Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by

M.Sc. Edoardo Romano
Born in Cernusco sul Naviglio (Milan, Italy)
Oral examination date: April 19, 2021

Engineering AraC to make it responsive to light instead of arabinose

Acknowledgements

I would start by thanking Barbara, for all the support and suggestions that she gave me during all the 5 years of my PhD. Her passion and energy helped to push the projects forward, especially in those moments when my project was stuck.

All the people I worked with gave me incredible support during these years, technical and emotional. From the first day, with my shirt decorated with bananas, to the lasts, when they had to bear my nervous and stressed mood.
I thank Anna, Linda and Pierre, for the funny moments, the laughs and the many beers together. And for the precious scientific discussions, of course!
Thank you also to Eethan, the living scientific encyclopedia, with (almost) always the right answer to my problems in the lab.
Thank you, Enoch, for the discussions, especially those in the microscopy room.
Emir, thank you for your help with the cloning, microscopy and the bacteriographs. I like your enthusiasm toward science, do not lose it!
Thank you, Mehmet, for the support to the protein function prediction of my project, I appreciated the point of view from someone with a "dry biology" background and the comments on my thesis.
Jeremy, Mirko, Giada, and Jara, thank you for the daily chats and jokes together. The daily routine was lighter with you around.
Thank you, Dagmar, oh Dagmar, for your patience and your smiles.
Lastly, thanks to the grumpy Daniel, my dear deskmate and benchmate, fellow of complains.

Vorrei anche ringraziare i miei amici rimasti in Italia. Anche se ci vediamo poco, quando ci troviamo è sempre una festa. Sono fortunato a non averli persi dopo più di 5 anni passati all'estero: è anche grazie a loro che ho mantenuto la tenacia per arrivare fino alla fine.
In particolare, Claudio, Stefano e Vito, grazie per la vostra vicinanza nonostante la lontananza.
Grazie anche a Laura, la mia supervisor durante il mio tirocinio di tesi magistrale in Unimi, che mi ha insegnato come pensare e come muovermi nell'ambito di un progetto di ricerca: quello che ho imparato in quell'anno mi è stato molto utile durante il dottorato.

Un grazie immenso alla mia famiglia e al nonno, senza i quali non sarei potuto arrivare dove sono ora, incoraggiandomi e sostenendomi sempre.

E infine, Margherita, grazie per essere sempre stata al mio fianco in questi anni, anche quando non avevo occhi che per la tesi, risollevandomi dai momenti di sconforto e spazzando via i cattivi pensieri.

Contributions

Leonard Ernst, a former bachelor student of the Di Ventura lab, cloned a 2-plasmid-based library composed of 5 VVD-AraC fusion proteins under IPTG induction and performed the first explorative experiments to demonstrate the light-responsiveness of the system.

Emir Bora Akmeriç, technical assistant in our laboratory, helped me during the last year of my doctorate in the cloning process of some constructs and with the induction and imaging of the genes with poor or unknown function. Moreover, he took care of the Blade Runner bacteriograph, repeating it until he made it flawless.

Mehmet Ali Öztürk, postdoc in our laboratory, contributed with structural biology analysis of the I_1-I_2 sequences rearrangements and collected, interpreted and catalogued the function and localization prediction outputs of the genes with poor or unknown function. He also provided valuable suggestions and criticisms regarding the aforementioned parts.

Navaneethan Palanisamy, postdoc in our laboratory, purified the BLADE FP6 and AraC proteins, performed EMSA and SEC.

Maximilian Hörner, from the Weber lab, provided technical assistance with the determination of the absorption spectrum of BLADE.

Abstract

Engineering methods to spatiotemporally control gene expression is one of the most important goals of synthetic biology. While some level of temporal and spatial control of gene expression can be achieved with chemicals added to the cell population, sophisticated regulation cannot be achieved in this way. The rapidly emerging field of optogenetics offers the solution to this problem because the light is a perfect spatially and temporally confinable external trigger, enabling the easy regulation of the gene expression process in individual cells within a population or layer of cells.

Here, I present the engineering of the natural L-arabinose-responsive bacterial transcriptional activator AraC in the non-photoresponsive bacterium *Escherichia coli* to render it light-inducible. Several optogenetic systems that regulate transcription in bacteria have already been developed. With my work, I aim to provide the scientific community with a tool for easily switching the induction trigger from L-arabinose to light in pre-existing L-arabinose-responsive plasmids and strains.

AraC activates transcription from the P_{BAD} promoter through dimer rearrangement after binding of L-arabinose. The dimerization domain was swapped with the blue light-triggered dimerizing protein VVD from *Neurospora crassa* to drive AraC dimerization with light and, consequently, control its ability to activate P_{BAD}. Initially, a small library of fusion constructs was created, whose expression was dependent on IPTG induction and the inducible promoter was cloned in a different plasmid.

I further engineered the system, removing the IPTG dependence for VVD-AraC expression and cloning the transcription factor in the same plasmid as the P_{BAD} and the reporter. Then I optimized the induction protocol and enlarged the initial library, obtaining higher and more reproducible induction levels.

I characterized this small family of novel **b**lue **l**ight-inducible **A**raC **d**imers in *E. coli*, named **BLADE**, to finely control gene expression in space and time. I compared BLADE with wild-type-AraC in terms of inducer catabolism, induction reversibility and population heterogeneity, highlighting the strengths and weaknesses of each. To showcase BLADE's ability to spatiotemporally control gene expression, I performed bacteriography, a method that relies on the selective passage of light through a photomask to reproduce, using bacteria, complex images. Among others, I reproduced the Blade Runner movie poster and Michelangelo's "Creation of Adam" fresco with unprecedented quality. I also investigated the mechanism of BLADE action *in vivo,* showing the formation of aggregates in the dark, which I speculate contribute to the tightness of the system.

I successfully demonstrated the applicability of BLADE in inducing with light gene expression from plasmids and strains that normally would respond to L-arabinose. These results prove that BLADE enables optogenetic experiments to be done with pre-existing L-arabinose-inducible systems without the need to clone, a distinctive feature that no other light-inducible system has.

I employed BLADE to regulate the expression of proteins involved in cell division (MinD and its mutant MinDΔ10) and cell shape (MreB and RodZ), showing that light can be applied to control bacterial cell morphology, which paves the way to more sophisticated studies of the effect of environmental factors on morphology in the future. Using BLADE to overexpress MinDΔ10, I demonstrate that minicell formation can be triggered at a specific time point, which offers the possibility to obtain minicells of similar metabolic activity, that bears potential benefits for the use of minicells as delivery vehicles. To showcase the advantage of light as an external trigger in medium

and high-throughput assays, I built a library of 117 constructs to characterize 39 *E. coli* genes with unknown or poorly defined function in terms of intracellular localization and effect on cell growth and morphology of their overexpression. I identified several proteins that, when overexpressed, affect cell growth, both positively and negatively; I also found two proteins whose overexpression leads to cell elongation and another one that exhibits a toxic effect. Lastly, through fusion to a fluorescent reporter, I determined their localization.

In conclusion, I believe that BLADE is a robust and effective optogenetic tool for the study of bacterial gene regulatory networks and gene function. I expect that its plug-and-play functionality, together with its tight induction control and its reliable performances, will allow its adoption in microbiology, synthetic biology and biotechnology.

Zusammenfassung

Die Entwicklung von Methoden zur räumlichen und zeitlichen Kontrolle der Genexpression ist eines der wichtigsten Ziele der synthetischen Biologie. Während ein gewisses Maß an zeitlicher und räumlicher Kontrolle der Genexpression mit Chemikalien erreicht werden kann, die der Zellpopulation zugesetzt werden, kann eine ausgefeilte Regulation auf diese Weise nicht erreicht werden. Das schnell aufkommende Feld der Optogenetik bietet die Lösung für dieses Problem, denn Licht ist ein perfekter räumlich und zeitlich eingrenzbarer externer Auslöser, der eine einfache Regulierung der Genexpression in einzelnen Zellen innerhalb einer Population oder Schicht von Zellen ermöglicht.

Hier stelle ich das Engineering des natürlichen L-Arabinose-responsiven bakteriellen Transkriptionsaktivator AraC in dem nicht-photoresponsiven Bakterium *Escherichia coli* vor, um ihn lichtinduzierbar zu machen. Es wurden bereits mehrere optogenetische Systeme entwickelt, die die Transkription in Bakterien regulieren. Mit meiner Arbeit möchte ich der wissenschaftlichen Gemeinschaft ein Werkzeug zur Verfügung stellen, mit dem der Induktionsauslöser in bereits existierenden L-Arabinose-responsiven Plasmiden und Stämmen einfach von L-Arabinose auf Licht umgestellt werden kann.

AraC aktiviert die Transkription vom P_{BAD}-Promotor durch Dimer-Umlagerung nach Bindung von L-Arabinose. Die Dimerisierungsdomäne wurde mit dem Blaulicht-aktiviertem dimerisierenden Protein VVD aus *Neurospora crassa* ausgetauscht, um die Dimerisierung von AraC mit Licht zu steuern und folglich seine Fähigkeit, P_{BAD} zu aktivieren, zu kontrollieren. Zunächst wurde eine kleine Bibliothek von Fusionskonstrukten erstellt, deren Expression von der IPTG-Induktion abhängig war und der induzierbare Promotor wurde in ein anderes Plasmid kloniert.

Ich baute das System weiter aus, entfernte die IPTG-Abhängigkeit für die VVD-AraC-Expression und klonierte den Transkriptionsfaktor in dasselbe Plasmid wie die P_{BAD} und den Reporter. Dann optimierte ich das Induktionsprotokoll und vergrößerte die ursprüngliche Bibliothek, um höhere und reproduzierbarere Induktionswerte zu erhalten.

Ich charakterisierte diese kleine Familie neuartiger Blaulicht-induzierbarer AraC-Dimere in E. coli, genannt BLADE, um die Genexpression in Raum und Zeit genau zu kontrollieren. Ich verglich BLADE mit Wildtyp-AraC in Bezug auf den Katabolismus des Induktors, die Reversibilität der Induktion und die Heterogenität der Population und hob die Stärken und Schwächen der beiden hervor. Um die Fähigkeit von BLADE zur räumlich-zeitlichen Kontrolle der Genexpression zu demonstrieren, führte ich die Bakteriographie durch, eine Methode, die auf dem selektiven Durchgang von Licht durch eine Photomaske beruht, um mit Hilfe von Bakterien komplexe Bilder zu reproduzieren. Unter anderem reproduzierte ich das Blade Runner-Filmplakat und Michelangelos Fresko "Die Erschaffung Adams" in noch nie dagewesener Qualität. Ich untersuchte auch den Wirkmechanismus von BLADE in vivo und zeigte die Bildung von Aggregaten im Dunkeln, die, wie ich vermute, zur Kontrollierbarkeit des Systems beitragen.

Ich demonstrierte erfolgreich die Anwendbarkeit von BLADE bei der Induktion der Genexpression mit Licht von Plasmiden und Stämmen, die normalerweise auf L-Arabinose reagieren würden. Diese Ergebnisse beweisen, dass BLADE es ermöglicht, optogenetische Experimente mit bereits existierenden L-Arabinose-induzierbaren Systemen durchzuführen, ohne dass diese geklont werden müssen - eine Besonderheit, die kein anderes lichtinduzierbares System hat.

Ich habe BLADE eingesetzt, um die Expression von Proteinen zu regulieren, die an der Zellteilung (MinD und seine Mutante MinD10) und an der Zellform (MreB und RodZ) beteiligt sind. Damit habe ich gezeigt, dass Licht zur Kontrolle der bakteriellen Zellmorphologie eingesetzt werden kann, was den Weg zu anspruchsvolleren Studien über den Einfluss von Umweltfaktoren auf die Morphologie in der Zukunft ebnet. Mit Hilfe der Überexpression von MinD10 durch BLADE zeige ich, dass die Bildung von Minizellen zu einem bestimmten Zeitpunkt ausgelöst werden kann, was die Möglichkeit bietet, Minizellen mit ähnlicher Stoffwechselaktivität zu erhalten, was potenzielle Vorteile für die Verwendung von Minizellen als Transportmittel mit sich bringt. Um den Vorteil von Licht als externem Auslöser in Mittel- und Hochdurchsatz-Assays zu demonstrieren, habe ich eine Bibliothek von 117 Konstrukten erstellt, um 39 *E. coli*-Gene mit unbekannter oder schlecht definierter Funktion hinsichtlich ihrer intrazellulären Lokalisierung und der Auswirkung ihrer Überexpression auf Zellwachstum und Morphologie zu charakterisieren. Ich identifizierte mehrere Proteine, die, wenn sie überexprimiert werden, das Zellwachstum sowohl positiv als auch negativ beeinflussen; ich fand auch zwei Proteine, deren Überexpression zu einer Zellextension führt und ein weiteres, das eine toxische Wirkung aufweist. Schließlich habe ich durch die Fusion mit einem Fluoreszenzreporter ihre Lokalisierung bestimmt.

Zusammenfassend glaube ich, dass BLADE ein robustes und effektives optogenetisches Werkzeug für die Untersuchung von bakteriellen Genregulationsnetzwerken und Genfunktionen ist. Ich erwarte, dass seine Plug-and-Play-Funktionalität zusammen mit seiner engen Induktionskontrolle und seinen zuverlässigen Leistungen seinen Einsatz in der Mikrobiologie, synthetischen Biologie und Biotechnologie ermöglichen wird.

Publications, presentations and posters

Publications

Romano E., Baumschlager A., Akmeriç E. B., Palanisamy E., Houmani M., Schmidt G., Öztürk M.A., Ernst L., Khammash M. and Di Ventura B. "Engineering AraC to make it responsive to light instead of arabinose" Revisions have been submitted to Nature Chemical Biology.

Oral presentations

"BLADE, a new light-inducible bacterial transcription factor in *E. coli*". Designer Biology, Newcastle upon Tyne, United Kingdom, 2019

Conference poster presentations

Romano E., Di Ventura B. "BLADE, a new light-inducible bacterial transcription factor"
Presented at:

- BIOSS / CIBSS retreat, Schluchsee, Germany, 2018
- DFG SPP 1926 Next Generation Optogenetics, Bonn, Germany, 2018
- Synthetic Biology - from Understanding to Application, Mosbach, Germany, 2018

Wehler P., Antwi E., Romano E., and Di Ventura B. "Quantitative analysis of gene expression control in mammalian cells and bacteria"
Presented at:

- e:Bio Innovation Competition in Systems Biology, Dresden, Germany, 2017
- Systems Biology of Human Disease Conference 2017, Heidelberg, Germany, 2017

Content

1. Introduction

1.1 Synthetic biology

Synthetic biology is an interdisciplinary area that brings several engineering principles to biology[1]. It has no accepted definition so far, and it is described in different ways depending on the specific context. All the definitions, however, have the same concept: the design of biological components and systems that do not already exist in the natural world, by the assembly of artificial or natural components or by synthesis[1].

The roots of synthetic biology come from Francois Jacob and Jacques Monod in 1961[2]. Their study in *E. coli* of the lac operon led them to postulate the existence of regulatory circuits that modulated the response of a cell to its environment. The ability to assemble new regulatory systems from molecular components was then predicted in a subsequent publication[3], but it was only in the 1970s and 1980s, after the development of molecular cloning and PCR, that genetic manipulation became widespread in microbiology research[4].

From the 2000s, the deep knowledge of the classic molecular biology workhorse *E. coli* was exploited to provide to the growing synthetic biology field the first circuit "parts". In the following months, the first work about engineered circuits was published. Collins and colleagues assembled a toggle switch, a synthetic, bistable gene-regulatory network, composed of two promoters arranged in a mutually inhibitory network[5]. In response to two external signals (IPTG and heat), the cells could toggle between two expression states. To mention another example, Elowitz and Leibler created a circuit composed of three negative feedback loops of repressor-promoter pairs. Activation of the circuit, named repressilator, generated periodic oscillation of the reporter protein[6]. In the following period, new circuits were assembled that included autoregulatory negative-feedback modules, to reduce noise and therefore have a narrow population-wide expression distribution[7], and positive-feedback modules, to obtain a binary response of a gene in a cell population with graded distribution[8].

A recent milestone of synthetic biology was the creation in 2010 of the world's first synthetic life form by the research group of Craig Venter, an engineered replica of the cattle bacterium *Mycoplasma mycoides*[9]. Then, in 2014, the group of Jef Boeke created the first synthetic *S. cerevisiae* chromosome, by deleting nonessential sequences as of subtelomeric regions, introns, transfer RNAs, transposons, and silent mating loci[10], and in 2016 the Venter group demonstrated that only 476 genes of *M. mycoides* were necessary for the bacterium to survive[11]. The field is rapidly evolving: in recent years, faster DNA synthesis at a cheaper price allows scientists to design and synthesize modified bacterial chromosomes. As well as synthesizing DNA *de novo*, synthetic biology can also make use of already existing sequences, by assembling them together to create new ones[12].

Other than basic research, synthetic biology applications are wide and encompass different applications: biofuel production, through microorganisms engineered with optimized biosynthetic pathways to efficiently convert biomass into biofuels[13], or biomaterial, through the production of silk proteins by engineered Salmonella secretion system[14], just to mention few examples, but also renewable chemicals, and compound of the health care sector as well[15,16].

1.2   Optogenetics

Tightly bound to synthetic biology, optogenetics is a new innovative approach that combines optical and genetic techniques to control cellular events using light[17]. The first optogenetic experiment was performed in 2002 from the Quail lab to control gene expression in yeast using the light-inducible interacting PhyA-PIF3 and PhyB-PIF3 pairs[18]. However, the first field that benefited from optogenetics was neurobiology, as this technique allowed using light-gated transport proteins to efficiently control membrane potentials of single cells[19].

In 2005, Karl Deisseroth and Edward Boyden demonstrated that, when expressed in neurons, the light-gated cation channel Channelrhodopsin-2 from green algae *Chlamydomonas reinhardtii,* discovered by Nagel *et. al* in 2003[20], could be used to modulate cell depolarization at milliseconds time scale using blue light pulses[21]. Channelrhodopsin-2 was then successfully expressed in freely moving rodents and activated directly within the intact brain using optic fibers[22,23,24].

This new way to control the neural action potential with single-cell precision provided a great advantage over the alternative techniques used to stimulate neural cells: the electrical stimulation allowed temporal precision but lacked cell-type specificity and is potentially risky[25], and the pharmacological approach provided cell specificity but lacked temporal precision[26]. These characteristics allowed controlling the individual elements of a neuronal circuit in order to determine the specific role of single neurons that compose them[27].

In contrast to the chemical inducers, which usually bind to specific receptors, light transfers information through photons, which provides unique properties: spatiotemporal precision and orthogonal inputs, briefly described below.

- **Temporal control:** Sensing the presence of small molecules in the medium can be easily achieved by a cell, for example through its binding to the corresponding receptor, while sensing its changes through time in the environment is a more challenging task. As small molecules are often relatively stable, sophisticated detection mechanisms are required to remain responsive to changes in the chemical signal. Light, however, has the property to instantaneously disappear, especially when related to timescales of cellular processes such as transcription. This is not the case for chemicals in the environment: for these, there may be an activation delay due to the active transport inside the cell[28], and its complete removal can require a certain delay to switch off the signalling.

- **Spatial control:** Electromagnetic radiation spans from 10^{-12} m of gamma radiation to 3-3000 m of long radio wavelengths, but the most familiar and used class of electromagnetic radiation for optogenetics is visible light, which ranges from 400 to 700 nm. Chemically induced systems are limited in spatial resolution due to difficulty in application and because of diffusion of the molecule in the medium. The superior spatiotemporal resolution of a light input compared to a chemical input can enable applications such as spatial patterning[29] or single-cell activation within a cell population[30].

- **Orthogonal input:** Small chemical inducers may require uptake and/or conversion to become the active signalling form, which can be involved in complex feedback regulation. Two examples are lactose and L-arabinose, two sugars that are used as a carbon source by *E. coli*. Besides, the cell-inducer interactions can be influenced by different factors such as growth phase, environmental conditions (*e.g* temperature) and the carbon sources available in the growth medium. Such factors can lead to unpredictable behaviour of the induction level and increased heterogeneity. Moreover, the induction efficiency can be further altered by factors like diffusion, degradation or catabolism of the inducer. Lastly, the lack of specificity could cause off-target effects[18] and in some cases, the inducer itself can be toxic[31]. These issues make some chemical inducers not practical in certain experimental or biotechnological settings, for example in industrial batch cultures.
  In contrast to this, light is a non-invasive orthogonal input for organisms such as *E. coli* that are non-photoresponsive. This allows for perturbations using a light input with little or no crosstalk, a feature difficult to achieve with small molecules. Overall, light-inputs can guarantee better predictability, robustness and homogeneous and rapid control.
  The potential side-effects caused by high intensity light, such as phototoxicity, have to be taken into consideration and avoided through appropriate light intensity, duration inputs and protocol adaptation[32].
  These features distinguish light from chemical inducers and hold the promise for more precise perturbation studies which helps to decipher dynamic interactive cellular networks and processes. Also, it enables regulation strategies for biotechnological applications that require fast changeable or spatially defined inputs.

Over the last fifteen years, the optogenetics field rapidly progressed, driven by the discovery and molecular characterization of natural photoreceptors as well as the development of new engineered photoreceptor variants able to control desired cellular events different than ion flux. Thereby, the field extended from the neurosciences towards cell biology applications and new cellular systems, including bacteria. In these approaches, light is used either as an additional input or as a replacement of a chemical molecule such as inducers for gene expression or hormones for cellular responses. In this case, light fulfils similar functions to small molecule signals, which have been used extensively in biological research and biotechnology. For example, small-molecule inducible gene expression systems are key components in synthetic biology[15] and biotechnological applications[33].

Today, there are optogenetic tools for controlling cellular processes as diverse as cell motility[34,35], kinase activity[36], apoptosis[37,38], protein degradation[39,40], epigenetics events[41] as well as gene expression in prokaryotes[42–48], yeast[49,50] and mammalian cells[19,40,51].

Clinical trials, that aim to restore vision using optogenetic approaches are currently emerging. To date, two studies (combined Phase I/IIa) have commenced using microbial opsins to restore vision in patients with end-stage retinitis pigmentosa (NCT02556736 and NCT03326336 at clinicaltrials.gov.). Vision has already been rescued in mice, and sensitivity to light of human retina from blind patients was partially restored *ex vivo*[52]. Other promising clinical applications are treatments of hearing-loss[53], diabetes[54] and cancer[55], so far all carried out in animal models. Despite the potential clinical prospects, optogenetics is mainly a tool for basic science, to study neuronal networks, circuits and cellular mechanisms. The knowledge acquired will then lead to new therapies and clinical treatments.

1.3 Photoreceptors and their classification

Light activation in biological systems can be either achieved through chemical modification with photosensitive groups and chemical effectors (chelators, isomers), or through genetically encoded photosensitive domains, or photoreceptors. Optogenetic refers to the latter approach. The former, such as photocaged molecules, have been used decades before the first optogenetic methods were developed. Because of their unique dynamic properties and their flexibility, however, optogenetic proteins have quickly caught up as they offer unique spatiotemporal control opportunities and are versatile for a broad range of implementation in different cellular functions. Unlikely most of the photoactivatable compounds, photoreceptor activation is reversible, does not require high energy wavelength for its activation, and the photoreceptor does not need to be constantly supplemented in the media.

Photoreceptors occur in all major kingdoms of life, from bacteria, microalgae and fungi to plants and humans. They mediate an array of functions, such as maintaining the circadian rhythm, phototaxis and phototropism, flowering, as well as the processing of visual cues in higher organisms[56]. They are essential components for optogenetic protein engineering: all photoactivatable proteins used in optogenetics derive from natural photoreceptors. Modularity is a distinctive trait of this class of protein: in general, the input sensing domains are physically and functionally separable from the catalytic activity or the output domains. This feature became the basis for synthetic biology approaches: through combinations with other protein domains, proteins with new functions can be created and expressed in an organism.

Light absorption in all photoreceptors is generally mediated by an organic chromophore carrying a conjugated π electron system. The absorbed light leads to electron jumps from a lower to a higher energy molecular orbital, causing a π - π * transitions. This excitation leads to a structural change of the chromophore that translates into a modification of the protein structure, from "dark" to "lit" state transition, which are both characterized by distinct protein conformations[57]. This process is named photocycle, and its timescale varies from milliseconds to hours, depending on the photoreceptor.

Generally, photoreceptors can be divided into four classes, dependent on their incorporated chromophore (Figure 1).
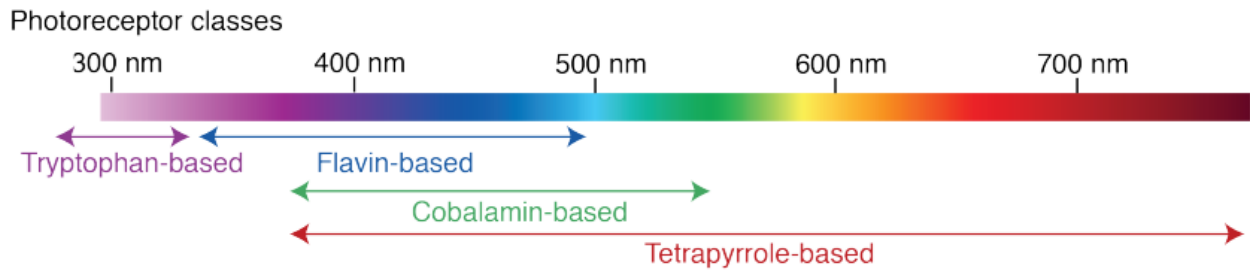
Figure 1 | Schematic overview of the four photosensor classes and their light wavelength ranges of activation. From Kianianmomeni, A. Optogenetics: Methods and Protocols, Methods in Molecular Biology. 1408, (2015).

1.3.1 Tryptophan-based

This class uses intrinsic tryptophan for the absorption of light, with the plant photoreceptor UVR8 as the main representative of this class, responsible to trigger changes in gene expression following UV-B light absorption[58]. UVR8 is a homodimer in the dark, and UV-B illumination reversibly disrupts salt bridges that maintain the dimer together, allowing heterodimer formation with the protein COP1 to initiate transcriptional responses[59]. The function in *Arabidopsis thaliana* is to promote plant development and survival under sunlight.

1.3.2 Cryptochromes

Cryptochromes are a key component of the circadian clock in both animals and plants but are also responsible for a variety of functions in all of life's kingdoms[60]. They show similarity to photolyases, as they contain the photolyase homology domain (PHR) which binds FAD chromophore[61]. In the dark, FAD is in the oxidized state and is reduced in natural semiquinone FADH* after blue or near-UV light illumination. This induces a negative charge in the vicinity of the flavin, leading to ATP release from the binding pocket and C-terminus partial detachment from the PHR at the N-terminal domain, which contains the FAD. This conformational change makes some residues accessible for phosphorylation, allowing the binding to other proteins[62,63]. A widely used cryptochrome is CRY2 from *A. thaliana,* which is monomeric in the dark and oligomerizes after blue light activation[64]. The half-life is in the timescale of minutes and can be tuned by mutating the PHR domain[65].

1.3.3 BLUF

Blue **L**ight **U**sing **F**AD domains were discovered in *R. sphaeroides*[66,67] and *E. gracilis*[68]. Their function is to aid in the adaptation of photosystems synthesis depending on light and oxygen conditions[69]. Blue light illumination allows the transfer of an electron and a proton from a conserved tyrosine to the FAD, leading to the formation of flavin and tyrosine radicals[70,71], which in turn might induce a rearrangement of hydrogen bonds in the flavin binding pocket[69].

1.3.4 Phototropins

Phototrophins are blue light receptors that mediate a large number of responses in a variety of higher plants such as nitrogen fixation, hypocotyl phototropism, regulation of the stomatal opening and chloroplasts movement[72-73]. The N-terminal of *A. sativa* phototropin I contains two similar photosensitive domains, **LOV1** and **LOV2**, members of the PAS (Per, ARNT, Sim) superfamily. As

co-factors, LOV domains employ flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD), which are non-covalently bound to the protein in the dark-adapted state[74,75,76]. In phototropin I, they regulate the activity of a C-terminal Ser/Thr kinase[73,77]. LOV proteins are more closely related to a subset of proteins within the PAS superfamily that are regulated by external signals such as light, oxygen, or voltage, hence the acronym LOV[78]. LOV domains are a highly diverse class of photoreceptors that are abundant in archaea, bacteria, fungi, protists and higher plants[79]. About ~ 6.700 LOV domain candidates have been bioinformatically identified so far[79].

In the dark, a conserved glutamine residue of *As*LOV2 is bound to FMN[74], through the C-terminal Jα helix interacting with the β sheet of the LOV core, thereby masking several hydrophobic residues of the central β-sheet[80]. Blue light irradiation induces the formation of a covalent adduct between the C(4a) carbon of the flavin chromophore and a conserved C450 residue within the LOV domain[81,82]. The Jα helix then undergoes a partial loss of structure that makes it more flexible[80,83]: this conformational change is crucial for the activation of phototropin kinase domain[80]. Relaxation to the dark state is spontaneously triggered by a base-catalyzed mechanism that eliminates the thioether bond between the *As*LOV2 C450 and the chromophore[84]. The half-life of the *As*LOV2 in the lit state is about 1 min[74,85]. The LOV1 domain undergoes an analogous photocycle to LOV2 but is not required for light-induced activation of kinase activity[75,77].

1.3.5 VVD

LOV-domain-containing proteins are not only restricted to plants. White Collar-1 (WC-1) is a fungal blue-light receptor that contains a single LOV domain and mediates phototropism and other light responses in *N. crassa* and other fungi[86]. VIVID (VVD) is a second *Neurospora* protein that mediates photoadaptive response[87] and resetting of circadian clock[88].

Both WC-1 and VVD contain LOV domains that bind FAD[89,90]. WC-1 interacts with its partner WC-2 to form the White Collar Complex (WCC)[91,92], while VVD, only 70 residues long, antagonizes the action of WCC in response to blue light[90].

The essential component for the blue light activity of VVD is the N-terminal cap (Ncap) composed of an α-helix (aα), a β-strand (bβ) and a short hinge that connects the terminus to the LOV core[93] (Figure 2). Analogously of the Jα helix of *As*LOV2 domains, the Ncap helix of VVD undergoes a structural rearrangement upon blue light illumination[93], resulting in a transition from monomer to dimer[94]: blue light irradiation causes adduct formation between the C4a carbon of FAD and the sulfhydryl group of a highly-conserved cysteine (C108)[90]. This induces the flavin N5 protonation, and in response, a conserved Gln residue at position 182 rotates, altering hydrogen bonds of the central beta sheet[95]. This rearrangement propagates to induce conformational changes in Ncap residues, leading to protein dimerization[94,96]. The crucial role of the ''flipping'' glutamine is to couple photochemical changes to signal transmission. Q182L mutation showed not to alter the spectral properties of the protein but impaired the Ncap conformational change[93]. The normal photocycle length of VVD is 3 hours[94], much slower compared to that of *As*LOV.
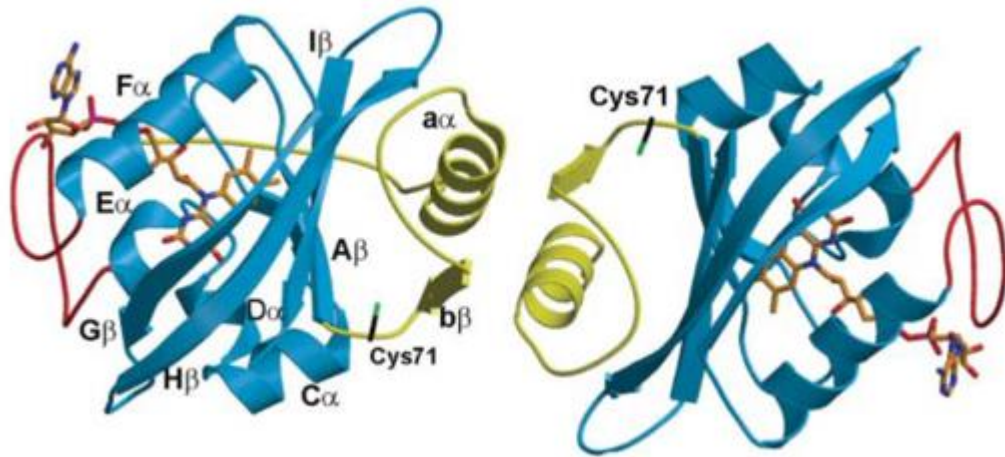
Figure 2 | Crystallographic dimer of VVD-36. Blue, PAS core; yellow, N-terminal cap; red, FAD insertion loop; orange, FAD adenosine moiety. Adapted from[93].

1.3.6 Cobalamin-based receptors

Cobalamin-binding domains (CBDs) are green light photoreceptors that utilize Vitamin B12 or cobalamin as a chromophore for photosensing in diverse bacteria[97–99]. The transcriptional regulator CarH from *Thermus thermophiles,* used in synthetic biology, controls the biosynthesis of carotenoids in response to light[100]. In the dark, CarH binds 5′-deoxyadenosylcobalamin (AdoCbl), one of the two forms of Vitamin B12, and forms tetramers, which bind operator DNA thus blocking transcription. Under wavelengths ranging from near-UV light up 530 nm, the CarH tetramer dissociates to the monomeric form and triggers dissociation of the 5'deoxyadenosyl group, weakening its affinity for DNA and allowing transcription. This reaction is irreversible[101], and this limits the applicability of CarH for optogenetic purposes.

1.3.7 Tetrapyrrole-based receptors

Phytochromes (Phy) are photoreceptors sensitive to red and far-red light found in bacteria, cyanobacteria, plants and fungi. They are involved in controlling gene expression, mediating functions like phototaxis and regulation of circadian rhythm in bacteria[102], and in promoting the development in plants such as germination and flowering[103,104] in response to NIR light.
Phytochromes are classified, depending on their activation light wavelength, to type I, activated by far-red light (730 nm), and type II, activated by red light (660 nm). Another important feature is that they can reversibly switch between a red-absorbing (Pr) and a far-red-absorbing (Pfr) state. There are also algal phytochromes[105] and cyanobacteriochromes[106] that can sense the full visible spectrum, reaching even the near-UV range.

Despite the diversity of light absorption, for the photoswitching to occur, all phytochromes employ a tetrapyrrole chromophore, with some differences depending on the organism. Phytochromes from cyanobacteria and green algae contain a tetrapyrrole chromophore in the reduced form called phycocyanobilin (PCB), while bacteria and fungi phytochromes use the oxidized form biliverdin (BV). In plant phytochromes another chromophore, called phytochromobilin, is present. Phycocyanobilin and phytochromobilin are endogenous to the respective organisms and need to be exogenously provided or genetically engineered when working in animal models[107].

The photocycle, the same for all the phytochromes, starts with the absorption of a photon, causing the shift from Pr to Pfr, while absorption of a second photon at a different wavelength or incubation in the dark over time causes reversion to the Pr state[76]. This feature distinguishes phytochromes over other photoreceptors that rely only on chromophore reversion in the dark, providing phytochromes with a superior temporal control[108]. In a suggested mechanism of photoactivation, far-red and NIR light induce a Z-E isomerization of the C_{15}-C_{16} double bond, causing a rotation of one of the four pyrrole rings (D-ring) composing the chromophore. This generates hydrogen bond rearrangements and consequent conformational change in the protein backbone, which are transferred to the output module that initiates the biological response[109,110]. Depending on the phytochrome, the structural change may lead to heterodimerization with the phytochrome-interacting factor (PIF3) or homodimerization in the case of Cph1 from cyanobacteria[111].

1.4 Inducible promoters in *E. coli*

1.4.1 The role of inducible promoters

Since the birth of molecular cloning, *E. coli* has been used as a host to introduce DNA sequences. The first attempt of molecular cloning traces back to 1973, when Herbert Boyer and Stanley Cohen showed that two fragments of bacterial DNA could be 'cut and pasted' together and returned to *E. coli*[4].

Nowadays, *E. coli* is one of the organisms of choice for the production of a large number of recombinant proteins and chemicals. Its fast growth kinetics (the doubling time in ideal conditions is 20 minutes) allows it to obtain a million *E. coli* cells from one parent cell within about 7 hours. Fast growth means that experiments can be done quickly and conveniently. Moreover, it has low demands for its growth medium and it can grow in absence of oxygen. Because of its important role in genetics and biotechnology, it was one of the earliest genome sequences to be completed, in 1997[112]. Nowadays, most of the regulatory networks and genes have been characterized[113,114].

Metabolic engineering gave an important contribution to the optimization of the expression of heterologous and endogenous systems, with the aim to increase productivity. The dynamic pathway regulation is already established for adjustment between cell growth and production, and inducible promoters are well suited to this purpose[115]. This dynamic control is especially important for complex pathways, like depletion of important metabolites or when the presence of toxic intermediates requires constant tuning of the enzyme concentration[116].

Basic research beneficed of time-controlled gene expression as well: to understand the function of a gene, or to express a recombinant protein, is it often important to titrate the transcription level to the desired quantity and in a specific time frame.

The transcriptional regulators can function as activators (positive regulation), repressors (negative regulation), or both (positive and negative regulation). The transcription is stimulated by binding to DNA of the inducer-activator complexes at specific sites upstream of promoters close to the sites of binding for the RNA polymerase. In the case of activators, there is an interaction with the polymerase or interference in the transcription initiation process in such a way that transcription becomes facilitated[33,117]. Conversely, in the absence of an inducer, the repressors bind to the operator, blocking

its binding to the DNA or inhibiting other steps of the transcription process[118]. When the inducer is added to the medium, it binds to the repressor, causing a conformational change that makes the repressor unable to bind to the operator, allowing transcription by RNA polymerase.

Ideally, expression systems should allow tight control of the expression level; the inducers should be inducible by cheap agents that are metabolically inert and nontoxic[31] and the induction should not present off-target effects[18]. The inducer should also not require particular uptake transport systems, to avoid delay in the induction[28] or inhomogeneous expression levels in the population[119]. A desirable feature is also to stop the transcription with high temporal precision by complete removal of the activating stimulus.

As it is difficult to find an induction system that fulfils all these requirements, there is a continuous ongoing search for new expression systems that may either have more universally preferable properties or are particularly useful for certain types of applications. In particular, there is a need for the development of gene expression systems that can be used for fine-tuning the expression of genes at low and physiologically relevant levels[120,121].

1.4.2 *E. coli araBAD* operon

One of the most often used chemically inducible prokaryotic gene expression systems by microbiologists and synthetic biologists is constituted by the DNA binding protein AraC and its cognate P$_{BAD}$ promoter. AraC is the transcriptional regulator of the L-arabinose operon, which in *E. coli* is responsible for the catabolism of L-arabinose. The arabinose operon and all its components have been studied in detail for the last 60 years, with major contributions from Robert Schleif[122–124]. AraC has been defined by Schleif himself as "my favourite protein" despite the struggle and research time required to detect and purify the protein, and to study its function[125]. He declared that work was worth the effort because of the discovery of the two regulatory mechanisms that emerged from his research: the DNA looping and the light-switch mechanism (described later in this paragraph), with potential more mechanisms to be discovered[125].

1.4.2.1 Function and mechanism

To uptake and catabolize L-arabinose, three operons in *E. coli* are present: the *BAD* operon, encoding three catabolic enzymes that convert L-arabinose to D-xylulose-5-phosphate which then enters the pentose phosphate pathway[123]; the high-affinity low-capacity transporter *FGH* operon, encoding the transporters that regulate L-arabinose uptake when its concentration in the extracellular environment is low[124,126], and the low-affinity high-capacity transporter *araE* operon, encoding a low-affinity transporter that acts at high extracellular L-arabinose[127,128] (Figure 3). If internal L-arabinose exceeds a threshold concentration, it activates AraC, which in turn upregulates by 50-fold the expression of *araFG* and, *araE*[129], and the genes for L-arabinose catabolism, *araBAD*.
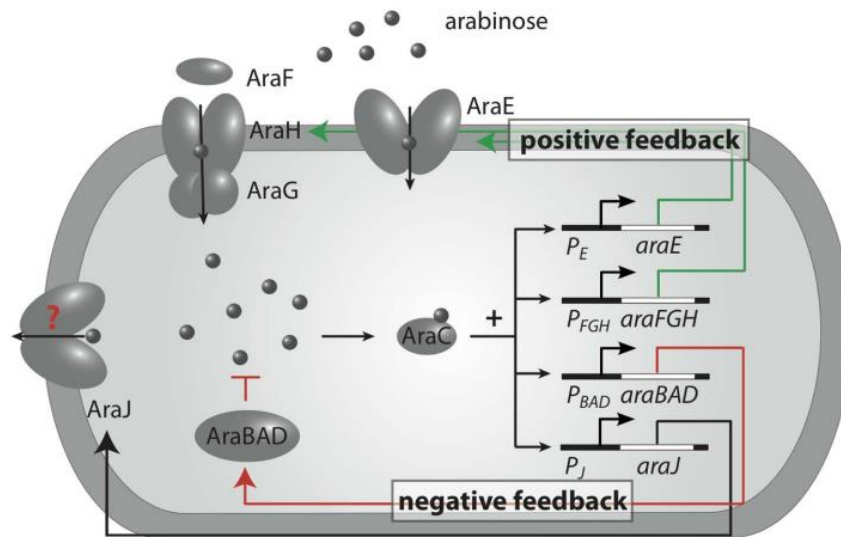
Figure 3 | Schematic of L-arabinose utilization in E. coli. *L-arabinose is imported via the arabinose transporters AraE and AraFGH. When the intracellular L-arabinose concentration is sufficiently high, it binds the transcriptional regulator AraC. This complex activates the promoters P_E, P_{FGH}, P_{BAD} and P_J, driving expression of araE, araFGH, araBAD and araJ, respectively. AraBAD encodes the proteins for L-arabinose catabolism. The negative autoregulation of AraC is not shown. Arrows indicate the arabinose transport and positive regulation, while the T-shaped arrow indicates L-arabinose catabolization. Adapted from*[130].

AraC is composed of an N-terminal dimerization domain (DD) and a C-terminal DNA-binding domain (DBD) connected via a linker (Figure 4a). This protein is always a homodimer, whether bound to L-arabinose or not[123,131]. Binding of L-arabinose triggers a conformational change in AraC, orienting the two DBDs in a way that favours their interaction with the I_1 and I_2 half-sites rather than the I_1 and O_2 half-sites (Figure 4b)[123,131].

In the absence of L-arabinose, P_{BAD} is repressed by AraC, the regulator of the system, bound to the distal I_1 and O_2 half-sites, which causes the formation of a DNA loop that sterically blocks the access of the RNA polymerase to the promoter (Figure 4a). In the presence of L-arabinose, transcription from the P_{BAD}, P_{FGH}, and P_E promoters is activated by AraC in the so-called "light switch mechanism"[123]. AraC additionally negatively feeds back on its own promoter P_C[123,131], divergently oriented from P_{BAD} (Figure 4b). Shortly after the addition of L-arabinose, P_C activity increases about 10-fold, returning to its pre-induction levels circa ten minutes later[132]. This transient derepression is consistent with the hypothesis that opening the DNA loop upon the addition of L-arabinose allows RNA polymerase free access to the P_C promoter until AraC can bind at the O_1 site. This binding is slow because of AraC's low concentration in the cytoplasm and its nonspecific binding to random DNA sequences as a result of the addition of L-arabinose. This hypothesis shows the advantage of DNA looping: it permits fast induction kinetics of the L-arabinose catabolic enzymes[131].
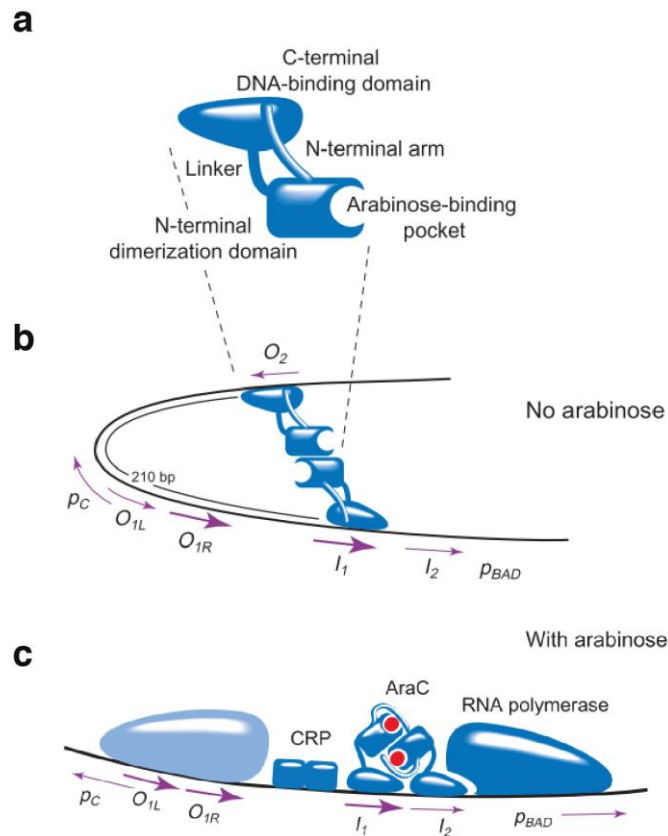
Figure 4 | Regulation of the arabinose operon by L-arabinose. a, The domain structure of the AraC monomer. b, P_{BAD} and P_C regulatory regions in the absence and c, presence of L-arabinose (red circle). The DNA sequences O_2, I_1 and I_2 are 17 bp half-sites each binding to one subunit of AraC. O_1 is composed of two half-sites, O_{1L} and O_{1R}, that allow the binding of two subunits. In the absence of L-arabinose, RNA polymerase is unable to bind P_{BAD} and P_C due to the 210 bp loop formed to AraC binding to O_2 and I_1. CRP, the cyclic AMP receptor protein, is hindered to bind its DNA site. When L-arabinose is present, AraC binds primarily to the adjacent I_1 and I_2 instead of looping. Therefore, the RNA polymerase has free access to P_{BAD}, and CRP is free to bind as well. At the P_C and O_1 sites, the RNA polymerase and AraC compete for the binding. Adapted from[131]

It has been measured that mRNA synthesis from P_{BAD} begins within three seconds of the addition of L-arabinose, and the protein products are induced up to 300 times their uninduced level by AraC[133]. Activation results from AraC binding to the adjacent I_1 and I_2 half-sites, which recruits the RNA polymerase. The gene induction shows the so-called "catabolic sensitivity", meaning that the inducibility is diminished in the presence of glucose[131]. This sensitivity is mediated by the amount of cyclic AMP, which modulates the activity of the cyclic AMP receptor protein, CRP, whose DNA binding site is located between $araO_1$/P_C and araI sites and that is required for full induction of both the P_{BAD} and P_C promoters [134–136].

Being AraC a regulatory protein, its in *vivo* levels are very low: direct physical measurement of AraC level reveals about 20 molecules per cell when L-arabinose is not present[137].

One of the drawbacks of the induction mediated by this promoter is the all-or-none autocatalytic induction at sub-saturating L-arabinose concentrations, in which a fraction of the cells become fully induced and the remainder are uninduced, an undesirable feature when heterologous genes are expressed[119,138]. The initial basal expression of the L-arabinose transporters *araFGH* and *araE* determines the rate of L-arabinose uptake[139]. If the internal L-arabinose concentration exceeds a

threshold, it activates transcription of AraC from the P_C promoter, which in turn binds to P_{FGH} and P_E, promoting transporters transcription, and to P_{BAD}, to promote and L-arabinose catabolism (Figure 3 and 4). Supplementation of L-arabinose in the medium would therefore lead to different rates of L-arabinose accumulation, causing distinct timing of gene induction. At a given time, the fraction of the cells population with low basal transporter expression would remain uninduced, while the rest, thanks to the positive feedback mechanism, would be fully induced. To achieve uniform expression among the population, a strategy is to allow constant L-arabinose transportation into the cell, in a way that is not dependent on the basal level of the transporters in absence of L-arabinose, by adjusting the AraE expression[140,141], or mutating the lactose transporter LacY to reduce its specificity and allow it to transport L-arabinose[142].

Moreover, L-arabinose, unlike IPTG used to induce the *lac* operon, is catabolized by the cell over time, therefore in order to maintain induction over time, supplementation in the media is required. To avoid this problem, D-fucose can be used instead: this non-metabolizable L-arabinose analogue is transported into the cell by AraE and AraFGH but act as an inhibitor, as it binds AraC but does not induce transcription[134,143]. By mutating AraC it is possible to make it inducible by D-fucose[144].

1.4.2.2 The first synthetic biology approach on AraC

In 1993, Bustos and Schlief studied AraC domain composition by constructing two types of chimera[145]. One was generated by fusing the dimerization domain of AraC, $AraC_{DD}$, to the DNA-binding domain of LexA, $LexA_{DBD}$, a protein associated with the repression of several genes involved in the response to DNA damage in *E. coli* (SOS response[146]); the other was composed by the AraC DNA-binding domain, $AraC_{DBD}$, fused to the leucine zipper from the eukaryotic transcriptional activator C/EBP[147], Zip_{DD}. They discovered that both chimeras, $AraC_{DD}$-$LexA_{DBD}$ and $AraC_{DBD}$-Zip_{DD}, could dimerize *in vivo* and regulate the expression of a reporter gene. This experiment showed the modularity of the protein, that includes independent domains capable of dimerizing and binding to the DNA.

The authors also replaced the I_2 half-site of P_{BAD} with I_1, obtaining I_1I_1. In a strain deprived of AraC, this synthetic promoter allowed constitutive induction, both with wild-type AraC and with the $AraC_{DBD}$-Zip_{DD} chimera. This experiment demonstrated that, with the I_1I_1 promoter, AraC could induce transcription even without L-arabinose, through the binding of a monomer to the I_1 proximal to the RNA polymerase binding site.

This result was expected and confirmed that the AraC occupancy of the half-site closer to the RNA polymerase binding site is fundamental for induction. The binding of AraC to I_2 is weaker than to I_1, and can only be stably bound when AraC is in its L-arabinose-bound conformation, binding to both half-sites. As the binding to I_1 is strong, the RNA polymerase recruitment was performed equally with and without L-arabinose.

1.5 Engineering of optogenetic proteins

Photoreceptors show an intrinsic modular architecture with the domain dedicated to photon absorption separated from the domain with effector function, like DNA binding, kinase activity, and dimerization[76] (see paragraph 1.3). This modularity is an advantage for the creation of optogenetic tools, as the photosensing domains can be isolated and engineered with other protein domains for synthetic biology approaches. The domain connection is usually mediated by linkers that can be of crucial importance for the correct functioning of the fusion protein, and therefore need to be tailored, in length and composition, based on the requirement of the specific modules[57,148].

The variety of photoreceptors is the basis of the wide library of photoreceptor-based optogenetic tools available nowadays. The choice of the light-sensing domain can be based either on structural consideration of the protein itself or by its distinctive characteristics as light sensitivity, activation, and reversion kinetic.

Engineering strategies for the development of novel optogenetic tools can be divided into two major categories: controlling protein function by light-inducible protein-protein interactions or by light-dependent allostery (Figure 5). In the first category, three types of activation can be distinguished: heterodimerization[149], homodimerization[45] and dissociation[47]. The second category includes steric hindrance of protein activity mediated by a photoreceptor fused to the protein of choice[150], the photocaging of a small peptide within the *As*LOV2 Jα helix[151,152] or the formation of local disorder from the LOV2 terminal helices[153].
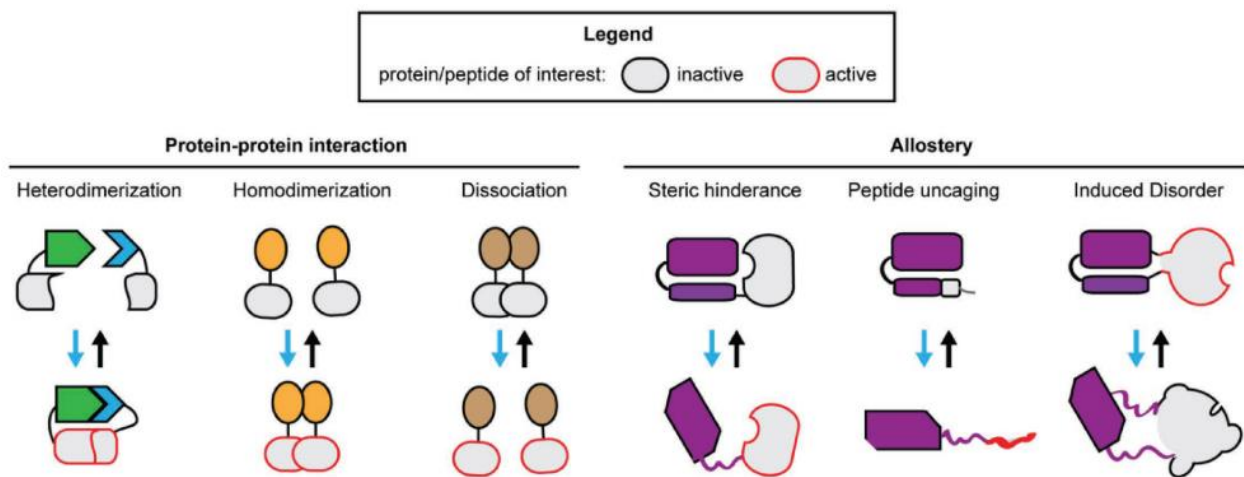
Figure 5 | Different strategies for engineering optogenetic regulation with LOV domains. Blue arrows indicate the transition to the lit state, while black arrows indicate the reversion in the dark. Adapted from[154].

1.5.1 Optogenetic transcriptional regulation in *E. coli*

The requirements for an ideal inducible promoter listed in paragraph 1.4 cannot be fulfilled by most chemical induction systems, a limitation that is not present for optogenetic regulators. Light-induced

expression systems are a versatile mean of regulation, not only for metabolic engineering but also for the study of organisms, genes, etc., as light is a rapid and nontoxic stimulus that naturally regulates many different cellular processes.

Below the currently available light-inducible transcription systems are described: they can be divided into systems composed of one or two distinct components.

1.5.1.1 Two-component systems (TCS)

In this category are included basic stimulus-response coupling mechanism to allow organisms to sense and respond to changes in many different environmental conditions[155]. They are present in all kingdoms of life but they are most common by far in bacteria[156,157]. The environmental signals are first detected by the sensor domain, causing a change of its conformation; the signal then propagates to the intracellular response regulator, usually a histidine kinase, changing its activity[157,158], which ultimately regulates the phosphorylation state of a response regulator. Below are described the main TCS.

- **Red/far-red light EnvZ/Cph1**

  This is the first optogenetic transcription regulator developed for *E. coli:* it consists of a light-sensitive sensor kinase and an intracellular response regulator[159]. Levskaya *et al*. created a chimera based on the sensor kinase EnvZ with the photosensor Cph1 instead of the native Omp sensory domain, changing the sensitivity of the system from osmotic pressure to 650 nm red light. The photosensor Cph1 requires the PCB chromophore, which is not produced by *E. coli,* therefore the authors had to introduce in the cell the genes for the endogenous PCB production[160].

- **Green/red light CcaS/CcaR**

  CcaS/CcaR is a TCS that responds to green and red light[161]. It is composed of the sensor histidine kinase CcaS and its response regulator CcaR. Green light induces CcaS autophosphorylation and subsequent phosphotransfer to CcaR, which then activates transcription; absorption of red light reverses the process[161]. The activation-inactivation wavelength allows orthogonality with the red-far red Cph8-based system[162]. An optimized version allowed reaching a dynamic range of 72-fold for the red-far red pair and 117-fold for the red-green pair[163].

- **UV/green light UirS/UirR**

  UirS/UirR[164] is a TCS from *Synechocystis* sp. PCC 6803 sensitive to UV and green light. The membrane-bound UirS binds and sequesters the transcription factor UirR to the membrane in the dark. UV-violet light phosphorylates and liberates UirR, allowing the promoter binding and the transcription initiation. Green light illumination shifts back UirS in the inactive state. The authors measured a dynamic range of about 4-fold[43].

- **Red/Near-Infrared Light BphP1/PpsR2**

  The use of near-infrared wavelengths (> 700 nm) in optogenetics are advantageous due to their low phototoxicity, spectral isolation from most photoproteins, and deeper tissue penetration. BphP1/PpsR2 is a synthetic TCS that binds the chromophore biliverdin, composed by the near-

infrared (760 nm) photosensory domain of *Rhodopseudomonas palustris* BphP1, that induces transcription by binding and following inhibition to the transcriptional repressor PpsR2. Red light (640 nm) then induces dissociation of BphP1 from PpsR2. The measured reporter fold-change between dark and light was 2.5x.

- **Blue Light FixL/FixJ**

  The TCS described so far require a tetrapyrrole chromophore, that needs to be either supplemented to the culture medium or synthesized through *E. coli* engineering. In 2009, Möglich *et al*.[165] used a LOV photosensor to create a blue-light induction system. This protein relies on the FMN chromophore, which is natively available in the cell. They replaced the chemosensor domain of the pair FixL/FixJ from *Bradyrhizobium japonicum* with the LOV domain of YtvA from *Bacillus subtilis.* The conformational change following blue light illumination triggers autophosphorylation of FixL that phosphorylates FixJ, allowing the binding to the FixK2 promoter[165].

  This system shows a 12-fold induction of the reporter expression between dark and light state[42]. By using the same system to control the expression of the λ phage repressor cI, they obtained blue light repression of transcription, with a light-dark fold-change of 430-fold.

- **Blue light pDusk/pDawn**

  Three years later from the development of FixL/FixJ, the group of Andreas Möglich developed another two-component system for light-regulated gene expression, by employing the previously created histidine kinase YF1[165] harbouring a LOV domain[42]. In the dark, YF1 phosphorylates FixJ that induces the expression from the FixK2 promoter. Blue light illumination reduces the kinase activity of YF1[165] and therefore the gene expression. The YF1/FixJ system was named pDusk. To make the system light-inducible instead of light-repressed, the authors inserted the λ phage repressor *cI* gene under the control of the pFixK2 promoter, which represses the pR promoter[166]. This configuration, called pDawn, allows blue light to repress the transcription of cI, allowing induction of the gene under pR. The gene expression increased of factor 19 after 1 hour of dark illumination with pDusk, and of factor 460 after 5 hours with pDawn.

- **OptoLac**

  This optogenetic protein, created by Lalwani *et al*.[167] offer, as the main novelty, the possibility to turn a chemically inducible gene expression system into a light-inducible one. The lacI repressor, a component of the lac operon, was controlled by the optogenetic transcription system pDawn. In the dark YF1 phosphorylates FixJ, that activates transcription of the λ phage repressor cI: this represses LacI expression, allowing transcription of the gene of interest, placed under a promoter containing the lac operator site LacO, bound by LacI. Blue light does not induce FixJ phosphorylation, therefore cI is not expressed and consequently, LacI expression can cause the repression of the gene of interest. To reduce the half-life of LacI and therefore increase the reporter expression they fused the C-terminal degradation tags from the 10Sa transfer-messenger RNA (SsrA)[168] to LacI. With this configuration, the construct with their higher dynamic range reaches 60.9-fold of gene expression in the dark, while keeping the circuit IPTG-inducible.

1.5.1.2 One-component systems

When in a signal transduction system, the input (light-sensitive domain) and output (effector domain) are present in the same protein we have a one-component system[169]. These response mechanisms, that lack the phosphotransfer domains, are evolutionarily older[169], more widely distributed among bacteria and archaea[170], and the domain repertoire is more diverse[169] than TCS. Below are described the one-component optogenetic systems.

- **LightOff and eLightOn**

    Chen *et al.*[171] created a blue light-inducible repressor by fusing VVD to the DNA-binding domain of the LexA repressor, from the *E. coli* SOS regulon[172]. The fusion protein was named LEVI, and the expression system LightOff. When VVD-LexA is illuminated, dimerizes and binds to the promoter, repressing the transcription; 2.5 hours after the stop of blue light illumination the measured fold-change reporter expression is 10,000x. The time took for the protein level induced by LightOff to reach its 50% maximal expression level was 3 h, therefore is slower than Cph8-OmpR and CcaS-CcaR systems (40 and 44 minutes, respectively) but nevertheless is faster than pDusk. The addition of a cI/pR, a NOT gate already employed to invert the induction condition of pDusk to pDawn, was employed to LightOff as well, leading to the formation of Inverted-LightOff, a blue light-inducible activator, at cost of a reduction of fold-change (but still beyond 1,000-fold induction), a lower light sensitivity, as the transcription of the cI repressor has to be fully repressed by high-intensity light to activate the reporter expression, and a delayed ON/OFF switch, as accumulation or degradation of the cI repressor requires time.

    More recently, Li *et al.*[47] switched from LEVI the photosensory domain VVD with the LOV domain from *R. sphaeroides* (*Rs*LOV), which dimerizes in the dark and monomerizes with blue light. This construct, named eLightOn, avoided the disadvantages of using the NOT gate of Inverted-LightOFF to obtain a blue light-inducible activator. The best-performing construct was obtained by random mutagenesis of the linker between LexA and RsLOV domain, with multiple rounds of sorting of mutants with the lower expression in the dark and higher expression after an illumination period. The selected construct reached 500-fold induction

- **EL222**

    Motta-Mena *et al.*[173] fused the VP16 activator domain from *Herpex simplex*[174] with the light-inducible DNA-binding protein EL222 from the bacterium *Erythrobacter litoralis* HTCC2594[175] generating the fusion protein called VP-EL222. EL222 is a natural light-controlled DNA-binding protein, therefore it was not engineered to be used in *E. coli*. The fusion protein includes a LOV domain and a helix-turn-helix (HTH) domain; blue light disrupts the LOV-HTH interactions, leading to EL222 dimerization and subsequent DNA binding[176,177]. The dark reversion is fast (τ ~ 11 s at 37 °C)[178] allowing precise control of gene expression in mammalian cells.

    EL222 was also used to control transcription in *E. coli* by Jayaraman *et al.*[45] They exploited the similarities of EL222 with DNA-binding protein family LuxR[177], and replaced the LuxR binding sequence (*lux* box) with the sequence bound by EL222. Blue light triggers the EL222 binding to the promoter recruiting the RNA polymerase similarly to LuxR type transcriptional activator, obtaining a fold-change of 5x. They also created a light-induced repression system, from a previous work demonstrating the conversion of LuxR into a transcriptional repressor[179]. This was

obtained by placing the EL222 binding site in between the -35 and -10 regions of the promoter, obtaining a 3-fold reduction of fluorescence by blocking the recruitment of RNA polymerase with blue light illumination[45].

- **LightOn**

  Wang et al.[180] described a new technique to spatiotemporally manipulate gene expression in mammalian cells and mice through the light-dimerization of VVD (see paragraph 1.3.5) coupled to a fragment of the transcription factor Gal4, named Gal4(65), comprising the DNA-binding domain and not the dimerization domain. Light-induced VVD dimerization caused Gal4(65) dimerization and DNA binding. The mutations N56K and C71V were inserted into VVD, and the transactivation domain of the transcription factor p65 was added to the fusion construct, obtaining the synthetic protein GAVPO. This protein, induced by blue light, can initiate transcription of genes containing Gal4-binding sites in their promoter region. The transactivator–promoter transgene system was named LightOn.

  In vitro studies showed that the level of gene expression induced by the LightOn system was between 200 and 300-fold, a level comparable to that driven by the human cytomegalovirus (HCMV) early promoter. The induction showed also to be tunable, as the authors obtained different responses of the mCherry expression with different blue light irradiances. By measuring the off expression kinetics of the reporter mRNA, the half-life induction state of LightOn was calculated to be 2 hours, similar to that of VVD[181].

- **Opto-T7RNAPs**

  Baumschlager et al. created a light-inducible T7 RNA polymerases[46] by splitting the protein into two inactive parts, which are fused to light-induced dimerization domains. Following previous studies showing the retention of activity after splitting and reconstituting the T7RNAP[182,183], the authors employed the VVD-based heterodimerizing domains called Magnet[149] to reconstitute the protein. Magnet is an engineered version of the photosensor VVD, which homodimerizes with blue light, obtained through the insertion of negative and positive charged amino acids into the dimer interface to create a "positive" pMag and a "negative" nMag domain, to have a heterodimerizable protein pair. This system has been used to reconstitute many split proteins[184–187]. Depending on the split position of the T7RNAP, different Opto-T7RNAPs variants were generated, each with their own basal expression levels and dynamic range. With this system, the authors managed to reach more than 300-fold induction of reporter expression.

- **AdoB12**

  Ortiz-Guerrero et al.[188] demonstrated that coenzyme B_{12} induces a light-responsive behaviour that regulates gene expression through a transcriptional factor in *E. coli*. It modulates the transcriptional activation of the repressor CarH binding to it in the dark and inducing its tetramerization, resulting in P_B promoter repression. Light disrupts the photosensitive $AdoB_{12}$-repressor complex, dismantling the tetramer and allowing the transcription from P_B.

- **Photocaged molecules**

  Photocaged molecules allow light-control through the photolytic release of photosensitive groups or isomerization of chemicals. Two of the most relevant inducers in biotechnology for bacterial transcription, IPTG and L-arabinose, have been photocaged in recent years. IPTG was photocaged with 6-nitropiperonal (NP) which sterically hinders binding to LacI[189]. The caging was then released with UV (365 nm) light. The lac promoter was activated 10-fold following uncaging of IPTG.

  L-arabinose was photocaged in a similar manner with 1-(6-nitrobenzo[d][1,3] dioxol-5-yl)ethanol (NBE) to inhibit binding to AraC[190]. UV light at 365 nm resulted in uncaging of the molecule and DNA transcription. Doxycycline, which regulates the eukaryote gene expression, was previously photocaged[191–193], but its applications in bacterial gene expression control are still not shown.

1.6 Bacterial photography

Bacterial photography, or bacteriography, is an image performed using genetically engineered bacteria that sense light and respond by producing a fluorescence protein or a dye, behaving like pixels of a digital photograph. It is an experiment commonly performed to show the spatial ability properties of proteins that regulate transcription with light.

To create the image, a lawn of bacteria expressing an optogenetic protein (corresponding to the sensor of digital cameras) is spread over an agar surface, and the light source, that induces the transcription or the repression of the reporter gene is placed on top of it. In between, the so-called "photomask" is positioned: this is the image to reproduce, printed on transparent support or directly composed by cutting out a shape from an opaque surface. The light passing through the photomask will then cast a shadow to the bacterial layer, that will faithfully reflect the picture of the photomask. Induction (or repression) of the reporter will then happen only in correspondence of the illuminated cells. The image is then taken with a fluorescence microscope, or by a normal camera, depending on the type of the reporter used.

The first example of this technique appeared in 2005 in the laboratory of Christopher Voigt[194], with the chimera Cph1-EnvZ (Cph8, Figure 6a), the first optogenetic transcription regulator for *E. coli* (see paragraph 1.5.1.1). With the development of new optogenetic proteins for transcription regulation, more examples of this technique were made, all with a similar protocol (Figure 6b-f).

A few years later the creation of Cph8, Voigt and his group created Bactograph (www.bactograph.org), a cheap bacteria photography kit that they deliver to high schools, universities and museums (**Figure 6**g). They use the same Cph1-EnvZ fusion construct employed in their original paper, that controlled the transcription of *bFMO*, which converts the naturally occurring molecule indol to the blue pigment indigo[195]. This allowed the visualization of the bacteriograph by eye, avoiding the use of a fluorescence microscope.

In 2017, the same group engineered the *E. coli* genome to insert a complex genetic circuit composed by different photosensors connected each other by NOT gates, allowing it to distinguish between red, green and blue light[196]: this system was demonstrated with the creation of a colour bacteriograph (Figure 6h).

However, the creation of bacteriographs is not a unique feature of optogenetic proteins: Kong *et al.*[197] exploited the double property of the antibiotic nisin to activate transcription of a reporter gene from the nisin-inducible promoter P_{nisA}, when present in low concentrations in the medium, and to kill the cells, when present in high concentrations, obtaining a circuit with a band-pass feature. By plating *E. coli* sensitive to nisin in an agar plate and covering it with with a photomask soaked with the antibiotic, the authors triggered the expression of the reporter gene only from the cells on the border of the photomask, that received the right amount of nisin that allowed to survive and transcribe the fluorophore. Nevertheless, the patterns obtained with this system are simple and not comparable with the bacteriographs achieved with optogenetic proteins.
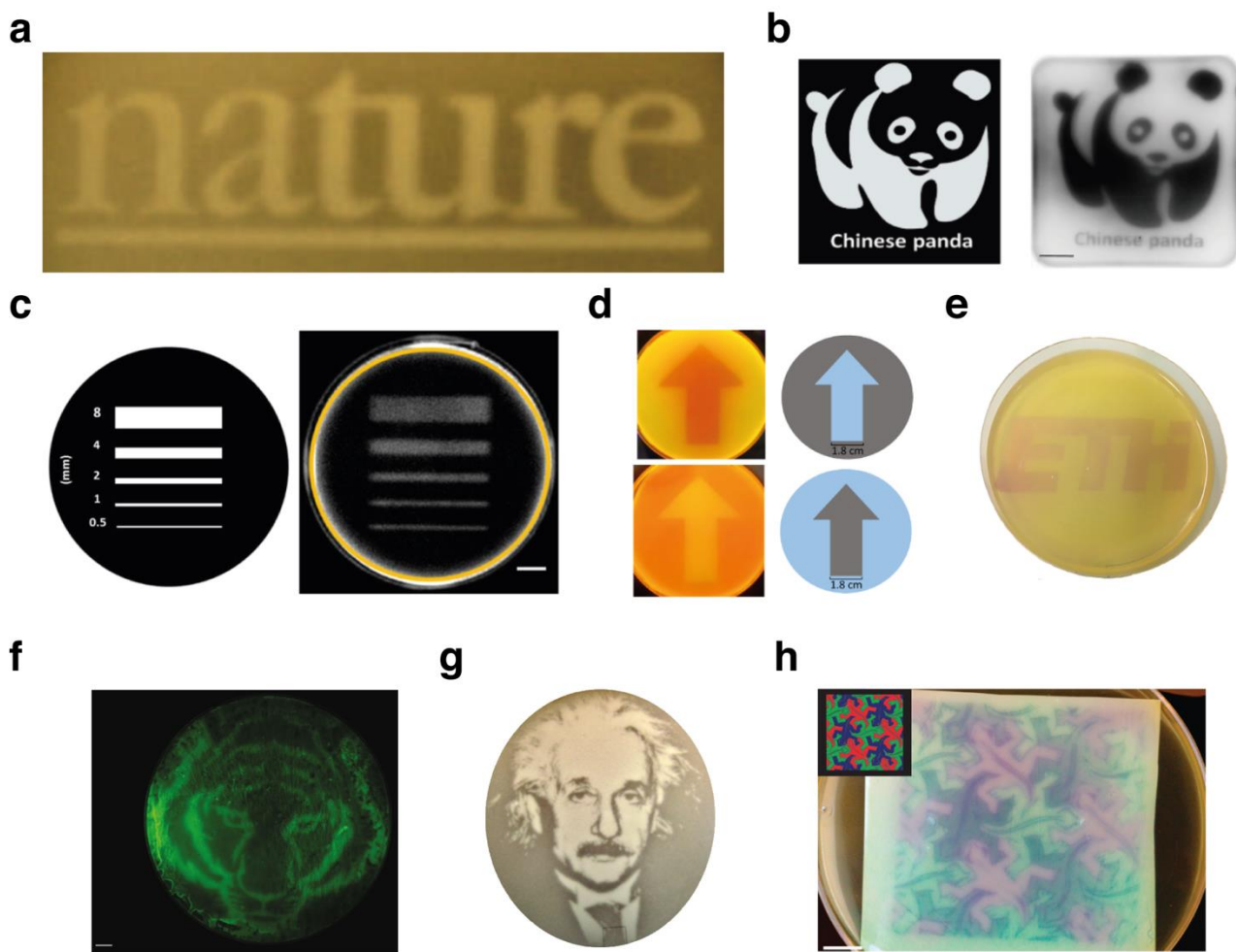
Figure 6 | Bacteriographs obtained through different light-inducible proteins in E. coli. a, Cph8, adapted from[194]. ***b****, LightOff, scale bar 2 cm, adapted from*[171]. ***c****, LightOn, scale bar 1 cm, adapted from*[47]. ***d****, EL222, adapted from*[45]. *Arrow thickness is 1.8 cm. **e**, Opto-T7RNAP, adapted from*[46]. ***f****, Opto-Lac, adapted from*[167]. ***g****, Cph8, from www.bactograph.org. **h**, Cph8, CcaS-CcaR, YF1-FixJ, scale bar: 1 cm, adapted from*[196].

1.7   The genes with unknown function in *E. coli:* cataloguing and characterization

The *E. coli* genome contains 4623 genes, of which 35% currently lack experimental evidence of function[113]. Historically, the unannotated genes are known as 'y-genes' in *E. coli,* as their names start with 'y'[198]. However this is not the official rule, as many poorly characterized genes do not have a name starting with 'y', and vice versa many y-genes that have been annotated did not receive new names[113]. The quality of the annotations in public databases can also greatly vary. Some are well-annotated with direct experimental evidence, while others have limited functional information. The different databases that collect the protein annotations, like EcoCyc[199], EcoGene[200], UniProt[201] and RefSeq[202] are not always updated and is not rare that recently characterized protein may appear only on some and not all the databases. In 2019, Ghatak *et al*. have analysed and catalogued all the genes with unknown or poorly characterized function, generating a consistent and updated list called 'y-ome'[198]. The constant advances in DNA sequencing are speeding up the characterization of *E. coli* genome, but as the definitive functional characterization of unknown genes still requires biochemical and genetic analysis in case-by-case studies, many genes remain so far uncharacterized. The characterization process, irrespective of the organism, can be performed by the deletion of target genes[203], or their overexpression[204,205] *in vivo*, followed by the phenotypic characterization of the organism.

Webservers that predict the function or the localization of an unknown protein are useful tools: with them is possible to narrow down the choice of the protein to characterize based on the predicted function or localization and plan the next experiments accordingly. As each webserver use a different algorithm, employing more than one is a valid solution to discard wrong predictions and obtain more realistic results. There is considerable activity today in the field of computational protein function prediction, as more accurate and reliable software are created every year: for example, Phyre predicts the three-dimensional structure of a protein sequence using the principles and techniques of homology modeling[206]. The algorithm of Argot (**A**nnotation **R**etrieval of **G**ene **O**ntology **T**erms) starts from the DNA/protein sequences and performs a BLAST search against the UniProt database and an HMMER3 search against Pfam, an annotated database of protein families. The results are used to generate GO terms[207]. Pannzer (**P**rotein **ANN**otation with **Z**-sco**RE**), uses the whole sequence similarity neighbourhood and weighted statistical testing in the annotation process to maximize the evidence for correct annotation; it is designed to predict the functional description and GO classes[208]. DeepGO plus employs deep convolutional neural networks (CNN) to learn sequence features and combine predictions with sequence similarity-based predictions, performing well in predicting cellular locations of proteins[209].

To decrease the size of the y-ome, in *E. coli* as in other organisms, new workflows will be necessary, combining computational genome annotation with systems biology modelling and new high-throughput experimental approaches[210–212] to establish reliable functional annotations across the genome.

1.8 Regulation of *E. coli* morphology and division

Cell shape in eukaryotes is primarily determined by the action of various filamentous structures, mostly actin filaments. Bacteria exhibit a wide diversity of shapes, and this characteristic has been used as an important taxonomic criterion. In the well-studied rod-shaped bacterium *E. coli*, several of mutants that have a spherical morphology have been isolated and characterized. Most of the mutations affect processes associated with peptidoglycan synthesis, composing the cell envelope (*e.g., pbpA* and *rodA*)[213,214]. The peptidoglycan sacculus is responsible for the bacteria morphology as well, as it provides mechanical stability to the bacteria and maintains the specific shape of the cell: its enlargement and division are essential for the bacterium growth[215].
On the other hand, another group of genes, *mreB*, *mreC*, and *mreD*, are not associated with cell envelope synthesis but are fundamental for the regulation of the cell shape determination[216].

Cell division is a crucial moment for the shape determination of the cell; the parental cell elongates, the chromosome is duplicated and segregated at the two poles, along with the proteins and all the component of the cytosol. Then the division complex accurately finds the mid-cell and assemble the septum that divides the cell into two identical daughter cells. A single error in the process, such as a different transcription level of a gene involved in this process, and the cell might not divide properly, becoming filamentous.
Cell shape can also greatly vary through the different growth phases, with cells becoming shorter as the population optical density increases over OD_{600} ~0.3 and nearly round in stationary phase[217]. Moreover, the steady-state cellular dimensions of many rod-shaped bacteria adjust in response to nutrient-determined changes in growth rate, with faster-growing cells having increased volume[217].
I will now briefly describe some of the genes involved in the determination of the cell morphology that I worked with in my thesis.

1.8.1 MinD and MinDΔ10

MinD is a membrane-bound ATPase, which, together with MinC and MinE, constitutes the Min system, machinery needed to place the divisome at mid-cell[218] and to aid chromosome segregation[219], preventing the formation of small cells lacking DNA[218]. It is present as a dimer, with the amphipathic helix at the C-terminus of each MinD monomer that allows the dimer to bind the lipid bilayer[220,221]. To exert its inhibitory function against FtsZ, the protein that starts divisome assembly, MinC must be recruited to the cytoplasmic membrane by MinD[222,223]. In *E. coli*, the MinC-MinD complex spontaneously oscillates from pole to pole due to MinE, that binds to MinD on the membrane and catalyzes the ATP hydrolysis to ADP, causing the release of the MinC-MinD complex from the membrane[224] (Figure 7). In the cytoplasm, ADP is phosphorylated again to ATP, so that MinD can bind an area of the membrane with lower MinE concentration, and the cycle continues.
An oscillatory cycle of MinD takes approximately 50-60 seconds at room temperature[225], therefore each pole is devoid of MinD, and consequently MinC, only for a short period. When averaged over time, MinC-MinD concentration is highest at the poles and minimal at mid-cell, causing the septum to form at mid-cell. There, FtsZ forms a ring structure that remains at the leading edge of the invaginating septum[226]. Once this ring is in place, it recruits a series of transmembrane proteins that ultimately trigger cytokinesis.
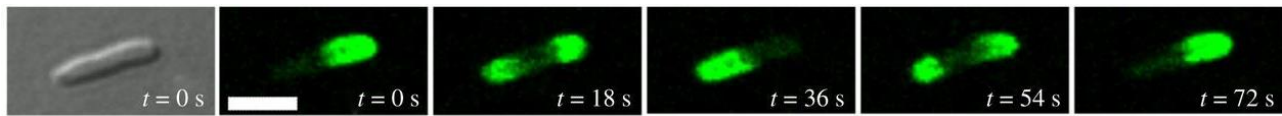
Figure 7 | Visualization of MinD oscillations. Snapshots of MinD-GFP in E. coli imaged by differential interference contrast microscopy. Scale bar, 3 μm. Adapted from[227]

Overexpression of MinD results in filamentation because endogenous MinE is not sufficient to displace all MinD molecules from the membrane, allowing the MinC-MinD complex to become homogeneously and stably membrane-bound, inhibiting FtsZ binding on the entire membrane surface and blocking cell division. Conversely, MinE overexpression suppresses the inhibition by MinC-MinD overexpression[218].

MinDΔ10 is a truncated form of MinD that lacks the C-terminal 10 amino acids constituting the membrane targeting sequence (MTS). Without the MTS, MinDΔ10 cannot associate with the membrane and therefore remains cytoplasmic. However, it maintains the ability to homodimerize[228] with the endogenous MinD. The heterodimer formed by MinD and MinDΔ10, with only one MTS, would not be able to stably bind to the membrane[229]. With MinD sequestered into the cytoplasm, endogenous MinC would no longer be recruited to the membrane, and FtsZ is free to start the assembly of the divisome at the poles, leading to the formation of small cells, with a diameter of about 400 μm, named minicells, together with variably long cells. With MinDΔ10 expressed in the cell, the nucleoid occlusion, the negative spatial regulator of Z-rings, would remain intact. This system ensures that the divisome would only form between nucleoids, so even in cells lacking Min proteins, its placement would occur only at mid-cell and the poles, and not randomly along the cell[230]. Like the parent cells, the minicells contain proteins, membranes, ribosomes, RNA, peptidoglycan, and often plasmids but no chromosome[231]. Therefore, they cannot divide or grow, but they can continue other cellular processes, such as replication and transcription of plasmid DNA, ATP synthesis, and mRNA translation.

1.8.2 MreB and RodZ

MreB is the bacterial actin homolog, necessary for the establishment and maintenance of rod shape and cell wall synthesis[232–234]. MreB preferentially locates to regions of small or even negative local Gaussian curvature[235]; it accumulates just underneath the cytoplasmic membrane, forming a spiral/banded-like pattern along the long axis of the cell[216]. MreB is also responsible for the maintenance of the overall arrangement of glycan strands in the cell wall, that are oriented helically relative to the cell's long axis[236,237]. To generate this long-range order, MreB guided growth causes cells to twist as they elongate[180]. More recently it has been observed that the helical pitch angle of MreB is a key determinant of cell diameter and that alterations to MreB affect cell shape through the MreB helical conformation[238]. Therefore, MreB polymers do not only act as a scaffold to cluster various aspects of the cell growth machinery, but also regulate specific aspects of the cell morphology.

MreB assembly is regulated by RodZ, a transmembrane protein that binds MreB, altering the conformational dynamics and intrinsic curvature of MreB polymers[239–241]. It consists of three domains: the N-terminal cytoplasmic domain, the transmembrane domain and the C-terminal

periplasmic domain. Most of the periplasmic domain is dispensable for RodZ function[240]: when MreB binds RodZ, its cytoplasmic domain undergoes to a structural change, which is then communicated to its periplasmic domain of RodZ. This structural change may also transmit information from MreB in the cytoplasm to PBP2 or RodA, proteins involved in the peptidoglycan synthesis located in the periplasm, through the transmembrane domain[242].

It has been previously established that overexpression of MreB or RodZ leads to cell elongation and thickening[234,239,243], while the loss of RodZ leads to misassembly of MreB into non-spiral structures and a consequent loss of cell shape[239].

1.9 Aim of this thesis

The AraC-controlled P$_{BAD}$ promoter is one of the inducible promoters most used by researchers in the field of microbiology and synthetic biology. While chemically inducible gene expression systems are precious tools to control biological processes for biotechnological applications and basic science, they have some limitations, that can be overcome using light instead of small molecules as the external trigger. Optogenetic gene expression systems enable sophisticated spatiotemporal regulation of a gene of interest and offer readily achievable reversibility, as light can be quickly and easily switched off.

Several light-inducible gene expression systems have been already developed for use in bacteria[42–48], and their dark/light fold-changes can be significantly high[46–48]. Nevertheless, with this work I aim to provide a valuable addition to the bacterial optogenetic tools, allowing researchers to use the same promoter, controlled by the well-known transcriptional regulator AraC, to perform either chemical or optogenetic induction.

Already other bacterial and eukaryotic transcriptional regulators have been made to respond to light[167,171,180]. When I started this project, OptoLac, the first tool that made a bacterial chemical induction system light-sensitive[167], was not published yet. I reasoned that, if AraC could become light-inducible, it would be possible for microbiologists and synthetic biologists to reversibly control, with a high spatiotemporal resolution, a great variety of biological processes that rely on gene expression in strains engineered having the P$_{BAD}$ promoter inserted in the place of an endogenous promoter[244–246]. With this tool, named BLADE, I aim to allow easy implementation of the optogenetic approach among those researchers that have already plasmids or strains responsive to L-arabinose in their repository: it would not be necessary to change the previous setup or clone new constructs, as a simple (co)-transformation of BLADE would be sufficient to make the transcription of the gene of interest light-sensitive. Alternatively, as I plan to clone BLADE in pBAD33, the plasmid typically used with wild-type AraC, the insertion of any genes of interest under light control would be straightforward, ensuring complete compatibility with other plasmids.

To characterize and show the applicability of this tool for the aforementioned purposes, I begin this thesis by reporting the engineering process of a small library of constructs. The library is composed of the light-responsive VVD photoreceptor from *N. crassa*, already employed to reconstitute Cre recombinase with light in *E. coli*[186], and the DNA-binding domain of AraC fused together through different linkers and driven by two distinct promoters. Light-mediated dimerization of VVD should

lead to AraC dimerization, which in turn should allow binding to the P$_{BAD}$ promoter and consequent induction of mCherry transcription.

I will start the characterization of these constructs by first measuring the expression level of the mCherry reporter in large cell populations by flow cytometry to have a quantitative fluorescence readout, and then by controlling the expression of MinD, an *E. coli* protein that requires tight regulation of expression. Then, I will create bacteriographs to demonstrate the spatial precision of BLADE.

Next, I aim to compare BLADE and AraC induction, highlighting the strengths and weaknesses of both systems. I will therefore measure the induction with different light intensities and L-arabinose concentrations, compare the reversibility to the uninduced state and perform kinetics of mCherry expression to discover the contribution of inducer catabolism. Lastly, I will compare the variability of the induction levels in the populations induced by BLADE and AraC.

I am also interested to investigate the mechanism of BLADE action *in vitro,* by analysing data from electrophoretic mobility shift assays, size exclusion chromatography, and *in vivo,* by GFP-tagging BLADE.

I will then express BLADE with pre-existing L-arabinose-responding plasmids and strains to verify the feasibility of replacing L-arabinose with light induction.

Finally, I aim to test BLADE with two distinct biological applications: through the transcription of genes controlling *E. coli* morphology and through the induction of a library of genes with an unknown or poorly defined function. The first one will demonstrate whether the induction with BLADE is strong enough to cause morphologic changes and tight enough to show no phenotype alteration in the dark, and the second one will present BLADE as a new alternative for the establishment of an inducible gene library, with the advantages of light as a trigger.

2. Results

2.1 BLADE engineering and characterization

2.1.1 Proposed VVD-AraC mechanism of function and its first engineering

Before the start of my PhD, during his bachelor thesis, Leonard Ernst constructed a small library of 5 chimeric proteins composed of VVD N-terminally fused to the DNA-binding domain of AraC (AraC$_{DBD}$). The idea behind these chimeras was that, in its monomeric form, the engineered VVD-AraC fusions would contact only the high-affinity I_1 half-site[247], but not the I_2 half-site, characterized by a lower affinity and needed to recruit the RNA polymerase (Figure 8b). Affinity comparison between I_1 and I_2 was already performed with two publications in 1993, with experiments showing that the replacement of I_2 with I_1 transformed P_{BAD} in a constitutive promoter[145,248]. Obtaining a light-inducible transcription factor (TF) would depend on finding the appropriate linker supporting the correct orientation of the two DBDs after dimer formation, allowing a correct I_1-I_2 binding (Figure 8). VVD was selected as the light dimerization domain because it had successfully been employed before to control with blue light the dimerization of proteins of interest[171,180,186,249] (see paragraph 1.3.5).
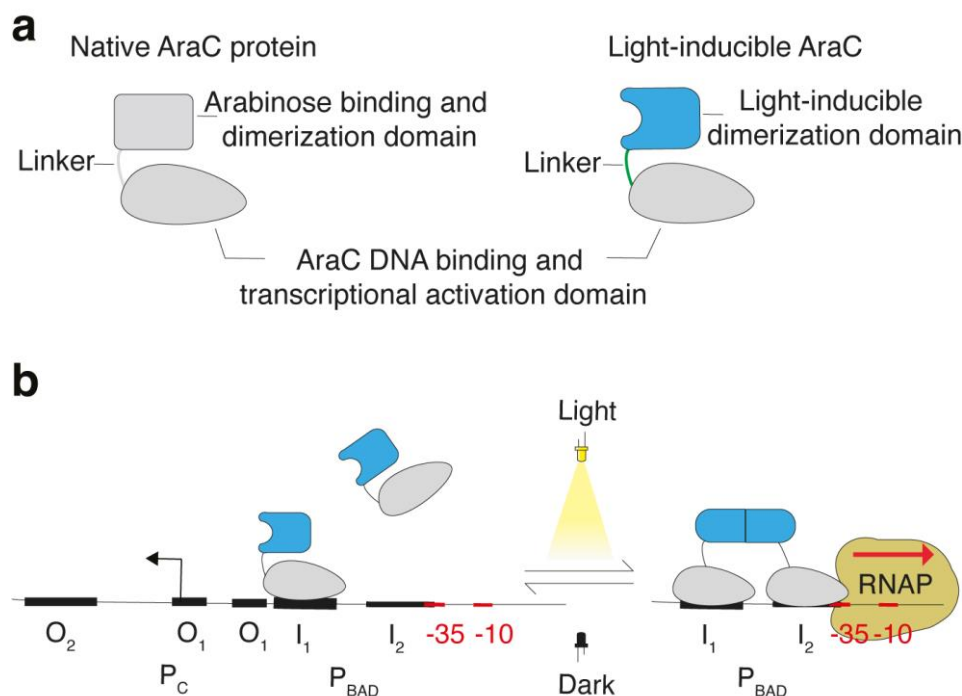
*Figure 8 | Comparison between native and light-inducible AraC. **a**, Domain composition of wild-type (left) and light-inducible (right) AraC. **b**, Expected mechanism of P_{BAD} activation by light-inducible AraC. Adapted from[250].*

His setup consisted of two plasmids: pBAD33, which he deprived of the *araC* gene, with mCherry under the inducible P_{BAD} promoter, and pDK12, in which he cloned the VVD-AraC constructs, under control of the IPTG-inducible pTrc promoter. Blue light-activated optogenetic gene expression tools

are commonly used in combination with red fluorescent proteins as reporters since these allow verifying if expression without activation occurs and to which extent[152,251]. mCherry was chosen as a fluorescent protein due to the excitation wavelength (590 nm)[252] far from VVD activation wavelength (450 nm)[90] to perform live-cell imaging at the microscope without risking to trigger VVD dimerization and following reporter transcription. VVD harboured the N56K and C71V mutations, shown to stabilize the dimer[180]. The five constructs differed in the linker between the two domains, which was taken from the natural linker (disordered region) present in the wild-type AraC protein (aa 168-178)[253] that connects AraC$_{DBD}$ to AraC$_{DD}$ (Figure 9, Supplementary Table 1). In addition, FP1 included a fragment of the C-terminal part of AraC$_{DD}$, FP4 a GS linker, and FP5 the linker from the bacteriophage lambda cI repressor protein[254] (Figure 9 Supplementary Table 1).
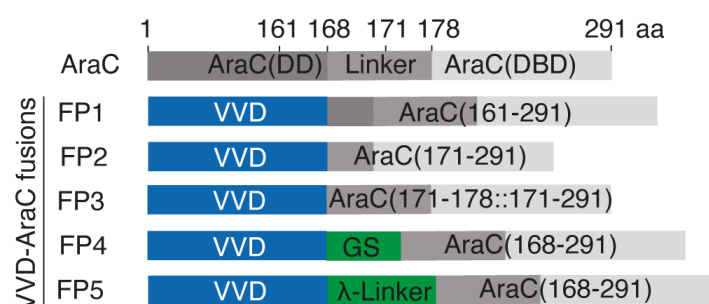
*Figure 9 | **The first library of VVD-AraC$_{DBD}$ chimeric proteins.** Domain composition of the VVD-AraC$_{DBD}$ fusion constructs created by Leonard Ernst, cloned into the pDK12 plasmid under IPTG induction.*

2.1.2 Establishment of a reproducible optogenetic setup

Leonard measured the fluorescence with a flow cytometer, an efficient way to obtain quantitative and reproducible results from large populations of cells. Despite obtaining promising preliminary results (the fold-change between dark and light induction reached 3-4x after 4 hours of blue light illumination), at the end of his project he encountered several reproducibility problems. However, his first data showed that the principle was valid. Therefore, I took the project on and set myself the goal to optimize the setup and proceed with more engineering and characterization of the library.

I started by working on the illumination and growth protocols in order to have reproducible results. First, I analysed the illumination setup used by Leonard. He employed battery-powered LEDs fixed in the lid of the falcon tubes. This was one of the causes of his irreproducible results: while by eye the LEDs seemed to emit the same light after repeated usage, a measurement of the light intensity showed a drastic decrease after their first usage. Moreover, the contact between the light and the batteries was not always optimal. As the light intensity has to be constant and uniform for every sample and in every experiment to ensure the same level of VVD-AraC activation, I set up a simple illumination device made of 6 LEDs fixed on a support (a plastic multiwell lid) connected to a power supply (Figure 10). This allowed me to precisely control the light intensity by changing the voltage applied to the LEDs. To avoid light contamination of the samples in the dark, I employed black plastic tubes, that perfectly shielded the content from light. Next, I standardized the induction protocol by illuminating the cells always at $OD_{600} = 0.4$ for 4 hours. Lastly, I incubated all the cultures with an

inhibitor of transcription and translation for 90 minutes after the induction, to obtain a complete mCherry maturation[255].
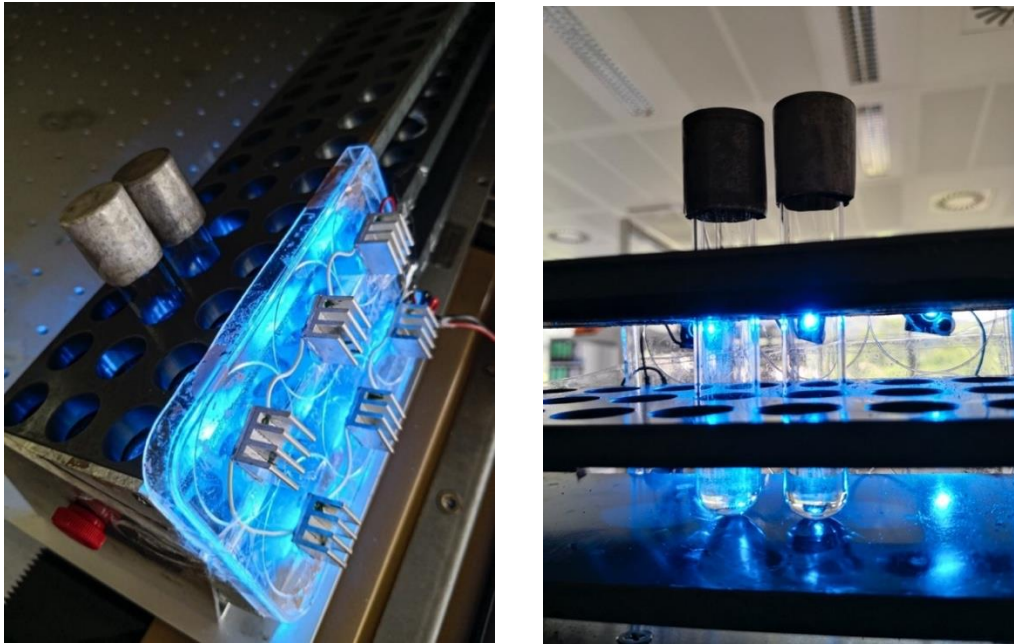
Figure 10 | Setup used to illuminate individual tubes in the 37 °C shaker incubator.

I decided to illuminate the bacterial cell cultures with 5 W/m^2 blue light. Before proceeding with the induction of the cultures, however, I verified that this intensity was not toxic for the cells. I performed a growth experiment measuring the absorbance at 600 nm of a culture maintained in the dark and of another one illuminated with 5 W/m^2 blue light (Figure 11). This experiment proved that light had no toxic effect since the growth of the two cultures was indistinguishable. Having assessed that this light intensity was not harmful to the cells, I proceeded with the characterization of the library.
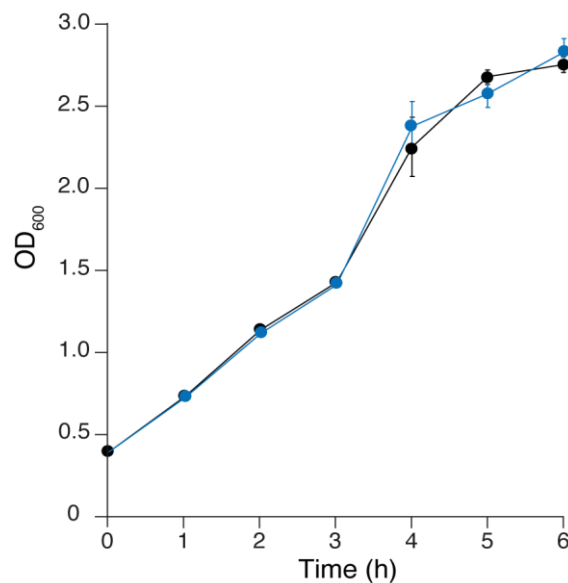
Figure 11 | 5 W/m^2 blue light is not toxic to the bacterial cells. Growth curves of E. coli MG1655 cells transformed with pBAD33 deprived of AraC, illuminated with blue light (5 W/m2; blue curve) or kept in the dark (black curve). Samples were collected every 60 min. Values represent mean ± s.d. of at least n=3 independent experiments.

I then induced the plasmids with the optimized setup, obtaining not only reproducible data but also a fold-change between dark and lit state of 9x for the construct FP4.

2.1.3 Creation of a single-plasmid setup

My next step was to use a constitutive promoter to induce the fusion constructs. The IPTG-inducible promoter was used in the preliminary experiments to test if the idea could work. This would, moreover, bring the advantage of keeping the system shut off in the absence of IPTG. However, the need for an optogenetic system to rely on IPTG for its functioning would be somewhat contradictory, as the process would then still depend on a chemical inducer. The second disadvantage of Leonard's setup was the need to co-transform the cells with two plasmids.

To overcome these issues, I introduced in the pBAD33 deprived of AraC created by Leonard the expression of the chimeric VVD-AraC$_{DBD}$ fusion constructs driven by the constitutive promoter J23101, leaving the *mcherry* gene downstream of the P$_{BAD}$ promoter as the reporter. I started to clone the FP4 construct, and after checking the sequence I realized that one or two nucleotides were deleted from the promoter sequence. Surprisingly, I did not manage, even after several attempts and with the help of my colleagues, to clone the correct J23101 sequence, as these mutations continued to arise. We hypothesized that they were introduced by the cell to weaken the expression of the promoter and diminish the burden caused by the high concentration of the chimeric construct. I decided to work with both the two mutant promoters, named J23101* and J23101** (Figure 12), as they offered two different transcription strength, and I cloned them as the promoter of all the remaining fusion constructs: I thought that two expression levels for each chimera of the library, other than the type of linker, would be an interesting parameter to study.

```
                   -35                               -10
J23101    5' -[TTTACA]GCTAGCTCAGTCCTAGG[TATAAT]GCTAGC - 3'
J23101*   5' -[TTTACA]- CTAGCTCAGTCCTAGG[TATA - T]GCTAGC - 3'
J23101**  5' -[TTTACA]GCTAGCTCAGTCCTAGG[TATA - T]GCTAGC - 3'
```
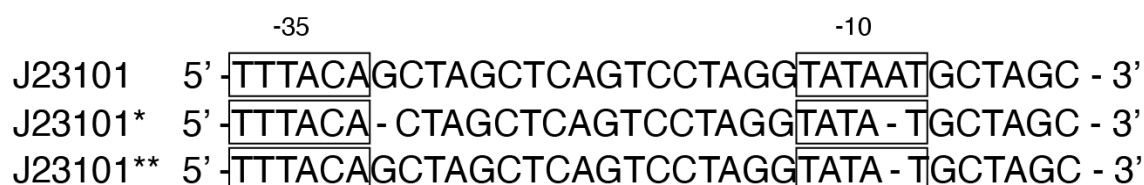
Figure 12 | Constitutive promoters used to drive the expression of the chimeric VVD-AraC fusion proteins. Alignment of the nucleotide sequences of the original J23101 promoter and the two variants that spontaneously arose in E. coli TOP10 cells after transformation of the plasmid bearing the FP4 construct. The -35 and -10 regions are indicated with a white box.

The new single-plasmid architecture with constitutive expression of the fusion constructs, other than being independent of IPTG induction, led to higher fold-change levels than the previous 2-plasmid system when the FP4 construct under J23101** was used (12x vs 9x).

I also decided to expand the library with 3 more constructs: I thought that a fusion construct with high structural similarity to the wild-type AraC could theoretically provide an optimal induction level. After comparison of the root mean square fluctuations (RMSF) of full-length AraC and FP4 model structures, generated through coarse-grained molecular dynamics simulations, performed by Mehmet Öztürk using CABS-flex 2.0 web-server[256], we found a high degree of flexibility in correspondence with the GS linker of FP4, which was not present in the wild-type AraC (Supplementary Figure 1a, b). A simple removal of the GS linker was sufficient to restore the fluctuation range of FP4 linker to

levels closer to the amino acids of AraC (Supplementary Figure 1c). This new construct was named FP6. I also cloned FP7, identical to FP4 but with 2 GS linkers, to be compared with the induction levels of FP4 and FP6, having one GS and no GS linker, respectively. Lastly, I cloned FP8, a construct without the linker to see if it remained light-activatable. In total, the library consisted of eight chimeras driven by two different promoters (Figure 13a and Supplementary Table 1).

We decided to name a generic member of this family of **B**lue **L**ight-inducible **A**raC **D**imers in *E. coli* **BLADE** and the pBAD33-derived corresponding expression plasmid pBLADE (Figure 13b).
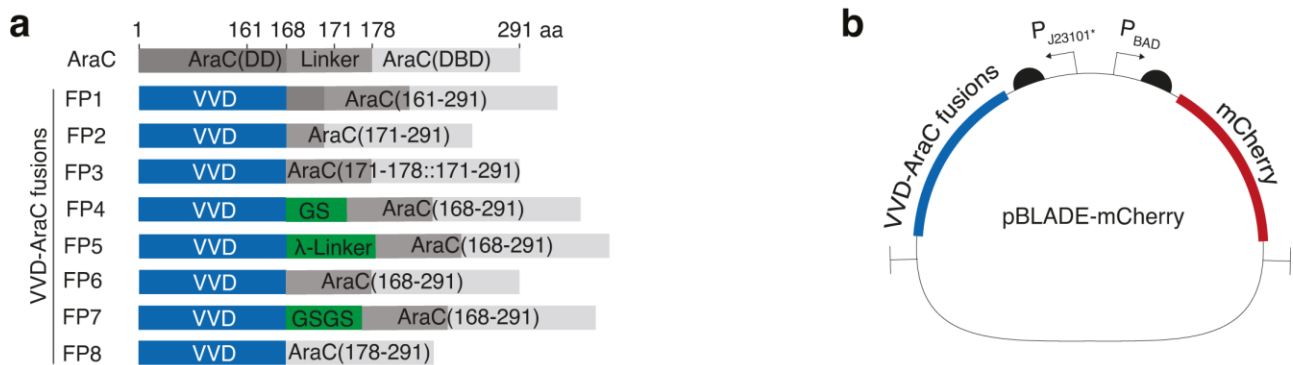
Figure 13 | Construction of the BLADE library. a, Domain composition of the engineered VVD-AraC$_{DBD}$ fusion constructs. Constructs FP1-FP5 are the same as in the first library constructed by Leonard Ernst. b, Simplified map of the plasmid for the expression of a gene of interest (here mCherry) under the control of BLADE. The promoter driving the expression of the VVD-AraC$_{DBD}$ fusion is either J23101* (depicted) or J23101**. Adapted from[250].

As a positive control, I constructed the same plasmid carrying full-length AraC in place of the VVD-AraC fusion construct under the same two constitutive promoters used for the library (Figure 14a). As negative control, I constructed a plasmid without any TF, which I named pReporter_only, (Figure 14b).
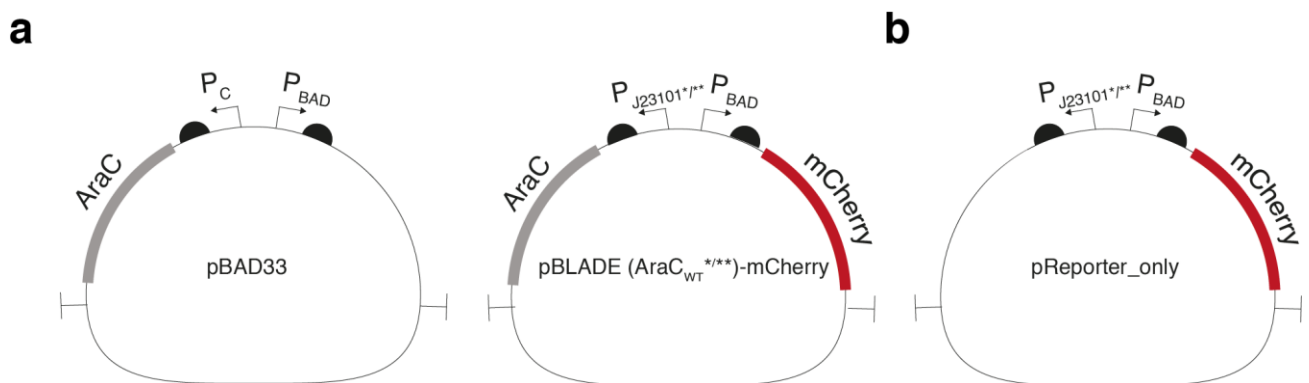
Figure 14 | Simplified maps of the original pBAD33 and the negative and positive control derivatives used in this study. a, The transcription factor AraC in pBAD33 is transcribed from the P$_C$ promoter, regulated via negative feedback from AraC. The positive controls pBLADE(AraC$_{WT}$*)-mCherry and pBLADE(AraC$_{WT}$**)-mCherry are obtained from pBAD33 with the constitutive promoters J23101* and J23101** in place of P$_C$ to drive the expression of AraC (Supplementary Figure 2). b, The negative control pReporter_only contains the P$_{BAD}$ promoter driving the expression of mCherry and the J23101* promoter but no transcription factor (Supplementary Figure 2).

2.1.4 Quantification of BLADE-mediated gene expression via fluorescence measurement

After transformation of *E. coli* MG1655 cells with the BLADE library, I either kept the cultures in the dark or illuminated them for 4 hours with 460 nm light (5 W/m^2). I then collected the samples and measured the corresponding fluorescence using a flow cytometer. I saw that all the 14 VVD-AraC constructs having an interdomain linker were light-inducible, despite showing a lower level of reporter expression than full-length AraC (Figure 15 and Supplementary Table **2**).

Different linkers influenced the amount of gene expression, with the construct devoid of the linker (FP8) showing no light sensitivity, and the construct harbouring the longest linker (FP5, 39 amino acids long) leading to a modest expression level and the lowest fold-change between dark and light (Figure 15 and Supplementary Table 2). With the weaker constitutive promoter (J23101*) driving expression of the VVD-AraC$_{DBD}$ fusion constructs, the levels of reporter expression in the dark were similar to those of the negative control, to which the values were normalized. The stronger constitutive promoter (J23101**) led to significantly higher expression of the reporter gene after blue light illumination for all constructs, albeit at the cost of increased leakiness in the dark (Fig. 8a and Supplementary Table 2). However, for some of the fusions, the light/dark fold-change was higher with this promoter (Supplementary Table 2).
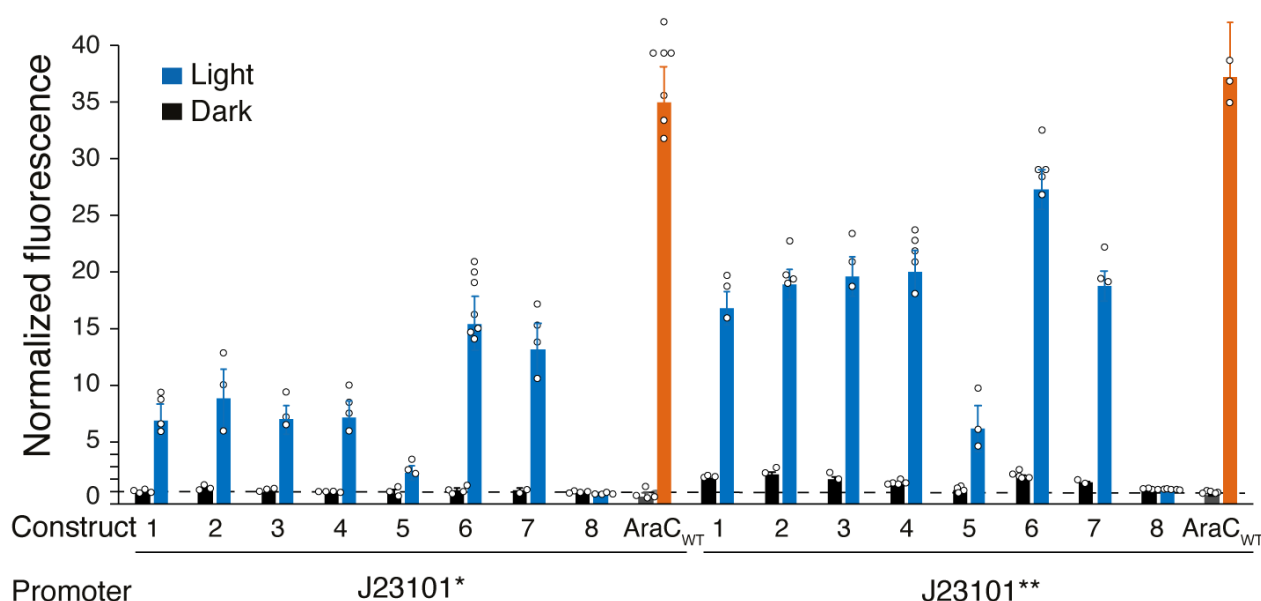
Figure 15 |The interdomain linker plays a role in the light-inducibility of BLADE. mCherry fluorescence intensity in E. coli MG1655 cells transformed with the library shown in Figure 13a grown for 4 h either in the dark or under 460 nm blue light at 5 W/m^2. Wild-type AraC, cloned under the same constitutive promoters, was used as a positive control (AraC$_{WT}$, see Supplementary Figure 2). The bars for AraC represent the values obtained without (grey) and with 0.1% (orange) L-arabinose for 4 h. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at least n=3 independent experiments.

I then measured mCherry levels obtained at different light intensities to assess the requirement of BLADE in terms of blue light. I found that the induction plateau was reached at 2 W/m^2 and 1 W/m^2 was already sufficient to obtain reporter gene expression levels close to the saturation value (Figure 16). However, as a precaution, and having assessed the non-toxicity of 5 W/m^2 of blue light for 6 hours, I kept inducing the samples at this light intensity.
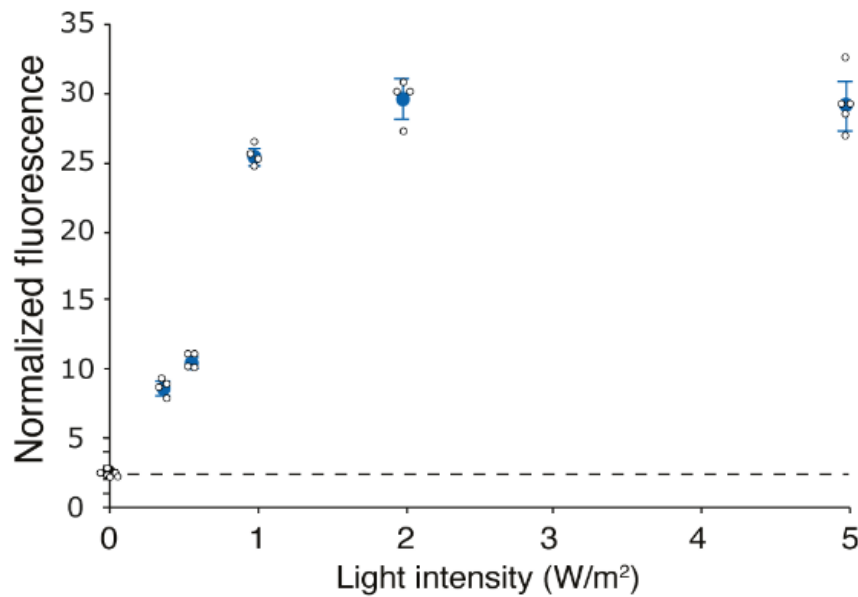
Figure 16 | Light intensity can be tuned to obtain different induction levels*. mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with the FP6 fusion driven by the J23101** promoter grown for 4 h under 460 nm light of the indicated light intensity (blue) or kept in the dark for 4 h (black). All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at least n=3 independent experiments.*

2.1.5 Visualisation of MinD oscillations

Next, I aimed to demonstrate the ability of BLADE to control the expression not only of fluorescent reporters but also functional *E. coli* proteins. I had to select a protein whose levels must be tightly controlled in order not to disturb the normal *E. coli* morphology. As explained in paragraph 1.8.1, MinD is involved in the determination of mid-cell, important for the generation of two equally-sized daughter cells. Importantly, an excessive overexpression of MinD would disrupt this process and induce cell filamentation. For this reason, I deemed MinD a good candidate to check the tightness of the BLADE system and I cloned *eyfp-minD* into pBLADE.

After 2 hours of induction, time-lapse fluorescence microscopy showed that eYFP-MinD oscillations were present only in cells illuminated with blue light and not in those kept in the dark when BLADE FP6 driven by the J23101* promoter (Figure 17a). The distribution of the cell length for both non-induced and induced samples was comparable to that of the same strain transformed with the negative control (Figure 17b). These data indicated that exists a BLADE family member that, under the right promoter, can be safely used to regulate the expression of a gene of interest that requires tightness. Conversely, BLADE FP4, when expressed by J23101**, did not manage to tightly control eYFP-MinD oscillations (Supplementary Figure 3). To challenge the system even more, I decided to clone a toxin in place of MinD.

a

b
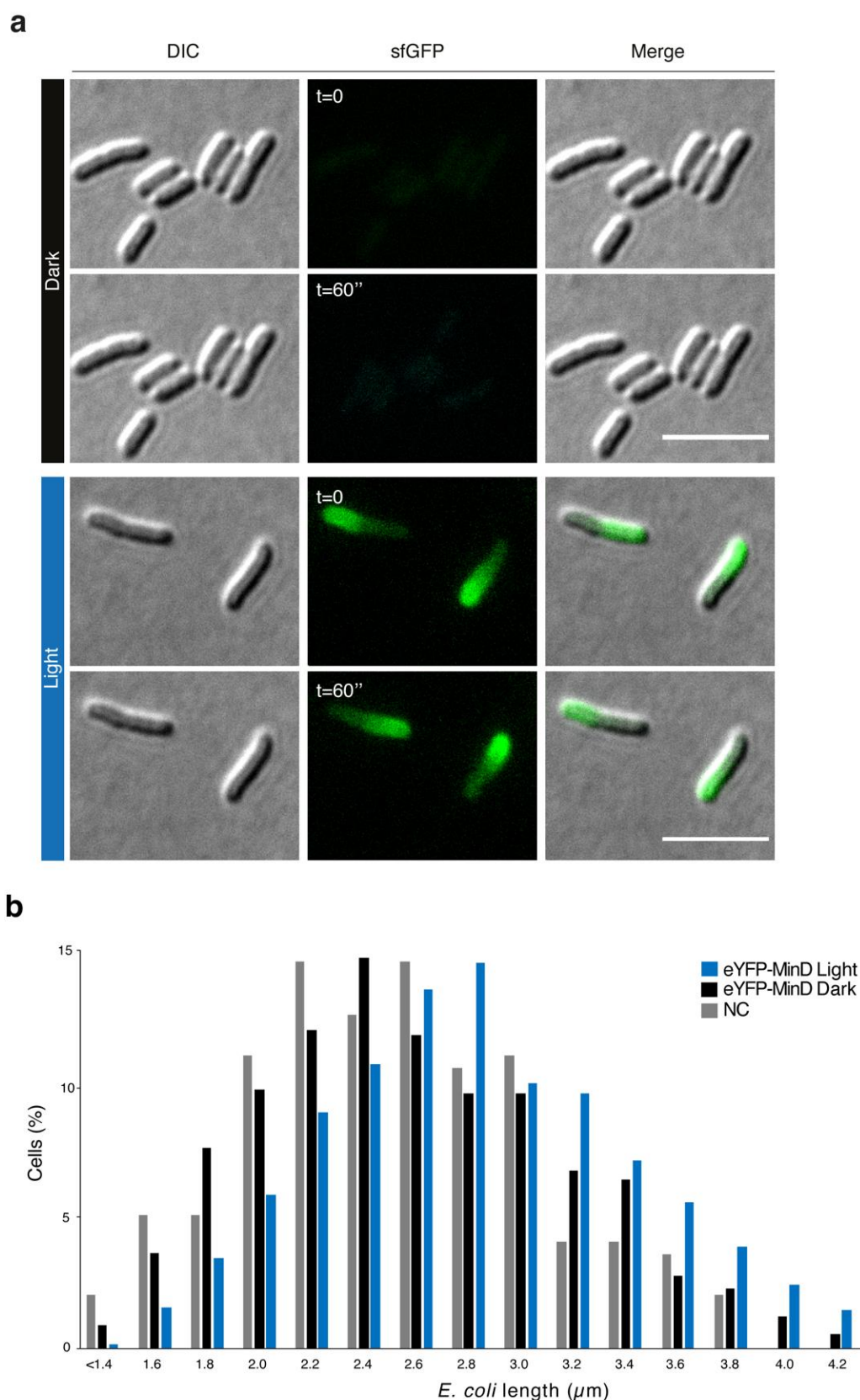
Figure 17 | BLADE can be used to express MinD. a, *E. coli MG1655 cells transformed with pBLADE(FP6*)-eYFP-MinD were illuminated with blue light (5 W/m²) for 2 h or kept in the dark for the same time and imaged with the fluorescence microscope. The images show two snapshots of a time-lapse microscopy experiment at the indicated time points. Scale bar, 5 μm.* **b**, *Quantification of the cell length distribution for MG1655 cells transformed with pBLADE(FP6*)-eYFP-MinD under the indicated conditions. The negative control NC (pReporter_only; see Supplementary Figure 2) was left in the dark.*

2.1.6 Expression of *ccdB*

As a toxin, I decided to take CcdB. This protein is a component of the CcdA/CcdB type II toxin-antitoxin system located on the F-plasmid of *E. coli*[257]. It affects the DNA gyrase, which performs double-strand breaks in the DNA backbone, relaxing positive supercoils and resealing the gaps. The CcdB protein traps the DNA gyrase with broken double-stranded DNA, causing DNA breakage and cell death[258].

Controlling the expression of the *ccdB* gene with BLADE would, beyond proving the tightness of the system beyond doubt, allow triggering cell death with light, which could be very useful. I first cloned the *ccdB* gene into pBLADE downstream of P_BAD, using DB3.1, a cell strain resistant to the toxic effect of this gene. However, after the transformation of pBLADE-CcdB into MG1655, I discovered that I could not induce cell death with light, as the coding sequence of CcdB acquired mutations to prevent the expression of the toxin. Interestingly, I encountered the same problem with *ccdB* cloned into pBAD33. These data suggest that the leakiness of the P_BAD promoter, albeit acceptable in many cases, is too high when toxic genes are to be controlled. The CcdB molecules present in the uninduced state are sufficiently damaging to impose a selective pressure on the cells to insert premature stop codons, point mutations or partial deletions to the *ccdB* gene sequence to grow normally. I concluded that to regulate the expression of very toxic genes, an extremely tight promoter is needed. Surprisingly, an example of optogenetic regulation of CcdB with light was reported in 2016: in that case, the transcriptional regulator was LexA-VVD (see paragraph 1.5.1.2), a fusion between VVD and the DBD of the LexA repressor of the *E. coli* SOS regulon. When compared to the AraC induction system, LexA-VVD showed a lower expression level in the absence of the inducer[171].

2.1.7 Spatial control of gene expression

One of the benefits of optogenetic induction is the ability to modulate gene expression in a spatially dependent fashion. To showcase how BLADE could be used to control the expression of a target gene only in selected cells, I cloned superfolder GFP (sfGFP)[259], a fluorescent protein with a maturation half-time of 6 minutes and brightness higher than mCherry, into pBLADE. I then transformed *E. coli* MG1655 cells with pBLADE-sfGFP and applied a sample to an agar pad, which was then subjected to confocal microscopy to expose a limited area (6.4 μm^2) to blue light every 5 minutes. After 3 hours, sfGFP was expressed up to 6.7-fold more in the illuminated cells compared to the surrounding non-illuminated cells (Figure 18).
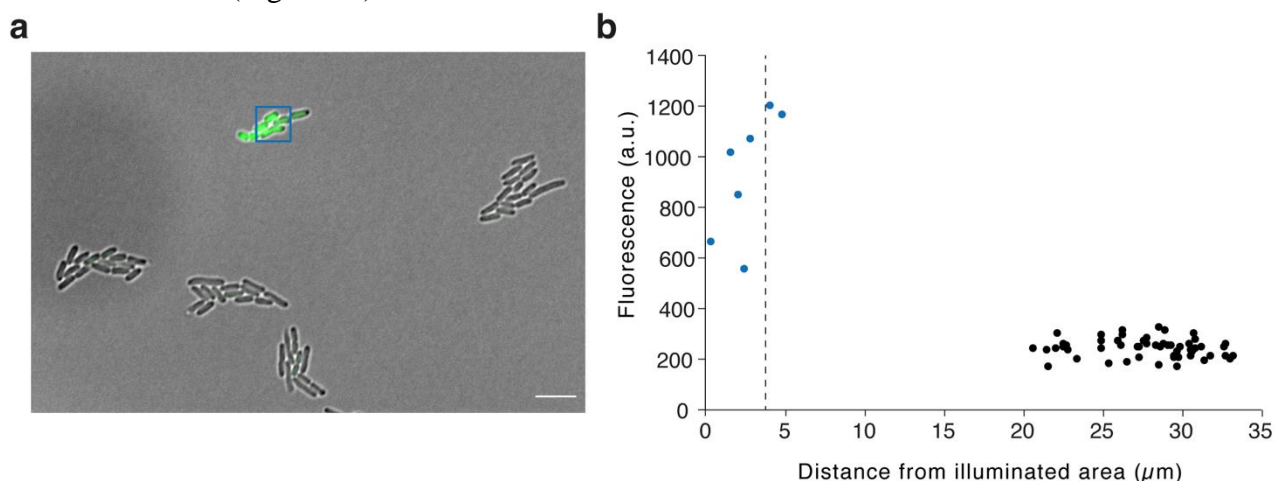
Figure 18 | Spatial control of gene expression is possible with BLADE. a, Representative microscopy image representing the overlay of the DIC and GFP fluorescence channels of E. coli MG1655 cells transformed with pBLADE(FP6**)-sfGFP took after 3h of blue light illumination of the area indicated by the blue square. Scale bar, 5 μm. b, Quantification of the single-cell GFP fluorescence within the illuminated microcolony (blue circles) and within non-illuminated microcolonies (dark circles) as a function of their distance from the centre of the illuminated area (dashed line).

Another interesting application of light-inducible TFs that relies on the possibility to shine a desired pattern of light on a plate is called bacterial photography[159] (see 1.6). To evaluate the efficacy of BLADE in this type of application, I covered a lawn of *E. coli* MG1655 cells transformed with pBLADE-sfGFP with a photomask depicting the Freiburg University logo (Figure 19a) and the protein model structure of BLADE (Supplementary Figure 4). I illuminated the plate with blue light overnight, then took several microscopy pictures, and stitched them together. The result was a bacteriograph of good quality, despite not all details were reproduced (e.g. the wings of the eagles on the left side were not visible) (Figure 19b).
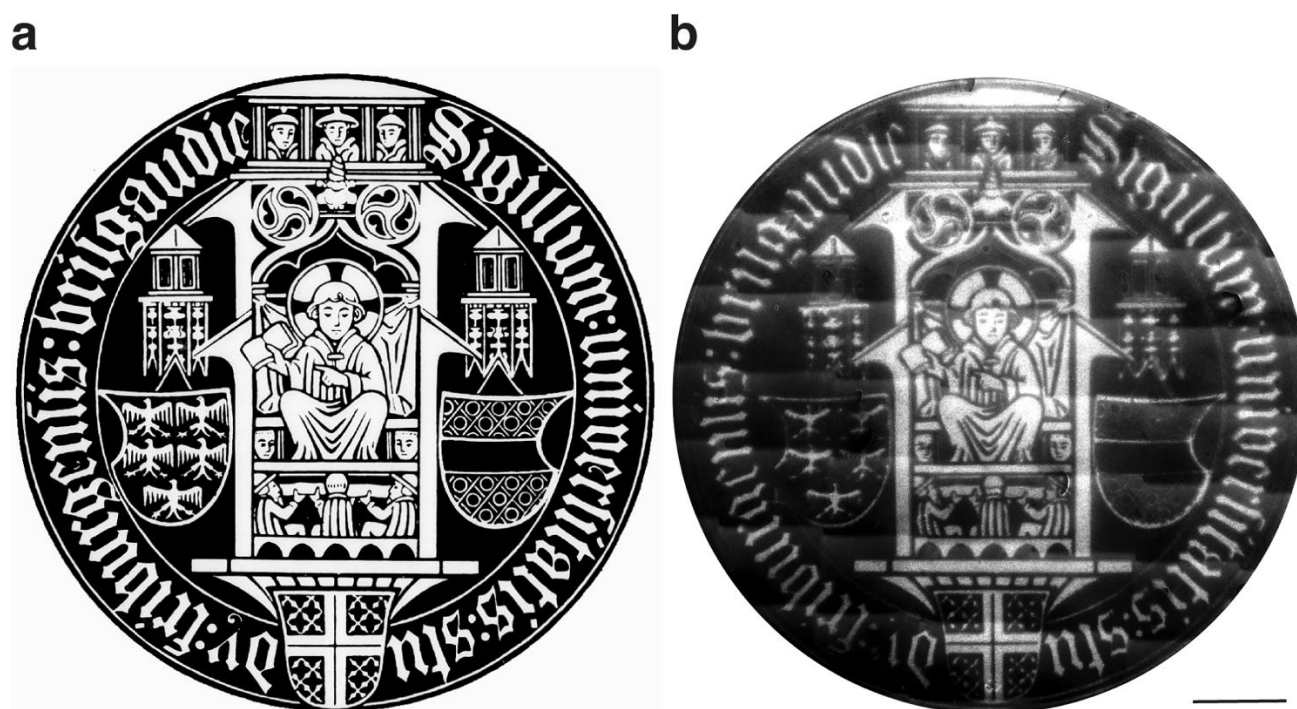
Figure 19 | Reproduction of the logo of the University of Freiburg using BLADE to spatially activate gene expression. a, Photomask used to produce the bacteriograph in b. b, Bacteriograph. A lawn of E. coli MG1655 cells transformed with pBLADE(FP6*)-sfGFP was grown overnight at 37 °C while being exposed to blue light through the photosmask in (a). 72 individual images were taken with a fluorescent microscope and stitched together via image analysis software. Scale bar, 1 cm.

To challenge BLADE with a more complex bacteriograph, I chose to reproduce the Blade Runner movie poster, given the obvious connection with the name of the tool, BLADE. I printed a black and white photomask (Figure 20a) and, with the help of Emir Bora Akmeriç, we employed the same protocol described for the previous figure. Thanks to the experience acquired with the previous bacteriograph, and after many trials (a non-homogeneous bacterial layer would result in irregularities of the final result), we managed to obtain a high-quality image (Figure 20b), with faithful reproduction of the smallest details (Figure 20c).

Figure 20 | Reproduction of the Blade Runner movie poster using BLADE to spatially activate gene expression. a, Photomask used to produce the bacteriograph in b (printed with permission from Warner Bros. Entertainment, Inc.). b, Bacteriograph. A lawn of E. coli MG1655 cells transformed with pBLADE(FP6*)-sfGFP was grown overnight at 37 °C while being exposed to blue light through the photosmask in (a). 110 individual images were taken with a fluorescent microscope and stitched together via image analysis software. Scale bar, 1 cm. c, Zoom in on two parts of the bacteriograph shown in (b). Scale bar, 300 μm.

Lastly, I aimed to create another complex image, potentially more difficult to obtain for the lack of high contrasts. I converted a picture of the "Creation of Adam" fresco by Michelangelo in black and white (Figure 21a) and I used it as a photomask for a bacteriograph (Figure 21b).

Surprisingly, also in this case BLADE managed to deliver a high level of details on both bright and dark areas of the image; moreover, as the photomask was printed by a different print shop, this time the ink dots that compose the image were more visible through a magnification of the bacteriograph (Figure 21c).
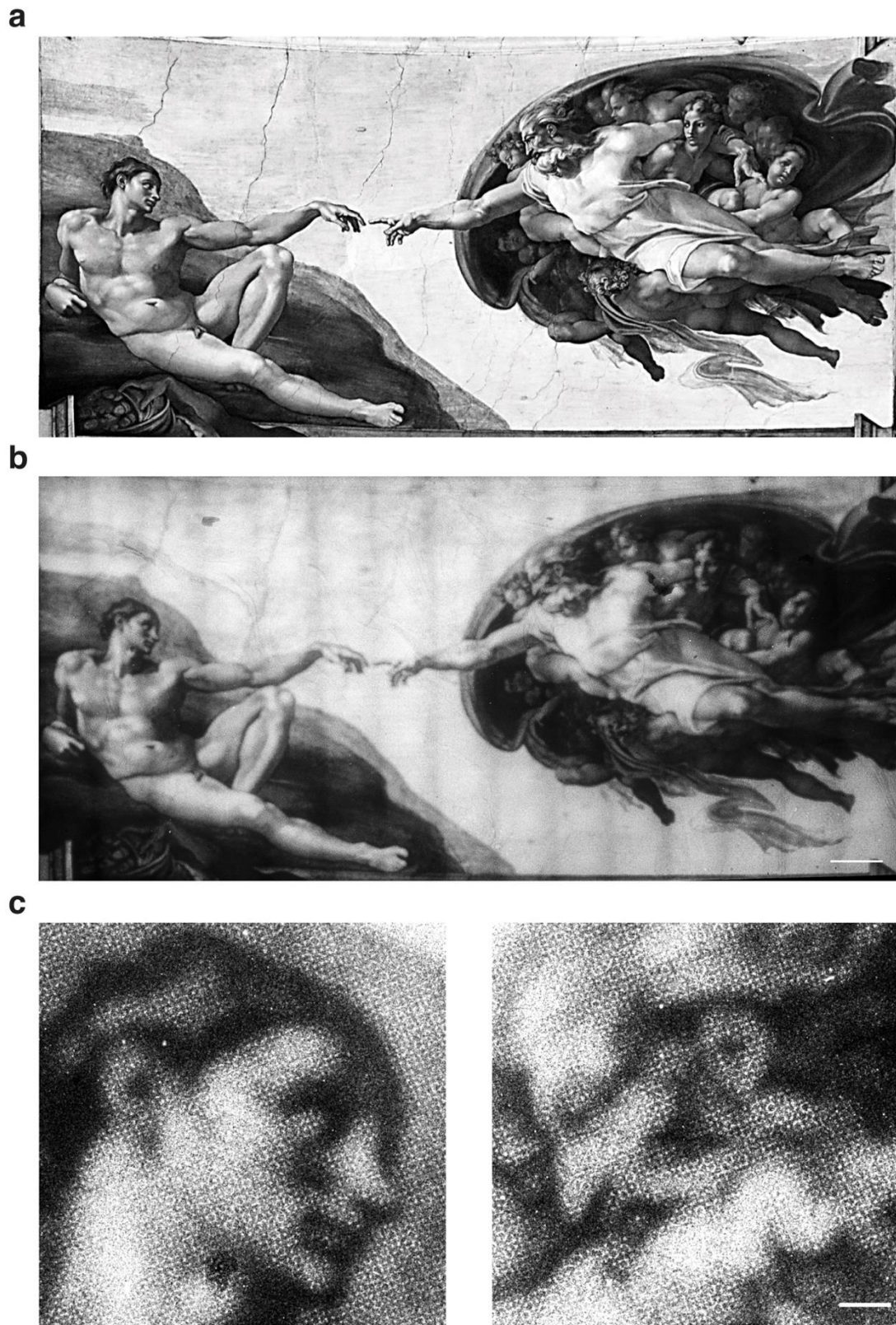
Figure 21 | Reproduction of Michelangelo's "Creation of Adam" using BLADE to spatially activate gene expression.
A lawn of E. coli MG1655 cells transformed with pBLADE(FP6)-sfGFP was grown overnight at 37 °C while being exposed to blue light. **a**, Photomask used to produce the bacteriograph in b. **b**, bacteriographs. Scale bar, 1 cm. **c**, The ink dots composing the photomask are visible with a zoom on two parts of the bacteriograph. Scale bar, 500 nm.*

2.1.8 Comparison between AraC and BLADE

In this work, I aim to demonstrate the advantages of light over L-arabinose to induce gene transcription. I have already compared the induction levels obtained with AraC and BLADE and showed that the chemically inducible AraC leads to higher reporter expression than BLADE (Figure 15). This parameter has relative importance, as in many cases the protein to be expressed must remain within physiological levels. As a matter of fact, during the engineering of BLADE, I focused more on keeping the basal expression as low as possible than increasing the induction level, as the leakiness value is crucial when the expression of a gene must be temporally defined. In the previous paragraph, I already showed the level of spatial control obtained using BLADE, which cannot be matched by a system based on a chemical inducer such as AraC. Aside from this feature, an optogenetic tool like BLADE offers additional advantages over AraC, described below.

2.1.8.1 Absence of inducer catabolism

As mentioned in paragraph 1.4.2, L-arabinose is catabolized by *E. coli* if richer carbon sources are not present in the growth medium. The medium used for all the experiments, tryptone broth, contains only tryptone as the energy source, thus it is to be expected that L-arabinose will be used up as carbon source once present in the medium. This was reflected in the kinetics of mCherry induction I obtained with pBAD33, whereby after the first hour of incubation with 0.01%, 0.001% and 0.0001% of L-arabinose, a progressive reduction of fluorescence over time is evident (Figure 22a). This reveals that the L-arabinose catabolism started already within the first hour. Therefore, the inducer concentration decreases and this, combined with the fast growth rate of the culture, leads to a quick depletion of the intracellular L-arabinose concentration. As a consequence, less L-arabinose is available for AraC to turn P_{BAD} on. More accurate measurements with sampling every 20 minutes, performed within the first 90 minutes of induction, revealed that the fluorescence level of the samples induced with 0.01% and 0.001% of L-arabinose peaked exactly after 1 hour (Figure 22b). Due to the higher inducer concentration, the culture supplemented with 0.1% L-arabinose allowed for the induction to proceed until the third hour, before it started decreasing.
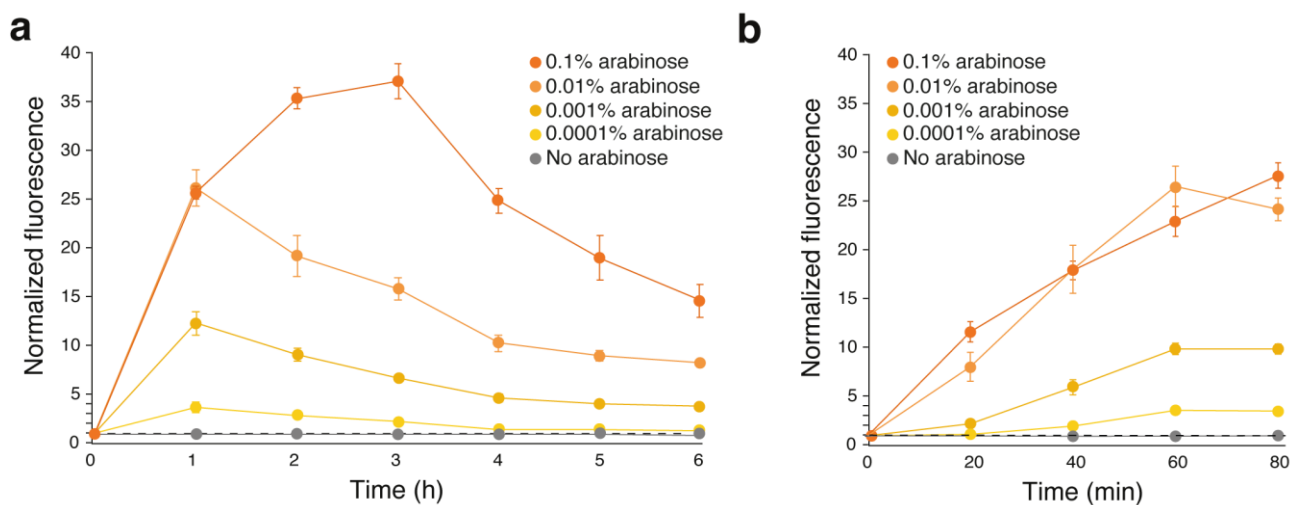
Figure 22 | L-Arabinose is quickly catabolized by the cell. *Kinetics of mCherry expression in E. coli MG1655 cells transformed with pBAD33 and induced at t_0 with different L-arabinose concentrations or no L-arabinose. Samples were taken **a**, every hour, **b** every 20 minutes. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (see Supplementary Figure 2; dashed line).*

To overcome the inducer catabolism problem, a solution would be to use a microfluidic system that continuously supplements the culture with L-arabinose. This would sustain the induction over time, but if changes in the catabolism levels happen during the induction, they will lead to L-arabinose concentration different than at the beginning. Lacking such an instrument in our lab, I supplemented the culture with the initial L-arabinose concentration every hour for 6 hours. I chose to induce the culture with 0.0001% L-arabinose because this low concentration is likely to be completely catabolized after a short time, and a constant supplementation may increase the induction level after the first hour of induction. Surprisingly, the plateau was reached already after the second hour of induction, with a fluorescence increase of about 25% from the first hour (Figure 23).
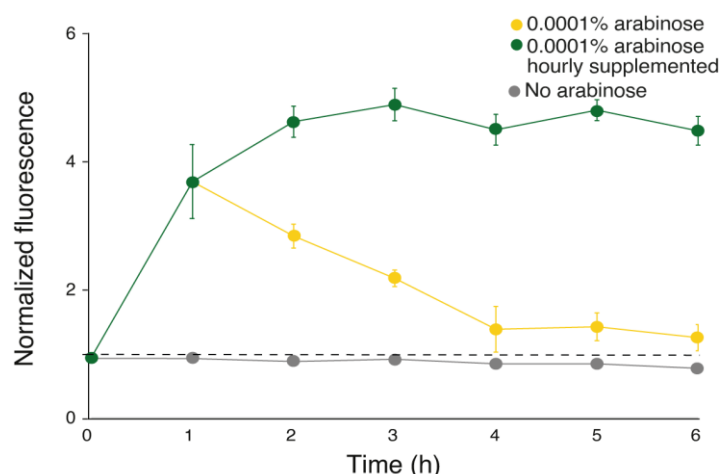
Figure 23 | Kinetics of AraC-mediated mCherry expression with L-arabinose supplemented every hour. E. coli MG1655 cells transformed with pBAD33-mCherry were supplemented at t_0 and after every hour with 0.0001% L-arabinose. Values represent mean ± s.d. of at least n=3 independent experiments. Samples were taken every hour. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line; see Supplementary Figure 2).

To compare these data to those obtained with light as the inducer, I measured the kinetics of mCherry expression from pBLADE under three different light intensities. I found that the half-maximum was reached between 2.5 and 3 hours of induction with light, while the levels plateaued after 5 hours of light induction at every light intensity (Figure 24). Differently to L-arabinose, light is not catabolized, therefore the reporter gene expression increased over time to eventually reach the steady-state.
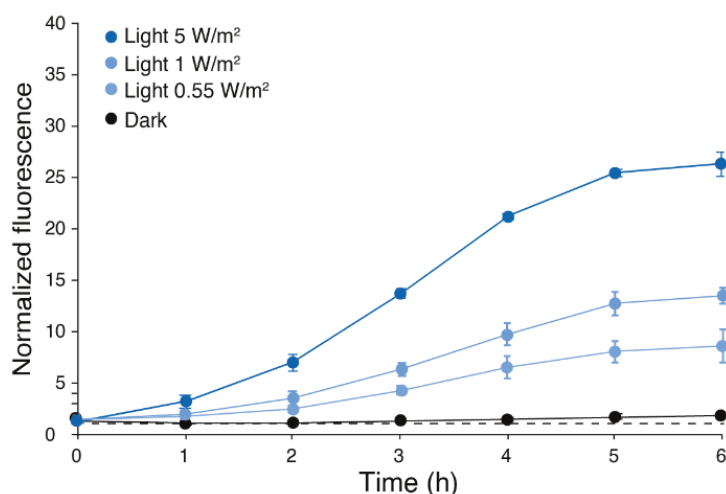
Figure 24 | Kinetics of BLADE-mediated mCherry expression under different light intensities. a, E. coli MG1655 cells transformed with the FP6 fusion driven by the J23101* promoter (pBLADE(FP6*)-mCherry) were grown for 4 h either in the dark or under 460 nm blue light illumination at the indicated light intensities. Values represent mean ± s.d. of at least n=3 independent experiments. Samples were taken every hour. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line; see Supplementary Figure 2).

Mutant *E. coli* strains that do not catabolize L-arabinose are reported in the literature: I used SKA703, a MG1655 strain characterized by having a*raCBAD*, *araE* and *araFGH* genomically deleted[260]. I transformed this strain with pBLADE(AraC*)-mCherry and observed, after 4 hours of induction with 0.1% L-arabinose, reporter gene expression levels much higher than those obtained with the same plasmid in MG1655 (Figure 25).
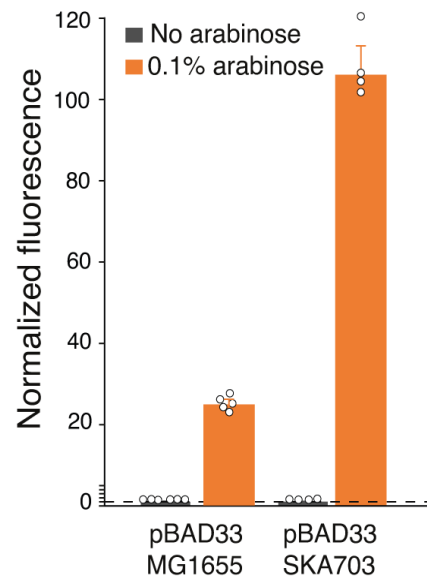
Figure 25 | In the absence of L-arabinose catabolism, the induction level is higher. mCherry fluorescence intensity in E. coli MG1655 or SKA703 cells transformed with pBLADE(AraC_{WT}*)-mCherry grown for 4 h either with 0.1% (orange bars) or without L-arabinose (grey bars). All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with the plasmid deprived of the transcription factor, pReporter_only (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at least n=3 independent experiments.

2.1.8.2 Complete reversibility

One of the main advantages of light as the inducer is the possibility to turn it off easily, without the need for potentially damaging and time-consuming washing steps, enabling complete reversible control of gene expression.

To compare the performance of AraC and BLADE in terms of reversible gene expression, I exposed *E. coli* MG1655 cells transformed with pBLADE-mCherry or pBAD33-mCherry to alternating 2 hours-cycles of induction (with blue light or L-arabinose, respectively) and rest (via incubation in the darkness or growth in a medium without L-arabinose, respectively), for a total of 3 induction cycles and measured mCherry levels via flow cytometry at the end of every phase.

To remove as much L-arabinose as possible, after every induction cycle, I gently pelleted and resuspended the culture with fresh medium without L-arabinose twice. The same expression levels were reached with BLADE after every illumination cycle, and the fluorescence level returned to the same initial value after every resting phase (Figure 26a). On the other hand, for the cultures

52

transformed with pBAD33-mCherry and induced with L-arabinose, the fluorescence level after the resting phases did not return to the initial levels (Figure 26b-d), especially when a high concentration of L-arabinose was used, leading to up to 5.1-fold-change in respect to the fluorescence at t_0 (Figure 26e).

e

	t_4/t_0	t_8/t_0
pBLADE - light	1.1x	1.3x
pBAD33 - 0.1% arabinose	5.0x	5.1x
pBAD33 - 0.01% arabinose	3.9x	4.3x
pBAD33 - 0.001% arabinose	2.0x	2.5x
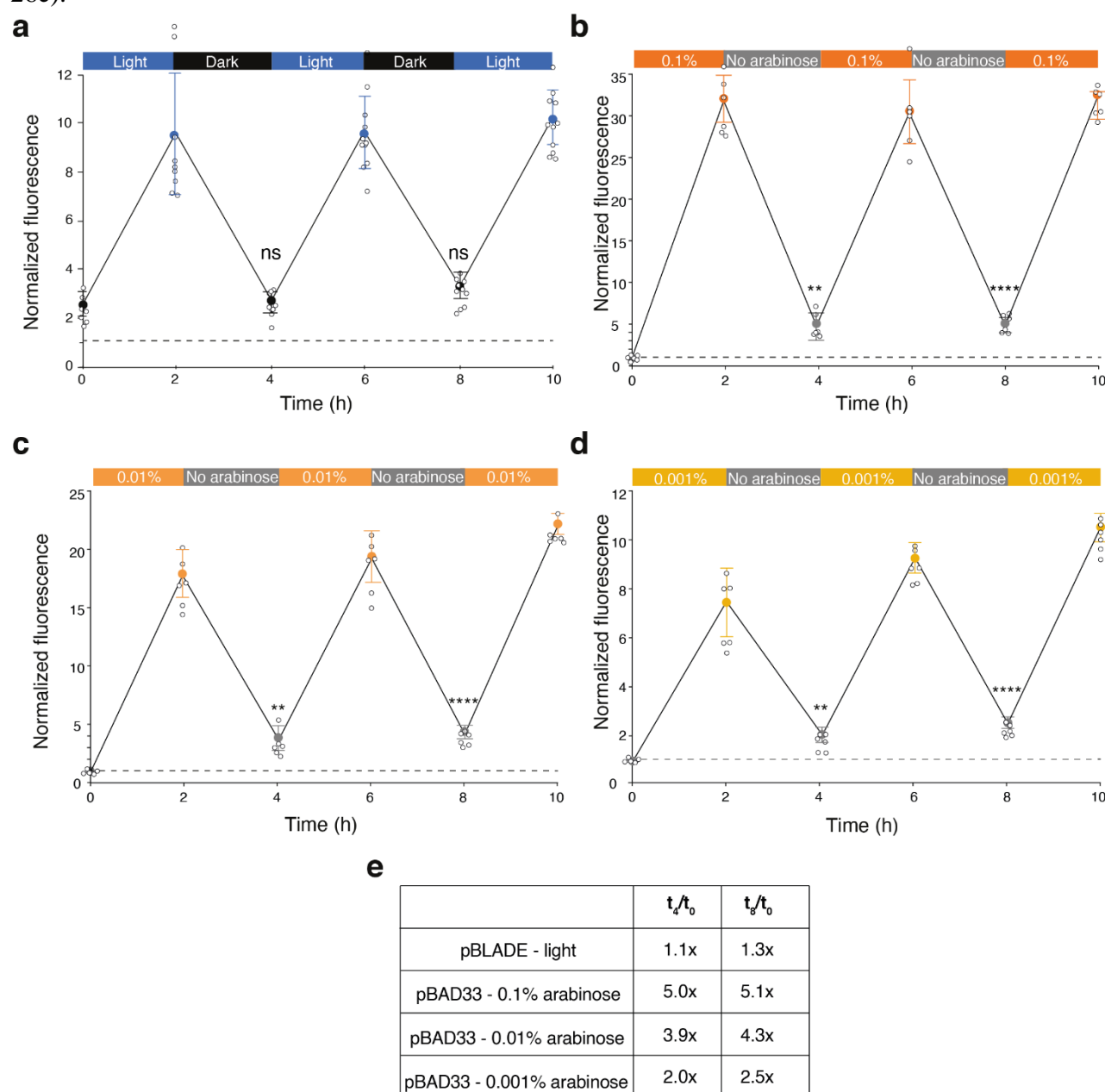
*Figure 26 | Fluorescence levels return to the initial value after BLADE induction is turned off. **a**, mCherry fluorescence intensity in E. coli MG1655 cells transformed with pBLADE-mCherry after repeated cycles of blue light exposure and darkness. **b-d**, mCherry fluorescence intensity in E. coli MG1655 cells transformed with pBAD33-mCherry after repeated cycles with the indicated L-arabinose concentration and without L-arabinose. **e**, Table reporting the fluorescence ratio of each uninduced cycle from **a-d** to the measurements at their t_0. **a-d**, All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line; see Supplementary Figure 2). Values represent mean ± s.d. of at least n=3 independent experiments. Not significant (ns), p-value >0.05; single asterisk (*), p-value <0.01; double asterisk (**), p-value <0.001; triple asterisk (***), p-value <0.0001; quadruple asterisk (****), p-value <0.00001. P-values were calculated by the two-tailed, homoscedastic Student's t-test. BLADE construct: FP6 fusion driven by the J23101* promoter.*

2.1.8.3 Lower population heterogeneity

Beyond the problem with the catabolism of L-arabinose, another issue arising when using low L-arabinose concentrations in a wild-type strain is the heterogeneity in reporter gene expression in the population (Figure 27a). This happens when the L-arabinose concentration in the medium is not high enough to allow each cell to import the same amount.

I found that the coefficient of variation (CV) increased for the samples induced with progressively lower L-arabinose concentrations. I induced pBAD33-mCherry with 4 different L-arabinose concentrations for 4 hours: while between 0.1% and 0.01% the difference in the expression heterogeneity did not change significantly, remaining at around CV levels of 60-70%, lowering the L-arabinose concentration to 0.001% and 0.0001% led an increase of the CV to 90% and 110%, respectively (Figure 27b).
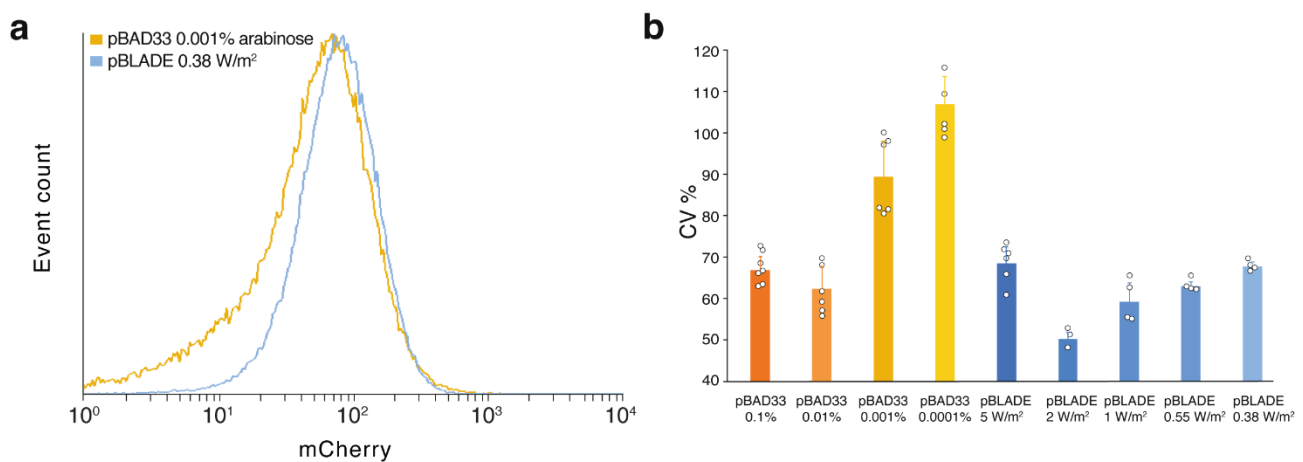
Figure 27 | Heterogeneity of reporter expression remains constant using BLADE with different light intensities. a, Representative histograms showing the distribution of the mCherry fluorescence within a population of E. coli MG1655 cells transformed with either pBAD33-mCherry (orange) or pBLADE-mCherry (pale blue) and induced for 4 h with either 0.001% L-arabinose or with 0.38 W/m² of 460 nm blue light. b, Coefficient of variation (CV) of the mCherry fluorescence levels measured by flow cytometry in MG1655 cells transformed with either pBAD33-mCherry (shades of orange) or pBLADE-mCherry (shades of blue) and induced for 4 h with the indicated concentrations of L-arabinose and light intensities, respectively. Values represent mean ± s.d. of at least n=3 independent experiments. BLADE construct: FP6 fusion driven by the J23101* promoter.

On the other hand, lower light intensities activating BLADE did not increase the gene expression heterogeneity, as the CV calculated on the population of cells subjected to 0.38 W/m² blue light was comparable to that of the population subjected to 5 W/m² blue light (Figure 27a), with the result that the variability never surpassed that obtained for the population induced with 0.1% L-arabinose (Figure 27b). In an ideal case, with no fluorescence variation among the cells, the CV value should be equal to 0. With my experiments, the measured CV was never below 50%: this variability was probably due to the intrinsic stochasticity of gene expression.

The histograms in Figure 28 show a comparison of representative mCherry fluorescence distributions in cell populations induced with L-arabinose or light. Even qualitatively, it is possible to appreciate the broadening of the distribution with the progressive decrease in the L-arabinose concentration, with a slight bistable population appearing in the population induced with 0.01% L-arabinose (Figure 28c). Conversely, the distributions remain comparable when lowering the light intensity.
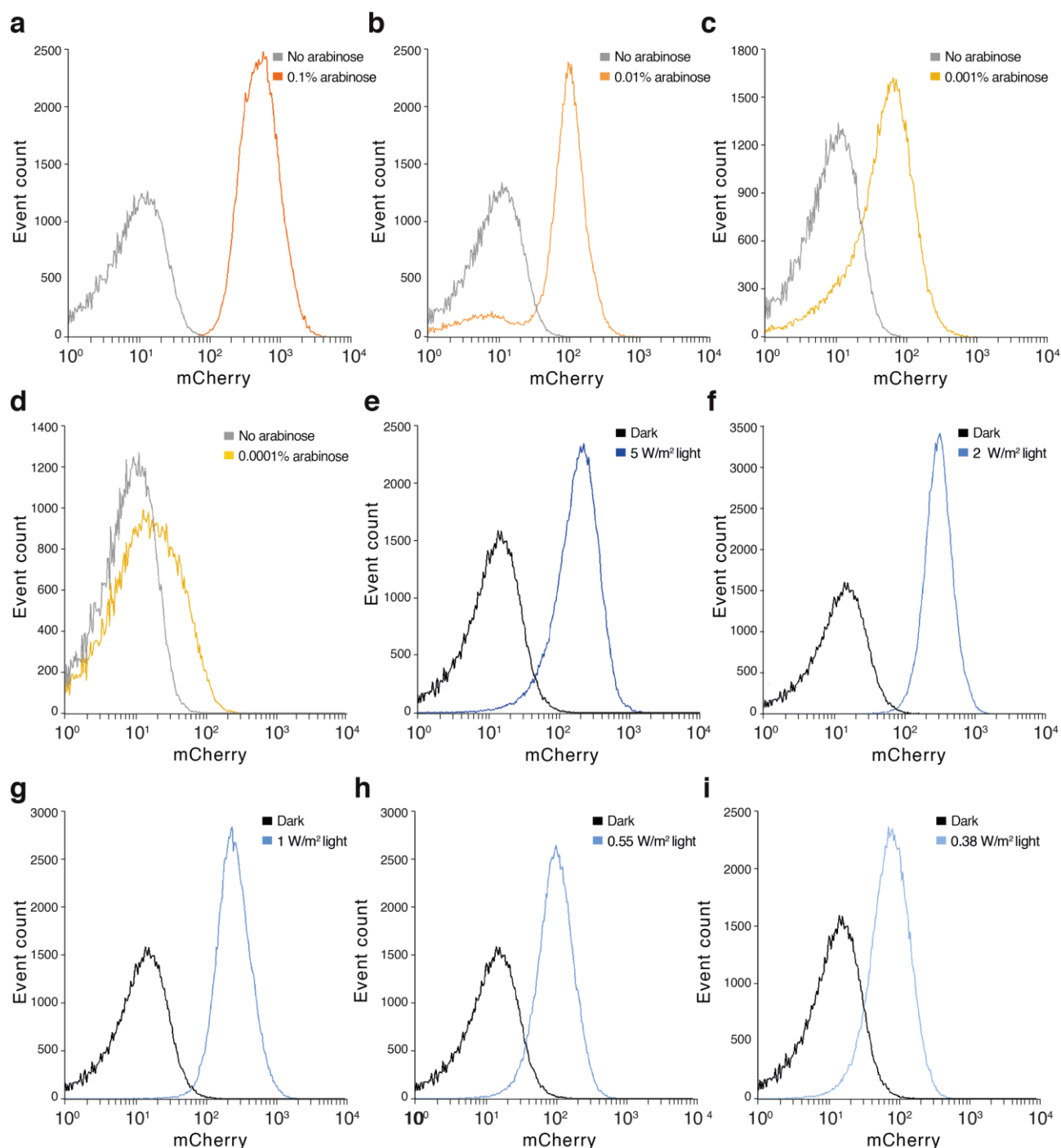
***Figure 28 | Examples of histograms showing the distribution of the mCherry fluorescence within a population of bacterial cells. a-d**, MG1655 cells transformed with pBAD33-mCherry analyzed by flow cytometry after 4 h of induction with the indicated L-arabinose concentration (orange) or in the absence of L-arabinose (grey). **e-i**, MG1655 cells transformed with pBLADE(FP6*)-mCherry analyzed by flow cytometry after 4 h of induction with the indicated 460 nm light intensities (blue) or 4 h in the dark (black).*

The CVs in Figure 27b are measured after 4 hours of induction. However, the induction peak for AraC is reached after 1 or 3 hours of induction, depending on the L-arabinose concentration used (Figure 22a). In order to have a proper characterization of the mCherry expression variability in the population at the time of maximum induction, I measured the CV after 3 hours of induction with 0.1% L-arabinose and after 1 hour when the culture was induced with concentrations ranging between 0.1%

55

and 0.0001% L-arabinose (Figure 29). Despite the different induction times, the coefficient of variation maintained the same growing trend and showed minimal differences from the values shown in Figure 27b.
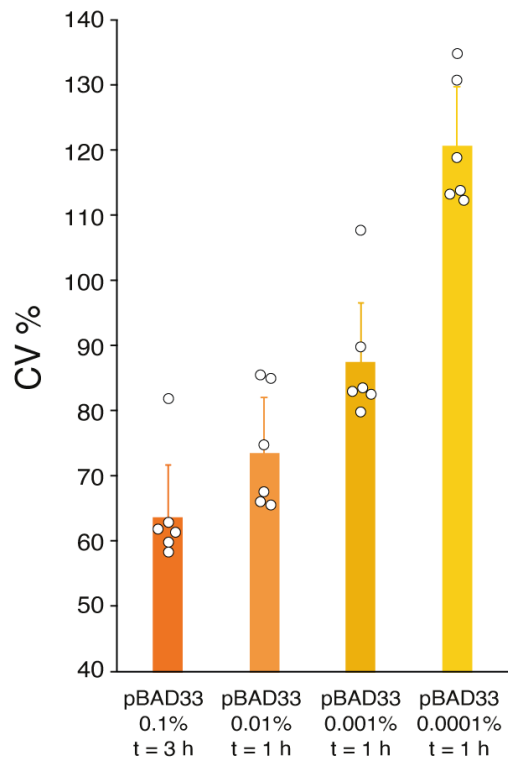
Figure 29| Variability in reporter gene expression in correspondence to the induction peak. Coefficient of variation (CV) of the mCherry fluorescence levels measured by flow cytometry in MG1655 cells transformed with pBAD33 and induced 4 h with the indicated concentrations of L-arabinose at the indicated time, corresponding to the induction peak. Values represent mean ± s.d. of at least n=3 independent experiments.

It is known that, for L-arabinose induction, bistability might occur, with cells within the population being either responding or non-responding , as described in paragraph 1.4.2. Despite the "all-or-none" response to L-arabinose induction being extensively described in the literature[119,130,138,139,141], both for strains deficient of enzymes for L-arabinose catabolism[119,261] and native strains[130,262], I did not often observe this phenomenon. It became particularly visible, though, when I performed the experiment to measure the kinetics of induction using pBAD33-mCherry and 0.0001% L-arabinose added every hour. In this case, the proportion of uninduced and induced subpopulations gradually shifted toward the induced subpopulation over time (Figure 30).

As seen in Figure 28, BLADE-mediated induction always showed a unimodal distribution, as light, despite being homogenously delivered to all cells, does not rely on transporters to be internalized.
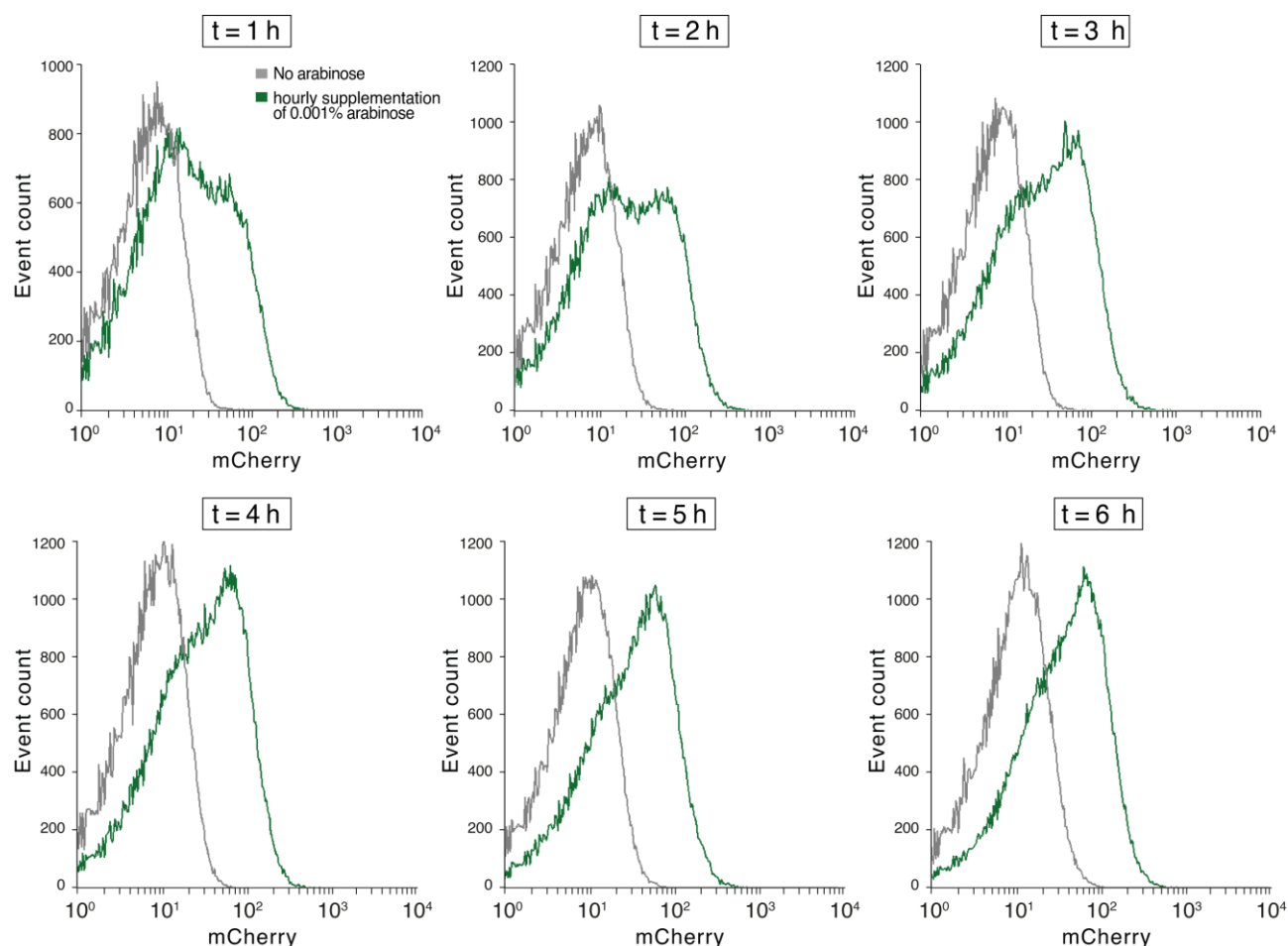
Figure 30 | All-or-none response in a population of cells induced with L-arabinose. MG1655 cells transformed with pBAD33-mCherry analyzed by flow cytometry every hour. The cultures were either hourly supplemented with 0.001% L-arabinose (green) or grew in the absence of L-arabinose (grey).

2.1.9 Investigating the mechanism of BLADE-mediated light-inducible gene expression

Wild-type AraC and BLADE are substantially different in their mode of action: BLADE has a monomeric conformation in the dark and becomes a dimer under blue light illumination, while AraC is always a dimer, binding the I_1 and O_2 half-sites in the absence of L-arabinose, and the I_1 and I_2 half-sites in the presence of it (Figure 4). With the following experiments, I aimed to analyse the binding of BLADE to P_{BAD} in the dark and after blue light illumination.

2.1.9.1 *In vitro* analysis of BLADE binding to I_1I_2

I started by comparing the binding of BLADE to the I_1I_2 half-sites of P_{BAD} performing electrophoretic shift mobility (EMSA). His-tagged BLADE (FP6) has been expressed and purified using the IMAC nickel column by Navaneethan Palanisamy. A 58-nucleotide HEX-labelled DNA fragment encompassing the I_1 and I_2 sites was used as a probe in the EMSA. The detection method was, therefore, fluorescence and not radioactivity. The advantage is that the detection is easy and non-hazardous, the disadvantage is its lower sensitivity. Given the high affinity of AraC$_{DBD}$ for I_1, we expected this binding to happen also with monomeric BLADE in the dark. Upon dimer formation

following blue light illumination, both I_1 and I_2 should be bound, with consequent recruitment of the RNA polymerase in the *in vivo* context. We, therefore, expected to see the formation of different complexes *in vitro* in the dark and lit states (BLADE-I_1 versus BLADE-I_1I_2).

Surprisingly, the EMSA showed a double band in all the samples incubated with BLADE; moreover, by looking at the unbound DNA that runs faster in the gel, especially at 200 and 250 nm of BLADE, it appeared that the protein in the dark state was bound better to the DNA than that exposed to light (Figure 31). Changing the experiment conditions (illumination and incubation time prior to loading, run time etc.) did not change the output, as 6 EMSAs out of 7 showed the same result. The stronger binding of BLADE for the DNA in the dark was unexpected, and this result clearly goes against the *in vivo* data obtained so far.
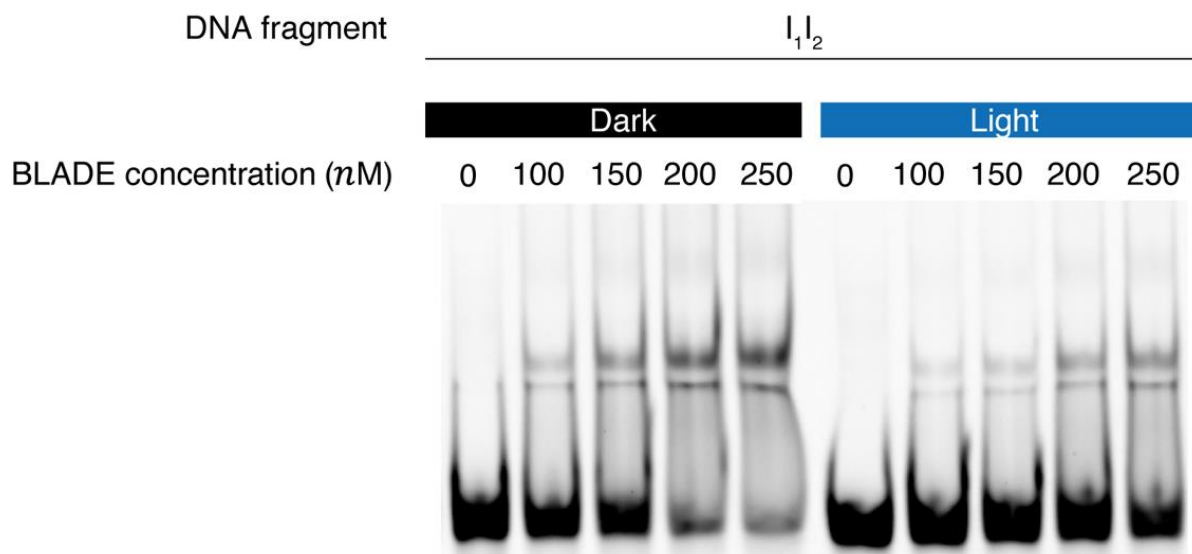
Figure 31 | BLADE binds better to a DNA fragment carrying the I_1 and I_2 sites in the dark. Electrophoretic shift mobility assay of BLADE FP6 incubated with 50 nM I_1I_2 DNA fragment shows a slightly stronger binding in the samples illuminated with blue light. The image was cut in half to have the dark samples on the left.

2.1.9.2 *In vitro* analysis of AraC binding to I_1I_2 and I_1

Given the puzzling results obtained with BLADE, we decided to start anew with a positive control testing the binding of wild-type AraC using the same assay. The his-tagged protein has been expressed and purified by Navaneethan Palanisamy. I would like to note that the scientific literature only reports EMSAs of AraC performed with synthetic I_1I_1 half-sites[247,263,264], but not with I_1I_2.
As arabinose-free AraC binds to the I_1 half-site (and to O_2 if present), while arabinose-bound AraC binds I_1 and I_2, we expected to see these different complexes as separate bands in the EMSA. Indeed, there were different bands in the gel depending on whether L-arabinose was added or not to the samples (Figure 32). Interestingly, while in the absence of L-arabinose the DNA appeared stuck in the well for high protein concentrations, a clearly defined band running into the gel was visible for the same protein concentrations provided L-arabinose was present. These results could be interpreted as the result of the aggregation of apo-AraC, which would retain the DNA in the well. At 250 nM AraC, two bands are clearly visible in the middle of the gel. We interpreted the upper band as the

AraC-I_1I_2 complex, since this band became the only visible for even higher AraC concentrations only in the presence of L-arabinose (Figure 32).

To confirm this hypothesis, we incubated AraC with L-arabinose and a DNA probe including only the I_1 half-site. Unexpectedly, the two bands were visible also in this case (Figure 32). More replicates are needed to confirm this data, however, at this stage, we concluded that our protocol did not allow resolving AraC-I_1 and AraC-I_1I_2 complexes. It is not surprising, therefore, that we were not able to resolve such complexes when using BLADE instead of AraC.
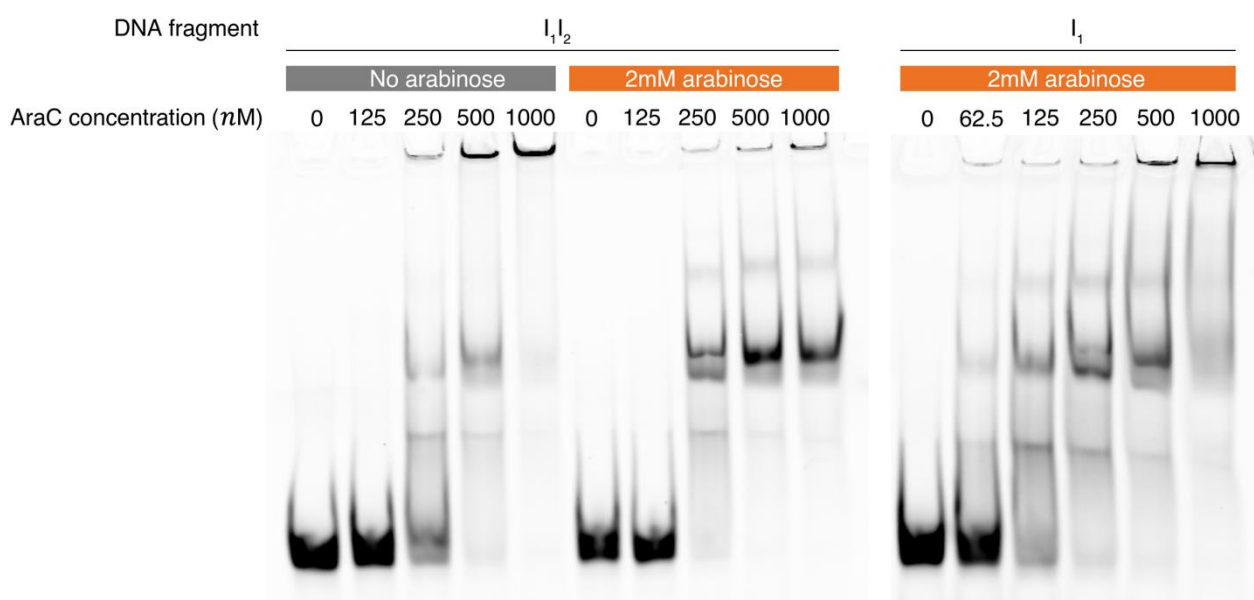
Figure 32 | The presence of L-arabinose changes the DNA binding behaviour of AraC. Left, Electrophoretic shift mobility of AraC incubated with 50 nM DNA probe encompassing the I_1I_2 half-sites in presence or absence of 2 mM L-arabinose. Right, Same assay but using 50 nM DNA probe bearing only the I_1 half-site.

2.1.9.3 *In vitro* analysis of BLADE binding to I_1 and I_2

We nonetheless decided to analyse the binding of BLADE to DNA fragments containing only the I_1 or the I_2 half-sites to gather some further insight.

As predicted, the presence of I_1 alone was sufficient for the binding to occur, both in the dark and lit states (Figure 33). The band patterns were identical to those obtained with BLADE incubated with a DNA probe containing the I_1I_2 half-sites (Figure 31). What seemed to be an inhibition of the binding when the protein was illuminated was also observed again. What caused the appearance of a third band in the samples at the highest concentration of BLADE is unknown, but this was not present in a replicate of the experiment. When we incubated BLADE with the DNA fragment containing only I_2, we saw no binding (Figure 33).
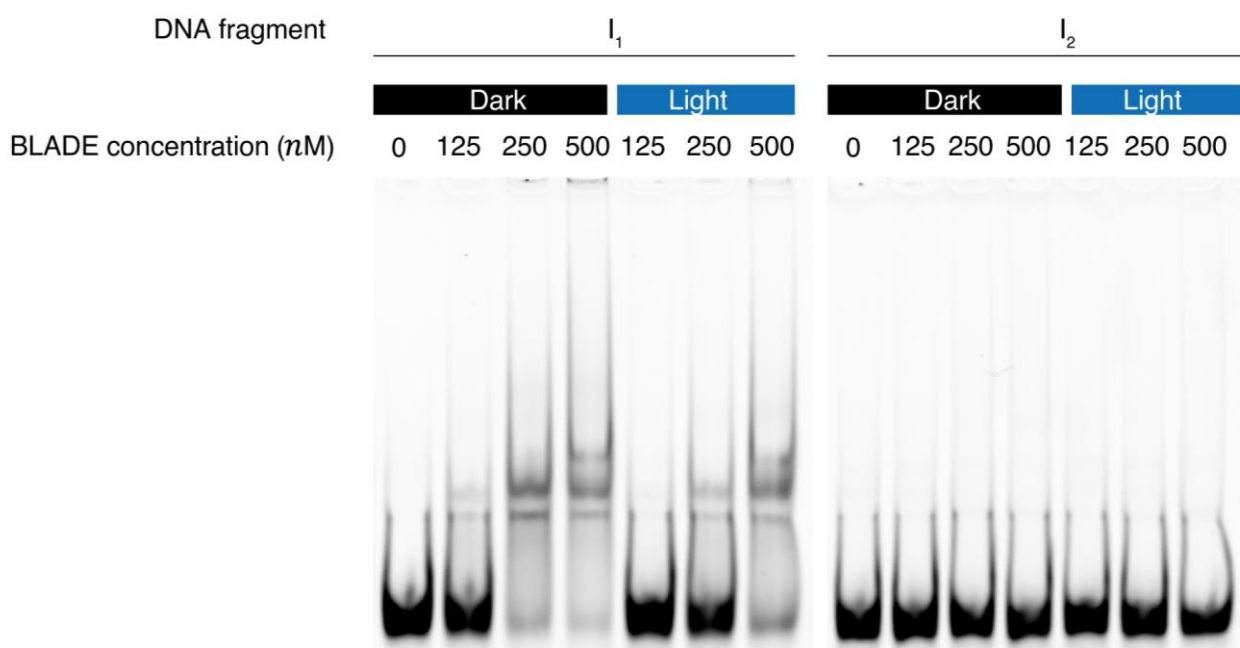
Figure 33 | BLADE binds to a DNA fragment containing only I₁ but does not bind if only I₂ is present. *Electrophoretic shift mobility of BLADE FP6 at the indicated concentrations incubated with the indicated DNA fragments at 50 nM.*

Taken together, the EMSA results were puzzling because of the better DNA binding observed with BLADE in the dark state. Since I know that, *in vivo*, BLADE is able to activate transcription only under blue light, I decided to use a different technique to observe the changes made to the protein by light. The EMSA is notoriously characterized by high variability and is influenced by many factors. I thought of using size exclusion chromatography (SEC) to detect the light-induced dimerization of BLADE.

2.1.9.4 *In vitro* analysis of BLADE dimerization

Before proceeding with size exclusion chromatography (SEC), we decided to verify that the purified protein could respond to light. With the technical help of Maximilian Hörner, I performed UV-vis spectroscopy with the BLADE FP6 protein incubated in the dark. Subsequently, the same sample was illuminated with blue light and analysed again to visualize the transition of the flavin adenine dinucleotide (FAD) chromophore from the blue light-absorbing state to the bleached one, indicative of the formation of the adduct. We had to maintain the protein at 4 °C because it was prone to precipitation when maintained at room temperature for more than 2 hours.

When BLADE was incubated in the dark for 4 days we could observe, as reported in the literature, a major peak of absorbance at 450 nm and two minor peaks at 428 and 478 nm[90] (Figure 34a, black curve). Upon 50 W/m² light incubation for 5 minutes at room temperature, an almost total loss of absorption in the blue light range was observed (Figure 34a, blue curve). We also noticed that, at 4 °C, the change of redox state from reduced (lit state) to oxidized (dark state) took several days: the same aliquot of sample induced previously, after 1 day of incubation in the dark started to show the two absorption peaks at 428 and 478 nm, even though not very pronounced (Figure 34b, black curve). Only after 4 days of dark incubation, the absorbance at those wavelengths was allowed to increase significantly (Figure 34a, black curve).
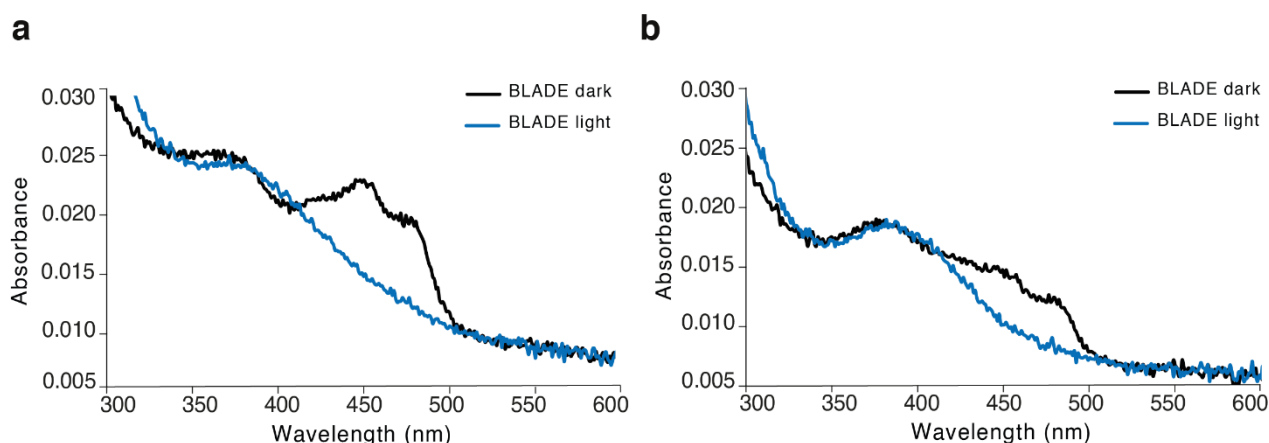
Figure 34 | Absorption spectra of the FAD cofactor within BLADE FP6. a, *Absorption spectra measured with the protein incubated 4 days in the dark at 4 °C and illuminated with blue light (455 nm; 50 W/m²) for 5 min at room temperature. **b**, Absorption spectra measured with the protein incubated 1 day in the dark at 4 °C and illuminated with blue light (455 nm; 50 W/m²) for 5 min at room temperature. **a,b**, The absorption spectrum of the blank (only medium) was subtracted from the dark and lit state spectra.*

After the confirmation of the light responsiveness of the purified protein, Navaneethan Palanisamy performed SEC at 4 °C with BLADE stored in the dark for 6 days. Then, he repeated the assay with the same sample illuminated with 50 W/m² blue light for 30 minutes, to resolve the monomeric and dimeric forms. The SEC nicely showed a shift between the curves of the dark and lit states (Figure 35). The control samples of known molecular weight, loaded on the same column, provided references of molecular weights that consented us to conclude that the two peaks approximately matched the sizes of a BLADE monomer and dimer, respectively (31.3 and 62.6 kDa).
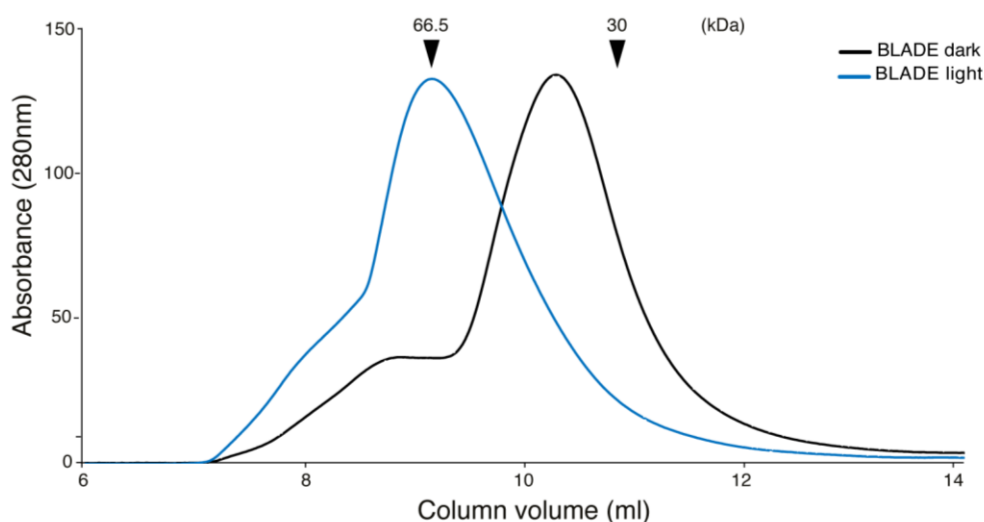
Figure 35 | *BLADE dimerizes under blue light illumination. SEC performed with purified BLADE FP6 kept in the dark or illuminated with 460 nm light (50 W/m²) for 30 minutes at 4 °C. BLADE monomer weights 31.3 kDa, and the dimer 62.5 kDa. The arrows at 66.5 and 30 kDa indicate the peaks of the molecular weight markers, corresponding to bovine serum albumin (BSA) and carbonic anhydrase (CA), respectively.*

2.1.9.5 The roles of I$_1$ and I$_2$ in the binding of BLADE to P$_{BAD}$

It is difficult to predict the 3D structure of the light-induced BLADE dimer, and it would be indeed very interesting to solve it. Theoretically, dimeric BLADE could assume a conformation similar to

61

either that of the apo- or holo-AraC or even another one, that would still allow for the interaction with the I_1 and I_2 half-sites. All the data I obtained point toward a conformation that leads to dimeric BLADE interacting with the I_1 and I_2 half-sites. In particular, the increase in reporter gene expression following illumination can only be explained with BLADE contacting the I_2 half-site and recruiting the RNA polymerase at the P_{BAD} promoter. The I_1 half-site is likely contacted also by monomeric BLADE, as seen in our *in vitro* assays and as observed in previous reports for a single DBD of AraC[247] (Figure 33). However, this is not sufficient for recruiting the RNA polymerase to the P_{BAD} promoter[265].

The data shown in Figure 15 indicate that full-length AraC is more efficient than BLADE in inducing reporter expression from P_{BAD}. One hypothesis to explain this behaviour is the higher binding affinity of the arabinose-bound AraC dimer compared to the light-induced BLADE dimer to I_1I_2. Despite having the same DBD and binding to the same sequence, the structure of dimeric BLADE might not allow for the correct positioning of the DBD on the I_1I_2 half-sites, resulting in a less stable binding to the P_{BAD} promoter and lower activation of reporter gene expression.

I already described the process to select the best linker in paragraph 2.1.1 and 2.1.3, but so far I did not try to optimize the DNA-binding domain of BLADE to better contact the I_1I_2 half-sites of P_{BAD}. With the help of Mehmet Öztürk, we attempted to find a DNA sequence that would let the DBD of BLADE to dimerize in back-to-back fashion: this could, in principle, increase the probability of dimerization efficiency compared to sequential binding, as BLADE is constructed with VVD fused to one end of the DBD structure (Figure 36a). He computationally analyzed alternative BLADE dimerization mechanisms by changing the order and direction of the I_1I_2 half-sites. For this purpose, he applied Brownian dynamics (BD)-based rigid-body docking simulations and generated alternative I_1I_2 structures (Figure 36a).

As the -35 region of the P_{BAD} promoter partially overlaps with I_2, we kept that sequence untouched when the half-site was reverted. We selected three combinations of DNA binding sites from the simulation output, with the hope that these would allow two VVD domains to come closer and thus increase BLADE dimerization efficiency (Figure 36a). We also included the combination I_1I_{2rev} as a negative control: we predicted that binding to that sequence could not happen when BLADE was in the dimeric state because of the too short interdomain linker.

Surprisingly, the mCherry levels between the dark and light samples demonstrated complete non-light responsiveness of BLADE with the new DNA binding sites, except for a modest fold-change when the I_1 binding site was reversed (Figure 36b). As previously mentioned, the I_1 site has a higher affinity for AraC$_{DBD}$ than I_2; when I_1 was replaced with I_2, the fluorescence levels remained below the basal level of the control plasmid without BLADE, pReporter_only, with wild-type I_1I_2 (dashed line). Having observed the lack of induction for the negative control sequence I_1I_{2rev}, we can also conclude that the I_2 half-site is contacted *in vivo* by BLADE in the presence of blue light and its presence (and right orientation) is fundamental for the recruitment of the RNA polymerase.

Further literature analysis showed that proper DNA bending[263], which is crucial for AraC activation, was not included in our rigid body BD docking and thus our analysis could not provide the full picture of the AraC DNA binding. Further computationally expensive full atom molecular dynamics simulations would be needed for a full understanding of the AraC-DNA interaction details.

a
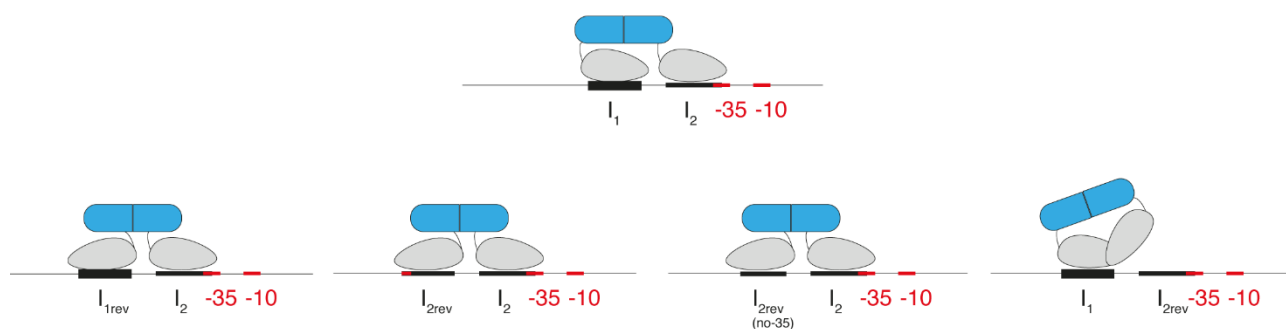
b
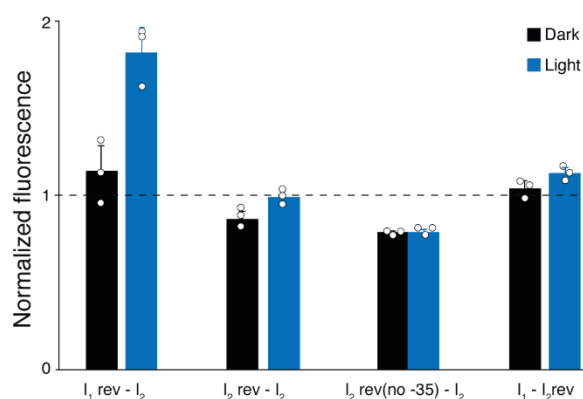
Figure 36 | Predictions and experimental validation of different arrangements of the I_1 and I_2 half-sites. a, Top, schematic representation of BLADE bound to wild-type I_1I_2 half-sites. Bottom, schematic representation of the 4 different predicted BLADE dimeric conformations binding to the rearranged half-sites. b, mCherry fluorescence intensity in E. coli MG1655 cells transformed with pBLADE(FP6*)-mCherry with the indicated DNA binding sites grown for 4 h either in the dark or under 460 nm light (5 W/m^2) illumination. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at least n=3 independent experiments. BLADE variant: FP6 driven by the J23101* promoter.

Following another strategy, in an attempt to increase the binding affinity of BLADE to P$_{BAD}$, I introduced a point mutation (A –> G) in the I_2 sequence, as reported in a paper from Reeder and Schlief[248]. The authors reported a 40% increase in the reporter level after induction with L-arabinose, compared with the induction with wild-type I_1I_2, while the uninduced levels remained unchanged. When I introduced the mutation in the P$_{BAD}$ promoter in pBLADE (yielding pBLADE_I_2*), the fluorescence levels increased by 220% after light induction compared to those obtained with pBLADE; however, there was also 85% higher fluorescence for the sample maintained in the dark (Figure 37). It is interesting to note that this point mutation, which presumably increases the affinity AraC$_{DBD}$ for I_2, allowed to reach a higher induced level than that obtained by pBAD33 induced with L-arabinose, albeit at the price of higher leakiness.
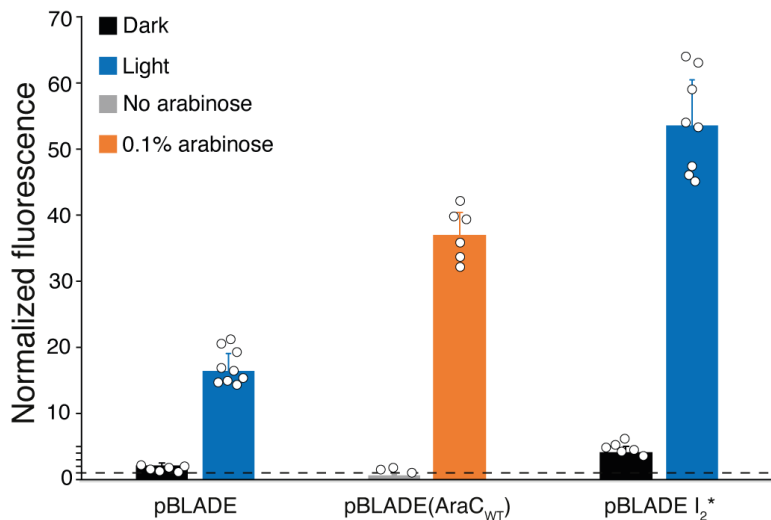
Figure 37 | A point mutation in the I₂ half-site dramatically increases the reporter expression level obtained with BLADE. *mCherry fluorescence intensity in E. coli MG1655 cells transformed with pBLADE(FP6*)-mCherry and pBLADE I₂*(FP6*)-mCherry grown for 4 h in the dark or light state (460 nm light at 5 W/m²) and pBLADE(AraC$_{WT}$*)-mCherry grown for 4 h without L-arabinose or with 0.1% L-arabinose. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at least n=3 independent experiments.*

2.1.9.6 Lowering the leakiness with additional copies of I₁I₂ on the plasmid

BLADE driven by the J23101* promoter minimally activates the reporter expression in the dark, in a way comparable to wt AraC. However, I wondered if it was possible to find a way to further reduce it. After discussions with our collaborator Mustafa Khammash, we reasoned that a solution could be to sequester BLADE away from P$_{BAD}$ by adding multiple copies of I₁I₂ in the pBLADE backbone in a separate location away from P$_{BAD}$. I, therefore, cloned up to 4 I₁I₂ DNA sequences into pBLADE-mCherry, and then tested these new plasmids in an induction experiment.

Despite the additional binding sequences, the leakiness did not change significantly from the original plasmid without extra I₁I₂ sites (Figure 38). To statistically assess whether there is a significant difference in the fluorescence levels in the sample with no extra copies and that with 4, the number of measurements should be increased. Nonetheless, even if the effect were significant, it would still be minor, thus I decided not to further pursue this method.
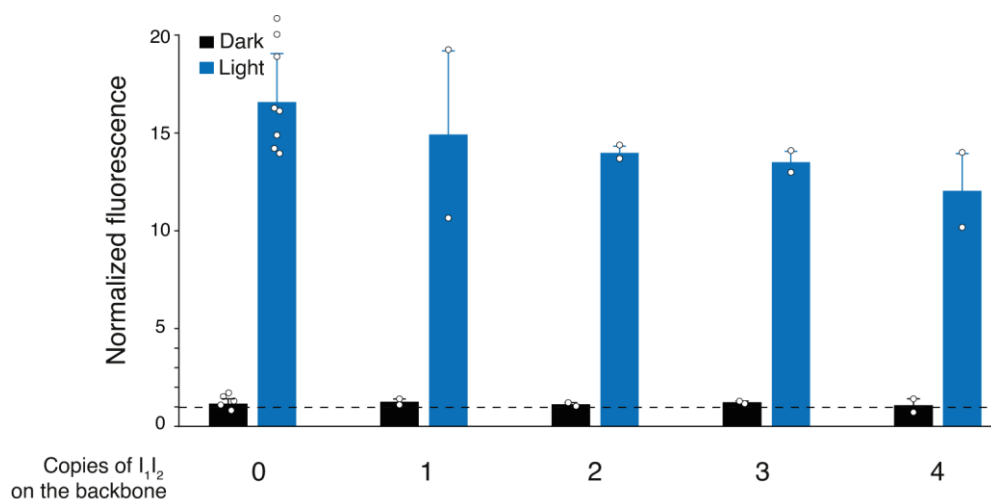
Figure 38 | The addition of further I_1I_2 binding sites decoupled from P_{BAD} into pBLADE does not improve the fold-change of induction. mCherry fluorescence intensity in E. coli MG1655 cells transformed with pBLADE(FP6*)-mCherry with additional I_1I_2 half-sites on the backbone grown for 4 h in the dark or under blue light (460 nm light at 5 W/m^2). All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at n = 2 independent experiments.

I imagined that the leakiness in the dark was already too low to be significantly reduced by additional copies of I_1I_2. That is why I thought to use pBLADE_I_2*-mCherry, which I showed to lead to a much higher level of reporter expression in the dark (Figure 37). I inserted 7 repetitions of I_1I_2 into pBLADE_I_2*-mCherry and measured the mCherry fluorescence levels for dark and light conditions. However, even with this leakier construct, I did not observe a reduction of the dark state fluorescence: both the fluorescence levels measured after 4 hours in the dark or after blue light illumination were identical to those obtained with the plasmid without extra copies of I_1I_2 (Figure 39).
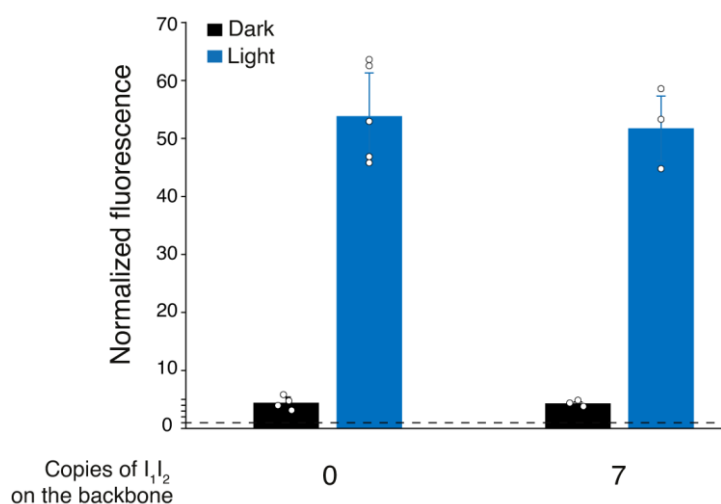
Figure 39 | The addition of further I_1I_2 binding sites decoupled from P_{BAD} into pBLADE_I_2* does not improve the fold-change of induction. mCherry fluorescence intensity in E. coli MG1655 cells transformed with pBLADE_I_2*(FP6*)-mCherry with 0 or 7 copies of the I_1I_2 half-sites grown for 4 h in the dark or light state (460 nm light at 5 W/m^2). All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with the plasmid deprived of the transcription factor (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at least n = 3 independent experiments.

2.1.9.7 The roles of O_1 and O_2 in the activation of P_{BAD} by BLADE

Located 210 base pairs from P_{BAD}, the O_2 half-site promotes looping formation with AraC in absence of L-arabinose (paragraph 1.4.2). I hypothesized that, when present, the genomic AraC would create the DNA loop in the absence of L-arabinose, shielding the I_1I_2 sites from BLADE dimers that might be formed in the dark, helping to keep the dark state induction low. Under blue light, the higher amount of BLADE dimers would displace the low number of AraC molecules from the I_1 site and manage to bind to P_{BAD}, recruiting the RNA polymerase.

To test this hypothesis, I removed both the O_1 and O_2 half-sites from pBLADE, and I induced the culture transformed with the resulting pBLADEΔO_1O_2-mCherry for 4 hours with blue light. As a control, I left another culture transformed with the same plasmid in the dark for the same time. Without O_1 and O_2, I observed an increase of leakiness in the dark state, which reached a similar level as that obtained from an empty pBLADE deprived of the O_1O_2 half-sites (pReporter_onlyΔO_1O_2,

dash-dot line; Figure 40). The data also showed that the loop unwinding was not relevant to increase gene expression after blue light illumination, as the induction levels after blue light illumination were comparable in the presence and absence of O_1O_2 (Figure 40). Despite the absence of the loop, I hypothesized that a certain degree of competition with AraC for I_1I_2 was maintained nonetheless, thanks to the binding of apo-AraC to the I_1 half-site.
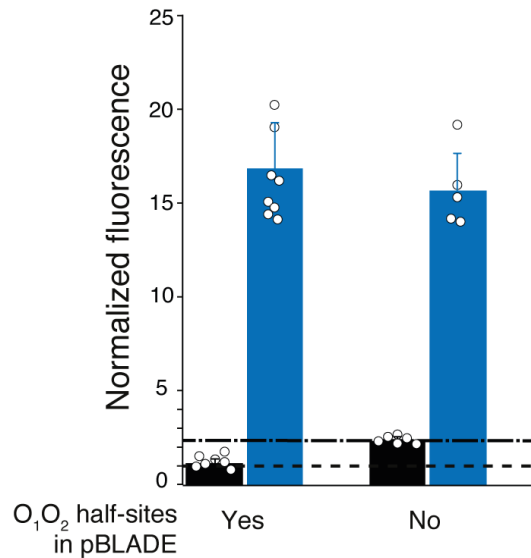
Figure 40 | The absence of O_1O_2 sites increases the leakiness of BLADE. *mCherry fluorescence intensity in E. coli MG1655 cells transformed with pBLADEΔO$_1$O$_2$(FP6*)-mCherry with or without the O_1O_2 half-sites grown for 4 h either in the dark or under 460 nm light (5 W/m^2) illumination. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with the plasmid deprived of the transcription factor (dashed line see Supplementary Figure 2). Dash-dot line indicates the fluorescence level of pReporter_onlyΔO$_1$O$_2$. Values represent mean ± s.d. of at least n = 3 independent experiments.*

2.1.9.8 AraC and BLADE compete for binding to P_{BAD}

So far, I have performed all the experiments with the wild-type strain MG1655 that codes for *araC* in the chromosome. To see what happens when endogenous *araC* is not present, I transformed pBLADE into SKA703, the strain lacking chromosomal *araC* unable to metabolize L-arabinose. I observed an even higher level of leakiness than that obtained with pBLADEΔO$_1$O$_2$ (Figure 41). With this strain, not only the loop was not formed, but also AraC absence left the I_1 site free to be bound from the few dimeric BLADE proteins that formed in the dark, increasing the chances of unwanted gene transcription. The absence of AraC seemed not to play a role in the level of induction under blue light.
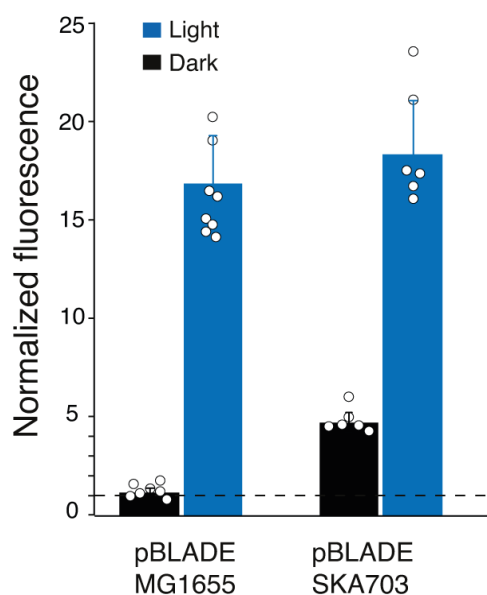
Figure 41 | The absence of genomic AraC increases the leakiness of BLADE. *mCherry fluorescence intensity in E. coli MG1655 or SKA703 cells transformed with pBLADE(FP6*)-mCherry with or without the O_1O_2 half-sites grown for 4 h either in the dark or under 460 nm light (5 W/m^2) illumination. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with the plasmid deprived of the transcription factor (dashed line, see Supplementary Figure 2). Values represent mean ± s.d. of at least n = 3 independent experiments.*

2.1.9.9 BLADE is found in intracellular aggregates in the dark

In *Neurospora crassa*, the organism in which it is naturally expressed, VVD is degraded in the dark[87,266,267]. I wondered whether VVD might trigger the degradation of BLADE in *E. coli* cells in the dark. This could contribute to regulate the system and maintain a high degree of tightness. To investigate this issue, I fused sfGFP to the C-terminus of BLADE in pBLADE(FP6)*-mCherry. I chose this terminus to avoid any interference with dimerization. I measured sfGFP levels with flow cytometry to compare BLADE levels in cells kept in the dark and exposed to blue light for 4 hours. The sfGFP levels were comparable in both conditions (Supplementary Figure 5a), therefore I discarded the hypothesis of VVD degradation.

I decided to take advantage of the construct to assess the localization of BLADE in living cells. Surprisingly, fluorescence microscopy revealed the presence of bright fluorescent foci in half of the cells kept in the dark (Figure 42a and Supplementary Figure 5b), while less than 20% of the illuminated cells showed foci (Figure 42a and Supplementary Figure 5b). To prove that are related to the light response of VVD, I mutated the adduct-forming cysteine to alanine (VVD^{C108A}) within BLADE. I expected the mutant to show aggregates also under blue light illumination since VVD^{C108A} is not responsive to light[90]. Indeed, 40% of the cells presented foci both in the dark and when illuminated with blue light for 4 hours (Figure 42b and Supplementary Figure 5b).
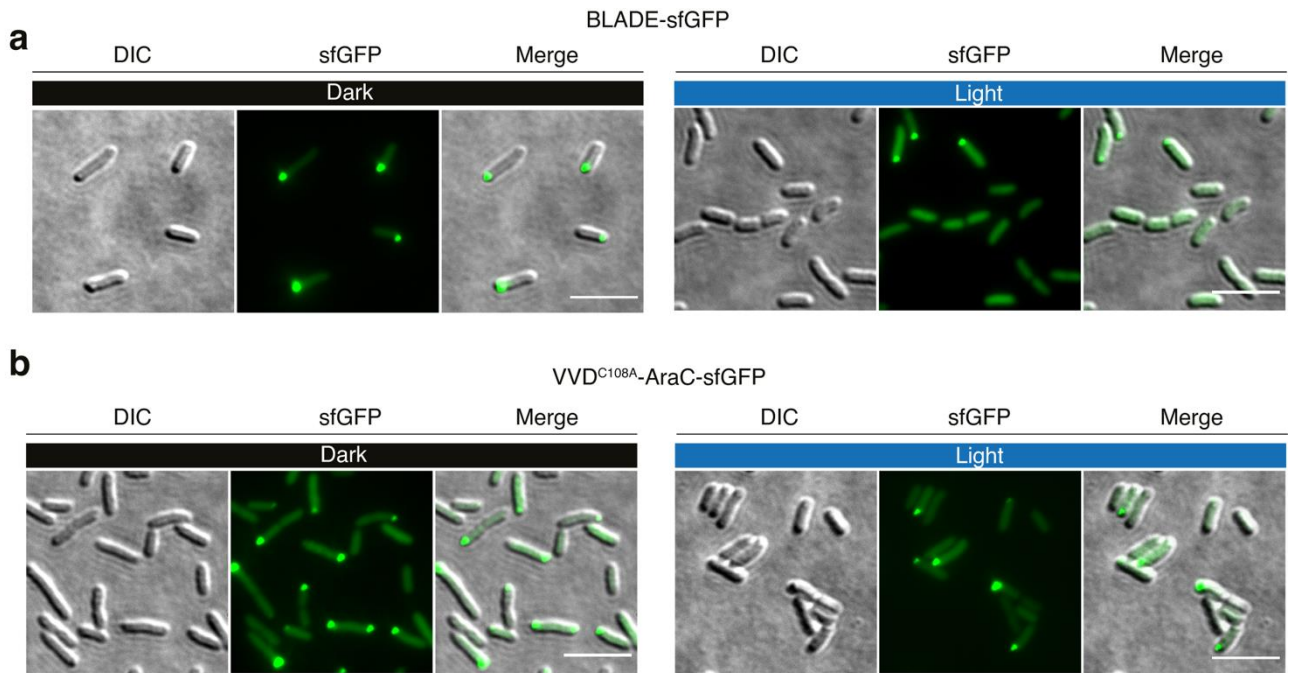
***Figure 42 | BLADE-mediated light-induced gene expression involves the formation of aggregates in the dark. a**, Representative microscopy images of E. coli MG1655 cells expressing BLADE C-terminally fused to sfGFP grown for 4 h in the dark or under 460 nm light (5 W/m²) light. **b**, Representative microscopy images of E. coli MG1655 cells expressing BLADE-sfGFP harbouring the C106A mutation in the VVD domain grown for 4 h in the dark or under 460 nm light (5 W/m²) light. BLADE variant: FP6 C-terminally fused to GFP driven by the J23101* promoter. Scale bar, 5 μm.*

To verify that the formation of these foci is an intrinsic property of VVD and not sfGFP, I fused sfGFP to the C-terminus of wt AraC and cloned the fusion into pBLADE, yielding pBLADE(AraC_WT), as well as pBAD33, yielding pBAD33(AraC_WT). I then imaged the cells at the microscope after 4 hours of induction with no or with 0.1% L-arabinose. As expected, for both constructs the fluorescence remained diffused in the cytoplasm, with no difference between the uninduced and induced samples (Figure 42c).

a

b
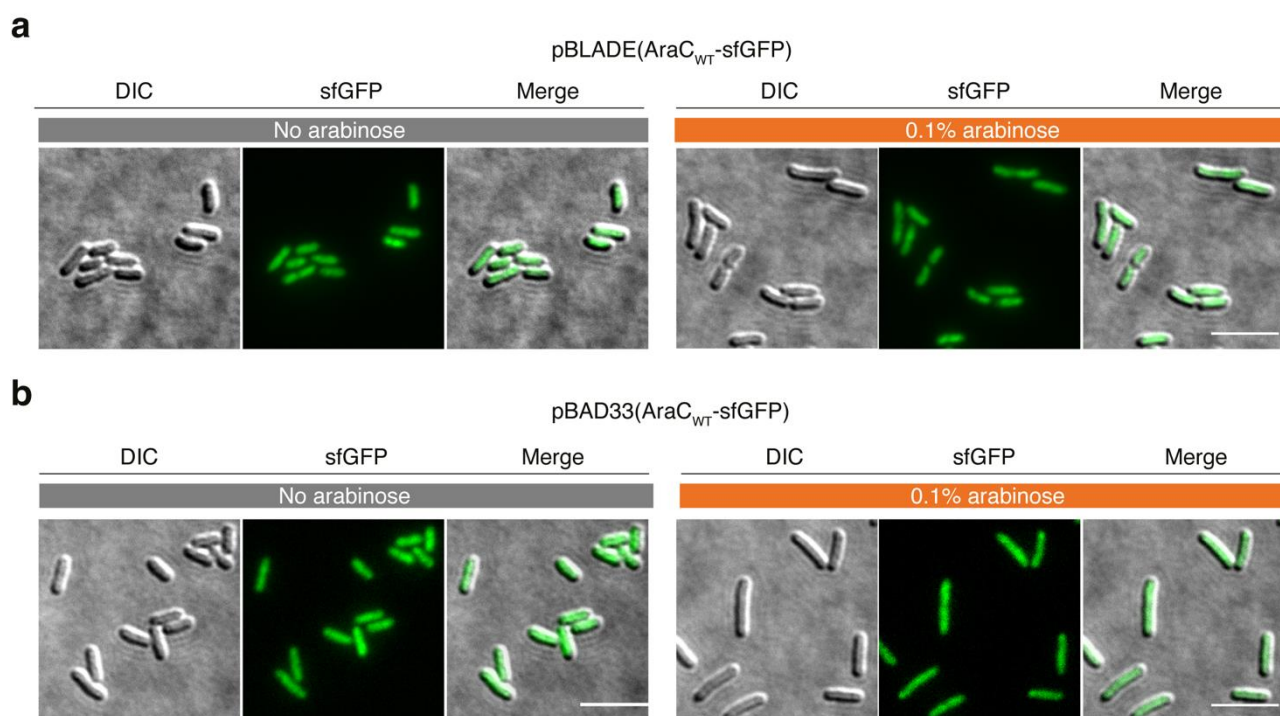
*Figure 43 | AraC-sfGFP is cytoplasmic. Representative microscopy images of E. coli MG1655 cells expressing AraC C-terminally fused to sfGFP grown for 4 h without L-arabinose or with 0.1% L-arabinose. **a**, pBLADE backbone, AraC-sfGFP is driven by the J23101* promoter. **b**, pBAD33 backbone, AraC-sfGFP is driven by the P_C promoter. Scale bar, 5 µm.*

As an attempt to investigate the nature of the foci, I performed fluorescence recovery after photobleaching (FRAP) experiments. Since there was no recovery after 15 minutes, I classified them as aggregates and not as functional liquid droplets (Figure 44).
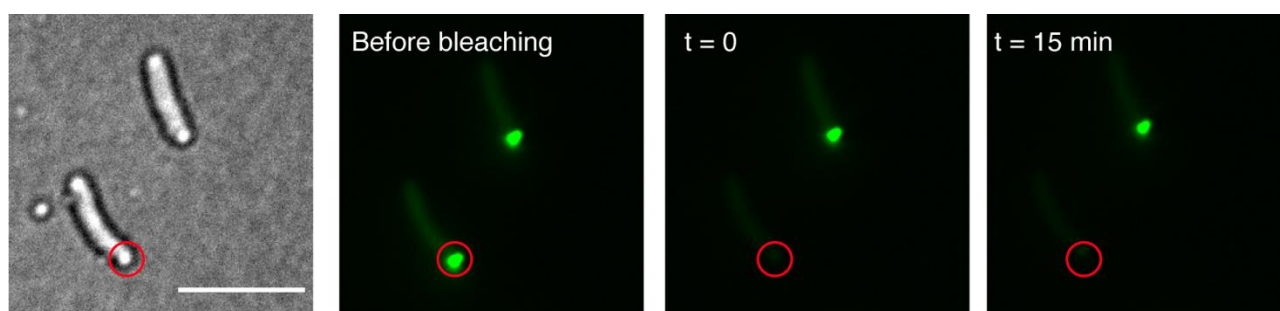
Figure 44 | The foci are protein aggregates. Representative images of fluorescence recovery after photobleaching (FRAP) experiment. Fluorescence does not recover within the time of the experiment, suggesting that the foci are aggregates rather than liquid droplets. BLADE variant: FP6 C-terminally fused to GFP driven by the J23101 promoter. Scale bar, 5 µm.*

It was shown in the literature that VVD can transition between locally unfolded and folded states, and that light shifts the half-life of the transition from about 5 minutes to 6 hours[268]. It was also suggested that simultaneous unfolding of several structural elements of VVD could lead to aggregation in the dark[268]. The aggregates I observed in *E. coli* could, therefore, be due to the VVD moiety in BLADE.

Finally, to investigate whether light could lead to the dispersion of previously formed foci, Emir Akmeriç performed time-lapse fluorescence microscopy to follow individual foci over time in illuminated cells. This experiment was important to follow the fate of the aggregates following light illumination: they could either be dispersed and disappear or remain intact but form less frequently in newborn cells. We found the second scenario to be true, as the aggregates did not disperse, but were instead asymmetrically segregated during cell division (Figure 45). Newborn cells contain either no foci or foci much smaller than those found in cells kept in the dark (Figure 45).
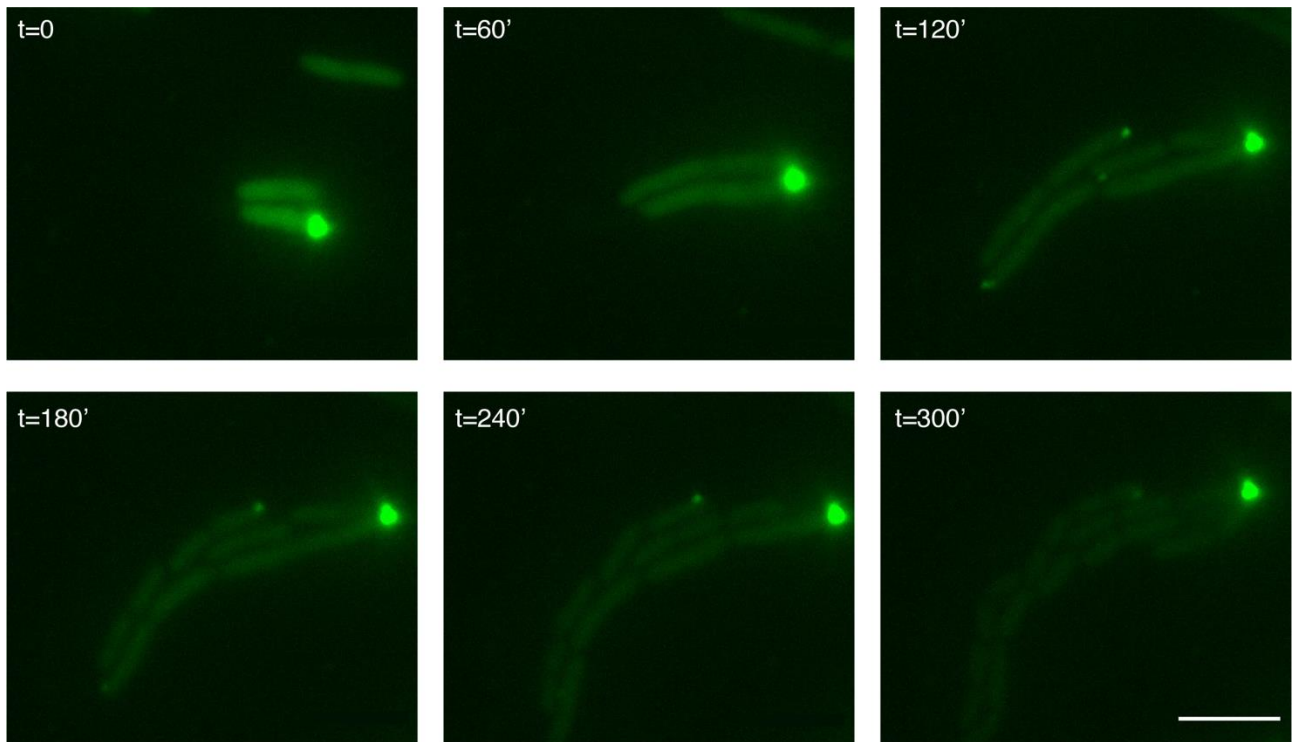
Figure 45 | BLADE-sfGFP foci do not disperse under blue light illumination and are asymmetrically segregated at cell division. Six frames were extracted at the indicated time points from a 7-hour time-lapse. BLADE variant: FP6 C-terminally fused to GFP driven by the J23101 promoter. Scale bar, 5 µm.*

2.1.10 BLADE is compatible with pre-existing L-arabinose-responsive plasmids and strains

As BLADE and AraC bind to the same promoter P_{BAD}, I intended to check the ability of BLADE to outcompete AraC for the activation of the P_{BAD} promoter in the absence of L-arabinose, a situation in which AraC is known to create the DNA loop that prevents the recruitment of the RNA polymerase (Figure 4). This would allow for the usage of previously constructed plasmids and strains with a simple (co)-transformation step to bring BLADE into the cells (Figure 46), providing the advantage to switch the induction method from L-arabinose to light, without changing anything else in the previous setup except the addition of a plasmid.
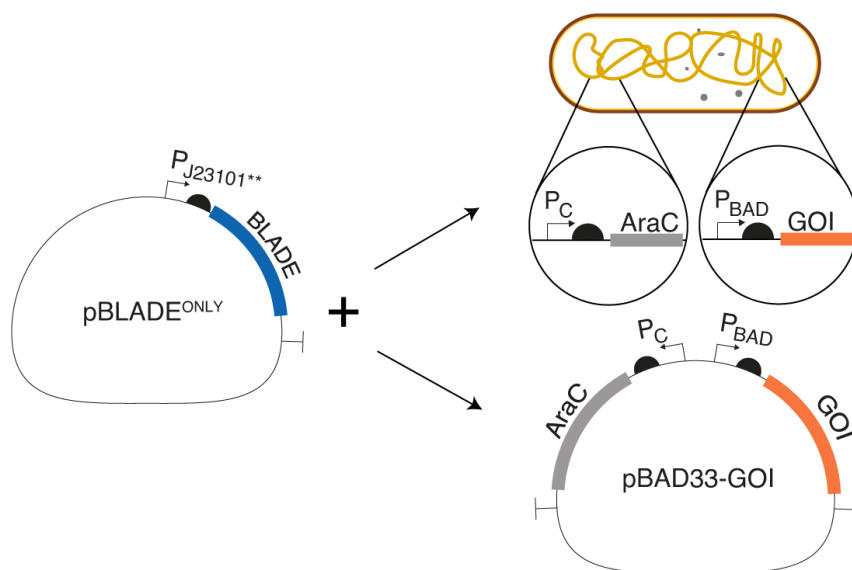
Figure 46 | Expression of BLADE in pre-existing L-arabinose-responsive plasmids and strains can make them light-sensitive. Schematic representation of the way in which BLADE can be used to control with light the expression of a gene of interest previously put under L-arabinose control, either genomically or on a plasmid.

2.1.10.1   BLADE induces the expression of mCherry located on the unmodified pBAD33

To test if a pre-existing pBAD33 plasmid carrying the gene of interest can become light-sensitive, I first cloned BLADE driven by the constitutive J23101** promoter into the pTRC99a plasmid, removing the IPTG-inducible pTrc promoter, yielding pBLADE$^{ONLY\_A}$ (Supplementary Figure 2). I then co-transformed it with pBAD33-mCherry into MG1655 cells. The mCherry reporter was activated with blue light, even though the induction levels were lower compared with the cells transformed only with pBLADE (Figure 46 and 47, Supplementary Table 3). The lower fluorescence could be explained by a competition between BLADE and AraC transcribed by pBAD33, that reduced the likelihood of the BLADE dimer to bind I_1I_2 half-sites. However, I also imagined that the use of two antibiotics, needed to maintain the aforementioned plasmids in the culture, could be responsible for a fitness decrease, and therefore a sub-optimal induction level. Having to counteract antibiotics can cause metabolic burden to the bacteria, leading to decreased fitness compared to a situation without antibiotics[269], even though it has been observed that the cells can rearrange their metabolic network to compensate for the fitness cost[270].

Since in our laboratory a new method has been recently developed to maintain in *E. coli* two plasmids with a single antibiotic[271], I decided to test if this system would lead to higher expression levels of my gene of interest, mCherry in this case. The method consists in two plasmids: pSiMPl_C, having a pTrc99a backbone that codifies for the C-terminal part of the antibiotic resistance gene (kanamycin in this case), and pSiMPl_N, having a pBAD33 backbone, for the N-terminal part. The two genes are fused to a split intein fragment: when a cell is transformed with both plasmids, the split inteins allow the protein to be reconstituted and become fully functional.

I cloned BLADE in pSiMPl_C and mCherry in pSiMPl_N, obtaining pSiMPl_C-BLADE and pSiMPl_N-mCherry, respectively (Supplementary Figure 2), and I induced the co-transformed culture for 4 hours with blue light or I maintained them in the dark. The obtained induction values

were similar to those obtained with the cells co-transformed with pBLADE$^{ONLY\_A}$ and pBAD33-mCherry, which were treated with two antibiotics instead of one (Figure 47, Supplementary Table 3). I also added a sample with cells cotransformed with pBLADE-mCherry and an empty pTrc99a, to analyse the induction levels of pBLADE when the cells are cultured in a medium with two antibiotics (Figure 47, Supplementary Table 3), obtaining similar induction levels to cells transformed only with pBLADE.

This comparison brought me to the conclusion that there is no additional burden associated with the presence of two antibiotics in the medium. The competition between BLADE and apo-AraC was then the only factor that reduced the fluorescence levels when BLADE was transformed with pBAD33, both of the samples in the dark and illuminated with blue light (Supplementary Table 3).
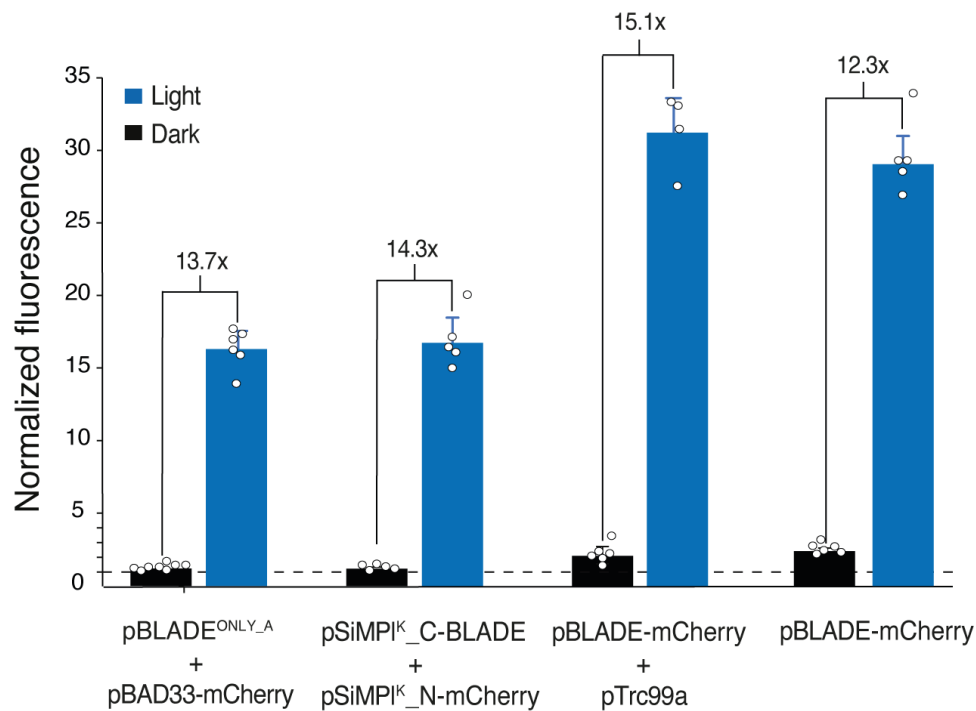
Figure 47 | BLADE is compatible with pre-existing plasmid. *mCherry fluorescence intensity in E. coli MG1655 cells co-transformed with the indicated plasmids grown for 4 h either in the dark or under 460 nm light (5 W/m^2) illumination. pBLADE$^{ONLY\_A}$, pTrc99a deprived of the pTrc promoter expressing only BLADE. pSiMPl^K_C-BLADE, pTrc99a expressing BLADE driven by J23101** promoter with the C-terminal half of the Kanamycin resistance cassette. SiMPl^K_N-mCherry, pBAD33 expressing mCherry under P$_{BAD}$ with the N-terminal half of the Kanamycin resistance cassette. All values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only (dashed line; see Supplementary Figure 2). Values represent mean ± s.d. of at least n = 3 independent experiments. BLADE variant: FP6 driven by the J23101** promoter.*

2.1.10.2 BLADE can be used with pre-existing L-arabinose-responsive strains

To demonstrate the potential of BLADE to induce the expression of genes under the P_{BAD} promoter from a chromosomal locus, I obtained from the Huang lab a strain (KC717) in which the endogenous promoter of the *rodZ* gene (see 1.8.2) was replaced with P_{BAD}[244]. In the absence of L-arabinose, chromosomal AraC inhibits transcription from P_{BAD}, therefore *rodZ* is not expressed and the cells display a spherical shape[239,240,244,272]. In the presence of L-arabinose, the endogenous AraC induces transcription from P_{BAD} and, consequently, *rodZ* is expressed, changing the cell morphology from spherical to rod-shaped[244].

I transformed KC717 cells either with a modified pBLADE from which the P_{BAD} promoter and the mCherry gene were eliminated (pBLADE$^{ONLY\_C}$; population A, **Figure 46**, Supplementary Figure 2) or with an empty pBAD33 deprived of *araC* and P_{BAD} (pCAM; Supplementary Figure 2), which was used to allow growing both strains in the presence of the same antibiotic (population B). I then kept both populations either uninduced (in the dark for population A, and without L-arabinose for population B) or induced them for 4 hours (with blue light for population A and with L-arabinose for population B) (Figure 48a). I found that population A recovered the rod shape similarly to population B (Figure 48a,b). To showcase the power of optogenetics to quickly switch the induction off, after induction I incubated the cells in the dark (population A) or washed L-arabinose off (population B) to recover the initial phenotype. While it was possible to obtain spherical cells again after 2 hours of dark incubation, the cells induced with L-arabinose did not recover the initial phenotype and rather became even more rod-shaped (Figure 48a,b).
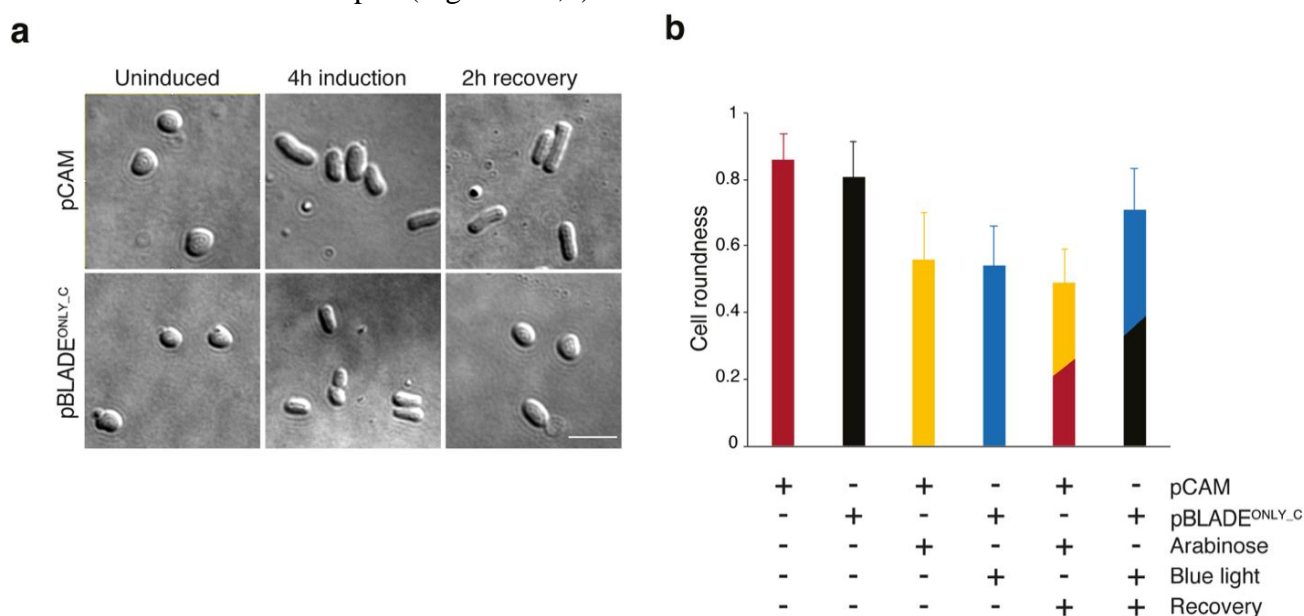
Figure 48 | BLADE is compatible with pre-existing L-arabinose-responsive strains. a, Representative DIC images of E. coli KC717 cells transformed with the indicated constructs at the indicated time points. pBLADE$^{ONLY\_C}$, pBAD33 deprived of the P_{BAD} promoter expressing only BLADE. pCAM, empty pBAD33 deprived of the P_{BAD} promoter. Induction indicates 460 nm light (5 W/m^2) for the cells transformed with pBLADE$^{ONLY\_C}$ and 0.2% L-arabinose for the cells transformed with pCAM. Scale bar, 5 μm Recovery indicates darkness for the cells transformed with pBLADE$^{ONLY\_C}$ and growth in a medium without L-arabinose for the cells transformed with pCAM. b, Quantification of the cell roundness for the samples and conditions in the upper panel. Values represent mean ± s.d. of n = 3 independent experiments. BLADE variant: FP6 driven by the J23101** promoter.

2.2 Biological applications of BLADE

2.2.1 Control of *E. coli* morphology and growth

Cell morphology impacts growth and survival in diverse environments[273,274]. Being able to change cell morphologies reversibly and with spatial precision could pave the way to experiments that could clarify its contribution to bacterial fitness and the adaptation in a particular environment, for example by studying in the same microscope field both wild-type and mutant phenotypes.

To demonstrate light control of cell morphology, I selected three *E. coli* proteins to overexpress: MinDΔ10, MreB and RodZ.

2.2.1.1 MinDΔ10 overexpression generates the minicell phenotype

As explained in the introduction (see paragraph 1.8.2), MinDΔ10 lacks the 10 C-terminal amino acids, constituting the membrane targeting sequence that allows MinD to bind to the membrane. Its heterodimerization with endogenous MinD prevents this latter one from binding to the membrane, which, in turn, leads to a failed recruitment of MinC there, with consequent formation of the FtsZ ring at the otherwise prohibited polar locations. This causes the formation of the so-called minicell phenotype, characterized by minicells lacking chromosomal DNA and cells of variable length (Figure 49a).

My aim was to control with light the formation of minicells: I first cloned the *minDΔ10* gene into pBLADE and transformed it into MG1655 *E. coli* cells, then I maintained them in the dark or exposed to 4 hours of blue light illumination. Within the induced population, more than half of the cells became minicells (Figure 49b,c); besides, the cell length variability in the population significantly increased, with more than 15% of the cells measuring between 3.6 to 5 µm (Figure 49b,c). The cells kept in the dark did not exceed 3.6 µm in length and were indistinguishable from those transformed with pReporter_only, used as a negative control (NC, Figure 49b,c). The phenotype was not caused by the illumination, as the cell length distributions of the negative control illuminated with blue light or kept in the dark were almost identical (Figure 49c).
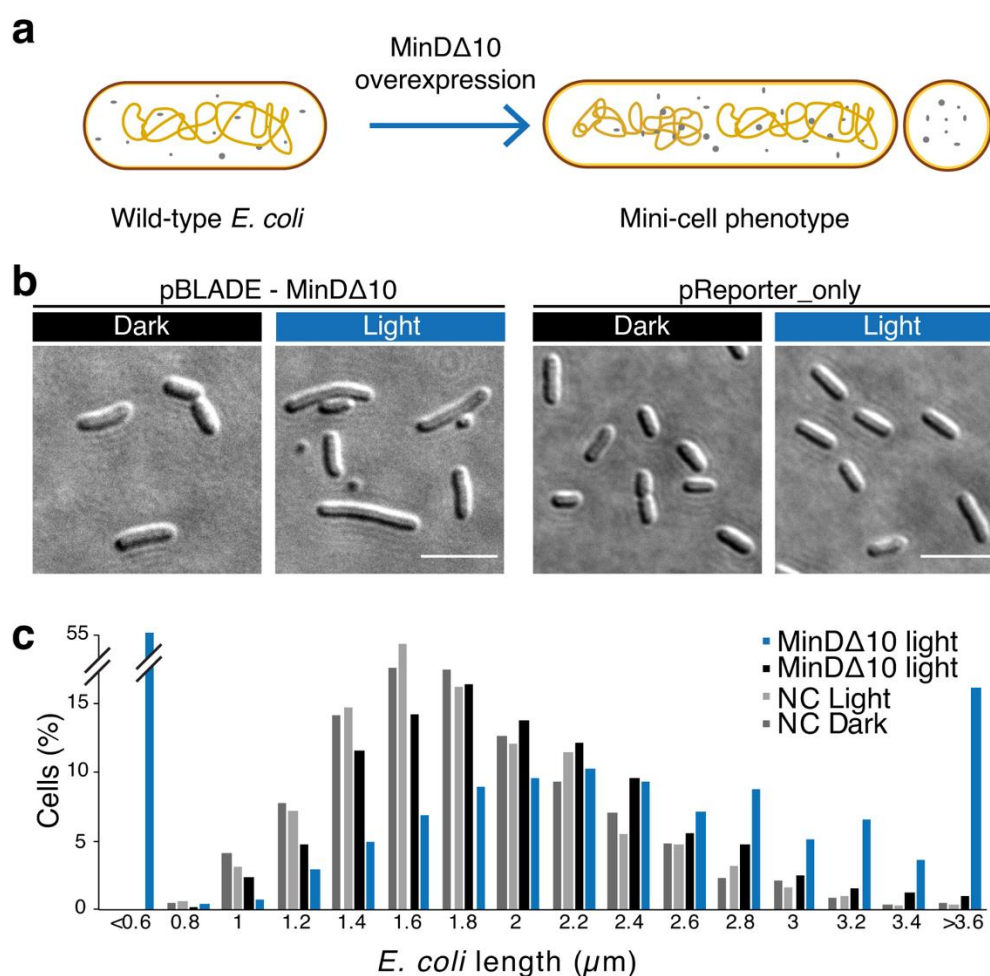
*Figure 49 | The minicell phenotype can be triggered with light using BLADE. a, Phenotype expected when overexpressing MinDΔ10. b, Representative DIC images of E. coli MG1655 cells transformed with the indicated constructs grown for 4 h either in the dark or under 460 nm light (5 W/m²) illumination. Scale bar, 5μm. c, Quantification of cell length distribution for the indicated samples and conditions. Cells smaller than 0.6 μm are classified as minicells. NC, pReporter_only (see Supplementary Figure 2). Values represent mean ± s.d. of n = 3 independent experiments. BLADE variant: FP4 driven by the J23101** promoter.*

2.2.1.2 MreB and RodZ overexpression increases cell length and thickness

MreB is the bacterial actin homolog, and together with RodZ is involved in the establishment and maintenance of the rod shape in *E. coli* (see paragraph 1.8.2). The overexpression of each of these proteins is known to make cells longer and thicker[234,239,243] (Figure 50a).

I cloned the *mreB* and *rodZ* genes into pBLADE individually, transformed each into MG1655 *E. coli* cells, then illuminated the cultures for 4 hours with blue light or kept them in the dark as a control. BLADE-induced MreB and RodZ overexpression led to cell elongation and thickening (Figure 50c,d). As seen also for MinDΔ10 expression, the low leakiness of BLADE did not cause conformational changes to the cells kept in the dark, which were indistinguishable from those of the negative control (Figure 50c,d).
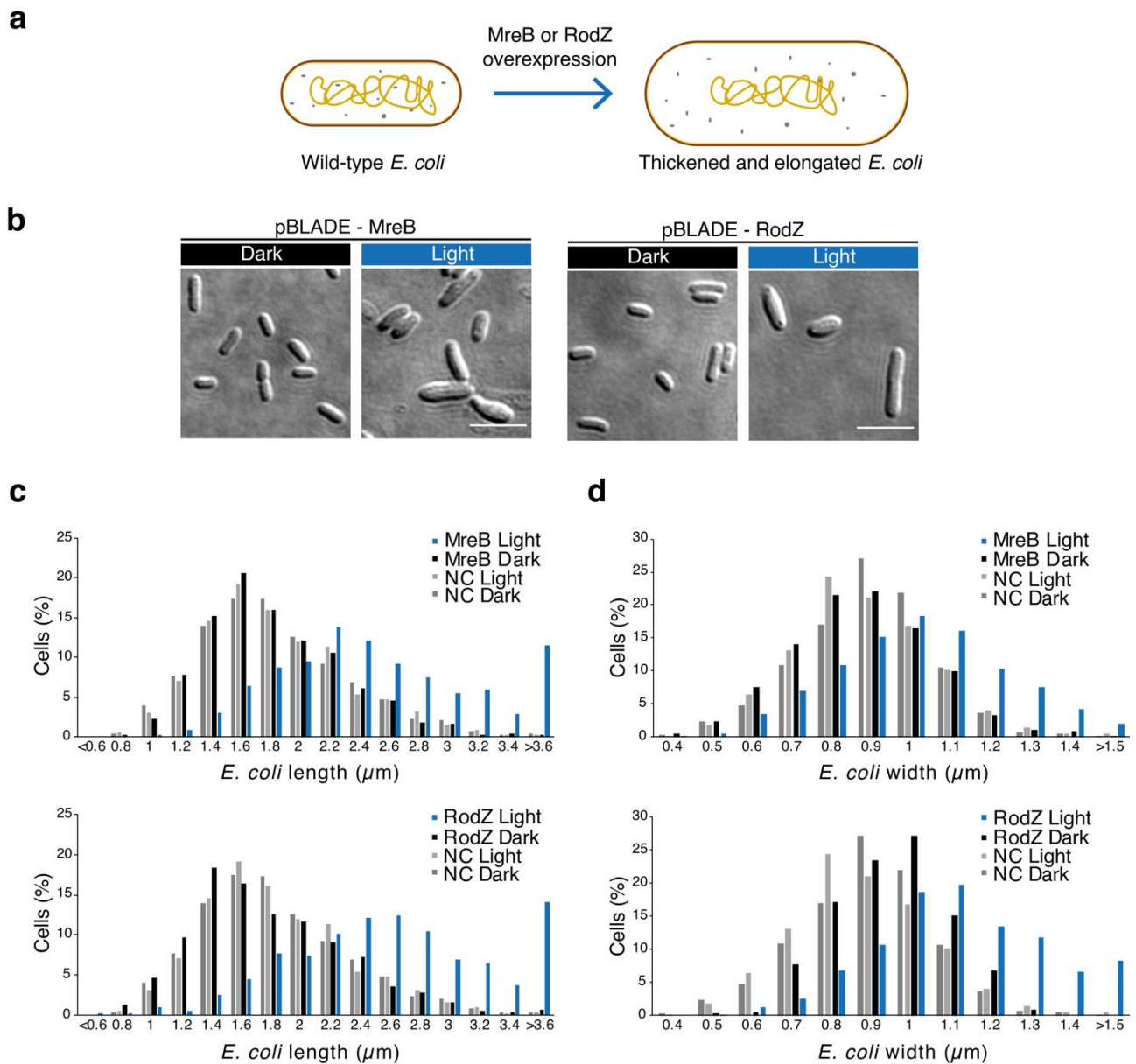
Figure 50 | BLADE-mediated MreB and RodZ overexpression induces cell elongation and thickening. a, Phenotypes expected when overexpressing MreB and RodZ. b, Representative DIC images of E. coli MG1655 cells transformed with the indicated constructs grown for 4 h either in the dark or under 460 nm light (5 W/m^2) illumination. Scale bar, 5 μm. c, d, Distribution of cell length and width for the indicated samples and conditions. NC, pReporter_only (see Supplementary Figure 2). c,d, Values represent mean ± s.d. of n=3 independent experiments. b-d, BLADE construct: FP4 driven by the J23101** promoter.

2.2.2 Characterization of *E. coli* genes with poor or unknown function

About 35% of the genes in *E. coli* still lacks characterization (see paragraph 1.7). To obtain more information about their functions, a first characterization can be performed through their overexpression, in order to observe the gene effects on bacterial growth and morphology. Another feature to assess is the intracellular localization of the proteins, through fusion with a fluorescent protein. Their overexpression can be performed with light as an inducer: light is a convenient choice for medium- to high-throughput applications due to its low cost, scalability, and reduced samples

handling. For these reasons, I thought of using BLADE to overexpress some of these genes to study their intracellular localization and the effects on cell morphology and growth.

2.2.2.1 Creation of the gene library and first characterization using *in silico* tools

I randomly selected 34 genes with no annotation in any gene database and included 5 additional genes, for which some information was available: *ydaT*, whose overexpression in *E. coli* reduced survival and led to cell elongation[275]; *ydiY*, a predicted outer membrane protein, which was shown to be induced by acid[276]; *ycbK* (renamed MepK), described as a murein hydrolase involved in cell wall synthesis[277]; *yehS*, a protein that, if downregulated, seemed to improve the growth of *E. coli* in presence of n-butanol and n-hexane[278]; and *yebE*, which was shown to be inducible by copper in a CpxA/CpxR-dependent manner[279], and whose protein product was predicted to be localized to the inner membrane[280].

Importantly, fluorescence microscopy-based localization studies have not been so far carried out for any of these 39 genes. Since fusion to a fluorescent protein could lead to alteration or impairment of the function of the gene products, I cloned each of the 39 genes in native form into pBLADE (Figure 51). However, in order to monitor the localization of the gene products in *E. coli*, I additionally cloned for each gene a fusion to sfGFP either at the N- or C-terminus (Figure 51). The choice to use both termini was made because the terminus at which the fluorescent protein is fused can play a role in the localization and in the propensity of the fusion protein to fold correctly[281–283]. In total, I constructed a library of 117 plasmids. I used those bearing the native genes for growth assays and differential interference contrast (DIC) microscopy, while those bearing the fusions to sfGFP were used for fluorescence microscopy (Figure 51).

Before performing experiments, Mehmet Öztürk, applied bioinformatics and computational structural biology approaches to predict the function and localization of the 39 selected genes (Figure 51). We used three different tools, namely Argot2.5[284], PANNZER2[208], and DeepGoPlus[209] (see 1.7), to predict protein function and localization based only on the amino acid sequence information, and one protein 3D modelling tool, called Phyre2[206], to use secondary and tertiary structure information to find the best template structure for 3D modelling (Supplementary Table 4). We generated a consensus table summarizing the predicted localization and function. To consider the prediction reliable, we considered only the predictions shared by at least two out of the four methods (Supplementary Table 5). A consensus was found for 14 out of 39 genes for functional prediction and 21 out of 39 for localization prediction.
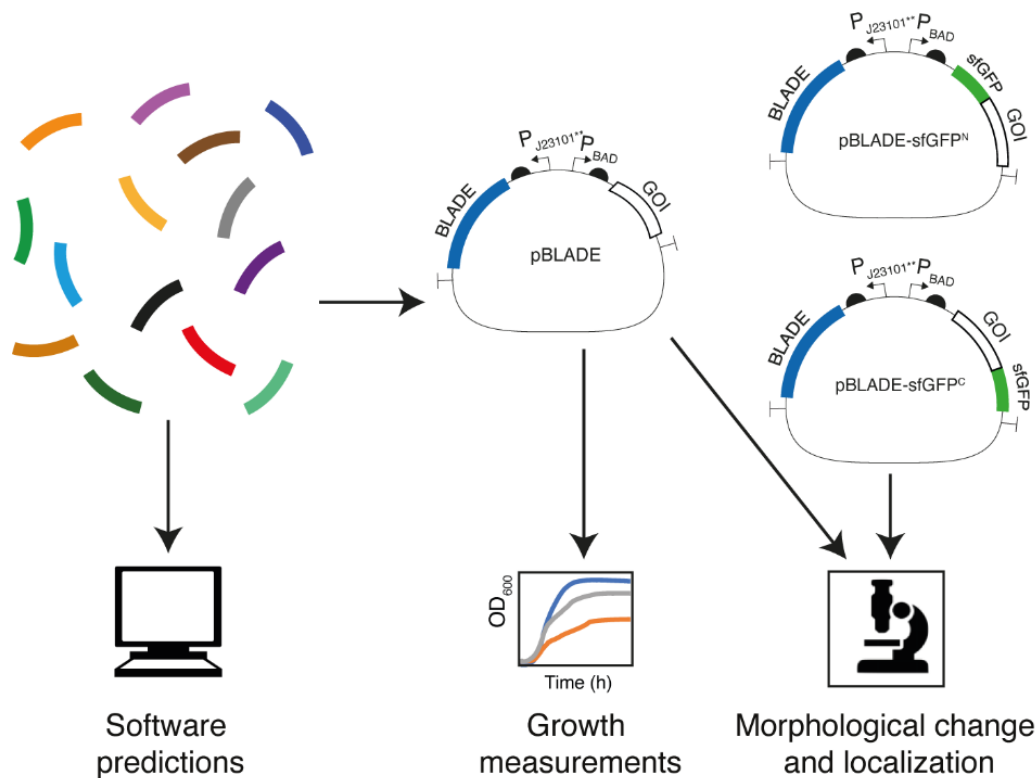
Figure 51 | Creation and characterization of the library of genes with poor or unknown function. *Overview of the workflow. Gene function is first predicted with online webservers, then the genes are cloned into pBLADE, alone or with sfGFP fused to them. Next, growth curves are collected and the effect of the overexpression of the genes on the bacterial morphology and their localization are studied. pBLADE-sfGFP^N, plasmid for sfGFP N-terminal fusion; pBLADE-sfGFP^C, plasmid for sfGFP C-terminal fusion.*

2.2.2.2 *in vivo* characterization of the library

The first experiment I performed was to analyse if the gene overexpression could affect bacterial growth. By measuring the absorbance of cultures illuminated with blue light with an automated plate reader for 20 hours, I found six genes whose products significantly affected the growth of MG1655 cells: three that had a positive effect (*yahC, yebE* and *yebY*) and three that had a negative effect (*yhhM, yjeO* and *ypaB*) (Figure 52 and Supplementary Figure 6). At least for YbeY, the functional prediction suggesting it to have transaminase activity (Supplementary Table 4) might explain why the cells overexpressing it grew faster.
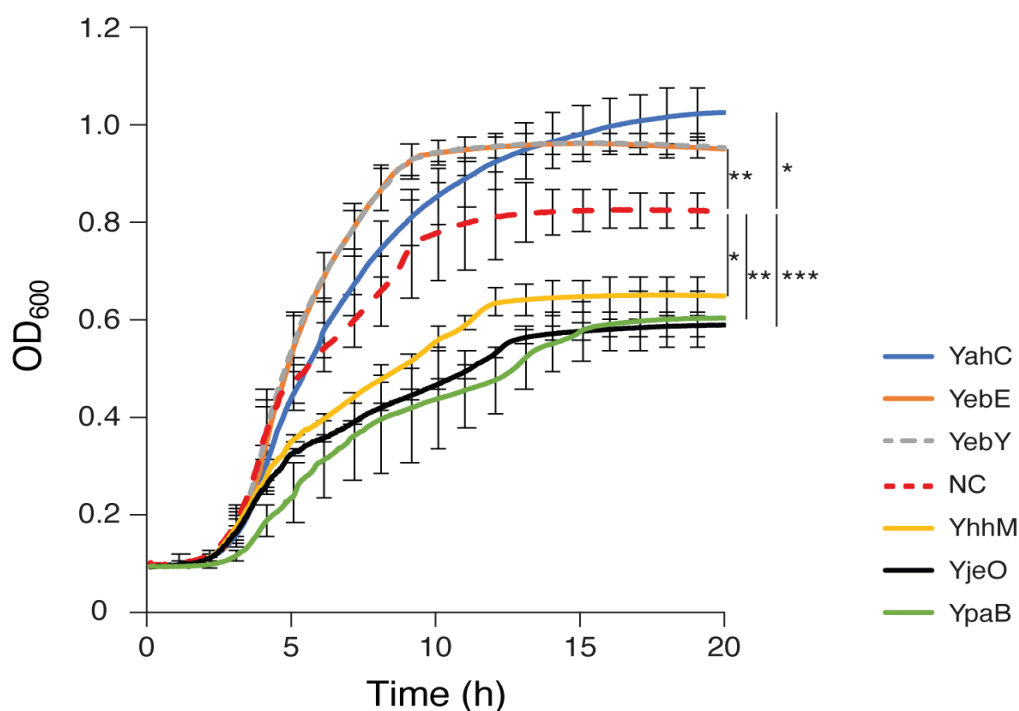
Figure 52 | Overexpression of six genes with unknown function alters bacterial growth. *Growth curves of E. coli MG1655 cells transformed with pBLADE carrying the genes encoding the indicated proteins. NC, cells transformed with the empty plasmid. Values represent mean ± s.d. of n=3 independent experiments. Single asterisk (*), p-value<0.5 (two-tailed, homoscedastic Student's t-test); double asterisk (**), p-value<0.01 (two-tailed, homoscedastic Student's t-test); triple asterisk (***), p-value<0.001 (two-tailed, homoscedastic Student's t-test). BLADE variant: FP6 driven by the J23101** promoter.*

Then, I assessed if any of the 39 genes caused morphological changes, by performing DIC microscopy on MG1655 cells exposed to light for 4 hours. The cultures were grown in a 96-well plate, to allow simultaneous induction of different samples. The plate was placed into a custom-made box with LEDs inside, illuminating the samples from the top (Supplementary Figure 7). The box was then positioned into a shaker at 16 °C, to reduce the propensity of the induced protein to aggregate. Most genes did not cause morphological alterations, however, two led to cell elongation (*ydaT* and *ydhL*) and one to cell lysis (*yhcF*; Figure 53a). These results confirm previous observations on the effect of *ydaT* overexpression on cell morphology found in the literature[275], and further indicate that *ydhL* could be involved in cell division.

In this assay, *yhcF* overexpression seemed to be toxic to the cell: to confirm this behaviour, I additionally measured the absorbance of the cell culture after 4 hours of growth in the incubator and found that it was indeed reduced compared to that of the cultures overexpressing *ydaT*, *ydhL* as well as of cells transformed with empty pBLADE as a control (Figure 53a).

Interestingly, *yhcF* overexpression did not cause growth defects in the assay performed in the 96-well plate (Supplementary Figure 6). However, the bacterial growth in a 96-well plate is slower than in a flask, due to lower oxygen exchange and shaking. Therefore, potentially other genes beyond those six might have some impact on cell growth.

a

b

Protein name	OD600 dark	OD600 light
YdaT	3.3	3.2
YdhL	3.2	3.0
YhcF	3.0	1.3
NC	3.8	3.8
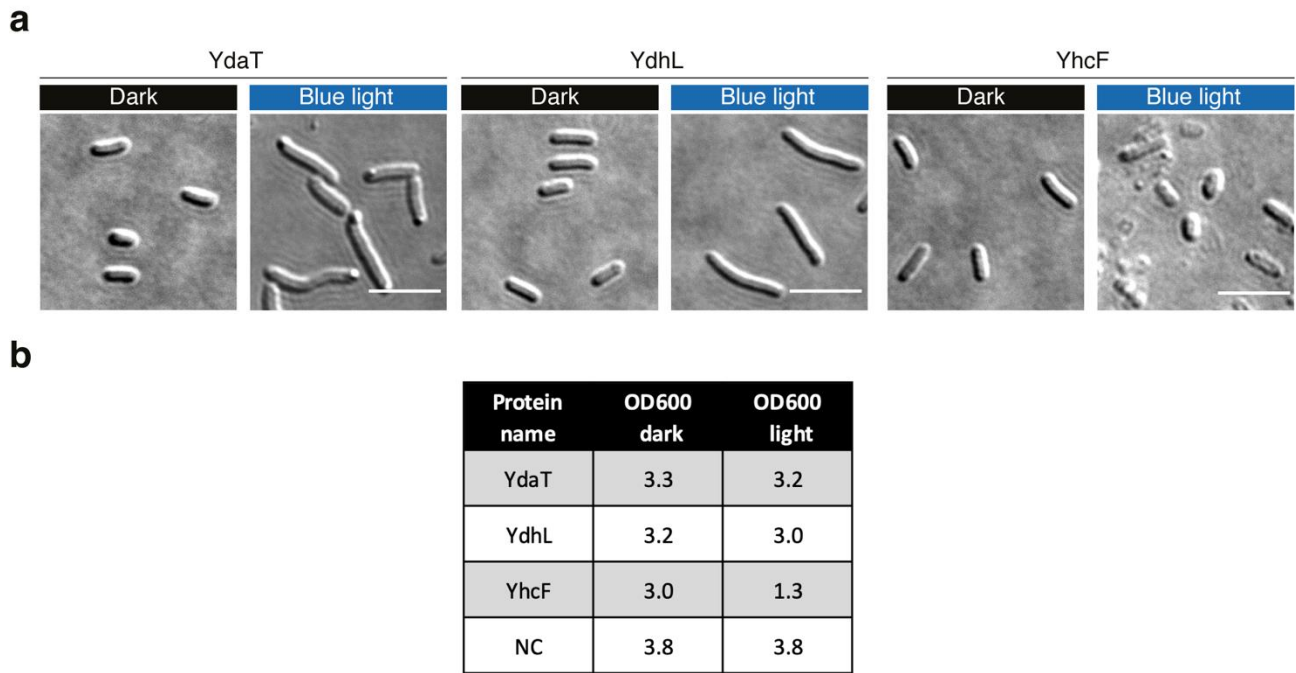
*Figure 53 | Overexpression of YdaT, YdhL and YhcF alter E. coli morphology. **a**, Representative DIC images of E. coli MG1655 cells transformed with pBLADE carrying the genes coding for the indicated proteins grown for 4 h either in the dark or under 460 nm light (5 W/m^2) illumination. **b**, OD$_{600}$ of the cultures overexpressing the indicated proteins grown in individual tubes in an incubator for 3 hours in the dark or under 460 nm light (5 W/m^2) illumination. Values represent a single experiment.*

Next, I performed fluorescence microscopy to study the localization of the protein products of the selected genes with poor or unknown function. Not all gene products tolerated fusions to either terminus, as expected (Supplementary Table 5): either the fluorescence was barely detectable for one of the two termini, or the localization changed depending to the terminus at which the fusion was made (Figure 54a,b, and Supplementary Table 5).
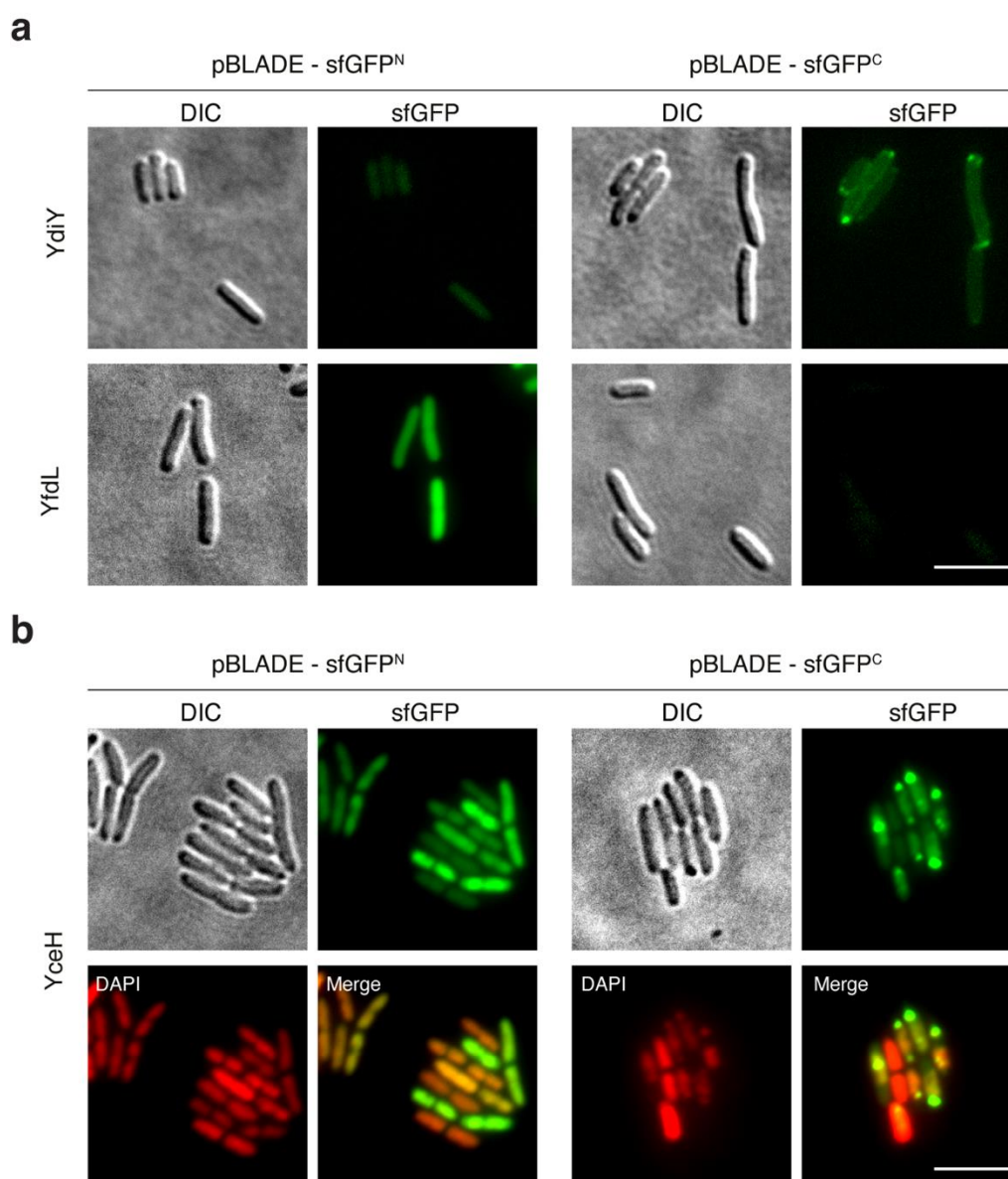
Figure 54 | Fusion to the fluorescent protein sfGFP may alter the expression and localization of a gene of interest.
a,b, Representative images of E. coli MG1655 cells transformed with the indicated constructs grown for 4 h under 455 nm light illumination (5 W/m^2). pBLADE-sfGFP^N, N-terminal fusion; pBLADE-sfGFP^C, C-terminal fusion. Scale bar, 5 μm.

I found 3 unknown proteins that co-localized with the nucleoid, 14 that localized to the cytoplasmic membrane, and 26 that formed foci (Figure 55b). I am aware that the localization alone is not sufficient to reveal the gene function, but still, it gives important information and could give hints on potential mechanisms of action. For example, since I found YdaT co-localized with the nucleoid (Figure 55a, b), having been reported to be a toxin[275], it may inhibit DNA gyrase binding to it. Other toxins that inhibit DNA gyrase and co-localize with the nucleoid have been already described[285,286].

DIC	sfGFP	Dye	Merge	Protein with same localization
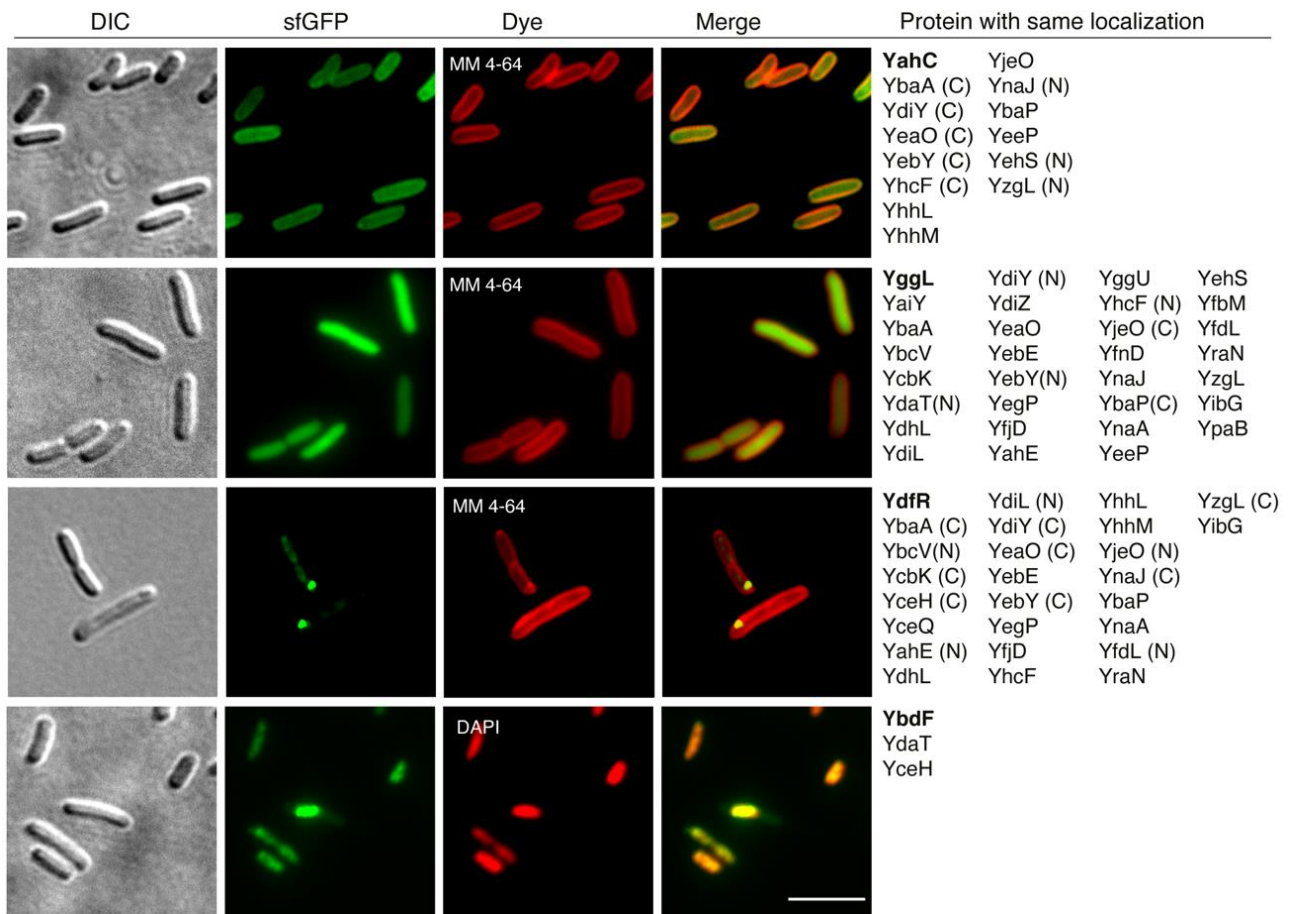
Figure 55 | Characterization of the intracellular localization of the protein products of the 39 genes with poor or unknown function. Representative images of E. coli MG1655 cells transformed with pBLADE-sfGFP[N] carrying the proteins indicated in bold in the right-most column grown for 4 h under 460 nm light (5 W/m[2]) illumination. (N), localization obtained only with N-terminal fusion; (C), localization obtained only with C-terminal fusion. The membrane was stained with the MM 4-64 dye and the nucleoid with DAPI. Scale bar, 5 µm. BLADE variant: FP6 driven by the J23101[**] promoter.

YebY, whose overexpression promoted faster growth (Figure 52), was localized to fluorescent clusters when C-terminally fused to sfGFP (Figure 56). Visualization of the cells through transmitted light microscopy highlighted swellings in correspondence of the fluorescent clusters (Figure 56 and Supplementary Figure 8). Conversely, YebY showed a weak cytoplasmic fluorescence when the sfGFP fusion was at the N-terminus (Figure 54 and Figure 55). Argot2.5[284] and PANNZER2[208] predicted YebY to have transaminase activity, while the closest known template that matched the predicted structure was that of a lipid-binding protein (Supplementary Table 4).

To verify whether the localization corresponded to a specific function or was just the result of aggregation (*e.g.,* due to misfolding), I performed fluorescence recovery after photobleaching (FRAP) with a confocal microscope (Figure 56), observing a recovery of the fluorescent cluster 7 minutes after its bleaching, arguing against the hypothesis of a nonfunctional aggregate.
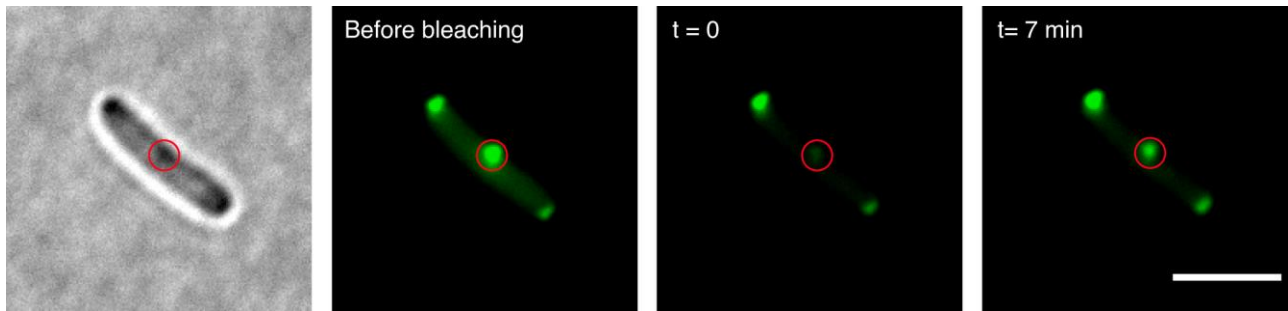
Figure 56 | The YebY-sfGFP clusters are dynamic. *Fluorescence recovery after photobleaching performed on a cell expressing YebY-sfGFP after 4 h of blue light illumination. The red circle points to the area that has been photobleached with blue light using a confocal microscope. After 7 minutes of incubation in the dark, the fluorescence is almost restored. Scale bar is 3 μm.*

3. Discussion

P_{BAD} is as an inducible promoter widely employed by microbiologists, synthetic biologists and biotechnologists. It is part of the L-arabinose operon, and its induction is regulated by the well-characterized AraC, one of the most studied bacterial transcriptional regulators.

In this work, I have developed, starting from preliminary experiments performed by Leonard Ernst (former student in our laboratory), a library of AraC-derived TFs, which we called BLADE, that activates transcription from the P_{BAD} promoter in response to blue light instead of L-arabinose. These constructs are based on the VVD photoreceptor from *Neurospora crassa*, fused to the DNA binding domain of AraC. I demonstrated by SEC that photoactivation triggers the dimerization of the VVD-AraC chimeric proteins, and showed that in vivo light leads to the binding to the I_1I_2 half-sites of the P_{BAD} promoter, mimicking the L-arabinose-bound dimer of AraC, that triggers the transcription of downstream genes.

In the following, I will discuss the role covered by the constitutive promoters driving BLADE expression, interdomain linkers connecting VVD and AraC$_{DBD}$ and DNA binding sites in the induction level achieved by BLADE, proposing potential optimizations that can guarantee an even tighter control of the induction or higher expression levels. I will then analyse the results obtained with the comparison of BLADE with AraC, highlighting the strengths and weaknesses of both systems. I will then discuss the relevance of the biological applications that I performed in this work. Then, I will analyze the advantages of using BLADE with pre-existing strains or plasmids and in library screenings.

Finally, I will provide new potential applications of BLADE in organisms different from *E. coli*, new ideas on future versions of BLADE and propose further assays to continue the characterization of the unknown gene library.

3.1 Characterization of the components involved in BLADE induction

3.1.1 The role of the linker within BLADE and the constitutive promoters driving its expression in BLADE-mediated gene expression

The induction experiments performed with the library of BLADE constructs show that the linker connecting VVD and AraC$_{DBD}$ has a role in the induction level of the construct; however, there is no clear rule to follow in selecting the optimal linker length to obtain the highest fold-change and the lowest leakiness of BLADE. The variant FP6, the one that reaches the highest fold-change (driven from the J23101* promoter), harbours a 10-amino acid linker, an intermediate length among the tested ones, that range from 0 to 39 (the sum of natural and synthetic linkers, Figure 13a and Supplementary Table 1). Nevertheless, I observed that no linker at all (FP8) and an excessive long linker between the two domains (FP5) are not ideal for the correct functioning of BLADE (Figure

15). I can therefore empirically limit the linker length that allows us to reach decent induction levels within 7 (FP2) and 14 (FP3-FP7) amino acids.

Ideally, the best linker should guarantee the same flexibility degree of the two AraC domains as in the wild-type protein. Unfortunately, the structure of the complete dimer has not been resolved yet; therefore, it was not possible to perform a direct comparison of the BLADE constructs and AraC. I then decided, with the help of Mehmet Öztürk, to perform a comparison of the structure flexibility (measured in terms of the root mean square fluctuation of the Cα atoms of the amino acids composing the protein) between the model structures of FP4, chosen as the best-performer construct of the library developed by Leonard Ernst, and the full-length AraC, to check the differences in the structure flexibility. The model structure of the two proteins, predicted from their amino acids composition, was used for the calculation (Supplementary Figure 1). This analysis allowed us to spot a high flexibility peak within the GS linker that was not present in AraC. After its removal, the new construct, named FP6, showed a fluctuation profile more similar to AraC. Interestingly, this construct led to a higher induction level than FP4 (Supplementary Table 2), suggesting that the flexibility of the linker region was detrimental for the binding stability of the dimer to P_{BAD}. This experiment showed that, other than the linker length, flexibility is an important parameter to play with to optimize transcription induction.

The choice of the constitutive promoter driving BLADE transcription is also important in the achievable level of reporter gene induction. Promoter J23101* is weaker than J23101**, and indeed BLADE transcribed by this promoter always led to lower induction levels of the gene downstream of P_{BAD} (Figure 15a). However, as the leakiness level in the dark is reduced in most of the constructs as well, this translated into a higher fold-change for the constructs expressed with this promoter. As described in paragraph 2.1.9.5, in the dark BLADE monomers probably bind the high-affinity I_1 site within P_{BAD}. If BLADE monomers are transcribed in higher concentration thanks to J23101**, the probability of dimer formation in the dark, with consequent binding to I_1I_2 and following transcription activation, is higher. The expression level obtained with a pBLADE harbouring a mutation on the I_2 half-site (Figure 37) confirmed this hypothesis: the better affinity of I_2 with $AraC_{DBD}$ increased the leakiness almost 4 times.

From the expression levels of Figure 15, I also observed that the fluorescence levels induced by the BLADE constructs are more similar to one another when their expression is driven by the stronger J23101** promoter (except for FP5). This could be explained by the fact that, although every construct binds with a different affinity to the I_1I_2 sites of P_{BAD}, the overproduction of the BLADE constructs can partially compensate for this difference, as more dimers are available to bind when those bound detach.

The fluorescence values measured for pBLADE(AraC$_{WT}$), driven by either J23101* or J23101** without L-arabinose are slightly lower than those of the negative control, pReporter_only, upon which all the fluorescence values have been normalized (Figure 15a and Supplementary Table 2). From the literature, it is known that the number of AraC molecules in the cell is around 20 when no L-arabinose is present[137], and the copy number of pBLADE, which harbours the p15a origin of replication (the same of pBAD33), is 20-30[287]. In cells transformed with pBLADE(AraC$_{WT}$)-mCherry, AraC transcribed from the plasmid could contribute, alongside the endogenous AraC, to form the DNA loop for all the plasmids of the cell, preventing the recruitment of RNA polymerase more effectively than the endogenous AraC only.

3.1.2 The importance of low leakiness level for a transcription factor

One of the strengths of the L-arabinose-inducible plasmid pBAD33 is its tightness in its uninduced state when compared to other systems like the IPTG-inducible plasmid pTrc99a. A minimal expression in absence of the inducer allows the plasmid to be used in contexts where the expression of a target gene must be precisely regulated, and unwanted expression avoided. This is why I took special care to engineer BLADE with minimal leakiness, prioritizing this parameter over a higher but less tight induction level.

Generally, leakiness in optogenetic gene expression systems is assessed by simply looking at the levels of reporter expression in the absence of the stimulus (in the dark for activators, in the light for repressors). However, this does not consider whether the expression in the uninduced state is already too high compared to the expression in the absence of the regulator. In the case of BLADE, minimal leakiness was demonstrated by comparing its activity in the dark with expression obtained with the same control plasmid deprived of the TF (Figure 15 and Supplementary Table 2). For example, I would state that the best-behaved construct of the library is BLADE FP6 driven by the J23101* promoter, which showed a dark state level only 1.15-fold higher than the control but maintained an induction fold-change of almost 17-fold. I put BLADE to the test by expressing several functional *E. coli* proteins whose overexpression causes morphological changes to the cells and showed that, in the dark, cells are indistinguishable from the control (Figure 49 and 50).

To cite a practical example about the importance of tightness, a strong MinD overexpression in *E. coli* would impair its pole-to-pole oscillation, hence cell division; when the levels of MinD are too high compared to the levels of MinE (whose task is to displace MinD from the membrane) the entire cell membrane would be occupied by the extra MinD copies, which in turn would recruit MinC everwhere, leading to FtsZ inhibition throughout the entire cell length. The experiment presented in Figure 17 indicates that eYFP-MinD oscillations were not detectable in the dark, as long as BLADE FP6* was used. Interestingly, BLADE FP4** appeared not to be suitable for this experiment: when I cloned mCherry-MinD in pBLADE with this variant, the leakiness level showed oscillations already in the dark state. When the cells were illuminated with blue light, mCherry-MinD fluorescence was more uniformly distributed on the membrane and the oscillations stopped or did not work properly anymore, leading to longer cells. (Supplementary Figure 3). The reason why I switched from mCherry to eYFP was only to have a brighter fluorescent protein.

The use of the stronger promoter J23101**, however, could be useful in other situations where the level of expression must be higher, and the leakiness control can be less stringent. For example, MinDΔ10 overexpression was controlled by BLADE FP4** (Figure 49): the truncated version of MinD must be present in amounts high enough to bind to the endogenous MinD molecules and prevent them from binding to the membrane. The phenotype of the cells transformed with pBLADE and kept in the dark was identical to the control cells transformed with pReporter_only. I imagine the existence of a MinDΔ10 concentration threshold above which the oscillations start to be impaired and the minicell phenotype is triggered. The use of BLADE driven by the J23101* promoter could be potentially unsuitable for MinDΔ10 overexpression because the lower level in the induced state could not overcome the concentration threshold or request a longer illumination time to reach it, and the lower leakiness in the dark would not make any difference. Similarly, for the expression of genes with poor or unknown function, I chose BLADE FP6**, because a high induction level would have

provided better visualization of the protein localization and clearer phenotypical changes caused by their overexpression.

3.1.3 Activation of reporter gene expression at the single-cell level

The most obvious advantage of light compared to a chemical is its high spatial precision: light coming from a laser can be confined to a small spot and therefore gene expression can be induced even in individual cells when using an optogenetic system. In this work, I have given several demonstrations of spatial control achievable with BLADE: I have shown micro-colonies activated with a confocal microscope (Figure 18) and I also have demonstrated that a photomask can create an induction pattern based on the light that gets blocked or transmitted (Figure 19, Figure 20 and Figure 21). The creation of a picture employing bacteria as the smallest item of information, like pixels in digital pictures, has been already performed[185,191,275,276], but none has ever reproduced complex images like the one I used in this work.

As the light was kept on overnight, to minimize the risk of sample overheating and fluorophore bleaching, the intensity of the LED on top of the plate was set to 1 W/m^2. This intensity was sufficient to activate reporter gene expression in the *E. coli* lawn underneath the photomask (Figure 16), as long as this was printed in black and white. A previous attempt to create the bacteriograph of the BLADE protein model structure (Supplementary Figure 4) performed with a coloured photomask failed to represent the DNA backbone, because that was represented in red, a wavelength far from the excitation spectrum of VVD[90].

The shades of sfGFP observed in the bacteriographs, however, were not created by the light passing through the shades of grey composing the photomask, illuminating the bacteria with different intensities. It was instead the extremely faithful reproduction of the shadows and lights cast by the black and transparent dots that, present in different concentrations on the photomask image, create the illusion of shades when seen from non-microscopic distances. This common printing technique, called dithering, can be easily seen from the magnifications of the "Creation of Adam" bacteriograph (Figure 21c).

Apart from the creative purpose, the spatial control experiments presented in this work demonstrated that BLADE could selectively induce transcription with single-cell precision, having uninduced and induced cells in the same field of view under the microscope. This would consent to directly compare the phenotype obtained by overexpression of the gene of interest, maintaining the same experimental conditions and limiting the human and biological errors that might arise when using two or more samples.

3.1.4 BLADE expression levels can be easily tuned

Another strength of AraC induction of P$_{BAD}$ is its wide range of expression modulation obtainable by titration of the L-arabinose concentration[289]. BLADE-mediated induction of P$_{BAD}$ can be tuned as well, not only by choosing the most suited construct/promoter combination from the library (Figure 15), but also using different light intensities (Figure 16). The induction curve over time follows a sigmoid shape at all the intensities tested (Figure 24), therefore it is easy to calculate the induction time and the light intensity needed to reach the desired expression value. As already mentioned, the *strongest* light setting used to illuminate the cells (continuous illumination with 5 W/m^2 blue light) was perfectly tolerated within the 6-hour time range of the experiment, as growth was not affected

(Figure 11). The reasons to reduce the light intensity is to regulate the induction level of the target gene, but also to prevent potential toxicity in long-term induction, as the outcome on the cell growth after the sixth hour has not been tested.

I always employed constant light, instead of pulsatile illumination, to activate BLADE, despite the 3-hour reversion kinetic of VVD[90] made unnecessary this continuous induction signal. With constant light, however, I aimed to activate immediately any new BLADE protein being expressed during the experiment. Nevertheless, a light pulse of a few seconds every 10 minutes could be enough to activate effectively the transcription of each cell within the culture. As the doubling time of *E. coli* during exponential phase is around 20 minutes, every cell, including the newborn ones, will induce BLADE dimerization in a short time.

It is worth mentioning that the light intensities used in this study are easy to obtain with commercial and cheap LEDs (Figure 10). In the absence of a power supply that allows tuning the voltage or current to change the LED brightness, a simple alternative would be to change the distance of the LED to the bacterial culture.

3.1.5 I_1 and I_2

The difference between the I_1 and I_2 half-sites of P_{BAD} lies in their different binding affinities for $AraC_{DBD}$. I_1 is always occupied by AraC but, without L-arabinose, the O_2 half-site is preferred over I_2. The presence of L-arabinose rearranges the AraC dimer, leading to I_2 binding and consequently to the RNA polymerase recruitment. BLADE functions differently: in the dark, it is mostly monomeric, but the different affinity of it towards the P_{BAD} half-sites is still crucial for the transcription activation: only the dimeric form of BLADE allows for proper binding of I_2 as its affinity for the monomer is too weak.

In our engineering process, I found two ways to increase the induction level: a higher concentration of the monomer, by using a strong promoter to express BLADE (J23101**, as described in paragraph 3.1.1), or an increased the affinity of I_2 for BLADE (Figure 37). Both cases lead to higher induction levels in both the dark and the lit states. An *in vitro* experiment to assess how the binding changes with different BLADE transcription levels is the Chromatin immunoprecipitation (ChIP). To ensure the highest precision in the detection of the binding site (each half-site is 17-bp long and separated by 4 nucleotides) this assay can be coupled with exonucleases (ChIP-exo) that degrade the unbound DNA, but not the DNA crosslinked with the $AraC_{DBD}$ of BLADE. Following sequencing of the DNA fragments will then reveal the involvement of I_1 and I_1I_2 to the binding. The sequencing of the DNA fragment bound by BLADE can reveal the involvement of I_1 and I_1I_2 to the binding, and how it changes with different concentration of BLADE.

As mentioned earlier, expression of AraC from the weak J23101* and the strong J23101** promoters in pBLADE did not cause particular changes to the expression levels of the reporter gene, both in the uninduced and induced states, revealing that this process was not influenced by the concentration of AraC in the cell (Figure 15 and Supplementary Table 2). However, by comparing the induction level of pBAD33-mCherry after 4 hours of induction in Figure 22a and pBLADE(AraC*/**)-mCherry in Figure 15, I could see an increase of about 10-fold when the synthetic promoters were used. The uninduced levels were comparable, a sign that the looping effectively prevented any unwanted transcription even with the presence of additional AraC molecules, but a constitutive AraC expression did increase the induction level. The P_C promoter, regulating AraC transcription in the ara operon, is

negatively regulated by AraC itself and therefore is active only for a short time. Transcription from the constitutive J23101* promoter probably allowed to reach a saturation level of AraC in the cell, therefore the stronger J23101** promoter did not lead to significant increases in the induction level (Figure 15).

Increasing the affinity of BLADE for I_2 is the second strategy to obtain a higher induction level. The point mutation reported in the literature to increase transcription by AraC only after the addition of L-arabinose[248] increased significantly also the leakiness of BLADE. While the DNA loop successfully prevented apo-AraC from binding to I_1I_2* in absence of L-arabinose, the stronger binding of BLADE dimer in the dark was not as effectively hindered (Figure 37). Due to the higher affinity of I_2*, it is also possible that a monomer of BLADE would be able to bind it; as it is known that the binding to the half-site proximal to the RNA polymerase recruitment site is responsible for transcription activation, this mutation would allow a monomer to induce the transcription in the dark. A simple experiment to prove this hypothesis is the introduction in BLADE of the mutation C71S or Y40K in the VVD domain, that will prevent dimer formation[96]: this will allow evaluating the monomer contribution to the binding of I_1I_2* by quantification of the fluorescence level. It would be also interesting to measure the fluorescence level of this mutant with the wild-type I_1I_2 half-sites, to evaluate the involvement of the monomer in the leakiness level.

Theoretically, a P_{BAD} composed of two I_1 half-sites should provide a stronger, but unregulated induction of gene expression, since already back in 1993, Reeder and Schlief showed that such promoter was insensitive to L-arabinose and constitutively active with apo-AraC[248]. Our collaborators, Armin Baumschlager and Mustafa Khammash, however, revealed that BLADE can induce even a synthetic promoter made of two I_1 half-sites under illumination. This configuration, as expected, led to very high induction levels, both in the dark and under blue light. However, through an accurate regulation of the levels of BLADE itself (it will be explained in paragraph 3.8), they found out that by decreasing the strength of the promoter driving BLADE expression approximately 100 times, it was possible not only to reduce the leakiness level but also to obtain high fold-change induction.

In an attempt to create a more suitable binding site for BLADE dimers, with the help of Mehmet Öztürk we created synthetic promoters with the I_1 and I_2 half-sites oriented in different ways (Figure 36a). However, this strategy led to non-significative induction levels or no induction at all (Figure 36b).
The explanation to this failed experiment can be found in a paper of 2007 from Rodgers and Schlief[263]: they computationally predicted the approximate orientations and positions of the DBD of AraC, finding either interaction of the DBD with the DD from the same monomer (cis) or of DBD and DD from different monomers (trans), as represented in Figure 57. In both cases, the prediction indicated that the two DNA binding domains are inclined toward each other and located on the same face of the pair of dimerization domains[263]. This is in contrast to the two-dimensional representation that is usually used to summarize the protein interactions to the DNA[263]. By not taking into account this DNA bending, our simulation could not reflect the *in vivo* situation entirely. Further computationally expensive full-atom molecular dynamics simulations would be needed for a complete understanding of the AraC-DNA interaction details.
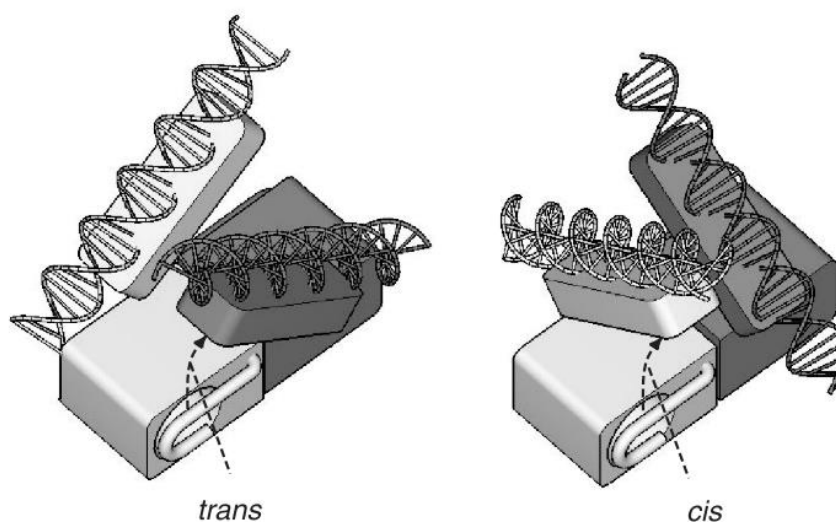
.*Figure 57 | Schematic representation of 2 models for the positioning of the two DBDs of AraC. Left, the DBD is placed near the N-terminal arm of the other subunit. Right, the interactions between DBD and DD come from the same domain. Adapted from*[263].

3.1.6 O$_2$

AraC binds to the O1 half-site shortly after the loop unwinding, following L-arabinose supplementation (Figure 4). During this short lag phase, the RNA polymerase is recruited at the P$_C$ site, allowing araC transcription. O$_2$, on the other side, is required for the loop formation in the absence of L-arabinose.

Despite being, strictly speaking, unrelated to P$_{BAD}$, I retained those regulatory elements when constructing pBLADE. Therefore, as L-arabinose was never added to the medium in my experiments with BLADE, the loop could naturally form, thanks to the chromosomal AraC transcribed by the cell. The loop is beneficial for maintaining a low leakiness in the dark: this was confirmed by the induction performed in the absence of O$_2$, and therefore, with no possibility for the loop to form, a condition that led to higher reporter fluorescence (Figure 40). Importantly, the measured induced levels were not significantly different compared to those obtained with the plasmid bearing the O$_2$ half-site, meaning that BLADE can easily bind to I$_1$I$_2$ when it is in dimeric form, after displacing the loop held by AraC bound to I$_1$ and O$_2$.

When AraC is transcribed at a higher level, for example with the contribution from pBAD33, the binding of dimeric BLADE is more hindered, and this translates in a lower induced level, but also a lower leakiness (Figure 47). These data demonstrate the involvement of O$_2$ in the containment of the leakiness and its involvement in future construct optimizations (see paragraph 3.2).

3.2 Strategies to lower the leakiness and increase the fold-change

As previously said, MinDΔ10, RodZ or MreB are not the right genes to assess the tightness of the BLADE, because they must be expressed beyond a critical level to cause a visible phenotype. eYFP-MinD, on the other side, was a good choice because MinD overexpression leads to cell elongation or even filamentation and death (see 1.8.1). However, the experiment I performed for which tightness was most critical has been the control of CcdB expression. Unfortunately, it turned out that the basal activity of BLADE in the dark was not sufficiently low in this case and the cells inserted mutations or deletions to inhibit the expression of *ccdB* to survive. Still, I obtained the same result when using the conventional arabinose-inducible AraC instead of BLADE, suggesting that the problem may be the P_{BAD} promoter itself. Interestingly, the P_{BAD} promoter has been used to control CcdB expression before, however in this case the ribosome binding site (RBS) upstream of the gene was substituted with a weaker one[290].

Thus, I assume that changing the RBS upstream of *ccdB* in pBLADE could be an effective way to control CcdB expression, therefore cell growth, with light. The failed control of CcdB made me realize that BLADE, despite being tight enough for many biological applications, can be made tighter. Below I describe some strategies that can be implemented in future versions of BLADE.

3.2.1 Additional I_1I_2 half-sites along the plasmid backbone

During a discussion with our collaborator Mustafa Khammash, the idea arose to insert multiple copies of the I_1I_2 half-sites on the plasmid backbone with the aim to sequester BLADE dimers potentially present in the dark away from P_{BAD}. Unfortunately, the results showed that the leakiness remained unchanged in the presence of multiple copies of I_1I_2, either using the wild-type promoter of the I_2* mutant. (Figure 38 and Figure 39). In the first case, as I used the BLADE FP6 variant, the leakiness was almost matching that of the negative control, therefore it would have been difficult to further decrease it, while for the second case the higher affinity for the mutated I_2 probably allowed the binding even of the BLADE monomers. Therefore, the strategy to sequester the dimers to reduce the leakiness might have not been the most appropriate.

To improve this strategy, I could clone on the backbone additional sequences of I_1I_1 that, as we have seen, lead to a stronger binding than I_1I_2. However, I_1I_1 would attract monomers as well, therefore this method would not be effective to contain the leakiness of I_1I_2*.

3.2.2 Co-expression of BLADE and AraC

Another method to reduce the leakiness whose effectiveness has been already demonstrated is through the transcription of plasmid-borne AraC together with BLADE: Figure 48c and Supplementary Table 3 show that, in the cells co-transformed with pBAD33 and pBLADE, AraC, transcribed by pBAD33, competes with BLADE for binding to P_{BAD}. Without L-arabinose, the DNA loop formed by AraC shielded the promoter and decreased the chance of BLADE dimers to bind in the dark. The competition also affected the binding to I_1I_2 after illumination, reducing the reporter transcription, but the overall fold-change in my experiment even increased slightly. Therefore, by cloning AraC driven by its natural promoter P_C into pBLADE, I speculate that expression in the dark could be further decreased. A more systematic characterization would involve a library of promoters controlling the expression of AraC in order to find the ideal transcription level that guarantees the lowest leakiness

or to prioritize the selection of the highest fold-change induction. To perform an even more rigorous characterization, I could find the best combination of promoters for AraC and BLADE to regulate the leakiness and the induction level of BLADE.

3.2.3 Creation of a larger linker library

I did not perform a systematic analysis to find the most suited linker between VVD and AraC$_{DBD}$, as I only tried seven linkers, different for length and amino acid composition, until I obtained a good induction level. This means that I only have a rough indication of how long a linker should be, leaving open the possibility that other linkers exist that would further increase the induction level, the tightness or the fold-change.

I could create a larger library of linkers, as already been employed to modulate the gene expression regulated by the blue light-repressed histidine kinase YF1[291]. In that case, the linker composition and length were chosen by a sequence prediction software, but the selection can also be performed manually by sequentially adding or removing nucleotides from a starting linker. Another alternative would be to generate a library of random linkers, simple to obtain by using different sets of primers with fixed lengths but random base pair composition, as performed for eLightOn[47]: the best constructs would be then chosen after multiple rounds of selections at the FACS for the bacteria showing the lowest reporter expression in the dark and highest expression after blue light illumination.

The data I obtained led me to imagine the existence of a "sweet spot", concerning the linker length and its flexibility, that would guarantee an induction as similar as possible to that of wild-type AraC; this new library, expanding the one presented in this work, would help to find it, contributing to a better understanding of the role of the linker in the induction process.

3.2.4 Creation of a P$_{BAD}$ promoter library

The binding to I_2, as often said, is essential for RNA polymerase recruitment at the P$_{BAD}$ site. I have also shown the drastic consequences that a point mutation to I_2 can have (Figure 37). A random library can be easily made with primers harbouring random nucleotides in correspondence to the I_2 or even I_1I_2 DNA sequence. The effect of varying the distance between the two half-sites can also be explored. A different DNA sequence could correspond to lower affinity for monomeric BLADE, thus reducing binding in the dark, or conversely could lead to stronger binding, as I showed already with the mutant binding site in Figure 37, that resulted in a reporter gene expression exceeding that obtained by AraC with wild-type P$_{BAD}$. The selection of the mutants would be performed based on the *E. coli* fluorescence level measured with the FACS as explained in the previous paragraph.

3.2.5 Effect of position and orientation of the genes in pBLADE

When I initially constructed pBLADE, I cloned J23101*/**-BLADE and P$_{BAD}$-mCherry in divergent orientation, that is, with the transcription of both genes that proceeds in opposite directions, away from each other, as a way to avoid transcriptional interference via terminator leakage[292] (Figure 13). I avoided to clone them in a convergent orientation, with the gene transcription happening in opposite directions but toward each other: this disposition would, in theory, suffer from abortive transcription events or transcriptional stalling from both genes, due to the polymerase leak through the terminators.

Tandem orientation, where gene transcription proceeds in the same direction for both genes, would cause the transcription to start on the downstream gene for the same reason.

However, Yeung and colleagues analysed the effects of intergenic compositional context, that is, the spatial arrangement and orientation of the genes on the DNA, indicating a role of supercoiling in the expression of the two genes that comprised their construct[293]. They theorized that the supercoiling propagated by proximal promoters generates expression interference: they demonstrated that, with the two genes arranged in opposite directions, induction of one gene caused leaky expression of the other. Moreover, the induction level was significantly lower compared to the same genes placed in convergent orientation, with the tandem one showing intermediate levels[293]. The supercoiling contribution was confirmed with supplementation of gyrase *in vitro*, that mitigated the expression differences between the two genes cloned in the divergent and tandem orientation.

An interesting experiment would, therefore, be to clone BLADE and mCherry in pBLADE with divergent and tandem orientation. This could potentially allow finding the optimal orientation leading to higher induction levels of P_{BAD}; nonetheless, it might well be that the leakiness would be higher too, as I have demonstrated how sensitive the system is to the concentration of BLADE in the cell. This will probably require some optimization, but I believe that the contribution of the intergenic compositional context deserves to be investigated.

3.3 Light overcomes the limitations of L-arabinose

In paragraph 2.1.8 I have compared the light-inducible transcription factor BLADE to its L-arabinose-inducible counterpart AraC. Despite reaching lower induction levels, at least with the wild-type P_{BAD} promoter (Figure 15), BLADE might be a better choice in specific experiments, for example where spatiotemporal precision is required.

3.3.1 Light is not catabolized by bacteria

As seen in Figure 22, L-arabinose is quickly catabolized by the cells. Light, on the other side, provided a constant stimulus, with no catabolism nor accumulation within the cell (Figure 24), allowing the induction to be finely titrated over time.

Inducer catabolism is not a general feature of all chemically inducible systems, as not all small molecules are catabolized (*e.g.*, IPTG) and special strains engineered not to catabolize L-arabinose exist[130,260]. I used one such strain that does not catabolize L-arabinose, and I indeed observed a higher expression level of the reporter gene than in the wild-type strain (Figure 25). However, in some cases, researchers may prefer or need to use a specific strain, that is already engineered harbouring a deletion of another gene of interest, and further introduction of deletions or mutations may be problematic and time-consuming.

Another way to overcome the problem of L-arabinose depletion is to use a microfluidic device that supplements the culture with a constant flux of L-arabinose, compensating the catabolized sugar over time. While this would allow for the usage of a wild-type strain, beside the need to create an assay *ad hoc*, the constant supplementation of L-arabinose could alter the growth rate, as the bacterium would eventually convert this sugar in energy. Therefore, the comparison between uninduced and induced samples would become problematic.

3.3.2 BLADE enables complete reversibility of reporter gene expression

Light can be easily switched on and off; I showed that BLADE allows for fully reversible phenotype control, with no influence from previous inductions. I have validated this both for the induction of a fluorescent protein (Figure 26a) and for functional genes important for cell morphology (Figure 48). In contrast to the fast and complete reversibility achievable with light, the chemical inducer L-arabinose, even after being washed off twice, remained inside the cells, which kept on transcribing the *mcherry* gene during the resting phase (*i.e.,* the phase of no induction) (Figure 26b-d). When using a statistical test to compare the fluorescence levels of the samples incubated with L-arabinose after the second and third induction cycles to those measured after the first cycle, I found no significant difference, probably due to the uncertainty of the measurements (Figure 26b-d). However, an increasing trend is visible, especially at low L-arabinose concentration (Figure 26d). I believe that, with larger sample size, the standard deviation would decrease, and the fluorescence levels measured after the second and third cycles would likely become significantly higher than those after the first induction cycle: as the retention of L-arabinose after the wash steps led to a residual transcriptional activity, providing a "memory" of the past inductions (**Figure 26**e), for the same reason the fluorescence values should be expected to be higher after every induction cycle.

This memory of previous L-arabinose inductions was observed also in the experiment with the KC717 strain harbouring P_{BAD} upstream of the *rodZ* gene in the genome, while light-induced BLADE-mediated transcription led to the recovery of the rod-shaped morphology (Figure 48). Despite two centrifugation and resuspension steps to remove L-arabinose from the medium, the residual intracellular L-arabinose continued to induce AraC-mediated transcription of *rodZ*, making the cells even more rod-shaped.

BLADE, not suffering from memory of past inductions, also has the advantage of saving time, reducing sample handling and cellular stress, because no centrifugations to remove the inducer are needed.

Reversibility would be even more problematic in experiments involving strains that do not catabolize L-arabinose since these would not be able to utilize residual molecules of the inducer after the washing steps as an energy source. For the same reason, using such strains to perform the cycles of induction-recovery with L-arabinose would likely lead to a higher level of fluorescence during the recovery phase than that observed in Figure 26b-d.

3.3.3 BLADE leads to homogeneous activation of gene expression in the cell population at every light intensity

Another key feature of BLADE is that it leads to a homogenous response in a cell population, in contrast to the heterogeneous activation of P_{BAD} at low L-arabinose concentrations (**Figure 27** and **Figure 28**). At every blue light intensity tested, the CV of the fluorescence distributions of cultures transformed with pBLADE remained comparable or lower than that of the distribution of the cultures transformed with pBAD33 and induced with the highest L-arabinose concentration. The exceptionally low variability measured for the culture illuminated with 2 W/m^2 blue light was probably obtained for stochastic reasons: by increasing the sample size, the CV would likely align with the values measured for the other light intensities. The possibility of light being shielded due to culture turbidity (at the end of the illumination time the density of the culture reaches OD$_{600}$ = ~ 2.0) did not have an impact on the final induction level, even when low light intensity was employed.

As mentioned in paragraph 1.4.2, all-or-none induction is an unwanted feature when inducing AraC-mediated transcription with L-arabinose, causing a heterogeneous expression level within the induced population. In uninduced cells, the stochastic background expression of the L-arabinose regulon leads to a wide distribution of proteins for the uptake of the inducer. The supplementation of subsaturating concentrations of L-arabinose would then lead to different rates of L-arabinose accumulation, causing heterogeneous timing of gene induction within the population[119]. Not relying on transporters to be internalized, light does not suffer from this problem (Figure 27 and 28). I encountered the all-or-none response mainly when I supplemented culture transformed with pBAD33 with 0.0001% L-arabinose every hour (Figure 30). With a single L-arabinose supplementation at the beginning of the induction, such low concentration was quickly catabolized (Figure 23); the broad curve, given the low fluorescence output after 4 hours of induction, partly overlapped with the uninduced sample curve (Figure 28d), possibly encompassing the L-arabinose-responsive and non-responsive populations. The constant presence of L-arabinose at a low concentration, on the other side, might have led to the bistable response (Figure 30).

This heterogeneity can be overcome with a higher L-arabinose flux inside the cell, obtained through overexpression of either AraE[141], the L-arabinose transporter, or LacY[142], a transporter with relaxed specificity. However, this requires the use of engineered strains or co-transformation with a plasmid encoding the transporter. The first solution, as previously discussed, is impractical in many experimental setups, while the second could cause compatibility problems in case other plasmids need to be co-transformed.

3.3.4 BLADE is not repressed by glucose

The binding site for the protein CAP is located between the O_1O_2 and I_1I_2 DNA sites (see 1.4.2). This protein is bound by cAMP in the absence of glucose and is responsible for the opening of the DNA loop created by AraC when it is bound to L-arabinose. Vice versa, in the presence of glucose, even with L-arabinose in the medium, it maintains the loop in position[131,294]. BLADE, as opposed to what happens with AraC, is not sensitive to the glucose inhibition: this means that *E. coli* cells expressing BLADE can be grown in culture media that contain glucose, a more convenient choice than media with complex sugars more difficult to catabolize, guaranteeing fast and long-term growth.

For example, to speed cell growth up during MinDΔ10 overexpression with BLADE and obtain a more evident minicell phenotype, I used LB without the risk of impairing the induction process (Figure 49). The KC717 cells are spherical when incubated without L-arabinose, therefore are fragile, and grow very slowly; in some cases, the culture failed to grow out of the glycerol stock during an overnight growth when the minimal tryptone broth was used. Using LB, I could guarantee the growth of the spherical KC717 cells. In this case, the high L-arabinose concentration used to induce transcription of RodZ (0.2%) allowed the use of LB, but the induction could not have worked properly if lower concentrations were employed.

3.4 *In vitro* analysis of BLADE DNA binding and dimerization

The EMSAs showed that BLADE binds to I_1 and I_1I_2 half-sites identically: what we observed, therefore, is likely only the binding to I_1, performed by the monomer of BLADE. Moreover, it seems to bind to I_1 and I_1I_2 better in the dark than in the lit state (Figure 31 and Figure 33): this is in contradiction with what we expected to see, given the *in vivo* results, that is, better binding in the lit state, when BLADE is in the dimeric form. However, it is unlikely that light induced the degradation of BLADE, as I never observed this behaviour in other experiments.

Contrarily to what happens *in vivo*, given its lower affinity compared to I_1, the condition of the EMSA seemed to have prevented the binding of BLADE to I_2. This led me to believe that the dimeric form of BLADE never formed in this assay.

Moreover, after performing an EMSA with AraC, we observed the same bands (albeit weaker) when AraC was incubated with I_1I_2 and I_1 in the presence of L-arabinose, meaning that the complex formed by AraC bound to both half-sites could not be detected (Figure 32). However, there is solid evidence of the role of I_2 in AraC-mediated induction of P_{BAD}[248,265,295]. Similarly, the EMSAs performed with BLADE incubated with I_1I_2 and I_1 gave very similar results (Figure 32 and Figure 33). These results convinced me that the assay prevented us to visualize the weaker I_2 binding.

We proved that purified BLADE could dimerize using SEC (Figure 35); since the illumination conditions were the same in this experiment and the EMSAs, and given the time required to VVD to revert to the dark monomeric state, it is unlikely that BLADE monomerized during the time of the assay. We also hypothesized that the FAD chromophore could have detached from VVD, preventing the dimerization process, during the sample incubation before the gel run. Supplementation of FAD in the reaction tube, however, did not make any difference in the output.

To be able to see the bands corresponding to a binding event with either only I_1 or I_1I_2, we could change the gel composition and percentage, the running buffer, and in general the running conditions. As a control, we could incubate BLADE and AraC with I_1I_1 and compare the obtained bands with those obtained with I_1 only to see if, at least when the two half-sites are bound with high affinity, the different complexes are detected.

An alternative assay that I could perform to verify the DNA binding is the Chromatin immunoprecipitation (ChIP). To ensure the highest precision in the detection of the binding site (each half-site is 17-bp long and separated by 4 nucleotides) this assay can be coupled with exonucleases (ChIP-exo), already proposed in paragraph 3.1.5 as a way to measure how the binding of BLADE to I_1 and I_1I_2 changes with different BLADE transcription levels. As I need single-nucleotide precision, I will need to run the sample into capillary electrophoresis to sequence the DNA.

3.5 The aggregates formed by BLADE in the dark likely contribute to keeping expression off before activation

Light-inducible transcription factors based on VVD have been so far employed without any fusion to a fluorescent protein (FP), as the scope of the experiment was the quantification of the reporter gene expression[47,171,180,249]. In my work, I also did not need at the beginning to include any FP in the BLADE construct, since visualization of its localization was not important, and the fusion may have not worked as well as the untagged version.

I was surprised to see BLADE localize to aggregates, predominantly in the dark (Figure 42a), even though the formation of aggregates of VVD in the dark was documented with previous *in vitro* analysis[268,296] as the simultaneous unfolding of VVD structural elements might increase their propensity of aggregation, similar to what observed for amyloid fibrils[297,298].

Despite not having analysed VVD alone, I believe that the aggregates reflect a property of VVD, not AraC$_{DBD}$. The evidence of this claim comes from the analysis of VVD^{C108A}, a mutant insensitive to light, which was found in aggregates under blue light illumination to the same extent as in the dark (Figure 42b), and from the results obtained with AraC-sfGFP, which was cytoplasmic, with or without L-arabinose (Figure 43).

As a further control, I could test if the aggregates are always absent by employing VVD^{T69W}, a mutant constantly in the dimeric state [96].

In *N. crassa,* VVD is involved in the regulation of the circadian clock together with FRQ and WC1. Its degradation in the dark is probably triggered by its unfolding, while the dimer formed with blue light is more stable[87,267,268]. I hypothesized that the aggregates found in *E. coli* might have a role similar to degradation, as the aggregated protein would be dysfunctional, with the advantage of a potential release of the protein, a faster process than the *de novo* expression. To confirm this hypothesis, I first observed that their fluorescence did not recover overtime after photobleaching (Figure 44). Then, to investigate the possibility of release of the aggregated proteins with light, I performed time-lapse microscopy and found that the aggregates formed in the dark were not dispersed after illumination; rather, the formation of new ones in newborn cells was prevented (Figure 45). This type of asymmetrical segregation has been previously observed for other types of aggregates as well[299].

The collected data suggest that the foci formed by BLADE in cells kept in the dark are non-functional aggregates incapable of dissolving and thus releasing functional proteins once light is present. Blue light illumination simply decreases the probability of their *de novo* formation. Their contribution to the observed fold-change remains to be established and I can only speculate that the aggregates contribute to the tightness of the system. Nonetheless, since only about half of the cells have aggregates in the dark (Supplementary Figure 5b), I believe that this is not the only reason why BLADE shows minimal activation of the P$_{BAD}$ promoter in the dark. The main mechanism remains the occupancy of the I$_2$ half-site, which occurs only following dimer formation upon blue light exposure.

3.6 The advantages of light in controlling the expression of functional genes

3.6.1 Minicell production with light could have clinical importance

I proved that BLADE is well suited to express the *minDΔ10* gene, whose product is a truncated form of MinD, a fundamental component of the cell division machinery. I demonstrated that the overexpression of this mutant protein in *E. coli* is able to trigger the appearance of the minicell phenotype, in line with previous results obtained by Linda Klauss, a former PhD student in our lab. When I regulated the expression of MinDΔ10 with BLADE, I found barely any minicell in the dark state, nor I did see morphological differences between the cells incubated in the dark and the control cells (Figure 49). Among the cells illuminated with blue light, I observed a large distribution of cell lengths, which was not seen in the cells kept the dark, and slightly more minicells than those counted from the culture of ΔMinB strain[300]: this strain, lacking the entire *minB* operon (coding for MinC, MinD and MinE), is normally used for the production of minicells: I counted about 40% of the cells in the population after overnight growth being minicells, while for BLADE driving MinDΔ10 expression the value reached 55% (Figure 50).

This result is important, not only for basic research purposes and to demonstrate the efficacy of BLADE *per se*: minicells are studied since 1967[301,302], but from 2007 the scientific community is working for their use as a delivery vehicle for targeted cancer therapy[303]. After several milestone experiments performed in animal models[280, 281], minicells have reached Phase I/II clinical trials as a new delivery vehicle for cancer therapy[282, 283]. The absence of chromosomal DNA prevents their replication, and they can remain metabolically active for up to a few days[308–310]. Thanks to their size (~ 400 nm in diameter), when intravenously administered, they can arrive at the cancer site, and enter through the fenestrations, that are larger than those found in normal tissue. Once there, the minicells can deliver their chemotherapeutic cargo directly on the target site if active binding to the cancer cells can be guaranteed by recognition mechanisms like antibodies against epidermal growth factor (EGF) receptor (EGFR) coated on the minicell surface[303,311].

The minicells generated from the Δ*minB* strain have a variable age, being produced during the entire overnight period, potentially leading to non-homogeneous quality between newer and older minicells. Using light to trigger minicell formation overcomes this problem since the minicells are formed during a much narrower time window. As the minicells have the same age, they would be all equally metabolically active: this would allow them to express genes part of synthetic circuits to selectively detect and kill only tumour cells. Such circuits have already been developed by synthetic biologists in the past[312–314].

I used the BLADE FP4** variant to induce MinDΔ10. Using a BLADE construct that reaches higher induction levels (*e.g.*, BLADE FP6**, BLADE_I$_2$* FP6*), or simply increasing the induction time, I could produce even more minicells. As already said, however, it is important to use a construct and a protocol that do not lead to the formation of minicells in the dark, as this would abrogate the advantage of using an inducible system to obtain minicells of similar age.

3.6.2 Comparison of chemically- and light-induced MreB and RodZ induction

We have seen that, as for MinDΔ10, overexpression of MreB and RodZ is tightly regulated by BLADE, with the phenotype in the dark and the control sample being indistinguishable. These two proteins are involved in rod shape maintenance in non-spherical organisms: while their absence causes spherical morphology[213,241] (Figure 48), their overexpression leads to cell elongation and widening, with higher size variability than the controls in the dark[244,247] (Figure 50).

MreB and RodZ overexpressed with BLADE led both to 1.5-fold increase in the mean length compared to the control, and 1.2 and 1.3-fold increase in the width, respectively (Figure 50). These phenotypes are less severe than those reported in the literature: MreB overexpression led to filamentous cells, with thickness 1.7-fold greater than the control sample, while RodZ overexpression doubled the length and increased the thickness of 1.3-fold[239,243]. The induction in those experiments was performed with IPTG-regulated promoters: I believe that I could induce the phenotype to the same extent described in the literature if I used a longer induction time or a different BLADE construct with a higher induction level (BLADE FP6** or BLADE_I_2* FP6*). As for MinDΔ10 overexpression, the wild-type morphology in the dark with those constructs could be maintained if the leakiness remains below the threshold necessary to alter the phenotype.

3.6.3 The use of BLADE in a medium-throughput library screening

Despite the increasing power of sequencing and bioinformatics technologies to rapidly collect and analyze data, the definitive functional characterization of unknown genes still requires genetic and functional analysis in case-by-case studies. This often involves the deletion or overexpression of target genes and phenotypic characterization of the cells.

3.6.3.1 BLADE allows easier and more convenient screening than AraC

To showcase the utility of light induction in medium to high-throughput studies, I used BLADE to overexpress 39 genes randomly selected from those with an unknown or poorly defined function. With the technical help of Mehmet, we applied four bioinformatics tools to first predict their function and localization. While it was possible to find a consensus prediction for 53% of the genes in case of localization, only for 35% of the genes we found a consensus for function prediction; even in this case, the prediction remained relatively vague (DNA or membrane binder, transporter, ligase, etc.; see Supplementary Table 4). This highlights how computer-based predictions cannot yet replace detailed biochemical characterization, which remains essential to assign a function to a protein.

This characterization could also have been performed with the pBAD33 plasmid and L-arabinose induction: in that case, to induce each sample I should have dispensed the desired L-arabinose concentration in the plate with a multichannel pipette, or, if it had been available, by using a pipetting robot. I was helped by our technical assistant Emir because, even with a library of *only* 117 constructs, there were more than a thousand samples in total to handle, if we consider all the preliminary experiments to set up the induction library protocol, the replicates and the growth curve measurements that we performed. The use of an optogenetic tool has been advantageous for us to speed the workload up and, at the same time, reduce the chances of human error, like cross-contamination and pipetting inaccuracy, avoiding the need for expensive automated pipetting setups.

Among the experiments performed for the characterization, the biggest advantage of a light-inducible system was shown in the 20-hour induction in the plate reader. Without a strain unable to catabolize

L-arabinose or a microfluidic system, the illumination with blue light was the simplest way to maintain the induction level constant over time.

3.6.3.2 The importance of pulsatile induction for gene characterization

Several regulatory factors have been observed undergoing repetitive activation pulses. Each of those activates and deactivates the regulator, through changes in its localization, modification state or concentration, on time scales ranging from minutes to hours[315–318].

Pulsatile expression of target genes involves different types of proteins, from alternative bacterial sigma factors to mammalian tumour suppressors like p53. This regulatory dynamic has been shown to regulate several processes, like signalling, differentiation and stress response[319]. For example, in *Bacillus subtilis*, pulsing is involved in at least three systems: sporulation initiation, which controls the transformation of cells into dormant spores[320], the general stress response pathway[315] and genetic competence, which allows cells to take up DNA[321]. Likewise, pulsing in yeast has been observed in pathways related to stress response, regulated by the transcription factors Msn2/4 and Crz1[316,322]. *E. coli* relies on the pulsatile dynamic as well: for instance, the *gadX* gene, encoding a transcription factor that regulates ~34 genes in the acid resistance system, exhibits high amplitude and duration pulses that persist well beyond the cell cycle[323]. Conversely, RecA, which plays a central role in SOS response and homologous recombination processes, shows large amplitude pulses, but with shorter durations[323].

Pulsatile dynamics in gene expression is also an important mechanism that cells utilize to transiently resist stress. This dynamic behaviour is spread across a broad range of stress response genes, including oxidative stress, multi-drug resistance, heat-shock response and can enable subpopulations of cells to withstand a temporary change in environmental conditions.

It would be interesting to investigate how the information encoded in transcription factor dynamics is transmitted to downstream genes, and the technologies that enable the manipulation of gene expression over time could help to identify how these signals are propagated[324,325].

Optogenetics is perhaps the most versatile technique now available to study and reproduce pulsatile activation dynamics. A pulsatile induction of the library of unknown genes with BLADE could help to better elucidate the behaviours of those genes whose activation requires a specific induction dynamic, rather than a continuous stimulus. BLADE, in addition to the frequency modulation, could also be used to tune the amplitude of the signal delivered to the culture, by a simple change of the light intensity. The induction protocol could be loaded on a Raspberry PI computer and connected to the power supply for an automated, cheap, and easy-to-use illumination setup.

Given the long dark state reversion of VVD, the pulse duration cannot be too short, preventing the application of dynamic pulses shorter than 3 hours. As the division time of *E. coli* is much shorter, mutations such as the I74V and I85V ones could be introduced into VVD to shorten the reversion time up to 600-fold, which is comparable with those rates observed for the fastest cycling phototropins[181].

The first example of an optogenetic protein used to generate pulsatile dynamics is VP-EL222, showing the possibility to tune induction variability, in *S. cerevisiae*[326], through pulse width modulation. Pulsatile induction was also shown to enhance the biosynthesis of products in engineered *S. cerevisiae* transformed with VP-EL222: this enables the possibility of new types of bioreactors for industrial application where pulsatile activation is used by optogenetic proteins to increase fermentation yield[50].

3.7 BLADE makes pre-existing L-arabinose-sensitive plasmids and strains light-responsive

We have engineered BLADE out of the natural transcription factor AraC, used worldwide among bacteriologists and synthetic biologists to induce gene transcription. The exclusive feature of BLADE is its compatibility with previously constructed strains, carrying the P_{BAD} promoter at an endogenous locus to drive the expression of a gene of interest, and plasmids, bearing the P_{BAD} promoter upstream of the gene of interest. A simple (co-)transformation step of a plasmid expressing BLADE with an ordinary pBAD33 or into a strain of choice allows for the regulation of the gene of interest with light without the need for any cloning.

When the need to use pre-existing plasmids and strains is not there, pBLADE should be used. I have constructed it using pBAD33 as a template (Figure 14a) to ensure compatibility with restriction enzymes previously employed with pBAD33 as well as with other plasmids that should be co-transformed with pBLADE.

Induction of mCherry was slightly lower when this was encoded on pBAD33 (rather than pBLADE) and BLADE was expressed from a separate plasmid; this could be due to the fact BLADE and apo-AraC transcribed from pBAD33 compete for binding to I_1 (**Figure 47** and Supplementary Table 3) or to the fact that the culture was grown in the presence of two antibiotics for the case of the two plasmids. Our laboratory recently showed that reducing the number of antibiotics from two to one can increase protein production[271]. To assess the contribution of having two antibiotics in the culture medium instead of one, I took two approaches: I simply added a second, empty plasmid to the cells transformed with pBLADE, which allowed me to grow both cultures with the same two antibiotics, and I re-cloned the parts of interest (namely, *blade* driven by J2301* and *mcherry* driven by P_{BAD}) into special plasmids, called SiMPl plasmids, that carry a split resistance cassette. A SiMPl plasmid pair can be used with a single antibiotic[271].

The results excluded that the need for neutralizing two antibiotics instead of one has a negative impact on mCherry production (**Figure 47** and Supplementary Table 3). Therefore, I conclude that competition with plasmid-borne AraC is the only factor altering the induction level.

BLADE can also be used to induce gene expression when P_{BAD} is inserted in the chromosome, providing a more physiological expression level compared to induction from the plasmid. I demonstrated this possibility by inducing genomic *rodZ* expression with an engineered strain[244]. Interestingly, not only I could trigger the same phenotype obtained with L-arabinose induction, but I also recovered the initial phenotype after only 2 hours of incubation in the dark (Figure 48b).

In 2005, an integration approach for *E. coli* has been developed, combining the lambda-red linear DNA recombination method with the site-directed insertion of a repression and expression cassette (RExBAD), to allow placing the P_{BAD} promoter upstream of any target gene in the genome[246]. P_{BAD} was chosen for its low leakiness, and indeed it allowed the researchers to create phenotypic knockouts, functionally equivalent to gene deletions. By combining this strategy with the transformation of the cells with pBLADE it would be possible to use light for functional analysis of any chromosomal gene. This can also include essential (like *ftsL*) or toxic genes (like *ccdb*), as demonstrated in the paper. I could not regulate the expression of *ccdB* with BLADE due to the intrinsic P_{BAD} leakiness that forced the cells to inhibit ccdB expression; in this case, however, being

P$_{BAD}$ on the chromosome, it would be present only in one copy, thus the leakiness level would be lower, allowing the cell to grow in the uninduced state.

The authors of the study point out that one application of RExBAD could be to specifically characterize genes with unknown or poorly characterized functions with high-throughput phenotypic analysis. As shown in this work, the use of BLADE in these types of applications can speed this process up and reduce the handling of the samples.

3.8 New BLADE characterizations and applications

As previously mentioned, we collaborated with Armin Baumschlager in the laboratory of Prof. Khammash at the ETH Zurich. Armin expanded the characterization of BLADE presented at the beginning of this work and provided an application in controlling cell growth on L-arabinose with blue light. Below I will give a brief overview of his work and compare his findings with mine.

3.8.1 Expanding the family of BLADE TFs

I spent a great amount of time engineering BLADE to optimize its performance, especially its leakiness in the dark. However, there are still ways to further improve it. Our collaborators have expanded the BLADE library by playing with different parameters. First, they inserted new linkers between VVD and AraC$_{DBD}$, then varied the position of AraC$_{DBD}$ within the fusion testing both termini, and tested a different photosensor namely the LOV domain of *Vaucheria frigida* Aureochrome1 (*Vf*Au1)[327,328], which is naturally found C-terminally to a bZip$_{DBD}$[328], and which, like VVD, homodimerizes upon blue light stimulation[327,329]. Lastly, they tested a wide range of BLADE expression levels to find the best synthetic promoter to use to obtain the highest fold-change in reporter gene expression.

They found out that placing AraC$_{DBD}$ at the C-terminus, as it is found in the library we constructed and in the native AraC protein, led to higher fold-changes for VVD-based constructs, mainly due to lower mCherry expression in the dark. Conversely, with *Vf*Au1, the opposite was true. When they varied the linker lengths, they started from ten amino acids from the endogenous disordered region (AraC 168-291), the same AraC fragment used for the BLADE constructs from FP4 to FP7 of our library (Figure 13a), and added new synthetic linkers to it. They found out that a total linker length ranging from 10 (equivalent in length to FP6) to 17 amino acids (equivalent in length to FP3) provided the highest fold-change for both photosensors, with no significant difference of induction within this range[250].

This range of optimal linker lengths is the same I found in the characterization that I performed, even though the linker composition partly differs from ours (Figure 13a). Similarly to what I observed with BLADE FP5 (Figure 13a and Figure 15), when Armin employed longer linkers (G$_6$A$_3$, G$_{20}$S$_5$) he obtained low induction of the reporter and low fold-change levels[250]. These independent analyses added solidity and reproducibility to the library characterization results that I obtained.

In the second part of his characterization, Armin induced the transcription of the *blade* coding sequence with an IPTG-inducible promoter[287] to achieve various levels of expression of BLADE, intending to find the expression level corresponding to the highest dark/light fold-change in reporter

expression. Then, he mapped the IPTG dose-response levels to a large set of constitutive promoters to show that, once the appropriate promoter strength is found, this can be obtained with a constitutive promoter dispensing of IPTG. Armin found out that, with increasing BLADE levels, the reporter expression levels in the dark and lit states increased with a sigmoidal trend, before reaching a plateau. Vice versa, by decreasing BLADE levels, the induction of the reporter in the dark and lit states reached the same basal level (Figure 58). The highest fold-change was reached at intermediate BLADE expression levels, by using a promoter with intermediate strength among those used.

These results expanded the library presented in this thesis, obtained with the two promoters J23101* and J23101** (Figure 15), confirming that BLADE concentration in the cell is crucial for the modulation of reporter gene transcription. This also revealed that if I had used much weaker or much stronger promoters, I would have obtained a lower fold-change, either with low or high induction levels.
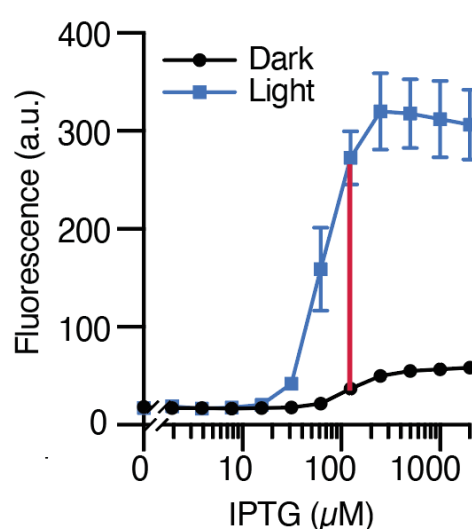
Figure 58 | IPTG dose-response curves obtained with cells transformed BLADE. The highest light/dark fold-change is indicated with a red line. Adapted from[250].

3.8.2 Making the endogenous L-arabinose metabolic pathway light-inducible

As mentioned many times, BLADE is based on the endogenous AraC protein, a TF that controls the expression of the genes necessary for the utilization of the sugar L-arabinose. AraC itself, as well as the promoters it regulates, is involved in complex feedback regulation schemes that control the expression of the genes *araB, araA* and *araD* necessary to catalyse the transformation of L-arabinose into D-xylulose-phosphate, which is then further processed via the pentose phosphate pathway[123] that ultimately leads to the creation of energy for the cell[125,330].

Armin used BLADE in a strain deprived of *araC* to activate transcription of the different endogenous genes necessary for transport and metabolism of L-arabinose, with the aim to induce growth when cells are grown with light but not in the dark. He incubated the cultures either in the dark or under saturating blue light in M9 medium containing trace amounts of amino acids and either 0 or 2% L-arabinose. With the medium containing 2% L-arabinose, activation of BLADE led to elevated growth in light-induced samples in comparison to the samples kept in the dark, that showed minimal growth, with a 10-fold light/dark difference when AraC$_{DBD}$ was placed at the N-terminus and 24-fold when

at the C-terminus within BLADE[250]. These results confirm the tightness and effectiveness of BLADE when involved in the regulation of biological processes.

In the next set of experiments, Armin focused on the tunability features of BLADE. Wild-type *E. coli* cells cultured in a medium containing L-arabinose grow at a certain fixed growth rate that depends on the concentration of the sugar. By using BLADE and light as an input, he showed that the growth can be modulated at will to any intermediate level through modulation of the light intensity at a fixed L-arabinose concentration. He induced the transcription of the *blade* coding sequence at four different levels and, with 2% L-arabinose in the medium as an energy source, he grew cells in the dark, or under different light intensities, obtaining intermediate cell densities only for illuminated cells.

This experiment demonstrated the fine level of tunability obtainable with BLADE, confirming our experiments performed at different light intensities (Figure 16, Figure 24). Moreover, the cell density obtained at the different light intensities plateaued at about 2 W/m2: these data are in agreement with our measurement of mCherry expression levels in Figure 16.

3.8.3 Expression of BLADE in other organisms

I always used *E. coli* as a model organism, but AraC-mediated induction has also been used in other organisms.

Pseudomonas researchers have used the inducible properties of the AraC regulator and the P$_{BAD}$ promoter cassette for inducible gene expression by integrating the AraC and P$_{BAD}$ into the chromosome by using a suicide vector or an integration-proficient one[331–334].

AraC is highly conserved across the *Enterobacteriaceae* family, which includes *E. coli* and *S. enterica*. The two helix-turn-helix of AraC$_{DBD}$ are 100% identical between these two organisms[335]; hence, AraC likely binds with similar DNA sequence specificity across all *Enterobacteriaceae* species, with the AraC binding sites *araBAD*, *araFGH*, *araE*, *ytfQ*, and *araTU* being conserved as well[335]. Given the property of *S. enterica* to spontaneously accumulate and proliferate in tumors[336,337], in 2014 an attenuated *S. enterica* strain, with the *ara* operon deleted to avoid L-arabinose catabolism, was used for tumour-specific expression of antitumor drugs[338]. The cells were transformed with a plasmid containing cytolysin, a cytotoxic gene capable of killing tumour cells, under the control of the P$_{BAD}$ promoter and intravenously injected in mice. Arabinose induction led to cell death specifically at the tumour site, demonstrating the promoter tightness also in a mammalian organism.

Acetic acid bacteria (AAB) oxidize a wide variety of carbohydrates and are therefore used industrially for oxidative biotransformations. However, so far the choice for low-cost, tight, and strongly inducible expression system is limited to the L-arabinose-inducible P$_{BAD}$ promoter; for instance, the P$_{BAD}$-AraC system has been utilized in the acetic acid bacterium *Gluconobacter oxydans*[339]. This organism uptakes L-arabinose with low efficiency into the cell, therefore a high concentration of L-arabinose is needed (1%). Moreover, some *G. oxydans* strains oxidize L-arabinose in L-arabinonic acid: this leads to a strong acidification of the medium and deep perturbations of the intracellular environment, causing upregulation and downregulation of genes.

I believe that BLADE would represent the ideal alternative in such cases. In particular, light would have the advantage of avoiding medium acidification, not only for *G. oxydans* but potentially for all the acetic acid bacteria. Reports in the literature of proteins bound to FAD from *Pseudomonas*,

Salmonella and various AAB demonstrate that the chromophore is produced by the aforementioned organisms[340–343], therefore its supplementation is not necessary when employing BLADE.

However, some parameters must be controlled to allow the expression of BLADE in organisms that are different from *E. coli*. First, the component of pBLADE, such as the J23101 promoter, the resistance gene and the origin of replication must be compatible with the host organism. If not, BLADE must be cloned in a different plasmid. Then, as for any heterologous protein expression, the codon usage of BLADE must be checked first, and in case optimized to guarantee mRNA stability and an optimal level of protein expression for the host organism. Lastly, there is the possibility of an incorrect expression, that could happen at any step, from the transcription to the post-translational modification. Performing microscopy with a fluorophore-tagged BLADE or a western blot with antibodies against VVD is therefore suggested.

3.9   Conclusion and outlook

BLADE is a single-component and single-plasmid optogenetic transcription factor that does not rely on external cofactor supplementation. I previously mentioned that other bacterial optogenetic transcription factors exist; however, BLADE possesses a series of unique features, which I tried to highlight in this thesis.

When I started this project, no other bacterial transcription factor was converted from chemical- to light-inducible yet, but in 2020, a system for the optogenetic regulation of the *lac* operon, called OptoLac, was published (see paragraph 1.5.1.1). However, there are differences between that system and BLADE. In OptoLac, the *lacI* repressor was not engineered to be light-sensitive, but its expression was put under light control using the optogenetic transcription system pDawn (see 1.5.1.1). The drawback of OptoLac, other than a higher complexity (it relies on four proteins to work: cI, LacI, and the dimerizable pair YF1-FixJ), is its delayed ON/OFF switch, due to the time required for the accumulation and degradation of the cI repressor[296]. Under certain circumstances, when fast dynamics are needed, this feature may become a limitation. Contrary to OptoLAC, BLADE responds quickly to a light input through dimerization: it does not require additional circuitry for light inducibility, which is achieved more directly and with less complexity.

3.9.1 Future modifications to BLADE

As BLADE is a modular protein, having the DD and DBD physically distinct, the photosensor can be easily exchanged with others, as our collaborator Armin Baumschlager did by replacing VVD with *Vf*Au1[250]. This could allow gaining new features. For example, to obtain light-inducible repression of the transcription it would be sufficient to fuse a protein that dimerizes in the darkness, like the LOV domain from *R. sphaeroides* used for the construction of eLighON[47].

At the moment, single-component proteins that dimerize with green or red light have not been discovered yet, therefore a version of BLADE sensitive to red-shifted wavelengths, harbouring a different photosensor, can not easily be created.

3.9.2 Future characterization of some unknown genes

The genes with poor or unknown function have been characterized only superficially; for some, however, I could reveal a likely role in cell growth, and for others in cell morphology. Through sfGFP fusion, I could study the intracellular localization of the protein products and, for some of them, I could infer that either the N- or C-terminus is involved in the localization or the correct folding (Figure 54). Of course, there are more *in vitro* experiments to perform to continue the characterization of some of the most promising genes, analyzed so far only *in vivo*; as this would drift from the main goal of the thesis, that is, the characterization of BLADE, I did not continue in this direction.

In the future, however, it would be interesting to better elucidate the function of some of the genes: for example, I could compare with microarray analysis the levels of the genes expressed by a wild-type strain of *E. coli* with a strain having the gene overexpressed with BLADE and a knockout strain (if available). This assay would allow me to identify the genes in *E. coli* that are influenced by the levels of the examined unknown gene.

The characterization can be further extended by studying the response to stress conditions, like different osmotic pressure or temperature, of the cells with the genes overexpressed.

Below, I describe more experiments that I could perform on three proteins of the library.

3.9.2.1 YdaT

ydaT is one of the few genes that I included in the library for which some prior information was available. Its protein product, YdaT, is a toxin that, when overexpressed, leads to cell elongation by impairing cell division[275]. In scientific literature, nothing is reported about its ability to bind to the DNA[275]. Interestingly, I observed the co-localization of YdaT with the nucleoid (Figure 55). It would be interesting to perform EMSA with the purified protein and incubate it with different DNA templates (*e.g.,* GC-rich, AT-rich templates) to gain some information on which type of sequence the protein is more prone to bind to. Even more conclusive would be to perform Chromatin Immunoprecipitation-sequencing (ChIP-seq) to detect the DNA binding sequence bound *in vivo* by YdaT. As the ΔydaT strain is not present in the Keio collection, the strain repository that includes mutants for all of the nonessential genes in *E. coli*[346], presumably the gene is essential for the survival of the bacterium.

3.9.2.2 YahC

YahC is one of the proteins that I found associated with the cytoplasmic membrane and whose overexpression lead to increased cell growth (Figure 52 and Figure 55). To examine protein-lipid binding, I could perform liposome co-sedimentation assays with purified YahC. To identify the type of lipids with which YahC interacts and its binding affinity towards them, I could establish a protein-lipid overlay (PLO), a technique widely used to study lipid-protein interactions[347–349]: serial dilutions of different lipids are spotted onto a membrane; then, the membrane is incubated with the purified protein, expressed with an epitope tag. Lastly, the binding affinity to the different phospholipids at the different concentrations is detected via immunoblotting with an antibody recognising the epitope tag.

Alternatively, to have a quantitative analysis of the lipids that interact better with YahC, I could perform the co-sedimentation assay with different batches of liposomes having a different lipid

composition. The higher the concentration of the sedimented protein, the better the binding with the membrane lipids will be.

Finally, by coupling pull-down with mass spectrometry[350], I could identify the lipids bound by YahC *in vivo*. To find the binding partners, I could perform co-immunoprecipitation (Co-IP)

It would also be interesting to observe the phenotype and the growth rate of the *yahC* mutant strain from the Keio collection to spot any abnormalities.

3.9.2.3 YebY

YebY is another interesting protein: it seems to increase the rate of cell growth (Figure 52) and is found in aggregates when sfGFP is C-terminally fused to it (Figure 56). sfGFP-YebY is cytoplasmic and appears to be expressed at low levels (Figure 55b and Supplementary Figure 5). The presence of swellings on the bacterial surface was not due to the fusion to sfGFP and the resulting aggregation, as they also appeared after the overexpression of native YebY: in this case, they were visible through transmitted light microscopy examination (Supplementary Figure 8).

I speculate the clusters are linked to the function of the protein. As a matter of fact, they are very dynamic, as assessed by FRAP (Figure 56). I hypothesize that these clusters may be liquid droplets, that is, membraneless compartments formed by liquid-liquid phase separation (LLPS)[351–354]. The ability of liquid droplets to recruit soluble proteins provides an example for future applications in metabolic engineering.

So far, researchers have typically used two approaches to concentrate sets of enzymes in a metabolic pathway: the first is to recruit multiple enzymes to a scaffold, which is present throughout a cell[355,356], and the second is the localization to endogenous organelles, such as mitochondria or peroxisomes[357–359]. Synthetic membraneless organelles are therefore a solution to expand the toolkit for locally concentrating proteins. To discover that YebY is directly responsible for the creation of droplets, I could express the protein *in vitro*, and then identify which part of the protein is involved in the formation of droplets.

To proceed with the characterization, I could look for a proportion between droplet formation and YebY concentration, and if the fluorescence of YebY-sfGFP could recover after photobleaching *in vitro*. I could then explore whether the cellular membraneless compartments could be targeted with cargo proteins, by fusing the entire sequence of YebY, or just the sequence responsible for the formation of liquid droplets, to a protein of interest, as already reported in literature[351,360]. It would also be interesting to understand from the secondary structure composition the causes that trigger the LLPS, following a similar approach as used for the spider silk protein I16[351].

4. Bibliography

1.   Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).

2.   Monod, J. & Jacob, F. General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 389–401 (1961).

3.   Jacob, F. & Monod, J. On the Regulation of Gene Activity. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 193–211 (1961).

4.   Cohen, S. N., Chang, A. C. Y., Boyer, H. W. & Helling, R. B. Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3240–3244 (1973).

5.   Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).

6.   Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).

7.   Becskel, A. & Serrano, L. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).

8.   Becskei, A., Séraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).

9.   Gibson, D. G. *et al.* Creation of a bacterial cell controlled by a chemically synthesized genome. *Science (80-.).* **329**, 52–56 (2010).

10.   Annaluru, N. *et al.* Total synthesis of a functional designer eukaryotic chromosome. *Science (80-.).* **344**, 55–58 (2014).

11.   Hutchison, C. A. *et al.* Design and synthesis of a minimal bacterial genome. *Science (80-.).* **351**, (2016).

12.   Yu, K., Liu, C., Kim, B. G. & Lee, D. Y. Synthetic fusion protein design and applications. *Biotechnol. Adv.* **33**, 155–164 (2015).

13.   Jagadevan, S. *et al.* Recent developments in synthetic biology and metabolic engineering in microalgae towards biofuel production. *Biotechnol. Biofuels* **11**, 1–21 (2018).

14.   Widmaier, D. M. *et al.* Engineering the Salmonella type III secretion system to export spider silk monomers. *Mol. Syst. Biol.* **5**, 1–9 (2009).

15.   Khalil, A. S. S. & Collins, J. J. J. Synthetic biology: applications come of age. *Nat. Rev. Genet.* **11**, 367–379 (2010).

16.   Konig, H., Frank, D., Heil, R. & Coenen, C. Synthetic Genomics and Synthetic Biology Applications Between Hopes and Concerns. *Curr. Genomics* **14**, 11–24 (2013).

17.   Deisseroth, K. Optogenetics. *Nat. Methods* **8**, 26–9 (2011).

18.   Shimizu-Sato, S., Huq, E., Tepperman, J. M. & Quail, P. H. A light-switchable gene promoter system. *Nat. Biotechnol.* **20**, 1041–4 (2002).

19.   Deubner, J., Coulon, P. & Diester, I. Optogenetic approaches to study the mammalian brain. *Curr. Opin. Struct. Biol.* **57**, 157–163 (2019).

20.   Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl. Acad. Sci.* **100**, 13940–13945 (2003).

21.   Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* **8**, 1263–8 (2005).

22.   Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. & De Lecea, L. Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* **450**, 420–424 (2007).

23.   Aravanis, A. M. *et al.* An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J. Neural Eng.* **4**, (2007).

24.   Arenkiel, B. R. *et al.* In Vivo Light-Induced Activation of Neural Circuitry in Transgenic Mice Expressing Channelrhodopsin-2. *Neuron* **54**, 205–218 (2007).

25.   Davis, N. J. Transcranial stimulation of the developing brain: A plea for extreme caution. *Front. Hum. Neurosci.* **8**, 8–11 (2014).

26.   Zhang, F., Lin, Y. A., Kannan, S. & Kannan, R. M. Targeting specific cells in the brain with nanomedicines for CNS therapies. *J. Control. Release* **240**, 212–226 (2016).

27.   Chow, B. Y. & Boyden, E. S. Physiology. Synthetic physiology. *Science* **332**, 1508–9 (2011).

28.   Bennett, M. R. *et al.* Metabolic gene regulation in a dynamically changing environment. *Nature* **454**, 1119–1122 (2009).

29.   Santos-Moreno, J. & Schaerli, Y. Using Synthetic Biology to Engineer Spatial Patterns. *Adv. Biosyst.* **3**, 1–15 (2019).

30.   Shemesh, O. A., Tanese, D., Zampini, V., Linghu, C. & Author, N. N. Temporally precise single-cell resolution optogenetics HHS Public Access Author manuscript. *Nat Neurosci* **20**, 1796–1806 (2017).

31.   Xie, J., Nair, A. & Hermiston, T. W. A comparative study examining the cytotoxicity of inducible gene expression system ligands in different cell types. *Toxicol. Vitr.* **22**, 261–266 (2008).

32.   Engesser, R., Timmer, J., Zurbriggen, M. D. & Weber, W. Orthogonal Optogenetic Triple-Gene Control in Mammalian Cells. (2014). doi:dx.doi.org/10.1021/sb500305v

33.   Brautaset, T., Lale, R. & Valla, S. Positively regulated bacterial expression systems. *Microb. Biotechnol.* **2**, 15–30 (2009).

34.   Redchuk, T. A. *et al.* Optogenetic regulation of endogenous proteins. *Nat. Commun.* **11**, (2020).

35.   Guglielmi, G., Falk, H. J. & De Renzis, S. Optogenetic Control of Protein Function: From Intracellular Processes to Tissue Morphogenesis. *Trends Cell Biol.* **26**, 864–874 (2016).

36.   Krishnamurthy, V. V. *et al.* Reversible optogenetic control of kinase activity during differentiation and

embryonic development. *Dev.* **143**, 4085–4094 (2016).

37. Hughes, R. M. *et al.* Optogenetic Apoptosis: Light-Triggered Cell Death. *Angew. Chemie Int. Ed.* **54**, 12064–12068 (2015).

38. Smart, A. D. *et al.* Engineering a light-activated caspase-3 for precise ablation of neurons in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E8174–E8183 (2017).

39. Mondal, P. *et al.* Repurposing Protein Degradation for Optogenetic Modulation of Protein Activities. *ACS Synth. Biol.* **8**, 2585–2592 (2019).

40. Baaske, J. *et al.* Dual-controlled optogenetic system for the rapid down-regulation of protein levels in mammalian cells. *Sci. Rep.* **8**, 1–10 (2018).

41. Lo, C. L., Choudhury, S. R., Irudayaraj, J. & Zhou, F. C. Epigenetic Editing of Ascl1 Gene in Neural Stem Cells by Optogenetics. *Sci. Rep.* **7**, 1–12 (2017).

42. 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).

43. Ramakrishnan, P. & Tabor, J. J. Repurposing Synechocystis PCC6803 UirS–UirR as a UV-Violet/Green Photoreversible Transcriptional Regulatory Tool in E. coli. *ACS Synth. Biol.* **5**, 733–740 (2016).

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

45. Jayaraman, P. *et al.* Blue light-mediated transcriptional activation and repression of gene expression in bacteria. *Nucleic Acids Res.* **44**, 6994–7005 (2016).

46. 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).

47. Li, X. *et al.* A single-component light sensor system allows highly tunable and direct activation of gene expression in bacterial cells. *Nucleic Acids Res.* **48**, e33–e33 (2020).

48. Ding, Q. *et al.* Light-powered Escherichia coli cell division for chemical production. *Nat. Commun.* **11**, (2020).

49. An-adirekkun, J. *et al.* A yeast optogenetic toolkit (yOTK) for gene expression control in Saccharomyces cerevisiae. *Biotechnol. Bioeng.* **117**, 886–893 (2020).

50. Zhao, E. M. *et al.* Optogenetic regulation of engineered cellular metabolism for microbial chemical production. *Nature* **555**, 683–687 (2018).

51. Mansouri, M., Strittmatter, T. & Fussenegger, M. Light-Controlled Mammalian Cells and Their Therapeutic Applications in Synthetic Biology. *Adv. Sci.* **6**, (2019).

52. Roska, B. & Sahel, J.-A. Restoring vision Where we are today. *Nature* **557**, (2018).

53. Wrobel, C. *et al.* Optogenetic stimulation of cochlear neurons activates the auditory pathway and restores auditory-driven behavior in deaf adult gerbils. *Sci. Transl. Med.* **10**, (2018).

54. Shao, J. *et al.* Smartphone-controlled optogenetically engineered cells enable semiautomatic glucose homeostasis in diabetic mice. *Sci. Transl. Med.* **9**, 1–14 (2017).

55. Xu, Y. *et al.* Optogenetic control of chemokine receptor signal and T-cell migration. *Proc. Natl. Acad. Sci. U. S.*

A. **111**, 6371–6376 (2014).

56. Kottke, T., Xie, A., Larsen, D. S. & Hoff, W. D. Photoreceptors Take Charge: Emerging Principles for Light Sensing. *Annu. Rev. Biophys.* **47**, 291–313 (2018).

57. Ziegler, T. & Möglich, A. Photoreceptor engineering. *Front. Mol. Biosci.* **2**, 1–25 (2015).

58. Christie, J. M. *et al.* Plant UVR8 Photoreceptor Senses UV-B by Tryptophan-Mediated Disruption of Cross-Dimer Salt Bridges. *Science (80-.).* **335**, 1492–1496 (2012).

59. Favory, J. J. *et al.* Interaction of COP1 and UVR8 regulates UV-B-induced photomorphogenesis and stress acclimation in Arabidopsis. *EMBO J.* **28**, 591–601 (2009).

60. Chaves, I. *et al.* The cryptochromes: Blue light photoreceptors in plants and animals. *Annu. Rev. Plant Biol.* **62**, 335–364 (2011).

61. Brautigam, C. A. *et al.* Structure of the photolyase-like domain of cryptochrome 1 from Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12142–12147 (2004).

62. Müller, P. & Bouly, J. P. Searching for the mechanism of signalling by plant photoreceptor cryptochrome. *FEBS Lett.* **589**, 189–192 (2015).

63. Yang, H.-Q. *et al.* The C Termini of Arabidopsis Cryptochromes Mediate a Constitutive Light Response dopsis CRY1 protein also affects anthocyanin produc-tion, as well as chalcone synthase gene expression, and mutant cry1 plants are deficient in these responses. *Cell* **103**, 815–827 (2000).

64. Mas, P., Devlin, P. F., Panda, S. & Kay, S. A. Functional interaction of phytochrome B and cryptochrome 2. *Nature* **408**, 207–211 (2000).

65. Taslimi, A. *et al.* Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nat. Chem. Biol.* **12**, 425–430 (2016).

66. Braatsch, S., Gomelsky, M., Kuphal, S. & Klug, G. A single flavoprotein, AppA, integrates both redox and light signals in Rhodobacter sphaeroides. *Mol. Microbiol.* **45**, 827–836 (2002).

67. Masuda, S. & Bauer, C. E. AppA is a blue light photoreceptor that antirepresses photosynthesis gene expression in Rhodobacter sphaeroides. *Cell* **110**, 613–623 (2002).

68. Iseki, M. *et al.* A blue-light-activated adenylyl cyclase mediates photoavoidance in Euglena gracilis. *Nature* **415**, 1047–51 (2002).

69. Masuda, S. Light detection and signal transduction in the BLUF photoreceptors. *Plant Cell Physiol.* **54**, 171–179 (2013).

70. Gauden, M. *et al.* Photocycle of the flavin-binding photoreceptor AppA, a bacterial transcriptional antirepressor of photosynthesis genes. *Biochemistry* **44**, 3653–3662 (2005).

71. Dragnea, V., Waegele, M., Balascuta, S., Bauer, C. & Dragnea, B. Time-resolved spectroscopic studies of the AppA blue-light receptor BLUF domain from Rhodobacter sphaeroides. *Biochemistry* **44**, 15978–15985 (2005).

72. Hepp, S. *et al. An Optogenetic Tool for Induced Protein Stabilization Based on the Phaeodactylum tricornutum Aureochrome 1a Light–Oxygen–Voltage Domain. Journal of Molecular Biology* (Elsevier Ltd, 2020). doi:10.1016/j.jmb.2020.02.019

73. Christie, J. M. Phototropin blue-light receptors. *Annu. Rev. Plant Biol.* **58**, 21–45 (2007).

74. Receptor, B. L. *et al.* Photochemical and Mutational Analysis of the FMN-Binding Domains of the Plant. 9401–9410 (2000).

75. Harper, S. M., Christie, J. M. & Gardner, K. H. Disruption of the LOV−Jα Helix Interaction Activates Phototropin Kinase Activity. *Biochemistry* **43**, 16184–16192 (2004).

76. Möglich, A., Yang, X., Ayers, R. A. & Moffat, K. Structure and function of plant photoreceptors. *Annu. Rev. Plant Biol.* **61**, 21–47 (2010).

77. Christie, J. M., Swartz, T. E., Bogomolni, R. A. & Briggs, W. R. Phototropin LOV domains exhibit distinct roles in regulating photoreceptor function. *Plant J.* **32**, 205–219 (2002).

78. Huala, E. *et al.* Arabidopsis NPH1: A protein kinase with a putative redox-sensing domain. *Science (80-.).* **278**, 2120–2123 (1997).

79. Glantz, S. T. *et al.* Functional and topological diversity of LOV domain photoreceptors. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1442–E1451 (2016).

80. Harper, S. M., Neil, L. C. & Gardner, K. H. Structural basis of a phototropin light switch. *Science* **301**, 1541–4 (2003).

81. Salomon, M. *et al.* An optomechanical transducer in the blue light receptor phototropin from Avena sativa. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 12357–12361 (2001).

82. Crosson, S. & Moffat, K. Photoexcited structure of a plant photoreceptor domain reveals a light-driven molecular switch. *Plant Cell* **14**, 1067–1075 (2002).

83. Halavaty, A. S. & Moffat, K. N- and C-Terminal Flanking Regions Modulate Light-Induced Signal Transduction in the LOV2 Domain of the Blue Light Sensor Phototropin 1 from A V ena sati V a †,‡. *Biochemistry* **46**, 14001–14009 (2007).

84. Alexandre, M. T. A., Arents, J. C., Van Grondelle, R., Hellingwerf, K. J. & Kennis, J. T. M. A base-catalyzed mechanism for dark state recovery in the Avena sativa phototropin-1 LOV2 domain. *Biochemistry* **46**, 3129–3137 (2007).

85. Swartz, T. E. *et al.* The Photocycle of a Flavin-binding Domain of the Blue Light Photoreceptor Phototropin. *J. Biol. Chem.* **276**, 36493–36500 (2001).

86. Ballario, P. *et al.* White collar-1, a central regulator of blue light responses in Neurospora, is a zinc finger protein. *EMBO J.* **15**, 1650–1657 (1996).

87. Heintzen, C., Loros, J. J. & Dunlap, J. C. The PAS protein VIVID defines a clock-associated feedback loop that represses light input, modulates gating, and regulates clock resetting. *Cell* **104**, 453–464 (2001).

88. Elvin, M., Loros, J., Dunlap, J. & Heintzen, C. The PAS/LOV protein VIVID supports a rapidly dampened daytime oscillator that facilitates entrainment of the Neurospora circadian clock. *Genes Dev.* **19**, 2593–2605 (2005).

89. Froehlich, A. C., Liu, Y., Loros, J. J. & Dunlap, J. C. White collar-1, a circadian blue light photoreceptor, binding to the frequency promoter. *Science (80-.).* **297**, 815–819 (2002).

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

91. Patel, A., Muñoz, A., Halvorsen, K. & Rai, P. Creation and validation of a ligation-independent cloning (LIC) retroviral vector for stable gene transduction in mammalian cells. *BMC Biotechnol.* **12**, 3 (2012).

92. Loros, J. J. & Dunlap, J. C. Genetic and Molecular Analysis of Circadian Rhythms in Neurospora. *Annu. Rev. Physiol.* **63**, 757–794 (2001).

93. Zoltowski, B. D. *et al.* Conformational Switching in the Fungal Light Sensor Vivid. *Science (80-.).* **316**, 1054–1057 (2007).

94. Zoltowski, B. D. & Crane, B. R. Light Activation of the LOV Protein Vivid Generates a Rapidly Exchanging Dimer. *Biochemistry* **47**, 7012–7019 (2008).

95. Zoltowski, B. D. & Gardner, K. H. Tripping the light fantastic: Blue-light photoreceptors as examples of environmentally modulated protein-protein interactions. *Biochemistry* **50**, 4–16 (2011).

96. 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**, ra50–ra50 (2011).

97. Takano, H. *et al.* Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in Thermus thermophilus. *J. Bacteriol.* **193**, 2451–2459 (2011).

98. Takano, H. *et al.* Role and function of LitR, an adenosyl B12-bound light-sensitive regulator of Bacillus megaterium QM B1551, in regulation of carotenoid production. *J. Bacteriol.* **197**, 2301–2315 (2015).

99. Pérez-Marín, M. C., Padmanabhan, S., Polanco, M. C., Murillo, F. J. & Elías-Arnanz, M. Vitamin B12 partners the CarH repressor to downregulate a photoinducible promoter in Myxococcus xanthus. *Mol. Microbiol.* **67**, 804–819 (2008).

100. Cheng, Z., Yamamoto, H. & Bauer, C. E. Cobalamin's (Vitamin B 12) Surprising Function as a Photoreceptor. *Trends Biochem. Sci.* **41**, 647–650 (2016).

101. Jost, M. *et al.* Structural basis for gene regulation by a B12-dependent photoreceptor. *Nature* **526**, 536–541 (2015).

102. Auldridge, M. E. & Forest, K. T. Bacterial phytochromes: More than meets the light. *Crit. Rev. Biochem. Mol. Biol.* **46**, 67–88 (2011).

103. Butler, W. L., Norris, K. H., Siegelman, H. W. & Hendricks, S. B. Detection, Assay, and Preliminary Purification of the Pigment Controlling Photoresponsive Development of Plants. *Proc. Natl. Acad. Sci.* **45**, 1703–1708 (1959).

104. Borthwick, H. A., Hendricks, S. B., Parker, M. W., Toole, E. H. & Toole, V. K. A Reversible Photoreaction Controlling Seed Germination. *Proc. Natl. Acad. Sci.* **38**, 662–666 (1952).

105. Rockwell, N. C. *et al.* Eukaryotic algal phytochromes span the visible spectrum. *Proc. Natl. Acad. Sci.* **112**, E1051–E1051 (2015).

106. Rockwell, N. C., Ohlendorf, R. & Möglich, A. Cyanobacteriochromes in full color and three dimensions. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 806–807 (2013).

107. Anders, K. & Essen, L. O. The family of phytochrome-like photoreceptors: Diverse, complex and multi-colored, but very useful. *Curr. Opin. Struct. Biol.* **35**, 7–16 (2015).

108. Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol.* **57**, 837–858 (2006).

109. Chernov, K. G., Redchuk, T. A., Omelina, E. S. & Verkhusha, V. V. Near-Infrared Fluorescent Proteins, Biosensors, and Optogenetic Tools Engineered from Phytochromes. *Chem. Rev.* **117**, 6423–6446 (2017).

Bibliography

110. Andel, F., Lagarias, J. C. & Mathias, R. A. Resonance Raman analysis of chromophore structure in the lumi-R photoproduct of phytochrome. *Biochemistry* **35**, 15997–16008 (1996).

111. Yeh, K. C., Wu, S. H., Murphy, J. T. & Lagarias, J. C. A cyanobacterial phytochrome two-component light sensory system. *Science (80-.).* **277**, 1505–1508 (1997).

112. Blattner, F. R. *et al.* The complete genome sequence of Escherichia coli K-12. *Science (80-.).* **277**, 1453–1462 (1997).

113. Ghatak, S., King, Z. A., Sastry, A. & Palsson, B. O. The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. *Nucleic Acids Res.* **47**, 2446–2454 (2019).

114. Fang, X. *et al.* Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10286–10291 (2017).

115. Tan, S. Z. & Prather, K. L. Dynamic pathway regulation: recent advances and methods of construction. *Curr. Opin. Chem. Biol.* **41**, 28–35 (2017).

116. Gupta, A., Reizman, I. M. B., Reisch, C. R. & Prather, K. L. J. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* **3**, (2017).

117. Kent, R. & Dixon, N. Contemporary Tools for Regulating Gene Expression in Bacteria. *Trends Biotechnol.* **38**, 316–333 (2020).

118. Rojo, F. Repression of transcription initiation in bacteria. *J. Bacteriol.* **181**, 2987–2991 (1999).

119. Siegele, D. A. & Hu, J. C. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 8168–8172 (1997).

120. Keasling, J. D. Gene-expression tools for the metabolic engineering of bacteria. *Trends Biotechnol.* **17**, 452–460 (1999).

121. Terpe, K. Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* **72**, 211–222 (2006).

122. Gross, J. & Englesberg, E. Determination of the order of mutational sites governing l-arabinose utilization in Escherichia coli B r by transduction with phage P1bt. *Virology* **9**, 314–331 (1959).

123. Schleif, R. AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* **34**, 779–796 (2010).

124. Hogg, R. W. & Englesberg, E. L-arabinose binding protein from Escherichia coli B-r. *J. Bacteriol.* **100**, 423–432 (1969).

125. Schleif, R. AraC protein: A love-hate relationship. *BioEssays* **25**, 274–282 (2003).

126. Horazdovsky, B. F. & Hogg, R. W. Genetic reconstitution of the high-affinity L-arabinose transport system. *J. Bacteriol.* **171**, 3053–3059 (1989).

127. Lee, N. L., Gielow, W. O. & Wallace, R. G. Mechanism of araC autoregulation and the domains of two overlapping promoters, P(c) and P(bad), in the l-arabinose regulatory region of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 752–756 (1981).

128. Stoner, C. & Schleif, R. The araE low affinity l-arabinose transport promoter. Cloning, sequence, transcription start site and DNA binding sites of regulatory proteins. *J. Mol. Biol.* **171**, 369–381 (1983).

114

129. Johnson, C. M. & Schleif, R. F. In vivo induction kinetics of the arabinose promoters in Escherichia coli. *J. Bacteriol.* **177**, 3438–3442 (1995).

130. Fritz, G. *et al.* Single cell kinetics of phenotypic switching in the arabinose utilization system of E. coli. *PLoS One* **9**, (2014).

131. Schleif, R. Regulation of the L-arabinose operon of Escherichia coli. *Trends Genet.* **16**, 559–565 (2000).

132. Hahn, S. & Schleif, R. In vivo regulation of the Escherichia coli araC promoter. *J. Bacteriol.* **155**, 593–600 (1983).

133. Hirsh, J. & Schleif, R. In vivo experiments on the mechanism of action of l-arabinose C gene activator and lactose repressor. *J. Mol. Biol.* **80**, 433–444 (1973).

134. GREENBLATT, J. & SCHLEIF, R. Arabinose C Protein: Regulation of the Arabinose Operon in vitro. *Nat. New Biol.* **233**, 166–170 (1971).

135. Haggerty, D. M. & Schleif, R. Kinetics of the onset of catabolite repression in Escherichia coli as determined by lac messenger ribonucleic acid initiations and intracellular cyclic adenosine 3',5' monophosphate levels. *J. Bacteriol.* **123**, 946–953 (1975).

136. Aiba, H. Autoregulation of the Escherichia coli crp gene: CRP is a transcriptional repressor for its own gene. *Cell* **32**, 141–149 (1983).

137. Lobell, R. B. & Schleif, R. F. DNA looping and unlooping by AraC protein. *Science (80-.).* **250**, 528–532 (1990).

138. Khlebnikov, A., Skaug, T. & Keasling, J. D. Modulation of gene expression from the arabinose-inducible araBAD promoter. *J. Ind. Microbiol. Biotechnol.* **29**, 34–37 (2002).

139. Megerle, J. A., Fritz, G., Gerland, U., Jung, K. & Rädler, J. O. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophys. J.* **95**, 2103–2115 (2008).

140. Széliová, D., Krahulec, J., Šafránek, M., Lišková, V. & Turňa, J. Modulation of heterologous expression from PBAD promoter in Escherichia coli production strains. *J. Biotechnol.* **236**, 1–9 (2016).

141. Khlebnikov, A., Risa, Skaug, T., Carrier, T. A. & Keasling, J. D. Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol.* **182**, 7029–7034 (2000).

142. Morgan-Kiss, R. M., Wadler, C. & Cronan, J. E. Long-term and homogeneous regulation of the Escherichia coli araBAD promoter by use of a lactose transporter of relaxed specificity. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7373–7377 (2002).

143. Wilcox, G. The interaction of L-arabinose and D-fucose with AraC protein. *J. Biol. Chem.* **249**, 6892–4 (1974).

144. Beverin, S., Sheppard, D. E. & Park, S. S. D-Fucose as a gratuitous inducer of the L-arabinose operon in strains of Escherichia coli B-r mutant in gene araC. *J. Bacteriol.* **107**, 79–86 (1971).

145. Bustos, S. a & Schleif, R. F. Functional domains of the AraC protein. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 5638–5642 (1993).

146. Podlesek, Z. & Žgur Bertok, D. The DNA Damage Inducible SOS Response Is a Key Player in the Generation of Bacterial Persister Cells and Population Wide Tolerance. *Front. Microbiol.* **11**, 1–8 (2020).

147. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. Isolation of a recombinant copy of the gene encoding C/EBP. *Genes Dev.* **2**, 786–800 (1988).

148. Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: Property, design and functionality. *Adv. Drug Deliv. Rev.* **65**, 1357–1369 (2013).

149. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat. Commun.* **6**, 1–8 (2015).

150. Cosentino, C. *et al.* Engineering of a light-gated potassium channel. *Science (80-.).* **348**, 707–710 (2015).

151. Niopek, D. *et al.* Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat. Commun.* **5**, 4404 (2014).

152. Niopek, D. *et al.* Optogenetic control of nuclear protein export. *Nat. Commun.* **7**, 1–9 (2016).

153. Dagliyan, O. *et al.* Engineering extrinsic disorder to control protein activity in living cells. **354**, (2016).

154. Hoffmann, M. D., Bubeck, F., Eils, R. & Niopek, D. Controlling Cells with Light and LOV. *Adv. Biosyst.* **2**, (2018).

155. Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-Component Signal Transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).

156. Salvado, B., Vilaprinyo, E., Sorribas, A. & Alves, R. A survey of HK, HPt, and RR domains and their organization in two-component systems and phosphorelay proteins of organisms with fully sequenced genomes. *PeerJ* **2015**, (2015).

157. Gao, R. & Stock, A. M. Biological Insights from Structures of Two-Component Proteins. *Annu. Rev. Microbiol.* **63**, 133–154 (2009).

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

159. Levskaya, A. *et al.* Synthetic biology: engineering Escherichia coli to see light. *Nature* **438**, 441–442 (2005).

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

161. 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).

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

163. 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).

164. Song, J. Y. *et al.* Near-UV cyanobacteriochrome signaling system elicits negative phototaxis in the cyanobacterium Synechocystis sp. PCC 6803. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10780–10785 (2011).

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

166. Shao, Q. *et al.* Coupling of DNA Replication and Negative Feedback Controls Gene Expression for Cell-Fate Decisions. *iScience* **6**, 1–12 (2018).

167. Lalwani, M. A. *et al.* Optogenetic control of the lac operon for bacterial chemical and protein production. *Nat. Chem. Biol.* (2020). doi:10.1038/s41589-020-0639-1

168. Tu, G.-F., Reid, G. E., Zhang, J.-G., Moritz, R. L. & Simpson, R. J. C-terminal Extension of Truncated Recombinant Proteins in Escherichia coli with a 10Sa RNA Decapeptide. *J. Biol. Chem.* **270**, 9322–9326 (1995).

169. Ulrich, L. E., Koonin, E. V & Zhulin, I. B. One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.* **13**, 52–56 (2005).

170. Capra, E. J. & Laub, M. T. Evolution of two-component signal transduction systems. *Annu. Rev. Microbiol.* **66**, 325–347 (2012).

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

172. Thliveris, A. T., Little, J. W. & Mount, D. W. Repression of the E coli recA gene requires at least two LexA protein monomers. *Biochimie* **73**, 449–456 (1991).

173. Motta-Mena, L. B. *et al.* An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat. Chem. Biol.* **10**, 196–202 (2014).

174. Sadowski, I., Ma, J., Triezenberg, S. & Ptashne, M. GAL4-VP16 is an unusually potent transcriptional activator. *Nature* **335**, 563–564 (1988).

175. Nash, A. I. *et al.* Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proc. Natl. Acad. Sci.* **108**, 9449–9454 (2011).

176. Rivera-Cancel, G., Motta-Mena, L. B. & Gardner, K. H. Identification of Natural and Artificial DNA Substrates for Light-Activated LOV−HTH Transcription Factor EL222. **51**, 19 (2012).

177. Zoltowski, B. D., Motta-Mena, L. B. & Gardner, K. H. Blue Light-Induced Dimerization of a Bacterial LOV−HTH DNA-Binding Protein. *Biochemistry* **52**, 2020 (2013).

178. Zoltowski, B. D., Nash, A. I. & Gardner, K. H. Variations in protein-flavin hydrogen bonding in a light, oxygen, voltage domain produce non-Arrhenius kinetics of adduct decay. *Biochemistry* **50**, 8771–8779 (2011).

179. Egland, K. A. & Greenberg, E. P. Conversion of the Vibrio fischeri transcriptional activator, LuxR, to a repressor. *J. Bacteriol.* **182**, 805–811 (2000).

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

181. Zoltowski, B. D., Vaccaro, B. & Crane, B. R. Mechanism-based tuning of a LOV domain photoreceptor. *Nat. Chem. Biol.* **5**, 827–834 (2009).

182. 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).

183. Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D. & Voigt, C. A. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase . *Mol. Syst. Biol.* **10**, 742 (2014).

184. Nihongaki, Y., Otabe, T., Ueda, Y. & Sato, M. A split CRISPR–Cpf1 platform for inducible genome editing and gene activation. *Nat. Chem. Biol.* **15**, 882–888 (2019).

185. 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).

186. Sheets, M. B., Wong, W. W. & Dunlop, M. J. Light-Inducible Recombinases for Bacterial Optogenetics. *ACS Synth. Biol.* **9**, 227–235 (2020).

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

188. Ortiz-Guerrero, J. M., Polanco, M. C., Murillo, F. J., Padmanabhan, S. & Elías-Arnanz, M. Light-dependent gene regulation by a coenzyme B12-based photoreceptor. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7565–7570 (2011).

189. Young, D. D. & Deiters, A. Photochemical activation of protein expression in bacterial cells. *Angew. Chemie - Int. Ed.* **46**, 4290–4292 (2007).

190. Binder, D. *et al.* Photocaged Arabinose: A Novel Optogenetic Switch for Rapid and Gradual Control of Microbial Gene Expression. *ChemBioChem* **17**, 296–299 (2016).

191. Cambridge, S. B. *et al.* Doxycycline-dependent photoactivated gene expression in eukaryotic systems. *Nat. Methods* **6**, 527–531 (2009).

192. Sauers, D. J. *et al.* Light-activated gene expression directs segregation of co-cultured cells in vitro. *ACS Chem. Biol.* **5**, 313–320 (2010).

193. Cambridge, S. B., Geissler, D., Keller, S. & Cürten, B. A caged doxycycline analogue for photoactivated gene expression. *Angew. Chemie - Int. Ed.* **45**, 2229–2231 (2006).

194. Levskaya, A. *et al.* Engineering Escherichia coli to see light. *Nature* **438**, 441–442 (2005).

195. Choi, H. S. *et al.* A novel flavin-containing monooxygenase from Methylophaga sp. strain SK1 and its indigo synthesis in Escherichia coli. *Biochem. Biophys. Res. Commun.* **306**, 930–936 (2003).

196. Fernandez-Rodriguez, J., Moser, F., Song, M. & Voigt, C. A. Engineering RGB color vision into Escherichia coli. *Nat. Chem. Biol.* **AOP**, 1–5 (2017).

197. Kong, W., Blanchard, A. E., Liao, C. & Lu, T. Engineering robust and tunable spatial structures with synthetic gene circuits. *Nucleic Acids Res.* **45**, 1005–1014 (2017).

198. Berlyn, M. K. B. Linkage Map of Escherichia coli K-12, Edition 10: The Traditional Map. *Microbiol. Mol. Biol. Rev.* **62**, 1554–1554 (1998).

199. Keseler, I. M. *et al.* EcoCyc: Fusing model organism databases with systems biology. *Nucleic Acids Res.* **41**, 605–612 (2013).

200. Zhou, J. & Rudd, K. E. EcoGene 3.0. *Nucleic Acids Res.* **41**, 613–624 (2013).

201. Bateman, A. *et al.* UniProt: A hub for protein information. *Nucleic Acids Res.* **43**, D204–D212 (2015).

202. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

203. Merlin, C., McAteer, S. & Masters, M. Tools for Characterization of Escherichia coli Genes of Unknown Function. *J. Bacteriol.* **184**, 4573–4581 (2002).

204. Sun, W. *et al.* Light-induced protein degradation in human-derived cells. *Biochemical and Biophysical*

Research Communications **487**, 2–7 (2017).

205. Tucker, D. L., Tucker, N. & Conway, T. Gene expression profiling of the pH response in Escherichia coli. *J. Bacteriol.* **184**, 6551–6558 (2002).

206. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

207. Lavezzo, E., Falda, M., Fontana, P., Bianco, L. & Toppo, S. Enhancing protein function prediction with taxonomic constraints - The Argot2.5 web server. *Methods* **93**, 15–23 (2016).

208. Törönen, P., Medlar, A. & Holm, L. PANNZER2: A rapid functional annotation web server. *Nucleic Acids Res.* **46**, W84–W88 (2018).

209. Kulmanov, M., Hoehndorf, R. & Cowen, L. DeepGOPlus: Improved protein function prediction from sequence. *Bioinformatics* **36**, 422–429 (2020).

210. Gao, Y. *et al.* Systematic discovery of uncharacterized transcription factors in Escherichia coli K-12 MG1655. *Nucleic Acids Res.* **46**, 10682–10696 (2018).

211. Fuhrer, T., Zampieri, M., Sévin, D. C., Sauer, U. & Zamboni, N. Genomewide landscape of gene–metabolome associations in Escherichia coli . *Mol. Syst. Biol.* **13**, 907 (2017).

212. Sévin, D. C., Fuhrer, T., Zamboni, N. & Sauer, U. Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli. *Nat. Methods* **14**, 187–194 (2017).

213. Spratt, B. G. Distinct penicillin binding proteins involved in the division, elongation, and shape of Escherichia coli K 12. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 2999–3003 (1975).

214. Tamaki, S., Matsuzawa, H. & Matsuhashi, M. Cluster mrdA and mrdB genes responsible for the rod shape and mecillinam sensitivity of Escherichia coli. *J. Bacteriol.* **141**, 52–57 (1980).

215. Höltje, J.-V. Growth of the Stress-Bearing and Shape-Maintaining Murein Sacculus of Escherichia coli. *Microbiol. Mol. Biol. Rev.* **62**, 181–203 (1998).

216. Kruse, T., Bork-Jensen, J. & Gerdes, K. The morphogenetic MreBCD proteins of Escherichia coli form an essential membrane-bound complex. *Mol. Microbiol.* **55**, 78–89 (2005).

217. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. Escherichia coli physiology in Luria-Bertani broth. *J. Bacteriol.* **189**, 8746–8749 (2007).

218. de Boer, P. A. J., Crossley, R. E. & Rothfield, L. I. A division inhibitor and a topological specificity factor coded for by the minicell locus determine proper placement of the division septum in E. coli. *Cell* **56**, 641–649 (1989).

219. Di Ventura, B. *et al.* Chromosome segregation by the Escherichia coli Min system. *Mol. Syst. Biol.* **9**, 1–12 (2013).

220. Wu, W., Park, K.-T., Holyoak, T. & Lutkenhaus, J. Determination of the structure of the MinD-ATP complex reveals the orientation of MinD on the membrane and the relative location of the binding sites for MinE and MinC. *Mol. Microbiol.* **79**, 1515–1528 (2011).

221. Szeto, T. H., Rowland, S. L., Rothfield, L. I. & King, G. F. Membrane localization of MinD is mediated by a C-terminal motif that is conserved across eubacteria, archaea, and chloroplasts. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15693–15698 (2002).

222.   De Boer, P. A. J., Crossley, R. E. & Rothfield, L. I. Roles of MinC and MinD in the site-specific septation block mediated by the MinCDE system of Escherichia coli. *J. Bacteriol.* **174**, 63–70 (1992).

223.   Hu, Z., Mukherjee, A., Pichoff, S. & Lutkenhaus, J. The MinC component of the division site selection system in Escherichia coli interacts with FtsZ to prevent polymerization. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 14819–14824 (1999).

224.   Di Ventura, B. & Sourjik, V. Self-organized partitioning of dynamically localized proteins in bacterial cell division. *Mol. Syst. Biol.* **7**, 1–13 (2011).

225.   Raskin, D. M. & De Boer, P. A. J. Rapid pole-to-pole oscillation of a protein required for directing division to the middle of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 4971–4976 (1999).

226.   Bi, E. & Lutkenhaus, J. FtsZ ring structure associated with division in Escherichia coli. *Nature* **354**, 161–164 (1991).

227.   Bonny, M., Fischer-Friedrich, E., Loose, M., Schwille, P. & Kruse, K. Membrane Binding of MinE Allows for a Comprehensive Description of Min-Protein Pattern Formation. *PLoS Comput. Biol.* **9**, (2013).

228.   Hu, Z. & Lutkenhaus, J. A conserved sequence at the C-terminus of MinD is required for binding to the membrane and targeting MinC to the septum. *Mol. Microbiol.* **47**, 345–355 (2003).

229.   Szeto, T. H., Rowland, S. L., Habrukowich, C. L. & King, G. F. The MinD membrane targeting sequence is a transplantable lipid-binding helix. *J. Biol. Chem.* **278**, 40050–40056 (2003).

230.   Yu, X. C. & Margolin, W. FtsZ ring clusters in min and partition mutants: Role of both the Min system and the nucleoid in regulating FtsZ ring localization. *Mol. Microbiol.* **32**, 315–326 (1999).

231.   McKenna, J. Minicells, Back in Fashion. *Virgin Aust. Voyer* **198**, pp108–112 (2016).

232.   Wachi, M. *et al.* Mutant isolation and molecular cloning of mre genes, which determine cell shape, sensitivity to mecillinam, and amount of penicillin-binding proteins in Escherichia coli. *J. Bacteriol.* **169**, 4935–4940 (1987).

233.   Doi, M. *et al.* Determinations of the DNA sequence of the mreB gene and of the gene products of the mre region that function in formation of the rod shape of Escherichia coli cells. *J. Bacteriol.* **170**, 4619–4624 (1988).

234.   Wachi, M. & Matsuhashi, M. Negative control of cell division by mreB, a gene that functions in determining the rod shape of Escherichia coli cells. *J. Bacteriol.* **171**, 3123–3127 (1989).

235.   Ursell, T. S. *et al.* Rod-like bacterial shape is maintained by feedback between cell curvature and cytoskeletal localization. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

236.   Deng, Y., Sun, M. & Shaevitz, J. W. Direct measurement of cell wall stress stiffening and turgor pressure in live bacterial cells. *Phys. Rev. Lett.* **107**, 7–10 (2011).

237.   Wang, S., Furchtgott, L., Huang, K. C. & Shaevitz, J. W. Helical insertion of peptidoglycan produces chiral ordering of the bacterial cell wall. *Proc. Natl. Acad. Sci. U. S. A.* **109**, (2012).

238.   Ouzounov, N. *et al.* MreB Orientation Correlates with Cell Diameter in Escherichia coli. *Biophys. J.* **111**, 1035–1043 (2016).

239.   Bendezú, F. O., Hale, C. A., Bernhardt, T. G. & De Boer, P. A. J. RodZ (YfgA) is required for proper assembly of the MreB actin cytoskeleton and cell shape in E. coli. *EMBO J.* **28**, 193–204 (2009).

240.   Alyahya, S. A. *et al.* RodZ, a component of the bacterial core morphogenic apparatus. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1239–1244 (2009).

241. Van Den Ent, F., Johnson, C. M., Persons, L., De Boer, P. & Löwe, J. Bacterial actin MreB assembles in complex with cell shape protein RodZ. *EMBO J.* **29**, 1081–1090 (2010).

242. Ikebe, R., Kuwabara, Y., Chikada, T., Niki, H. & Shiomi, D. The periplasmic disordered domain of RodZ promotes its self-interaction in Escherichia coli. *Genes to Cells* **23**, 307–317 (2018).

243. Kruse, T., Møeller-Jensen, J., Løbner-Olesen, A. & Gerdes, K. Dysfunctional MreB inhibits chromosome segregation in Escherichia coli. *EMBO J.* **22**, 5283–5292 (2003).

244. Colavin, A., Shi, H. & Huang, K. C. RodZ modulates geometric localization of the bacterial actin MreB to regulate cell shape. *Nat. Commun.* **9**, 1–11 (2018).

245. Rengby, O. & Arnér, E. S. J. Titration and conditional knockdown of the prfB gene in Escherichia coli: Effects on growth and overproduction of the recombinant mammalian selenoprotein thioredoxin reductase. *Appl. Environ. Microbiol.* **73**, 432–441 (2007).

246. Roux, A., Beloin, C. & Ghigo, J. M. Combined inactivation and expression strategy to study gene function under physiological conditions: Application to identification of new Escherichia coli adhesins. *J. Bacteriol.* **187**, 1001–1013 (2005).

247. Timmes, A., Rodgers, M. & Schleif, R. Biochemical and physiological properties of the DNA binding domain of AraC protein. *J. Mol. Biol.* **340**, 731–738 (2004).

248. Reeder, T. & Schleif, R. AraC protein can activate transcription from only one position and when pointed in only one direction. *Journal of Molecular Biology* **231**, 205–218 (1993).

249. Xu, X. *et al.* A Single-Component Optogenetic System Allows Stringent Switch of Gene Expression in Yeast Cells. *ACS Synth. Biol.* **7**, 2045–2053 (2018).

250. Romano, E. *et al.* An inducible AraC that responds to blue light instead of arabinose. (2020). doi:10.1101/2020.07.14.202911

251. De Mera, R. M. M. F. *et al.* A simple optogenetic MAPK inhibitor design reveals resonance between transcription-regulating circuitry and temporally-encoded inputs. *Nat. Commun.* **8**, (2017).

252. Cloin, B. M. C. *et al.* Efficient switching of mCherry fluorescence using chemical caging. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 7013–7018 (2017).

253. Eustance, R. J. & Schleif, R. F. The linker region of AraC protein. *J. Bacteriol.* **178**, 7025–7030 (1996).

254. Bell, C. E., Frescura, P., Hochschild, A. & Lewis, M. Crystal structure of the λ repressor C-terminal domain provides a model for cooperative operator binding. *Cell* **101**, 801–811 (2000).

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

256. Kuriata, A. *et al.* CABS-flex 2.0: A web server for fast simulations of flexibility of protein structures. *Nucleic Acids Res.* **46**, W338–W343 (2018).

257. Dao-Thi, M. H. *et al.* Crystallization of CcdB in complex with a GyrA fragment. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 1132–1134 (2004).

258. Bernard, P. *et al.* The F Plasmid CcdB Protein Induces Efficient ATP-dependent DNA Cleavage by Gyrase. *J. Mol. Biol.* **234**, 534–541 (1993).

259. 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).

260. Aoki, S. K. *et al.* A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* **570**, 533–537 (2019).

261. Afroz, T., Biliouris, K., Kaznessis, Y. & Beisel, C. L. Bacterial sugar utilization gives rise to distinct single-cell behaviours. *Mol. Microbiol.* **93**, 1093–1103 (2014).

262. Mäkelä, J. *et al.* In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter. *Nucleic Acids Res.* **41**, 6544–6552 (2013).

263. Rodgers, M. E. & Schleif, R. DNA tape measurements of AraC. *Nucleic Acids Res.* **36**, 404–410 (2008).

264. Seedorff, J. & Schleif, R. Active role of the interdomain linker of AraC. *J. Bacteriol.* **193**, 5737–5746 (2011).

265. Lee, N., Francklyn, C. & Hamilton, E. P. Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 8814–8818 (1987).

266. Malzahn, E., Ciprianidis, S., Káldi, K., Schafmeier, T. & Brunner, M. Photoadaptation in Neurospora by Competitive Interaction of Activating and Inhibitory LOV Domains. *Cell* **142**, 762–772 (2010).

267. Hunt, S. M., Elvin, M., Crosthwaite, S. K. & Heintzen, C. The PAS/LOV protein VIVID controls temperature compensation of circadian clock phase and development in Neurospora crassa. *Genes Dev.* **21**, 1964–1974 (2007).

268. Lee, C. T., Malzahn, E., Brunner, M. & Mayer, M. P. Light-induced differences in conformational dynamics of the circadian clock regulator VIVID. *J. Mol. Biol.* **426**, 601–610 (2014).

269. Andersson, D. I. & Levin, B. R. The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* **2**, 489–493 (1999).

270. Händel, N., Schuurmans, J. M., Brul, S. & Ter Kuilea, B. H. Compensation of the metabolic costs of antibiotic resistance by physiological adaptation in escherichia coli. *Antimicrob. Agents Chemother.* **57**, 3752–3762 (2013).

271. Palanisamy, N. *et al.* Split intein-mediated selection of cells containing two plasmids using a single antibiotic. *Nat. Commun.* **10**, (2019).

272. Shiomi, D., Sakai, M. & Niki, H. Determination of bacterial rod shape by a novel cytoskeletal membrane protein. *EMBO J.* **27**, 3081–3091 (2008).

273. Yang, D. C., Blair, K. M. & Salama, N. R. Staying in Shape: the Impact of Cell Shape on Bacterial Survival in Diverse Environments. *Microbiol. Mol. Biol. Rev.* **80**, 187–203 (2016).

274. Smith, W. P. J. *et al.* Cell morphology drives spatial patterning in microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E280–E286 (2017).

275. Silencing, C. G. CRISPR-Cas-Mediated Gene Silencing Reveals RacR To Be a Negative Regulator of YdaS and YdaT Toxins in Escherichia coli K-12. **2**, 1–11 (2017).

276. Stancik, L. M. *et al.* pH-Dependent Expression of Periplasmic Proteins and Amino Acid Catabolism in Escherichia coli. *J. Bacteriol.* **184**, 4246–4258 (2002).

277. Chodisetti, P. K. & Reddy, M. Peptidoglycan hydrolase of an unusual cross-link cleavage specificity contributes to bacterial cell wall synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7825–7830 (2019).

278. Otoupal, P. B. & Chatterjee, A. CRISPR gene perturbations provide insights for improving bacterial biofuel tolerance. *Front. Bioeng. Biotechnol.* **6**, 1–15 (2018).

279. Yamamoto, K. & Ishihama, A. Characterization of copper-inducible promoters regulated by CpxA/CpxR in Escherichia coli. *Biosci. Biotechnol. Biochem.* **70**, 1688–1695 (2006).

280. Raivio, T. L., Leblanc, S. K. D. & Price, N. L. The Escherichia coli Cpx envelope stress response regulates genes of diverse function that impact antibiotic resistance and membrane integrity. *J. Bacteriol.* **195**, 2755–2767 (2013).

281. Brieger, A. *et al.* C-terminal fluorescent labeling impairs functionality of DNA mismatch repair proteins. *PLoS One* **7**, (2012).

282. Shiomi, D. & Margolin, W. The C-terminal domain of MinC inhibits assembly of the Z ring in Escherichia coli. *J. Bacteriol.* **189**, 236–243 (2007).

283. Weill, U. *et al.* Assessment of GFP Tag Position on Protein Localization and Growth Fitness in Yeast. *J. Mol. Biol.* **431**, 636–641 (2019).

284. Falda, M. *et al.* Argot2: A large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics* **13**, (2012).

285. Yuan, J. *et al.* Vibrio cholerae ParE2 poisons DNA gyrase via a mechanism distinct from other gyrase inhibitors. *J. Biol. Chem.* **285**, 40397–40408 (2010).

286. Hallez, R. *et al.* New toxins homologous to ParE belonging to three-component toxin-antitoxin systems in Escherichia coli O157:H7. *Mol. Microbiol.* **76**, 719–732 (2010).

287. 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**, 1203–10 (1997).

288. 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).

289. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

290. Chen, Z. & Zhao, H. A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res.* **33**, 1–7 (2005).

291. Ohlendorf, R., Schumacher, C. H., Richter, F. & Möglich, A. Library-Aided Probing of Linker Determinants in Hybrid Photoreceptors. *ACS Synth. Biol.* **5**, 1117–1126 (2016).

292. Larson, M. H., Greenleaf, W. J., Landick, R. & Block, S. M. Applied Force Reveals Mechanistic and Energetic Details of Transcription Termination. *Cell* **132**, 971–982 (2008).

293. Yeung, E. *et al.* Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Syst.* **5**, 11-24.e12 (2017).

294. Lobell, R. B. & Schleif, R. AraC-DNA looping: orientation and distance dependent loop breaking by cycic AMP receptor protein. *J. Mol. Biol.* **218**, 45–54 (1991).

295. Seabold, R. R. & Schleif, R. F. Apo-AraC actively seeks to loop. *J. Mol. Biol.* **278**, 529–538 (1998).

296. Hernández-Candia, C. N., Casas-Flores, S. & Gutiérrez-Medina, B. Light induces oxidative damage and protein stability in the fungal photoreceptor Vivid. *PLoS One* **13**, 1–16 (2018).

Bibliography

297. Vendruscolo, M., Knowles, T. P. J. & Dobson, C. M. Protein solubility and protein homeostasis: A generic view of protein misfolding disorders. *Cold Spring Harb. Perspect. Biol.* **3**, 1–12 (2011).

298. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).

299. Lindner, A. B., Madden, R., Demarez, A., Stewart, E. J. & Taddei, F. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 3076–3081 (2008).

300. Akerlund, T., Bernander, R. & Nordström, K. Cell division in Escherichia coli minB mutants. *Mol. Microbiol.* **6**, 2073–2083 (1992).

301. Slater, M. & Schaechter, M. Control of cell division in bacteria. *Bacteriol. Rev.* **38**, 199–221 (1974).

302. Adler, H. I., Fisher, W. D., Cohen, A. & Hardigree, A. A. MINIATURE escherichia coli CELLS DEFICIENT IN DNA. *Proc. Natl. Acad. Sci.* **57**, 321–326 (1967).

303. MacDiarmid, J. A. *et al.* Bacterially Derived 400 nm Particles for Encapsulation and Cancer Cell Targeting of Chemotherapeutics. *Cancer Cell* **11**, 431–445 (2007).

304. Kaitu'u-Lino, T. J. *et al.* Targeted nanoparticle delivery of doxorubicin into placental tissues to treat ectopic pregnancies. *Endocrinology* **154**, 911–919 (2013).

305. MacDiarmid, J. A. *et al.* Targeted doxorubicin delivery to brain tumors via minicells: Proof of principle using dogs with spontaneously occurring tumors as a model. *PLoS One* **11**, 1–23 (2016).

306. Whittle, J. R. *et al.* First in human nanotechnology doxorubicin delivery system to target epidermal growth factor receptors in recurrent glioblastoma. *J. Clin. Neurosci.* **22**, 1889–1894 (2015).

307. Solomon, B. J. *et al.* A First-Time-In-Human Phase i Clinical Trial of Bispecific Antibody-Targeted, Paclitaxel-Packaged Bacterial Minicells. *PLoS One* **10**, 1–17 (2015).

308. Roozen, K. J., Fenwick, R. G. & Curtiss, R. Synthesis of ribonucleic acid and protein in plasmid-containing minicells of Escherichia coli K-12. *J. Bacteriol.* **107**, 21–33 (1971).

309. Libby, R. T., Shaw, J. E. & Reeve, J. N. Expression of coliphage T7 in aging anucleate minicells of Escherichia coli. *Mech. Ageing Dev.* **27**, 197–206 (1984).

310. Carleton, H. A., Lara-Tejero, M., Liu, X. & Galán, J. E. Engineering the type III secretion system in non-replicating bacterial minicells for antigen delivery. *Nat. Commun.* **4**, 1–8 (2013).

311. Alfaleh, M. A. *et al.* Targeting mesothelin receptors with drugloaded bacterial nanocells suppresses human mesothelioma tumour growth in mouse xenograft models. *PLoS One* **12**, 1–21 (2017).

312. Nissim, L. & Bar-Ziv, R. H. A tunable dual-promoter integrator for targeting of cancer cells. *Mol. Syst. Biol.* **6**, 1–9 (2010).

313. Nissim, L. *et al.* Synthetic RNA-Based Immunomodulatory Gene Circuits for Cancer Immunotherapy. *Cell* **171**, 1138-1150.e15 (2017).

314. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science (80-.).* **333**, 1307–1311 (2011).

315. Locke, J. C. W., Young, J. W., Fontes, M., Jimenez, M. J. H. & Elowitz, M. B. Stochastic Pulse Regulation in Bacterial Stress Response. *Science (80-.).* **334**, 366–369 (2011).

124

316.   Hao, N. & O'Shea, E. K. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. Mol. Biol.* **19**, 31–39 (2012).

317.   Purvis, J. E. *et al.* P53 Dynamics Control Cell Fate. *Science (80-.).* **336**, 1440–1444 (2012).

318.   Yissachar, N. *et al.* Dynamic Response Diversity of NFAT Isoforms in Individual Living Cells. *Mol. Cell* **49**, 322–330 (2013).

319.   Levine, J. H., Lin, Y. & Elowitz, M. B. Functional Roles of Pulsing in Genetic Circuits. *Science (80-.).* **342**, 1193–1200 (2013).

320.   Levine, J. H., Fontes, M. E., Dworkin, J. & Elowitz, M. B. Pulsed feedback defers cellular differentiation. *PLoS Biol.* **10**, (2012).

321.   Süel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. Tunability and noise dependence in differentiation dynamics. *Science (80-.).* **315**, 1716–1719 (2007).

322.   Cai, L., Dalal, C. K. & Elowitz, M. B. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**, 485–490 (2008).

323.   Sampaio, N. M. V, Blassick, C. M., Lugagne, J.-B. & Dunlop, M. J. Cell-to-cell heterogeneity in Escherichia coli stress response originates from pulsatile expression and growth. *bioRxiv* (2020). doi:10.1101/2020.09.14.297101

324.   Lugagne, J. B. & Dunlop, M. J. Cell-machine interfaces for characterizing gene regulatory network dynamics. *Curr. Opin. Syst. Biol.* **14**, 1–8 (2019).

325.   Sampaio, N. M. V. & Dunlop, M. J. Functional roles of microbial cell-to-cell heterogeneity and emerging technologies for analysis and control. *Curr. Opin. Microbiol.* **57**, 87–94 (2020).

326.   Benzinger, D. & Khammash, M. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat. Commun.* **9**, (2018).

327.   Toyooka, T., Hisatomi, O., Takahashi, F., Kataoka, H. & Terazima, M. Photoreactions of aureochrome-1. *Biophys. J.* **100**, 2801–2809 (2011).

328.   Takahashi, F. *et al.* AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19625–19630 (2007).

329.   Grusch, M. *et al.* Spatio-temporally precise activation of engineered receptor tyrosine kinases by light. *EMBO ...* **33**, 1713–1727 (2014).

330.   Aidelberg, G. *et al.* Hierarchy of non-glucose sugars in Escherichia coli. *BMC Syst. Biol.* **8**, 133 (2014).

331.   Ma, L., Jackson, K. D., Landry, R. M., Parsek, M. R. & Wozniak, D. J. Analysis of Pseudomonas aeruginosa conditional psl variants reveals roles for the psl polysaccharide in adhesion and maintaining biofilm structure postattachment. *J. Bacteriol.* **188**, 8213–8221 (2006).

332.   Boles, B. R., Thoendel, M. & Singh, P. K. Rhamnolipids mediate detachment of Pseudomonas aeruginosa from biofilms. *Mol. Microbiol.* **57**, 1210–1223 (2005).

333.   Baynham, P. J., Ramsey, D. M., Gvozdyev, B. V., Cordonnier, E. M. & Wozniak, D. J. The Pseudomonas aeruginosa ribbon-helix-helix DNA-binding protein AlgZ (AmrZ) controls twitching motility and biogenesis of type IV pili. *J. Bacteriol.* **188**, 132–140 (2006).

334.   Damron, F. H., McKenney, E. S., Schweizer, H. P. & Goldberg, J. B. Construction of a Broad-Host-Range Tn 7

-Based Vector for Single-Copy P BAD -Controlled Gene Expression in Gram-Negative Bacteria. *Appl. Environ. Microbiol.* **79**, 718–721 (2013).

335. Stringer, A. M. *et al.* Genome-scale analyses of Escherichia coli and Salmonella enterica AraC reveal noncanonical targets and an expanded core regulon. *J. Bacteriol.* **196**, 660–671 (2014).

336. Ganai, S., Arenas, R. B., Sauer, J. P., Bentley, B. & Forbes, N. S. In tumors Salmonella migrate away from vasculature toward the transition zone and induce apoptosis. *Cancer Gene Ther.* **18**, 457–466 (2011).

337. Pawelek, J. M., Low, K. B. & Bermudes, D. Bacteria as tumour-targeting vectors. *Lancet Oncol.* **4**, 548–556 (2003).

338. Hong, H. *et al.* Targeted deletion of the ara operon of Salmonella typhimurium enhances L-arabinose accumulation and drives P BAD -promoted expression of anti-cancer toxins and imaging agents. *Cell Cycle* **13**, 3112–3120 (2014).

339. Fricke, P. M. *et al.* A tunable l-arabinose-inducible expression plasmid for the acetic acid bacterium Gluconobacter oxydans. *Appl. Microbiol. Biotechnol.* **104**, 9267–9282 (2020).

340. Meneely, K. M. & Lamb, A. L. Biochemical characterization of a flavin adenine dinculeotide-dependent monooxygenase, ornithine hydroxylase from Pseudomonas aeruginosa, suggests a novel reaction mechanism. *Biochemistry* **46**, 11930–11937 (2007).

341. Ostrowski, J. *et al.* Characterization of the flavoprotein moieties of NADPH-sulfite reductase from Salmonella typhimurium and Escherichia coli. Physicochemical and catalytic properties, amino acid sequence deduced from DNA sequence of cysJ, and comparison with NADPH-cytochrom. *J. Biol. Chem.* **264**, 15796–15808 (1989).

342. Kawai, S., Goda-Tsutsumi, M., Yakushi, T., Kano, K. & Matsushita, K. Heterologous overexpression and characterization of a flavoprotein-cytochrome c complex fructose dehydrogenase of Gluconobacter japonicus NBRC3260. *Appl. Environ. Microbiol.* **79**, 1654–1660 (2013).

343. Noman, A. E. *et al.* A novel strain of acetic acid bacteria Gluconobacter oxydans FBFS97 involved in riboflavin production. *Sci. Rep.* **10**, 1–17 (2020).

344. Zheng, M. *et al.* DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. *J. Bacteriol.* **183**, 4562–4570 (2001).

345. Hemm, M. R. *et al.* Small stress response proteins in Escherichia coli: Proteins missed by classical proteomic studies. *J. Bacteriol.* **192**, 46–58 (2010).

346. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol. Syst. Biol.* **2**, (2006).

347. Dowler, S., Kular, G. & Alessi, D. R. Protein lipid overlay assay. *Sci. STKE* **2002**, 1–11 (2002).

348. Baron, C. L. & Malhotra, V. Role of diacylglycerol in PKD recruitment to the TGN and protein transport to the plasma membrane. *Science (80-.).* **295**, 325–328 (2002).

349. Munnik, T. & Wierzchowiecka, M. Lipid-Binding Analysis Using a Fat Blot Assay. in (eds. Munnik, T. & Heilmann, I.) **1009**, 253–259 (Humana Press, 2013).

350. Frick, M. & Schmidt, C. Mass spectrometry—A versatile tool for characterising the lipid environment of membrane protein assemblies. *Chem. Phys. Lipids* **221**, 145–157 (2019).

351. Wei, S. P. *et al.* Formation and functionalization of membraneless compartments in Escherichia coli. *Nat.*

Chem. Biol. **16**, 1143–1148 (2020).

352.   Feric, M. *et al.* Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **165**, 1686–1697 (2016).

353.   Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7189–7194 (2015).

354.   Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science (80-.).* **357**, (2017).

355.   Dueber, J. E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).

356.   Delebecque, C. J., Lindner, A. B., Silver, P. A. & Aldaye, F. A. Organization of intracellular reactions with rationally designed RNA assemblies. *Science (80-.).* **333**, 470–474 (2011).

357.   Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* **31**, 335–341 (2013).

358.   DeLoache, W. C., Russ, Z. N. & Dueber, J. E. Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. *Nat. Commun.* **7**, (2016).

359.   Hammer, S. K. & Avalos, J. L. Harnessing yeast organelles for metabolic engineering. *Nat. Chem. Biol.* **13**, 823–832 (2017).

360.   Schuster, B. S. *et al.* Controllable protein phase separation and modular recruitment to form responsive membraneless organelles. *Nat. Commun.* **9**, 1–12 (2018).

361.   Kentner, D., Thiem, S., Hildenbeutel, M. & Sourjik, V. Determinants of chemoreceptor cluster formation in Escherichia coli. *Mol. Microbiol.* **61**, 407–417 (2006).

362.   Thiem, S., Kentner, D. & Sourjik, V. Positioning of chemosensory clusters in E. coli and its relation to cell division. *EMBO J.* **26**, 1615–1623 (2007).

363.   Yu, X. *et al.* WebSDA: A web server to simulate macromolecular diffusional association. *Nucleic Acids Res.* **43**, W220–W224 (2015).

5. Materials and methods

The following Materials and methods paragraph is mostly adapted from the manuscript Romano, Baumschlager *et al.*[250], currently under revision, for which I had myself written the original text for the corresponding Materials and methods parts.

5.1 Strains, media and reagents

The strains used in this study are listed in Supplementary Table 6. The cultures were always grown in autoclaved Tryptone Broth (TB; 10 g l^{-1} Tryptone, 5 g l^{-1} NaCl, 1 mM NaOH) except for Figure 19-21, 48 and 49, for which autoclaved Lysogeny Broth (LB) was used, and Figure 18 and 45, where Tethering Buffer (10 mM potassium phosphate, 0.1 mM EDTA, 1 mM L-methionine and 10 mM sodium lactate; pH 7.0) was used. In experiments in which plasmids had to be maintained, the medium was supplemented with either 34 μg ml^{-1} chloramphenicol or 100 μg ml^{-1} ampicillin and 34 μg ml^{-1} chloramphenicol, or kanamycin 50 μg ml^{-1} (Sigma-Aldrich Chemie GmbH). Rifampicin and doxycycline were purchased from Sigma-Aldrich Chemie GmbH.

5.2 Constructions of plasmids

Below, I describe how I performed the clonings of all the constructs described in this thesis. The sequences of the primers can be found in the Supplementary Table 7 and 8, while the list of all vectors used and constructed in this thesis is shown in Supplementary Table 9.

5.2.1 Library of BLADE constructs and controls

The fusion VVD-AraC proteins FP1-5 were first subcloned into pDK12[361] using the NcoI and NotI restriction sites. The *vvd* gene carrying the N56K and C71V mutations and coding for a VVD protein missing the first 36 amino acids was PCR-amplified out of plasmid pGAVPO (gift from Yi Yang; East China University of Science and Technology). The *araC* fragments were amplified from pBAD33. To clone the fusions, a two-step protocol was followed. In the first step, the two parts were separately PCR-amplified. After the purification of the PCR products, the two fragments were fused together in the second PCR step and then cloned into pDK12 with NcoI and NotI restriction enzymes, yielding plasmids pDK12(FP1-5).

Next, the DNA sequences coding for the fusion proteins FP1-FP5 were PCR-amplified using primers 1 and 2 and cloned into pBAD33 (previously deprived of AraC via PCR using primers 3 and 4, yielding the negative control plasmid pReporter_only) linearized with ClaI via NEBuilder® HiFi DNA Assembly (New England Biolabs). The J23101 promoter was included in the forward primer. The obtained pBAD33-derived plasmids are called pBLADE(FP1-5).

The *mCherry* gene codon-optimized for expression in *E. coli* was synthesized (IDT) and cloned into pBLADE with SacI and HindIII restriction enzymes. I subsequently generated pBLADE(FP7)-mCherry by inserting an additional GS linker between VVD and AraC with a site-directed mutagenesis protocol using pBLADE(FP4)-mCherry as a template with primers 5 and 6. The primers,

designed with the QuikChange Primer Design website, were used to amplify the plasmid. The PCR reactions contained 8% of DMSO to allow proper annealing of the primers to the template DNA. Similarly, pBLADE(FP6)-mCherry was generated by removal of the GS linker from the same template with the same protocol previously described, using primers 7 and 8.

pBLADE(FP8)-mCherry was generated by the removal of the endogenous disordered region of AraC by amplification and re-circularization of pBLADE(FP6)-mCherry plasmid with primers 9 and 10 via NEBuilder® HiFi DNA Assembly (New England Biolabs). To construct the positive control with full-length AraC, pBLADE was linearized with ClaI, the *araC* gene was PCR-amplified with primers 11 and 12 and then cloned into the backbone via NEBuilder® HiFi DNA Assembly (New England Biolabs), yielding pBLADE(AraC$_{WT}$)-mCherry.

The mutations and deletions leading to promoters J23101* and J23101** (Figure 12) generated spontaneously during the growth of bacterial cultures transformed with pBLADE(FP4)-mCherry. pBLADE(FP4*)-mCherry and pBLADE(FP4**)-mCherry were then amplified with primers 13 and 14. All the other fusion constructs were amplificated with primers 15 and 16 and inserted in these backbones via NEBuilder® HiFi DNA Assembly (New England Biolabs), yielding pBLADE(FP1*/FP8*)-mCherry and pBLADE(FP1**/FP8**)-mCherry. The promoters were cloned into pBLADE(AraC$_{WT}$)-mCherry by overlapping PCR with primer pairs 17-18, and 29-30 via NEBuilder® HiFi DNA Assembly (New England Biolabs) yielding pBLADE(AraC$_{WT}$*/AraC$_{WT}$**)-mCherry.

5.2.2 Library of genes with unknown or poorly known function

The 39 genes were PCR-amplified from genomic DNA isolated from E. coli MG1655 using the primer pairs listed in Supplementary Table 8. The backbone pBLADE(FP6**) was amplified with primers 83 and 84 to insert the first 16 genes, and with primers 85 and 86 to insert the others. The primer pair was changed to maintain start and stop codon on the plasmid backbone.

To create the library with the fusion to sfGFP, the first 16 genes in the list in Supplementary Table 8 were amplified with primers that included a GS linker at their N or C-terminus, and cloned in the backbone pBLADE(FP6**)-sfGFP amplified with primers 87 and 88 (sfGFP at the N-terminal fusions) and 89 and 34 (sfGFP at the C-terminal fusions). For the remaining genes, primers 38 and 90 (N-terminal fusions) and 85 and 91 (C-terminal fusions) were used.

5.2.3 MinDΔ10, RodZ and MreB

The DNA sequence coding for MinDΔ10 was amplified out of pBDV-13[219] with primers 39 and 40 and cloned via NEBuilder® HiFi DNA Assembly (New England Biolabs) into pBLADE(FP4**)-mCherry previously amplified with primers 41 and 42, yielding pBLADE-MinDΔ10. The *rodZ* and *mreB* genes were PCR-amplified from genomic DNA isolated from *E. coli* MG1655 using primers 43-44 and 45-46 respectively. Cloning into pBLADE(FP4**) was achieved via NEBuilder® HiFi DNA Assembly (New England Biolabs) after amplification of the plasmid backbone with primers 33 and 34, yielding pBLADE-RodZ and pBLADE-MreB, respectively.

5.2.4 Plasmids for co-transformation

To clone pBAD33-mCherry, the same ribosome binding site of P_{BAD} from pBLADE (AGGAGG) was inserted with primers 19 and 20 via NEBuilder® HiFi DNA Assembly (New England Biolabs). Then, the *mcherry* gene was cloned into the plasmid with the SacI and HindIII restriction enzymes. pBLADE$^{ONLY\_C}$ (pBAD33 backbone deprived of the P_{BAD} promoter with BLADE FP6**) was cloned via NEBuilder® HiFi DNA Assembly (New England Biolabs) by amplification of pBLADE(FP6**)-mCherry with primers 47 and 48, followed by digestion of the linearized plasmid with SacI and ligation.

pBLADE$^{ONLY\_A}$ (pTrc99a backbone deprived of the pTrc promoter with BLADE FP6**) and SiMPl_C-BLADE were obtained via PCR-amplification of pTRC99a and SiMPL_C respectively with primers 21 and 22 followed by insertion of the FP6** fragment (amplified with primers 23 and 24) downstream of the *lacI* gene via NEBuilder® HiFi DNA Assembly (New England Biolabs). Then, the pTrc promoter was removed and a T2 terminator sequence was inserted by amplifying the backbone with primers 25 and 26 to isolate BLADE from the *lacI* gene. BLADE was then amplified with primers 27 and 28 to insert 100 bp of spacing from the T2 terminator via NEBuilder® HiFi DNA Assembly (New England Biolabs).

pCAM was cloned by removal of P_{BAD}-mCherry from pReporter_only through linearization and following self-circularization with primers 49 and 50. The *mcherry* gene was cloned into SiMPl_N with SacI and HindIII restriction enzymes, yielding to SiMPl_N-mCherry.

5.2.5 Plasmids with different I_1I_2, multiple copies of I_1I_2 and without O_1O_2

To obtain pBLADE(FP6*)_I_1rev_I_2-mCherry, pBLADE(FP6*)_I_2rev_I_2-mCherry, pBLADE(FP6*)_I_1rev(no-35)_I_2-mCherry, and pBLADE(FP6*)_I_1_I_2rev-mCherry, the plasmid pBLADE(FP6*)-mCherry was amplified with overlapping PCR with primers 69-70, 71-72, 73-74 and 53-54 respectively, via NEBuilder® HiFi DNA Assembly (New England Biolabs).

The mutation in the I_2 half-site was inserted with primers 75 and 76, designed with the QuikChange Primer Design website, using pBLADE(FP6*)-mCherry as a template, to obtain pBLADE_I_2*-mCherry.

To insert the multiple copies of the I_1I_2 half-sites, the plasmid pBLADE(FP6*)-mCherry was amplified with primers 77 and 78, while the inserts were amplified with primers 79 and 80 from pre-existing plasmids harbouring multiple copies of the I_1I_2 half-sites. The plasmids pBLADE_nI_1I_2-mCherry, where n are numbers between 1 and 4, were obtained via NEBuilder® HiFi DNA Assembly (New England Biolabs). The plasmid pBLADE_I_2*_7I_1I_2-mCherry was obtained by amplification of plasmid pBLADE_I_2*-mCherry with primers 77 and 78, while the insert was amplified with primers 79 and 80 from a pre-existing plasmid harbouring 7 copies of the I_1I_2 half-sites via NEBuilder® HiFi DNA Assembly (New England Biolabs).

The O_1O_2 half-sites have been removed by amplification and following circularization of pBLADE(FP6*)-mCherry and pReporter_only with primers 81 and 82 via NEBuilder® HiFi DNA Assembly (New England Biolabs), yielding to pBLADE$^{\Delta O1O2}$ (FP6*)-mCherry and pReporter_only ΔO1O2, respectively.

5.2.6 AraC$_{DBD}$-sfGFP

For the fusion of sfGFP to the C-terminal of BLADE FP6 and AraC, the pBLADE(FP6*)-mCherry, pBLADE(AraC$_{WT}$*)-mCherry and pBAD(AraC$_{WT}$)-mCherry backbones were amplified with primers 55 and 56, the *sfgfp* gene was amplified from pBLADE(FP6*)-sfGFP with primers 57 and 58, that carried a GS linker. The cloning, performed with NEBuilder® HiFi DNA Assembly (New England Biolabs), yielded pBLADE((FP6-sfGFP)*)-mCherry, pBLADE((AraC$_{WT}$-sfGFP)*)-mCherry, and pBAD(AraC$_{WT}$-sfGFP)-mCherry, respectively.

The C108A mutation in VVD was inserted using primers 59 and 60, designed with the QuikChange Primer Design website, with pBLADE((FP6-sfGFP)*)-mCherry as a template. The PCR reaction, containing 8% of DMSO, yielded to pBLADE((VVD^{C108A}-AraC-sfGFP)*)-mCherry.

5.2.7 Probes for EMSAs

The DNA fragments for the EMSA containing the I$_1$I$_2$, I$_1$ or I$_2$ sequences were obtained with primers 63-64, 65-66 and 67-68, respectively. The primers 63, 65 and 67 were tagged with HEX at the 5'. They were annealed by mixing of each pair in NEB buffer 2 and heating up the mixture to 98°C. After 2 minutes, the heat block was switched off while the sample was left in there, in order to cool the samples gradually down.

5.2.8 Other clonings

The *ccdb* gene was amplified from pDONR221 with primers 61 and 62 and cloned into pBLADE(FP6*)-mCherry amplified with primers 33 and 37 via NEBuilder® HiFi DNA Assembly (New England Biolabs), yielding to pBLADE-CcdB.

The DNA sequence coding for eYFP-MinD was PCR-amplified out of pSR-4[362] with primers 31 and 32 and cloned into pBLADE(FP6*) via NEBuilder® HiFi DNA Assembly (New England Biolabs) after having amplified the backbone (pBLADE(FP6*)-mCherry) with primers 33 and 34, yielding pBLADE-eYFP-MinD.

The *sfgfp* gene was PCR-amplified with primers 35 and 36 from plasmid pHR-scFv-GCN4-sfGFP-GB1-NLS-dwPRE (gift from Ron Vale; Addgene plasmid # 60906; http://n2t.net/addgene:60906; RRID:Addgene 60906) and cloned in pBLADE(FP6*)-mCherry via NEBuilder® HiFi DNA Assembly (New England Biolabs) after amplification of the backbone with primers 37 and 38 via NEBuilder® HiFi DNA Assembly (New England Biolabs), yielding pBLADE(FP6*)-sfGFP.

For protein purification, the BLADE FP6 construct was PCR-amplified with primers 51 and 52 from pBLADE(FP6*)-mCherry and cloned into pET28a with BamHI and NotI restriction enzymes, yielding pET28a-FP6. AraC was extracted from pBAD33 and cloned into pET28a with BamHI and HindIII restriction enzymes, yielding pET28a-AraC. These two clonings were performed by Navaneethan Palanisamy.

All PCRs were performed using the Phusion Flash High Fidelity PCR Master Mix (Thermo Scientific). Oligonucleotides were ordered at Sigma Aldrich. The cloning was performed using chemically competent *E. coli* TOP10 cells (Thermo Scientific).

5.3 Bacterial growth

The cultures were handled under safe red light whenever containing light-sensitive constructs. The cultures were incubated overnight in TB or LB medium and grown at 37°C in an incubator shaking at 250 rpm, in black plastic tubes (Argos Technologies LiteSafe® 15 ml) if containing light-sensitive samples, in transparent glass tubes otherwise. The following morning, the cultures were diluted to $OD_{600} = 0.1$ and let grow until $OD_{600} = 0.4$. Half of the culture was then transferred in transparent glass tubes and induced either with blue light or with arabinose for 4 hours.

5.4 Light illumination systems

To illuminate the glass tubes in the shaker, six high-power 460 nm LEDs type CREE XP-E D5-15 (LED-TECH.DE) were used (Supplementary Fig. 16). The LEDs were connected to a power supply (Manson HCS-3102) that allowed to tune the voltage, hence the light intensity. The light intensity reaching the cultures was 5 W/m^2 as measured with a LI-COR LI-250A Light Meter.

For the bacterial photography, and the induction of the library of genes with unknown or poorly defined function, I used a custom-made light box with, among others, 6 blue (455 nm) LEDs (Supplementary Fig. 28). To avoid the generation of a blurred image in the bacteriograph, all the LEDs except for the one in the centre were obscured with coloured tape. The average light intensity reaching the plate was 15W/m^2 with 6 LEDs and 1.3 W/m^2 with one LED.

5.5 Flow Cytometry

The fluorescence was measured using the LSR Fortessa flow cytometer (BD Biosciences). Samples were centrifuged at 4000 g for 4 min to remove the glycerol-containing solution, then the pellets were resuspended in PBS. Data analysis was performed using the open-source FCSalyzer software. This software was also used to calculate the mean fluorescence and the coefficient of variation (CV) for each sample.

The mCherry fluorescence was excited with a 561 nm laser (50 mW), and emission was detected using a 610/20-nm filter pass (PMT voltage set to 750 V). The GFP fluorescence was excited with 488 nm laser (100 mW), and emission was detected using a 530/30-nm filter pass (PMT voltage set to 405 V). A forward scatter height (FSC-H) threshold of 1,400 was used to gate for living cells and eliminate debris. 10^5 events per sample were recorded for each experiment. The cell density of the samples was manually regulated by the addition of PBS in order to have less than $2*10^4$ events/s recorded by the machine. To compensate any variable that can alter the measurement of the fluorescence by the flow cytometer, each experiment was normalized with the fluorescence value of the negative control grown the same day of the experiment.

5.6 Characterization of the FP1-FP8 VVD-AraC(DBD) fusion constructs

Chemically competent *E. coli* MG1655 cells were transformed with pBLADE(FP1*/FP8*)-mCherry, pBLADE(FP1**/FP8**)-mCherry, pReporter_only (negative control), or pBLADE(AraC$_{WT}$*/AraC$_{WT}$**)-mCherry (positive controls) or co-transformed with pBLADE$^{ONLY\_A}$ (based on pTrc99a) and pBAD33-mCherry, pBLADE(FP6**)-mCherry and pTrc99a, or SiMPl^K_C-BLADE and SiMPl^K_N-mCherry. Overnight cultures of cells transformed with the FP1-FP8 fusions were diluted to $OD_{600} = 0.1$, let grow in the dark to $OD_{600} = 0.4$ and split into two cultures, one of which was kept in the dark and one of which was illuminated for 4 hours. The overnight culture of the negative control was diluted to $OD_{600} = 0.1$, and let grow for the same amount of time as all other cultures (circa 5 h 30 min). The overnight cultures of the positive controls were diluted to $OD_{600} = 0.1$, let grow to $OD_{600} = 0.4$ and split into two cultures, one of which was left without arabinose and one of which was induced with 0.1% arabinose for 4 hours.

After the induction time, 200 µl of each sample were collected, mixed with 200 µl of a transcription and translation inhibition solution (500 µg ml^{-1} rifampicin and 50 µg ml^{-1} doxycycline in phosphate buffered saline) and incubated in the dark 90 min at 37°C with 110 rpm shaking. This protocol allows obtaining a full maturation of almost all the mCherry proteins translated at the end of the induction time[255]. After the incubation with the inhibitor, samples were diluted 1:1 with 60% glycerol and frozen at -80°C.

5.7 Dynamic control of gene expression

The overnight cultures transformed with pBLADE(FP6**)-mCherry, pReporter_only (negative control) and pBAD33-mCherry were diluted in TB to $OD_{600} = 0.05$ in dark tubes and glass transparent tubes, respectively, and let grow until $OD_{600} = 0.15$. 200 µl of each sample were collected, mixed with 200 µl of a transcription and translation inhibition solution (500 µg ml^{-1} rifampicin and 50 µg ml^{-1} doxycycline in phosphate buffered saline), incubated in the dark 90 min at 37°C with 110 rpm shaking, diluted 1:1 with 60% glycerol, and frozen at -80°C. The rest of the culture was transferred in transparent glass tubes: the culture with BLADE was illuminated with blue light as described (see "Light illumination systems") for 2 h, while the culture with pBAD33 was split into three different tubes and induced with different arabinose concentrations for 2 h. Then, aliquots from all the samples were taken and frozen with 60% glycerol 1:1. The remaining culture transformed with pBLADE(FP6*)-mCherry was diluted to $OD_{600} = 0.15$ with pre-warmed TB and transferred to a dark tube, while the cultures transformed with pBAD33-mCherry were centrifuged at 6000 *g* for 4 min and resuspended with the same volume of TB. The centrifugation and resuspension steps were repeated a second time to further remove the arabinose from the medium. Then, the cultures were transferred to another glass tube. All the cultures were subjected to a total of three cycles.

5.8 Measurement of the kinetics of BLADE- and AraC- mediated mCherry expression

Chemically competent *E. coli* MG1655 cells were transformed with pBLADE(FP6*)-mCherry, pReporter_only and pBAD33-mCherry. The overnight cultures were diluted to $OD_{600} = 0.1$ and let grow in the dark to $OD_{600} = 0.4$. 100 µl of each sample were collected, mixed with 100 µl of a transcription and translation inhibition solution (500 µg ml^{-1} rifampicin and 50 µg ml^{-1} doxycycline in phosphate-buffered saline), incubated in the dark 90 min at 37°C with 110 rpm shaking, diluted 1:1 with 60% glycerol, and frozen at -80°C. Then, the culture transformed with pBLADE was split into 4 tubes, of which 3 were induced with blue light of different intensities and one was kept in the dark. The cultures transformed with pBAD33-mCherry were split into 4 cultures, of which 3 were induced with different arabinose concentrations and one was kept without arabinose. Every hour for 6 hours, 100 µl of each sample was collected, mixed with 100 µl of the transcription and translation inhibition solution, incubated in the dark 90 min at 37°C with 110 rpm shaking, diluted 1:1 with 60% glycerol, frozen at -80°C and subsequently analyzed with the flow cytometer.

5.9 Light intensity titration

Chemically competent *E. coli* MG1655 cells were transformed with pBLADE(FP6**)-mCherry or pReporter_only. The overnight culture of the cells transformed with pBLADE(FP6**)-mCherry was diluted and split into 5 independent cultures, each of which was induced with blue light of different intensity (which was tuned adjusting the voltage in the power supply connected to the LEDs) for 4 hours. The overnight culture of the cells transformed with pReporter_only was diluted and grown in the dark for 4 hours. 200 µl of each sample were then collected, mixed with 200 µl of the transcription and translation inhibition solution, incubated in the dark 90 min at 37 °C with 110 rpm shaking, diluted 1:1 with 60% glycerol, frozen at -80 °C and subsequently analyzed with the flow cytometer.

5.10 Bacterial photography

For the bacteriograph of Blade Runner movie poster, chemically competent *E. coli* MG1655 cells were transformed with pBLADE(FP6*)-sfGFP. The overnight culture was diluted in LB to $OD_{600} = 0.1$ and grown for approximately 6 hours. A 96-well lid (12.7 x 8.5cm) was filled with 30-40 ml of 1% LB-agar and let solidify. 1 ml of the culture was then mixed with 9 ml of 0.4% agar at 42 °C (measured with infrared thermometer TFA Dostmann (Wertheim-Reicholzheim, Germany)) and plated on top of the solidified agar in the 96-well lid. The plate was covered with a transparent plexiglass sheet (12.5 x 8.3 cm) with the Blade Runner movie poster sticker.

For the bacteriograph of Michelangelo's Creation of Adam and the BLADE protein model structure, the same protocol was followed, but this time 350-400 ml of 1 % LB-agar were poured in an 18 x 15 cm case, for a total photomask area of 15.5 x 9 cm. The plates were then placed in a 37 °C incubator under blue light overnight. The next morning, the plates were imaged with a Zeiss Axio Zoom.V16 stereo zoom microscope equipped with PlanNeoFluar Z 1.0x objective, zoom 0.7x, AxioCam MR R3 camera, and the 38 HE filter set (Ex BP 470/40, FT 495, Em BP 525/50; sfGFP).

The Blade Runner bacteriograph is composed of 110 tiles, while Michelangelo's Creation of Adam and the BLADE protein model structure are composed of 160 tiles each. The composed pictures have been stitched together with the ZEN Blue software.

5.11 DIC and fluorescence microscopy

5 µl of the bacterial culture were applied to a thin agarose pad composed of 1% agarose for microscopy at room temperature and of 1% agarose and 0.1% LB in Tethering buffer (10 mM potassium phosphate, 0.1 mM EDTA, 1 mM L-methionine and 10 mM sodium lactate; pH 7.0) for long-term microscopy at 37°C. Images were acquired on a Zeiss Axio Observer Z1/7 fluorescence microscope equipped with an Alpha Plan-Apochromat 100x/1.46 Oil DIC (UV) M27 objective, filter sets 38 HE (Ex BP 470/40, FT 495, Em BP 525/50; sfGFP), 108 HE (Ex BP 423/44, DBS 450+538, Em DBP 467/24+598/110; MM 4-64), 96 HE (Ex BP 390/40, FT 420, Em BP 450/40; DAPI), 64 HE (Ex BP 587/25, FT 605, Em BP 647/70; mCherry) and an Axiocam 506 Mono camera.

To image the library of genes with unknown or poorly defined function in a fast and efficient way, the samples (circa 5 µl) were applied to a 96-well lid, which was filled with 1% agarose, let solidify and covered with two 75 x 50 mm glass coverslips (Carl Roth GmbH, Karlsruhe). Before imaging, samples were incubated for 5 min with 1.2 µg ml^{-1} of the membrane dye MM 4-64 (AAT Bioquest Sunnyvale, CA) and 0.5 µg ml^{-1} of 4′,6-diamidino-2-phenylindole (DAPI, Sigma-Aldrich Chemie GmbH).

The induction of gene expression in selected cells within a population of MG1655 cells transformed with pBLADE(FP6**)-sfGFP was performed on a Zeiss LSM 800 confocal microscope. An area of 6.4 µm^2 was illuminated with a 488 nm diode laser (10 mW) at 0.1% intensity, with a frame average of 8, resulting in 0.36 µs of light per pixel. The illumination was given in pulses of 5 min for a duration of 3 h.

5.12 FRAP

FRAP were performed on a Zeiss LSM 800 confocal microscope. An overnight culture of MG1655 cells transformed with pBLADE((FP6-sfGFP)*)-mCherry was diluted in the morning in fresh TB medium to OD$_{600}$ = 0.1, and grown until it reached OD$_{600}$ = 0.4. 5 µl of the culture were then applied to a thin 1% agarose pad. After selecting a cell with a bright fluorescent spot, the area to bleach within the cell (whole spot) was manually set (the whole focus) and bleached with a single 1 s pulse of a 488 nm diode laser (10 mW) at 50% intensity. An image in the GFP channel (filter set 38 HE: Ex BP 470/40, FT 495, Em BP 525/50) was taken 15-min post bleaching to measure the recovery of the fluorescent signal.

An overnight culture of MG1655 cells transformed with pBLADE(FP6**)-yebY-sfGFP, was induced as described in paragraph 5.14, and then I followed the same bleaching and imaging protocol previously described. The image in the GFP channel was taken 7 minutes after bleaching.

5.13 Induction of rodZ in KC717 cells

Strain KC717 (kind gift of KC Huang, Stanford University) was grown in LB medium supplemented with 0.2% arabinose (to maintain the cells rod-shaped) during the transformation of chemically competent KC717 cells and DNA extraction procedures. The blue light and arabinose induction were performed as described above. The recovery phase of the cultures induced with arabinose was performed by centrifuging it at 6000 g for 4 min and resuspending it with the same volume of LB. The centrifugation and resuspension steps were repeated a second time and the culture was then

diluted to $OD_{600} = 0.1$. The recovery phase of the culture transformed with pBLADE$^{ONLY\_C}$ (based on pBAD33) was performed by dilution of the culture exposed to blue light $OD_{600} = 0.1$ and incubation in the dark.

5.14 Light-induced expression of genes with unknown or poorly defined function

Chemically competent MG1655 cells were transformed with the 117 pBLADE plasmids constituting the library to characterize the 39 genes with unknown or poorly defined function. Cultures were grown in the dark overnight in LB in non-treated 96-well plates (VWR, Radnor, PA) at 37°C with 110 rpm shaking. The following morning a Scienceware® replicator (96-well; Merck KGaA, Darmstadt, Germany) was used to transfer about 5 µl of each culture into a fresh 96-well plate with 145 µl of TB in each well. The diluted cultures were incubated at 18°C with 110 rpm shaking for 1 h in the dark and then they were induced with blue light (455 nm, 5 W/m^2) for 4 h.

5.15 Computational prediction of function and localization of 39 genes with unknown or poorly defined function

We randomly selected 34 genes out of the y-ome, defined as the group of genes lacking to date experimental evidence of function[113]. We manually checked that the selected genes were not mentioned in any publication using several search engines. As controls for our pipeline, we included 5 genes for which some information was available (*ydaT*[275]; *ydiY*[276]; *ycbK* (MepK)[277]; *yehS*[278]; and *yebE*[279,280]). We retrieved the amino acid sequences of the proteins encoded by all 39 genes in FASTA format and submitted them to the following webservers: Argot2.5[284] PANNZER2[208], DeepGoPlus[209] and Phyre2[206]. The consensus localization and function were calculated as the output provided by at least 2/4 prediction tools.

5.16 Measurement of bacterial growth

The growth curves of the cells transformed with the library of genes with unknown or poorly defined function were measured on a Synergy H4 Hybrid plate reader (BioTek) in 96-well plates. The cultures were grown in the dark overnight in LB in a 96-well plate at 37°C with 110 rpm shaking. The following morning the cultures were diluted to OD_{600} 0.1 in a fresh 96-well plate with 120 µl of LB. To prevent evaporation of the medium, also the unused wells of the plate were filled with the same amount of LB and the lid was sealed with parafilm. The plate was then illuminated with blue light (460 nm) and the OD_{600} of the culture was measured every 2 min in constant shaking for 20 h. The overnight cultures of three selected members of the library (y*daT, ydhL, yhcF*) were diluted in LB to OD_{600} 0.1 and grown until they reached OD_{600} 0.4. Each culture was then split into two tubes, one of which was kept in the dark and one of which was illuminated for 4 h at 37°C with shaking at 250 rpm. The OD_{600} was measured at the end of the growth with the OD600 DiluPhotometer™ (Implen).

5.17 BLADE (FP6) expression and purification (performed by Navaneethan Palanisamy)

Chemically competent *E. coli* Rosetta (DE3) cells carrying the pLysS plasmid were freshly transformed with pET28a-FP6 and cultivated overnight in LB medium supplemented with 50 µg ml^{-1} kanamycin. LB medium (1 l) containing kanamycin was inoculated using the pre-culture to obtain $OD_{600} = 0.1$. The culture was grown at 37 ^{o}C until $OD_{600} = 0.5$, after which 1 mM IPTG and 5 µM FAD were added, and the culture was grown for 16 hours at 18 ^{o}C under constant blue light. Cells were collected by centrifugation and the pellet was re-suspended in 30 ml of lysis buffer (50 mM potassium phosphate pH 8.0, 300 mM NaCl and 10 mM imidazole pH 8.0) supplemented with a cOmplete™ protease inhibitor cocktail tablet (Roche). Cell lysis was performed by sonication and the lysate was centrifuged at 20,000 rpm for 20 min at 4°C. The supernatant was then co-incubated with 1 ml of HisPur™ Ni-NTA Resin (Thermo Scientific) for 2 h at 4 ^{o}C.

Protein purification was performed by the gravity flow method. The bound proteins were washed twice with 5 ml of wash buffer (lysis buffer + 10 % glycerol + 20 mM imidazole) and finally eluted with 1.5 ml of elution buffer (50 mM potassium phosphate pH 7.5, 300 mM NaCl, 500 mM imidazole pH 8.0 and 10 % glycerol). The elution buffer was replaced with a storage buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl and 10 % glycerol) using an Amicon® Ultra-4 regenerated cellulose NMWL 10 kDa centrifugal filter unit (Merck). The protein was then stored as 50 µl aliquots at -80 ^{o}C. We verified that the purified protein could respond to light by measuring the absorption spectrum (Supplementary Fig. 12).

5.18 Electrophoretic mobility shift assays (performed by Navaneethan Palanisamy)

To visualize the binding of AraC to the I_1I_2 and I_1 half-sites, 10 µl reaction volume containing the DNA probes, tagged with HEX at the 5' (50 nM), and different concentration of the purified AraC protein in a buffer containing 20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 10 % glycerol and 0.1 mM EDTA at pH 8.0 was incubated for 30 min at RT, without or with 2 mM arabinose. The samples were then loaded onto 6% native polyacrylamide gels and subjected to electrophoresis for 40 minutes (I_1I_2 probe) or 45 minutes (I_1 probe) at 150 V at RT. Fluorescence was detected with Amersham Typhoon.

To visualize the binding of BLADE (FP6) to I_1I_2, I_1 and I_2 half-sites, the purified protein was incubated at 4 °C either 10 min under 50 W/m^2 blue light and then 1 hour in the dark, or 70 minutes in the dark. Then, 10 µl reaction volume containing the DNA probes, tagged with HEX at the 5' (50 nM) and different concentrations of the purified BLADE protein were mixed in a buffer containing 20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 10 % glycerol and 0.1 mM EDTA pH 8.0 and incubated for 15 minutes. The samples were then loaded onto 6% native polyacrylamide gels and subjected to electrophoresis for 45 minutes at 150 V and 4 °C. Fluorescence was detected with Amersham Typhoon.

5.19 Spectroscopy (performed by Maximilian Hörner)

The absorption spectrum of the FAD cofactor bound to VVD within BLADE (FP6) was measured exciting the sample in the 300-600 nm range using a Multiskan GO (Thermo Scientific) plate reader. The protein sample was incubated 4 days at 4 °C in the dark in a buffer solution (25 mM HEPES, 150 mM NaCl, 10% glycerol, 0.1% EDTA; pH 7.5) and then diluted to 0.5 mg ml^{-1} prior to the measurement of the absorption spectrum in the dark state. The same sample was then illuminated with blue light (455 nm; 50 W/m^2) for 5 min at room temperature and the absorption spectrum in the lit state was measured. The absorption spectrum of the blank (only medium) was subtracted from the dark and lit state spectra.

5.20 SEC (performed by Navaneethan Palanisamy)

Purified BLADE (FP6) was thawed and stored in complete darkness at 4 °C for 6 days. The sample (1 ml of protein with a concentration of 0.5 mg ml^{-1}) was loaded onto a Superdex™ 75 Increase 10/300 GL (GE Healthcare Lifesciences) column at 4 °C. The running buffer consisted of 20 mM HEPES-NaOH pH 7.5, 150 mM NaCl and 10 % glycerol, and the flow rate was adjusted to 0.25 ml min^{-1}. Dimerization of BLADE FP6 was triggered by incubating the protein under constant blue light (455 nm; 50 W/m^2) for 30 min at 4 °C, prior to injection. During the run, the column was either illuminated with constant blue light (460 nm; 8 W/m^2, lit sample) or kept in complete darkness (dark sample). Bovine serum albumin (BSA) and carbonic anhydrase (CA) were used as size markers at a concentration of 0.5 mg ml^{-1} each.

## 5.21 Model structure of the alternative I_1I_2 half-sites (performed by Mehmet Öztürk)

The model structure of the synthetic I_1I_2 half-sites (I_1rev_I_2, I_2rev_I_2, I_1rev(no-35)_I_2, and I_1_I_2rev) were generated by using the "DNA sequence to structure" webserver (http://www.scfbio-iitd.res.in/software/drugdesign/bdna.jsp). The DNA-binding domain of AraC was downloaded from the RSCB database (PDB id: 2K9S). Brownian dynamics (BD) simulations were applied by using the WebSDA[363] web server with default parameters to dock AraC DNA domain to alternative DNA sequences.

5.22 Quantification of cell length, width, and roundness

The cell length and width were calculated by first staining the cell with the membrane dye MM 4-64 (AAT Bioquest Sunnyvale, CA) to visualize the cell contour, and then manually measuring the long and short axes of the cell, respectively, using the straight-line 'Selection' tool of Fiji. At least 500 cells were measured for each sample. The histograms were generated in Excel by the Analysis ToolPak's Histogram option. The roundness was manually calculated with the oval 'Selection' tool on unstained cells, using Fiji. At least 200 cells were measured for each sample.

6. **Appendix**

6.1 Supplementary Figures
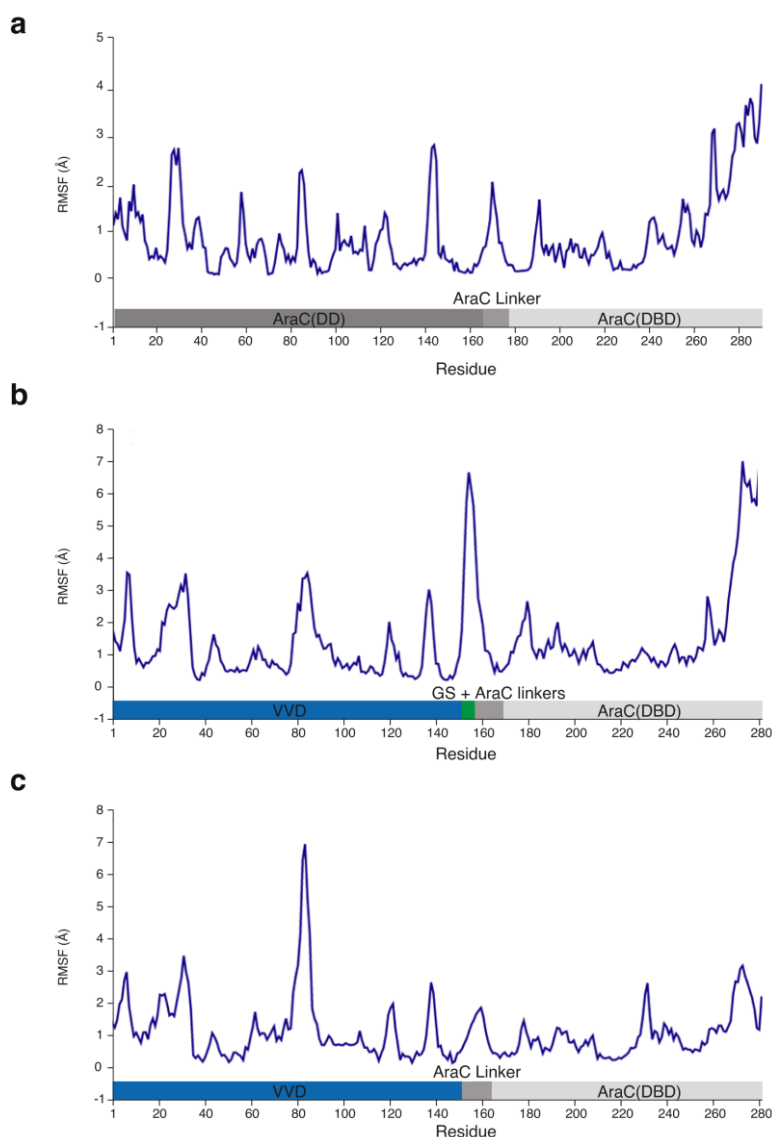

*Supplementary Figure 1 | Root mean square fluctuation (RMSF) values of all the residues of three different proteins. Output of CABS-flex 2.0 webserver after submission of the model structures of **a**, full-length AraC **b**, BLADE FP4 **c**, BLADE FP6. The high level of flexibility (7 Å) measured around the residue 150 of (**b**) is in correspondence with the flexible GS linker.*

|	*blade* gene	Promoter driving BLADE expression	P$_{BAD}$ promoter
pBLADE-mCherry	Yes (FP1-FP8)	J23101* or J23101**	Yes
pBLADE(AraC$_{WT}$*/)-mCherry**	No	J23101* or J23101**	Yes
pBLADE$^{ONLY\_C}$	Yes (FP6)	J23101**	No
pReporter_only	No	No	Yes
pBAD33-mCherry	No	No	Yes
pSiMPI^K_C-BLADE	Yes (FP6)	J23101**	No
pSiMPI^K_N-mCherry	No	No	Yes
pCAM	No	No	No
pTRC99a	No	No	No
pBLADE$^{ONLY\_A}$	Yes (FP6)	J23101**	No

Supplementary Figure 2 | Characteristics of some of the plasmids used in this work. *The list shows only a selection of the plasmids constructed and used in this study. A complete list of all plasmids is given in Supplementary Table 9*
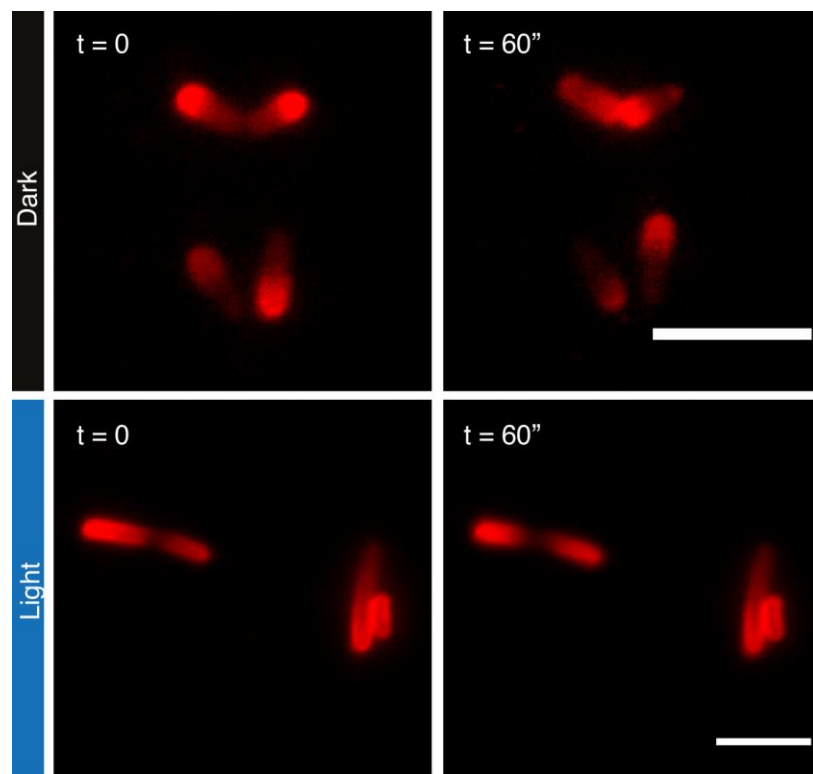
*Supplementary Figure 3 | The leakiness of pBLADE driven by J2101** promoter prevents mCherry-MinD oscillation visualization.* *E. coli MG1655 cells transformed with pBLADE(FP4**)-mCherry-MinD were illuminated with blue light (5 W/m^2) for 2 hours or kept in the dark for the same time and imaged at the fluorescence microscope. The images show two snapshots of time-lapse microscopy experiments at the indicated time points. Oscillations are already visible in the dark sample (first row), due to leakiness. 2 hours of blue light illumination (second row) led to significant mCherry-MinD overexpression that disrupted oscillations. The elongated phenotype is a consequence of MinD overexpression. Scale bar, 5 μm.*

a
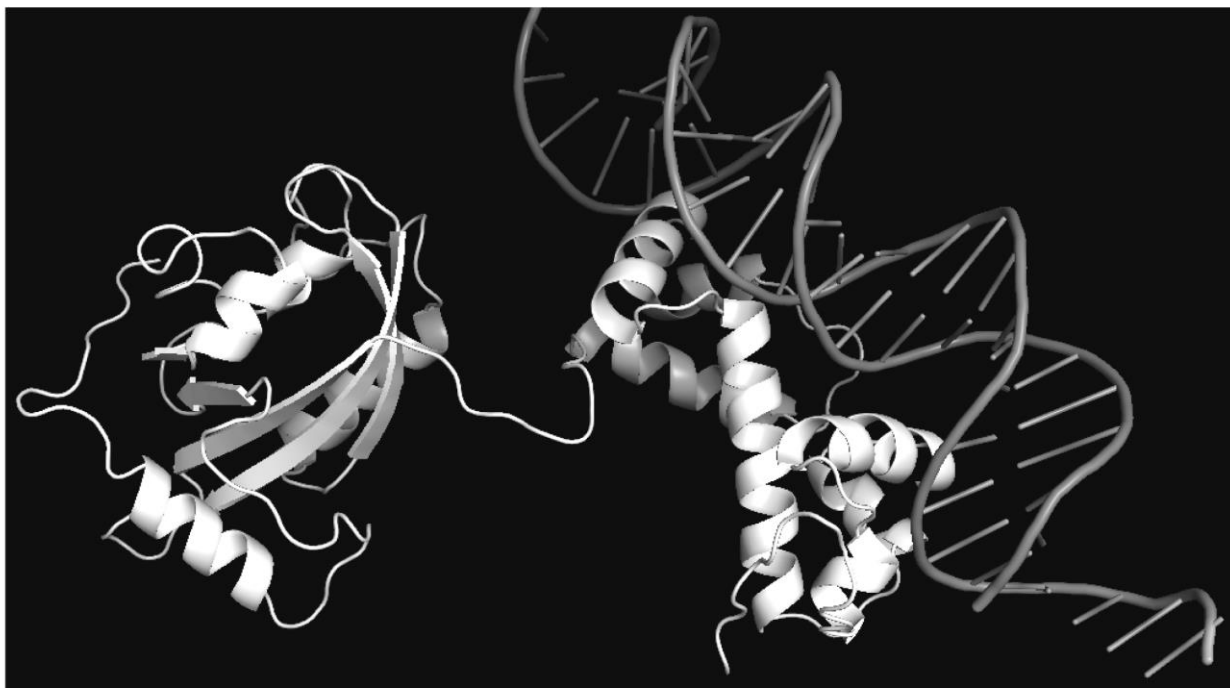
b
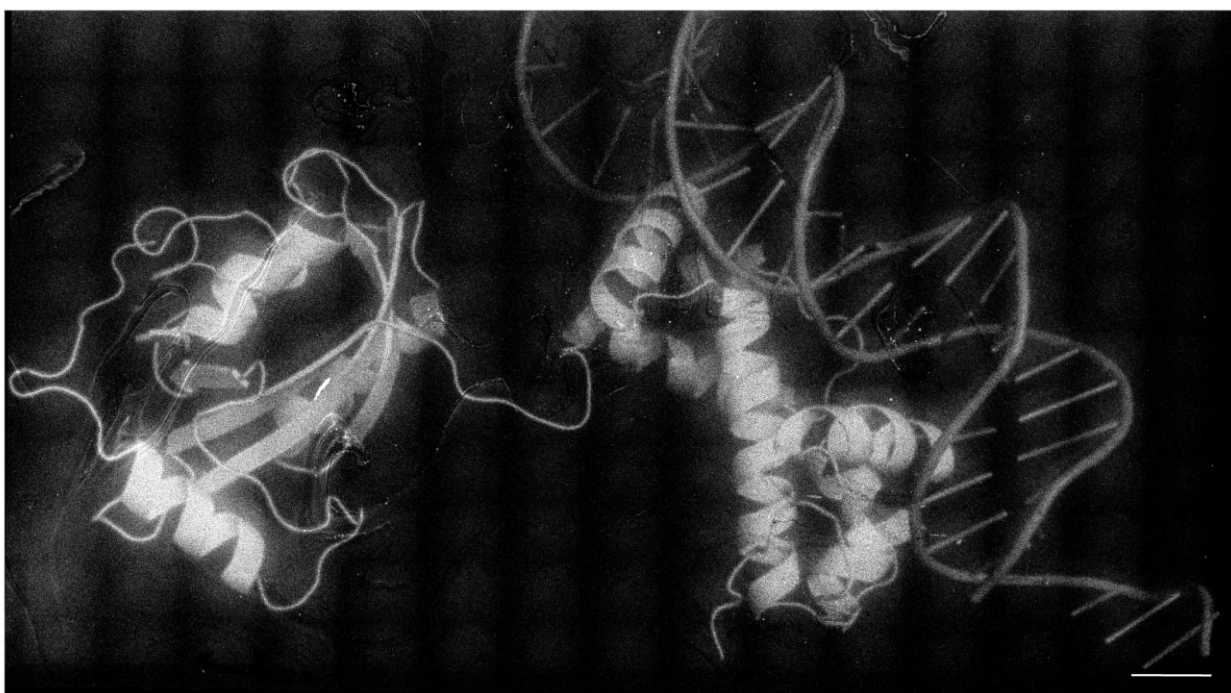
Supplementary Figure 4 | Reproduction of the protein model structure of BLADE, using BLADE to spatially activate gene expression. *a, Photomask used to produce the bacteriograph in b. **b**, Bacteriograph. A lawn of E. coli MG1655 cells transformed with pBLADE(FP6*)-sfGFP was grown overnight at 37 °C while being exposed to blue light through the photosmask in (**a**). 110 individual images were taken with a fluorescent microscope and stitched together via image analysis software. Scale bar, 1 cm.*
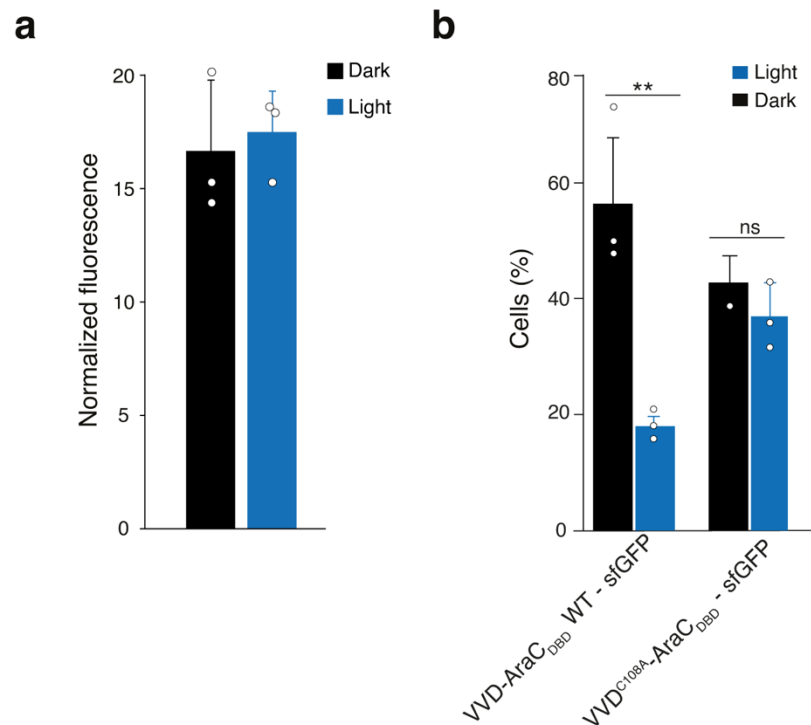
Supplementary Figure 5 | The mechanism of BLADE-mediated gene expression involves the formation of aggregates in the dark. a, *MG1655 cells transformed with pBLADE_I$_{2\ rev}$-mCherry grown 4 h either in the dark or under 460 nm light (5 W/m^2) illumination were analyzed by flow cytometry. The values were normalized to the mCherry fluorescence intensity measured in E. coli MG1655 cells transformed with pReporter_only, see Supplementary Fig. 2; dashed line).* ***b,*** *Quantification of the number of cells showing aggregates for the indicated constructs and conditions. Values represent mean ± s.d. of n=3 independent experiments. Values represent mean ± s.d. of n=3 independent experiments. Not significant (ns), p-value <0.05; double asterisk (**), p-value <0.01. P-values were calculated by the two-tailed, homoscedastic Student's t-test.*
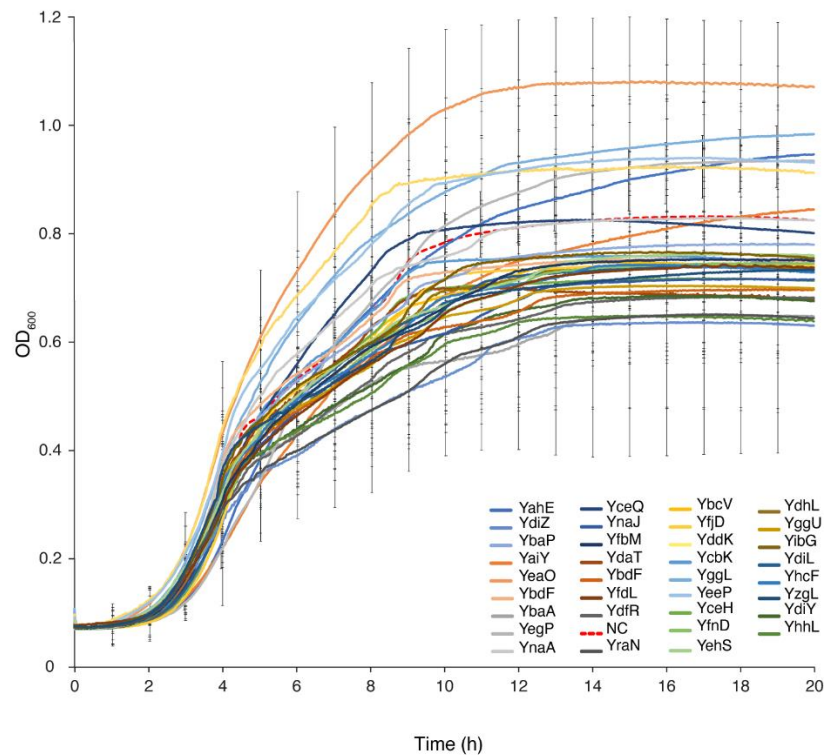
Supplementary Figure 6 | Most analyzed genes with unknown or poorly defined function do not alter the growth of bacterial cells. a, Growth curves of E. coli MG1655 cells transformed with pBLADE carrying the genes coding for the indicated proteins for which the two-tailed, homoscedastic Student's t test resulted in non-significant growth alteration from the cells transformed with the empty plasmid (NC). Values represent mean ± s.d. of n=3 independent experiments.

Supplementary Figure 7 | Custom-made lightbox used to illuminate 96-well plates. Left, front view. Right, bottom view.
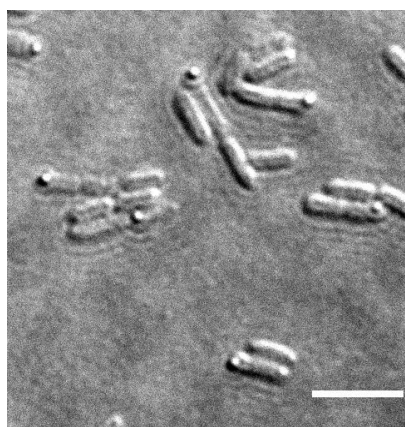
Supplementary Figure 8 | Overexpression of YebY generates swellings on E. coli cells. Representative image of E. coli MG1655 cells taken with DIC channel transformed with pBLADE-yebY has grown for 4 h under 455 nm light illumination (5 W/m^2). Swellings are visible in most of the cells. Scale bar, 5 μm.

6.2   Supplementary Tables

	FP1	FP2	FP3	FP4	FP5	FP6	FP7	FP8
Natural linker length (aa)	17	7	7 + 7	10	10	10	10	0
Synthetic linker length (aa)	0	0	0	2	29	0	4	0
Total linker length (aa)	17	7	14	12	39	10	14	0

Supplementary Table 1 | Type of linker and their length relative to the constructs of the library **represented in** *Figure 13a and Figure 15.*

	FP1*	FP2*	FP3*	FP4*	FP5*	FP6*	FP7*	FP8*	AraC$_{WT}$*
Uninduced state fluorescence	0.97 ± 0.07	1.27 ± 0.17	1.10 ± 0.08	0.98 ± 0.02	0.95 ± 0.25	1.15 ± 0.24	1.20 ± 0.28	0.93 ± 0.06	0.67 ± 0.29
Induced state fluorescence	7.66 ± 1.48	9.67 ± 2.72	7.73 ± 1.28	8.00 ± 1.44	2.81 ± 0.61	16.67 ± 2.49	14.26 ± 2.45	0.94 ± 0.06	37.10 ± 3.40
Fold change	7.87	7.61	7.03	8,1	2.96	14.48	11.90	1.00	55.68

	FP1**	FP2**	FP3**	FP4**	FP5**	FP6**	FP7**	FP8**	AraC$_{WT}$**
Uninduced state fluorescence	2.27 ± 0.05	2.68 ± 0.19	2.28 ± 0.22	1.71 ± 0.13	1.13 ± 0.19	2.37 ± 0.22	1.90 ± 0.14	1.05 ± 0.04	0.75 ± 0.08
Induced state fluorescence	18.07 ± 1.56	20.25 ± 1.36	21.03 ± 1.88	21.5 ± 1.98	6.90 ± 2.14	29.10 ± 1.85	20.13 ± 1.46	1.10 ± 0.06	39.47 ± 5.12
Fold change	7.97	7.57	9.24	12.56	6.13	12.30	10.60	1.06	52.28

Supplementary Table 2 | Quantification of the mCherry fluorescence levels induced by the BLADE library and the positive controls in their uninduced and induced state. *Top: constructs transcribed with J23101* promoter. Bottom: constructs transcribed with J23101** promoter. These values have been used to compose the graph of Figure 15.*

	pTrc99a-BLADE + pBAD33-mCherry	SiMPl^K_N-BLADE + SiMPl^K_C-mCherry	pBLADE-mCherry + pTrc99a	pBLADE-mCherry
Antibiotics in the medium	Chloramphenicol + Ampicilline	Kanamycin	Chloramphenicol + Ampicilline	Chloramphenicol
Dark state fluorescence	1.19 ± 0.07	1.17 ± 0.08	2.07 ± 0.62	2.37 ± 0.22
Light state fluorescence	16.30 ± 1.12	16.73 ± 1.68	31.27 ± 2.27	29.1 ± 1.85
Fold-change	13.7	14.26	15.13	12.30

Supplementary Table 3 | Quantification of the mCherry fluorescence levels of BLADE in a 2-plasmid system. *These values have been used to compose the graph of Figure 47.*

Protein name	PDB ID	Phyre 2 Modelling template function	Phyre 2 Coverage(%) / identity (%)	Argot 2.5 (Score: 0-500 weak, 500-1000 intermediate, >1000 strong) Molecular function	Biological process	Cellular component	Pannzer 2 (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong) Description	Biological process	Molecular function	Cellular component	DeepGO Plus (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong) Cellular component	Biological process
YahC	5A40, 5A43	Dual topology fluoride ion channel	47-61 / 14-13	Oxidoreductase activity (538)	Oxidation-reduction (1094)	Integral component of membrane (3741)	Inner membrane protein (0.43)	NA	NA	Integral component of membrane (0.43)	Plasma membrane (0.777)	NA
YahE	1GPR	Glucose permease-like	21 / 20	NA	NA	Integral component of membrane (325)	NA	NA	NA	NA	NA	Metabolic (0.316)
YaiY	NA	NA	NA	NA	NA	Integral component of membrane (1236)	Inner membrane protein yaiY (0.51)	NA	NA	Cell periphery (0.35), intrinsic component of membrane (0.30)	Cell periphery (0.904)	NA
YbaA	2OKQ	NA	99 / 100	Alkaline phosphatase (2171), hydrolyase (634)	Dephosphorylation (4173), methylation (1146)	NA	RNA signal recognition particle 4.5S RNA(0.48)	NA	NA	NA	NA	NA
YbcV	2HH8	NA	93 / 80	NA	Single-organism development (84)	Viral envelope (6085)	DUF1398 domain-containing protein (0.56)	NA	NA	NA	NA	NA
YcbK	1LBU, 5OPZ	Muramoyl-pentapeptide carboxypeptidase, l-ala d-glu endopeptidase	63-32 / 26-21	Proline-tRNA ligase (422), ATP binding (349)	Proteolysis(239), cell wall macromolecule catabolic process(98)	Integral component of membrane (562)	DUF882 domain-containing protein (0.66)	Proteolysis (0.61)	Peptidase (0.62)	NA	NA	NA
YceH	5U8J, 3BZ6	Membrane biogenesis, winged helix	79-37 / 87-46	N-acetyltransferase (1078)	NA	NA	UPF0502 protein YceH (0.83)	Oxidation-reduction (0.46)	Oxidoreductase (0.48)	Cytosol (0.58)	Cytosol (0.608)	NA
YceQ	NA	NA	NA	NA	NA	NA	DUF2655 domain-containing protein (0.96)	NA	NA	NA	NA	Regulation of cellular process (0.304)
YdaT	NA	NA	NA	Transcription factor sequence-specific DNA binding (144)	Regulation of transcription (182), DNA-templated (151)	Intracellular membrane-bounded organelle (205)	Phage regulatory protein (0.24)	NA	NA	NA	NA	NA

Protein name	Phyre 2			Argot 2.5 (Score: 0-500 weak, 500-1000 intermediate, >1000 strong)			Pannzer 2 (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong)				DeepGO Plus (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong)	
	PDB ID	Modelling template function	Coverage(%) / identity (%)	Molecular function	Biological process	Cellular component	Description	Biological process	Molecular function	Cellular component	Cellular component	Biological process
YdfR	NA	NA	NA	NA	Chordate embryonic development (324)	Microtubule cytoskeleton (102)	Qin prophage (0.80)	NA	NA	NA	Intracellular (0.325)	NA
YdhL	4DJA	Prokaryotic photolyase phrb	51 / 24	Oxidoreductase (384)	Oxidation-reduction process(1728)	Integral component of membrane (55)	Oxidoreductase (0.49)	NA	NA	NA	Intracellular (0.323)	Regulation of cellular process
YdiL	1S4K	Lambda repressor-like DNA-binding domain	99 / 56	DNA binding (3144)	NA	NA	Putative HTH domain DNA-binding protein (0.73)	NA	DNA binding (0.56)	NO	NA	Single-organism process (0.304)
YdiY	2IWV 3QRA 3NB3	Outer membrane porin	50-57-61 16-11-16	Receptor (1133)	Transport (543), cell adhesion (254)	Membrane (341)	Acid-inducible outer membrane protein (0.78)	NA	NA	NA	NA	NA
YdiZ	2LLZ	YjdK monomer	85 / 18	Endoribonuclease (17462)	RNA phosphodiester bond hydrolysis, endonucleolytic (20186)	Integral component of membrane (419)	Type V toxin-antitoxin system endoribonuclease antitoxin GhoS (0.79)	RNA phosphodiester bond hydrolysis, endonucleolytic (0.70)	Endoribonuclease (0.71)	NA	NA	Response to DNA damage (0.510) Response to stress (0.537)
YeaO	1E8C	MurD-like peptide ligases, peptide-binding domain	33 / 13	Methyltransferase (1366)	Methylation (5126)	Cytoplasm (995)	Cytoplasmic protein YeaO (0.76)	NA	NA	NA	NA	NA
YebE	2OU3, 2IXU	Tellurite resistance protein	57 / 17	ATPase (177)	NA	Plasma membrane (2902)	Inner membrane protein YebE (0.44)	NA	NA	Integral component of membrane (0.43), plasma membrane (0.35)	Plasma membrane (0.709)	NA
YebY	2MVO	Lipid binding protein	78 / 25	Transaminase (1805)	NA	Cytosol (477), membrane (210)	Putative histidinol-phosphate/aromatic aminotransferase and cobyric acid decarboxylase (0.89)	NA	Transaminase activity	NA	Cytosol (0.538)	NA
YegP	2K49	YegP-like	100 / 65	Phosphorus-oxygen lyase (117)	Cyclic nucleotide biosynthetic process (212)	Integral component of membrane (132)	UPF0339 protein YegP (0.80)	Double-strand break repair (0.76)	NA	NA	Intracellular (0.316)	NA

Protein name	Phyre 2			Argot 2.5 (Score: 0-500 weak, 500-1000 intermediate, >1000 strong)			Pannzer 2 (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong)				DeepGO Plus (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong)	
	PDB ID	Modelling template function	Coverage(%)/ identity (%)	Molecular function	Biological process	Cellular component	Description	Biological process	Molecular function	Cellular component	Cellular component	Biological process
YbaP	2G5G	Iron transport, membrane lipoprotein	45 / 12	Ligase (13706)	Regulation of translational fidelity (2521)	Integral component of membrane (1790)	Ligase (0.61)	NA	Ligase activity (0.63)	NA	NA	NA
YbdF	3H9X	DNA binding	86 / 28	Methyltransferase (118)	DNA repair (341)	NA	MmcQ/YjbR family DNA-binding protein	NA	DNA binding (0.56)	NA	Intracellular (0.334)	NA
YnaA	2CH7, 5XG2	Methyl-accepting chemotaxis, chromosome partition	98-98 10-12	Cytoskeletal protein binding (311)	Cell cycle process (81)	Integral component of membrane (147), cytoskeletal part (287)	Putative phage tail length tape measure protein (0.80)	Intermediate filament cytoskeleton organization (0.61)	NA	Integral component of membrane (0.41)	Intracellular (0.355)	Cellular component organization (0.300)
YddK	4U06	Membrane	95 / 11	NA	NA	Integral component of membrane (82)	Glycoprotein with ribonuclease inhibitor domain (0.80)	NA	Ligase (0.63)	NA	Cytosol (0.572)	NA
YeeP	3IEV	GTP-binding	63 / 27	GTP binding (7211)	Ribosomal large subunit assembly (5356)	Cytoplasm (1500)	Ferrous iron transport B family protein (0.47)	NA	GTP binding (0.66)	NA	NA	NA
YehS	3SIB	DNA binding	85 / 13	Glutamate synthase (NADPH) (627)	Oxidation-reduction (1712)	Integral component of membrane (353)	YehS (0.84)	NA	NA	NA	NA	NA
YfbM	6MK7	Membrane protein	50 / 9	NA	NA	NA	YfbM (0.24)	NA	NA	NA	NA	NA
YfdL	4MTM 6CL5	Membrane protein	76 / 25	Carbohydrate binding (128)	Carbohydrate metabolic process (245)	Extracellular region (245)	Putative RNA polymerase beta (0.71)	NA	NA	NA	NA	NA
YraN	1GEF 2EO0	Restriction endonuclease-like DNA binding	58-54 23-16	Nuclease (1551)	Nucleic acid phosphodiester bond hydrolysis (7594)	NA	UPF0102 protein (0.84)	Nucleic acid phosphodiester bond hydrolysis (0.63)	Nuclease activity (0.64)	NA	NA	NA
YzgL	1OED	Membrane protein	77 / 23	Cytochrome-c oxidase (139)	Hydrogen ion transmembrane transport (150)	Integral component of membrane (1049)	Membrane protein ATP binding protein of ABC transporter (0.73)	NA	NA	Integral component of membrane (0.43)	Plasma membrane (0.328)	NA
YibG	3E4B, 4BWR	Protein binding / antigen	94 / 15	NA	Protein phosphorylation (150)	NA	NA	NA	NA	NA	NA	NA
YpaB	5TF3	Membrane	82 / 25	NA	NA	NA	UpfO (0.89)	NA	NA	NA	NA	NA

Protein name	Phyre 2			Argot 2.5 (Score: 0-500 weak, 500-1000 intermediate, >1000 strong)				Pannzer 2 (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong)				DeepGO Plus (Score: 0-0.3 weak, 0.3-0.6 intermediate, 0.6-1 strong)	
	PDB ID	Modelling template function	Coverage(%) / identity (%)	Molecular function	Biological process	Cellular component	Description	Biological process	Molecular function	Cellular component		Cellular component	Biological process
YfjD	4HG0	Magnesium and cobalt efflux protein	50 / 24	Flavin adenine dinucleotide binding (6068)	Oxidation-reduction process (18488)	Integral component of membrane (5487)	CBS/transporter protein (0.68)	NA	Flavin adenine dinucleotide binding (0.66)	Integral component of membrane (0.43)		Plasma membrane (0.713)	NA
YggL	3QD7	Small muts-related protein	87 / 14	Glutaminase activity (3309), DNA binding	Glutamine metabolic process (4433)	Cytosol (1449)	DUF469 domain-containing protein (0.79)	NA	NA	NA		Cytosol (0.733)	NA
YggU	1YH5	NA	98 / 98	Hydrogen-translocating pyrophosphatase (108)	Developmental process involved in reproduction (153)	Intracellular membrane-bounded organelle (215)	UPF0235 protein YggU (0.84)	NA	NA	Integral component of membrane (0.43)		Intracellular (0.350)	NA
YhcF	6ERJ, 5AFO, 2JTY	Part of pilus, cell adhesion	92-90-92 / 16-11-15	Protein-glutamine glutaminase (413)	Cellular response to DNA damage stimulus (9352)	Pilus (691), outer membrane-bounded periplasmic space (316)	Transcriptional regulator (0.44)	NA	NA			NA	Response to DNA damage stimulus (0.507)
YhhL	6IGZ	Integral component of membrane	88 / 23	NA	NA	Integral component of membrane (4903)	Conserved predicted inner membrane protein (0.36)	NA	NA	Integral component of membrane (0.43)		Membrane (0.317)	NA
YhhM	2L0C	Integral component of membrane	69 /85	Symporter (353)	Transmembrane transport (646)	Integral component of membrane (5492)	Receptor (0.47)	NA	NA	Integral component of membrane (0.43)		NA	NA
YjeO	NA	NA	NA	Chitinase (266)	Cell wall macromolecule catabolic process (173)	Integral component of membrane (4380)	Inner membrane protein yjeO (0.51)	NA	NA	Integral component of membrane (0.43)		Plasma membrane (0.873)	NA
YfnD	NA	NA	NA	NA	NA	Integral component of membrane (290)	Periplasmic protein (0.67)	NA	NA	NA		Extracellular region (0.384)	Regulation of cellular process (0.304)
YnaJ	NA	NA	NA	Translation elongation factor (460)	Mannitol metabolic process (677)	Integral component of membrane (4781)	Predicted inner membrane protein (0.51)	NA	NA	Integral component of membrane (0.43)		Cell periphery (0.385), membrane (0.443)	NA

Supplementary Table 4 | Function and localization prediction of the 39 genes of the library.

Protein name	Phenotype or fluorescence for N-terminal tag	Phenotype or fluorescence for C-terminal tag	Consensus (at least 2 methods) Function Prediction	Consensus (at least 2 methods) Localization Prediction
YahC	Membrane fluorescence	Weak membrane fluorescence	NA	Membrane
YahE	Cytoplasm fluorescence with fluorescent clusters	Weak cytoplasm fluorescence	NA	NA
YaiY	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	Cell periphery / membrane
YbaA	Cytoplasmic fluorescence	Cytoplasmic fluorescence with few fluorescent membrane clusters	NA	NA
YbcV	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence	NA	NA
YcbK	Cytoplasmic fluorescence	Cytoplasmic fluorescence with fluorescent clusters	Peptidase	NA
YceH	Nucleoid binding	Nucleoid binding with clusters on some cells	NA	Cytosol
YceQ	Fluorescent clusters	Fluorescent clusters	NA	NA
YdaT	Nucleoid binding and Cytoplasmic fluorescence	Nucleoid binding	NA	NA
YdfR	Fluorescent clusters	Bright fluorescent clusters	NA	NA
YdhL	Longer cells, Cytoplasmic fluorescence with clusters	Long cells, Cytoplasmic fluorescence with fluorescent clusters	Oxidoreductase	NA
YdiL	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence	DNA binding	DNA binding
YdiY	Very weak cytoplasmic fluorescence	Weak membrane fluorescence with fluorescent clusters at poles and mid-cell	Transporter	Membrane
YdiZ	Cytoplasmic fluorescence	Cytoplasmic fluorescence	Endoribonuclease	NA
YeaO	Cytoplasmic fluorescence	Cytoplasmic fluorescence with few fluorescent membrane clusters	NA	cytoplasmic
YebE	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with few fluorescent clusters	NA	Membrane
YebY	Weak Cytoplasmic fluorescence	Weak membrane fluorescence with bright polar fluorescent clusters	Transaminase activity	NA
YegP	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with fluorescent clusters	NA	NA
YfjD	Cytoplasmic fluorescence with fluorescent clusters	Membrane fluorescence with fluorescent clusters	Transporter protein, FAD binding	Membrane
YggL	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	Cytosol
YggU	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	NA
YhcF	Cytoplasmic fluorescence, some cells with polar fluorescent clusters	High number of fluorescent membrane clusters	Cellular response to DNA damage stimulus	Pilus
YhhL	Membrane fluorescence with polar fluorescent clusters	Membrane fluorescence with fluorescent clusters mostly at the poles and mid cells	NA	Membrane
YhhM	Membrane fluorescence with membrane fluorescent clusters	Membrane fluorescence with membrane clusters	NA	Membrane
YjeO	Membrane fluorescence, with membrane fluorescent clusters	Weak membrane and Cytoplasmic fluorescence	NA	Membrane
YfnD	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	Periplasmic
YnaJ	Weak membrane and Cytoplasmic fluorescence	Weak Cytoplasmic fluorescence with fluorescent clusters	NA	Membrane
YbaP	Membrane fluorescence, some cells have fluorescent clusters	Cytoplasmic fluorescence, weak membrane fluorescence with clusters	Ligase	Membrane
YbdF	Nucleoid binding	Nucleoid binding	DNA binding	DNA binding
YnaA	Longer cells, Cytoplasmic fluorescence with polar fluorescent clusters	Longer cells, cytoplasmic fluorescence with polar fluorescent clusters	Cytoskeleton	Cytoskeleton
YddK	No fluorescence	No fluorescence	NA	Membrane
YeeP	Cytoplasmic fluorescence in cells with high GFP levels, weak membrane fluorescence in cells with low GFP levels	Cytoplasmic fluorescence in cells with high GFP levels, weak membrane fluorescence in cells with low GFP levels	GTP binding	NA
YehS	Weak membrane and cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	NA
YfbM	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	NA
YfdL	Cytoplasmic fluorescence with fluorescent clusters	Very weak cytoplasmic fluorescence	NA	Membrane
YraN	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with fluorescent clusters	Nuclease activity	NA
YzgL	Cytoplasmic fluorescence in cells with high GFP levels, weak membrane fluorescence in cells with low GFP levels	Cytoplasmic fluorescence with fluorescent clusters	Transporter protein	Membrane
YibG	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with fluorescent clusters	NA	NA
YpaB	Longer cells, Cytoplasmic fluorescence	Longer cells, low cytoplasmic fluorescence	NA	NA

Protein name	Phenotype or fluorescence for N-terminal tag	Phenotype or fluorescence for C-terminal tag	Consensus (at least 2 methods) Function Prediction	Consensus (at least 2 methods) Localization Prediction
YahC	Membrane fluorescence	Weak membrane fluorescence	NA	Membrane
YahE	Cytoplasm fluorescence with fluorescent clusters	Weak cytoplasm fluorescence	NA	NA
YaiY	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	Cell periphery / membrane
YbaA	Cytoplasmic fluorescence	Cytoplasmic fluorescence with few fluorescent membrane clusters	NA	NA
YbcV	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence	NA	NA
YcbK	Cytoplasmic fluorescence	Cytoplasmic fluorescence with fluorescent clusters	Peptidase	NA
YceH	Nucleoid binding	Nucleoid binding with clusters on some cells	NA	Cytosol
YceQ	Fluorescent clusters	Fluorescent clusters	NA	NA
YdaT	Nucleoid binding and Cytoplasmic fluorescence	Nucleoid binding	NA	NA
YdfR	Fluorescent clusters	Bright fluorescent clusters	NA	NA
YdhL	Longer cells, Cytoplasmic fluorescence with clusters	Long cells, Cytoplasmic fluorescence with fluorescent clusters	Oxidoreductase	NA
YdiL	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence	DNA binding	DNA binding
YdiY	Very weak cytoplasmic fluorescence	Weak membrane fluorescence with fluorescent clusters at poles and mid-cell	Transporter	Membrane
YdiZ	Cytoplasmic fluorescence	Cytoplasmic fluorescence	Endoribonuclease	NA
YeaO	Cytoplasmic fluorescence	Cytoplasmic fluorescence with few fluorescent membrane clusters	NA	cytoplasmic
YebE	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with few fluorescent clusters	NA	Membrane
YebY	Weak Cytoplasmic fluorescence	Weak membrane fluorescence with bright polar fluorescent clusters	Transaminase activity	NA
YegP	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with fluorescent clusters	NA	NA
YfjD	Cytoplasmic fluorescence with fluorescent clusters	Membrane fluorescence with fluorescent clusters	Transporter protein, FAD binding	Membrane
YggL	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	Cytosol
YggU	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	NA
YhcF	Cytoplasmic fluorescence, some cells with polar fluorescent clusters	High number of fluorescent membrane clusters	Cellular response to DNA damage stimulus	Pilus
YhhL	Membrane fluorescence with polar fluorescent clusters	Membrane fluorescence with fluorescent clusters mostly at the poles and mid cells	NA	Membrane
YhhM	Membrane fluorescence with membrane fluorescent clusters	Membrane fluorescence with membrane clusters	NA	Membrane
YjeO	Membrane fluorescence, with membrane fluorescent clusters	Weak membrane and Cytoplasmic fluorescence	NA	Membrane
YfnD	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	Periplasmic
YnaJ	Weak membrane and Cytoplasmic fluorescence	Weak Cytoplasmic fluorescence with fluorescent clusters	NA	Membrane
YbaP	Membrane fluorescence, some cells have fluorescent clusters	Cytoplasmic fluorescence, weak membrane fluorescence with clusters	Ligase	Membrane
YbdF	Nucleoid binding	Nucleoid binding	DNA binding	DNA binding
YnaA	Longer cells, Cytoplasmic fluorescence with polar fluorescent clusters	Longer cells, cytoplasmic fluorescence with polar fluorescent clusters	Cytoskeleton	Cytoskeleton
YddK	No fluorescence	No fluorescence	NA	Membrane
YeeP	Cytoplasmic fluorescence in cells with high GFP levels, weak membrane fluorescence in cells with low GFP levels	Cytoplasmic fluorescence in cells with high GFP levels, weak membrane fluorescence in cells with low GFP levels	GTP binding	NA
YehS	Weak membrane and cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	NA
YfbM	Cytoplasmic fluorescence	Cytoplasmic fluorescence	NA	NA
YfdL	Cytoplasmic fluorescence with fluorescent clusters	Very weak cytoplasmic fluorescence	NA	Membrane
YraN	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with fluorescent clusters	Nuclease activity	NA
YzgL	Cytoplasmic fluorescence in cells with high GFP levels, weak membrane fluorescence in cells with low GFP levels	Cytoplasmic fluorescence with fluorescent clusters	Transporter protein	Membrane
YibG	Cytoplasmic fluorescence with fluorescent clusters	Cytoplasmic fluorescence with fluorescent clusters	NA	NA
YpaB	Longer cells, Cytoplasmic fluorescence	Longer cells, low cytoplasmic fluorescence	NA	NA

Supplementary Table 5 | Phenotype or fluorescence of the 39 genes and their consensus with the predictions.

Strain name	Genotype	Source
Top10	mcrA, Δ(mrr-hsdRMS-mcrBC), Phi80lacZ(del)M15, ΔlacX74, deoR, recA1, araD139, Δ(ara-leu)7697, galU, galK, rpsL(SmR), endA1, nupG	Thermo Scientific
MG1655	F- lambda- ilvG- rfb-50 rph-1	Laboratory stock
KC717	NO34 + ProdZ<>(frt araC P_{BAD})	K. C. Huang (Colavin et. al, 2018)
Rosetta™ (DE3) pLysS	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysSRARE (CamR)	Laboratory stock
SKA703	ΔaraCBAD ΔlacIZYA ΔaraE ΔaraFGH attB::lacYA177C ΔrhaSRT ΔrhaBADM Tn7::tetR	Aoki, S. K. et al., Nature (2019)
DB 3.1	F- gyrA462 endA1 glnV44 Δ(sr1-recA) mcrB mrr hsdS20(rB-, mB-) ara14 galK2 lacY1 proA2 rpsL20(Smr) xyl5 Δleu mtl1	Laboratory stock

Supplementary Table 6 | List of strains used in this thesis.

#	Sequence 5'-->3'
1	CGCGATCGATTTCTAGAGTTTACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTACTAGAGACTAGAGAAGGAGATATACCATGCATACGCTCTACG
2	GTAGCCGTCAAGTTGTCATAA ATCGATCG CGTTAGGCTACATCGGTAGCCTAAAAAAATCGATCGCG
3	CGGATGCGCCATGACAGGACTTTTCATACTCCCGCCATTCAG
4	CCTGTCATGGCGCATCCGTTGGTAACGAATCAGACAATTGACG
5	GAAACGGAAGGCAGCGGCAGCAACGAGTCGCTCCAT
6	ATGGAGCGACTCGTTGCTGCCGCTGCCTTCCGTTTC
7	GGATGGAGCGACTCGTTTTCCGTTTCGCAC
8	GTGCGAAACGGAAAACGAGTCGCTCCATCC
9	CGCGTACCCGTTCCGTTTCGCACTGGAAAC
10	CGAAACGGAACGGGTACGCGAGGCTTGTC
11	CGCGATCGAT TTCTAGAGTTTACAGCTAGCTCAGTCCTAGGTATAATGCTAGCTACTAGAGACTAGAGAAGGAGATATACCATGGCTGAAGCGCAAAATG
12	GTAGCCGTCAAGTTGTCATAAATCGATCGCGTTAGGCTACATCGGTAGCCTAAAAAAATCGATCGCG
13	GCCGGGAGCGTAGAGCGTATG
14	ATCGATCGCGTTAGGCTACATC
15	TGTAGCCTAACGCGATCGATTTATGACAACTTGACGGCTAC
16	ATACGCTCTACGCTCCCGGCGGTTATGACATTATGGGCTATC
17	GTCTCTAGTAGCTAGCATATACCTAGGACTGAGCTAGTGTAAACTCTAGAAATGCATAATGTGCCTGTCAAATG
18	ATTATGCATTTACTAGAGACTAGAGAAGGAGATATAC
19	GCGAATTCGAGCTCGGTACCTAAGGAGGATTTAAAGATGGTGAGCAAGGGCGAG
20	GGTACCGAGCTCGAATTCG
21	AGCTAGTGTAAACACCGACAAACAACAGATAAAA
22	TGTCATAAATCGATCAACGCTCTCCTGAGTAGGAC
23	TGTTGTTTGTCGGTGTTTACACTAGCTCAGTCC T
24	TCAGGAGAGCGTTGATCGATTTATGACAACTTGAC
25	AACGAGTCGCTCCATCCA
26	TGTCTGATTCGTTACCAACGAAAAGGCCATCCGTCAGGATGGCCTTCTAGCTCATTTCAGAATATTTGCC
27	CGTTGGTAACGAATCAGACAATTG
28	GGTGGATGGAGCGACTCGTTTTCCGTTTCGCACTGGAAAC
29	CCTAGGACTGAGCTAGCTGTAAAC
30	ACAGCTAGCTCAGTCCTAGGATAATGCTAGTTACTAGAGACTAGAGAAGGAGATATACCATGGCTGAAGCGCAAAATGATC
31	TACCTAAGGAGGATTTAAAGATGGTGAGCAAGGGCGAG
32	CCGCCAAAACAGCCAAGCTTTTATCCTCCGAACAAGCG
33	AAGCTTGGCTGTTTTGGC
34	CTTTAAATCCTCCTTAGGTACC
35	TACCTAAGGAGGATTTAAAGATGAGCAAAGGAGAAGAACTTTTCAC
36	CCAAAACAGCCAAGCTTTTATTATTTGTAGAGCTCATCCATGCC
37	CTTTAAATCCTCCTTAGGTACCGAGC
38	TAAAAGCTTGGCTGTTTTGGCG
39	TACCTAAGGAGGATTTAAAGGTGGCACGCATTATTGTTGTTAC
40	CCAAAACAGCCAAGCTTTTACTCTTCTTCAATGAAGCGGAAAGG
41	TAAAAGCTTGGCTGTTTTGGCG
42	CTTTAAATCCTCCTTAGGTACCGAGC
43	TACCTAAGGAGGATTTAAAGATGAATACTGAAGCCACGCAC
44	CCGCCAAAACAGCCAAGCTTTTACTGCGCCGGTGATTG
45	TACCTAAGGAGGATTTAAAGATGTTGAAAAAATTTCGTGGCATGTTTTC
46	CCGCCAAAACAGCCAAGCTTTTACTCTTCGCTGAACAGGTC
47	CTTTAAATCCTCCTTAGGTACC
48	GCGC GAGCTC GGAGTAGCATAGGGTTTGCAGAATC
49	TTTTGCGGCCGCCTTTAAATCCTCCTTAGGTACCGAGCT
50	TTTTGCGGCCGCCTTTAAATCCTCCTTAGGTACCGAGCT

#	Sequence 5'-->3'
51	AAAGGATCCATGCATACGCTCTACGCTCCCGG
52	TTTGCGGCCGCTTATGACAACTTGACGGCTACATCATTCAC
53	TCCATAAGATCATCCTAGGCGATCTGACGCTTTTTATCGCAAC
54	ATCGCCTAGGATGATCTTATGGATAAAAATGCTATGG
55	GCTGCCTGACAACTTGAC
56	TAAATCGATCGCGTTAGGC
57	AGCCTAACGCGATCGATTT ATTTGTAGAGC TCATCCATGC
58	CCGTCAAGTTGTCAGGCAGCAGCAAAGGAGAAGAACTTTTC
59	GGTGACTGAAGAAAACGGGCGTTTCTCCCCAAGACCTC
60	GAGGTCTTGGGGAGAAACGCCCGTTTTCTTCAGTCACC
61	TACCTAAGGAGGATTTAAAGATGCAGTTTAAGGTTTACAC
62	CCGCCAAAACAGCCAAGCTTTTATATTCCCCAGAACATCAG
63	HEX - TGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTAT
64	ATAAAAAGCGTCAGGTAGGATCCGCTAATCTTATGGATAAAAATGCTATGGCATAGCA
65	HEX - TGCTATGCCATAGCATTTTTATCCATAAGATAGTTTGCTATCCGATCTCGCTTTTTAT
66	ATAAAAAGCGAGATCGGATAGCAAACTATCTTATGGATAAAAATGCTATGGCATAGCA
67	HEX - TGCTATGCCAAGTTTGCTATCCGATCT AGATTAGCGGATCCTACCTGACGCTTTTTAT
68	ATAAAAAGCGTCAGGTAGGATCCGCTAATCTAGATCGGATAGCAAACTTGGCATAGCA
69	TGCTATGCCAATACCTATTTTTACGATAGATTAGCGGATCCTACC
70	ATCGTAAAAATAGGTATTGGCATAGCAAAGTGTGAC
71	TGCTATGCCAAGTCCATCCTAGGCGATAGATTAGCGGATCCTACC
72	ATCGCCTAGGATGGACTTGGCATAGCAAAGTGTGAC
73	TGCTATGCCACATCCTAGGCGATAGATTAGCGGATCCTACC
74	ATCGCCTAGGATGTGGCATAGCAAAGTGTGAC
75	GATAAAAAGCGTCAGGTAGGATCTGCTAATCTTATGGATAAAAATGC
76	GCATTTTTATCCATAAGATTAGCAGATCCTACCTGACGCTTTTTATC
77	GTAAATAGGCGGTGGTGCGTAACGGCAAAAG
78	TTCTTCATGCCCGTCAGTAGCTGAACAGG
79	CTACTGACGGGCATGAAGAATCTGCTTAG
80	ACGCACCACCGCCTATTTACCAACAGTAC
81	TTCAGAGAAGTATTTGCACGGCGTCACAC
82	CGTGCAAATACTTCTCTGAATGGCGGGAG
83	AAGCTTGGCTGTTTTGGC
84	CTTTAAATCCTCCTTAGGTACC
85	CATCTTTAAATCCTCCTTAGGTACCG
86	TAAAAGCTTGGCTGTTTTGGCG
87	AAGCTTGGCTGTTTTGGC
88	TTTGTAGAGCTCATCCATGC
89	AGCAAAGGAGAAGAACTTTTC
90	GCTGCCTTTGTAGAGCTCATCC
91	GGCAGCAGCAAAGGAGAAGAAC
92	TACCTAAGGAGGATTTAAAGATGAACGGACTCACGGCAAC
93	CCGCCAAAACAGCCAAGCTTTTAATACCCGTCGCGTTTAGTC
94	GCATGGATGAGCTCTACAAAGGCAGCAACGGACTCACGGCAACG
95	AAAAGTTCTTCTCCTTTGCTGCTGCCATACCCGTCGCGTTTAGTC
96	TACCTAAGGAGGATTTAAAGGTGTGGGCGCTAACTGCG
97	CCGCCAAAACAGCCAAGCTTTTAACTGACATAATTCCCTCCGTAGG
98	GCATGGATGAGCTCTACAAAGGCAGCTGGGCGCTAACTGCGGAT
99	AAAAGTTCTTCTCCTTTGCTGCTGCCACTGACATAATTCCCTCCGTAGG
100	TACCTAAGGAGGATTTAAAGATGGCTGATTTCACCCTGTC

#	Sequence 5'-->3'
101	CCGCCAAAACAGCCAAGCTTTTACTGGCGGCGATGTCG
102	GCATGGATGAGCTCTACAAAGGCAGCGCTGATTTCACCCTGTCAAAATC
103	AAAAGTTCTTCTCCTTTGCTGCTGCCCTGGCGGCGATGTCGCCA
104	TACCTAAGGAGGATTTAAAGATGAAGTATGTTGATGGTTTTG
105	CCGCCAAAACAGCCAAGCTTCTATTCGTCGATGATTGAC
106	GCATGGATGAGCTCTACAAAGGCAGCAAGTATGTTGATGGTTTTGTG
107	AAAAGTTCTTCTCCTTTGCTGCTGCCTTCGTCGATGATTGACTC
108	TACCTAAGGAGGATTTAAAGATGGCTCAGGTTGCCATTTTTAAAG
109	CCGCCAAAACAGCCAAGCTTTTAGCGTGGAAAGATTTGTGTAGTG
110	GCATGGATGAGCTCTACAAAGGCAGCGCTCAGGTTGCCATTTTTAAAG
111	AAAAGTTCTTCTCCTTTGCTGCTGCCGCGTGGAAAGATTTGTGTAG
112	TACCTAAGGAGGATTTAAAGATGGACAAATTCGACGCTAATC
113	CCGCCAAAACAGCCAAGCTTCTACCAGTGCCGTGCTGG
114	GCATGGATGAGCTCTACAAAGGCAGCGACAAATTCGACGCTAATCGC
115	AAAAGTTCTTCTCCTTTGCTGCTGCCCCAGTGCCGTGCTGGCCC
116	TACCTAAGGAGGATTTAAAGATGAAATATCAACTTACTGCCC
117	CCGCCAAAACAGCCAAGCTTTTAATCTCCCAGATGGGC
118	GCATGGATGAGCTCTACAAAGGCAGCAAATATCAACTTACTGCCCTTG
119	AAAAGTTCTTCTCCTTTGCTGCTGCCATCTCCCAGATGGGCCAG
120	TACCTAAGGAGGATTTAAAGGTGTCGGTTGCCCGTTTTTC
121	CCGCCAAAACAGCCAAGCTTTCACCAGCAAGAAGTGAAAAAAC
122	GCATGGATGAGCTCTACAAAGGCAGCTCGGTTGCCCGTTTTTCATG
123	AAAAGTTCTTCTCCTTTGCTGCTGCCCCAGCAAGAAGTGAAAAAACTG
124	TACCTAAGGAGGATTTAAAGATGAAAATCAAGCATGAGC
125	CCGCCAAAACAGCCAAGCTTTTAATGAACAATGACAGAATCG
126	GCATGGATGAGCTCTACAAAGGCAGCAAAATCAAGCATGAGCACATC
127	AAAAGTTCTTCTCCTTTGCTGCTGCCATGAACAATGACAGAATCGTC
128	TACCTAAGGAGGATTTAAAGATGACTCAAGACTATGAACTGG
129	CCGCCAAAACAGCCAAGCTTTCATAGATATTTATGATGAGGACG
130	GCATGGATGAGCTCTACAAAGGCAGCACTCAAGACTATGAACTGGTTG
131	AAAAGTTCTTCTCCTTTGCTGCTGCCTAGATATTTATGATGAGGACG
132	TACCTAAGGAGGATTTAAAGGTGGCGGAGCAATTAGAG
133	CCGCCAAAACAGCCAAGCTTTCAAAAGAGTGATGGTTGC
134	GCATGGATGAGCTCTACAAAGGCAGCGCGGAGCAATTAGAGTTCTTTC
135	TACCTAAGGAGGATTTAAAGATGGCGGAGCAATTAGAG
136	TACCTAAGGAGGATTTAAAGATGAACGCTTATGAACTCC
137	CCGCCAAAACAGCCAAGCTTTTAACAAAGCCGCTCCAG
138	GCATGGATGAGCTCTACAAAGGCAGCAACGCTTATGAACTCCAGG
139	AAAAGTTCTTCTCCTTTGCTGCTGCCACAAAGCCGCTCCAGATC
140	TACCTAAGGAGGATTTAAAGATGAAGCTTTTGAAGACAGTTC
141	CCGCCAAAACAGCCAAGCTTTTACATGCTGTAGCCCAG
142	GCATGGATGAGCTCTACAAAGGCAGCAAGCTTTTGAAGACAGTTCC
143	AAAAGTTCTTCTCCTTTGCTGCTGCCCATGCTGTAGCCCAGCGATAAG
144	TACCTAAGGAGGATTTAAAGATGGCCAGTGGCGATCTTG
145	CCGCCAAAACAGCCAAGCTTTTATTTTCTGTTGCTATTCCATTCCTC
146	CTAAGGAGGATTTAAAGATGGCCAGTGGCGATCTTGTCCG
147	CCAAAACAGCCAAGCTTTTATTTTCTGTTGCTATTCCATTCC
148	TACCTAAGGAGGATTTAAAGATGAATATTCAGTGCAAACG
149	CCGCCAAAACAGCCAAGCTTTCACAAGCTACGTAGCCAG
150	CTAAGGAGGATTTAAAGATGAATATTCAGTGCAAACGCGTTTATG

#	Sequence 5'-->3'
151	TCTTCTCCTTTGCTGCTGCCCAAGCTACGTAGCCAGTCGG
152	TACCTAAGGAGGATTTAAAGATGGCTAACTGGTTAAATCAACTG
153	CCGCCAAAACAGCCAAGCTTTTATTCCGCCAAAGTGCG
154	GCATGGATGAGCTCTACAAAGGCAGCGCTAACTGGTTAAATCAACTGC
155	AAAAGTTCTTCTCCTTTGCTGCTGCCTTCCGCCAAAGTGCGTTTTTG
156	CTAAGGAGGATTTAAAGATGATGAAAAAAAGTATTCTGGCG
157	CCAAAACAGCCAAGCTTTTAACAGAGTTTTTCAGCTCGTTC
158	ATGAGCTCTACAAAGGCAGCATGAAAAAAAGTATTCTGGCGTTTC
159	TCTTCTCCTTTGCTGCTGCCACAGAGTTTTTCAGCTCGTTC
160	CTAAGGAGGATTTAAAGATGGCTGGTTGGTTTGAACTCAG
161	CCAAAACAGCCAAGCTTTTACGTATTGTCTTTCACCGTCTG
162	ATGAGCTCTACAAAGGCAGCGCTGGTTGGTTTGAACTCAG
163	TCTTCTCCTTTGCTGCTGCCCGTATTGTCTTTCACCGTCTG
164	CTAAGGAGGATTTAAAGATGGAACACATTTCTACTACTACGTTGATCATTATTC
165	CCAAAACAGCCAAGCTTTTACTCCGCCACACTCTCGCG
166	ATGAGCTCTACAAAGGCAGCGAACACATTTCTACTACTACGTTGATCATTATTC
167	TCTTCTCCTTTGCTGCTGCCCTCCGCCACACTCTCGCG
168	CTAAGGAGGATTTAAAGATGGCAAAGAACCGTAGCCGTCG
169	CCAAAACAGCCAAGCTTTTAGTCCCACCAAACGTCGAAAAG
170	ATGAGCTCTACAAAGGCAGCGCAAAGAACCGTAGCCGTCG
171	TCTTCTCCTTTGCTGCTGCCGTCCCACCAAACGTCGAAAAG
172	CTAAGGAGGATTTAAAGATGAATGCCGTAACAGTTAATGATG
173	CCAAAACAGCCAAGCTTTTAATTAATTAACGCCGCGATTTC
174	ATGAGCTCTACAAAGGCAGCAATGCCGTAACAGTTAATGATG
175	TCTTCTCCTTTGCTGCTGCCATTAATTAACGCCGCGATTTC
176	CTAAGGAGGATTTAAAGATGAACAATGTAAAATTACTGATTGCC
177	CCAAAACAGCCAAGCTTTTACAAGTAGACTACGCTTATCACAG
178	ATGAGCTCTACAAAGGCAGCAACAATGTAAAATTACTGATTGCC
179	TCTTCTCCTTTGCTGCTGCCCAAGTAGACTACGCTTATCACAG
180	CTAAGGAGGATTTAAAGATGCTGATTAATATTGGTCGTTTGTTAATG
181	CCAAAACAGCCAAGCTTTTATTTTTTTAACTTTAAATTTCTTCTGCC
182	ATGAGCTCTACAAAGGCAGCCTGATTAATATTGGTCGTTTGTTAATG
183	TCTTCTCCTTTGCTGCTGCCTTTTTTTAACTTTAAATTTCTTCTGCC
184	CTAAGGAGGATTTAAAGATGAGCAAACCACCTCTTTTCTTTATTG
185	CCAAAACAGCCAAGCTTTTATTGTTCGCCTACAAAGCTGAC
186	ATGAGCTCTACAAAGGCAGCAGCAAACCACCTCTTTTCTTTATTG
187	TCTTCTCCTTTGCTGCTGCCTTGTTCGCCTACAAAGCTGAC
188	CTAAGGAGGATTTAAAGATGAGTGCGCGCATGTTTGTCTTATG
189	CCAAAACAGCCAAGCTTTTACAAACAAAACTGATAGCGCAG
190	ATGAGCTCTACAAAGGCAGCAGTGCGCGCATGTTTGTC
191	TCTTCTCCTTTGCTGCTGCCCAAACAAAACTGATAGCGCAGAAAAAAC
192	CTAAGGAGGATTTAAAGATGAAACTCTCAACCTGCTGTGCCG
193	CCAAAACAGCCAAGCTTTTACTGCGGTTCGGCAGGCGC
194	ATGAGCTCTACAAAGGCAGCAAACTCTCAACCTGCTGTGCCG
195	TCTTCTCCTTTGCTGCTGCCCTGCGGTTCGGCAGGCGC
196	CTAAGGAGGATTTAAAGATGGCTCAGGTTGCCATTTTTAAAG
197	CCAAAACAGCCAAGCTTTTAGCGTGGAAAGATTTGTGTAG
198	ATGAGCTCTACAAAGGCAGCGCTCAGGTTGCCATTTTTAAAG
199	TCTTCTCCTTTGCTGCTGCCGCGTGGAAAGATTTGTGTAG
200	CTAAGGAGGATTTAAAGATGGATCTGTTGTACCGGGTAAAAACG

#	Sequence 5'-->3'
201	CCAAAACAGCCAAGCTTTTAGCGCAACATTTGCGGCAAATTC
202	ATGAGCTCTACAAAGGCAGCGATCGTTGTACCGGGTAAAAACG
203	TCTTCTCCTTTGCTGCTGCCGCGCAACATTTGCGGCAAATTC
204	CTAAGGAGGATTTAAAGATGGATAAGCAATCACTGCACGAAAC
205	CCAAAACAGCCAAGCTTTTAGCCTGGACGCACTCTTTTTTG
206	ATGAGCTCTACAAAGGCAGCGATAAGCAATCACTGCACGAAAC
207	TCTTCTCCTTTGCTGCTGCCGCCTGGACGCACTCTTTTTTG
208	CTAAGGAGGATTTAAAGATGGCAGAGCAAACCTCGCGTC
209	TCTTCTCCTTTGCTGCTGCCCGCTGAACCAGCGGTTAC
210	ATGAGCTCTACAAAGGCAGCGCAGAGCAAACCTCGCGTC
211	TCTTCTCCTTTGCTGCTGCCCGCTGAACCAGCGGTTAC
212	CTAAGGAGGATTTAAAGATGTCACTGACCTTATATTACACAATC
213	CCAAAACAGCCAAGCTTTTAAGTTTTGCTCTGTTTTATTATGG
214	ATGAGCTCTACAAAGGCAGCATCACTGACCTTATATTACACAATC
215	TCTTCTCCTTTGCTGCTGCCAGTTTTGCTCTGTTTTATTATGG
216	CTAAGGAGGATTTAAAGATGGTGACGCAGGCCGACAAAAC
217	CCAAAACAGCCAAGCTTTTAGAAGAAAATCCAGTTCCATACCGC
218	ATGAGCTCTACAAAGGCAGCGTGACGCAGGCCGACAAAAC
219	TCTTCTCCTTTGCTGCTGCCGAAGAAAATCCAGTTCCATACCGC
220	TAAGGAGGATTTAAAGATGCTAAGTAACGATATTCTGCGC
221	AAAACAGCCAAGCTTTTAGCCTTTTTTCACATGCTGG
222	GCTCTACAAAGGCAGCCTAAGTAACGATATTCTGCGC
223	TCCTTTGCTGCTGCCGCCTTTTTTCACATGCTGG
224	CTAAGGAGGATTTAAAGATGGGAATGATTGGCTATTTTGCAG
225	CCAAAACAGCCAAGCTTTTACCCAACAATAACCGTCAGC
226	ATGAGCTCTACAAAGGCAGCGGAATGATTGGCTATTTTGCAG
227	TCTTCTCCTTTGCTGCTGCCCCCAACAATAACCGTCAGC
228	CTAAGGAGGATTTAAAGATGTTGCAGGAGGAGCTTTGC
229	CCAAAACAGCCAAGCTTTTAACACCCTAATGCGAAAAATG
230	ATGAGCTCTACAAAGGCAGCTTGCAGGAGGAGCTTTGC
231	TCTTCTCCTTTGCTGCTGCCACACCCTAATGCGAAAAATG
232	CTAAGGAGGATTTAAAGATGGCTACAGTACCAACAAGGTC
233	CCAAAACAGCCAAGCTTTTATGAGTGGTCATTAAAGGCATC
234	ATGAGCTCTACAAAGGCAGCGCTACAGTACCAACAAGGTC
235	TCTTCTCCTTTGCTGCTGCCTGAGTGGTCATTAAAGGCATC
236	CTAAGGAGGATTTAAAGATGCAAAACAGAAAATGGATTTTGACC
237	CCAAAACAGCCAAGCTTTTAGGGAATTACCGCAAAGCC
238	ATGAGCTCTACAAAGGCAGCCAAAACAGAAAATGGATTTTGACC
239	TCTTCTCCTTTGCTGCTGCCGGGAATTACCGCAAAGCC
240	CTAAGGAGGATTTAAAGATGAAAGCATGCTTGTTACTATTTTTTTATTTC
241	CCAAAACAGCCAAGCTTTTACGGATTCTCCTTATTCTTCCTTC
242	ATGAGCTCTACAAAGGCAGCAAAGCATGCTTGTTACTATTTTTTTATTTC
243	TCTTCTCCTTTGCTGCTGCCCGGATTCTCCTTATTCTTCCTTC
244	CTAAGGAGGATTTAAAGATGACCCTCTTGCAAGTGCATAAC
245	CCAAAACAGCCAAGCTTTTAAGAAATCGCCGAACAGTTATTTTTAAC
246	ATGAGCTCTACAAAGGCAGCACCCTCTTGCAAGTGCATAAC
247	TCTTCTCCTTTGCTGCTGCCAGAAATCGCCGAACAGTTATTTTTAAC

Supplementary Table 7 | List of primers used to clone the constructs presented in this thesis.

Gene name	Primer	Gene name	Primer	Gene name	Primer	Gene name	Primer
yahC	92-93	ydhL	131-132	yggU	170-171	yddK	210-211
sfGFP-yacH	94-93	sfGFP-ybaA	133-132	sfGFP-ybaA	172-171	sfGFP-ybaA	212-211
yahC-sfGFP	92-95	ybaA-sfGFP	131-133	ybaA-sfGFP	170-173	ybaA-sfGFP	210-213
yahE	96-97	ydiL	134-135	yhcF	174-175	yeeP	214-215
sfGFP-yahE	98-97	sfGFP-ybaA	136-135	sfGFP-ybaA	176-175	sfGFP-ybaA	216-215
yahE-sfGFP	96-99	ybaA-sfGFP	134-137	ybaA-sfGFP	174-177	ybaA-sfGFP	214-216
yaiY	100-101	ydiY	138-139	yhhL	178-179	yehS	217-218
sfGFP-yaiY	102-101	sfGFP-ybaA	140-139	sfGFP-ybaA	180-179	sfGFP-ybaA	219-218
yaiY-sfGFP	100-103	ybaA-sfGFP	138-141	ybaA-sfGFP	178-181	ybaA-sfGFP	217-220
ybaA	104-105	ydiZ	142-143	yhhM	182-183	yfbM	221-222
sfGFP-ybaA	106-105	sfGFP-ybaA	144-143	sfGFP-ybaA	184-183	sfGFP-ybaA	223-222
ybaA-sfGFP	104-107	ybaA-sfGFP	142-145	ybaA-sfGFP	182-185	ybaA-sfGFP	221-224
ybcV	108-109	yeaO	146-147	yjeO	186-187	yfdL	225-226
sfGFP-ybaA	110-109	sfGFP-ybaA	148-147	sfGFP-ybaA	188-187	sfGFP-ybaA	227-226
ybaA-sfGFP	108-111	ybaA-sfGFP	146-149	ybaA-sfGFP	186-189	ybaA-sfGFP	225-228
ycbK	112-113	yebE	150-151	yfnD	190-191	yraN	229-230
sfGFP-ybaA	114-113	sfGFP-ybaA	152-151	sfGFP-ybaA	192-191	sfGFP-ybaA	231-230
ybaA-sfGFP	112-115	ybaA-sfGFP	150-153	ybaA-sfGFP	190-193	ybaA-sfGFP	229-231
yceH	116-117	yebY	154-155	ynaJ	194-195	yzgL	232-233
sfGFP-ybaA	118-117	sfGFP-ybaA	156-155	sfGFP-ybaA	196-195	sfGFP-ybaA	234-233
ybaA-sfGFP	116-119	ybaA-sfGFP	154-157	ybaA-sfGFP	194-197	ybaA-sfGFP	232-235
yceQ	120-121	yegP	158-159	ybaP	198-199	yibG	236-237
sfGFP-ybaA	122-121	sfGFP-ybaA	160-159	sfGFP-ybaA	200-199	sfGFP-ybaA	238-237
ybaA-sfGFP	120-122	ybaA-sfGFP	158-161	ybaA-sfGFP	198-201	ybaA-sfGFP	236-239
ydaT	123-124	yfjD	162-163	ybdF	202-203	ypaB	240-241
sfGFP-ybaA	125-124	sfGFP-ybaA	164-163	sfGFP-ybaA	204-203	sfGFP-ybaA	242-241
ybaA-sfGFP	123-126	ybaA-sfGFP	162-165	ybaA-sfGFP	202-205	ybaA-sfGFP	240-243
ydfR	127-128	yggL	166-167	ynaA	206-207		
sfGFP-ybaA	129-128	sfGFP-ybaA	168-167	sfGFP-ybaA	208-207		
ybaA-sfGFP	127-130	ybaA-sfGFP	166-169	ybaA-sfGFP	206-209		

Supplementary Table 8 | Primer pairs from Supplementary Table 7 used to amplify the 39 unkown and poorly characterized genes.

Name	Backbone	Origin of Replication	Antibiotic resistance	Transcription factor promoter	Transcription factor gene	Inducible promoter	Induced gene
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp1	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp2	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp3	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp4	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp5	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp7	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp8	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp1	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp2	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp3	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp4	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp5	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp7	pBAD	mcherry
pBLADE–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	blade fp8	pBAD	mcherry
pBLADE (AraC$_{WT}$*)–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	arac $_{WT}$	pBAD	mcherry
pBLADE (AraC$_{WT}$**)–mCherry	pBAD33	p15A	Chloramphenicol	J23101**	arac $_{WT}$	pBAD	mcherry
pBLADE–sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	sfgfp
pBLADE((VVD-AraC$_{DBD}$-sfGFP)*)–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6-sfgfp	pBAD	mcherry
pBLADE((VVD$_{C108A}$-AraC$_{DBD}$-sfGFP)*)–mCherry	pBAD33	p15A	Chloramphenicol	J23101*	vvd^{C108A}-arac$_{DBD}$-sfgfp	pBAD	mcherry
pBLADE(AraC$_{WT}$-sfGFP)-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	aracwt-sfgfp	pBAD	mcherry
pBLADE_I$_1$rev_I$_2$-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	synthetic pBAD with I$_1$rev_I$_2$	mcherry
pBLADE_I$_2$rev_I$_2$-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	synthetic pBAD with I$_2$rev_I$_2$	mcherry
pBLADE_I$_1$rev(no-35)_I$_2$-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	synthetic pBAD with I$_1$rev(no-35)_I$_2$	mcherry
pBLADE_I$_1$_I$_2$rev-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	synthetic pBAD with I$_1$_I$_2$rev	mcherry
pBLADE_1I1I2-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	mcherry
pBLADE_2I1I2-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	mcherry
pBLADE_3I1I2-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	mcherry
pBLADE_4I1I2-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	mcherry
pBLADE_I2*_7I1I2-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	mcherry
pBLADEΔO$_1$O$_2$-mCherry	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	mcherry
pReporter_only ΔO$_1$O$_2$	pBAD33	p15A	Chloramphenicol	Pc	-	pBAD	mcherry
pReporter_only	pBAD33	p15A	Chloramphenicol	Pc	-	pBAD	mcherry
pBLADE-CcdB	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	ccdb
pBLADE–MinDΔ10	pBAD33	p15A	Chloramphenicol	J23101**	blade fp4	pBAD	mindδ10
pBLADE–MreB	pBAD33	p15A	Chloramphenicol	J23101**	blade fp4	pBAD	mreb
pBLADE–RodZ	pBAD33	p15A	Chloramphenicol	J23101**	blade fp4	pBAD	rodz
pBLADE$^{ONLY\_A}$	pTRC99	pBR322	Ampicillin	J23101**	blade fp6	-	-
pBLADE$^{ONLY\_C}$	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	-	-
pCAM	pBAD33	p15A	Chloramphenicol	-	-	-	-
SiMPl_C-BLADE	pTRC99	pBR322	Ampicillin	J23101**	blade fp6	-	-
SiMPl_N-mCherry	pBAD33	p15A	Chloramphenicol	-	-	pBAD	mcherry
pBLADE–eYFP-MinD	pBAD33	p15A	Chloramphenicol	J23101*	blade fp6	pBAD	eyfp-mind
pBLADE-mCherry-MinD	pBAD33	p15A	Chloramphenicol	J23101**	blade fp4	pBAD	mcherry
pBAD33-mCherry	pBAD33	p15A	Chloramphenicol	Pc	arac $_{WT}$	pBAD	mcherry
pTRC99a-BLADE	pTRC99	pBR322	Ampicillin	J23101**	blade fp6	-	-
pTRC99a	pTRC99	pBR322	Ampicillin	-	laci	pTrc	-
pET28a-FP6	pET28a	F1	Kanamycin	T7	blade fp6	pTrc	fp6
pBLADE–YahC	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yahc
pBLADE–YahE	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yahe
pBLADE–YaiY	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yaiy
pBLADE–YbaA	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ybaa
pBLADE–YbcV	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ybcv
pBLADE–YcbK	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ycbk
pBLADE–YceH	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yceh
pBLADE–YceQ	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yceq
pBLADE–YdaT	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydat

Name	Backbone	Origin of Replication	Antibiotic resistance	Transcription factor promoter	Transcription factor gene	Inducible promoter	Induced gene
pBLADE–YdfR	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydfr
pBLADE–YdhL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydhl
pBLADE–YdiL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydil
pBLADE–YdiY	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydiy
pBLADE–YdiZ	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydiz
pBLADE–YeaO	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yeao
pBLADE–YebE	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yebe
pBLADE–YebY	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yeby
pBLADE–YegP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yegp
pBLADE–YfjD	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yfjd
pBLADE–YggL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yggl
pBLADE–YggU	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yggu
pBLADE–YhcF	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yhcf
pBLADE–YhhL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yhhl
pBLADE–YhhM	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yhhm
pBLADE–YjeO	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yjeo
pBLADE–YfnD	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yfnd
pBLADE–YnaJ	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ynaj
pBLADE–YbaP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ybap
pBLADE–YbdF	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ybdf
pBLADE–YnaA	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ynaa
pBLADE–YddK	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yddk
pBLADE–YeeP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yeep
pBLADE–YehS	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yehs
pBLADE–YfbM	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yfbm
pBLADE–YfdL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yfdl
pBLADE–YraN	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yran
pBLADE–YzgL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yzgl
pBLADE–YibG	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yibg
pBLADE–YpaB	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ypab
pBLADE–YahC-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yahc
pBLADE–YahE-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yahe
pBLADE–YaiY-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yaiy
pBLADE–YbaA-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ybaa
pBLADE–YbcV-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ybcv
pBLADE–YcbK-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ycbk
pBLADE–YceH-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yceh
pBLADE–YceQ-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yceq
pBLADE–YdaT-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydat
pBLADE–YdfR-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydfr
pBLADE–YdhL-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydhl
pBLADE–YdiL-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydil
pBLADE–YdiY-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydiy
pBLADE–YdiZ-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ydiz
pBLADE–YeaO-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yeao
pBLADE–YebE-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yebe
pBLADE–YebY-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yeby
pBLADE–YegP-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yegp
pBLADE–YfjD-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yfjd
pBLADE–YggL-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yggl
pBLADE–YggU-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yggu
pBLADE–YhcF-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yhcf
pBLADE–YhhL-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yhhl
pBLADE–YhhM-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yhhm
pBLADE–YjeO-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yjeo
pBLADE–YfnD-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	yfnd
pBLADE–YnaJ-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	ynaj
pBLADE–YbaP-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YbaP
pBLADE–YbdF-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YbdF
pBLADE–YnaA-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YnaA
pBLADE–YddK-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YddK
pBLADE–YeeP-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YeeP
pBLADE–YehS-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YehS
pBLADE–YfbM-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YfbM
pBLADE–YfdL-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YfdL
pBLADE–YraN-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YraN
pBLADE–YzgL-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YzgL
pBLADE–YibG-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YibG

Name	Backbone	Origin of Replication	Antibiotic resistance	Transcription factor promoter	Transcription factor gene	Inducible promoter	Induced gene
pBLADE–YpaB-sfGFP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YpaB
pBLADE–sfGFP-YahC	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YahC
pBLADE–sfGFP–YahE	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YahE
pBLADE–sfGFP–YaiY	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YaiY
pBLADE–sfGFP–YbaA	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YbaA
pBLADE–sfGFP–YbcV	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YbcV
pBLADE–sfGFP–YcbK	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YcbK
pBLADE–sfGFP–YceH	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YceH
pBLADE–sfGFP–YceQ	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YceQ
pBLADE–sfGFP–YdaT	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YdaT
pBLADE–sfGFP–YdfR	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YdfR
pBLADE–sfGFP–YdhL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YdhL
pBLADE–sfGFP–YdiL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YdiL
pBLADE–sfGFP–YdiY	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YdiY
pBLADE–sfGFP–YdiZ	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YdiZ
pBLADE–sfGFP–YeaO	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YeaO
pBLADE–sfGFP–YebE	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YebE
pBLADE–sfGFP–YebY	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YebY
pBLADE–sfGFP–YegP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YegP
pBLADE–sfGFP–YfjD	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YfjD
pBLADE–sfGFP–YggL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YggL
pBLADE–sfGFP–YggU	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YggU
pBLADE–sfGFP–YhcF	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YhcF
pBLADE–sfGFP–YhhL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YhhL
pBLADE–sfGFP–YhhM	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YhhM
pBLADE–sfGFP–YjeO	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YjeO
pBLADE–sfGFP–YfnD	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YfnD
pBLADE–sfGFP–YnaJ	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YnaJ
pBLADE–sfGFP–YbaP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YbaP
pBLADE–sfGFP–YbdF	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YbdF
pBLADE–sfGFP–YnaA	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YnaA
pBLADE–sfGFP–YddK	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YddK
pBLADE–sfGFP–YeeP	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YeeP
pBLADE–sfGFP–YehS	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YehS
pBLADE–sfGFP–YfbM	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YfbM
pBLADE–sfGFP–YfdL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YfdL
pBLADE–sfGFP–YraN	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YraN
pBLADE–sfGFP–YzgL	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YzgL
pBLADE–sfGFP–YibG	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YibG
pBLADE–sfGFP–YpaB	pBAD33	p15A	Chloramphenicol	J23101**	blade fp6	pBAD	YpaB

Supplementary Table 9 | List of vectors used in this thesis.

6.3 List of Figures

6.4   List of Tables

6.5   Abbreviations

Amino acids

Nucleobases

A	Alanine	L	Leucine
R	Arginine	K	Lysine
N	Asparagine	M	Methionine
D	Aspartic acid	F	Phenylalanine
C	Cysteine	P	Proline
E	Glutamic acid	S	Serine
Q	Glutamine	T	Threonine
G	Glycine	W	Tryptophan
H	Histidine	Y	Tyrosine
I	Isoleucine	V	Valine

A	Adenine
C	Cytosine
G	Guanine
T	Thymine
U	Uracyl

Units

(k) Da	(kilo) Dalton		ms	Millisecond
°C	Degrees celsius		nm	Nanometer
Å	Ångström		nM	Nanomolar
bp	Base pairs		OD_{600}	Optical density at 600 nm
g	G-force		rpm	Rounds per minute
g	Gram		s	Second
h	Hour		V	Volt
m^2	Square meter		W	Watt
min	Minute		λ	Wavelength (lambda)
mol	Mole		μ	Micro

Further abbreviations

AdoCbl	Deoxyadenosylcobalamin
ADP	Adenine diphosphate
Argot	Annotation retrieval of gene ontology terms
As	*Avena sativa*
ATP	Adenine triphosphate
BD	Brownian dynamics
BLADE	Blue light araC dimerization in *E. coli*
BLUF	Blue-light sensors using flavin adenine dinucleotide
BSA	Bovine serum albumin
BV	Biliverdin
CA	Carbonic anhydrase
CBD	Cobalamin-binding domains
CNN	Convolutional neural networks
CRP	Cyclic AMP receptor protein
CV	Coefficient of variation
DBD	DNA binding domain
DD	Dimerization domain
DIC	Differential interference contrast
DMSO	Dimethyl sulfoxide
DNA	Desoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic shift mobility
FAD	Flavin adenine dinucleotide
Fiji	Fiji is not imagej
FMN	Flavin mononucleotide
FRAP	Fluorescence recovery after photobleaching
FSC	Forward scatter
HCMV	Human cytomegalovirus
HEPES	N-2-hydroxyethylpiperazine-N-ethanesulfonic acid
IDT	Integrated DNA technologies
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Infrared light
LB	Lysogeny broth

LOV	Light-oxygen-voltage
MTS	Membrane targeting sequence
NaCl	Sodium chloride
NaOH	Sodium hydroxide
Nc	*Neurospora crassa*
NC	Negative control
NEB	New england biolabs
NIR	Near infrared light
ns	Not significant
Pannzer	Protein annotation with Z-score
PAS	Per-ARNT-Sim; period circadian protein - aryl hydrocarbon receptor nuclear translocator protein - single-minded protein
PCB	Phytocyanobilin
PCR	Polymerase chain reaction
phot1	Phototropin1
PHR	Photolyase homology domain
Phy	Phytochromes
RBS	Ribosome binding site
RMSF	Root mean square fluctuations
RNA	Ribonucleic acid
RT	Room temperature
s	Seconds
s.d.	Standard deviation
SEC	Size-exclusion chromatography
sfGFP	Superfolder green fluorescent protein
TB	Tryptone broth
TCS	Two-component systems
TF	Transcription factor
UV	Ultraviolet light
Vf	*Vaucheria frigida*
VVD	Vivid
WC	White collar
wt	Wild-type