Engineering light-inducible transcription factors for control of gene expression

A thesis submitted to Cardiff University
for the degree of Doctor of Philosophy by:

Sara Royo Miguel

Supervisor: Rudolf K. Allemann

Cardiff University
School of Chemistry
June 2024

Abstract

Light-oxygen-voltage (LOV) photoreceptors are ubiquitous blue light sensing domains that regulate the activity of various effectors, such as DNA binding domains. Illumination of a LOV domain results in the formation of a cysteinyl-flavin adduct between the protein and the chromophore. This signal is propagated to adjacent effectors and modulates their activity. LOV domains have been used for optogenetic tools, to control biological processes non-invasively with high spatiotemporal resolution. This thesis focuses on the study and engineering of LOV domain-based transcription factors for the development of optogenetic tools.

Straightforward evaluation of function is necessary in protein engineering to assess novel constructs. Fluorescent RNA aptamers are RNA sequences that selectively bind and enhance the fluorescence of small molecules. They overcome some of the limitations of fluorescent proteins as reporter systems. In Chapter 2, the use of the Pepper aptamer as a reporter of transcription in optogenetic systems was explored. An improved synthesis of the fluorogen was established and photoinduced isomerisation was observed, deeming this system unsuitable to work in combination with LOV domain-based tools.

The photoactivatable transcription factor EL222 has been extensively used to control gene expression by placing a gene after its cognate promoter. In Chapter 3, a novel approach was used to design EL222 chimeras with different DNA binding affinity that maintained photoactivation. EL222 was engineered to mimic the master biofilm regulator CsgD. One EL222 variant was able to bind the CsgD promoter naturally present in *E. coli* and induce biofilm formation under illuminated conditions.

Rational design of optogenetic tools requires detailed knowledge of the signal transduction mechanism between adduct formation and effector activation. Seven variants of aureochrome LOV domain were analysed to understand the role of Asn_{194}, hypothesised to be involved in the hydrogen bond network that links chromophore and effector. Changes in photocycle kinetics were recorded and the need for a hydrogen bond donor and acceptor amino acid in position 194 to maintain light induced dimerisation was identified. Additionally, three novel crystal structures of AuLOV variants were obtained.

Acknowledgements

Firstly, I would like to thank my supervisor, Prof. Rudolf K. Allemann, for giving me the opportunity to carry out a PhD in his research group, for providing the excellent research facilities that made this work possible and for his continued supervision. I thank the School of Chemistry, Cardiff University for funding my PhD. I would particularly like to thank Dr Pierre Rizkallah for all the help with X-ray crystallography. Additionally, I would like to thank Dr. Luke Johnson for his guidance and advice during my PhD.

I would like to thank the current members of the Allemann group: Dr David Miller, Dr. Prabhakar Srivastava, Prof. Luciana Gonzaga De Oliveira, Marco Kubiak and Oscar Wickstead. And the previous members: Gwawr, Alan, Rob, Douglas, Martin, Irene, Erin, Alice, Maks, Adura, Gareth, Yu, Jenny, Florence, Zulfa and Huw. It was a pleasure to work with you, thank you for all the help and for making the Allemann group a nice place to work. A special thank you to Gwawr, for all your support and advice and for your visits to Cardiff even after you had finished your PhD.

Thank you to the members of the Pickett, Luk and Rhys groups for joining us on Monday cake time and making the beginning of the weeks more enjoyable. And thank you to Liz and Danny for the lunch and coffee breaks that made me tackle the rest of the day with more energy.

I'm extremely grateful all my friends and family for their support and encouragement throughout this time. To those in Cardiff, and to my Spanish friends for always making time for me when I go visit, especially Blanca, Dani, Silvia and Nacho. And most importantly, thank you to my partner, Georgi, for always being there for me, for keeping me calm and for making me believe that everything will be okay.

Table of contents

List of figures

List of tables

List of abbreviations

A	Absorbance
ANOVA	Analysis of variance
APS	Ammonium persulfate
ATP	Adenosine triphosphate
AuLOV	Aureochrome 1a LOV domain from *Ochromonas danica*
BFP	Blue fluorescent protein
BL	Blue light
BLUF	Blue light using FAD sensor
bp	Base pair
BSA	Bovine serum albumin
bZIP	Basic region leucine zipper
CFP	Cyan fluorescent protein
CV	Crystal violet
DFHBI	(5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2,3-dimethyl-4H-imidazol-4-one
DFHBI-1T	(5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2-methyl-3-(2,2,2-trifluoroethyl)-4H-imidazol-4-one
DFHO	(Z-)4-(3,5-Difluoro-4-hydroxybenzylidene)-1-methyl-5-oxo-4,5-dihydro-1H-imidazole-2-carbaldehyde oxime
dH_2O	Deionised water
DLS	Diamond Light Source
DMHBI	(Z)-4-(4-Hydroxy-3,5-dimethoxybenzylidene)-1,2-dimethyl-1H-imidazol-5(4H)-one
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDO	1,2-ethanediol
EDTA	Ethylenediaminetetraacetic acid
EPS	Extracellular polymeric substance
FACS	Fluorescence Activated Cell Sorting

FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
FRET	Förster-resonance energy transfer
FT	Flowthrough (Ni-NTA chromatography)
GC	Gas chromatography
GFP	Green Fluorescent Protein
HBC	(4-((2-hydroxyethyl)(methyl)amino)-benzylidene)-cyanophenylacetonitrile
HBC_{620}	5-[(2-hydroxyethyl)methylamino]thieno[3,2-b]thiophene-2-(p-cyanophenylacetonitrile)
HBI	p-hydroxybenzylidene-2,3-dimethyl-imidazolinone
HEPES	2[4-(2-Hydroxyethyl)-1-piperazin-1yl]ethanesulfonic acid
HLH	Helix-loop-helix
HPLC	High pressure liquid chromatography
HSD	Honestly significant difference
HTH	Helix-turn-helix
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISC	Intersystem crossing
LB	Luria Broth
LC–MS	Liquid chromatography mass spectrometry
LED	Light-emitting diode
LID	Light inducible dimer
LOV	Light-oxygen-voltage
MES	2-(N-morpholino)ethanesulfonic acid
MOPS	3-(N-morpholino)propanesulfonic acid
MS	Mass spectroscopy
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at 600 nm
PAGE	Polyacrylamide gel electrophoresis
PAS	Per-ARNT-Sim
PCR	Polymerase chain reaction
PEG	Polyethylene glycol

PMSF	Phenylmethyl sulfonyl fluoride
PNIPAM	Poly(N-isopropylacrylamide)
RFP	Red fluorescent protein
RNA	Ribonucleic acid
SD	Standard deviation
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SEC	Size exclusion chromatography
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
sfGFP	Superfolder green fluorescent protein
STAS	Sulphate transporter anti-sigma antagonist
TAE	Tris.Acetate EDTA
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TEMED	N,N,N',N'-Tetramethylethylenediamine
TEV	*Tobacco Etch Virus*
TLC	Thin layer chromatography
TRIS	Trisaminomethane
UV	Ultraviolet
V_e	Elution volume
WT	Wild type
YFP	Yellow fluorescent protein

Amino acids

Alanine	Ala	A	Methionine	Met	M
Cysteine	Cys	C	Asparagine	Asn	N
Aspartic acid	Asp	D	Proline	Pro	P
Glutamic acid	Glu	E	Glutamine	Gln	Q
Phenylalanine	Phe	F	Arginine	Arg	R
Glycine	Gly	G	Serine	Ser	S
Histidine	His	H	Threonine	Thr	T
Isoleucine	Ile	I	Valine	Val	V
Lysine	Lys	K	Tryptophan	Trp	W
Leucine	Leu	L	Tyrosine	Tyr	Y

Chapter 1 - Introduction

1.1 Photocontrol

Controlling biological processes requires either harnessing natural stimuli and the associated feedback or engineering novel responses to natural or unnatural applied signals. Exogenous addition of chemical inducers such as isopropyl β-D-1-thiogalactopyranoside (IPTG), L-arabinose or L-rhamnose has primarily been used to activate inducible systems. However, their high cost is a major drawback in large scale processes.[1] The widely used inducer IPTG often amounts to 50% of total media costs, similar or higher than the carbon sources necessary for cell growth.[2] Additionally, chemical inducers may present a toxicity hazard if present in the final product.[3] Other constraints include diffusion effects, irreversibility of induction and unwanted interactions with the host organism. These limitations have led to the development of systems that respond to other extracellular factors, such as light. The use of light as an inducer presents several advantages:

- Temporal control: light offers a rapid response as the time required for optical activation is frequently on the order of milliseconds – seconds, while chemical inducer activation is typically minutes – hours. This is particularly useful when studying fast-acting signalling pathways and cellular processes.[4] The rate at which chemical inducers act is usually limited by diffusion and any metabolisation steps needed to convert the inducer to the active form and by the potential degradation by the organism. Additionally, the use of small molecules as inducers generally originates irreversible responses since their removal can be challenging or unfeasible. In contrast, the exposure to light is highly tuneable, its intensity can be modulated, and illumination can be easily stopped or used intermittently.[5]

- Spatial control: light can be applied to bulk populations or focused on a small area to induce single cells or subcellular compartments. This high spatial resolution can be achieved using photomasks or digital projection systems. Most chemical inducers are small diffusible molecules, which hinders spatial control.[6]

- Orthogonal input: for non-photoresponsive organisms, light allows a remote and non-invasive activation method that does not interfere with other biological processes. Light is broadly biocompatible, although there is a risk of phototoxicity when high intensities and prolonged exposure are used. Alternatively, many chemical inducers are metabolised by the organism and exhibit off-target effects.[5,7]

These unique features of light make optical activation and the development of novel optogenetic tools desirable for control of biological processes in complex environments. Photocontrol of biological systems can be attained by employing naturally photo-responsive proteins or by attaching photo-switchable moieties to proteins, DNA or RNA of interest. The latter photochromic compounds undergo a reversible transformation upon photoirradiation and later revert to the initial state thermally and/or photochemically.[8] When coupled with proteins, they can serve as activators/deactivators of molecular function and interactions as their electronic and geometrical changes affect the conjugated biomolecule.[9] Examples of existing photochromic molecules capable of altering protein function include spiropyrans[10], stilbenes[11], diarylethenes[12] and azobenzenes[13] (Figure 1).

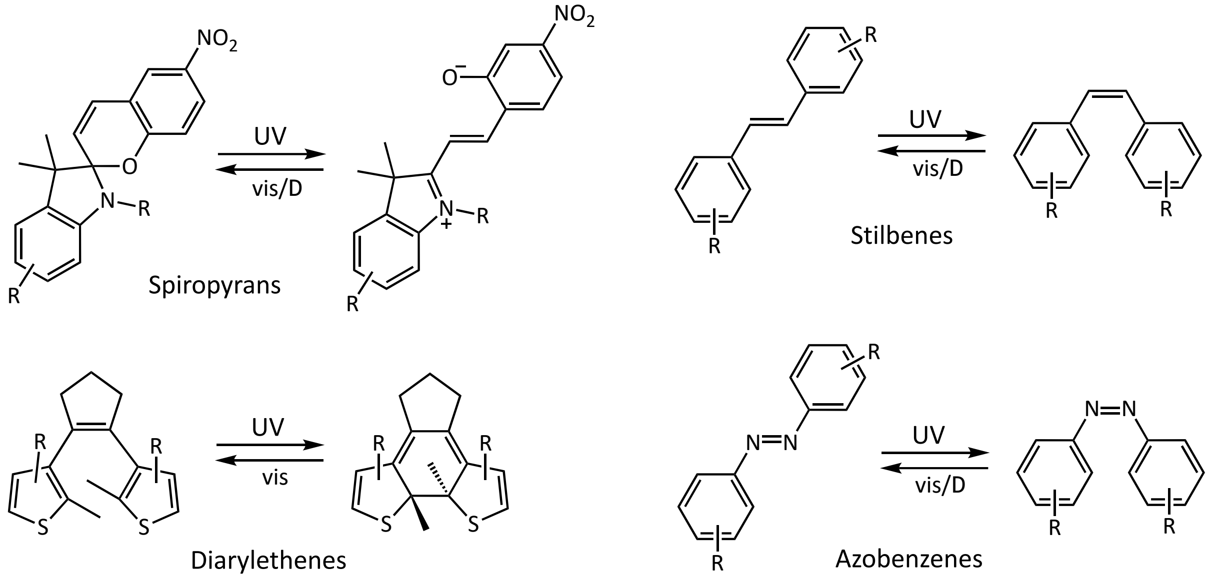


Figure 1. Examples of photochromic molecules and the structural changes they undergo upon absorption of light.

Light sensing proteins contain chromophores that, when irradiated with light of a suitable wavelength, undergo a photochemical change that affects the whole protein. Several families of photoreceptors can be distinguished depending on the chromophore they contain and the photochemical mechanism by which they react to light and include rhodopsins, phytochromes, xanthopsins, UVR8, cryptochromes, LOV (light-oxygen-voltage) domains and BLUF (blue light using FAD sensor) proteins (Figure 2).[14] Phytochromes are plant photoreceptors with a linear tetrapyrrole cofactor that are responsible for red light induced responses in gene expression, tactic responses or resetting of the circadian clock.[15] The green light photoreceptors rhodopsins contain the polyene retinal cofactor and are found in both prokaryotic and eukaryotic organisms, where they are involved in phototaxis and enabling vision in dim light.[16] Xanthopsins utilise coumaric acid as a cofactor and initiate a negative phototactic response to blue light.[17] The photoactivation of phytochromes, rhodopsins and xanthopsins causes the cis-trans isomerisation of their cofactor.[14] Cryptochromes, LOV domain and BLUF are blue light sensing flavoproteins.[18] Cryptochromes can be found in animals and plants, where they are involved in regulation of the circadian clock and in plant photomorphogenesis.[19] LOV domains appear in prokaryotic and eukaryotic organisms, and they play a key role in environmental adaptation and cellular signalling. BLUF domains are primarily bacterial photoreceptors that can be involved in a variety of functions including photophobic responses or regulation of photosynthesis.[18] UVR8 is an unusual photoreceptor that does not contain an external chromophore but interacts with UV-B light through a series of tryptophan residues in its structure and induces photomorphogenic response and protection against UV and other stresses.[20]

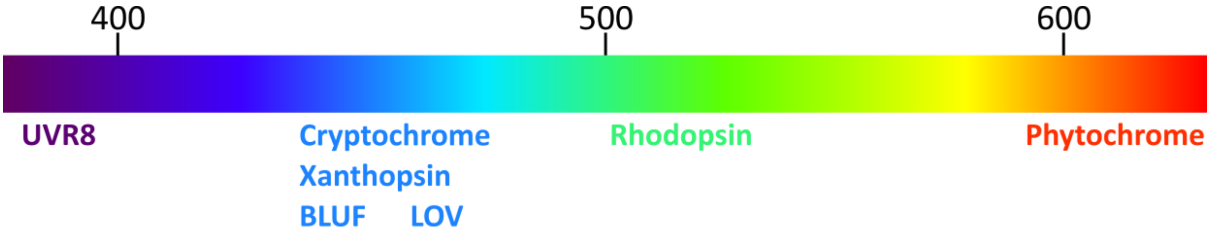


Figure 2. Absorption wavelengths of the major photoreceptor protein families.

1.2 LOV domains

LOV domains are photosensors that absorb blue and UVA light through a flavin cofactor;[21] either riboflavin, flavin mononucleotide or flavin adenine dinucleotide.[22] They belong to the Per-ARNT-Sim (PAS) superfamily, signalling proteins that sense oxygen, redox potential, light, and some other stimuli and are widely distributed across the different kingdoms.[23] LOV domains were originally discovered in phototropins,[24] a class of proteins present in plants and algae that control responses related to photosynthesis and plant growth in low light environments.[25] LOV photoreceptors have also been found in bacteria,[26] fungi[27] and archaea.[28] Although some standalone LOV domains exist in bacteria,[29] they are usually located at the N-terminus of larger proteins and have one or more effector domains linked to their C-terminus. A great variety of effector domains have been identified, including kinases,[30] DNA binding domains,[27] sulphate transporter anti-σ antagonists[31] and phosphodiesterases.[32]

LOV domains therefore confer the ability to control a diverse range of functions by light in nature and demonstrate a potential for expanding this to novel functions.

1.2.1 Structure of LOV domains

The structures of LOV domains are highly conserved between species and comprise a classical PAS fold. LOV domains consist of around 110 amino acids in an antiparallel β-sheet with five strands (Aβ, Bβ, Gβ, Hβ and Iβ) and four α-helices (Cα, Dα, Eα and Fα). The PAS fold creates a pocket for binding the flavin cofactor through hydrogen bonds, electrostatic interactions, and Van der Waals contacts (Figure 3). The flavin-binding pocket is nonpolar around the dimethylbenzene side of the isoalloxazine ring and polar on the pyrimidine side and the ribityl phosphate chain.[33] The cofactor is oriented so that the C_{4a} position of the flavin ring system is close to the sidechain of a highly conserved cysteine residue of the Eα helix.[34] The LOV core is usually linked to N- or C- terminal helical extensions that can be undocked or packed against it.[35] There is a recurring motif on the C-terminal end of LOV domains, an α helix (Jα) with a conserved amphipathic pattern that forms hydrophobic interactions with the solvent exposed side of the β-sheet, directly underneath the flavin cofactor.[36]

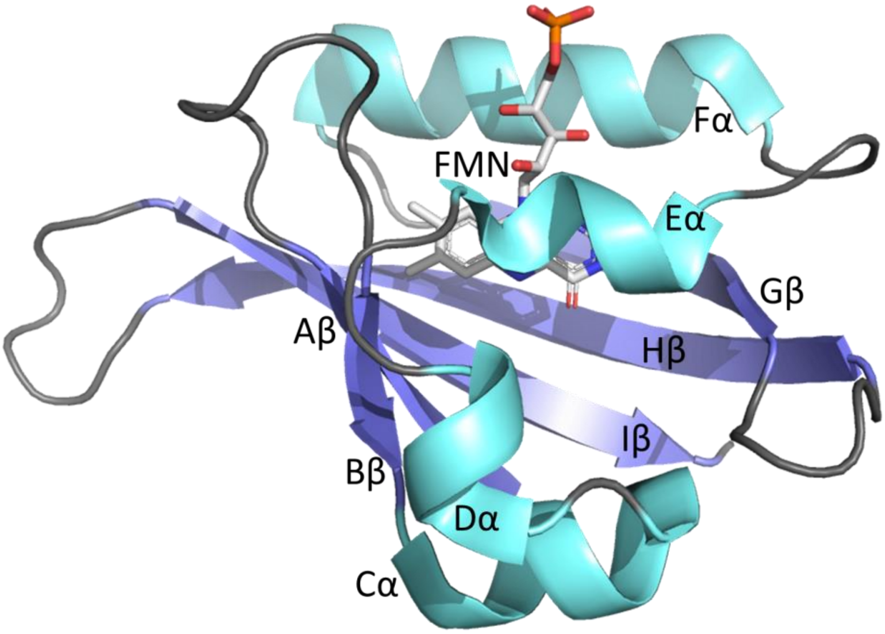


Figure 3. Structure of a LOV domain core in the light state with α-helices shown in light blue, β-strands in dark blue and loops in grey (adapted from EL222 PDB ID: 3P7N).

1.2.2 Interaction with light

Exposing LOV domains to blue light results in the formation of a covalent adduct between C_{4a} of the flavin cofactor and the sulphur atom of the conserved cysteine sidechain of the LOV domain (Figure 4).[37] This bond formation and concurrent protonation of N_5 causes a change in hydrogen bonding around the cofactor binding pocket. A larger scale conformational change propagates, resulting in signal transduction to the effector domain. The formation of the cysteinyl-flavin C_{4a} adduct changes the absorption spectrum of the protein (Figure 5).[38] The dark state has three absorbance maxima between 400 and 500 nm (approximately at 425, 450 and 475 nm) which are lost upon illumination and replaced with a new absorbance maximum at 390 nm.

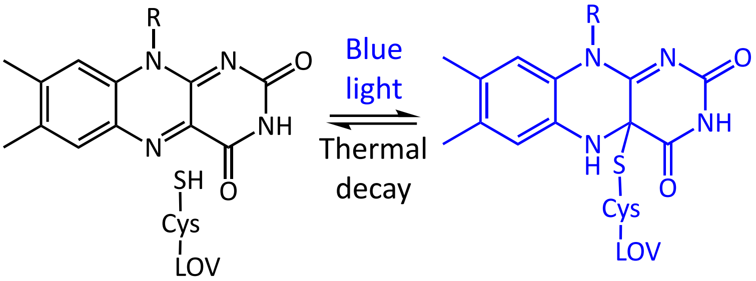


Figure 4. Formation of the cysteinyl-flavin C_{4a} adduct upon blue light illumination.

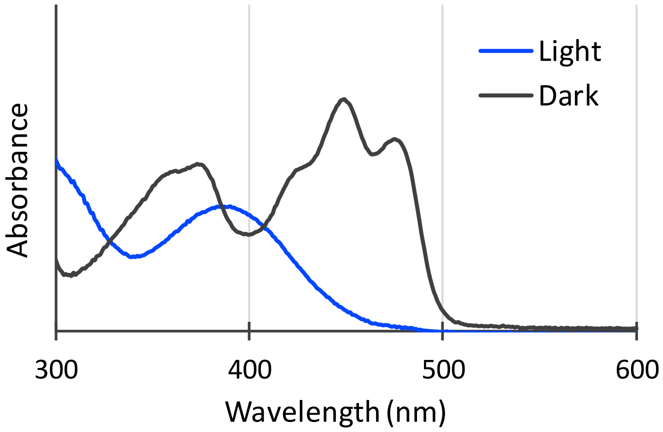


Figure 5. Absorption of a LOV domain in dark (black) and light (blue) states. Spectra obtained from measurements of AuLOV N194Y.

The formation of the adduct occurs by excitation of the flavin cofactor followed by intersystem crossing (ISC) to give the triplet state (S1→T1).[39] From this point, the proposed mechanism is a radical formation of the adduct[40] of which there are two possibilities. Firstly, an electron transfer to give a radical anion followed by ISC and bond formation, with a proton transfer to give the final adduct (Figure 6).[41] Secondly, a hydrogen transfer to generate a neutral radical pair and the subsequent ISC and recombination to yield the adduct (Figure 7).[41, 42]

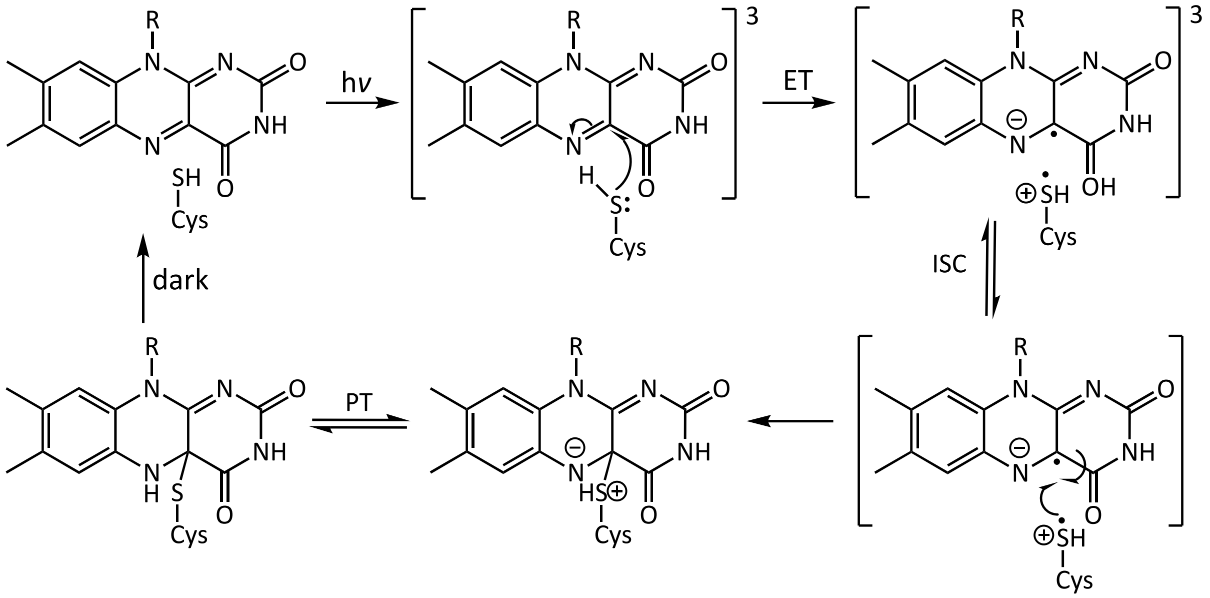


Figure 6. Adduct formation by radical anion mechanism.

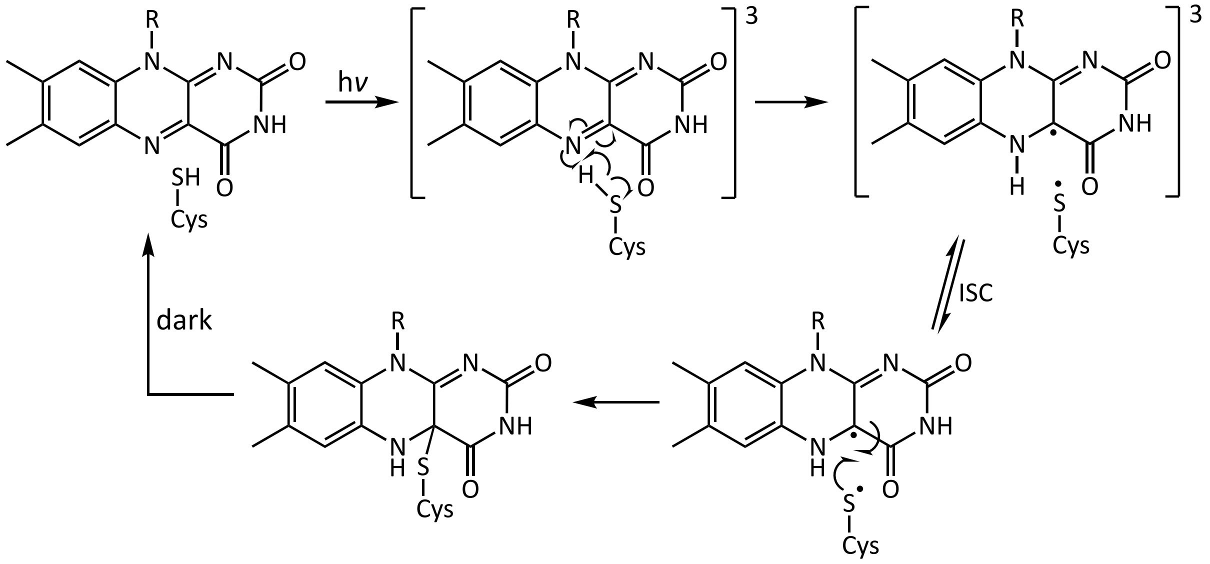


Figure 7. Adduct formation by neutral radical pair mechanism.

The adduct undergoes thermal decay to the dark state though a base catalysed deprotonation of the N_5 of the flavin cofactor by an internal (amino acid) or an external (solvent) base. The S–C_{4a} bond breaks and the C_{4a}–N_5 double bond is regenerated. The kinetics of this step depend on the temperature and the amino acid sequence of each LOV domain.[43]

1.2.3 Signalling mechanisms

The formation of the cysteinyl-flavin adduct disrupts the planarity of the isoalloxazine ring and leads to a ~10° tilting of the flavin ring (Figure 8 left), which in turn affects neighbouring amino acids and modifies the H-bonding network.[44] A noteworthy change is the rotation of a glutamine residue in Iβ strand of the β-sheet (Figure 8 right); in the dark, the amine sidechain forms a H-bond with O_4 of the flavin ring but upon illumination it is lost, and a H-bond between the carbonyl oxygen with the protonated N_5 is formed. This rearrangement in the hydrogen bond network triggers further structural changes, initiating downstream signalling. Residues that interact with the nonpolar side of the ring display smaller changes following adduct formation.[45] Additionally, there is an increase in electronic density in the flavin ring, especially around the N_5 and C_{4a} positions, which affects nearby residues with diffuse electron clouds.[46] Following this, there are five main mechanisms of signal transduction from the LOV domain to the effector domain *via* large-scale conformational changes that follow adduct formation.

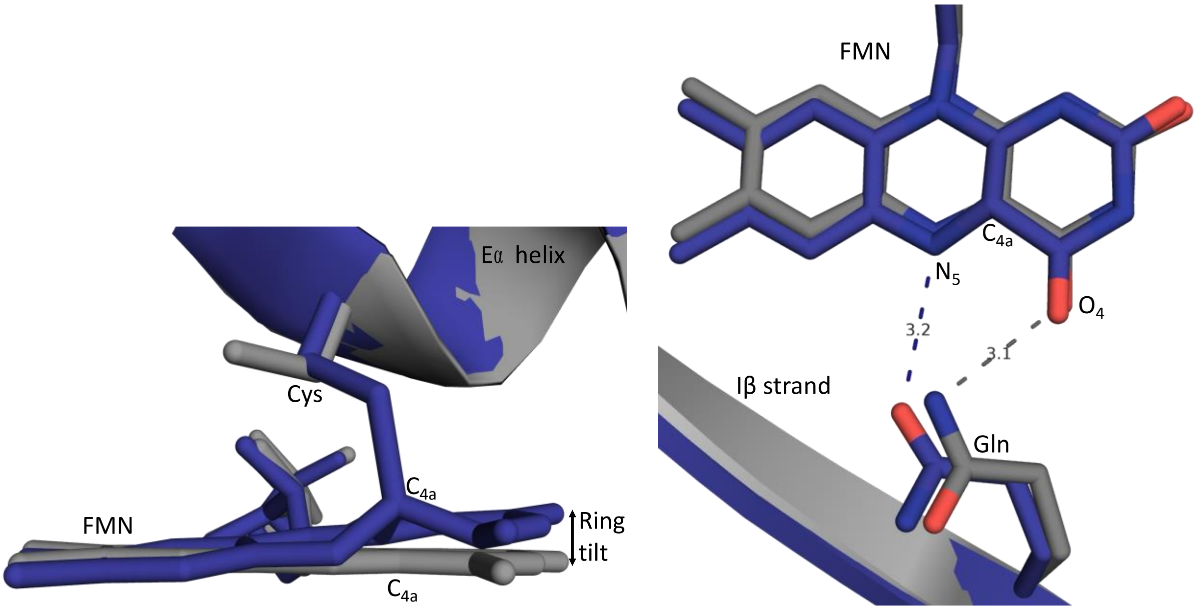


Figure 8. Tilting of the flavin ring system (left) and rotation of glutamine residue (right) from the dark state (grey) to the light state (blue). Adapted from crystal structures of YtvA in dark (PDB ID: 2PR5) and light (PDB ID: 2PR6) states.

Unfolding

The first mechanism involves unfolding of a segment of the protein and the consequent release and activation of an effector. This behaviour has been observed in phototropins, which consist of two LOV domains (LOV1 and LOV2) near the N-terminus followed by a kinase domain in the C-terminus.[22] LOV2 plays a more important role in the regulation of the kinase domain.[47] The formation of the adduct leads to unfolding of the Jα helix and subsequent activation of the kinase effector domain.[48] The most studied example is the phototropin-1 LOV2 domain of *Avena sativa* (*As*LOV2), where photoactivation of the kinase domain produces autophosphorylation and induces responses including phototropism,[49] stomatal opening,[50] chloroplast movement[51] and leaf positioning.[52]

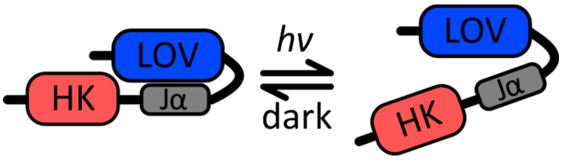


Figure 9. Unfolding mechanism of LOV domain activation, represented by AsLOV2.

Dimerisation

The conformational changes that occur after illumination can enable interactions between two monomers and cause dimer formation. The protein vivid (VVD) from fungus *Neurospora crassa* undergoes homodimerisation when illuminated. VVD is a short cytosolic protein that contains a LOV domain[53] and a N-terminal extension formed by an α-helix (aα) and a β strand (bβ) that, in the dark state, is packed against the LOV core β-sheet.[54] Photoactivation leads to conformational changes in the N-cap, dissociating from the LOV domain core, which allows homodimerisation of the VVD protein.[55] It regulates the response of the organism to varying intensities of light by controlling the activity of White Collar Complex, the main photoreceptor for the circadian system of *Neurospora crassa.*[56]

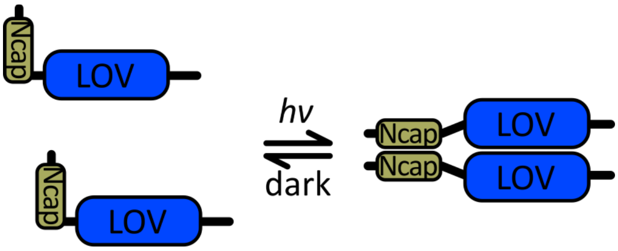


Figure 10. Homodimerisation mechanism of LOV domain activation, represented by VVD.

Conformational changes can also make available a binding region selective to a certain biomolecule, leading to heterodimerisation. This is observed in partner proteins FKF1 and GIGANTEA in *Arabidopsis thaliana.*[57] FKF1 consists of a N-terminal LOV domain, an F-box (related to ubiquitin-mediated degradation) and a Kelch repeat (involved in protein-protein interactions).[58] Illumination causes conformational changes that allow the FKF1 protein to bind to GIGANTEA, a nuclear membrane protein. This heterodimer is involved in flowering regulation depending on daylight length.[59]

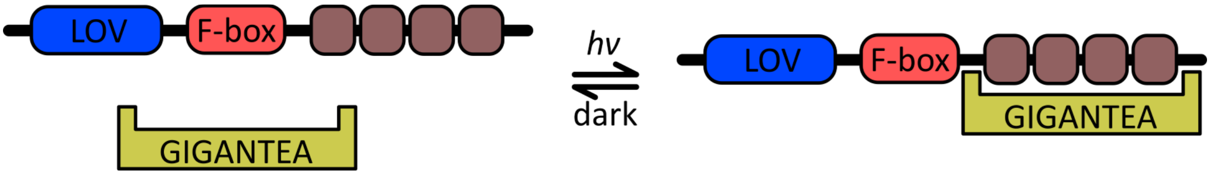


Figure 11. Heterodimerisation mechanism of LOV domain activation, represented by FKF1.

Dissociation

Although less frequent, the opposite response has also been observed, for example in the gram-negative bacterium *Rhodobacter sphaeroides*, where *Rs*LOV is a small protein that contains a LOV domain, an N-terminal aα helix and a C-terminal helix-turn-helix (HTH) dimerisation motif which includes the Jα helix. It is a dimer in the dark state, with the HTH motive as the interaction surface; upon illumination, the conformational changes affect the dimerisation interface, and the dimer dissociates into monomers.[60] *Rs*LOV is involved in regulation of responses to photooxidative stress and chemotaxis and in photosynthesis and carbohydrate metabolism.[61]

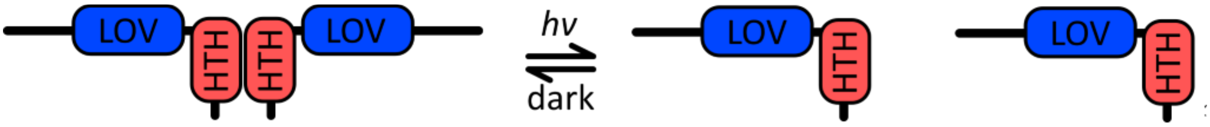


Figure 12. Dissociation mechanism of LOV domain activation, represented by RsLOV2.

Unfolding and dimerisation

The first two behaviours can appear in combination when the formation of the adduct causes the release of an effector, which makes possible the interactions needed for dimerisation.[34] This behaviour occurs in protein EL222 from bacterium *Erythrobacter litoralis* HTCC2594. It is formed by an N-terminal LOV domain linked by a Jα helix to a C-terminal HTH DNA binding domain. In the dark state, there are interactions between the LOV core β-sheet and a HTH helix that inhibit the DNA binding ability of the protein.[62] The light-induced changes in conformation disrupt the interactions between the β-sheet and the α helix that holds both domains together. The unfolded monomer can dimerise by forming LOV-LOV and HTH-HTH interactions and bind to DNA, although the proportion of dimer is very low in absence of DNA. Once illumination is ceased, the protein reverts to the dark state and the DNA is released.[63] EL222 is believed to be involved in photoprotection and repair of photodamaged DNA.[64]

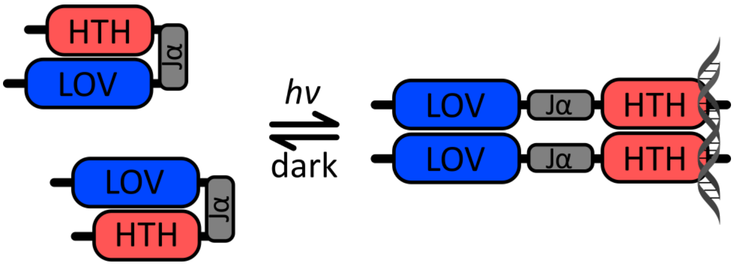


Figure 13. Unfolding and dimerisation mechanism of LOV domain activation, represented by EL222.

Tilting and rotation

Light can also induce conformational changes in a segment of the protein that causes rotation. An example of this response appears in the YtvA protein found in *Bacillus subtilis,* which acts as a positive regulator of the general stress response.[65] YtvA contains an N-terminal segment and a LOV domain linked to a C-terminal STAS (sulphate transporter anti-sigma antagonist) domain by a Jα helix.[45] YtvA is dimeric in solution in both dark and light states. Blue light exposure causes structural changes in the linker Jα helices where the two monomers rotate relative to each other.[45] This causes a conformational change in the nucleotide binding pocket of the STAS domain, which can now bind ATP or GTP.[66]

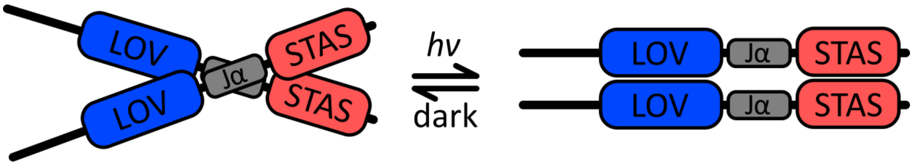


Figure 14. Tilting and rotation mechanism of LOV domain activation, represented by YtvA.

1.3 Optogenetic tools

Natural photoswitches can be engineered and introduced into organisms to control biological processes; using a technique known as optogenetics.[67] The performance of optogenetic tools is influenced by the nature of the photosensor. LOV photoreceptors are versatile thanks to their small size and the possibility to modulate parameters such as the dynamic range, the quantum efficiency of photoexcitation and the kinetics of the photocycle by introducing certain mutations.[68] This can optimise the performance of the original photoswitch for specific tasks and give way to a great variety of tools with different properties. However, the light-induced conformational changes within the LOV domain and the mechanism of signal transduction that leads to effector activation have not been fully elucidated. Progress in solving these questions is still needed to improve the scope of novel optogenetic tool design.

Current LOV domain based optogenetic tools mainly derive from a few model proteins. *Avena sativa* LOV2 domain has been used to photocage different effectors that are not naturally associated with the protein. These engineered variants exploit the unfolding of the Jα-helix to activate the linked effectors with illumination (Figure 15). For instance, *As*LOV2 was fused with *E. coli* tryptophan repressor protein (TrpR) to confer optical control of transcription in LovTAP (LOV and tryptophan activated protein). In the dark, the contacts between the LOV domain and TrpR render it inactive. Upon illumination, the unfolding of the Jα-helix allows

TrpR to bind DNA.[69] Initially, the difference in DNA affinity between light and dark states was modest (5-fold) but, by stabilising the docked conformation of the Jα-helix, undesired DNA affinity in the dark state was reduced and the dynamic range was improved (70-fold difference).[24] Similarly, a fusion protein of *As*LOV2 and *E. coli* dihydrofolate reductase (LOV-DHFR) was made to control the reduction of dihydrofolate to tetrahydrofolate. However, the chimera catalytic rate was significantly slower than the natural enzyme and only achieved a 2-fold activation with light due to the interference with a region critical for enzymatic function.[70] *Bacillus subtilis* YtvA LOV domain has also been used to achieve photocontrol of different effector domains. This LOV domain was fused to the histidine kinase domain of the oxygen-sensing protein FixL from nitrogen fixing bacterium *Bradyrhizobium japonicum* (*Bj*FixL). FixL consists of two PAS domains and a histidine kinase that comprises a phosphoacceptor (DHp) and a catalytic (CA) subdomain. In absence of oxygen, FixL first undergoes autophosphorylation at a conserved histidine in the DHp subdomain and then transfers the phosphate moiety to the response regulator FixJ. This in turn activates the expression of genes involved in nitrogen metabolism. Kinase activity of *Bj*FixL is strongly repressed in the presence of oxygen; by substituting the oxygen-sensing for a light-sensing domain, the engineered chimera YF1 attained a 1000-fold reduction of kinase activity upon illumination.[35]

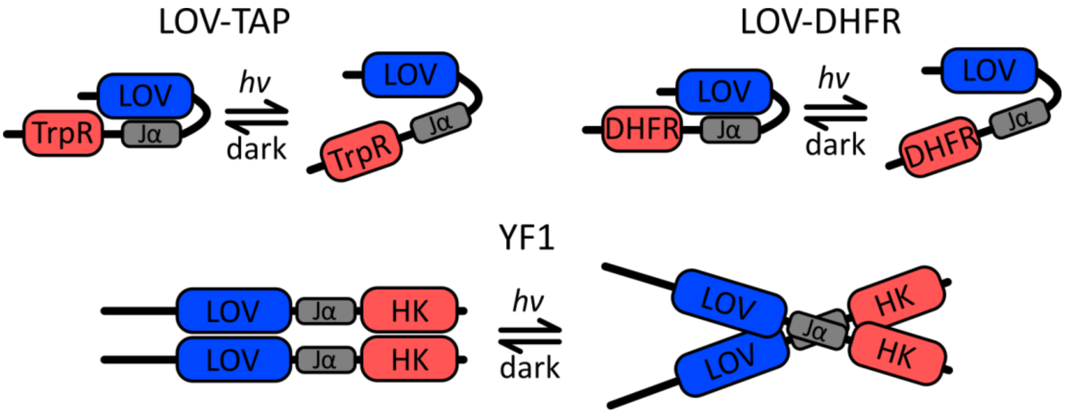


Figure 15. Optogenetic tools based on the undocking (top) or rotation (bottom) of Jα helix.

Another approach to designing optogenetic tools involves the photocontrol of formation or disruption of intermolecular interactions. Several light inducible dimers (LIDs) have been designed based on LOV domains, including *As*LOV2, VVD and FKF1.[71] TULIPs (tuneable, light controlled interacting protein tags) was developed based on the interaction of an engineered PDZ domain with a peptide recognition sequence that was linked to *As*LOV2 (LOVpep).[72] The peptide is caged in the dark state by the docking of the Jα helix into *As*LOV2. Illumination causes the peptide to become exposed so that it can interact with its PDZ binding partner (Figure 16). PDZ are small globular domains that bind to specific sequences on the C-terminus of other proteins, playing an important role in signal transduction.[73] This system can be used to recruit proteins to specific regions of a cell and to control signalling pathways.[72]

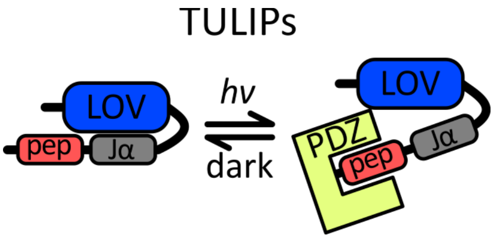


Figure 16. Mode of operation of TULIPs. The light activation of the LOV domain releases the caged peptide for interaction with its binding partner PDZ.

Another example of LIDs that exploit the natural affinity of binding partners is 'improved light inducible dimer' (iLID). This tool is based on the fusion of *As*LOV2 and SsrA,[71] a short peptide degradation tag found in *E. coli* that binds the protein SsrB, which then stimulates degradation of the tagged protein.[74] In the dark state, steric hinderance prevents SsrA and SsrB from interacting, and the undocking of the Jα-helix with illumination allows them to bind (Figure 17). The initial system only showed a twofold increase in affinity in the light state. By stabilising the docking of the Jα helix in the dark and introducing other mutations followed by screening, the system was optimised to achieve a 50-fold change in affinities between dark and light states. The recognition between its components can be used to sequester molecules to the organelle of choice and, as a result, to control their function through localisation.[71]

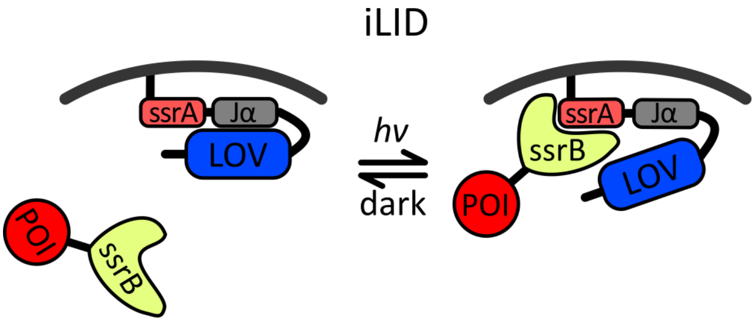


Figure 17. Mode of operation of iLID (adapted from [75]). Illumination leads to the release of ssrA, which can then interact with its binding partner ssrB.

'Magnets' are alternative LIDs based on the protein VVD where the N-caps have been modified to give two distinct variants which heterodimerise.[76] The use of the natural photoswitch to control interactions between two distinct proteins (P1, P2) had a reduced yield due to the homodimerisation that also took place. The N-cap interaction surfaces were modified with positively charged amino acids (pMag) or negatively charged amino acids (nMag) to promote heterodimerisation through electrostatic interactions (Figure 18). The reversion kinetics of VVD were originally too slow to be precisely controlled so the LOV domain was optimised, resulting in a remarkably fast and efficient photoswitch that allows excellent spatiotemporal control.

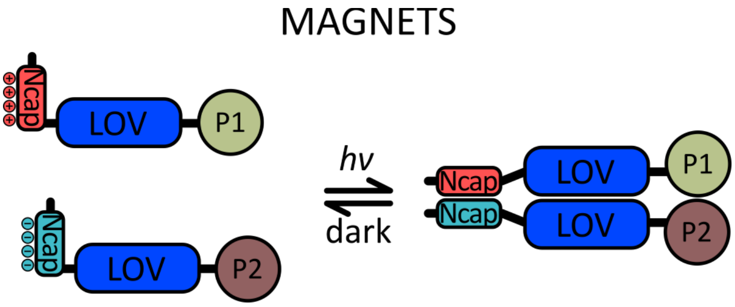


Figure 18. Mode of operation of magnets. Illumination allows selective electrostatic interactions between the N-caps of the two variants of VVD.

An instance of dissociation-based LOV domain control is LOVTRAP, which is based on the interaction between *As*LOV2 and Zdk.[77] Zdk (Zdark) derives from the Z domain of protein A, a protein present in *Staphylococcus* that binds immunoglobulin. Zdk was optimised to selectively bind the dark state of AsLOV2. Illumination causes a 150-fold decrease affinity of Zdk for *As*LOV2 and therefore dissociation. To use this system, one member should be linked to a protein of interest (POI) and the other anchored to a site away from where POI acts. In the dark state, the interactions between *As*LOV2 and Zdk do not allow the POI to get to its site of action. Upon illumination and subsequent dissociation, the POI can migrate to exert its function (Figure 19).

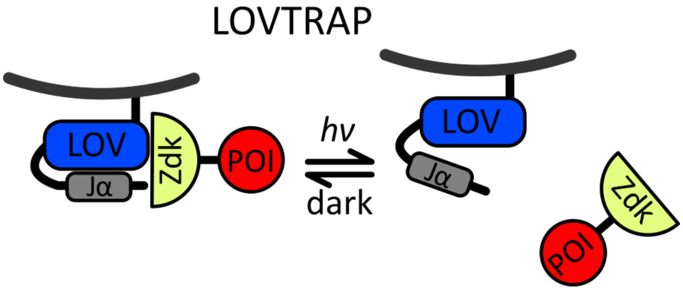


Figure 19. Mode of operation of LOVTRAP. The affinity of Zdk for AsLOV2 is greatly reduced upon illumination.

In addition to directly controlling protein activity and localisation of certain proteins, LOV photoreceptors can be exploited to control transcription. For example, the use of plasmids pDusk and pDawn allows for light-regulated gene expression (Figure 20).[78] The two systems are based on fusion protein YF1, consisting of the LOV domain of YtvA and the histidine kinase domain of FixL[35] (Figure 15). In pDusk YF1 and FixJ are constitutively expressed, and target genes can be introduced downstream of the FixK2 promoter using the multiple-cloning site.

In the dark, YF1 is active and phosphorylates FixJ. The phosphorylated FixJ binds to the FixK2 promoter and induces transcription of the desired gene. In the presence of light, YF1 does not phosphorylate FixJ, which cannot bind to the promoter to induce transcription. The system pDusk represses gene transcription upon illumination. To achieve the activation of transcription in the presence of light instead, a gene-inversion cassette was introduced into pDusk to create pDawn. In this modified system, FixJ is also phosphorylated in the dark but, when it binds to the promoter, it induces transcription of the gamma phage repressor cI. This represses the expression from the strong gamma promoter pR. When illuminated, the repressor is not expressed, and transcription of the desired gene occurs.

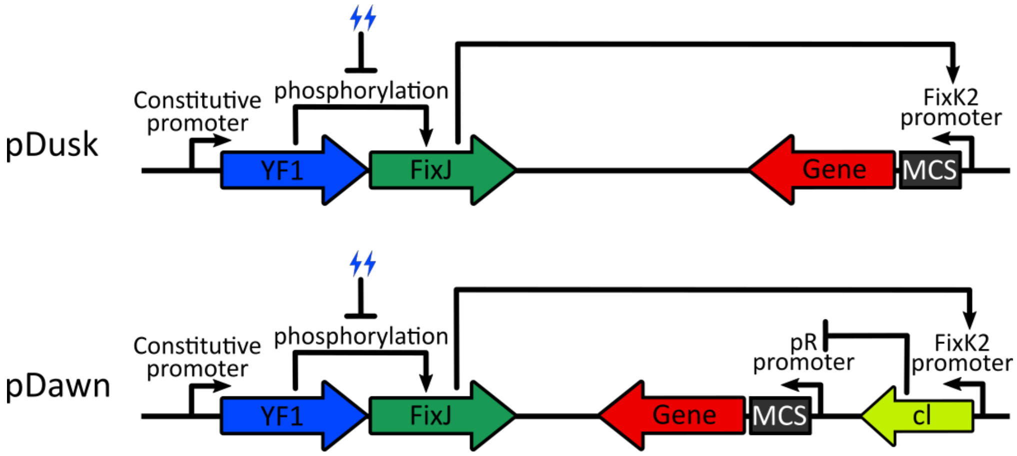


Figure 20. pDusk and pDawn systems designed to control gene expression with blue light.

1.4 Transcription

Transcription is the process whereby RNA polymerases synthesise RNA from DNA templates. Regulation of transcription is vital for cellular response to stimuli and tissue specific gene expression in higher organisms.[79] Transcription factors are proteins that control recruitment of RNA polymerase and transcription of individual genes in relation to external stimuli. They have DNA-binding domains to recognise specific sequences (enhancers or promoters). Some of them bind to a promoter sequence close to the start site and are involved in the formation of the transcription initiation complex; others bind to regulatory sequences and modulate transcription by stimulating or repressing transcription of the gene.[80] The regulation of transcription is critical for the correct functioning of an organism, as misregulation can result in diseases, whether caused by a mutation in DNA encoding the transcription factor, a mutation in the target DNA sequence that prevents the transcription factor from binding effectively or from incorrect regulation of transcription factor synthesis or activity.

The recognition of DNA by a transcription factor is a complex process dictated by the structure of the transcription factor and on the chemical interactions the amino acids can form with the DNA. Most proteins bind in the major groove of DNA since its deeper and wider character allows differentiation between the hydrogen bond pattern of each distinct base pair and consequently more specific interactions. These interactions can include hydrogen bonds, hydrophobic interactions, van der Waals forces and salt bridges. However, most DNA binding proteins recognise a family of DNA rather than a unique DNA sequence.[81] Transcription factors are classified according to the topology of the DNA binding domains and the mode of interacting with DNA. Several LOV photoreceptors are found in nature linked to DNA binding domains and act as light regulated transcription factors. The main classes of transcription factors found coupled to LOV domains are helix-turn-helix, zinc finger and basic domains.

Helix-turn-helix (HTH) domains are the predominant class of transcription factors in prokaryotes and are also present in eukaryotic organisms. The basic HTH domain comprises three α-helices with a sharp turn between the second and third helices. The third helix has amphipathic character, with a hydrophobic face that forms interactions with the other two helices, and a hydrophilic face that recognises the DNA and fits into the major groove. HTH domains typically bind DNA as dimers, with each monomer recognising a half-site.[82] Due to the open configuration of the tri-helical core, HTH domains can have one or more additional helices[83] or contain a β-strand hairpin unit[84] which can provide additional interface for contact with the minor groove of DNA. (Figure 21 left). EL222, the model protein for the unfolding and dimerisation signal transduction mechanism, contains a C-terminal HTH DNA binding domain linked to an N-terminal LOV domain *via* the Jα helix (Figure 21 right). Apart from the three α-helices of a basic HTH domain, EL222 contains a fourth α-helix that interacts with the LOV domain in the dark state, hindering dimerisation and DNA binding.[62]

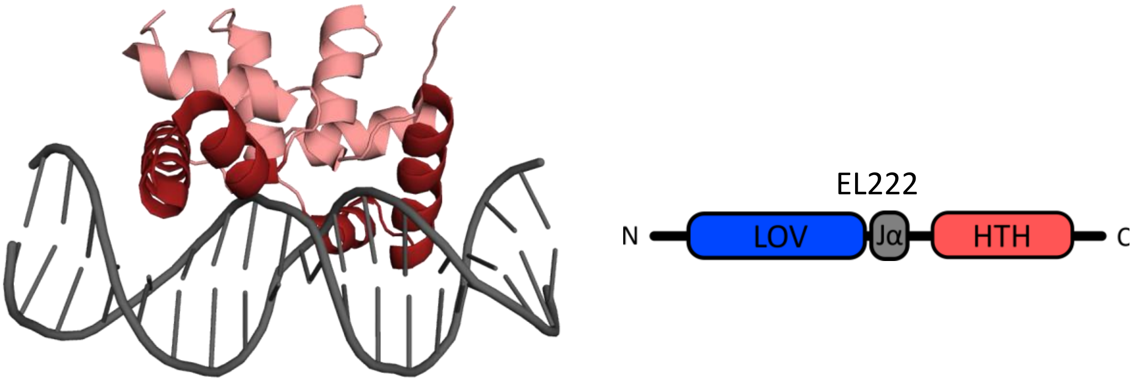


Figure 21. HTH DNA binding domain from quorum sensing protein TraR from Agrobacterium tumefaciens bound to its DNA target (PDB ID: 1H0M) and domain architecture of EL222.

The zinc finger DNA binding domain is the most commonly found in the eukaryotic kingdom, although it is also widespread in prokaryotes. Eukaryotic zinc finger motifs consist of a two stranded β-sheet hairpin and an α-helix held together by a zinc ion,[85] which is coordinated either by cysteine residues or by a combination of cysteine and histidine residues. The classical ββα motif consists of two cysteines from the hairpin turn and two histidines (C2H2) from the inside face of the α-helix tetrahedrally coordinated to a zinc ion. They are usually found in tandem, as the repeated units achieve more specific binding. The polar residues of the α-helix interact with three base pairs in the major groove of DNA and they are arranged forming an almost continuous stretch of α-helices around the DNA molecule (Figure 22 left).[86] White Collar-1 (WC-1) is the main photoreceptor for the *Neurospora* genus and is involved in the regulation of the circadian rhythm. This protein contains a LOV domain and two other PAS domains, along with a nuclear localisation signal (NLS) and a Cys4 zinc finger DNA binding domain (Figure 22 right). WC-1 heterodimerises with WC-2 to form the White Collar Complex and initiate transcription of genes related to carotenogenesis, phototropism, asexual development of spores and shifts in the circadian rhythm of sporulation.[87–89]

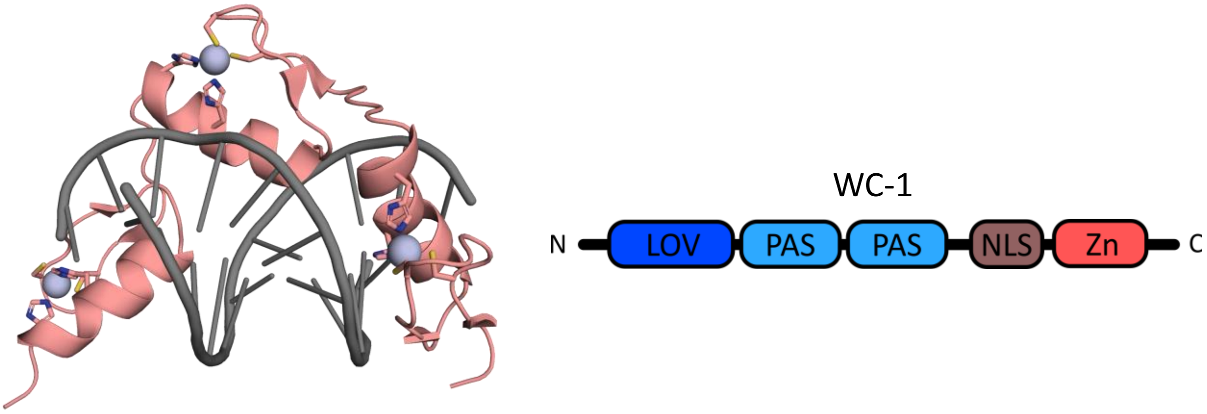


Figure 22. Classical C2H2 zinc fingers from early growth response protein 1 (Zif268) from Mus musculus bound to its cognate DNA (PDB ID: 1A1L) and domain architecture of WC-1.

Basic transcription factors exist in both prokaryotic and eukaryotic organisms and interact with DNA in dimeric form through a region that is rich in basic amino acids. This region is unstructured in solution but, when the positive charges are neutralised by the phosphate backbone of the DNA, the protein adopts an α-helical structure that inserts into the major groove of DNA.[90] Basic region leucine zipper (bZIP) motifs are a simple example of this class, consisting of a contiguous α-helix with two structural features, a basic region in the N-terminus that binds DNA and a repeating heptad region rich in leucine in the C-terminus that is involved in the dimerisation (Figure 23 left). To bind DNA, two subunits dimerise through hydrophobic interactions between the leucine residues to create a coiled coil structure. This positions each basic region in contact with half of the palindromic site in the DNA major groove.[91] Aureochromes are blue-light regulated transcription factors found in *Stramenopile* algae. They contain an unstructured N-terminal region, a bZIP domain and a C-terminal LOV domain (Figure 23 right); an inversion of the usual effector-sensor topology, where an A'α helix is linking the sensor and effector instead of the Jα helix.[92] Aureochromes play an important role in photomorphogenesis, acclimation to bright light and development of sexual reproduction organs.

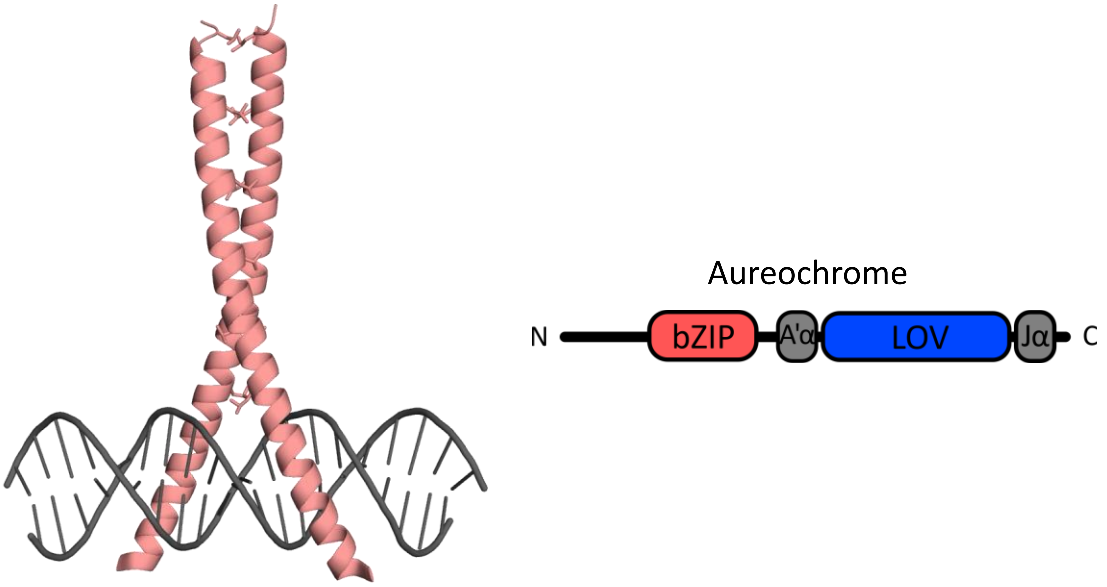


Figure 23. bZIP domain of the proto-oncogene Jun from Homo sapiens bound to its DNA target (PDB ID: 1JNM) and domain architecture of Aureochrome.

1.5 Bacterial biofilm formation

Biofilm formation is an example of an important response to external stimuli and has a very complex regulation that utilises numerous transcription factors. Biofilm formation occurs as a response to environmental stress factors such as extreme temperature, extreme pH, limited nutrients, UV radiation or antimicrobial agents and confers bacteria an increased chance of survival. It can also occur as a means to remain in a favourable habitat, such as a nutrient rich environment or a host.[93]

Biofilms are complex structured three-dimensional bacterial communities held together by a self-produced extracellular polymeric substance (EPS) matrix that are attached to abiotic or biological surfaces.[94] The formation of bacterial biofilms is a complex process that typically includes four phases: reversible and irreversible attachment, maturation and dispersal (Figure 24). In the initial step, planktonic bacteria are transported to a suitable surface, where motility and adhesion organelles allow interactions by overcoming the effects of repulsive forces.[95] The properties of the surface can affect bacterial attachment and biofilm development, including hydrophobicity, surface charge and whether the surface is biotic or abiotic.

These surface-bacteria interactions strengthen until they become irreversible, a process mediated mainly by three types of organelles.[96] Type 1 fimbriae are filamentous proteins found in the outer membrane of some *E. coli*. They bind to mannose present in eukaryotic cells, abiotic surfaces and the EPS, therefore promoting cell-cell and cell-surface interactions.[96] Curli fimbriae are amyloid fibres that aggregate at the cell surface and interact with certain host proteins of the EPS, promoting colonisation of host cells. They also contribute to cell aggregation and interactions with abiotic surfaces.[97] Conjugative pili are protein filaments that enable horizontal gene transfer by forming efficient contacts between donor and recipient cells. They also promote binding to abiotic surfaces and stabilise the biofilm structure.[98] During this step, cell to cell interactions strengthen, with a self-recognising outer membrane protein named Antigen 43 playing an important role in auto-aggregation and interspecies cell aggregation.[99]

Biofilm maturation consists of the production of an extracellular polymeric substance (EPS) matrix and the structured three-dimensional growth of the biofilm. The matrix is an essential feature of biofilms as it promotes structure, stability and intracellular interactions; provides protection against desiccation, toxic molecules and host immune responses; and enables nutrient accumulation and transport for the survival of the cells.[96]

Dispersion is the last phase in biofilm formation, where the immobilised bacteria are separated from the biofilm and return to the planktonic mode of growth, and can occur through a passive or an active route.[100] Detachment is the passive release of cells due to mechanical stress and can appear as abrasion (loss of cells caused by collision of particles with the biofilm), erosion (removal of small portions of the biofilm by liquid friction), sloughing (removal of large sections of the biofilm by friction) and grazing (loss of biofilm mass by feeding of eukaryotic organisms).[101] Active dispersal occurs as a response to steep gradients formed within the biofilm or environmental changes such as nutritional availability, oxygen depletion, concentration of toxic products and quorum sensing. The responses originated from such stresses may include enzymatic degradation of the matrix, induction of motility, production of surfactants or cell lysis of subpopulations.[102]

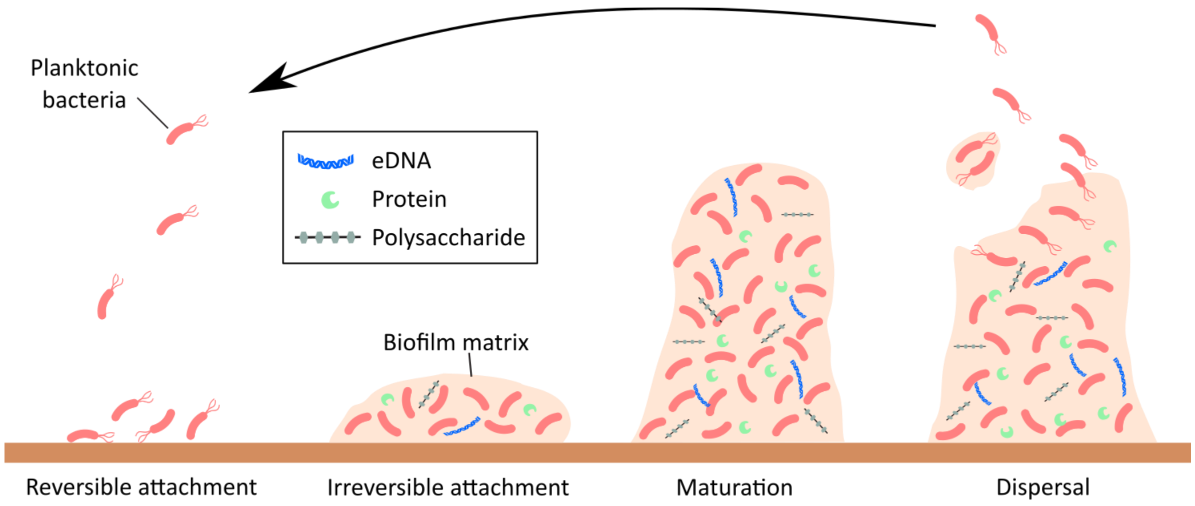


Figure 24. Stages of biofilm formation. First, planktonic cells reversibly attach to a surface. This attachment becomes irreversible as the extracellular matrix starts being produced. The biofilm then grows in a structured three-dimensional manner. Finally, cells detach and become planktonic.

1.5.1 The biofilm matrix

The matrix is the main component of a biofilm rather than the microbial cells, which can account for as little as 5%. The matrix is a highly hydrated heterogeneous substance that contains polysaccharides, proteins, amyloids, nucleic acids and lipids. The exact composition of EPS varies depending on the environmental conditions, bacterial motility and communication, and what microorganisms are present.[103]

Polysaccharides are a major constituent of the EPS matrix that provide structural support to the biofilm. In *E. coli*, three main exopolysaccharides play an important role in biofilm formation. The polymer β-1,6-N-acetyl-D-glucosamine (PGA) is involved in cell-cell and cell-surface adhesion and therefore stabilises the biofilm.[104] Colanic acid is a polyanionic heteropolysaccharide of glucose, fucose, galactose, and glucuronic acid that is involved in the development of the three-dimensional architecture upon initial surface attachment.[105] Cellulose, a β-1,4 D-glucose homopolymer, is found in only certain strains of *E. coli*. When present, cellulose closely interacts with curli fibres, conferring cohesion and elasticity and allowing the formation of complex morphologies. In the absence of cellulose, biofilms are more brittle and form less complex structures.[106]

Considerable amounts of proteins are also present in the biofilm matrix, and play roles in scaffolding, adhesion and degradation. In addition to their role in the attachment phase, curli fibres are the main structural component of the EPS matrix in *E. coli*, providing structural support and cell– cell and cell–surface adhesion.[97] Carbohydrate binding proteins also provide structural integrity by stabilising the polysaccharide network. Enzymatic degradation of biopolymers can take place to create smaller carbon and energy sources for uptake or for biofilm dispersion. Enzymes can also act as virulence factors during host colonisation and infection.[107]

Extracellular nucleic acids have been found in the biofilm matrix of many bacteria, although the amount, its origin, and its localisation vary between species. While they are generally involved in the cell-surface interactions during initial attachment and provide structural stability through cell-cell interaction, in some cases they also take part in biofilm defence by serving as a nutrient source, chelating cationic antimicrobial substances and promoting horizontal gene transfer.[108,109]

Along with these hydrophilic constituents, hydrophobic molecules are present in the biofilm matrix. Lipids can be present on their own, as phospholipids or as glycolipids and they are usually involved in initial attachment, especially to hydrophobic surfaces.[103] Due to their biosurfactant properties, glycolipids can participate in the conservation of nutrient channels, therefore maintaining the biofilm architecture.[110]

The architecture of the biofilm is dependent on the microorganisms present and their state, the EPS composition, the availability of nutrients and the environmental conditions.[111] These factors determine either a smooth or rough morphology, the level of porosity and the three-dimensional distribution of the cell dense areas, pores and channels. The presence of channels allows the flow of nutrients, enzymes, metabolites and waste through the biofilm. Due to bacterial metabolic activity, gradients of oxygen, nutrients, waste product and signalling molecules can develop, leading to varying levels of heterogeneity.[112]

1.5.2 Curli biogenesis in *E. coli*

Curli fimbriae are nonbranching fibres rich in β-sheet that, when present in *E. coli*, are located on the surface of the cells. They are essential for initial surface attachment in the biofilm formation process and form a network around the immobilised bacteria, anchoring the cells to the surface and mediating cell-cell interactions in the mature biofilm.[113] The process of curli assembly in the surface of the cell is very complex and highly regulated by the action of many stress-sensing proteins.[114]

There are seven curli subunit genes (csg) from two oppositely transcribed operons, csgDEFG and csgBAC, involved in curli production (Figure 25).[115] The major curli constituent CsgA is secreted from the cell in monomeric soluble form and polymerises at the cell surface.[116] CsgB, the minor curli subunit, is anchored to the bacterial surface after secretion and acts as a nucleator for the polymerisation of CsgA into the fibers.[117, 118]

The other five proteins are involved in the regulation, secretion and assembly of the curli subunits. CsgC is a periplasmic protein that interacts with CsgA to maintain it in soluble form before its secretion as intracellular amyloid fibre formation would cause cytotoxicity.[119] CsgD is a transcriptional regulator located in the cytoplasmic membrane that controls the transcription of the csgBAC operon[115] and of more than twenty other genes for adaptation to the biofilm mode of life, favouring biofilm formation and repressing cell motility.[120, 121] Due to its importance regulating biofilm formation, the transcription of CsgD itself is regulated by various transcription factors that monitor different environmental conditions.[114] CsgE is found in the periplasmic side of the outer membrane forming interactions with CsgG. It acts as a specificity factor for secretion of curli related proteins through the CsgG channel.[122] CsgF is secreted to the outer surface of the outer membrane[123] where it interacts with CsgG and stabilises and promotes the anchoring of CsgB to the membrane, thus acting as a chaperone in the nucleation process.[124] CsgG is a lipoprotein located in the outer membrane[125] that oligomerises forming a channel that transports CsgA, CsgB and CsgF from the periplasm to the extracellular space.[123]

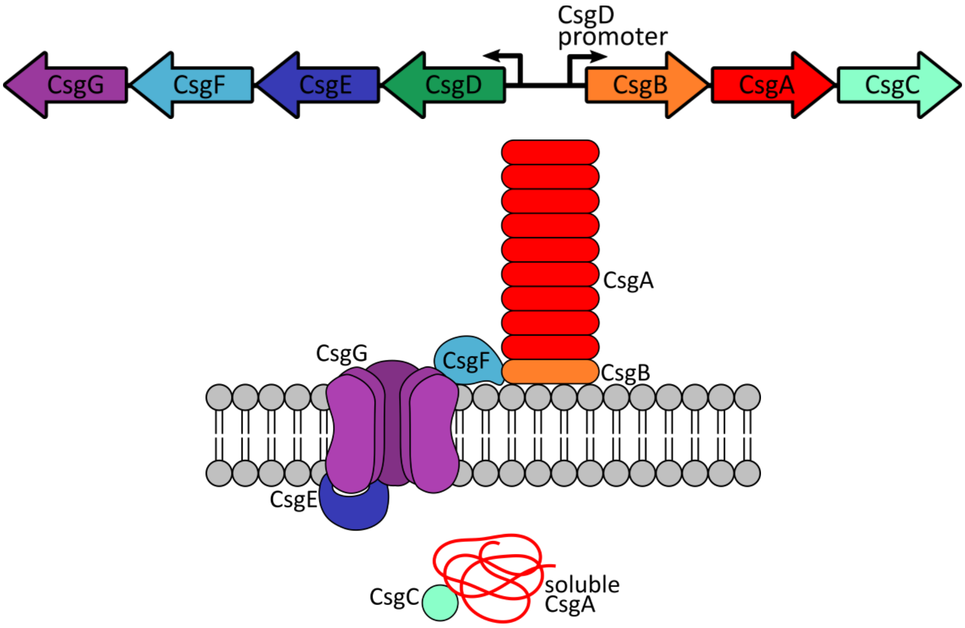


Figure 25. Curli csgDEFG and csgBAC operons (top). Schematic representation of the curli assembly process across the outer membrane (bottom).

Curli fibres are composed of two subunits, CsgA and CsgB (Figure 26). Both subunits contain a 22 amino acid long N-terminal recognition motif for interaction with CsgG and CsgE in the secretion process,[123] and an amyloidogenic motif that folds to form five repeats of β-sheet-turn-β-sheet.[116] The same amino acid can be found at the same point of the motif in all five stacked repeats. In CsgB, the fifth repeating unit is somewhat different from the other four as it contains charged amino acids that are essential for its interaction with the outer membrane.[118]

CsgA

R1 S E L N I Y Q Y G G G N S A L A L Q T D A R N
R2 S D L T I T Q H G G G N G A D V G Q - G S D D
R3 S S I D L T Q R G F G N S A T L D Q W N G K N
R4 S E M T V K Q F G G G N G A A V D Q - T A S N
R5 S S V N V T Q V G F G N N A T A H Q Y

CsgB

R1 Q A A I I G Q A G T N N S A Q L R Q G G S K
R2 L L A V V A Q E G S S N R A K I D Q T G D Y
R3 N L A Y I D Q A G S A N D A S I S Q G A Y G
R4 N T A M I I Q K G S G N K A N I T Q Y G T Q
R5 K T A I V V Q R Q S Q M A I R V T Q R

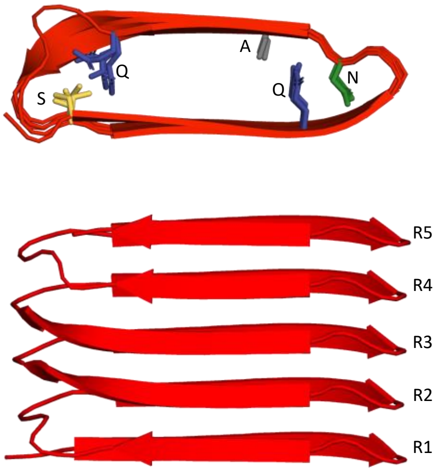


Figure 26. Amino acid alignment of the amyloidogenic motif of CsgA and CsgB (left) and structure of CsgA with five repeats of β-sheet-turn-β-sheet from a top and side view (right). PDB ID: 8ENQ.

1.5.3 Controlling biofilm formation

Biofilms confer microorganisms an increased protection from external factors, including UV radiation, oxidation, absence of water and oxygen, changes in temperature and pH and presence of metal ions and antibiotics.[93] Due to these properties, biofilm formation can be detrimental when it occurs in healthcare settings, in the food industry or in drinking water distributions systems. However, they also present beneficial roles in a variety of fields[126]. Biofilms have been used in agriculture to promote plant growth and increase the yield of crops and ornamental flowers by increasing nutrient availability or protecting roots from pathogens and abiotic stresses.[127, 128] Bioremediation employing biofilms can reach high efficiency due to the increased adaptability and tolerance to contaminants. Algal, fungal and bacterial biofilms have been employed to remove pollutants such as xenobiotics, heavy metals, oil spills or pesticides.[129,130] Also due to their high resistance and activity, biofilms have gained importance in the production of biochemicals in industrial manufacturing.[131] The treatment of wastewater with biofilms is being studied in hospital settings for removal of pharmaceuticals[132] and in domestic settings mainly for removal of phosphorous and nitrogen.[133] Biofilms also serve as anticorrosion agents in marine environments by acting as a physical barrier and producing secondary metabolites that inhibit steel corrosion.[134]

Understanding the biofilm formation process is essential for controlling their growth and dispersal. Several approaches have been developed to promote biofilm growth for biotechnological applications, including modifying the environmental conditions, adding signalling molecules or manipulating genes.[135] Due to the dynamic nature of the biofilm formation process, a control method capable of achieving high spatial and temporal precision is needed. The use of light as an inducer of biofilm formation presents several advantages, as it provides reversible, non-invasive and spatiotemporal control that can be tuned by adjusting the time, area, and intensity of illumination.

Optogenetic tools for biofilm formation can lead to an improved understanding of naturally existing biofilms, and to the improvement and design of new devices. Several examples of light-controlled biofilm formation have been described in the literature. The optogenetic tool 'Magnets' has been used to promote specific bacterial adherence to substrates with high spatiotemporal control (Figure 27 top). One binding partner was expressed in the surface of *E. coli* cells while the other was immobilised on a substrate. Illumination led to the heterodimerisation of pMag and nMag and the surface interaction promoted biofilm formation.[136] The 'Magnets' tool has also been employed to promote bacterial aggregation, where nMag and pMag were expressed in the surface of different bacteria (Figure 27 bottom). Upon illumination, both binding partners interact, bringing the bacteria close and causing cell aggregation and consequent biofilm formation. Under dark conditions, nMag and pMag dissociate from each other, and the biofilm formation process is reversible unless the illumination time surpasses two hours, due to the secondary interactions formed between the bacteria.[137] This photocontrolled aggregation has been used to improve the efficiency of biochemical production. L-threonine is an essential amino acid that is widely produced by fermentation. By utilising the 'Magnets' system to induce biofilm formation, a significant improvement in L-threonine production was achieved and the duration of the fermentation cycle and the batch numbers were reduced.[131]

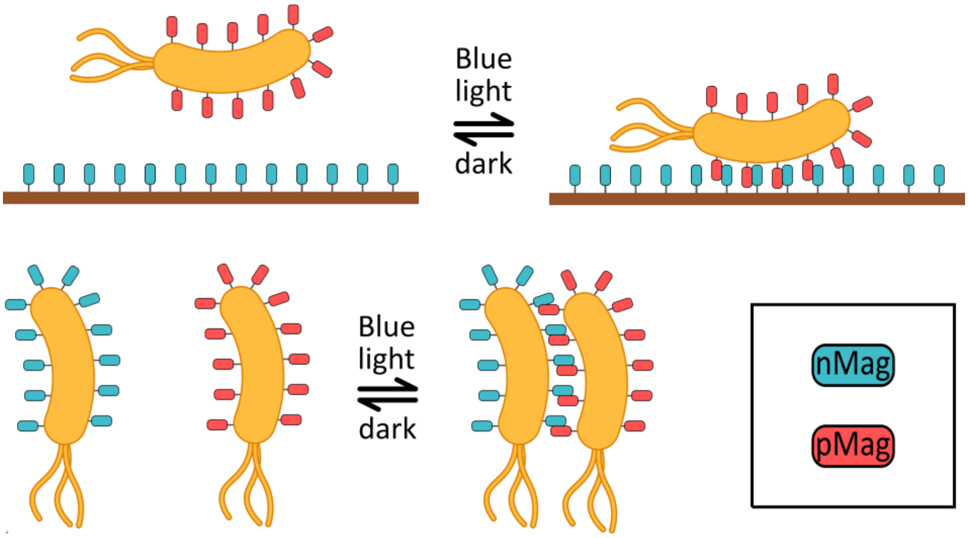


Figure 27. Biofilm formation tools based on 'Magnets' that promote surface interaction (top) or bacterial aggregation (bottom) upon illumination.

A technique named 'Biofilm Lithography' was developed to control biofilm structure by projecting illumination patterns in engineered bacteria to induce biofilm formation only in the illuminated regions. This patterning tool is based on the light dependent expression of the adhesin Antigen 43 under the control of the pDawn system. This protein promotes cell-substrate interaction, which facilitates biofilm formation (Figure 28). This method allows very high spatial resolution (micrometre scale) on a variety of surfaces, with no surface pre-treatment or patterning necessary. This technique can help understand and explore the dynamics of natural biofilms and lead to the development of new biomaterials or microbial communities that could be used in diagnostic and microfluidic devices.[138]

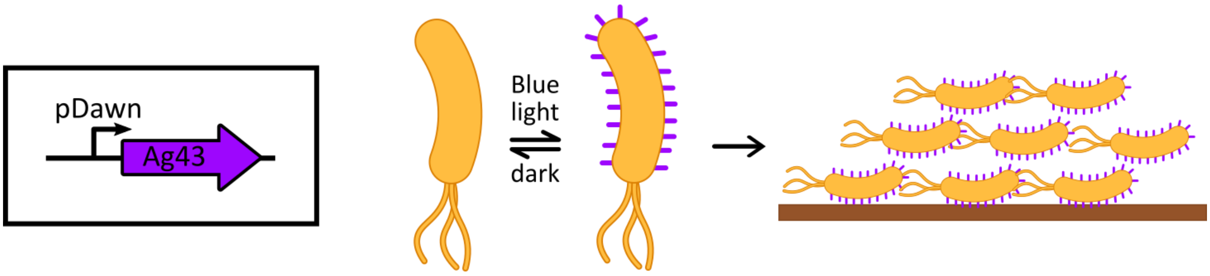


Figure 28. Biofilm lithography technique based on light dependent expression of the adhesin Ag43.

The biofilm lithography method was used to control the expression of the aggregation proteins CdrAB and induce biofilm formation in the electrochemically active bacteria *Shewanella oneidensis*. The photoinduced patterning of conductive biofilms on electrode surfaces enables further study in the field of bioelectronic devices, where biological systems are integrated as electrical components in circuits.[139]

Optical control of biofilm formation has also been achieved utilising the EL222 photoreceptor to control the exopolysaccharide production in the soil bacteria *S. meliloti*. It was accomplished by introducing constitutive expression of EL222 and placing the wgaAB gene under the transcriptional control of EL222 (Figure 29). This gene is required for EPS synthesis, which promotes autoaggregation of *S. meliloti* and the formation of structured biofilms. This system enables spatial control of biofilm formation, regulating its thickness, biomass, and structure.[140]

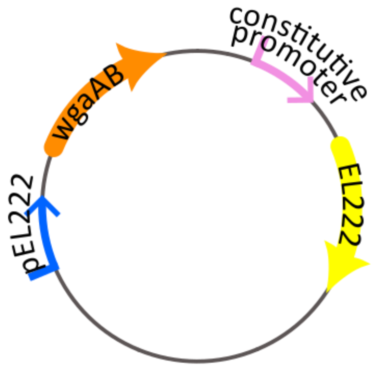


Figure 29. Plasmid map of the system used to introduce light controlled expression of wgaAB genes in S. meliloti, which leads to EPS synthesis and biofilm formation.

1.6 Fluorescent markers

In protein engineering, straightforward assessment of protein expression, localisation and function is necessary to rapidly evaluate novel designs. Genetically encodable fluorescent markers enable non-invasive and sensitive monitoring of such processes and can allow for high-throughput methods.

1.6.1 Green Fluorescent Protein (GFP)

Green fluorescent protein (GFP) is a versatile tool that enables the study of molecular processes in biological systems including gene expression, protein localisation and protein–protein interactions.[141] It was first discovered in the bioluminescent jellyfish *Aequorea victoria.*[142] It was later cloned[143] and expressed in living organisms,[141] which gave rise to many biological applications including reporting gene expression, monitoring of protein localisation and dynamic interactions and as an active indicator of the cellular or organelle environment. GFP consists of a 238 amino acid polypeptide folded to form an 11-stranded β-barrel with a single α-helix inside the cylinder, which is capped with short helical and loop segments[144,145] (Figure 30). The fluorescence of GFP originates from the p-hydroxybenzylidene-2,3-dimethyl-imidazolinone (HBI) chromophore encapsulated in the centre of the β-barrel and rigidly stabilised by hydrogen bonds.[146] The intramolecular cyclisation of Ser$_{65}$-Tyr$_{66}$-Gly$_{67}$ from the central α-helix followed by oxidation and dehydration gives the mature chromophore[143] (Figure 31). As the protein folds, the amino group of Gly is brought close to the carbonyl carbon of Ser, allowing the nucleophilic attack to form a cyclic intermediate. The carbonyl carbon of Ser then undergoes oxidation and formation of a double C=N bond followed by dehydration to form the mature chromophore.[147] The order of the oxidation and dehydration is disputed and may be reversed depending on the available oxygen concentration.[148]

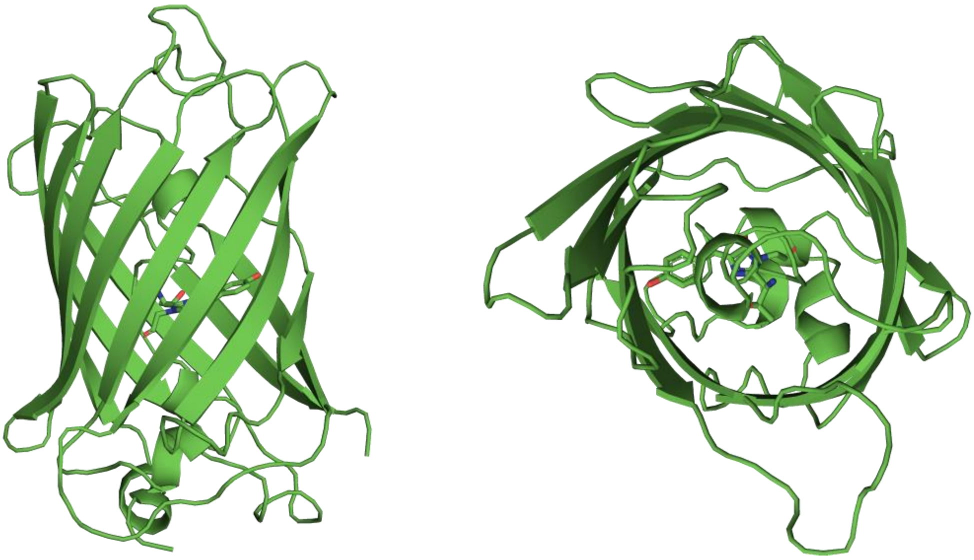


Figure 30. Side (left) and top (right) view of the GFP β-barrel structure (PDB ID: 1EMA).

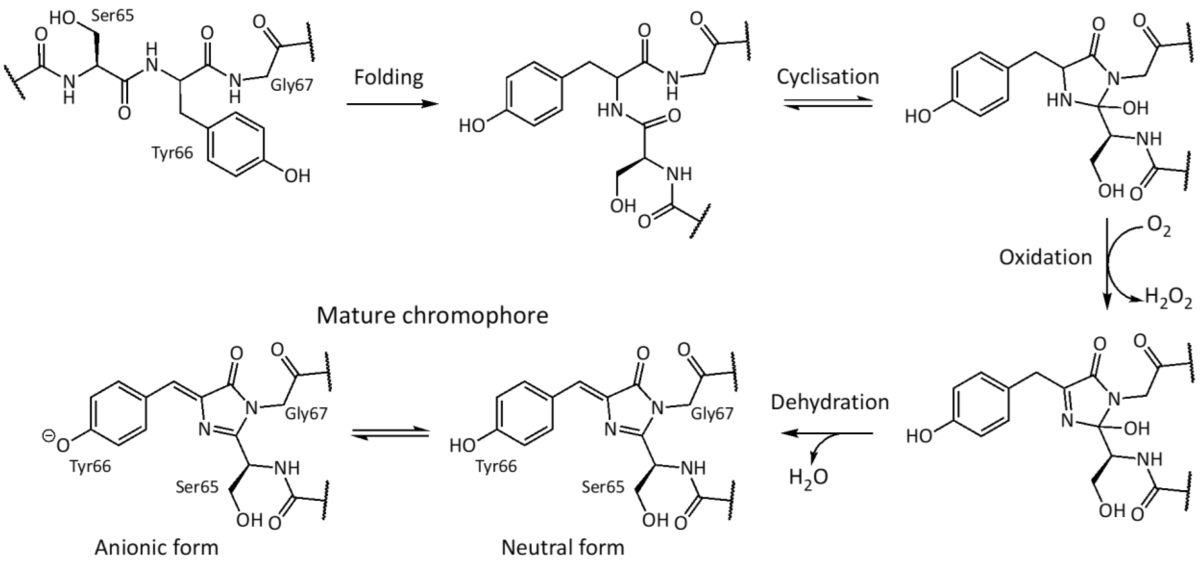


Figure 31. Mechanism of formation of the GFP chromophore. The oxidation and dehydration steps may be reversed depending on the oxygen concentration available. The neutral chromophore can be deprotonated into the anionic form.

The protonation state of the phenol group of the GFP chromophore dictates the absorbance properties.[149] The neutral chromophore absorbs with a λ_{max} of 395 nm and emits at 508 nm, whereas the anionic form absorbs at 475 nm and emits at 503 nm.[141] Both ionisation forms exist in an equilibrium in native GFP giving rise to two excitation peaks, with the one corresponding to the neutral phenol being the major peak.[150] The relative sizes of the excitation peaks vary with conditions such as pH, temperature and UV irradiation.[151]

The mutation of Serine to Threonine (S65T) causes a fourfold increase in the speed of chromophore formation (120 min for WT GFP to 27 min for S65T), which mitigates the effects of the slow forming chromophore for *in vivo* applications.[152] This mutation also abolishes the first excitation peak, while amplifying sixfold and slightly shifting the second excitation peak (489 nm) due to the stabilisation of the anionic form of the chromophore through hydrogen bonds. This increased brightness is a further improvement compared to WT GFP.[151] Ten additional mutations identified by a directed evolution approach gave rise to superfolder GFP, a variant with increased brightness and stability and more robust folding that displays higher resistance to chemical denaturation.[153]

Mutagenesis of GPF in the chromophore or adjacent amino acids leads to significant changes in the excitation and absorption wavelength (Figure 32). These variants have been used as distinct markers to monitor multiple biochemical processes simultaneously or as donors/acceptors for Förster-resonance energy transfer (FRET). The tyrosine that becomes part of the chromophore can be mutated to other aromatic amino acids to obtain variants with excitation and emission peaks at shorter wavelengths. A histidine in position 66 leads to blue fluorescent protein (λ_{ex} = 380 nm, λ_{em} = 445 nm) and a tryptophan, to cyan fluorescent protein (λ_{ex} = 440 nm, λ_{em} = 475 nm).[149, 154] Additionally, the mutation of threonine 203 to a polar aromatic residue (Tyr) results in π-stacking with the phenol group of the chromophore. This lowers the energy of the excited state causing a red shift, giving yellow fluorescent protein (λ_{ex} = 515 nm, λ_{em} = 525 nm).[155] However, the fluorescence intensity of these variants is significantly lower than that of WT GFP, and other mutations were needed to increase their brightness and stability.

A new fluorescent protein containing a GFP-like core with an additional double C=N bond was later discovered in the corallimorpharian *Discosoma*. Due to the extended π-electron conjugation, the absorbance and emission wavelengths are significantly red-shifted (λ_{ex} = 558 nm, λ_{em} = 583 nm) and the protein was named red fluorescent protein (RFP).[156, 157]

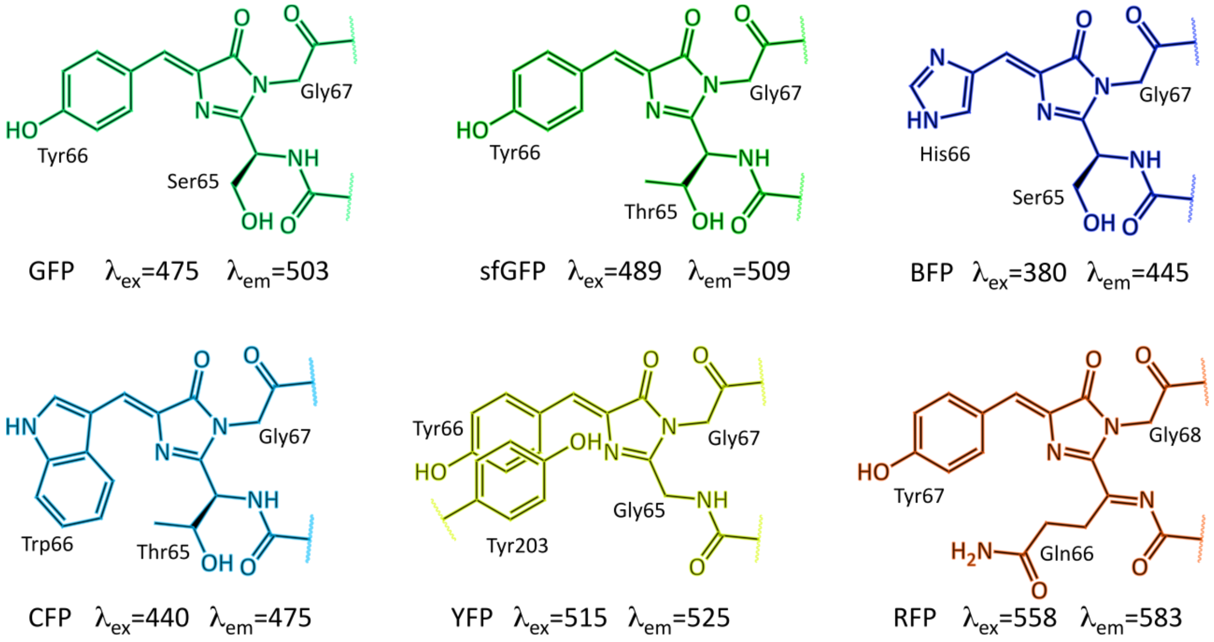


Figure 32. Chromophore structures of GFP and its derivatives along with their wavelengths of excitation and emission in nm.

Green fluorescent protein and its derivatives have been extensively used to monitor gene expression.[158] Many characteristics of GFP and its improved variants make them ideal for use as a reporter system. GFP is a small and compact protein that is able to generate its internal chromophore without the need for external cofactors or substrates other than molecular oxygen.[143] It is non-toxic to cells, stable at physiological ranges of temperature and pH and resistant to the presence of detergents, some organic salts and many proteases. It is easy to visualise in cells and generally stays active when fused to other proteins.[153] However, it also has some disadvantages that have encouraged the search for alternative reporter systems. The correct folding and forming of the chromophore are dependent on the presence and concentration of oxygen, which prevents its use in anaerobic conditions.[143] Although small in size (~26 kDa), it can have detrimental effects on the function or localisation of some co-expressed proteins.[159] For applications which require low quantities of GFP, the sensitivity is limited due to lack of signal amplification since each molecule of GFP produces only one chromophore. Therefore, strong promoters are needed to achieve sufficient fluorescence.[150] Additionally, the wavelength necessary for GFP excitation also causes autofluorescence, from other components that are naturally present in the cell, or the media used.[160] While fluorescent proteins with different excitation wavelengths do not have this problem, their brightness is lower than that achieved with GFP. GFP also has a tendency to oligomerise, which may lead to incorrect protein localisation or function.[161]

1.6.2 Fluorescent RNA aptamers

The fluorescence of GFP stems from the rigidity of the chromophore confined in the protein β-barrel. The isolated HBI chromophore is not fluorescent in solution due to the rotation of the phenyl group and the E-Z isomerisation of the double bond. By mimicking this rigidity effect, chromophore analogues have been developed through different methods over the past few years. This results in improved fluorescence, increased range of emission wavelengths, and introduces new potential applications.[162] These methods for increasing rigidity include introducing covalent bonds, [163] complexation with metal ions,[164] fusing to macromolecules such as polymers[165] or nucleic acids[166] to achieve organised stacking, and constraining in a scaffold such as metal organic frameworks[167] or nucleic acids[168].

This latter method using nucleic acids as scaffolds has given rise to the promising field of fluorescent aptamers. Aptamers are short, single-stranded oligonucleotides or peptides that can selectively bind to a specific target, such as ions, small molecules, carbohydrates, proteins or even living cells. Depending on their tertiary structure, they bind targets with high selectivity and specificity since they can wrap tightly around small molecules and serve as a scaffold or fit into gaps of much bigger molecules.[169] Typical motifs present in nucleic acid aptamers are stems, loops, hairpins, knots or G-quadruplex structures and the intermolecular interactions between aptamer and bases are mainly shape complementarity, stacking interactions, electrostatic interactions, or hydrogen bonds.[170] The role of the scaffold is to increase the rigidity of the ligand and therefore favour the radiative fluorescence decay pathway and prevent the non-radiative decay that was caused by the bond rotation when in solution. [168] This emerging field has found many applications in RNA imaging.

RNA plays a key role in cellular metabolism, being involved in a broad range of processes such as protein translation, gene regulation and catalysis of chemical reactions. Therefore, RNA imaging in living cell is essential to better understand the variety of RNA functions. Until recently, this remained a challenging task due to the drawbacks of the traditionally used techniques (Figure 33). Fluorescence *in situ* hybridisation enables the detection of single RNA molecules in fixed cells with the use of fluorescent complementary probes, but the need to fix the cells and remove unhybridised probes narrows the application to static measurements.[171] The MS2-GFP system consists of the MS2 RNA binding protein fused to GFP, and the MS2 binding stem-loop sequence introduced into the target RNA.[172] Even though it allows spatiotemporal measurements in live cells, it generates background fluorescence and the large modifications introduced could interfere with the normal function of the target RNA.[173] Molecular beacons exploit the use of FRET, where an RNA bound fluorophore, quenched when in a stem-loop structure, fluoresces after binding to target RNA sequence and separating from the quencher.[174] Although due to their small size they are unlikely to disturb normal function, they are exogenous probes that need to be generated *in vitro* and then introduced into cells.

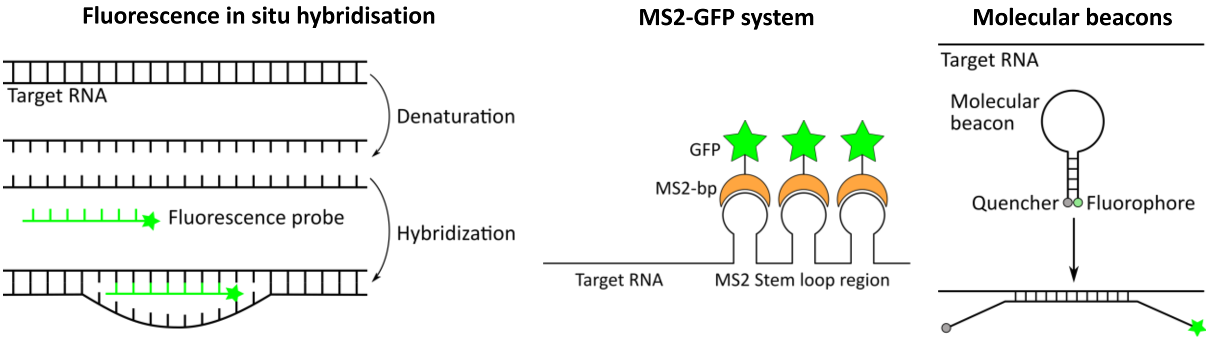


Figure 33. Three traditional techniques of RNA imaging: fluorescence in situ hybridisation, MS2-GFP system and molecular beacons.

The use of endogenously produced RNA would be optimal for live imaging but, since intrinsically fluorescent RNAs are not known, a fluorogen acts as the fluorescent reporter. Fluorescent RNA aptamers have gained importance in this field. The RNA aptamer sequences can bind cell-permeable small molecules and significantly enhance their fluorescence upon binding (Figure 34). Ideally, the fluorogen would be non-toxic and have low intrinsic fluorescence, a large fluorescence enhancement and high affinity and selectivity for the specific RNA sequence.[168] The aptamer sequences are selected through Systematic Evolution of Ligands by Exponential Enrichment (SELEX) (Figure 34). A large pool of RNA sequences is tested for binding to the fluorogen. Those RNA sequences that bind are amplified and enriched, and a new cycle is started. Multiple cycles results in the exponential increase of binding ligands.[175,176]

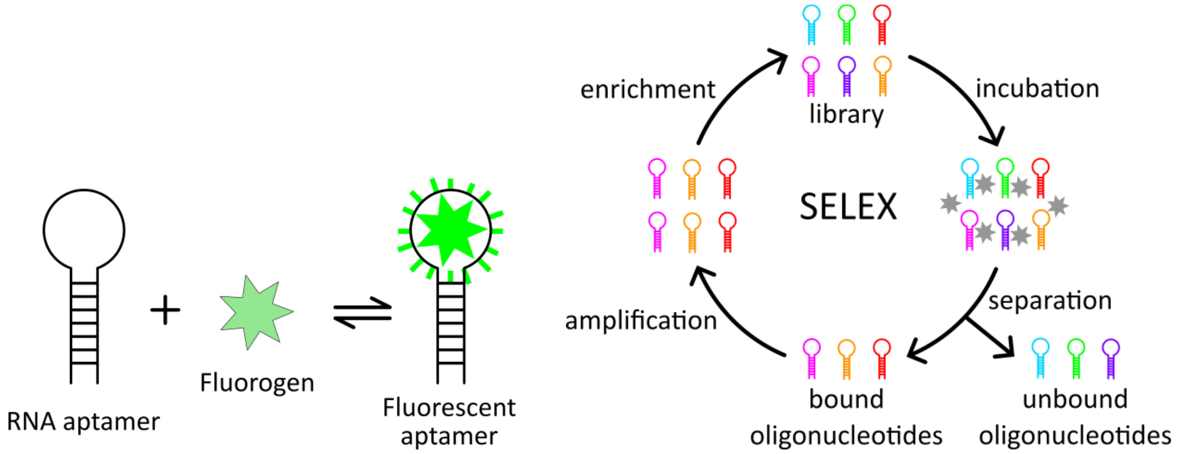


Figure 34. In fluorescent RNA aptamers, the fluorescence of a fluorogen is enhanced upon binding to the RNA (left). SELEX selection protocol for RNA sequences with high affinity to the fluorogen (right).

The first fluorescence aptamer was selective for the triphenylmethane dye malachite green (MG). This aptamer enhanced the fluorescence of malachite green (Figure 35) by three orders of magnitude. Although the nonspecific binding to other cell components and its cytotoxicity prevented *in vivo* applications, it paved the way for developing biologically compatible fluorescent aptamers.[177]

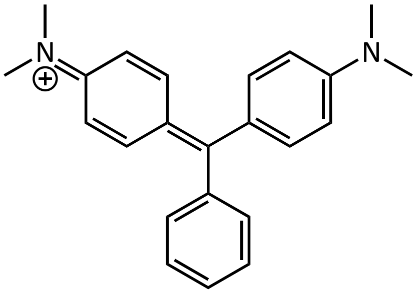


Figure 35. Structure of the dye malachite green.

The next fluorescent aptamers were developed using fluorogens derived from the GFP chromophore (Figure 36), since the ability of a tightly binding RNA should restrict molecular motions and restore their fluorescence. The aptamer Spinach[168] consists of DFHBI (3,5-difluoro-4-hydroxybenzylidene imidazolinone) and a 98-nucleotide long RNA sequence, which constrains the fluorogen in a planar conformation between a G-quadruplex, a base triple and solvent molecules.[178,179] While it has been successfully used in *E. coli* and mammalian cells, Spinach needs a tRNA scaffold to fold correctly *in vivo* and shows significantly lower brightness than *in vitro*. Six mutations lead to Spinach2, an aptamer with almost identical photophysical properties that shows increased melting temperature and improved folding in vivo, leading to a higher brightness in cell imaging.[180]

A modified SELEX protocol, which includes expression in *E. coli* and FACS screening in the presence of the fluorogen at 37°C, was used to create the Broccoli aptamer. It is made of a 49 nucleotide RNA that folds forming a G-quadruplex structure, and the improved fluorogen DFHBI-1T, with an additional trifluoromethyl group that increases brightness and causes a slight redshift in fluorescence excitation and emission. These improvements in the selection process ensure increased fluorescence upon binding, compatibility with cellular expression, high thermostability and low magnesium dependence.[181]

A new fluorogen was later developed to contain an oxime, a stable version of the N-acyl imine from the RFP fluorophore, to avoid the photoisomerisation that decreased the photostability of DFHBI based aptamers.[182] Corn aptamer was obtained by applying the modified SELEX protocol on DFHO (3,5-difluoro-4-hydroxybenzylidene imidazolinone-2-oxime), resulting in a 36-nucleotide long RNA sequence that stabilises DFHO in the interface of an RNA homodimer, with a four-tiered G-quadruplex on each side flanked by adenine flaps.[183] Corn exhibits significantly red-shifted fluorescence and increased photostability. However, due to its dimeric nature and low thermostability, it is not suitable for imaging in live cells.[184]

The fluorophores of Spinach, Broccoli and Corn are mostly deprotonated at physiological pH, but HBI derivatives with higher pKa can be present as the phenol form and show large red shifts. An RNA derived from the selection process of Spinach aptamer was optimised to bind 3,5-dimethoxy-4-hydroxybenzylidene imidazolinone (DMHBI), giving rise to the Chili aptamer.[185] A positively charged trimethylammonium phenyl derivative (DMHBI$^+$) greatly increases fluorescence enhancement by preventing the non-radiative deactivation of the excited state. The introduction of an oxime (DMHBO$^+$) does not vary the fluorescence enhancement compared to Chili-DMHBI but causes a bathochromic shift, leading to the longest wavelength emitting aptamer in the HBI-binding family.[186] This RNA aptamer forms a two-tiered G-quadruplex that holds the ligand in a slightly twisted conformation. The oxime moiety forms additional hydrogen bonds with the phosphate backbone, causing a higher affinity and thermostability in Chili-DMHBO$^+$.[187]

To avoid the complex G-quadruplex folding, a new approach consisting in evolving a naturally occurring well-folding RNA aptamer using SELEX was developed. This resulted in Squash, evolved from an adenine-binding aptamer to bind DFHBI-1T without a G-quadruplex. Squash exhibits a high folding efficiency *in vivo,* and improved photostability and fluorophore affinity compared to Broccoli and Corn, and it also induces the fluorescence of DFHO.[188,189]

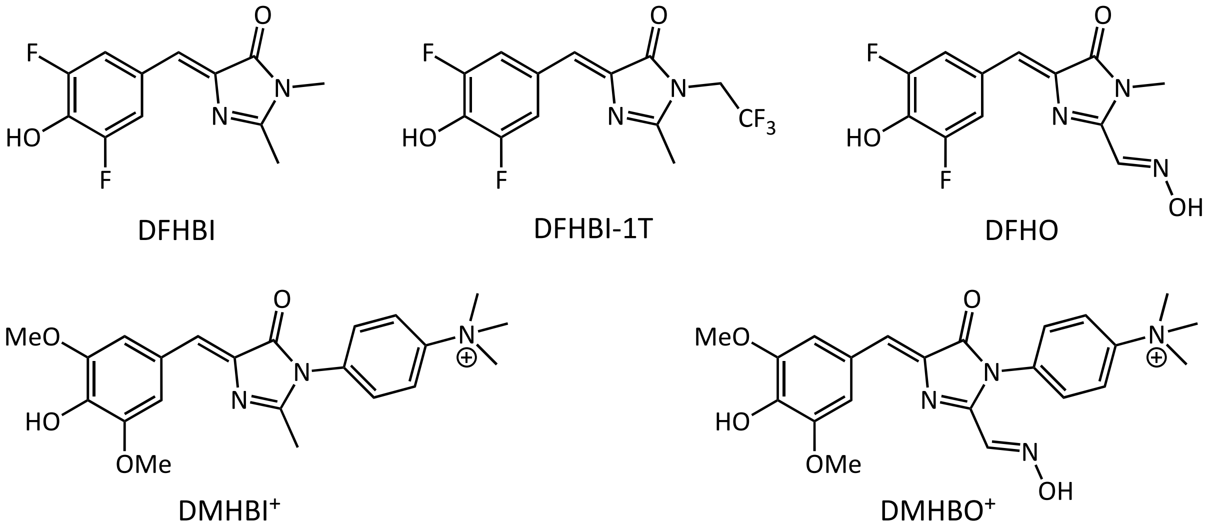
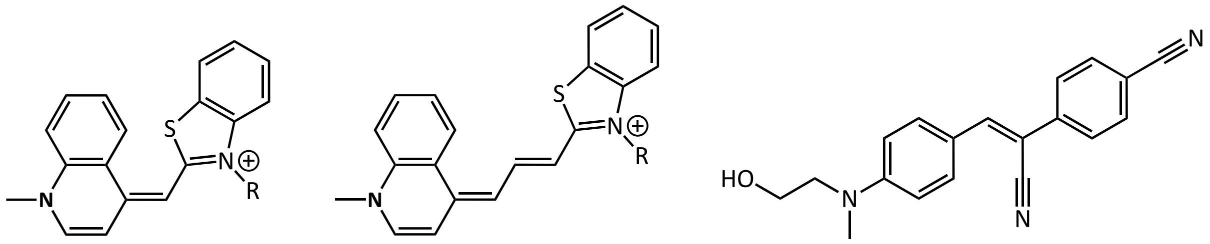


Figure 36. HBI derivatives used in fluorescent aptamers.

Fluorescent aptamers based in distinct fluorogens instead of GFP mimics have also been developed (Figure 37). Thiazole Orange (TO1) is a small, non-toxic, cell permeable cyanine fluorophore that exhibits very low fluorescence in solution but becomes strongly fluorescent when the structure is rigidified. A PEG-Biotin moiety was added to destabilise nucleic acid intercalation and decrease cellular background fluorescence. An RNA sequence was then selected, aiming for very tight binding and high fluorescent enhancement.[190] The resulting Mango aptamer forms a three-tiered G-quadruplex that stabilises the TO1-Biotin fluorogen in its top face.[191]

A GFP inspired synthetic dye, (4-((2-hydroxyethyl)(methyl)amino)-benzylidene)-cyanophenyl-acetonitrile (HBC), containing a rigid electron acceptor and a strong electron donor that fluoresces upon constraint was used to select a RNA sequence that greatly enhanced its fluorescence. The optimisation of this sequence lead to the Pepper aptamer, which shows remarkable fluorescence enhancement both *in vitro* and *in vivo* and increased thermal and folding stability.[192] By tuning the π-conjugation and modifying the electron donor/acceptor ability of the dye, a series of analogues covering a broad spectral range that maintained high affinity for Pepper RNA were designed. The Pepper aptamer binds all analogues in a similar manner, it forms a four-sided box binding pocket in the centre of the overall structure that holds the fluorogen in a quasi-planar conformation. It does not contain G-quadruplexes, leading to the increased stability and folding efficiency in live cells.[193]



TO1-Biotin TO3-Biotin HBC

Figure 37. Other fluorogens used in fluorescent aptamers.

The nature of fluorescent aptamers gives them certain advantages over the use of fluorescent proteins for imaging applications: they have low background fluorescence in the absence of RNA, they tend to have low to no interference with cellular function due to their small size, they can be reversibly thermally denatured, and they are easy and quick to produce inside the cell. Thanks to these benefits, in addition to being used for direct imaging of RNA transcripts,[168] they have also been exploited as the recognition element in biosensors for *in vivo* monitoring of small molecule metabolites or of bimolecular interactions (RNA–RNA or RNA–protein) utilising FRET based systems.[194,195] The improvement in photostability of the more recent aptamers warrants further investigation on their ability to assess the activity of novel engineered constructs in optogenetic systems.

1.7 Project aims

The overall aim of the project was to harness light as a stimulus using LOV domain-based transcription factors as tools. These proteins, key players in the field of optogenetics, allow the study and control of biological processes in a non-invasive and precise manner. In this work, the LOV domain-containing protein EL222 was engineered to alter its affinity for DNA and be able to activate the transcription of a target gene. Additionally, the fluorescent RNA aptamer Pepper was evaluated as a reporter of transcription in optogenetic systems. Finally, the signal transduction mechanism of Aureochrome LOV domain was studied to achieve a better understanding of the light activated structural changes, which could assist with the rational design of novel and improved optogenetic tools

Chapter 2 - Investigating the use of pepper RNA aptamer as a reporter of transcription in optogenetic systems

2.1 Introduction

Fluorescent RNA aptamers consist of a fluorogen, and an RNA sequence selected to enhance its fluorescence.[168] *In vitro* applications of fluorescent RNA aptamers include identification of ligands/metabolites in a solution, visualisation of RNA in native gels and affinity purification of tagged RNA or proteins. *In vivo* applications encompass real time monitoring of gene transcription, interaction studies of RNA with RNA, proteins or small molecules; detection of RNA sequences; or monitoring the localisation and dynamics of metabolites and aptamer-tagged proteins.[195] While fluorescent RNA aptamers overcome some limitations of fluorescent proteins, some key properties need to be considered (Table 1) such as brightness, thermal and photostability, affinity, wavelength of emission, and robust expression and correct folding in cells. Limited brightness, thermostability or photostability would hinder their use for general imaging applications *in vivo*, particularly for low abundance RNAs. To avoid this, a strong interaction between the RNA and the fluorogen is desired. A broad spectral range coverage is desired in aptamers to avoid overlap with cell autofluorescence, which falls in the blue-green range, and to be able to use two distinct systems simultaneously.[160]

Many aptamers contain G-quadruplexes in their structure, which consist of two or more tiers of planar guanine tetrads coordinated to potassium ions. Their planar arrangement is beneficial for binding the fluorogen and restricting conformational changes, inducing the activation of fluorescence.[196] However, the presence of these motifs in aptamers can lead to difficulties in folding due to their tendency to be polymorphic.[197] This, along with the high dependence on potassium ions, can affect their use for *in vivo* applications.

	Length (nt)	Fluorogen	$\lambda_{ex\ max}$ (nm)	$\lambda_{em\ max}$ (nm)	Fold activation	K_d (nM)	Photostability (s)	Tm (°C)	Year
MGA[177]	38	MG	630	650	2360	117	n.d.	45	2003
Spinach[168]	98	DFHBI	467	501	2000	537	<1	34	2011
Broccoli[181]	49	DFHBI-1T	472	507	960	360	<1	48	2014
Corn[184]	36 (x2)	DFHO	505	545	420	70	10	33	2017
Chili[185]	52	DMHBI+	400	540	625	63	n.d.	43	2019
	52	DMHBO+	455	592	125	12	n.d.	50	2019
Squash[188]	83	DFHBI-1T	452	503	1000	45	<1	48	2021
	83	DFHO	495	562	500	54	3	50	2021
Mango[190]	29	TO1-Biotin	510	535	1100	3	5	40	2014
	29	TO3-Biotin	637	658	35	7	n.d.	40	2014
Pepper[192]	43	HBC$_{485}$	443	485	691	8.0	>3600	66	2019
	43	HBC$_{497}$	435	497	16601	6.7	<1	58	2019
	43	HBC$_{508}$	458	508	9091	27.0	<1	45	2019
	43	HBC$_{514}$	458	514	4748	12.0	<1	54	2019
	43	HBC$_{525}$	491	525	585	3.8	<1	63	2019
	43	HBC$_{530}$	485	530	3595	3.5	<1	55	2019
	43	HBC$_{599}$	515	599	708	18.0	>4700	55	2019
	43	HBC$_{620}$	577	620	12600	6.1	>30000	66	2019

Table 1. Properties of different fluorescent RNA aptamer systems. The colour of each row represents the emission wavelength of that aptamer system.

2.1.1 Pepper aptamer system

Among the aptamer systems developed and described in Table 1, Pepper stands out for its broad range of emission maxima, high affinity between RNA and fluorogen, high fold activation, lack of G-quadruplex in its structure, high melting temperature and, for three of the fluorogens, remarkable photostability. The Pepper aptamer system (Figure 38) is comprised of a fluorogen, (4-((2-hydroxyethyl)(methyl)amino)-benzylidene)-cyanophenyl acetonitrile or HBC$_{530}$, that has a rigid electron acceptor and a strong electron donor; and a 43 nucleotide long RNA oligo identified by SELEX to bind said fluorogen.[192]

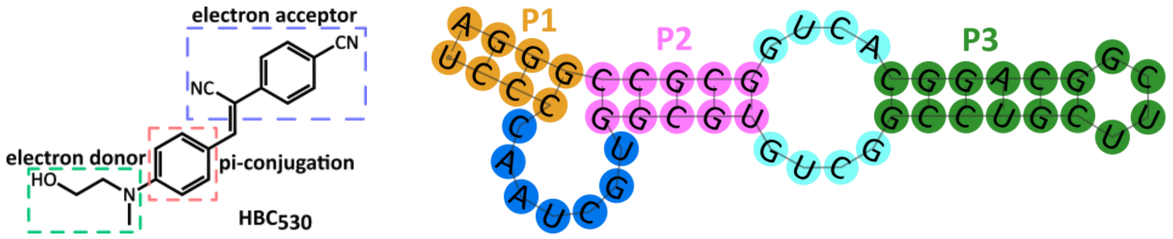


Figure 38. Pepper aptamer system. Structure of (4-((2-hydroxyethyl)(methyl)amino)-benzylidene)-cyanophenyl acetonitrile and secondary structure of the Pepper RNA.

Tuning the π-conjugation and modifying the electron donor/acceptor ability of the fluorogen gives rise to a series of analogues covering a broad spectral range (Figure 39). Increasing the conjugation of the system reduces the HOMO-LUMO gap, and therefore produces a redshift. This approach yielded HBC$_{599}$ and HBC$_{620}$. On the contrary, decreasing the electron density of the π-conjugation system and the increase in the HOMO-LUMO gap causes a blueshift (HBC$_{497}$ and HBC$_{508}$). This latter effect can also be achieved by decreasing the electron withdrawing capability of the electron acceptor group (HBC$_{525}$) or decreasing the electron donating ability of the electron donor (HBC$_{485}$ and HBC$_{514}$). These analogues maintained high affinity for Pepper RNA, and some even showed improved fluorescence activation, photostability and thermostability.[192]

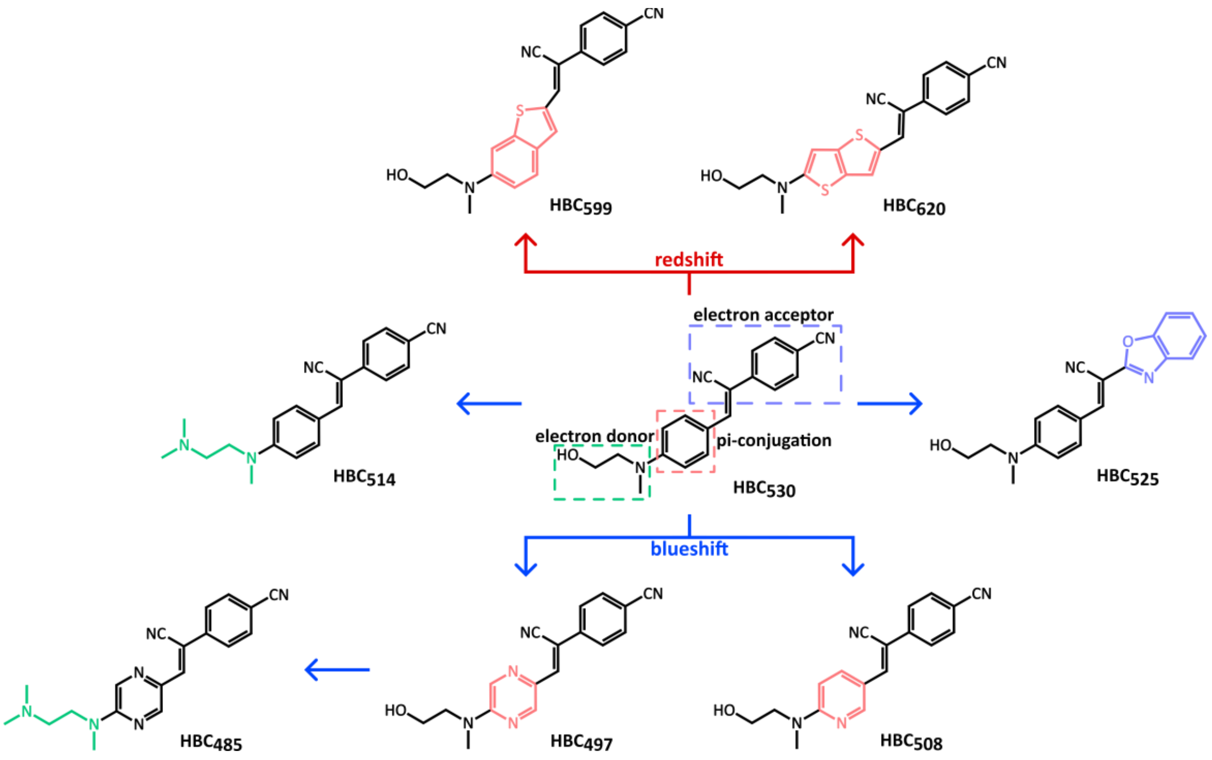


Figure 39. Strategies of fluorogen modification to obtain analogues with different spectral properties. The colour of the arrows indicates a redshift (red) or blueshift (blue) in emission wavelength.

The pepper RNA adopts a helical structure with one protruded region and includes several magnesium ions that participate in the tertiary folding. It contains three stems connected by three junctions (Figure 40).[193] Junction 3/2 extrudes from the helix and is positioned by the major groove of stem P2. The fluorogen is intercalated in the middle of the structure in a quasiplanar conformation. It binds at the intersection of P2 and the junctions, in a four-sided box shaped binding pocket formed by eight nucleobases: the top face is a noncanonical base pair (G10, U40), the bottom face is a tetrad (G41, C43, U8, U42), one side is a guanine (G9) and a cytosine (C33), and the other side is a phosphate. Those analogues containing a terminal –OH group form a hydrogen bond with a nitrogen of the guanine 41. However, the longer length of thienothiophene disrupts that hydrogen bond formation in HBC_{620}. The substitution of the hydroxy group for an amine makes the ligand lose that hydrogen bonding interaction.

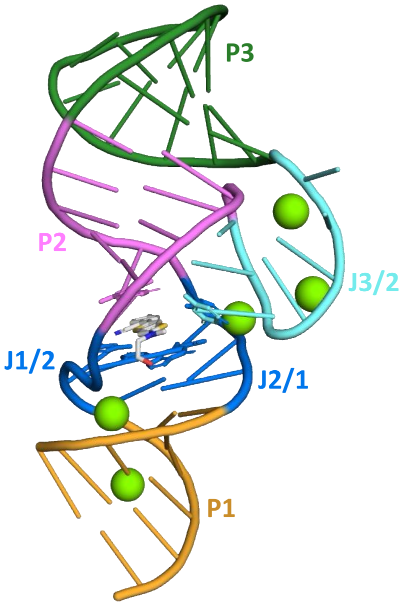


Figure 40. Tertiary structure of Pepper aptamer in complex with HBC_{620} (PDB ID: 7EOP).

Apart from direct reporting of transcription, this aptamer system has been used to visualise protein tethering and trans-localisation in mammalian cells,[192] as a biosensor to recognise specific sequences of endogenous intracellular RNA[198] and to measure the dynamic changes in the levels of SAM and other metabolites in real time in live cells.[199]

2.1.2 Aim

The aim of the work described in this chapter was to study the ability of the Pepper aptamer system to report transcription initiated by a LOV-domain based transcription factor, namely EL222. The presence of FMN in LOV domains grants them a weak intrinsic fluorescence that is lost upon adduct formation.[200] The excitation (λ_{ex} = 450 nm, 475 nm) and emission (λ_{em} = 490 nm, 535 nm) peaks overlap with GFP autofluorescence. While GFP has successfully been used for reporting on LOV based optogenetic tools,[5,72,140,201] the use of RNA aptamers may improve results by avoiding fluorescence overlap and overcoming the previously mentioned limitations of fluorescence proteins.

The HBC_{620} – Pepper RNA system possesses excitation and emission wavelengths that do not overlap with cell autofluorescence and the intrinsic fluorescence of LOV domains. The system also presents high affinity between fluorogen and RNA, remarkable fluorescence activation upon binding and increased photostability, necessary when illumination is required to activate transcription.[192] These properties make it worth investigating for reporting transcription in optogenetic systems.

2.2 Results

2.2.1 Synthesis of HBC_{620}

The synthesis of fluorogen 5-[(2-hydroxyethyl)methylamino]thieno[3,2-b]thiophene-2-(p-cyanophenylacetonitrile) (compound **9**, HBC_{620}) was initially attempted following the synthetic route outlined in by Chen *et al*,[192] using thieno[3,2-b]thiophene as a starting material and introducing the substituents to the aromatic system as depicted in Figure 41.

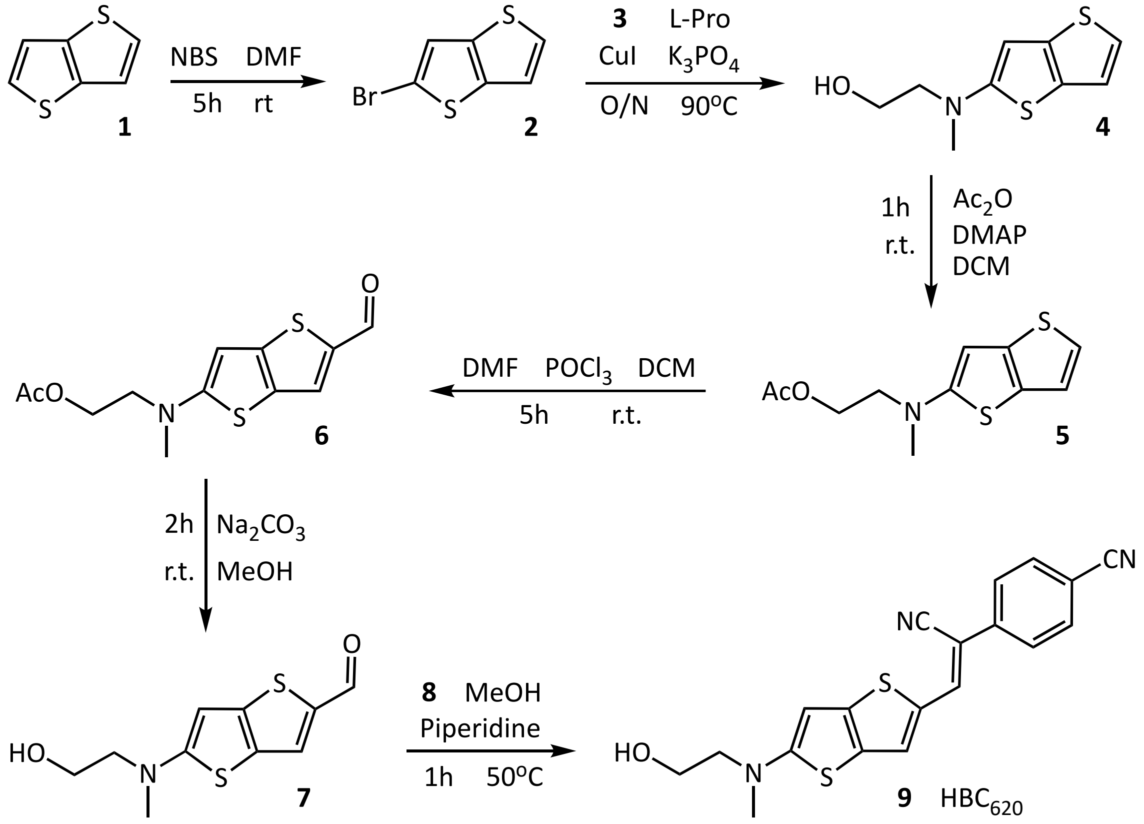


Figure 41. Synthetic route to HBC_{620} described by Chen et al.[192]

The first substituent is introduced on the alpha carbon of thieno[3,2-b]thiophene, the most reactive position towards electrophilic substitution. This can be illustrated by the position of the negative charge in the resonance forms (Figure 42).[202]

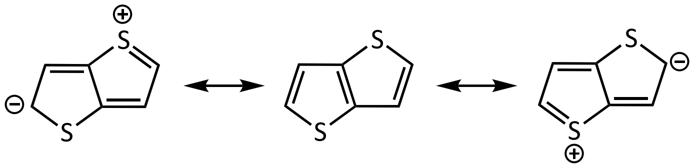


Figure 42. Resonance forms of thieno[3,2-b]thiophene.

The result of the bromination of compound **1** was analysed by GC-MS (Figure 43). Compound **2** (11.57 min) was obtained as the main product of this reaction (69% yield). However, a small amount of starting material **1** (7.51 min) was left unreacted and a small fraction of the reaction progressed further to yield the dibrominated compound (15.23 min).

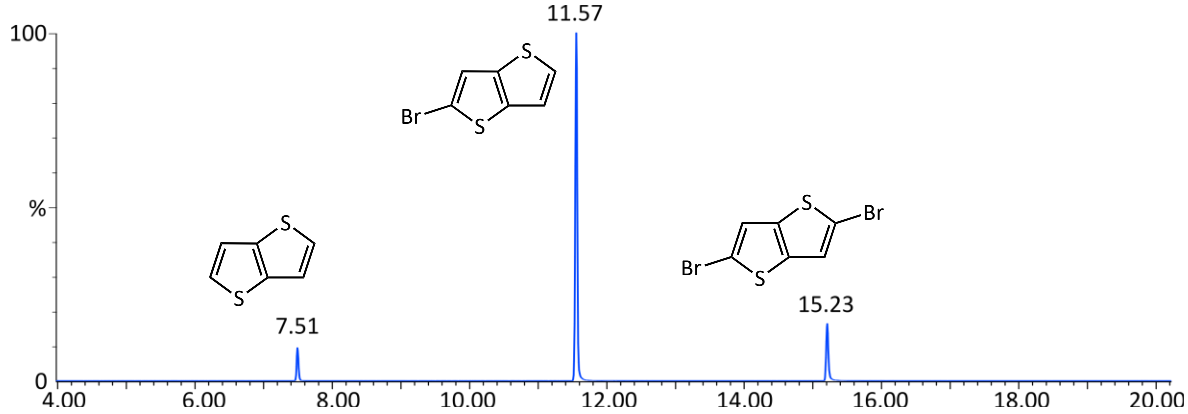


*Figure 43. GC-MS spectrum after bromination of compound **1** without further purification.*

The Ullmann type C-N coupling reaction resulted in a very low yielding (5%) formation of compound **4**. A possible cause for the reduced yield is the light, air, and moisture sensitivity of copper iodide. The purification of CuI by boiling in NaI solution[203] before performing the coupling reaction increased the yield to 25%, but this was still distant from the 79% reported in the literature.[192] The inert atmosphere required for this reaction was achieved by using an argon balloon since a Schlenk line was not available. It has been proven that air slowly diffuses into the balloon over time, not fully maintaining the inert atmosphere[204] and this non-optimal set up could be the cause of the low yield.

The protection of the alcohol group was required due to incompatibility with the Vilsmeier-Haack reagents, as their addition in the presence of the deprotected alcohol would lead to its chlorination.[205] Acetic anhydride was used to convert the primary alcohol into an ester to avoid this undesired reactivity.

The fourth step was a Vilsmeier-Haack formylation to introduce an aldehyde onto the aromatic system. After multiple attempts and testing different conditions, the aldehyde group could not be directed to the position described in the literature.[192] Instead, it was mainly being directed to the carbon adjacent to the amino substituent (Figure 44). This was identified by NMR thanks to the multiplicity of the aromatic protons, doublets with a coupling constant of 5.25 Hz, characteristic of adjacent protons in thienothiophene.[206] (Figure 45). A small amount of dialdehyde product, with the additional aldehyde group in the α position of the second thiophene ring was also formed (Figure 46). The amino substituent has an activating effect of the aromatic system due to the presence of the free electron pair, and it mainly directs electrophiles to its adjacent carbon in the aromatic system. The slight excess of $POCl_3$ allowed the formation of the dialdehyde product, with the additional aldehyde group on the next most reactive position (α carbon of the second ring).

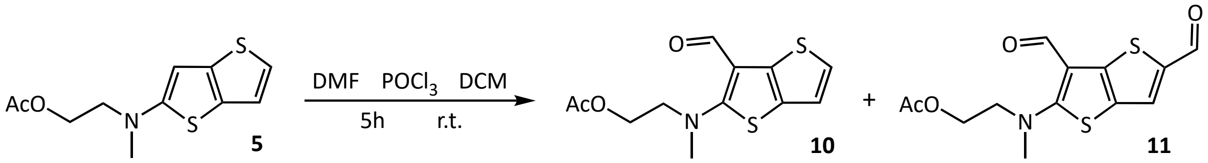


*Figure 44. Results of the Vilsmeier-Haack formylation on compound **5**.*

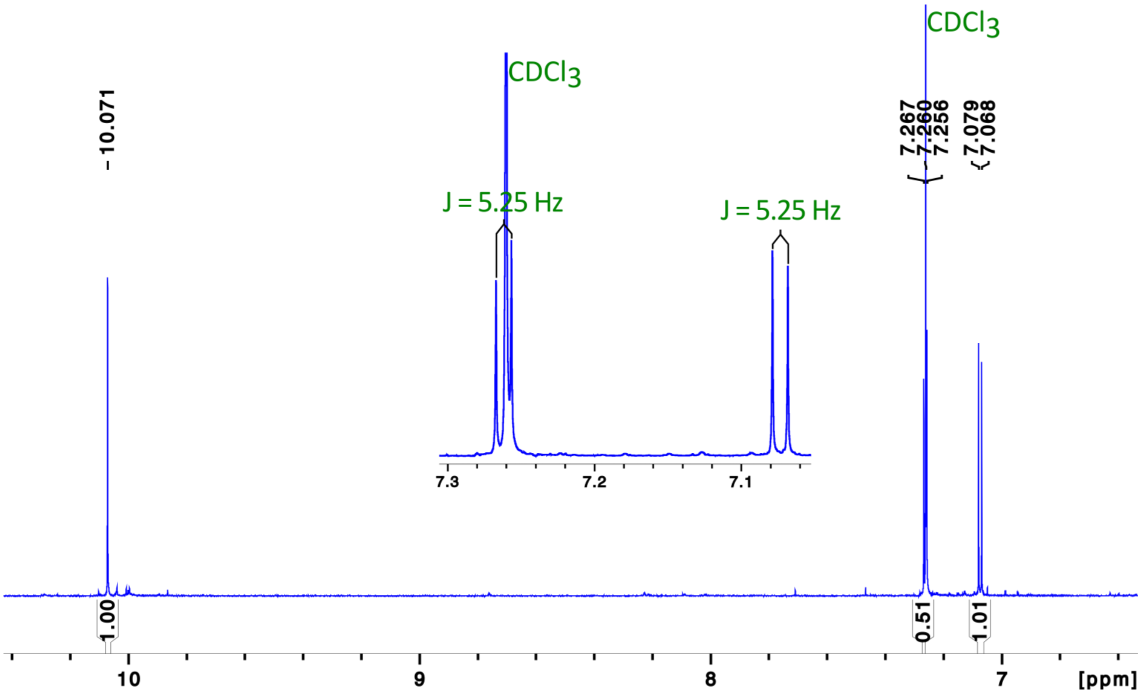


*Figure 45. ^{1}H-NMR spectrum (7 – 10 ppm) of compound **10**. The multiplicity and coupling constant of the aromatic hydrogens indicates the remaining protons are adjacent.*

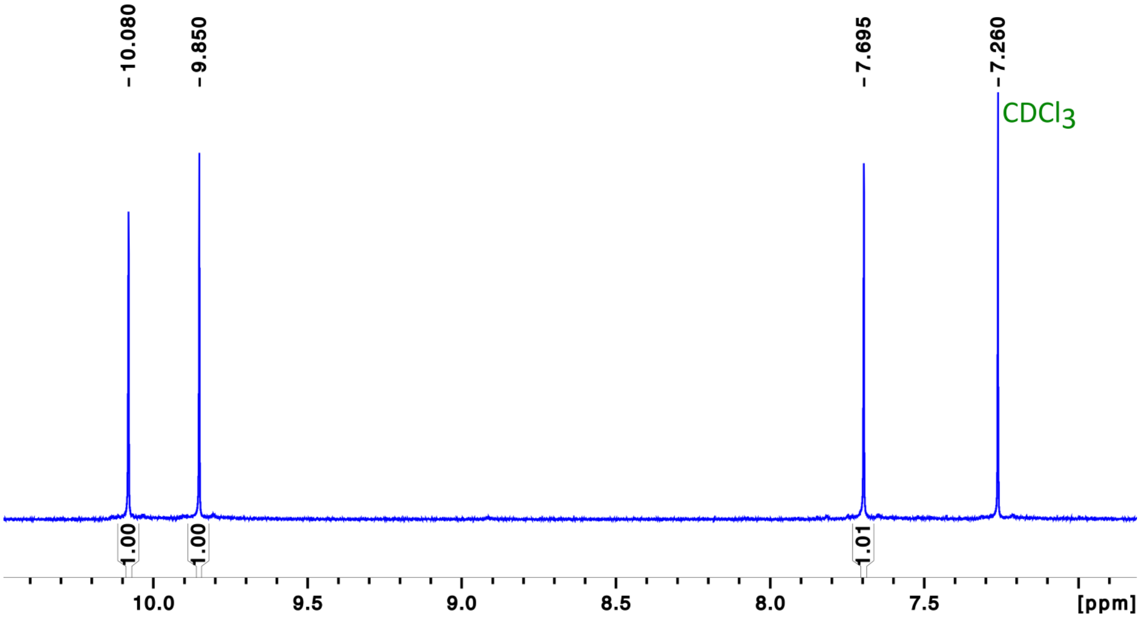


*Figure 46. ^{1}H-NMR spectrum (7 – 10 ppm) of compound **11**. The two peaks at ~10ppm mark the presence of two aldehyde groups.*

This problem led to the design of a new synthetic route (Figure 47), beginning with a Vilsmeier–Haack formylation, followed by a bromination to achieve the correct positioning of the substituents. The aldehyde group has a deactivating effect on the aromatic system due to its electron withdrawing character, the adjacent carbon is disfavoured, and the bromine atom can be introduced in the desired position. The reversion of the order of these two steps would result in an incorrect product. While the bromine atom also has a deactivating effect, it favours electrophilic substitution in the adjacent carbon thanks to its ability to donate a free electron pair to form a pi bond. Once the substituents have been correctly positioned, they can be modified to obtain the final product. The order of the next two reactions is likely to be irrelevant but it was decided to perform the amination reaction first as it might be beneficial. Since the reagent (2-methylamino)ethanol is in excess, an equivalent could form an iminium ion with the aldehyde before the amination occurs and favour the reaction due to the aromatic system now being less electron rich. The final Knoevenagel condensation adds p-cyanophenylacetonitrile (**8**).

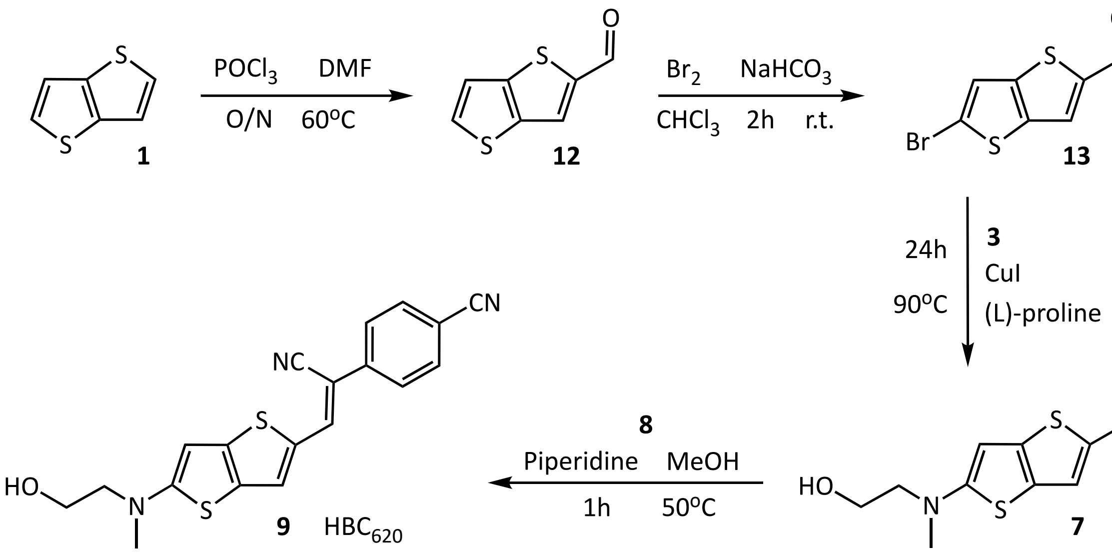


Figure 47. New synthetic route to compound HBC$_{620}$.

This new synthetic route gave better results than the previous reported synthesis. It comprised only four steps instead of six since the protection and deprotection steps were no longer needed, and the substituents were directed to the correct positions (characterisation in Appendix). However, the yield for the amination of the brominated substrate was still low due to the limitations in the inert atmosphere set up. Additionally, some degradation could be seen on silica plates due to the appearance of additional spots with time. It is therefore possible that, while purifying via silica column, part of the product was lost, contributing to the low yield.

The purification of the target compound **9** proved challenging due to the recurrent coelution of the starting material and product in multiple solvent systems. Compounds **7** and **9** were separated in TLC using 3% MeOH in DCM as the mobile phase, but they coeluted in flash chromatography with the same solvent system. Methanol is a protic solvent and the formation of hydrogen bonds with the compounds can lead to an unpredictable elution in flash chromatography that does not always correlate to the results seen on TLC. Acetonitrile has been described as an effective replacement for methanol in normal phase chromatography due to its polar and aprotic character, that results in a higher predictability when creating a flash chromatography method based on TLC results.[207] Ultimately, the solvent system 5% MeCN in DCM was able to separate compounds **7** and **9** both in TLC and flash chromatography. However, the duplicated appearance of the signals on the ^{1}H-NMR spectrum of the purified compound indicated the presence of two species which were identified as isomers (Figure 48).

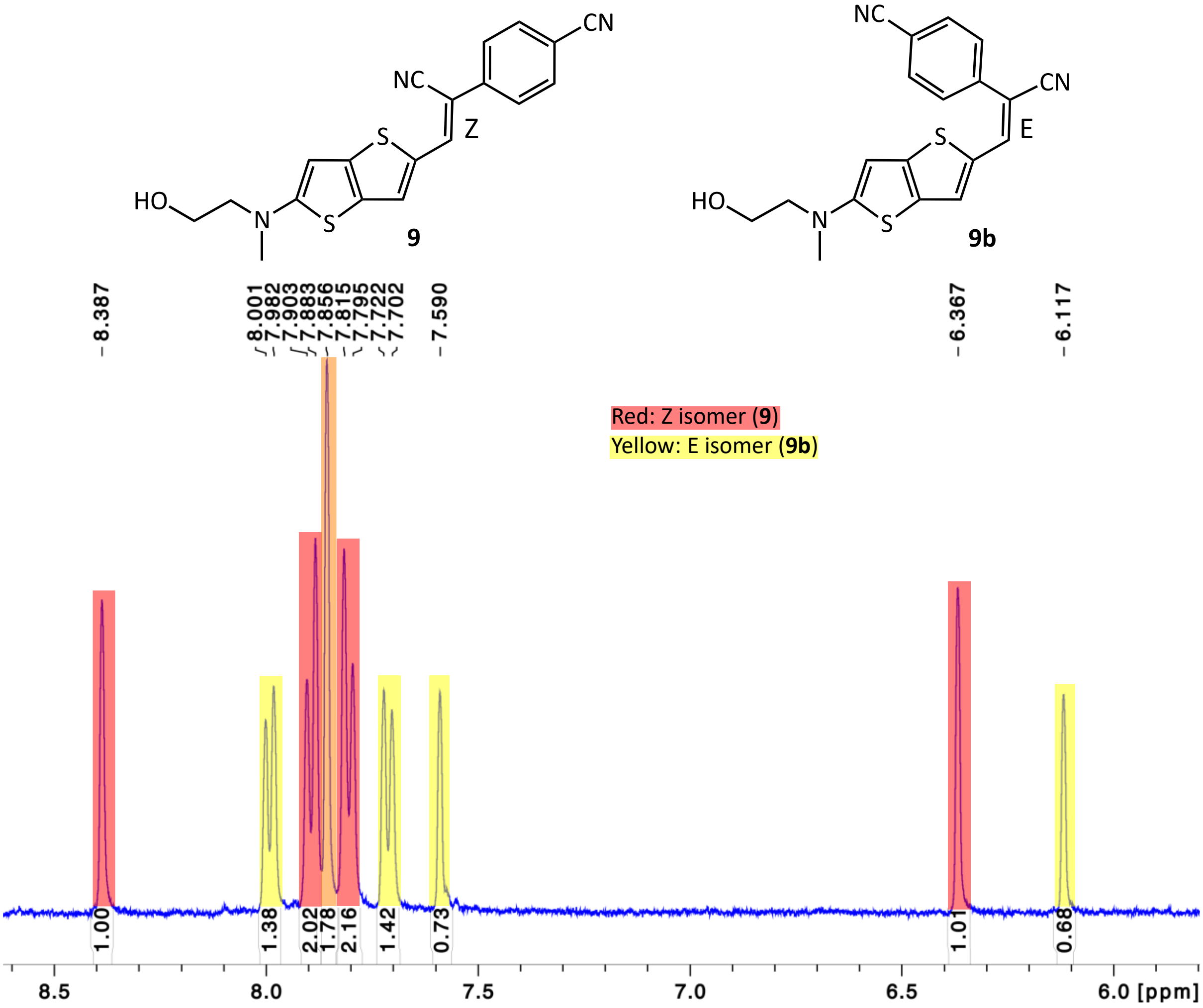


Figure 48. ^{1}H-NMR spectrum (6 – 8 ppm) of the two isomers and their chemical structures. The duplication in the expected NMR signals indicated the presence of isomers.

Initially, attempts were made to isomerise **9b** (E) to **9** (Z), to utilise the amount of **9b** that had already been synthesised. Iodine and p-toluenesulfonic acid, known to catalyse isomerisation reactions,[208,209] were individually added to a solution of the isomer mixture and the changes were monitored by NMR. However, no changes in isomer ratio were visible. Secondly, the reaction conditions of the final step of the synthesis were varied to reduce the proportion of **9b** obtained. Different temperatures (0°C, 5°C, 25°C and 50°C) and different quantities of catalytic piperidine (0.5 and 1 drops) were tested. For the reactions at 0°C, no product was visible on TLC after 1h, so one reaction was carried out at 5°C overnight. The results obtained from these reactions could not be rationalised (Table 2).

	5°C	**r.t**	**50°C**
0.5 drop	-	35% Z	-
1 drop	50% Z	80% Z	60% Z

*Table 2. Percentage of **9** obtained in Knoevenagel reaction in different conditions.*

After failing to obtain exclusively compound **9**, separation of the two isomers became the objective. This separation could not be achieved using standard TLC, so silver coated TLC, shown to be effective in separation of complex isomeric mixtures, was tested.[210] Though silver coated TLC led to visible albeit small separation of the isomers, this was not reproducible in silver coated microscale (pipette) column chromatography.

Since the isomers could not be manually separated, HPLC was next attempted. The sample was first analysed by a preliminary run using a rapid 0-100% gradient of acetonitrile in water (Figure 49 left). The presence of a shoulder on the peak implied that separation by HPLC would be possible after narrowing the gradient. An optimised run was carried out with a gradient of 60-100% acetonitrile in water, which resulted in baseline separation of the two isomers (Figure 49 right).

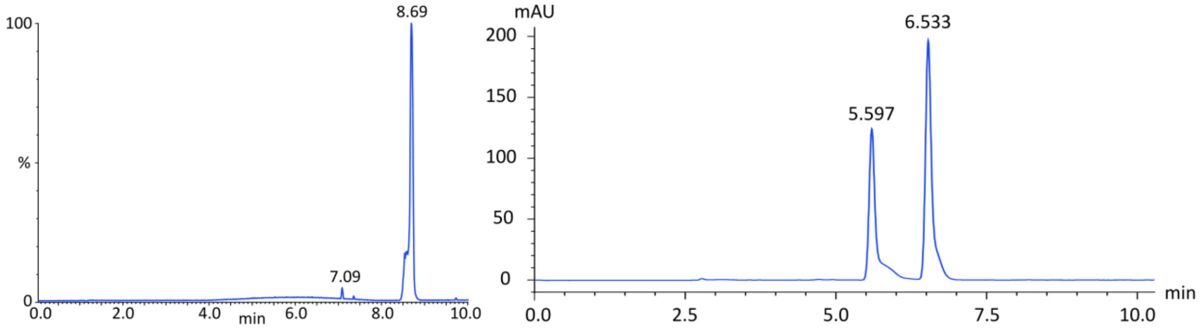


*Figure 49. Chromatograms of the separation of **9** and **9b** in gradients of water and acetonitrile. Left: 0-100%, right: 60-90%.*

The fractions of each isomer were collected separately, with the first fraction presenting a red colour and the second, a light pink. The solvents were removed, and the samples were analysed by ^{1}H-NMR in d$_6$-DMSO. Both fractions showed a mixture of isomers with different proportions rather than the expected single isomers (Figure 50 left). For simplicity, the Z/E isomer ratio was compared in the 6.0 – 6.5 ppm region in ^{1}H-NMR. After being exposed to ambient light overnight, both samples had the same colour and had reached the same Z/E ratio (Figure 50 middle). This suggested that isomerisation was taking place over time. To investigate if light was causing the cis/trans isomerisation, both samples were kept in the dark (covered in foil) overnight. The increased proportion of the Z isomer (**9**) in both samples (Figure 50 right) indicated that the compound photoisomerised. The E isomer was favoured under white light conditions and the Z isomer was favoured in the dark.

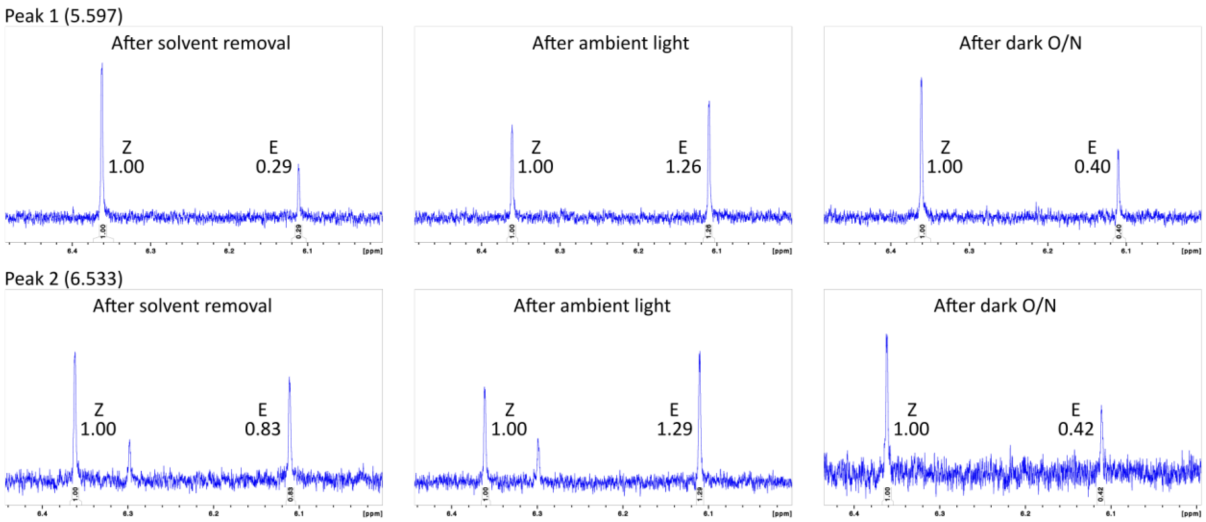


Figure 50. Isomer ratio on both HPLC fractions after solvent removal (left). Illumination with ambient light overnight (middle) led to an increase in E isomer, reaching similar levels in both samples. After relaxation in the dark overnight (right), the amount of E isomer was reduced.

The activity of this photoswitch was further studied using ^{1}H-NMR spectroscopy. Solutions of **9** kept in the dark were illuminated with red (Figure 51) and blue (Figure 52) light and the extent of isomerisation was compared. Both red and blue light had an effect on the isomerisation upon illumination. Complete isomerisation to **9a** (E) was not observed but a maximum Z : E ratio (1 : ~1.2) was reached in all tested conditions (red light, blue light and ambient light). Red light induced isomerisation more rapidly as illumination for one hour was enough to reach the maximum ratio, whereas ambient or blue light necessitated multiple hours to reach this level of isomerisation.

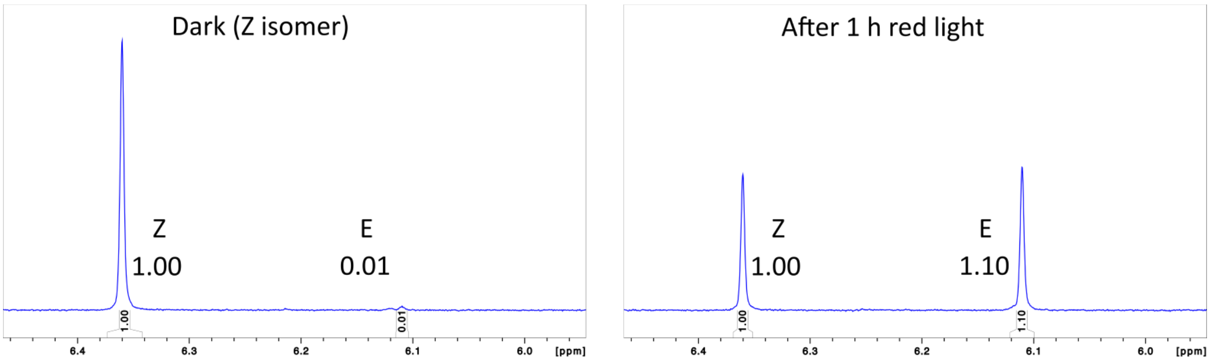


Figure 51. Change in isomer ratio after one hour of red-light illumination. A significant amount of E isomer was observed, almost reaching the maximum Z : E ratio.

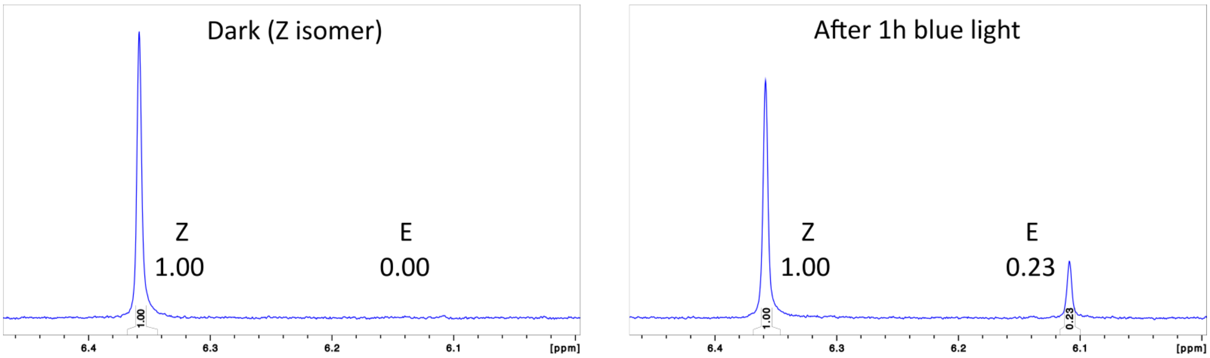


Figure 52. Change in isomer ratio after one hour of blue-light illumination. A small proportion of E isomer was formed, showing a slower conversion than with red light.

The rate of the conversion from the maximum isomeric mixture to the Z isomer in the dark was followed by NMR and the half-life was found to be 6.5 h (Figure 53). After 70h in the dark at room temperature, the sample had been completely converted to the Z isomer. Therefore, to isolate the isomer of interest after performing the final reaction and purification, it was needed to leave the isomeric mixture in a small amount of solvent system in the dark for three days before completing solvent removal, also in dark conditions.

The effect of an acid on the speed of the conversion was analysed by addition of acetic acid and followed by ^{1}H-NMR (Figure 53). The presence of acetic acid in the solution slowed the conversion to the Z isomer, with a half-life of 20 h and only reaching complete conversion after over 190 hours (8 days). Due to the acidic pH of the solution, the nitrogen atom is likely to be protonated, which slows the isomerisation and suggests that the lone pair of electrons of the nitrogen is involved in the photoisomerisation mechanism.

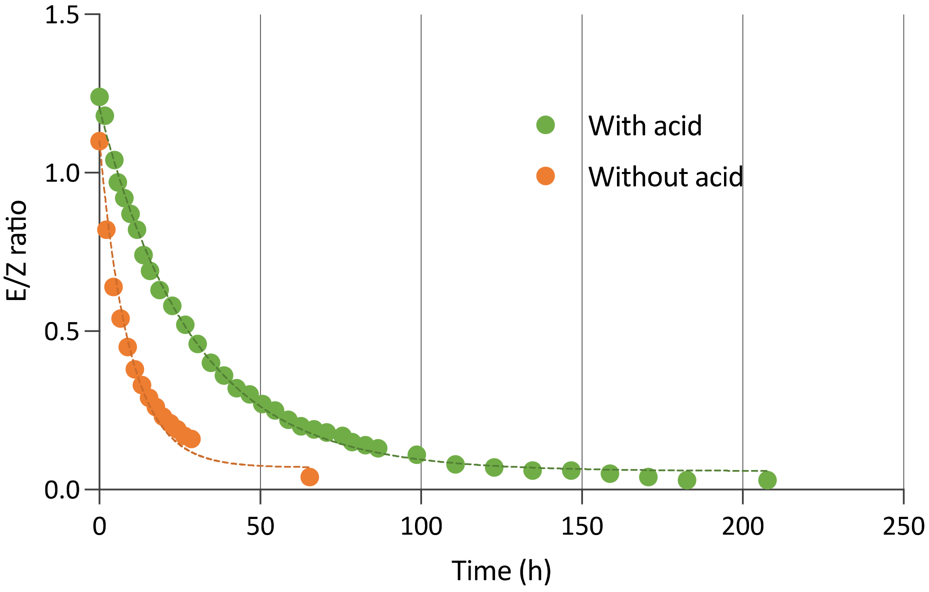


Figure 53. Relaxation from E/Z isomer mixture (illuminated solution) to Z isomer (dark conditions) in the presence and absence of acetic acid in d_6-DMSO. The thermal conversion to the Z isomer is slowed in the presence of acid.

The isomerisation data obtained in the polar aprotic solvent DMSO could not be extrapolated to the biological assays since a water-based solution is used in the measurement of fluorescence. Unfortunately, the limited solubility of **9** in water did not allow NMR spectra to be recorded in D_2O solution. To assess the effect of light on **9** in water, the biological assays were carried out. They would better indicate whether it is possible to use the Pepper aptamer as a reporter of transcription in *E. coli* and the effect of light on its fluorescence enhancement.

2.2.2 Reporting transcription *in vivo*

A reporter system based on the rhamnose promoter was established to test the use of Pepper aptamer as a reporter of transcription. The rhamnose promoter is tightly regulated, with a low basal level of expression in the absence of L-rhamnose that can be further repressed by the addition of glucose. It is also tuneable and varying concentrations of inducer result in different levels of expression.[211]

Golden Gate Assembly was used to introduce the Pepper RNA sequence into a plasmid containing the rhamnose promoter (Figure 54). Firstly, primers were designed to amplify the plasmid backbone to give a double stranded linear product with the rhamnose promoter and the rrnB terminator on separate ends. A BsaI restriction site was introduced on each end. The pepper fragment was designed to contain a BsaI restriction site on each end and four tandem repeats of the Pepper RNA sequence, as this has been shown to increase the fluorescence. [192] The two fragments were joined using Golden Gate assembly. This technique is a one pot cloning method that uses BsaI, a type IIS restriction enzyme which cleaves outside its recognition sequence leaving an overhang. T4 ligase then links segments with complementary overhangs with the absence of the restriction site in the product.[212] The resulting circular plasmid was transformed, amplified and sequenced to verify the correct addition of the Pepper gene.

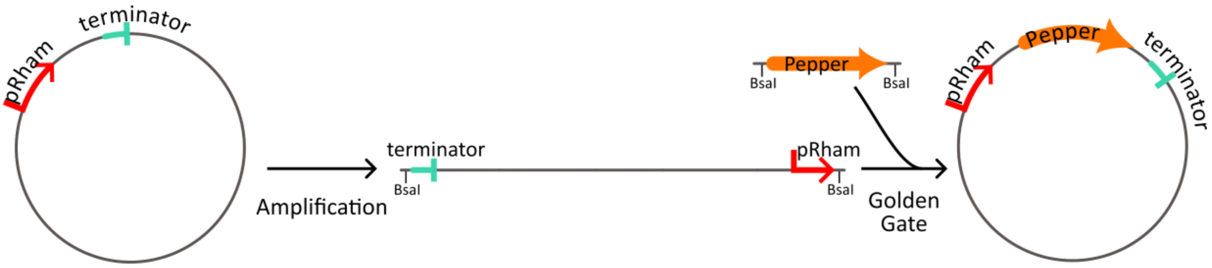


Figure 54. Golden Gate Assembly procedure used to introduce the Pepper gene into a plasmid backbone.

In vivo experiments were carried out in the *E. coli* cloning cell line XL1-blue. Cells transformed with the plasmid were grown either in the presence or absence of L-rhamnose and the cell-permeable fluorogen **9** was added before fluorescence measurements. The fluorescence of the cells was normalised against the fluorescence of XL1-blue cells with HBC_{620}. The cells induced with L-rhamnose show considerably higher fluorescence than those not induced (Figure 55). The low fluorescence of the uninduced cells indicates only a minimal amount of Pepper RNA was produced, whereas the high fluorescence of the induced culture corresponds to a high degree of RNA transcription. This experiment corroborates the validity of the Pepper-HBC_{620} system to detect the presence of Pepper RNA produced in cells.

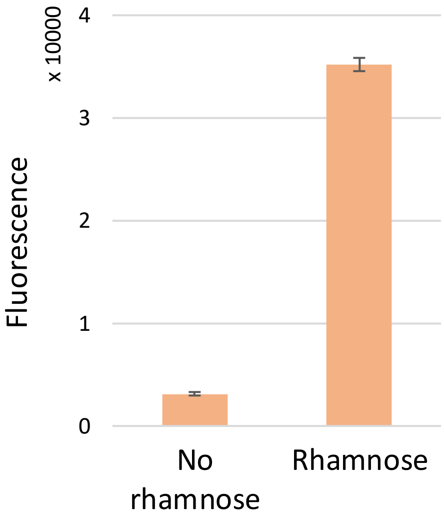


Figure 55. Fluorescence measurements of the pepper aptamer system with and without L-rhamnose induction, showing a low and high level of fluorescence respectively.

To investigate the sensitivity of the aptamer system for following the transcriptional changes within XL1-blue cells, fluorescence was measured after induction with different L-rhamnose concentrations. A linear increase of fluorescence was observed until 0.09% w/v of L-rhamnose, at which point the fluorescence plateaued due to the promoter reaching maximum activation of transcription (Figure 56). This indicates that the fluorescence of the Pepper aptamer system is dependent on the amount of RNA transcribed inside the cells.

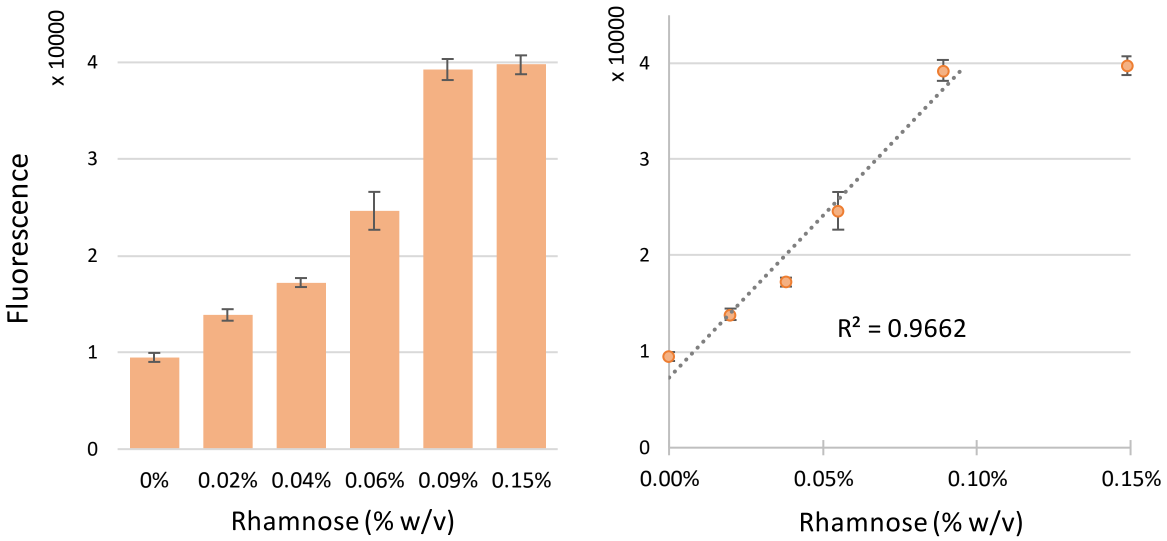


Figure 56. Induction of RNA production with different chemical inducer concentrations. A linear increase of fluorescence with increasing rhamnose concentration was observed until saturation.

The Pepper aptamer system has only been described to bind the Z isomer (**9**). It was hypothesised that the photoisomerisation of the fluorogen could affect the fluorescence signal and therefore the pepper aptamer may not be an adequate reporter of gene expression for light induced systems. To test this hypothesis, the fluorescence signal was measured before and after illumination in induced and uninduced samples in the presence of HBC_{620} (Figure 57).

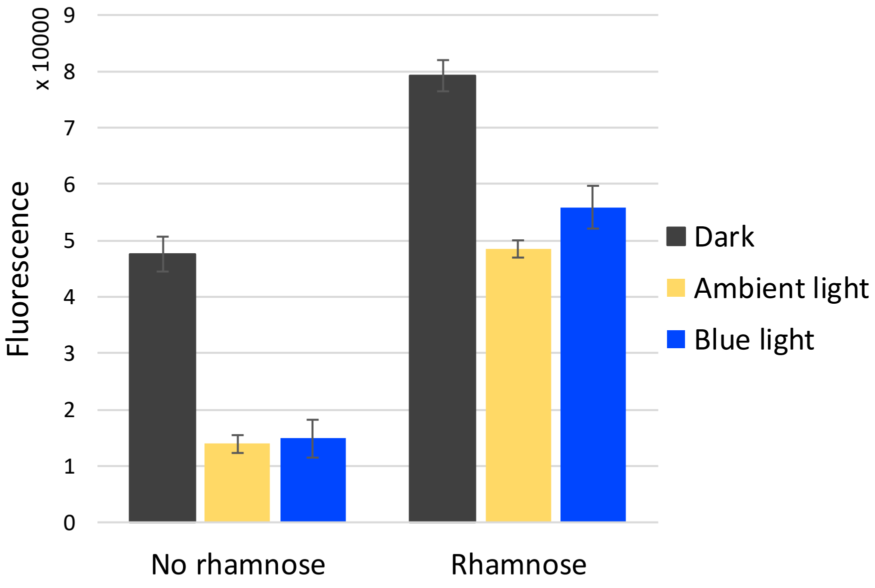


Figure 57. Effect of blue light and ambient light illumination in induced (rhamnose) or uninduced (no rhamnose) cultures of the pepper aptamer system. Illumination causes a reduction in fluorescence in both induced and uninduced samples.

Illumination of the samples with ambient (white) light or blue light caused a similar and noticeable reduction in fluorescence intensity. When comparing samples that have the same illumination treatment, Pepper RNA production can be assessed. The rhamnose induced sample showed higher fluorescence than the noninduced sample for each of the illumination conditions. However, comparing samples kept in different illumination conditions does not provide useful information regarding the expression of the Pepper RNA sequence. For instance, uninduced light and dark samples display a substantial difference in response independent of Pepper RNA expression, and the fluorescence of the uninduced dark sample and the induced light samples are comparable due to the photoisomerisation.

Next the Pepper aptamer system was tested using a plasmid where expression of the RNA sequence was under the control of the EL222 promoter (Figure 58). EL222 is constitutively expressed via the Anderson promoter J23106 of medium strength. EL222 is activated upon blue light illumination and can then bind to its promoter to induce Pepper RNA transcription. The Pepper gene was introduced in place of RFP in the pEBLindv2 plasmid[213] utilising Golden Gate assembly.

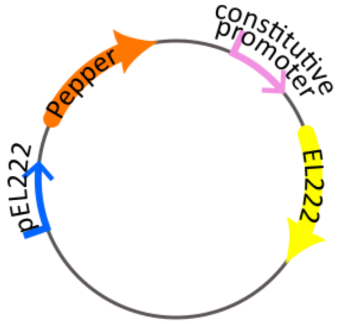


Figure 58. Plasmid map for the light-controlled expression of Pepper aptamer.

For these experiments, XL1-blue cells without the plasmid were used as a blank. Cells transformed with the plasmid were grown both in the dark (covered with foil) and constantly illuminated with blue light and the fluorescence was measured in the presence of HBC_{620}. The cells grown in the dark were expected to give low fluorescence as the Pepper RNA was not anticipated to be transcribed without light activation of EL222. Those grown under illumination were expected to give significantly higher fluorescence due to the activation of the transcription factor and production of the Pepper RNA. However, the results obtained (Figure 59) displayed the opposite behaviour. The fluorescence intensity was higher for the dark sample than the blue light, likely due to the photoisomerisation of **9** to **9b**. The illumination needed for activation of EL222 likely lead to the isomerisation of the fluorogen, which could result in a decrease in the fluorescence of the illuminated sample that the results are reversed. To check this, a sample of the dark culture was briefly illuminated prior to fluorescent measurement (Figure 59). Upon illumination, the fluorescence of the dark sample was reduced to a similar level to that of the induced sample.

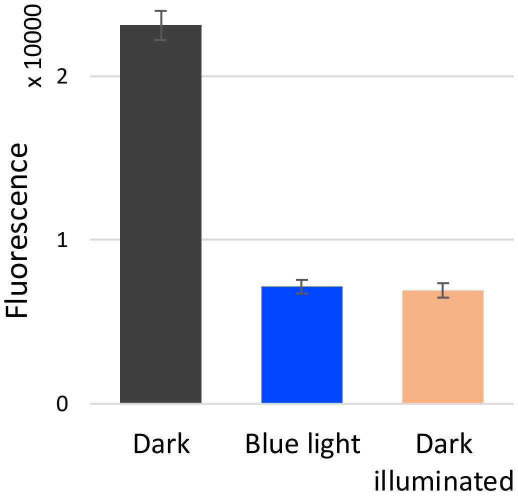


Figure 59. Fluorescence of the pepper aptamer system with (blue light) and without (dark) induction of RNA transcription. The sample induced with blue light showed lower fluorescence. Illumination of an uninduced sample caused a reduction in fluorescence.

This reduction in fluorescence upon illumination indicates that the photoisomerisation of the fluorogen is limiting when measuring RNA transcription under different illumination conditions. However, other potential causes for this opposite behaviour of the system were assessed. Correct aptamer activity (fluorescence production upon RNA and fluorogen binding) had been confirmed by exploiting the rhamnose promoter. EL222 expression was corroborated by growing cultures on a larger scale (500 mL) and centrifuging. The bright yellow colour of the cell pellet indicated its presence. Finally, a different plasmid was tested to ensure the transcription of Pepper RNA from the EL222 promoter or the concentration of EL222 produced from the constitutive promoter were not the cause for the unexpected behaviour of the reporter. The Pepper gene was introduced using Golden Gate Assembly into the plasmid backbone pAP05.[140] In this plasmid, the weaker Anderson promoter J23105 leads to the constitutive expression of EL222. Blue light illumination activates EL222 and induces Pepper RNA transcription. Due to the tetracycline resistance gene present on this plasmid, XL1-blue cells could not be used, and experiments were carried out using DH5α cells. However, The fluorescence measurements gave similar results to the previous plasmid, showing higher fluorescence for the dark sample (Figure 60).

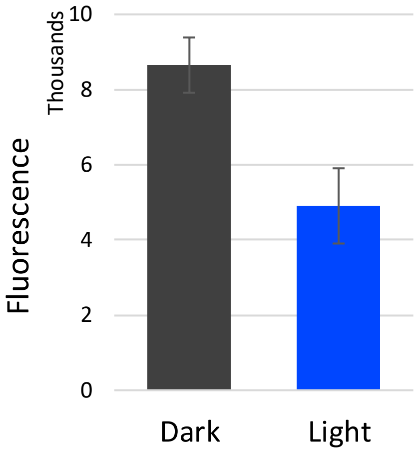


Figure 60. Fluorescence of the pepper aptamer system with (blue light) and without (dark) induction of RNA transcription. The sample induced with blue light showed lower fluorescence.

2.3 Conclusions and future work

In this chapter, the Pepper-HBC_{620} aptamer system was studied as a reporter of transcription. The fluorogen HBC_{620} was synthesised from thieno[3,2-b]thiophene. The route described in the literature did not yield the indicated product and a new synthetic route to the HBC_{620} fluorogen was successfully performed. Aside from generating the correct product, this new route required two less steps by avoiding requiring protection and deprotection of the alcohol intermediate.

Once synthesised and purified, the fluorogen was found to be a photoswitch. ^{1}H-NMR studies in DMSO indicated that the Z isomer, described to bind the Pepper RNA, was favoured in the dark. Illumination with red, blue or ambient light caused conversion to the E isomer, reaching a maximum Z/E ratio of 1/1.2. Addition of acetic acid to the illuminated isomeric mixture slowed the thermal conversion to the Z isomer, pointing to the involvement of the lone pair of the nitrogen atom in the isomerisation mechanism.

The effect of illumination in reporting of transcription was studied due to the intent of using the Pepper-HBC_{620} aptamer system in a blue-light induced system that would require different illumination states for different samples. *In vivo* assays utilising the rhamnose promoter corroborated the ability of the Pepper-HCB_{620} aptamer system to detect RNA transcription in cells. However, the use of a blue-light dependent promoter based on protein EL222 indicated that this system is not a valid reporter of gene expression for light induced systems due to the photoisomerisation of the fluorogen and concurrent loss of fluorescence. Due to this outcome, GFP was used with its known limitations as a reporter of gene expression in the work described in Chapter 3.

Chapter 3 - Engineering the DNA binding domain of EL222 to create photoresponsive transcription factors

3.1 Introduction

The protein EL222 is one of the smallest light-induced transcription factors, containing only the essential photosensory LOV domain and helix-turn-helix DNA binding domain linked via the Jα helix (Figure 61). The small size and simple structure of this protein is advantageous for protein engineering, and its quick activation and deactivation kinetics are important for optogenetic control of processes that require precise temporal control of gene expression.[214]

In EL222, the Jα helix serves as a linker between both domains but interacts more closely with the HTH domain.[62] Within the tri-helical core of HTH domains, the recognition of a specific sequence of DNA is mainly mediated by two α helices. The 3α helix binds to the major groove of DNA through hydrogen bonds and Van der Waals interactions with exposed bases. The 2α helix stabilises the interaction between protein and DNA but does not play a significant role in its recognition.[215] The 4α helix, which typically comprises the dimerisation interface for HTH domains, interacts with the outer surface of the LOV β-sheet in the dark state, interfering with DNA binding.[62] Light-induced conformational changes cause a reorientation of the domains relative to each other that frees the 4α helix and allows dimerisation and DNA binding.[63]

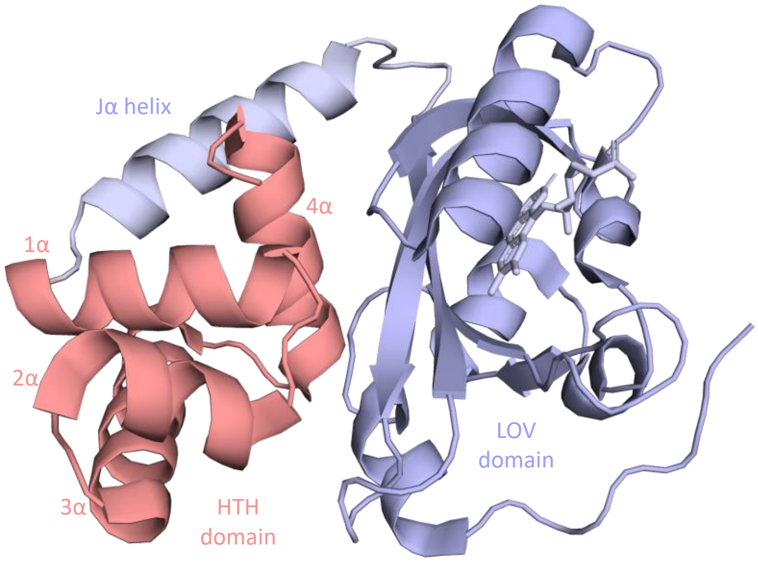


Figure 61. Structure of EL222 in the dark state showing the LOV domain in blue and the helix-turn-helix domain in pink linked by the Jα helix. The 4α helix of the DNA binding domain and the β-sheet of the LOV domain form interactions that prevent DNA binding in the dark state (PDB ID: 8A5R).

HTH DNA binding domains are widespread amongst bacteria and regulate a multitude of valuable processes,[83] including nitrogen fixation (FixJ), antibiotic resistance (TetR), biofilm formation (CsgD) and carbohydrate metabolism (HexR). The formation of biofilms is key in the survival of bacterial communities under environmental stresses. Achieving spatial and temporal control of this process can be beneficial for a variety of biotechnological applications.[126] Such control can be attained using light as an inducer, with the development of photoresponsive transcription factors. The existence of direct methods of measuring biofilm formation, and therefore straightforward evaluation of novel constructs, makes this process suitable for development of optogenetic tools.

CsgD is the transcriptional regulator that controls biofilm development. It contains an N-terminal regulatory domain[216] and a C-terminal HTH DNA binding domain[115] (Figure 62). The regulatory domain contains a conserved aspartate residue that is susceptible to phosphorylation, although the cognate histidine kinase sensor has not been found yet. The unphosphorylated transcription factor has high affinity for its DNA targets, but phosphorylation of the aspartate residue reduces its DNA binding affinity.[216,217]

CsgD is a key regulator of the transition between planktonic to biofilm lifestyle in *E. coli*. It directly activates the transcription of the csgBAC operon leading to the formation of curli, the main component of *E. coli* biofilms. It also positively regulates genes that favour biofilm adaptation such as cellulose production[121] and stress resistance.[218] Additionally, it represses negative determinants for biofilm formation,[120] including genes related to cell motility[114] and metabolism.[121] CsgD regulation is complex and occurs both at transcriptional and post-transcriptional level, depending on external and internal factors such as temperature,[219] osmolarity,[220] nutrient availability,[221] and growth phase[222].

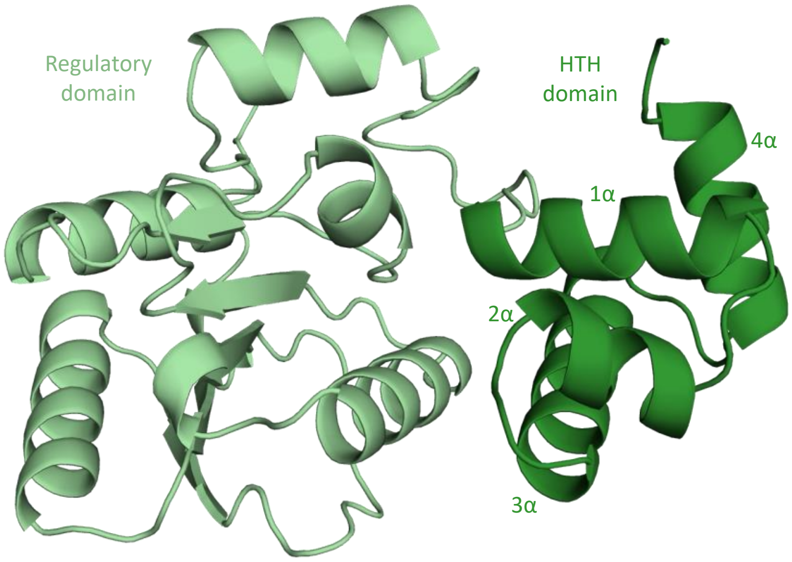


Figure 62. Predicted structure of CsgD showing the N-terminal regulatory domain in light green and the C-terminal helix-turn-helix domain in dark green (AlphaFold prediction P52106).

Evaluation of biofilm formation can be performed through visualisation of biomass after its staining with an appropriate dye such as crystal violet (CV). CV is a triphenylmethane dye (Figure 63), which is basic and positively charged, and interacts with negatively charged molecules in the bacterial cell wall and in the biofilm extracellular matrix.

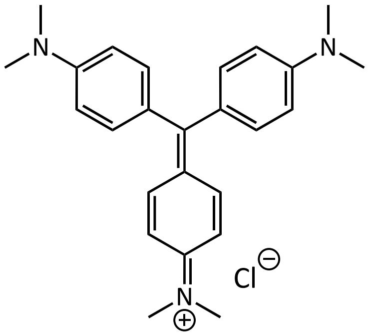


Figure 63. Structure of the dye crystal violet.

This crystal violet assay involves growing bacterial cultures in static conditions in microtiter plates, allowing the formation of biofilms. The removal of the liquid culture followed by extensive washing ensure the complete removal of planktonic bacteria. An aqueous CV solution is then applied to bind to the remaining biofilm. Following removal of excess CV, 30% acetic acid can be used to solubilise the CV that stayed bound to the biofilm (Figure 64). Measuring the absorbance at 570 nm allows the evaluation of the biofilm formation; a higher absorbance signifies a higher level of biofilm formation.[223]

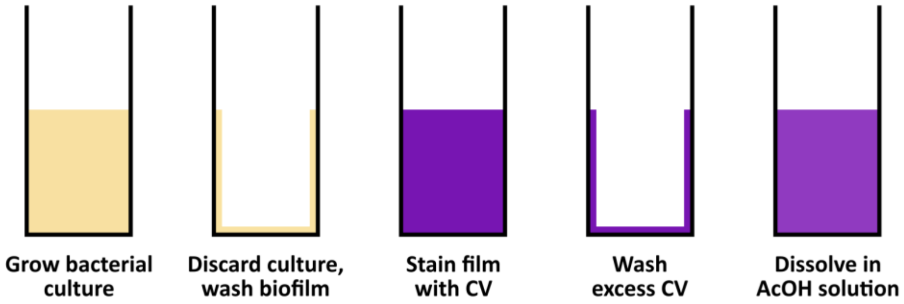


Figure 64. Steps of the crystal violet biofilm formation assay.

This assay is widely used[224–227] to evaluate the early stages of bacterial biofilm formation since it is easy to use, rapid, low cost and relatively high throughput. However, care should be taken when carrying out the washing steps[228] as washing too vigorously can lead to detachment of the biofilm and insufficient washing can lead to incomplete removal of planktonic cells (1st wash) or excess crystal violet (2nd wash). To overcome these shortcomings, it is essential to have positive and negative controls on each test plate, to do at least four replicate wells, and primarily compare results within a plate or plates set up at the same time.[229]

3.1.1 Aim

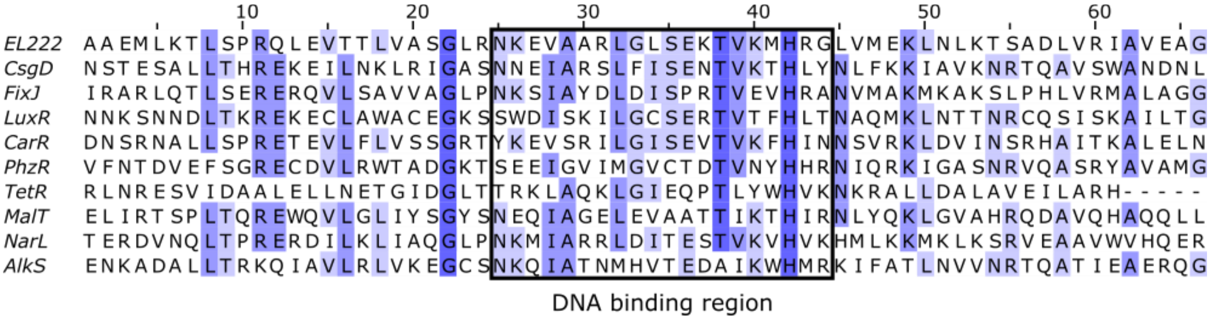
The aim of this project was to determine whether the DNA specificity of the HTH domain in EL222 could be altered to create novel photoresponsive transcription factors. It was hypothesised that this would establish the structural requirements of EL222 DNA binding and photoactivation and establish synthetic rules by which HTH domains can be exchanged to introduce optical control. By comparing the DNA binding domain of EL222 to other HTH transcription factors, the CsgD transcription factor and promoter were selected as a model system. The goal was to regulate curli expression, an essential step in biofilm formation, with blue light by replacing the DNA binding region of EL222 for that of CsgD. To assess DNA binding and transcription under different illumination states, GFP expression and biofilm formation assays were used.

3.2 Results

3.2.1 HTH domain alignments

It has previously been shown that the sequences of the different HTH motifs are poorly conserved. The orientation of the 3α helix in the major groove of DNA is also variable amongst different complexes and depends on the structure of the rest of the protein.[81] The HTH domains of ten LuxR-type proteins were aligned for comparison. The sequences aligned include EL222 (photoprotection), CsgD (biofilm formation), FixJ (nitrogen fixation), LuxR (bioluminescence), CarR (biosynthesis of carbapenem antibiotics), PhzR (biosynthesis of phenazine antibiotics), TetR (tetracycline resistance), MalT (catabolism of malto-oligosaccharides), NarL (nitrate assimilation) and AlkS (alkene degradation).

The twenty amino acid long DNA binding region was the most conserved region of the full helix-turn-helix domain. These similarities in the region crucial for recognition of DNA sequences may allow for manipulation of the affinity for DNA by performing amino acid substitutions. However, these substitutions are likely to affect protein structure and stability due to poor conservation in the rest of the domain and optimisation may be required.



```
EL222  AAEMLKTLSPRQLEVTTLVASGLRNKEVAARLGLSEKTVKMHRGLVMEKLNLKTSADLVRIAVEAG
CsgD   NSTESALLTHREKEILNKLRIGASNNEIARSLFISENTVKTHLYNLFKKIAVKNRTQAVSWANDNL
FixJ   IRARLQTLSERERQVLSAVVAGLPNKSIAYDLDISPRTVEVHRANVMAKMKAKSLPHLVRMALAGG
LuxR   NNKSNNDLTKREKECLAWACEGKSSWDISKILGCSERTVTFHLTNAQMKLNTTNRCQSISKAILTG
CarR   DNSRNALLSPRETEVLFLVSSGRTYKEVSRILGISEVTVKFHINNSVRKLDVINSRHAITKALELN
PhzR   VFNTDVEFSGRECDVLRWTADGKTSEEIGVIMGVCTDTVNYHHRNIQRKIGASNRVQASRYAVAMG
TetR   RLNRESVIDAALELLNETGIDGLTTRKLAQKLGIEQPTLYWHVKNKRALLDALAVEILARH-----
MalT   ELIRTSPLTQREWQVLGLIYSGYSNEQIAGELEVAATTIKTHIRNLYQKLGVAHRQDAVQHAQQLL
NarL   TERDVNQLTPRERDILKLIAQGLPNKMIARRLDITESTVKVHVKHMLKKMKLKSRVEAAVWVHQER
AlkS   ENKADALLTRKQIAVLRLVKEGCSNKQIATNMHVTEDAIKWHMRKIFATLNVVNRTQATIEAERQG
```

Figure 65. Alignment of helix-turn-helix domains of ten LuxR-type proteins with the DNA binding region highlighted. The intensity of blue indicates the degree of conservation per amino acid. Sequences are generally most conserved in the DNA binding region.

The sequences of the domains were obtained from UniProt KnowledgeBase: EL222 from *Erythrobacter litoralis* (Q2NB98), CsgD from *Escherichia coli* (P52106), FixJ from *Rhizobium meliloti* (P10958), LuxR from *Aliivibrio fischeri* (P12746), CarR from *Pectobacterium carotovorum* (Q46751), PhzR from *Pseudomonas chlororaphis* (P54303), TetR from *Escherichia coli* (P0ACT4), MalT from *Escherichia coli* (P06993), NarL from *Escherichia coli* (P0AF28) and AlkS from *Pseudomonas oleovorans* (P17051).

3.2.2 Comparison of the DNA binding domains of EL222 and CsgD

EL222 and CsgD contain HTH domains with four helices, although in EL222, the Jα helix linker to the LOV domain interacts closely with the DNA binding domain. The twenty amino acids from helices 2α and 3α comprise the region involved in the recognition of DNA. Comparing the amino acid sequence of the DNA binding region of both proteins shows 50% identity, with only a ten amino acid difference between CsgD and EL222 (Table 3). The high identity between the DNA recognition region of the two proteins and the existing direct assays for measuring biofilm formation make CsgD suitable for the study of this novel method of EL222 engineering.

	DNA binding region																			
EL222	N	K	E	V	A	A	R	L	G	L	S	E	K	T	V	K	M	H	R	G
CsgD	N	N	E	I	A	R	S	L	F	I	S	E	N	T	V	K	T	H	L	Y
	Stabilization of interaction										Interaction with DNA									

Table 3. Comparison of the sequences in the DNA binding regions of EL222 and CsgD. Highlighted in red are the amino acids that differ.

Since the DNA binding domain of CsgD has not been structurally determined, an AlphaFold[230,231] prediction was used to compare with the structure of EL222 (Figure 66). There is a high similarity in the tertiary structure of both domains, indicating that the contacts of the external amino acid sidechains with DNA, and not the protein conformation, are what cause the affinity for different DNA sequences.

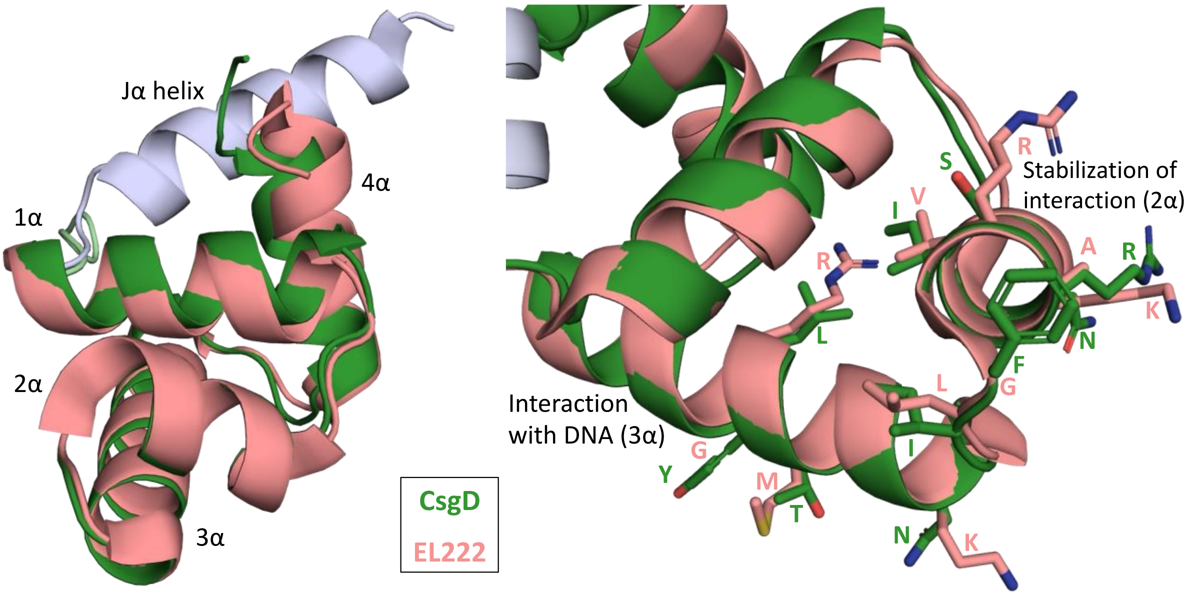


Figure 66. Superposed structures of the HTH domain of CsgD (green, AlphaFold ID: P52106) and of EL222 (pink) including the Jα helix (blue, PDB ID: 3P7N). On the right, the DNA binding region is highlighted with the differing amino acid sidechains displayed for comparison.

3.2.3 Optimisation of the reporter assays

Prior to engineering the DNA binding domain of EL222, two methods for studying the transcription factor activity were tested and optimised. Firstly, GFP expression was used as a reporter of transcription. The plasmid used for this assay (pAP05) was a gift from Prof. Yong Ku Cho and Dr. Azady Pirhanov, University of Connecticut.[140] It provides the constitutive expression of EL222 from the moderate-strength promoter J23105, and the light-regulated expression of GFP, controlled by the EL222 promoter (Figure 67). EL222 is produced both in dark and light conditions due to the constitutive expression. However, it only binds to the promoter to initiate GFP expression when illuminated with blue light. By substituting the EL222 promoter for the CsgD promoter and mutating the DNA binding domain of EL222, a comparison of the GFP expression in dark and light conditions would indicate the ability of the variants to bind the CsgD promoter and initiate transcription.

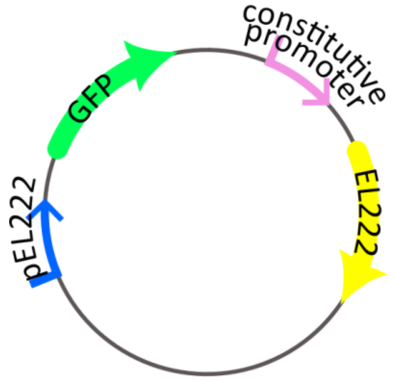


Figure 67. Plasmid pAP05 used for light dependent GFP expression. It provides the constitutive expression of EL222 and the light-regulated expression of GFP.

To obtain a satisfactory dynamic range in GFP fluorescence when using the pAP05 plasmid to report EL222 activity between illuminated cells and those kept in the dark, the cell growth conditions were first optimised (Figure 68, Table 4). All trial experiments were performed using WT EL222, one set of cells was constantly illuminated with blue light (L), and another was covered in foil (D). Two different cell lines were tested (BL21 Star DE3 and DH5α) and grown in two sets of conditions (37°C until $OD_{600} = 1.7$, and 37°C until $OD_{600} = 0.4$ + 20°C for a further 20 h). The effect of lysozyme addition prior to imaging was tested in cells grown at 37°C since the lysis of the cell wall and release of GFP could affect the fluorescence L/D ratio.

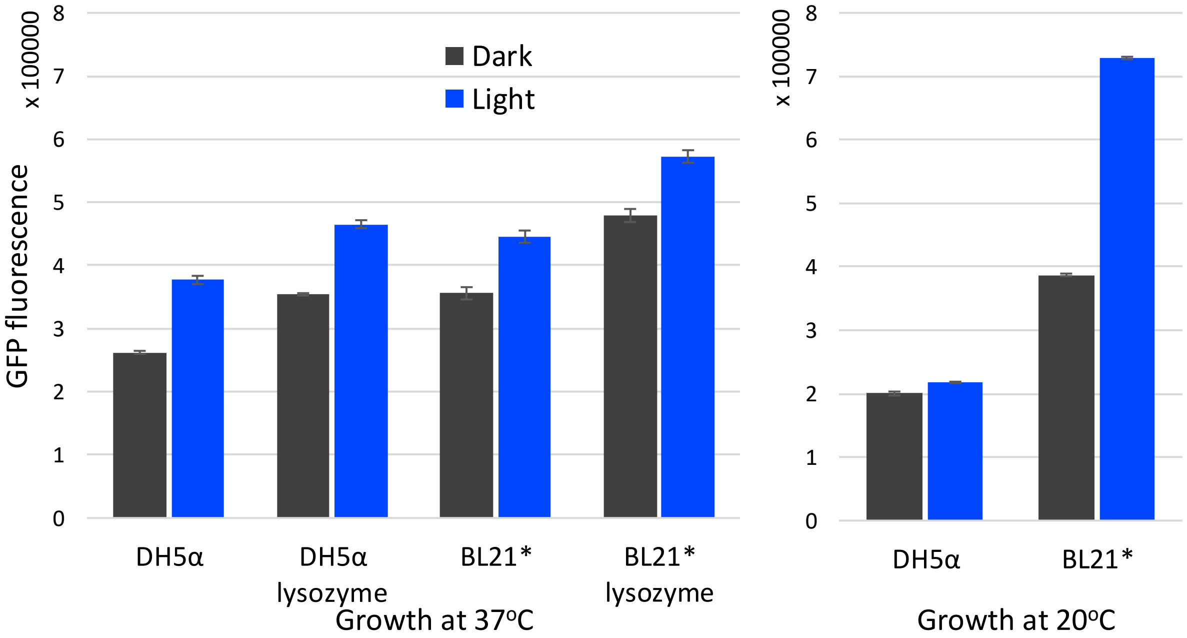


Figure 68. Optimisation of the GFP fluorescence assay using WT EL222. Two cell lines were grown in dark (black) and light (blue) conditions at different temperatures (37°C left, 20°C right) and the effect of lysozyme was analysed.

Temperature	37°C				20°C	
Cell line	DH5α	DH5α	BL21*	BL21*	DH5α	BL21*
Additional		lysozyme		lysozyme		
Ratio L/D	1.44 ± 0.03	1.31 ± 0.05	1.25 ± 0.13	1.20 ± 0.06	1.09 ± 0.01	1.88 ± 0.09

Table 4. Fluorescence light/dark ratios obtained in the different conditions tested.

For the first set of conditions, growth at 37°C, the use of DH5α cells gave a better L/D fluorescence ratio than BL21 Star DE3. The addition of lysozyme did not improve the L/D fluorescence ratio as, aside from increasing the fluorescence of cells grown under blue light, it also increased the fluorescence of cells grown in the dark. Overall, the best fluorescence L/D ratio was obtained from the second set of conditions, growth at 20°C, using BL21 Star DE3 cells. However, OD_{600} measurements indicated that the growth of BL21 Star DE3 cells under blue light for 24h was impaired as those cultures reached an OD_{600} of 1.25, considerably lower than those grown in the dark. Due to the detrimental effect of long-term exposure of BL21 Star DE3 to blue light, these conditions were not deemed appropriate. Instead, growth of DH5α cells at 37°C until OD_{600} = 1.7 was considered the best alternative and used for the following GFP expression experiments, since there was no apparent effect on growth rate due to illumination.

The second reporter method relies on detection of biofilm formation *via* the crystal violet assay (Figure 64). Two *E. coli* strains were used, BW25113 as a positive control since it has all the necessary machinery to produce biofilm; and JW1023-1 as a negative control, where the CsgD gene has been knocked out so that it is unable produce biofilm. The activity of the controls did not depend on illumination conditions. A successful EL222 variant capable of mimicking CsgD would restore biofilm formation ability in JW1023-1 when grown under illumination since it would bind to the CsgD promoter naturally present in the cells and induce transcription of the curli biogenesis genes. However, if the engineered variant was not able to bind to the CsgD promoter, there would be no increase in biofilm formation in the illuminated sample.

It has been shown that the capability of *E. coli* to form biofilm is dependent on the growth medium used.[232] To determine the most adequate conditions for cell growth, the crystal violet assay was initially performed on the positive and negative controls grown in four different media (Figure 69). Growth in minimal media (M9 and M63) resulted in a higher measured absorbance for the positive control than for the negative control, indicating biofilm formation occurred in the positive control. In contrast, growth in rich media (LB and LB supplemented with glucose) did not lead to a difference in biofilm formation between positive and negative controls, suggesting that the use of rich media is not suitable for this assay. These results agree with previous studies that show biofilm formation increases at low-nutrient conditions.[233] Since there was no substantial difference between the minimal media, M9 medium was chosen for ensuing experiments.

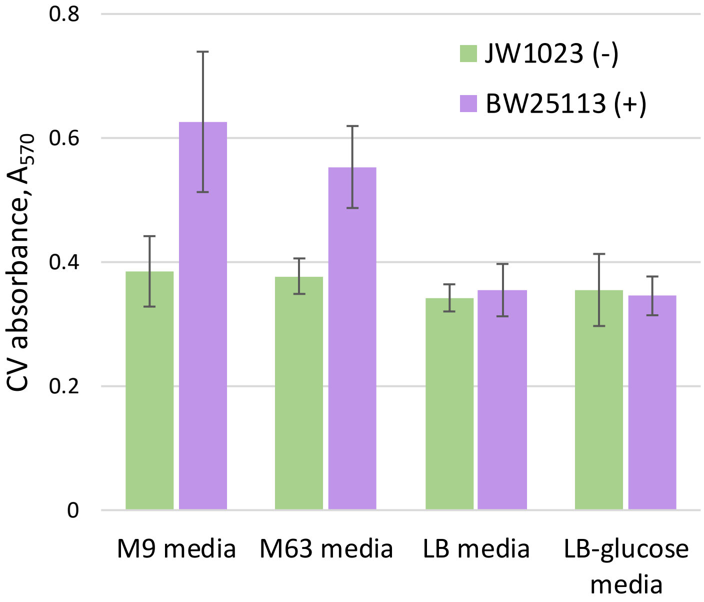


Figure 69. Crystal violet assay of positive (purple) and negative (green) controls in different media. Minimal media was suitable for this assay as the positive control produced more biofilm.

3.2.4 Ruling out interactions between WT EL222 with pCsgD

The EL222 promoter sequence comprises 12 nucleotides[64] $\frac{\mathrm{G}}{\mathrm{A}}\mathrm{GNC}\frac{\mathrm{T}}{\mathrm{C}}\frac{\mathrm{T}}{\mathrm{A}}\frac{\mathrm{T}}{\mathrm{A}}\frac{\mathrm{G}}{\mathrm{A}}\mathrm{GNC}\frac{\mathrm{T}}{\mathrm{C}}$ and the CsgD promoter is 11 nucleotides long[120] $\mathrm{CGGG}\frac{\mathrm{T}}{\mathrm{G}}\mathrm{GA}\frac{\mathrm{T}}{\mathrm{G}}\mathrm{N}\frac{\mathrm{T}}{\mathrm{G}}\mathrm{A}$. Due to these differences, EL222 would not be expected to bind to the CsgD promoter. However, the ability of wild type EL222 to bind to the CsgD promoter and activate transcription was tested to confirm this hypothesis.

For the GFP reporter assay, the EL222 promoter was substituted for the CsgD promoter using KLD cloning (Figure 70). The DNA was amplified to contain the CsgD promoter sequence (CGGGTGATCGA) instead of the EL222 promoter sequence (AGCCTTTAGTCC). The KLD enzyme mix, containing kinase, ligase and DpnI, was used on the amplified fragment with blunt ends. The process involved phosphorylation and circularisation of the amplified DNA and removal of the template DNA.

Figure 70. KLD cloning method used for the change in promoter.

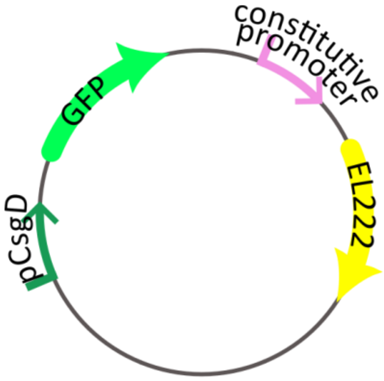


Figure 71. Resulting plasmid containing WT EL222 and pCsgD. It provides the constitutive expression of EL222 and the expression of GFP under the CsgD promoter.

The resulting plasmid (Figure 71) was used on the GFP assay alongside pAP05 for comparison (Figure 72 left). Wild type EL222 was not able to induce GFP expression under control of the CsgD promoter in light conditions and therefore supports that WT EL222 is not able to interact with the CsgD promoter.

The ability of WT EL222 to induce biofilm formation was tested in cell strain JW1023 transformed with pAP05, containing WT EL222. The biofilm formation assay was performed along with the positive and negative controls for comparison in dark conditions and under blue light illumination (Figure 72 right). No increase in biofilm formation was observed in the illuminated sample of JW1023 cells containing WT EL222, confirming that the native protein was not able to bind to the CsgD promoter present in JW1023 to induce biofilm formation.

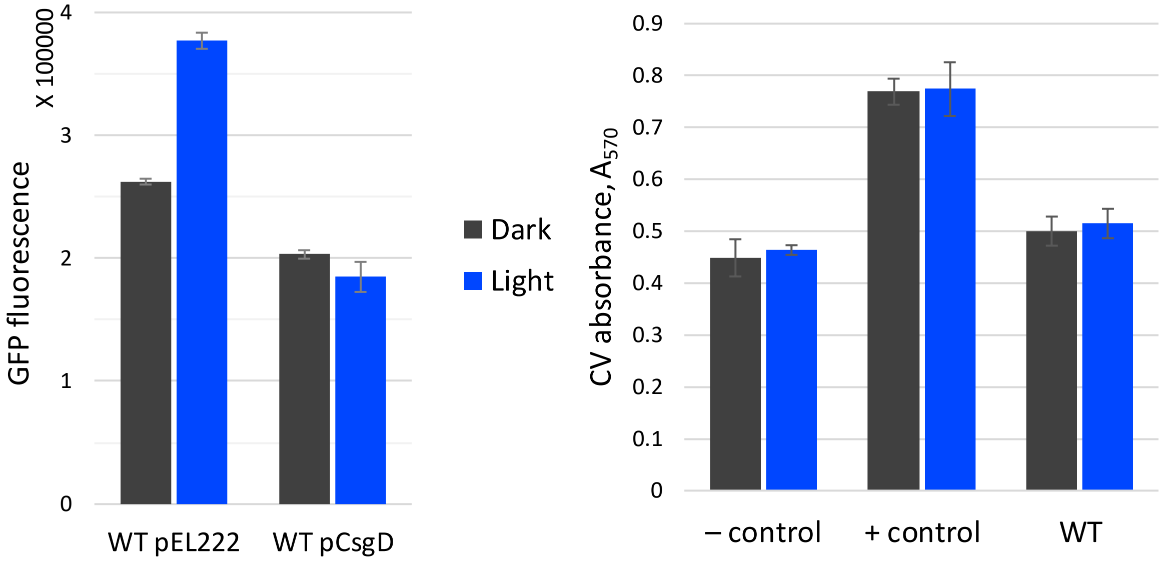


Figure 72. GFP expression (left) and biofilm formation (right) assays testing the capacity of WT EL222 to bind to the CsgD promoter. WT EL222 was not able to initiate GFP transcription or biofilm formation upon illumination, indicating it was not able to activate the CsgD promoter.

3.2.5 Engineering and testing the EL222 variants

The twenty amino acid long DNA binding region of EL222 HTH domain was exchanged to that of CsgD using HiFi DNA assembly. The DNA was amplified to remove the twenty amino acids long DNA recognition region from the HTH domain of EL222, and an insert was created to contain the recognition region from CsgD with complementary overlaps on both ends. The HiFi assembly method (Figure 73) relies on three different enzymes for seamless assembly of multiple DNA fragments with varied overlaps. First, an exonuclease creates single-stranded 3´ overhangs that facilitate the annealing of fragments that share complementarity at one end. Then, the polymerase fills in gaps within each annealed fragment and finally, the DNA ligase seals nicks in the assembled DNA. This resulted in a double-stranded fully sealed DNA molecule that contains the engineered EL222 (Figure 74).

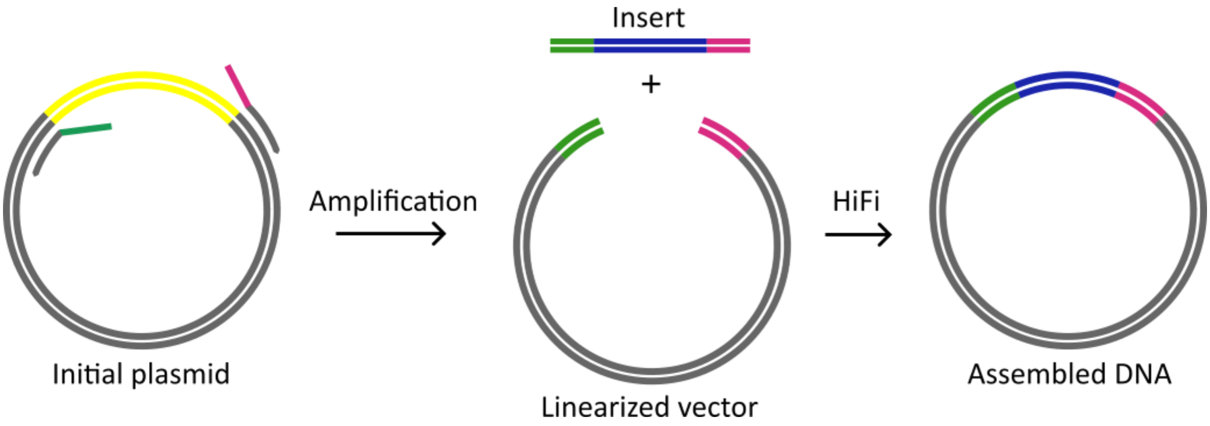


Figure 73. HiFi DNA assembly procedure used to exchange the DNA binding region of EL222 for that of CsgD.

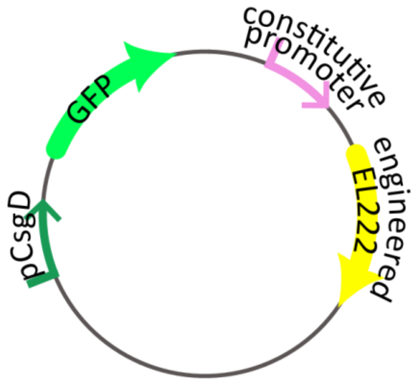


Figure 74. Resulting plasmid containing engineered EL222 and pCsgD. It provides the constitutive expression of engineered EL222 and the expression of GFP under the CsgD promoter.

The activity of this EL222 variant containing 10 mutations, termed EL222::CsgD, (Table 5) was tested using the two described assays (Figure 75). In both assays, the EL222 chimera showed no difference in response between light and dark conditions. EL222::CsgD showed no evidence of binding to the CsgD promoter to activate transcription.

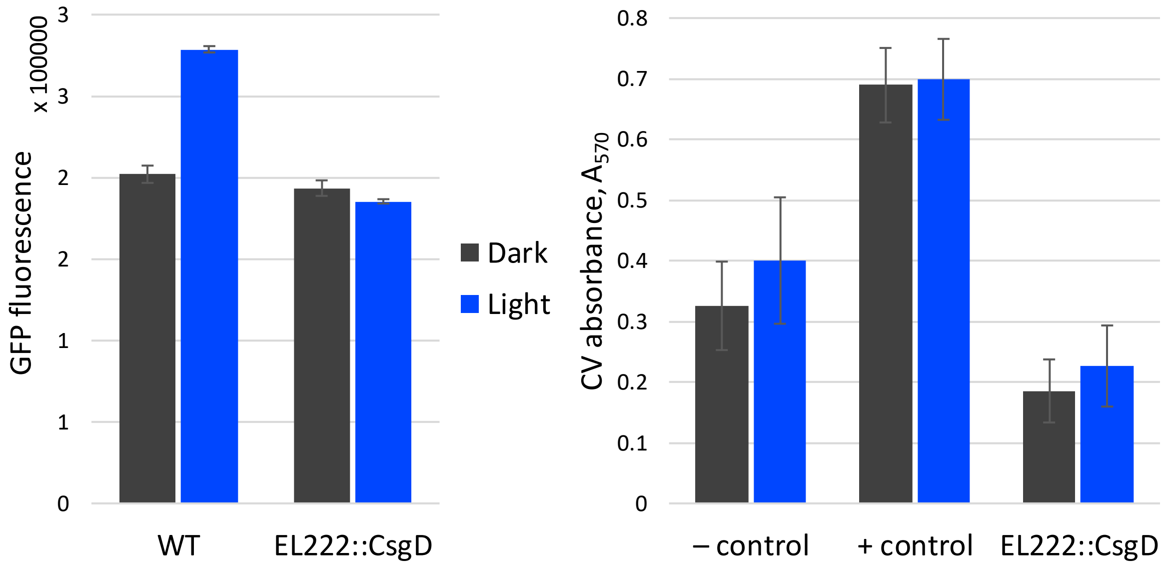


Figure 75. GFP expression (left) and biofilm formation (right) assays testing the ability of EL222::CsgD to bind to the CsgD promoter. This variant showed no evidence of binding to the CsgD promoter since there was no light induced GFP expression or biofilm formation.

Since this chimera was unable to activate transcription upon illumination, further engineering was needed. The exchange of the HTH region in EL222 with CsgD may result in either destabilisation of the HTH domain or of the interactions with the LOV domain. The amino acid sidechains that point towards the interior of the protein are likely to be a significant factor in the destabilisation of the protein tertiary structure (Figure 76). Such destabilisation could lead to inefficient folding and solubility problems that would lead to an inactive protein.

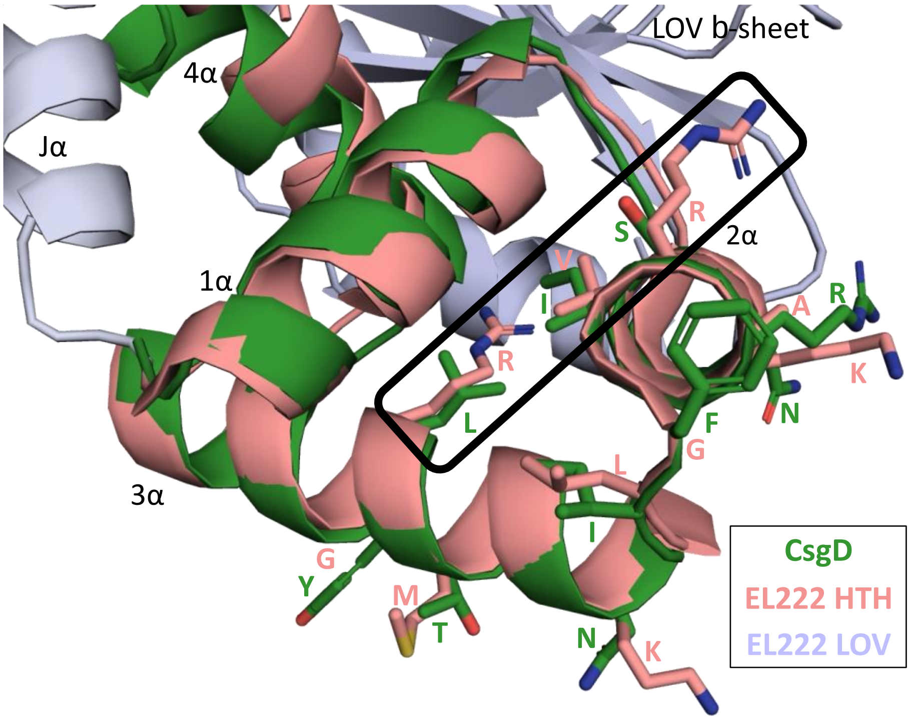


Figure 76. Comparison of the HTH domains of EL222 and CsgD with the differing amino acids in the DNA binding region represented. The three amino acids thought to be involved in protein stabilisation because of their position towards the interior of the protein are highlighted.

There are three mutated amino acids which point towards the interior in EL222::CsgD (I186, S189, L201) and differ to WT EL222 (V186, R189 and R201). The change of valine for isoleucine is unlikely to impact activity due to the sidechain similarity in size and non-polar character. Conversely, the introduction of serine, a small polar neutral amino acid, and leucine, a medium sized non-polar amino acid in the place of arginine, a large positively charged amino acid are more likely to cause significant disruptions in the protein. Reverting S189 or L201 to arginine could improve the stability of the domain without affecting the DNA binding affinity for the CsgD promoter. Two new variants were created using site directed mutagenesis (SDM) on the chimera, the first containing the mutation S189R and the second containing L201R. These EL222::CsgD$_{R1}$ and EL222::CsgD$_{R2}$ variants contain nine mutations respective to WT EL222, as one amino acid from the EL222::CsgD variant was reverted to its original identity in EL222 (Table 5). The single point mutations were performed using site directed mutagenesis. Primers containing the desired mutation were designed and the DNA was amplified. This resulted in a circular plasmid containing the mutation (Figure 77).

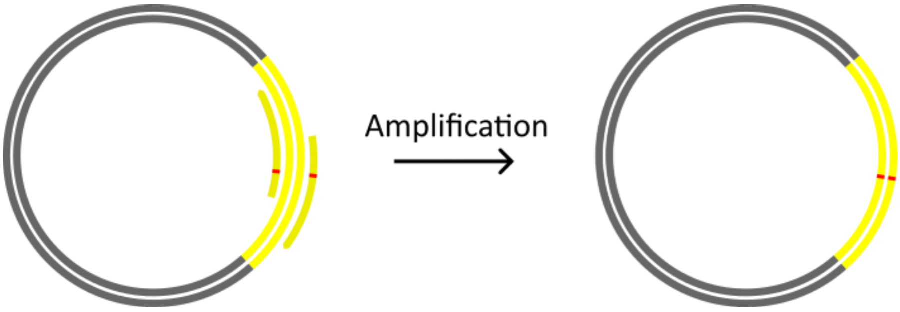


Figure 77. Site directed mutagenesis process used to introduce point mutations in EL222.

The two variants were tested for activity alongside the controls (Figure 78). EL222::CsgD$_{R1}$ was able to induce biofilm formation under illumination conditions. Therefore, the EL222 protein was successfully engineered to target the CsgD promoter present in JW1023 cells. However, the activation of transcription was not seen with the GFP expression assay. It has been shown that EL222 only effectively binds to DNA at a concentration of 0.25−0.75 μM.[64] The significant mutations of the DNA binding region of EL222 could have affected transcription and protein folding, which could lead to a protein concentration outside of the dynamic range. If the concentration of EL222 in the cells was too low, transcription would not be activated even in the presence of light. On the other hand, if the concentration of EL222 was too high, it would be activated both in dark and light conditions. Both scenarios would lead to the loss of light activated GFP expression. Unfortunately, this means that the GFP expression assay may not give a positive response even in the event of the engineered transcription factor having affinity for the promoter.

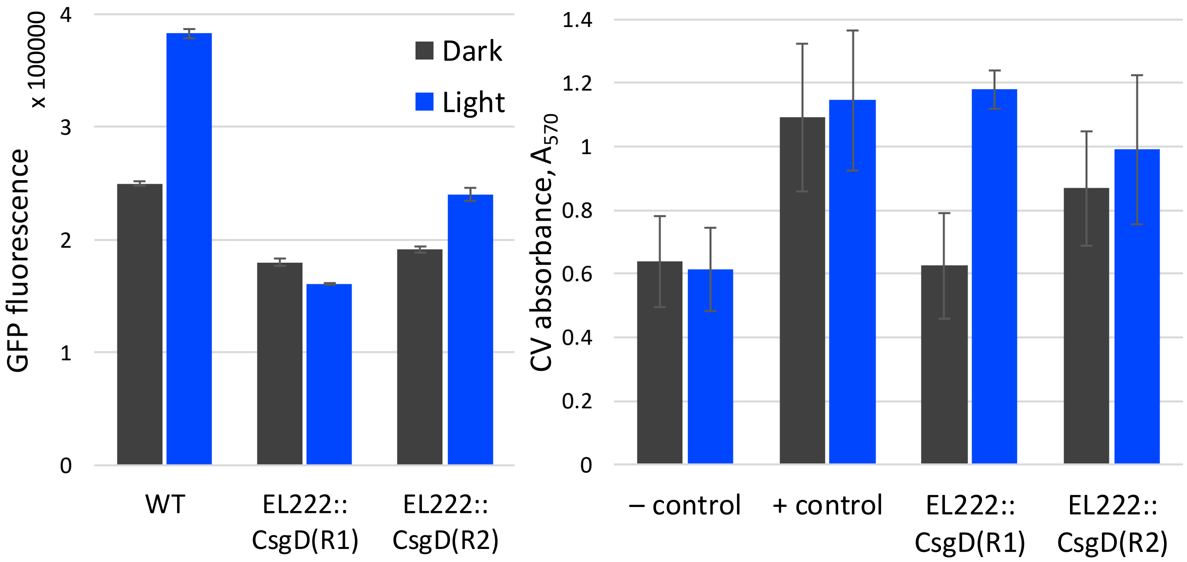


Figure 78. GFP expression (left) and biofilm formation (right) assays testing the ability of two EL222 variants to activate the CsgD promoter. EL222::CsgD$_{R1}$ induced biofilm formation upon illumination. EL222::CsgD$_{R2}$ induced a small degree of biofilm formation in both conditions and a small increase in GFP expression upon illumination.

The variant EL222::CsgD$_{R2}$ produced measurable biofilm under both light and dark states. The GFP expression assay displayed a small increase in fluorescence for the illuminated sample, which may indicate some affinity for the CsgD promoter but a small dynamic range of light activation. To investigate this result further, another variant was created via SDM where both amino acids (S189 and L201) were reverted to the WT amino acid. This new variant, referred to as EL222::CsgD$_{RR}$ (Table 5), contained eight mutations from WT EL222. This variant was not able to bind to the CsgD promoter and induce transcription since it showed no difference in response between light and dark conditions in any of the assays (Figure 79).

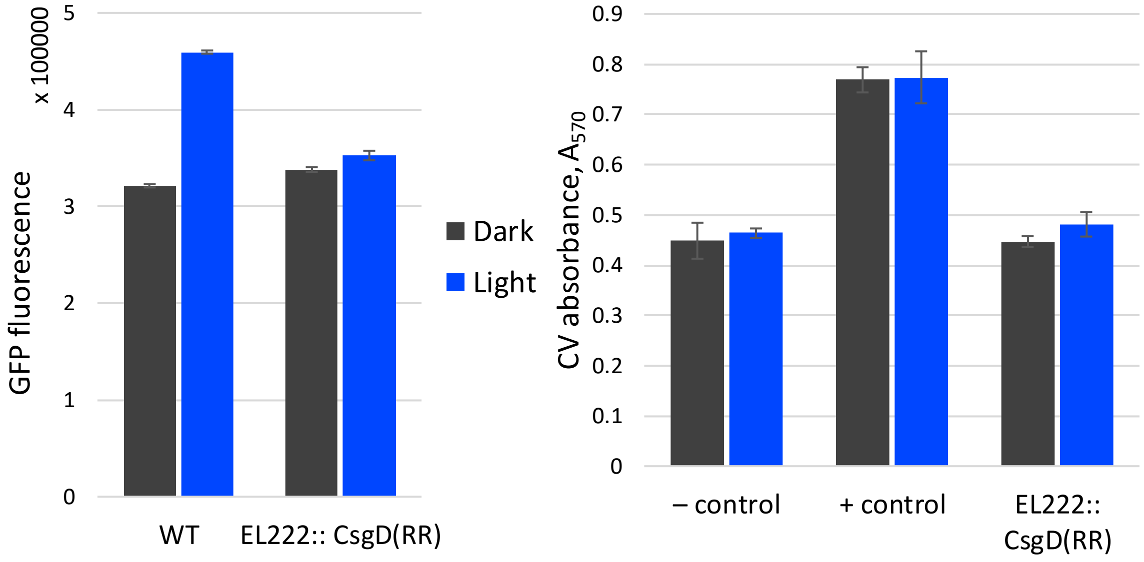


Figure 79. GFP expression (left) and biofilm formation (right) assays testing the ability of EL222::CsgD$_{RR}$ to bind to the CsgD promoter. This variant showed no evidence of binding to the CsgD promoter since there was no light induced GFP expression or biofilm formation.

3.2.6 Isolating the variants

The expression vector pETM11 containing EL222 or its engineered variants was used in BL21 DE3 cells with the intention of isolating the four EL222 variants tested for activity (Table 5). Expression and purification of the variants could provide information on their stability and analysis could provide structural insight. The expression plasmid, containing His$_6$-tagged EL222 under the control of the T7 promoter (Figure 80) was modified by HiFi DNA assembly and site directed mutagenesis to obtain the four variants.

Name	**No. mutations**	**DBD amino acid sequence**
WT	0	NKEVAARLGLSEKTVKMHRG
EL222::CsgD	10	NNEIARSLFISENTVKTHLY
EL222::CsgD$_{R1}$	9	NNEIARRLFISENTVKTHLY
EL222::CsgD$_{R2}$	9	NNEIARSLFISENTVKTHRY
EL222::CsgD$_{RR}$	8	NNEIARRLFISENTVKTHRY

Table 5. DNA binding region sequences of the engineered EL222 variants with mutations highlighted.

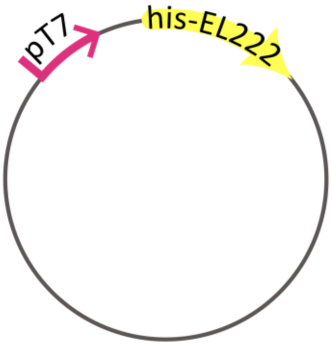


Figure 80. Plasmid used for expression of EL222 and its engineered variants. Expression of His$_6$-tagged EL222 and variants is controlled with the T7 promoter.

The four variants were expressed in BL21 DE3 cells, the cells were lysed, and the soluble fraction was purified using immobilised metal chelate affinity chromatography. The resulting fractions were analysed by SDS-PAGE (Figure 81). The three variants that were not able to induce biofilm formation were not successfully expressed or were found in the insoluble fractions of *E. coli*. This lack of expression or solubility is likely to play a part in their inability to induce transcription from the CsgD promoter and likely indicates that protein folding is compromised. Conversely, the active variant EL222::CsgD$_{R1}$ was successfully isolated, its His$_6$-tag was cleaved using TEV protease, and further purification was carried out via size exclusion chromatography.

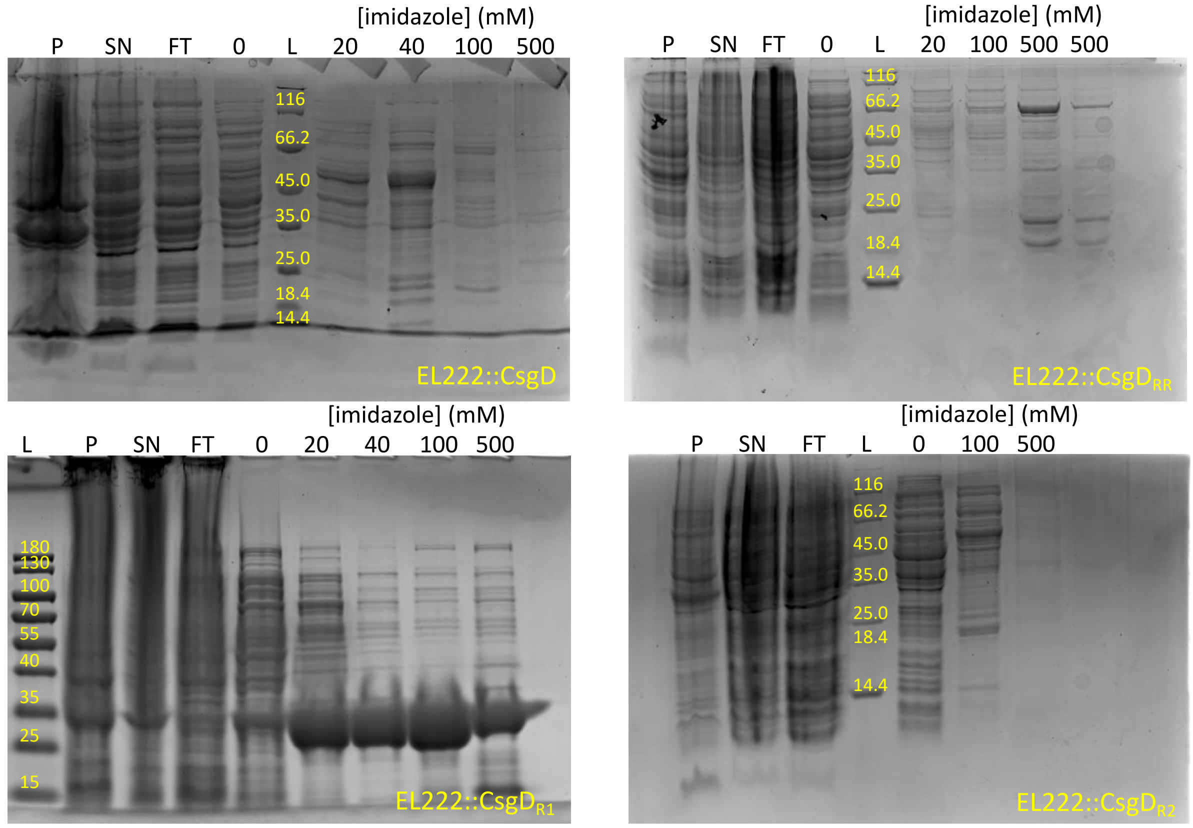


Figure 81. SDS-PAGE gels after attempts at purification of the four EL222 variants. EL222::CsgD$_{R1}$ was soluble and able to be purified. Small amounts of protein could be seen in the insoluble fraction (P) for EL222::CsgD and EL222::CsgD$_{R2}$. No EL222::CsgD$_{RR}$ was observed.

The amino acid Arg_{189} is therefore thought to be important for the stabilisation and solubility of EL222. Structural insight into the active variant would provide information about the requirements for structural stabilisation of the DNA binding domain. This would aid the goal of successfully engineering EL222 to target different DNA sequences while maintaining light dependent activation. To study the structure of EL222::$CsgD_{R1}$, attempts to crystallise it were carried out using the vapour diffusion sitting drop method and two commercial screens: PACT Premier, a systematic screen that tests the effect of pH and ions using PEG as the precipitant, and JCSG Plus, a sparse-matrix screen based on previously successful crystallisation conditions of various proteins. These two screens are complementary, comprising a thorough strategy to identify new crystallisation conditions.[234] Unfortunately, these attempts did not yield protein crystals in any of the conditions tested. Since precipitation was not observed, the protein concentration used (6.5 mg/mL) was likely too low for successful crystallisation.

3.3 Conclusions and future work

In this chapter, the LOV photoreceptor EL222 was engineered to bind to the promoter sequence of CsgD, the master regulator of biofilm formation. Four variants, with eight to ten amino acid mutations, were created and tested. One variant, EL222::$CsgD_{R1}$, was successfully modified to have affinity for the promoter sequence of CsgD. Introduction of this engineered variant into an *E. coli* strain equipped with the necessary machinery for biofilm formation excluding the CsgD gene, lead to light-induced biofilm formation.

This engineered variant contains nine mutations in the helix-turn-helix domain, specifically in the region responsible for DNA recognition, that result in the change of the DNA sequence it is able to bind to in order to induce transcription. Directly modifying the DNA binding affinity of EL222 through mutations in the DNA recognition region provides a new method of achieving light control of gene expression that exploits the protein directly, in contrast to utilising its existing promoter to express a gene of interest.

Attempts to isolate the four engineered variants were carried out. However, only the variant able to induce biofilm formation was soluble, which suggested instability or folding problems for the other variants. Structural studies of EL222::$CsgD_{R1}$ would be useful to better understand the role of the mutated amino acids in structural stabilisation of the DNA binding domain and DNA affinity. Using a higher protein concentration in crystallisation studies would increase the likelihood of obtaining protein crystals. Co-crystallisation of EL222::$CsgD_{R1}$ with the CsgD promoter in the presence of blue light could lead to crystals of the transcription factor-DNA complex.

By optically controlling the activity of the newly designed CsgD mimic, control of biofilm formation was achieved. This optogenetic tool for biofilm formation can allow further studies of *E. coli* biofilms and lead to the design of new devices for beneficial uses of biofilms, such as biocatalysis, bioremediation or waste-water treatment.

Furthermore, this method of engineering EL222 serves as a proof of concept and opens the possibility of modifying EL222 to mimic other existing helix-turn-helix domains to target their cognate promoters. However, the biofilm formation process targeted in the present work possesses direct measuring assays. Analysing the DNA binding capability of other EL222 variants to initiate processes that do not have direct assays would be more complex. In these scenarios, the GFP reporter system would need to be improved to expand its dynamic range or a new method would need to be devised that is not as dependent on the concentration of engineered proteins.

Chapter 4 - Elucidating the role of Asn_{194} in aureochrome1a LOV domain signal transduction

4.1 Introduction

The combination of a LOV photoreceptor with a bZIP DNA binding domain in aureochromes allows the possibility of controlling gene expression using light, making this protein noteworthy in the field of optogenetics. Signal transduction is driven by light catalysed cysteinyl-flavin adduct formation and concurrent protonation of the N_5 position. Understanding the structural rearrangements that occur in the protein in response to the changes in the chromophore is needed to establish the mechanism of light activation, improve photoreceptor activities and employ LOV domains in biotechnology.

In aureochromes, the usual effector-sensor topology is inverted and an A'α helix links the sensor and effector instead of the Jα helix.[92] Aureochromes are dimeric both in dark and light states due to the interactions between the bZIP domains. However, in the dark, the A'α helix partially obscures the LOV dimerisation site and the affinity for DNA is reduced due to the interactions between the LOV and bZIP domains. Blue light illumination causes the stepwise unfolding of A'α and Jα helices and the dissociation of the LOV domain from the bZIP domain. This leads to dimerisation of the LOV domain and an increase in affinity of the bZIP domain for DNA (Figure 82).[235] Isolated aureochrome LOV domains containing the surrounding A'α and Jα helices have been reported to dimerise upon blue light illumination.[236]

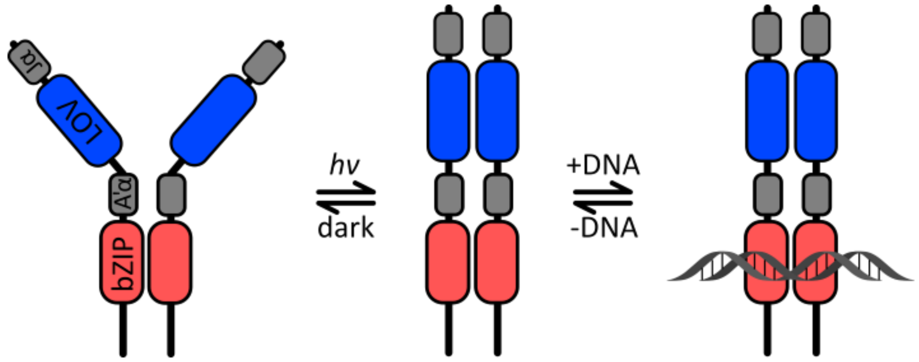


Figure 82. Blue light activation mechanism and DNA binding of aureochrome.

The formation of the cysteinyl flavin adduct upon illumination and the rotation of a glutamine residue to maintain the hydrogen bond with the flavin cofactor have been established in LOV domains.[45] However, further rearrangements in sidechain orientation through the hydrogen bond network leading to effector activation are still unclear. Previous studies on the aureochrome 1a LOV domain from the alga *Ochromonas danica* (AuLOV) indicate that an asparagine residue (Asn$_{194}$) is involved in the hydrogen bond network that links the flavin and the key glutamine residue (Gln$_{293}$) with the A'α helix.[237] X-ray structures obtained from crystals grown in dark conditions, grown in the dark and illuminated before harvesting, and grown in blue light, revealed changes in sidechain orientation of these two key amino acids (Figure 83).

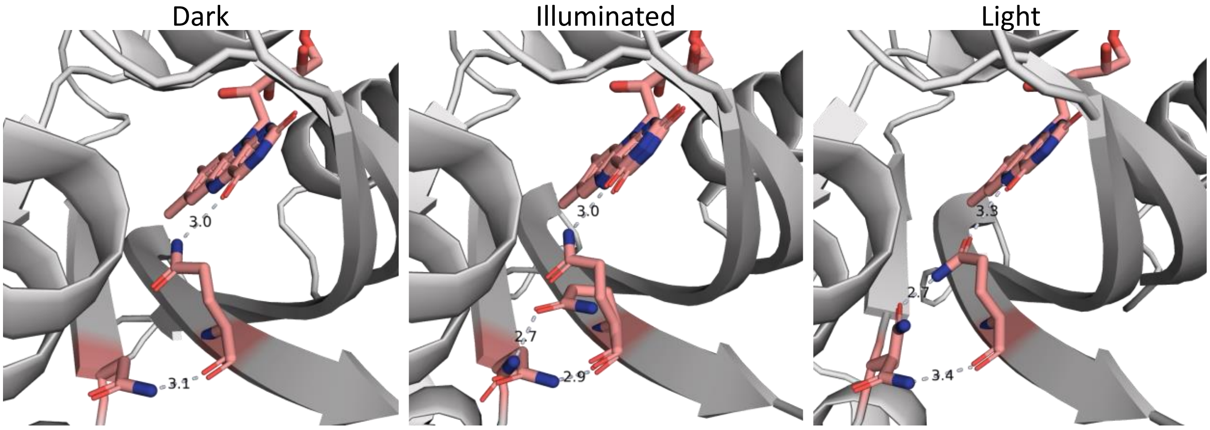


Figure 83. Crystal structures of AuLOV in three different states: dark (left), illuminated (middle) and light (right). PDB IDs: 6i20, 6i21 and 6i22.

Thanks to these crystal structures, the effect of the adduct formation was observed on the first two amino acids of the hydrogen bond network (Figure 84).[237] In the dark state, Gln$_{293}$ acts as a hydrogen bond donor to O$_4$ and N$_5$ of FMN, and Asn$_{194}$ forms a hydrogen bond with the carbonyl group of the peptide chain of Gln$_{293}$. Upon illumination and adduct formation, the Gln$_{293}$ sidechain flips to maintain the hydrogen bond with the now protonated N$_5$. This affects the neighbouring amino acid Asn$_{194}$, which also alters its sidechain orientation to form a new hydrogen bond between its carbonyl group and the amino group of Gln$_{293}$. These sidechain rotations likely cause the conformational changes in the A'α helix that result in effector activation. Another state, thought to be an intermediate in activation, was observed in the crystals harvested following illumination. The Gln$_{293}$ 'swings' away from the flavin but the conformation of Asn$_{194}$ is not changed. However, crystal lattice constraints may not allow widespread conformational changes in the A'α or Jα helices in these structures and more research is needed to follow the sidechain rotations in more distal amino acids.[237]

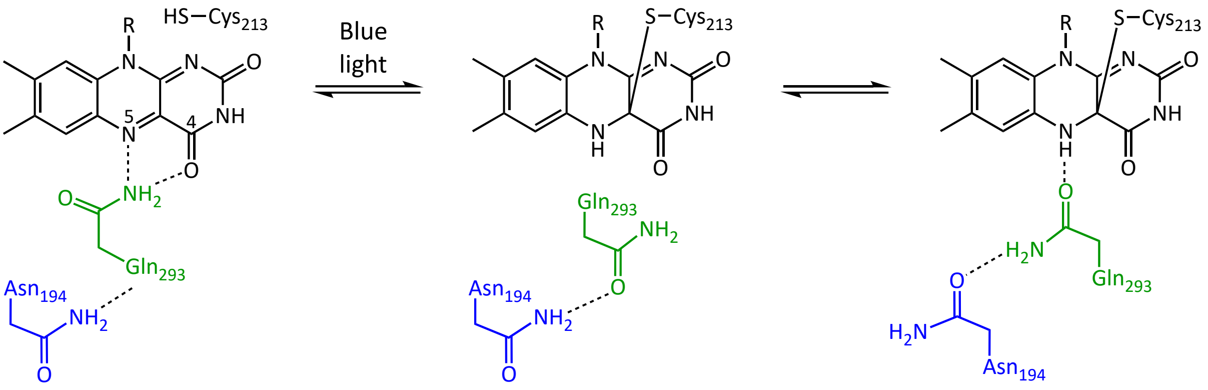


Figure 84. Change in orientation of the first two amino acids in the hydrogen bond network after illumination (Gln293 in green and Asn194 in blue).

4.1.1 Aim

The aim of the work described in this chapter was to characterise the role of asparagine 194 in signal transduction between the chromophore and the A'α helix of aureochrome 1a LOV domain from *O. danica* (AuLOV). This key amino acid was substituted for other residues with different properties and the effect of these substitutions on relaxation kinetics (UV-vis spectroscopy), oligomeric state (size exclusion chromatography) and sidechain orientation (protein crystallography) was studied. These techniques were performed in the presence and absence of blue light to understand how the conserved hydrogen bond network acts to transduce the signal from the chromophore to distal regions of the protein.

4.2 Results

4.2.1 Protein synthesis

To investigate the mechanism of signal transduction of *O. danica* aureochrome1a, AuLOV N194 variants were produced. The expression vector pETM11 containing AuLOV with a His_6-tag was used for protein expression.

Asparagine is a polar uncharged amino acid with the capacity to accept and donate hydrogen bonds. Seven mutations were chosen to investigate the effect of size, polarity, charge and hydrogen bonding ability on signal transduction between the chromophore and the A'a helix (Table 6).

Amino acid	Size	Polarity	Charge	H-bond ability
Asparagine (N)	medium	polar	neutral	donor/acceptor
Aspartate (D)	medium	polar	negative	acceptor
Histidine (H)	medium	polar	positive	donor/acceptor
Lysine (K)	large	polar	positive	donor
Leucine (L)	large	nonpolar	neutral	none
Glutamine (Q)	large	polar	neutral	donor/acceptor
Serine (S)	small	polar	neutral	donor/acceptor
Tyrosine (Y)	large	polar	neutral	donor/acceptor

Table 6. Characteristics of asparagine and the seven amino acids it was substituted by in AuLOV.

The point mutations at position 194 were performed using site directed mutagenesis. Primers containing the desired mutation were designed and the DNA was amplified, leading to the modified circular plasmids. The resulting DNA was sequenced to confirm correct introduction of the mutations (Figure 85). The extra point mutation observed in N194H (cytosine to adenine) was silent as the modified codon also coded for alanine.

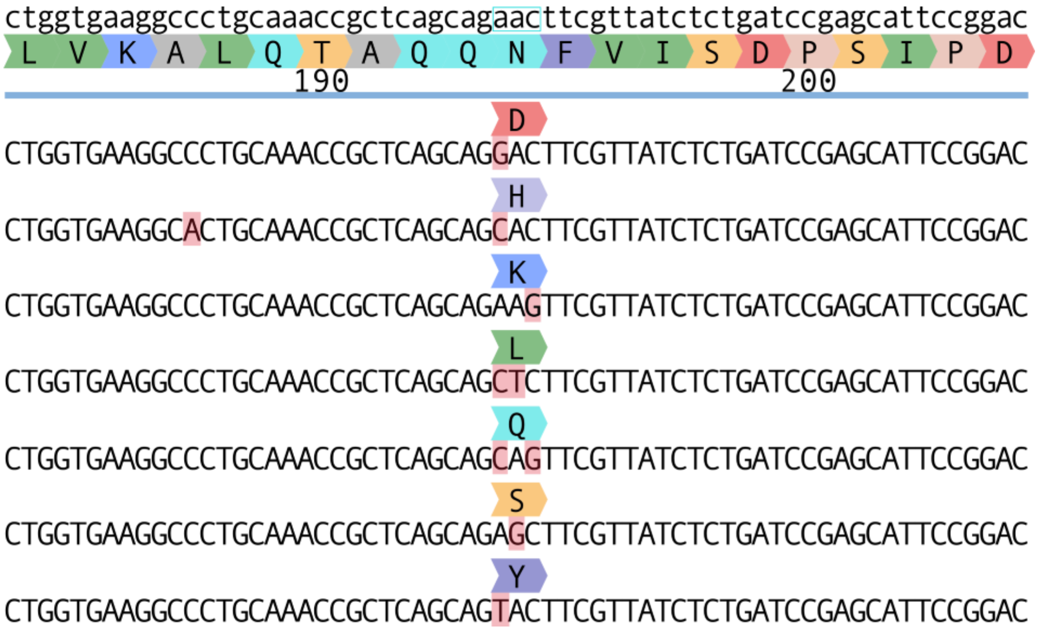


Figure 85. Sequencing results of the site directed mutagenesis performed on AuLOV to substitute N194 for seven different amino acids.

Wild type AuLOV and the seven variants were expressed and purified using BL21(AI) *E. coli* cells grown in autoinduction media. The variants were purified by Ni^{2+}-NTA affinity chromatography. The His_6-tag was removed from the fractions containing protein using TEV-His_6 protease before performing a second column for separation from the cut His_6-tag and TEV protease. Presence of protein could be assessed visually due to the characteristic yellow colour of LOV domains and the purity was determined by SDS-PAGE (Figure 86).

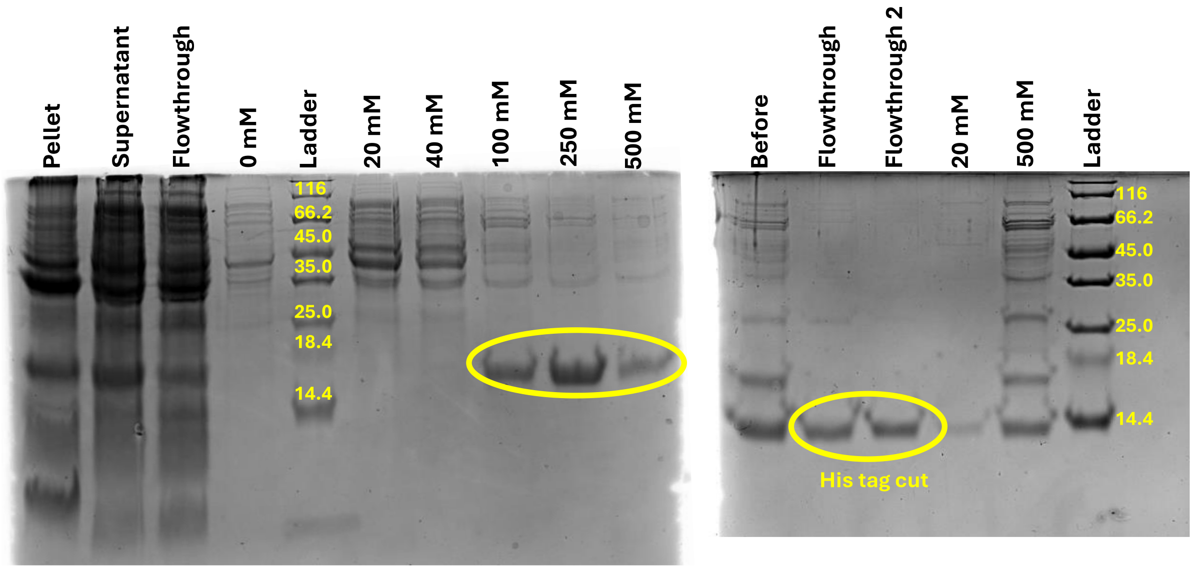


Figure 86. Example SDS-PAGE gels of Ni-NTA column purification of AuLOV N194Y (left) and second Ni-NTA column after His$_6$-tag removal using TEV-protease (right).

Following purification of the AuLOV variants, the protein purification buffer was exchanged for assay buffer (20 mM MES, 0.3 mM TCEP, 5 mM NaN_3, 1 mM EDTA, pH 6). The photochemical behaviour, oligomeric state and macromolecular structure were then assessed to determine the effect of the amino acid substitution on signal transduction.

4.2.2 UV-vis spectroscopy

The flavin-cysteinyl adduct thermally decays to regenerate the dark state at a rate that can vary by several orders of magnitude between LOV domains from different species. The thermal relaxation rate is limited by the deprotonation of the N_5 position of the FMN cofactor, which depends on the availability of a base.[68] The base can be internal, a basic amino acid, or external, such as the solvent or a base added to the buffer.[238] Due to this, changes in the chromophore environment vary the rate of thermal reversion. The conserved glutamine and the neighbouring asparagine, Gln$_{293}$ and Asn$_{194}$ in AuLOV, have been shown to have a significant effect on the photocycle kinetics in other LOV domains.[239] The asparagine is thought to control water access and coordination in the vicinity of the chromophore.[38]

Photocycle reversion kinetics were measured by UV-vis spectroscopy in the range 300-600 nm. The spectra were recorded periodically until the variant reached the dark state (Figure 87). The kinetics of the variants were monitored at 450 nm, the characteristic maximum of the non-covalently bound flavin in the dark state. Relaxation of the WT AuLOV takes place with a half-life of 112 min. The AuLOV variants remained photoreactive, they were able to form the flavin-cysteinyl adduct when illuminated with blue light and then relax to the ground state when kept in dark conditions for sufficient time. However, the substitution of asparagine for other amino acids caused changes in the rate of reversion to the dark state (Figure 88).

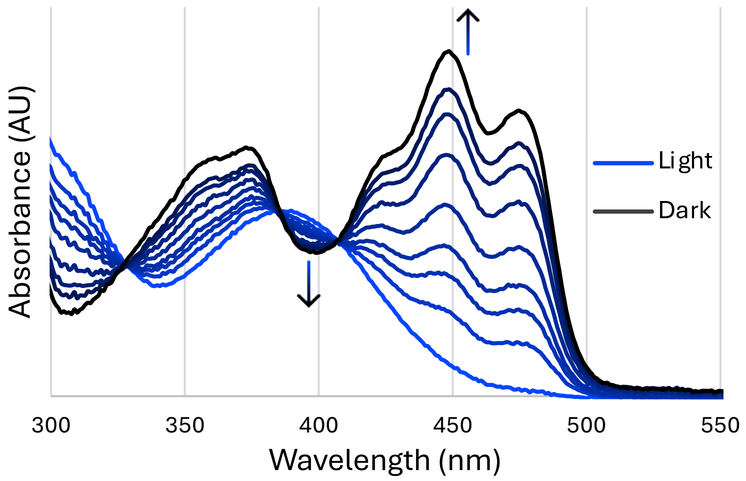


Figure 87. Absorption spectra of AuLOV N194Y. The sample was illuminated with blue-light and spectra were taken at 5 – 15 min intervals. The light state maximum at 390 nm decreases and the dark state maximum at 450 nm increases along with relaxation.

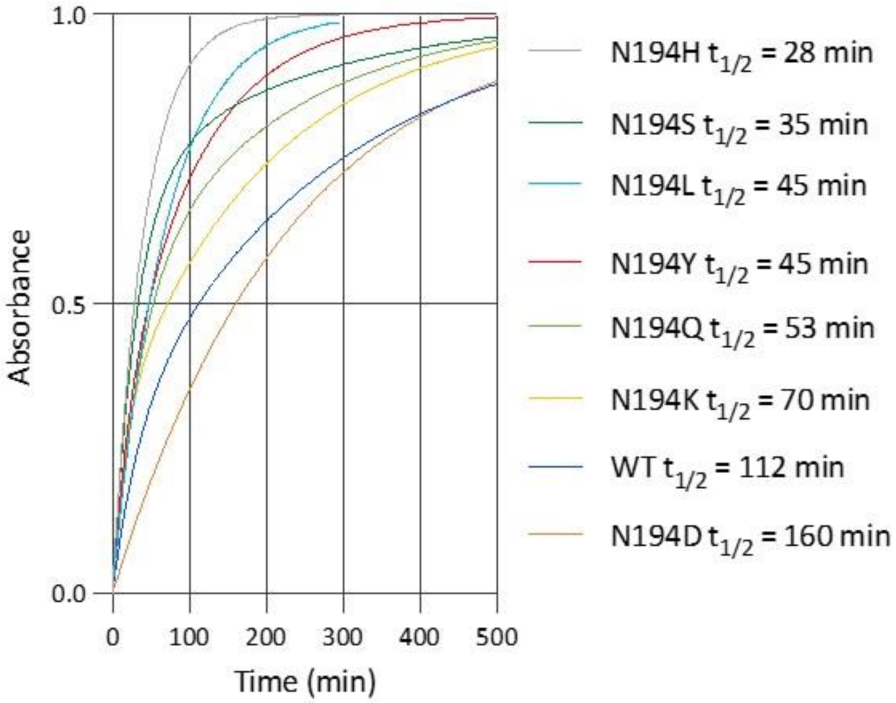


Figure 88. Relaxation kinetics of AuLOV WT and the seven variants followed at 450 nm (the dark state maximum) starting immediately after blue-light illumination with a 450 nm LED light.

The presence of aspartate slowed the thermal decay ($t_{1/2}$ > 112 min). Due to the acidic character of aspartate, a base was less available for deprotonation of the flavin, which hindered the rate limiting step.

The remaining variants showed a quicker relaxation ($t_{1/2}$ < 112 min), indicating that a base was able to deprotonate the N_5 position more readily than in the wild-type protein and/or solvent exposure of the chromophore was altered by the changes. The basic character of lysine and histidine likely facilitated the deprotonation of the cofactor. The hydroxy sidechains of serine and tyrosine possibly aided in the deprotonation through proton transfer. The inability of leucine to form hydrogen bonds and the larger size of glutamine were likely to render water molecules more accessible for deprotonation.

4.2.3 Size exclusion chromatography

Size exclusion chromatography was used to assess the capacity of the variants to dimerise in response to blue light illumination. SEC allows the separation of molecules in terms of size and is sensitive to changes in oligomeric state and shape. The Superdex 75 10/300 GL column, able to separate molecules with molecular weights between 3 and 70 kDa, is suitable for the AuLOV variants that have a molecular weight of ~15 kDa. Samples of 50 µM were used for all the variants to avoid concentration effects. The absorbance was monitored at 390 nm for the light state maximum of the cysteine-FMN adduct.

Attempts to generate a calibration curve using assay buffer were unsuccessful, possibly due to its acidic character (pH 6.0) or low salt concentration that increased the likelihood of aggregation of the protein standard mixture. A calibration curve was generated using the phosphate buffer recommended by the manufacturer (100 mM sodium phosphate, 100 mM sodium sulphate, 0.05 % sodium azide, pH 6.7). However, this meant that the apparent sizes of the samples could not be calculated accurately since the use of a different buffer can affect the elution volumes. Due to this, the elution volumes of the variants were compared to those of wild type in dark and light conditions.

The protein standard mixture contained four proteins and a low molecular weight marker. Thyroglobulin represented the void volume of the column and was not included in the calculation of the calibration curve (Table 7, Figure 89).

Protein	**MW (Da)**	**Log (MW)**	**V_e**
Thyroglobulin bovine	670000	5.826	9.00
y-globulins from bovine blood	150000	5.176	9.41
Albumin chicken egg grade VI	44300	4.646	11.67
Ribonuclease A type I-A from bovine pancreas	13700	4.137	14.18
p-aminobenzoic acid	137	2.137	29.61

Table 7. Protein standards of known molecular mass and elution volumes in phosphate buffer.

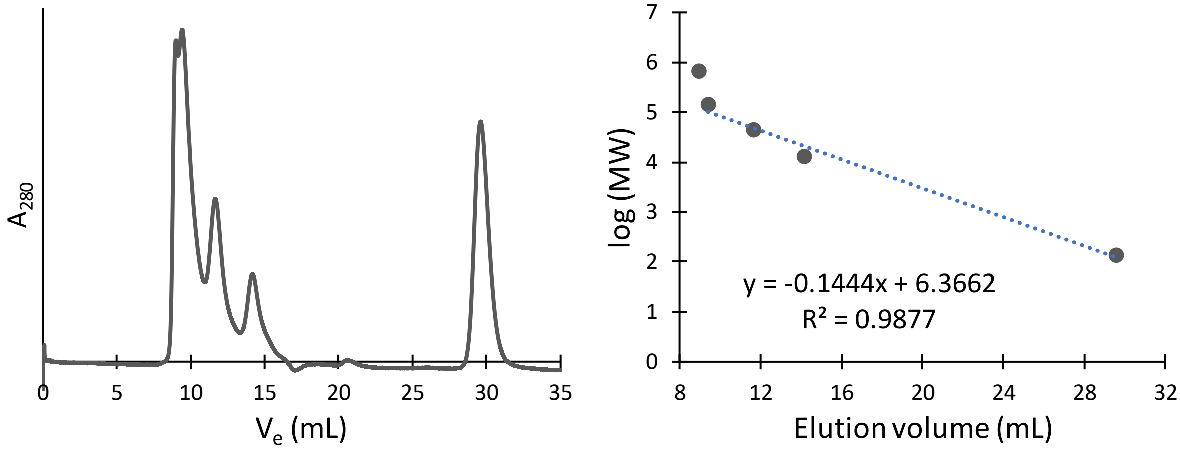


Figure 89. Chromatogram (left) and calibration curve (right) of protein standard mixture in phosphate buffer. The calculated equation does not include thyroglobulin as it represents the void volume.

Isolated aureochrome LOV domains are involved in a rapidly exchanging equilibrium between fully monomeric and fully dimeric form.[240] Dark-state WT AuLOV eluted at 11.97 mL, whereas the light-state elution maxima shifted to 11.34 mL. This 0.63 mL change in elution volume agrees with the shifting of the equilibrium towards the dimeric state upon illumination (Figure 90).[237]

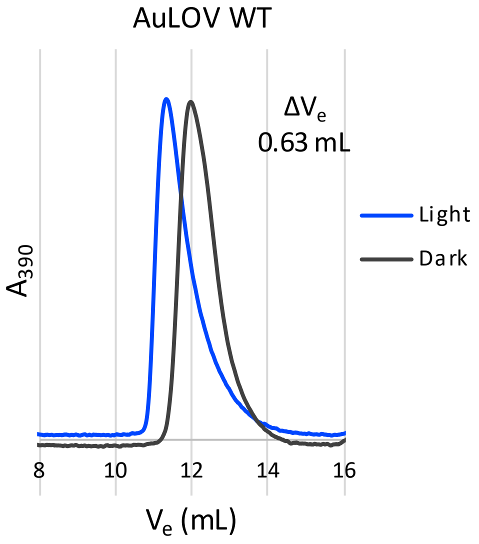


Figure 90. Size exclusion chromatogram of AuLOV in dark (black) and light (blue) states. The dark state maximum at 11.97 mL shifted to 11.34 mL when sample was illuminated for analysis, confirming an increase in apparent mass (dimerisation).

The AuLOV variants eluted at different volumes both in dark and light states. The difference in elution volumes between illumination states within each variant (ΔV_e) also changed, indicating different oligomerisation abilities due to effects of the amino acid substitutions on the signal transduction mechanism (Figure 91).

Variant	V_e Light	V_e Dark	ΔV_e
N194K	11.90	12.01	0.11
N194S	11.67	11.83	0.15
N194L	11.52	11.71	0.19
N194D	11.62	11.82	0.19
N194Q	11.30	11.71	0.41
wt	11.34	11.97	0.63
N194Y	11.02	11.63	0.62
N194H	11.01	11.74	0.72

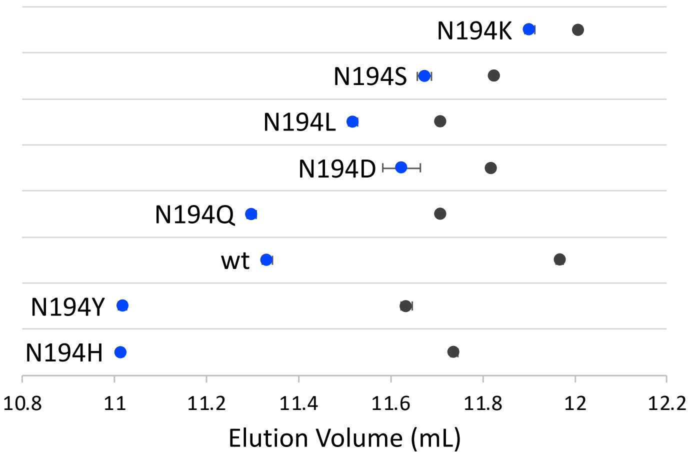


Figure 91. Elution volumes of the AuLOV variants in dark (black) and light (blue) conditions and ΔV_e for each of the variants. Higher difference in elution volumes between dark and light states indicates a higher degree of dimerisation upon illumination.

The presence of medium or large amino acids capable of acting as a hydrogen bond donor and acceptor was necessary to observe notable changes in elution volumes upon illumination (Figure 92). Since histidine, tyrosine and glutamine have similar H-bonding abilities to asparagine, they were able to participate in the network that connects the chromophore with the A'α helix and induce a change in oligomerisation state. The histidine and tyrosine variants have smaller elution volumes than wild type in both states, this higher apparent mass of the proteins signifies that the equilibrium is further shifted towards the dimeric form. AuLOV N194Q has a similar light elution volume as WT, but a smaller dark elution volume, indicating a smaller dynamic range that could be caused by the difference in size of the two amino acids. The change in elution volume for the histidine variant was larger than that of wild type. This increased dynamic range could potentially be exploited if it translated to a larger fold activation of DNA binding in the full aureochrome protein.

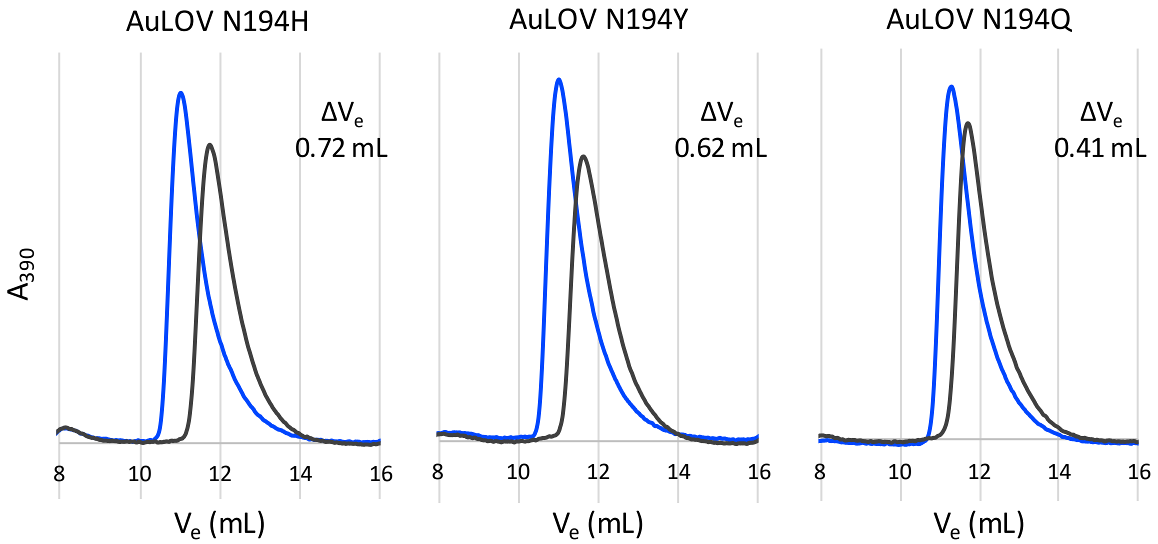


Figure 92. Size exclusion chromatogram of three AuLOV variants in dark (black) and light (blue) states. Amino acids with H-bond donor and acceptor ability maintain light dependent dimerisation.

The amino acid serine in position 194 abolished the ability of the AuLOV to dimerise, showing only a 0.15 mL change in elution volume between light and dark states. Although the hydroxy sidechain can be both a hydrogen bond donor and an acceptor, its small size is likely to prevent hydrogen bonding and the signal cannot be communicated to the A'α helix. Both these states were intermediate to the WT in dark and light and corresponded to an equilibrium intermediate between monomer and dimer.

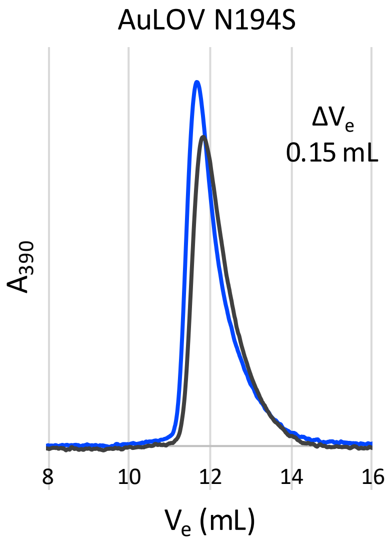


Figure 93. Size exclusion chromatogram of AuLOV N194S variant in dark (black) and light (blue) states. An amino acid with hydrogen bond donor and acceptor capability but small size leads to loss of light dependent dimerisation.

The presence of residues with nonpolar sidechains (leucine) or only able to act as H-bond donor (lysine) or acceptor (aspartate) abolished the light induced oligomerisation. Aside from the small ΔV_e, these three variants had different elution volumes. The equilibrium in AuLOV N194K was mostly towards the monomeric state due to the higher elution volumes observed, similar to the dark elution volume of WT AuLOV. On the other hand, the aspartate and leucine variants had an oligomeric state between WT dark and light conditions, representing a more balanced ratio of monomer and dimer in the equilibrium.

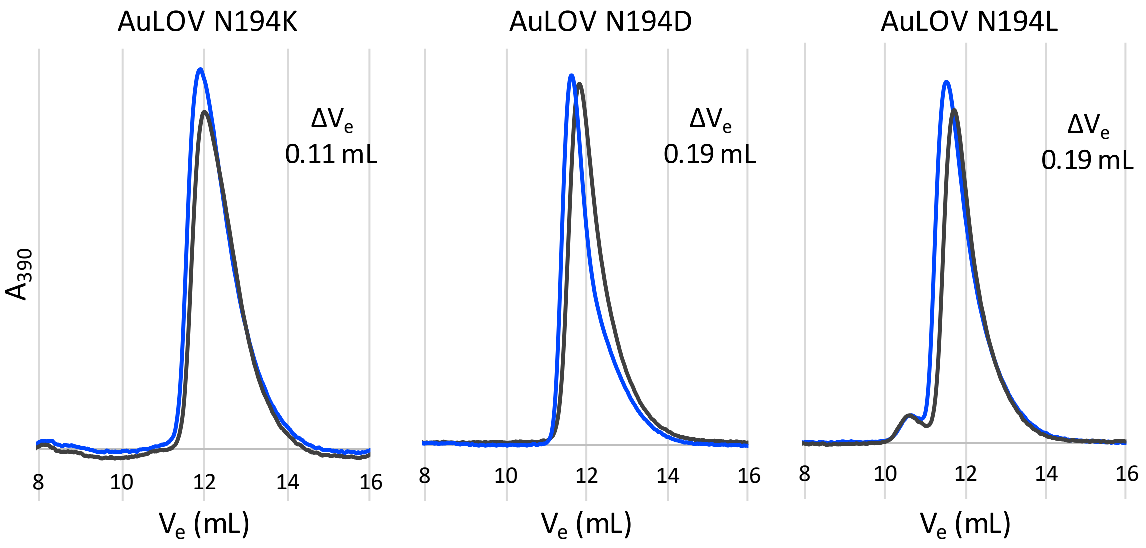


Figure 94. Size exclusion chromatogram of three AuLOV variants in dark (black) and light (blue) states. Amino acids without both H-bond donor and acceptor capability abolish the light dependent dimerisation ability.

4.2.4 Protein crystallography

X-ray crystallography was used to investigate the structural differences in the hydrogen bond network and sidechain orientation in different illumination states with atomic resolution.

Crystallisation was attempted for the seven variants under different illumination conditions using the sitting drop method. Crystals were grown under dark conditions and harvested either in dark conditions (dark state) or illuminated with blue light prior to harvesting (illuminated state). A screen containing 15 to 20 % w/v PEG 2000, 0.1 M NH_4Cl and 0.1 M sodium acetate (pH 4 to 5.5) or 0.1 M disodium citrate (pH 4 to 5.5) was prepared based on dark crystallisation conditions of WT AuLOV.[237,241] Crystals were also grown and harvested under blue light conditions (light state) using a crystallisation screen based on one that yielded light state crystals. [237,241] This screen contained 0.1 M tris-acetate buffer (pH 6.5 to 8.0) and 1.5 to 3.0 M disodium malonate and was supplemented with hexamminecobalt (III) chloride (10 mM) as an additive.

An additional AuLOV N194D variant containing 2,6-difluorotyrosine in place of Tyr_{182} (AuLOV N194D*) was used in the crystallisation trials. This variant had previously been created to study the role of the A'α helix using ^{19}F-NMR but was found not be a suitable method for assessing the oligomeric state of the protein. Due to its availability in the laboratory, it was set to crystallise in light and dark conditions using the aforementioned screens for each illumination condition.

Although all variants were set to crystallise both in dark and light conditions, not all yielded crystals (Table 8). Crystals of variant N194D were obtained both in dark and light growth conditions. In addition, crystals of variants N194D*, N194Q and N194Y were obtained in dark conditions and crystals of variants N194H and N194S in light conditions. Two variants (N194L and N194K) were not successfully crystallised in any of the conditions. AuLOV N194K formed microcrystals that were used as seeds in subsequent trials.

Variant	**D**	**D***	**H**	**K**	**L**	**Q**	**S**	**Y**
Dark	✓	✓				✓		✓
Light	✓		✓				✓	

Table 8. Summary of the crystals obtained from the crystallisation trials of all variants in both crystallisation conditions.

The crystals obtained in dark conditions were harvested in dim red light conditions (dark state) and upon blue light illumination (illuminated state). Those grown under blue light were harvested whilst illuminated (light state). The harvested crystals were frozen in liquid nitrogen and sent to Diamond Light Source (DLS) for data collection. Cryoprotection was not required during the harvesting process due to the presence of PEG and sodium malonate in the crystallisation screens. The X-ray diffraction and data processing that took place in DLS was successful for some of the samples and led to different resolutions (Table 9).

Variant	**D**	**D***	**H**	**Q**	**S**	**Y**
Dark	✓, ~2Å	✗	╱	✗	╱	✓, ~3Å
Illuminated	✓, ~2Å	✓, >1Å	╱	✗	╱	✗
Light	✓, >3Å	╱	✗	╱	✓, ~3Å	╱

Table 9. Summary of the datasets obtained and their resolution from the crystals sent to Diamond. No crystals had been obtained for the variants and conditions in crossed out cells. Crystals from cells marked with ✓ led to diffraction datasets and with ✗ did not.

Only diffraction datasets with resolutions of 2 Å and below were used since high resolution was needed to determine the sidechain orientation with certainty. From the six datasets that were obtained, those with low resolution (≥ 3 Å) belonging to N194D (light), N194S (light) and N194Y (dark) were not used. The datasets that were used corresponded to AuLOV N194D in the dark and illuminated states and AuLOV N194D* in the illuminated state (Figure 95).

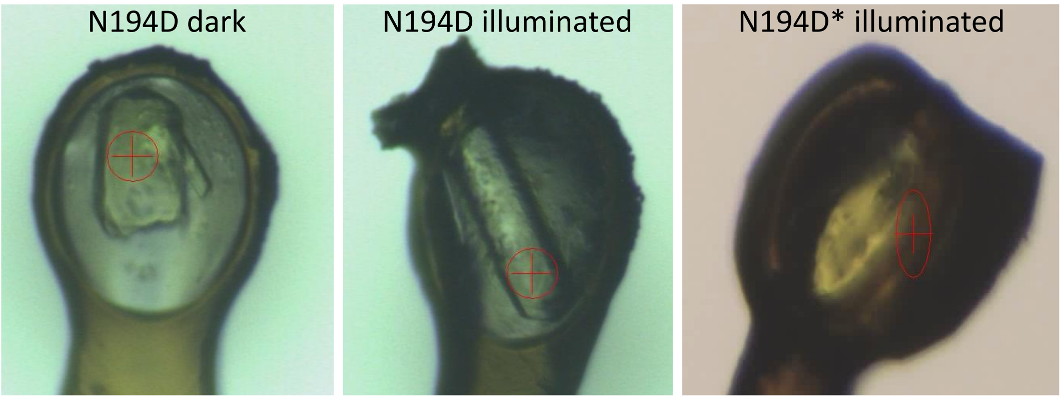


Figure 95. Crystals that led to good quality datasets.

Initially, the cell parameters obtained from the data processing (Table 10) were compared to the three structures of WT AuLOV (Table 11). The three datasets had parameters belonging to a hexagonal lattice type, like the wild type illuminated structure.

	N194D dark	**N194D illuminated**	**N194D* illuminated**
a, b, c (Å)	106.1, 106.1, 67.03	106.58, 106.58, 66.98	150.18, 91.51, 99.84
h, k, l	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Lattice type	Hexagonal	Hexagonal	Hexagonal
Space group	P 6_2 2 2	P 3_1 2 1	P 3_1 2 1

Table 10. Cell parameters of the datasets obtained.

	Dark (6i20)	**Illuminated (6i21)**	**Light (6i22)**
a, b, c (Å)	65.68, 104.04, 105.94	106.39, 106.39, 67.49	150.18, 91.51, 99.84
h, k, l	90.0, 90.0, 90.0	90.0, 90.0, 120.0	90.0, 131.34, 90.0
Cell type	Orthorhombic	Hexagonal	Monoclinic
Space group	P 2_1 2_1 2_1	P 6_4 2 2	C 1 2 1

Table 11. Cell parameters of the wild type AuLOV structures.[241]

The crystal structures were solved using CCP4i2 and Coot software.[242,243] Molecular replacement was carried out using aureochrome 1a LOV domain from *Ochromonas danica* in the dark or illuminated state as a template (PDB entries 6i20, 6i21). Structural refinement was achieved using alternate cycles of Coot and REFMAC5. The flavin chromophore, solvent molecules, ions and, when needed, 2,6-difluorotyrosine were added where positive electron density peaks appeared.

The crystals of N194D variant grown and harvested in dark conditions diffracted with 1.81 Å resolution, in space group P 6_2 2 2 and it contained a monomer per asymmetric unit (Figure 96).

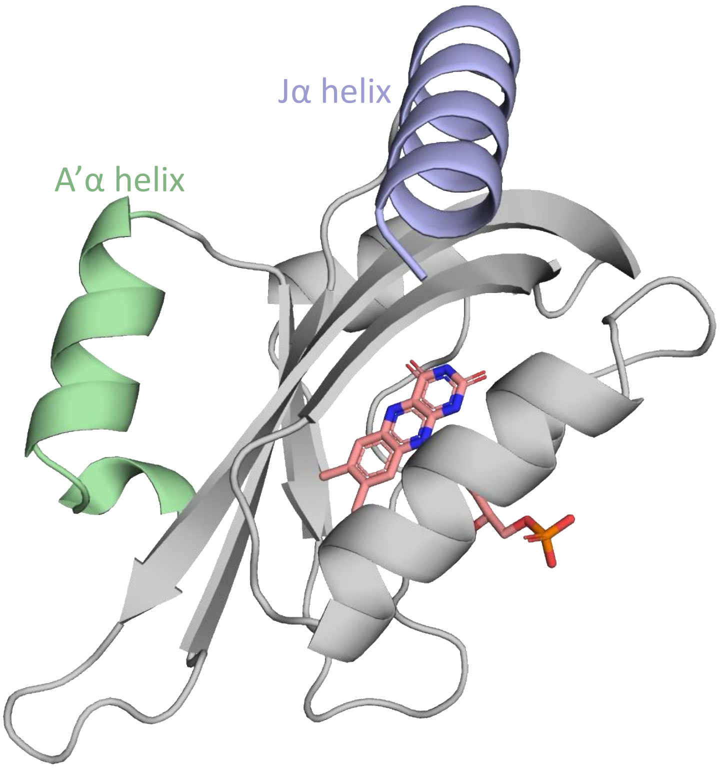


Figure 96. Structure of dark state AuLOV N194D. The asymmetric unit contains one monomer. The A'α helix (green) and Jα helix (blue) are positioned along the β-sheet surface.

The N194D variant crystals grown in the dark and illuminated while harvesting diffracted with 2.06 Å resolution, in space group P 3_1 2 1 and contained two monomers per asymmetric unit (Figure 97).

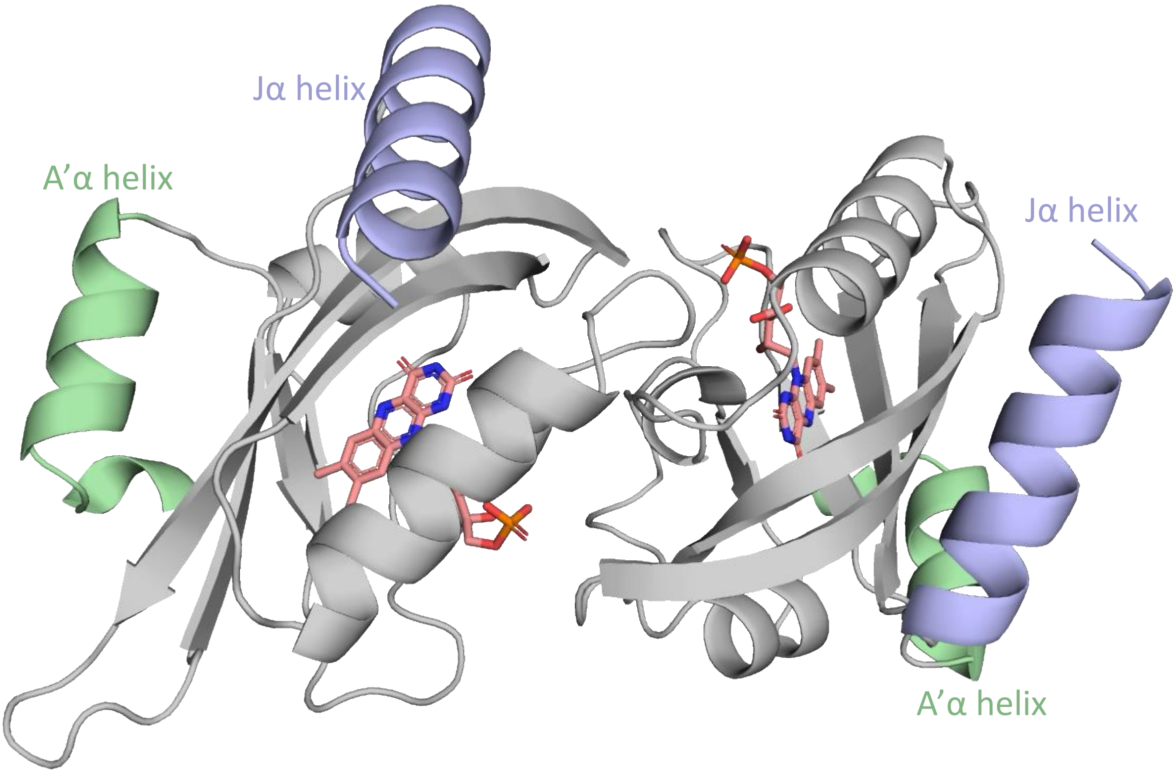


Figure 97. Structure of illuminated state AuLOV N194D. The asymmetric unit contains two monomers. The A'α (green) and Jα helices (blue) are positioned along the β-sheet surface.

Crystals of AuLOV N194D*, containing 2,6-difluorotyrosine in position 182, grown in dark conditions and illuminated before harvesting diffracted with 1.26 Å resolution, in space group P 3_1 2 1 and contained two monomers per asymmetric unit (Figure 98).

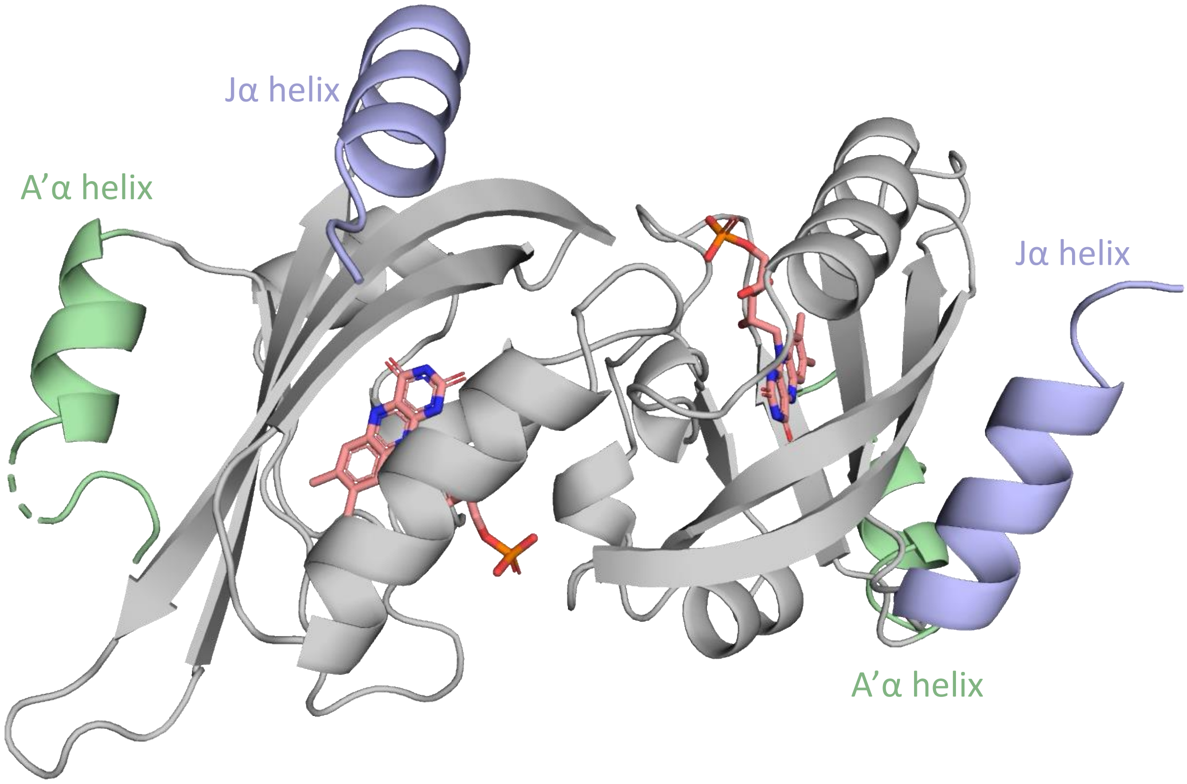


Figure 98. Structure of illuminated state AuLOV N194D. The asymmetric unit contains two monomers. The A'α (green) and Jα helices (blue) are positioned along the β-sheet surface.*

The LOV domain core of the three structures appeared to be identical, with the exception of 2,6-difluorotyrosine in AuLOV N194D* being in a slightly different conformation than Tyr_{182}. The three structures were most similar to the illuminated state of the wild type although two changes were visible. Firstly, there was no H-bond between the sidechain of Asp_{194} and main chain of Gln_{293} since the measured distance was larger than 3.3 Å. Secondly, the Gln_{293} did not swing away from the flavin in the illuminated states (Figure 99). Since the carboxylic group of the aspartate was not able to form the hydrogen bond with the amine group of the glutamine, this change in sidechain orientation cannot take place. The "swing" state is thought to be an intermediate and if it cannot form, it may not be able to proceed to the light state conformation, explaining the inability of this variant to dimerise upon illumination.

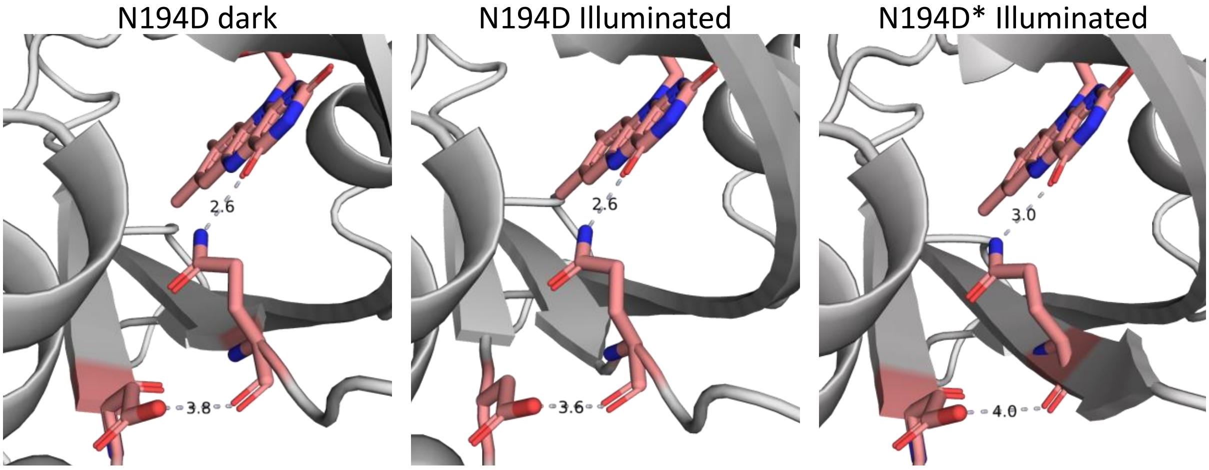


Figure 99. Crystal structures of three AuLOV variants: N194D dark (left), N194D illuminated (middle) and N194D illuminated (right). The illuminated structures do not have the "swing" glutamine conformation that was present in WT AuLOV.*

Further attempts to crystallise N194D in light conditions were not successful in yielding crystals big enough for single crystal diffraction but produced microcrystals. Since this variant showed impaired dimerisation upon illumination, being able to obtain and compare crystal structures for all three illumination states with the wild-type protein would provide more insight into signal transduction and why it does not occur for this aspartate variant. Due to the interest in obtaining a high-resolution crystal structure for AuLOV N194D in light state, the microcrystals were measured in the state-of-the-art VMXm beamline in DLS (Figure 100). This beamline enables data collection from crystals 2-10 µm in size, too small for single crystal diffraction experiments. The samples were prepared on transmission electron microscopy grids that had been modified to hold a negatively charged surface. The crystal suspensions were applied to the grid and blotted to remove excess liquid, which improved the signal to noise ratio and reduced the presence of ice rings. The loaded grids were cooled using a plunge freezing instrument and mounted on special holders to load onto the beamline. Data was collected under vacuum and at low temperatures to reduce background noise. Each microcrystal was individually focused, and data was recorded at different angles. Diffraction data from different microcrystals were then combined to provide complete datasets. Unfortunately, this data collection did not yield better results than the single crystal diffraction experiments and the resolution reached was > 3 Å.

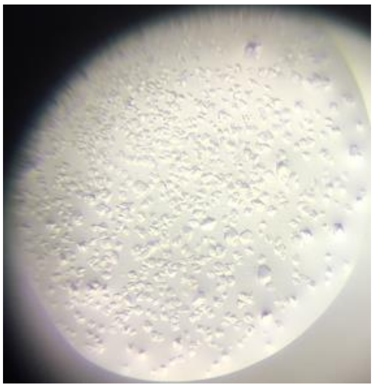


Figure 100. Small crystals obtained from AuLOV N194D under blue light conditions that were taken to VMXm beamline at DLS for data collection.

4.3 Conclusions and future work

Previous studies have shown the involvement of Asn_{194} in the hydrogen bond network linking the chromophore with the A'α helix. Hence, further investigation on the mechanism of signal transduction was warranted utilising site directed mutagenesis.[237,241] Seven variants of aureochrome1a LOV domain from *O. danica* with substitutions of N194 were isolated. The amino acids were selected to test the effect of size, charge and hydrogen bonding ability in signal transduction.

All the variants were able to interact with blue light, forming the cysteinyl-flavin adduct and decaying to the dark state. However, the rate of the relaxation varied greatly. An amino acid of acidic character in position 194 slowed the photocycle kinetics, whereas the rest of amino acids accelerated the thermal relaxation due to increased basicity, hydrogen transfer ability or solvent accessibility. This supports the importance of Asn_{194} in the rate limiting deprotonation step.[38,239]

Size exclusion chromatography assays illustrated the need for the amino acid in position 194 to be medium to large size and be able to act as a hydrogen bond donor and acceptor for successful light induced dimerisation. A small amino acid able to act as hydrogen bond donor and acceptor (serine) abolished the ability to dimerise upon illumination, highlighting the importance in size for signal transduction. The presence of a nonpolar residue or a residue only able to act either as a H-bond donor or acceptor also supressed the light induced dimerisation.

The AuLOV variant N194H showed a larger degree of dimerisation in light conditions than WT AuLOV. The N194H variant of the full aureochrome1a would need to be studied to see if this increased ΔV_e translated to a larger fold activation of DNA binding upon illumination. Improved dynamic ranges of activity in the dark and light states are key for the success of LOV domain-based optogenetic tools.

Crystals of six variants were obtained, harvested and sent for data collection. Three datasets were obtained with high resolution, suitable to observe the changes in sidechain orientation that are to be followed. Crystal structures were obtained of N194D in dark and illuminated states and of N194D* (containing 2,6-difluorotyrosine) in illuminated state. They showed that the hydrogen bond between the sidechain of amino acid 194 and the main chain of Gln_{293} had been broken and that the glutamine did not have a "swing" conformation in the illuminated states. Since this conformation is thought to be an intermediate, it would hinder formation of the light state conformation. This is consistent with the inability of N194D to dimerise upon illumination seen with SEC.

Due to difficulties in the crystallisation process, a structure of AuLOV N194D could not be obtained in blue light conditions. Achieving this would be beneficial as it would show whether undergoing any conformational changes upon illumination is possible for this variant. Additionally, obtaining crystal structures of the N194H variant in the different illumination states would also provide more insight into the signal transduction mechanism. Since this variant dimerised upon illumination with a bigger dynamic range than WT, crystal structures would reveal if the changes in sidechain orientation that take place are the same as in WT or if there is more than one possible mechanism.

Elucidating the mechanism of signal transduction in AuLOV could lead to the development of new optogenetic tools and improvements on the dynamic ranges of DNA binding affinities. The slow cycling character of this protein would be excellent for applications where an extended light activation was needed with small amounts of light.

Chapter 5 - Materials and methods

5.1 General

Chemicals were purchased from Sigma-Aldrich UK, Fisher Scientific Ltd, Apollo Scientific UK, Fluorochem Ltd or Melford Laboratories Ltd and were used without further purification unless otherwise specified.

SYBR safe DNA gel stain, FastDigest Green buffer (10x), and DNA ladder were obtained from Thermo Fisher Scientific Ltd. PrimeSTAR HS DNA polymerase (premix) was obtained from Takara Bio Europe. Golden Gate cloning reagents (T4 reaction buffer, T4 DNA ligase, BSA and BsaI-HFv2), KLD reaction buffer, Dpn1 and enzyme mix and HiFi DNA assembly Master Mix were purchased from New England BioLabs Inc. Easy Pure® Plasmid MiniPrep Kit was obtained from TransGen Biotech. Oligonucleotide primers and all other biological reagents were obtained from Sigma-Aldrich UK. DNA sequencing was completed by Eurofins Genomics. PACT Premier and JCSG Plus crystallography screens, 2-drop crystallisation plates and mounted LithoLoops for protein crystallisation were purchased from Molecular Dimensions. Biofilm assays were performed in U bottom 96 well microtiter plates from Thermo Fisher Scientific (Part no. 612U96). Fluorescence was measured in black flat-bottomed 96 well plate from Thermo Scientific (Part no. 137101).

PCR and Golden Gate assembly were performed indistinctively in Techne Touchgene Gradient Thermal Cycler or Biometra Thermocycler T-Gradient Thermoblock. Gel images were visualised using a BioRad ChemiDoc MP imaging system. DNA concentration was measured via a Thermo Fisher Scientific NanoDrop 3300 Fluorospectrometer. UV/vis spectroscopy was completed with a Shimadzu UV-2600 UV-Vis spectrophotometer. Fluorescence was measured using a VICTOR X5 Multilabel Plate Reader (PerkinElmer) or a FLUOstar Omega Microplate Reader (BMG Labtech). Size exclusion chromatography was performed in a Cytiva ÄKTA pure protein purification system. Crystallisation trials were performed using an Oryx 4 robot (Douglas Instruments).

^{1}H and ^{13}C-NMR spectra were measured on a Bruker DRX 600 MHz with cryoprobe, Bruker Avance 500 MHz and Bruker Ascend 400 MHz NMR spectrometers. GCMS was performed using a PerkinElmer Clarus 680 gas chromatograph and a Clarus SQ 8 C mass spectrometer. LCMS was carried out by the university service, using a Waters Synapt G2-Si mass spectrometer with electrospray ionisation and time of flight detection. HRMS was carried out by the university service, using a Xevo G2-Xs QTof mass spectrometer. Flash chromatography was performed on a Biotage Isolera™ Four using Biotage® SNAP Ultra HP-Sphere™ 25 µM columns 10 – 100 grams. HPLC was carried out on a Shimadzu preparative scale HPLC system equipped with a C_{18} column, an FRC-10A Fraction collector and an SPD-M40 photo diode array detector.

Different illumination conditions for dark/light experiments were achieved using LED lamps. For dark conditions, samples were covered in aluminium foil. Dim red light conditions were used when handling dark samples, working in a room without windows whenever possible and using a red LED lamp (623 nm). Illumination of samples with blue light was achieved with a blue LED lamp (450 nm).

5.2 Chemical methods

5.2.1 Synthesis of HBC_{620} (Adapted from [192])

2-Bromothieno[3,2-b]thiophene (2)

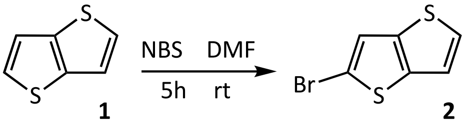


*Figure 101. Synthesis of compound **2**.*

N-bromosuccinimide (0.64 g, 3.57 mmol, 1 eq.) was added to a solution of thieno[3,2-b]thiophene (0.5 g, 3.57 mmol, 1 eq.) in dimethylformamide (20 mL). The mixture was stirred at room temperature for 5 hours under inert atmosphere. It was then poured into water (100 mL) and extracted with dichloromethane (3 x 100 mL). The combined organic phases were dried with anhydrous sodium sulfate, filtered and the solvent evaporated. The light purple solid obtained (69% yield, **2**) was used without further purification.

GC-MS: 11.57 min, m/z = 217.98; calculated 217.89 for $C_6H_3BrS_2$.

2-[(2-hydroxyethyl)methylamino]thieno[3,2-*b*]thiophene (4)

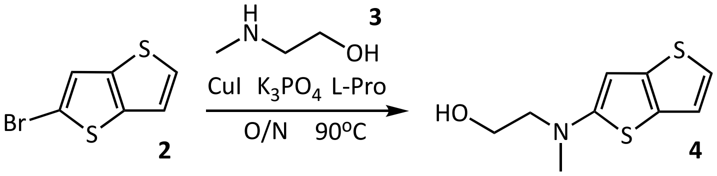


*Figure 102. Synthesis of compound **4**.*

2-Bromothieno[3,2-b]thiophene (~0.5 g) was dissolved in 2-methylaminoethanol (10 mL) and copper (I) iodide (76 mg, 0.4 mmol, cat.), potassium phosphate (0.829 g, 6 mmol) and (L)-proline (92 mg, 0.8 mmol, cat.) were added. The mixture was stirred overnight at 90 °C. The reaction mixture was cooled down to room temperature and water (50 mL) was added. It was extracted with dichloromethane (3 x 500 mL), dried with anhydrous sodium sulfate and filtered. The solvent was removed, and the product was purified by silica column (Hex/EA = 1:1, R_f = 0.5), giving compound **4** (65 mg, 24 %) as a yellow liquid that turns blue with time.

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.07 (dd, *J* = 5.2, 0.6 Hz, 1H), 7.03 (d, *J* = 5.2 Hz, 1H), 6.17 (d, *J* = 0.5 Hz, 1H), 3.82 (t, *J* = 5.5 Hz, 2H), 3.39 (t, *J* = 5.5 Hz, 2H), 2.99 ppm (s, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 160.4, 138.9, 126.7, 121.0, 119.5, 96.3, 59.9, 58.3, 41.2 ppm.
HRMS (TOF AP$^+$) calculated 214.04; found 214.04 for $C_9H_{12}NOS_2$ [M+H]$^+$.

5-[(2-acetoxyethyl)methylamino]thieno[3,2-*b*]thiophene (5)

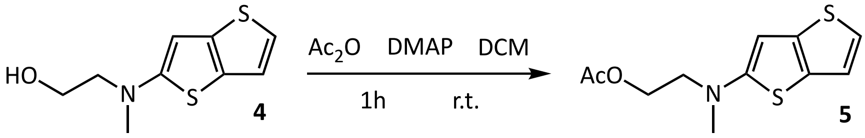


*Figure 103. Synthesis of compound **5**.*

Acetic anhydride (29 µL, 0.31 mmol, 1 eq.) was added to a solution of compound **4** (65 mg, 0.31 mmol, 1 eq.) and 4-dimethylaminopyridine (44.7 mg, 0.37 mmol, 1.2 eq.) in dry dichloromethane (1 mL) at 0 °C. The mixture was allowed to reach room temperature and stirred for 1 hour. The reaction mixture was quenched with water (1 mL) and extracted with dichloromethane (3 x 5 mL). The organic phase was dried with anhydrous sodium sulfate, filtered and the solvent evaporated under vacuum. It was purified by silica column (DCM, R_f = 0.74) to give a yellow compound (**5**, 48.3 mg, 61% yield).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.06 (dd, *J* = 5.2, 0.6 Hz, 1H), 7.02 (d, *J* = 5.2 Hz, 1H), 6.12 (d, *J* = 0.5 Hz, 1H), 4.30 (t, *J* = 5.7 Hz, 2H), 3.52 (t, *J* = 5.4 Hz, 2H), 3.05 (s, 3H), 2.03 ppm (s, 3H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 171.06, 159.62, 139.17, 126.39, 120.78, 119.49, 95.45, 61.42, 54.46, 40.81, 21.03 ppm.

5-[(2-acetoxyethyl)methylamino]thieno[3,2-*b*]thiophene-2-carbaldehyde (6)

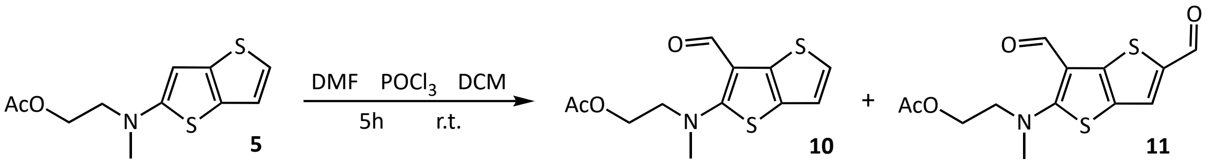


*Figure 104. Synthesis of compounds **10** and **11**.*

A solution of compound **13** (128 mg, 0.5 mmol, 1 eq.) in dry dichloromethane (5 mL) and dimethylformamide (0.5 mL, 6.5 mmol, 13 eq.) under inert atmosphere was cooled to 0 °C, and phosphorus oxychloride (57 µL, 0.6 mmol, 1.2 eq.) was added. The mixture was warmed to room temperature and stirred for 5 hours. The reaction mixture was quenched with a saturated aqueous solution of sodium carbonate (5 mL), extracted with dichloromethane (3 x 10 mL), dried with anhydrous sodium sulfate and filtered. The solvent was removed, and the crude product was purified by silica column (Hex/EA = 1:2). Compounds **10** and **11** were obtained from this reaction.

Monoaldehyde (**10**): ^{1}**H NMR** (500 MHz, $CDCl_3$) δ 10.00 (s, 1H), 7.19 (d, *J* = 5.2 Hz, 1H), 7.01 (d, *J* = 5.2 Hz, 1H), 4.33 (t, *J* = 5.6 Hz, 2H), 3.73 (t, *J* = 5.6 Hz, 2H), 3.27 (s, 3 H), 1.98 ppm (s, 3H).

Dialdehyde (**11**): ^{1}**H NMR** (500 MHz, $CDCl_3$) δ 10.02 (s, 1H), 9.79 (s, 1H), 7.64 (s, 1H), 4.38 (t, *J* = 5.5 Hz, 2H), 3.83 (t, *J* = 5.5, 2H), 3.38 (s, 3H), 1.98 ppm (s, 3H); **LRMS** (TOF ES$^+$) calculated for $C_{13}H_{13}NO_4S_2$ [M+H]$^+$: 312.03, found 312.04.

5.2.2 Improved synthesis of HBC_{620}

Thieno[3,2-b]thiophene-2-carbaldehyde (12) (Adapted from [244])

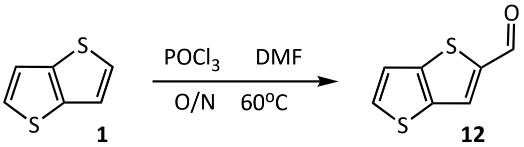


*Figure 105. Synthesis of compound **12**.*

A solution of thieno[3,2-b]thiophene (500 mg, 3.57 mmol, 1 eq.) in DMF (3 mL) was cooled to 0 °C. A mixture of phosphorus oxychloride (1 mL, 10.7 mmol, 3 eq.) and DMF (3 mL) was cooled to 0 °C and added dropwise to the previous solution. The reaction mixture was allowed to reach room temperature and consequently stirred at 60 °C overnight under inert atmosphere. The reaction mixture was poured over ice and water. Saturated aqueous sodium carbonate solution was added until pH reached 8-9. The mixture was extracted with dichloromethane (3 x 75 mL) and the organic phase was washed with water (3 x 150 mL), dried over anhydrous sodium sulfate and filtered. The solvent was evaporated under vacuum with the help of heptane to give compound **12** (560 mg, 3.32mmol, 93%) as an orange-brown solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 9.97 (s, 1H), 7.95 (d, *J* = 0.7 Hz, 1H), 7.70 (d, *J* = 5.3 Hz, 1H), 7.34 ppm (dd, *J* = 5.3, 0.7 Hz, 1H).
^{13}C NMR (126 MHz, $CDCl_3$) δ 183.77, 145.90, 145.58, 139.34,134.05, 129.27, 120.33 ppm.
HRMS (TOF ES+) Calculated for $C_7H_4OS_2$ $[M+H]^+$: 167.97, found 167.97

5-bromothieno[3,2-b]thiophene-2-carbaldehyde (13) (Adapted from [245])

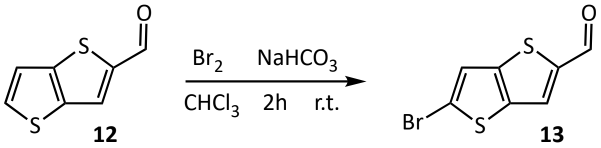


*Figure 106. Synthesis of compound **13**.*

To a solution of thieno[3,2-b]thiophene-2-carbaldehyde (263 mg, 1.56 mmol, 1 eq.) in chloroform (5 mL) was added sodium bicarbonate (155 mg, 1.84 mmol, 1.2 eq.) and the solution was cooled to 0 °C. Bromine (80 µL, 1.56 mmol, 1 eq.) was added dropwise and the mixture was stirred at room temperature for 2 hours under inert atmosphere. The reaction mixture was extracted with water (20 mL) and the aqueous phase was washed with dichloromethane (3 x 15 mL). The organic extracts were combined, dried with anhydrous magnesium sulfate and filtered. The solvents were evaporated under vacuum and the grey-orange residue was purified by recrystallisation in cyclohexane to give product **13** (207 mg, 0.84 mmol, 54%) as an orange solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 9.97 (s, 1H), 7.84 (d, *J* = 0.7 Hz, 1H), 7.36 ppm (d, *J* = 0.7 Hz, 1H)
^{13}C NMR (126 MHz, $CDCl_3$) δ 183.42, 144.88, 144.68, 139.48, 128.31, 123.05, 121.08 ppm.
HRMS (TOF ES$^+$) Calculated for $C_7H_3BrOS_2$ $[M+H]^+$: 246.88, found 246.87.

5-[(2-hydroxyethyl)methylamino]thieno[3,2-*b*]thiophene-2-carbaldehyde (7) (Adapted from [192])

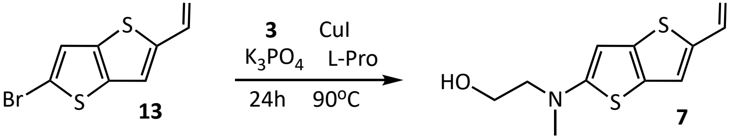


*Figure 107. Synthesis of compound **7**.*

Compound **13** (207 mg, 0.84 mmol, 1 eq.) was dissolved in 2-methylaminoethanol (5 mL) and copper (I) iodide (17.5 mg, 0.09 mmol, 0.11 eq.), potassium phosphate (302 mg, 1.42 mmol, 1.7 eq.) and (L)-proline (21.2 mg, 0.18 mmol, 0.22 eq.) were added. The mixture was stirred at 90 °C for 24 hours under inert atmosphere. After the reaction mixture was cooled to room temperature, water (20 mL) was added. It was extracted with dichloromethane (3 x 150 mL), dried over anhydrous sodium sulfate and filtered. The solvent was evaporated under vacuum and the residue was purified by silica column (Hex/EA = 1:2, R_f = 0.35), giving compound **7** (31.5 mg, 0.13 mmol, 15%) as a yellow solid.

^{1}H NMR (500 MHz, $CDCl_3$) δ 9.72 (s, 1H), 7.66 (s, 1H), 6.08 (s, 1H), 3.92 (t, *J* = 5.4 Hz, 2H), 3.54 (t, *J* = 5-4 Hz, 2H), 3.14 ppm (s, 3H).

^{13}C NMR (126 MHz, $CDCl_3$) δ 181.51, 165.71, 150.22, 138.02, 130.10, 126.01, 93.92 ppm.

HRMS (TOF ES^+) Calculated for $C_{11}H_{12}N_1O_2S_2$ $[M+H]^+$: 242.02, found 242.03.

HBC$_{620}$ (9) (Adapted from [192])

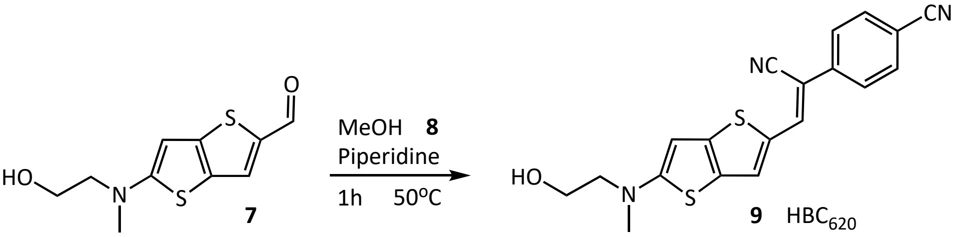


*Figure 108. Synthesis of compound **9**.*

Compound **7** (31.5 mg, 0.13 mmol, 1 eq.) and 4-cyanophenylacetonitrile (**8**) (40.8 mg, 0.29 mmol, 2.2 eq.) were dissolved in dry methanol (5 mL). Piperidine (1 drop, cat.) was added and the reaction mixture was stirred for 1 hour at 50 °C. The mixture was cooled to room temperature and the solvent was evaporated under vacuum. The solid was purified by silica column (5% MeCN in DCM, Rf = 0.36) and the fractions containing product were kept in the dark overnight before evaporating to give the product (**9**) as a red solid (14.24 mg, 0.039 mmol, 30%).

^{1}H NMR (600 MHz, d_6-DMSO) δ 8.39 (s, 1H), 7.89 (d, *J* = 0.7 Hz, 2H), 7.85 (s, 1H), 7.80 ppm (d, *J* = 0.7 Hz, 2H), 6.36 (s, 1H), 4.88 (t, *J* = 5.4 Hz, 1H), 3.66 (q, *J* = 5.4 Hz, 2H), 3.46 (t, *J* = 5.4 Hz, 2H), 3.10 ppm (s, 3H).

^{13}C NMR (151 MHz, d_6-DMSO) δ 165.32, 148.50, 139.32, 138.51, 132.88, 130.85, 130.23, 125.06, 124.97, 118.82, 118.75, 108.98, 96.40, 93.35, 58.01, 57.05, 40.42 ppm.

HRMS (TOF ES^+) Calculated for $C_{19}H_{15}N_3OS_2$ $[M+]^+$: 365.07, found 365.07.

5.2.3 Silver coated chromatography

Thin layer chromatography

A TLC plate was dipped in 1% silver nitrate solution in acetonitrile three times, drying in between, and kept in the dark. The compound was loaded and eluted in 5% acetonitrile in dichloromethane.

Column chromatography

Silica was added to a round bottom flask that contained 5% silver nitrate solution in acetonitrile and evaporated under vacuum. The silica was azeotroped with acetone four times. The column was packed, the isomer mixture was loaded in DCM and the column was run.

5.2.4 HPLC

HBC_{620} isomer mixture was dissolved in water/acetonitrile (1:1) and 10 mL were injected. A gradient run was performed (60-100% acetonitrile in water in 15 min with a 10 mL/min flowrate). The fractions were collected and transferred to a round-bottom flask. Solvents were removed before analysing with NMR.

5.2.5 Isomerisation studies of HBC_{620}

Samples were prepared in the dark by dissolving HBC_{620} (1 mg/mL) in d_6-DMSO. ^{1}H-NMR spectra were recorded in the dark and following illumination of the sample with red (623 nm) and blue (450 nm) LED light. Bruker DRX 600 MHz and Bruker Ascend 400 MHz NMR spectrometers were used for this experiment.

Thermal reversion in the dark was followed by ^{1}H-NMR in the absence or presence of acetic acid (6 µL). A spectrum was recorded every 2 h until full thermal conversion to the Z isomer (**9**) was achieved using a Bruker Ascend 400 MHz NMR spectrometer. Data was fitted to a first order decay using Prism10 software.

5.3 Biological methods

5.3.1 Polymerase Chain Reaction (PCR)

PrimeSTAR premix (25 µL), plasmid (1 µL), 10 µM solution of forward (3-5 µL) and reverse primers (3-5 µL) and DMSO (0-3 µL) were added to a PCR tube. The solution was topped up to 50 µL with deionised water. The following program was used on the thermocycler:

Step	*Temperature (°C)*	*Time (s)*	*Cycles*
Initial denaturation	98	120-180	1
Denaturation	98	10-30	35
Annealing	50-65	5-15	
Elongation	72	70s per kbp of DNA	
Final extension	72	600	1

Table 12. Thermocycler protocol followed for PCR reactions.

The annealing temperature depends on the melting temperature of the primers (around 5°C below the Tm). The elongation temperature depends on the DNA polymerase used, in this case PrimeSTAR HS DNA Polymerase. The elongation time depends on the size of the desired PCR product.

When primers had a melting temperature higher than 72°C, two-step PCR was carried out. In this method, the annealing and elongation steps are combined.

Step	*Temperature (°C)*	*Time (s)*	*Cycles*
Initial denaturation	98	180	1
Denaturation	98	10	30
Annealing + Elongation	68	60s per kbp of DNA	
Final extension	72	600	1

Table 13. Thermocycler protocol followed for two-step PCR reactions.

The result of the PCR reaction was digested with DpnI (0.5 µL) for 2 h at 37°C to eliminate the parental plasmid from the mixture. No further purification was performed for SDM applications, and 5 µL of the reaction mixture was used to transform XL1-Blue cells.

5.3.2 Agarose gel electrophoresis

Tris-acetate EDTA (TAE) buffer (50x)

Tris-Base (242 g, 2 M), glacial acetic acid (57.1 mL, 1 M) and disodium EDTA (100 mL, 5 mM) were dissolved in deionised water (80% of the appropriate volume). The pH was adjusted to 8 and the solution was topped up to the final volume (1 L) with deionised water.

TAE (1x) buffer was prepared by diluting 50xTAE buffer (20 mL) in deionised water (980 mL).

Agarose gel 1%

Agarose powder (1 g, 1%) was dissolved in 1xTAE buffer (100 mL) while heating. SYBR safe dye (5 µL) was added once the solution was at ~50°C. The solution was poured into the casting tray (Bio-Rad Sub Cell GT Mini) containing a well comb and was allowed to solidify at room temperature (30 min).

Protocol

The solidified gel was placed in the electrophoresis unit (Bio-Rad) and the later was filled with 1xTAE buffer until the gel was covered. DNA samples (5 µL) were mixed with FastDigest Green Buffer (1 µL) and carefully loaded into the gel wells. DNA ladder (5 µL) was loaded in one of the wells to identify the size of DNA samples. The gel was run at a constant amplitude of 85 mA for 60 minutes. The gel was visualised using a BioRad ChemiDoc MP imaging system.

5.3.3 Gel extraction of DNA

The QIAquick Gel Extraction using a Microcentrifuge protocol was followed. A DNA fragment was cut from the gel using a scalpel and transferred into a weighted Eppendorf tube. QG buffer (300 µL/100 mg gel) was added, and the samples were incubated at 50°C for 10 min or until the gel was dissolved. Isopropanol (100 µL/100 mg gel) was added, and the sample was inverted five times. The sample was transferred to a spin column, centrifuged (1 min, 13000 rpm) and the flowthrough was discarded. PE buffer (750 µL) was added, left to stand for 5 min, centrifuged (1 min, 13000 rpm) and the flowthrough discarded. This step was repeated. After discarding the flowthrough, the column was centrifuged (1 min, 13000 rpm) to completely remove residual ethanol. The column was placed on an Eppendorf tube and deionised water (30 µL) was added carefully to the centre of the membrane. The column was left for 1 min and centrifuged (1 min, 13000 rpm). The Eppendorf tube containing the DNA was stored at -20°C.

5.3.4 Golden Gate Assembly

BSA (1.5 µL), T4 buffer (1.5 µL), DTT 0.1M (0.1 µL), backbone DNA (75 ng) and insert DNA (75 or 150 ng) were added to a PCR tube. The solution was topped up to 14 µL with deionised water. Lastly, the enzymes BsaI-HFv2 (0.5 µL) and T4 ligase (0.5 µL) were added, giving a total volume of 15 µL. The reaction mixture was subjected to the following program on the thermocycler:

Step	***Temperature (°C)***	***Time (min)***	***Cycles***
Incubation	37	60	1
Activation of BsaI	37	5	26
Activation of ligase	16	5	
Inactivation of BsaI	50	5	1
Inactivation of ligase	80	5	1

Table 14. Thermocycler protocol used for Golden Gate Assembly.

5.3.5 KLD ligation

PCR product (1 µL), KLD reaction buffer (5 µL), KLD enzyme mix (1 µL) and deionised water (3 µL) were added to a PCR tube. The reaction mixture was incubated at 25°C for 5 min. The product is suitable for immediate transformation or storage at -20°C.

5.3.6 HiFi DNA assembly

Linear vector, obtained from PCR amplification (0.1 pmol); insert, from reverse complement primer dimer (0.2 pmol), HiFi assembly Master Mix (5 µL) and dH_2O (final volume 10 µL) were mixed in a PCR tube and incubated at 50°C for 45 min. The product is suitable for immediate transformation or storage at -20°C.

5.3.7 Site Directed Mutagenesis (SDM)

PrimeSTAR premix (25 µL), plasmid (0.2 µL), 10 µM solution of forward primer (0.2 µL), 10 µM solution of reverse primer (0.2 µL) and DMSO (2 µL) were added to a PCR tube. The solution was topped up to 50 µL with deionised water. The following protocol was used on the thermocycler:

Step	*Temperature (°C)*	*Time (s)*	*Cycles*
Initial denaturation	98	120-180	1
Denaturation	98	15-30	30
Annealing	50-65	10-20	
Elongation	72	60s per kbp of DNA	
Final extension	72	600	1

Table 15. Thermocycler protocol used for site directed mutagenesis.

The result of the PCR reaction was digested with DpnI (1 µL) in presence of 10x Fast Digest Buffer (5 µL) to eliminate the template plasmid from the mixture. 10 µL of the mixture were used with no further purification to transform in a cloning cell line (DH5α or XL1-Blue).

5.3.8 Plasmids and primers specifications

Gene(s)	**Vector**	**Antibiotic resistance**
Pepper	pRham	Kanamycin
EL222	pEBLindv2	Kanamycin
EL222, sfGFP	pAP05	Tetracycline
His$_6$-EL222	pETM11	Kanamycin
His$_6$-TEV	pET28a	Kanamycin
His$_6$-AuLOV	pETM11	Kanamycin

Table 16. List of plasmids used in this project.

Name	Sequence
Pepper	uugccauguguaugugggguucgcccacauacucugaugaucccccaaucguggcgugucggccucucccaaucguggcgugucggccucucccaaucguggcgugucggccucucccaaucguggcgugucggccucucuucggagaggcacuggcgccggagaggcacuggcgccggagaggcacuggcgccggagaggcacuggcgccgggaucauucauggcaa
EL222 (pAP05, pEBLindv2)	MLDMGQDRPIDGSGAPGADDTRVEVQPPAQWVLDLIEASPIASVVSDPRLADNPLIAINQAFTDLTGYSEEECVGRNCRFLAGSGTEPWLTDKIRQGVREHKPVLVEILNYKKDGTPFRNAVLVAPIYDDDDELLYFLGSQVEVDDDQPNMGMARRERAAEMLKTLSPRQLEVTTLVASGLRNKEVAARLGLSEKTVKMHRGLVMEKLNLKTSADLVRIAVEAGI
sfGFP	MRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYPDHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGITHGMDELYK
His$_6$-EL222	MKHHHHHHPMSDYDIPTTENLYFQGAMGADDTRVEVQPPAQWVLDLIEASPIASVVSDPRLADNPLIAINQAFTDLTGYSEEECVGRNCRFLAGSGTEPWLTDKIRQGVREHKPVLVEILNYKKDGTPFRNAVLVAPIYDDDDELLYFLGSQVEVDDDQPNMGMARRERAAEMLKTLSPRQLEVTTLVASGLRNKEVAARLGLSEKTVKMHRGLVMEKLNLKTSADLVRIAVEAGI
His$_6$ - AuLOV	MKHHHHHHPMSDYDIPTTENLYFQGAMDYSLVKALQTAQQNFVISDPSIPDNPIVYASQGFLTLTGYALSEVLGRNCRFLQGPETDPKAVEKVRKGLERGEDTTVVLLNYRKDGSTFWNQLFIAALRDGEGNVVNYLGVQCKVSEDYAKAFLKNEENEK

Table 17. Gene sequences of Pepper RNA (lowercase) and proteins (uppercase) used in this project. His$_6$-tags are underlined when present.

Name	Sequence
Pepper 4x insert	aagtccggtctcctgttgccatgtgtatgtgggttcgcccacatactctgatgatccccaatcgtggcgtgtcggcctctcccaatcgtggcgtgtcggcctctcccaatcgtggcgtgtcggcctctcccaatcgtggcgtgtcggcctctcttcggagaggcactggcgccggagaggcactggcgccggagaggcactggcgccggagaggcactggcgccgggatcattcatggcaagggagaccggactt

Table 18. Pepper insert used for Golden Gate Assembly. The BsaI sites are underlined.

Oligonucleotide primers were designed in the online software Benchling following these criteria when possible:

- Primers should be 20 to 45 bp-long
- Primers should start and end with G or C and ideally have a GC content of 40 to 60%
- The melting temperature should be lower than 72°C, ideally between 55 and 65°C
- Forward and reverse primer should have similar melting temperatures
- Primers should contain at least 15 base pairs matching the plasmid.
- Primers for SDM should have 15 bp overlap with each other and contain the mutation in the centre of the overlapping area.

- Specificity and secondary structure should be checked for unwanted interactions

Plasmid	FWD primer	REV primer
pRham	cgatgtggtctcccaagccaggcatcaaataaaacga	cgatgtggtctccaacacagttccctactctcgca
pEBLindv2	tgtccggtctcccaagctgaaggtcgtcactccacc	aagtccggtctccaacatatttattcgactataacaaaccat
pAP05	agggaaggtctcccaaggggatcctctagagtcga	agggaaggtctccaacaggtaccatatctccttcttaatga

Table 19. Primers used in Golden Gate Assembly to introduce the Pepper 4x insert in pRham, pEBLindv2 and pAP05. The BsaI recognition sites are underlined.

Name	FWD primer	REV primer
pCsgD	GAGTTAatgttagcgaagaaaatggtttg	ACCCGaccggctctgtgtgaaattg

Table 20. Primers used in KLD cloning to exchange EL222 promoter for CsgD promoter. In capital letters are the bases belonging to the modified promoter.

Name	FWD primer	REV primer
pAP05	ccgtcaagaCgcaccTcTATctggtgatggaaaagctc	GAACGcgcGaTttcGttgttgcgcaagcccgatgccacc
Insert	CgaaAtCgcgCGTTCgctcTTcAtCtcggagaaTaccgtcaagaCgcacc	ggtgcGtcttgacggtAttctccgaGaTgAAgagcGAACGcgcGaTttcG

Table 21. Primers used in HiFi DNA assembly to substitute EL222 DNA binding region for CsgD DNA binding region. Capital letters represent the mutations and underline represents the overlap needed for assembly.

Name	Variant	FWD primer	REV primer
EL222::CsgD$_{R1}$	S189R	cgaaatcgcgcgt**AG**Gctcttcatctcggag	ctccgagatgaagagC**CT**acgcgcgatttcg
EL222::CsgD$_{R2}$	L201R	ccgtcaagacgcacC**G**Ctatctggtgatgg	ccatcaccagataG**C**Ggtgcgtcttgacgg

Table 22. Primers used for site-directed mutagenesis (5' to 3') of EL222::CsgD with the mutations in bold and the corresponding codon in capital letters. The primers for EL222::CsgD$_{R2}$ were used on EL222::CsgD$_{R1}$ to obtain the double mutation variant EL222::CsgD$_{RR}$.

Variant	FWD primer	REV primer
N194D	cagcag**G**ACttcgttatctctgatccgagc	aacgaaGT**C**ctgctgagcggtttgcagggcc
N194H	cagcag**C**ACttcgttatctctgatccgagc	aacgaaGT**G**ctgctgagcggtttgcagggcc
N194K	cagcagAA**G**ttcgttatctctgatccgagc	aacgaa**C**TTctgctgagcggtttgcagggcc
N194L	cagcag**CT**Cttcgttatctctgatccgagc	aacgaaG**AG**ctgctgagcggtttgcagggcc
N194Q	cagcag**C**A**G**ttcgttatctctgatccgagc	aacgaa**C**T**G**ctgctgagcggtttgcagggcc
N194S	cagcagA**G**Cttcgttatctctgatccgagc	aacgaaG**C**Tctgctgagcggtttgcagggcc
N194Y	cagcag**T**ACttcgttatctctgatccgagc	aacgaaGT**A**ctgctgagcggtttgcagggcc

Table 23. Primers used for site-directed mutagenesis (5' to 3') of AuLOV with the mutations in bold and the corresponding codon in capital letters.

5.3.9 Antibiotic stock solution preparation

Antibiotics were prepared as sterile stock solutions and were diluted 1:1000 into media to yield the correct concentrations.

Kanamycin

Kanamycin sulphate (0.5 g) was dissolved in deionised water (10 mL). The solution was filter sterilised through a 0.2 µm syringe filter into Eppendorf tubes and stored at -20°C.

Tetracycline

Tetracycline sulphate (0.1 g) was dissolved in ethanol (3 mL) and deionised water (7 mL). The solution was filter sterilised through a 0.2 µm syringe filter into Eppendorf tubes and stored at -20°C.

Chloramphenicol

Chloramphenicol (0.25 g) was dissolved in ethanol (10 mL). The solution was filter sterilised through a 0.2 µm syringe filter into Eppendorf tubes and stored at -20°C.

Ampicillin

Ampicillin (1 g) was dissolved in deionised water (10 mL). The solution was filter sterilised through a 0.2 µm syringe filter into Eppendorf tubes and stored at -20°C.

5.3.10 Media preparation

LB medium

Sodium chloride (10 g/L), tryptone (10 g/L) and yeast extract (5 g/L) were dissolved in deionised water. The solution was autoclaved (120°C, 20 min) and allowed to cool down to room temperature before use.

M9 medium

Disodium hydrogen phosphate (64 g/L), potassium dihydrogen phosphate (15 g/L), sodium chloride (2.5 g/L) and ammonium chloride (5 g/L) were dissolved in deionised water. The M9 salts solution was autoclaved (120°C, 20 min) and allowed to cool down to room temperature before use.

Filter sterilised solutions of glucose (2 M, 1.1 mL), trace metals (1 mL), magnesium sulphate (1 M, 200 µL) and calcium chloride (1 M, 10 µL), and autoclaved water (78 mL) were added to M9 salts solution (20 mL) immediately before use.

M63 medium

Ammonium sulphate (2 g/L), potassium dihydrogen phosphate (13.6 g/L) and iron (II) sulphate (0.5 mg/L) were dissolved in deionised water. The M63 salts solution was autoclaved (120°C, 20 min) and allowed to cool down to room temperature before use.

Filter sterilised solutions of glycerol (50%, 400 µL) and magnesium sulphate (1 M, 100 µL) were added to M63 salts solution (100 mL) immediately before use.

Autoinduction medium

Sodium chloride (0.5 g/L), tryptone (10 g/L), yeast extract (5 g/L), magnesium sulfate (2 g/L), ammonium chloride (0.5 g/L) and glycerol (5 mL/L) were dissolved in deionised water. The solution was autoclaved (120°C, 20 min) and allowed to cool down to room temperature before use. Disodium hydrogen phosphate (6 g/L), potassium dihydrogen phosphate (3 g/L), glucose (1 g/L) and arabinose (2 g/L) are added in solid form immediately before use.

5.3.11 LB agar plate preparation

Sodium chloride (10 g/L), tryptone (10 g/L), yeast extract (5 g/L) and agar (15 g/L) were dissolved in deionised water. The suspension was autoclaved (120°C, 20 min) and allowed to cool down to ~50°C before use. 1 µL/mL of each antibiotic necessary was added and the solution was poured into sterile Petri dishes. The plates were stored at 4°C.

5.3.12 Bacterial strains

Several *E. coli* strains were used during the course of this project, present in the Allemann group library. XL1-Blue and DH5α cloning strains were used for DNA amplification and cloning. BL21 (AI), BL21 (DE3) and BL21 (DE3) Star were used for protein expression. XL1-Blue, BL21 (DE3) Star and DH5α cell lines were used for fluorescence experiments. BW25113 and JW1023 were used in the biofilm formation assays.

It is important to note that XL1-Blue and BL21 (AI) strains are resistant to tetracycline and JW1023 is resistant to kanamycin.

5.3.13 Competent cell preparation

CM1 buffer

Sodium acetate (10 mM) and sodium chloride (5 mM) were dissolved in deionised water (80% of desired volume) and the pH adjusted to 5.6 using acetic acid. Manganese chloride (II) (20 mM) was added, and the solution made up to the desired volume with deionised water. The solution was filter sterilised through a 0.2 µm syringe filter before use.

CM2 buffer

Sodium acetate (10 mM), calcium chloride (20 mM), manganese chloride (II) (5 mM) and glycerol (5% v/v) were dissolved in deionised water. The solution was filter sterilised through a 0.2 µm syringe filter before use.

Protocol

A colony from an agar plate of the cell strain was inoculated into a Sterilin tube containing LB (10 mL). If needed, 10 µL of antibiotic stock were added. It was incubated overnight at 37°C, 275 rpm. The overnight culture (2 mL) was used to inoculate LB media (50 mL) containing 50 µL of the required antibiotic(s). The resulting culture was incubated for two hours at 37°C, 275 rpm. The cells were placed on ice for 10 min, transferred to a Falcon tube and centrifuged at 4000 rpm for 10 min at 4°C. The supernatant was discarded, and the cell pellet was resuspended in CM1 buffer (20 mL) added through a syringe filter. Cells were left on ice for 20 min and centrifuged at 4000 rpm for 5 min at 4°C. The cell pellet was resuspended in CM2 buffer (1.5 mL) added though a syringe filter. The solution was distributed in 50 µL aliquots into Eppendorf tubes and flash-frozen by submerging in liquid nitrogen. These aliquots were stored at -80°C.

5.3.14 Super-competent cell preparation

Tfb1 buffer

Potassium acetate (30 mM), rubidium chloride (100 mM), calcium chloride (10 mM) and glycerol (15% v/v) were dissolved in deionised water (80% of desired volume) and the pH adjusted to 5.8 using dilute acetic acid. Manganese chloride (II) (50 mM) was added and the solution was made up to the final volume with deionised water. The solution was filter sterilised through a 0.2 µm syringe filter before use.

Tfb2 buffer

3-(N-morpholino)propanesulfonic acid (MOPS, 10 mM), calcium chloride (75 mM), rubidium chloride (10 mM) and glycerol (15% v/v) were dissolved in deionised water (80% of desired volume) and the pH adjusted to 6.5 using dilute sodium hydroxide. The solution was made up to the final volume with deionised water. The solution was filter sterilised through a 0.2 µm syringe filter before use.

Protocol

A colony from an agar plate of the cell strain was inoculated into a Sterilin tube containing LB (10 mL). If needed, 10 µL of antibiotic stock were added. It was incubated overnight at 37°C, 275 rpm. The overnight culture (0.5 mL) was used to inoculate LB media (50 mL) containing 50 µL of the required antibiotic(s). The resulting culture was incubated at 37°C, 275 rpm to an OD_{600} of 0.6 – 0.7. The cells were placed on ice for 15 min, transferred to a Falcon tube and centrifuged at 5000 rpm for 5 min at 4°C. The supernatant was discarded, and the cell pellet was resuspended in Tfb1 buffer (20 mL) added through a syringe filter. Cells were left on ice for 15 min and centrifuged at 5000 rpm for 5 min at 4°C. The cell pellet was resuspended in Tfb2 buffer (2 mL) added though a syringe filter. The solution was left on ice for 15 min and distributed in 50 µL aliquots into Eppendorf tubes and flash-frozen by submerging in liquid nitrogen. These aliquots were stored at -80°C.

5.3.15 Transformation of (super)competent cells

To transform *E. coli*, 2 µL (10 µL for KLD, Golden Gate and SDM products) of the appropriate plasmid were added to a microcentrifuge tube of competent cells under sterile conditions. The cells were left on ice for 1 hour. The cells were then placed in a water bath at 42°C for 45 seconds and placed back on ice for 5 minutes. Under sterile conditions, LB medium (1 mL) was added, and the microcentrifuge tube was placed in a shaker at 37°C for 1.5 h to allow the cells to grow. It was centrifuged for 1 min at 6000 rpm and, immediately after, the majority of the supernatant was poured off under sterile conditions. The cells were resuspended in the remaining liquid and the solution was spread onto an agar plate with appropriate antibiotics. The agar plates were incubated at 37°C overnight and then stored at 4°C.

A negative control of an aliquot of competent cells which contained no plasmid was subjected to the same protocol and spread on agar plates with and without the appropriate antibiotic to ensure no microbial contamination.

A positive control of a known plasmid was added to an aliquot of competent cells and subjected to the same protocol and plated on the appropriate antibiotic agar plate to ensure cell competency.

5.3.16 Glycerol stock preparation

A colony was picked from the agar plate and added to LB medium (10 mL) that contained 10 µL of required antibiotic stock. This was incubated overnight at 37°C, 275 rpm. 0.5 mL of the overnight culture was added to a sterile Eppendorf tube along with 0.5 mL of 50% v/v glycerol solution. The Eppendorf tubes were stored at -80°C.

5.3.17 Bacterial plasmid DNA isolation

DNA was purified using the TransGen Biotech 'EasyPure® Plasmid MiniPrep Kit' following the protocol provided by the manufacturers. LB media (10 mL) containing appropriate antibiotic (10 µL) was inoculated with a single colony from the agar plate (or an aliquot of glycerol stock) and incubated overnight at 37°C, 275 rpm. This starter colony was centrifuged at 4000 rpm for 10 min at 4°C and the supernatant was discarded. Resuspension buffer (250 µL) was added to the cell pellet and the suspension was transferred to an Eppendorf tube. Lysis buffer (250 µL) was added, and the sample was mixed by inversion. Neutralisation buffer (350 µL) was added, and the sample was inverted 5 times and left for 5 min. The sample was centrifuged (10 min, 12000 rpm) and the supernatant was transferred to a spin column. The column was centrifuged (1 min, 12000 rpm) and the flowthrough discarded. Wash buffer (800 µL) was added to the column, left to stand for 5 min, centrifuged (1 min, 12000 rpm) and the flowthrough discarded. This step was repeated. After discarding the flowthrough, the column was centrifuged (1 min, 12000 rpm) to completely remove residual wash buffer. The column was placed on an Eppendorf tube and deionised water (50 µL) was added carefully to the centre of the membrane. The column was left for 1 min and centrifuged (1 min, 12000 rpm). The Eppendorf tube containing the plasmid was stored at -20°C.

5.3.18 Protein expression

BL21 (DE3) *E. coli* cells were used to express TEV protease. A starter culture (5 mL) was used to inoculate LB media (500 mL) supplemented with kanamycin (500 µL) and was grown at 37°C, 275 rpm. The cell cultures were induced with arabinose when they reached an OD_{600} of 0.6 and incubated overnight at 20°C. The culture was then centrifuged at 4000 rpm for 15 min and, if needed, the cell pellets were stored at -20°C before purification.

BL21 (AI) *E. coli* cells were used to express AuLOV and EL222 variants. 5 mL of a starter culture were added to 500 mL of autoinducing media supplemented with the appropriate antibiotics

(500 µL) at 20°C. The cells were incubated at 20°C for 24 h. The cells were harvested and, if needed, stored at -20°C before purification.

5.3.19 Protein purification

Protein	***Lysis buffer***	***Elution buffer***	***Assay buffer***
EL222	50 mM Tris-HCl 100 mM NaCl pH 8	50 mM Tris-HCl 100 mM NaCl 500 mM imidazole pH 8	
AuLOV	20 mM HEPES 20 mM NaCl 0.3 mM TCEP pH 7.8	20 mM HEPES 20 mM NaCl 0.3 mM TCEP 500 mM imidazole pH 7.8	20 mM MES 0.3mM TCEP 5mM NaN_3 1mM EDTA pH 6
TEV protease	50 mM TRIS base 500 mM NaCl 5% glycerol 0.3 mM TCEP pH 8	50 mM TRIS base 500 mM NaCl 5% glycerol 0.3 mM TCEP 500 mM imidazole pH 8	

Table 24. Buffers used in protein purification and analysis.

A protease inhibitor (PMSF), lysozyme and ~20 ml of lysis buffer were added to the cell pellet and the cells were suspended. They were sonicated at 40% power for five minutes (5 seconds pulse on and 15 seconds off, with a total time of 20 min) using a sonicator Sonic Vibra-cell. The lysate was centrifuged for 45 minutes at 16000 rpm to remove the insoluble components of the cell. The supernatant, containing the protein, was purified using a Ni^{2+}-NTA column. The column, kept in 20% ethanol solution, was washed with deionised water and lysis buffer before loading the protein solution. Buffer with increasing concentrations of imidazole was used to wash (50 mL 20 mM, 20 mL 40 mM) and elute (10 mL 250 mM, 10 mL 500 mM) the protein. The fractions were assessed by SDS-PAGE and, if needed, TEV-His_6 protease was added to the fractions that contained the >95% pure protein to remove the histidine tag. The solution was dialyzed overnight to remove the imidazole from the solution using a 7000Da cut-off dialysis membrane.

When TEV-His6 protease was used, a second Ni^{2+}-NTA column was performed on the dialyzed solution to separate the protein from the histidine tag and the protease. It was eluted using a step gradient buffer with 0, 20 and 500 mM imidazole. SDS-PAGE was used to verify the presence of protein in the flow-through. Once confirmed, the protein was concentrated using Vivaspin centrifugal concentrators with a 10 kDa molecular weight cut off membrane. Vivaspins were used in a centrifuge forcing the buffer across the membrane, decreasing the volume of the protein solution and concentrating it. The Vivaspin was also used for buffer exchange into the assay buffer for AuLOV variants.

5.3.20 SDS-PAGE

Resolving buffer
1.5M TRIS-HCl pH 8.8

Stacking buffer
0.5 M TRIS-HCl pH 6.8

Resolving gel (12% bisacrylamide)
- 1.7 mL H_2O
- 1.25 mL resolving buffer
- 50 µL 10% Sodium dodecyl sulfate
- 2 mL 30% bisacrylamide (w/v)
- 50 µL 10% (w/v) ammonium persulfate
- 10 µL TEMED

Stacking gel (6% bisacrylamide)
- 1.3 mL H_2O
- 0.65 mL resolving buffer
- 25 µL 10% Sodium dodecyl sulfate
- 0.5 mL 30% bisacrylamide (w/v)
- 50 µL 10% (w/v) ammonium persulfate
- 10 µL TEMED

Loading dye
- 4% SDS
- 20% glycerol
- 10% β-mercaptoethanol
- 0.004% bromophenol blue
- 0.125 M TRIS-HCl pH 6.8

Cathode buffer (middle tank)
- 0.1 M TRIS
- 0.1 M Tricine
- 0.1% SDS

Anode buffer (outer tank)
- 0.2 M TRIS, pH 8

Protocol

The resolving gel solution was made in a universal tube and poured between the two glass plates on a casting frame, leaving some space at the top. A small volume of propan-2-ol was added to remove bubbles and to level the resolving gel. After the gel solidified, the propan-2-ol was poured off. The stacking gel was then prepared and poured on top of the resolving gel. The gel comb was placed immediately after.

The samples were prepared by mixing protein sample and loading dye in 1:1 ratio and heating to 80°C for five minutes. The gel was placed in the electrophoresis tank and filled with cathode buffer (middle tank) and anode buffer (outer tank). The comb was removed and 5-10 µL samples were loaded with a micropipette in each well, including a protein ladder to indicate the molecular weights. A voltage was applied to run the gel (60 V for 10 min, 180 V for 50 min). Once run, the gel was boiled in water to destain the loading dye and then placed in Coomassie brilliant blue R250 stain. The stain was poured off and water was added to destain and make the protein bands visible. Pictures of the gels were taken with a Bio-Rad ChemiDoc MP Imaging System.

5.3.21 UV-Vis spectroscopy

The concentration of the protein solution and the transition between light and dark states were measured by UV-vis spectroscopy, using a Shimadzu UV-2600 spectrophotometer with a CPS-100 temperature-controlled cell positioner set at 20°C and the software UVProbe.

AuLOV variants

Assay buffer (1 mL) was used to record the baseline (600 nm to 200 nm). 20 µL of protein solution in assay buffer were added before illuminating the cuvette with blue light (450 nm) for one minute and the spectra were recorded. The absorption at 388 nm (light state maximum, $\varepsilon = 9135\ M^{-1}\ cm^{-1}$) was used to calculate the concentration.

To measure the relaxation of the protein from the light to the dark state, the spectra were recorded from 600 to 200 nm every 5, 10 or 15 minutes one hundred times. The absorption at 450 nm (dark state maximum) over time was used to determine the relaxation kinetics. Data was fitted using Prism10 software.

EL222

Due to the short relaxation time of this protein, it was not possible to measure it in the light state. The protein sample was kept in the dark overnight before measuring and the room was kept in dim red light conditions. The absorption at 450 nm ($\varepsilon=13600\ M^{-1}\ cm^{-1}$) was used to calculate the concentration.

5.3.22 Size exclusion chromatography

The analysis of AuLOV by size exclusion chromatography (SEC) was carried out using a Cytiva ÄKTA pure chromatography system equipped with a Superdex^{TM} 75 10/300 GL column and UNICORN^{TM} 7 software. 100 µL (50 µM) of protein solution were injected each time and analysed by a two-step program (sample application and elution at 0.8 mL/min for 1.5 column volumes). Degassed AuLOV assay buffer was used for elution. The detection took place at three different wavelengths: 280 nm due to the aromatic amino acids in proteins, 390 nm for the light state maximum of the cysteine-FMN adduct, and 450 nm is the maximum of the oxidised-FMN in the dark state. The fractions were collected in a 96 deep well plate.

When analysing the protein in the light state, the sample was exposed to blue light for a minute immediately before injecting and the column was continuously illuminated with blue light. For the analysis of the dark state, the sample was left to relax overnight covered in aluminium foil. The room was kept in dim red-light conditions and the column was covered in aluminium foil. Four repeats were done for each condition.

Analysis of the protein standard mixture took place in the buffer recommended by the manufacturer (100 mM sodium phosphate, 100 mM sodium sulphate, 0.05 % sodium azide, pH 6.7) due to the incompatibility with AuLOV assay buffer.

5.3.23 Protein crystallisation

Dark state crystals of AuLOV were set up under dim red-light conditions and grown in the dark at 20°C by the sitting-drop vapor diffusion method. 80 µL of precipitant (0.1 M sodium acetate/citrate pH 4.0 to 5.5, 0.1 M ammonium chloride, 15-20% polyethylene glycol 2000) were placed in each well (Figure 109 left). 1 µL of protein (3 – 5 mg/mL) was mixed with 1 – 2 µL of precipitant in the drop. In some cases, crushed crystals obtained from $N194K_{F\text{-}tyr}$ (0.5 µL) were used as nucleation points.

Crystallisation trials for AuLOV light state crystals were performed using an Oryx 4 robot (Douglas Instruments) and WaspRun software. Protein solution was illuminated with 450 nm light prior to and during crystallisation by sitting-drop vapor diffusion. 80 µL of precipitant (1.5 to 3.0 M disodium malonate pH 7, 0.1 M Tris.OAc pH 6.5 to 8.0) were placed in each well (Figure 109 right). 0.5 µL of protein in crystallography buffer (5 – 10 mg/mL) was mixed with 0 .5 to 1.0 µL precipitant in each drop. 0.1 µL of 100 mM $Co^{3+}(NH_3)_6$ was used as an additive.

DARK		PEG 2000					
		15%	16%	17%	18%	19%	20%
pH	4	Sodium acetate 0.1M Ammonium chloride 0.1M					
	4.5						
	5						
	5.5						
pH	4	Sodium citrate 0.1M Ammonium chloride 0.1M					
	4.5						
	5						
	5.5						

LIGHT		pH					
		6.5	7	7.25	7.5	7.75	8
Sodium malonate [M]	1.5	Tris acetate 0.1M Additive: 0.1µL $Co^{3+}(NH_3)_6$ 100mM					
	2						
	2.5						
	3						
	0.15	Tris acetate 0.1M Additive: 0.1µL $Co^{3+}(NH_3)_6$ 100mM					
	0.2						
	0.25						
	0.3						

Figure 109. Crystallisation conditions for AuLOV variants in dark (left) and light (right) states.

Crystal harvesting was carried out with the help of Dr. Pierre Rizkallah, senior lecturer in Structural Biology. The well containing crystals was opened using a scalpel and 1 µL of precipitant was transferred from the well to the drop to avoid drying. Crystals were extracted with 0.1 – 0.3 mm mounted LithoLoops, flash frozen using liquid nitrogen and placed on a frozen puck. Once a puck was filled, it was placed into a shipping cane in a cryogenic Dewar. Harvested crystals were sent to Diamond Light Source (Oxfordshire, UK) for data collection. Dark state crystals were harvested under dim red light. Illuminated state crystals were obtained by illuminating dark state crystals with blue light for one minute before harvesting. Lit-state crystals were harvested while being illuminated with blue light.

Diffraction datasets were collected at Diamond Light Source (Oxfordshire, UK). Crystals were kept at 100 K during data collection in a cold nitrogen stream. Data reduction was completed with Xia2 using XDS, or DIALS or Autoproc at Diamond Light Source. Crystal structures were solved using CCP4i2 and Coot software. Molecular replacement was carried out using Aureochrome 1a LOV domain from *Ochromonas danica* in the dark, illuminated or light state as a template (PDB entries 6i20, 6i21 and 6i22 respectively). Structural refinement was achieved using alternate cycles of Coot and REFMAC5. The unnatural amino acid, FMN, solvent molecules or ions were added where positive electron density peaks appeared.

Crystallisation trials for EL222::CsgD_{R1} were performed using the Oryx 4 robot under dim light conditions and grown in the dark at 20°C by the sitting-drop vapor diffusion method. PACT Premier and JCSG Plus commercial screens were used. 80 µL of precipitant were placed in each well. 0.75 µL of protein in crystallography buffer (6.5 mg/mL) was mixed with 0.5 or 0.75 µL precipitant in each drop. Upon plate preparation, the protein was kept in dark conditions.

Seeding protocol (adapted from[246])

Crystals too small for harvesting were used to make seed stocks. The well was opened with a scalpel and reservoir solution was transferred to an Eppendorf tube (50 µL) and to the drop (5 µL). The crystals were crushed using a crystal crusher. The crystal crusher was washed in the Eppendorf with precipitant and the solution with the crushed crystals was transferred to the Eppendorf tube. The seed stock was kept on ice and used immediately or stored at -80°C to avoid redissolution of the microcrystals.

Microcrystals in VMXm beamline

The sample preparation and data collection took place in Diamond Light Source Versatile Macromolecular Crystallography microfocus (VMXm) beamline facilities.[247] CryoTEM grids were glow discharged to ensure the sample distributes evenly throughout the grid and mounted in the plunge freezer. 2 µL of reservoir solution were added to the back of the grid and 2 µL of the crystal solution were added to the front of the grid. The sample was blotted, plunge-freezed in liquid ethane and transferred to liquid nitrogen. For the diffraction experiments, the grids were loaded onto the sample holder and transferred to the VMXm endstation. The beam was focused and aligned on each crystal for measurements.

5.3.24 In-vivo Pepper aptamer experiments

Imaging buffer

HEPES (40 mM), potassium chloride (100 mM) and magnesium chloride (5 mM) were dissolved in deionised water (80% of desired volume) and the pH adjusted to 7.4. The solution made up to the desired volume with deionised water. The solution was stored at 4°C.

HBC imaging buffer

HBC_{620} (1 mg) was dissolved in the minimum quantity of DMSO, and imaging buffer (100 mL) was added. The solution was aliquoted and stored at -80°C.

Protocol

A colony/glycerol stock aliquot was grown in LB medium (10 mL) that contained 10 µL of the required antibiotic(s). It was incubated overnight at 37°C, 275 rpm. The overnight culture (5 mL) was used to inoculate LB media (50 mL) supplemented with the appropriate antibiotic(s) (50 µL) in a conical flask or a 75 cm^3 cell culture flask for light dependent experiments. The cells were incubated at 37°C, 275 rpm until all flasks had reached OD_{600} of 1.5. When necessary, the culture was induced using rhamnose (0.2% w/v) or light (constant blue light illumination during growth) when it reached OD_{600} of 0.6.

The cultures were diluted with deionised water to an OD_{600} of 1.5 and the contents of each flask were transferred to Falcon tubes (25 mL on each). The Falcon tubes were centrifuged for 5 min at 4000 rpm, the supernatant was discarded, and the pellet was resuspended in HBC imaging buffer (1 mL).

The solution was incubated for 60 min at room temperature and loaded on a black 96 well plate (100 µL/well). The fluorescence was measured using a VICTOR X5 Multilabel Plate Reader (PerkinElmer) with a 531/25 excitation filter and a 610 emission filter.

Experiments were conducted in triplicate and data was presented in graph form as arithmetic mean ± standard deviation.

5.3.25 In-vivo green fluorescence protein production assay

Cell culture flasks

A colony/glycerol stock aliquot was grown in LB medium (10 mL) that contained 10 µL of the required antibiotic(s). It was incubated in dark conditions overnight at 37°C, 275 rpm. The overnight culture (5 mL) was used to inoculate LB media (50 mL) supplemented with the appropriate antibiotic(s) (50 µL) in a conical flask or a 75 cm^3 cell culture flask for light dependent experiments. When needed, cultures were grown in dark (covered in aluminium foil) or light (constant blue light illumination) conditions. The cells were incubated at 37°C, 275 rpm until all flasks had reached OD_{600} of 1.5.

The cultures were diluted with deionised water to an OD_{600} of 1.5 and the contents of each flask were transferred to Falcon tubes (25 mL on each). The Falcon tubes were centrifuged for 5 min at 4000 rpm, the supernatant was discarded, and the pellet was resuspended in imaging buffer (1 mL). The solution was incubated for 60 min at room temperature and loaded on a black 96 well plate (100 µL/well). The fluorescence was measured using a VICTOR X5 Multilabel Plate Reader (PerkinElmer) with a 485 excitation filter and a 535 emission filter.

Experiments were conducted in triplicate and data was presented in graph form as arithmetic mean ± standard deviation.

96 well plate

A colony/glycerol stock aliquot was grown in LB medium (10 mL) that contained 10 µL of the required antibiotic(s). It was incubated overnight in dark conditions at 37°C, 275 rpm. The starter culture was diluted 1:100 into a black flat-bottomed 96 well plate and incubated for 24h at 37°C in dark (covered in aluminium foil) or blue light (constant illumination during growth) conditions. The fluorescence was measured using a VICTOR X5 Multilabel Plate Reader (PerkinElmer) with a 485 nm excitation filter and a 535 nm emission filter.

5.3.26 Crystal violet assay for biofilm formation (adapted from [223])

This microtiter dish assay measures the extent of biofilm formation using the dye crystal violet. The positively charged crystal violet dye interacts with negatively charged molecules found in the extracellular matrix of biofilms.

A starter culture grown in dark was diluted 1:100 into a U-bottomed 96 well plate and incubated for 24h at 37°C in dark (covered in aluminium foil) or blue light (constant illumination during growth) conditions. The multiwell plates were sealed with Breathe Easier sealing membranes during incubation. After incubation, the plate was emptied by shaking out the liquid. The plate was gently washed with deionised water by submerging in a small tub of water and shaking out the water. This step was carried out twice. Crystal violet solution (125 µL, 0.1%) was added into each well and incubated at room temperature for 15 min. The plate was washed three times and dried at room temperature for a few hours. Acetic acid solution (125 µL, 30%) was added to each well and incubated at room temperature for 15 min. 100 µL of the solubilised crystal violet were added to a new flat bottomed 96 well plate and the absorbance (570 nm) was quantified using a FLUOstar Omega Microplate Reader (BMG Labtech). Positive and negative controls were included on each plate for comparison. Five replicate wells were done per control and sample. Data was presented in graph form as arithmetic mean ± standard deviation.

References

1 V. M. Cardoso, G. Campani, M. P. Santos, G. G. Silva, M. C. Pires, V. M. Gonçalves, R. de C. Giordano, C. R. Sargo, A. C. L. Horta and T. C. Zangirolami, Cost analysis based on bioreactor cultivation conditions: Production of a soluble recombinant protein using *Escherichia coli* BL21(DE3), *Biotechnol. Rep.*, 2020, **26**, e00441.

2 R. D. G. Ferreira, A. R. Azzoni and S. Freitas, Techno-economic analysis of the industrial production of a low-cost enzyme using *E. coli*: The case of recombinant β-glucosidase, *Biotechnol. Biofuels*, 2018, **11**, 81.

3 R. S. Donovan, C. W. Robinson and B. R. Glick, Optimizing inducer and culture conditions for expression of foreign proteins under the control of the lac promoter, *J. Ind. Microbiol.*, 1996, **16**, 145–154.

4 R. G. Simas, A. Pessoa Junior and P. F. Long, Mechanistic aspects of IPTG (isopropylthio-β-galactoside) transport across the cytoplasmic membrane of *Escherichia coli*—a rate limiting step in the induction of recombinant protein expression, *J. Ind. Microbiol. Biotechnol.*, 2023, **50**, kuad034.

5 E. Romano, A. Baumschlager, E. B. Akmeriç, N. Palanisamy, M. Houmani, G. Schmidt, M. A. Öztürk, L. Ernst, M. Khammash and B. Di Ventura, Engineering AraC to make it responsive to light instead of arabinose, *Nat. Chem. Biol.*, 2021, **17**, 817–827.

6 L. Benedetti, A. E. S. Barentine, M. Messa, H. Wheeler, J. Bewersdorf and P. De Camilli, Light-activated protein interaction with high spatial subcellular confinement, *PNAS U.S.A.*, 2018, **115**, E2238–E2245.

7 P. Gorostiza and E. Y. Isacoff, Optical switches for remote and noninvasive control of cell signaling, *Science (1979)*, 2008, **322**, 395–399.

8 H. Bouas-Laurent and H. Dürr, Organic photochromism (IUPAC Technical Report), *Pure Appl. Chem.*, 2001, **73**, 639–665.

9 W. Szymański, J. M. Beierle, H. A. V Kistemaker, W. A. Velema and B. L. Feringa, Reversible photocontrol of biological systems by the incorporation of molecular photoswitches, *Chem. Rev.*, 2013, **113**, 6114–6178.

10 T. Sakata, Y. Yan, G. Marriott and J. A. Spudich, Optical switching of dipolar interactions on proteins, *PNAS*, 2005, **102**, 4759–4764.

11 N. J. V Lindgren, M. Varedian and A. Gogoll, Photochemical regulation of an artificial hydrolase by a backbone incorporated tertiary structure switch, *Chem. - Eur. J.*, 2009, **15**, 501–505.

12 M. Irie, Photochromism of diarylethene molecules and crystals, *Proc. Jpn. Acad., Ser. B*, 2010, **86**, 472–483.

13 R. J. Mart and R. K. Allemann, Azobenzene photocontrol of peptides and proteins, *Chem. Commun.*, 2016, **52**, 12262–12277.

14 M. A. Van Der Horst and K. J. Hellingwerf, Photoreceptor proteins, 'star actors of modern times': A review of the functional dynamics in the structure of representative members of six different photoreceptor families, *Acc. Chem. Res.*, 2004, **37**, 13–20.

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

16 O. P. Ernst, D. T. Lodowski, M. Elstner, P. Hegemann, L. S. Brown and H. Kandori, Microbial and animal rhodopsins: structures, functions, and molecular mechanisms, *Chem. Rev.*, 2014, **114**, 126–163.

17 R. Kort, W. D. Hoff, M. Van West, A. R. Kroon, S. M. Hoffer, K. H. Vlieg, W. Crielaand, J. J. Van Beeumen and K. J. Hellingwerf, The xanthopsins: a new family of eubacterial blue-light photoreceptors, *EMBO J.*, 1996, **15**, 3209–3218.

18 B. D. Zoltowski and K. H. Gardner, Tripping the light fantastic: blue-light photoreceptors as examples of environmentally modulated protein-protein interactions, *Biochemistry*, 2011, **50**, 4.

19 C. Lin and T. Todo, The cryptochromes, *Genome Biol.*, 2005, **6**, 220.

20 G. I. Jenkins, The UV-B photoreceptor UVR8: From structure to physiology, *Plant Cell*, 2014, **26**, 21–37.

21 K. S. Conrad, C. C. Manahan and B. R. Crane, Photochemistry of flavoprotein light sensors, *Nat. Chem. Biol.*, 2014, **10**, 801–809.

22 J. M. Christie, M. Salomon, K. Nozue, M. Wada and W. R. Briggs, LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): Binding sites for the chromophore flavin mononucleotide, *PNAS U. S. A.*, 1999, **96**, 8779–8783.

23 B. L. Taylor and I. B. Zhulin, PAS Domains: internal sensors of oxygen, redox potential, and light, *Microbiol. Mol. Biol. Rev.*, 1999, **63**, 479–506.

24 D. Strickland, X. Yao, G. Gawlak, M. K. Rosen, K. H. Gardner and T. R. Sosnick, Rationally improving LOV domain-based photoswitches, *Nat. Methods*, 2010, **7**, 623–626.

25 E. Kaiserli, S. Sullivan, M. A. Jones, K. A. Feeney and J. M. Christie, Domain swapping to assess the mechanistic basis of Arabidopsis phototropin 1 receptor kinase activation and endocytosis by blue light, *Plant Cell*, 2009, **21**, 3226–3244.

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

27 M. S. Kritsky, T. A. Belozerskaya, V. Y. Sokolovsky and S. Y. Filippovich, Photoreceptor apparatus of the fungus *Neurospora crassa*, *Mol. Biol.*, 2005, **39**, 514–528.

28 U. Krauss, B. Q. Minh, A. Losi, W. Gärtner, T. Eggert, A. von Haeseler and K.-E. Jaeger, Distribution and phylogeny of Light-Oxygen-Voltage-blue-light-signaling proteins in the three kingdoms of life, *J. Bacteriol.*, 2009, **191**, 7234–7242.

29 S. Endres, J. Granzin, F. Circolone, A. Stadler, U. Krauss, T. Drepper, V. Svensson, E. Knieps-Grünhagen, A. Wirtz, A. Cousin, P. Tielen, D. Willbold, K.-E. Jaeger and R. Batra-Safferling, Structure and function of a short LOV protein from the marine phototrophic bacterium *Dinoroseobacter shibae*, *BMC Microbiol.*, 2015, **15**, 30.

30 W. R. Briggs and E. Huala, Blue-light photoreceptors in higher plants, *Annu. Rev. Cell Dev. Biol.*, 1999, **15**, 33–62.

31 S. Akbar, T. A. Gaidenko, C. Min Kang, M. O'Reilly, K. M. Devine and C. W. Price, New family of regulators in the environmental signaling pathway which activates the general stress transcription factor σB of *Bacillus subtilis*, *J. Bacteriol.*, 2001, **183**, 1329–1338.

32 Z. Cao, E. Livoti, A. Losi and W. Gärtner, A blue light-inducible phosphodiesterase activity in the cyanobacterium *Synechococcus elongatus*, *Photochem. Photobiol.*, 2010, **86**, 606–611.

33 S. Crosson and K. Moffat, Structure of a flavin-binding plant photoreceptor domain: Insights into light-mediated signal transduction, *PNAS U. S. A.*, 2001, **98**, 2995–3000.

34 J. Herrou and S. Crosson, Function, structure and mechanism of bacterial photosensory LOV proteins, *Nat. Rev. Microbiol.*, 2011, **9**, 713–723.

35 A. Möglich, R. A. Ayers and K. Moffat, Design and signaling mechanism of light-regulated histidine kinases, *J. Mol. Biol.*, 2009, **385**, 1433–1444.

36 S. M. Harper, L. C. Neil and K. H. Gardner, Structural basis of a phototropin light switch, *Science (1979)*, 2003, **301**, 1541–1544.

37 R. Fedorov, I. Schlichting, E. Hartmann, T. Domratcheva, M. Fuhrmann and P. Hegemann, Crystal structures and molecular mechanism of a light-induced signaling

switch: The phot-LOV1 domain from *Chlamydomonas reinhardtii*, *Biophys. J.*, 2003, **84**, 2474–2482.

38 J. P. Zayner and T. R. Sosnick, Factors that control the chemistry of the LOV domain photocycle, *PLoS One*, 2014, **9**, e87074.

39 E. Schleicher, R. M. Kowalczyk, C. W. M. Kay, P. Hegemann, A. Bacher, M. Fischer, R. Bittl, G. Richter and S. Weber, On the reaction mechanism of adduct formation in LOV domains of the plant blue-light receptor phototropin, *J. Am. Chem. Soc.*, 2004, **126**, 11067–11076.

40 T. Kottke, J. Heberle, D. Hehn, B. Dick and P. Hegemann, Phot-LOV1: photocycle of a blue-light receptor domain from the green alga *Chlamydomonas reinhardtii*, *Biophys. J.*, 2003, **84**, 1192–1201.

41 C. Bauer, C.-R. Rabl, J. Heberle and T. Kottke, Indication for a radical intermediate preceding the signaling state in the LOV domain photocycle, *Photochem. Photobiol.*, 2011, **87**, 548–553.

42 C. W. M. Kay, E. Schleicher, A. Kuppig, H. Hofner, W. Rüdiger, M. Schleicher, M. Fischer, A. Bacher, S. Weber and G. Richter, Blue light perception in plants. Detection and characterization of a light-induced neutral flavin radical in a C450A mutant of phototropin, *J. Biol. Chem.*, 2003, **278**, 10973–10982.

43 T. E. Swartz, S. B. Corchnoy, J. M. Christie, J. W. Lewis, I. Szundi, W. R. Briggs and R. A. Bogomolni, The photocycle of a flavin-binding domain of the blue light photoreceptor phototropin, *J. Biol. Chem.*, 2001, **276**, 36493–36500.

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

45 A. Möglich and K. Moffat, Structural basis for light-dependent signaling in the dimeric LOV domain of the photosensor YtvA, *J. Mol. Biol.*, 2007, **373**, 112–126.

46 A. Pudasaini, K. K. El-Arab and B. D. Zoltowski, LOV-based optogenetic devices: Light-driven modules to impart photoregulated control of cellular signaling, *Front. Mol. Biosci.*, , DOI:10.3389/fmolb.2015.00018.

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

48 S. M. Harper, J. M. Christie and K. H. Gardner, Disruption of the LOV-Jα helix interaction activates phototropin kinase activity, *Biochemistry*, 2004, **43**, 16184–16192.

49 E. Huala, P. W. Oeller, E. Liscum, I. S. Han, E. Larsen and W. R. Briggs, *Arabidopsis* NPH1: A protein kinase with a putative redox-sensing domain, *Science (1979)*, 1997, **278**, 2120–2123.

50 T. Kinoshita, M. Doi, N. Suetsugu, T. Kagawa, M. Wada and K.-I. Shimazaki, phot1 and phot2 mediate blue light regulation of stomatal opening, *Nature*, 2001, **414**, 656–660.

51 T. Sakai, T. Kagawa, M. Kasahara, T. Swartz, J. Christie, W. Briggs, M. Wada and K. Okada, *Arabidopsis* nph1 and npl1: Blue light receptors that mediate both phototropism and chloroplast relocation, *PNAS U. S. A.*, 2001, **98**, 6969–6974.

52 K. Sakamoto and W. R. Briggs, Cellular and subcellular localization of Phototropin 1, *Plant Cell*, 2002, **14**, 1723–1735.

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

54 B. D. Zoltowski, C. Schwerdtfeger, J. Widom, J. J. Loros, A. M. Bilwes, J. C. Dunlap and B. R. Crane, Conformational switching in the fungal light sensor vivid, *Science (1979)*, 2007, **316**, 1054–1057.

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

56 A. T. Vaidya, C.-H. Chen, J. C. Dunlap, J. J. Loros and B. R. Crane, Structure of a light-activated LOV protein dimer that regulates transcription, *Sci. Signal.*, 2012, **4**, ra50.

57 M. Yazawa, A. M. Sadaghiani, B. Hsueh and R. E. Dolmetsch, Induction of protein-protein interactions in live cells using light, *Nat. Biotechnol.*, 2009, **27**, 941–945.

58 D. C. Nelson, J. Lasswell, L. E. Rogg, M. A. Cohen and B. Bartel, FKF1, a clock-controlled gene that regulates the transition to flowering in *Arabidopsis*, *Cell*, 2000, **101**, 331–340.

59 M. Sawa, D. A. Nusinow, S. A. Kay and T. Imaizumi, FKF1 and GIGANTEA complex formation is required for day-length measurement in *Arabidopsis*, *Science (1979)*, 2007, **318**, 261–265.

60 K. S. Conrad, A. M. Bilwes and B. R. Crane, Light-Induced subunit dissociation by a LOV domain photoreceptor from *Rhodobacter sphaeroides*, *Biochemistry (Easton)*, 2013, **52**, 378–391.

61 S. Metz, A. Jäger and G. Klug, Role of a short light, oxygen, voltage (LOV) domain protein in blue light-and singlet oxygen-dependent gene regulation in *Rhodobacter sphaeroides*, *Microbiology (N Y)*, 2012, **158**, 368–379.

62 A. I. Nash, R. McNulty, M. E. Shillito, T. E. Swartz, R. A. Bogomolni, H. Luecke and K. H. Gardner, Structural basis of photosensitivity in a bacterial light-oxygen-voltage/ helix-turn-helix (LOV-HTH) DNA-binding protein, *PNAS U. S. A.*, 2011, **108**, 9449–9454.

63 B. D. Zoltowski, L. B. Motta-Mena and K. H. Gardner, Blue light-induced dimerization of a bacterial LOV-HTH DNA-binding protein, *Biochemistry*, 2013, **52**, 6653–6661.

64 G. Rivera-Cancel, L. B. Motta-Mena and K. H. Gardner, Identification of natural and artificial DNA substrates for light-activated LOV-HTH transcription factor EL222, *Biochemistry*, 2012, **51**, 10024.

65 T. A. Gaidenko, T. J. Kim, A. L. Weigel, M. S. Brody and C. W. Price, The blue-light receptor YtvA acts in the environmental stress signaling pathway of *Bacillus subtilis*, *J. Bacteriol.*, 2006, **188**, 6387–6395.

66 M. Avila-Pérez, J. Vreede, Y. Tang, O. Bende, A. Losi, W. Gärtner and K. Hellingwerf, In vivo mutational analysis of YtvA from *Bacillus subtilis*. Mechanism of light activation of the general stress response, *J. Biol. Chem.*, 2009, **284**, 24958–24964.

67 L. de Mena, P. Rizk and D. Rincon-Limas, Bringing light to transcription: The optogenetics repertoire, *Front. Genet.*, 2018, **9**, 518.

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

69 D. Strickland, K. Moffat and T. R. Sosnick, Light-activated DNA binding in a designed allosteric protein, *PNAS U. S. A.*, 2008, **105**, 10709–10714.

70 J. Lee, M. Natarajan, V. C. Nashine, M. Socolich, T. Vo, W. P. Russ, S. J. Benkovic and R. Ranganathan, Surface sites for engineering allosteric control in proteins, *Science (1979)*, 2008, **322**, 438–442.

71 G. Guntas, R. A. Hallett, S. P. Zimmerman, T. Williams, H. Yumerefendi, J. E. Bear and B. Kuhlman, Engineering an improved light-induced dimer (iLID) for controlling the localization and activity of signaling proteins, *PNAS U. S. A.*, 2015, **112**, 112–117.

72 D. Strickland, Y. Lin, E. Wagner, C. M. Hope, J. Zayner, C. Antoniou, T. R. Sosnick, E. L. Weiss and M. Glotzer, TULIPs: tunable, light-controlled interacting protein tags for cell biology, *Nat. Methods*, 2012, **9**, 379–384.

73 C. P. Ponting, Evidence for PDZ domains in bacteria, yeast, and plants, *Protein Sci.*, 1997, **6**, 464–468.

74 D. A. Wah, I. Levchenko, T. A. Baker and R. T. Sauer, Characterization of a specificity factor for an AAA+ ATPase: Assembly of SspB Dimers with ssrA-Tagged Proteins and the ClpX Hexamer, *Chem. Biol.*, 2002, **9**, 1237–1245.

75 M. D. Hoffmann, F. Bubeck, R. Eils and D. Niopek, Controlling cells with light and LOV, *Adv. Biosyst.*, 2018, **2**, 1800098.

76 F. Kawano, H. Suzuki, A. Furuya and M. Sato, Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins, *Nat. Commun.*, 2015, **6**, 6256.

77 H. Wang, M. Vilela, A. Winkler, M. Tarnawski, I. Schlichting, H. Yumerefendi, B. Kuhlman, R. Liu, G. Danuser and K. M. Hahn, LOVTRAP, An pptogenetic system for photo-induced protein dissociation, *Nat. Methods*, 2016, **13**, 755–758.

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

79 D. S. Latchman, Transcription factors: An overview, *Int. J. Biochem. Cell Biol.*, 1997, **29**, 1305–1312.

80 D. S. Latchman, Transcription factors: an overview, *Int. J. Exp. Pathol.*, 1993, **74**, 417–422.

81 J. Locker, *Transcription factors*, BIOS, Oxford, UK, 2001.

82 R. G. Brennan and B. W. Matthews, The helix-turn-helix DNA binding motif, *J. Biol. Chem.*, 1989, **264**, 1903–1906.

83 L. Aravind, V. Anantharaman, S. Balaji, M. M. Babu and L. M. Iyer, The many faces of the helix-turn-helix domain: Transcription regulation and beyond, *FEMS Microbiol. Rev.*, 2005, **29**, 231–262.

84 R. G. Brennan, The winged-helix DNA-binding motif: Another helix-turn-helix takeoff, *Cell*, 1993, **74**, 773–776.

85 N. P. Pavletich and C. O. Pabo, Zinc finger-DNA recognition: Crystal structure of a Zif268-DNA complex at 2.1 Å, *Science (1979)*, 1991, **252**, 809–817.

86 S. A. Wolfe, L. Nekludova and C. O. Pabo, DNA recognition by (Cys2His2) zinc finger proteins, *Annu. Rev. Biophys. Biomol. Struct.*, 2000, **29**, 183-212.

87 A. C. Froehlich, Y. Liu, J. J. Loros and J. C. Dunlap, White Collar-1, a circadian blue light photoreceptor, binding to the frequency promoter, *Science (1979)*, 2002, **297**, 815–819.

88 Q. He, P. Cheng, Y. Yang, L. Wang, K. H. Gardner and Y. Liu, White Collar-1, a DNA binding transcription factor and a light sensor, *Science (1979)*, 2002, **297**, 840–843.

89 P. Ballario, P. Vittorioso, A. Magrelli, C. Talora, A. Cabibbo and G. Macino, White collar-1, a central regulator of blue light responses in *Neurospora*, is a zinc finger protein, *EMBO J.*, 1996, **15**, 1650–1657.

90 P. Stegmaier, A. Kel and E. Wingender, Systematic DNA-binding domain classification of transcription factors, *Genome informatics*, 2004, **15**, 276–286.

91 T. Ellenberger, Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domains, *Curr. Opin. Struct. Biol*, 1994, **4**, 12–21.

92 F. Takahashi, D. Yamagata, M. Ishikawa, Y. Fukamatsu, Y. Ogura, M. Kasahara, T. Kiyosue, M. Kikuyama, M. Wada and H. Kataoka, AUREOCHROME, a photoreceptor required for photomorphogenesis in stramenopiles, *PNAS U.S.A.*, 2007, **104**, 19625–19630.

93 K. K. Jefferson, What drives bacteria to produce a biofilm?, *FEMS Microbiol. Lett.*, 2004, **236**, 163–173.

94 R. M. Donlan, Biofilms: Microbial life on surfaces, *Emerging Infect. Dis.*, 2002, **8**, 881–890.

95 R. Belas, Biofilms, flagella, and mechanosensing of surfaces by bacteria, *Trends Microbiol.*, 2014, **22**, 517–527.

96 C. Beloin, A. Roux and J. M. Ghigo, *Escherichia coli* biofilms, *Curr. Top. Microbiol. Immunol.*, 2008, **322**, 249–289.

97 M. M. Barnhart and M. R. Chapman, Curli biogenesis and function, *Annu. Rev. Microbiol.*, 2006, **60**, 131–147.

98 P. M. Silverman and M. B. Clarke, New insights into F-pilus structure, dynamics, and function, *Integr. Biol.*, 2010, **2**, 25–31.

99 K. Kjaergaard, M. A. Schembri, H. Hasman and P. Klemm, Antigen 43 from *Escherichia coli* Induces Inter- and Intraspecies Cell Aggregation and Changes in Colony Morphology of *Pseudomonas fluorescens*, *J. Bacteriol.*, 2000, **182**, 4789–4796.

100 K. P. Rumbaugh and K. Sauer, Biofilm dispersion, *Nat. Rev. Microbiol.*, 2020, **18**, 571–586.

101 J. Bryers, in *Physiological Models in Microbiology*, 2018, pp. 109–144.

102 E. Karatan and P. Watnick, Signals, regulatory networks, and materials that build and break bacterial biofilms, *Microbiol. Mol. Biol. Rev.*, 2009, **73**, 310–347.

103 H.-C. Flemming and J. Wingender, The biofilm matrix, *Nat. Rev. Microbiol.*, 2010, **8**, 623–633.

104 X. Wang, 3rd Preston James F and T. Romeo, The pgaABCD locus of *Escherichia coli* promotes the synthesis of a polysaccharide adhesin required for biofilm formation, *J. Bacteriol.*, 2004, **186**, 2724–2734.

105 P. N. Danese, L. A. Pratt and R. Kolter, Exopolysaccharide production is required for development of *Escherichia coli* K-12 biofilm architecture, *J. Bacteriol.*, 2000, **182**, 3593–3596.

106 D. O. Serra, A. M. Richter and R. Hengge, Cellulose as an architectural element in spatially structured *Escherichia coli* biofilms, *J. Bacteriol.*, 2013, **195**, 5540–5554.

107 J. N. C. Fong and F. H. Yildiz, Biofilm matrix proteins, *Microbiol. Spectrum*, , DOI:10.1128/microbiolspec.MB-0004-2014.

108 C. B. Whitchurch, T. Tolker-Nielsen, P. C. Ragas and J. S. Mattick, Extracellular DNA required for bacterial biofilm formation, *Science (1979)*, 2002, **295**, 1487.

109 V. Sanchez-Torres, T. Maeda and T. K. Wood, Global regulator H-NS and lipoprotein NlpI influence production of extracellular DNA in *Escherichia coli*, *Biochem. Biophys. Res. Commun.*, 2010, **401**, 197–202.

110 M. E. Davey, N. C. Caiazza and G. A. O'Toole, Rhamnolipid surfactant production affects biofilm architecture in *Pseudomonas aeruginosa* PAO1, *J. Bacteriol.*, 2003, **185**, 1027–1036.

111 I. W. Sutherland, The biofilm matrix – an immobilized but dynamic microbial environment, *Trends Microbiol.*, 2001, **9**, 222–227.

112 P. S. Stewart and M. J. Franklin, Physiological heterogeneity in biofilms, *Nat. Rev. Microbiol.*, 2008, **6**, 199–210.

113 C. Prigent-Combaret, G. Prensier, T. T. Le Thi, O. Vidal, P. Lejeune and C. Dorel, Developmental pathway for biofilm formation in curli-producing *Escherichia coli* strains: role of flagella, curli and colanic acid, *Environ. Microbiol.*, 2000, **2**, 450–464.

114 H. Ogasawara, K. Yamamoto and A. Ishihama, Role of the biofilm master regulator CsgD in cross-regulation between biofilm formation and flagellar synthesis, *J. Bacteriol.*, 2011, **193**, 2587–2597.

115 M. Hammar, A. Arnqvist, Z. Bian, A. Olsén and S. Normark, Expression of two csg operons is required for production of fibronectin- and Congo red-binding curli polymers in *Escherichia coli* K-12, *Mol. Microbiol.*, 1995, **18**, 661–670.

116 M. Hammar, Z. Bian and S. Normark, Nucleator-dependent intercellular assembly of adhesive curli organelles in *Escherichia coli*, *PNAS*, 1996, **93**, 6562–6566.

117 Z. Bian and S. Normark, Nucleator function of CsgB for the assembly of adhesive surface organelles in *Escherichia coli*, *EMBO J.*, 1997, **16**, 5827–5836.

118 N. D. Hammer, J. C. Schmidt and M. R. Chapman, The curli nucleator protein, CsgB, contains an amyloidogenic domain that directs CsgA polymerization, *PNAS*, 2007, **104**, 12494–12499.

119 M. L. Evans, E. Chorell, J. D. Taylor, J. Åden, A. Götheson, F. Li, M. Koch, L. Sefer, S. J. Matthews, P. Wittung-Stafshede, F. Almqvist and M. R. Chapman, The bacterial curli system possesses a potent and selective inhibitor of amyloid formation, *Mol Cell*, 2015, **57**, 445–455.

120 E. Brombacher, C. Dorel, A. J. B. Zehnder and P. Landini, The curli biosynthesis regulator CsgD co-ordinates the expression of both positive and negative determinants for biofilm formation in *Escherichia coli*, *Microbiology (N Y)*, 2003, **149**, 2847–2857.

121 E. Brombacher, A. Baratto, C. Dorel and P. Landini, Gene expression regulation by the curli activator CsgD protein: modulation of cellulose biosynthesis and control of negative determinants for microbial adhesion, *J. Bacteriol.*, 2006, **188**, 2027–2037.

122 A. A. Nenninger, L. S. Robinson, N. D. Hammer, E. A. Epstein, M. P. Badtke, S. J. Hultgren and M. R. Chapman, CsgE is a curli secretion specificity factor that prevents amyloid fibre aggregation, *Mol. Microbiol.*, 2011, **81**, 486–499.

123 L. S. Robinson, E. M. Ashman, S. J. Hultgren and M. R. Chapman, Secretion of curli fibre subunits is mediated by the outer membrane-localized CsgG protein, *Mol. Microbiol.*, 2006, **59**, 870–881.

124 A. A. Nenninger, L. S. Robinson and S. J. Hultgren, Localized and efficient curli nucleation requires the chaperone-like amyloid assembly protein CsgF, *PNAS*, 2009, **106**, 900–905.

125 H. Loferer, M. Hammar and S. Normark, Availability of the fibre subunit CsgA and the nucleator protein CsgB during assembly of fibronectin-binding curli is limited by the intracellular concentration of the novel lipoprotein CsgG, *Mol. Microbiol.*, 1997, **26**, 11–23.

126 M. H. Muhammad, A. L. Idris, X. Fan, Y. Guo, Y. Yu, X. Jin, J. Qiu, X. Guan and T. Huang, Beyond risk: Bacterial biofilms and their regulating approaches, *Front. Microbiol.*, 2020, **11**, 928.

127 K. Das, M. V. S. Rajawat, A. K. Saxena and R. Prasanna, Development of *Mesorhizobium ciceri*-based biofilms and analyses of their antifungal and plant growth promoting activity in Chickpea hallenged by Fusarium wilt, *Indian J. Microbiol.*, 2017, **57**, 48–59.

128 A. Kanchan, K. Simranjit, K. Ranjan, R. Prasanna, B. Ramakrishnan, M. C. Singh, M. Hasan and Y. S. Shivay, Microbial biofilm inoculants benefit growth and yield of chrysanthemum varieties under protected cultivation through enhanced nutrient availability, *Plant Biosyst.*, 2019, **153**, 306–316.

129 A. Mitra and S. Mukhopadhyay, Biofilm mediated decontamination of pollutants from the environment, *AIMS bioeng.*, 2016, **3**, 44–59.

130 S. Mishra, Y. Huang, J. Li, X. Wu, Z. Zhou, Q. Lei, P. Bhatt and S. Chen, Biofilm-mediated bioremediation is a powerful tool for the removal of environmental pollutants, *Chemosphere*, 2022, **294**, 133609.

131 W. Sun, S. Shi, J. Chen, W. Zhao, T. Chen, G. Li, K. Zhang, B. Yu, D. Liu, Y. Chen, H. Ying and P. Ouyang, Blue light signaling regulates *Escherichia coli* W1688 biofilm formation and L-Threonine production, *Microbiol. Spectrum*, 2022, **10**, e0246022–e0246022.

132 G. T. H. Ooi, K. Tang, R. K. Chhetri, K. M. S. Kaarsholm, K. Sundmark, C. Kragelund, K. Litty, A. Christensen, S. Lindholst, C. Sund, M. Christensson, K. Bester and H. R. Andersen, Biological removal of pharmaceuticals from hospital wastewater in a pilot-scale staged moving bed biofilm reactor (MBBR) utilising nitrifying and denitrifying processes, *Bioresour. Technol.*, 2018, **267**, 677–687.

133 C.-C. Tang, Y. Tian, Z.-W. He, W. Zuo and J. Zhang, Performance and mechanism of a novel algal-bacterial symbiosis system based on sequencing batch suspended biofilm reactor treating domestic wastewater, *Bioresour. Technol.*, 2018, **265**, 422–431.

134 T. Liu, Z. Guo, Z. Zeng, N. Guo, Y. Lei, T. Liu, S. Sun, X. Chang, Y. Yin and X. Wang, Marine bacteria provide lasting anticorrosion activity for steel via biofilm-induced mineralization, *ACS Appl. Mater. Interfaces*, 2018, **10**, 40317–40327.

135 M. Mukhi and A. S. Vishwanathan, Beneficial biofilms: A mini-review of strategies to enhance biofilm formation for biotechnological applications, *Appl. Environ. Microbiol.*, 2021, **88**, e0199421.

136 F. Chen and S. V Wegner, Blue light switchable bacterial adhesion as a key step toward the design of biofilms, *ACS Synth. Biol.*, 2017, **6**, 2170–2174.

137 F. Chen and S. V Wegner, Blue-light-switchable bacterial cell–cell adhesions enable the control of multicellular bacterial communities, *ACS Synth. Biol.*, 2020, **9**, 1169–1180.

138 X. Jin and I. H. Riedel-Kruse, Biofilm Lithography enables high-resolution cell patterning via optogenetic adhesin expression, *PNAS*, 2018, **115**, 3698–3703.

139 F. Zhao, M. S. Chavez, K. L. Naughton, C. M. Niman, J. T. Atkinson, J. A. Gralnick, M. Y. El-Naggar and J. Q. Boedicker, Light-induced patterning of electroactive bacterial biofilms, *ACS Synth. Biol.*, 2022, **11**, 2327–2338.

140 A. Pirhanov, C. M. Bridges, R. A. Goodwin, Y.-S. Guo, J. Furrer, L. M. Shor, D. J. Gage and Y. K. Cho, Optogenetics in *Sinorhizobium meliloti* enables spatial control of exopolysaccharide production and biofilm structure, *ACS Synth. Biol.*, 2021, **10**, 345–356.

141 M. Chalfie, G. Euskirchen, W. W. Ward and D. C. Prasher, Green fluorescent protein as a marker for gene expression, *Science (1979)*, 1994, **263**, 802–805.

142 Y. Shimomura, Osamu; Johnson, Frank H. ; Saiga, Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea.*, *J. Cell. and Comp. Physiol.*, 1962, **59**, 223–240.

143 D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast and M. J. Cormier, Primary structure of the *Aequorea victoria* green-fluorescent protein, *Gene*, 1992, **111**, 229–233.

144 M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington, Crystal structure of the *Aequorea victoria* Green Fluorescent Protein, *Science (1979)*, 1996, **273**, 1392–1395.

145 F. Yang, L. G. Moss and G. N. Phillips, The molecular structure of green fluorescent protein, *Nat. Biotechnol.*, 1996, **14**, 1246–1251.

146 C. W. Cody, D. C. Prasher, W. M. Westler, F. G. Prendergast and W. W. Ward, Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein, *Biochemistry (Easton)*, 1993, **32**, 1212–1218.

147 L. Zhang, H. N. Patel, J. W. Lappe and R. M. Wachter, Reaction progress of chromophore biogenesis in green fluorescent protein, *J. Am. Chem. Soc.*, 2006, **128**, 4766–4772.

148 B. G. Reid and G. C. Flynn, Chromophore formation in Green Fluorescent Protein, *Biochemistry (Easton)*, 1997, **36**, 6786–6791.

149 R. Heim, D. C. Prasher and R. Y. Tsien, Wavelength mutations and posttranslational autoxidation of Green Fluorescent Protein, *PNAS*, 1994, **91**, 12501–12504.

150 R. Y. Tsien, The green fluorescent protein, *Annu. Rev. Biochem.*, 1998, **67**, 509–544.

151 K. Brejc, T. K. Sixma, P. A. Kitts, S. R. Kain, R. Y. Tsien, M. Ormo and S. J. Remington, Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein, *PNAS*, 1997, **94**, 2306–2311.

152 R. Heim, A. B. Cubitt and R. Y. Tsien, Improved green fluorescence, *Nature*, 1995, **373**, 663–664.

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

154 R. Heim and R. Y. Tsien, Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer, *Curr. Biol.*, 1996, **6**, 178–182.

155 M. Ormö, A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington, Crystal structure of the *Aequorea victoria* Green Fluorescent Protein, *Science (1979)*, 1996, **273**, 1392–1395.

156 M. V Matz, A. F. Fradkov, Y. A. Labas, A. P. Savitsky, A. G. Zaraisky, M. L. Markelov and S. A. Lukyanov, Fluorescent proteins from nonbioluminescent Anthozoa species, *Nat. Biotechnol.*, 1999, **17**, 969–973.

157 L. A. Gross, G. S. Baird, R. C. Hoffman, K. K. Baldridge and R. Y. Tsien, The structure of the chromophore within DsRed, a red fluorescent protein from Coral, *PNAS*, 2000, **97**, 11990–11995.

158 M. Zimmer, Green Fluorescent Protein (GFP): Applications, structure, and related photophysical behavior, *Chem. Rev.*, 2002, **102**, 759–782.

159 W. Margolin, The price of tags in protein localization studies, *J. Bacteriol.*, 2012, **194**, 6369–6371.

160 M. Monici, Cell and tissue autofluorescence research and diagnostic applications, *Biotechnol. Annu. Rev.*, 2005, **11**, 227–256.

161 O. V Stepanenko, M. I. Sulatsky, E. V Mikhailova, I. M. Kuznetsova, K. K. Turoverov, O. V Stepanenko and A. I. Sulatskaya, New findings on GFP-like protein application as fluorescent tags: Fibrillogenesis, oligomerization, and amorphous aggregation, *Int. J. Biol. Macromol.*, 2021, **192**, 1304–1310.

162 J. Kong, Y. Wang, W. Qi, M. Huang, R. Su and Z. He, Green fluorescent protein inspired fluorophores, *Adv. Colloid Interface Sci.*, 2020, **285**, 102286.

163 M. S. Baranov, K. A. Lukyanov, A. O. Borissova, J. Shamir, D. Kosenkov, L. V Slipchenko, L. M. Tolbert, I. V Yampolsky and K. M. Solntsev, Conformationally locked chromophores as models of excited-state proton transfer in Fluorescent Proteins, *J. Am. Chem. Soc.*, 2012, **134**, 6025–6032.

164 A. Baldridge, K. M. Solntsev, C. Song, T. Tanioka, J. Kowalik, K. Hardcastle and L. M. Tolbert, Inhibition of twisting of a green fluorescent protein-like chromophore by metal complexation, *Chem. Commun. (Cambridge, U. K.)*, 2010, **46**, 5686–5688.

165 Y. Zheng, G. Li, H. Deng, Y. Su, J. Liu and X. Zhu, Temperature-induced fluorescence enhancement of GFP chromophore containing copolymers for detection of *Bacillus thermophilus*, *Polym. Chem.*, 2014, **5**, 2521–2529.

166 T. Kanamori, A. Takamura, N. Tago, Y. Masaki, A. Ohkubo, M. Sekine and K. Seio, Fluorescence enhancement of oligodeoxynucleotides modified with green fluorescent protein chromophore mimics upon triplex formation, *Org. Biomol. Chem.*, 2017, **15**, 1190–1197.

167 E. A. Dolgopolova, T. M. Moore, W. B. Fellows, M. D. Smith and N. B. Shustova, Photophysics of GFP-related chromophores imposed by a scaffold design, *Dalton Trans.*, 2016, **45**, 9884–9891.

168 J. S. Paige, K. Y. Wu and S. R. Jaffrey, RNA mimics of green fluorescent protein, *Science (1979)*, 2011, **333**, 642.

169 M. Mascini, I. Palchetti and S. Tombelli, *Angew. Chem.*, 2012, 51, 1316–1332.

170 T. Hermann and D. J. Patel, Adaptive recognition by nucleic acid aptamers, *Science (1979)*, 2000, **287**, 820–825.

171 A. M. Femino, F. S. Fay, K. Fogarty and R. H. Singer, Visualization of single RNA transcripts in situ, *Science (1979)*, 1998, **280**, 585–590.

172 E. Bertrand, P. Chartrand, M. Schaefer, S. M. Shenoy, R. H. Singer and R. M. Long, Localization of ASH1 mRNA Particles in Living Yeast, *Mol Cell*, 1998, **2**, 437–445.

173 J. Dictenberg, Genetic encoding of fluorescent RNA ensures a bright future for visualizing nucleic acid dynamics, *Trends Biotechnol.*, 2012, **30**, 621–626.

174 S. Tyagi and F. R. Kramer, Molecular beacons: probes that fluoresce upon hybridization, *Nat. Biotechnol.*, 1996, **14**, 303–308.

175 A. D. Ellington and J. W. Szostak, In vitro selection of RNA molecules that bind specific ligands, *Nature*, 1990, **346**, 818–822.

176 C. Tuerk and L. Gold, Systematic Evolution of Ligands by Exponential Enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science (1979)*, 1990, **249**, 505–510.

177 J. R. Babendure, S. R. Adams and R. Y. Tsien, Aptamers switch on fluorescence of triphenylmethane dyes, *J. Am. Chem. Soc.*, 2003, **125**, 14716–14717.

178 H. Huang, N. B. Suslov, N.-S. Li, S. A. Shelke, M. E. Evans, Y. Koldobskaya, P. A. Rice and J. A. Piccirilli, A G-quadruplex–containing RNA activates fluorescence in a GFP-like fluorophore, *Nat. Chem. Biol.*, 2014, **10**, 686–691.

179 K. D. Warner, M. C. Chen, W. Song, R. L. Strack, A. Thorn, S. R. Jaffrey and A. R. Ferré-D’Amaré, Structural basis for activity of highly efficient RNA mimics of green fluorescent protein, *Nat. Struct. Mol. Biol.*, 2014, **21**, 658–663.

180 R. L. Strack, M. D. Disney and S. R. Jaffrey, A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA, *Nat. Methods*, 2013, **10**, 1219–1224.

181 G. S. Filonov, J. D. Moon, N. Svensen and S. R. Jaffrey, Broccoli: Rapid selection of an RNA mimic of Green Fluorescent Protein by fluorescence-based selection and directed evolution, *J. Am. Chem. Soc.*, 2014, **136**, 16299–16308.

182 K. Y. Han, B. J. Leslie, J. Fei, J. Zhang and T. Ha, Understanding the photophysics of the Spinach–DFHBI RNA aptamer–Fluorogen complex to improve live-cell RNA imaging, *J. Am. Chem. Soc.*, 2013, **135**, 19033–19038.

183 K. D. Warner, L. Sjeklоća, W. Song, G. S. Filonov, S. R. Jaffrey and A. R. Ferré-D’Amaré, A homodimer interface without base pairs in an RNA mimic of red fluorescent protein, *Nat. Chem. Biol.*, 2017, **13**, 1195.

184 S. Wenjiao, S. F. Grigory, K. Hyaeyeong, H. Markus, L. Xing, D. M. Jared and R. J. Samie, Imaging RNA polymerase III transcription using a photostable RNA–fluorophore complex, *Nat. Chem. Biol.*, 2017, **13**, 1187–1194.

185 C. Steinmetzger, N. Palanisamy, K. R. Gore and C. Höbartner, *Chem. - Eur. J.*, 2019, 25, 1931–1935.

186 C. Steinmetzger, I. Bessi, A.-K. Lenz and C. Höbartner, Structure-fluorescence activation relationships of a large Stokes shift fluorogenic RNA aptamer, *Nucleic Acids Res.*, 2019, **47**, 11538–11550.

187 M. Mieczkowski, C. Steinmetzger, I. Bessi, A.-K. Lenz, A. Schmiedel, M. Holzapfel, C. Lambert, V. Pena and C. Höbartner, Large Stokes shift fluorescence activation in an RNA aptamer by intermolecular proton transfer to guanine, *Nat. Commun.*, 2021, **12**, 3549.

188 S. K. Dey, G. S. Filonov, A. O. Olarerin-George, B. T. Jackson, L. W. S. Finley and S. R. Jaffrey, Repurposing an adenine riboswitch into a fluorogenic imaging and sensing tag, *Nat. Chem. Biol.*, 2022, **18**, 180–190.

189 L. Truong, H. Kooshapur, S. K. Dey, X. Li, N. Tjandra, S. R. Jaffrey and A. R. Ferré-D'Amaré, The fluorescent aptamer Squash extensively repurposes the adenine riboswitch fold, *Nat. Chem. Biol.*, 2022, **18**, 191–198.

190 E. V Dolgosheina, S. C. Y. Jeng, S. S. S. Panchapakesan, R. Cojocaru, P. S. K. Chen, P. D. Wilson, N. Hawkins, P. A. Wiggins and P. J. Unrau, RNA mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking, *ACS Chem. Biol.*, 2014, **9**, 2412–2420.

191 R. J. Trachman, N. A. Demeshkina, M. W. L. Lau, S. S. S. Panchapakesan, S. C. Y. Jeng, P. J. Unrau and A. R. Ferré-D'Amaré, Structural basis for high-affinity fluorophore binding and activation by RNA Mango, *Nat. Chem. Biol.*, 2017, **13**, 807–813.

192 X. Chen, D. Zhang, N. Su, B. Bao, X. Xie, F. Zuo, L. Yang, H. Wang, L. Jiang, Q. Lin, M. Fang, N. Li, X. Hua, Z. Chen, C. Bao, J. Xu, W. Du, L. Zhang, Y. Zhao, L. Zhu, J. Loscalzo and Y. Yang, Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs, *Nat. Biotechnol.*, 2019, **37**, 1287–1293.

193 K. Huang, X. Chen, C. Li, Q. Song, H. Li, L. Zhu, Y. Yang and A. Ren, Structure-based investigation of fluorogenic Pepper aptamer, *Nat. Chem. Biol.*, 2021, **17**, 1289–1295.

194 L. Roszyk, S. Kollenda and S. Hennig, Using a specific RNA–protein interaction to quench the fluorescent RNA Spinach, *ACS Chem. Biol.*, 2017, **12**, 2958–2964.

195 M. F. Al Mazid, O. Shkel, Y. Kharkivska and J.-S. Lee, Application of fluorescent turn-on aptamers in RNA studies, *Mol. Omics*, 2021, **17**, 483–491.

196 L. Truong and A. R. Ferré-D'Amaré, From fluorescent proteins to fluorogenic RNAs: Tools for imaging cellular macromolecules, *Protein Sci.*, 2019, **28**, 1374–1386.

197 M. T. Banco and A. R. Ferré-D'Amaré, The emerging structural complexity of G-quadruplex RNAs, *RNA*, 2021, **27**, 390–402.

198 Q. Wang, F. Xiao, H. Su, H. Liu, J. Xu, H. Tang, S. Qin, Z. Fang, Z. Lu, J. Wu, X. Weng and X. Zhou, Inert Pepper aptamer-mediated endogenous mRNA recognition and imaging in living cells, *Nucleic Acids Res.*, 2022, **50**, e84–e84.

199 M. Fang, H. Li, X. Xie, H. Wang, Y. Jiang, T. Li, B. Zhang, X. Jiang, Y. Cao, R. Zhang, D. Zhang, Y. Zhao, L. Zhu, X. Chen and Y. Yang, Imaging intracellular metabolite and protein changes in live mammalian cells with bright fluorescent RNA-based genetically encoded sensors, *Biosens. Bioelectron.*, 2023, **235**, 115411.

200 M. Kasahara, T. E. Swartz, M. A. Olney, A. Onodera, N. Mochizuki, H. Fukuzawa, E. Asamizu, S. Tabata, H. Kanegae, M. Takano, J. M. Christie, A. Nagatani and W. R. Briggs, Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from *Arabidopsis*, Rice, and *Chlamydomonas reinhardtii*, *Plant Physiol.*, 2002, **129**, 762–773.

201 Y. I. Wu, D. Frey, O. I. Lungu, A. Jaehrig, I. Schlichting, B. Kuhlman and K. M. Hahn, A genetically encoded photoactivatable Rac controls the motility of living cells, *Nature*, 2009, **461**, 104–108.

202 M. P. Cava and M. V Lakshmikantham, eds. A. R. Katritzky and C. W. B. T.-C. H. C. Rees, Pergamon, Oxford, 1984, pp. 1037–1083.

203 W. L. F. Armarego, Elsevier, United States, 8th Editio., 2017, p. 1.

204 K.-J. Hwang, S. H. Lee, H. J. Kim, J.-Y. Lee and J. S. Kim, A balloon filled with nitrogen gas does not satisfy the air- or moisture-free reaction condition, *Bull. Korean Chem. Soc.*, 2010, **31**, 515–516.

205 D. R. Hepburn and H. R. Hudson, Factors in the formation of isomerically and optically pure alkyl halides. Part XI. Vilsmeier reagents for the replacement of a hydroxy-group by chlorine or bromine, *J. Am. Chem. Soc.*, 1976, 754–757.

206 B. Gestblom, P. Nielsen, S. Forsén, I. Andersen and J. Munch-Petersen, Long-range proton spin couplings in thieno[3,2-b]thiophene, *Acta Chem. Scand.*, 1963, **17**, 280–281.

207 B. Bickler and E. Denton, in *ACS National Meeting*, Biotage, San Francisco, 2017.

208 R. S. Schwartz, H. Yokokawa and E. W. Graham, Kinetics of the acid-catalyzed isomerization of phenylbutenes in glacial acetic acid, *J. Am. Chem. Soc.*, 1972, **94**, 1247–1249.

209 A. E. Settle, L. Berstis, S. Zhang, N. A. Rorrer, H. Hu, R. M. Richards, G. T. Beckham, M. F. Crowley and D. R. Vardon, Iodine-catalyzed isomerization of cimethyl muconate, *ChemSusChem*, 2018, **11**, 1768–1780.

210 C. M. Williams and L. N. Mander, Chromatography with silver nitrate, *Tetrahedron*, 2001, **57**, 425–447.

211 L. Marschall, P. Sagmeister and C. Herwig, Tunable recombinant protein expression in E. coli: promoter systems and genetic constraints, *Appl. Microbiol. Biotechnol.*, 2017, **101**, 501–512.

212 C. Engler, R. Kandzia and S. Marillonnet, A one pot, one step, precision cloning method with high throughput capability, *PLoS One*, 2008, **3**, e3647–e3647.

213 P. Jayaraman, J. W. Yeoh, S. Jayaraman, A. Y. Teh, J. Zhang and C. L. Poh, Cell-free optogenetic gene expression system, *ACS Synth. Biol.*, 2018, **7**, 986.

214 P. Jayaraman, K. Devarajan, T. K. Chua, H. Zhang, E. Gunawan and C. L. Poh, Blue light-mediated transcriptional activation and repression of gene expression in bacteria, *Nucleic Acids Res.*, 2016, **44**, 6994–7005.

215 S. C. Harrison and A. K. Aggarwal, DNA recognition by proteins with the Helix-Turn-Helix motif, *Annu. Rev. Biochem.*, 1990, **59**, 933–969.

216 Y. Wen, Z. Ouyang, B. Devreese, W. He, Y. Shao, W. Lu and F. Zheng, Crystal structure of master biofilm regulator CsgD regulatory domain reveals an atypical receiver domain, *Protein Sci.*, 2017, **26**, 2073–2082.

217 K. Zakikhany, C. R. Harrington, M. Nimtz, J. C. D. Hinton and U. Römling, Unphosphorylated CsgD controls biofilm formation in *Salmonella enterica* serovar Typhimurium, *Mol. Microbiol.*, 2010, **77**, 771–786.

218 C.-H. Yan, F.-H. Chen, Y.-L. Yang, Y.-F. Zhan, R. A. Herman, L.-C. Gong, S. Sheng and J. Wang, The transcription factor CsgD contributes to engineered *Escherichia coli* resistance by regulating biofilm formation and stress responses, *Int. J. Mol. Sci.*, 2023, **24**, 13681.

219 U. Gerstel, C. Park and U. Römling, Complex regulation of csgD promoter activity by global regulatory proteins, *Mol. Microbiol.*, 2003, **49**, 639–654.

220 C. Prigent-Combaret, E. Brombacher, O. Vidal, A. Ambert, P. Lejeune, P. Landini and C. Dorel, Complex regulatory network controls initial adhesion and biofilm formation in *Escherichia coli* via regulation of the csgD gene, *J. Bacteriol.*, 2001, **183**, 7213–7223.

221 M. K. Thomason, F. Fontaine, N. De Lay and G. Storz, *Mol. Microbiol.*, 2012, 84, 17–35.

222 A. Olsén, A. Arnqvist, M. Hammar, S. Sukupolvi and S. Normark, The RpoS Sigma factor relieves H-NS-mediated transcriptional repression of csgA, the subunit gene of fibronectin-binding curli in *Escherichia coli*, *Mol. Microbiol.*, 1993, **7**, 523–536.

223 G. A. O'Toole, Microtiter dish biofilm formation assay, *JoVE*, 2011, **47**, 2437.

224 Y. Wu, R. Wang, M. Xu, Y. Liu, X. Zhu, J. Qiu, Q. Liu, P. He and Q. Li, A Novel Polysaccharide Depolymerase Encoded by the Phage SH-KP152226 Confers Specific Activity Against Multidrug-Resistant *Klebsiella pneumoniae* via Biofilm Degradation, *Frontiers Microbiol.*, 2019, **10**, 2768.

225 B. Pitts, M. A. Hamilton, N. Zelver and P. S. Stewart, A microtiter-plate screening method for biofilm disinfection and removal, *J. Microbiol. Methods*, 2003, **54**, 269–276.

226 M.-C. Wang, C.-C. Tseng, A.-B. Wu, J.-J. Huang, B.-S. Sheu and J.-J. Wu, Different roles of host and bacterial factors in *Escherichia coli* extra-intestinal infections, *Clin. Microbiol. Infect.*, 2009, **15**, 372–379.

227 K. Thongbhubate, Y. Nakafuji, R. Matsuoka, S. Kakegawa and H. Suzuki, Effect of spermidine on biofilm formation in *Escherichia coli* k-12, *J. Bacteriol.*, 2021, **203**, 1.

228 J. Azeredo, N. F. Azevedo, R. Briandet, N. Cerca, T. Coenye, A. R. Costa, M. Desvaux, G. Di Bonaventura, M. Hébraud, Z. Jaglic, M. Kačániová, S. Knøchel, A. Lourenço, F. Mergulhão, R. L. Meyer, G. Nychas, M. Simões, O. Tresse and C. Sternberg, Critical review on biofilm methods, *Crit. Rev. Microbiol.*, 2017, **43**, 313–351.

229 J. H. Merritt, D. E. Kadouri and G. A. O'Toole, Growing and analyzing static biofilms, *Curr. Protoc. Microbiol.*, 2006, **0**, 1B.1.1-1B.1.17.

230 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli and D. Hassabis, Highly accurate protein structure prediction with AlphaFold, *Nature*, 2021, **596**, 583–589.

231 M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis and S. Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models, *Nucleic Acids Res.*, 2022, **50**, D439–D444.

232 D. W. Nielsen, J. S. Klimavicz, T. Cavender, Y. Wannemuehler, N. L. Barbieri, L. K. Nolan and C. M. Logue, The impact of media, phylogenetic classification, and *E. coli*

pathotypes on biofilm formation in extraintestinal and commensal *E. coli* from humans and animals, *Frontiers Microbiol.*, 2018, **9**, 902.

233 P. Naves, G. del Prado, L. Huelves, M. Gracia, V. Ruiz, J. Blanco, V. Rodrguez-Cerrato, M. C. Ponte and F. Soriano, Measurement of biofilm formation by clinical isolates of *Escherichia coli* is method-dependent, *J. Appl. Microbiol.*, 2008, **105**, 585–590.

234 D. Egan, R. Meged, I. Berry, J. Newman, T. S. Walter, M. Ben Jelloul, J. L. Sussman, D. I. Stuart and A. Perrakis, Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2005, **61**, 1426–1431.

235 U. Heintz and I. Schlichting, Blue light-induced LOV domain dimerization enhances the affinity of Aureochrome 1a for its target DNA sequence, *Elife*, 2016, **5**, e11860.

236 E. Herman and T. Kottke, Allosterically regulated unfolding of the A ʹalpha helix exposes the dimerization site of the blue-light-sensing Aureochrome-LOV domain, *Biochemistry*, 2015, **54**, 1484–1492.

237 M. E. Kalvaitis, L. A. Johnson, R. J. Mart, P. Rizkallah and R. K. Allemann, A noncanonical chromophore reveals structural rearrangements of the Light-Oxygen-Voltage domain upon photoactivation, *Biochemistry*, 2019, **58**, 2608.

238 J. Kennis, M. Alexandre, J. Arents, R. van Grondelle and K. Hellingwerf, A base-catalyzed mechanism for dark state recovery in the *Avena sativa* phototropin-1 LOV2 domain, *Comp. Biochem. Physiol., Part A: Mol. Integr. Physiol.*, 2007, **146**, S227–S227.

239 A. I. Nash, W.-H. Ko, S. M. Harper and K. H. Gardner, A conserved Glutamine plays a central role in LOV domain signal transmission and its duration, *Biochemistry (Easton)*, 2008, **47**, 13842–13849.

240 T. Toyooka, O. Hisatomi, F. Takahashi, H. Kataoka and M. Terazima, Photoreactions of Aureochrome-1, *Biophys. J.*, 2011, **100**, 2801–2809.

241 M. E. Kalvaitis, PhD Thesis, Cardiff University, 2019.

242 J. Agirre, M. Atanasova, H. Bagdonas, C. B. Ballard, A. Baslé, J. Beilsten-Edmands, R. J. Borges, D. G. Brown, J. J. Burgos-Mármol, J. M. Berrisford, P. S. Bond, I. Caballero, L. Catapano, G. Chojnowski, A. G. Cook, K. D. Cowtan, T. I. Croll, J. É. Debreczeni, N. E. Devenish, E. J. Dodson, T. R. Drevon, P. Emsley, G. Evans, P. R. Evans, M. Fando, J. Foadi, L. Fuentes-Montero, E. F. Garman, M. Gerstel, R. J. Gildea, K. Hatti, M. L. Hekkelman, P. Heuser, S. W. Hoh, M. A. Hough, H. T. Jenkins, E. Jiménez, R. P. Joosten, R. M. Keegan, N. Keep, E. B. Krissinel, P. Kolenko, O. Kovalevskiy, V. S. Lamzin, D. M. Lawson, A. A.

Lebedev, A. G. W. Leslie, B. Lohkamp, F. Long, M. Malý, A. J. McCoy, S. J. McNicholas, A. Medina, C. Millán, J. W. Murray, G. N. Murshudov, R. A. Nicholls, M. E. M. Noble, R. Oeffner, N. S. Pannu, J. M. Parkhurst, N. Pearce, J. Pereira, A. Perrakis, H. R. Powell, R. J. Read, D. J. Rigden, W. Rochira, M. Sammito, F. Sánchez Rodr\'\iguez, G. M. Sheldrick, K. L. Shelley, F. Simkovic, A. J. Simpkin, P. Skubak, E. Sobolev, R. A. Steiner, K. Stevenson, I. Tews, J. M. H. Thomas, A. Thorn, J. T. Valls, V. Uski, I. Usón, A. Vagin, S. Velankar, M. Vollmar, H. Walden, D. Waterman, K. S. Wilson, M. D. Winn, G. Winter, M. Wojdyr and K. Yamashita, The CCP4 suite: integrative software for macromolecular crystallography, *Acta Crystallogr. , Sect. D*, 2023, **79**, 449–461.

243 L. Potterton, J. Agirre, C. Ballard, K. Cowtan, E. Dodson, P. R. Evans, H. T. Jenkins, R. Keegan, E. Krissinel, K. Stevenson, A. Lebedev, S. J. McNicholas, R. A. Nicholls, M. Noble, N. S. Pannu, C. Roth, G. Sheldrick, P. Skubak, J. Turkenburg, V. Uski, F. von Delft, D. Waterman, K. Wilson, M. Winn and M. Wojdyr, CCP4i2: the new graphical user interface to the CCP4 program suite, *Acta Crystallogr. , Sect. D*, 2018, **74**, 68–84.

244 J. Podlesný, O. Pytela, M. Klikar, V. Jelínková, I. V Kityk, K. Ozga, J. Jedryka, M. Rudysh and F. Bures, Small isomeric pushpull chromophores based on thienothiophenes with tunable optical (non)linearities, *Org. Biomol. Chem.*, 2019, **17**, 3623–3634.

245 M. Manuela, M. Raposo, C. Herbivo, V. Hugues, G. Clermont, M. C. R. Castro, A. Comel and M. Blanchard-Desce, Synthesis, fluorescence, and two-photon absorption properties of push-pull 5-arylthieno[3,2-b]thiophene derivatives, *Eur. J. Org. Chem.*, 2016, **2016**, 5263–5273.

246 P. D. Shaw Stewart, S. A. Kolek, R. A. Briggs, N. E. Chayen and P. F. M. Baldock, Random Microseeding: A Theoretical and Practical Exploration of Seed Stability and Seeding Techniques for Successful Protein Crystallization, *Cryst. Growth Des.*, 2011, **11**, 3432–3441.

247 A. D. Crawshaw, E. V Beale, A. J. Warren, A. Stallwood, G. Duller, J. Trincao and G. Evans, A sample preparation pipeline for microcrystals at the VMXm beamline, *JoVE*, 2021, e62306.

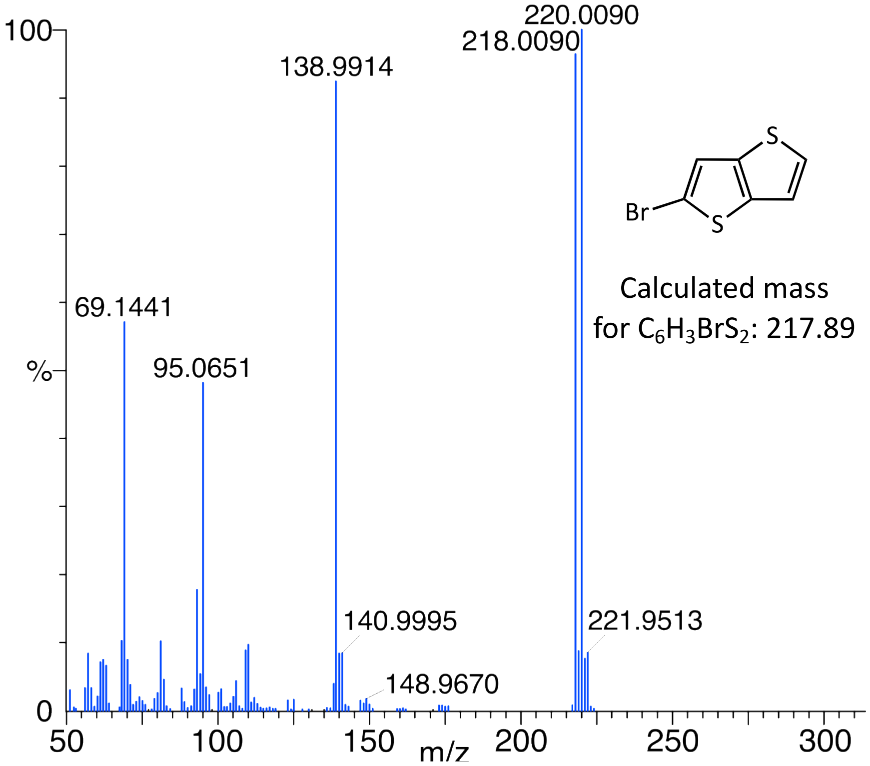
Appendix

Chapter 2. Studying pepper RNA aptamer as a reporter of transcription in optogenetic systems

Synthesis of HBC_{620}

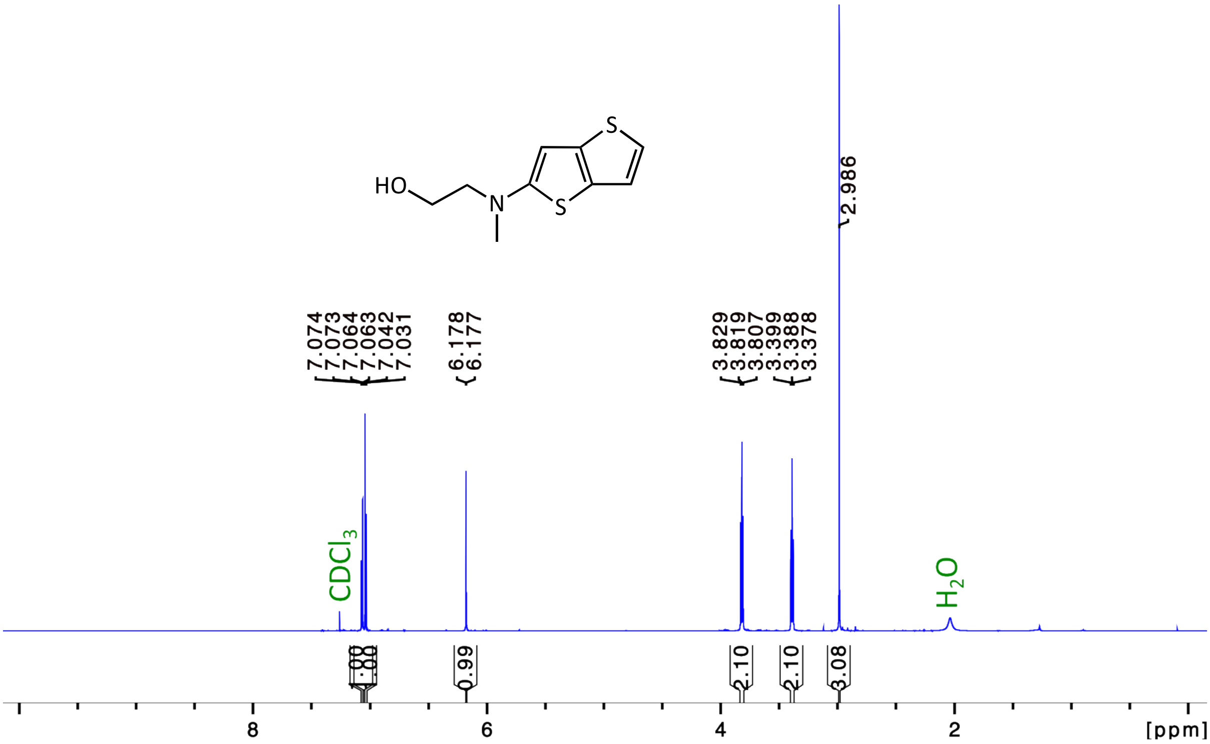
Characterisation of compounds from original HBC_{620} synthetic route[192]

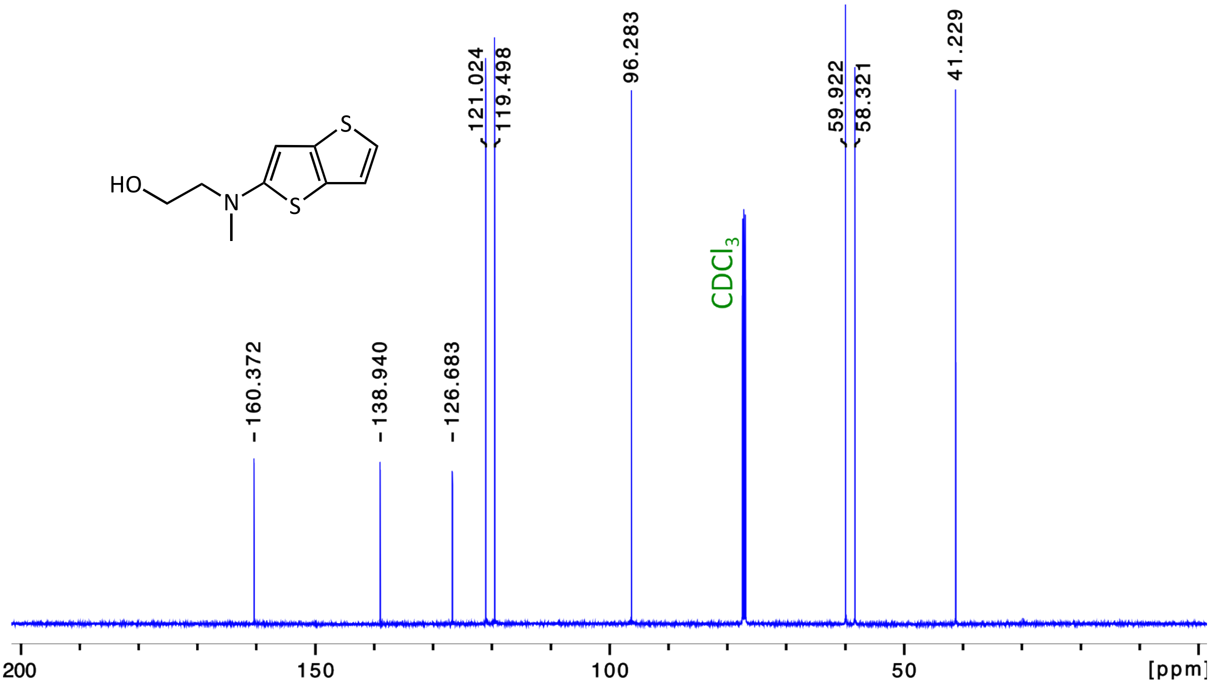
Mass spectrum of 2-Bromothieno[3,2-b]thiophene (2)

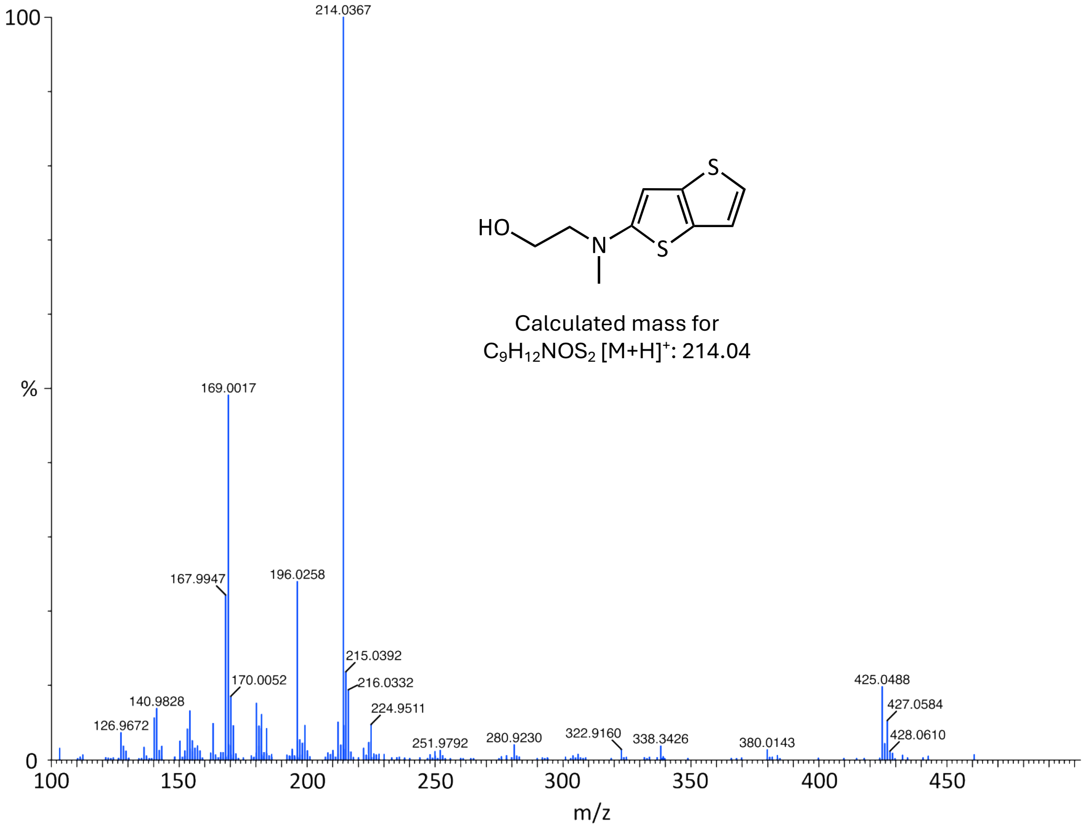


Calculated mass for $C_6H_3BrS_2$: 217.89

^{1}H-NMR, ^{13}C-NMR and MS of 2-[(2-hydroxyethyl)methylamino]thieno[3,2-*b*] thiophene (4)

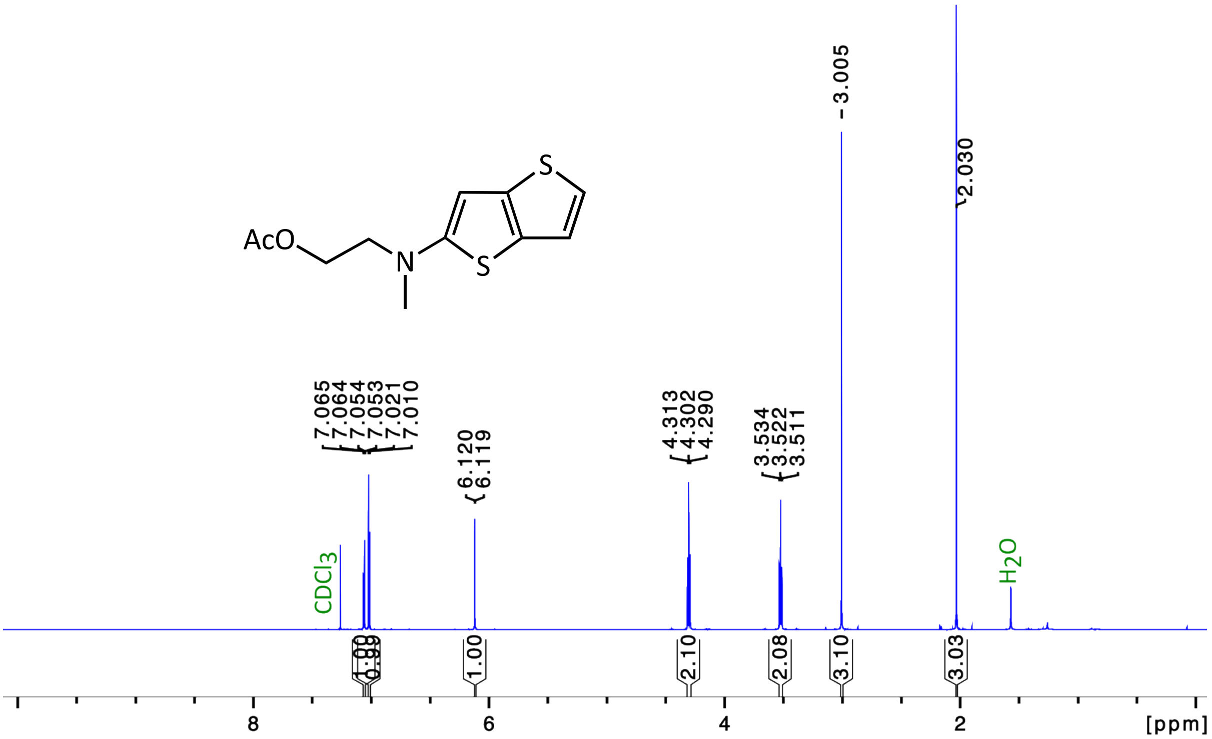


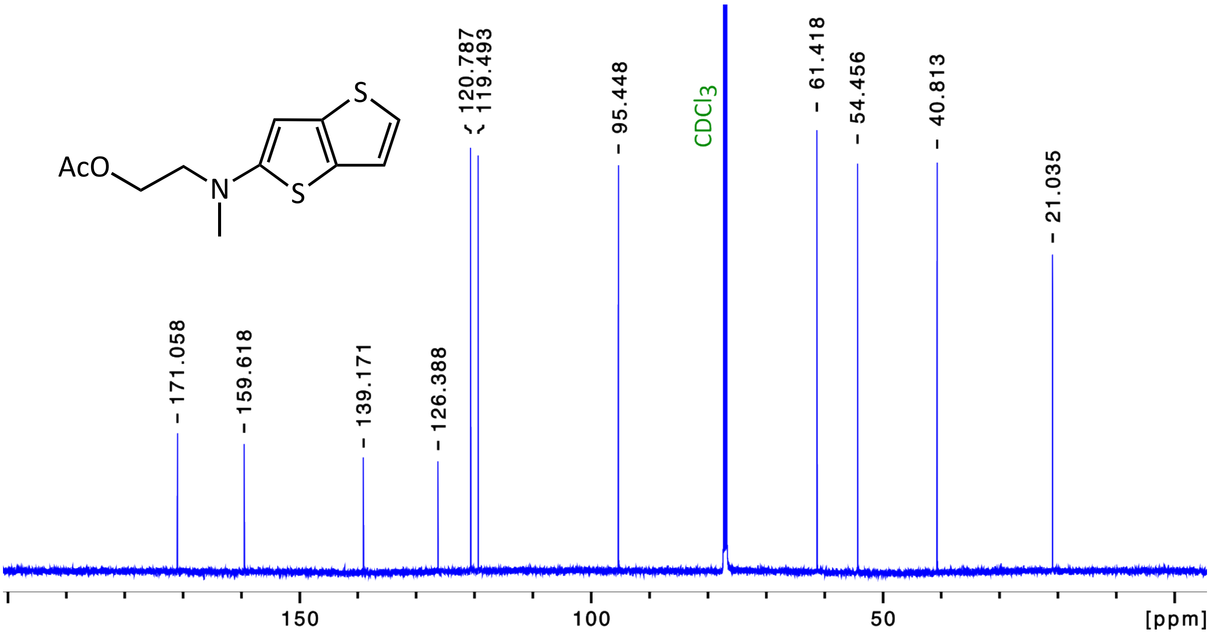




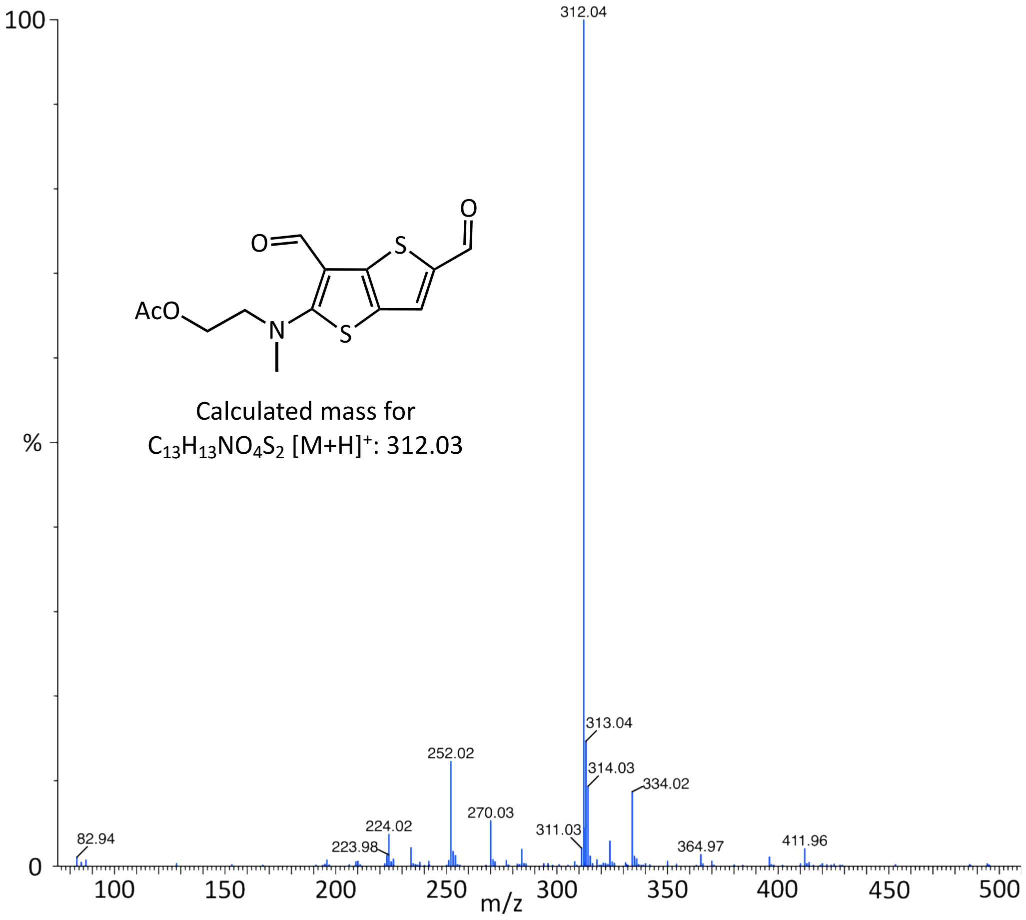
Calculated mass for
$C_9H_{12}NOS_2$ $[M+H]^+$: 214.04

^{1}H-NMR and ^{13}C-NMR of 5-[(2-acetoxyethyl)methylamino]thieno[3,2-*b*]thiophene (5)





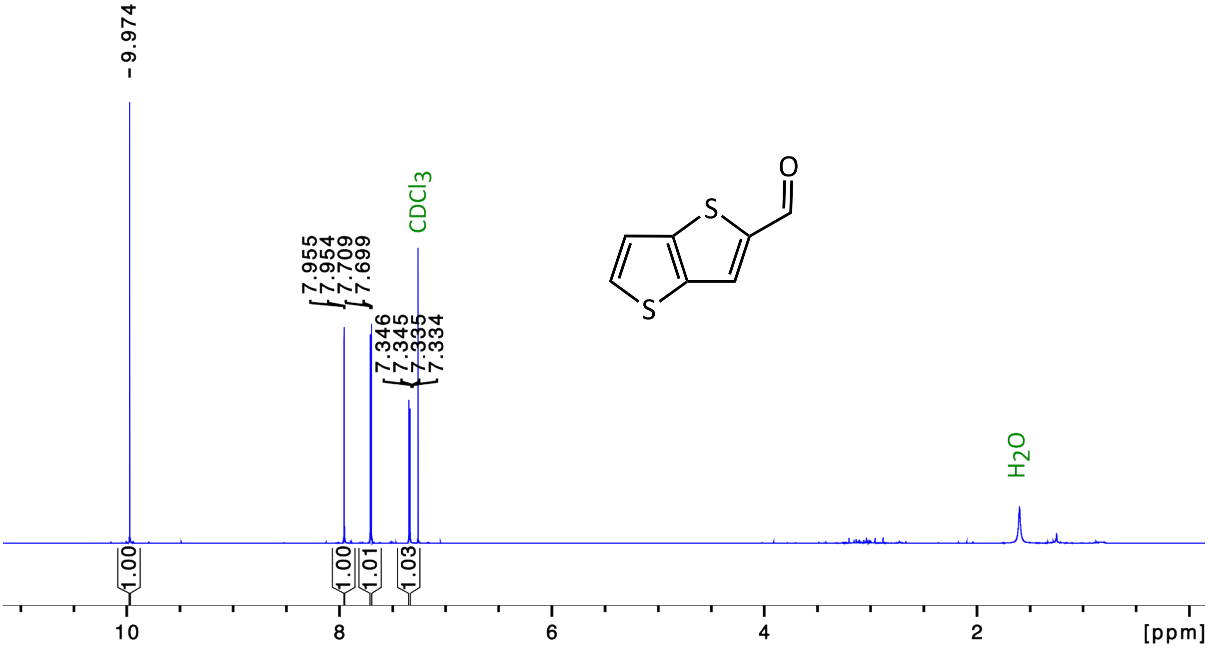
MS of 5-[(2-acetoxyethyl)methylamino]thieno[3,2-*b*]thiophene-2,4-dicarbaldehyde (11)

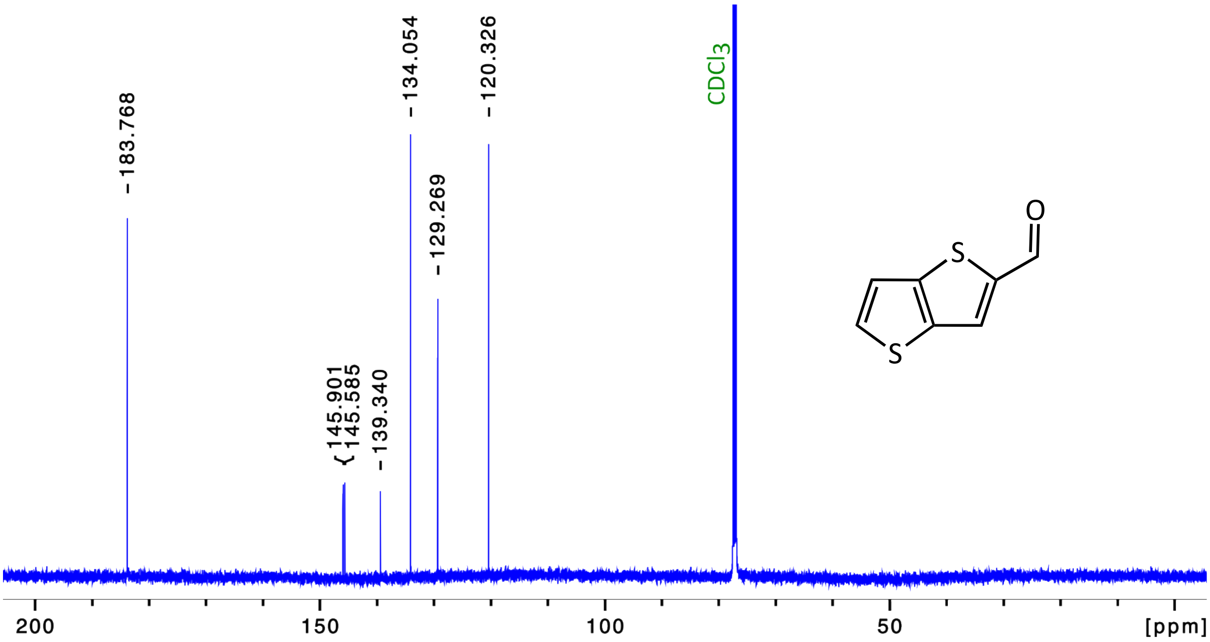


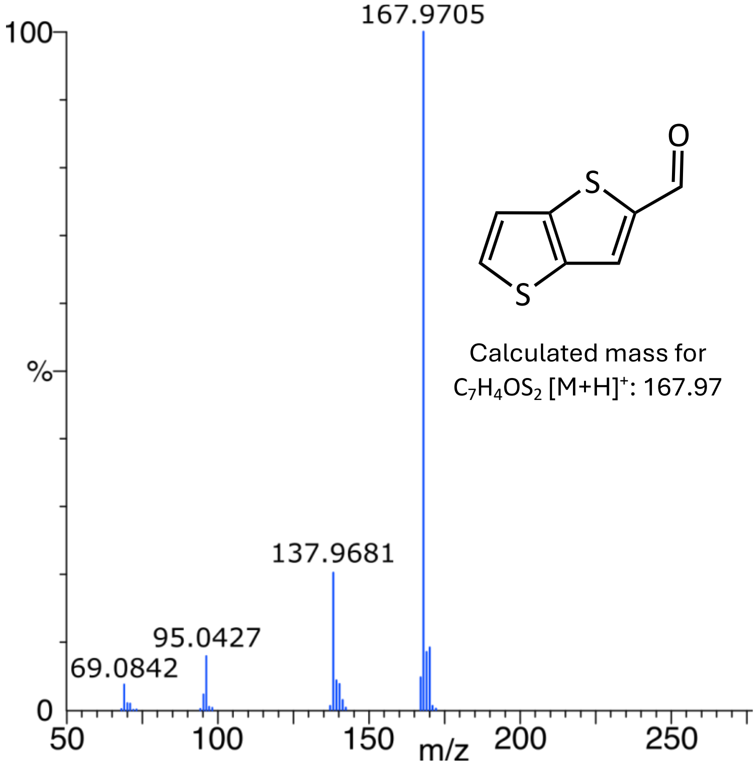
Calculated mass for $C_{13}H_{13}NO_4S_2$ $[M+H]^+$: 312.03

Characterisation of compounds from the improved synthetic route

^{1}H-NMR and ^{13}C-NMR of thieno[3,2-b]thiophene-2-carbaldehyde (12)

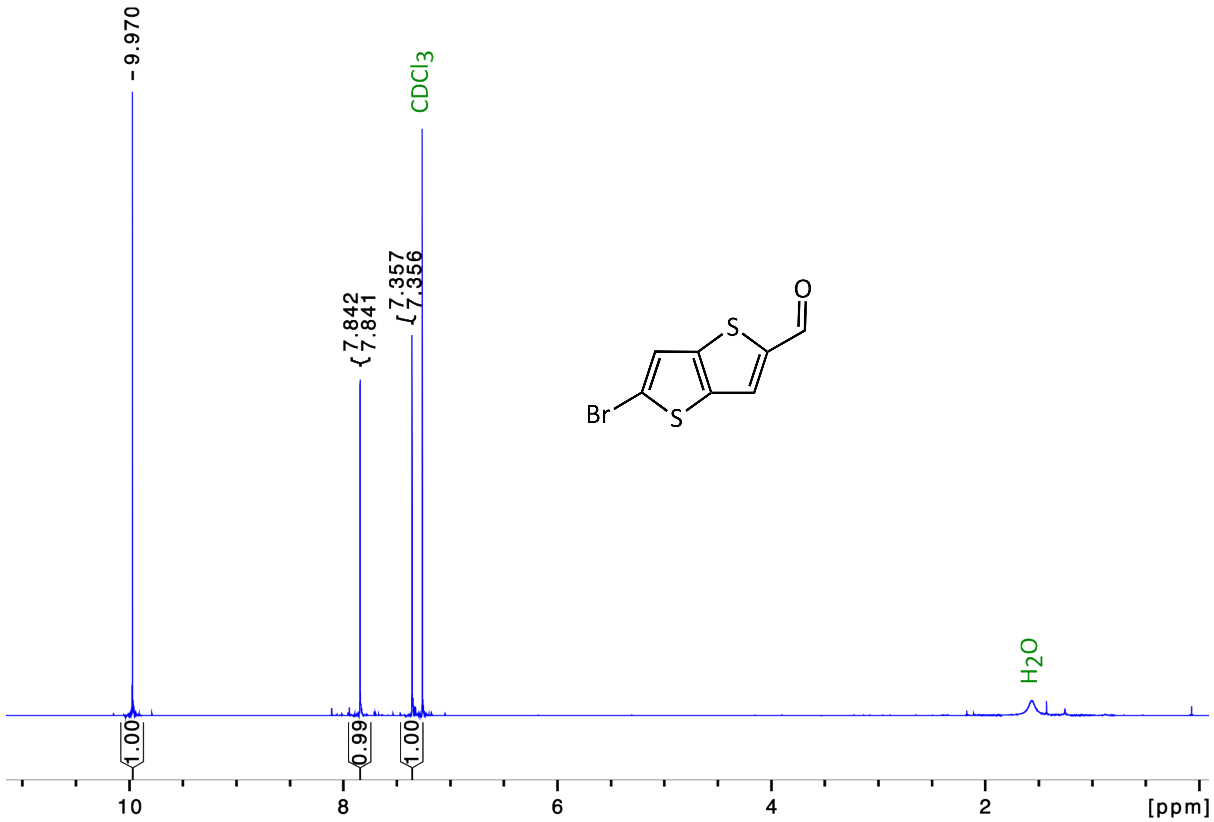


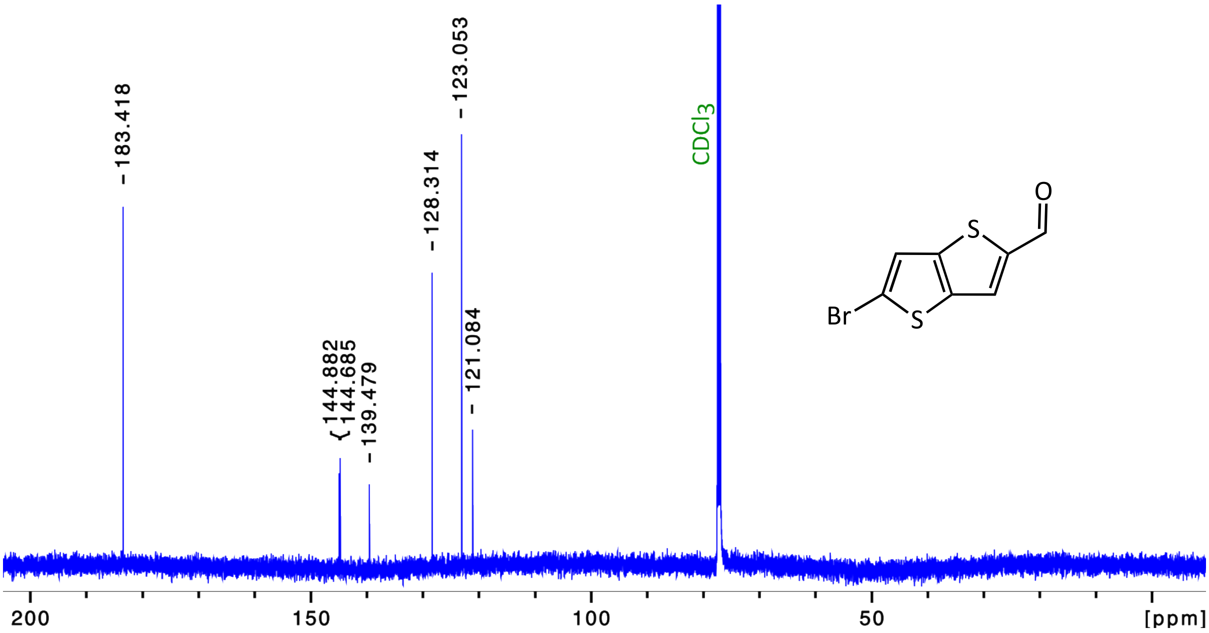


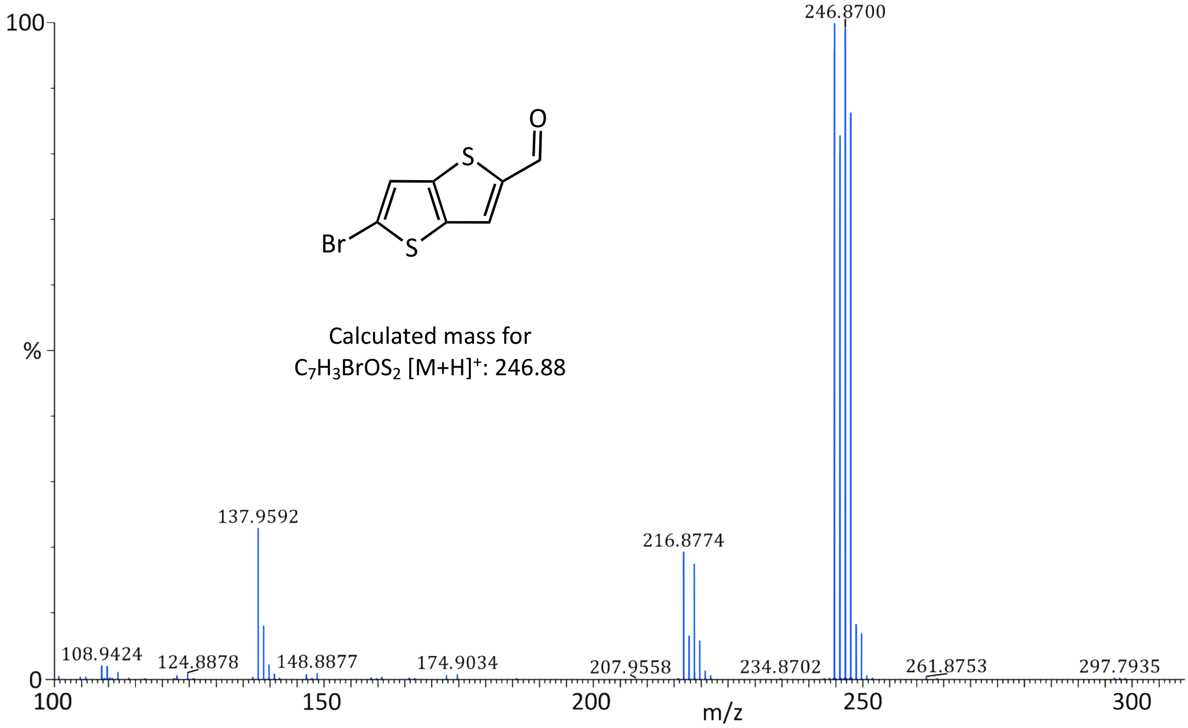


Calculated mass for $C_7H_4OS_2$ $[M+H]^+$: 167.97

^{1}H-NMR, ^{13}C-NMR and MS of 5-bromothieno[3,2-b]thiophene-2-carbaldehyde (13)

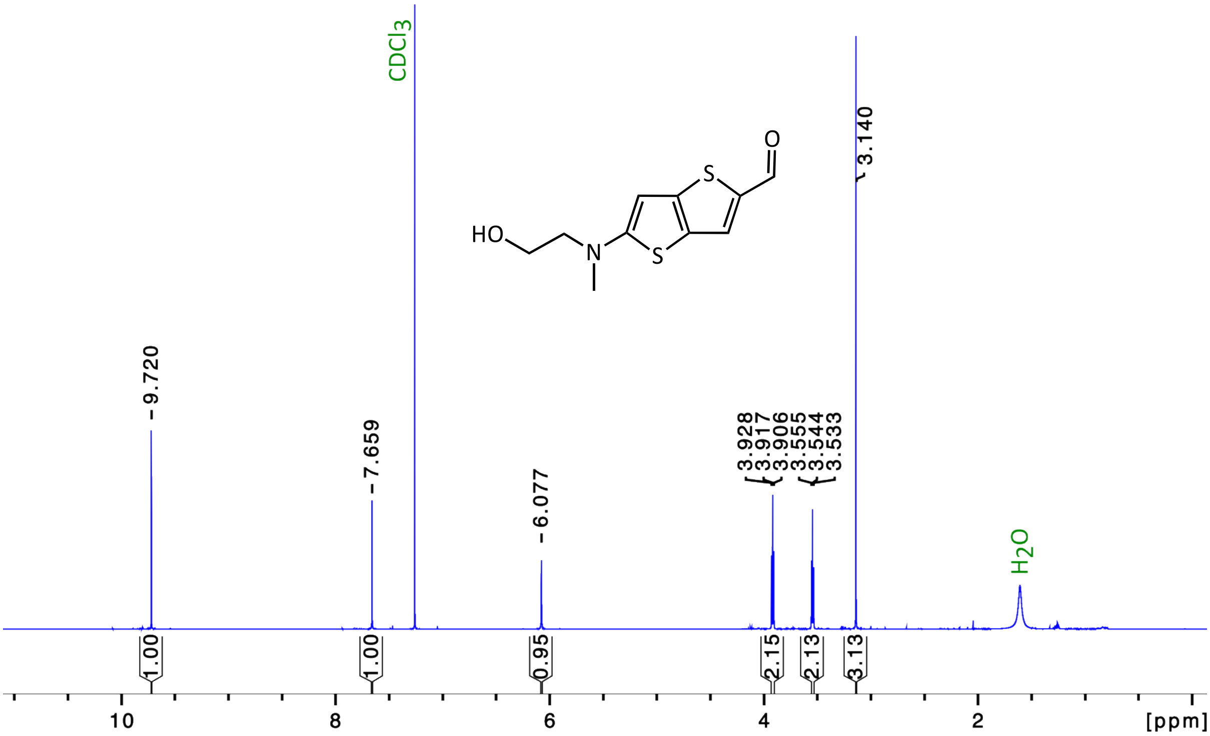


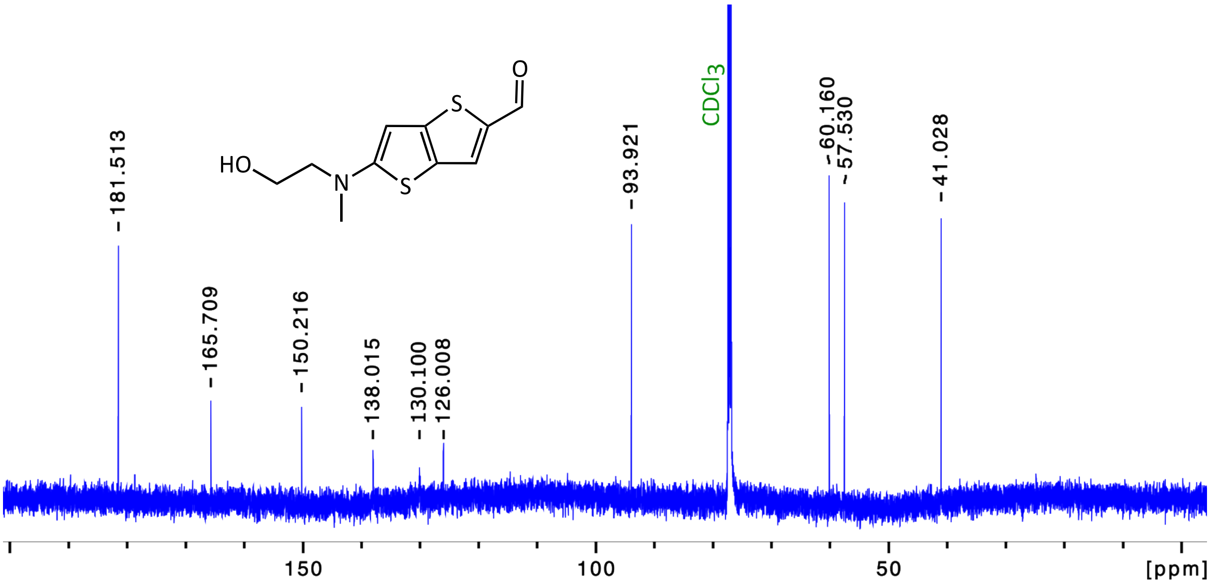


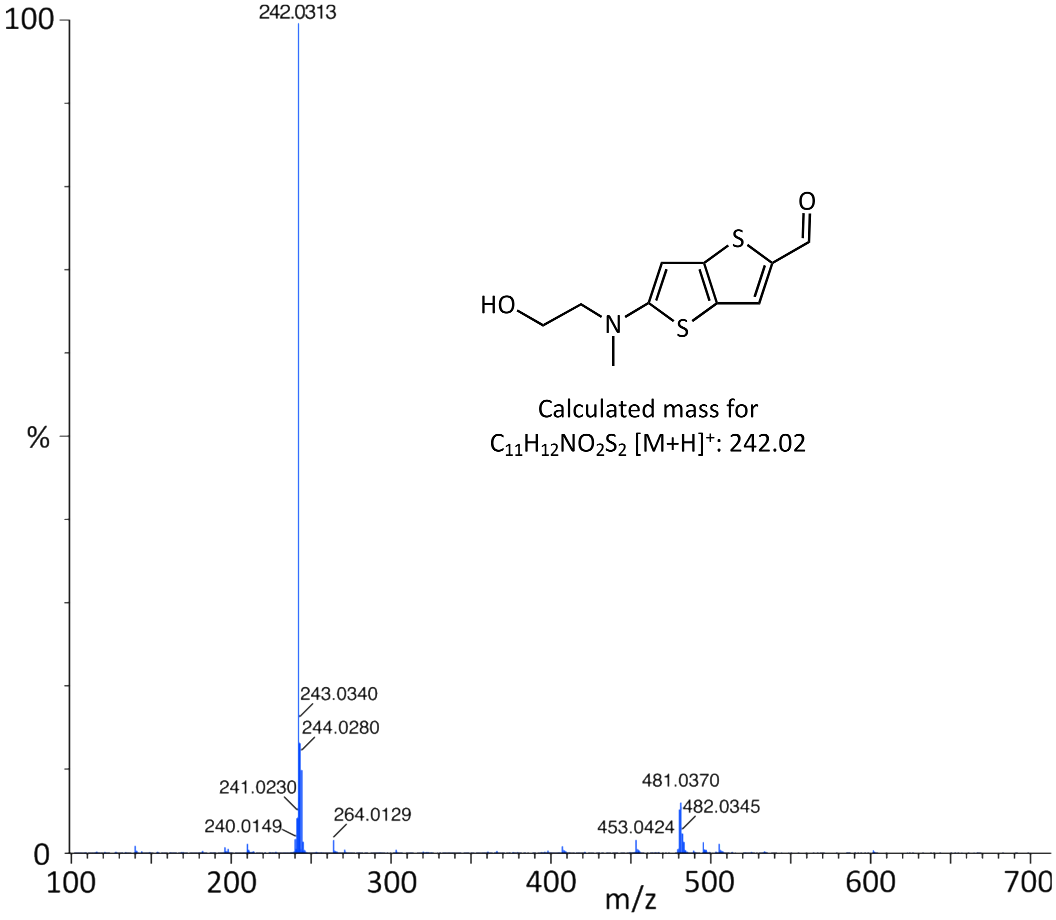


Calculated mass for
$C_7H_3BrOS_2$ $[M+H]^+$: 246.88

^{1}H-NMR, ^{13}C-NMR and MS of 5-[(2-hydroxyethyl)methylamino]thieno[3,2-*b*]thiophene-2-carbaldehyde (7)

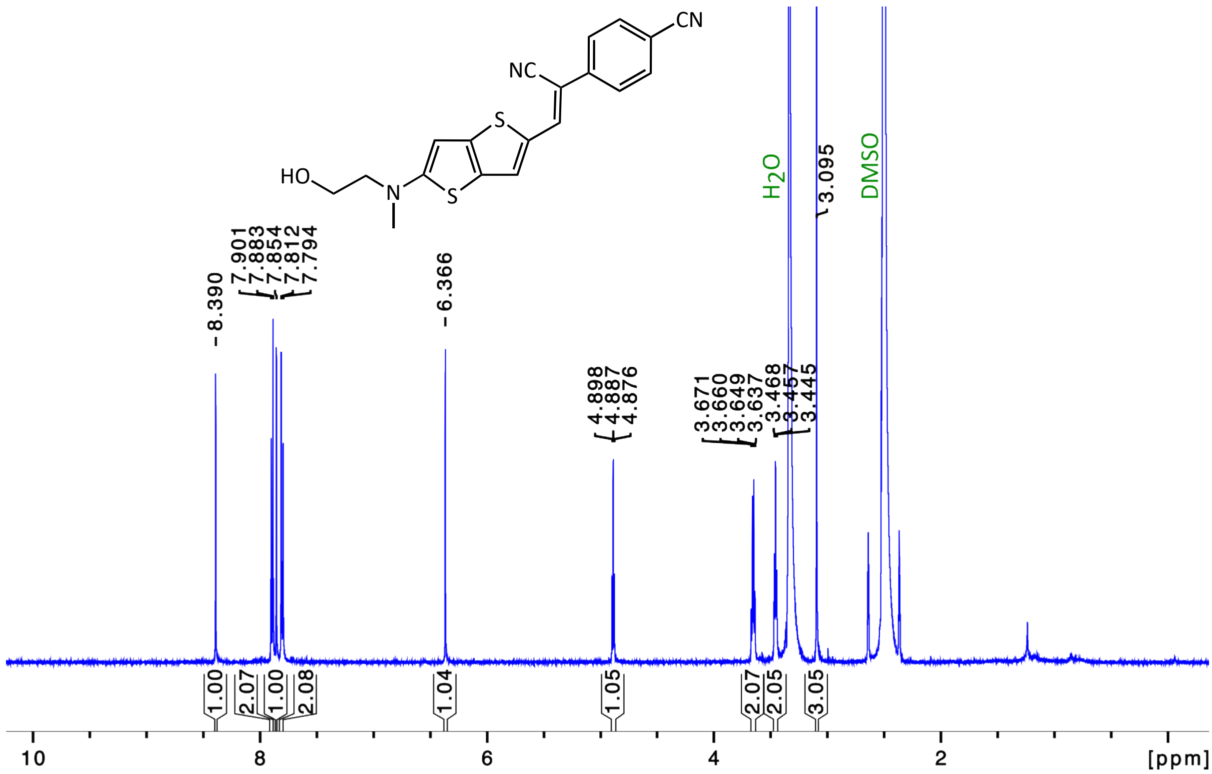


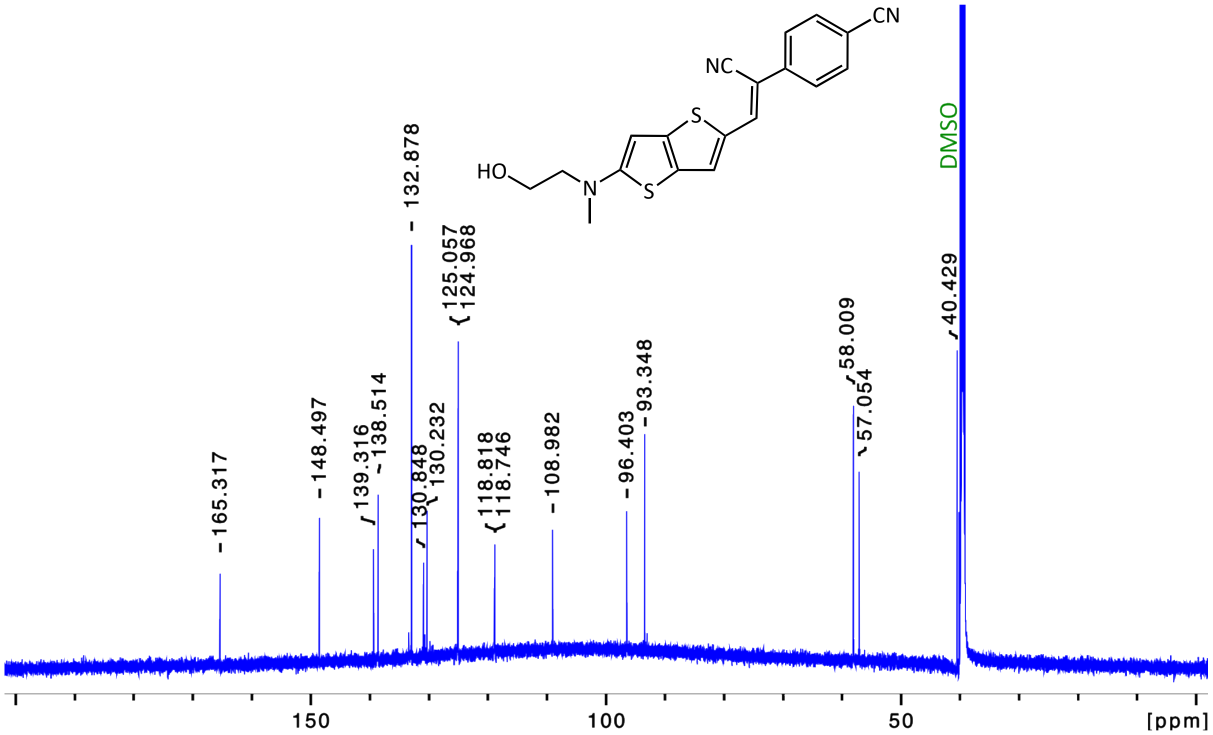


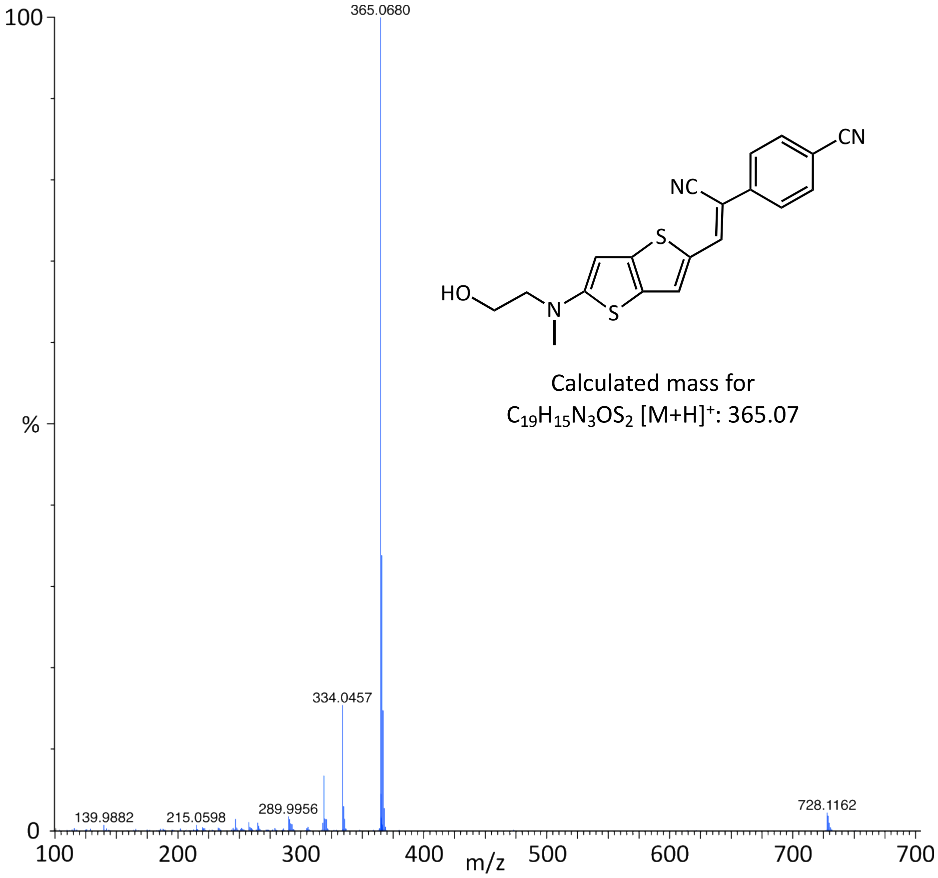


Calculated mass for $C_{11}H_{12}NO_2S_2$ $[M+H]^+$: 242.02

^{1}H-NMR, ^{13}C-NMR and MS of HBC$_{620}$ (9)



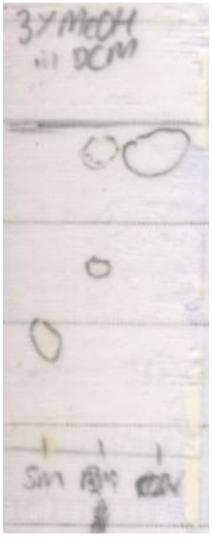




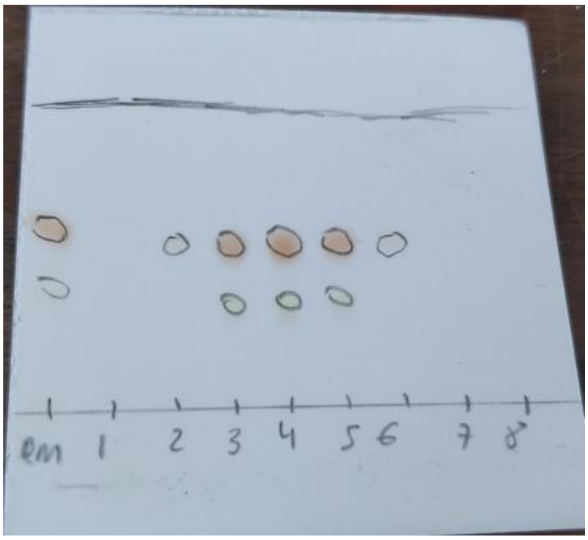
Calculated mass for
$C_{19}H_{15}N_3OS_2$ $[M+H]^+$: 365.07

Purification of HBC$_{620}$ (9)

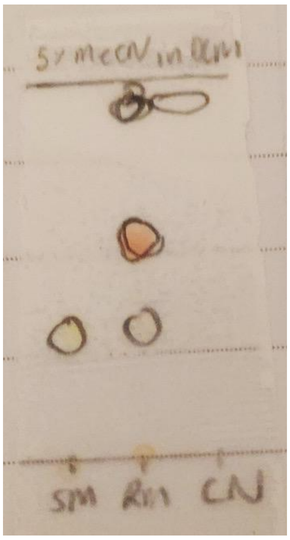
Separation in 2% MeOH in DCM



Coelution in silica column

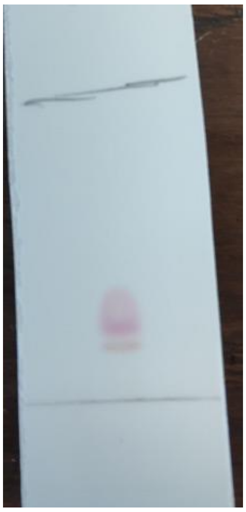


Separation in 5% MeCN in DCM

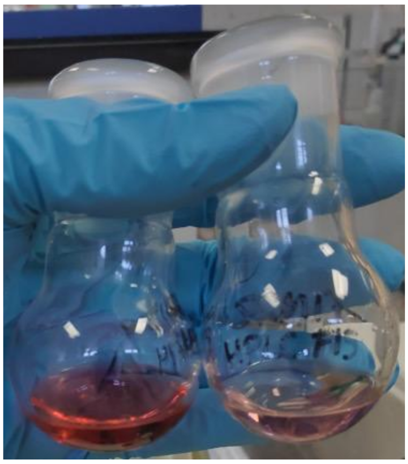


Study of HBC$_{620}$ isomers (9 and 9a)

TLC coated in silver nitrate

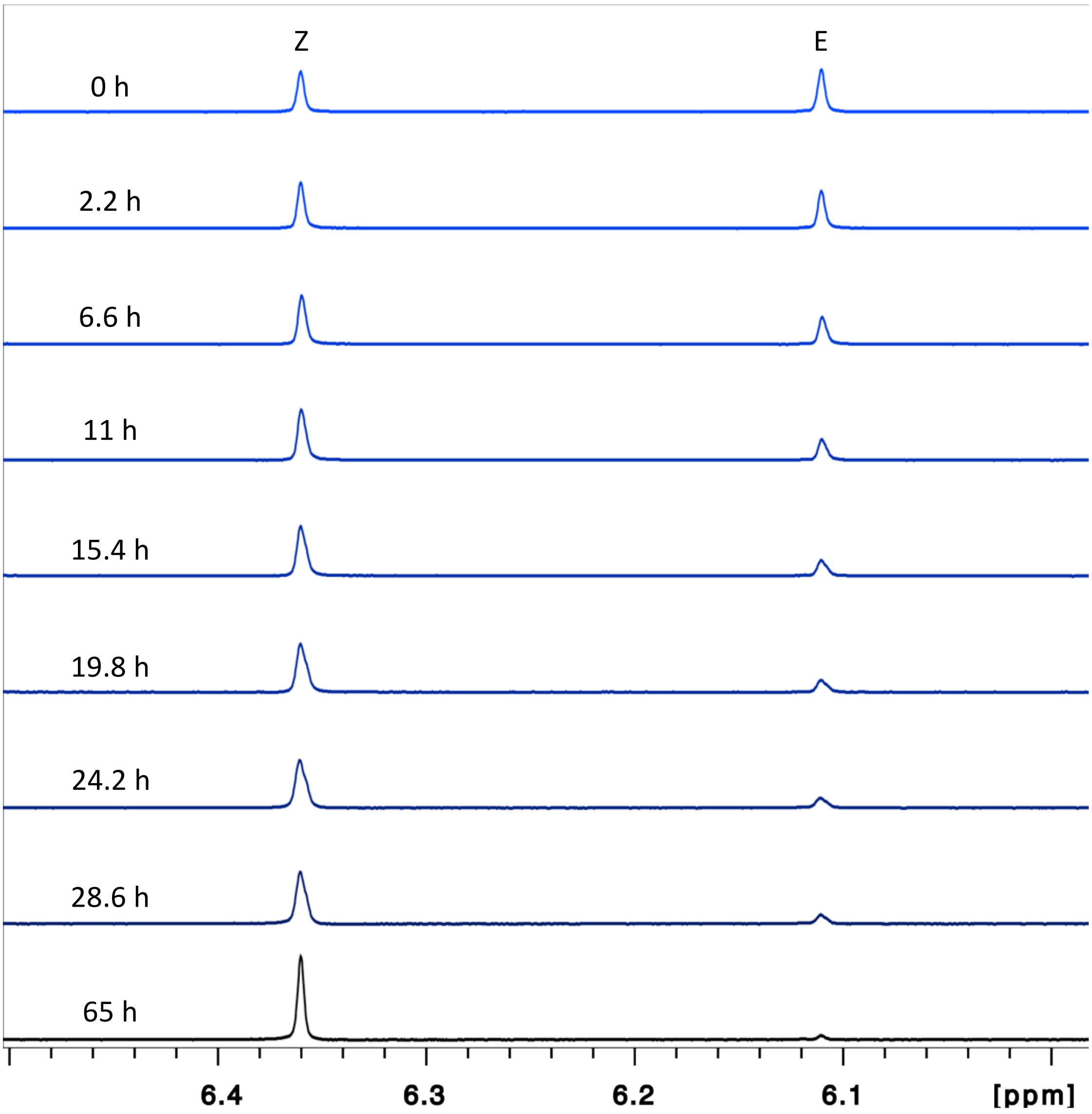


Isomer separation in HPLC

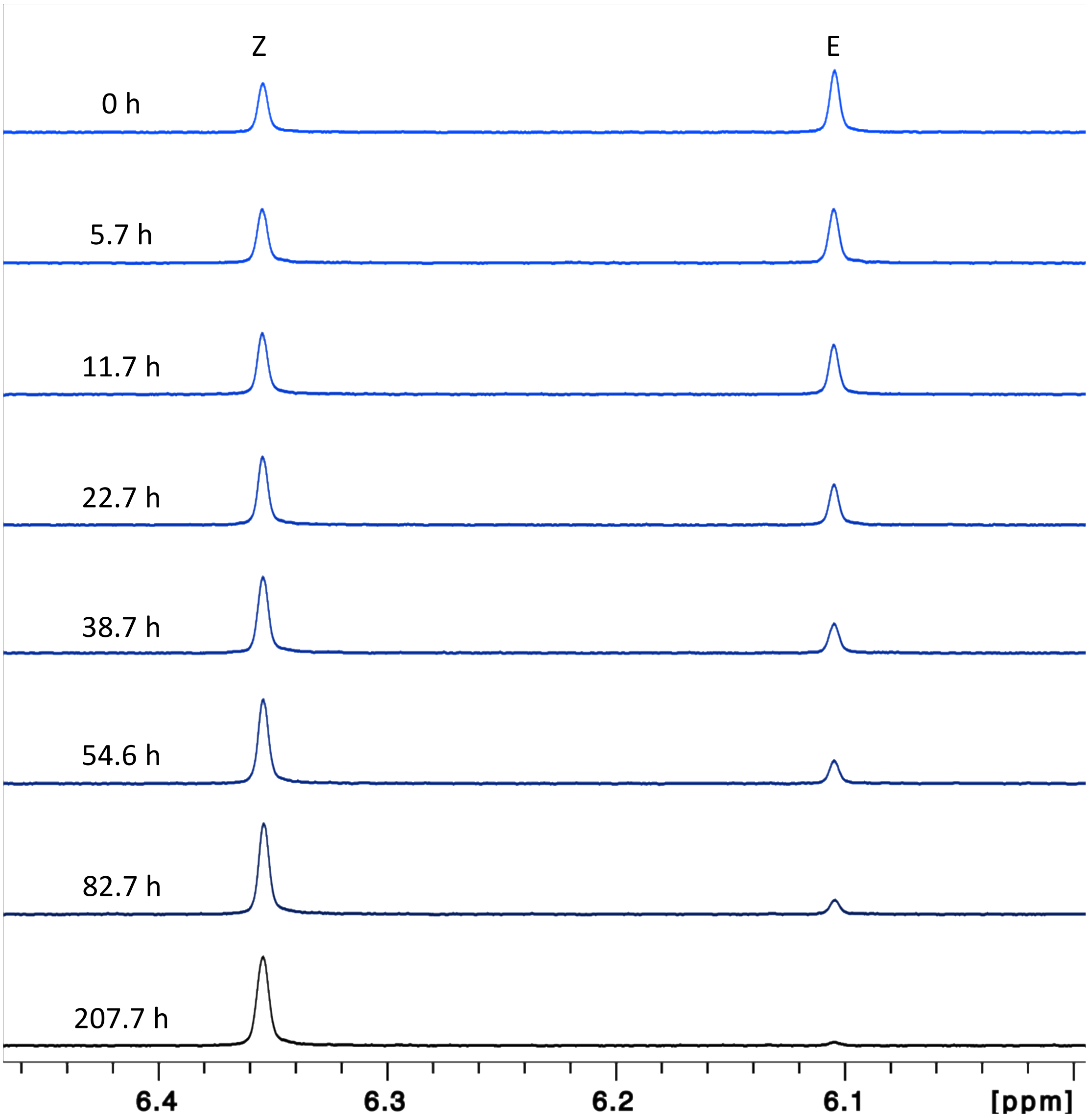


Peak 1 Peak 2

Relaxation of illuminated HBC_{620} to the dark state.

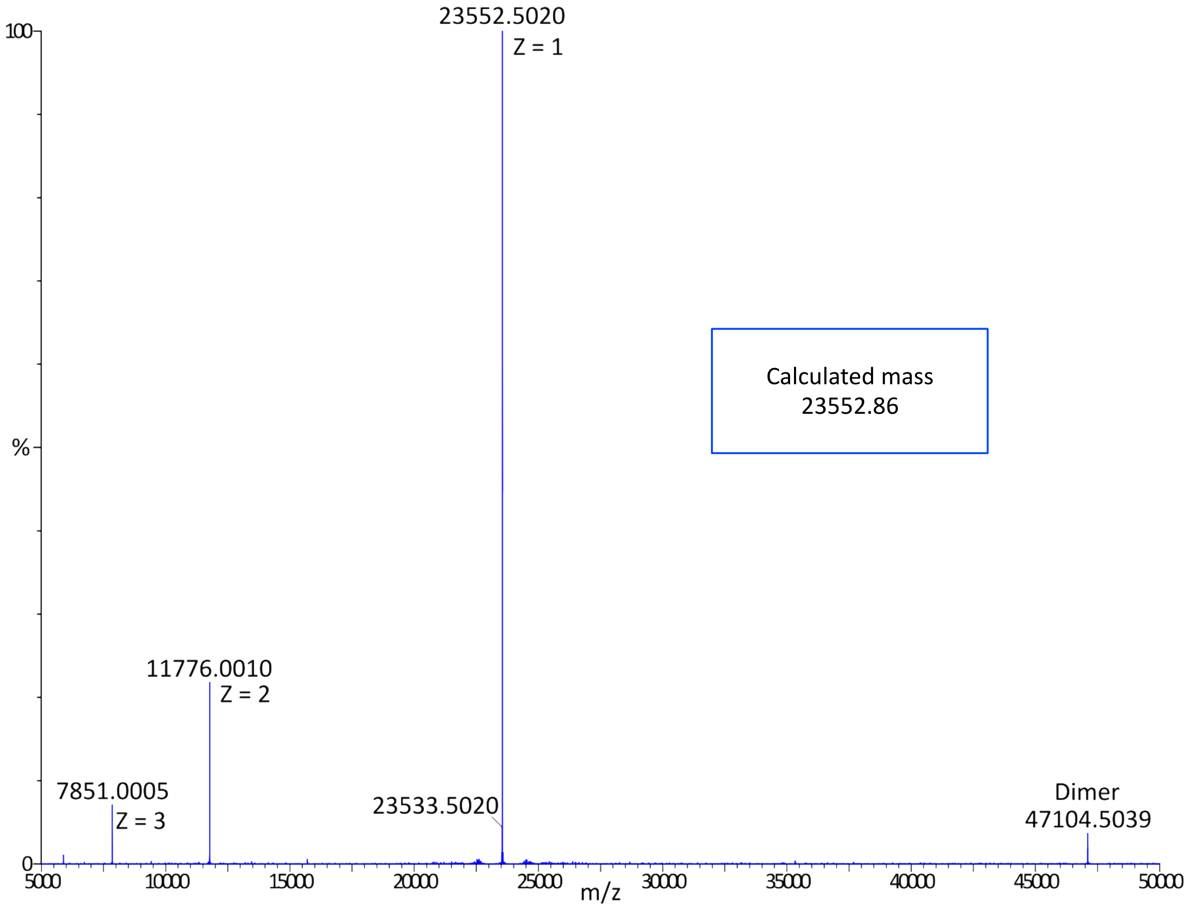


Relaxation of illuminated HBC_{620} to the dark state with acetic acid.



Chapter 3. Engineering the DNA binding domain of EL222 to create photoresponsive transcription factors

Mass spectrum of EL222::CsgD_{R1}



Please note there is a 1 AMU error for this experiment.

Statistical analysis of optimisation of GFP fluorescence

One-way ANOVA followed by Tukey HSD (Honestly Significant Difference) test were performed on the light/dark ratios (Figure 68, Table 4) to analyse the effect of cell strain and growth conditions on GFP fluorescence. This statistical analysis supports the conclusions drawn from the optimization of fluorescence experiments.

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1.179243	5	0.235849	63.44557	3.28E-08	3.105875
Within Groups	0.044608	12	0.003717			
Total	1.223851	17				

The P-value, lower than 0.05, indicated there was statistical significance between the groups. Tukey HSD test was used to determine which factor had statistical significance on fluorescence ratio. There is a statistical significance when the absolute difference of the means is higher than the calculated HSD value.

Tukey HSD test

Groups	6	*Degrees of freedom*	12
Exp per group	3	*q value (table)*	4.75
Total exp	18	*HSD*	0.167205

Comparison	*Mean difference*	*Significance*	*Effect*
C1-C2	0.122687194	no	Effect of lysozyme on DH5α at 37°C
C1-C3	0.184018883	yes	Effect of cell strain at 37°C
C3-C4	0.056315686	no	Effect of lysozyme on BL21* at 37°C
C1-C5	0.349418532	yes	Effect of temperature on DH5 α
C3-C6	0.629265094	yes	Effect of temperature on BL21*
C5-C6	0.794664743	yes	Effect of cell strain at 20°C

Two-way ANOVA on biofilm formation assays

A two-way ANOVA was performed to analyse the effect of light and EL222 variant on biofilm formation. This statistical analysis supports the conclusions drawn from the biofilm assays.

WT EL222

There was not a statistically significant interaction between the effects of light and EL222 variant ($p = 0.642$, $p > 0.05$). Illumination conditions (sample) did not have a statistically significant effect on biofilm formation ($p = 0.209$, $p > 0.05$). The EL222 variant (columns) had a statistically significant effect on biofilm formation ($p = 8.27 \cdot 10^{-18}$, $p < 0.05$).

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.001599	1	0.001599	1.665746	0.209125	4.259677
Columns	0.587789	2	0.293895	306.2199	8.27E-18	3.402826
Interaction	0.000867	2	0.000434	0.451784	0.641794	3.402826
Within	0.023034	24	0.00096			
Total	0.613289	29				

EL222::CsgD

There was not a statistically significant interaction between the effects of light and EL222 variant ($p = 0.637$, $p > 0.05$). Illumination conditions (sample) did not have a statistically significant effect on biofilm formation ($p = 0.154$, $p > 0.05$). The EL222 variant (columns) had a statistically significant effect on biofilm formation ($p = 2.64\cdot10^{-12}$, $p < 0.05$).

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.006208	1	0.006208	2.215225	0.153966	4.413873
Columns	0.925418	2	0.462709	165.1058	2.64E-12	3.554557
Interaction	0.002591	2	0.001295	0.462191	0.637172	3.554557
Within	0.050445	18	0.002803			
Total	0.984662	23				

EL222::CsgD$_{R1}$

There was a statistically significant interaction between the effects of light and EL222 variant ($p = 2.06\cdot10^{-4}$, $p < 0.05$). Illumination conditions (sample) had a statistically significant effect on biofilm formation ($p = 2.82\cdot10^{-3}$, $p < 0.05$). The EL222 variant (columns) had a statistically significant effect on biofilm formation ($p = 5.53\cdot10^{-8}$, $p < 0.05$).

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.235853	1	0.235853	11.07228	0.002816	4.259677
Columns	1.546458	2	0.773229	36.29969	5.53E-08	3.402826
Interaction	0.52584	2	0.26292	12.34294	0.000206	3.402826
Within	0.51123	24	0.021301			
Total	2.819381	29				

EL222::CsgD$_{R2}$

There was not a statistically significant interaction between the effects of light and EL222 variant ($p = 0.848$, $p > 0.05$). Illumination conditions (sample) did not have a statistically significant effect on biofilm formation ($p = 0.879$, $p > 0.05$). The EL222 variant (columns) had a statistically significant effect on biofilm formation ($p = 3.89\cdot10^{-7}$, $p < 0.05$).Table

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.000644	1	0.000644	0.02366	0.87904	4.259677
Columns	1.581695	2	0.790848	29.05297	3.89E-07	3.402826
Interaction	0.009035	2	0.004518	0.165962	0.848042	3.402826
Within	0.653301	24	0.027221			
Total	2.244676	29				

EL222::CsgD$_{RR}$

There was not a statistically significant interaction between the effects of light and EL222 variant ($p = 0.485$, $p > 0.05$). Illumination conditions (sample) did not have a statistically significant effect on biofilm formation ($p = 0.080$, $p > 0.05$). The EL222 variant (columns) had a statistically significant effect on biofilm formation ($p = 8.12 \cdot 10^{-23}$, $p < 0.05$).

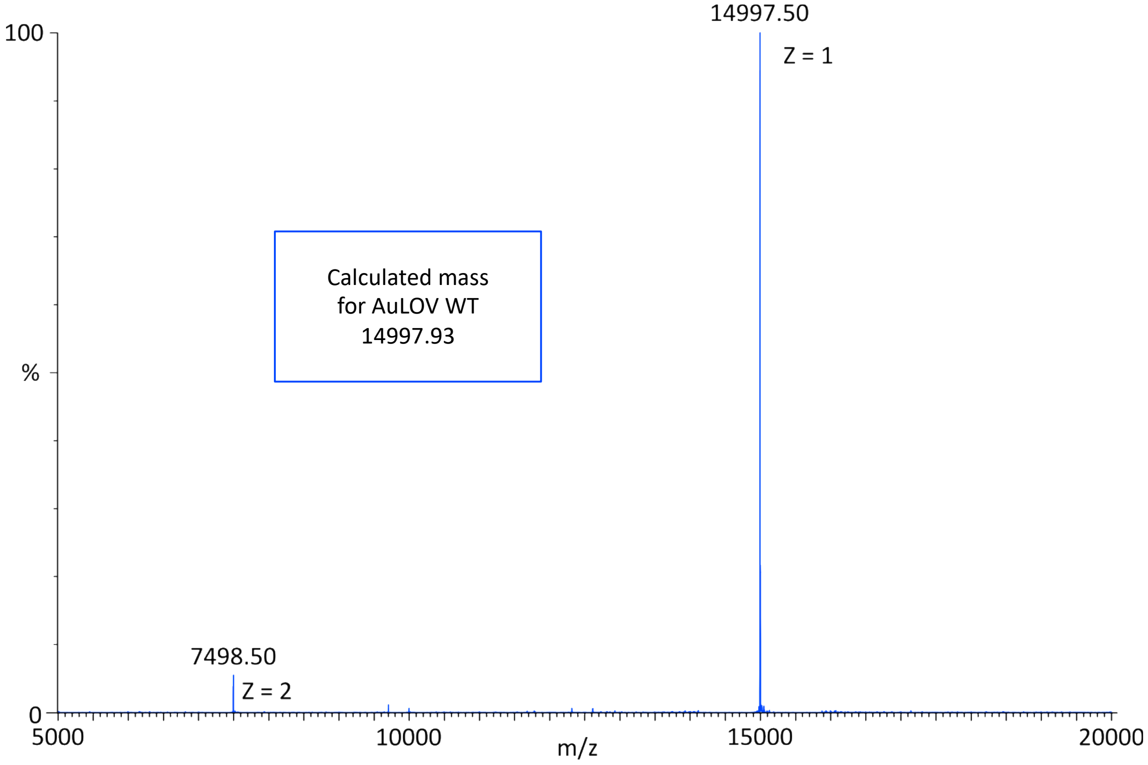
ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.002952	1	0.002952	3.279435	0.080178	4.170877
Columns	0.774982	2	0.387491	430.455	8.12E-23	3.31583
Interaction	0.001334	2	0.000667	0.741079	0.485122	3.31583
Within	0.027006	30	0.0009			
Total	0.806274	35				

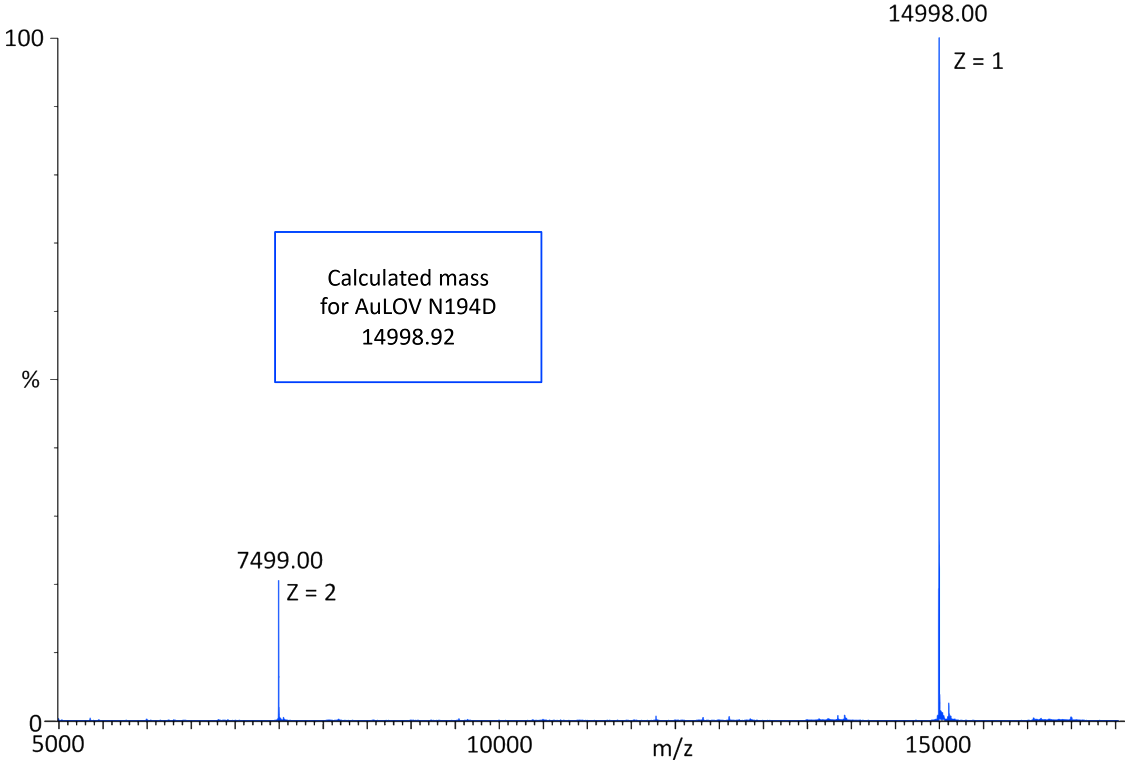
Chapter 4. Elucidating the role of Asn$_{194}$ in aureochrome LOV domain signal transduction

Mass spectra of the AuLOV variants

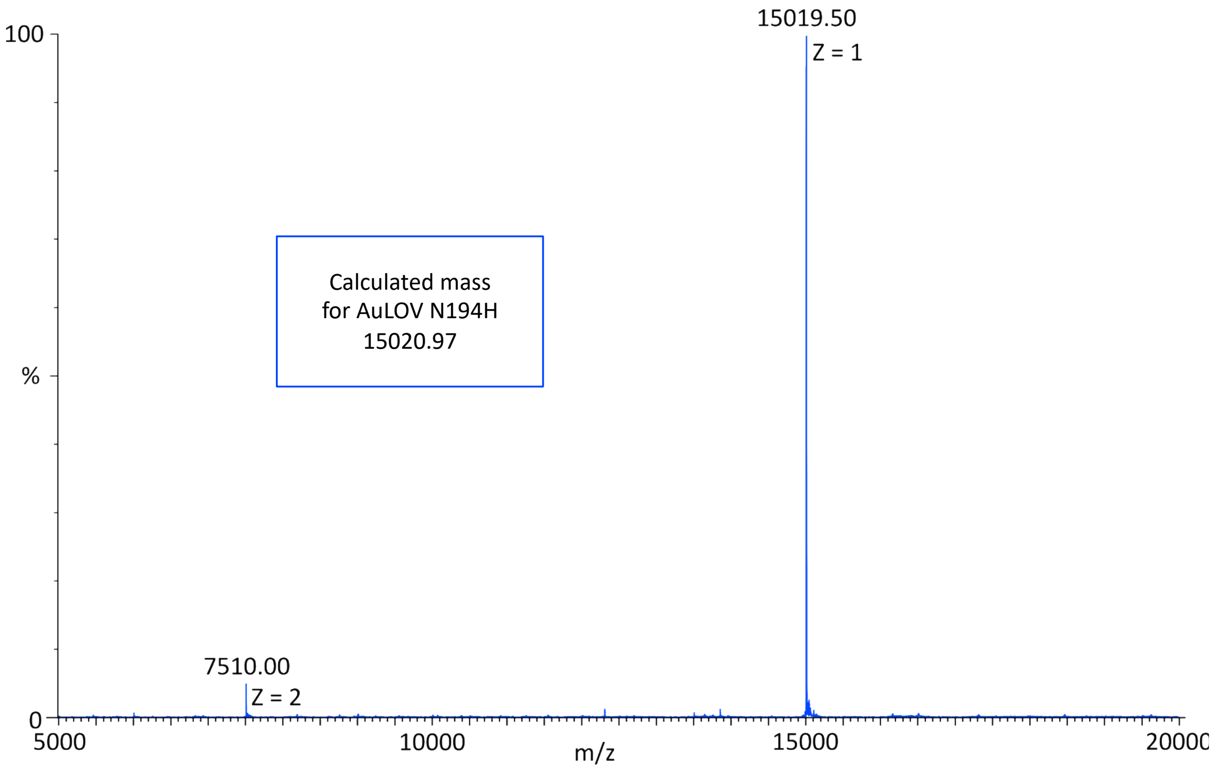
AuLOV WT



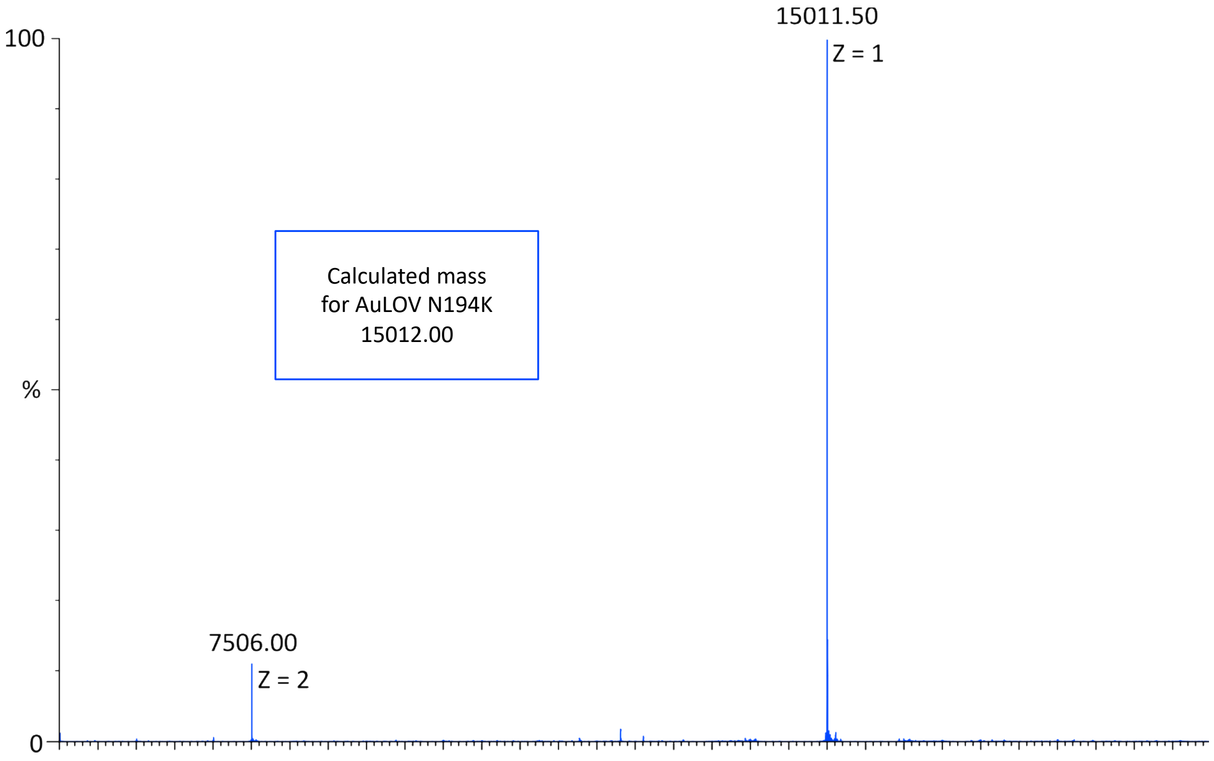
AuLOV N194D



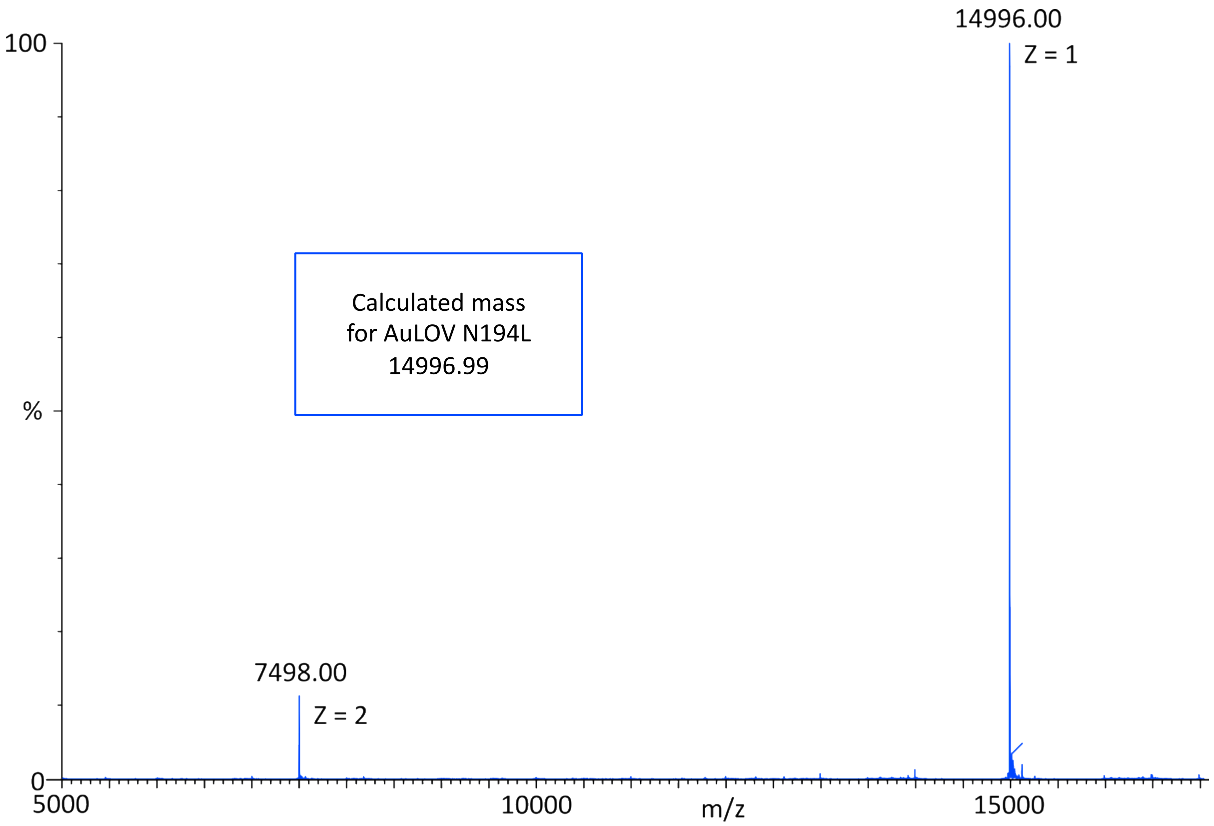
AuLOV N194H



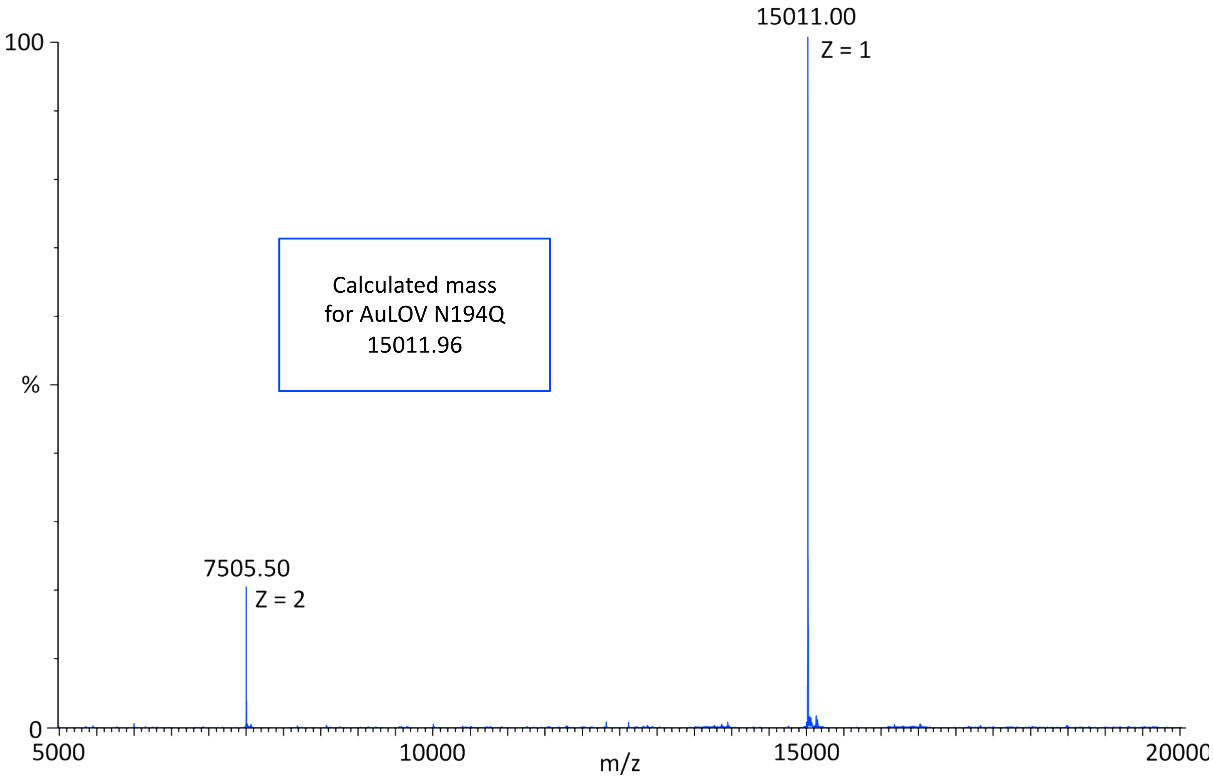
AuLOV N194K



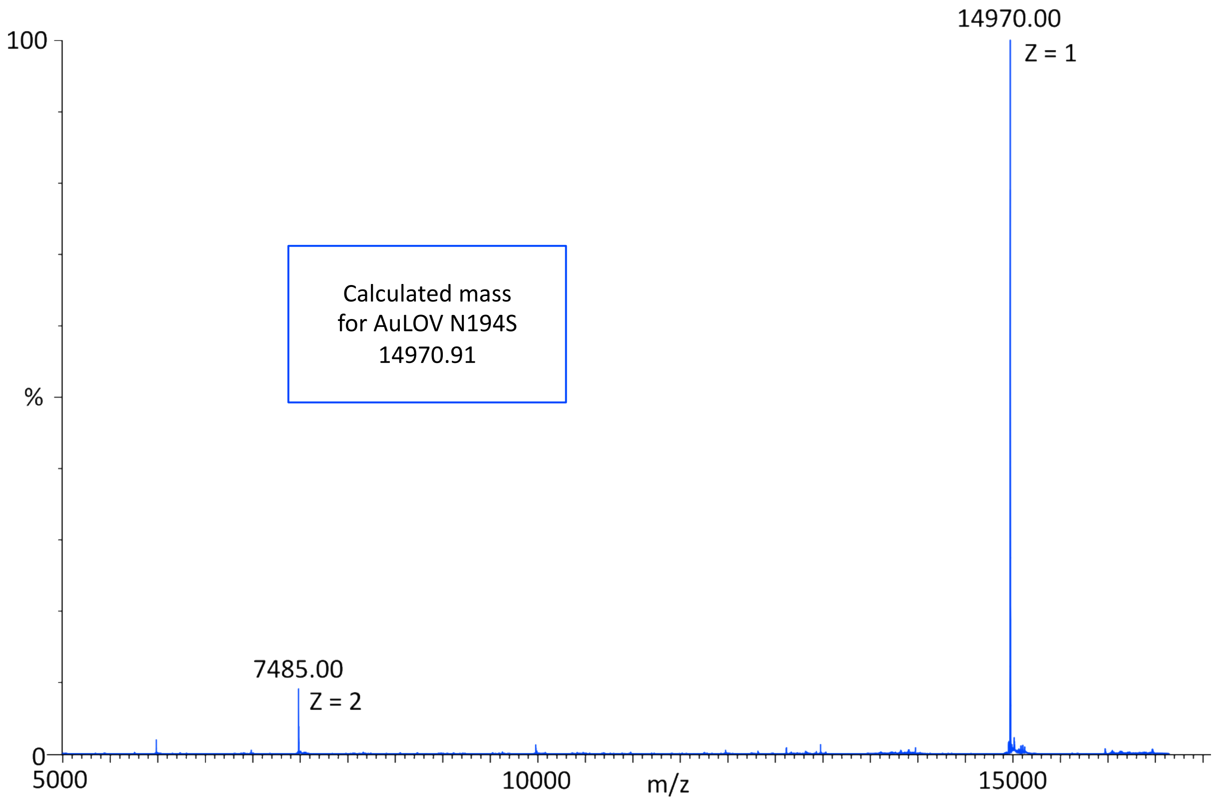
AuLOV N194L



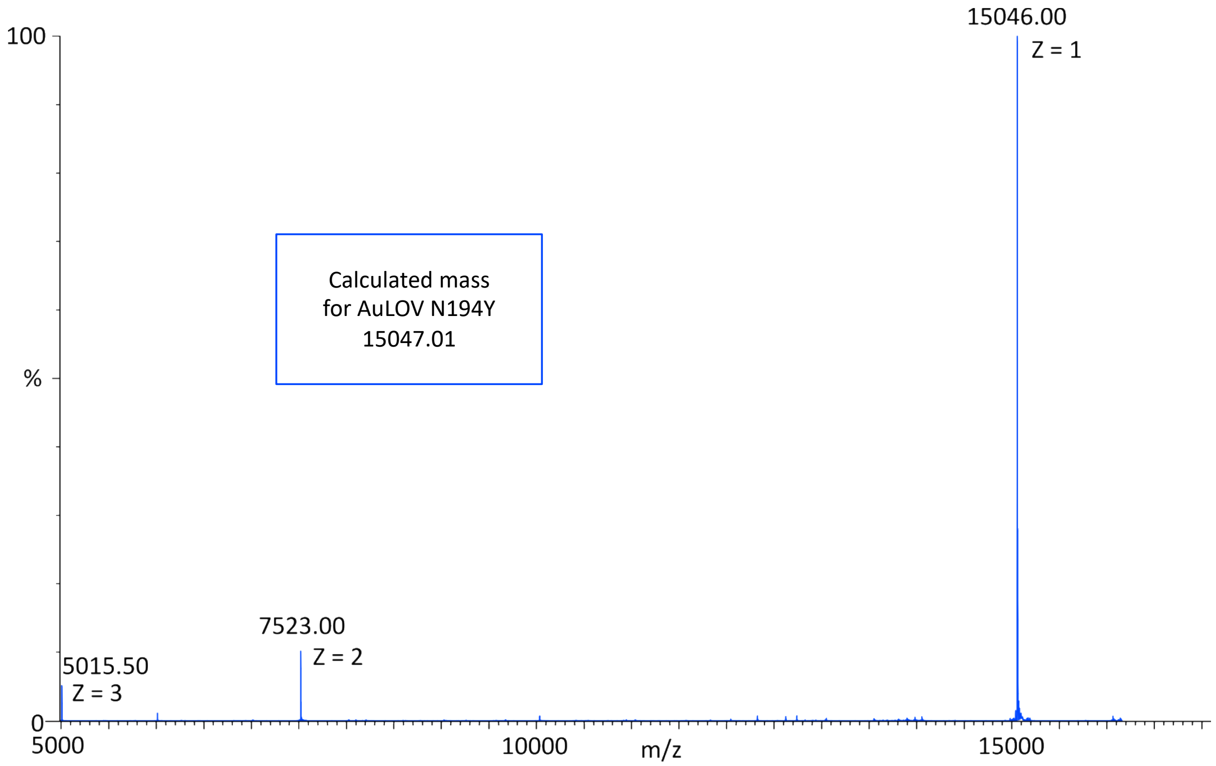
AuLOV N194Q



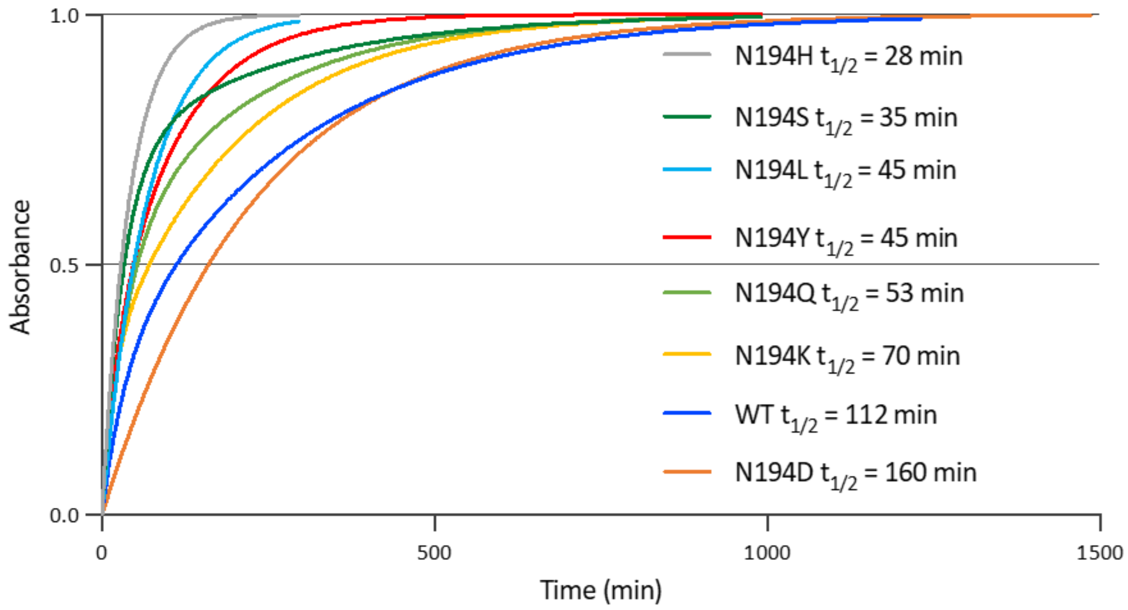
AuLOV N194S



AuLOV N194Y



Complete UV-vis spectra of AuLOV variants relaxation followed at 450 nm



Estimated masses of AuLOV variants from calibration curve

Variant	**V_e Light**	**MW light**	**V_e Dark**	**MW dark**	**ΔV_e**	**ΔMW**
N194K	11.90	44447	12.01	42851	0.11	1596
N194S	11.67	47979	11.83	45493	0.15	2486
N194L	11.52	50433	11.71	47345	0.19	3087
N194D	11.62	48783	11.82	45645	0.19	3139
N194Q	11.30	54260	11.71	47345	0.41	6915
WT	11.34	53721	11.96	43569	0.63	10153
N194Y	11.02	59554	11.64	48460	0.62	11094
N194H	11.01	59752	11.74	46875	0.72	12877

The apparent masses obtained using the calibration curve are different to the theoretical masses of AuLOV monomer (~15 kDa) and dimer (~30 kDa). This can be explained by the use of different buffers due to the incompatibility of the protein standard mix with AuLOV assay buffer.

Table of parameters of crystal structures

	N194D dark	**N194D illuminated**	**N194D* illuminated**
Data Collection			
Diamond Beamline	i24	i24	i24
Wavelength (Å)	0.6199	0.6199	0.9763
Crystallisation conditions	0.1 M NaOAc 0.1 M NH_4Cl 16% PEG 2000 pH 4.5	0.1 M NaOAc 0.1 M NH_4Cl 16% PEG 2000 pH 4.5	0.1 M NaOAc 0.1 M NH_4Cl 18% PEG 2000 pH 4.8
Crystal Data			
a, b, c (Å)	105.79, 105.79, 67.18	106.58, 106.58, 66.98	104.01, 104.01, 66.90
h, k, l	90.0, 90.0, 120.0	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Space group	$P\ 6_2\ 2\ 2$	$P\ 3_1\ 2\ 1$	$P\ 3_1\ 2\ 1$
Resolution (Å)	54.10 – 1.81	38.00 – 2.06	66.90 – 1.26
Outer shell	1.84 – 1.81	2.10 – 2.06	1.28 – 1.26
R_{merge}	0.159 (10.419)	0.172 (4.350)	0.049 (1.552)
R_{meas}	0.160 (10.483)	0.176 (4.457)	0.051 (1.690)
$CC_{1/2}$	1.000 (0.313)	0.999 (0.319)	0.999 (0.377)
$I/\sigma i$	11.5 (0.1)	11.95 (0.74)	17.6 (1.0)
Completeness (%)	100.0 (100)	100.0 (100)	99.38 (88.99)
Multiplicity	78.8 (82.0)	20.55 (21.02)	10.5 (6.4)
Total measurements	1635505 (83371)	564714 (29242)	1169714 (31568)
Unique reflections	20757 (1017)	27477 (1391)	111689 (4949)
Refinement			
Non-H atoms	1084	2148	2100
R_{work} reflections	17955	27453	111648
R_{free} reflections	1736	1397	5522
R_{work}/R_{free}	0.210/0.239	0.192/0.254	0.174 / 0.190
Rms deviations			
Bond lengths (Å)	0.0138	0.0110	0.0159
Bond angles ()	2.483	1.82	2.127
Mean B value ($Å^2$)	55.3	52.51	17.4
Ramachandran outliers			
Favoured	(96.99 %)	(99.25 %)	251 (98.05 %)
Allowed	(3 %)	1 (0.37 %)	4 (1.56 %)
Outliers	0 (0%)	1 (0.38 %)	1 (0.39 %)