the tool's selection parameters to distinct experimental needs, which are likely to change for each new bioluminescent protein being developed and along the protein engineering process itself.

For this project to be successful, bioluminescence transduction had to be restricted within the cell to avoid non-specific selection of mutants due to transactivation of neighbouring bacteria. In most cases, bioluminescence activations had been contained intracellularly by physically-coupling the bioluminescent protein to the light-sensitive protein with a peptide linker, ensuring a close-contact delivery of light. Alternatively, distant activations had also been achieved by expressing the two components separately to control proteins inside the cell[134], on the cell surface[135], and on the surface of other cells[21,132]. Both strategies appear to be highly suitable for compartmentalising activations at the single-cell level due to the light absorption and dispersion effects of cell membranes. For instance, one of these studies showed that it was impossible to activate a transmembrane sensor that contained the chromophore on the cytoplasmic side of the membrane when a bright bioluminescent protein was fused at the extracellular terminus[132]. Additionally, the study focused on distant intracellular activations

showed that while unnecessary for the soluble light-sensors tested, a mitochondrial outer-membrane sensor failed to get induced unless the bioluminescent protein was co-localised at the same subcellular structure. Therefore, it is unlikely that bioluminescence will cross two cell membranes at sufficient levels to transactivate other photoreceptors, as long as these are not exposed on the cell surface.

Despite the groundwork established by these proof-of-concept studies, some aspects relevant to this project remain still unanswered. First of all, there is no available data on the persistence of bioluminescence-induced protein expression in fast-dividing cells like *E. coli*. In fact, only one study so far has attempted to control gene expression directly with bioluminescence, reporting the systematic activation of four distinct light-induced transcriptional circuits in mammalian cells[134]. Second, this study did not test whether the transcriptional activation of these systems could produce relevant biological responses using genes other than bioluminescent reporter proteins. Even the highest fold-change reported (40-fold) is relatively uninformative in the absence of protein activity context in each separate uninduced and induced state. Finally, bacterial production of bioluminescence might change during growth due to varying levels of bioluminescent enzymes and luciferin uptake capacity, which might antagonise with optimal circuit activation conditions. The only example of a bioluminescence activation in bacteria engineered to date used an autonomous bioluminescence operon to control a light-inducible protein and worked in carefully designed non-dividing conditions[135], thus avoiding the experimental constrains that characterise this project. It is a mercury biosensor that expressed the bacterial bioluminescence operon in response to the heavy metal to trigger bacterial aggregation through photoactivatable cell adhesion molecules. As a result, both bioluminescence and bacterial sedimentation can be used as readouts for mercury detection, which synergistically interact to provide enhanced sensitivity by concentrating light signal and facilitating transactivation in closer proximity.

Hence, the bioluminescence transduction designs presented in this chapter were guided by these preliminary studies with additional considerations on the interplay between bacterial growth, gene expression dynamics, and bioluminescence production. Subsequent fitness acquisition experiments contextualising the physiological relevance of the responses reported herein can be found in *Chapter 5*.

4.1.2 Selection criteria for suitable bioluminescent components

The bioluminescent proteins tested in this chapter to engineer bioluminescence transduction were mainly shortlisted according to their spectral compatibility with the light-sensitive transcriptional circuits available in the laboratory. The photoreceptors regulating CcaS-CcaR and pDusk/pDawn have green and blue light-absorbing chromophores respectively with relatively broad absorption peaks. The PCB chromophore of CcaS has an absorption maximum at 536 nm that extends from 450 to 600 nm and should be more efficiently activated by green-emitting bioluminescent reactions, although blue and yellow emissions might also prove effective[103] (**Figure 15A**). Instead, the FMN cofactor of YF1 has an absorption maximum at 449 nm with a shoulder peak at 475 nm that terminates at 500 nm, and it is only suited to sense blue bioluminescence[138] (**Figure 15B**). Accordingly, an array of spectrally-compatible bioluminescent proteins was selected from the literature to activate these photoreceptors, considering previous evidence of heterologous expression in prokaryotic hosts, availability of luciferin compounds, protein evolution potential, and perceived research utility (**Figure 15C**).

The *Renilla reniformis* luciferase (RLuc), from the eponymous North-American soft coral, was an interesting candidate because it had been extensively characterised, but its evolution potential still seemed relatively unexplored. This small protein (36 kDa) catalyses the oxidation of coelenterazine to produce blue light in an ATP-independent manner, and it is competitively inhibited by the reaction by-product[139,140]. It has been cloned and subjected to exhaustive sequence analysis in *E. coli* to yield large datasets of mutants that have been used to construct improved variants with novel light-emitting properties. Initial engineering efforts

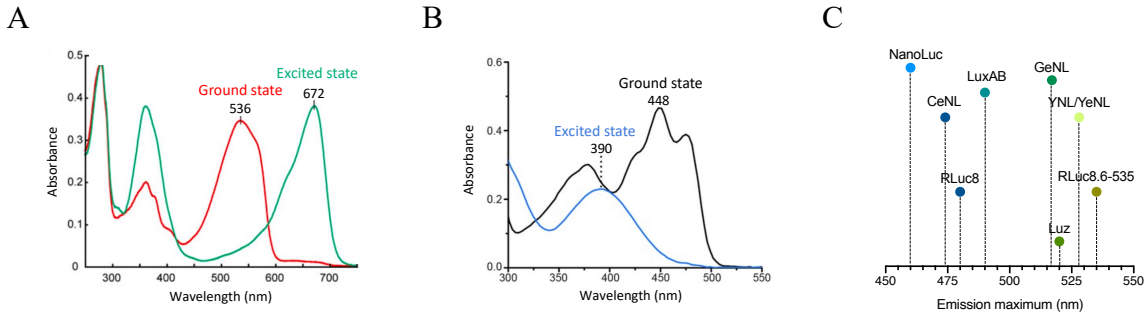


Figure 15 | Spectral compatibility between light-sensing and light-emitting components.
A Absorption spectrum of the chromophore-binding domain of CcaS photoreceptor in the target ground state (red) and its resulting excited state upon green-light illumination (green); reproduced from Hirose Y *et al,* 2010. **B** Absorption spectra of a typical LOV domain in the target ground state (dark) and its resulting excited state upon blue-light illumination (blue); reproduced from Heintz U & Schlichting I, 2016. **C** Emission maxima catalysed by bioluminescent proteins shortlisted to activate CcaS and YF1 photoreceptors, resulting from their native or co-evolved luciferin substrates.

focused on stabilising the protein for technological applications, resulting in an 8-mutation variant named RLuc8 with 200-fold enhanced stability and 4-fold higher light output compared to the native enzyme[37]. This stabilised variant allowed resolving the protein crystal structure in its luciferin-bound form, which was essential for the identification of key residues influencing the spectral properties of the protein[141]. Subsequent directed evolution studies used this information to develop green-emitting variants with varying kinetics for live-imaging and reporter assay applications[40,41]. The most promising of these variants, RLuc8.6-535, retained the intracellular stability of RLuc8 with a 49 nm bathochromic shift in its spectral emission and slightly higher light production (~ 40%). Except for some later studies that either combined[142] or identified additional point-mutations with increased thermostability[143–145], no further significant improvements have been achieved beyond RLuc8.6-535. This RLuc variant is fifteen mutations away from the *wild-type* sequence and only 6-times brighter, suggesting that its evolution potential is still largely unexplored. Moreover, a recent study reconstructed the ancestral sequence of this protein[146], offering a highly evolvable template, and showed that its enzymatic activity could be recovered with just two point-mutations while avoiding the naturally-selected structural features that allow by-product inhibition in modern RLuc[147]. For these reasons, both the blue RLuc8 and the green RLuc8.6-535 were selected to establish a bioluminescence transduction system with YF1 and CcaS photoreceptors.

The synthetic luciferase NanoLuc, derived from the deep-sea shrimp *Oplophorus gracilirostris* luciferase (OLuc), was considered due to its robust properties for the development of novel bioluminescent technologies. This protein was evolved in the laboratory from the 19-kDa catalytic subunit of OLuc alongside a battery of synthetic luciferin analogues both to optimise its monomeric expression and adapt the catalytic site to a substrate with improved light-emitting features [38]. The resulting bioluminescent reaction was 2.5 million-fold brighter than the template protein with its native luciferin and resistant to catalytic autoinhibition, providing a bright and sustained blue light-emission. The sixteen mutations that conform NanoLuc mostly improved its stability, making it exceptionally tolerant to fusions and fragmentation for the construction of bioluminescence-based biosensors[45,148,149], imaging probes[65,150,151], and immunological assays[152]. Moreover, most bioluminescence activations engineered to this date have used NanoLuc as a light-emitter[132–134,136]. Therefore, even if the evolution potential of NanoLuc is apparently lower than in other luciferases, its compatibility with this selection tool could still assist the creation of other technologies based on this protein.

The two autonomous bioluminescent systems known to date were also selected for their low-cost sustained signal production and barely explored engineering potential. The bacterial system is a fairly conserved multicistronic operon that uses long-chain aliphatic aldehydes and reduced FMN to generate cyan light. It is constituted by a heterodimeric luciferase, *luxAB*, three metabolic genes involved in fatty aldehyde production, *luxCDE*, and a complementary FMN reductase gene that recycles the redox cofactor to increase light production efficiency[153]. The fungal system is formed by a hydrophobic luciferase named *Luz* that uses 3-hydroxyhispin as a substrate to produce green bioluminescence. The luciferin can be produced from caffeic acid in two metabolic steps catalysed by the hispidin synthase and hispidin-3-hydroxylase genes; and its oxidated form can be reconverted into caffeic acid by a fourth enzyme called caffeoyl pyruvate hydrolase[28]. Both systems hold a great potential as reporter technologies, especially for utilisation in whole organisms in which non-toxic and efficient luciferin delivery can be problematic[154,155]. However, they have been minimally engineered so far: the brightest *lux* operon version evolved to date contains around 3 mutations per gene[61], while only the *wild-type* sequence is available for the fungal operon[28]. Importantly, all proteins from the bacterial operon have been structurally resolved, which might aid future engineering attempts[34,156,157]. Hence, these two pathways could potentially sustain long photoreceptor activations, as previously achieved with the bacterial operon[135], and be evolved using cheaper and simpler selection protocols.

Finally, four additional bioluminescent constructs termed Nano-Lanterns were chosen for their bright and variegated colour emissions. These probes consist of luciferases fused to fluorescent proteins, in which the energetic state generated by the bioluminescent reaction is transferred to the chromophore of the fluorescent protein and released as a photon of a different wavelength[64]. This process occurs via a non-radiative phenomenon known as bioluminescence resonance energy transfer (BRET) that is highly sensitive to the distance between donor and acceptor molecules. As a result, the chemiluminescent properties of luciferases are coupled to the higher quantum yield and spectral diversity of fluorescent properties, resulting in more efficient and spectrally-tunable light emissions. The first generation of Nano-Lanterns was engineered using RLuc8 variants[158] and was shortly followed by a second array of NanoLuc constructs named "enhanced Nano-Lanterns"[65]. Importantly, both Nano-Lantern types had been used to engineer bioluminescence transduction distantly[134] and in a close-contact setup[159]. The selected probes were Cyan enhanced Nano-Lantern (CeNL) based on mTurquoise2, Green

enhanced Nano-Lantern (GeNL) based on mNeonGreen, and Yellow Nano-Lantern (YNL) and Yellow enhanced Nano-Lantern (YeNL) both based on Venus.

4.2 Results and discussion

4.2.1 Distant delivery of bioluminescence

In *Chapter 3*, it was demonstrated that bacterial growth influences the signalling dynamics of CcaS-CcaR and pDusk/pDawn in a circuit-specific manner. As a result, these circuits respond differently to identical light stimuli depending on the division rate and the growth stage of the culture. Although this should minimally constrain their utilisation under sustained bioluminescent production, most light-emitting reactions available are restricted in time due to their dependency on luciferin supply. Anticipating this limitation, their optimal activation conditions were identified using external light pulses to provide a reliable framework for engineering bioluminescence transduction.

The intracellular conditions to activate these light-sensing transcriptional circuits with bioluminescence were unknown and unpredictable, both in terms of photonic and spatial requirements. In terms of light production, bioluminescent reactions provide different light outputs depending on the enzymatic properties of each luciferase, and the light-emitting properties and cell permeability of each luciferin. For this reason, several spectrally-compatible bioluminescent reactions were shortlisted to offer a wide range of available bioluminescent pulses for activations (**Figure 15C**). In regard to spatial requirements, light-sensors had been previously activated using both distant and close-contact delivery of bioluminescence.

In most cases, bringing the light-emitter in proximity to the light-sensor should increase the light delivery to the chromophore. However, fusing both components together might disrupt the activity of any of the two proteins, making it necessary to independently validate their functionality in each fusion protein design. Instead, distant delivery of bioluminescence is usually achieved by simply co-expressing both genes in the same strain, which should facilitate the screening of potentially activating bioluminescent reactions. Although probably less efficient, this approach would both accelerate the engineering process and motivate the wide-spread utilisation of the selection tool by other researchers. Therefore, a co-expression approach was initially chosen for its higher simplicity and modularity prospects.

4.2.1.1 Co-expression of bioluminescence is insufficient to control CcaS or YF1 activity

To co-express each bioluminescent protein with its target light-sensitive circuit, a constitutive transcription unit was engineered that could systematically express any of these genes. It was observed that promoter strength influenced considerably the metabolic burden of

bioluminescent genes, causing toxicity at high transcription levels despite having selected relatively weak translation conditions. Four sequences from the Anderson Promoter Collection were used to tune their expression: the consensus promoter sequence and three mutant variants with approximately 25, 50, and 75% attenuated activity (*Materials and Methods*). Only the weakest promoter allowed non-toxic expression of the bioluminescent array. This transcription unit was used to co-express each of the five green bioluminescent genes with CcaS-CcaR (nnLuz, RLuc8.6-535, YNL, GeNL, and YeNL) and three of the blue bioluminescent genes with pDawn (RLuc8, NanoLuc, and CeNL). The bacterial lux operon was reserved to engineer an autonomous bioluminescence transduction system (*Section 4.2.1.2)*.

The bioluminescent emissions of these strains were optimised by choosing appropriate luciferin substrates, vehicle controls, and luciferin doses:

Luciferins were selected to diversify the features of the bioluminescent pulses. The native substrate of nnLuz, 3-hydroxyhispin, was the only compound available for this luciferase. An esterified coelenterazine formulation named ViviRen™ was chosen for RLuc-based constructs to produce sustained bioluminescent pulses. This conjugation protects the luciferin from oxidation in the medium and allows its progressive release as it is transported into cells and de-esterified by endogenous enzymes. Instead, NanoGlo® (Promega) was selected for NanoLuc-based constructs to maximise photon flux over signal duration, as it was the brightest non-lytic reagent available for these proteins.

Vehicle controls were designed to affect growth comparably to the luciferin-treated cultures. This was especially important for CcaS-CcaR strains to avoid differential accumulation of basal circuit activity between conditions, which could lead to misinterpretation of artifacts that did not correspond with specific circuit activation. Both 3-hydroxyhispidin and ViviRen™ were available as lyophilized solids and were resuspended in DMSO, allowing the appropriate selection of the vehicle control. However, the composition of NanoGlo® solution is undisclosed and its solvent matrix could not be reproduced in the laboratory. Consequently, NanoGlo® aliquots were heat-inactivated and validated each time to obtain a non-functional reagent that affected growth similarly to the commercial substrate.

Finally, bioluminescent pulses were characterised under optimal circuit activation conditions for multiple luciferin doses to select treatments that maximised light production without significantly compromising bacterial growth.

In CcaS-CcaR strains, the brightest bioluminescent pulse was achieved with GeNL (**Figure 16A**, blue), which was 4-times brighter than YeNL (purple), 50-times brighter than RLuc8.6-535 (red) and YNL (yellow), and 200-times brighter than nnLuz (green). To assess the capacity of these pulses to distantly activate CcaS-CcaR, strains were grown in the dark until optimal conditions and then treated with their respective luciferin or vehicle solutions. Two vehicle control groups were included in each test to contextualise bioluminescence effects on gene expression with the dark state of the circuit and its full-activation under green light illumination. None of the available bioluminescent pulses was able to distantly activate CcaS photoreceptor in sufficient levels to raise gene expression above the dark state baseline, except for GeNL (**Figure 16B-F**). This NanoLuc-based construct induced a small increase in signal production (1.5-fold) that became transiently detectable at around 3 hours post-induction (**Figure 16E**). Since GeNL had produced the brightest bioluminescent pulse, although a short one, the lack of general activation was attributed to insufficient photon production.

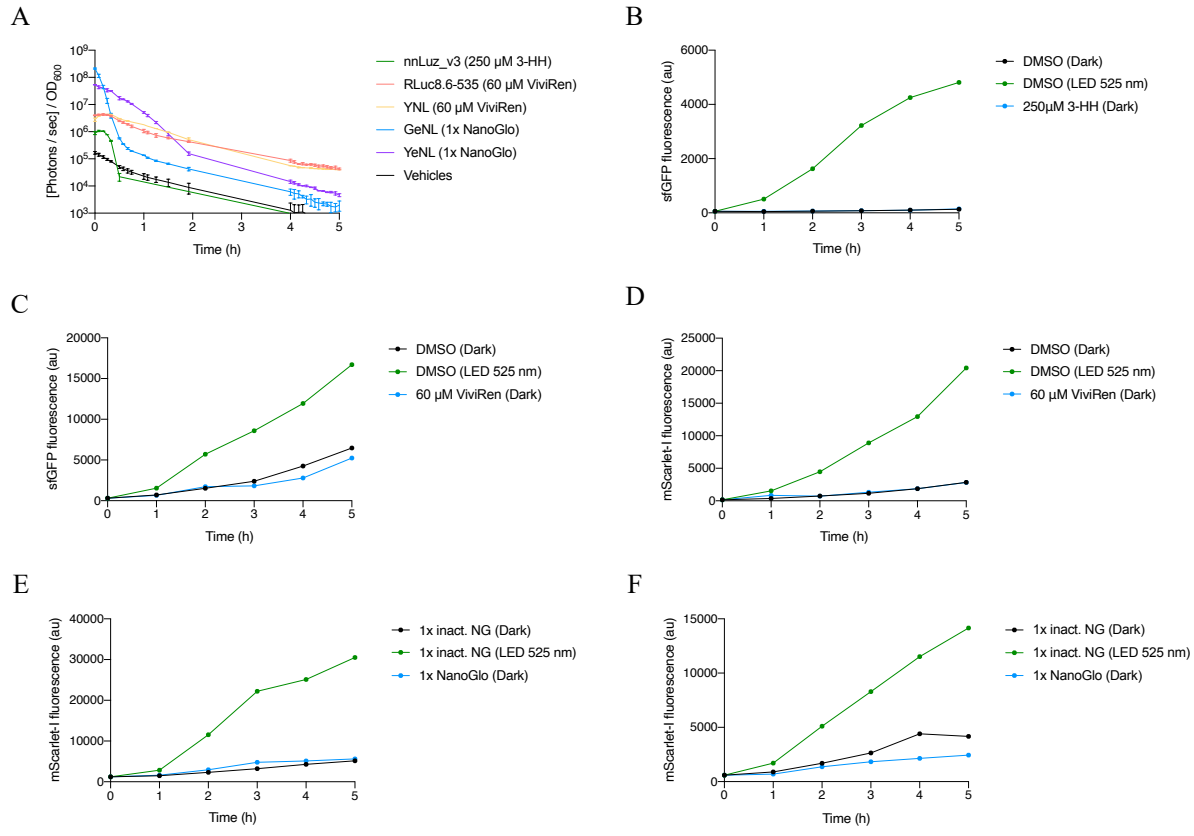


Figure 16 | Distant intracellular emission of green bioluminescent pulses failed to activate CcaS-CcaR.
A Optimised bioluminescent pulses produced by each strain co-expressing CcaS-CcaR and one of the candidate bioluminescent proteins under optimal circuit activation conditions. Individual values correspond to the mean ± SD of two biological replicates. **B-F** Temporal production of green fluorescence in CcaS-CcaR strains co-expressing either nnLuz_v3 (**B**), RLuc8.6-535 (**C**), YNL (**D**), GeNL (**E**), or YeNL (**F**) when treated with vehicle control (dark), vehicle control and saturating green light (green), or luciferin (blue). Time-points correspond to the averaged fluorescence of 10,000 cells analysed by flow cytometry.

In pDawn strains, bioluminescent pulses were characterised simultaneously with their concomitant circuit activation in a microplate reader with controlled culture conditions. Bioluminescent pulses were consistently brighter than those obtained with similar proteins in CcaS-CcaR strains (**Figure 17**, left). Most likely, these differences arise from higher luciferase expression capacity in pDawn strains, as they are not burdened by chromophore biosynthesis and the expression of associated metabolic genes, because both circuits are expressed from plasmids with similar replicability. Despite producing photon outputs that were 6-to-24-times higher than the brightest pulse achieved in CcaS-CcaR strains (GeNL), none of these strains was able to distantly activate pDawn in a significant manner (**Figure 17**, right). The highest circuit fold-inductions registered were caused by RLuc8 and the brightest of CeNL pulses, and were all less than 2-times higher than the gene expression baseline of the dark control.

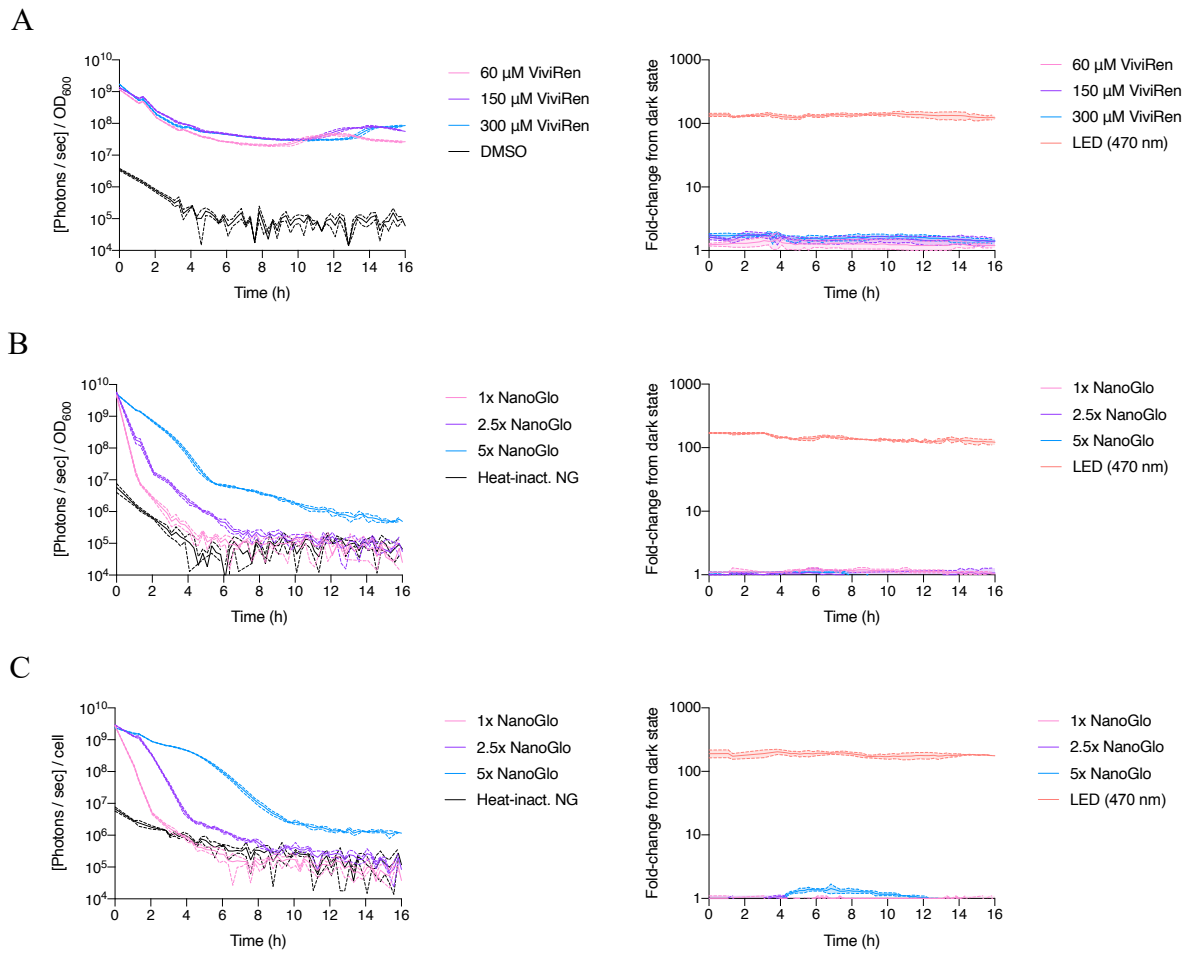


Figure 17 | Distant intracellular emission of blue bioluminescent pulses failed to derepress pDawn.
Bioluminescent pulses (left) produced by each strain co-expressing pDawn and one of the candidate bioluminescent proteins in response to distinct luciferin doses (1x, pink; 2.5x, purple; 5x, blue), or their respective vehicle controls (black) are shown with their associated fold-induction responses of pDawn output promoter (right): RLuc8 (**A**), NanoLuc (**B**), and CeNL (**C**). Culture aliquots were grown overnight under saturating blue light illumination to display maximal signal accumulation, as positive controls could not be generated inside the microplate reader. Individual values correspond to the mean ± SD of three biological replicates. Fold-change measurements correspond to the average of luciferin/vehicle or LED/vehicle ratios of those replicates.

In conclusion, intracellular delivery of distant bioluminescent pulses failed to activate both CcaS-CcaR and pDawn under optimal circuit activation conditions. It appears that photon outputs obtained with the bioluminescent expression conditions chosen were insufficient to activate the photoreceptors controlling these circuits. Although the membrane-bound location of CcaS might complicate chromophore accessibility to light pulses produced in the cytoplasm, YF1 is a soluble photoreceptor and could neither be activated with considerably brighter bioluminescent emissions. These results suggested that transactivation of neighbouring bacteria was unlikely for these two light-sensitive circuits, as it was already difficult to achieve intracellular activation using some of the brightest bioluminescent genes in the literature. In the best of cases, gene expression perturbations were transiently induced, which might indicate that activating bioluminescence levels could not be sustained for long enough to effectively control the activity of these photoreceptors. Therefore, either brighter and longer bioluminescent pulses were required or photon delivery efficiency to the photoreceptors had to be improved.

4.2.1.2 Metabolic limitations cause insufficient iLux activity to distantly activate CcaS or YF1

As shown above, engineering distant bioluminescent activations of light-sensing circuits in bacteria appeared more challenging than reported for mammalian cells[134]. The only distant bioluminescent activation in bacteria reported so far used the bacterial *lux* operon to induce the photodimerization of VVD molecules on the cell surface[135]. This operon contains the necessary genes to produce a sustained bioluminescence glow without external addition of luciferin and was bright enough to activate extracellular proteins. Recently, an improved version of this system named iLux was engineered in *E. coli* that is 7-times brighter and allows imaging single bacterial and mammalian cells with high resolution over several days[61,154]. Moreover, iLux should be spectrally-compatible with both CcaS-CcaR and pDusk/pDawn since its light emission peaks at around 490 nm. Accordingly, it was attempted to activate both circuits by co-expressing this autonomous bioluminescence system with each of them.

The genetic arrangement of iLux is similar to other bacterial *lux* operons. It contains the heterodimeric luciferase genes (*iluxAB*) that use reduced aldehydes and reduced FMN to produce light, and four metabolic enzymes that synthesise and recycle these compounds (*iluxCDE* and *frp)* in an ATP- and NADPH-dependent process (**Figure 18A**). The original iLux plasmid expresses these genes as a multicistronic unit from a pBR322 vector (~ 15-20 copies per cell), which is incompatible for co-expression with CcaS-CcaR or pDusk/pDawn. Since

shuttling the operon into any of these circuits resulted in no viable colonies, it was cloned into a different plasmid with a compatible origin of replication, pSC101 (~ 5 copies per cell).

The autonomous activity of these genes complicates the design of negative controls to isolate the effects of bioluminescence on gene expression. Despite being expressed from the IPTG-inducible promoter P_{TAC}, constitutive bioluminescence is observed independently of inducer supplementation, and it could not be used to create a dark condition (*data not shown*). In any case, different operon expression conditions would probably affect bacterial growth unequally and complicate the interpretation of results. Instead, a catalytically-inefficient luciferase mutant that would grow comparably to the functional operon would be more appropriate as a negative control. For this purpose, the point-mutation H44A was introduced in *iluxA* as it had been reported to reduce the quantum yield of the luciferase by 6 orders of magnitude[160]. This essential histidine participates in a multiresidue electrostatic network in the catalytic pocket of the luciferase that interacts with the flavin ring of FMN[157]. Alanine substitution alters this structure and promotes the formation of a dark luciferin intermediate that decays without emitting light. Therefore, this mutant should allow expressing the operon without producing bioluminescence while impacting bacterial metabolism similarly to the functional operon.

To validate the negative control strain, both the functional iLux operon and the H44A mutant were grown overnight in a microplate reader and had their growth and bioluminescent outputs recorded. As expected, these two strains grew comparably and only the functional iLux

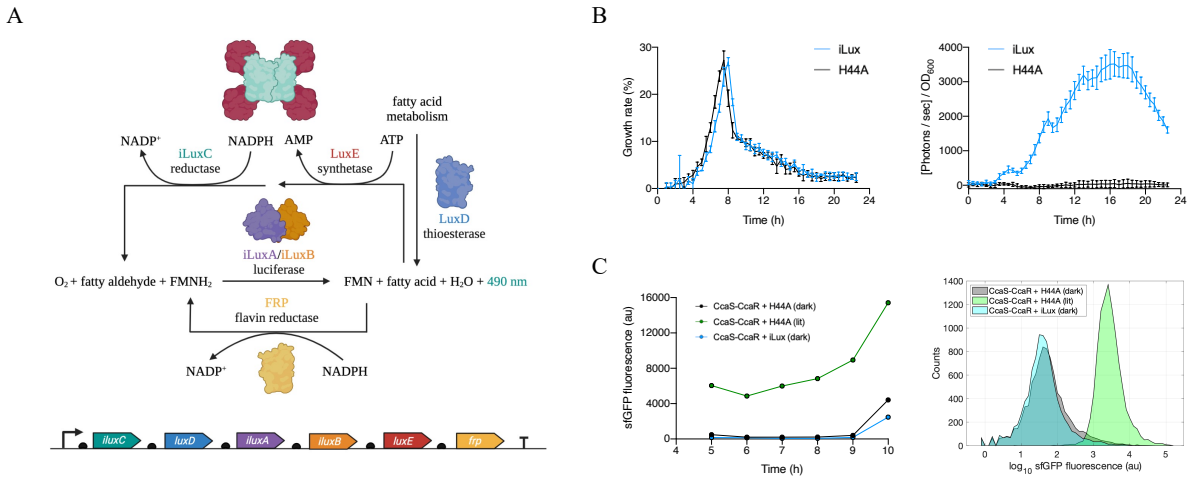


Figure 18 | Co-expression of the bacterial iLux operon with CcaS-CcaR is insufficient to induce circuit gene expression. **A** Schematic representation of the bioluminescent enzymatic pathway expressed from iLux, adapted from Tian Q *et al*, 2022. **B** Growth curves (left) and bioluminescence production over time (right) of strains expressing either the functional iLux operon (blue) or the catalytically-inefficient *iluxA* point-mutant H44A (black). Individual values correspond to the mean ± SD of six biological replicates. **C** Fluorescence levels produced by CcaS-CcaR during optimal activation conditions in response to iLux H44A co-expression in the dark (black) or under external saturating green light (green), or functional iLux co-expression (blue). Time-points correspond to the averaged fluorescence of 10,000 cells analysed by flow cytometry (right).

operon emitted detectable bioluminescence levels (**Figure 18B**). The bioluminescent strain showed a progressive production of light across growth that peaked as cultures transitioned into stationary phase, and began to decay at a similar rate several hours after. Potentially, this growth-dependent behaviour originates from iLux dependency on ATP and NADPH to recycle its co-factors, which are expected to be less available during fast-dividing conditions and starvation.

This result suggested that iLux and CcaS-CcaR might be incompatible for intracellular activation, as bright light production and fast-dividing conditions appeared to be mutually exclusive. To confirm this hypothesis, each operon was co-expressed with the circuit and the H44A mutant was used to generate dark and LED-induced cultures to contextualise a potential bioluminescent activation. Fluorescence levels remained undifferentiated from the dark state condition across the optimal window for circuit activation, confirming that iLux co-expression is insufficient to activate CcaS-CcaR (**Figure 18C**).

Light production in iLux appeared more compatible with pDawn, as bioluminescence levels peaked at the same time that this circuit becomes responsive to light, and remained at near-maximal activity for around 6 hours (**Figure 19A**). To assess whether the photon output of iLux was sufficient to distantly activate pDawn, the experimental set-up of the previous experiment was reproduced with this circuit using external blue light as a positive control. Fluorescence levels were measured after 18 hours of culture to guarantee complete activity overlap between iLux and pDawn. However, gene expression levels from pDawn output promoter remained completely unaltered by iLux co-expression (**Figure 19B**). Therefore, iLux bioluminescence in the current expression conditions is too dim to distantly activate pDawn.

In an attempt to enhance iLux activity for a potential pDawn activation, its bioluminescent activity was investigated under distinct expression conditions. Its light output was measured across growth in the original vector (pBR322, ~ 15-20 copies per cell), in the vector with origin compatibility for co-expression with pDawn (pSC101, ~ 5 copies per cell), and in the strain co-expressing iLux with pDawn. It was found that iLux activity is sensitive to both the number of operon copies in the cell and the metabolic burden caused by other heterologous genes (**Figure 19C**). The modification of the vector copy-number caused a 28-fold reduction in photon output and its co-expression with pDawn an additional 20-fold decrease, resulting in approximately 560-fold bioluminescence loss in total.

Since it had not been possible to increase iLux expression levels earlier, it was sought to recover iLux activity by relieving the metabolic burden caused by pDawn co-expression. In *E. coli*, the primary metabolic pathway providing ATP, reducing agents, and fatty acid precursors is the glycolysis, which should contribute to some extent to both aldehyde synthesis and the reduction of bioluminescent co-factors. Accordingly, a range of glucose concentrations was used to supplement iLux cultures with the intention to boost their light production. At low glucose concentrations, supplementation increased iLux activity by around 20%, although it was still insufficient to significantly derepress pDawn (**Figure 19D**). Higher glucose supplementation was detrimental for light production, leading up to a 50% reduction on bioluminescence emission, which occurred with an equivalent decrease in bacterial growth. The toxicity observed for glucose doses above 0.1% was attributed to culture acidification caused by anaerobic glycolysis in an unbuffered medium such as LB.

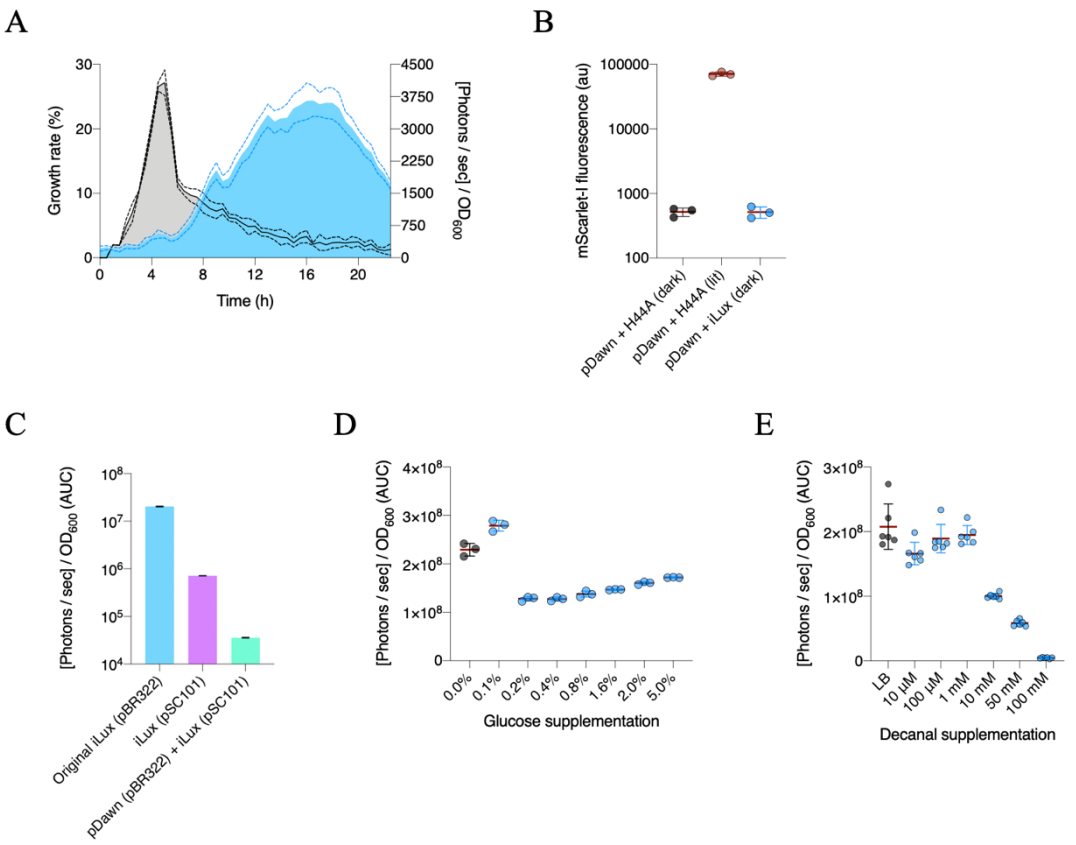


Figure 19 | iLux co-expression fails to derepress pDawn in the presence of glucose and decanal supplementation.
A Bacterial growth (grey) and iLux bioluminescence (blue) are mutually exclusive due to the metabolic dependency of the bacterial operon to produce light. Individual values correspond to the mean ± SD of six biological replicates. **B** Fluorescence levels produced by pDawn in response to iLux H44A co-expression in the dark (black) or under external saturating blue light (red), or functional iLux co-expression (blue). Bars correspond to the mean ± SD of three biological replicates obtained by averaging the fluorescence of 10,000 cells analysed by flow cytometry. **C** Bioluminescent pulses generated by the iLux operon expressed from the original vector (blue), the vector compatible for co-expression with pDawn (purple), and the strain co-expressing pDawn and iLux (green). Bars correspond to the mean ± SD of twelve biological replicates. **D-E** Bioluminescent pulses generated by the iLux operon in the presence of a range of glucose and decanal concentrations, respectively. Bars correspond to the mean ± SD of three and six biological replicates in each case. Bioluminescent emissions in **C-E** are expressed as the area under the curve of 24-hour experiments tracking photon emission and culture density at 5-minute intervals.

Traditionally, bacterial bioluminescence has been externally controlled using the aldehyde decanal in the absence of the metabolic genes *luxCDE*. As a final attempt to increase iLux activity, a titration of decanal concentrations was applied to iLux cultures to potentially decrease the metabolic demand of the bioluminescent pathway. Decanal supplementation had no noticeable effects on light production at micromolar concentrations and became increasingly toxic across the millimolar range (**Figure 19E**). Therefore, it seems that *luxD* activity is sufficient to create excess aldehyde levels without significantly burdening bacterial metabolism.

Collectively, these results indicate that ATP and NADPH availability for aldehyde and FMN recycling are the limiting factors in the iLux bioluminescent pathway. As a result, light production is significantly lower during fast-dividing conditions and starvation, in which bacteria might prioritise other more advantageous biochemical processes needed to reproduce or survive. The expression of additional heterologous genes also reduces iLux activity by presumably decreasing the metabolic capacity of the strain. Finally, it was found that light production can be boosted by increasing iLux expression levels and through glucose supplementation, but additional challenges were encountered that required further optimisation. Since previous results suggested that an improvement of several orders of magnitude in light output was probably required to distantly activate these light-sensing circuits with iLux, co-expression approaches were abandoned. Instead, it was deemed more appropriate to prioritise optimising photon delivery efficiency by co-localising bioluminescent production with the photoreceptors.

4.2.2 Close-contact delivery of bioluminescence

In the previous section, an array of spectrally-compatible bioluminescent genes was systematically co-expressed with CcaS-CcaR or pDawn to induce gene expression using intracellular light pulses. However, unlocalised light production was not sufficient to distantly activate the photoreceptors controlling these circuits, even for some of the brightest genes in the literature. Although bioluminescence expression could be further optimised, this light delivery strategy would limit considerably the utility of the selection tool, as most genes with improvement potential would not be detectable inside the cell. For this reason, it was decided to bring the bioluminescent proteins closer to the photoreceptors to increase the light delivery efficiency.

As already mentioned, two approaches had been used to engineer close delivery of bioluminescence in the past: subcellular localisation tags and physical linkage of proteins. The first option allows modular co-expression of bioluminescent genes without interfering with photoreceptor function. Nevertheless, the necessary distance for effective bioluminescent control probably varies according to the brightness of each reaction. Moreover, YF1 location was not certain to rationally target the bioluminescent proteins, and relocating photoreceptors could destabilise their light-regulation, as observed for CcaS (*Section 3.3.2.3*). The second option should be compatible for a larger range of bioluminescent signals, but it risked both lacking modularity and disrupting the function of any of the two components in the absence of a universal fusion design. Overall, the fusion protein approach was followed as it maximised the chances of engineering a bioluminescence transduction pathway.

4.2.2.1 Intramolecular distance determines the bioluminescent activation of CcaS

Prior to this project, the only relevant knowledge that could be identified to aid the design of CcaS fusion proteins was confined to a single observation. One report had tagged the N terminus of this photoreceptor with a fluorescent protein to quantify its expression levels, but a functional validation of the fusion protein had not been provided[58]. Preliminary efforts in the laboratory had failed at fusing two distinct bioluminescent proteins to this terminus for different reasons. In the first design, the globular luciferase NanoLuc impaired CcaS regulation, causing constitutive signalling activity in both dark and illuminated conditions (*data not shown*). In the second design, the transmembrane luciferase nnLuz was fused without interfering with either of the two protein activities, but the bioluminescent emission was too dim to elicit a detectable photoreceptor activation (*data not shown*). In the absence of further evidence, these observations and the photoreceptor sequence analysis performed earlier in this project (*Section 3.3.2.3*) were used to design a panel of fusion proteins to screen for potentially self-activating CcaS constructs (**Table 7**). Except for two cases, all other fusions contained the luciferase GeNL, as it had provided the brightest green bioluminescent pulse (**Figure 16A**).

Fusion 1 replaced the *wild-type* nnLuz sequence in the preliminary construct for a brighter evolved version (nnLuz_v4; not published) to potentially solve the bioluminescent limitation. A five residue peptide linker optimised for BRET (SGLSR) was used to ensure that the light-emitter would remain as close as possible to the light-sensing domain of CcaS[161].

Fusions 2 and 3 attempted to fuse the globular protein GeNL to the N terminus of CcaS without removing the photoreceptor from the membrane by respectively introducing its

sequence between nnLuz and CcaS, or targeting GeNL to the membrane using nnLuz N-terminal transmembrane helix. In both cases, a long flexible linker (GHGTGSTGSGSS) was used to attach GeNL to CcaS to facilitate the N-terminal insertion of the photoreceptor into the membrane[162].

Fusions 4-6 were designed with the soluble CcaS variant (ΔN23-CcaS) to allow the fusion of two globular luciferases near the chromophore. The first two fusions iterated GeNL with each of the two linkers previously described; and the third one contained RLuc8.6-535, as an alternative bioluminescent source, fused with the shorter linker. These constructs were included for the possibility that molecular contacts provided by the luciferases in the tertiary structure might be able to replace the stabilisation provided by the membrane and recover CcaS activity in this variant.

Fusion 7 was a circular permutation of CcaS that allowed fusing GeNL close to the light-sensing domain without altering the N terminus of the photoreceptor. The redundant internal sequence that separates the light-sensing domain from the histidine-kinase domain in CcaS was replaced by GeNL at the exact same positions used to truncate miniCcaS#10. The short linker flanked GeNL on each side to confer rigidity for signal transduction along the photoreceptor axis.

Fusions 8-10 explored the capacity of GeNL to activate CcaS from the C terminus. The first two fusions tested which linker would provide a better tridimensional arrangement to avoid interfering with CcaS autophosphorylation and phosphotransfer to CcaR; whereas the last one used the internally truncated miniCcaS#10 photoreceptor variant to reduce the distance between the C-terminally fused GeNL and the N-terminal light-sensing domain of CcaS.

Table 7 | Panel of fusion proteins designed to identify potentially self-activating CcaS constructs

Fusion 1	nnLuz_v4-SGLRS-CcaS
Fusion 2	nnLuz_v1-SGLRS-GeNL-GHGTGSTGSGSS-CcaS
Fusion 3	nnLuz_v1(N1-39)-SGLRS-GeNL-GHGTGSTGSGSS-CcaS
Fusion 4	GeNL-SGLRS-ΔN23CcaS
Fusion 5	GeNL-GHGTGSTGSGS-ΔN23CcaS
Fusion 6	RLuc8.6(535)-SGLRS-ΔN23CcaS
Fusion 7	CcaS(N1-221)-SGLRS-GeNL-SGLRS-CcaS(531-C753)
Fusion 8	CcaS-SGLRS-GeNL
Fusion 9	CcaS-GHGTGSTGSGSS-GeNL
Fusion 10	miniCcaS#10-SGLRS-GeNL

The functional validation of these potentially self-activating photoreceptors started with a bioluminescence analysis to confirm that light production was not disrupted in any fusion design. Strains expressing the non-fused bioluminescent components were used as positive controls for comparison to infer the retained activity of each construct. All fusion proteins retained bioluminescence activity at sufficient levels to differentiate their emissions from detector noise: 20- to 800-times higher depending on the construct (**Figure 20A**). Their light outputs were 2-3 orders of magnitude lower than those of their respective non-fused controls. However, this is likely an underestimation of their activity since fused proteins are expected to be expressed at lower levels due to the additional 753 residues corresponding to CcaS photoreceptor. As expected, the most affected designs were those attempting to fuse globular proteins to the N-terminus of CcaS and the circularly-permuted construct (Fusion 2, 3 and 7). Moreover, targeting GeNL to the plasma membrane for N-terminal fusion to CcaS altered the conformation of this Nano-Lantern, causing partial BRET loss from NanoLuc to mNeonGreen

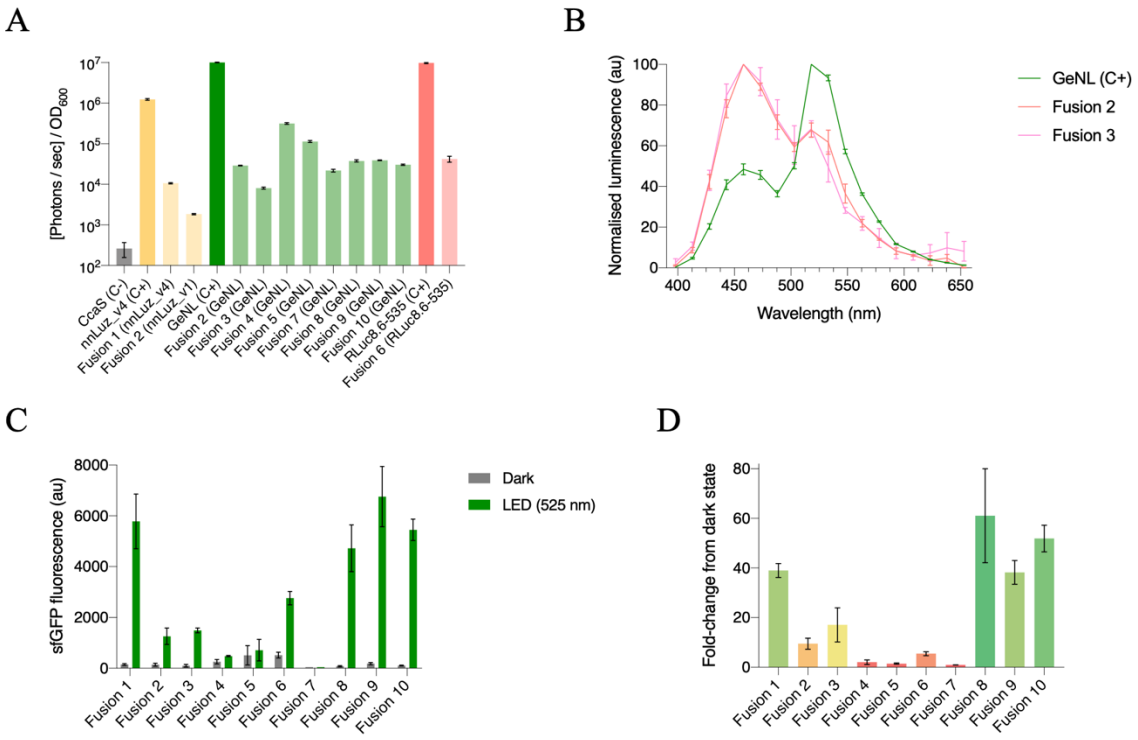


Figure 20 | Functional validation of a panel of potentially self-activating CcaS constructs.
A Optimised bioluminescence levels produced by the luciferase components present in each fusion protein under optimal circuit activation conditions: nnLuz (yellow), GeNL (green), and RLuc8.6-535 (red). Non-fused luciferases were used as positive controls and non-fused CcaS as a negative control for bioluminescence comparison. Bars correspond to the mean ± SD of three biological replicates. **B** Spectral signature of GeNL without fusing (green) and fused to the N-terminus of CcaS in Fusion 2 (red) and Fusion 3 (pink). Individual values correspond to the mean ± SD of three biological replicates. **C** Fluorescence levels produced by each fusion protein under optimal circuit activation conditions in response to either darkness (grey) or saturating green light-illumination (green). **D** Fold-induction responses corresponding to the ratio lit/dark fluorescence levels obtained in this validation, expressed with an additional green-yellow-red colour scale. Bars correspond to the mean ± SD of three biological replicates obtained by averaging the fluorescence of 10,000 cells analysed by flow cytometry.

(**Figure 20B**). In contrast, those fusions that respected the solubility of the luciferases retained higher bioluminescent activity (Fusions 1, 4-6). Finally, C-terminal fusions had intermediate light outputs compared to these two former groups, which might be a consequence of the protein interactions occurring in this domain.

Next, the light-sensitivity of these fusion proteins was tested under optimal circuit activation conditions using either darkness or an external light source. Four constructs exhibited apparently intact light-regulation: the N-terminal nnLuz_v4 fusion and the three C-terminal GeNL fusion proteins. (**Figure 20B**; Fusions 1, 8-10). Their signalling activity was tightly-regulated in the dark and led to high phosphotransfer to CcaR in response to green light, resulting in large fold-induction responses (47 ± 13-fold). All fusion proteins containing globular luciferases attached to the N terminus of CcaS had reduced light-responsiveness, suggesting that they either altered signal transduction from the sensory domain or blocked light delivery to the chromophore (Fusions 2-6). From these constructs, fusions expressing full-length CcaS were the only ones to partially retain the dynamic range of the photoreceptor (**Figure 20C**; Fusions 2 and 3). Instead, those N-terminal fusions that stabilised the unspecific activation of soluble CcaS disrupted its green light-inducibility (Fusions 4-6). Lastly, the circular permutation of CcaS abolished its signalling activity in both states, denoting the formation of a non-functional protein structure upon folding (Fusion 7). Therefore, Fusions 4-7 were discarded due to compromised light-sensitivity, and the other six were used to screen for bioluminescent activations.

To assess the capacity of these CcaS constructs to sense their own light emissions, strains were grown until optimal activation conditions and treated with their respective luciferin or vehicle solutions. For each strain, a vehicle control culture was illuminated with external green light from the time of induction to compare bioluminescence activations to a saturating gene expression response. Moreover, aliquots from all cultures were transferred into a microplate reader upon induction to register their bioluminescent pulses.

Bioluminescence data showed that despite light production being dimmer than in co-expression experiments, GeNL pulses were consistently stable across the 3-hour experiment rather than steeply decaying within minutes (**Figure 21A**; Fusions 2,3,8-10). It is possible that the presumably lower bioluminescent expression levels in fused conditions are associated with a slower luciferin consumption rate, resulting in longer signal duration at the expense of dimmer emissions. In contrast, nnLuz emissions lasted around 45 minutes independently of expression conditions or luciferase variants, varying only in brightness across experiments

(**Figure 21A**; Fusion 1). Since 3-hydroxyhispidin was used in excess in this project, fungal bioluminescence duration was most likely limited by the chemical instability of the purified native luciferin.

Flow cytometry analysis revealed that only two constructs elicited significant increases in gene expression in response to luciferin: Fusion 1 (nnLuz_v4-CcaS) and Fusion 10 (miniCcaS#10-GeNL) (**Figure 21B**). Fusion 1 caused a small activation (~ 2-fold) that appeared to peak early and was detectable at both 1.5 and 3 hours post-induction, which was consistent with the duration of nnLuz_v4 bioluminescent pulse. Instead, Fusion 10 caused a continuous activation of the circuit that was 7-times higher than the vehicle control at 1.5 hours post-induction and increased to 37-times higher by the end of the experiment. Given the lack

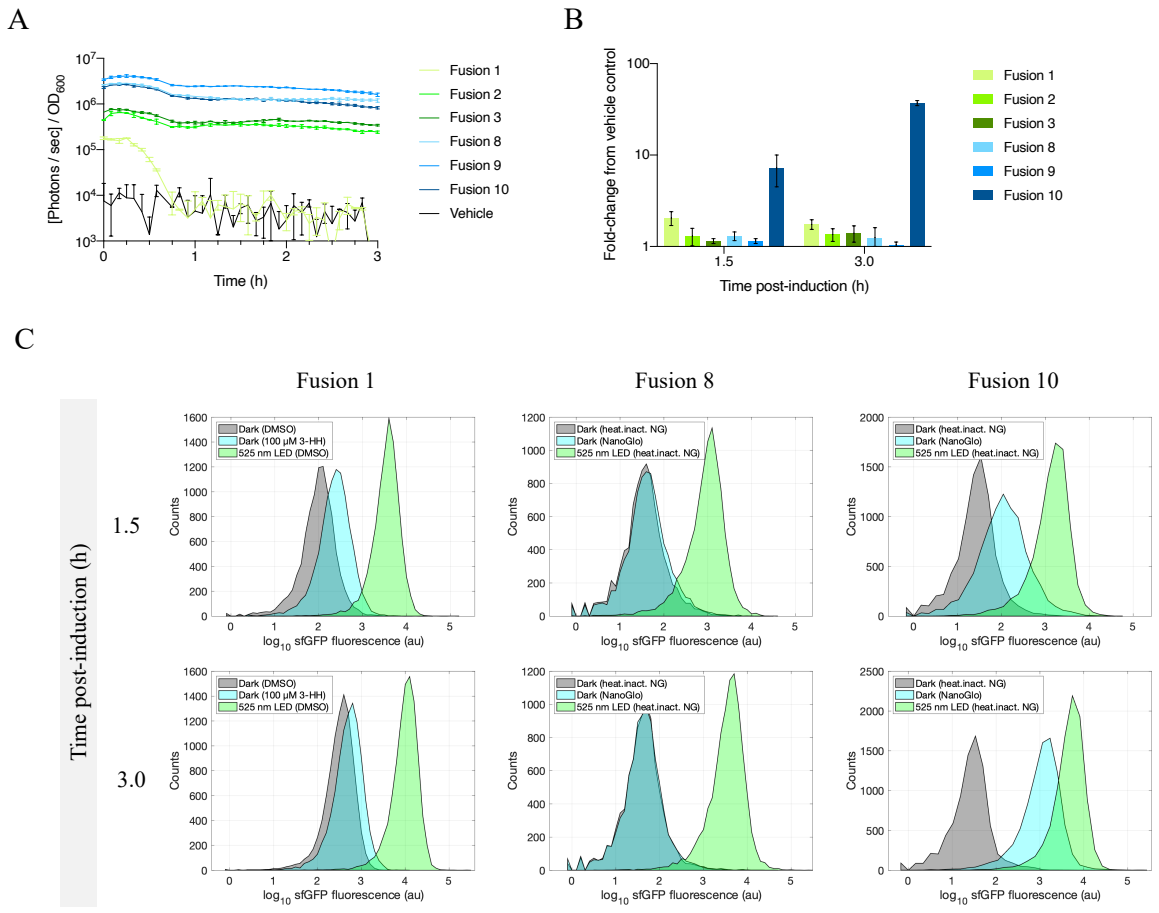


Figure 21 | Screening of potentially self-activating CcaS constructs.
A Optimised bioluminescent pulses produced by previously validated fusion proteins containing nnLuz_v4 (Fusion 1) or GeNL (Fusion 2 and 3) fused to the N terminus of CcaS, or GeNL fused to the C terminus of CcaS (Fusion 8 and 9) or miniCcaS#10 (Fusion 10) under optimal circuit activation conditions. **B** Fold-induction responses in fluorescence expression from the vehicle control group associated to each bioluminescent pulse at 1.5 and 3 hours post-induction. Bars correspond to the mean ± SD of two biological replicates in each case, obtained by averaging the fluorescence of 10,000 cells analysed by flow cytometry. **C** Representative flow cytometry histograms for Fusion 1 (left), Fusion 8 (middle), and Fusion 10 (right) showing the fluorescence distribution of 10,000 cells produced in the dark in response to 1x heat-inactivated NanoGlo® (grey) or 1x NanoGlo®, or saturating green light illumination and 1x heat-inactivated NanoGlo® (green).

of fluorescence saturation or bioluminescence decay in GeNL emissions, it is possible that this activation further amplifies at later time-points.

Remarkably, only two constructs were able to sense bioluminescence despite it being produced from the same photoreceptor structure. These results suggest that close molecular proximity to the light-sensing domain of CcaS might be essential for bioluminescent activation of the circuit, but perhaps not always sufficient. For example, the dimmer emission of nnLuz_v4 was sufficient to activate CcaS activity from the N terminus (Fusion 1), while brighter proteins were unsuccessful when fused to the C terminus (Fusion 8 and 9). However, GeNL also failed to activate CcaS when fused at the same position as nnLuz_v4, despite being 4-times brighter (Fusion 2 and 3). These two GeNL constructs had slightly compromised tridimensional folding, as shown by their activity profiles, and might adopt spatial orientations that are inefficient for light delivery to the chromophore. Another example was GeNL fusion to the C terminus in Fusions 8 and 10, which led to completely different activation outcomes by just changing the length of CcaS, which was 291 residues shorter in the latter construct (**Table 7**). Their identical light pulses failed to induce gene expression in Fusion 8 while causing a high sustained activation in Fusion 10 (**Figure 21C**). Internal truncation of CcaS in Fusion 10 has two effects: it brings GeNL closer to the N-terminal sensory domain and causes a predicted 180-degree switch in its orientation[106]. Although this orientation switch could be important for activation, Fusion 9 contained a longer flexible linker sequence between full-length CcaS and GeNL that should allow higher mobility and still was not able to activate the photoreceptor (**Figure 21B**; Fusion 9). Therefore, evidence is not sufficient to determine the contribution of each variable in each scenario, but intramolecular distance appeared to have a greater influence on the bioluminescent activation of CcaS when luciferases folded properly.

In conclusion, the improved fungal luciferase nnLuz_v4 manged to activate CcaS from its N terminus, overcoming the bioluminescent limitation previously encountered. Still, the rapid signal decay caused, most likely, by luciferin instability resulted in a fold-induction that was deemed impractical for selection purposes (~ 2-fold). Although this problem might be solved by using synthetic luciferin analogues with enhanced stability or the fungal luciferin biosynthesis genes, this project direction was de-prioritised. Instead, the construct containing GeNL fused to the C-terminal end of miniCcaS#10 (Fusion 10) was given preference to further engineer bioluminescence transduction using CcaS-CcaR circuit.

4.2.2.2 CcaS: a modular, tunable platform for bioluminescence transduction with extended spectral compatibility

The transduction of bioluminescence into gene expression using CcaS requires co-localising light production to the immediate environment of its light-sensing domain. In the previous section, this was achieved by fusing a bioluminescent protein to a photoreceptor site that allowed efficient light delivery to the chromophore without disrupting the activity of either of the two proteins. The resulting activation of CcaS-CcaR circuit in response to luciferin was high and sustained, which offered promising prospects to engineer fitness acquisition for selection purposes. However, for this *in vivo* selection tool to be able to have a tangible impact in the development of novel bioluminescent technologies, it should fulfil a number of conditions that were unclear in the current design. Ideally, it should be easy to implement by other users, adaptable to the needs of each new experiment, and compatible with as many bioluminescent proteins as possible.

To begin with, the fusion approach used to transduce bioluminescence limits the reusability of this tool, because a functional fusion topology must be identified for each new luciferase. It was observed that the C-terminal end of CcaS showed high tolerance to fusions, probably because it is exposed to the solvent and allows proper folding of the components without interfering with their activities. Additionally, internal photoreceptor truncation in miniCcaS#10 appeared to bring the C terminus closer to the light-sensing domain and facilitate bioluminescence delivery from that end. For these reasons, it was hypothesised that the C terminus of miniCcaS#10 might constitute a modular docking site to fuse bioluminescent proteins for signal transduction in a plug-and-play manner.

To test this first hypothesis, the vector containing Fusion 10 was modified to replace GeNL coding sequence for a golden gate-like cloning site that would allow introducing any bioluminescent gene at the C terminus of miniCcaS#10 in a one-step cloning reaction. An identical vector was also created containing the longer peptide linker previously used to assess the influence of linker length on CcaS bioluminescent activation from that site. Then, RLuc8.6-535 was shuttled into each vector and assessed for its capacity to induce gene expression in response to luciferin under optimal circuit conditions. Both of these proteins retained the original miniCcaS#10 dynamic range in response to external green light (~ 160-fold at 3h post-induction) and elicited near-maximal activations when treated with luciferin (**Figure 22**). Interestingly, RLuc8.6-535 bioluminescence was able to generate significantly larger gene expression inductions than that produced by GeNL in the same fusion design, 72-fold

compared to 37-fold respectively, despite being 3-times dimmer. Moreover, the presence of a longer linker improved substantially the bioluminescent activation efficiency of the construct, resulting in higher and faster fluorescence production over time that was comparable to the effects of external green light (~ 120-fold). Taken together, these results provide additional evidence to support that the C-terminus of miniCcaS#10 is especially suited to engineer bioluminescent activations, and the fact that intramolecular distance determines the light delivery efficiency to CcaS chromophore.

The second condition for this selection tool to be useful is that it can be tuned to suit different experimental needs. In some instances, low bioluminescence transduction might be necessary to detect activity improvements in already bright luciferases, while the opposite might facilitate the screening of dim luciferases. In fact, this parameter might have to be adjusted multiple times in the evolution process of a protein to suit the changing needs of different selection rounds. Hence, three variables affecting bioluminescence transduction were explored to gain control over the fitness acquisition function: bacterial growth, induction length, and luciferin concentration.

As shown in *Chapter 3*, the ability of CcaS-CcaR to transduce an identical light stimulus varies during bacterial growth depending on the dilution speed of its signalling components. Moreover, gene expression differences between illuminated and dark cultures are proportional to their division rates at the time of induction, due to the accumulation of basal circuit activity. Accordingly, the induction point selected in the growth curve should allow

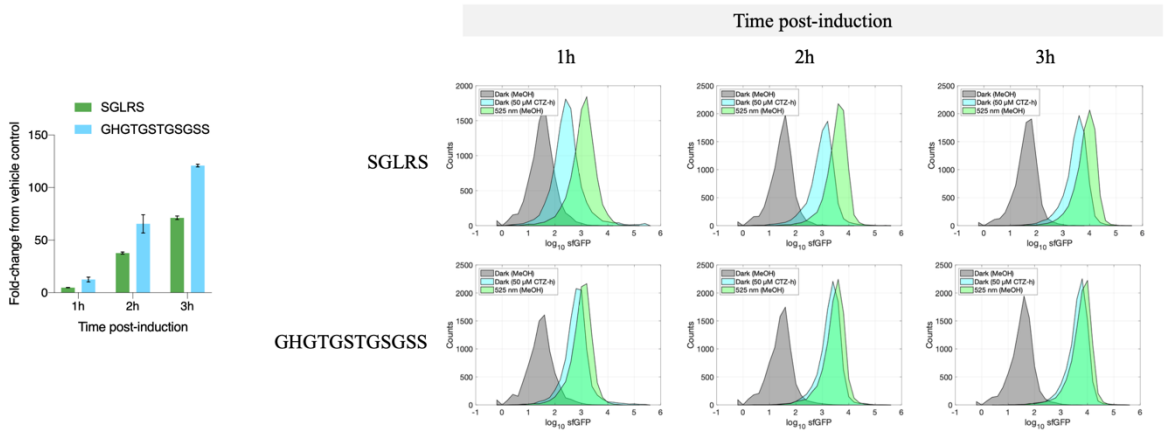


Figure 22 | Bioluminescence transduction efficiency in miniCcaS#10-RLuc8.6-535 improves with longer peptide linker. Fold-induction responses (left) corresponding to the ratio luciferin/vehicle of fluorescent levels produced by miniCcaS#10-RLuc8.6-535 fusion construct when using two peptide linkers of different length (green, SGLRS; blue, GHGTGSTGSGSS). Bars correspond to the mean ± SD of two biological replicates obtained by averaging the fluorescence of 10,000 cells analysed by flow cytometry and calculating the fold-change between treatment conditions in each replicate. Representative flow cytometry histograms (right) showing the distribution of fluorescence expression over time in 10,000 cells treated with either vehicle control in the dark (grey), 50 µM of coelenterazine-h (cyan), or vehicle control under saturating green light (green).

adjusting the transduction efficiency of the circuit to modulate fitness acquisition. In turn, the length of induction should define the levels of bioluminescence transduction that can occur before introducing the selective pressure. Finally, luciferin concentration should provide additional external control over fitness acquisition by regulating photon production inside cells.

The degree of control offered by these circuit variables was assessed using the two miniCcaS#10 C-terminal fusions to GeNL and RLuc8.6-535 in parallel experiments. For bacterial growth and induction length, cultures were grown to distinct points in the exponential phase and interrogated with either excess luciferin, vehicle control, or vehicle control and external green light, and their fluorescence production was monitored over time. In line with previous characterisation data, faster-dividing cultures generated larger gene expression differences between induced and non-induced cultures that amplified over time independently of bacterial growth rate (**Figure 23A**). As previously observed, fluorescence production was

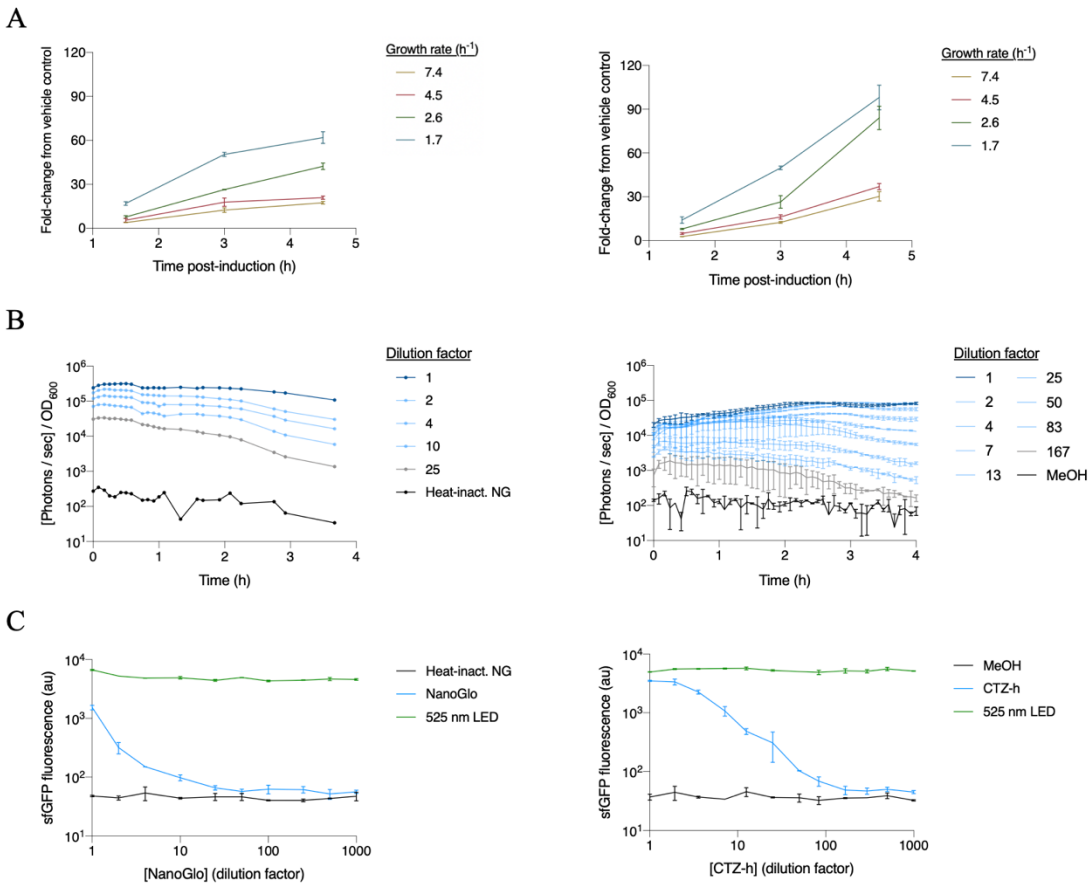


Figure 23 | Self-activation profiles of two miniCcaS#10 sensors fused C-terminally to different bioluminescent proteins. Both miniCcaS#10-GeNL (left) and miniCcaS#10-RLuc8.6-535 (right) allowed tuning circuit activation across the whole dynamic range using three externally controlled variables: bacterial growth, induction length, and luciferin concentration. **A** Fold-induction responses in gene expression obtained using optimal luciferin dosage at distinct culture growth rates. **B** Bioluminescent pulses obtained for a titration of luciferin concentrations expressed as the dilution factor applied in each condition: optimal dose (dark blue), detectable doses (light blue), non-detectable doses (grey), and vehicle control (black). **C** Fluorescence levels produced in response to each luciferin dilution (blue), the respective vehicle controls in the dark (black) or external green light (green) after 3 hours of induction. Bars correspond to the mean ± SD of two biological replicates obtained by averaging the fluorescence of 10,000 cells analysed by flow cytometry (**A**, **C**) or whole-well luminescence (**B**).

consistently twice higher for miniCcaS#10-RLuc8.6-535 across time-points, resulting in gradually higher fold-induction responses. In a follow-up experiment, the dose-response curves of these constructs were measured by performing a luciferin titration under optimal circuit activation conditions. Their respective substrates were tested using serial dilutions across 4 logs of luciferin concentration, starting from the highest dose recommended by the manufacturer. Fluorescence analysis showed that miniCcaS10#-RLuc8.6-535 induced gene expression in response to bioluminescent pulses up to 10-times dimmer than miniCcaS#10-GeNL (**Figure 23B**). As a result, the GeNL construct was only responsive across 1 order of magnitude of NanoGlo concentration, whereas the RLuc8.6-535 construct showed sensitivity until a 100-fold dilution of its optimal CTZ-h dosage (**Figure 23C**). As previously mentioned, the higher bioluminescence transduction efficiency of the RLuc8.6-535 construct probably originates from its longer peptide linker, which seems to improve light delivery to miniCcaS#10 chromophore.

Collectively, these results demonstrate that both constructs can be externally controlled using these three variables to tune bioluminescence-mediated gene expression across the whole 120-fold dynamic range of CcaS-CcaR circuit (**Figure 24**). Moreover, each variable should theoretically favour the selection of specific enzymatic traits: bacterial growth affects

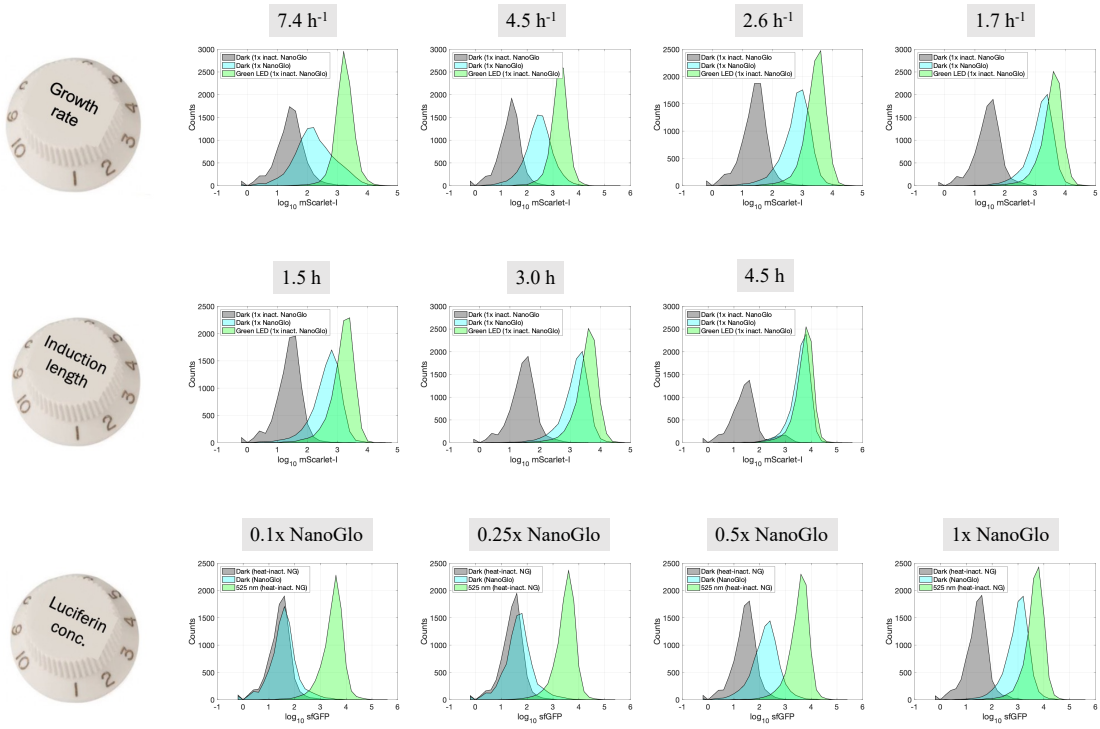


Figure 24 | Three externally controlled variables for tuning bioluminescence transduction in CcaS-CcaR circuit. The growth-dependent signalling activity of CcaS-CcaR can be used to influence transduction efficiency at the time of induction (top). Induction length permits regulating the amount of bioluminescence transduction that occurs before selection (middle). The luciferin concentration can be used to modulate photon production inside cells (bottom).

transduction efficiency and should thereby allow selecting overall brighter bioluminescent proteins; induction length restricts the temporal activation of the circuit and should benefit variants with faster kinetics; and luciferin concentration limits the number of bioluminescent reactions and should thus preferably select proteins with improved quantum yields. Therefore, bioluminescence transduction in this circuit should be relatively adaptable to the needs of a wide range of selection experiments.

In the last instance, the potential for the wide application of this technology depends on the diversity of bioluminescent proteins that might be compatible with it. Most natural bioluminescent reactions emit around the blue-green region of the visible light spectrum, although yellow and red emissions also exist with lower occurrence. As mentioned earlier, the ground state of CcaS chromophore absorbs light from 450 to 600 nm with varying efficiency that peaks at 536 nm (**Figure 15A)**, and might allow sensing blue and yellow emissions despite its higher affinity for green photons. Additionally, its excited state absorbs light at orange and red wavelengths. Therefore, CcaS-CcaR should theoretically be spectrally compatible with any bioluminescent reaction as long as fitness acquisition is coupled to the right signalling trajectory.

To infer the scope of CcaS-CcaR in transducing different bioluminescent reactions, a blue-emitting protein was used to validate CcaS spectral compatibility with the most occurrent bioluminescent wavelengths in nature. For practical reasons, the already characterised miniCcaS#10-GeNL construct was repurposed to emit blue light by inactivating the spectral tuning of NanoLuc by mNeonGreen that occurs in GeNL (**Figure 25A**). This was achieved by substituting the tyrosine involved in the formation of the chromophore ring in mNeonGreen for a glycine (Y59G) to cancel its photochemical properties, and thus recover the original blue NanoLuc emission[163]. Moreover, a key residue in NanoLuc sequence was also mutated to generate a catalytically-inactive luciferase variant (R167A; unpublished) to confirm that photoreceptor activation is due to bioluminescence absorption.

As hypothesised, mNeonGreen Y59A was unable to accept BRET from NanoLuc in GeNL, and displayed the blue spectral signature of the luciferase, while NanoLuc R167A remained enzymatically inactive in the presence of luciferin (**Figure 25B**). Despite producing less than 10% of green photons, the blue-peaked emission of NanoLuc was able to generate fluorescence levels that were comparable to the functional GeNL construct (**Figure 25C**). Instead, NanoLuc R167A elicited identical fluorescence levels to the vehicle-treated dark control. Although it is not possible to differentiate the contribution of each photon wavelength

in this activation, the dose-response curve of miniCcaS#10-GeNL suggests that photon absorption is limiting in the current conditions (**Figure 23C**; left). It is quite likely that blue-light absorption by CcaS is sufficiently efficient to compensate for the loss of green photons in GeNL Y59A, as both constructs produced comparable fluorescence levels. Therefore, the ground state of CcaS is similarly efficient at transducing blue and green bioluminescent reactions, at least in this fusion topology.

In conclusion, these results demonstrate that miniCcaS#10 can be used as a modular platform for engineering bioluminescence transduction in a user-friendly manner with a wide dynamic range, external tunability, and extended spectral compatibility. Accordingly, it should be able to suit a wide variety of light-emitting proteins and experimental needs, and potentially facilitate the development of novel bioluminescent technologies.

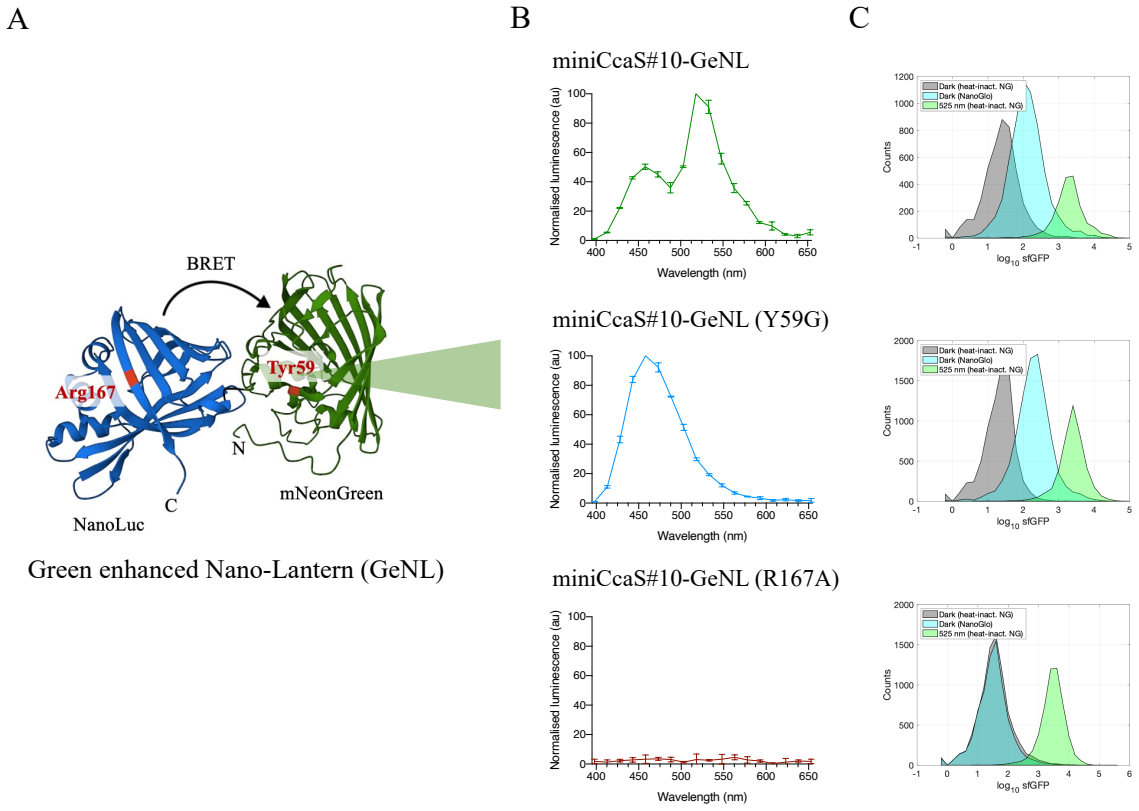


Figure 25 | CcaS photoreceptor transduces blue and green bioluminescent signals at comparable efficiencies.
A Predicted structure of the BRET-based bioluminescent GeNL obtained using Evolutionary Scale Modelling[208]. The targeted residues in NanoLuc (blue) and mNeonGreen (green) are highlighted in red. **B** Bioluminescence emission spectra of GeNL (top), GeNL with inactive mNeonGreen (Y59A; middle), and GeNL with inactive NanoLuc (R167A; bottom) produced in the presence of NanoGlo®. **C** Representative flow cytometry histograms showing the fluorescence distribution of 10,000 cells after 3 hours of induction with either 1x heat-inactivated NanoGlo® (grey), 1x NanoGlo® (cyan), or 1x heat-inactivated NanoGlo® and external green light (green) under optimal circuit activation conditions.

4.3 Conclusions

In the second stage of this project, the aim was to repurpose two light-controlled transcriptional circuits to sense intracellular bioluminescence instead of external light. For this purpose, a diverse array of bioluminescent proteins containing distinct luciferin-luciferase pairs, enzymatic mechanisms, protein solubilities, and spectral properties, was used to investigate distinct bioluminescence transduction strategies in bacteria.

Initially, it was attempted to deliver bioluminescence to the photoreceptors by simply co-expressing each protein with its spectrally-compatible light-sensing circuit, as it had previously been achieved in mammalian cells[134]. This strategy promised high modularity for further tool implementation since circuit components are physically decoupled and cannot interfere with each other. However, the light outputs obtained after optimising luciferase expression and luciferin dosage were insufficient to exert control over these circuits under optimal activation conditions. The same bioluminescent constructs used by those authors were present in the array, which suggests that bioluminescent expression capacity in bacteria might be considerably lower than in mammalian cells and/or the intracellular light-sensitivity of their genetic circuits was higher.

Then, an improved version of the bacterial bioluminescent operon was co-expressed with these circuits to ensure constant luciferin supply inside cells. This was motivated by the fact that a dimmer version of this operon had been able to activate an extracellular photoactivatable protein in the only distant bioluminescent activation reported in bacteria to date[135]. Again, different challenges were encountered that hindered the distant transduction of bioluminescence. On one side, it was not possible to clone the operon into any circuit vector and the only available co-expression conditions reduced bioluminescence expression considerably. On the other side, the energy requirements of the operon constrained its optimal activity to specific growth stages and was further impacted by the metabolic burden imposed by the light-sensing circuits, which nutritional supplementation could not rescue sufficiently to induce gene expression.

Finally, close-contact delivery of bioluminescence was engineered by screening a panel of fusion protein designs with CcaS photoreceptor. It was discovered that the C-terminal end of CcaS constitutes a tolerant fusion site that can accommodate distinct bioluminescent proteins without altering the signalling activity of the photoreceptor. However, intramolecular distance to the light-sensing domain of CcaS limited bioluminescent activation, as miniaturised

photoreceptor sequence was required for self-activation. This hypothesis was confirmed by optimising the linker sequence, which showed that bioluminescence transduction efficiency was significantly improved in the presence of a longer linker peptide. Furthermore, multiple tool utility parameters were tested to demonstrate that this bioluminescence transduction system can be easily repurposed to sense blue and green light-emitting proteins with comparable efficiency, and its activity levels can be externally tuned across a wide dynamic range to suit different experimental needs.

Overall, it was concluded that this circuit offered promising features for the development of novel bioluminescent technologies. Its characterisation data should also contribute to guide future experimental designs and its fine-tuned utilisation. For these reasons, it was selected for coupling bioluminescence transduction to fitness acquisition and testing distinct selection strategies in *Chapter 5*.

5. Translating bioluminescent signals into bacterial fitness

The selective potential of distinct antibiotic-based strategies was investigated to develop an *in vivo* selection platform for bioluminescent bacteria using a luciferin-inducible CcaS-CcaR circuit. Regardless of the selectable gene tested, protein accumulation dynamics determined the selection efficacy of this circuit. Although final optimisations are still needed, engineered protein degradation impeded unspecific resistance acquisition caused by basal gene expression, but translation rates used were insufficient to rescue bacteria upon bioluminescence induction.

5.1 Introduction

5.1.1 Design considerations for a versatile *in vivo* selection platform

In the previous chapter, a bioluminescence transduction pathway was engineered in bacteria that converted intracellular light signals into measurable gene expression levels using a fluorescent reporter gene. The third stage of this project aimed at deploying this pathway to develop an *in vivo* platform capable of isolating specific bioluminescent genotypes from diverse bacterial populations. Amongst other aspects, this involved deciding on selection strategies, identifying genetic designs for their implementation, and establishing a protocol for efficient selection in proof-of-concept experiments.

Three types of *in vivo* selection strategies exist for enriching cell populations with a desired function, depending on whether the transduction mechanism confers a reproductive advantage (growth-based selection), cell viability (survival-based selection), or the expression of reporter gene (screening methods)[164]. Growth-based strategies increase the replication rate of cells with desired genotypes until they outcompete the rest of the population; survival-based strategies compromise the viability of all clones that fail to acquire sufficient fitness expression; and screening methods enable the rational isolation of genotype variants by quantifying the activity of genes of interest in individual cells (**Figure 26**).

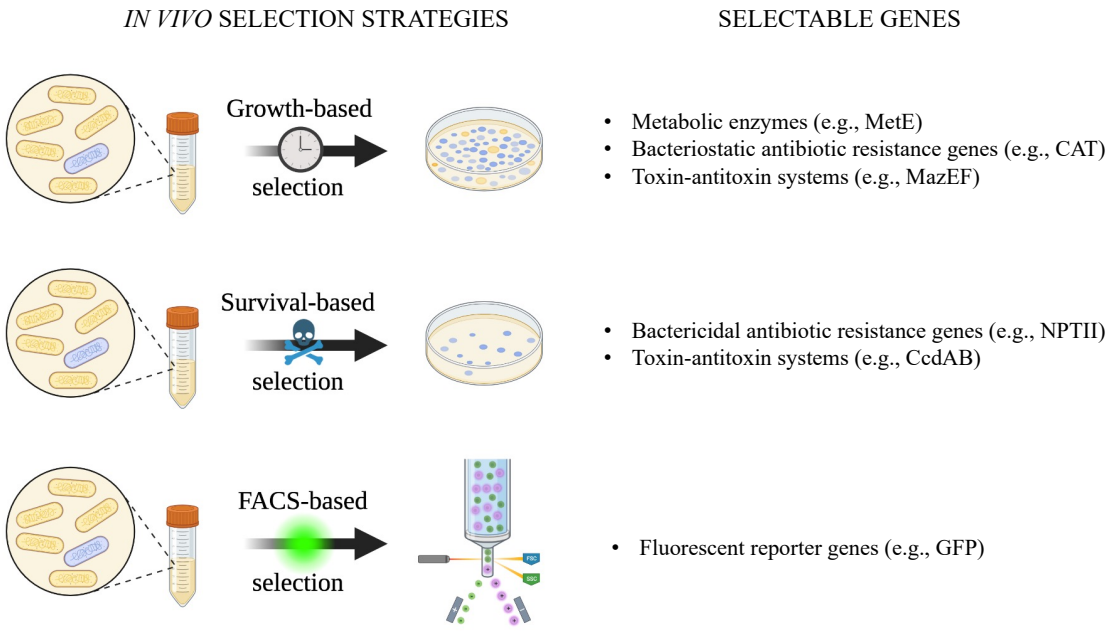


Figure 26 | Main strategies for *in vivo* selection of genotypes in directed evolution experiments.
Mutant cell libraries can be enriched for a specific function by rigorously linking the expression of genotypes into selectable phenotypes. This is usually achieved with synthetic transduction circuits that confer fitness advantages or the expression of a quantifiable protein to each individual cell as a function of the activity of the gene of interest.

These three strategies use phenotypic information to rescue desired genotypes, and hence rely on robust genotype-to-phenotype translation for effective selection. This function is performed by the engineered transduction pathway and errors or interferences with its activity originate false negatives and false positives in the selection process. Although both selection errors can occur due to transduction variability and other stochastic factors, false positives can also appear as a result of semi-predictable variables such as cheater mutations, phenotypic fitness emergence, and circuit background activity. Furthermore, it is important to consider the dynamic range of selection offered by distinct selective pressures and their counteracting fitness genes (or detection instruments and reporter genes), and the external control provided by the selective pressure to modulate its intensity and temporal effects.

Growth-based selection strategies are especially suitable for engineering continuous evolution experiments, which allow complex exploration of fitness landscapes at large scale and high sequence depth[165]. These mechanisms require fitness acquisition and nutritional resources to last until selection is effectuated and external control of their selective pressures is not always possible. Additionally, they might be more susceptible to the apparition of false positives due to accumulation of basal circuit activity over time. Survival-based strategies have faster selection dynamics, rely less on circuit activity over time, and their selective pressures are usually chemical challenges that can be externally controlled. Unlike the previous type, they might be less compatible for continuous experiments due to the emergence of unspecific resistance mechanisms, and require being tuned in each selection round to obtain accurate recovery rates. Finally, screening methods are simpler to design but usually depend on laborious pipelines or expensive instruments to achieve high-throughput or automation[166]. One of the most successful screening techniques in protein engineering is fluorescence-activated cell sorting (FACS), which can automate the analysis of virtually any population size at the single-cell level and isolate variants of interest based on their fluorescence intensity[167–169]. Hence, each selection strategy offers advantages and disadvantages that should be considered in relation to the properties of the transduction pathway and the protein engineering goals.

Four types of output genes were identified during this project to potentially engineer any of the selection strategies mentioned above: metabolic enzymes, toxin-antitoxin systems, antibiotic resistance genes, and fluorescent proteins.

(1) Metabolic enzymes provide a basis for growth-based selection strategies. For instance, CcaS-CcaR has been used twice to express the final enzyme in methionine biosynthesis (*metE*) in response to external light, allowing the control of bacterial growth in a

basic medium lacking this essential amino acid[73,74]. However, an important concern with engineering metabolic selections using CcaS-CcaR is that accumulation of basal circuit activity would lead to unspecific growth, ultimately reducing the selection efficiency of the system. In fact, the first study reported a 4-fold dynamic range in growth control as a result of this limitation, while the second study used a turbidostat to dilute unspecific growth. Furthermore, metabolic strategies depend on selective pressures that cannot be externally controlled, which reduces their utility.

(2) Toxin-antitoxin systems typically consist of dual expression modules in which one gene disrupts an essential cellular process (toxin) while the other gene counteracts its activity to recover bacterial homeostasis (antitoxin)[170]. One of these systems was used to control bacterial growth by multiplexing pDusk and pDawn, inhibiting growth in the dark and restoring it in response to blue light, and could theoretically be repurposed as a selection mechanism[75]. Unlike metabolic enzymes, toxin expression can be engineered to allow external control of the selective pressure strength using an inducible promoter. Although certain toxin-antitoxin systems target protein synthesis and could address the basal activity accumulation of CcaS-CcaR, they were not considered due to time constrains, as they required additional circuit engineering.

(3) Antibiotic resistance genes are a broad and diverse group of genetically-encoded defense mechanisms against compounds that are produced to kill (bactericidal) or inhibit the growth of bacteria (bacteriostatic) and other microorganisms[171]. These antibiotics and resistance genes constitute biochemical selective pressures and fitness adaptations that co-evolved to generate competitive advantages amongst microorganisms. Apart from their diversity, they offer high molecular specificity as they recognise or protect key regulators of essential cell processes, such as DNA replication, transcription, translation, cell wall synthesis, and primary metabolic routes. In fact, distinct species have evolved antibiotics with different targets and binding sites, and often multiple resistance mechanisms also exist for each compound[172]. Moreover, their clinical importance and utility in research as selection markers for plasmid maintenance have resulted in extensive knowledge being available about their diversity, mechanisms of action, kinetics, interactions, and laboratory use[173,174]. As a result, these compound-gene pairs have a great potential for engineering both growth- and survival-based selection strategies in a multiplicity of ways that should suit most design specifications, such as the need for protein synthesis inhibition in CcaS-CcaR. For example, they have been used to engineer light-control of bacterial growth in whole cultures using pDawn[60] and in

bacterial consortia using Opto-T7RNA*(563)[130], and bacterial survival at the single-cell level using CcaS-CcaR[121].

(4) Fluorescent proteins are especially useful as reporter genes for the development of high-throughput screening methods due to their compatibility with FACS. As explained in *Chapter 1*, bioluminescent protein engineering is mainly limited by the low throughput and lack of automation of current screening methodologies. While expensive robotic workstations can facilitate the process, these are not economically viable for most laboratories and still require a significant degree of supervision. This project aimed to develop a bioluminescent selection platform that relied on biological computation rather than expensive devices to democratise protein engineering in this field and enable *in vivo* evolution of bioluminescent proteins. Nevertheless, the bioluminescence transduction pathway already translated bioluminescent signals into fluorescence and could be used to leverage the screening power of FACS as an additional selection mode in parallel with the original plan.

For these reasons, it was decided to assess multiple antibiotic-based selection mechanisms and a FACS screening protocol to build a versatile toolkit of selection genes that could meet different needs in bioluminescent protein engineering.

5.1.2 Shortlisted antibiotic compounds and resistance genes

The antibiotic selection mechanisms tested in this chapter aimed to provide a set of selective pressures as varied as possible to maximise the chances of identifying useful selection strategies. Antibiotics were selected to block distinct cell processes, sometimes using different molecular targets, and resistance mechanisms were diversified to influence culture dynamics in multiple ways (**Table 8**). Six antibiotic-resistance gene pairs were ultimately selected for laboratory implementation following a shortlisting process available in detail in *Chapter 2*:

Ampicillin resistance (AmpR). The antibiotic chosen for this selection mechanism was a semi-synthetic ampicillin derivative with higher resistance to inactivation and improved stability called carbenicillin[175]. Like other β-lactam antibiotics, it is a bactericidal compound that enters through the outer membrane pores of Gram-negative bacteria and irreversibly binds to the catalytic site of peptidoglycan transpeptidases with its β-lactam ring. This inhibits the last step in the formation of the bacterial cell wall, leading to osmotic imbalances and cell lysis. The resistance gene selected, TEM-116, is a more efficient mutant variant of the first AmpR gene described TEM-1, which catalyses the hydrolysis of the β-lactam ring of the antibiotic[176].

Kanamycin resistance (KanR). Kanamycin is an aminoglycoside with fast bactericidal action that binds irreversibly to the small ribosomal subunit 30S and inhibits protein synthesis[177]. It specifically interferes with the ribosome decoding site, leading to the mistranslation of messenger RNAs and production of toxic peptides that insert and permeabilise the cell membrane[178]. In turn, this promotes further antibiotic entry and creates a positive feedback that rapidly compromises bacterial viability. The resistance gene selected was neomycin phosphotransferase II (*nptII*), which inactivates the antibiotic in the cytoplasm via phosphorylation[179].

Zeocin resistance (ZeoR). Zeocin is the commercial formula for a broad-spectrum glycopeptide antibiotic from the bleomycin family, also known as phleomycin D1. This molecule intercalates between DNA base pairs and forms a complex with metal ions that catalyses the production of reactive oxygen species, which cause single and double DNA strand breaks and lead to cell death[180]. The cytotoxic effects can be suppressed with the overexpression of any of the *ble* genes, a family of zeocin-binding proteins that stoichiometrically sequester the antibiotic and prevent DNA damage[181].

Tetracycline resistance (TcR). Tetracycline is a bacteriostatic that passively diffuses through the bacterial membranes and reversibly binds to the 30S ribosomal subunit at a highly conserved site[182,183]. This interaction blocks translation at the elongation step by physically impeding the aminoacyl-transfer RNA (tRNA) from contacting the messenger RNA (mRNA)-ribosome complex. Multiple genetically-encoded TcR mechanisms are known to exist, involving active drug efflux, enzymatic inactivation, and ribosome protection[184]. For this project, the tetracycline efflux pump gene *tetC* was chosen to increase bacterial competition in selection experiments.

Erythromycin resistance (EmR). Erythromycin is a bacteriostatic that reversibly binds to the large ribosomal subunit 50S and inhibits protein synthesis by interfering with peptide bond formation[185]. Recent research has shown that erythromycin effects are sequence- and context-specific, depending on macrolide-stalling motifs in the nascent peptide, which can cause premature ribosomal release[186], translation arrest[187], frameshifting[188], or allow complete synthesis[189]. The most prevalent EmR genes are ribosomal RNA methyltransferases that modify the drug binding site to protect the ribosomes, such as the gene used in this project, *ermC*[190].

Trimethoprim resistance (TmpR). Trimethoprim is lipophilic weak base that reversibly inhibits bacterial dihydrofolate reductase (*dhfr*), blocking the production of tetrahydrofolate[191]. This co-factor is an essential precursor of three nucleotides (thymine, adenine, and guanine) and two amino acids (methionine and histidine), and its depletion incapacitates several processes including DNA replication and protein synthesis, hence suppressing bacterial growth and division. The only TmpR mechanism identified to date relies on dhfr mutations that render the enzyme insensitive to trimethoprim. For instance, substitutions at position Leu 28 in *E. coli dhfr* gene for bulky or polar amino acids, such L28R, alter the enzyme structure in a detrimental way for trimethoprim binding[192].

Table 8 | Antibiotic selection mechanisms shortlisted for *in vivo* bioluminescence enrichment

Antibiotic	Effect	Mechanism of action	Resistance mechanism
Carbenicillin	Bactericidal	Inhibits cell wall synthesis	Antibiotic hydrolysis
Kanamycin	Bactericidal	Mistranslation and toxic peptide production	Antibiotic phosphorylation
Zeocin	Bactericidal	ROS-induced DNA damage	Stoichiometric binding
Tetracycline	Bacteriostatic	Inhibits protein elongation	Active drug efflux
Erythromycin	Bacteriostatic	Inhibits translation (various)	Ribosome methylation
Trimethoprim	Bacteriostatic	Inhibits folate biosynthesis	Insensitive dhfr mutant

5.2 Results and discussion

5.2.1 A bicistronic reporter for dual selection

5.2.1.1 Translation inhibition is required to block basal resistance acquisition in CcaS-CcaR

As mentioned above, the bioluminescence transduction circuit engineered in the previous chapter produced a fluorescent marker as a function of intracellular light. The main objective of this project was to replace the fluorescent reporter in this circuit for any fitness gene that allowed isolating specific bioluminescent genotypes *in vivo* without screening their activities. Still, the circuit as it was could potentially leverage the power of FACS for high-throughput screening and separation of specific bacterial populations as an additional bioluminescence selection mode. Accordingly, it was hypothesised that a bicistronic reporter simultaneously expressing the fitness gene and the fluorescent marker in response to bioluminescence could combine both selection strategies in a single genetic design.

The bicistron structure was designed respecting the original circuit output gene sequence, using the same RBS for both genes (BBa_B0034) at the same distance from their start codons. In each case, the antibiotic resistance gene occupied the first position and sfGFP preceded the terminator sequence (BBa_B0015) without additional spacing between genes, except for two stop codons. Six bicistron variations were assembled comprising each of the shortlisted antibiotic resistance genes: AmpR, KanR, ZeoR, TcR, EmR, and TmpR.

To identify the most suitable antibiotic selection mechanism for CcaS-CcaR, the basal resistance conferred by the unspecific accumulation of circuit activity was tested using a gradient of antibiotic concentrations. Each strain expressing one of the six bicistronic units was grown in the dark until late-exponential phase to dilute active CcaS and basal resistance, and then treated with either fresh medium, 0.05x, 0.1x, 0.5x, 1x, 5x or 10x the respective minimum inhibitory concentration (MIC). Except for relatively high doses of kanamycin and tetracycline, the rest of antibiotics failed to reduce unspecific bacterial growth compared to untreated cultures at any concentration tested (**Figure 27**).

As expected, only antibiotics inhibiting translation were able to stop proliferation in the dark, confirming that accumulation of basal circuit activity was sufficient to confer antibiotic resistance in CcaS-CcaR. For instance, strains expressing the AmpR or ZeoR genes were unaffected by doses 10-fold higher than the MIC, whereas tetracycline and kanamycin could gradually suppress unspecific growth to negligible levels in strains containing KanR or TcR as

circuit output genes. Still, complete growth inhibition in the dark required 5-times the MIC for both antibiotics, which indicated that resistance expression levels probably needed tuning down to improve the selection properties of these strategies.

Erythromycin and trimethoprim failed to control bacterial growth despite suppressing gene expression, most likely because their resistance mechanisms provide drug-insensitive molecules to sustain the targeted cell processes. Pre-existing ErmC should be able to protect new ribosomes not yet methylated by the time of antibiotic addition, while trimethoprim-insensitive DHFR should maintain folate metabolism while more resistance gene and other essential genes are being expressed. Although expression and degradation rates of drug-resistant DHFR could be engineered to minimise unspecific folate production, ribosomal methylation by ErmC might be long lasting and lead to resistance accumulation even in the presence of very low protein expression levels.

Therefore, although optimisation of output gene expression seemed desirable, it was proceeded to test the preliminary utility of kanamycin and tetracycline as potential selection mechanisms with their respective partially-resistant strains.

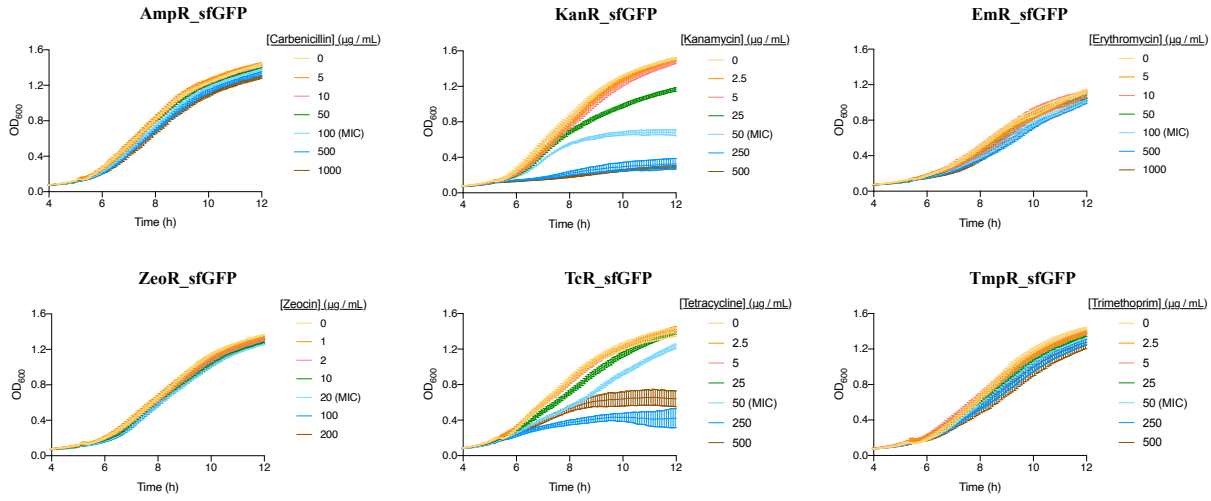


Figure 27 | Unspecific bacterial growth caused by basal antibiotic resistance acquisition in CcaS-CcaR.
Optical density curves of six luciferin-inducible CcaS-CcaR strains growing in the dark under distinct antibiotics and antibiotic concentrations. Each strain expressed a bicistronic reporter with sfGFP and an antibiotic resistance gene (AmpR, KanR, EmR, ZeoR, TcR, or TmpR) and was treated with a gradient of antibiotic concentrations that ranged from 0.05x to 10x the MIC at late-exponential phase (OD ~ 0.12). Individual values correspond to the mean ± SD of four biological replicates.

5.2.1.2 The bicistronic design compromised the light-inducibility of both selection genes

For these two antibiotic resistance genes to allow antibiotic selection of bioluminescent bacteria, they had to be able to rescue cell growth upon induction with luciferin under inhibiting antibiotic concentrations. The first circuit variation tested contained the CcaS-GeNL

photoreceptor with the KanR_sfGFP bicistronic reporter. To assess both the magnitude and temporal dynamics of resistance acquisition in response to bioluminescence, the strain was induced with 1x NanoGlo (luciferin) or 1x heat-inactivated NanoGlo (vehicle) and then treated with distinct antibiotic concentrations at different time-points while monitoring culture density.

Sub-inhibitory doses of kanamycin failed to generate significant growth differences between luciferin- and vehicle-treated cultures, independently of the amount of bioluminescence produced before antibiotic treatment (**Figure *28***; light blue and blue). The untreated controls (yellow) confirmed that this lack of antibiotic selection was not the result of basal resistance being too high to allow for significant outgrowth upon induction, but due to an apparent absence of bioluminescence-mediated fitness acquisition. The only exception was 50 µg / mL at 2-hours post-induction after approximately 8-hours of antibiotic treatment, which

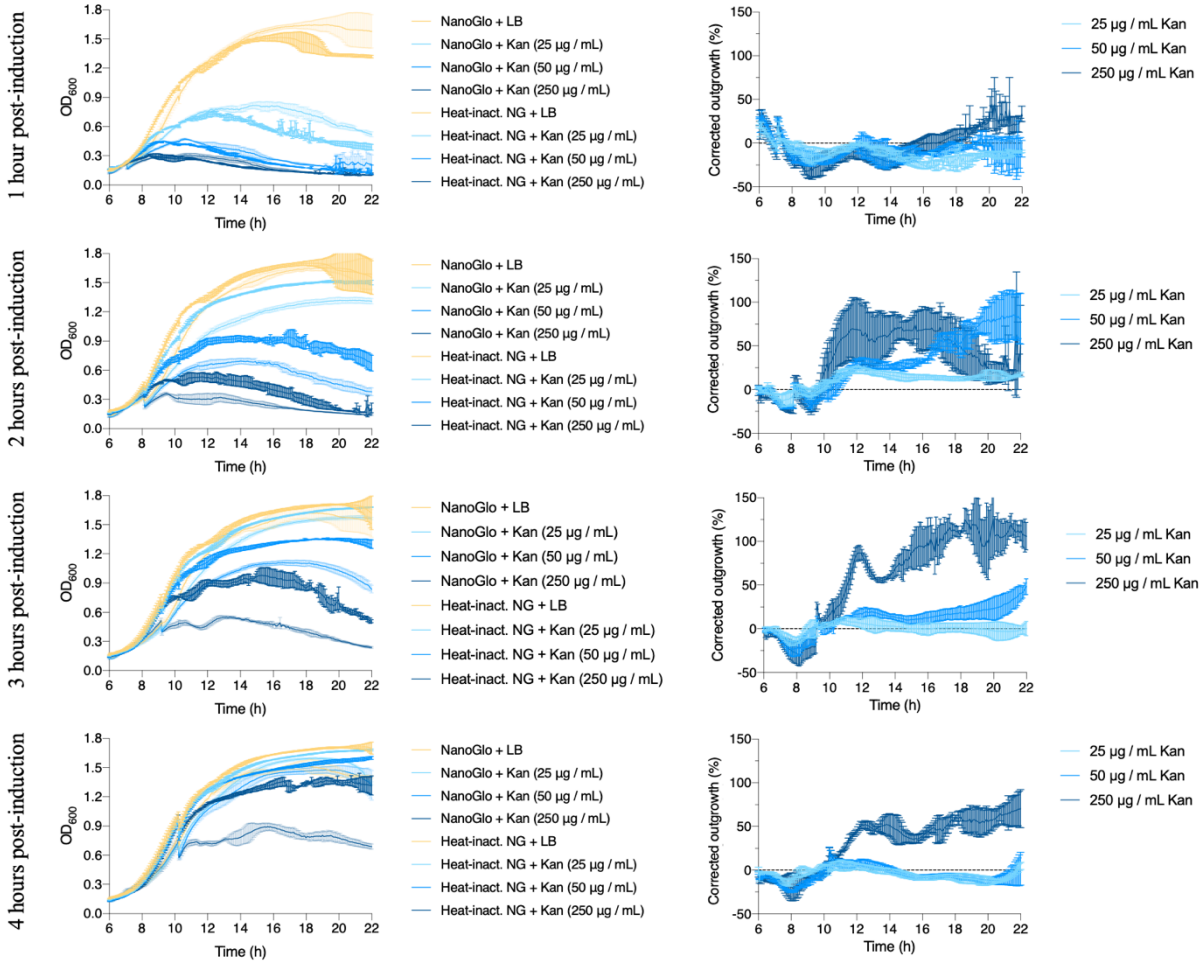


Figure 28 | Bioluminescent activation of CcaS-CcaR with the bicistronic reporter KanR/sfGFP confers subtle growth advantages under inhibitory kanamycin concentrations.
Optical density curves (left) of cultures induced at late-exponential phase with luciferin (opaque) or vehicle solution (transparent) and then treated with either LB medium (yellow), 25 µg / mL (light blue), 50 µg / mL (blue), or 250 µg / mL (dark blue) of kanamycin at distinct time-points. Outgrowth (right) of the luciferin-treated cultures from their respective vehicle controls, expressed as a percentage and corrected by the artifact difference observed between untreated groups upon LB administration. Individual values correspond to the mean ± SD of three biological replicates.

seemed to occur because the luciferin-treated group was lysing at a slower pace than the vehicle control. Instead, the previously titrated MIC for basal CcaS-CcaR resistance (250 µg / mL) showed a subtle bioluminescence-mediated outgrowth from 2-hours post-induction onwards, although these differences were lower and transitory at the earlier time-point (**Figure *28***; dark blue). This result was consistent with previous characterisation data showing that luciferin-induced gene expression in this circuit takes around 3 hours to significantly differentiate from the vehicle control (**Figure 23A**; left).

Bioluminescence-mediated growth differences using the KanR/sfGFP bicistronic reporter appeared to have been minimal if not inexistent. However, kanamycin is a bactericidal antibiotic and OD measurements cannot distinguish viable from dead cells unless lysis occurs, which seemed to have happened at different rates between groups under some conditions. For this reason, a bacterial viability assay was performed following a previously published method to verify that hypothesis[69]. Essentially, the experiment was repeated taking aliquots at different time-points and plating them as droplets on LB agar after generating a dilution series across several orders of magnitude. Colony formation at distinct 10-fold dilutions allowed estimating total cell viability numbers and establish comparisons across conditions (**Figure *29*A**).

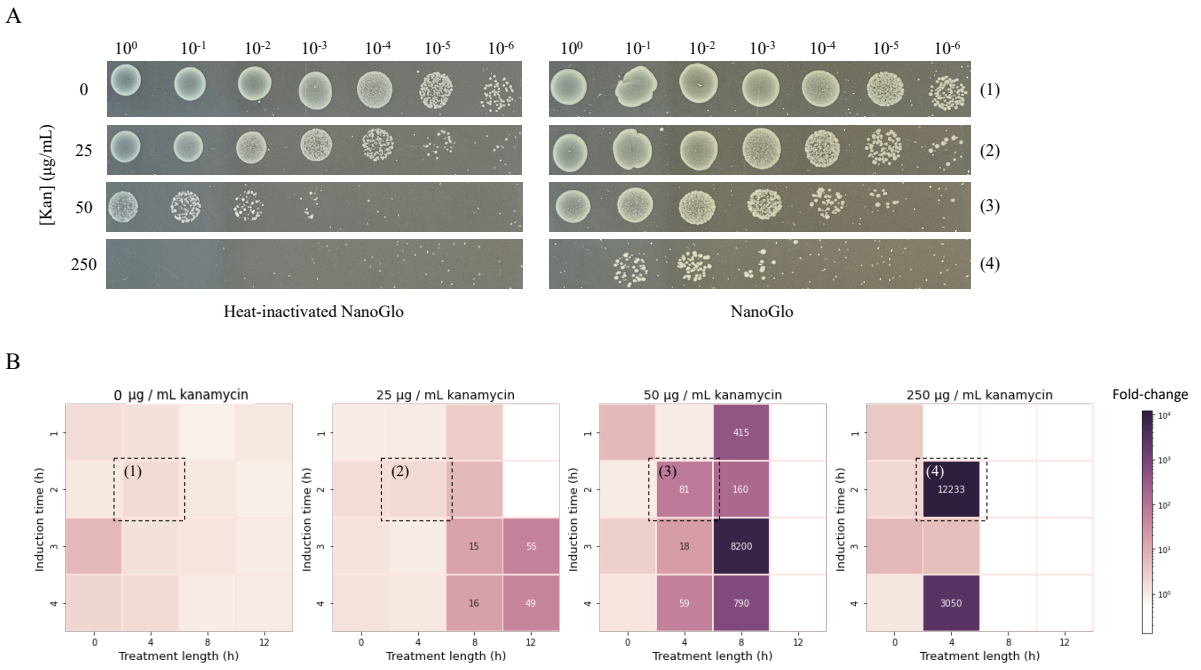


Figure 29 | Bioluminescence delays cell viability loss in kanamycin-treated cultures expressing KanR / sfGFP bicistron. **A** Bacterial viability assay showing culture droplets plated on LB agar after serial 10-fold dilutions for luciferin- (right) and vehicle-treated (left) groups after applying distinct kanamycin selective pressures. **B** Fold-change differences in cell viability between the luciferin and vehicle groups across distinct post-induction times, kanamycin concentrations, and treatment lengths. Fold-change measurements correspond to the average of two biological replicates, represented with a purple-to-salmon colourmap and the numeric value for values ≥ 15. Blank squares represent culture obliteration in both conditions.

As hypothesised, viability differences between groups were considerably larger than those previously registered, confirming that fitness acquisition was underestimated with OD measurements (**Figure *29*B**). These viability differences were not caused by the two different treatments, as demonstrated by cultures grown in the absence of kanamycin, and should thus account for antibiotic selection of bioluminescent bacteria. Generally, lower antibiotic doses required longer treatment lengths to cause smaller survival differences between groups, while higher doses elicited larger differences faster. This was reasonable considering that weaker cytotoxic pressures should take longer to exert a less stringent selection. Moreover, largest viability differences between groups occurred before cultures in both conditions became entirely obliterated (blank squares), suggesting that bioluminescence-mediated resistance to kanamycin was just sufficient to slightly delay cell death compared to the vehicle controls.

Next, the capacity of this selection strategy to rescue bioluminescent bacteria from mixed cultures with non-bioluminescent bacteria was tested. Several of the selection conditions characterised above were used to enrich mixed cultures containing the functional GeNL circuit and the catalytically-inactive GeNL variant (**Figure *25*;** NanoLuc R167A). Despite best efforts, no significant enrichments were achieved using both low and high kanamycin concentrations and distinct treatment lengths (*data not shown*). In general, early recovery points resulted in no selection and later time-points caused a general lack of cell viability. Before discarding these circuits, the selective capacity of tetracycline with the TcR/sfGFP bicistronic unit was assessed as an alternative strategy. However, the bioluminescence advantage was again minor both in terms of bacterial growth and viability.

Based on the previously measured induction of gene expression in CcaS-CcaR in response to luciferin (60-fold), it was expected that bioluminescence-mediated competitive advantages would be noticeable under antibiotic pressures that challenged basal resistance levels. Therefore, it was hypothesised that modifications to the circuit output gene structure might have altered the inducibility of the antibiotic resistance genes.

To validate the bicistronic design, both the function of sfGFP and KanR was assessed under constant darkness or external green light illumination. First, green fluorescence levels were measured by flow cytometry using an empty reporter strain, the original sfGFP reporter, and the KanR/sfGFP bicistron. Fluorescence histograms confirmed that the inducibility of sfGFP in this dual reporter was compromised, with around 46.5% higher background levels in the dark and only 2-fold induction in response to saturating green light (**Figure *30*A**). Then, a bacterial viability assay confirmed that full CcaS-CcaR induction with external green light was

not able to confer any additional kanamycin tolerance to the bicistronic strain across three sub-lethal antibiotic doses (**Figure *30*B**). Hence, it was concluded that the current bicistronic design impaired the light-inducibility of both genes, rendering it unsuitable for any selection strategy.

The reason for the loss of inducibility of CcaS-CcaR output gene in the bicistronic design remained unresolved. On one side, the 5' untranslated region used to control the expression of each gene was identical to the original circuit, which worked reliably independently of the CDS expressed in monocistronic designs (*Section 5.2.2*). On the other side, this effect could not be attributed to sequence-specific context effects because it was observed in both KanR and TcR genes. Still, genetic context is known to influence regulatory regions in often unpredictable ways, and it is possible that the simple presence of a second gene downstream of the leading gene could disrupt the expression dynamics of the transcriptional unit[193,194]. For instance, the second gene could cause structural DNA changes that altered the accessibility to the regulatory element recognised by CcaR in $P_{cpcG2\text{-}172}$, leading to only basal promoter activity. Moreover, it is also very likely that the regulatory mechanisms involved in the translation of multicistronic mRNAs, such as translation coupling effects or the independent recruitment of ribosomes, might interfere with the original circuit expression

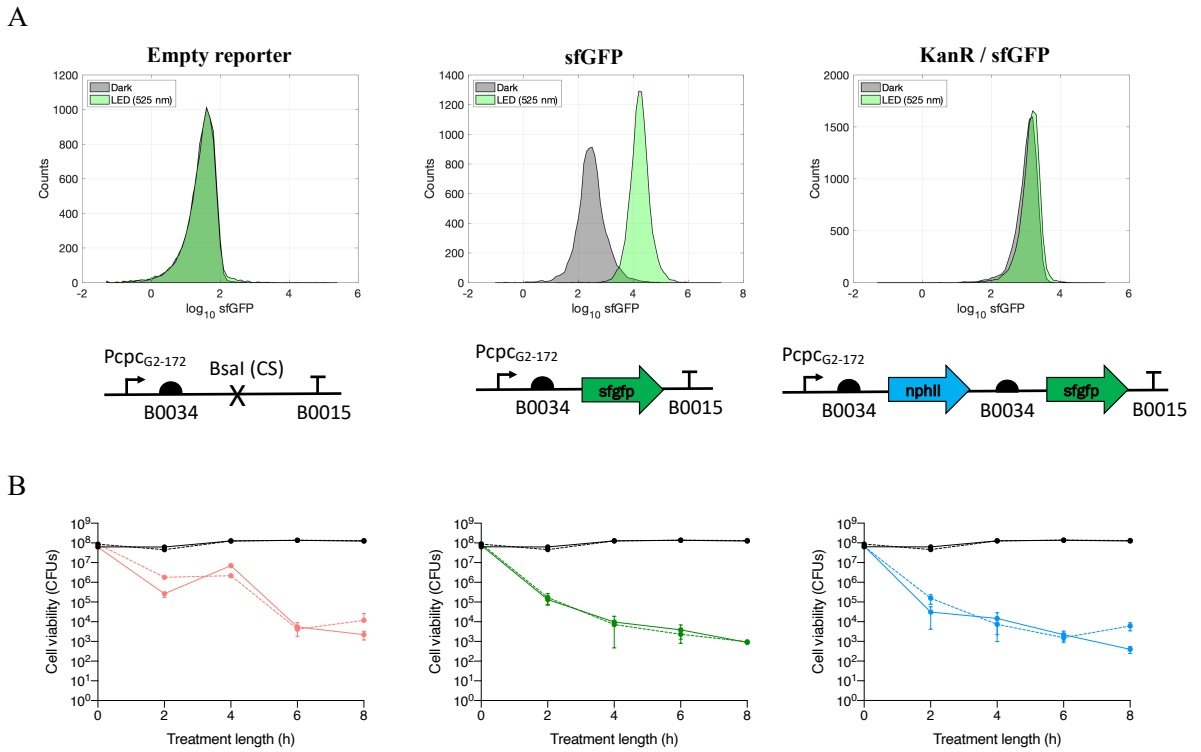


Figure 30 | Compromised light-inducibility of both antibiotic resistance gene and sfGFP in the bicistronic design. **A** Representative flow cytometry histograms showing the fluorescence distribution of 10,000 cells exposed to either constant darkness (grey) or saturating green light (green), expressing an empty reporter (left), the original sfGFP reporter (middle), or the bicistronic reporter (right) (N = 4) **B** Viability counts for cultures expressing the bicistronic reporter strain challenged with 100 µg / mL (red), 200 µg / mL (green), or 300 µg / mL (blue) of kanamycin under constant darkness (dashed) or saturating green light (solid). Individual values correspond to the mean ± SD of two biological replicates.

dynamics[195,196]. In conclusion, it appears that monocistronic expression of two genes fused with either stable or post-translationally cleaved linkers might have constituted a more reliable strategy to generate dual reporters without disrupting CcaS-CcaR expression dynamics[197].

5.2.2 Single reporter genes for bioluminescent selection

5.2.2.1 Protein accumulation dynamics determine antibiotic selection efficacy with CcaS-CcaR

After the failure to modify the structure of CcaS-CcaR output gene for bicistronic expression, selection strategies were engineered using single reporter constructs. To avoid interfering with the gene expression dynamics of this circuit, the coding sequence of sfGFP was replaced by each antibiotic resistance gene using a scarless cloning method. Since the bicistronic units had shown relatively high basal antibiotic resistance, the RBS sequence was systematically varied to optimise the translation efficiency of these genes using the Community RBS Collection from the iGEM Parts Registry (www.parts.igem.org). Five RBS variants were tested with measured translation efficiencies of 100%, 60%, 30%, 7%, and 1% relative to the original sequence.

To assess the influence of these RBS sequences with distinct translation rates on basal antibiotic resistance, the five strains with KanR genes and a non-resistant strain were grown in the dark across a gradient of kanamycin concentrations. Unexpectedly, all strains carrying the KanR gene grew comparably to the non-treated controls, showing complete insensitivity to

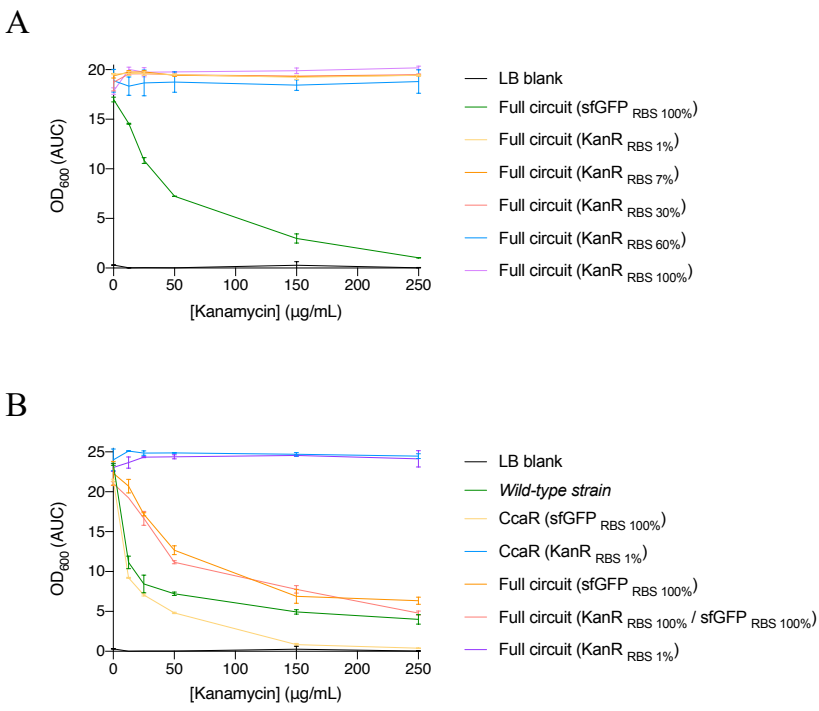


Figure 31 | Dissecting the factors contributing to basal antibiotic resistance in CcaS-CcaR circuit. **A** Impact of translation efficiency on basal kanamycin resistance acquisition in strains expressing full miniCcaS-CcaR with either KanR or sfGFP as output genes under a gradient of kanamycin concentrations. **B** Growth-inhibition caused by a gradient of kanamycin concentrations on distinct bacterial strains expressing either no heterologous genes, the transcriptional components of the circuit with either sfGFP (yellow) or KanR (blue), or the full circuit with either sfGFP (orange), the previously tested KanR / sfGFP bicistronic reporter (red), or KanR (purple) as output genes. Individual values represent the mean ± SD of two biological replicates obtained by integrating the area under the curve (AUC) of 20-hour OD curves.

kanamycin doses up to 5-times the MIC independently of RBS sequence (**Figure *31*A**). In contrast, the non-resistant strain exhibited progressively compromised growth with increasing antibiotic concentrations, confirming the efficacy of the treatment. This finding was particularly surprising given that basal expression of KanR in the bicistronic unit had not conferred complete growth protection against any of the concentrations tested in this experiment.

Further investigations revealed that, even in the absence of CcaS photoreceptor, the transcriptional leakage from CcaR-specific promoter with the weakest RBS sequence was sufficient to confer complete resistance to such high doses of kanamycin (**Figure *31*B**; blue). These results also showed that strains expressing the bicistronic unit engineered in the previous section had comparable basal resistance levels to an identical strain lacking the KanR gene (red and orange). This suggests that KanR expression must be severely compromised in the dual output gene design, which would explain the discrepancy with these last two experiments. Finally, it was discovered that maintaining CcaS plasmid conferred a reproductive advantage under sub-inhibitory kanamycin doses compared to both the *wild-type* strain (green) and the strain expressing only the transcriptional components of the circuit (yellow). Since this plasmid is spectinomycin-selectable and spectinomycin is closely related to aminoglycosides such as kanamycin, it is plausible that its resistance gene, aminoglycoside adenylyl transferase (*aadA*), has some level of cross-reactivity with kanamycin.

Before discovering that the antibiotic resistance levels originating from the bicistronic unit were not representative of the basal activity of these genes in this circuit, it was already expected that their translation rates would need to be adjusted to identify a practical range of enzymatic activities for antibiotic selection using CcaS-CcaR. However, simultaneously decreasing the translation efficiency of KanR gene by 100-fold using the weakest RBS variant and removing the photoreceptor unspecific activation of the circuit failed to sensitise the strain to high kanamycin doses. The CcaR transcriptional module had been engineered to exhibit tightly-regulated activity levels in the absence of CcaS, which this project confirmed to be remarkably stable during bacterial growth (**Figure 9C**). Furthermore, these RBS sequences have been characterised and widely used by the synthetic biology community and should, in principle, be reliable. Therefore, it was unlikely that basal gene expression could explain alone the antibiotic resistance levels observed in the dark.

The last option considered in engineering these selection strategies was that bacterial antibiotic resistance proteins might have evolved to be highly stable to minimise the necessary

gene expression for long-term survival. Accordingly, it was hypothesised that increasing the degradation rate of these genes should reduce protein accumulation and bring antibiotic resistance levels closer to the real-time expression of the resistance genes. For this purpose, the *E. coli* ssrA degradation tag was introduced at the C terminus of each of the six genes. This tag contains a specific sequence recognised by the adaptor protein SspB, which helps attach the targeted peptide to the proteosome complex CIpXP through the last three amino acids (LAA)[66]. Moreover, substrate delivery from the adaptor to the proteosome can be enhanced by inserting residues in-between their respective binding sites[198]. Consequently, a second degradation tag with four extra residues (LAA+4) was also tested to reduce the basal resistance of these strains.

The effect of these degradation tags on basal antibiotic resistance was studied using the full luciferin-inducible circuit containing the weakest RBS variant. Strains expressing each antibiotic resistance gene with or without each degradation tag were grown in the dark and treated with either a vehicle solution or their respective antibiotics at the MIC. All strains expressing tagged genes exhibited varying degrees of antibiotic sensitivity, independently of the degradation tag used (**Figure *32***). The original LAA tag was generally more effective at reducing protein activity compared to LAA+4, even though the latter sequence was designed to enhance protein degradation. Notably, these results were achieved with a moderate antibiotic pressure and regardless of the antibiotic target and the resistance mechanism involved. Therefore, it was confirmed that the stability of these antibiotic resistance proteins was hindering the conditional selection of these strains by causing resistance accumulation over time, which protected bacteria against high antibiotic doses despite low basal expression levels.

In agreement with previous data, antibiotics targeting protein synthesis were the most effective at reducing unspecific bacterial growth, except for erythromycin. Kanamycin, tetracycline, and trimethoprim fully suppressed growth in the dark for strains expressing LAA-tagged resistance genes. Among these three resistance mechanisms, KanR displayed the highest basal resistance in the absence of a degradation tag. In contrast, lowering the translation efficiency of TcR and TmpR was already sufficient to sensitise these strains to their antibiotics, which the degradation tags potentiated further.

This experiment also revealed that the strain used in this project had a partial innate resistance to erythromycin, partially explaining the low efficacy of this treatment despite inhibiting translation. After investigating the genome of the strain, a loss-of-function mutation in the outer membrane porin gene *tonA* was identified, which has been reported to be an important entry route for erythromycin into *E. coli* and might explain the reduced treatment

efficacy[199]. Finally, antibiotics targeting cell processes other than protein synthesis were less efficient at controlling bacterial growth in the dark. Nevertheless, the degradation tags managed to reduce bacterial fitness, causing delayed growth for carbenicillin and an apparent loss of cell viability for zeocin at later culture stages.

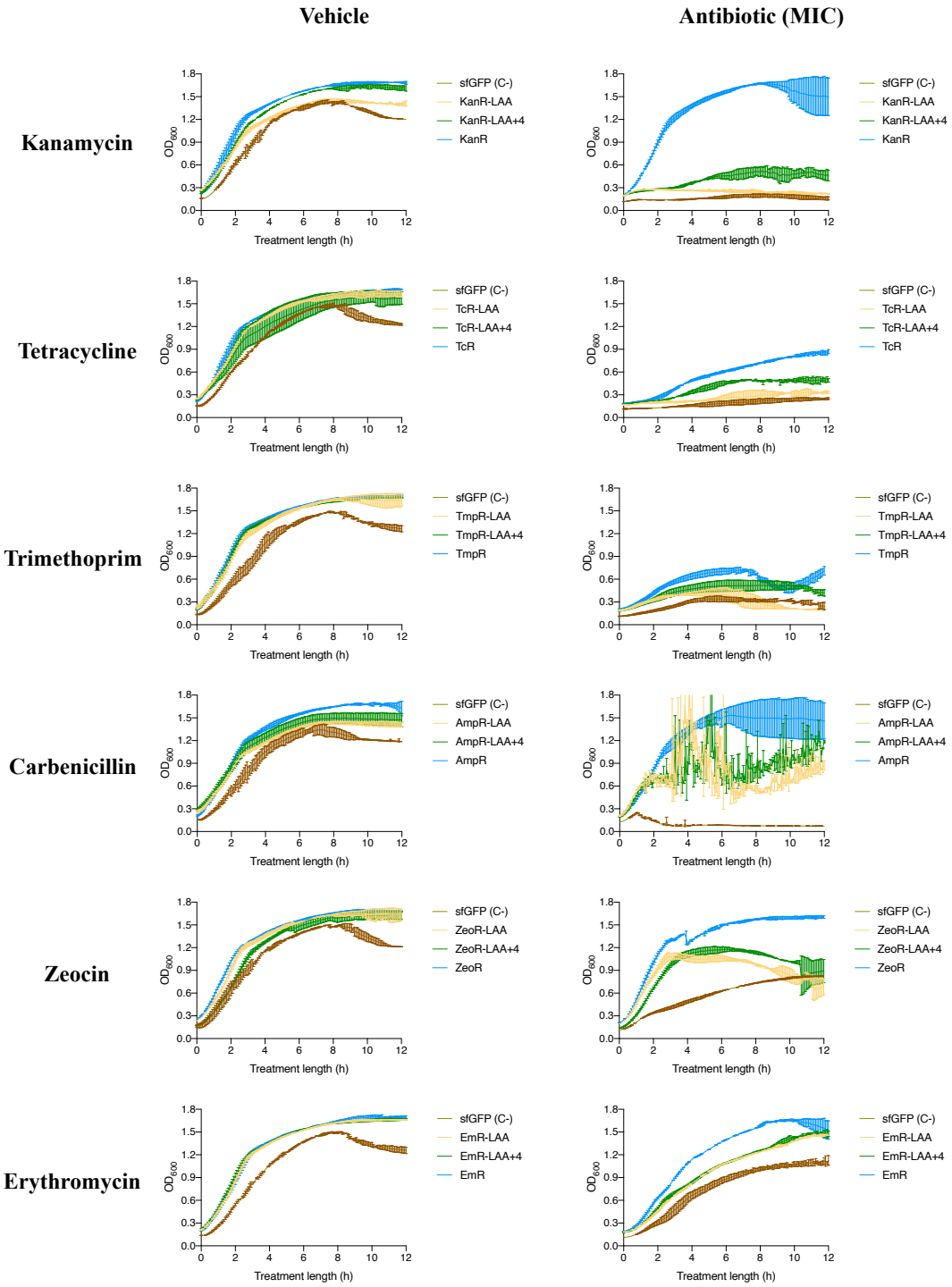


Figure 32 | Protein accumulation enhances the unspecific acquisition of antibiotic resistance in CcaS-CcaR circuit. Bacterial growth of strains expressing the full luciferin-inducible CcaS-CcaR circuit with either sfGFP (brown), or the antibiotic resistance gene either untagged (blue), or tagged with LAA (yellow) or LAA+4 (green). Cultures were treated with PBS as a vehicle solution (left) or the minimal inhibitory concentration of each antibiotic (right) at late-exponential phase to ensure maximal intracellular dilution. Individual values correspond to the mean ± SD of two biological replicates.

To evaluate the fitness acquisition capacity of the best performing circuits in response to bioluminescence, the growth of their respective strains was monitored in the dark with either 1x NanoGlo (luciferin) or 1x heat-inactivated NanoGlo (vehicle) and a gradient of antibiotic concentrations. For strains expressing KanR with a LAA degradation tag, luciferin induction generated slight growth differences between conditions under sub-inhibitory doses of kanamycin (**Figure *33*A**). Although the bioluminescence-mediated growth advantage was modest, it should be noted that actual fitness differences may be larger in terms of bacterial viability, as previously demonstrated. In contrast, TcR and TmpR genes tagged with degradation sequences failed to confer observable protection against any of the antibiotic pressures tested upon luciferin induction (**Figure *33*B** and **C**). Finally, the untagged TmpR gene enabled unimpeded growth of bioluminescent cultures across all trimethoprim concentrations tested, at the expense of relatively high basal resistance (**Figure *33*C**; right).

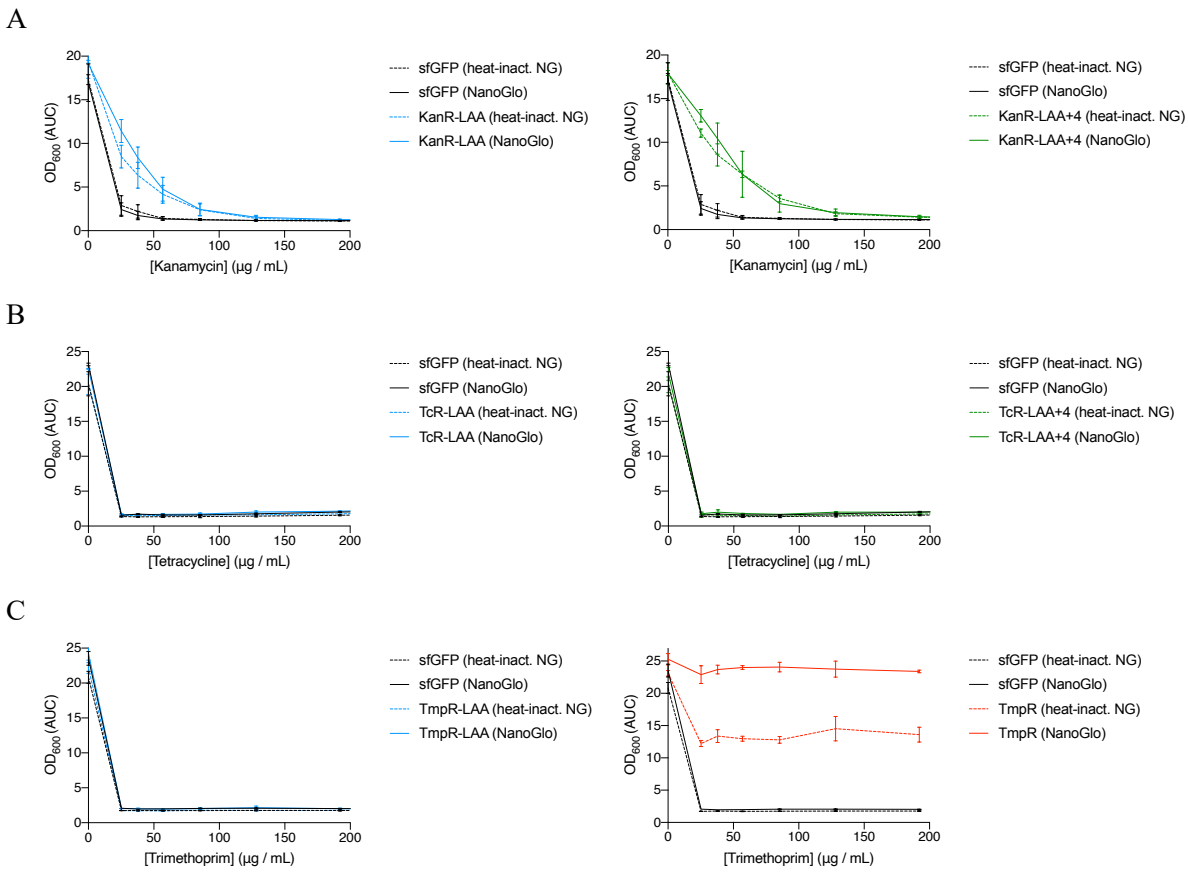


Figure 33 | Antibiotic resistance translation rate is too low to induce fitness advantages with a LAA degradation tag. Growth advantages induced by 1x NanoGlo under a gradient of antibiotic concentrations in strains expressing the luciferin-inducible CcaS-CcaR circuit with distinct antibiotic resistance genes using a relatively weak RBS sequence (BBa_B0033) and two distinct LAA degradation tags: KanR (**A**), TcR (**B**), and TmpR (**C**). Individual values represent the mean ± SD of two biological replicates obtained by integrating the area under the curve (AUC) of 20-hour OD curves.

Overall, these results indicated that while degradation tags were necessary to suppress the unspecific growth caused by basal resistance accumulation, the translation rate of the weakest RBS variant might be too low to rescue these strains using a bioluminescence pulse. It seems that the production and degradation rates of these antibiotic resistance genes need to be simultaneously tuned to identify a range of activities that it is suitable for selecting bioluminescent bacteria using CcaS-CcaR. Therefore, it would be convenient to test the control over antibiotic resistance provided by these three genes using the other RBS variants studied earlier in this project in combination with the LAA degradation tag.

5.2.2.2 FACS-based screening of bioluminescent proteins using CcaS-CcaR

In parallel with the antibiotic selection strategies, the luciferin-inducible CcaS-CcaR circuit containing the fluorescent reporter sfGFP with the original RBS sequence was used in an attempt to develop a FACS-based selection protocol for bioluminescent bacteria. As explained in *Section 5.1*, this powerful technique can perform high-throughput screening and sorting of single cells based on their fluorescence levels, and it has partially influenced the greater advances achieved in fluorescent protein evolution compared to their bioluminescent counterparts. Since the repurposed CcaS-CcaR circuit already converted bioluminescent signals into fluorescence gene expression, it was hypothesised that it could potentially allow access to this screening technology without significant further engineering. Therefore, although this was not the primary objective of the project, it was still worth investigating as it could provide an alternative means of selection that was preferrable to existing methodologies in terms of efficiency, simplicity, and accessibility.

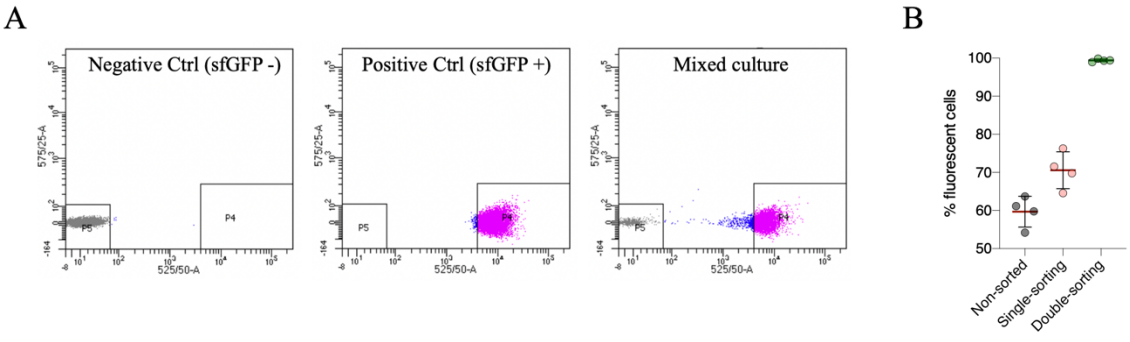


Figure 34 | Double sorting of bacterial cultures is required to achieve specific fluorescent selection using FACS. **A** Pure cultures constitutively expressing either a non-fluorescent protein (left) or sfGFP (middle) were used to establish selection parameters to exclusively sort fluorescent bacteria in mixed cultures (right). Individual values correspond to the fluorescence signal of individual cells recorded with a 525/50 nm light filter. **B** Percentage of fluorescent bacteria in initial non-sorted mixed cultures (black), single-sorted samples (red), and double-sorted samples (green). Bars correspond to the mean ± SD of four biological replicates.

Initially, two strains constitutively expressing either sfGFP or RLuc8.6-535 as a non-fluorescent control were used to identify FACS settings that allowed sorting fluorescent *E. coli* cells with high specificity. Cultures expressing each gene were grown to similar optical densities and sequentially analysed to define fluorescent parameters discriminating non-fluorescent bacteria (**Figure *34*A**). Then, the efficiency of distinct sorting strategies was assessed using mixed cultures containing approximately equal proportions of each population. All sorting attempts failed consistently at specifically recovering fluorescent cells with 25-35% of false positives depending on the settings (**Figure *34*B**; red).

Given their distinctive fluorescent profiles, it was hypothesised that double events containing a positive and a negative bacteria in the same droplet might account for the unspecific recovery rates observed. Since it was not possible to exclude this type of events by bacterial size, it was hypothesised that the probability of a non-fluorescent cell to evade selection in two consecutive rounds of sorting should be relatively low. In agreement with this hypothesis, double-sorted samples displayed > 99% purity on average, independently of the settings used (**Figure *34*B**; green).

After optimising the FACS conditions for bacterial sorting, these were used to assess whether bacteria could also be selected based on their bioluminescence activity using the luciferin-inducible CcaS-CcaR circuit. To test this, two strains expressing miniCcaS#10 fused to either functional GeNL or the catalytically-inactive GeNL mutant were grown until optimal activation conditions and induced with 1x NanoGlo separately and after mixing at approximately equal proportions. All cultures were analysed to validate the luciferin induction of fluorescence and fluorescence plots were used to design distinct fluorescent thresholds for cell sorting (**Figure *35***). However, none of the selection thresholds used succeeded in either enriching mixed cultures with bioluminescent bacteria or recovering a reasonable proportion of viable cells after double sorting (**Table 9**).

Table 9 | Double sorting of mixed bioluminescent populations using FACS and CcaS-CcaR

Group	(+)	(–)	(%)	Sorted cells	Recovered cells
sfGFP (non-sorted)	122	91	57.28	-	-
sfGFP (sorted)	5,529	0	100.00	6,100	5,529 (90.64%)
GeNL (non-sorted)	81	15	84.38	-	-
GeNL (top 0.1%)	4	2	66.67	2,750	6 (0.22%)
GeNL (top 0.4%)	84	12	87.50	9,000	114 (1.27%)
GeNL (top 1.6%)	75	9	89.29	10,000	89 (0.89%)

Two main limitations challenged the isolation of bioluminescent bacteria using FACS and CcaS-CcaR. On one hand, the intrapopulation variability caused by CcaS-CcaR gene expression results in overlapping fluorescence distributions between its basal and induced states, which makes it impossible to establish selection conditions without sorting non-bioluminescent bacteria (**Figure *35***). Most likely, this occurs because basal gene expression in the current circuit design allows certain bacteria to accumulate sufficient fluorescence to be selected even in the most stringent conditions. On the other hand, the low recovery numbers obtained after double-sorting these strains caused most bacteria to be lost independently of their bioluminescent activities. This problem might be caused by two factors or the combination of their effects: the toxic side effect of NanoGlo probably makes a larger fraction of bacteria more susceptible to lose viability upon double sorting; and electronic noise at the tail of the fluorescence distribution causing unreal events to be counted as sorted cells, as indicated by the recovery rate of the distinct gates (**Table 9**).

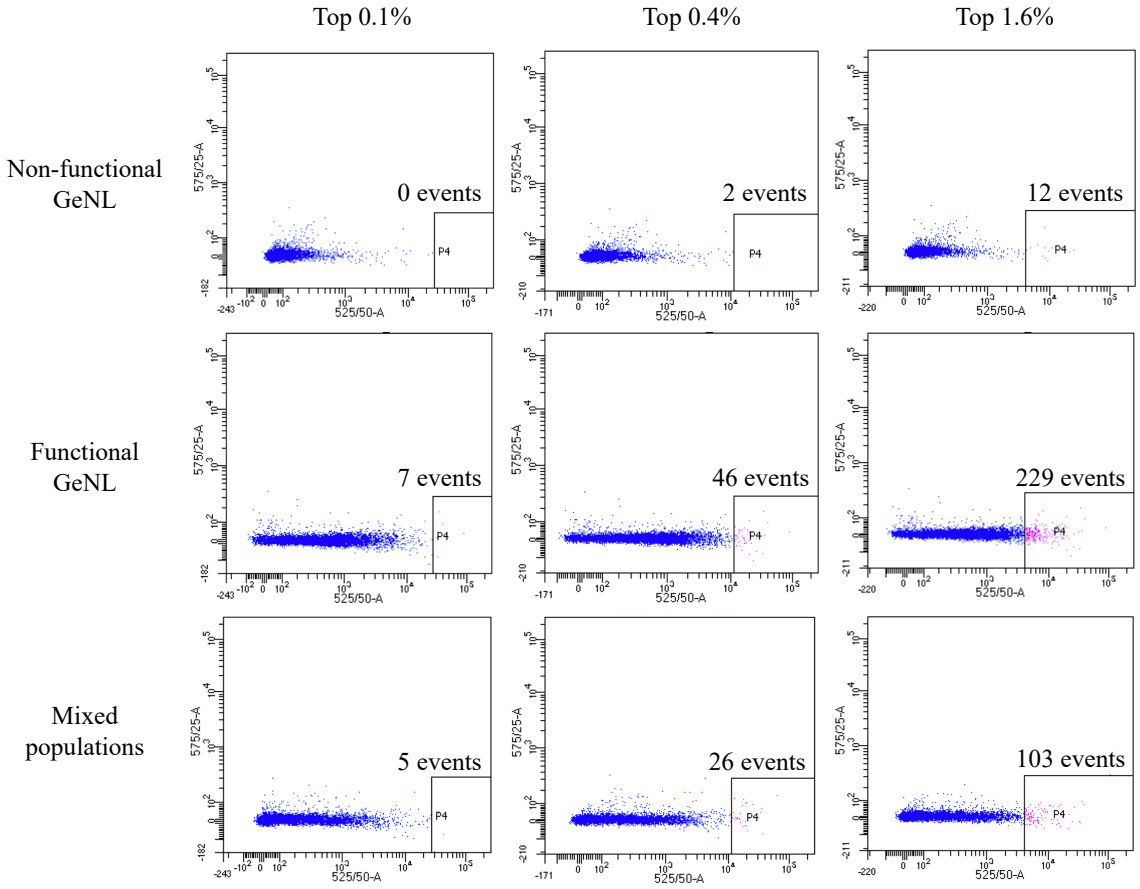


Figure 35 | Overlapping fluorescence distributions in CcaS-CcaR impede FACS selection of bioluminescent bacteria. The luciferin-inducible CcaS-CcaR circuit containing either a functional or a non-functional GeNL variant with the sfGFP reporter gene under a relatively strong RBS sequence (BBa_B0034) was used to translate bioluminescent signals into fluorescence levels. Failure to efficiently select bioluminescent bacteria using FACS was primarily attributed to an overlap in the fluorescence distributions produced by the basal (non-functional GeNL) and induced (functional GeNL) states of the circuit. Individual values correspond to the fluorescence signal of individual cells recorded with a 525/50 nm light filter.

Multiple protocol changes were tested to minimise the loss of bacterial viability during sorting, including collecting cells into distinct buffers, at different temperatures and sorting speeds, and performing a recovery step in rich medium before platting the cells, but it all was of no avail. In any case, this problem could be mitigated better, if not completely, by simultaneously tunning the production and degradation rates of sfGFP to avoid unspecific signal accumulation in negative cells. This would allow less strict gating upon sorting, being able to recover a larger fraction of the sample while reducing the proportion of false events produced by electronic noise. Nevertheless, this selection strategy was put on hold given its unsatisfactory results and the amount of troubleshooting required to potentially make it work.

5.3 Conclusions

In the last stage of this project, it was aimed to couple the activity of distinct selectable genes to the luciferin-inducible CcaS-CcaR circuit to finally engineer an *in vivo* selection platform for bioluminescent bacteria. For this purpose, antibiotic resistance genes were preferentially chosen as they allowed the systematic targeting of cell processes for distinct growth- and survival-based selection strategies while conferring external control over the selective pressure.

Initially, a bicistronic unit expressing both the antibiotic resistance gene and a fluorescent reporter was created to potentially enable the dual selection of bacteria using antibiotics or FACS to suit a wider range of experimental needs. As predicted early in this project, antibiotics inhibiting protein synthesis were more efficient at suppressing unspecific bacterial growth caused by the accumulation of basal circuit activity. However, the bicistronic design built impaired the light-inducibility of both genes, causing insufficient resistance acquisition in response to bioluminescence for practical selection purposes.

These antibiotic resistance genes were tested again in a single reporter format, revealing that their basal expression was sufficient to confer complete antibiotic insensitivity. This effect persisted even after decreasing the translation efficiency of these genes by 100-fold and removing the photoreceptor unspecific activation of the circuit. It was found that protein accumulation allowed the acquisition of resistance despite minimal gene expression levels, which was solved by increasing the turnover rate of these genes with a LAA degradation tag. Nevertheless, the selection of these strains using bioluminescence was relatively inefficient because the translation rate of these genes had been excessively reduced to allow significant fitness acquisition upon induction. In parallel, the fluorescent reporter was used in an attempt to develop a FACS-based selection protocol for bioluminescent bacteria with the initially repurposed CcaS-CcaR circuit. Similarly, the basal accumulation of fluorescence in non-bioluminescent bacteria allowed these cells to be selected even in the most stringent conditions.

In conclusion, it appears that regardless of the selection strategy followed, it is essential to simultaneously tune the production and degradation rates of the selection genes to reduce the unspecific accumulation of gene expression caused by CcaS-CcaR while retaining enough circuit production capacity to specifically select bacteria upon bioluminescence induction. Unfortunately, both project directions had to be put on hold for the writing of this document, but the current state of the project does not appear to be far from completion.

6. Concluding remarks

Personal insights and scientific discussion of the hurdles and catalysts that have influenced the engineering process of this *in vivo* bioluminescence selection platform. Alongside, the current state of the technology is summarised together with the final optimisations needed to complete its development, and its future utility and limitations.

6.1 Thesis conclusion

The work presented in this thesis expounds the research and development stages that led to the prototyping of an *in vivo* bioluminescence selection platform for *E. coli*. The current technology consists of a repurposed light-sensing circuit that transduces intracellular bioluminescence activity into the expression of distinct antibiotic resistance genes. Although further circuit optimisation is still needed, the current version already confers modest but detectable growth and survival advantages to glowing cultures compared to those restricted of light under two separate antibiotic pressures. Overall, this document constitutes a manual for the informed utilisation of this platform and contains a wide range of valuable insights for the repurposing of any of the light-emitting and light-sensing technologies tested along the project.

6.2 Towards a more comprehensive synthetic circuit utilisation

The first aim of this project was to identify light-controlled transcriptional circuits that could be repurposed to transduce bioluminescence inside bacterial cells. Despite the current emphasis in scientific reproducibility and standardisation of biological parts and circuits in synthetic biology, this proved to be challenging for two different reasons.

First, poor description and annotation of genetic sequences complicated the reconstruction of unavailable circuits from synthetic DNA fragments. For instance, it was found that pDusk/pDawn contains the RBS of FixJ embedded in the last codons of YF1, which was accidentally modified in our assembled version when codon-optimising the bicistron. This probably altered the translation rate of FixJ and, thereby, the ratio of YF1/FixJ protein levels, significantly reducing the dynamic range of the circuit. Additionally, it was not possible to reconstitute the light-regulation of four T7RNAP-based circuits, despite using intact published sequences in a varied collection of vector backbones. Since all these circuits failed to work in our hands, it is most likely that some unidentified critical elements of the genetic, host, and/or environmental context in which they were developed were not accurately reproduced. This highlights the lack of knowledge on the factors influencing the signalling dynamics of these circuits, which is connected and aggravated by the following problem.

Second, the limited public availability of experimental considerations for the proper utilisation of these circuits. In my personal experience, the reproducibility and utility of these circuits for novel applications would greatly benefit from more comprehensive methodological descriptions in their reports, whether the specific molecular mechanisms governing their

signalling dynamics are known or not. It is more helpful to understand why things are made in a specific way than knowing how things are made to the greatest of details in one particular case. In my opinion, the developer of a new technology should be aware that highly-specific experimental conditions might not be suitable to other researchers or constrict their scientific creativity. Hence, methodological descriptions should help future users understand how to adapt these protocols to meet their own needs if possible. This could be easily achieved without necessarily having longer method sections by providing generic guidelines for circuit utilisation and warnings about pitfalls and limitations (**Table 10**).

Table 10 | Guidelines and warnings for comprehensive CcaS-CcaR circuit utilisation

Warnings
⚠ Basal photoreceptor activation and low reversal kinetics in the dark lead to unspecific circuit activity accumulation in slow-dividing cultures in the absence of red light.
⚠ Cultures start accumulating unspecific signalling activity after exponential phase exit (OD ~ 0.2-0.3), which progressively diminishes the circuit's dynamic range and quality of results
⚠ N-terminal association of CcaS to the membrane is required for ground-state stability
Guidelines
☐ Dynamic range is directly proportional to the growth rate of bacteria
☐ Cell growth dilutes active CcaS molecules via cell division and *de novo* synthesis of the photoreceptor in the inactive state.
☐ Prepare seeding aliquots with previously "*washed-out*" bacterial cells by growing cultures to late-exponential phase from very low seeding concentration (OD = 10^{-4}-10^{-5}) or maintaining them in fast division conditions in a turbidostat overnight (OD < 0.1).
☐ Start cultures with seeding aliquots at OD = 10^{-3} for 4-6-hour experiment or OD = 10^{-4} for up to 8-hour experiment (always preferable to measure exponential phase exit from distinct seeding concentrations to decide on experimental conditions).
☐ Optimal circuit induction occurs at near-maximal growth rates, with a 2-hour window around the exponential phase exit (OD ~ 0.2).
☐ miniCcaS#10 variant has enhanced ground-state stability and a wider range of light intensities that differentially activate it
☐ CcaS fusions are advised at the C-terminus for globular proteins and the N-terminus for transmembrane proteins (although it is a sensitive fusion site).
☐ Clone luciferases to the C-terminus of miniCcaS#10 with a 6-12 residue linker peptide for optimal bioluminescence transduction (shorter intramolecular distance to the chromophore).

In certain cases, important methodological discrepancies were found in the literature that further confounded the utilisation of these circuits. The clearest example were the T7RNAP-based circuits which despite all sharing a practically identical sensor design, one research group used them in strict exponential phase conditions while the other group extended their protocols long into the stationary phase. Even more concerning was the lack of consensus regarding the thermostability of their light-sensing domains. Some articles claim that VVD/Magnet domains have low thermostability at 37ºC and require long preincubations at 28ºC to fold properly before utilisation[88,89,93,94]. However, several other publications reporting technologies based on these domains ignored this requirement, including both articles for the T7RNAP-based sensors[49,52,90–92,130].

Other times, the omission of circuit regulatory details or experimental explanations concealed important circuit limitations. For instance, no data is available about CcaS-CcaR signalling dynamics in the absence of red light or outside steady-state exponential phase conditions, nor explanations are provided on the need for terminating experiments at the reported ODs or preparing seeding aliquots the way it is reported. These methodological constrains mitigate the photoreceptor unspecific activity and its slow-reversal kinetics and leverage the dilutive effect of bacterial growth on active CcaS molecules for optimal circuit performance. Another example is the lack of explanation given for consistently delivering light pulses to pDawn at an OD ~ 0.4 and its apparent stability at late-stationary phase. This circuit is irresponsive to blue light until that culture density and its signal production capacity progressively dissipates to negligible levels as cells enter stationary phase. In these regards, this project has contributed to clarify these and other important aspects governing the signalling dynamics of these two circuits for a more informed and flexible utilisation.

6.3 Tool overview and upcoming work

The tool presented in these pages was developed by repurposing an optimised version of the cyanobacterial two-component system CcaS-CcaR. After precisely characterising its optimal light-induction conditions during bacterial growth, its photoreceptor was engineered to sense bioluminescence emissions produced intracellularly. Despite best efforts, it was not possible to obtain bioluminescence expression levels bright enough to distantly activate CcaS photoreceptor with some of the best performing luciferases available in the literature. Instead, photon production had to be co-localised to the immediate environment of the chromophore for bioluminescence transduction to occur. As a result, the bioluminescence-sensing capacity

of this circuit is restricted intracellularly and discriminates individual bacterial light emissions, which is a pre-requisite for efficient *in vivo* bioluminescence selection.

Multiple aspects were advantageous about using CcaS for engineering a bioluminescence transduction platform:

First, it was relatively easy to identify a tolerant fusion site that could accommodate distinct luciferase proteins without altering the signalling properties of the photoreceptor. Despite being the less intuitive position, the C-terminal end, located at the opposite site of the chromophore, was sufficiently exposed to the solvent to incorporate GeNL and RLuc8.6-535 while still allowing efficient bioluminescence activation. However, this was only possible due to the availability of an internally-truncated photoreceptor variant named miniCcaS#10, because the intramolecular distance that separates both ends in full-length CcaS is too far apart for the chromophore to detect bioluminescence production. This was corroborated when optimising the linker peptide length, which revealed that longer sequences increased the bioluminescence transduction efficiency for the same luciferase activity. Instead, the N-terminal end was inaccessible due to its insertion into the membrane, which stabilises the photoreceptor inactive-state; and circularly-permuting CcaS at internal non-functional domains interfered with the signal transduction from the light-sensing domain to the signalling domain.

Second, this photoreceptor fusion topology allowed inducing gene expression at almost comparable levels to saturating LEDs irradiances for both luciferases, while offering multiple control variables to externally tune circuit activity across its whole dynamic range (~ 120-fold). In fact, bioluminescence transduction in optimal conditions was sensitive enough to detect light pulses obtained by diluting up to 100-times the recommended luciferin concentration for the dimmest of the two luciferases. Thus, it is expected that this tool will be able to evolve relatively dim luciferase templates, and also detect small incremental gains in bioluminescence activity for bright proteins by decreasing luciferin concentration. Potentially, it might also allow favouring the selection of specific enzymatic properties, such as catalytic efficiency or turnover rate, by using suboptimal induction conditions and providing shorter induction lengths, respectively.

Third, the absorption properties of PCB in CcaS light-sensing domain endow this photoreceptor with extended spectral compatibility. At least in the current fusion design, CcaS was able to sense blue and green light-emitting proteins with comparable efficiency, which comprise the most occurrent bioluminescent wavelengths in nature. Arguably, yellow

emissions might be similarly effective at activating the photoreceptor as estimated from its absorption spectrum, although this hypothesis was not tested in this study. Furthermore, PCB excited state absorbs light at orange and red wavelengths. Therefore, this tool should theoretically be spectrally compatible with any bioluminescent reaction as long as fitness acquisition is coupled to the right CcaS-CcaR signalling trajectory.

Engineering an *in vivo* selection strategy was more complicated, although significant advances were achieved for multiple selection mechanism, which are expected to be completed after the submission of this thesis. Antibiotic resistance genes were preferentially chosen for this task, as they allow the systematic targeting of cell processes for distinct growth- and survival-based selection strategies while conferring external control over the selective pressure. Additionally, a fluorescence reporter gene was also used to develop a FACS-based selection protocol for bioluminescent bacteria.

The main obstacle faced when trying to differentiate bioluminescent bacteria with any of these genes was the accumulation of basal circuit expression. Regardless of the selection strategy, non-bioluminescent bacteria were able to become sufficiently similar at the phenotype level to bioluminescent bacteria to impede antibiotic or fluorescent selection. This problem was partially solved by increasing the degradation rate of these genes with a ssrA degradation tag. Still, only those antibiotics inhibiting protein synthesis, and thereby blocking basal expression in non-bioluminescent bacteria, were able to supress unspecific fitness acquisition. Unfortunately, the translation efficiency of these genes had been reduced by approximately 100-fold while trying to identify the origin of this phenotypic inconsistency, which complicated specific fitness acquisition in response to bioluminescence. Nevertheless, it was still sufficient to modestly differentiate bioluminescent bacteria based on growth and survival.

The successful antibiotic selection mechanisms identified in this project were kanamycin, tetracycline, and trimethoprim. While they all suppress protein synthesis, they target different molecular processes, cause bactericidal or bacteriostatic effects, and are overcome by resistance mechanisms that are enzymatically very distinct: antibiotic inactivation, efflux, and desensitisation respectively. Therefore, it is expected that they will constitute a versatile group of selection strategies to meet different needs in bioluminescence engineering once specific fitness acquisition is fully optimised. For instance, the potent bactericidal effect of kanamycin could be implemented as a filtering mechanism for high-throughput screening of bioluminescent proteins, while the bacteriostatic effect of tetracycline or trimethoprim might be more appropriate for continuous evolution experiments.

The current state of the technology shows that the production and degradation rate of these genes need to be simultaneously tuned to identify conditions in which circuit expression confers minimal selective advantages while its bioluminescence induction can phenotypically differentiate bacteria for their selection. As a starting point, the RBS sequences already available in the laboratory will be systematically tested to increase the production rate of each antibiotic resistance gene and the fluorescence reporter gene in combination with the ssrA tag. It is expected that current fitness differences will become greater by increasing the translation efficiency of these genes while retaining the degradation tag to suppress unspecific fitness acquisition.

A possible setback that might be encountered when trying to rescue bioluminescent bacteria after increasing the translation efficiency of these genes is that the degradation efficiency of the consensus ssrA tag is too high to ensure specific fitness persistence until selection has been effectuated. If that were the case, weaker degradation tags variants exist that could be used to identify a range of activities that it is suitable for selection. Alternatively, the output promoter of CcaS-CcaR could be transferred into a lower copy-number plasmid or integrated in the genome to achieve finer control over gene expression. Right now, this promoter is expressed from a ColE1 vector backbone, which has been quantified at 25-30 copies per cell[200] and even 50-70 copies per cell[201]. Perhaps fewer gene copies would remove the need for using a degradation tag to suppress basal fitness acquisition, and facilitate tunning this circuit function.

As soon as these variables are optimised to maximise the phenotypic differences between a bioluminescent strain from a non-bioluminescent strain using any of these selection strategies, a selection protocol will be developed using bioluminescent mutant libraries. Multiple GeNL mutant libraries of distinct bioluminescent diversity were constructed for this purpose (not presented in this thesis) using a commercial error-prone PCR kit. Therefore, all necessary materials are ready for testing distinct selection protocols with antibiotics and/or FACS and assessing their enrichment capacity until identifying optimal selection conditions. These final experiments will assist the development of a manual for the comprehensive utilisation of this tool under distinct experimental requirements. If time allows, an attempt will also be made to evolve a luciferase with promising technological potential to showcase the utility of this tool.

6.4 Future perspectives and limitations

As mentioned in the introduction, the public availability of this *in vivo* bioluminescence selection platform once fully developed should allow screening millions of luciferase mutants by simply growing cells in the presence of luciferin and an antibiotic. This is expected to accelerate bioluminescence engineering in multiple ways: (1) speeding up the engineering process; (2) removing economic barriers that only justify the involvement of high-budget or specialised laboratories; (3) allowing more complex genotype sampling strategies and higher diversity of screened mutants in each evolution round. Collectively, this should result into a more active and diverse bioluminescence engineering community with higher exploration capacity of luciferase evolution landscapes.

In the future, this tool could be combined with an *in vivo* mutagenesis system to fully-automate the directed evolution pipeline of bioluminescent proteins using bacterial cells and enable continuous evolution experiments[165]. Essentially, these systems consist of engineered molecular machinery that can be targeted to specific genetic sequences inside cells to increase their mutation rate. The most appropriate for *E. coli* are MutaT7 systems, which consists of a T7RNAP fused to a nucleobase deaminase that modifies the genetic sequence of interest every time it transcribes it[202]. Additionally, this selection platform could be coupled to genetic logic gates for the development of bioluminescence biosensors or more complex protein engineering. These circuits only allow a certain outcome to occur when certain conditions are fulfilled, such as the presence and/or absence of multiple stimuli[203]. This would permit increasing the complexity of the selection algorithm in the current selection platform and confer fitness only when bioluminescence is produced under specific conditions.

The main limitation of our selection platform is its susceptibility to the apparition of cheating phenotypes via the unspecific acquisition of antibiotic resistance. For instance, this can occur by remodelling the outer membrane lipid and protein composition to alter its permeability, restricting or decreasing antibiotic transport into the cytoplasm[204]. Another evasion mechanism is the emergent expression of fitness that results from transient gene expression variation, such as stochastic transcriptional and translational bursting[205]. Nevertheless, these evasion phenomena are usually adaptative responses that require gradual exposure to an environmental pressure and should only be concerning for continuous evolution experiments, as they are unlikely to arise in staged selection rounds.

Finally, the utility of directed evolution methods might be reconsidered once greater understanding of protein structure-function relationship becomes available for the efficient rational design of luciferases. Artificial intelligence might help us achieve this faster, but current bioluminescence datasets still appear to be insufficient to replace directed evolution approaches. During the writing of this document, a report was published that had used deep learning for the *de novo* design of luciferases[206]. This approach allowed the computational design of artificial small, monomeric luciferases with affinity for specified luciferin substrates, native and synthetic, that could be optimally expressed in *E. coli* and human cells. While these proteins showed remarkable enzymatic properties, they still required substrate-binding pocket optimisation via site-saturation mutagenesis using large, fully-randomised libraries at multiple key residues to achieve comparable catalytic efficiencies to native luciferases.

Most likely, computer-aided design will soon make directed evolution methods less central in protein engineering, which is a highly-desirable scenario. However, it is my opinion that directed evolution will not become redundant but act synergistically with *de novo* protein design, as they allow exploring the genotype space in different but complementary ways. Artificial intelligence is able to extract complex information patterns from large datasets to extrapolate similar outcomes, which allows the discovery of remote fitness peaks in an evolutionary landscape[207]. These unknown genotype regions would otherwise be too far to reach using random or site-saturation mutagenesis approaches, perhaps even impossible due to the sequence constrains imposed by the evolution history of native proteins. Still, deep learning is limited by its exposure to real or predicted data, but might not be able to precisely determine the effects of small variations in a completely novel landscape it has just predicted. Instead, directed evolution allows the small scale exploration of local fitness peaks for sequence optimisation without any knowledge requirements. Therefore, directed evolution can assist deep learning approaches in two ways: optimising predicted designs and providing real insights into predicted landscapes to further train these models and expand their understanding beyond the natural boundaries of evolution.

Bibliography

1. Lee, J. Bioluminescence: the First 3000 Years (Review). *Journal of Siberian Federal University. Biology* **1**, 194–205 (2008).
2. Shimomura, O. *Bioluminescence: Chemical principles and methods*. *Bioluminescence: Chemical Principles and Methods* (World Scientific Publishing Co. Pte. Ltd, 2006). doi:10.1142/6102.
3. Haddock, S. H. D., Moline, M. A. & Case, J. F. Bioluminescence in the Sea. *Ann Rev Mar Sci* **2**, 443–493 (2010).
4. Fan, F. & Wood, K. V. Bioluminescent assays for high-throughput screening. *Assay Drug Dev Technol* **5**, 127–136 (2007).
5. Ura, S., Ueda, H., Kazami, J., Kawano, G. & Nagamune, T. Single cell reporter assay using cell surface displayed Vargula luciferase. *J Biosci Bioeng* **92**, 575–579 (2001).
6. Hickson, J. *et al.* Noninvasive molecular imaging of apoptosis in vivo using a modified firefly luciferase substrate, Z-DEVD-aminoluciferin. *Cell Death Differ* **17**, 1003–1010 (2010).
7. Wu, C. *et al.* Preparation of biotinylated Cypridina luciferase and its use in bioluminescent enzyme immunoassay. *Anal Chem* **79**, 1634–1638 (2007).
8. Shama, G. & Malik, D. J. The uses and abuses of rapid bioluminescence-based ATP assays. *International Journal of Hygiene and Environmental Health* vol. 216 115–125 Preprint at https://doi.org/10.1016/j.ijheh.2012.03.009 (2013).
9. Uebelhoer, L. S. *et al.* High-throughput, luciferase-based reverse genetics systems for identifying inhibitors of Marburg and Ebola viruses. *Antiviral Res* **106**, 86–94 (2014).
10. Girotti, S., Ferri, E. N., Fumo, M. G. & Maiolini, E. Monitoring of environmental pollutants by bioluminescent bacteria. *Analytica Chimica Acta* vol. 608 2–29 Preprint at https://doi.org/10.1016/j.aca.2007.12.008 (2008).
11. Machleidt, T. *et al.* NanoBRET-A Novel BRET Platform for the Analysis of Protein-Protein Interactions. *ACS Chem Biol* **10**, 1797–1804 (2015).
12. Dimond, A., Van de Pette, M. & Fisher, A. G. Illuminating Epigenetics and Inheritance in the Immune System with Bioluminescence. *Trends Immunol* **41**, 994–1005 (2020).
13. Luker, K. E. *et al.* In vivo imaging of ligand receptor binding with Gaussia luciferase complementation. *Nat Med* **18**, 172–177 (2012).
14. Iwano, S. *et al.* Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science (1979)* **359**, 935–939 (2018).
15. Khalil, A. S. & Collins, J. J. Synthetic biology: Applications come of age. *Nat Rev Genet* **11**, 367–379 (2010).

16. Proshkina, G. M., Shramova, E. I., Shilova, O. N., Ryabova, A. V. & Deyev, S. M. Phototoxicity of flavoprotein miniSOG induced by bioluminescence resonance energy transfer in genetically encoded system NanoLuc-miniSOG is comparable with its LED-excited phototoxicity. *J Photochem Photobiol B* **188**, 107–115 (2018).

17. Zenchak, J. R. *et al.* Bioluminescence-driven optogenetic activation of transplanted neural precursor cells improves motor deficits in a Parkinson's disease mouse model. *J Neurosci Res* **98**, 458–468 (2020).

18. Yu, S. P. *et al.* Optochemogenetic stimulation of transplanted ips-npcs enhances neuronal repair and functional recovery after ischemic stroke. *Journal of Neuroscience* **39**, 6571–6594 (2019).

19. Petersen, E. D. *et al.* Restoring Function After Severe Spinal Cord Injury Through BioLuminescent-OptoGenetics. *Front Neurol* **12**, 1–12 (2022).

20. Sureda-Vives, M. & Sarkisyan, K. S. Bioluminescence-driven optogenetics. *Life* **10**, 1–11 (2020).

21. Prakash, M. *et al.* Selective control of synaptically-connected circuit elements by all-optical synapses. *Commun Biol* **5**, 1–13 (2022).

22. Love, A. C. & Prescher, J. A. Seeing (and Using) the Light: Recent Developments in Bioluminescence Technology. *Cell Chem Biol* **27**, 904–920 (2020).

23. Kaskova, Z. M., Tsarkova, A. S. & Yampolsky, I. V. 1001 lights: Luciferins, luciferases, their mechanisms of action and applications in chemical analysis, biology and medicine. *Chem Soc Rev* **45**, 6048–6077 (2016).

24. Fleiss, A. & Sarkisyan, K. S. A brief review of bioluminescent systems (2019). *Current Genetics* vol. 65 877–882 Preprint at https://doi.org/10.1007/s00294-019-00951-5 (2019).

25. Tsarkova, A. S. Luciferins Under Construction: A Review of Known Biosynthetic Pathways. *Front Ecol Evol* **9**, 1–9 (2021).

26. Purtov, K. V. *et al.* The Chemical Basis of Fungal Bioluminescence. *Angewandte Chemie - International Edition* **54**, 8124–8128 (2015).

27. Kaskova, Z. M. *et al.* Mechanism and color modulation of fungal bioluminescence. *Sci Adv* **3**, (2017).

28. Kotlobay, A. A. *et al.* Genetically encodable bioluminescent system from fungi. *Proc Natl Acad Sci U S A* **115**, 12728–12732 (2018).

29. Petushkov, V. N. *et al.* A novel type of luciferin from the siberian luminous earthworm fridericia heliota: Structure elucidation by spectral studies and total synthesis. *Angewandte Chemie - International Edition* **53**, 5566–5568 (2014).

30. Mitani, Y. *et al.* Novel gene encoding a unique luciferase from the fireworm Odontsyllis undecimdonta. *Sci Rep* **8**, 1–7 (2018).

31. Schultz, D. T. *et al.* Luciferase of the Japanese syllid polychaete Odontosyllis umdecimdonta. *Biochem Biophys Res Commun* **502**, 318–323 (2018).

32. Kotlobay, A. A. *et al.* Bioluminescence chemistry of fireworm Odontosyllis. *Proc Natl Acad Sci U S A* **116**, 18911–18916 (2019).

33. Purtov, K. V. *et al.* Luciferin–Luciferase System of Marine Polychaete Chaetopterus variopedatus. *Dokl Biochem Biophys* **486**, 209–212 (2019).

34. Tian, Q. *et al.* Cryo-EM structure of the fatty acid reductase LuxC-LuxE complex provides insights into bacterial bioluminescence. *J Biol Chem* **298**, 102006 (2022).

35. Oba, Y. *et al.* Selected Least Studied but not Forgotten Bioluminescent Systems. *Photochem Photobiol* **93**, 405–415 (2017).

36. Tisi, L. C. *et al.* Development of a thermostable firefly luciferase. *Anal Chim Acta* **457**, 115–123 (2002).

37. Loening, A. M., Fenn, T. D., Wu, A. M. & Gambhir, S. S. Consensus guided mutagenesis of Renilla luciferase yields enhanced stability and light output. *Protein Engineering, Design and Selection* **19**, 391–400 (2006).

38. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem Biol* **7**, 1848–1857 (2012).

39. Cui, B. *et al.* Engineering an enhanced, thermostable, monomeric bacterial luciferase gene as a reporter in plant protoplasts. *PLoS One* **9**, 1–11 (2014).

40. Loening, A. M., Wu, A. M. & Gambhir, S. S. Red-shifted Renilla reniformis luciferase variants for imaging in living subjects. *Nat Methods* **4**, 641–643 (2007).

41. Loening, A. M., Dragulescu-Andrasi, A. & Gambhir, S. S. A red-shifted Renilla luciferase for transient reporter-gene expression. *Nat Methods* **7**, 5–6 (2010).

42. Smirnova, N. A. *et al.* Development of Neh2-luciferase reporter and its application for high throughput screening and real-time monitoring of Nrf2 activators. *Chem Biol* **18**, 752–765 (2011).

43. Fan, F. *et al.* Novel genetically encoded biosensors using firefly luciferase. *ACS Chem Biol* **3**, 346–351 (2008).

44. Remy, I. & Michnick, S. W. A highly sensitive protein-protein interaction assay based on Gaussia luciferase. *Nat Methods* **3**, 977–979 (2006).

45. Dixon, A. S. *et al.* NanoLuc Complementation Reporter Optimized for Accurate Measurement of Protein Interactions in Cells. *ACS Chem Biol* **11**, 400–408 (2016).

46. Bloom, J. D. & Arnold, F. H. In the light of directed evolution: pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences of the United States of America* vol. 106 Suppl 9995–10000 Preprint at https://doi.org/10.1073/pnas.0901522106 (2009).

47. Holland, I. & Davies, J. A. Automation in the Life Science Research Laboratory. *Front Bioeng Biotechnol* **8**, 1–18 (2020).

48. Zhou, L., Lei, X. H., Bochner, B. R. & Wanner, B. L. Phenotype MicroArray analysis of Escherichia coli K-12 mutants with deletions of all two-component systems. *J Bacteriol* **185**, 4956–4972 (2003).

49. Baumschlager, A., Aoki, S. K. & Khammash, M. Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control. *ACS Synth Biol* **6**, 2157–2167 (2017).

50. Castillo-Hair, S. M. *et al.* FlowCal: A User-Friendly, Open Source Software Tool for Automatically Converting Flow Cytometry Data from Arbitrary to Calibrated Units. *ACS Synth Biol* **5**, 774–780 (2016).

51. Baumschlager, A., Aoki, S. K. & Khammash, M. Dynamic Blue Light-Inducible T7 RNA Polymerases (Opto-T7RNAPs) for Precise Spatiotemporal Gene Expression Control. *ACS Synth Biol* **6**, 2157–2167 (2017).

52. Han, T., Chen, Q. & Liu, H. Engineered Photoactivatable Genetic Switches Based on the Bacterium Phage T7 RNA Polymerase. *ACS Synth Biol* **6**, 357–366 (2017).

53. Chen, Y. J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods* **10**, 659–664 (2013).

54. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS One* **6**, (2011).

55. Shis, D. L. & Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc Natl Acad Sci U S A* **110**, 5028–5033 (2013).

56. Datta, S., Costantino, N. & Court, D. L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).

57. Bassalo, M. C. *et al.* Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synth Biol* **5**, 561–568 (2016).

58. Schmidl, S. R., Sheth, R. U., Wu, A. & Tabor, J. J. Refactoring and optimization of light-switchable Escherichia coli two-component systems. *ACS Synth Biol* **3**, 820–831 (2014).

59. Ong, N. T. & Tabor, J. J. A Miniaturized Escherichia coli Green Light Sensor with High Dynamic Range. *ChemBioChem* **19**, 1255–1258 (2018).

60. Ohlendorf, R., Vidavski, R. R., Eldar, A., Moffat, K. & Möglich, A. From dusk till dawn: One-plasmid systems for light-regulated gene expression. *J Mol Biol* **416**, 534–542 (2012).

61. Gregor, C., Gwosch, K. C., Sahl, S. J. & Hell, S. W. Strongly enhanced bacterial bioluminescence with the ilux operon for single-cell imaging. *Proc Natl Acad Sci U S A* **115**, 962–967 (2018).

62. Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79–88 (2006).

63. Bindels, D. S. *et al.* MScarlet: A bright monomeric red fluorescent protein for cellular imaging. *Nat Methods* **14**, 53–56 (2016).

64. Saito, K. *et al.* Luminescent proteins for high-speed single-cell and whole-body imaging. *Nat Commun* **3**, (2012).

65. Suzuki, K. *et al.* Five colour variants of bright luminescent protein for real-time multicolour bioimaging. *Nat Commun* **7**, 13718 (2016).

66. McGinness, K. E., Baker, T. A. & Sauer, R. T. Engineering Controllable Protein Degradation. *Mol Cell* **22**, 701–707 (2006).

67. Sharan, S. K., Thomason, L. C., Kuznetsov, S. G. & Court, D. L. Recombineering: A homologous recombination-based method of genetic engineering. *Nat Protoc* **4**, 206–223 (2009).

68. Gerhardt, K. P. *et al.* An open-hardware platform for optogenetics and photobiology. *Sci Rep* **6**, (2016).

69. Liu, Y. *et al.* Functional assessment of hydrophilic domains of late embryogenesis abundant proteins from distant organisms. *Microb Biotechnol* **12**, 752–762 (2019).

70. Steel, H., Habgood, R., Kelly, C. & Papachristodoulou, A. In situ characterisation and manipulation of biological systems with Chi.Bio. *PLoS Biol* **18**, 1–12 (2020).

71. Liu, Z. *et al.* Programming bacteria with light-sensors and applications in synthetic biology. *Front Microbiol* **9**, (2018).

72. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946–950 (2009).

73. Davidson, E. A., Basu, A. S. & Bayer, T. S. Programming microbes using pulse width modulation of optical signals. *J Mol Biol* **425**, 4161–4166 (2013).

74. Milias-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P. & Khammash, M. Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat Commun* **7**, (2016).

75. Lalwani, M. A., Kawabe, H., Mays, R. L., Hoffman, S. M. & Avalos, J. L. Optogenetic Control of Microbial Consortia Populations for Chemical Production. *ACS Synth Biol* **10**, 2015–2029 (2021).

76. Bremer, H. & Dennis, P. P. Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* **3**, (2008).

77. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139**, 1366–1375 (2009).

78. Navarro Llorens, J. M., Tormo, A. & Martínez-García, E. Stationary phase in gram-negative bacteria. *FEMS Microbiol Rev* **34**, 476–495 (2010).

79. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat Methods* **11**, 508–520 (2014).

80. Chait, R., Ruess, J., Bergmiller, T., Tkačik, G. & Guet, C. C. Shaping bacterial population behavior through computer-interfaced control of individual cells. *Nat Commun* **8**, (2017).

81. Tabor, S. Expression Using the T7 RNA Polymerase/Promoter System. *Curr Protoc Mol Biol* **11**, (1990).

82. Iost, I., Guillerez, J. & Dreyfus, M. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J Bacteriol* **174**, 619–622 (1992).

83. Miroux, B. & Walker, J. E. Production of Proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol* **260**, 289–298 (1996).

84. Zoltowski, B. D. & Crane, B. R. Light activation of the LOV protein vivid generates a rapidly exchanging dimer. *Biochemistry* **47**, 7012–7019 (2008).

85. Zoltowski, B. D. *et al.* Conformational switching in the fungal light sensor vivid. *Science (1979)* **316**, 1054–1057 (2007).

86. Vaidya, A. T., Chen, C. H., Dunlap, J. C., Loros, J. J. & Crane, B. R. Structure of a light-activated LOV protein dimer that regulates transcription. *Sci Signal* **4**, 1–8 (2011).

87. Schwerdtfeger, C. & Linden, H. VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation. *EMBO Journal* **22**, 4846–4855 (2003).

88. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat Commun* **6**, (2015).

89. Benedetti, L. *et al.* Optimized vivid-derived magnets photodimerizers for subcellular optogenetics in mammalian cells. *Elife* **9**, 1–49 (2020).

90. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* **33**, 755–760 (2015).

91. Kawano, F., Okazaki, R., Yazawa, M. & Sato, M. A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nat Chem Biol* **12**, 1059–1064 (2016).

92. Morikawa, K. *et al.* Photoactivatable Cre recombinase 3.0 for in vivo mouse applications. *Nat Commun* **11**, (2020).

93. Furuya, A., Kawano, F., Nakajima, T., Ueda, Y. & Sato, M. Assembly Domain-Based Optogenetic System for the Efficient Control of Cellular Signaling. *ACS Synth Biol* **6**, 1086–1095 (2017).

94. Benedetti, L. *et al.* Light-activated protein interaction with high spatial subcellular confinement. *Proc Natl Acad Sci U S A* **115**, E2238–E2245 (2018).

95. Chen, X. *et al.* An extraordinary stringent and sensitive light-switchable gene expression system for bacterial cells. *Cell Res* **26**, 854–857 (2016).

96. Wang, X., Chen, X. & Yang, Y. Spatiotemporal control of gene expression by a light-switchable transgene system. *Nat Methods* **9**, 266–269 (2012).

97. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res* **40**, 8773–8781 (2012).

98. Bowers, L. M., Lapoint, K., Anthony, L., Pluciennik, A. & Filutowicz, M. Bacterial expression system with tightly regulated gene expression and plasmid copy number. *Gene* **340**, 11–18 (2004).

99. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: A one-pot DNA shuffling method based on type ils restriction enzymes. *PLoS One* **4**, (2009).

100. Busby, S. & Ebright, R. H. Transcription activation by catabolite activator protein (CAP). *J Mol Biol* **293**, 199–213 (1999).

101. Tabor, J. J., Levskaya, A. & Voigt, C. A. Multichromatic control of gene expression in escherichia coli. *J Mol Biol* **405**, 315–324 (2011).

102. Hirose, Y., Narikawa, R., Katayama, M. & Ikeuchi, M. Cyanobacteriochrome CcaS regulates phycoerythrin accumulation in Nostoc punctiforme, a group II chromatic adapter. *Proc Natl Acad Sci U S A* **107**, 8854–8859 (2010).

103. Hirose, Y., Shimada, T., Narikawa, R., Katayama, M. & Ikeuchi, M. Cyanobacteriochrome CcaS is the green light receptor that induces the expression of phycobilisome linker protein. *Proc Natl Acad Sci U S A* **105**, 9528–9533 (2008).

104. Gambetta, G. A. & Lagarias, J. C. Genetic engineering of phytochrome biosynthesis in bacteria. *Proc Natl Acad Sci U S A* **98**, 10566–10571 (2001).

105. Hirose, Y. *et al.* Green/red cyanobacteriochromes regulate complementary chromatic acclimation via a protochromic photocycle. *Proc Natl Acad Sci U S A* **110**, 4974–4979 (2013).

106. Nakajima, M., Ferri, S., Rögner, M. & Sode, K. Construction of a Miniaturized Chromatic Acclimation Sensor from Cyanobacteria with Reversed Response to a Light Signal. *Sci Rep* **6**, 4–11 (2016).

107. Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R. & Tabor, J. J. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat Methods* **11**, 449–455 (2014).

108. Fernandez-Rodriguez, J., Moser, F., Song, M. & Voigt, C. A. Engineering RGB color vision into Escherichia coli. *Nat Chem Biol* **13**, 706–708 (2017).

109. Schmidl, S. R. *et al.* Rewiring bacterial two-component systems by modular DNA-binding domain swapping. *Nat Chem Biol* **15**, 690–698 (2019).

110. Olson, E. J., Tzouanas, C. N. & Tabor, J. J. A photoconversion model for full spectral programming and multiplexing of optogenetic systems. *Mol Syst Biol* **13**, 926 (2017).

111. Mahajan, T. & Rai, K. A novel optogenetically tunable frequency modulating oscillator. *PLoS One* **13**, 1–29 (2018).

112. Castillo-Hair, S. M., Baerman, E. A., Fujita, M., Igoshin, O. A. & Tabor, J. J. Optogenetic control of Bacillus subtilis gene expression. *Nat Commun* **10**, 1–11 (2019).

113. Hueso-Gil, A., Nyerges, Á., Pál, C., Calles, B. & de Lorenzo, V. Multiple-Site Diversification of Regulatory Sequences Enables Interspecies Operability of Genetic Devices. *ACS Synth Biol* **9**, 104–114 (2020).

114. Larsen, B. *et al.* Highlighter: an optogenetic actuator for light-mediated, high resolution gene expression control in plants. *bioRxiv* 2022.10.28.514161 (2022).

115. Wang, H. & Yang, Y. T. Mini Photobioreactors for in Vivo Real-Time Characterization and Evolutionary Tuning of Bacterial Optogenetic Circuit. *ACS Synth Biol* **6**, 1793–1796 (2017).

116. Shih, S. C. C., Soffer, G. & Perry, J. M. Real-time optogenetics system for controlling gene expression using a model-based design. *Anal Chem* **93**, 3181–3188 (2021).

117. Tandar, S. T., Senoo, S., Toya, Y. & Shimizu, H. Optogenetic switch for controlling the central metabolic flux of Escherichia coli. *Metab Eng* **55**, 68–75 (2019).

118. Senoo, S., Tandar, S. T., Kitamura, S., Toya, Y. & Shimizu, H. Light-inducible flux control of triosephosphate isomerase on glycolysis in Escherichia coli. *Biotechnol Bioeng* **116**, 3292–3300 (2019).

119. Wang, S. *et al.* Development of Optogenetic Dual-Switch System for Rewiring Metabolic Flux for Polyhydroxybutyrate Production. *Molecules* **27**, 1–12 (2022).

120. Hartsough, L. A. *et al.* Optogenetic control of gut bacterial metabolism to promote longevity. *Elife* **9**, 1–16 (2020).

121. Lugagne, J.-B., Blassick, C. M. & Dunlop, M. J. Deep model predictive control of gene expression in thousands of single cells. *bioRxiv* 2022.10.28.514305 (2022).

122. Balleza, E., Kim, J. M. & Cluzel, P. Systematic characterization of maturation time of fluorescent proteins in living cells. *Nat Methods* **15**, 47–51 (2018).

123. David, M. *et al.* Cascade regulation of nif gene expression in Rhizobium meliloti. *Cell* **54**, 671–683 (1988).

124. Wright, G. S. A. *et al.* Architecture of the complete oxygen-sensing FixL-FixJ two-component signal transduction system. *Sci Signal* **11**, (2018).

125. Möglich, A., Ayers, R. A. & Moffat, K. Design and Signaling Mechanism of Light-Regulated Histidine Kinases. *J Mol Biol* **385**, 1433–1444 (2009).

126. Diensthuber, R. P., Bommer, M., Gleichmann, T. & Möglich, A. Full-length structure of a sensor histidine kinase pinpoints coaxial coiled coils as signal transducers and modulators. *Structure* **21**, 1127–1136 (2013).

127. Gleichmann, T., Diensthuber, R. P. & Möglich, A. Charting the signal trajectory in a light-oxygen-voltage photoreceptor by random mutagenesis and covariance analysis. *Journal of Biological Chemistry* **288**, 29345–29355 (2013).

128. Łoś, M. *et al.* Effective inhibition of lytic development of bacteriophages λ, P1 and T4 by starvation of their host, Escherichia coli. *BMC Biotechnol* **7**, 1–6 (2007).

129. Slomińska, M., Neubauer, P. & Wegrzyn, G. Regulation of bacteriophage λ development by guanosine 5'-diphosphate- 3'-diphosphate. *Virology* **262**, 431–441 (1999).

130. Gutiérrez Mena, J., Kumar, S. & Khammash, M. Dynamic cybergenetic control of bacterial co-culture composition via optogenetic feedback. *Nat Commun* **13**, 1–16 (2022).

131. Berglund, K., Birkner, E., Augustine, G. J. & Hochgeschwender, U. Light-Emitting Channelrhodopsins for Combined Optogenetic and Chemical-Genetic Control of Neurons. *PLoS One* **8**, (2013).

132. Kim, C. K., Cho, K. F., Kim, M. W. & Ting, A. Y. Luciferase-lov bret enables versatile and specific transcriptional readout of cellular protein-protein interactions. *Elife* **8**, (2019).

133. Shramova, E. I., Proshkina, G. M., Chumakov, S. P., Khodarovich, Y. M. & Deyev, S. M. Flavoprotein miniSOG Cytotoxisity Can Be Induced By Bioluminescence Resonance Energy Transfer. *Acta Naturae* **8**, 118–125 (2016).

134. Parag-Sharma, K. *et al.* Engineered BRET-Based Biologic Light Sources Enable Spatiotemporal Control over Diverse Optogenetic Systems. *ACS Synth Biol* **9**, 1–9 (2020).

135. Chen, F., Warnock, R. L., Van Der Meer, J. R. & Wegner, S. V. Bioluminescence-Triggered Photoswitchable Bacterial Adhesions Enable Higher Sensitivity and Dual-Readout Bacterial Biosensors for Mercury. *ACS Sens* **5**, 2205–2210 (2020).

136. Naim, N. *et al.* Luminescence-activated nucleotide cyclase regulates spatial and temporal cAMP synthesis. *Journal of Biological Chemistry* **294**, 1095–1103 (2019).

137. Gomez-Ramirez, M., More, A. I., Friedman, N. G., Hochgeschwender, U. & Moore, C. I. The BioLuminescent-OptoGenetic in vivo response to coelenterazine is proportional, sensitive, and specific in neocortex. *J Neurosci Res* **98**, 471–480 (2020).

138. Losi, A., Polverini, E., Quest, B. & Gärtner, W. First evidence for phototropin-related blue-light receptors in prokaryotes. *Biophys J* **82**, 2627–2634 (2002).

139. Matthews, J. C., Hori, K. & Cormier, M. J. Purification and Properties of Renilla reniformis Luciferase. *Biochemistry* **16**, 85–91 (1977).

140. Lorenz, W. W., McCann, R. O., Longiaru, M. & Cormier, M. J. Isolation and expression of a cDNA encoding Renilla reniformis luciferase. *Proc Natl Acad Sci U S A* **88**, 4438–4442 (1991).

141. Loening, A. M., Fenn, T. D. & Gambhir, S. S. Crystal Structures of the Luciferase and Green Fluorescent Protein from Renilla reniformis. *J Mol Biol* **374**, 1017–1028 (2007).

142. Rahnama, S., Saffar, B., Kahrani, Z. F., Nazari, M. & Emamzadeh, R. Super RLuc8: A novel engineered Renilla luciferase with a red-shifted spectrum and stable light emission. *Enzyme Microb Technol* **96**, 60–66 (2017).

143. Shigehisa, M. *et al.* Stabilization of luciferase from Renilla reniformis using random mutations. *Protein Engineering, Design & Selection* **30**, 7–13 (2016).

144. Fanaei Kahrani, Z., Emamzadeh, R., Nazari, M. & Rasa, S. M. M. Molecular basis of thermostability enhancement of Renilla luciferase at higher temperatures by insertion of a disulfide bridge into the structure. *Biochim Biophys Acta Proteins Proteom* **1865**, 252–259 (2017).

145. Ishibashi, M. *et al.* Expression and characterization of the Renilla luciferase with the cumulative mutation. *Protein Expr Purif* **145**, 39–44 (2018).

146. Chaloupkova, R. *et al.* Light-Emitting Dehalogenases: Reconstruction of Multifunctional Biocatalysts. *ACS Catal* **9**, 4810–4823 (2019).

147. Schenkmayerova, A. *et al.* Engineering the protein dynamics of an ancestral luciferase. *Nat Commun* **12**, (2021).

148. Yang, J. *et al.* Coupling optogenetic stimulation with NanoLuc-based luminescence (BRET) Ca++ sensing. *Nat Commun* **7**, 13268 (2016).

149. Inagaki, S. *et al.* Genetically encoded bioluminescent voltage indicator for multi-purpose use in wide range of bioimaging. *Sci Rep* **7**, 42398 (2017).

150. Schaub, F. X. *et al.* Fluorophore-NanoLuc BRET reporters enable sensitive In Vivo optical imaging and flow cytometry for monitoring tumorigenesis. *Cancer Res* **75**, 5023–5033 (2015).

151. Chu, J. *et al.* A bright cyan-excitable orange fluorescent protein facilitates dual-emission microscopy and enhances bioluminescence imaging in vivo. *Nat Biotechnol* **34**, 760–767 (2016).

152. Boute, N. *et al.* NanoLuc Luciferase - A multifunctional tool for high throughput antibody screening. *Front Pharmacol* **7**, 1–11 (2016).

153. Brodl, E., Winkler, A. & Macheroux, P. Molecular Mechanisms of Bacterial Bioluminescence. *Computational and Structural Biotechnology Journal* vol. 16 551–564 Preprint at https://doi.org/10.1016/j.csbj.2018.11.003 (2018).

154. Gregor, C. *et al.* Autonomous bioluminescence imaging of single mammalian cells with the bacterial bioluminescence system. *Proc Natl Acad Sci U S A* **116**, 26491–26496 (2019).

155. Mitiouchkina, T. *et al.* Plants with genetically encoded autoluminescence. *Nat Biotechnol* **38**, 944–946 (2020).

156. Fisher, A. J., Thompson, T. B., Thoden, J. B., Baldwin, T. O. & Rayment, I. The 1.5-Å resolution crystal structure of bacterial luciferase in low salt conditions. *Journal of Biological Chemistry* **271**, 21956–21968 (1996).

157. Campbell, Z. T., Weichsel, A., Montfort, W. R. & Baldwin, T. O. Crystal structure of the bacterial luciferase/flavin complex provides insight into the function of the β subunit. *Biochemistry* **48**, 6085–6094 (2009).

158. Takai, A. *et al.* Expanded palette of Nano-lanterns for real-time multicolor luminescence imaging. *Proc Natl Acad Sci U S A* **112**, 4352–4356 (2015).

159. Tung, J. K., Gutekunst, C. A. & Gross, R. E. Inhibitory luminopsins: Genetically-encoded bioluminescent opsins for versatile, scalable, and hardware-independent optogenetic inhibition. *Sci Rep* **5**, 1–14 (2015).

160. Xin, X., Xi, L. & Tu, S. C. Functional Consequences of Site-Directed Mutation of Conserved Histidyl Residues of the Bacterial Luciferase α Subunit. *Biochemistry* **30**, 11255–11262 (1991).

161. Van der Krogt, G. N. M., Ogink, J., Ponsioen, B. & Jalink, K. A comparison of donor-acceptor pairs for genetically encoded FRET sensors: Application to the Epac cAMP sensor as an example. *PLoS One* **3**, (2008).

162. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **22**, 1567–1572 (2004).

163. Clavel, D. *et al.* Structural analysis of the bright monomeric yellow-green fluorescent protein mNeonGreen obtained by directed evolution. *Acta Crystallogr D Struct Biol* **72**, 1298–1307 (2016).

164. Tizei, P. A. G., Csibra, E., Torres, L. & Pinheiro, V. B. Selection platforms for directed evolution in synthetic biology. *Biochem Soc Trans* **44**, 1165–1175 (2016).

165. Molina, R. S. *et al.* In vivo hypermutation and continuous evolution. *Nature Reviews Methods Primers* **2**, (2022).

166. Wang, Y. *et al.* Directed Evolution: Methodologies and Applications. *Chem Rev* **121**, 12384–12444 (2021).

167. Yang, G. & Withers, S. G. Ultrahigh-throughput FACS-based screening for directed enzyme evolution. *ChemBioChem* **10**, 2704–2715 (2009).

168. Subach, F. V., Piatkevich, K. D. & Verkhusha, V. V. Directed molecular evolution to design advanced red fluorescent proteins. *Nat Methods* **8**, 1019–1026 (2011).

169. Wellner, A. *et al.* Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nat Chem Biol* **17**, 1057–1064 (2021).

170. Jurėnas, D., Fraikin, N., Goormaghtigh, F. & Van Melderen, L. Biology and evolution of bacterial toxin–antitoxin systems. *Nat Rev Microbiol* **20**, 335–350 (2022).

171. Aminov, R. I. & Mackie, R. I. Evolution and ecology of antibiotic resistance genes. *FEMS Microbiol Lett* **271**, 147–161 (2007).

172. Peterson, E. & Kaur, P. Antibiotic resistance mechanisms in bacteria: Relationships between resistance determinants of antibiotic producers, environmental bacteria, and clinical pathogens. *Front Microbiol* **9**, 1–21 (2018).

173. Levy, S. B. & Bonnie, M. Antibacterial resistance worldwide: Causes, challenges and responses. *Nat Med* **10**, S122–S129 (2004).

174. Darby, E. M. *et al.* Molecular mechanisms of antibiotic resistance revisited. *Nat Rev Microbiol* (2022) doi:10.1038/s41579-022-00820-y.

175. Bush, K. & Bradford, P. A. β-Lactams and β-Lactamase Inhibitors: An Overview. *Cold Spring Harb Perspect Med* (2016) doi:10.1101/cshperspect.a025247.

176. Jeong, S. H. *et al.* Molecular characterization of extended-spectrum beta-lactamases produced by clinical isolates of Klebsiella pneumoniae and Escherichia coli from a Korean nationwide survey. *J Clin Microbiol* **42**, 2902–2906 (2004).

177. Krause, K. M., Serio, A. W., Kane, T. R. & Connolly, L. E. Aminoglycosides : An Overview. *Cold Spring Harb Perspect Med* (2016) doi:10.1101/cshperspect.a027029.

178. Davis, B. D., Chen, L. & Tai, P. C. Misread protein creates membrane channels: An essential step in the bactericidal action of aminoglycosides. *Proc Natl Acad Sci U S A* **83**, 6164–6168 (1986).

179. Sarwar, M. & Akhtar, M. Cloning of aminoglycoside phosphotransferase (APH) gene from antibiotic-producing strain of Bacillus circulans into a high-expression vector, pKK223-3. Purification, properties and location of the enzyme. *Biochemical Journal* **268**, 671–677 (1990).

180. Stubbe, J. & Kozarich, J. W. Mechanisms of Bleomycin-Induced DNA Degradation. *Chem Rev* **87**, 1107–1136 (1987).

181. Gatignol, A., Durand, H. & Tiraby, G. Bleomycin resistance conferred by a drug-binding protein. *FEBS Lett* **230**, 171–175 (1988).

182. Nikaido, H. & Thanassi, D. G. Penetration of lipophilic agents with multiple protonation sites into bacterial cells: Tetracyclines and fluoroquinolones as examples. *Antimicrob Agents Chemother* **37**, 1393–1399 (1993).

183. Pioletti, M. *et al.* Crystal structures of complexes of the small ribosomal subunit with tetracycline, edeine and IF3. *EMBO Journal* **20**, 1829–1839 (2001).

184. Thaker, M., Spanogiannopoulos, P. & Wright, G. D. The tetracycline resistome. *Cellular and Molecular Life Sciences* **67**, 419–431 (2010).

185. Dunkle, J. A., Xiong, L., Mankin, A. S. & Cate, J. H. D. Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc Natl Acad Sci U S A* **107**, 17152–17157 (2010).

186. Tenson, T., Lovmar, M. & Ehrenberg, M. The mechanism of action of macrolides, lincosamides and streptogramin B reveals the nascent peptide exit path in the ribosome. *J Mol Biol* **330**, 1005–1014 (2003).

187. Sothiselvam, S. *et al.* Macrolide antibiotics allosterically predispose the ribosome for translation arrest. *Proc Natl Acad Sci U S A* **111**, 9804–9809 (2014).

188. Gupta, P., Kannan, K., Mankin, A. S. & Vázquez-Laslop, N. Regulation of gene expression by macrolide-induced ribosomal frameshifting. *Mol Cell* **52**, 629–642 (2013).

189. Kannan, K., Vázquez-Laslop, N. & Mankin, A. S. Selective protein synthesis by ribosomes with a drug-obstructed exit tunnel. *Cell* **151**, 508–520 (2012).

190. Weisblum, B. Erythromycin resistance by ribosome modification. *Antimicrob Agents Chemother* **39**, 577–585 (1995).

191. Estrada, A., Wright, D. L. & Anderson, A. C. Antibacterial antifolates: From development through resistance to the next generation. *Cold Spring Harb Perspect Med* **6**, 1–10 (2016).

192. Lombardo, M. N., G-Dayanandan, N., Wright, D. L. & Anderson, A. C. Crystal Structures of Trimethoprim-Resistant DfrA1 Rationalize Potent Inhibition by Propargyl-Linked Antifolates. *ACS Infect Dis* **2**, 149–156 (2016).

193. Köbbing, S., Blank, L. M. & Wierckx, N. Characterization of Context-Dependent Effects on Synthetic Promoters. *Front Bioeng Biotechnol* **8**, 1–13 (2020).

194. Tietze, L. & Lale, R. Importance of the 5′ regulatory region to bacterial synthetic biology applications. *Microb Biotechnol* **14**, 2291–2315 (2021).

195. Quax, T. E. F. *et al.* Differential translation tunes uneven production of operon-encoded proteins. *Cell Rep* **4**, 938–944 (2013).

196. Huber, M. *et al.* Translational coupling via termination-reinitiation in archaea and bacteria. *Nat Commun* **10**, (2019).

197. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep* **7**, 1–9 (2017).

198. Hersch, G. L., Baker, T. A. & Sauer, R. T. SspB delivery of substrates for ClpXP proteolysis probed by the design of improved degradation tags. *Proc Natl Acad Sci U S A* **101**, 12136–12141 (2004).

199. Braun, M., Killmann, H. & Braun, V. The β-barrel domain of FhuAΔ5-160 is sufficient for TonB-dependent FhuA activities of Escherichia coli. *Mol Microbiol* **33**, 1037–1049 (1999).

200. Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. & Helinski, D. R. Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA. *Proc Natl Acad Sci U S A* **71**, 3455–3459 (1974).

201. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**, 1203–1210 (1997).

202. Moore, C. L., Papa, L. J. & Shoulders, M. D. A Processive Protein Chimera Introduces Mutations across Defined DNA Regions in Vivo. *J Am Chem Soc* **140**, 11560–11564 (2018).

203. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).

204. Delcour, A. H. Outer membrane permeability and antibiotic resistance. *Biochim Biophys Acta Proteins Proteom* **1794**, 808–816 (2009).

205. Ciechonska, M. *et al.* Emergent expression of fitness-conferring genes by phenotypic selection. *PNAS Nexus* **1**, 1–13 (2022).

206. Yeh, A. H. W. *et al.* De novo design of luciferases using deep learning. *Nature* **614**, 774–780 (2023).

207. Somermeyer, L. G. *et al.* Heterogeneity of the GFP fitness landscape and data-driven protein design. *Elife* **11**, 1–28 (2022).

208. Lin, Z. *et al.* Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science (1979)* **379**, 1123–1130 (2023).