

Cardiff University
School of Biosciences



School of Biosciences

Ysgol y Biowyddorau

**Hybrid Beetle Luciferases derived from Firefly Luciferases and Murine
Medium-Chain Fatty Acyl-CoA Synthetases for In Vivo Imaging**

James Alexander Long

16th of November 2020 (initial submission)

19th of May 2021 (resubmission – with minor corrections)

Cardiff School of Biosciences
The Sir Martin Evans Building
Museum Avenue
CF10 3AX

A thesis submitted for the degree of Doctor of Philosophy for Cardiff University
(~41,500 words)

Abstract

The firefly luciferase (Fluc) bioluminescence system has been widely studied and used by biomedical researchers for in vitro assays as well as a wide variety of optical imaging applications in small animals. However, Fluc is immunogenic to mice which limits its applications, especially in longitudinal experiments requiring time courses extending over more than four weeks.

Murine acyl-CoA synthetases (ACSs) are structurally and functionally similar to Flucs and as native mouse proteins are inherently non-immunogenic. Both ACS and Fluc enzymes are composed of N and C-terminal domains (NTD and CTD).

A strategy of hybridisation between *Photinus pyralis* (*Ppy*) Fluc and several medium-chain ACSs (ACSMs) was employed with the aim of creating hybrids with minimal immunogenicity whilst maintaining brightness when ultimately expressed in BALB/c mice.

Several hybrid variants were created with regions of importance substituted into an ACSM backbone, but all the constructs tested failed to produce any bioluminescence when expressed in *E. coli*.

Hybrids created containing the NTD of *Ppy* Fluc and the CTD of murine ACSMs 1, 2, and 3 were named PpA1, PpA2, and PpA3 respectively. Light emission of these hybrids was respectively ~0.9%, ~0.004%, and ~0.2% that of WT *Ppy* Fluc when expressed in *E. coli*, but PpA1 was ~10% and PpA3 was ~0.2% as bright as when expressed in HEK cells. The PpA1 hybrid was selected as the best candidate and was expressed in immunocompromised NSG™ mice, producing ~45% the light output of WT *Ppy* Fluc.

Further modifications to PpA1 included a construct which replaced a major epitope in the *Ppy* Fluc NTD (G160 – V168) of PpA1 with the equivalent sequence from a highly similar luciferase with lower predicted immunogenicity; when expressed in HEK cells this construct was marginally brighter than PpA1, emitting light with λ_{\max} 10 nm red-shifted.

Acknowledgements

This work was initially funded by the National Centre for the Replacement, Refinement & Reduction of Animals in Research (NC3Rs).

I would like to thank the Jim Murray lab research group – of which I was a part – for treating me so kindly, especially Joanne Kilby for always taking the time to answer my questions and provide me with the materials I required.

An enormous amount of gratitude must go to my mentor, Dr Amit Jathoul, for his creating of this PhD project, accepting me as the student, and his guidance and assistance with the planning and execution of experiments; no detail was too fine to escape Amit's attention. I am grateful that – despite his constantly saturated schedule – he was always willing to make time for my questions and requests. I am also grateful for his guidance with the writing and proofreading of this PhD thesis.

A special thank you must go to Prof Jim Murray for his remarkable ability to see clarity in seemingly confounding situations and to help me find a logical path forward. I am also grateful for his guidance with the writing and proofreading of this PhD thesis.

*Further thanks go to Prof Paul Kemp for his moral support with settling into the PhD lifestyle during the first several months of the project; Dr Mark Young for his marking of my various reports throughout the project, and for allowing me to sample his reagents before ordering them in bulk; Lainey Williamson for successfully expressing *Arachnocampa Richardsae* glowworm luciferase; Jack Bate for producing the list of candidate luciferases for epitope substitution; and Harley Worthy for his help with protein modelling.*

Finally, thank you to my supportive family and my wonderful partner Ruth for their encouragement and always being there in times of weakness. Further thanks must go to Ruth for proofreading the final thesis for errors.

Table of Contents

1 Contents

1	Chapter 1 – Introduction.....	1
1.1	A Brief History of Bioluminescence.....	1
1.2	Bioluminescent Systems in Nature	1
1.3	Beetle Luciferase/Luciferin System.....	3
1.3.1	Beetle Luciferin reaction	4
1.3.2	Firefly luciferase (Fluc)	6
1.4	Applications of Fluc.....	7
1.4.1	In vitro applications.....	7
1.4.1.1	ATP detection.....	8
1.4.1.2	Enzyme-linked immunosorbent assays (ELISAs)	8
1.4.2	In vivo applications	8
1.5	Engineered Improvements to the Firefly Luciferase System to date for Mammalian Applications.....	9
1.5.1	Generic protein engineering strategies	9
1.5.1.1	Directed evolution (random) approach	9
1.5.1.2	Rational design approach and molecular dynamics	10
1.5.1.3	Semi-rational design approach	10
1.5.2	Engineered luciferins for more efficient tissue penetration (luciferin analogues) – the near-infrared window in biological tissue (bio-optical window)	10
1.5.3	Thermal and pH stability.....	14
1.5.4	Gene silencing	14
1.5.5	Hypoxia resistance	15
1.6	Limitations of the Firefly Luciferase System for Mammalian Applications	15
1.6.1.1	Effects upon stem cells	16

1.6.1.2	Anti-luciferase immune response.....	16
1.7	Current Options for Countering the Anti-Luciferase Immune Response.....	18
1.7.1	Transgenic reporter-tolerised animals.....	18
1.7.2	Immunocompromised animal strains	19
1.7.3	Immunosuppressive Drugs.....	20
1.8	Potential for Luciferases with Reduced Immunogenicity	20
1.9	Acyl-CoA Synthetases (ACSs)	21
1.9.1	Purpose and function of Acyl-CoA synthetases	21
1.9.2	Genetic, Functional, and Structural Similarities between Acyl-CoA Synthetases and Firefly Luciferases.....	21
1.9.3	Evidence of the Bioluminescent Potential of Acyl-CoA Synthetases	24
1.10	Aims and Objectives.....	24
2	Chapter 2 – Materials and Methods	28
2.1	Materials	28
2.1.1	Materials related to purified DNA	28
2.1.1.1	Reagents and buffers	28
2.1.1.2	Oligonucleotide primers	29
2.1.1.3	Double-stranded oligonucleotide gene fragments (gBlocks)	32
2.1.2	Materials related to purified protein.....	33
2.1.2.1	Buffer list for protein purification.....	33
2.1.2.2	Reagents for Coomassie staining of SDS-PAGE gels.....	34
2.1.2.3	Reagents for silver staining of SDS-PAGE gels	34
2.1.3	Materials related to <i>E. coli</i> cells	34
2.1.3.1	<i>E. coli</i> strains and plasmids	34
2.1.3.2	Culture media.....	35
2.1.4	Materials related to mammalian cells	35
2.1.4.1	Culture media.....	35
2.1.4.1.1	Preparation of heat-inactivated foetal calf serum (FCS)	35

2.1.4.1.2	Preparation of complete cell culture media	35
2.1.4.2	Polyethylenimine (PEI) transfection agent	36
2.1.5	Materials related to bioluminescence screening.....	36
2.1.5.1	Luciferin stocks.....	36
2.1.5.2	Luciferin for <i>E. coli</i> colony BLI screening.....	37
2.1.5.3	Luciferin for mammalian cell culture and in vivo BLI screening	37
2.1.5.4	Haemoglobin phantom	37
2.2	Methods.....	38
2.2.1	Methods related to purified DNA	38
2.2.1.1	Ordering codon-optimised DNA oligonucleotides (gBlocks).....	38
2.2.1.2	DNA concentration.....	38
2.2.1.3	DNA concentration quantification (NanoDrop spectrophotometer)	38
2.2.1.4	Agarose gel electrophoresis.....	39
2.2.1.4.1	Running samples via agarose gel electrophoresis	39
2.2.1.4.2	Post-run staining of agarose gels.....	39
2.2.1.4.3	Agarose gel analysis	39
2.2.1.4.4	Agarose gel extraction and purification of DNA	39
2.2.1.5	Cleaning (PCR-cleaning) of DNA.....	40
2.2.1.6	Restriction digestion of DNA.....	40
2.2.1.7	T4 DNA ligation	40
2.2.1.8	Gibson assembly	40
2.2.1.9	Slicing by overlap extension (SOE) PCR.....	40
2.2.1.10	DNA Sanger sequencing.....	41
2.2.1.11	General polymerase chain reaction (PCR)	41
2.2.1.12	Error-prone polymerase chain reaction mutagenesis of DNA.....	42
2.2.2	Methods related to purified protein.....	42
2.2.2.1	Desalting of proteins.....	42
2.2.2.2	Concentration of proteins.....	42

2.2.2.3	Running of proteins on SDS-PAGE gels	42
2.2.2.4	Coomassie staining of SDS-PAGE gels	43
2.2.2.5	Silver staining of SDS-PAGE gels.....	43
2.2.2.6	Imaging and ImageJ analysis of (Coomassie) stained SDS-PAGE gels.....	43
2.2.2.7	Bradford protein assay.....	44
2.2.3	Methods related to <i>E. coli</i>	44
2.2.3.1	Plasmid purification	44
2.2.3.2	Growth and maintenance of <i>E. coli</i> strains	44
2.2.3.3	Transformation of <i>E. coli</i>	45
2.2.3.4	Selection of sequence-correct colonies	45
2.2.3.5	Glycerol stock creation.....	45
2.2.3.6	Over-expression of enzymes (IPTG induction).....	45
2.2.3.6.1	IPTG induction of colonies	45
2.2.3.6.2	IPTG induction of liquid culture	46
2.2.3.7	Protein purification (Ni-NTA column)	46
2.2.3.7.1	Small-scale purification (for up to 5 mL of 0.7 OD ₆₀₀ LB culture).....	46
2.2.3.7.2	Large-scale purification (for up to 300 mL of 0.7 OD ₆₀₀ LB culture)	47
2.2.4	Methods related to mammalian cells (HEK)	47
2.2.4.1	Trypsinisation of cells.....	47
2.2.4.2	Counting of cells.....	48
2.2.4.3	Subculturing (passaging) of mammalian cells.....	48
2.2.4.4	Flow cytometry of cells	49
2.2.4.4.1	Preparation of cells for flow cytometer.....	49
2.2.4.4.2	General procedure for setting up flow cytometers	49
2.2.4.4.3	Running cells through LSR Fortessa for analysis.....	50
2.2.4.4.4	Running cells through BD FACSAria for flow-assisted cell sorting (FACS)	50
2.2.4.5	Cryopreservation of cell stocks	51
2.2.4.6	Transfection of cells	51

2.2.4.7	Viral transduction of cells	52
2.2.5	Methods related to in vivo work.....	55
2.2.5.1	Preparation of NSG mice for implantation	55
2.2.5.2	Implantation of NSG mice with HEK cells in Matrigel.....	55
2.2.5.3	Imaging of mice using the PhotonIMAGER.....	56
2.2.6	Methods related to bioluminescence measurement	57
2.2.6.1	PhotonIMAGER	57
2.2.6.2	BMG Labtech CLARIOstar for precise spectra.....	57
2.2.6.3	Purified protein screening.....	57
2.2.6.4	<i>E. coli</i> BLI screening.....	58
2.2.6.4.1	<i>E. coli</i> colony BLI screening	58
2.2.6.4.2	<i>E. coli</i> crude lysate BLI screening	58
2.2.6.5	Mammalian cell culture BLI and FLI screening.....	59
2.2.6.5.1	Mammalian cell culture FLI screening	59
2.2.6.5.2	Mammalian cell culture BLI screening.....	59
2.2.7	Methods related to <i>Ari</i> GWluc expression, lysate preparation, and BLI screening.....	59
2.2.7.1	<i>Ari</i> GWluc expression, lysate preparation, and BLI screening according to Trowell <i>et al.</i> supplementary	60
2.2.7.1.1	<i>Ari</i> GWluc expression according to Trowell <i>et al.</i> supplementary	60
2.2.7.1.2	<i>Ari</i> GWluc lysate preparation according to Trowell <i>et al.</i> supplementary	60
2.2.7.1.3	<i>Ari</i> GWluc BLI screening according to Trowell <i>et al.</i> supplementary.....	60
2.2.7.2	Successful <i>Ari</i> GWluc expression	61
3	Chapter 3 – Creation and In Vitro Testing of Random and Rational Mutants of Murine Acyl-CoA Synthetase 3.....	62
3.1	Chapter Summary	62
3.2	Overall Introduction.....	63
3.2.1	Random mutagenesis.....	64
3.2.2	Substitution of the active site consensus sequence	65

3.2.3	Substitution of substrate determining residues	65
3.2.4	Substitution of the R223 loop	66
3.3	Results and Discussion	66
3.3.1	Selection of the luciferase parent enzyme (comparison of <i>Ari</i> GWluc and <i>Ppy</i> Fluc) ..	66
3.3.1.1	Comparison of <i>Ari</i> GWluc to <i>Ppy</i> Fluc via pET16b Expression	70
3.3.1.1.1	Design.....	70
3.3.1.1.2	Cloning	73
3.3.1.1.3	Screening.....	73
3.3.1.2	Comparison of <i>Ari</i> GWluc to <i>Ppy</i> Fluc via pCold™ Expression (only for <i>Ari</i> GWluc) .	75
3.3.1.2.1	Design.....	75
3.3.1.2.2	Cloning	75
3.3.1.2.3	Screening.....	77
3.3.1.3	Mini Discussion	80
3.3.2	Selection of the ACS parent	82
3.3.2.1	Cloning of ACSM3 into pET16b and transformation into <i>E. coli</i> and screening with D-LH_2	86
3.3.3	Design, cloning, and screening of ACSM3 without mitochondrial targeting sequence	86
3.3.3.1	Design.....	87
3.3.3.2	Cloning	88
3.3.3.3	Screening.....	88
3.3.4	Modelling and LCMS analysis of the adenylation of D-LH_2 by WT ACSM3	89
3.3.5	Random Mutagenesis of ACSM3.....	91
3.3.5.1	Cloning of random mutagenic libraries.....	92
3.3.5.2	Screening of the random mutagenic libraries.....	92
3.3.6	<i>Ppy</i> Fluc active site consensus sequence substitution.....	94
3.3.6.1	Design.....	94
3.3.6.2	Cloning	94
3.3.6.3	Screening.....	95

3.3.7	Substrate Determining Residues (SDRs)	96
3.3.7.1	Design.....	96
3.3.7.2	Cloning	99
3.3.7.3	Screening.....	100
3.3.8	<i>Ppy</i> Fluc R223 – T235 Loop	100
3.3.8.1	Modelling of structural differences between the R223 loop in <i>Ppy</i> Fluc and equivalent loop in ACSM3.....	100
3.3.8.2	ACSM3 with the active site consensus sequence and R223 loop of <i>Ppy</i> Fluc substituted	104
3.3.8.2.1	Design.....	104
3.3.8.2.2	Cloning	105
3.3.8.2.3	Screening.....	105
3.3.8.3	ACSM3 with the active site consensus sequence, SDRs, and R223 loop of <i>Ppy</i> Fluc substituted	106
3.3.8.3.1	Design.....	106
3.3.8.3.2	Cloning	106
3.3.8.3.3	Screening.....	107
3.3.9	Further protein modelling.....	107
3.4	Further Discussion.....	110
3.4.1	Could a more optimal ACS been selected?	111
3.4.2	Was epPCR adequate, and could it have been improved in throughput?.....	111
3.4.3	Unexplored techniques for identifying bioluminescent ACSM3 mutants	113
3.5	Conclusions	114
4	Chapter 4 – Creation and In Vitro Testing of the <i>Photinus Pyralis</i> / Murine Acyl-CoA Synthetase 1, 2, and 3 Hybrid Constructs.....	115
4.1	Chapter Summary	115
4.2	Overall Introduction.....	115
4.3	Results and Discussion	117
4.3.1	Design, cloning, and screening of the <i>E. coli</i> chimera constructs	117

4.3.1.1	Design.....	117
4.3.1.2	Cloning	122
4.3.1.3	Screening of the <i>E. coli</i> chimera constructs.....	122
4.3.1.3.1	Bioluminescence screening of colonies	122
4.3.1.3.2	Screening of purified proteins	125
4.3.1.3.2.1	Quantification of purified bioluminescent proteins	125
4.3.1.3.2.1.1	SDS-PAGE gel analysis and quantification	125
4.3.1.3.2.1.2	Bradford assay	127
4.3.1.3.2.2	BLI screening with D-LH_2	128
4.3.2	Design, cloning, and screening of the mammalian chimera constructs.....	130
4.3.2.1	Design.....	130
4.3.2.2	Cloning	133
4.3.2.3	BLI screening	134
4.3.2.3.1	BLI screening with D-LH_2	134
4.3.2.3.2	BLI screening with infraluciferin (iLH ₂) and 6-amino-D-luciferin (ALH ₂)	135
4.3.3	Design, cloning, and screening of the N-terminal P2A-linked mCherry chimera constructs.....	137
4.3.3.1	Design.....	137
4.3.3.2	Cloning	140
4.3.3.3	Screening.....	140
4.3.3.3.1	Bioluminescence screening using the PhotonIMAGER.....	140
4.3.3.3.2	Fluorescence screening using the PhotonIMAGER.....	141
4.3.3.3.3	Fluorescence screening via flow cytometry.....	141
4.4	Further Discussion.....	143
4.4.1	Would ACSLs be more suitable for hybridisation with <i>Ppy</i> Fluc?	144
4.4.2	Potential explanations for increased PpA1 expression relative to <i>Ppy</i> Fluc.....	144
4.4.3	Non-optimal ACSM1 used for hybridisation	145
4.5	Conclusions	145

5	Chapter 5 – Creation and In Vivo Testing of Virally-Transduced HEK 293 Cell Lines of <i>Photinus Pyralis</i> / Murine Acyl-CoA Synthetase 1 Hybrid Construct	147
5.1	Chapter Summary	147
5.2	Overall Introduction (animal ethical considerations and experimental design)	147
5.3	Results and Discussion	150
5.3.1	Creation and preparation of virally-transduced cell lines for implantation	150
5.3.2	Implantation of HEK into NSG mice	151
5.3.3	FLI EGFP and BLI screening of implanted NSG mice with D-LH ₂	152
5.4	Further Discussion.....	160
5.4.1	How does the increased expression of PpA1 relative to Ppy Fluc impact in vivo use?	160
5.4.2	How were final data impacted by non-optimal mouse placement?	161
5.4.3	How were final data impacted by premature acquisition termination?	162
5.5	Conclusions	162
6	Chapter 6 – Further engineering enhancements to the <i>Photinus Pyralis</i> / murine acyl-CoA synthetase 1 hybrid construct	164
6.1	Chapter Summary	164
6.2	Overall Introduction.....	165
6.2.1	Substitution of the <i>Ppy</i> Fluc major BALB/c epitope.....	165
6.2.2	Fragmented ACSM substitution of the <i>Ppy</i> Fluc N-terminal domain.....	165
6.3	Results and Discussion	167
6.3.1	Design, cloning, and screening of microdomain-exchanged <i>Ppy</i> Fluc / ACSM1 constructs	167
6.3.1.1	Design.....	167
6.3.1.2	Cloning	168
6.3.1.3	Screening.....	171
6.3.2	Design, cloning, and screening of epitope-substituted <i>Ppy</i> Fluc mutants.....	172
6.3.2.1	Design.....	172
6.3.2.2	Cloning	179

6.3.2.3	Screening.....	179
6.4	Further Discussion.....	185
6.4.1	Potential light output of the A1PpPpPpPpPpA1 construct.....	186
6.4.2	Impact of Hb filtration.....	186
6.5	Conclusions	186
7	Chapter 7 – General Discussion, Conclusions, and Ideas for Future Work.....	188
7.1	Strategy Overview.....	188
7.2	Technical and Methodological Challenges and Complications.....	191
7.3	Concluding Remarks.....	194
7.4	Future Work	194
8	References	197
9	Appendix	210
9.1	Expanded Figures from Chapter 3	210
9.2	Expanded Figures from Chapter 4	227
9.3	Expanded Figures from Chapter 6	240
9.4	Generic Figures from Various Chapters	247

List of Figures

Figure Number	Title
1.1	Luciferase systems identified to date
1.2	Natural orange, green, and red bioluminescence
1.3	The two stages of the bioluminescent D-luciferin reaction
1.4	Possible paths for luciferin oxidation resulting in various colour emission
1.5	Stereo surface representation of firefly luciferase
1.6	Cartoon representations of Fluc in the adenylating and oxidising conformations
1.7	Absorption spectra of oxy- and deoxyhaemoglobin
1.8	Structure and λ_{\max} of beetle luciferin (D-LH ₂) and a selection of analogues
1.9	Activation of the adaptive immune system and MHC class I antigen-presentation pathway
1.10	Comparison of ACS and Fluc respective reactions
1.11	Structural alignment of insect ACS and Fluc N-terminal domains
2.1	Example of FSC-A vs SSC-A scatter plot, gated into FSC-A vs FSC-H scatter plot
2.2	Flow chart of processes required for viral transduction of adherent cells
3.1	The Fluc active site consensus sequence in multiple Flucs
3.2	Phylogenetic tree of <i>Ari</i> GWluc, various other luciferases, and luciferase-like enzymes
3.3	Normalised emission spectra of <i>Ppy</i> Fluc and <i>Ari</i> GWluc
3.4	Map of WT <i>Ppy</i> Fluc within pET16b plasmid
3.5	<i>Ari</i> GWluc linker identification via 2D protein sequence alignment with <i>Ppy</i> Fluc
3.6	Map of entire <i>Ari</i> GWluc insert, for insertion into pET16b plasmid
3.7	Maps of NTD and CTD of <i>Ari</i> GWluc gBlock oligonucleotides, for Gibson assembly and insertion into pET16b plasmid
3.8	Silver-stained SDS-PAGE of <i>Ari</i> GWluc and <i>Ppy</i> Fluc protein purification fractions
3.9	Map of WT <i>Ari</i> GWluc within pCold TF plasmid
3.10	Silver-stained SDS-PAGE of <i>Ari</i> GWluc, <i>Ppy</i> Fluc, ACSM1 (MTS), ACSM3 (MTS), ACSM1 (no MTS), and ACSM3 (no MTS).
3.11	<i>Ppy</i> Fluc-relative specific activity of <i>Ari</i> GWluc
3.12	Normalised pH dependence of total light output of <i>Ppy</i> Fluc and <i>Ari</i> GWluc
3.13	Normalised spectra of <i>Ari</i> GWluc and <i>Ppy</i> Fluc at pH 7.8
3.14	Possible <i>Ari</i> GWluc luciferin biosynthetic and bioluminescence pathway
3.15	Small molecules involved in the bioluminescence of <i>Ari</i> GWluc and Fluc

3.16	Structural alignment of <i>Ppy</i> Fluc and ACSM3 model
3.17	ACSM3 active site identification via protein sequence alignment derived from 3D alignment of ACSM3 and <i>Ppy</i> Fluc
3.18	Map of ACSM3 gBlock oligonucleotide, for insertion into pET16b plasmid
3.19	Putative mitochondrial targeting sequences of ACSM1 and ACSM3
3.20	Sequence of the gBlock oligonucleotide for removal of the ACSM1 MTS
3.21	Sequence of the gBlock oligonucleotide for removal of the ACSM3 MTS
3.22	Computer model-derived binding energies of luciferin to ACSM3
3.23	Identification of luciferyl adenylate in HPLC portion of LCMS
3.24	DSLR photographs of representative error-prone PCR <i>Ppy</i> Fluc colonies, lifted from a petri dish
3.25	Map of ACSM3 backbone with the active site consensus sequence of <i>Ppy</i> Fluc substituted
3.26	Upstream and downstream SOE PCR of ACSM3 backbone to introduce <i>Ppy</i> Fluc active site consensus sequence of <i>Ppy</i> Fluc, with primers shown
3.27	Structural alignment of <i>Ppy</i> Fluc and 1AMU
3.28	Identification of the primary (Khurana et al. 2010) substrate-determining residues of <i>Ppy</i> Fluc via 3D protein sequence alignment to 1AMU.
3.29	Identification of the primary (Khurana et al. 2010) and secondary (Oba et al. 2009) substrate-determining residues of ACSM3 via 3D protein sequence alignment to <i>Ppy</i> Fluc.
3.30	Map of ACSM 3 backbone with <i>Ppy</i> Fluc: active site consensus sequence substitution and all SDRs
3.31	SOE PCRs of ACSM3 (with <i>Ppy</i> Fluc active site) to introduce all <i>Ppy</i> Fluc SDRs, with primer annealing positions shown
3.32	Spatial relationship between luciferyl adenylate and the R223 (and equivalent) loops of <i>Ppy</i> Fluc (oxidising conformation) and ACSM3 threaded model
3.33	Solvent shielding provided by the R223 loop of the oxidising conformation of <i>Ppy</i> Fluc for luciferyl adenylate
3.34	Solvent shielding provided by the R223-equivalent loop of ACSM3 threaded model for luciferyl adenylate
3.35	ACSM3 R223-equivalent loop identification via protein sequence alignment derived from 3D alignment of ACSM3 and <i>Ppy</i> Fluc
3.36	Map of ACSM 3 backbone with <i>Ppy</i> Fluc: active site consensus sequence substitution and R223 loop
3.37	Upstream and downstream SOE PCR of ACSM3 (with <i>Ppy</i> Fluc active site) backbone to introduce R223 loop of <i>Ppy</i> Fluc, with primers shown
3.38	Map of ACSM 3 backbone with <i>Ppy</i> Fluc: active site consensus sequence substitution, all SDRs, and R223 loop

- 3.39 Upstream and downstream SOE PCR of ACSM3 (with *Ppy* Fluc active site and all SDRs) backbone to introduce R223 loop of *Ppy* Fluc, with primers shown
- 3.40 Influence of K443 and Q448 residues on the luciferyl adenylate binding pocket of *Ppy* Fluc
- 3.41 Influence of K476 and R481 residues (equivalent to K443 and Q448 in *Ppy* Fluc) on the “luciferyl adenylate binding pocket” of ACSM3
- 3.42 Alternating positions of K443 and Q448 residues between oxidising and adenylating conformations of *Ppy* Fluc
- 3.43 Presumptive visualisation of area of sequence space probed by the various protein engineering approaches
- 4.1 Structure-derived protein sequence alignment of *Ppy* Fluc and ACSM3 (via TM-align)
- 4.2 2D Protein sequence alignment of ACSM1, ACSM2, and ACSM3 (via Clustal Omega)
- 4.3 Map of gBlock DNA sequence encoding the entire ACSM1 with *PacI* mutation, for insertion into pET16b
- 4.4 Maps of gBlock DNA sequences encoding the NTD and CTD of ACSM2 with *PacI* mutation, for insertion into pET16b
- 4.5 Maps of gBlock DNA sequences encoding the NTD and CTD of ACSM3 with *PacI* mutation, for insertion into pET16b
- 4.6 Map of gBlock DNA sequence encoding the NTD of *Ppy* Fluc (Pp00) (with *PacI*) with a stop codon and *Bam*HI site for insertion into pET16b
- 4.7 DSLR image of D-LH_2 -sprayed PpPp *E. coli* colonies
- 4.8 PpPp-relative brightness of PpA1, PpA2, PpA3, and Pp00
- 4.9 Normalised spectra of PpPp, PpA1, PpA2, and PpA3
- 4.10 Coomassie-stained SDS-PAGE gel of PpPp, PpA1, PpA2, PpA3, and Pp00
- 4.11 ImageJ analysis values of Coomassie-stained SDS-PAGE gel of PpPp, PpA1, PpA2, PpA3, and Pp00
- 4.12 Combined ImageJ analysis values of Coomassie-stained SDS-PAGE gel of PpPp, PpA1, PpA2, PpA3, and Pp00
- 4.13 Bradford assay protein concentration values of construct protein purification batches of PpPp, PpA1, PpA2, PpA3, and Pp00
- 4.14 Bradford assay-derived expression efficiencies of of PpPp, PpA1, PpA2, PpA3, and Pp00
- 4.15 PpPp-relative specific activities of PpA1, PpA2, PpA3, and Pp00
- 4.16 Normalised spectra of PpPp, PpA1, PpA2, and PpA3
- 4.17 Plasmid map of pCCL 305 with a Fluc inserted
- 4.18 Map of gBlock DNA sequence encoding the entire WT *Ppy* Fluc with *PacI* and *Hind*III, for insertion into pCCL 305
- 4.19 Maps of gBlock DNA sequences encoding the entire ACSM1 with *PacI* or *Hind*III, for insertion into pCCL 305
- 4.20 Maps of gBlock DNA sequences encoding the entire ACSM3 with *PacI* or *Hind*III, for insertion into pCCL 305

- 4.21 PpPp-relative light output of PpA1 (HindIII), PpA1 (PacI), PpA3 (HindIII), and PpA3 (PacI)
- 4.22 Normalised spectra of PpPp, PpA1 (HindIII), PpA1 (PacI), PpA3 (HindIII), and PpA3 (PacI)
- 4.23 PpPp-relative light output of PpA1 (HindIII) and PpA3 (HindIII), with ALH₂ (left) and iLH₂ (right)
- 4.24 Normalised spectra of PpPp, PpA1 (HindIII), PpA3 (HindIII) with D-LH₂, ALH₂, and iLH₂
- 4.25 EGFP and mCherry absorption and emission spectra with 488 nm laser as available on the LSRFortessa
- 4.26 EGFP and mCherry absorption and emission spectra with 590 nm laser as available on the LSRFortessa
- 4.27 Taken from (Kim et al. 2011). Diagram of the 2A system (A) and sequence of P2A (B)
- 4.28 Map of gBlock DNA sequence encoding mCherry and P2A for splicing into *Ppy* Fluc in pCCL 305 via the flanking BamHI and BstEII sites
- 4.29 PpPp-relative light output of PpA1 (both with N-terminal-linked mCherry).
- 4.30 EGFP, mCherry, and PE-Texas Red® absorption and emission spectra with 568 nm laser as available on the LSRFortessa
- 4.31 Quadrant analysis of HEK cells for FITC and mCherry fluorescence
- 4.32 mCherry mean fluorescence indexes of FITC-positive: non-transformed, *Ppy* Fluc, and PpA1 transformed cells
- 5.1 FITC histogram of HEK cells for implantation
- 5.2 Mean fluorescent indexes of HEK cells for implantation
- 5.3 D-LH₂ acquisition overlaid on bright-field image of mice of Cage A, with ROIs
- 5.4 D-LH₂ acquisition overlaid on bright-field image of mice of Cage B, with ROIs
- 5.5 D-LH₂ acquisition overlaid on bright-field image of mice of Cage C, with ROIs
- 5.6 Longitudinally acquired relative (to brightest data point) light emission of all mice
- 5.7 Longitudinally acquired PpPp-relative light emission of PpPp and PpA1 mice, averaged
- 5.8 PpPp-relative total light emission of PpA1 mice
- 5.9 Normalised spectra of PpPp and PpA1 mice from Cage A
- 6.1 Arrangement of the four structural subdomains derived from the eight sequential “microdomains” of *Ppy* Fluc
- 6.2 Map of gBlock DNA sequence encoding the 8μD *Ppy* Fluc parent
- 6.3 Map of gBlock DNA sequence encoding the 8μD ACSM1 parent
- 6.4 Primer binding locations for μD amplification from *Ppy* Fluc gBlock oligo parent
- 6.5 Primer binding locations for μD amplification from ACSM1 gBlock oligo parent

6.6	Gel electrophoresis of purified μ D PCR products of <i>Ppy</i> Fluc and ACSM1
6.7	Light output of PpPpPpPpPpPpPpPpPp (<i>Ppy</i> Fluc), PpPpPpPpPpPpA1 (PpA1), and A1PpPpPpPpPpPpPp
6.8	Three BALB/c (and one C57BL/6) epitopes identified within <i>Ppy</i> Fluc and their respective binding scores
6.9	Major <i>Ppy</i> Fluc epitope (GFQSMYTFV) highlighted displayed in VMD viewer
6.10	The homology of the ten chosen candidate luciferases to <i>Ppy</i> Fluc
6.11	Alignment of the ten candidate luciferases with <i>Ppy</i> Fluc at the GFQSMYTFV region (marked in red)
6.12	Map of gBlock oligo of μ D3 of <i>Ppy</i> Fluc with the <i>Psc</i> Fluc epitope replacement substituted
6.13	<i>Ppy</i> Fluc-relative light output of PpPsPp, PpPpA1, and PpPsA1
6.14	Low resolution (10 nm divisions) spectra screen of PpPpPp, PpPsPp, PpPpA1, and PpPsA1
6.15	High resolution (1nm divisions) spectra screen of PpPpPp, PpPsPp, PpPpA1, and PpPsA1
6.16	PpPpPp-relative light output of PpPsPp, PpPpA1, and PpPsA1; all <i>without</i> Hb phantom
6.17	PpPpPp-relative light output of PpPsPp, PpPpA1, and PpPsA1; all <i>with</i> Hb phantom
6.18	Normalised spectra of PpPpPp, PpPsPp, PpPpA1, and PpPsA1; all <i>without</i> Hb phantom
6.19	Normalised spectra of PpPpPp, PpPsPp, PpPpA1, and PpPsA1; all <i>with</i> Hb phantom
7.1	Flow chart of the progression and termination of constructs created within this project
9.1	An example of a negative screen via the Biospace Lab PhotonIMAGER (bright-field photograph with photon acquisition superimposed)

List of Tables

Table Number	Title
1.1	λ_{\max} of beetle luciferin (D-LH ₂) and a selection of analogues
1.2	Categories of fatty acids and corresponding acyl-CoA synthetases by chain length
1.3	Comparison of ACS and Fluc respective reactions
2.1	Manufacturers of the general reagents and buffers used
2.2	Exhaustive list of primers ordered for the project
2.3	Quantities of reagents required for transient transduction of adherent cells
3.1	Exhaustive list of constructs created and detailed within Chapter 3.
3.2	<i>Ppy</i> Fluc active site homology with a number of acyl-CoA synthetases
3.3	Comparison of ACS and Fluc respective reactions
3.4	Primer pairs to introduce all <i>Ppy</i> Fluc SDRs to ACSM (with <i>Ppy</i> Fluc active site), with length of resulting PCR products shown
4.1	The protein sequence mutations required to introduce a <i>PacI</i> site into the linkers of ACSM1, ACSM2, and ACSM3 at the equivalent position as in <i>Ppy</i> Fluc cDNA linker
4.2	Constructs combining the <i>Ppy</i> Fluc and ACSM domains
4.3	Residue mutations required to introduce <i>PacI</i> or <i>HindIII</i> into ACSM1 and ACSM3
5.1	State of HEK cells for implantation in terms of number and viability
5.2	Identity and status of each mouse, and the imaging acquisition time
5.3	Advantages of using the Safe-harbour system over lentivirus for transduction of mammalian cells
6.1	Desired PCR amplification products for <i>Ppy</i> Fluc and ACSM1 microdomains
6.2	Details of the ten chosen candidate luciferases
6.3	Predicted equivalent epitopes of <i>Ppy</i> Fluc major BALB/c epitope from candidate luciferases
6.4	The impact of Hb filtration on PpPpPp, PpPsPp, PpPpA1, and PpPsA1 emission
6.5	Five experiments comparing the relative light output of novel luciferases to WT <i>Ppy</i> Fluc

List of Abbreviations

Abbreviation	Meaning
μ D	Microdomain
β -ME	β -mercaptoethanol
λ_{\max}	Maximal wavelength
ACS	Acyl-CoA synthetase (often "FACS" in literature)
ACSL	Acyl-CoA synthetase long-chain family member
ACSM	Acyl-CoA synthetase medium-chain family member
ACSM1	Acyl-CoA synthetase medium-chain family member 1
ACSM2	Acyl-CoA synthetase medium-chain family member 2 mitochondrial isoform 3
ACSM3	Acyl-CoA synthetase medium-chain family member 3 mitochondrial
ALH ₂	amino luciferin (6'-amino-D-luciferin)
AMP	Adenosine monophosphate
ANL	Acyl-CoA synthetases, nonribosomal peptide synthetase adenylation domains, and luciferase enzymes
<i>Alu</i>	<i>Arachnocampa Luminosa</i>
APC	Antigen-presenting cell
APS	Ammonium persulphate
<i>Ari</i>	<i>Arachnocampa richardsae</i>
ARRIVE Guidelines	(Animal Research: Reporting of In Vivo Experiments) guidelines
ATP	Adenosine triphosphate
BCR	B-cell receptor
BL21 cell	BL21 (DE3) <i>E. coli</i> cell (New England Biolabs)
BLAST	Basic local alignment search tool
BLASTp	Protein basic local alignment search tool
BLI	Bioluminescent imaging
bp	Base pair
BRET	Bioluminescence Resonance Energy Transfer
BSA	Bovine serum albumin
CD(number)	Cluster of differentiation (number)
CLIP	(MHC) Class II-associated invariant chain peptide
CoA	Coenzyme A
CP	Cyclophosphamide
CTD	C-terminal domain
CTLA	Cytotoxic T-lymphocyte-associated protein 4
D-ALH ₂	D enantiomer of amino luciferin (6'-amino-D-luciferin)
DAMP	Damage-associated molecular pattern
DC	Dendritic cell
D-LH ₂	D enantiomer of beetle luciferin
D-LH ₂ -AMP	D enantiomer of luciferyl adenylate
dH ₂ O	Distilled water

DH5 α cell	DH5-Alpha <i>E. coli</i> cell (New England Biolabs)
DMSO	Dimethylsulphoxide
dNTPs	deoxyribonucleotide triphosphates
d.p.	Decimal place
DPBS	Dulbecco's phosphate-buffered saline
DSLR	Digital single-lens reflex camera
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDA	Experimental Design Assistant
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
epPCR	Error-prone polymerase chain reaction
ER	Endoplasmic reticulum
FA	Fatty acid
FCS	Foetal calf serum (sometimes "serum")
FF	Firefly
FITC	Fluorescein isothiocyanate
FLI	Fluorescent imaging
Fluc	Firefly luciferase
FSC	Forward-scattered light
g	"g-force", "gram", or "gauge" depending on context
GFP	Green fluorescent protein
GH	Glowing Head
GROMACS	Groningen machine for chemical simulations
Gluc	Gussia luciferase
GWluc	Glowworm luciferase
Hb	Haemoglobin (both oxy- and deoxyhaemoglobin)
HEK cell	Human embryonic kidney 293 cell
HEPA	High efficiency particulate air
HF [®]	High-Fidelity [®]
His-tag	Polyhistidine-tag
h ν	Photon energy
IDT	Integrated DNA Technology, Inc.
IgH	Immunoglobulin heavy chain
iLH ₂	infra-luciferin
I _{max}	Maximum observed intensity
IMD	Imidazole
IMDM	Iscoe's Modified Dulbecco's Medium
IP	Intraperitoneal
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRES	Internal ribosome entry site
kb	Kilobase
k _{cat}	Catalytic rate constant
k _m	Michaelis constant
KO	Knockout
L-AMP	dehydroluciferyl-adenylate

L-CoA	dehydroluciferyl-CoA
L-LH ₂	L enantiomer of beetle luciferin
LB	Lysogeny broth
LH ₂	Beetle luciferin
LH ₂ -AMP	Luciferyl adenylate
LO	Oxyluciferin (ground state)
LO*	Oxyluciferin (electronically excited)
LPS	Lipopolysaccharide
Luc	Beetle luciferase
MCC	Multiple cell clone
Mg ²⁺	Monoatomic magnesium dication
MHC	Major histocompatibility complex
MHC class I	Major histocompatibility complex class I
MHC class II	Major histocompatibility complex class II
Midiprep	Plasmid DNA midipreparation
Miniprep	Plasmid DNA Minipreparation
MM	Michaelis-Menten
MTS	“Mitochondrial targeting signal” or “Mitochondrial targeting sequence” (dependent on context: former for protein, latter for DNA)
NC3Rs	National Centre for the Replacement Refinement & Reduction of Animals in Research
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NK cell	Natural killer cell
Ni	Nickel
NIR	Near-infrared
NSG™	NOD (non-obese diabetic) scid (severe combined immunodeficiency) gamma (null allele of the IL2 receptor common gamma chain)
NT	Non-transformed
NTA	Nitrilotriacetic acid
NTD	N-terminal domain
Oligo	Oligonucleotide
PAGE	polyacrylamide gel electrophoresis
PAMP	Pathogen-associated molecular pattern
PB	Polybrene (Hexadimethrine bromide)
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
ph	Photons
PI	Protease inhibitor
PCD	Programmed cell death
PEI	Polyethylenimine
PET	Positron emission tomography
PMT	Photomultiplier tube
P _{Pi}	Inorganic pyrophosphate
<i>Ppy</i>	<i>Photinus pyralis</i>

<i>Psc</i> (<i>Psn</i> in certain literature)	<i>Photinus scintillans</i>
rcf	Relative centrifugal force
rpm	Revolutions per min
SC	Subcutaneous
SCC	Single cell clone
SDR	Substrate determining residue
SDM	Site directed mutagenesis
SDS	sodium dodecyl sulphate
SEM	Standard error of the mean
SIR	Substrate-interacting residue
SOC	Super optimal broth with catabolite repression
SOE	Splicing by overlap extension
SSC	Side-scattered light
sr	Steradian
SPECT	Single photon emission computed tomography
TAP	Transporter associated with antigen processing
TEM	Tris-acetate EDTA MgSO ₄ buffer
TF	Trigger factor
TC	Cytotoxic T cell
T _H cell	T helper cell
T _H 1	T helper cell class 1
T _H 2	T helper cell class 2
T _H 0 cell	Naive T cell
T reg	Regulatory T cell
Tris	tris(hydroxymethyl)aminomethane
UV	Ultraviolet
UVGI	Ultraviolet germicidal irradiation
VMD	Visual Molecular Dynamics
WT	Wild type

1 Chapter 1 – Introduction

1.1 A Brief History of Bioluminescence

Humans have been fascinated by bioluminescent organisms for millennia. Perhaps the most striking examples are fireflies and glowworms which are native to many parts of Europe, Asia, and the Americas. More subtle examples such as the dim blue glow of decaying fish and the dim green glow of decaying wood were both noted by the Greek philosopher Aristotle (384-322 BC), however it was not until the 18th and 19th centuries that these were attributed to bacteria and fungi respectively. In 1667, the Anglo-Irish chemist Robert Boyle documented that bioluminescence required air (oxygen had not yet been discovered). Since about 500 BC it was assumed that bioluminescence was a form of phosphorescence (fluorescence with prolonged emission), but in 1885 this was disproven by German scientist Emil du Bois-Reymond when he extracted two mixtures from bioluminescent clams and found that the heat-inactivated mixture could be added to the non-heated mixture to restore bioluminescence (McElroy and Strehler 1954); du Bois-Reymond had crudely isolated luciferin and active luciferase and correctly assumed that it was a reaction between a chemical moiety and an oxidising enzyme. The first luciferin structure determined was that of bacterial luciferin by M.J. Cormier in 1954 (Strehler et al. 1954), and the properties of highly-purified bacterial luciferase were determined in 1955 by Green and McElroy (McElroy and Green 1955). Firefly luciferases (Flucs) would be studied in depth over the following decades, revealing them to be among the most efficient.

Bioluminescence occurs when a small molecule 'luciferin' is oxidised via an enzyme 'luciferase'. Luciferase is from the Latin 'light bearing'. Bioluminescence is a biologically enhanced form of chemiluminescence, the defining property is that it must be catalysed by a luciferase.

1.2 Bioluminescent Systems in Nature

Bioluminescence systems are most common in the ocean where the emitted light is often within the blue range of ~440-478nm as this allows more efficient transmission through

water. Green-red emission may be achieved indirectly through the use of emitter-bound accessory chromophores (such as green fluorescent protein) as is the case in the *Aequorea* genus. The close proximity between the luciferase and chromophore allows a phenomenon known as bioluminescence resonance energy transfer (BRET) to occur, resulting in highly efficient red-shifting of the emitted light (Morise et al. 1974). Green-red emission may also be achieved directly by an altered luciferase or entirely different luciferin system.

Despite only ~20% of bioluminescent systems occurring terrestrially (as supposed to ~80% in the ocean) they are perhaps the most investigated, characterised, and engineered; this is largely due to their greater brightness. It is also red-shifted compared to its ocean counterparts. In the depths of the ocean, background light is extremely low and bioluminescent organisms compete mostly with each other's emissions, whereas terrestrial bioluminescent organisms must be visible on moonlit nights and at dawn and dusk.

Several bioluminescence systems exist, defined by their different types of luciferin substrates (Figure 1.1). The diversity of bioluminescent systems and their independent occurrence in all known phyla of life (except plants and archaea) is an example of convergent evolution and suggests that it can be a highly beneficial trait. Further bioluminescence systems may yet to be identified, potentially including the *Arachnocampa* fungus gnat genus which has cross-reactivity with beetle luciferin but is believed to use a novel substrate (Watkins et al. 2018), and sharks since their luminescence mechanisms remain elusive (Duchatelet et al. 2019).

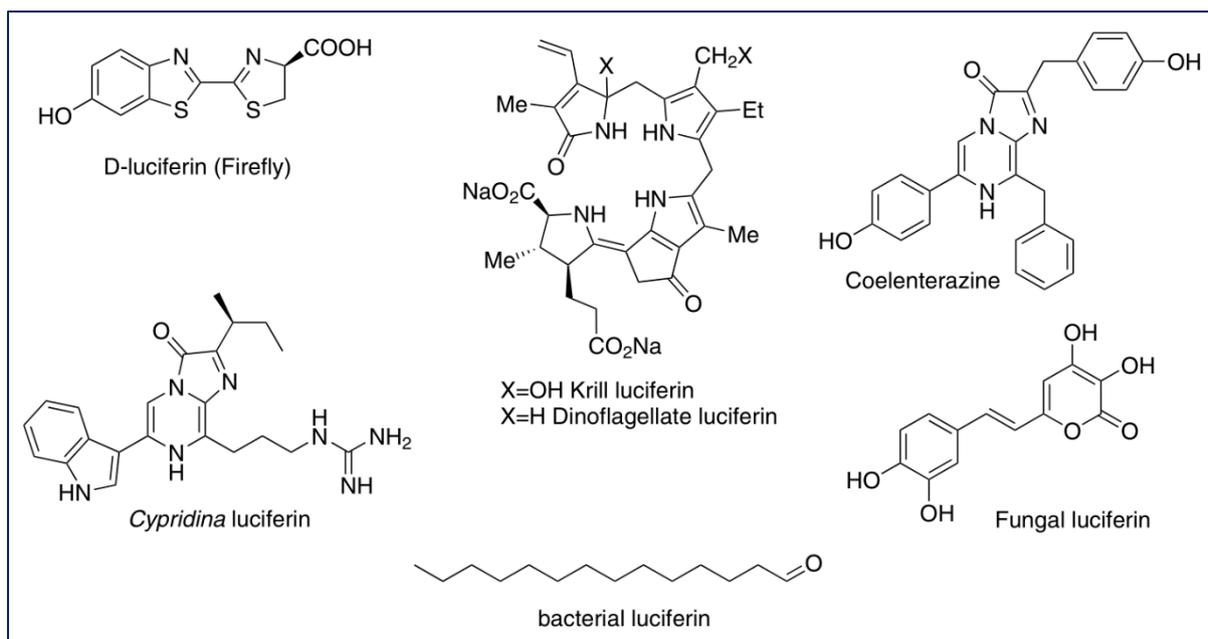


Figure 1.1 taken from (Fleiss and Sarkisyan 2019). **Luciferase systems identified to date.**

The seven luciferins for which at least one luciferase has been identified, namely: firefly, krill, dinoflagellate, coelenterazine (used by renilla luciferases), fungal, bacterial, and Cypridina. The remaining two luciferins (with no identified luciferase counterpart) are earthworm (*Diplocardia*) and limpet (*Latia*).

1.3 Beetle Luciferase/Luciferin System

Despite the name, fireflies (Lampyridae) are actually a family of beetle (Coleoptera) and can be found world-wide, especially in North America and East Asia. Most fireflies emit light in the green-to-orange range. All beetle bioluminescence uses the D enantiomer of the identical molecule beetle luciferin (D-LH₂) (Figure 1.1), despite this many different colours are possible. For instance *Phrixothrix hirtus* railroad worm can produce both red and green light separately using the same luciferin but with separate luciferases, and is also the only known example of native (not engineered or bound to a fluorophore) red bioluminescence (Figure 1.2). The most blue-shifted (yet still green) light from beetle luciferin is emitted from *Amydetes fanestratus* fireflies (Viviani et al. 2011).



Figure 1.2 taken from (Bechara and Stevani 2018). **Natural orange, green, and red bioluminescence.** a. firefly (*Macrolampis omissa*); b. click beetle (*Pyrophorus* sp.); c. railroad worm (*Phrixothrix hirtus*).

1.3.1 Beetle Luciferin reaction

Beetle luciferases are bi-functional enzymes which catalyse both an adenylation and an oxidation reaction (Figure 1.3) in the following reaction involving LH₂, adenosine tri and monophosphate (ATP and AMP respectively), inorganic pyrophosphate (PPi), Luciferyl-AMP (LH₂-AMP), oxygen (O₂), carbon dioxide (CO₂), and excited oxyluciferin (LO*).

1. $\text{LH}_2 + \text{ATP} \leftrightarrow \text{LH}_2\text{-AMP} + \text{PPi}$
2. $\text{LH}_2\text{-AMP} + \text{O}_2 \rightarrow \text{LO}^* + \text{CO}_2 + \text{AMP}$

The excited oxyluciferin (LO*) is electronically excited and highly unstable, visible light emission results from the rapid loss of energy via a fluorescence pathway.

3. $\text{LO}^* \rightarrow \text{LO} + \text{light}$

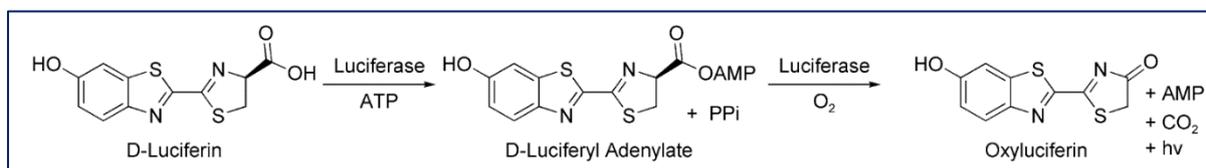


Figure 1.3 taken from (Mofford et al. 2014b). **The two stages of the bioluminescent D-luciferin reaction.** LH₂ is adenylated to LH₂-AMP which is then oxidised to LO* giving light.

LH₂-AMP or its esters can be oxidised to chemiluminesce in the absence of a luciferase, however it is extremely dim and red-shifted in water (Viviani and Ohmiya 2006). Visible light

photons (especially if green rather than red) are of relatively high energy, therefore this requires that the excited state oxyluciferin emit in an environment highly favourable to photon emission. The molecular basis for the emitted colour remains an area of active research, and there is presently no consensus. Several hypotheses exist regarding it, one is that an oxyluciferin (dioxetanone) intermediate emits as different enantiomeric forms via three pathways, with distinctly different colours for each (Figure 1.4). Thus, different luciferases may favour different light-emitting pathways.

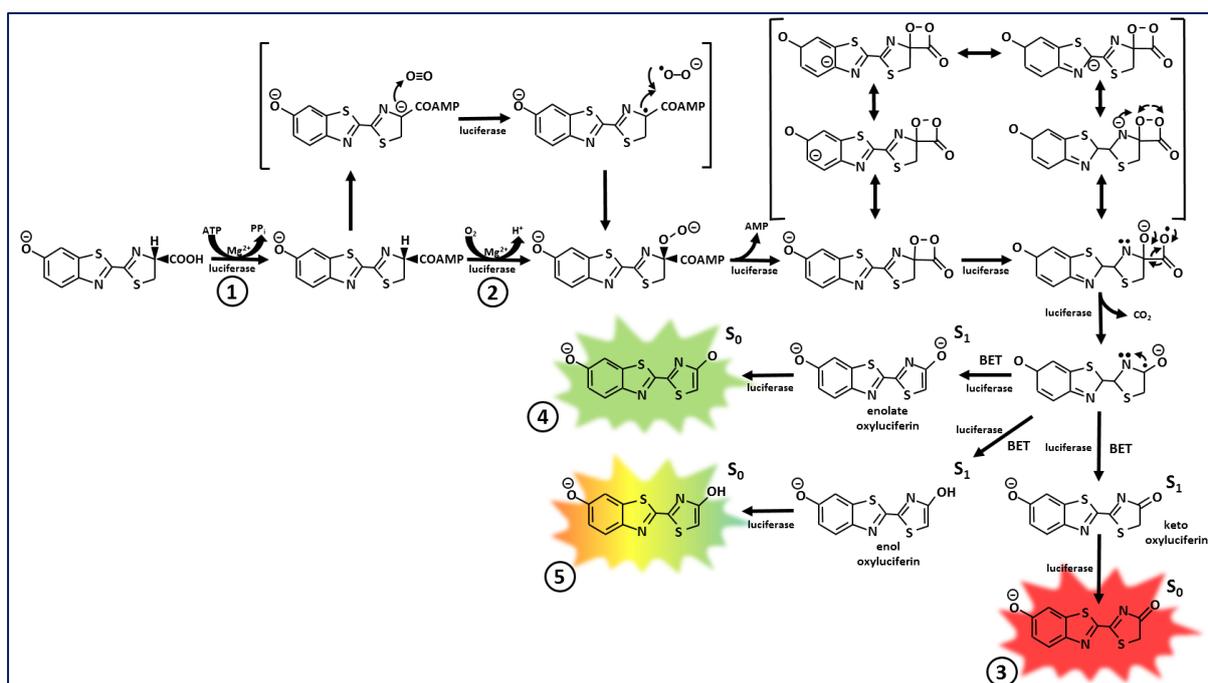


Figure 1.4 adapted from (Bechara and Stevani 2018). Possible paths for luciferin oxidation resulting in various colour emission. 1. Luciferase uses ATP to activate the luciferin carboxylic group as a luciferyl-adenylate anhydride, this is the first step of the simplified two-step reaction discussed earlier. 2. Luciferase inserts molecular oxygen into the luciferin α -carbon to form an α -hydroperoxide. 3. The keto form of oxyluciferin emits in the red (λ_{\max} 615 nm). 4. The enolate form of oxyluciferin emits in the green (λ_{\max} 540 nm). 5. Enol form of oxyluciferin emits in the green to orange (λ_{\max} up to 600 nm). BET stands for bioluminescence (resonance) energy transfer.

1.3.2 Firefly luciferase (Fluc)

During the bioluminescent reaction the enzyme does not merely accelerate the rate of oxidation – the active site of the luciferase protects the emitting species from solvent (water molecules). As a result the emitted photons have higher energy and are thus blueshifted with a higher rate of the oxidation reaction, thus the reaction is brighter even when the number of luciferin molecules used is controlled for. The quantum yield of a bioluminescent reaction is defined as the probability of photon production per chemical reaction (Ando et al. 2008). One general consensus in the literature is that the hydrophobic active site of the Fluc shields the luciferin from solvent to improve light yield and blueshift emission (Maghami et al. 2010). The active site of Fluc can be viewed below (Figure 1.5).

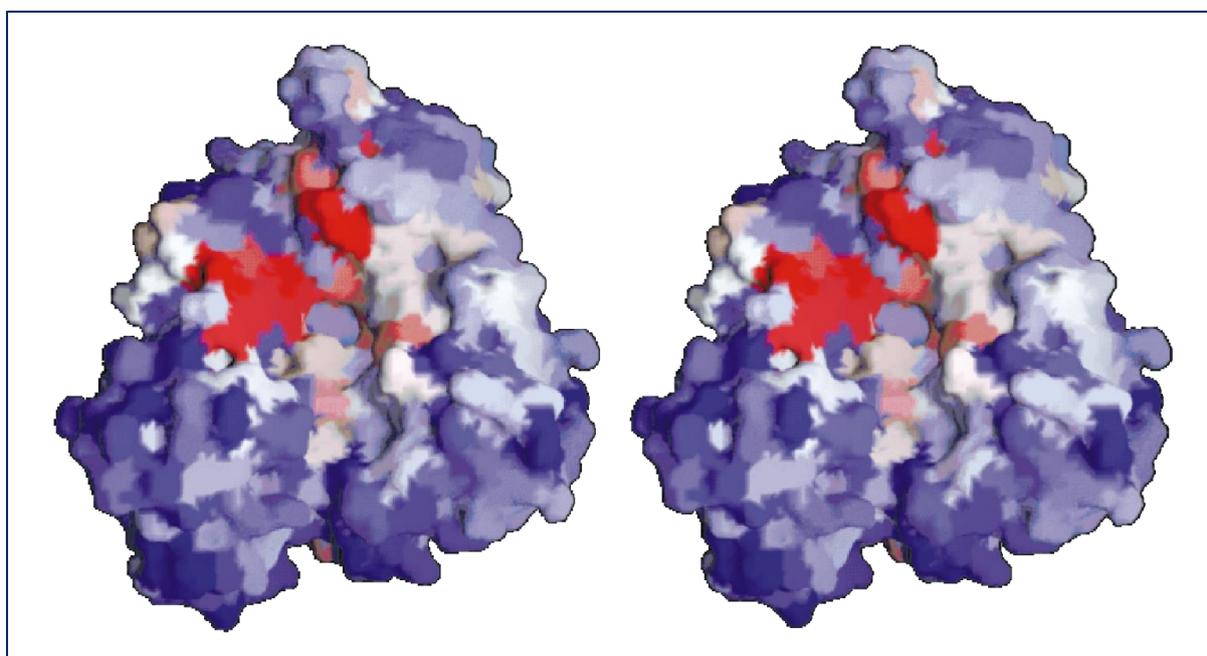


Figure 1.5 taken from (Conti et al. 1996). Stereo surface representation of firefly luciferase. Viewing angle is of the concave molecular surface of the large N-terminal domain of firefly luciferase, looking down onto the Y-shaped system of valleys. The residues highlighted in red represent the highly conserved residues (notably clustered around the active site), blue represent highly variable residues, and shades in-between these colours indicate a range of variability.

Fluc has an open conformation which allows D-LH₂ and ATP to bind, this is known as its ground state or adenylate-forming conformation, and allows the first (adenylating) step of the

reaction to occur. Then – through the use of a flexible linker between its N and C-terminal domains – it changes its conformation and closes around the substrate into the oxidising conformation, which completes the final (oxidising) step of the reaction (Figure 1.6). This second, closed conformation ensures that solvent (water) is excluded from the active site during emission.

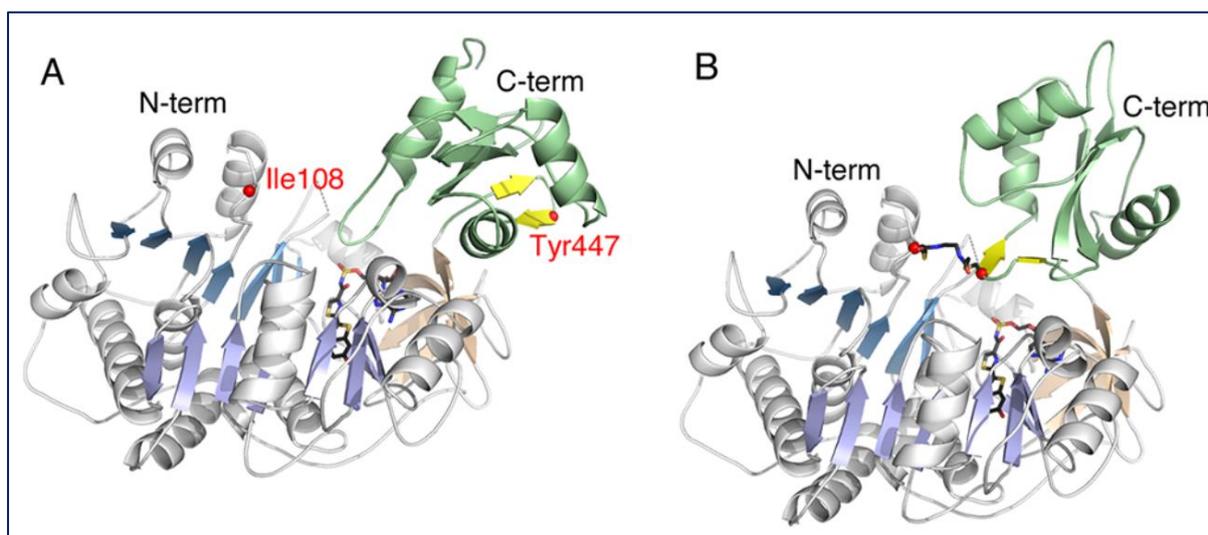


Figure 1.6 taken from (Sundlov et al. 2012). **Cartoon representations of Fluc in the adenylating and oxidising conformations.** A: the adenylating conformation; B. the oxidising conformation. For the purpose of X-ray crystallography of the oxidising conformation the residues Y447 and I108 have been covalently cross-linked.

1.4 Applications of Fluc

Miners have to contend with flammable and explosive atmospheres, and as such jars of fireflies and decaying fish skins were used as a safe form of lighting before this was superseded by the invention of the safety lamp and later the electric lamp. Nevertheless this probably represents the first systemic harnessing of bioluminescence by humans, and although abandoned it was not to be the last such case...

1.4.1 In vitro applications

The high brightness of terrestrial bioluminescence systems made them attractive for researchers with intentions of recombinant expression and as detection kits, these are discussed here:

1.4.1.1 ATP detection

As the luciferase reaction requires ATP to proceed, one use for Flucs are as ATP detectors. ATP is a biological molecule and can give a direct measure of microbiological concentration in food, for instance. When the constituents of the luciferase reaction (minus ATP) are combined with a food sample, the bioluminescence is a direct measure of contamination. One notable example of such a detector is the bioluminescent assay in real-time (BART) which uses the ATP synthesised by ATP sulfurylase during DNA amplification, which when combined with PCR can link light output to amplification of a specific sequence (Gandelman et al. 2010);(Kiddle et al. 2012).

1.4.1.2 Enzyme-linked immunosorbent assays (ELISAs)

ELISA relies upon the detection of the secondary antibody ultimately bound to the antigen. Commonly this is achieved by tagging it with a fluorescent compound or a peroxidase (chemiluminescence), however both of these may suffer from background emission. When even greater sensitivity is required a Fluc may be used in-place of these.

1.4.2 In vivo applications

As a genetically-encodable reporter, Fluc can be transformed into a wide variety of tissues, including animals, insects, and plants. These applications are often the same which fluorescent proteins (another kind of genetically-encodable biomarker) find use for, however an advantage that luciferase has over fluorescent proteins is the greater sensitivity. Consequently many animal studies which would require that a transformed animal tissue/organ be excised, need not be when a Fluc is used as the signal can penetrate through to the surface and be captured by the camera (Choy et al. 2003).

One common application is to link the transformed Fluc to a gene-of-interest, this allows light emission to directly correlate to the expression of the gene-of-interest. A second use is as a tissue-type marker, in which a small animal is implanted with a transplanted tissue (commonly a tumour), Fluc allows longitudinal and spatial monitoring of the transplant. This second application allows – for instance – cancer researchers to test the effectiveness of a new therapeutic on a Fluc-transformed tumour over the course of several weeks.

A further reason that terrestrial bioluminescence systems (like Flucs, as supposed to marine luciferases) have found use is that the native organisms tend to emit light of longer wavelengths, bringing emission closer to the bio-optical window and accordingly better tissue penetration (discussed in further detail below), and therefore makes them more suitable for mammalian applications. It should be noted that there is not a universally-applicable firefly luciferase mutant since different applications may require different enzyme properties.

1.5 Engineered Improvements to the Firefly Luciferase System to date for Mammalian Applications

The wild-type *Ppy* Fluc system can be used for imaging live animals, however it has a number of disadvantages including: thermal and pH instability, light output which is strongly attenuated by tissue, and immunogenicity. Alterations to both the substrate and Fluc itself have been researched to suit different applications. Protein engineering techniques – and how they were exploited to achieve the improvements made to Fluc to date – are also discussed in this section.

1.5.1 Generic protein engineering strategies

There are two approaches to how a protein engineer can set about improving (or otherwise altering) the function of an enzyme such as Fluc – random and rational. These are briefly discussed in further detail below.

1.5.1.1 Directed evolution (random) approach

Directed evolution mimics how proteins are engineered in nature by Darwinian evolution, i.e. through random mutation and natural selection. The difference is that the researcher controls both the degree and the way in which the mutations are delivered, and the criteria by which the mutants are selected. A plethora of implementations exist, with many revolving around different utilisations of the polymerase chain reaction (PCR). It is even possible to mimic sexual reproduction with PCR in which a number of “parent” proteins are shuffled to produce a library of “offspring”. These methods of library generation require heterogeneous cloning and transformation of the created DNA into cells for expression and high-throughput screening (HTS). Fluc lends itself well to this method as enzyme expression and quantification

is extremely simple. The mutations imparted to a protein through directed evolution tend to be small, and thus the improvements made for each round of HTS tend to be incremental. Unlike natural Darwinian evolution, protein engineers have the luxury of foresight, and when such radical changes are required a less blind approach may be sought.

1.5.1.2 Rational design approach and molecular dynamics

A rational design approach is typically taken when the protein is already understood to a degree, such as the crystal structure being known and thus the effect of modifications (up to a point) can be predicted. For instance, a protein's water solubility may be increased by substituting solvent-exposed residues with hydrophobic side chains for those with polar side chains – arginine being the most polar. It is an assumption of protein engineering (and especially of enzyme engineering) that the overwhelming majority of protein sequences will be non-functional for a given application, and thus the more drastically a protein is altered, the higher the probability that the modification is deleterious. The crystal structure of *Ppy* Fluc is known, however large mutations may alter its overall structure and it is currently implausible to computationally predict how a protein as large as Fluc would fold, which greatly limits the scope of rational mutations available to engineers.

1.5.1.3 Semi-rational design approach

Semi-rational design is a blend of the above two approaches, for instance this could come in the form of restricting random mutagenesis to a specific region of the protein, or by creating a set of rationally designed candidates and screening then selecting the best for further modification.

1.5.2 Engineered luciferins for more efficient tissue penetration (luciferin analogues) – the near-infrared window in biological tissue (bio-optical window)

Mammalian tissues are virtually all vascular and therefore perfused with blood which contains haemoglobin (Hb), and muscle tissue contains myoglobin – a distant cousin of Hb. Both Hb and myoglobin contain the red pigment haem and thus have similar absorbance spectra and strongly absorb light of wavelengths shorter than 600 nm. Hb (and myoglobin) can be either oxygenated (oxyhaemoglobin or oxy-Hb HbO₂) or deoxygenated (deoxyhaemoglobin or

deoxy-Hb or Hb). Although Hb can refer to either haemoglobin generally or deoxyhaemoglobin, for this thesis Hb refers to the former. Oxy- and deoxyhaemoglobin are both present in varying concentrations in mammalian tissues, but have different absorbance spectra (Figure 1.7).

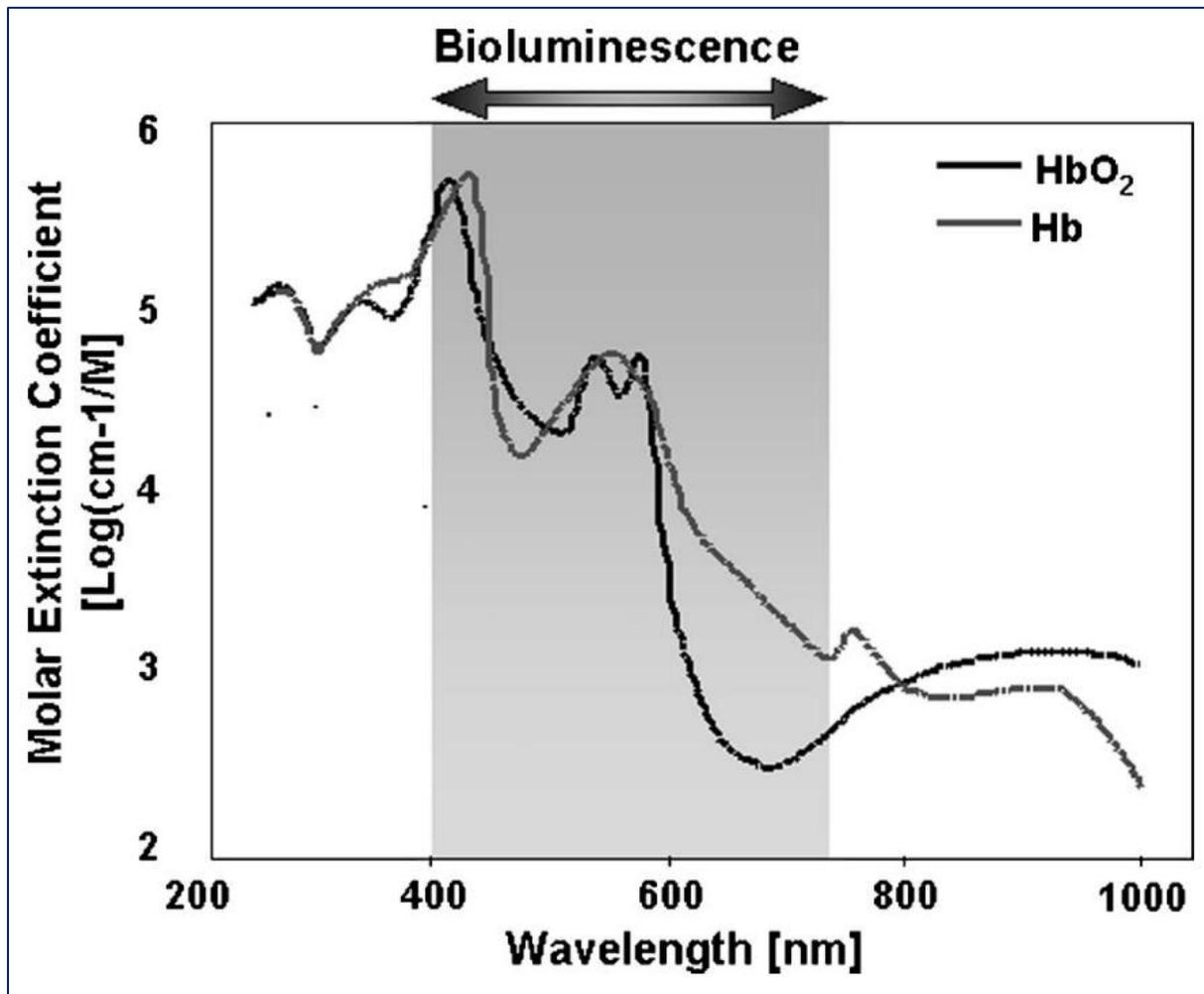


Figure 1.7 taken from (Zhao et al. 2005). Absorption spectra of oxy- and deoxyhaemoglobin. Optimal transmission of photons through mammalian tissues occurs between 600 and 900 nm based on the absorption spectra of tissues. The peak emission spectra range of well-characterised luciferases is also shown (shaded in grey) and is greatly influenced by haemoglobin absorption.

Water, of which mammals are more than 50% by mass, begins to strongly attenuate light at 900-1400nm (depending on tissue type) and this therefore sets the upper limit of the bio-optical window. Between $\lambda = 600-800$ nm, the absorption of light by Hb decreases by a factor

of approximately 50 (Jathoul et al. 2014). Further, longer wavelengths of light scatter less as they pass through tissue, therefore red-shifted luciferases give sharper resolution BLI of the bioluminescent cells, and more reliably represent the structures beneath the skin. Given these advantages of red-emitting bioluminescence it is of great potential interest for researchers for in vivo imaging. Since relatively few naturally occurring beetle luciferases emit in this window, two approaches exist to shifting emission to longer wavelengths, namely – engineering the luciferin and/or the luciferase, these are discussed here.

The beetle luciferin biosynthetic pathway remains elusive to researchers and as such the D-LH_2 used by researchers and other end-users is produced chemically. Chemists are able to produce a plethora of synthetic analogues including (but not limited to) 6-amino- D-luciferin and infra-luciferin, which have λ_{max} of 610 nm and 706 nm respectively (Mofford et al. 2014a), whereas D-LH_2 has an λ_{max} of 565 nm. A selection of beetle luciferin analogues is detailed in Figure 1.8 and table 1.1.

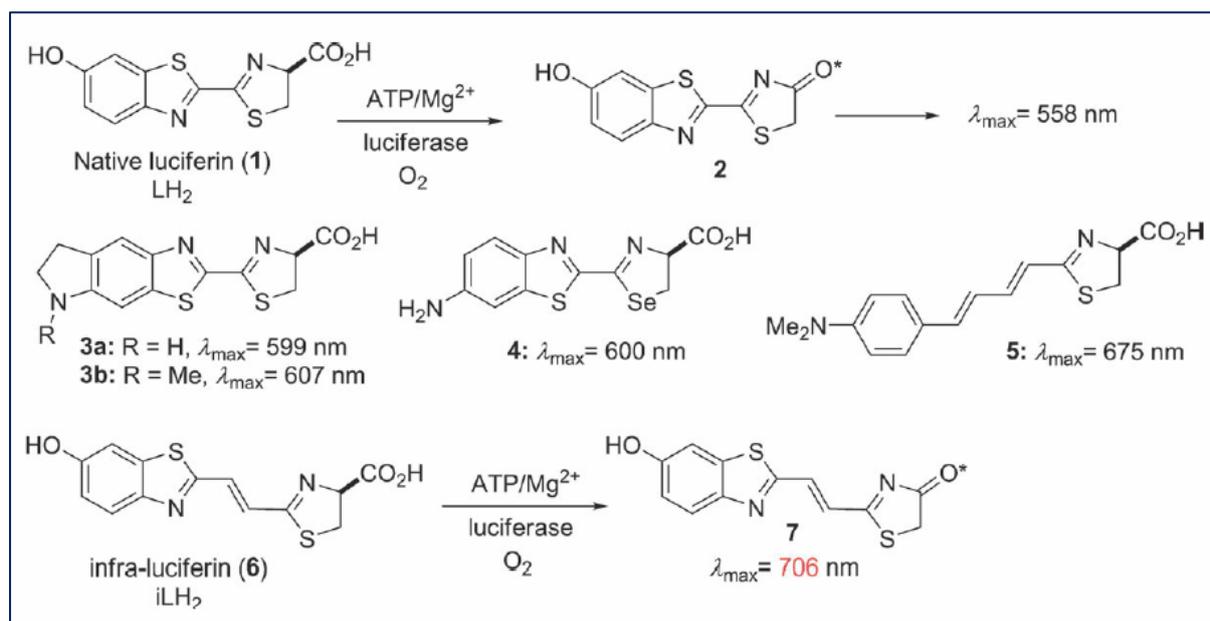


Figure 1.8 taken from (Jathoul et al. 2014). Structure and λ_{max} of beetle luciferin (D-LH_2) and a selection of analogues. The full name – and further details – of each is listed in table 1.1 (below).

No. if in Figure 1.8	Luciferin analogue	λ_{\max}
1 and 2	Beetle luciferin	558 nm
3a	Cyclic aminoluciferin R = H (hydrogen)	599 nm
3b	Cyclic aminoluciferin R = Me (methyl group)	607 nm
4	seleno-D-aminoluciferin	600 nm
5	A 4-(dimethylamino)phenyl derivative conjugated to a thiazoline group	675 nm
6 and 7	infra-luciferin	706 nm*
-	6-amino-D-luciferin	610 nm
-	AkaLumine™ (TokeOni)	677 nm

Table 1.1. λ_{\max} of beetle luciferin (D-LH₂) and a selection of analogues. The structure of each luciferin is shown in Figure 1.8 (above), with the exception of 6-amino-D-luciferin and AkaLumine™. The asterisk (*) indicates that the λ_{\max} was achieved with a mutant Fluc engineered for use with iLH₂.

Although many of these analogues result in dramatic red-shifting of the λ_{\max} , they are also often patent-protected and are therefore expensive to purchase. For instance, one of the most respected analogues for in vivo applications currently is AkaLumine™ (TokeOni) which at the time of writing can be purchased for £13,000 per gram from Merck & Co., Inc. In contrast, beetle luciferin can be purchased – at the time of writing – for £361 per gram from Promega Corporation. Such luciferin analogues are therefore cost-prohibitive for many research projects to depend on, including this one.

With the creation of the beetle luciferin analogues came luciferin-analogue-optimised luciferases. These are enzymes engineered for use with a specific luciferin analogue, which can both increase the brightness and further red-shift the light emitted. A notable example of such an enzyme is Akaluc™, specifically made for use with Akalumine™ (Iwano et al. 2018). It is also possible to red-shift the emitted light from D-LH₂. The red-shifting of beetle and firefly luciferases – such as Click Beetle Red (CBR) (Zhao et al. 2005), x11 Fluc (Jathoul et al. 2012), and DeltaFlucs (Halliwell et al. 2018) – has typically been achieved through site-directed mutagenesis (SDM).

1.5.3 Thermal and pH stability

Fireflies are insects and thus “cold-blooded” – normally emitting light at dawn, dusk, and at night when air temperatures are relatively cool. The firefly biochemistry is accordingly adapted to relatively cool temperatures, in contrast to mammals where the optimal temperature for enzymes is 37 °C. Although Flucs may be brighter at elevated temperatures – due to a higher substrate turnover rate (k_{cat}) – (as is commonly found with other enzymes) the enzyme is unstable and denatures relatively rapidly (Kitayama et al. 2003). Broadly speaking, both thermal and pH stability engineering involve increasing the rigidity of the non-moving parts of the enzyme, and increasing its overall solubility. Many of the changes are difficult to predict and many advances have been made through random mutagenesis. Fluc mutants of relevance are “x11Fluc” and “x12Fluc” (Jathoul et al. 2012) these are based on WT *Ppy* Fluc, and the number in the name refers to the number of residue substitution mutations between it and WT *Ppy* Fluc. “x11Fluc” and “x12Fluc” are identical apart from one substitution. The majority of the mutations (9 out of 12, or 9 out of 11 respectively) occur on the surface of the folded protein. Other technologies to improve stability include the circularisation of proteins whereby the N and C termini are bonded (with an isopeptide bond) post-translationally (Si et al. 2016).

1.5.4 Gene silencing

There are multiple mechanisms by which genes can be silenced (Meister and Tuschl 2004) such as small interfering RNA and microRNA (miRNA). As an exogenous gene, this is another disadvantage which once reduced the applicability of Fluc in certain models. Unlike most immune responses which occur at the protein level, this occurs at the DNA level. Thus it affects both in vitro and in vivo studies. There are some enhancements currently available for the suppression of gene silencing such as codon optimisation, and the removal of regulatory signals and consensus glycosylation sites. For mammalian expression the circumvention of gene silencing has proven relatively simple, however for certain plant species expression-specific cDNA sequences may be required (Chou and Moyle 2014).

1.5.5 Hypoxia resistance

The reaction catalysed by Fluc requires molecular oxygen and in vivo this is provided by the inherent oxygenated nature of tissue in living creatures. However, tumours – commonly labelled with Fluc by researchers – are often hypoxic due to their rapid uncontrolled growth and thus there is an inability of blood vessels to spawn (angiogenesis) and efficiently transport oxygenated blood inside. As a result, the longitudinal light signal from such a luciferase-transformed tumour may eventually cease increasing linearly with tumour growth. As a result, unless/until the tumour is excised the researcher may be lead into believing that tumour growth has plateaued when it in fact still growing rapidly. This is an inherent limitation of the luciferase system. Some attempts to alleviate this include promoters for the luciferase gene which are linked to hypoxia, i.e. as oxygen decreases, more luciferase is produced to compensate (Coulet et al. 2003);(Lehmann et al. 2009).

1.6 Limitations of the Firefly Luciferase System for Mammalian Applications

So far we have discussed the improvements made to the firefly luciferase system with respect to colour, thermal and pH stability, and brightness. Biocompatibility refers to the performance of an agent in a specific biological system and is a broad term with regards to luciferases in mammalian systems. Crucially, it includes compatibility with the host immune system – i.e. not stimulating an unwanted immune response. Due to the current *incompatibility* of the Fluc system, it is often uncompetitive when compared to other imaging technologies such as PET (positron emission tomography) / SPECT (single photon emission computed tomography) which is more expensive and less safe (for both the researchers and animals due to ionising radiation). PET / SPECT probes are not directly genetically encodable – unlike the luciferase system where the luciferin only emits when in contact with the localised luciferase, PET / SPECT radiolabelled probes emit constantly regardless of location but can be localised via the use of genetically-encoded probe-binding ligands. Perhaps the most commonly used biocompatible alternatives to Flucs are fluorescent proteins which, despite being genetically encodable, have a far reduced signal-to-noise ratio and are thus less sensitive.

1.6.1.1 Effects upon stem cells

Stem cells present a particularly sensitive tissue-type for labelling with genetic reporters. There is conflicting evidence about whether the current Fluc system may affect the viability, growth, and differentiation of implanted stem cells (Brutkiewicz et al. 2007);(Tiffen et al. 2010);(Lee et al. 2009);(Bradbury et al. 2007), the most serious cases of the latter result in the formation of teratomas – tumours made up of several different types of tissue.

1.6.1.2 Anti-luciferase immune response

Fluc expression minimally affects cell growth in most cell types commonly used by researchers (Tiffen et al. 2010), however in BALB/c and C57/BL6 strains of mice, cells transformed with the WT *Ppy* Fluc gene have been reported to be eventually identified by the adaptive immune system, which normally takes approximately four weeks to occur (Podetz-Pedersen et al. 2014) but can vary on the degree of antigen presentation; some cancers are well-shielded from the immune system whereas others are “leaky”. This response may be noticeable to the researcher as a decrease in longitudinal bioluminescence intensity. The immune response to Fluc is noted as “one of the most serious problems preventing Fluc deployment in mammalian systems” (Zhang and Wu 2007), to date there has been no effort to address it through protein engineering.

As an internally-expressed protein, it is solely the T_H1 response which is relevant for potential immunological rejection of Fluc in mammals. Accordingly, Fluc is processed via the proteasome and presented via “TAP” on the major histocompatibility complex (MHC) class I to naive T cells, which begins the mounting of a T_H1 immune response eventually resulting in the killing of luciferase-expressing cells by cytotoxic (CD8+) T cells (Dubey 2012) (Figure 1.9). A protein’s entire primary sequence must be free of identifiably-foreign epitopes in order to remain undetected by the T_H1 pathway. Several epitopes in WT *Ppy* Fluc are responsible for the immunogenicity in the BALB/c and C57BL/6 mice strains, and have been identified (Limberis et al. 2009). As stated, the length of time post-implantation or post-transformation for the response to be noticeable varies between individual mice, but is commonly observed from the four-week mark (Podetz-Pedersen et al. 2014).

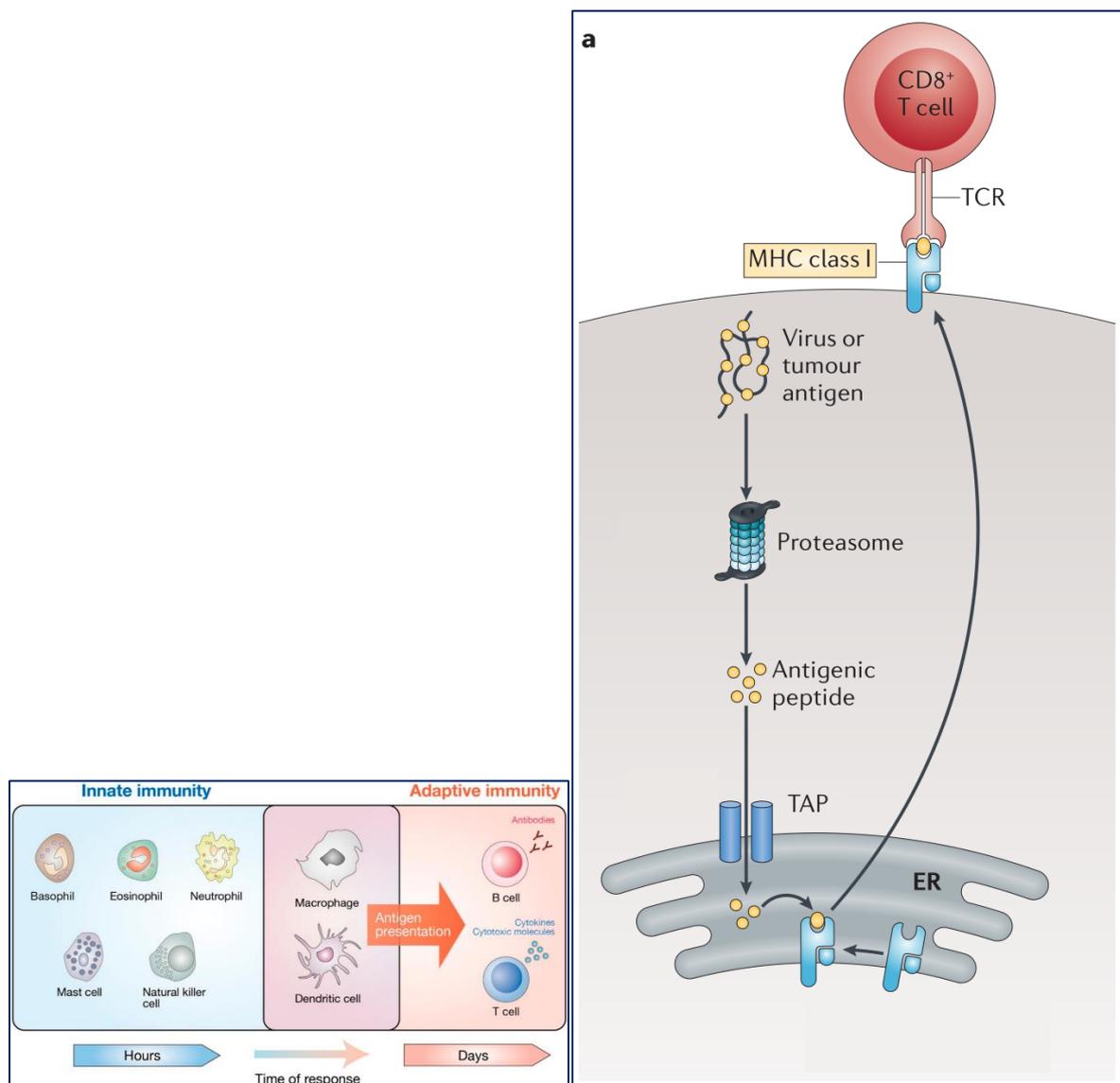


Figure 1.9 taken from (Yamauchi and Moroishi 2019) and adapted from (Kobayashi and Van Den Elsen 2012). **Left: activation of the adaptive immune system; right: the MHC class I antigen-presentation pathway.** Left boxed: antigen presenting cells (APCs) such as macrophages and dendritic cells present processed antigens on the major histocompatibility complex (MHC) class I to T cells, initiating an adaptive cellular immune response. Right boxed: intracellular antigens are processed into peptides by the immunoproteasome, these are then delivered to the endoplasmic reticulum (ER) by the transporter associated with antigen processing (TAP) where they are loaded onto the MHC class I for presentation to T cells via the T cell receptor (TCR).

For in vivo imaging applications the immunogenicity of Fluc presents a time-limit beyond which most animals in the cohort will become sensitised and destroy any transformed tissue.

For instance, in a hypothetical cancer drug study Fluc-transformed tumours may eventually be destroyed by the immune system regardless of whether or not an anti-cancer drug is used. Therefore the researcher can either use immunocompromised mice or make certain that the study can be completed before the immune response starts. For studies which last long enough for sensitisation – in at least some mice – to occur, large cohort sizes must be used to compensate for the variability (and correspondingly high statistical random error) in sensitisation time for mice. These complications result in a limited scope for the use of the luciferase system, less reliable data, and ethical concerns. Thus, less potently immunogenic fluorescence options may be chosen instead (Ansari et al. 2016).

Anti-luciferase antibodies may be produced in Fluc-transformed animals. For instance Fluc-transformed tumours often lyse and leak free Fluc protein due to structural defects. Despite this, the antibodies are essentially inconsequential as luciferase-transformed cells are internally inaccessible to antibodies. There are some applications where this T_H2 response is of importance such as membrane-bound luciferases (Pellegatti et al. 2005).

1.7 Current Options for Countering the Anti-Luciferase Immune Response

1.7.1 Transgenic reporter-tolerised animals

Through a process known as “central tolerance”, mammals resist the development of autoimmune T and B cells. During foetal development, this process results in the immune system registering all antigens present as “self”, which can be exploited by researchers to “trick” the immune system into registering Fluc as “self”, as is discussed here.

“Glowing Head” (GH) mice (Day et al. 2014) are transgenic mice which express Fluc (and GFP) in the anterior pituitary gland of the brain. As a result the immune system registers Fluc as a self-protein during immune system development and becomes centrally and peripherally tolerant to it. Thus, Fluc-transformed homografts can be implanted without rejection. Fluc-transformed xenografts such as HEK cells will still stimulate an immune response as the MHC itself is recognised as foreign, as are many of the human peptides expressed on it.

The homograph tissue can be longitudinally imaged and the light intensity compared to that of the head, which thus serves as a benchmark and a positive control. However, it could present problems for imaging less bright tissues as the bright output of the head may cause both the sensor sensitivity and aperture of the camera to reduce, which may therefore reduce the number of photons captured from the implant. This could be potentially mitigated by covering the head with an opaque material.

However, creating transgenic animals such as the GH mouse is expensive, time-consuming, and requires advanced techniques and equipment. These include embryo transfer, transduction of the gene into the embryo (which must be inserted at a specific locus for localised expression, as in GH mice), and implantation of the embryo to a pseudo-pregnant mother. To create a transgenic line, multiple transgenic animals must be created to allow breeding. Thus, for many researchers, it is inconvenient or infeasible to create such a transgenic line for their preferred strain of animal.

1.7.2 Immunocompromised animal strains

Immunocompromised mice exist on somewhat of a spectrum of immunocompetence; they may have specific features of the innate and/or adaptive immune system knocked-out or altered in some way. One commonly used immunocompromised mouse strain in research is the NOD scid gamma (NSG) mouse. NSG mice lack natural killer cells and many cytokine signalling pathways, and crucially for luciferases they effectively have a totally non-functioning adaptive immune system (Shultz et al. 2005). NSG mice represent one of the most immunocompromised strains created to date, despite this they are relatively healthy and long-lived.

However, immunocompromised mice (such as NSG mice) are more expensive and difficult to experiment with than their more immunocompetent counterparts. Often they must be maintained in a perfectly sterile environment to protect them from pathogenic infection. Thus, they must be housed and handled in airtight isolators, the air must be high efficiency particulate air (HEPA) filtered, and their food, bedding, cages etc. must be either gamma-irradiated or autoclaved. Unless carried out shortly before sacrifice, experimental treatments

must also be sterile which can prove problematic as some reagents may be damaged by autoclaving; irradiation is often not easily available.

Immunocompromised animal strains may be unsuitable for studies in which the immune response to the implanted tumour itself (not the Fluc) is of interest. However it may be possible to use strains in which the components of interest are still functional whilst the anti-Fluc response is sufficiently suppressed.

1.7.3 Immunosuppressive Drugs

One of the most commonly administered immunosuppressive drugs for animal studies (and in humans for autoimmune diseases) is cyclophosphamide (CP). Drugs like CP are used for suppressing the immune response against allografts and xenografts in both human medicine and animal research. It is currently the only viable option for long-term (on the order of weeks) implantation of luciferase in otherwise immunocompetent mouse strains such as BALB/c. Immunosuppressed animals bear the same limitations as genetically immunocompromised animal strains, that of the need for sterility and the potential irrelevance for studies involving the immune system.

1.8 Potential for Luciferases with Reduced Immunogenicity

The exact strategy the engineer chooses to employ will be determined largely on the nature of the protein and what is currently understood about it. Certain proteins are difficult to screen and therefore do not lend themselves well to the high-throughput screening often required by directed evolution, whereas others – including luciferases – do. It is an assumption of protein engineering (and especially of enzyme engineering) that the overwhelming majority of protein sequences will be non-functional for a given application, and thus the more drastically a protein is altered randomly, the higher the probability that the modification is deleterious. Thus, engineering an inherently non-immunogenic protein into a Fluc is seemingly a high-risk strategy.

The alternative approach is to engineer Fluc itself for reduced immunogenicity. Two potential strategies which are explored here are to: engineer the protein to remove potentially

immunogenic peptides, or to engineer Fluc to be more similar to a cognate mammalian enzyme – Acyl-CoA synthetases are obvious candidates.

1.9 Acyl-CoA Synthetases (ACSs)

1.9.1 Purpose and function of Acyl-CoA synthetases

Acyl-CoA synthetases (ACSs) (formally known as acyl-CoA ligases) are present in all organisms and serve to activate fatty acids to generate fatty acyl-CoA. The fatty acyl-CoA is then beta-oxidised by separate enzymes to generate acetyl-CoA which enters the citric acid cycle to generate metabolic energy. Due to the varying carbon chain lengths of fatty acids, organisms express an assortment of ACSs which have specificity for a range of lengths (with many cases of overlap). The mammalian equivalents are shown below (table 1.2).

Name prefix of acyl-CoA synthetase enzyme	Initialisation of acyl-CoA synthetase enzyme	Carbon chain length of fatty acid substrate
Short-chain	ACSS	C2 – C4
Medium-chain	ACSM	C4 – C12
Long-chain	ACSL	C12 – C20
“Bubblegum”	ACSBG	C14 – C24
Very-long-chain	ACSVL	C18 – C26

Table 1.2. Categories of fatty acids and corresponding acyl-CoA synthetases by chain length.

Note that many ACSs also have activity with fatty acids outside their designated range.

1.9.2 Genetic, Functional, and Structural Similarities between Acyl-CoA Synthetases and Firefly Luciferases

ACSs and Flucs are members of the ANL (acyl-CoA synthetases, nonribosomal peptide synthetase adenylation domains, and luciferase enzymes) superfamily. It is believed that beetle luciferin may have existed prior to its use as the substrate in bioluminescence, serving its function as an antioxidant. In bioluminescent beetles the protein sequence but also the cDNA of the luciferase and ACS is highly homologous, and it is believed that Flucs have evolved from an ancestral ACS via a gene duplication event, after which one ACS gene copy could serve its original essential metabolic function whilst the other was allowed the redundancy to develop its new function as a bioluminescent enzyme. Thus, the stage was set for firefly

luciferase to evolve. The reaction catalysed by ACS bears a remarkable similarity to that of Fluc, both involve an initial energisation of the substrate via the ligation of adenosine monophosphate (AMP) and release of inorganic pyrophosphate (PP_i). This is then followed by oxidation for Fluc or thioesterification for ACS. These reactions are compared below (Figure 1.10);(table 1.3).

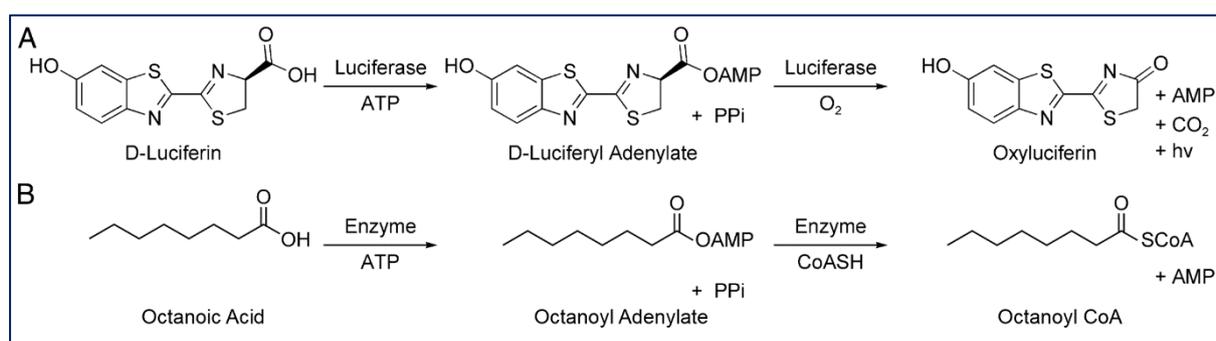


Figure 1.10 adapted from (Mofford et al. 2014b). **Comparison of ACS and Fluc respective reactions.** (A) Firefly luciferase catalyses the formation of an activated AMP ester of its native substrate, D-LH₂. Subsequent oxidation within the luciferase binding pocket generates an excited-state oxyluciferin molecule that is responsible for light emission. (B) Fatty acyl-CoA synthetases catalyse the formation of activated AMP esters from fatty acids such as octanoic acid. AMP is then displaced by CoASH to form the fatty acyl-CoA product.

Name of enzyme	Reaction step 1	Reaction step 2
Firefly luciferase	$\text{LH}_2 + \text{ATP} \rightleftharpoons \text{LH}_2\text{-AMP} + \text{PP}_i$	$\text{LH}_2\text{-AMP} + \text{O}_2 \rightleftharpoons \text{LO} + h\nu + \text{CO}_2 + \text{AMP}$
Acyl-CoA synthetase	$\text{FA} + \text{ATP} \rightleftharpoons \text{FA-AMP} + \text{PP}_i$	$\text{FA-AMP} + \text{CoA} \rightleftharpoons \text{FA-CoA} + \text{AMP}$

Table 1.3. Comparison of ACS and Fluc respective reactions. The description for this table is the same as that of Figure 1.10.

Interestingly, beetle luciferin is also capable of coenzyme A ligation as shown in Figure 1.4. The presence of CoA increases the brightness of Fluc in vitro (Fraga et al. 2005), however this is not due to an allosteric effect, rather it is due to CoA quenching dehydroluciferyl-adenylate (L-AMP) – a bioluminescence by-product which strongly inhibits further bioluminescence – to dehydroluciferyl-CoA (L-CoA). It is unknown from the literature whether murine ACSs adenylate D-LH₂, however it is known that Flucs are bi-functional enzymes capable of fatty acid activation, and that D-LH₂-AMP is weakly spontaneously chemiluminescent at

physiological pH (Viviani and Ohmiya 2006). As stated, the protein sequences of beetle Fluc and ACSs are highly homologous. Both proteins are composed of a larger NTD and smaller CTD connected by a flexible linker. The flexible linker allows all enzymes in the ANL superfamily to perform a 140° conformation change following substrate adenylation. Flucs have two binding conformations allowing adenylation and oxidation of the substrate respectively; ACSs have two which allows adenylation and thioesterification respectively. The overall structures of insect Flucs and ACSs are extremely similar and the substrate binding site is analogous (Figure 1.11).

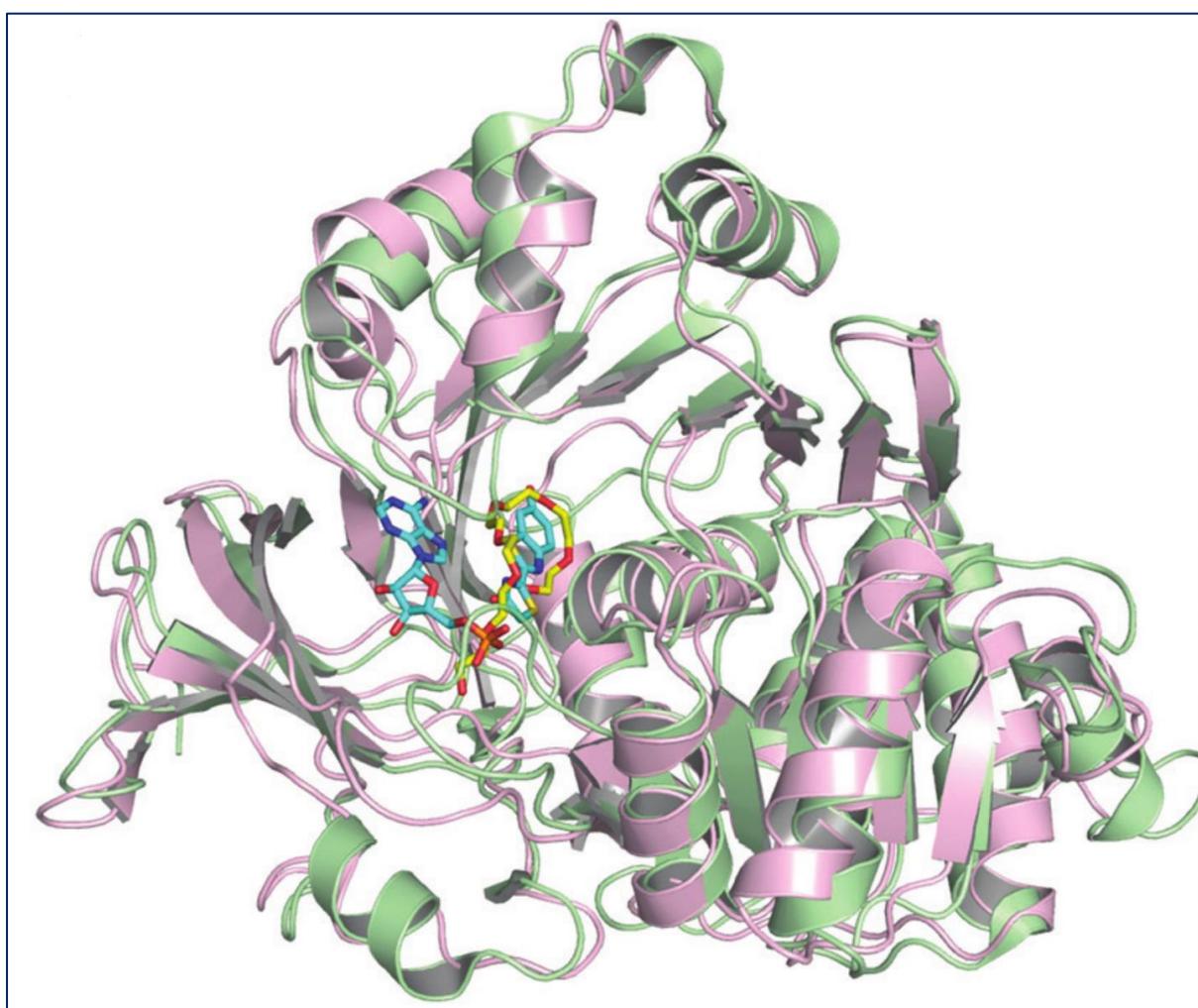


Figure 1.11 taken from (Prado et al. 2016). **Structural alignment of insect ACS and Fluc N-terminal domains.** The N-terminal domain (NTD) of *Zophobas morio* protoluciferase (an ACS) (in pink) structurally aligned with the NTD of *Luciola cruciata* Fluc (in green). The former is complexed with a fatty acid (carbon atoms in yellow), and the latter is complexed with AMP and oxyluciferin (carbon atoms of both in cyan).

1.9.3 Evidence of the Bioluminescent Potential of Acyl-CoA Synthetases

An ancestral beetle ACS evolved into the range of Flucs displayed by bioluminescent beetles today, this is supported by the similar biochemistry of the respective reactions (Wood 1995), and the presence of conserved residues between current ACSs and Flucs with mutagenic studies confirming that the function of such enzymes can be altered with specific mutations (Viviani 2002). Two insect ACSs from the fruit fly (*Drosophila melanogaster*) and click beetle (*Agrypnus binodulus*) were successfully converted to Flucs by the substitution of a highly-conserved serine (S345) from Flucs into the equivalent positions (Oba et al. 2009). Indeed, similar enzymes occur naturally – beetle ACSs exist which are putative protoluciferases in mealworms (*Tenebrio molitor* and *Zophobas morio*), these display dim bioluminescence with D-LH₂ and ATP (Viviani and Bechara 1996);(Viviani et al. 2009) indicating substrate promiscuity. Thus, it was theorised that other insect ACSs may have the potential for mutation into functional luciferases with the potential for high brightness achieved through directed evolution (Viviani et al. 2013). It was demonstrated that naturally non-bioluminescent fruit fly (*Drosophila Melanogaster*) ACS CG6178 can indeed exhibit bioluminescence with certain synthetic luciferin analogues such as cyclic luciferin – a luciferin that more-closely resembles the native fatty acid substrate (Mofford et al. 2014b) (Mofford et al. 2017). A hybrid “chimeric” protein composed of the N-terminal domain of *Photinus pyralis* Fluc and the C-terminal domain of *Drosophila melanogaster* ACS (CG6178) retained 4% of the activity of the WT *Ppy* Fluc. The N-terminal domain of *Ppy* Fluc is 0.03% as active as WT *Ppy* Fluc (Zako et al. 2003). Despite the successes in bringing bioluminescence to otherwise dark insect ACSs, there is evidence that insect ACSs are “special” and are not orthologues of other (for instance mammalian) ACSs (Oba et al. 2005). Thus, the potential for the conversion of mammalian ACSs to Flucs remains uncertain.

1.10 Aims and Objectives

For tissue transplant experiments in small animals, luciferases are presently considered the “gold standard” owing to their high sensitivity and dynamic range allowing accurate quantification of transplant growth. This accuracy – and the ability to quantify tissue without excision – allows longitudinal experiments with relatively small cohort sizes, *reducing* the total number of animals needed for a given experiment. The system is relatively non-invasive

thereby minimising the suffering endured by the animal; for transplant experiments the transplant is usually labelled *ex vivo* and then injected, imaging itself requires only that the animal be injected with the luciferin immediately prior. Luciferase labelling is a well-established and highly-*refined* technology with purpose-built imaging equipment available – such as the Biospace Lab PhotonIMAGER (used in this project).

Reduction of the number of animals required and ethical refinement of experimental procedures constitute two of the three “Rs”, with the third being replacement of animals entirely. Improvement of these three Rs is a goal for researchers conducting animal experiments. The applicability of luciferase for small animal experiments has been made possible by several engineering advancements which are discussed here.

Photinus pyralis (*Ppy*) firefly luciferase (Fluc) has been the subject of various protein engineering studies aimed at enhancing its properties for optical imaging applications. The thermal stability and colour have been separately addressed leading to significant improvements for biomedical applications (Jathoul et al. 2012);(Baggett et al. 2004). However, to date there have been no reported attempts to reduce the inherent immunogenicity of Fluc to mouse (*Mus musculus*) strains through enzyme engineering.

The aim of this work is twofold: to use protein engineering to explore the potential of mammalian enzymes – specifically murine ACS – to function as beetle luciferases, and thus create a luciferase with reduced immunogenicity (T_H1 response) within immunocompetent BALB/c mice.

If such a luciferase is generated through this strategy, it would ideally have the following properties for it to be of use to researchers conducting *in vivo* experiments with immunocompetent BALB/c mice:

- ≥10% as bright as WT *Ppy* Fluc with D-LH₂
- Thermostable at 37 °C – for efficient expression in mammals
- Redshifted in its light emission, ideally within the bio-optical window – for efficient transmission through mammalian tissue
- More efficient than *Ppy* Fluc in terms of initial protein expression in mammalian cells

- Not silenced in different mammalian cell types
- Minimal in its effect on cell growth or function
- Compatible with infra-luciferin – to further redshift emission

A firefly luciferase with reduced immunogenicity and the properties listed above would have immediate application for improving the use of animals in science in terms of ethics, informativeness, reproducibility, and/or sensitivity of *in vivo* studies. Any sensitive reporter that is highly homologous to a native cellular mouse or human enzyme will be of great benefit to research using small animal models. It will not itself represent an antigen and can therefore be used more accurately to determine the *in vivo* responses to different genes or agents. Such a reporter system will extend the duration of studies in leaky mice. As immune attack will hinder the normal function of transplanted cells *in vivo* and could reduce the efficacy of cellular therapies and their detection *in vivo*, many studies can then also be carried out in non-immune compromised mice and bring greater clinical relevance. However, this work would impact wider areas of biomedical research, for example, small molecule, gene and T-cell chimeric antigen receptor (CAR) therapies or prophylactics targeting Fluc-tagged cells will benefit from improved signal to noise ratios (i.e. reduced cell death *in vivo* caused by the murine immune system and not the agent). This would otherwise skew data compared to *in vitro* results and may lead to over estimation of therapeutic effect, which can then impact on subsequent clinical success rates, often at great cost to invested parties and leading to a requirement for further animal studies to prove effectiveness. This could in turn refine and reduce testing in primates and other animals.

As described earlier, the immunogenic epitopes in Fluc for both BALB/c and C57BL/6 mice strains have been reported (Limberis et al. 2009) and one approach to constructing a non-immunogenic luciferase would be to either mutate to replace, or delete to remove these regions from Fluc. However, even if this is successful in preserving bioluminescence function, there is no guarantee that new epitopes for other mouse strains have not been generated. There is precedent for the use of non-immunogenic mammalian enzymes as reporter genes, however these currently rely on chemiluminescence with low signal-to-noise ratio (Mælandsmo et al. 2005);(Christou and Parks 2011). Therefore, a potentially superior approach relies on engineering murine ACS proteins to contain bioluminescence function,

which structurally maps to regions of WT *Ppy* Fluc. Rational recombinant proteins will be cloned and screened for light output in *E. coli* and/or HEK cells, viable mutants may undergo further cycles of rational and/or random mutations with screening. Academically, this work will also improve our understanding of the structure-function of *Ppy* Fluc, specifically the regions of critical importance that are required for bioluminescence (determinants).

2 Chapter 2 – Materials and Methods

2.1 Materials

2.1.1 Materials related to purified DNA

2.1.1.1 Reagents and buffers

Many common items used throughout this project and not ordered from Sigma (with the exception of TruPAGE pre-cast gels) are shown in table 2.1 below.

Reagent or buffer	Manufacturer	Storage temperature
Agarose powder	Bioline	RT
BL21 (DE3) cells	New England Biolabs	-80 °C
Coloured protein molecular weight markers	New England Biolabs	-20 °C
Diamond nucleic acid dye	Promega	-20 °C
DH5 α cells	New England Biolabs	-80 °C
DNA Smart Ladder molecular weight marker	Eurogentec and Bioline	4 °C
Foetal calf serum	Gibco	-80 °C
HisPur Ni-NTA resin	Thermo Scientific	4 °C
Hybond nylon membrane	GE Healthcare	RT
Iscove's Modified Dulbecco's Medium	Gibco	4 °C
Penicillin-streptomycin antibiotic solution	Gibco	4 °C
Phusion polymerase	New England Biolabs	-20 °C
Q5 polymerase	New England Biolabs	-20 °C
SafeView dye	NBS Biologicals	4 °C
10% TruPAGE pre-cast gels	Sigma	4 °C
<i>Taq</i> polymerase	Qiagen	-20 °C
Ultrapure water	Millipore	RT

Table 2.1. Manufacturers of the general reagents and buffers used. Those not listed within this table are listed below, or elsewhere where relevant.

Restriction enzymes and associated buffers were obtained from New England Biolabs, with the High-Fidelity (HF[®]) versions of enzymes being preferred. Chemical and biochemical reagents were obtained from Sigma Chemical Co. unless otherwise stated. Stock 10 mM deoxyribonucleotide triphosphates (dNTPs) were prepared by premixing adenine, cytosine,

thiamine, and guanine triphosphates from a Qiagen PCR kit in equal proportions and diluting appropriately; solutions were further diluted if required.

2.1.1.2 Oligonucleotide primers

Primers were obtained from Integrated DNA Technologies as a lyophilised stock and were dissolved in ultrapure water to 100 μ M for long-term storage stocks and 10 μ M for working stocks, and both stored at -20 °C. Generic T7 primers were used to sequence genes within the pET16b vector; these were included with the Eurofins sequencing service and not ordered into the lab.

The primers were named in the format JL (for James Long) followed by a three digit number (indicating the sequential number the primer was ordered) and finally either an F or R (for forward direction (5' to 3') or reverse direction (3' to 5')). For example JL002R, this was the second primer ordered from IDT and amplifies in the 3' to 5' direction.

A comprehensive list of all the primers ordered (but not necessarily used) throughout this project is shown in table 2.2 below. Specific primer names and purpose are cited in the text of the results chapters.

Name	Primer sequence
JL001F	GGTCGTCATATGGTGATGTTACTTCGCGC
JL002R	GCAGCCGGATCCAGTTACATTTTATGTTGTG
JL005F	GGTCGTCATATGGCGTGT
JL006R	GCAGCCGGATCCAGTTAC
JL007F	ATAGCTCGTTCGGGCTTAGATTAAGCGTCAATGGC
JL008R	GCCATTGACGCTTAATCTAAGCCCGAACGAGCTAT
JL009F	TGGAACACGTCCGACACAGGACATGGATTTTCTGCTTTTACCTCTGTATTCTCG CCCTGG
JL010R	CCAGGGCGAGAATACAGAGGTAAGCAGAAAATCCATGTCCTGTGTCTGGACGT GTTCCA
JL011F	TCTAAATTTCCGATTACGGTGTTTTTAAGCGCGCCGACTG
JL012R	CAGTCGGCGCGCTTAAAAACACCGTAATCGGAAATTTAGA
JL013F	ATTTAACAGTCTTAAACACGAAGCATCCGGAGGGGAGCCTATTAAT
JL014R	ATTAATAGGCTCCCCTCCGGATGCTTCGTGTTTAAAGACTGTTAAAT
JL015F	TGGGCTGACCGAGACAACCTTCTGCTATCTGCGGGAACCTCAAAGG
JL016R	CCTTTGAAGTCCCAGATAGCAGAAGTTGTCTCGGTCAGCCCA

JL017F	AGCTCGTTTCGGGCTTGGATTAAGCGTCAATAGAGATCCTATTTTTGGCAATCAA ATCATTCCGGATACTGTTATGTGGAACACGTCCGACACAGGATGG
JL018R	CCATCCTGTGTTCGGACGTGTTCCACATAACAGTATCCGGAATGATTTGATTGCC AAAAATAGGATCTCTATTGACGCTTAATCCAAGCCCGAACGAGCT
JL019F	AGCTCGTTTCGGGCTTGGATTAAGCGTCAATAGAGATCCTATTTTTGGCAATCAA ATCATTCCGGATACTGCGATTTTAAGTACAGGATGGGCAAAGTCTGCTTGGAGC TCT
JL020R	AGAGCTCCAAGCAGACTTTGCCCATCCTGTACTTAAAATCGCAGTATCCGGAAT GATTTGATTGCCAAAAATAGGATCTCTATTGACGCTTAATCCAAGCCCGAACGA GCT
JL021F	GCCCTGGACACAGGGATTTTTGCGTGTTCACATTACTTAC
JL022R	GTAAGTAATGTGCAAACACGCAAATCCCTGTGTCCAGGGC
JL023F	CCGGGCTGGACATCAGAGAGGGTTATGGGCTGAC
JL024R	GTCAGCCCATACCCTCTCTGATGTCCAGCCCGG
JL025F	CATAGCTCGTTCGGGCTTAGATTAAGCGTCAATAGAGATCC
JL026R	GGATCTCTATTGACGCTTAATCTAAGCCCGAACGAGCTATG
JL027F	ACTGCGATTTTAAGTACAGGACATGGATTTTCTGCTTTTACCTCTGTATTCTCG CCCTGG
JL028R	CCAGGGCGAGAATACAGAGGTAAAAGCAGAAAATCCATGTCCTGTACTTAAAAT CGCAGT
JL029F	GATATACCATGGGCCATCATCATCATCATCATCATCAT
JL030F	AAGTCTTTAATTAAAAGCAGCGGTACCGTATCGG
JL031F	GACCGCTTGAAGTCTTTAATTAAATAAAAAGGATCCCAGGTGGCCCCCGC
JL032R	GCGGGGGCCACCTGGGATCCTTTTTATTTAATTAAAGACTTCAAGCGGTC
JL033F	CTATGAGGGTTATGGGCTGACCGAAACAAC TAGTCTTATCTGCGGGAAC T
JL034R	AGTTCCCGCAGATAAGACTAGTTGTTTTCGGTCAGCCCATACCCTCATAG
JL035F	CCAGGGGGATCCGCCGCCATGGATGGAAGACGCCAAAAACATAAAGAAAGGC
JL036R	TCCTGATGATGATCCTCCTGATGATGATCCTGTTGTGACCCACTCCTTTTTGCG TA
JL037F	GGATCATCATCAGGAGGATCATCATCAGGAATGGTGAGCAAGGGCGAGG
JL038R	GTTGACGCGGCCGCTTTACTTGTACAGCTC
JL039R	TCCTGATGATGATCCTCCTGATGATGATCCGGCAGCAGCGGAGTTTTCC
JL040F	CCAGGGGGATCCGCCGCCATGGATGGTGATGTTACTTCGCGCACG
JL041F	CCAGGGGGATCCGCCGCCATGGATGCACCATCTTTGGAAGATCCCGC
JL042F	AATCACCGACCTCTCTCCCCAGGGGGATCCGCCGCCATGGATGGAAGACGCCAA AACAT
JL043R	AATCCAGAGGTTGATTGTTCGACGCGGCCGCTTTACTTGTACAGCTCGTCCATGC CGAGAG
JL044F	AATCACCGACCTCTCTCCCCAGGGGGATCCGCCGCCATGGATGCACCATCTTTG GAAGAT
JL045F	AATCACCGACCTCTCTCCCCAGGGGGATCCGCCGCCATGGATGGTGATGTTACT TCGCGC
JL046F	GCGGCCGCGTCGACAATCAACCTCTG
JL047R	GGATCCCCCTGGGGAGAGAGGTCGGT
JL048F	CATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGGAAGACGCCAAA AACATA

JL049R	GCCAACTCAGCTTCCTTTTCGGGCTTTGTTAGCAGCCGGATCCAGTTACATTTTCAATTT
JL050F	CATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATATGGTGATGTTACTTCGCGCA
JL051R	GCCAACTCAGCTTCCTTTTCGGGCTTTGTTAGCAGCCGGATCCAGTTACATTTTATGTTGT
JL052F	TAACAAAGCCCGAAAGGAAGCTGAGTTGG
JL053R	TTCGATATGGCCGCTGCTGTGATGATG
JL054F	GAAAGGGATCATCATCAGGACGCGTTGGATCATCATCAGGAATGG
JL055R	TTCCATGGCGGCGGATCCCC
JL056F	CGGCCATATCGGATCCGCCGCCATGGAAGACGCCAAAAACATAA
JL057F	CGGCCATATCGGATCCGCCGCCATGGTGATGTTACTTCGCG
JL058F	CGGCCATATCGGATCCGCCGCCATGCACCATCTTTGGAAGATC
JL059R	CATATAACGTAACGCGTCCTGATGATGATCCCTTTCCGCCCTTCTTGGC
JL060R	CATATAACGTAACGCGTCCTGATGATGATCCTGTTGTGACCCACTCCTT
JL061R	CATATAACGTAACGCGTCCTGATGATGATCCGGCACGACCGGAGGTTTT
JL062F	CCAACTACGAGGATCCGC
JL063R	GTTCTCCTGCGGTACCATC
JL064F	TCTGGAGGATGGTACCGC
JL065R	GATTCGTGTTGAGCCCATAGC
JL066F	GAAACGCTATGGGCTCAACA
JL067R	GGAATGATCTGGTTACCAAAGATCG
JL068F	TCCGATCTTTGGTAACCAGATCA
JL069R	TTTCCACAGCCCCG
JL070F	AGATGACAAACCCGGGG
JL071R	CATGATCATGGGACCGCG
JL072F	TTGTGTCCGCGGTCC
JL073R	ACCCGTCTTTATCGATGAGAGC
JL074F	GAACGCTCTCATCGATAAAGACG
JL075R	ATATTTAATTAAGCTTTTCAACCGGTTCGAC
JL076F	ACCGGTTGAAAAGCTTAATTAATATAAAGGC
JL077R	AGTAACCTAGTCTAGACTCGACTTTACT
JL078F	CCAACTACGAGGATCCGC
JL079R	CACTCGGAAGGTACCTTC
JL080F	TCCCGAAGAAGGTACCTTCG
JL081R	CGCCTTGCTGGAGCC
JL082F	ACAAATATGTGGGCTCCAGC
JL083R	GACGTTTTCAGGTTACCCAATTTCC
JL084F	TAGGAAATTGGGTAACCTGAAAACG
JL085R	TGCCAATGCTCCCGG
JL086F	GAAAATTAAACCCGGGAGCATTG
JL087R	AAAAAGCCCCGGACCTC
JL088F	ACCAGAACGAGGTCCGG
JL089R	AAAAATCGCCATCGATTACTTCGG
JL090F	ATCCGAAGTAATCGATGGCG

JL091R	TGCATTTATTAAGCTTTTCCCTCTACC
JL092F	GTAGAGGGAAAAGCTTAATAAATGCAAGT
JL093R	AGTAACCTAGTCTAGACTCGACTTTACA
JL094F	TGCAGTCAGACTTAGCCTCCAACCTACGAGGATCCGC
JL095R	CGATCGTCGAATCGTAGCAGTAACCTAGTCTAGACTCGACTTTACT
JL096F	TGCAGTCAGACTTAGCCT
JL097R	CGATCGTCGAATCGTAGC
JL098F	TGCAGTCAGACTTAGCCTCCAACCTACGAGGATCCGC
JL099R	CGATCGTCGAATCGTAGCAGTAACCTAGTCTAGACTCGACTTTACA
JL100R	ATGATCGCCTTGCTGGAG
JL101F	GGGCTGGACATCCGCCAAGGTTATGGGCTGACCGAAACAACCTAGTGCCATCTGC GGGAAC
JL102R	GTTCCCGCAGATGGCACTAGTTGTTTCGGTCAGCCCATAACCTTGGCGGATGTC CAGCCC
JL103F	ATTTAACAGTCTTAAACACTGTGCATCCGGAGGGGAGCCTATTAAT
JL104R	ATTAATAGGCTCCCCCTCCGGATGCACAGTGTTAAGACTGTTAAAT
JL105F	TGGAACACGTCCGACACAGGACATGGATTTTCTGCTTTTAGCACGGTATTCTCG CCCTGG
JL106R	CCAGGGCGAGAATACCGTGCTAAAAGCAGAAAATCCATGTCCTGTGTGGACGT GTTCCA
JL107R	CCCTTTGAAGTTCCCGCAGATGGCACTAGTTGTTTCGGTCAGCCCATAACCCTC ATAGA
JL108F	GTTCCGGGCTTAGATTAAGCGTCAATGGCGATCCTATTTTTGGCAATCAAATCAT TCCGGATACTGTTATGTGGAACACGTCCGACACAGGA
JL109R	TCCTGTGTGGACGTGTTCCACATAACAGTATCCGGAATGATTTGATTGCCAAA AATAGGATCGCCATTGACGCTTAATCTAAGCCCGAAC
JL110F	CGGGCTTGGATTAAGCGTCAATGGCGATCCTATTTTTGGCAATCAAATCATTCC GGATACTGTTATGTGGAACACGTCCGACACAG
JL111R	CTGTGTGGACGTGTTCCACATAACAGTATCCGGAATGATTTGATTGCCAAAA TAGGATCGCCATTGACGCTTAATCCAAGCCCG
JL112F	TCTATGAGGGTTATGGGCTGACCGAAACAACCTAGTGCCATCTGCGGGAACCTCA AAGGG

Table 2.2. Exhaustive list of primers ordered for the project.

2.1.1.3 Double-stranded oligonucleotide gene fragments (gBlocks)

Whole and partial genes were ordered as double-stranded DNA from Integrated DNA Technologies as a lyophilised stock and were dissolved in ultrapure water to the manufacturers' recommended concentrations (included with each order) for long term storage and working stocks.

2.1.2 Materials related to purified protein

2.1.2.1 Buffer list for protein purification

Storage buffer (200 mL)

10 x TEM (1 M tris-acetate, 20 mM EDTA, and 100 mM MgSO ₄ At pH 7.8)	20 mL
50% (v/v) glycerol	40 mL
Ultrapure water (pH adjusted to 7.8 after addition of water)	Up to 200 mL

Buffer A (400 mL)

Phosphate buffered saline (PBS) tablets	2 tablets
Glycerol	3.8 g
Ultrapure water (pH adjusted to 8.0 after addition of water)	up to 400 mL

4 M IMD (5mL)

IMD	1.36 g
Buffer A (pH adjusted to 8.0 with H ₃ PO ₄ after addition of water)	5 mL

Lysis buffer

Buffer A	20 mL
4 M IMD	100 µL
25 x EDTA-free protease inhibitor (PI) (1 EDTA-free PI tablet dissolved in 2 mL dH ₂ O)	800 µL
100% Triton X-100 (pH adjusted to 8.0 after addition of water)	200 µL
1 M β-mercaptoethanol (β-ME) solution (β-ME added to chilled solution)	200 µL

20 mM IMD solution (10 mL)

Buffer A	10 mL
4 M IMD	50 µL
(pH adjusted to 8.0 after addition of water)	
1 M β-ME solution (added to chilled solution)	100 µL

Constituents	IMD concentration (mM)			
	50	200	300	500
Buffer A	10 mL	15 mL	15 mL	10 mL
4 M IMD	125 µL	750 µL	1125 µL	1250 µL
β-ME (added to chilled solution)	100 µL	150 µL	150 µL	100 µL

2.1.2.2 Reagents for Coomassie staining of SDS-PAGE gels

Methanol fixer solution (1 L)

400 mL methanol, 100 mL acetic acid, and 500 mL distilled water

Coomassie staining solution (1 L)

2.5 g Coomassie Blue R-250, 400 mL methanol, 100 mL acetic acid, and 500 mL distilled water

Destain solution (1 L)

50 mL methanol, 75 mL acetic acid, and 875 mL distilled water

2.1.2.3 Reagents for silver staining of SDS-PAGE gels

Ethanol fixer solution (1 L)

400 mL ethanol, 100 mL acetic acid, and 500 mL distilled water

0.02% sodium thiosulphate (200 mL)

0.04 g Na₂S₂O₃, 200 mL distilled water

Silver nitrate solution (1 L)

1 g AgNO₃, 1 L distilled water

3% sodium carbonate solution (250 mL)

7.5 g Na₂CO₃, 250 mL distilled water

2.1.3 Materials related to *E. coli* cells

2.1.3.1 *E. coli* strains and plasmids

Glycerol stocks of *E. coli* BL21 containing pET16b vector encoding WT *Ppy* Fluc gene were constructed previously (Law et al. 2006). These were streaked onto agar plates, picked into LB broth, grown, and made into new glycerol stocks.

2.1.3.2 Culture media

Both solid agar and liquid broth was kindly prepared by support staff within the Murray Lab. Stocks of sterile 100 mg/mL carbenicillin in ultrapure water and 34 mg/mL chloramphenicol in ethanol were prepared by dissolving the antibiotic powder in appropriate solvent and storing at $-20\text{ }^{\circ}\text{C}$. LB agar plates with 100 $\mu\text{g}/\text{mL}$ carbenicillin (for all *E. coli* cells) optionally combined with 25 $\mu\text{g}/\text{mL}$ chloramphenicol (only for BL21 cells) were prepared by the addition of antibiotic stock to melted hand-hot LB agar, and the agar was then poured to create 100 mm plates of 12 mL in volume. Super optimal broth with catabolite repression (SOC) medium was obtained from New England Biolabs DH5 α competent cell kits. Stock solutions of 200 mg/mL isopropyl β -D-1-thiogalactopyranoside (IPTG) were added to LB broth or agar to a final concentration of 1 mM (238 $\mu\text{g}/\text{mL}$).

2.1.4 Materials related to mammalian cells

2.1.4.1 Culture media

2.1.4.1.1 Preparation of heat-inactivated foetal calf serum (FCS)

Litre bottles of FCS were thawed overnight at $4\text{ }^{\circ}\text{C}$, and once thawed the bottles were inverted to homogenise the solution and (inside a laminar flow hood using a 60 mL syringe) filtered through a 0.2 – 0.22 μm filter into 50 mL aliquots in 50 mL skirted centrifuge tubes. The aliquots were then placed in a $56\text{ }^{\circ}\text{C}$ water bath for 30 mins to inactivate the complement enzymes. Following this the aliquots were stored at $-80\text{ }^{\circ}\text{C}$ until required for creating complete cell culture media.

2.1.4.1.2 Preparation of complete cell culture media

500 mL bottles of Iscove's Modified Dulbecco's Medium (IMDM) were opened inside a laminar flow hood. A 50 mL aliquot of heat-inactivated FCS (pre-thawed at $4\text{ }^{\circ}\text{C}$ the day prior) was added to each through a 0.2 – 0.22 μm filter using a 60 mL syringe. 5.5 mL of 100 X (10,000 U/mL) of penicillin-streptomycin antibiotic solution was added to each. The final mixture was then homogenised by pipetting up and down with a 50 mL serological pipette; this is preferred as inverting will wet the inside of the lids which is considered poor aseptic technique. The final bottles of complete cell culture media was stored at $4\text{ }^{\circ}\text{C}$ until required; the bottles were preheated in a $37\text{ }^{\circ}\text{C}$ water bath for 30 mins prior to use.

2.1.4.2 Polyethylenimine (PEI) transfection agent

A Cold Spring Harbor protocol (Polyethylenimine (PEI), linear (1 mg/mL). 2008) was followed and is adapted henceforth. 500 mg of linear 25 kDa PEI powder was added to 450 mL of ultrapure water in a large beaker. A magnetic stir bar was placed in and set to stir at a moderate speed on a magnetic stirrer plate (the heat-plate was not turned on). A pH meter probe was placed into the solution to monitor pH in real-time. 12 M hydrochloric acid was added drop-wise to the solution until the pH was <2. The solution was stirred for 3 hrs and the pH monitored throughout this time; typically 12 M hydrochloric acid was added drop-wise as the pH rose above 2 and ~0.8 mL of 12 M hydrochloric acid was required in total. Typically a small amount of PEI remained undissolved. 10 M sodium hydroxide was added drop-wise until the pH was 7.0. It is expected that ~0.5 mL will be required to neutralise the solution. The solution was transferred to a 500 mL measuring cylinder and the volume adjusted to 500 mL total by the addition of ultrapure water. The entire volume was then (inside a laminar flow hood using a 60 mL syringe) filtered through a 0.2 – 0.22 μ M filter into ~9 x 50 mL aliquots in 50 mL skirted centrifuge tubes, and the remaining volume aliquoted into a 10 mL aliquot in a 15 mL centrifuge tube. The aliquots were then all stored at –80 °C. PEI transfection agent is known to increase its transformation efficiency (for unknown reasons) with increased freeze-thaw cycles (increasing up to at least four cycles) (Reed et al. 2006) and thus the 10 mL aliquot was freeze-thawed at –80 °C and RT three times before being stored at –80 °C before use.

2.1.5 Materials related to bioluminescence screening

2.1.5.1 Luciferin stocks

D-luciferin (D-LH₂) potassium salt was obtained from Regis or Promega. 6'-amino-D-luciferin (D-ALH₂) solid was obtained from Marker Gene Technologies Inc. or Sigma Chemical Co. Infra-luciferin (iLH₂) was obtained from Jim Anderson's group at University College London. The stock concentration of D-LH₂ was 10 mg/mL; the luciferin analogues were considered too precious to store as liquid solutions (presumed to increase the rate of degradation) and were prepared from powder as required. All luciferin stocks were stored in light-proof vials at –20 °C until use.

2.1.5.2 Luciferin for *E. coli* colony BLI screening

It is known that the optimal pH of the D-LH₂ solution used for *E. coli* screening is pH 5.0 (Wood and DeLuca 1987). This is likely because D-LH₂ possesses normally (at neutral pH) negatively charged carboxylate and phenolate groups which protonate at this lowered pH, allowing dramatically greater cell membrane permeability. An external pH of 5.0 is not acidic enough to meaningfully damage the cells, further the cells are typically able to maintain pH homeostasis of the cytoplasm and thus the luciferase is not affected.

0.1 M sodium citrate buffer was created by dissolving 12.5 g of sodium citrate dehydrate and 11.3 g of citric acid into 800 mL of TEM buffer. This was then adjusted to pH 5.0 by the addition of citric acid or sodium hydroxide. Appropriate volumes of 0.5 M D-LH₂ sodium citrate solution can be created as-required for colony screening by combining sodium citrate and D-LH₂ stocks.

2.1.5.3 Luciferin for mammalian cell culture and in vivo BLI screening

Reagents were required to be sterile if used for screening mammalian cells and/or live mice. As beetle luciferin is damaged by both high temperatures and light (especially ultraviolet light), autoclaving and ultraviolet germicidal irradiation (UVGI) are not viable. Further, PAMPs (such as LPS) are not removed by these methods and can cause immune responses for in vivo experiments such as septic shock. To reduce the possibility of bacterial growth, samples were diluted appropriately shortly before use in sterile Dulbecco's phosphate-buffered saline (DPBS) inside a laminar flow hood. The prepared solution was then filtered with a 0.2 – 0.22 µM filter to further reduce the possibility of pathogenic contamination.

2.1.5.4 Haemoglobin phantom

Haemoglobin light-filtration phantoms were created by first gently (to reduce air bubbles) dissolving agarose in water using a microwave oven at a w/v ratio of 1:9 (10% final mixture). Once the mixture had cooled to 60 °C it was degassed using a vacuum chamber until bubbles stopped forming. Defibrinated horse blood was similarly degassed at RT. The agarose gel was re-melted if it had solidified. The melted agarose and the horse blood were heated in a 90 °C water bath until the temperature of both was also 90 °C. The agarose was gently mixed in a 1:1 ratio with the horse blood and immediately poured into square 150 x 150 mm petri

dishes to a depth of 10 mm. Surface bubbles were removed by popping with a pipette tip. The phantoms can be stored at $-20\text{ }^{\circ}\text{C}$ until required, they must be thawed at RT so that no frost deposits on the surface which could increase light scatter.

2.2 Methods

2.2.1 Methods related to purified DNA

2.2.1.1 Ordering codon-optimised DNA oligonucleotides (gBlocks)

Constructs were designed using a combination of Microsoft Word and Snapgene Viewer. The protein sequence of the construct was entered into IDT's codon optimisation tool and the tool run. If any undesired restriction sites were present in the generated DNA sequence then the tool was re-run until removed. There is a random pool of optimised codons which can be chosen for each residue, thus multiple optimisations of the same protein sequence will generate a highly diverse DNA sequence. Any required features such as desired restriction sites were edited into the sequence prior to ordering from IDT.

2.2.1.2 DNA concentration

DNA was concentrated by evaporation using an Eppendorf Concentrator Plus centrifugal concentrator at $45\text{ }^{\circ}\text{C}$ and $2,000\text{ x g}$ until no liquid remained. The resulting pellets were resuspended in appropriate volumes of $65\text{ }^{\circ}\text{C}$ ultrapure water.

2.2.1.3 DNA concentration quantification (NanoDrop spectrophotometer)

Purified DNA in ultrapure water could be quantified using a Thermo Fisher Scientific NanoDrop 1000 spectrophotometer. The machine was "blanked" using $1\text{ }\mu\text{L}$ of ultrapure water, then $1\text{ }\mu\text{L}$ of sample loaded and analysed automatically. The machine gives a reading of DNA concentration in units of $\text{ng}/\mu\text{L}$, but this reading was not considered reliable if below $15\text{ ng}/\mu\text{L}$; samples were DNA-concentrated if below this threshold. The machine also gives an absorbance trace reading; DNA absorbs optimally at λ_{max} of 260 nm , salts commonly at 230 nm , and protein commonly at 280 nm . There is no strict threshold of acceptability of $230/260$ and $280/260\text{ nm}$ values, but a quality purified sample will typically have values of no more than 0.5 for both.

2.2.1.4 Agarose gel electrophoresis

2.2.1.4.1 Running samples via agarose gel electrophoresis

Agarose gels were prepared from 1% w/v agarose powder in tris-acetate-EDTA (TAE) buffer. SafeView was added to hand-hot agarose gel to give a concentration of 5 μL / 100 mL, and the buffer to give a concentration of 10 μL / 100 mL. DNA was mixed with NEB Purple 6x loading dye and loaded into each well of gels. Appropriate DNA molecular weight markers were loaded onto to outer lane wells with samples in between to allow evaluation of band sizes. Samples were loaded and run on an agarose gel at 100 volts until the loading dye had travelled two thirds across the gel.

2.2.1.4.2 Post-run staining of agarose gels

For DNA of extremely low concentration, the low signal-to-noise ratio of SafeView was inadequate to visualise it. Diamond Nucleic Acid Dye has far greater sensitivity. A pre-run gel (regardless of pre-staining with SafeView) was immersed in a 1:10,000 solution of Diamond dye and TAE buffer for 30 mins with gentle agitation at RT. The gel was rinsed with TAE before imaging.

2.2.1.4.3 Agarose gel analysis

Images of the electrophoretically separated DNA bands which appeared on the gel were captured using an Integrated Scientific Solutions U: Genius Gel Doc. If DNA was to be used for cloning downstream then a blue light transilluminator was placed inside the gel doc to minimise DNA degradation, otherwise the stock UV transilluminator was used. Band intensity could be compared to the known quantity in the ladder lane for rough approximation.

2.2.1.4.4 Agarose gel extraction and purification of DNA

DNA was extracted and purified from agarose gels using a Qiagen QIAquick gel extraction kit to manufacturer's instructions. A blue light transilluminator was used to visualise the bands during extraction to minimise DNA degradation. Disposable scalpel blades were used to excise each band to minimise contamination. The included columns isolate DNA of sizes 70 – 10,000 bp in length, and exclude virtually all other contaminants. 50 μL of 65 °C water was typically used for the final elution.

2.2.1.5 Cleaning (PCR-cleaning) of DNA

“PCR clean-up” kits are typically used to isolate amplified DNA from PCRs. However they can also be used to purify DNA from restriction digestions and other reactions. The QIAquick PCR purification kit from QIAGEN was used for this project. The included columns isolate DNA of sizes 100 – 10,000 bp in length, and exclude virtually all other contaminants. The manufacturer’s instructions were followed for use of the kit, and 50 µL of 65 °C water was typically used for the final elution.

2.2.1.6 Restriction digestion of DNA

DNA was digested in 25 µL – 150 µL reactions using a variety of type II restriction enzymes, and specific enzyme combinations are detailed in the results chapters. After digestion, DNA was mixed at a ratio of 1:6 with NEB Purple 6x loading dye and heated to 80 °C for 10 mins to disassociate restriction enzymes from DNA; this step is particularly important for BamHI (HF) as this enzyme can otherwise remain bound to DNA during agarose gel electrophoresis (affecting motility), and may also interfere with T4 ligation.

2.2.1.7 T4 DNA ligation

The T4 Ligation Kit from NEB was used. The manufacturer’s instructions were followed. DNA for ligation was quantified via the NanoDrop spectrophotometer to allow ideal fragment ratios to be adhered to. Ligation mixtures were incubated at 4 °C overnight, then at least 10 mins at RT the following day, before heat inactivation at 65 °C for 10 mins. The latter two steps were performed using a PCR machine for precise temperature control.

2.2.1.8 Gibson assembly

The Gibson Assembly Cloning Kit from NEB was used. The manufacturer’s instructions and recommended fragment ratios were followed. A maximum of three fragments at once were assembled using this method within this project (detailed in Chapter 3). A PCR machine was used for the 50 °C incubation.

2.2.1.9 Slicing by overlap extension (SOE) PCR

Each fragment to be ultimately spliced together was individually PCR-amplified, and the specific primers used are detailed within the relevant results chapters. Once fragments were

2.2.1.12 Error-prone polymerase chain reaction mutagenesis of DNA

Error-prone polymerase chain reactions (epPCRs) were carried out with terminal primers containing respective restriction site sequences for later restriction digestion and ligation reactions. *Taq* polymerase was used for the reaction according to manufacturer's instructions, and carried out in an Eppendorf thermal cycler. The reaction volumes were 50 μ L each and were "spiked" with 0.4 mM $MnCl_2$ to increase the error-rate during polymerisation – this is further detailed within Chapter 3.

2.2.2 Methods related to purified protein

2.2.2.1 Desalting of proteins

For large protein purifications, a PD-10 Desalting Kit (General Electric Healthcare) was used according to the manufacturer's instructions to desalt (remove IMD from) purified protein fractions. The desalted protein was eluted using storage buffer and stored as 0.5 mL aliquots at -80 °C.

For small scale protein purifications, a Zeba Desalting Kit (Thermo Fisher Scientific) was used according to the manufacturer's instructions to desalt (remove IMD from) purified protein fractions. The desalted protein was eluted using storage buffer and stored at -80 °C.

2.2.2.2 Concentration of proteins

Amicon Ultra centrifugal filters of ≤ 50 kDa (usually 30 kDa) were used since the constructs created in this project were typically ≥ 60 kDa. These kits were used according to the manufacturer's instructions. The desalted protein was concentrated by eluting using 5% as much storage buffer as initial diluent, and stored at -80 °C. The size of filter used was dependent upon the volume of initial purified protein; typically large scale protein purifications used the 4 mL variant, whereas small scale used the 0.5 mL variant.

2.2.2.3 Running of proteins on SDS-PAGE gels

Crude enzyme lysate and purified fractions were analysed by Coomassie or silver nitrate staining of 10% SDS-PAGE gels. 10 μ L samples were premixed with 10 μ L of 2 x SDS-reducing sample buffer (SB) and heated to 95 °C for 10 mins before loading into wells. 10 – 15 μ L protein molecular weight markers were run in the outer lanes. Gels were immersed in 1 x

TEA-Tricine running buffer (included with the TruPAGE kit) and run at 300 V until the sample buffer pigment visibly ran beyond the end of the gel.

The ladders used within this thesis are: Biotechrabbit Tricolor (Broad Range) and NEB Blue Protein Standard (Broad Range). The former is typically used on gels to be silver stained, whilst the latter for those to be Coomassie stained.

2.2.2.4 Coomassie staining of SDS-PAGE gels

Gels were fixed in methanol fixing solution for ≥ 30 mins, then immersed in Coomassie staining solution for 2 – 4 hrs (until gel was uniform and no longer visible in the staining solution), both at RT with gentle agitation. Gels were destained in destaining solution overnight at RT with gentle agitation, and stored in 7% acetic acid at 4 °C.

2.2.2.5 Silver staining of SDS-PAGE gels

Gels were fixed in ethanol fixing solution for ≥ 1 hr, then washed in distilled water overnight. Gels were sensitised in 0.02% sodium thiosulphate solution for exactly 1 min, then washed three times in distilled water for 20 seconds each. Gels were stained in silver nitrate solution (with 20 μ L of 35% formaldehyde added per 100 mL immediately prior) for 20 mins. Gels were again washed three times in distilled water for 20 seconds each, then transferred to a new staining tray and washed with distilled water once more for 1 min. Gels were developed in 3% sodium carbonate solution (with 50 μ L of 35% formaldehyde added per 100 mL immediately prior), the solution changed immediately if gels turned yellow. When staining was sufficient gels were washed with distilled water for 20 seconds, and staining was terminated with 5% acetic acid for 5 mins. Gels were stored in 1% acetic acid at 4 °C.

2.2.2.6 Imaging and ImageJ analysis of (Coomassie) stained SDS-PAGE gels

Gels were imaged in dry staining trays inside an office document scanner, or were captured using a DSLR camera or high-quality mobile phone camera.

ImageJ analysis could only be performed on Coomassie stained gels (not silver stained gels due to non-linear stain intensity). The image of the gel was opened using ImageJ and the bands of interest selected by outlining in rectangular boxes. The boxed bands were plotted

as a histogram of band intensity, with the isolated bands clearly visible as peaks on the trace. The straight line tool was used to mark the areas of the peaks relative to the surrounding trace, and this was then selected using the wand tool. The areas of the peaks (the relative intensity of the bands) were then exported as arbitrary units for data processing.

2.2.2.7 Bradford protein assay

The Bio-Rad Quick Start Bradford Protein Assay kit was used to quantify total protein content, and was used according to manufacturer's instructions. Once the standard plot BSA solutions and the samples had been prepared in triplicate on a clear 96-well plate, they were immediately quantified using a BMG Labtech CLARIOstar. Absorbance was measured at 595 nm in triplicate and protein concentrations of samples were determined by linear regression of BSA standard plots.

2.2.3 Methods related to *E. coli*

2.2.3.1 Plasmid purification

Plasmids were purified from *E. coli* DH5 α , DG1, or BL21 cells using Qiagen QIAprep Spin Miniprep Kits. These were used to harvest up to 20 μ g of high copy number plasmid DNA from 2.5 mL *E. coli* cultures; Qiagen midiprep kits were used to harvest up to 150 μ g of high copy plasmid DNA from 100 mL *E. coli* cultures, both to manufacturer's protocol. DNA was eluted using 65 °C ultrapure water. DNA concentrations and purities were estimated by applying 1.5 μ L volumes onto a Nanodrop spectrophotometer.

2.2.3.2 Growth and maintenance of *E. coli* strains

E. coli from glycerol stocks were spread onto LB agar plates with appropriate antibiotics and grown at 37 °C overnight. Single colonies were then picked from these plates and grown in 5 mL LB broth with appropriate antibiotics at 37 °C with shaking at ~200 rpm overnight. Aseptic technique was practiced throughout. Strains were stored at –80 °C as 1 mL glycerol stocks (250 μ L water, 250 μ L glycerol, 500 μ L cultured broth with antibiotic(s)). pET and pCCL family plasmid-transformed cells were maintained short term using 100 μ g/mL carbenicillin LB broth and storing at 4 °C.

2.2.3.3 Transformation of *E. coli*

Plasmid DNA (including ligations) was transformed into *E. coli* DH5 α or DG1 cells by heat shock according to respective manufacturer's instructions. Only purified (high concentration) plasmid DNA was transformed into *E. coli* BL21 (DE3) cells by heat shock according to manufacturer's instructions. Transformed cells were spread onto LB agar containing the suitable antibiotic selection and incubated overnight.

2.2.3.4 Selection of sequence-correct colonies

The following day after plating (ideally morning) the agar plates were assessed. For ligations, those which had relatively few colonies compared to the vector-alone ligation plate were assumed to have inefficient ligation, and those with many assumed to be efficient. The inefficient plates likely had a smaller ratio of correct to incorrect construct colonies present and therefore a greater number had to be screened to isolate a correct colony. 5-10 single colonies from each plate were picked (loosely depending on how efficient the ligation was assumed to be) into separate 5 mL volumes of LB media in 15 mL centrifuge tubes and stored at RT until incubation overnight at 37 °C shaking at ~200 rpm. The following morning, 4.5 mL of each of the 5 mL volumes of media were DNA miniprep according the method included in the Qiagen Miniprep Kit and the final elution volumes were 50 mL. The remaining 0.5 mL of media could be stored as a bacterial glycerol stock.

2.2.3.5 Glycerol stock creation

Liquid *E. coli* culture was mixed at a ratio of 1:1 with 50% glycerol (in ultrapure water). 1 mL aliquots were stored in screw-top Cryovials directly in a -80 °C freezer (i.e. no liquid nitrogen or alcohol and dry ice snap-freezing) for future use.

2.2.3.6 Over-expression of enzymes (IPTG induction)

2.2.3.6.1 IPTG induction of colonies

12 μ L of 1 mM IPTG was spread onto a blank agar plate made from 12 mL of solid agar. Hybond nylon membrane was cut to the size of the agar plates and used to lift colonies. Cut membrane was then placed on the IPTG plate with the colonies facing upwards. The colonies were then typically left to induce at RT for 3 hrs. The IPTG solution cannot be added directly

to colonies as it causes cross-contamination of colonies, precluding later isolation. The IPTG membrane could be directly screened via FLI or BLI in the PhotonIMAGER.

2.2.3.6.2 IPTG induction of liquid culture

Constructs in the pET16b vector were over-expressed by controlled growth and induction of IPTG. BL21 (DE3) *E. coli* cells containing these constructs were streaked from glycerol stocks onto LB agar antibiotic plates and grown overnight at 37° C. Single colonies of each were picked with a sterile pipette tip into 5 mL LB broth with 100 mg/mL carbenicillin and 34 mg/mL chloramphenicol and grown overnight at 37 °C, shaking at ~200 rpm. Each culture was then added to 400 mL of LB broth with 100 mg/mL carbenicillin and 34 mg/mL chloramphenicol and grown at 37 °C, shaking at ~200 rpm until the optical density of cultures (at 600 nm) reached 0.6 – 0.7 AU to ensure cells were in the log phase of growth. A Nanodrop spectrophotometer was used to measure the cell density. IPTG was then added to each to a final concentration of 1 mM and cultures were placed at RT for 3 hrs, shaking at ~200 rpm. Each was then rapidly chilled by submerging in ice before centrifugation (prechilled to 4 °C) at 14700 x g for 30 mins. Supernatants were discarded, pellets of each were weighed and stored at –80 °C until purification.

2.2.3.7 Protein purification (Ni-NTA column)

Deca-His-tagged protein purification was carried out on nickel-NTA (NiNTA) resin as detailed here.

2.2.3.7.1 Small-scale purification (for up to 5 mL of 0.7 OD₆₀₀ LB culture)

E. coli pellets of over-expressed constructs were resuspended in 5 mL lysis buffer per gram of pellet and incubated with 10 µL benzonase nuclease for 15 min on ice before centrifugation (pre-chilled to 4 °C) at 40,000 x g for 10 min. For each construct to be purified, a 2 mL microcentrifuge tube was filled with 0.5 mL of Ni-NTA resin. 1 mL volumes of supernatant (remainder stored to –80 °C) were loaded into each tube (pre-equilibrated in 20 mM imidazole (IMD) in buffer A) and inverted several times, centrifuged briefly and the supernatant collected. Then, 1 mL of 50 mM IMD solution in buffer A was applied to elute non-specifically bound protein, the tube inverted several times, centrifuged briefly and the supernatant collected. His-tagged constructs were then eluted in 1 mL sets of 200 mM, 300

mM, and 500 mM IMD in buffer A, all as with 50 mM. Immediately, these eluted fractions were BLI screened and the optimal fractions desalted by applying them to Zeba Desalting columns. 50 μ L of each non-desalted fraction was kept on ice and assayed by the addition of ATP and D-LH_2 and imaged via PhotonIMAGER.

2.2.3.7.2 Large-scale purification (for up to 300 mL of 0.7 OD₆₀₀ LB culture)

E. coli pellets of over-expressed constructs were resuspended in 5 mL lysis buffer per gram of pellet and incubated with 10 μ L benzonase nuclease for 15 min on ice before centrifugation (pre-chilled to 4 °C) at 40,000 x g for 90 min. For each construct to be purified an emptied PD-10 column was filled with 5 mL of Ni-NTA resin. 10 mL volumes of supernatant (remainder stored to -80 °C) were loaded onto each column (pre-equilibrated in 20 mM imidazole (IMD) in buffer A) and the flow-through was reapplied 1-2 times before 50 mM IMD solution in buffer A was applied to elute non-specifically bound protein. His-tagged constructs were then eluted in one 2.5 mL sets of 200 mM, three sets 300 mM IMD in buffer A, and three sets 500 mM IMD in buffer A. Immediately, these eluted fractions were BLI screened and the optimal fractions desalted by applying them to PD-10 columns. 200 μ L of each non-desalted fraction were kept on ice and assayed by the addition of ATP and D-LH_2 and imaged via PhotonIMAGER.

2.2.4 Methods related to mammalian cells (HEK)

2.2.4.1 Trypsinisation of cells

The flask media was aspirated and set aside in a 50 mL centrifuge tube, and enough 0.25% trypsin solution was added to coat the flask area. Cells were checked for dissociation frequently, and upon total dissociation, aspirated into a centrifuge tube. Cells were centrifuged at 300 x g for 5 mins. The supernatant was then discarded and the cells resuspended in pre-warmed culture media.

If the cells had overgrown (beyond 80% confluency) then clumping may occur where the cells are adhering to one another in addition to the flask surface. The highly confluent film also shields the surface of the flask from the trypsin, preventing proper trypsinisation. When this occurs the cells may be mechanically separated from the flask surface by sharply slapping the

flask against the hand. The cells may detach as a film rather than individual cells and thus require further separation once resuspended via vortexing then straining through a 40 µm strainer.

2.2.4.2 Counting of cells

10 µL of resuspended cells were mixed with 10 µL of 0.4% trypan blue solution. 10 µL of this mixture was pipetted into a Luna™ (Logos Biosystems) cell counting slide and inserted onto a Luna™ automated bright-field cell counter. A preview bright-field image of the cells appears on the display, the focus knob was adjusted so that the cells were sharp and had high contrast. The Luna™ automated counting was performed by simply pressing “bright-field counting” followed by “count” on the touchscreen. The device then returns the number of viable, dead, and total cells per mL which can be used to calculate the total number of viable cells harvested.

If no Luna™ cell counting slides were available, the cells were counted and assessed using a traditional glass haemocytometer and bright-field microscope. The cells were mixed with 0.4% trypan blue solution as above and 10 µL of the mixture pipetted into one of the counting chambers of the haemocytometer (with the cover slip placed on top prior). The cells within one 1mm x 1mm section of the grid of the haemocytometer were then observed with a bright-field microscope and manually counted with a tally counter. The volume of this section is 100 nL and thus the number of live, dead, and/or total cells harvested could be calculated.

2.2.4.3 Subculturing (passaging) of mammalian cells

Cells were observed with an optical microscope and checked for health by comparing their morphology to images of ideal cells, and their confluency was also estimated. Cells were not allowed to grow beyond 80% confluency. The doubling time of HEK cells is 34 hrs, and thus cells must be subcultured when they reached 60% or else they would grow beyond 80% confluency before the following morning.

Following trypsinisation and assessment, a proportion (typically ~10%) were mixed with an appropriate volume of pre-warmed culture media and reflasked, then incubated at 37 °C, 5% CO₂, and 100% humidity.

2.2.4.4 Flow cytometry of cells

2.2.4.4.1 Preparation of cells for flow cytometer

Cells were lifted, pelleted, and resuspended in 10 mL RT DPBS and counted, then pelleted and resuspended in cold (4 °C) DPBS at a concentration of 1×10^7 cells/mL. Cells were kept on ice until run through the flow cytometer.

2.2.4.4.2 General procedure for setting up flow cytometers

Non-transformed (NT) cells were run through at low flow speed. During this, the voltages of the forward-scattered light (FSC) and side-scattered light (SSC) photomultiplier tubes (PMT) were adjusted incrementally so that the NT cells were in the central order of magnitude for both. A scatter plot was generated with FSC-A on the x-axis and SSC-A on the y-axis, then the cells in the centre of the plot were gated (and those in the lower left excluded); this isolates live cells. The gated cells were plotted onto a second scatter plot with FSC-A on the x-axis and FSC-H on the y-axis, and the large population of cells gated; this isolates single cells (Figure 2.1). The live single cells were then plotted as histograms with the lasers of choice (for instance fluorescein isothiocyanate (FITC)) on the x-axis. The various laser PMTs were adjusted incrementally until the NT cells were registered as a very weak positive signal on the histogram preview (lowest log). The PMTs typically have a dynamic range of 4 orders of magnitude, so it is unlikely that positive cells will be off the scale. Experiments performed on flow cytometers revolved around minimising the time spent using them, as their availability is limited and use is charged by the hour.

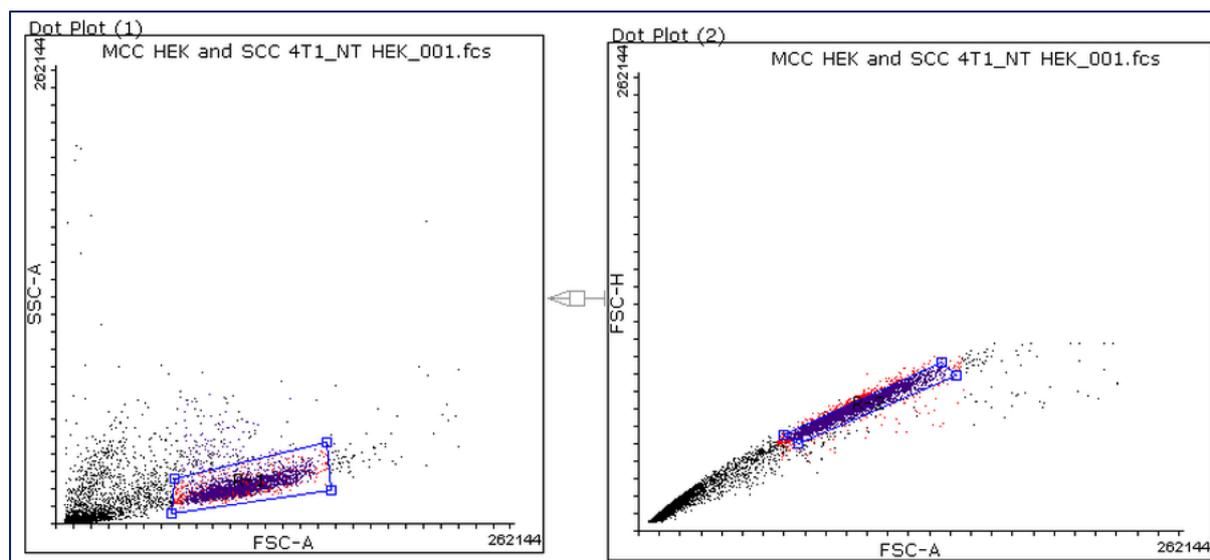


Figure 2.1. Example of FSC-A vs SSC-A scatter plot (left), gated into FSC-A vs FSC-H scatter plot (right). The title of the sample is not of relevance, rather it is a representative sample.

2.2.4.4.3 Running cells through LSR Fortessa for analysis

Transformed cells were run through at high flow speed and data acquired for each sample until a defined number of counts had passed through. The LSR Fortessa is not an aseptic flow cytometer, and as such the remaining sample volumes cannot be subcultured. The acquired data were exported as .FCS files for analysis in FlowJo software.

2.2.4.4.4 Running cells through BD FACSAria for flow-assisted cell sorting (FACS)

The available FACS technician was used for setting up the FACSAria for sorting into either 5 mL tubes or microplates. The storage format for the cells was water-cooled at 4 °C during sorting. The technician tuned the voltage of the sorting plates for the desired format. Depending on what purity of positively transformed cells were desired, the sorting gate (for instance FITC or mCherry) was set such that virtually no NT cells were included in the positively sorted cells. For transformations with very high transformation efficiencies a stricter gate could be used, whereas applying such a gate to a low efficiency transformation may lead to an insufficient number of positively sorted cells, and/or the sort itself taking an extremely long time. The sorted cell solutions were mixed with 4 °C culture media and reflasked appropriately.

2.2.4.5 Cryopreservation of cell stocks

Cells were inspected, lifted, counted, centrifuged, and the media discarded. The remaining cell pellet was then resuspended in 1 mL of Bambanker™ cryopreservation reagent (Nippon Genetics) per 1×10^7 of cells. The resuspension was then pipetted into 1 mL Nunc® CryoTube® vials as 0.5 mL aliquots. These aliquots were then directly stored in a -80 °C freezer (i.e. no liquid nitrogen or alcohol and dry ice snap-freezing) for future use.

2.2.4.6 Transfection of cells

Transfection was performed either with PEI or Genejuice®. The Genejuice® protocol was followed for both agents, and is adapted henceforth; PEI was used in volumes 1.5 x that of Genejuice®. The commonest format which cells were transfected was as 24-well plates, and the following method details how to transfect this format, however table 2.3 (below) can be used to adapt it for different formats.

	Tissue Culture Format								
	Plate format (wells/plate)					Dish format (mm)			
	96	48	24	12	6	35	60	100	150
Number of adherent cells (1×10^5)	0.2	0.4	0.8	1	3	3	10	15	34
Volume of complete growth medium in well or dish (mL)	0.13	0.25	0.5	1	3	3	5	10	23
Volume of serum-free medium in transfection mixture (μL)	6.3	12.5	25	50	100	100	250	500	1125
Volume of GeneJuice® transfection reagent (μL)	0.36	0.75	1.5	3	6	6	12	30	68
Volume of PEI transfection reagent (μL)	0.54	1.13	2.25	4.5	9	9	18	45	101
Amount of plasmid DNA (μg)	0.12	0.25	0.5	1	2	2	6	10	23

Table 2.3 transposed and expanded from the manufacturer's (Sigma-Aldrich) transfection protocol. Quantities of reagents required for transient transduction of adherent cells.

The following was all performed aseptically in a tissue culture hood. 8×10^4 HEK cells were plated in complete growth medium per well of a 24-well plate and incubated at 37 °C (humidified, 5% CO_2) overnight. Ideally cells should be 50 – 80% confluent at the point of transfection. For each well to be transfected, 25 μ L of serum-free medium was added into a sterile tube. 1.5 μ L of GeneJuice® or 2.25 μ L of PEI per well was added to the tube and

vortexed for several seconds to homogenise thoroughly. The mixture was incubated at RT for 5 mins. 0.5 µg of plasmid DNA per well was added to the tube, and the mixture gently homogenised by gentle pipetting up and down then incubated at RT for 15 mins. The whole volume of the mixture (divided by the number of wells it was intended for) was added drop-wise to cells in complete growth medium. The drops were distributed evenly over the entire surface of the dish to ensure that the cells were transfected as uniformly as possible. The cells were incubated at 37 °C (100% humidity, 5% CO₂) for at least 24 hrs.

2.2.4.7 Viral transduction of cells

There are many methods to transduce cells, and one of the most commonly used is a lentiviral transduction plasmid set. These have been developed over many years and several generations of the systems are available. The third generation system requires two packaging plasmids, namely “REV” (pRSV-Rev) and “RRE” (pMDLg_pRRE), and an envelope plasmid “ENV” (pMD2.G) along with a transfer plasmid which is the same pCCL 305 plasmid used for transient transfections discussed elsewhere in this project. The older second generation system requires a packaging plasmid “Pax2” (psPAX2), an envelope plasmid “VSV-G” (VSV-G) along with the same transfer plasmid. The utilisation of both systems is illustrated in Figure 2.2.

The second generation system is simpler and considered more reliable, and may be more efficient at both viral particle production and transduction, based on Dr Amit Jathoul’s experience working with both systems. However, in the interests of safety, and to be in-line with other researchers’ practices within the same lab (such as Dr Amit Jathoul), the third generation system was preferred, with the second generation system being available as a fall-back option.

1 x 10⁷ HEK cells were seeded per flask into three 150 mm TC-coated petri dishes per construct for the production of viral particles (table 2.3). Seeding at this density allows 48 hrs until confluency without transfection. 24 hrs later, each plate of cells was transfected with third generation lentiviral plasmids (9 µg “ENV”, 12.5 µg “RRE”, and 6.25 µg “Rev”) and the construct plasmid (30 µg pCCL 305), and at this point the cells were ~70% confluent which is optimal for transfection. The supernatants from the viral production flasks were harvested

and replaced with fresh complete cell media at 48 hrs, 72 hrs, and 96 hrs post-transfection. The supernatants were stored at 4 °C until the final harvest, at which point they were spin-concentrated by centrifugation in either a Beckman Coulter Allegra 64R or Sigma 3K18 centrifuge. For each construct, 50 mL of viral supernatant was centrifuged at 100,000 x g for 2 hrs and the newly-created supernatant carefully poured off (leaving ~0.5 mL) and discarded. A pellet was typically observed at the base of each tube, which is composed mostly of cells but the viral particles will be bound-up within it. A further <50 mL was loaded into the same centrifuge tube and the process repeated until the entire pooled volume was concentrated; it was then stored at -80 °C to kill the HEK cells and preserve viral particles. Filtration through a 0.22µm nylon membrane is another option, however due to the small volumes it was feared that the filter would retain a meaningful volume of viral particles. Lentiviral particles are susceptible to damage from freeze-thaw cycles (it is generally understood that each freeze-thaw cycle reduces the titre 2-3 fold), however the HEK cells and any possible contaminants (such as bacteria) are reliably killed by either method. The thawed viral supernatant volumes were first topped up to that of the largest volume (~5 mL) before being combined with 2 µL of Polybrene (Hexadimethrine bromide) per 1 mL of supernatant. Two T75 flasks of HEK cells per construct – which had been plated 24 hrs prior with 1×10^7 cells per flask – had the cell culture media aspirated off, then the transduction mixture was added evenly to these cells. A third T75 flask was left unaltered as the non-transduced negative control.

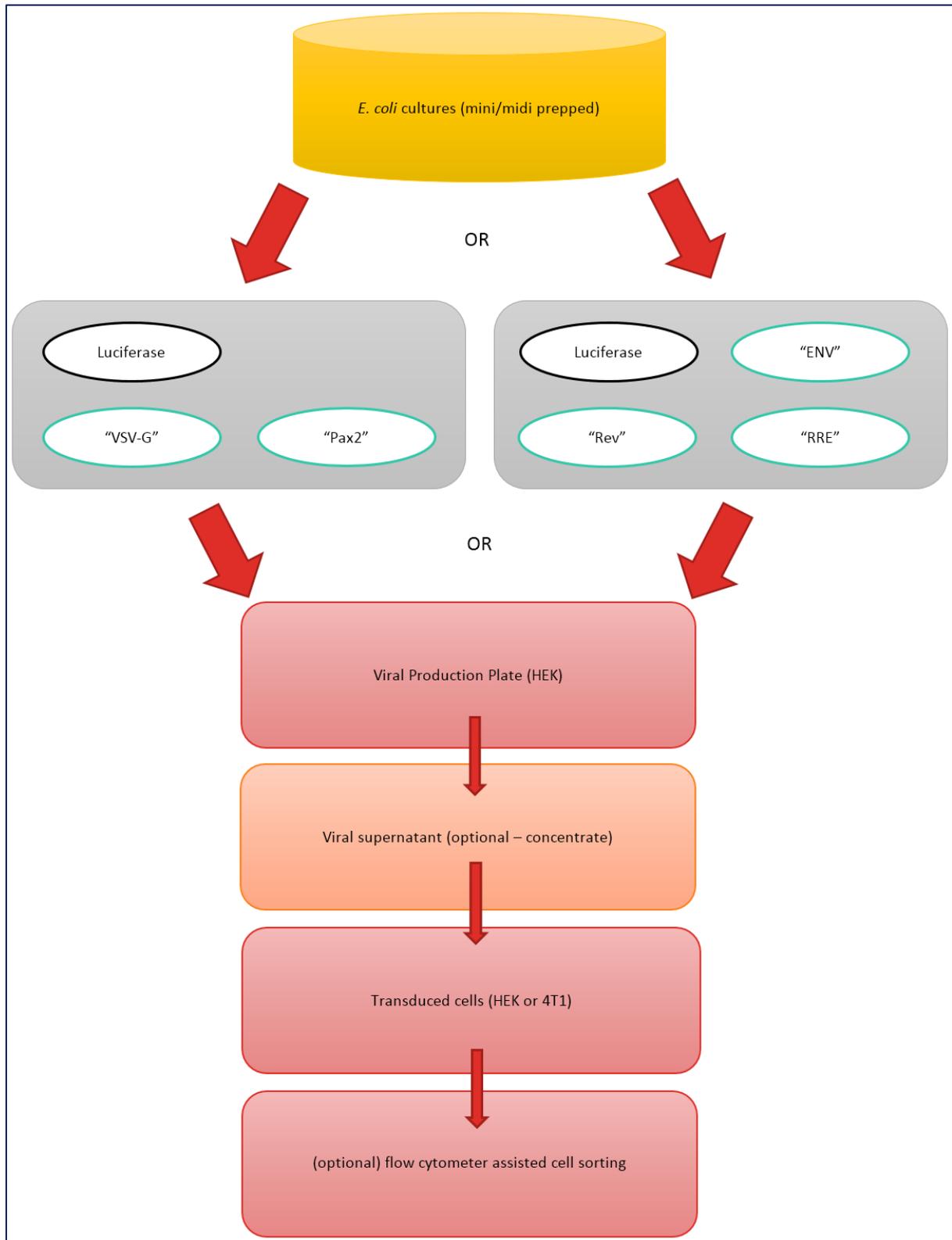


Figure 2.2. Flow chart of processes required for viral transduction of adherent cells. 4T1 cells were not ultimately used for this project.

2.2.5 Methods related to in vivo work

All animal work undertaken during this project and detailed within this thesis was carried out in accordance with The Animals (Scientific Procedures) Act 1986 (ASPA), and the European Directive 2010/63/EU. Further, the NC3Rs' ARRIVE guidelines were followed where possible. Both James A. Long and Dr Amit Jathoul were Home Office licenced prior to any animal work being performed.

2.2.5.1 Preparation of NSG mice for implantation

Prior to implantation, NSG mice were placed in a negative pressure isolator on arrival in the animal house for quarantine. After 24 hrs the mice were transferred to a positive pressure isolator which allows for aseptic procedures to be performed on the animals inside as the front of the isolator has sealed gloves fitted.

Along with my mentor (Dr Amit Jathoul, also Home Office licensed) I used the positive isolator gloves to label the tails of the mice with rings using marker pen, which is a more ethical alternative to ear clipping. The number and colour of rings is unique to each mouse.

The mice were then acclimatised in the positive pressure isolator for 7 days, to ensure they are at maximal health and minimal stress upon implantation and to reduce ensuing shock. Further, stressed animals often over-groom themselves as a coping mechanism and this can lead to them focussing their efforts on the sensitive implantation site, leading to damage.

2.2.5.2 Implantation of NSG mice with HEK cells in Matrigel

A Matrigel vial was thawed at 4 °C overnight. The high concentration variant of Matrigel is 18-22 mg/mL and can be diluted to >4 mg/mL (i.e. as little as >22% Matrigel stock) without fully inhibiting gel formation. Each mouse was implanted with 100 µL of Matrigel and cell mixture in a 1:1 ratio, i.e. 50 µL of Matrigel and 50 µL of cells. Prior to mixing with Matrigel, the cells were resuspended in complete media to appropriate concentrations. The Matrigel was combined with cells according to manufacturer's instructions.

The syringe needle gauge is advised by Corning to be between 21 G and 25 G as this is narrow enough to minimise trauma to the mouse, whilst being wide enough to minimise cell shearing

and potential Matrigel clumping and blocking of the needle. Given the animal ethics-driven nature of this PhD project, the lower gauge of 21 G was chosen.

Once the Matrigel™ syringes were prepared they were stored on ice and transferred into the positive isolator through the airlock according to the aseptic procedure. The mice were then each scruffed and held taught with the non-dominant hand, using the other hand they were implanted. For each mouse the needle tip was inserted into the subcutaneous area of the right rear flank, and the needle was then carefully swayed to open up a pocket during injection as recommended by Corning to maximise the contact area of the injected cells with the mouse's own tissue and promote growth. The full 100 µL volume of the syringe was injected slowly (to minimise cell shearing) but gradually (to avoid the solidifying Matrigel blocking the needle). The mice were then returned to their cages.

2.2.5.3 Imaging of mice using the PhotonIMAGER

Mice were imaged in sets of up to five, as it is the maximum capacity of the PhotonIMAGER.

The heat plate of the PhotonIMAGER was preheated to 37 °C. The mice were anaesthetised initially with 5% isoflurane in 100% oxygen at 2 L/min inside a sealed chamber. Less than one min after *all* the mice were unconscious the gas mixture was reduced to 2% isoflurane in 100% oxygen at 2 L/min and rerouted to the nosecone manifold. Mice were immediately transferred to the PhotonIMAGER nosecone manifold, and this allows up to 5 mice to be kept anaesthetised whilst they are being imaged. A black plastic spacing jig was placed underneath the mice to keep them separated and prevent light bleed between mice. The built-in heat plate was *maintained* at 37 °C to prevent hypothermia which is the commonest cause of death for animals during anaesthesia.

One at a time the mice were quickly (to prevent the regaining of consciousness) removed from the nosecones, injected intraperitoneally at the midline – with 200 µL of 20 mg/mL D-LH₂ (roughly 150 mg/kg assuming a male weight of 25 g) – and immediately placed back into their nosecones. Once all the mice had been injected with D-LH₂, the acquisition was performed. The time between the first and last injection is typically less than one minute,

which is acceptable as peak light output typically occurs 8 mins post-injection (Khalil et al. 2013).

2.2.6 Methods related to bioluminescence measurement

2.2.6.1 PhotonIMAGER

Samples were placed on the baseplate and raised to an appropriate distance from the lens. A bright field image was captured followed by a non-filtered (PMT) acquisition whose duration typically varied inversely to the brightness of samples. For experiments where kinetics were of interest a defined longitudinal screen was acquired. Spectra were obtained by successive PMT acquisitions through the available bandpass filters with 25 nm divisions, a second non-filtered acquisition was captured immediately after the filtered acquisitions to allow for light output reduction (over the course of the screen) compensation. Data were analysed using Biospace Lab's M3 Vision software and exported as excel spreadsheets for later analysis.

2.2.6.2 BMG Labtech CLARIOstar for precise spectra

The PhotonIMAGER uses a filter wheel which is capable of producing spectra with resolutions of 25 nm, whereas the CLARIOstar has a resolution up to 1 nm as it possesses a monochromator and thus it was used for bioluminescence screening where a high resolution spectra was desired. However, the CLARIOstar is less sensitive than the PhotonIMAGER and can only accept microtitre plates. The sensor is presumed to be the size of a well of a 384-well microtitre plate, and it views the centre of the well regardless of well size. Thus, when imaging adherent cells, the cells were scraped into the centre of the well to maximise the light received by the sensor. The data were exported from the included MARS software as excel spreadsheets for later analysis.

2.2.6.3 Purified protein screening

The protocol from a paper (Mofford et al. 2017) was adapted for screening purified proteins. Following enzyme relative concentration being ascertained via Coomassie stained SDS-PAGE, the enzymes were diluted with storage buffer to that of the lowest sample. For each construct bioluminescence was initiated by adding 30 µL of purified enzyme in storage buffer to 30 µL

of 200 μM D-LH₂ in substrate buffer [20 mM Tris (pH 7.8), 0.1 mM EDTA, 8 mM MgSO₄, and 4 mM ATP] in a black 96-well plate. For non-desalted non-quantified protein fractions, 30 μL of eluted fraction was added to 30 μL of 200 μM D-LH₂ in substrate buffer. Imaging was performed 1 min after enzyme addition using either the PhotonIMAGER or CLARIOstar.

2.2.6.4 *E. coli* BLI screening

2.2.6.4.1 *E. coli* colony BLI screening

This method for *E. coli* colony screening is adapted from a paper (Wood and DeLuca 1987). Expression of plasmid-encoded proteins was induced with IPTG as stated earlier within this chapter. The membrane was lifted and placed colony-side up on petri-dish lids and bioluminescence was initiated by lightly spraying colonies several times with 0.5 mM D-LH₂ in citrate buffer, pH 5. A DSLR in a darkroom was used to acquire a bright-field image of the colonies before spraying, and a BLI of the colonies immediately after spraying for a 30 sec exposure (aperture and ISO adjusted automatically for *Ppy* Fluc and used for other constructs). The colonies were then immediately transferred to the PhotonIMAGER and a more sensitive BLI acquired 1 min after the initial spray.

Recombinant libraries could be secondarily screened by picking colonies of interest with sterile pipette tips and spotting them into triplicate rows on further LB agar antibiotic plates. *E. coli* transformed with WT *Ppy* Fluc were also spotted onto the same plate to allow a relative light output comparison. Thus, secondary screening allowed a light colour and output evaluation to be made as well as confirming that the colony of interest had been isolated before culturing in LB broth for storage and further experiments.

2.2.6.4.2 *E. coli* crude lysate BLI screening

50 μL of transformed and induced cultures were mixed with 10 μL of 20 mM EDTA in 1 M Na₂HPO₄ (pH 7.8). Aliquots of the mixture were frozen at $-80\text{ }^{\circ}\text{C}$, thawed, then mixed with 300 μL freshly prepared lysis mix (1.25 mg/mL lysozyme, 2.5 mg/mL bovine serum albumin (BSA) in PBS) and incubated at RT for 10 mins. 10 μL aliquots of the cell lysate were mixed with 90 μL of 25 mM Tris-acetate buffer (pH 7.8) containing, 2 mM ATP, 4 mM magnesium acetate and 0.4 mM D-LH₂. Light output and spectra were acquired using the PhotonIMAGER.

2.2.6.5 Mammalian cell culture BLI and FLI screening

2.2.6.5.1 Mammalian cell culture FLI screening

D-LH₂ is fluorescent, and is very similar in absorption and emission wavelengths to that of EGFP, one of the fluorescent proteins used for mammalian transformation in this project. Thus, fluorescent imaging of mammalian cells must be performed prior to the addition of D-LH₂ for BLI screening. The culture media contains phenol red as a pH indicator, which attenuates light and is also fluorescent, and thus it was replaced with DPBS for imaging purposes.

Cells were removed from the incubator and the culture media aspirated, equivalent volumes of DPBS were added, aspirated, and DPBS added again. The cell plates/flasks were then placed in the PhotonIMAGER and a FLI acquisition pre-programmed for the chosen fluorescent protein performed. If asepsis was not required then the plate lid was removed for imaging, if there was substantial light bleed from bright wells into dim wells then bright wells could be covered with opaque material to occlude excitation.

2.2.6.5.2 Mammalian cell culture BLI screening

Cells were removed from the incubator and the culture media aspirated, then equivalent volumes of DPBS were added, aspirated, and DPBS added again. Stock D-LH₂ was added – to a final concentration of 0.1 mM – to wells to induce bioluminescence and the plates agitated (without swirling) to mix the solution. The plates were then placed in the PhotonIMAGER and BLI acquisition begun 1 min after D-LH₂ addition. If asepsis was not required then this was performed in the PhotonIMAGER room, if asepsis was required then this was done in a laminar flow hood and BLI acquisition started immediately after – ensuring the same lag period for all constructs.

2.2.7 Methods related to *Ari* GWluc expression, lysate preparation, and BLI screening

In Chapter 3 (discussed in detail there) the method for expressing *Ari* GWluc according to that outlined in the supplementary document of its paper (Trowell et al. 2016) was followed but failed to produce protein (checked via SDS-PAGE) or bioluminescence. Thus, the paper's suggestions for expression were ignored in favour of a different approach utilising a vector known as pCold TF (Chapter 3). Both the paper's method and the pCold method for

expression are detailed here. The paper's method for lysate preparation and bioluminescence screening were followed with success and are also detailed.

2.2.7.1 Ari GWluc expression, lysate preparation, and BLI screening according to Trowell *et al.* supplementary

2.2.7.1.1 Ari GWluc expression according to Trowell *et al.* supplementary

In the *Ari* GWluc supplementary paper Trowell *et al.*: “2 mL cultures of transformed BL21 (DE3) cells were incubated overnight in LB media containing 1 M glucose at 37 °C and 200 rpm. The following day 50 mL of LB (100 µg/mL ampicillin) was inoculated with a 1:50 dilution of each culture and incubated at 37 °C until an Abs600 >0.6 was achieved. A 10 mL aliquot was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.4 mM final concentration) and a second aliquot was not induced (added equivalent amount of water instead). Cultures were incubated at 20 °C and 150 rpm for 48 hours at which time the Abs600 was again measured. 5 mL of pETDuet-1: control vector, pETDuet-1:FFLuc and pETDuet-1:GWluc cultures were transferred into clean centrifuge tubes and centrifuged at 4 °C for 10 minutes at 3,000 rpm. Pellets were resuspended in ice cold phosphate buffer saline (PBS, pH 7.4, 137 mM NaCl, 2.7 mM KCl, and 4.3 mM Na₂HPO₄) and centrifuged again at 4 °C for 10 minutes at 3,000 rpm. The supernatant was removed and the pellets resuspended in ice cold PBS.”

2.2.7.1.2 Ari GWluc lysate preparation according to Trowell *et al.* supplementary

In the *Ari* GWluc supplementary paper Trowell *et al.*: “40 µL of BL21(DE3) non-transformed cells were mixed with 50 µL of the transformed cultures, and 10 µL of 20 mM EDTA in 1 M Na₂HPO₄ (pH 7.8), was added. Aliquots of the cell mix were snap frozen on dry ice and stored at –80 °C. Cell aliquots were mixed with 300 µL freshly prepared lysis mix (1x luciferase cell culture lysis reagent (CCLR, Promega), 1.25 mg/mL lysozyme (Sigma), 2.5 mg/mL bovine serum albumin (BSA) (Sigma) and incubated at room temperature for 10 minutes.”

2.2.7.1.3 Ari GWluc BLI screening according to Trowell *et al.* supplementary

In the *Ari* GWluc supplementary paper Trowell *et al.*: “Synthetic D-luciferin (Sigma) was assayed using the cell lysates (above). 10 µL aliquots of the cell lysate were mixed with 90 µL

of 25 mM Tris-acetate buffer (pH 7.8) containing, 2 mM ATP, 4 mM magnesium acetate and 0.4 mM D-luciferin (Sigma). Total light output was measured.”

2.2.7.2 Successful *Ari* GWluc expression

BL21 cells transformed with *Ari* GWluc in pCold (Chapter 3) were used to express the enzyme in 300 mL liquid LB as in Trowell *et al.* except that the cells were grown at 150 rpm and expressed using 0.4 mM IPTG at 15 °C and 150 rpm for 18 hrs. [These conditions were found by expression of *Ari* GWluc colonies across a matrix of conditions of 0.1, 0.4, and 1 mM IPTG; 16, 18, and 22 hrs; and 15 °C; then probing with 0.5 mM D-LH₂ at pH 5.0 (Chapter 3)]. *Ppy* Fluc was expressed under the same conditions in parallel.

Cell lysate preparation and bioluminescence screening were performed as in Trowell *et al.*, except that a PhotonIMAGER was used for acquisition.

3 Chapter 3 – Creation and In Vitro Testing of Random and Rational Mutants of Murine Acyl-CoA Synthetase 3

3.1 Chapter Summary

In this work it was hypothesised that a murine ACS enzyme could be converted to a functional beetle luciferase with a relatively small number of rational mutations. This began with selecting a functional luciferase to act as a template from which the ACS could inherit bioluminescence function. *Photinus pyralis* (*Ppy*) Fluc was chosen as the most-researched Fluc, after experimentally discarding another potential blue dipteran luciferase from *Arachnocampa richardsae* (*Ari*). ACSM3 was selected as the ACS owing to its relatively high homology to *Ppy* Fluc in areas of the active site and other regions of potential importance. However, random mutagenesis of ACSM3 failed to create any bioluminescent *E. coli* colonies. Rationally designed mutants of ACSM3 were then created and screened in *E. coli* colonies including an ACSM3 mutant with the active site of *Ppy* Fluc substituted, an ACSM3 mutant with the R223 – T235 region of *Ppy* Fluc substituted, and an ACSM3 mutant with substrate-determining residue substitutions from *Ppy* Fluc. However, all combinations of the above were non-bioluminescent. Protein modelling revealed that WT ACSM3 appears to have two residues (equivalent to K443 and Q448 in *Ppy* Fluc) near the active site that could create a steric clash with luciferyl adenylate. The novel constructs created in this chapter are shown below in table 3.1 and explained throughout.

ACSM3 Backbone Constructs
ACSM3 epPCR libraries
ACSM3 with <i>Ppy</i> Fluc active site
ACSM3 with <i>Ppy</i> Fluc active site + <i>Ppy</i> Fluc SDRs
ACSM3 with <i>Ppy</i> Fluc R223 loop + <i>Ppy</i> Fluc active site
ACSM3 with <i>Ppy</i> Fluc active site + <i>Ppy</i> Fluc SDRs + <i>Ppy</i> Fluc R223 loop

Table 3.1. Exhaustive list of constructs created and detailed within Chapter 3. DNA sequencing was not performed on the epPCR libraries and as such their exact identities are not known.

3.2 Overall Introduction

As discussed in Chapter 1, Flucs are immunogenic (Podetz-Pedersen et al. 2014) with the specific epitopes identified in the N-terminal domain (NTD) of *Ppy* Fluc (Limberis et al. 2009). It is not yet possible to design proteins as complicated as enzymes purely computationally, thus advances are currently made incrementally and often blindly, and rely heavily on experimentally screening candidates in the lab. One possible solution to the immunogenicity would be to perform site-directed mutagenesis (SDM) of the epitopic regions. However – given the further aforementioned drawbacks with the Fluc system (Chapter 1) – a preferable solution could be to convert a native (and thus inherently biocompatible and non-immunogenic) murine protein into a functional luciferase. Insect ACSs (such as from *Zophobus morio*, *Agrypnus binodulus*, and *Drosophila melanogaster*) have previously been converted into relatively bright beetle luciferases by single residue mutations (Viviani et al. 2013); (Oba et al. 2009) and the introduction of oxygenase function was found to be of crucial importance (Prado et al. 2016). As with all ANL superfamily enzymes, this second reaction step is made possible through a 140° rotation of the C-terminal domain (CTD) and corresponding conformation change of the enzyme (Sundlov et al. 2012); the active site is present at the interface between these domains.

Other prior work provides further precedent for the effects of single amino acid changes on luciferase function, including increase of pH tolerance (Law et al. 2006), increase in brightness through single residue deletions (Halliwell et al. 2018), and thermostable and colour-shifted mutants produced through site directed mutagenesis (SDM) (Branchini et al. 2007). Thus, single residue substitution is one potential strategy for the conversion of murine ACSs into luciferases. However, given that murine ACSs are more phylogenetically distant from their insect counterparts, endowing bioluminescence function may require unpredicted mutations and/or substitution of a greater proportion of Fluc. A number of strategies were considered to explore these possibilities:

3.2.1 Random mutagenesis

Random mutagenesis is a method of protein engineering which simulates the evolution of proteins in nature, i.e. by iteration of random mutation and controlled selection pressure cycles. The optimal methods for luciferase engineering use a small number of mutations introduced to a given protein per round of mutagenesis. Given that single residue mutations may be sufficient to endow bioluminescence function to non (or weakly) bioluminescent ACSs (Viviani et al. 2013);(Oba et al. 2009), this method was considered viable since even low levels of emitted light are readily detected. Further, an unlimited number of rounds of mutagenesis is possible, allowing the accumulation of beneficial mutations which has proven successful in the creation of many Fluc mutants – such as the detergent-stable Ultra-Glo™ Luciferase (Auld et al. 2009).

The greatest drawback of this method is that the vast majority of the mutants created will be deleterious, especially when large numbers of mutations are made per round. This is because each unique protein sequence exists conceptually in a sequence space for which only a minute number will serve a specific useful function. For a dipeptide one can imagine a grid with 20 amino acids along the x and y axes, and 400 unique dipeptide grid squares (20 x 20). For tripeptides a third (z) axis is added and the number of unique tripeptides rises to 8,000 (20 x 20 x 20). The number of iterations increases exponentially with length such that there are 3.69×10^{715} unique 550-residue proteins – the length of *Ppy* Fluc. For comparison, there are a mere 10^{80} hydrogen atoms in the observable universe (Eddington number).

identified the following residues, the equivalent positions in *Ppy* Fluc are: R218, F247, T251, L286, G339, S347, and A348. The S347 residue is of particular importance as, in the Oba *et al.* paper, it alone was capable of endowing bioluminescence function to the two insect ACSs tested. Note that of the latter seven residues only one (T251) was not predicted to be a SDR by the other (Khurana *et al.* 2010) paper. Thus, it is conceivable that a murine ACS could be converted to a functional luciferase by substitution of these SDRs.

3.2.4 Substitution of the R223 loop

The loop between the residues 223 – 235 in *Ppy* Fluc (and other beetle luciferases at this equivalent position) is of known importance (Viviani *et al.* 2008) and is referred to as the R223 loop henceforth. For Flucs this is of importance for retaining a hydrophobic environment inside the active site and retaining bright green emission and pH-insensitivity, and mutations to it often result in the deterioration of these properties. Although the sequence identity of this loop is not crucial for retaining a degree of bioluminescence function in Flucs, it is believed to stabilise the active site and serve as an anchorage site at the bottom of the luciferin binding site.

3.3 Results and Discussion

3.3.1 Selection of the luciferase parent enzyme (comparison of *Ari* GWluc and *Ppy* Fluc)

Owing to a long history of coleopteran *Photinus pyralis* (*Ppy*) firefly (Fluc) engineering in The Murray Lab and an abundance of literature information, *Ppy* Fluc was initially selected as one potential bioluminescent template or “parent” enzyme. However, during the initiation stage of the project, a paper (Trowell *et al.* 2016) was made available (as an accepted manuscript) which detailed the characterisation of a luciferase from a fungus gnat (*Arachnocampa richardsae* – *Ari*).

This dipteran enzyme is potentially more homologous to murine ACSs (Figure 3.2) and is responsible for the blue bioluminescence generated by *Arachnocampa richardsae* glowworm luciferase (*Ari* GWluc) in certain Australian caves. Trowell *et al.* claimed that *Ari* GWluc uses D-LH₂ as a substrate and emits blue light – as the organism does in nature (Figure 3.3). This claim was unprecedented as it was previously claimed by another paper (Silva *et al.* 2015)

that the enzyme was not an ACS and did not use D-LH₂. When Flucs are randomly mutated by substitutional mutagenesis most bioluminescent mutants are dimmer and often display a characteristic bathochromic (red) shift (Shapiro et al. 2005); (Tisi *et al.* 2002). The active site of Flucs is hydrophobic and this is easily disrupted by changes to both nearby or more distant residues, causing solvent exposure of the excited state oxyluciferin and leading to non-radiative losses and redshifts (Branchini et al. 1999). In this way, the colour of a given Fluc mutant may indicate the capacity for deleterious mutations before bioluminescence function is totally lost. Thus, it was speculated that the active site of *Ari* GWluc could be more hydrophobic and protective from solvent, or more stable to higher energy colour forms. This could allow greater tolerance for mutagenesis than Fluc and have a higher likelihood that hybrid constructs with murine ACS would retain bioluminescence function.

Although untested, since this enzyme sequence was more homologous it was decided to synthesise the gene and briefly characterise it to test its suitability for hybridisation with a murine ACS.

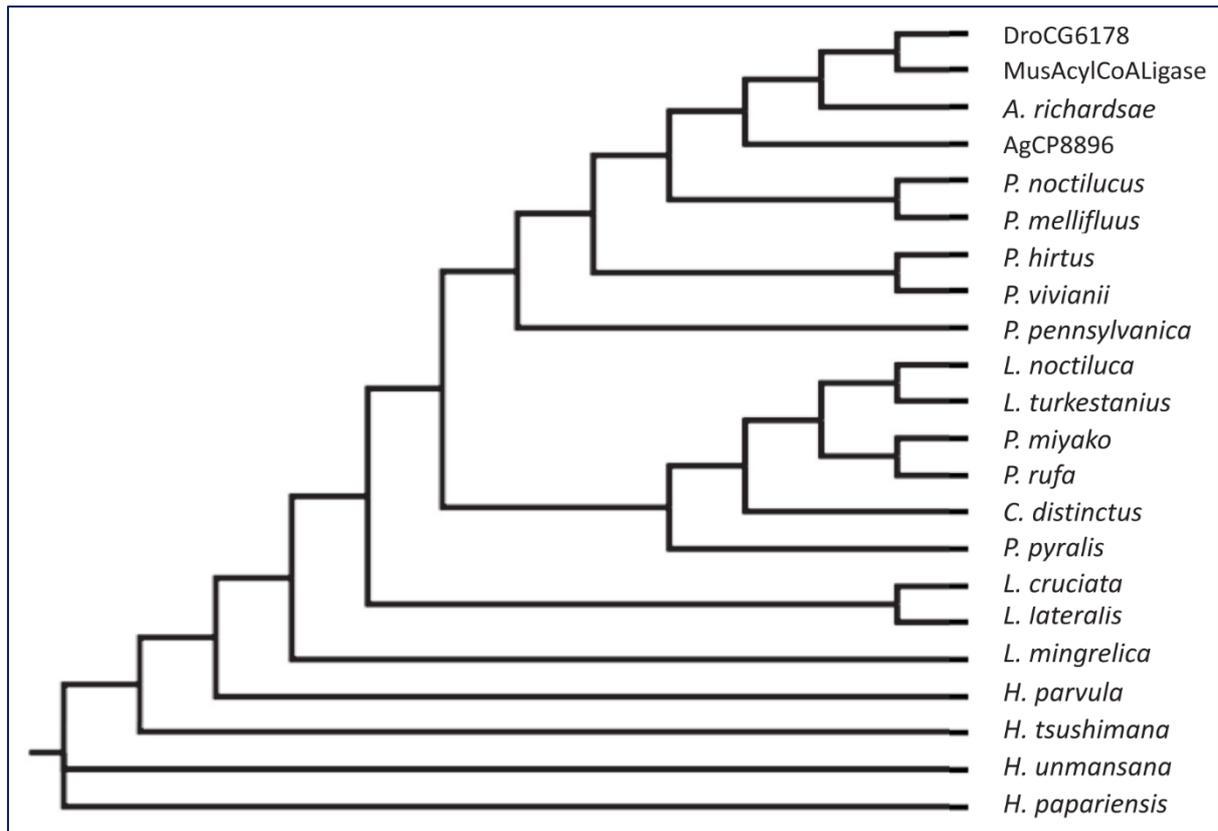


Figure 3.2 taken from (Trowell et al. 2016). **Phylogenetic tree of *Ari* GWluc, various other luciferases, and luciferase-like enzymes.** Of the many enzymes listed, of note are: *P. pyralis* (*Ppy* Fluc), *A. richardsae* (*Ari* GWluc), and Murine (Mus) Acyl CoA Ligase (ACS).

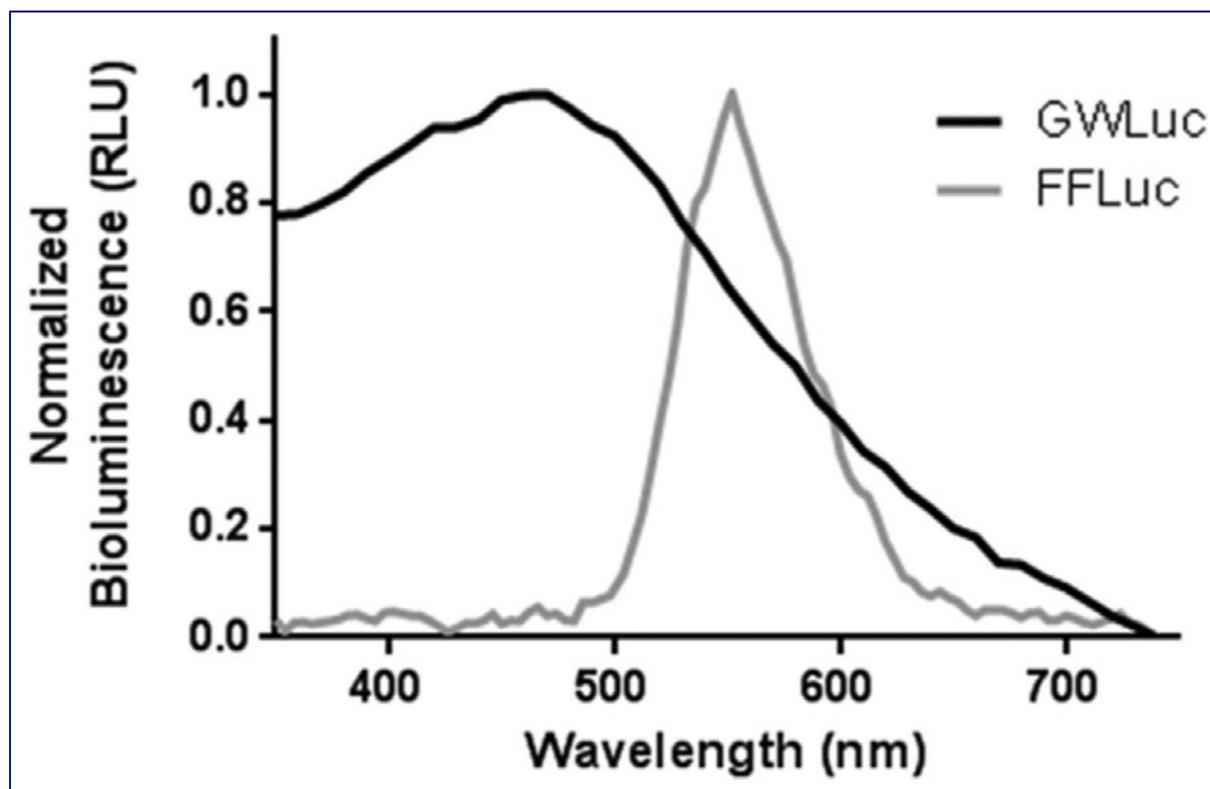


Figure 3.3 taken from (Trowell et al. 2016). Normalised emission spectra of *Ppy* Fluc and *Ari* GWLuc. An absolute emission spectra Figure was not available from the paper.

There is extremely limited information available for *Ari* GWLuc. I held some doubts about the claims of Trowell *et al.*, since in the emitted spectra the scale cuts off at approximately 350 nm in the ultraviolet (UV), specifically the UVA range which produces DNA-damaging free radicals; UVB damages DNA directly and ranges from 280 – 315 nm. It seems unlikely that an organism could survive continual exposure to an internal mutagen. Thus, it seems likely that the apparent spectrum from Trowell *et al.* was the consequence of high relative background noise, implying a very dim signal. After gene synthesis of the published sequence codon-optimised for expression in *E. coli*, recombinant expression and testing of the enzyme with D-LH₂ was carried out to test the claims of the paper and gauge its viability as a parent luciferase. The viability of *Ari* GWLuc as a parent luciferase depended on its brightness, colour, and kinetics compared to the gold standard Fluc – WT *Ppy* Fluc – which is extremely well characterised in the literature. If successful, both parent enzymes would be screened in parallel as templates with which to contribute a bioluminescence function to murine ACS. The immunogenicity of *Ari* GWLuc is unknown, but given that *Ppy* Fluc is known to be immunogenic it was decided to be worth exploring.

3.3.1.1 Comparison of *Ari* GWluc to *Ppy* Fluc via pET16b Expression

3.3.1.1.1 Design

The creation of the *Ppy* Fluc-containing pET16b plasmid used for this work is explained in Chapter 2 (see 2.1.3.1) and its map is shown below (Figure 3.4).

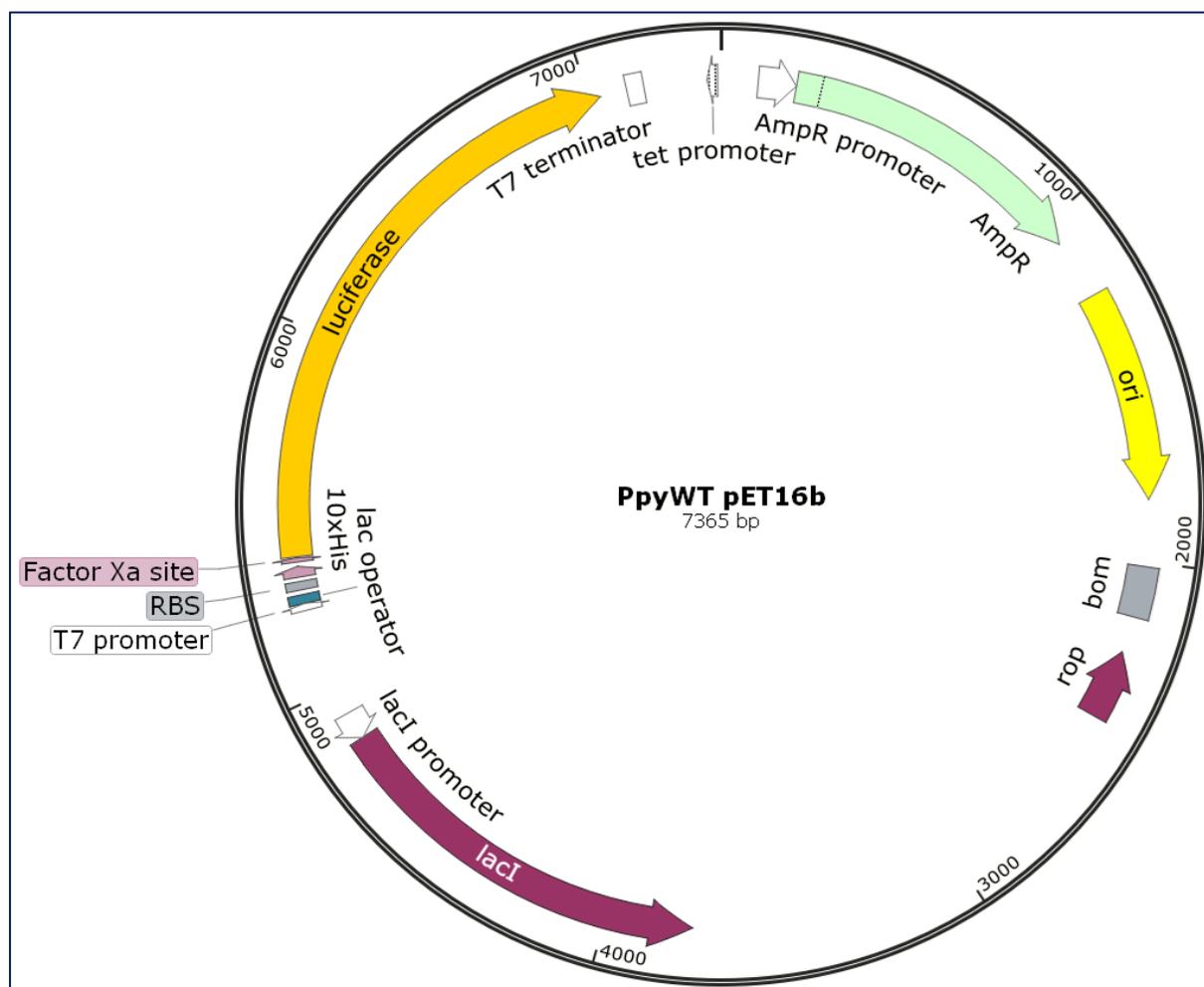


Figure 3.4. Map of WT *Ppy* Fluc within pET16b plasmid. The insert (in this case *Ppy* Fluc) is flanked by NdeI and BamHI restriction sites (not shown). An expanded version of the *Ppy* Fluc portion of this Figure – with DNA and protein sequences shown – can be found in the appendix (Figure 3.4EX).

Trowell *et al.* expressed *Ari* GWluc in pETDuet-1™ which is similar to pET16b, with the difference that pETDuet-1™ can express two constructs. As Fluc expression is well established in pET16b it was used for expression of both *Ppy* Fluc and *Ari* GWluc with 10x-N-terminal His-

tags. The presence of this tag does somewhat reduce the specific activity of Flucs but is still the gold-standard for purification (Law et al. 2006).

For *Ari* GWluc, the complimentary DNA sequence from the paper (Trowell et al. 2016) was codon optimised for expression in *E. coli* using IDT DNA's codon optimisation tool (see 2.2.1.1), after the addition of flanking NdeI and BamHI restriction enzyme sites required for cloning into pET16b. As a putative ACS superfamily luciferase, *Ari* GWluc is likely composed of two domains as for Fluc, and likely amenable to domain-exchange for chimerisation with other luciferases (explained in further detail in Chapter 4). The linker of *Ari* GWluc was putatively identified by 2D protein sequence alignment with *Ppy* Fluc, as the latter is known; the linker sequence regions were highly homologous (Figure 3.5). However, the PacI site within the linker used by Branchini *et al.* for convenient domain exchange with *Ppy* Fluc (Branchini et al. 2014) could not be mutated silently into the linker region of *Ari* GWluc (as is possible with *Ppy* Fluc) and so it was instead split into the two domains with homologous overhangs to allow ligation via Gibson assembly. *Ari* GWluc was designed (Figure 3.6) and the two fragmented sequences were then ordered as gBlock oligos from IDT DNA (Figure 3.7).

Ppy_Fluc	--MEDAKNIKKGPAPFYPLEDGT--AGEQLHKAMKRYALVPGTIAFTDAHIEVNITYAEY	56
Ari_GWluc	MACTSVNNIVYGPKPTFDVLKEANSYGEY---AFKRLRARGDEVSVIDALTGEEIRASDI	57
	..:** ** * : : . : ** *:** . :. ** :* :	
Ppy_Fluc	FEMSVRLAEAMKRYGLNTNHRIVVCSENSLQFFMPVLGALFIGVAVAPANDIYNERELLN	116
Ari_GWluc	YAKTVRTAECLQAYGIRKGDVRVGICSDTMIEYYYIVMGTMAVGAIICPIIISWTEADMNH	117
	: :** **.: : **.....* : **.: : : : *:** :.* .:* :.* :. :	
Ppy_Fluc	SMNISQPTVVFVSKKGLQKILNVQKKLPPIIQKIIIMDSKTDYQGFQSMYTFVTSHLPPGF	176
Ari_GWluc	AFNISCPTVVFVSKSILPTIARIAKRNPYVKDIIIVFDDNAPEKPLISFKDFLANPKVPSK	177
	::*** **.*****. * .* : : * * : :.***:*.: : : * : *:. * . *	
Ppy_Fluc	NEYDFVPESEFDRDKTIALIMNSSGSTGLPKGVALPHRTACVRFSHARDPIFGNQIIPDTA	236
Ari_GWluc	PHFDCEPQ--DMENTIATVLLTSGTTGISKGVVAISQYNLIHFMS-L-----DTKTYKKGL	229
	.* * : * :**** : : :****: ***** : . :.* :. :. : .	
Ppy_Fluc	ILSVVPFHGGFMFTTLGYLICGRVVLMYRFEELFLRSLQDYKIQSALLVPTLFSFFA	296
Ari_GWluc	FLCVAQYSNAFGFTALMRRAFNGTRVLHLPRYDEKSYLECVQKFKVNYISVHPPLMLSIA	289
	:.*. : :.**: : : : * ** : : *:** : *..*..** : * * : *	
Ppy_Fluc	KSTLIDKYDLSNLHEIASGGAPLSKEVGEAVAKRFHLPGIRQGYGLTETTSAILITPEGD	356
Ari_GWluc	KKPEIANYDLSSLERIYCSGTTVSVRILYQVAERIGVKVVRQFYGSSECLAVV-AQSDEF	348
	*. * :****.*..* ..* : * . : **:* : : ** ** :* :. : :	
Ppy_Fluc	DKPGAVGKVVFFFEAKVVDLDTGKTLGVNQRGELCVRGPMIMSGYVNNPEATNALIDKDG	416
Ari_GWluc	CTKGSVGTLMPGIIGKVIHPETGALLGPNERGFLKFKANSTMYGYFNNPEASKVVKDEEG	408
	. *:**.:* : .**.: ** ** *:** * .. * **.*****.:. :**	
Ppy_Fluc	WLHSGDIAYWDEDEHFFIVDRLKSLIKYKGYQVAPAELESILLQHPNIFDAGVAGLPDDD	476
Ari_GWluc	YVNTGDAGYNERFEWVVDRLKDIVMVDGVAVAPTEMETTILLHPDIIDACVIGISDGE	468
	:::**.* :**.* ..* :****.:. * **:*:* : * **:*:** * * : * . :	
Ppy_Fluc	AGELPAVVVLEHGKTMTEKEIVDYVASQVTTAKKLRGGVVVDEVPKGLTGKLDARKIR	536
Ari_GWluc	GGEVLF AFL-TKTRKEVTEKDVMDFVAEKLFPYKHLKGGCQFVDEIPKNPAGKMLRRILR	527
	.*: * : : * :****.:****.: *:**.* ****:* . ** : * : *	
Ppy_Fluc	EILIKAKKGGK	547
Ari_GWluc	GTL-----	530
	*	

Figure 3.5. *Ari GWluc* linker identification via 2D protein sequence alignment with *Ppy Fluc*.

Despite the large phylogenetic distance (as shown in Figure 3.2) the linker sequences (boxed in red) are highly homologous.

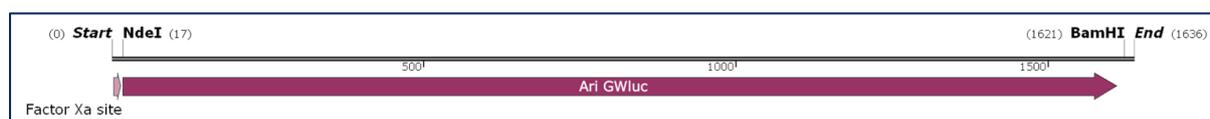


Figure 3.6. Map of entire *Ari GWluc* insert, for insertion into pET16b plasmid. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 3.6EX).

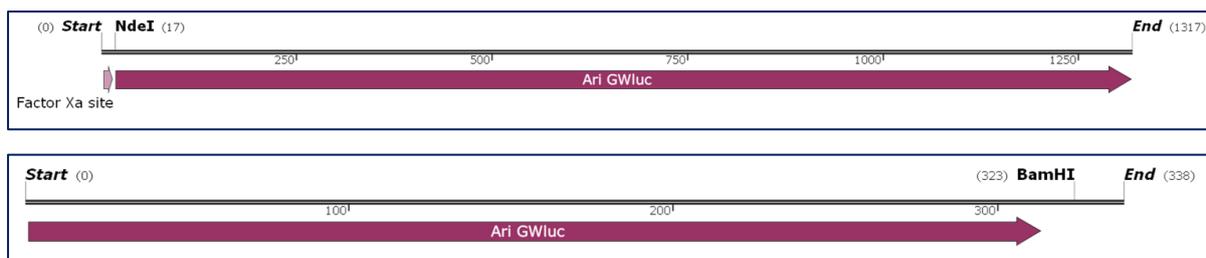


Figure 3.7. Maps of NTD (above) and CTD (below) of *Ari GWLuc* gBlock oligonucleotides, for Gibson assembly and insertion into pET16b plasmid. An expanded version of this Figure – with DNA and protein sequences shown – can be found in the appendix (Figures 3.7EXa and 3.7EXb).

3.3.1.1.2 Cloning

The *Ari GWLuc* NTD and CTD oligos were Gibson assembled and ligated into pET16b vector (see 2.2.1.7), the ligation mixture transformed into DH5 α cells and spread onto agar plates (see 2.2.3.3). Colonies were picked, grown in liquid LB and DNA miniprepmed with a small volume preserved at 4 °C. The minipreps were sent for Sanger sequencing (see 2.2.1.0) and the sequence-correct constructs retransformed into BL21 cells (see 2.2.3.3).

3.3.1.1.3 Screening

The exact protocol of Trowell *et al.* was followed for the expression and screening of the *Ari GWLuc* protein (see 2.2.7.1). The lysate was screened with D-LH₂ but no bioluminescence signal was detected (data not shown, see Figure 9.1).

Briefly after subcloning, enzymes were over-expressed as previously described (see 2.2.3.5). The *Ari GWLuc* lysate and *Ppy Fluc* were examined using silver-stained SDS-PAGE (see 2.2.2.5), and the *Ppy Fluc* band was visible at ~5 kDa larger than the expected size of ~63.3 kDa, however no band for *Ari GWLuc* at the expected similar size of 60.3 kDa (nor ~65.3 kDa if assuming the same size discrepancy as *Ppy Fluc*) was observed (Figure 3.8).

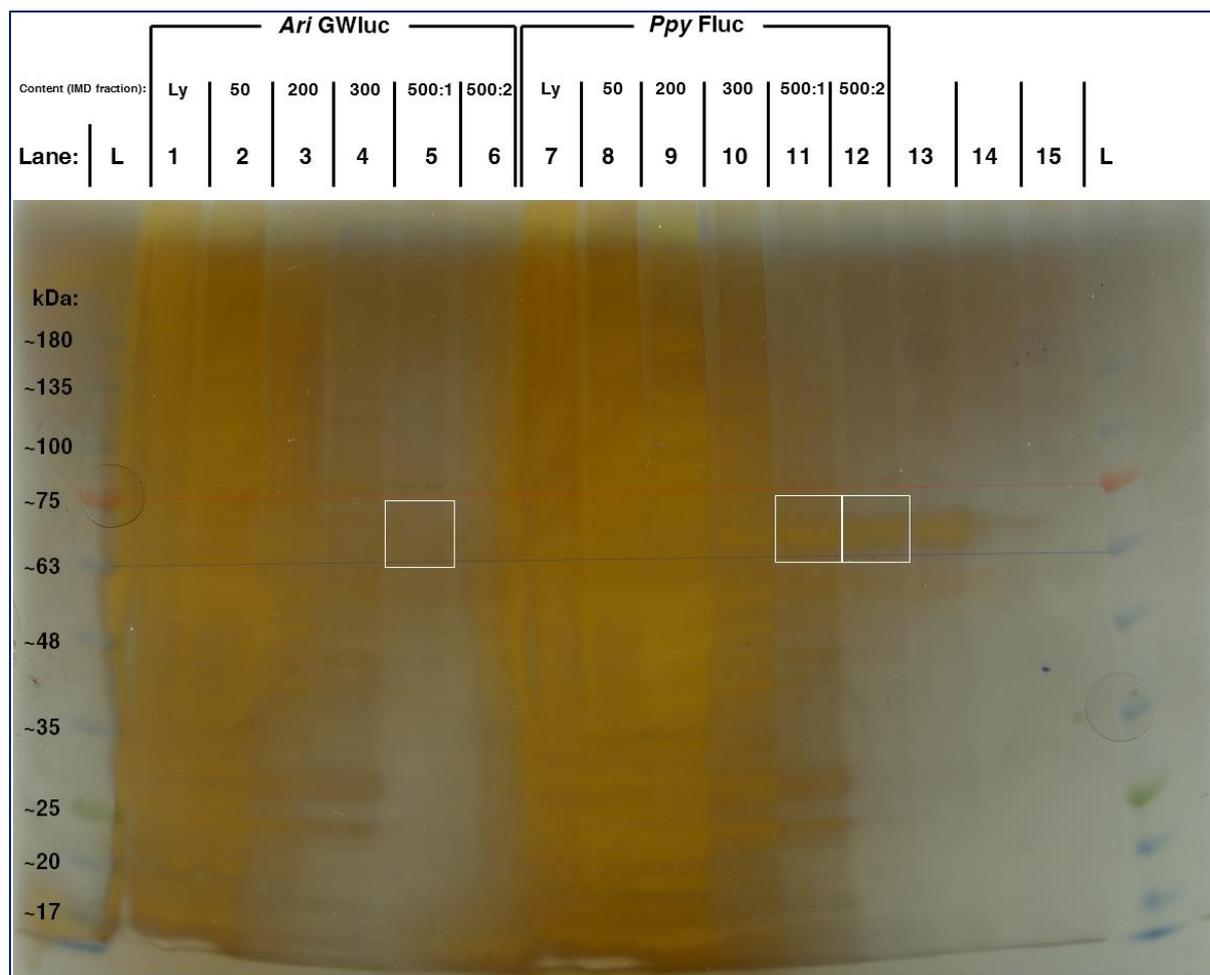


Figure 3.8. Silver-stained SDS-PAGE of *Ari GWluc* and *Ppy Fluc* protein purification fractions.

L stand for ladder, which is Biotech rabbit TriColor (Broad Protein) and is traced together for each band with a correspondingly coloured line. Ly stands for lysate. The IMD fractions are 50mM, 200mM, 300mM, 500mM wash 1, and 500mM wash 2. There is substantial spill-over of lane 7 into 6, and lane 13 into 14. The white box in lane 5 is where *Ari GWluc* would appear if present. The white box in lane 12 indicates the *Ppy Fluc* band. Size predictions of constructs (using SnapGene): *Ari GWluc* = 60.3 kDa, *Ppy Fluc* = 63.3 kDa. As *Ppy Fluc* appeared 5 kDa larger than theoretically predicted on this particular gel-run, *Ari GWluc* was assumed to appear around the 65 kDa mark if present.

As *Ari GWluc* failed to express using pET16b in BL21 *E. coli* when the aforementioned conditions were followed, this cast further doubt on the claims made by Trowell *et al.* in which the authors apparently successfully expressed *Ari GWluc* in *E. coli* and found the lysate to bioluminesce with D-LH₂, producing bluish light. Thus, as the lysate neither produced bands

of the expected size on SDS-PAGE nor bioluminescence, it was not purified. Instead, expression using a different plasmid expression vector was attempted.

3.3.1.2 Comparison of *Ari* GWluc to *Ppy* Fluc via pCold™ Expression (only for *Ari* GWluc)

3.3.1.2.1 Design

Ari GWluc failed to express in pET16b in BL21 cells, however other expression vectors are available. pCold™ TF (Takara Bio Inc., GenBank Accession No: AB213654) is an *E. coli* vector specifically designed for difficult-to-express proteins, and was successfully used to express the luciferase-like ACS found in mealworms (Viviani et al. 2013). pCold™ TF expresses trigger factor (TF) which is a 48 kDa ribosome-associated chaperone protein that assists with the folding of the co-expressed protein. TF is expressed as a soluble 6xHis-tagged protein. *Ari* GWluc itself was to be expressed as a 10xHis-tagged protein.

3.3.1.2.2 Cloning

Ari GWluc was ordered as a gBlock oligo for insertion into pCold via Gibson Assembly (Figure 3.9). Note that this design introduces a double 10x HisTag unintentionally.

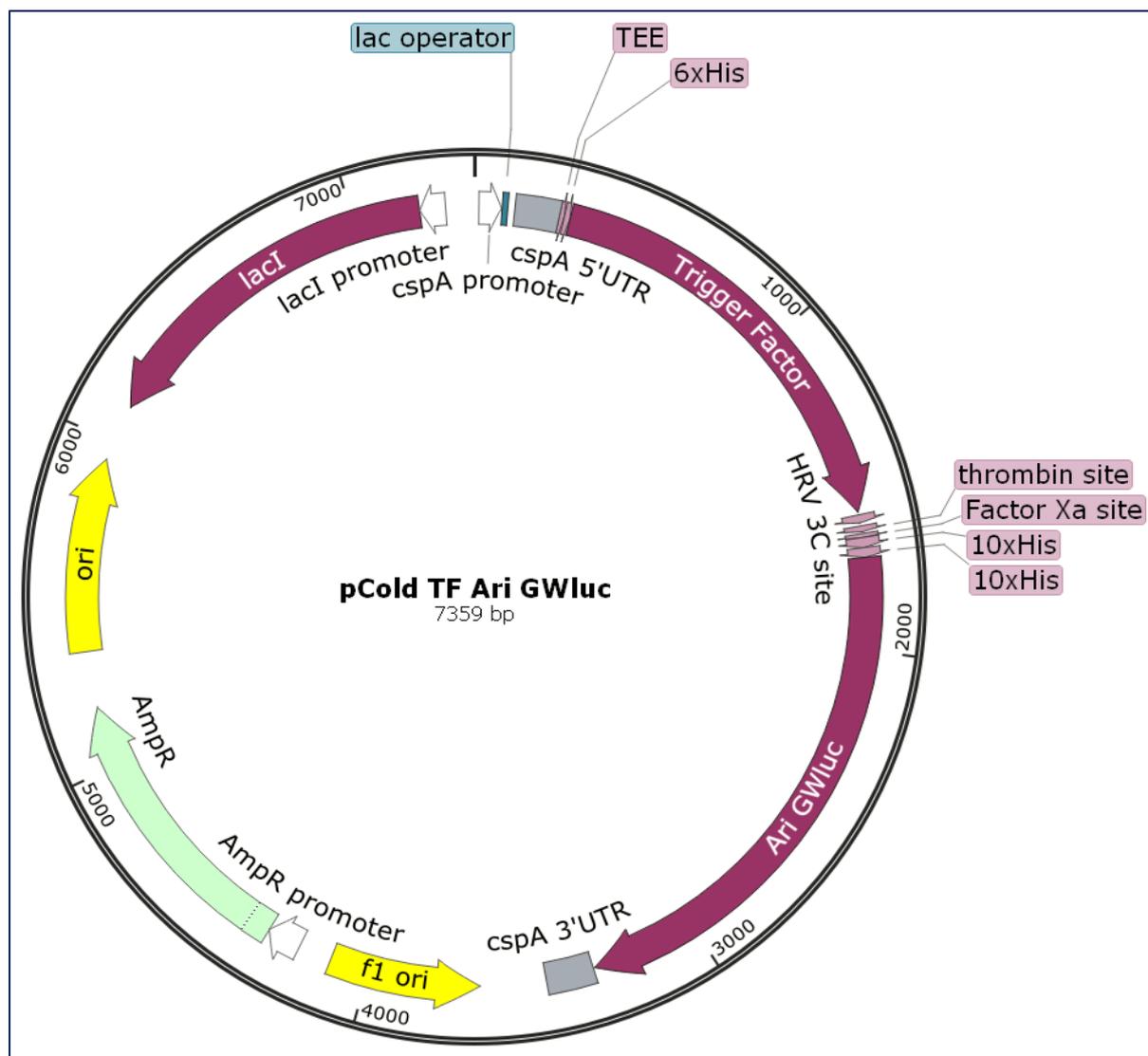


Figure 3.9. Map of WT *Ari GWluc* within pCold TF plasmid. The insert (in this case *Ari GWluc*) is flanked by *NdeI* at the N-terminus, but no restriction site exists at the C-terminus, hence why Gibson assembly is used for insertion.

The *Ari GWluc* oligo was Gibson assembled into pCold vector (see 2.2.1.8), the ligation mixture transformed into DH5 α cells and spread onto agar plates (see 2.2.3.3). Colonies were picked, grown in liquid LB and DNA miniprep with a small volume preserved at 4 °C. The minipreps were sent for Sanger sequencing (see 2.2.1.10) and the sequence-correct constructs retransformed into BL21 cells (see 2.2.3.3).

3.3.1.2.3 Screening

The expression of *Ari* GWluc was only of interest for this PhD project if the expressed protein was also bioluminescent with D-LH_2 . *Ari* GWluc expression was attempted using a matrix of conditions with varying IPTG concentrations, incubation times and temperatures and each combination was screened with D-LH_2 using the PhotonIMAGER (see 2.2.6.4). Bioluminescence signal was observed in some conditions, with the optimal being 0.4 mM IPTG for 18 h at 15 °C. These conditions were used for bulk expression of the protein which was then purified using the Ni-NTA method (see 2.2.3.7.2). The purified protein was intended for bioluminescence screening with D-LH_2 and was thus immediately desalted (see 2.2.2.1) to prevent damage to the purified protein from the imidazole. The purified and desalted protein was run on two SDS-PAGE gels (see 2.2.2.3) and Coomassie stained (see 2.2.2.4) to confirm its size and approximate yield and purity. The protein concentration was too low to visualise via Coomassie staining so a silver stain (see 2.2.2.5) was performed on an identical gel (Figure 3.10). A band approximating to the correct size was observed at ~61.7 kDa, the design of the gBlock resulted in *Ari* GWluc unintentionally having a double 10xHis-tag which increased the band size on the gel. The band for trigger factor was also observed at ~48.2 kDa. Although silver staining has a potentially high dynamic range, as the stain saturates the bands very quickly it is difficult to avoid overexposed bands. Further, unlike Coomassie in which the colour intensity increases linearly with protein concentration and exposure, silver staining rapidly becomes dark brown and then gradually fades to translucent orange. Thus, protein quantification from stained gel images is much more complex and not feasible in this instance. Further, the presence of high background staining and several other impurity bands in the *Ari* GWluc fraction would make a Bradford assay inaccurate.

The desalted protein fraction was screened with D-LH_2 for bioluminescence activity relative to *Ppy* Fluc (from pET16b) (see 2.2.6.3). The specific activity of *Ari* GWluc was found to be ~0.0007% that of *Ppy* Fluc (Figure 3.11). Across a pH range of 7.15 to 8.15, and without controlling for protein concentration, the pH dependence of *Ari* GWluc is broadly similar to *Ppy* Fluc (Figure 3.12). The spectra for the optimal pH of both enzymes (pH 7.8) was also found to be highly similar (Figure 3.13); although *Ari* GWluc does appear somewhat blueshifted this is likely an artefact caused by low brightness overall, which results from the very dim background “light” (sensor noise present for all bandpass filters, but noticeable in

the 472 – 522 nm and 647 – 822 nm range) being much more elevated in order to normalise the peak emission with *Ppy* Fluc. Further, the λ_{\max} for both *Ari* GWluc and *Ppy* Fluc is strongly within the 547 nm bandpass filter, suggesting an almost identical λ_{\max} .

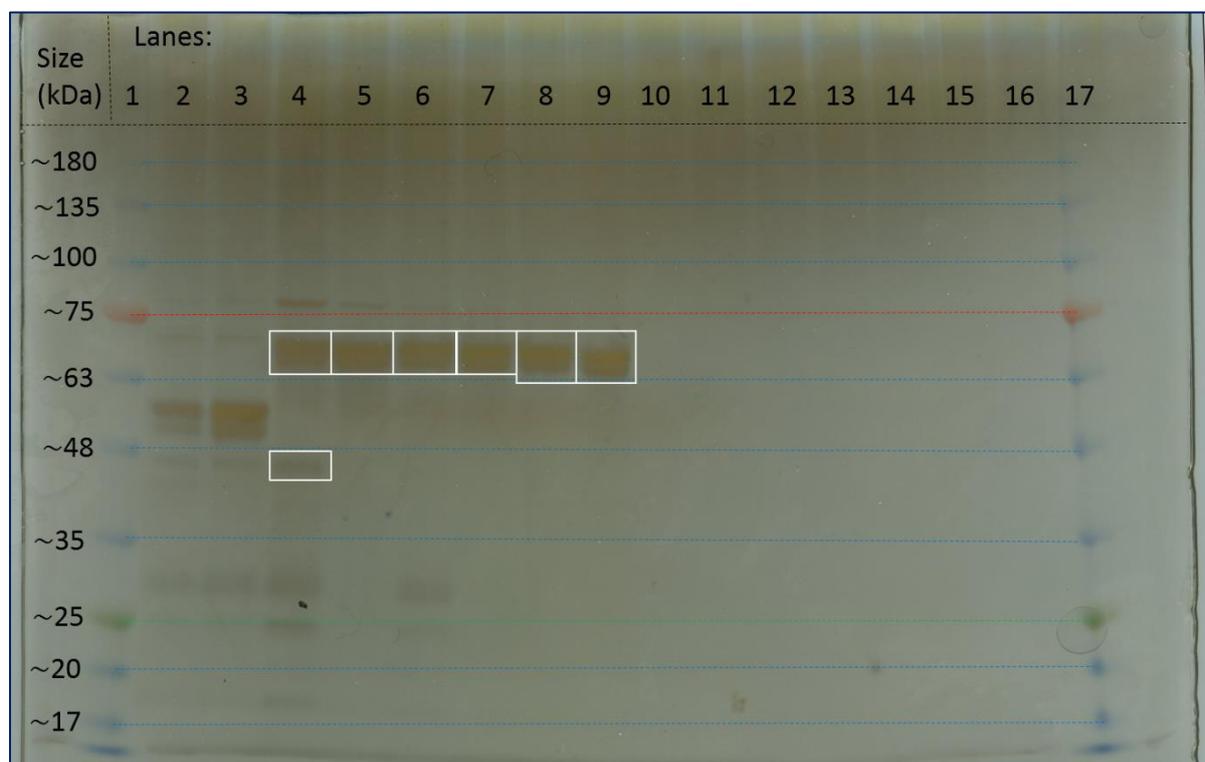


Figure 3.10. Silver-stained SDS-PAGE of *Ari* GWluc, *Ppy* Fluc, ACSM1 (MTS), ACSM3 (MTS), ACSM1 (no MTS), and ACSM3 (no MTS). Lanes 1 and 17 contain Biotechrabbit TriColor (Broad Range) Protein Ladder which are traced together for each band with a correspondingly coloured dotted line. Lanes 4 – 9 are boxed in white as follows: lane 4 contains a desalted fraction of *Ari* GWluc and co-purified trigger factor; lane 5 contains a desalted fraction of *Ppy* Fluc; lanes 6 – 7 contain ACSM1 and ACSM3 (respectively) with MTSs; lanes 8 – 9 contain ACSM1 and ACSM3 (respectively) with MTSs removed. Lanes 1 – 2 contain other constructs. Lanes 10 – 16 are empty. Size predictions of constructs (using SnapGene): *Ari* GWluc = 61.7 kDa, Trigger Factor = 48.2 kDa, *Ppy* Fluc = 63.3 kDa, ACSM1 (with MTS) = 67.3 kDa, ACSM1 (no MTS) = 64.0 kDa, ACSM3 (with MTS) = 68.2 kDa, ACSM3 (no MTS) = 64.9 kDa.

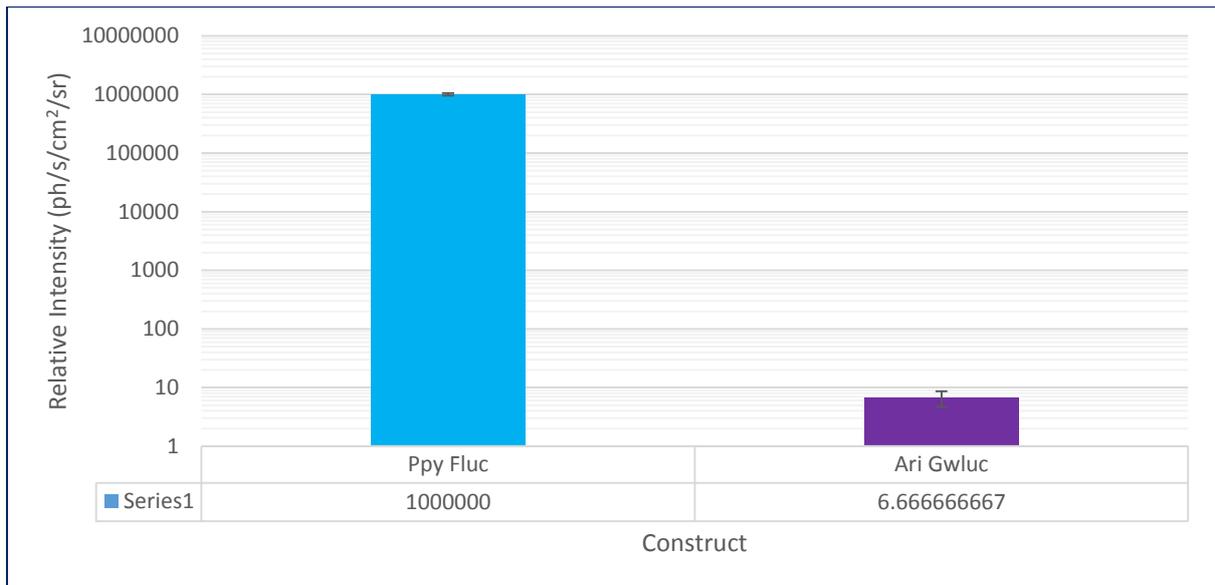


Figure 3.11. *Ppy Fluc*-relative specific activity of *Ari GWluc*. The log scale is relative to a *Ppy Fluc* brightness of 1,000,000, resulting in the *Ari GWluc* value being greater than 1 thus allowing error bars to display correctly.

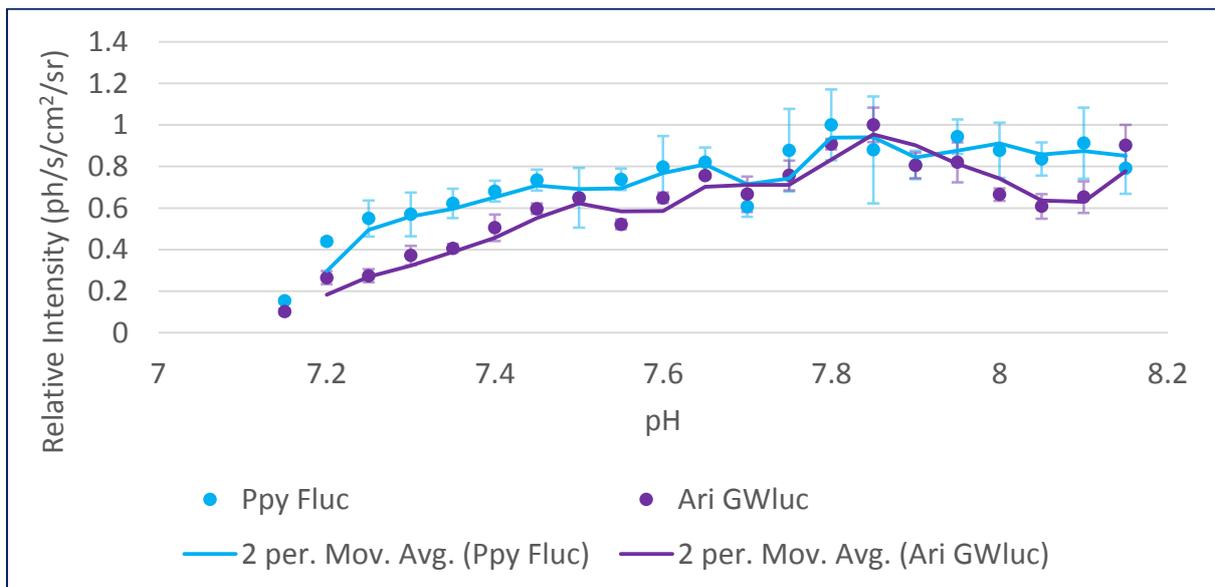


Figure 3.12. Normalised pH dependence of total light output of *Ppy Fluc* and *Ari GWluc*.

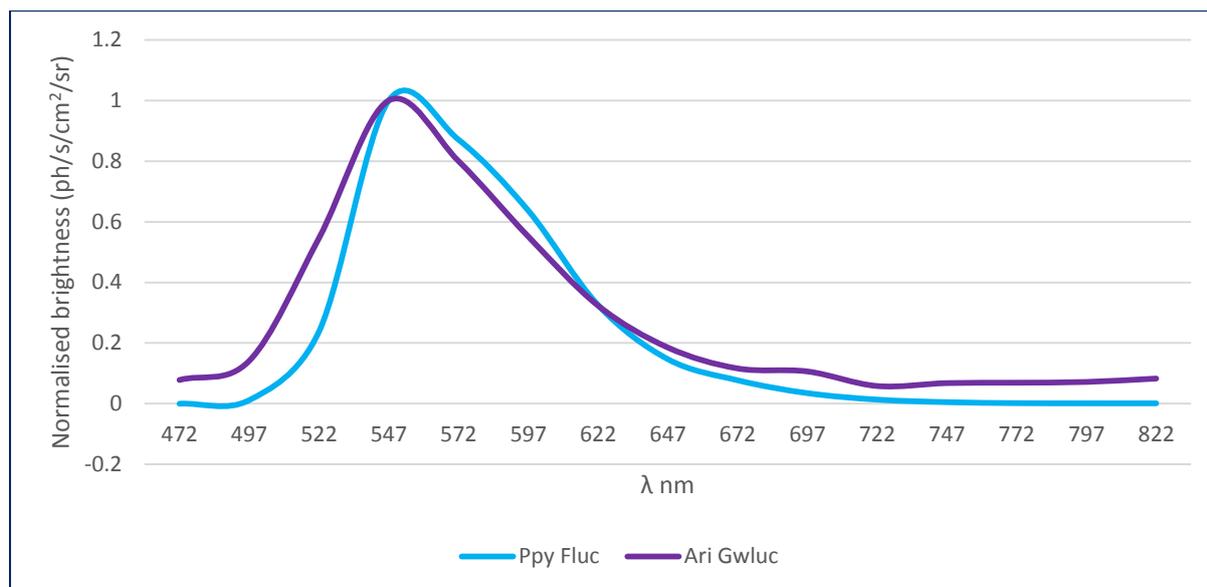


Figure 3.13. Normalised spectra of *Ari GWluc* and *Ppy Fluc* at pH 7.8. The apparent marginal blue-shift of *Ari GWluc* is likely an artefact of its low light output (figure 3.11).

3.3.1.3 Mini Discussion

Although *Ari GWluc* was successfully expressed it appeared to have very low activity with D-LH₂ and exhibited a broadly similar pH spectrum to *Ppy Fluc*. In 2018, another paper was published (Watkins et al. 2018) which found that a close relative of *Arachnocampa richardsae* – *Arachnocampa luminosa* – did not produce bioluminescence when its crudely extracted luciferase was combined with D-LH₂, but that bright blue bioluminescence was observed when combined with *A. luminosa* hot extract (luciferin and denatured luciferase) suggesting a different luciferin is present in the fungus gnat. Further, the researchers found that the crude luciferase extract produces bioluminescence (of similar colour and intensity to the organism) when combined with ATP-Mg²⁺, tyrosine, and xanthurenic acid at once, implying that a luciferin synthetase is present in the crude extract which uses tyrosine and xanthurenic acid as substrates to produce this novel luciferin (Figure 3.14). Thus, it appears that the *Arachnocampa* genus may have its own distinct luciferin (Figure 3.15) which is structurally similar enough to beetle luciferin that a small degree of cross-reactive bioluminescence is possible.

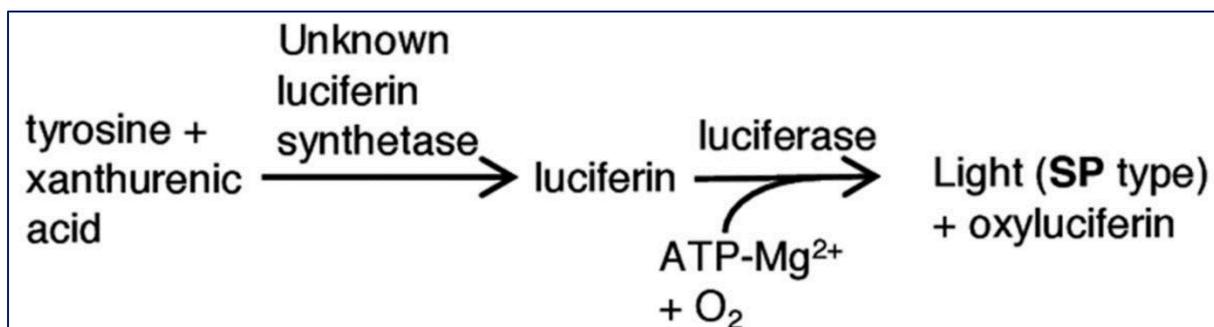


Figure 3.14 taken from (Watkins et al. 2018). Possible *Ari* GWLuc luciferin biosynthetic and bioluminescence pathway. SP stands for slow peak, as the luciferase is bottlenecked by the luciferin synthetase, thus slowing the emission kinetics.

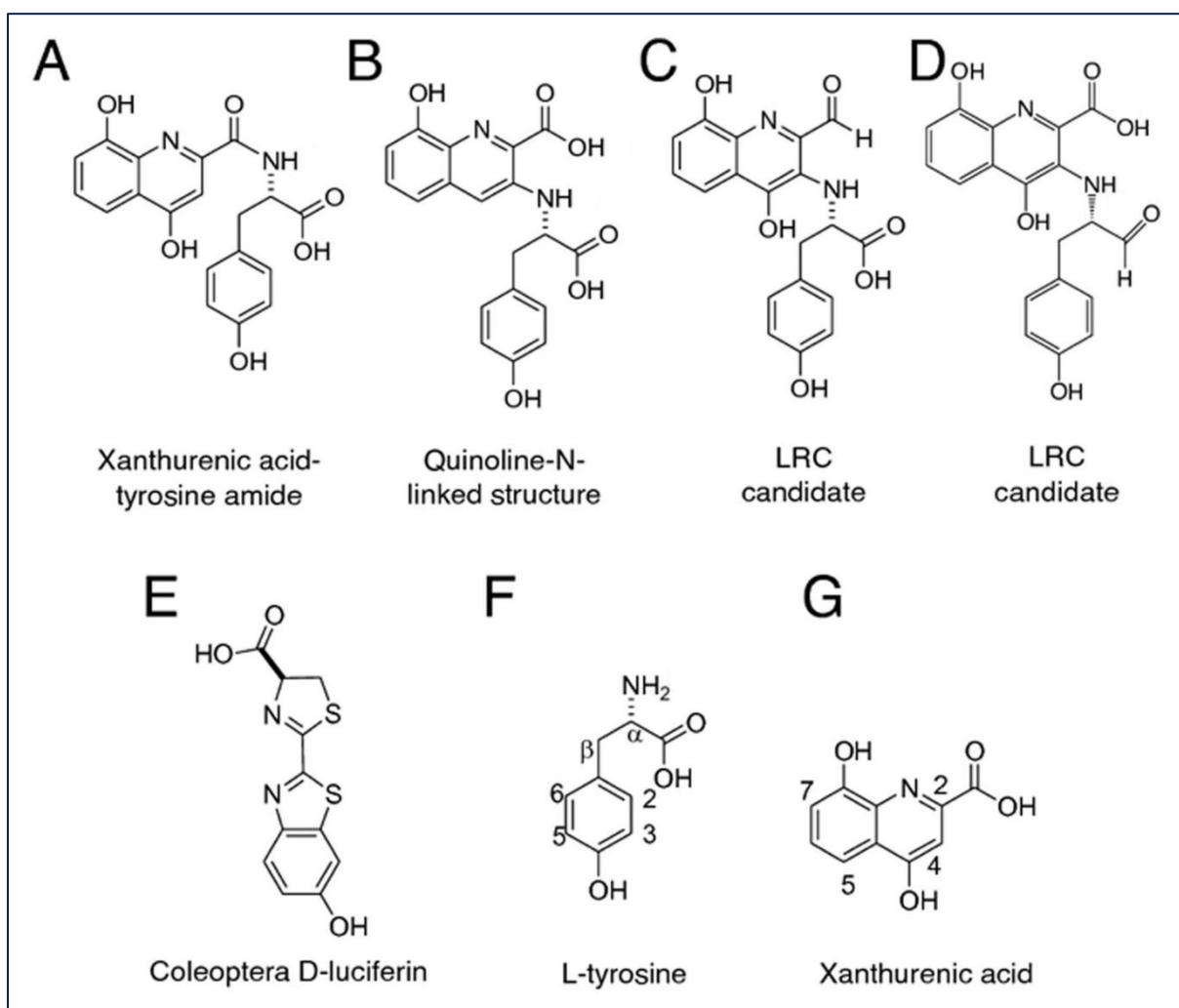


Figure 3.15 adapted from (Watkins et al. 2018). Small molecules involved in the bioluminescence of *Ari* GWLuc and Fluc. LRC stands for Luciferin-related compound.

Although *Ppy* Fluc is known to be immunogenic, it is extremely well characterised, with many thermo-stable, pH-stable, and red-shifted mutants available and known to perform well in vivo. Hence it was chosen as the luciferase parent for this project.

3.3.2 Selection of the ACS parent

With *Ppy* Fluc now chosen as the bioluminescent parent protein of choice, the protein basic local alignment search tool (BLASTp) from The National Center for Biotechnology Information (NCBI) was used to search for homologous murine proteins. As was expected the most homologous proteins returned (of those whose function was identified) were all ACSs, and these were predominantly ACSMs.

A Clustal Omega 2D protein alignment of murine ACSM3 and *Ppy* Fluc revealed a poor homology (~25%), however proteins may have radically different primary sequences while maintaining similar tertiary structure (Chavelas-Adame et al. 2011)

Threading (also known as fold recognition) is a technique used to generate a 3D protein file (usually a .pdb file) from a 2D protein by using another known .pdb file of assumed similar structure. I.e. the 2D protein 1 is “threaded” onto the 3D protein 2 to create a predicted 3D protein 1. Human ACSMs are assumed to be highly similar in structure to their murine counterparts, given their similar primary sequences. Thus, a human ACSM (2VZE.pdb) was used as a 3D template, and the primary sequence of ACSM3 threaded onto it using the online tool Swiss Model (<https://swissmodel.expasy.org/interactive>).

The predicted 3D ACSM3 was then aligned to *Ppy* Fluc (4g37.pdb) using the online tool “TM-align” (<https://zhanglab.ccmb.med.umich.edu/TM-align/>). The alignment score was 0.85034 out of 1 and thus the resulting 2D alignment was treated with high confidence. From the visual display of the alignment created in PyMOL (Figure 3.16) it appears that the N-terminal domain (NTD) aligns well but the C-terminal domain (CTD) is tilted slightly out of overall alignment which one would expect due to the flexible linker allowing a range of positions the CTD may have relative to the NTD, and the two crystal structures compared happen to have different NTD-to-CTD conformation. Despite the appearance of low CTD homology the

program gave high alignment scores for the CTD residues, nevertheless only regions within the NTD alignment are of interest for design and cloning.

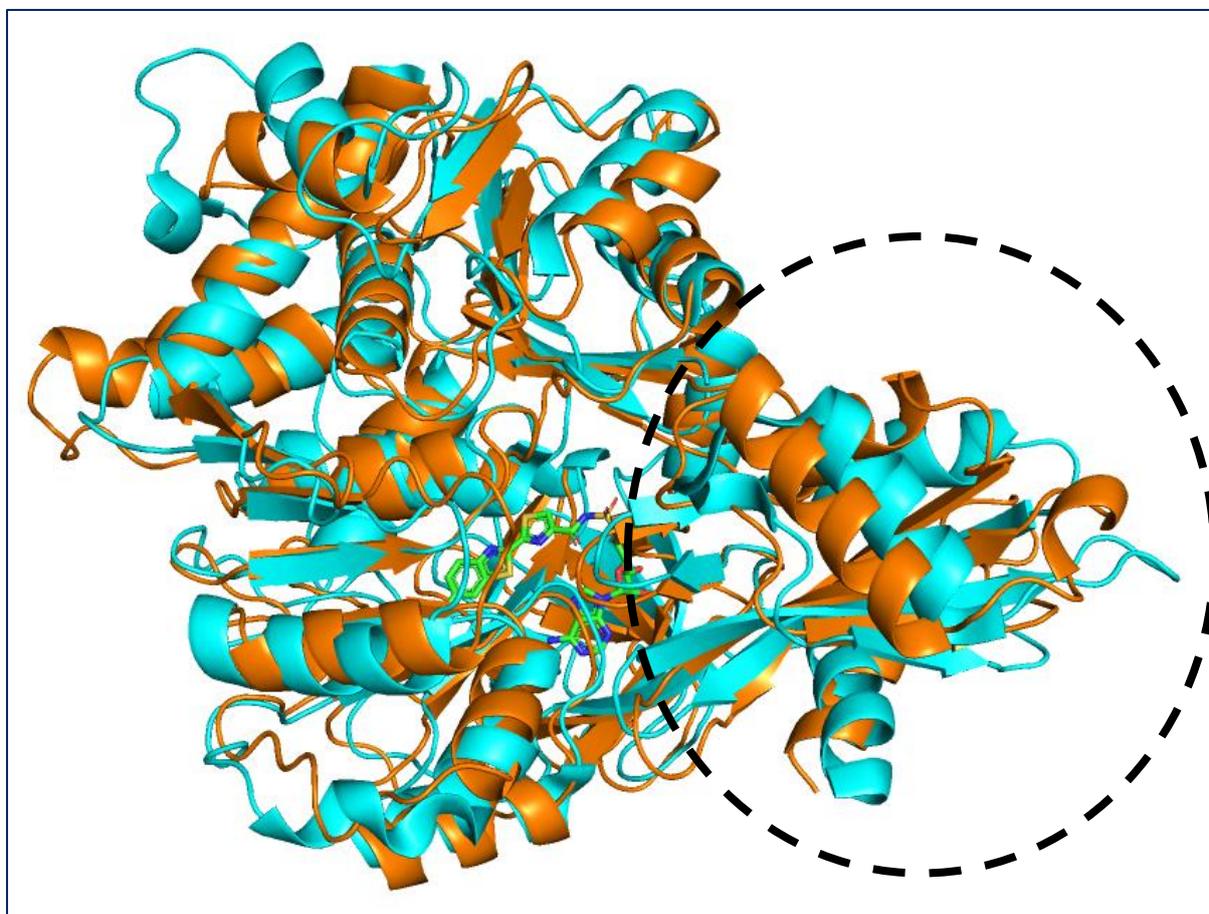


Figure 3.16. Structural alignment of *Ppy Fluc* and ACSM3 model. *Ppy Fluc* in cyan and ACSM3 in orange. CTD crudely circled. Luciferyl adenylate (carbons in green) bound to active site of *Ppy Fluc*.

ACs are translated in the cytosol with N-terminal mitochondrial targeting sequence (MTS) and shuttled into the mitochondria by heat shock protein 60 (HSP60). The MTS-mediated binding of HSP60 prevents folding in the cytosol, and it is assumed that the pre-mitochondrial form is inactive or less active. As with many other mitochondrial proteins, the MTS is cleaved by a mitochondrial processing peptidase (MPP) (Regev-Rudzki et al. 2008). Due to the potential for reduced activity, the pre-mitochondrial forms were excluded from the BLAST results. Isoforms of both ACs and Flucs may exist due to alternative splicing of transcribed chromosomal DNA. This is because complementary DNA does not often exist as a single uninterrupted sequence in the genome, often there are introns which allow the chromosomal

DNA (as exons) to be transcribed as short fragments which have a number of possible splicing combinations (Wood et al. 1989).

Many of the most homologous ACSs were examined by multiple protein sequence alignments for homology, particularly in important regions known in the literature of bioluminescent ACS enzymes. Of particular interest was the equivalent sequence to the region spanning R339 – A348 (GYGLTET TSA) in *Ppy Fluc* as it is a highly conserved motif forming part of the active site of *Fluc*, and was recently found to be the primary determinant of bioluminescence in insect ACS (Viviani et al. 2013);(Oba et al. 2009). The highest-scoring murine proteins are compared with WT *Ppy Fluc*, along with *Zophobas Morio* protoluciferase (positive control) and bovine serum albumin (BSA) (negative control) (table 3.2). Bovine serum albumin is the gold standard for negative control proteins, however it should be noted that it may possess some bioluminescent activity with luciferyl adenylate (Viviani and Ohmiya 2006).

As a potentially large number of modifications were planned (e.g. active site consensus sequence exchange, random mutagenesis, substrate determining residue substitutions, and others which are all covered in detail later in this chapter) it was deemed unfeasible within the timeframe of this PhD project to carry these out on multiple murine ACSMs in parallel, as such the optimal ACSM was chosen. ACSM2 appeared marginally more appropriate as it has the greatest overall homology to *Ppy Fluc* (via NCBI pBLAST), however ACSM3 has the highest homology to the *Ppy Fluc* region of the active site consensus sequence (GYGLTET TSA) (Figure 3.17). Little is known in the literature about ACSM2 whereas ACSM3 has been characterised (Fujino et al. 2001). One overarching engineering strategy of this project was to – where possible – stay within what is known from the literature, for this reason ACSM3 was chosen.

The murine ACSM3 template used for this chapter was generated using IDT DNA's codon optimisation tool. The resulting DNA sequence was then ordered as a gBlock oligo (Figure 3.18).

<i>Ppy Fluc</i>	AKNIKKGP-APFYPLEDGTAGEQLHKAMKRYAL---VPGTIAFTDAHI-EVNITY-AEYFEMSVRLAEAMKRY-GLNTNHRIVVSSENSLOFFMPVLGALFIGVAVAPANDCYNERELLN	120
ACSM3	-----IE---IPEYFNFAKVDLQWNTMEKAGKRLSNPAFWIDGNGEELRWSFEELGLLSRKFANILTEACSLQRGDRVMVILPKIPEWLANVAACLRTGTVLIPGTTQLTQKDILY	120
<i>Ppy Fluc</i>	SMNISQPTVVVFSKGLQKILNVQKLPKLPKIIIMDSKTDYQGFQSMYTFVTSHLPPGFNEYDFVPESFDRDKTIALIMNSS----PKGVALPHRALAVRFSHARDPIFGNQIAPDTAI	240
ACSM3	RLQSSKAKCIITDDTLAPAVDAVAAKCENLHSLKLVISQ-HSREGWGNLKEMMKY-A---SDS--HTCVDTKHD-EMMAIYFTSGTTGPPKMIGHTHSSFGGLSVNGRF-WDLIA-SOVM	240
<i>Ppy Fluc</i>	LSVVPFHGFGMFTTL-GYLISGFRVVLN-Y-RFEEELFLRSLQDYKIQSALLVPTLFSFLAKSTLIDKYDLSNLHEIASGGAPLSKEVGEAVAKRFHLPGIRQGYGLTETTSALLITPKG	360
ACSM3	WNTSDTGWAKSAWSSVFSWPTQGACVFAHYLPRFESTSILQTLKFPITVFCAPTAYRMLVQN-DMSSYKFNSLKHCVSAGEPINPEVMEQRKKTGL-DIYEGYGQTET-VLTCGNFKG	360
<i>Ppy Fluc</i>	D-DKPGAVGVVPPFEAKVVDLDTGKTLGVNQRGELSVR----GPMHMSGYVNNPEATNALIDKDGWLHSGDIAYNDEDEHFFIVDRKLSLIKYKGCQVAPAELESILLQHPNIFDAGVA	480
ACSM3	MKIKPGSMGKPSAFDVKILD-ENGATLPPGQEGDIALQVLPERPFGLFTHYVDPNSKTASTLRGS-FYITGDRGYMDEGYFWFVARSDDLIKSSGYRIGPFVESALIEHPSIAESAVV	480
<i>Ppy Fluc</i>	GLPDDDAGELPAAVVLEHGKTMT-EKEIVDYVASQVT---TAKKLRGGVVFVDEVPKGL-KLDARKIREILIKAK	554
ACSM3	SSPDPIRGEVVKAFIVLNPDKSHDQEQLKKEIQEHVKKTTAPYKYPRKVEFIEELPKTVSGKVKRNLKKEWV--	554

Figure 3.17. ACSM3 active site identification via protein sequence alignment derived from 3D alignment of ACSM3 and *Ppy Fluc*. Active site sequence is boxed in red. For *Ppy Fluc* the residue sequence M1 – D3 is omitted from the N-terminus and K544 – L550 from the C-terminus; for ACSM3 the residue sequence M3 – K45 is omitted from the N-terminus and T579 – T580 from the C-terminus.

Protein name and species	<i>Ppy Fluc</i> active site consensus sequence	Homology of the equivalent sequence	Proportion of sequences which were aligned
<i>Photinus pyralis</i> firefly luciferase (NCBI: P08659.1)	GYGLTETTSA	100%	
<i>Zophobas Morio</i> protoluciferase with bright I327T/S substitution (Viviani et al. 2013)	GYGMSE [T / S] GLL	>29.07%	68%
<i>Zophobas Morio</i> protoluciferase (4W8O.pdb)	GYGMSEIGLL	29.07%	68%
Murine acyl-CoA synthetase 2, mitochondrial isoform 3 (NCBI: NP_666309.1)	IYGQTETGLI	25.00%	92%
Murine acyl-CoA synthetase 3, mitochondrial (NCBI: NP_058566.3)	GYGQTETVLI	24.17%	89%
Bovine serum albumin (NCBI: AAN17824.1) – as negative reference	n/a	35.56	7%

Table 3.2. *Ppy Fluc* active site homology with a number of acyl-CoA synthetases. This table is not an exhaustive list of murine ACSs examined for the project using this method.

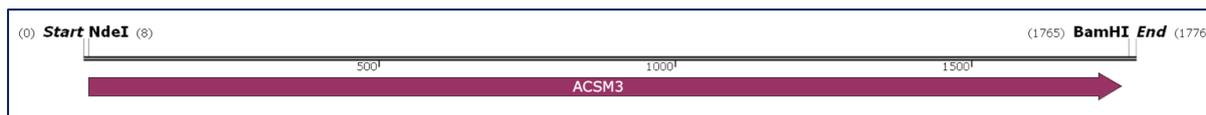


Figure 3.18. Map of ACSM3 gBlock oligonucleotide, for insertion into pET16b plasmid. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 3.18EX)

3.3.2.1 Cloning of ACSM3 into pET16b and transformation into *E. coli* and screening with Δ -LH₂

The ACSM3 gBlock was digested with NdeI and BamHI and subcloned into pET16b and transformed into DH5 α then BL21 cells, which were screened for bioluminescence with Δ -LH₂ using the PhotonIMAGER (see 2.2.6.4.1). It is assumed that ACSM3 is non-bioluminescent, this was expected as mice are non-bioluminescent when injected with Δ -LH₂. The ACSM3 colonies were non-bioluminescent as expected (data not shown, see Figure 9.1). However, this could be because the *E. coli* cells fail to express the mammalian protein. To examine this, ACSM3 and *Ppy* Fluc were overexpressed as liquid culture and screened as crude lysates. Furthermore, a novel small-scale optimised nINTA protein purification was performed to isolate the His-tagged proteins (see 2.2.3.7.1). ACSM3 was purified with *Ppy* Fluc in parallel for optimal fraction desalting via BLI probing of fractions. Purified and desalted fractions were run via SDS-PAGE and silver-stained, and bands of the expected size (~68 kDa) were very faintly visible. Due to the apparent low concentration the protein was 25 x concentrated (see 2.2.2.2) and run again via SDS-PAGE and silver-stained, bands were clearly visible (Figure 3.10). ACSM1 was also assessed in the same way (Figure 3.10), and is explained in Chapter 4.

3.3.3 Design, cloning, and screening of ACSM3 without mitochondrial targeting sequence

Both ACSM1 and ACSM3 contain predicted mitochondrial targeting sequences (MTSs) (Figure 3.19), and these were assumed to be irrelevant for expression in *E. coli* and thus were left intact as a precaution as discussed earlier in this chapter. There is extremely poor expression of constructs containing the NTD of ACSM1, ACSM2, and ACSM3 (as explained further in chapter 4). Following a discussion with protein engineer Dr D. Dafydd Jones, it was considered that this sequence could be interfering with protein folding. It was crucial at this stage in the project to ascertain whether ACSM3 has any bioluminescent activity, as if so directed evolution would have a very high chance of success. Therefore it was decided that expression

screening without this sequence would be carried out. The predicted MTS for ACSM2 is not known in the literature, further ACSM2 was the poorest performing ACSM CTD when domain-exchanged with *Ppy* Fluc (also discussed in chapter 4), and for these reasons it was omitted.

hMACS1	..	MQWLMR	FRTLWG	IHKSPHNIHPAPSQLRCRSL	SEFGAPRWN.DYEVP	PEEFNFASYVL	56
mMACS1	..	MQWLMK	FQ	ICKVLOGFSLSPTQLHRRIF	SRVGAPRWN.DHDS	PEEFNFASDVL	52
hSa	..	MLRHAKC	FQR.LA	FGSVRALHKDHRTATPQNF	SNYESMKQDFKLG	PEYFNFAKDVL	57
mSa		MVNLLEARC	FQR.LA	PDPMRVLYKDYRTATPQNF	SNYESMKQDFKIEI	PEYFNFAKDVL	59

Figure 3.19 taken from (Fujino et al. 2001). Putative mitochondrial targeting sequences of ACSM1 and ACSM3. h/mMACS1 refers to human and murine ACSM1 respectively; h/mSa refers to human and murine ACSM3 respectively. The putative mitochondrial targeting sequences are boxed, identical residues are highlighted in solid black.

3.3.3.1 Design

The MTSs of ACSM1 and ACSM3 were to be removed by replacement via restriction digestion with a truncated segment. Thus, the DNA sequences for these constructs was examined for unique commonly-exploited restriction digestion sites as little downstream of the C-terminal end of the MTSs as possible. For ACSM1 this was a KpnI site, and for ACSM3 this was a PstI site. The ACSM DNA sequences were kept otherwise identical to those which they were to modify. The ordered gBlock oligo is shown below (Figures 3.20 and 3.21).

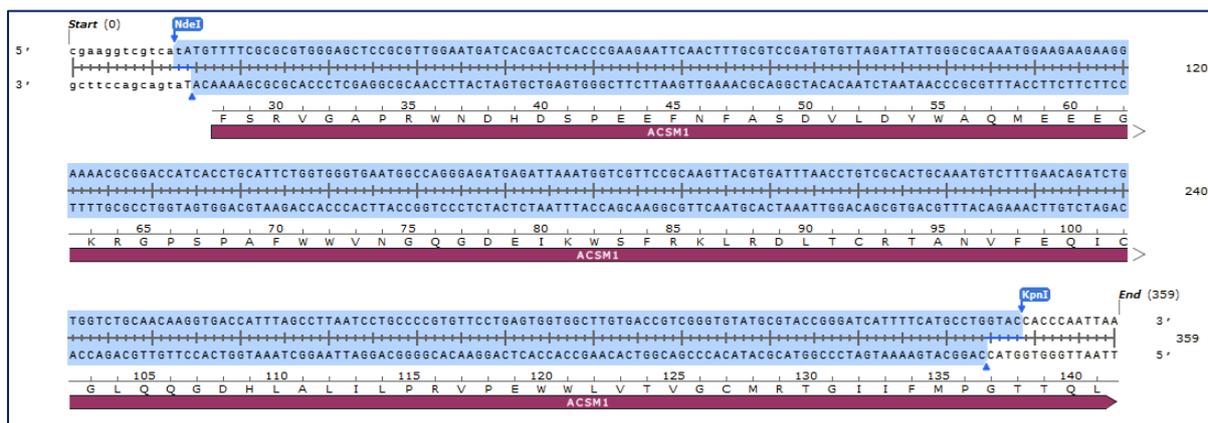


Figure 3.20. Sequence of the gBlock oligonucleotide for removal of the ACSM1 MTS. What is normally R30 in ACSM1 becomes R4 in MTS-free ACSM1.

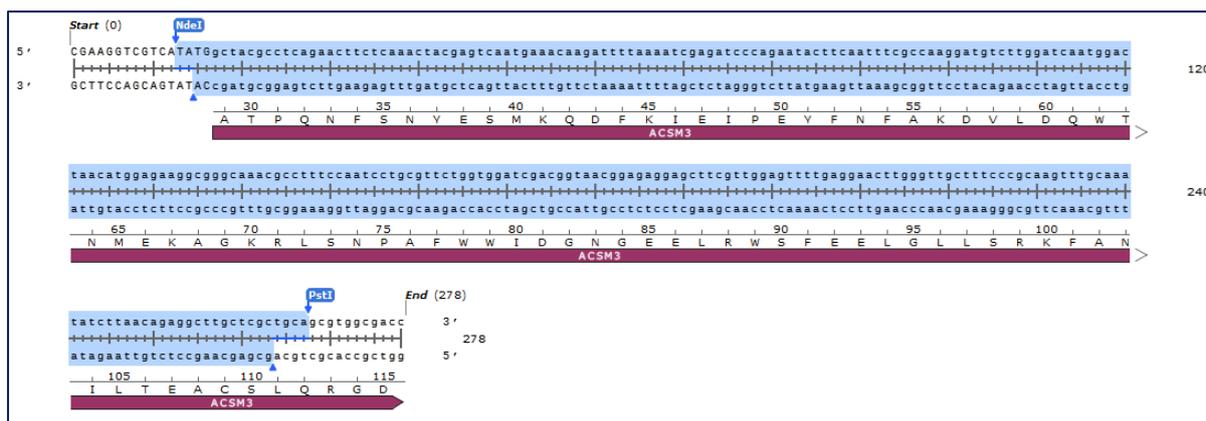


Figure 3.21. Sequence of the gBlock oligonucleotide for removal of the ACSM3 MTS. What is normally T30 in ACSM3 becomes T3 in MTS-free ACSM3.

3.3.3.2 Cloning

ACSM1 and ACSM3 in pET16b were digested with NdeI + KpnI and NdeI + PstI respectively and pre-digested MTS-removal gBlocks inserted via ligation (see 2.2.1.6). This was then transformed into DH5 α cells, DNA miniprep and retransformed into BL21 cells (see 2.2.3.3 and 2.2.3.1).

3.3.3.3 Screening

ACSM1 and ACSM3 without MTSs were expressed in LB broth followed by a small-scale Ni-NTA protein purification (see 2.2.3.7.1). As with the MTS-containing construct, ACSM3 was purified with *Ppy* Fluc in parallel for optimal fraction desalting via BLI probing of fractions. Purified and desalted fractions were run via SDS-PAGE and silver-stained, and bands of roughly the expected sizes (~64.0 kDa and ~64.9 kDa respectively) were very faintly visible. Due to the apparent low concentration, ACSM3 was 25 x concentrated (see 2.2.2.2) and run again via SDS-PAGE and silver-stained, bands were clearly visible (Figure 3.10). ACSM1 was also assessed in the same way (Figure 3.10). Due to silver staining the bands could not be quantified, but appeared by eye during staining to be very similar in intensity to the MTS-containing counterparts. Thus, as expected, it appears that the MTS has no effect on ACSM expression in *E. coli*.

3.3.4 Modelling and LCMS analysis of the adenylation of D-LH₂ by WT ACSM3

As stated above, ACSM3-transformed BL21 cells were found to be non-bioluminescent despite expressing low levels of the protein, and thus it can be assumed that ACSM3 fails to ultimately catalyse the bioluminescent oxidation reaction with D-LH₂. However, as the conversion from luciferin to oxyluciferin is enzymatically a two-step reaction with a luciferyl adenylate (D-LH₂-AMP) intermediate (table 3.3), it is not possible to infer which of the steps the bottleneck occurs.

Enzyme	Rx Step 1	Rx Step 2
Firefly luciferase	$\text{LH}_2 + \text{ATP} \rightleftharpoons \text{LH}_2\text{-AMP} + \text{PP}_i$	$\text{LH}_2\text{-AMP} + \text{O}_2 \rightleftharpoons \text{LO} + h\nu + \text{CO}_2 + \text{AMP}$
Acyl-CoA synthetase	$\text{FA} + \text{ATP} \rightleftharpoons \text{FA-AMP} + \text{PP}_i$	$\text{FA-AMP} + \text{CoA} \rightleftharpoons \text{FA-CoA} + \text{AMP}$

Table 3.3. Comparison of ACS and Fluc respective reactions. Duplicated from table 1.3 (Chapter 1).

The binding energy of D-LH₂ to *Ppy* Fluc is 7.5 kcal/mole – the benzothiazole ring portion contributes 6.0 kcal/mole and the thiazoline ring portion contributes the remaining 1.5 kcal/mole (Denburg et al. 1969). The binding of the substrate in this one specific position in the active site – energetically favoured in Flucs – is necessary for the initiation of adenylation; it is thus highly improbable for D-LH₂ to bind with similar energy to a protein without a compatible active site.

The human ACSM-threaded model of ACSM3 created earlier within this chapter was modelled in the program Autodock Vina according to a video tutorial which is too extensive to transpose here (<https://www.youtube.com/watch?v=k6tgCeDIwEk>). The entire ACSM3 model was sampled for docking sites, and the highest energy site found to be inside the active site with a binding energy of 7.3 kcal/mole (Figure 3.22), thus very close to that of *Ppy* Fluc.

From the modelling it thus appeared highly plausible that ACSM3 is capable of adenylation of D-LH₂. A liquid chromatography-mass spectroscopy (LCMS) assay was undertaken to detect the presence of D-LH₂-AMP post-reaction with D-LH₂, as it is an established technique (Figure 3.23). Purified and desalted *Ppy* Fluc, ACSM3 (25 fold concentrated), and BSA (negative control) were reacted with D-LH₂ (see 2.2.6.3); as a qualitative test, protein concentration was

not controlled for but proteins were added in equal volumes. BSA was added at a negative control at the same concentration as the most concentrated sample to be tested. The reactions were allowed to progress for one hour, then the protein removed via filtration through a ≤ 50 kDa centrifugal filter and the liquid placed on ice.

A LCMS researcher kindly offered to assist with this assay, and carried out the running of samples through the machine and performed the analysis. A D-LH_2 -AMP peak was found in small quantities in both *Ppy* Fluc and ACSM3, but not in BSA samples. An oxyluciferin peak was found in the *Ppy* Fluc sample as was expected, however not for ACSM3 confirming that no bioluminescence occurs. Large peaks of unreacted D-LH_2 were found in all samples as was expected. These data indicate that ACSM3 is capable of D-LH_2 adenylation but not oxidation. These results were found via a brief analysis by the LCMS researcher using the software on the LCMS-connected computer. The LCMS data was exported as raw files with the intention of further analysis (suitable for this thesis) at a later date. The processing of these data requires extremely expensive proprietary software and technical expertise. However, the LCMS researcher was not contactable near the time of writing, and as such no Figures or exact values have been produced from this experiment. The raw files are available at a link below (https://cf-my.sharepoint.com/:f:/g/personal/longja_cardiff_ac_uk/ErN1VW30-o1HgRBWCVBGOewBrCFWie0xigdulqJSjTvtgA) using the password “TheGoldStandard”.

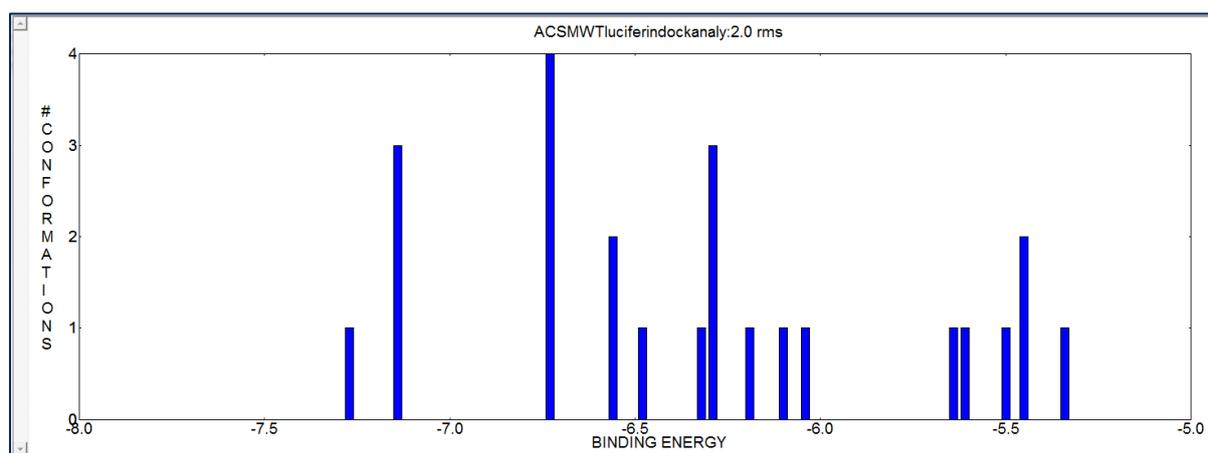


Figure 3.22. Computer model-derived binding energies of luciferin to ACSM3. The “conformations” (y-axis) indicates the number of docking positions with similar energy.

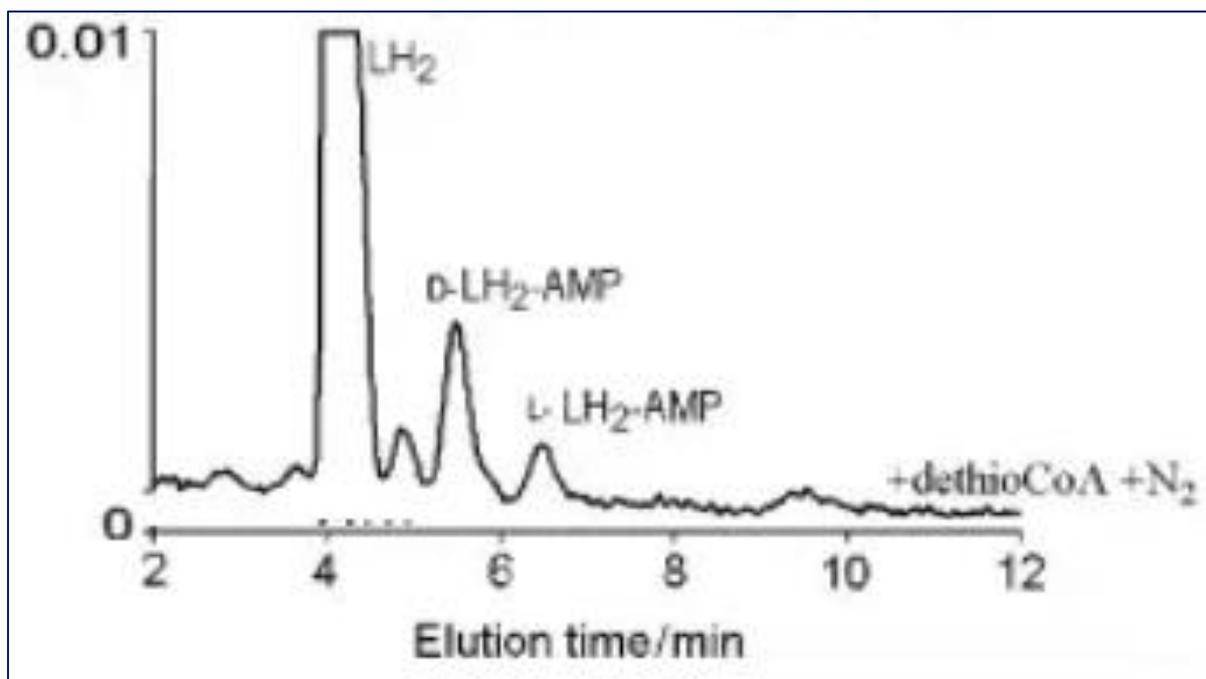


Figure 3.23 adapted from (Fraga et al. 2004). **Identification of luciferyl adenylate in HPLC portion of LCMS.** The D-LH₂-AMP peak occurs with an absorbance of 336 nm at approximately 5m30s over an elution time of 12 mins.

3.3.5 Random Mutagenesis of ACSM3

As ACSM3 appeared to be capable of adenylation, it was presumed to be close (in terms of fitness landscape) to having bioluminescence function. Thus, random mutagenesis directed evolution may produce a bioluminescent mutant. The chosen method of creating the random mutagenic libraries was error-prone PCR (epPCR); this is PCR in which the error-rate of the polymerase is increased. The error-rate of *Taq* polymerase is normally 1 error per 1×10^4 bp amplified. This can be increased with the addition of manganese(II) chloride (MnCl_2) to the reaction mixture (Lin-Goerke et al. 1997). This method has long been used to create mutants in Fluc engineering. It was previously found that 0.4 mM MnCl_2 was optimal (Jathoul, A., PhD thesis, University of Cambridge [not publically available]) as it maintained a useable transformation efficiency of 44.6% (relative to 0.0 mM MnCl_2) while allowing a high phenotypic index of 27.3%, with 0.0 mM giving 7.3%. The phenotypic index is a measure of several qualities of the transformed colonies, namely: colour, brightness, and size. 0.8 mM MnCl_2 gave a similar phenotypic index of 27.2% however with a much reduced transformation efficiency of 29.7% relative to 0.0 mM MnCl_2 .

3.3.5.1 Cloning of random mutagenic libraries

WT *Ppy* Fluc was used as a positive control for the epPCR technique, as successful mutagenic libraries will display a variety of coloured colonies when screened for bioluminescent activity. 0.4 mM MnCl₂ was used in the PCR mixture. The reagent quantities used to perform the epPCR are detailed in (see 2.2.1.12). Both *Ppy* Fluc and ACSM3 were subjected to epPCR.

A small fraction of both PCR products was run on an agarose gel (see 2.2.1.4). A bright band was observed at ~1.7 kb indicating successful amplification. Thus the PCR products were cleaned of primers and then terminally restriction digested with NdeI and BamHI (see 2.2.1.6). The digested PCR products were again cleaned and T4 ligated into pET16b empty vector (see 2.2.1.5 and 2.2.1.7).

3.3.5.2 Screening of the random mutagenic libraries

For monoclonal constructs the normal protocol is to transform into DH5α cells to maximise transformation efficiency and ensure that the correct construct can be isolated. Once isolated, grown, and DNA miniprep the construct is retransformed into low-efficiency BL21 cells for protein expression and screening. However, libraries should ideally be transformed directly into BL21 cells to maintain maximal diversity, but this has the trade-off of lower transformation efficiency and thus fewer colonies. This inefficiency can be compensated for by simply increasing the volume of cells and amount of DNA.

The ligation mixtures were transformed into BL21 cells and spread onto five agar plates. The following day the colonies were lifted and induced with IPTG (see 2.2.3.6.1). The colonies were screened for bioluminescence with D-LH₂ under both a DSLR briefly then inside the PhotonIMAGER precisely one min later (see 2.2.6.4.1).

The *Ppy* Fluc library displayed clear evidence of mutagenesis as many colonies were of different colours ranging from green-yellow to red, and many appeared non-bioluminescent (Figure 3.24). The ACSM3 library did not display any bioluminescent activity (data not shown, see Figure 9.1).

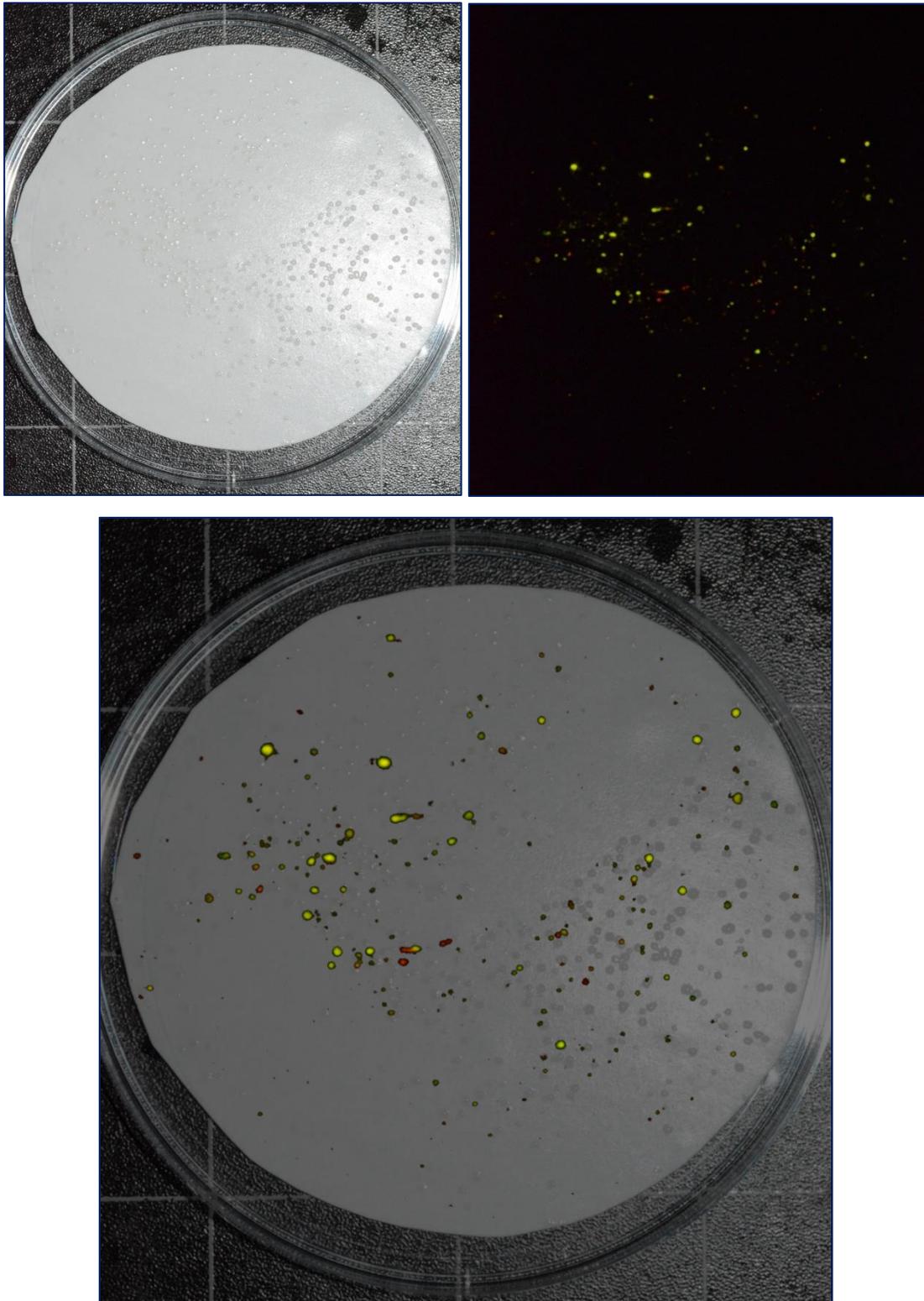


Figure 3.24. DSLR photographs of representative error-prone PCR *Ppy Fluc* colonies, lifted from a petri dish. Top left: flash-illuminated image of colonies. Top right: long-exposure image of same colonies immediately following spraying with $D-LH_2$. Bottom centre: superimposed image (of top left and top right images). Evidence of mutagenesis is present in the form of colour diversity ranging from native yellow-green, to red.

3.3.6 *Ppy* Fluc active site consensus sequence substitution

As previously discussed in Chapter 1, ACSMs (both murine and human) are known to be functionally similar to beetle luciferases and the structure of human ACSs are highly homologous to Flucs (figure 3.16). Thus, there are equivalent regions between these two enzymes, which can be exchanged through engineering.

3.3.6.1 Design

Based on the 2D protein sequence alignment produced by TM-align (figure 3.17), the region G369 – L377 of ACSM3 was to be converted to GYGLTETTSA. The DNA sequence for the mutated region was codon-optimised for *E. coli* using IDT's codon optimisation tool. The design is shown below (figure 3.25).

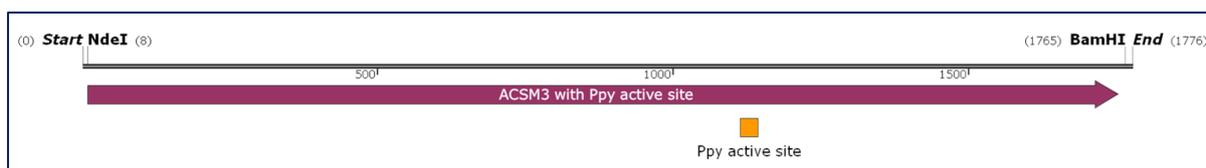


Figure 3.25. Map of ACSM3 backbone with the active site consensus sequence of *Ppy* Fluc substituted. An expanded version of this figure – with DNA and protein sequence shown – can be found in the appendix (figure 3.25EX).

3.3.6.2 Cloning

Splicing by overlap extension (SOE) (also known as overlap extension polymerase chain reaction) is a technique used to introduce specific mutations at specific points in a DNA sequence (Heckman and Pease 2007). It requires two PCRs to be performed: the first introduces the desired mutation(s) but creates at least two fragments with homologous overhangs, and the second PCR splices these fragments together. The spliced DNA can then be inserted into a plasmid for expression.

Primers were designed to substitute the *Ppy* Fluc active site consensus sequence into ACSM3 using the splicing by overlap extension method (SOE) (Figure 3.26). DNA of ACSM3 in pET16b was prepared and PCR amplified using the SOE primers to create the two overlapping fragments (see 2.2.1.9). The PCR products were checked via agarose gel electrophoresis (see

2.2.1.4) and bright bands of ~1200 bp and ~500 bp were observed, indicating successful amplification, they were then PCR-cleaned (see 2.2.1.5). A second PCR was then performed to splice the two SOE fragments using the terminal primers (see 2.2.1.11). The final PCR product was checked via agarose gel electrophoresis (see 2.2.1.4) and a bright band of ~1.7 kb, indicating successful amplification was observed, it was then PCR-cleaned (see 2.2.1.5). The DNA was then terminally digested with NdeI and BamHI and ligated into pre-digested pET16b (see 2.2.1.6 and 2.2.1.7). The ligation mixture was then transformed into DH5 α cells (see 2.2.3.3), which were picked, partially DNA miniprep, and the highest-concentration sequence-correct culture stored at -80 °C as glycerol stock.

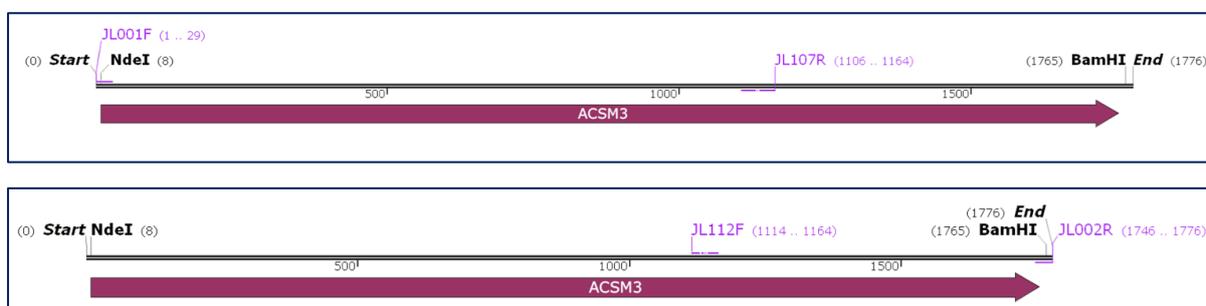


Figure 3.26. Upstream and downstream PCR of ACSM3 backbone to introduce active site consensus sequence of *Ppy Fluc*, with primers shown. The upper sub-figure is the upstream PCR, the lower sub-figure is the downstream PCR. An expanded version of this Figure – with DNA and protein sequences shown – can be found in the appendix (Figures 3.26EXa and 3.26EXb).

3.3.6.3 Screening

The miniprep harvested from the DH5 α cells was transformed into BL21 cells (see 2.2.3.3) and spread onto agar plates. The following day the colonies were lifted and induced with IPTG (see 2.2.3.6.1). The colonies were screened for bioluminescence with D-LH₂ (see 2.2.6.4.1) using the PhotonIMAGER, however no bioluminescence activity was detected (data not shown, see Figure 9.1).

3.3.7 Substrate Determining Residues (SDRs)

3.3.7.1 Design

The primary of two SDR papers (Khurana et al. 2010) uses the consensus sequence for each enzyme family, therefore the listed ACSM SDRs will not necessarily be the same as in ACSM3. The SDRs in ACSM3 were to be found by the use of TM-align (as was used for the active site consensus sequence substitution). First, it was tested whether the results given by TM-align would match those given by the paper. Therefore, two proteins for which the SDRs were already known – 1AMU and *Ppy* Fluc – were aligned and the SDRs compared (Figures 3.27 and 3.28). The results matched the alignment of the primary paper (Khurana et al. 2010), thus the alignment of *Ppy* Fluc and ACSM3 (previously performed for the active site consensus sequence substitution work in this chapter) was used with confidence to identify the primary SDRs of ACSM3. It was also used to find the SDRs from the secondary paper (Oba et al. 2009), many of which overlap with the SDRs from the primary paper (Khurana et al. 2010). Thus, the mutations to be made to ACSM3 were: G249R, W274H, A275G, K276F, W279F, S281T, C318L, V344A, and A346G. The position of these mutations within ACSM3 (with the active site consensus sequence of *Ppy* Fluc) can be visualised below (Figure 3.29).

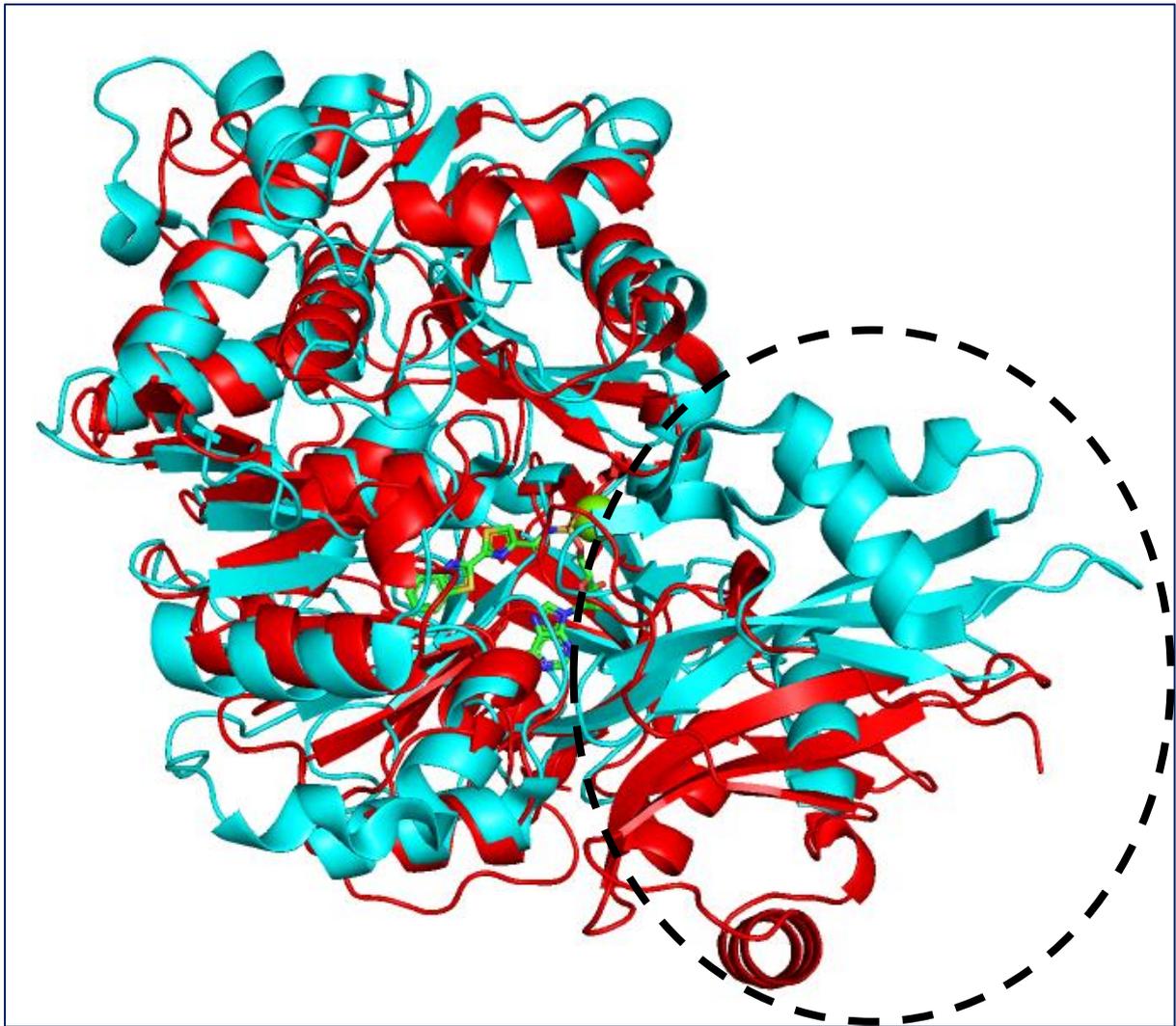


Figure 3.27. Structural alignment of *Ppy* Fluc and 1AMU. *Ppy* Fluc (4g37.pdb) in blue, 1AMU (1amu.pdb) in red. CTDs crudely circled. Luciferyl adenylate (carbons in green) bound to active site of *Ppy* Fluc. The TM-score for this alignment was 0.85768, indicating high structural homology.

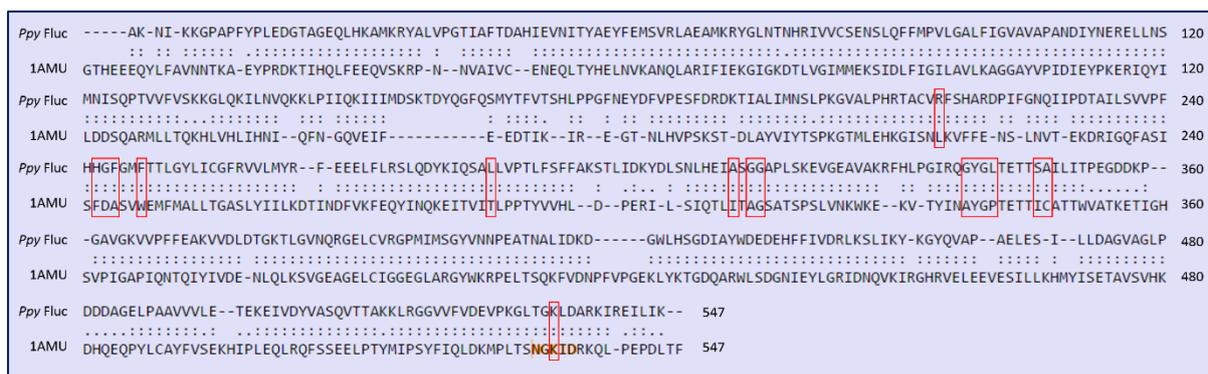


Figure 3.28. Identification of the primary (Khurana et al. 2010) substrate-determining residues of *Ppy Fluc* via 3D protein sequence alignment to *1AMU*. The SDRs from the primary SDR paper (Khurana et al. 2010) are outlined in red. [The slight orange outline around certain *1AMU* residues is an image artefact and has no meaning].

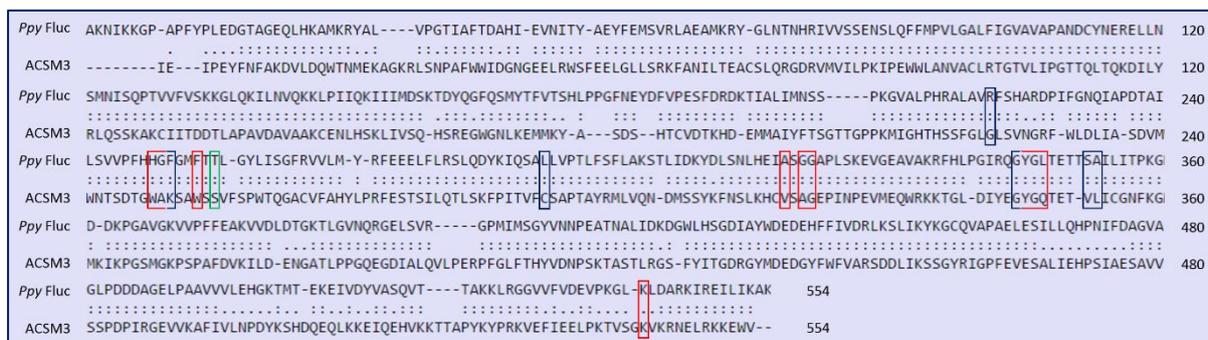


Figure 3.29. Identification of the primary (Khurana et al. 2010) and secondary (Oba et al. 2009) substrate-determining residues of *ACSM3* (with *Ppy Fluc* active site) via 3D protein sequence alignment to *Ppy Fluc*. The primary SDRs from are outlined in red, the secondary SDRs are outlined in green, and residue positions which are from both papers are outlined in navy.

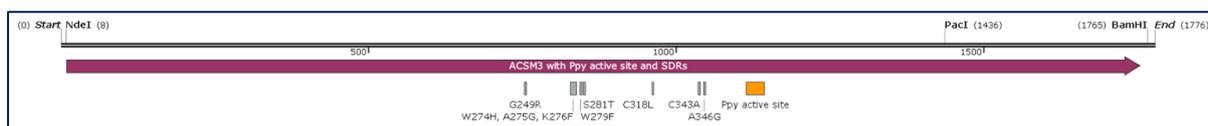


Figure 3.30. Map of *ACSM3* backbone with *Ppy Fluc*: active site consensus sequence substitution and all SDRs. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 3.30EX).

3.3.7.2 Cloning

As with the active site consensus sequence substitution work, primers were designed to substitute the *Ppy* Fluc active site consensus sequence into ACSM3 using SOE, shown below (figure 3.31). The lengths of the fragments produced by each primer pair is also shown below (table 3.4).

The PCR products were checked via agarose gel electrophoresis (see 2.2.1.4) and bright bands of approximately the expected sizes (table 3.4) were observed, indicating successful amplification. The DNA was thus PCR-cleaned (see 2.2.1.5). A second terminal PCR was then performed to splice these five SOE fragments into one (see 2.2.1.9). This final PCR product was checked via agarose gel electrophoresis (see 2.2.1.4) and PCR-cleaned (see 2.2.1.5). It was then terminally digested with NdeI and BamHI (see 2.2.1.6) and ligated into pre-digested pET16b (see 2.2.1.7). The ligation mixture was then transformed into DH5 α cells (see 2.2.3.3), which were picked, partially DNA miniprep, and the highest-concentration sequence-correct culture stored at -80°C as glycerol stock (see 2.2.3.5).

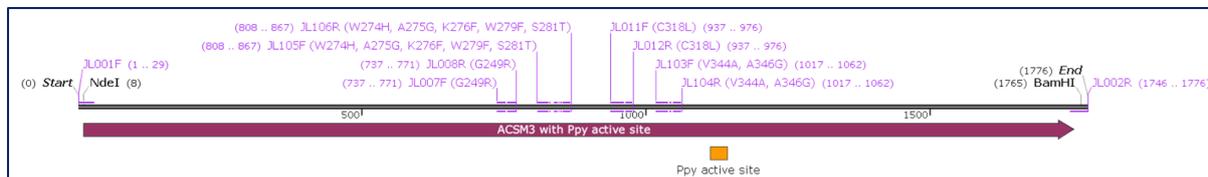


Figure 3.31. SOE PCRs of ACSM3 (with *Ppy* Fluc active site) to introduce all *Ppy* Fluc SDRs, with primer annealing positions shown. Note that these five PCRs are performed separately, the primer pairs and the length of the resulting PCR product is shown in table 3.4 below. An expanded version of this Figure – with DNA and protein sequences shown – can be found in the appendix (Figure 3.31EX).

Forward primer – reverse primer	Length (bp)
JL001F – JL008R	771
JL007F – JL106R	131
JL105F – JL012R	169
JL011F – JL104R	126
JL103F – JL002R	760

Table 3.4. Primer pairs to introduce all *Ppy* Fluc SDRs to ACSM (with *Ppy* Fluc active site), with length of resulting PCR products shown. Primer annealing positions can be found above in Figure 3.31.

3.3.7.3 Screening

The DNA miniprep harvested from the DH5 α cells was transformed into BL21 cells (see 2.2.3.3) and spread onto agar plates. The following day the colonies were lifted and induced with IPTG (see 2.2.3.6.1). The colonies were screened for bioluminescence with D -LH₂ using the PhotonIMAGER (see 2.2.6.4.1). No bioluminescent activity was detected (data not shown, see Figure 9.1).

3.3.8 *Ppy* Fluc R223 – T235 Loop

It was hypothesised that the *Ppy* Fluc sequence substitution for this region could bring bioluminescence function to both ACSM3 mutants created up until this point – 1. ACSM3 with *Ppy* Fluc active site consensus sequence, 2. ACSM3 with both the *Ppy* Fluc active site consensus sequence and all SDRs.

3.3.8.1 Modelling of structural differences between the R223 loop in *Ppy* Fluc and equivalent loop in ACSM3

The human ACSM-threaded model of ACSM3 (created previously within this chapter) was structurally aligned with the D -LH₂-AMP-bound oxidising conformation of *Ppy* Fluc (4g37.pdb) using the program PyMOL (Schrödinger). The alignment allowed the positioning of D -LH₂-AMP within the active site of ACSM3. The spatial relationship between the R223-equivalent loop of ACSM3 and D -LH₂-AMP could then be examined and compared to that of *Ppy* Fluc.

Overall the structures between the R223-equivalent loop of ACSM3 and that of *Ppy* Fluc were found to be similar, however the ACSM3 loop has a much sharper turn at the point of contact

with D-LH₂-AMP (Figure 3.32). This sharper turn is at least partly required since the ACSM3 loop is two residues shorter than that of *Ppy* Fluc. Loops in *Ppy* Fluc (as in other enzymes) are more tolerant to additions than deletions (Halliwell et al. 2018) and therefore this could explain the lack of bioluminescence in the mutants created of ACSM3 thus far. W257 of the R223-equivalent loop of ACSM3 comes into closer contact with D-LH₂-AMP than the spatially equivalent G228 of *Ppy* Fluc, but does not make steric clash with the substrate. However, the replacement of the smallest possible residue glycine (G228) with the largest possible tryptophan (W257), is considered a major change. This may be due to the preference of ACSM3 for medium-chain fatty acids in the shorter range (C4) (Fujino et al. 2001), and thus the R223-equivalent loop must reach deeper into the active site to make contact. Thus, it may be the case that W257 prevents larger substrates such as D-LH₂-AMP from positioning correctly.

The residue G228 is invariant in beetle luciferases, indicating a role of critical importance. Viviani *et al.* found that changes to it had a dramatic effect on spectral emission. For instance, replacement with an alanine (a medium-sized nonpolar side chain) caused dramatic red-shift. Since ACSM3 W257 is spatially equivalent to *Ppy* Fluc G228, it may be the case that replacement with such a large side chain disrupts the active site to the point of disabling oxygenase function.

The solvent shielding provided by the R223 equivalent loop and that of *Ppy* Fluc were also examined (Figures 3.33 and 3.34). The ACSM3 R223-equivalent loop appears to provide marginally greater solvent shielding than that of *Ppy* Fluc. The reactions catalysed by both enzymes are hydrophobic and require exclusion of water, thus this is expected.

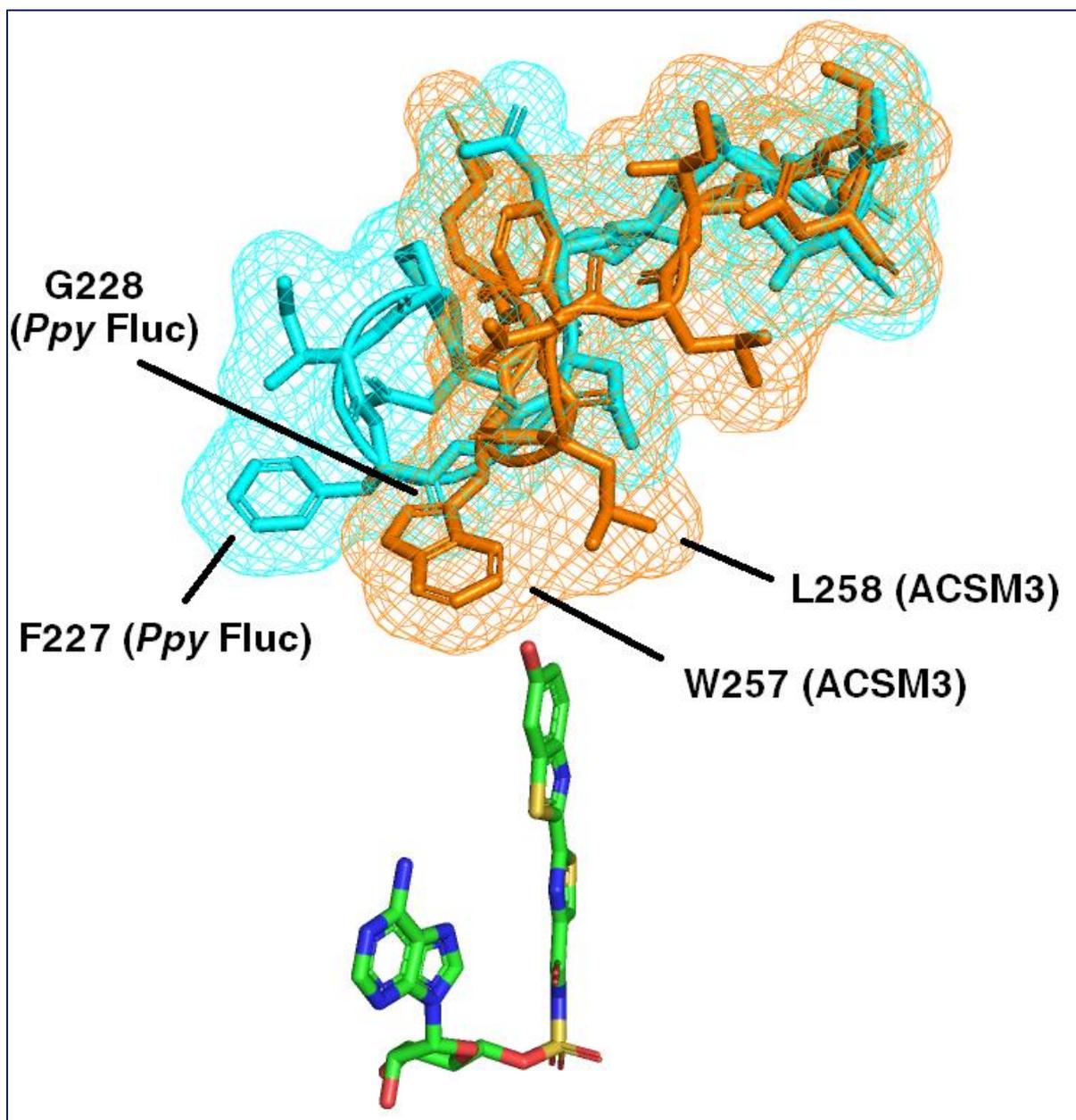


Figure 3.32. Spatial relationship between luciferyl adenylate and the R223 (and equivalent) loops of *Ppy Fluc* (oxidising conformation) and ACSM3 threaded model. The loops are displayed as mesh, stick, and ribbon; with *Ppy Fluc* in cyan and ACSM3 in orange. The *Ppy Fluc* equivalent of ACSM3 L258 is a glycine (G228), and is barely visible. Luciferyl adenylate is displayed as stick.

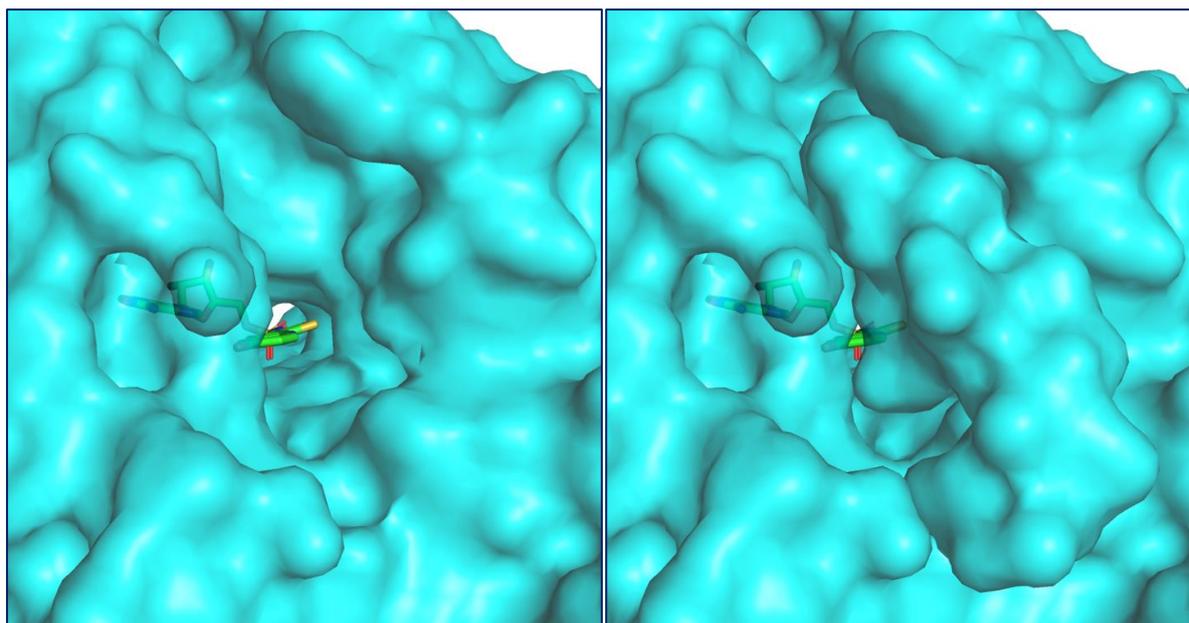


Figure 3.33. Solvent shielding provided by the R223 loop of the oxidising conformation of *Ppy Fluc* for luciferyl adenylate. With the R223 loop removed (left) and with the R223 loop present (right). Luciferyl adenylate is visible (carbons in green).

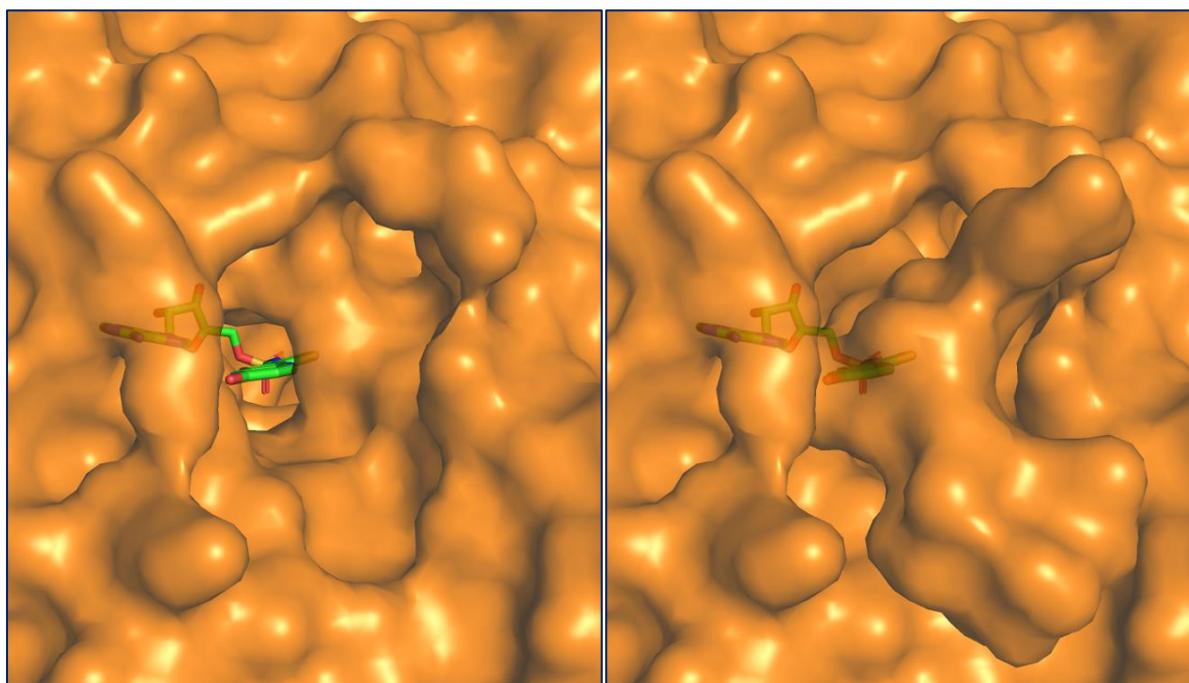


Figure 3.34. Solvent shielding provided by the R223-equivalent loop of ACSM3 threaded model for luciferyl adenylate. With the R223-equivalent loop removed (left) and with the R223 loop present (right). Luciferyl adenylate is visible (carbons in green).

3.3.8.2 ACSM3 with the active site consensus sequence and R223 loop of *Ppy* Fluc substituted3.3.8.2.1 Design

For consistency with other prior work in this chapter, the TM-align-derived 3D alignment of ACSM3 with *Ppy* Fluc was again used to confirm the equivalent residues in ACSM3 for the *Ppy* Fluc R223 loop (Figure 3.35).

It was found (as it was using PyMOL) that the *Ppy* Fluc R223 loop is R255 – D264 in ACSM3, notably this region is truncated by two residues. Based on the 2D alignment produced by TM-align, the region G254 – D264 of ACSM3 was to be converted to R223 – T235 of *Ppy* Fluc. The design is shown below (Figure 3.36).

<i>Ppy</i> Fluc	AKNIKKGP-APFYPLEDGTAGEQLHKAMKRYAL---VPGTIAFTDAHI-EWNITY-AEYFEMSVRLAEAMKRY-GLNTHNRIVVSSSENSLQFFMPVLGALFIGVAVAPANDCYNERELLN	120
ACSM3	-----IE---IPEYFNFAKDVLQWNTNMEKAGKRLSNPAFWIDGNGEELRWSFEELGLLSRKFANILTEACSLQRGDRVMVILPKIPEWLANVACLRGTGTVLIPGTTQLTQKDILY	120
<i>Ppy</i> Fluc	SMNISQPTVVFVSKKGLQKILNVQKLPPIIQKIIIMDSKTDYQGFQSMYFVTSHLPPGFNEYDFVPESFDRDKTIALIMNSS----PKGVALPHRALAVRFSHA	240
ACSM3	RLQSSKAKCIITDDTLAPAVDAVAAKCENLHSLIVSQ-HSREGWNLKEMMKY-A---SDS--HTCVDTKHD-EMMAIYFTSGTTGPPKMIGHTHSFGGLSVN	240
<i>Ppy</i> Fluc	LSVVPFHGHGFMFTTL-GYLISGFRVVLH-Y-RFEEELFLRSLQDYKIQSALLVPTLFSFLAKSTLIDKYDLNHLHEIASGGAPLSKEVGEAVAKRFHLPGRQGYGLTETTSAILITPKG	360
ACSM3	WNTSDTGWAKSAWSSVFSWQACVFAHYLPRFESTSILQTLKFPITVFCAPTAYRMLVQN-DMSSYKFNSLKHCVSAGEPINPEVMEQWRKKTGL-DIYEGYQTET-VLICGNFKG	360
<i>Ppy</i> Fluc	D-DKPGAVGKVPFFFEAKVVDLDTGKTLGVNQRGELSVR----GPMIMSGYVNNPEATNALIDKDGWLHSGDIAYNDEDEHFFIVDRLKSLIKYKGCQVAPAELESILLQHPNIFDAGVA	480
ACSM3	MKIKPGSMGKPSPAFDVKILD-ENGATLPPQEGDIALQVLPERPFLFTHYVONPSKTASTLRGS-FYITGDRGYMDEGDFWFVARSDDLKSSGYRIGPFVEVESALIEHPSIAESAVV	480
<i>Ppy</i> Fluc	GLPDDDAGELPAAVVLEHGKMT-EKEIVDYVASQVT---TAKKLRGGVVFVDEVPKGL-KLDARKIREILIKAK	554
ACSM3	SSPDPIRGEVVKAFIVLNPDKSHDQEQLKKEIQEHVKKTTAPYKYPRKVEFIEELPKTVSGVKRNELRKKKVV--	554

Figure 3.35. ACSM3 R223-equivalent loop identification via protein sequence alignment derived from 3D alignment of ACSM3 and *Ppy* Fluc. The *Ppy* Fluc R223 loop and its equivalent position in ACSM3 is boxed in red. For *Ppy* Fluc the residue sequence M1 – D3 is omitted from the N-terminus and K544 – L550 from the C-terminus; for ACSM3 the residue sequence M3 – K45 is omitted from the N-terminus and T579 – T580 from the C-terminus.

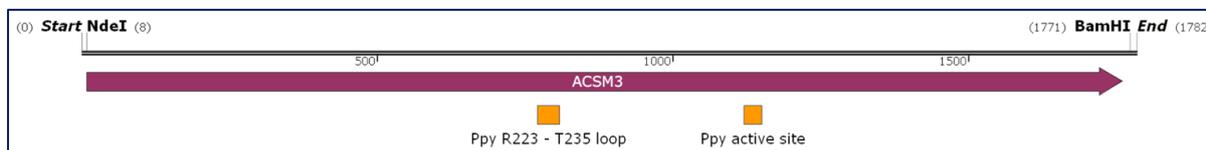


Figure 3.36. Map of ACSM3 backbone with *Ppy* Fluc: active site consensus sequence substitution and R223 loop. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 3.36EX).

3.3.8.2.2 Cloning

Primers were designed to substitute the R223 – T235 loop of *Ppy* Fluc into ACSM3 using the SOE method (Figure 3.37).

Miniprep DNA of ACSM3 in pET16b was PCR amplified using the SOE primers to create the two overlapping fragments. The PCR products were checked via agarose gel electrophoresis (see 2.2.1.4) and bright bands both of ~800 bp were observed, indicating successful amplification. The DNA was thus PCR-cleaned (see 2.2.1.5). A second PCR was then performed to splice the two SOE fragments (see 2.2.1.9). The final PCR product was checked via agarose gel electrophoresis (see 2.2.1.4) and a bright band of ~1.7 kb was observed, indicating successful amplification. The DNA was then PCR-cleaned (see 2.2.1.5). It was then terminally digested with NdeI and BamHI (see 2.2.1.6) and ligated into pre-digested pET16b (see 2.2.1.7). The ligation mixture was then transformed into DH5 α cells, which were picked, partially DNA miniprep, and the highest-concentration sequence-correct culture stored at -80 °C as glycerol stock (see 2.2.3.5).

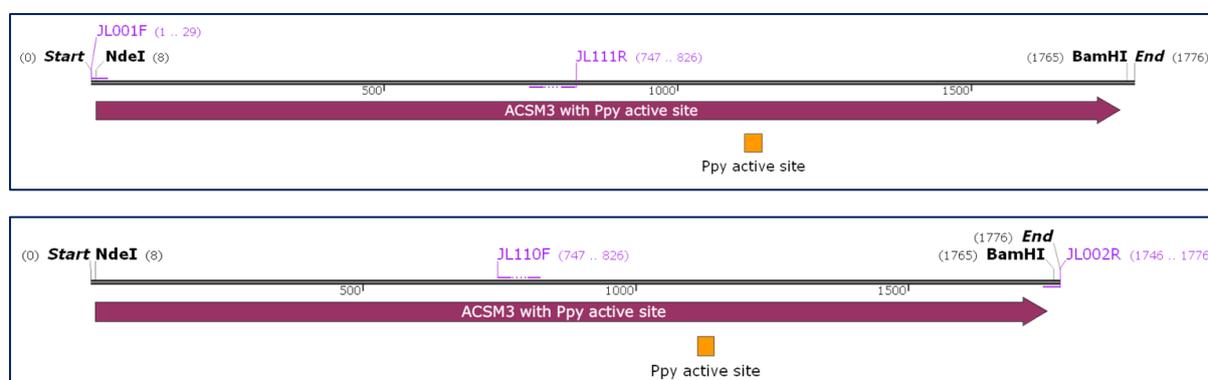


Figure 3.37. Upstream and downstream SOE PCR of ACSM3 (with *Ppy* Fluc active site) backbone to introduce R223 loop of *Ppy* Fluc, with primers shown. The upper sub-figure is the upstream PCR, the lower sub-figure is the downstream PCR. An expanded version of this Figure – with DNA and protein sequences shown – can be found in the appendix (Figures 3.37EXa and 3.37EXb).

3.3.8.2.3 Screening

The DNA miniprep harvested from the DH5 α cells was transformed into BL21 cells (see 2.2.3.3) and spread onto agar plates. The following day the colonies were lifted and induced

with IPTG (see 2.2.3.6.1). The colonies were screened for bioluminescence with α -LH₂ using the PhotonIMAGER (see 2.2.6.4.1). No bioluminescent activity was detected (data not shown, see Figure 9.1).

3.3.8.3 ACSM3 with the active site consensus sequence, SDRs, and R223 loop of *Ppy* Fluc substituted

3.3.8.3.1 Design

The ACSM3-equivalent position of the *Ppy* Fluc R223 loop was again used to design a construct of ACSM3 with the *Ppy* Fluc: active site consensus sequence, SDRs, and R223 loop. It is shown below (Figure 3.38).



Figure 3.38. Map of ACSM3 backbone with *Ppy* Fluc: active site consensus sequence substitution, all SDRs, and R223 loop. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 3.38EX).

3.3.8.3.2 Cloning

Primers were designed to substitute the R223 – T235 loop of *Ppy* Fluc into ACSM3 using the SOE method (Figures 3.41).

Miniprep DNA of ACSM3 in pET16b was PCR amplified using the SOE primers to create the two overlapping fragments. The PCR products were checked via agarose gel electrophoresis (see 2.2.1.4) and bright bands were observed at ~800 bp for both, indicating successful amplification. The DNA was then PCR-cleaned (see 2.2.1.5). A second PCR was then performed to splice the two SOE fragments (see 2.2.1.9). The final PCR product was checked via agarose gel electrophoresis (see 2.2.1.4) and a bright band of ~1.7 kb was observed, indicating successful amplification. The DNA was thus PCR-cleaned (see 2.2.1.5). It was then terminally digested with NdeI and BamHI (see 2.2.1.6) and ligated into pre-digested pET16b (see 2.2.1.7), the ligation mixture was then transformed into DH5 α cells (see 2.2.3.3), which were picked, partially DNA miniprep, and the highest-concentration sequence-correct culture stored at -80 °C as glycerol stock (see 2.2.3.5).

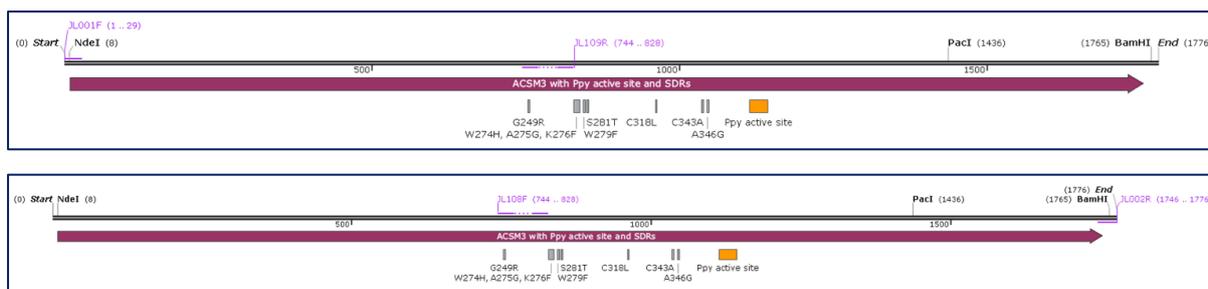


Figure 3.39. Upstream and downstream SOE PCR of ACSM3 (with *Ppy* Fluc active site and all SDRs) backbone to introduce R223 loop of *Ppy* Fluc, with primers shown. The upper sub-figure is the upstream PCR, the lower sub-figure is the downstream PCR. An expanded version of this Figure – with DNA and protein sequences shown – can be found in the appendix (Figures 3.39EXa and 3.39EXb).

3.3.8.3.3 Screening

The DNA miniprep harvested from the DH5 α cells was transformed into BL21 cells (see 2.2.3.3) and spread onto agar plates. The following day the colonies were lifted and induced with IPTG (see 2.2.3.6.1). The colonies were screened for bioluminescence with D-LH₂ using the PhotonIMAGER (see 2.2.6.4.1). No bioluminescent activity was detected (data not shown, see Figure 9.1).

3.3.9 Further protein modelling

The ACSM3 model created earlier within this chapter was created by threading ACSM3 onto a human ACSM whose structure was known. This process exchanges the side chains of the human ACSM for that of ACSM3, but the backbone conformation is virtually unchanged. To more accurately predict the structure, the protein must be molecularly simulated and a period of time elapsed to allow, for instance, regions of induced tension to relax.

As was found in the D-LH₂ LCMS work earlier in this chapter, WT ACSM3 does appear to adenylate D-LH₂ and thus it is the binding of D-LH₂-AMP to an accurately simulated ACSM3 model which is of interest.

GROMACS – GRONingen MACHine for Chemical Simulations (originally developed by Groningen University) – is a molecular dynamics package which can be used to perform

structure relaxation and energy minimisation of macromolecules including proteins. The simulations and relaxation/energy-minimisation of protein models using this program were performed externally on a powerful purpose-built computer by a researcher highly familiar with the software. The ACSM3 model previously created was supplied to the external researcher as a .pdb with D-LH₂-AMP optimally bound via Autodock, and the relaxed/energy-minimised ACSM3 models returned as .pdb files which could be opened with PyMOL or VMD programs for visualisation.

The binding pocket of both enzymes around D-LH₂-AMP was visualised as surface view and examined for areas of steric clash. In ACSM3, D-LH₂-AMP was found to protrude through the region of the binding pocket created by the residues K476 and R481 (Figure 3.40). In *Ppy* Fluc, D-LH₂-AMP – as expected – does not clash with any region of the binding pocket, which closely conforms to the molecule. The equivalent residues of ACSM3 K476 and R481 in *Ppy* Fluc are K443 and Q448 (Figure 3.41).

It should be noted that the residue Y447 (adjacent to Q448) was used to cross-link *Ppy* Fluc and hold it in the oxidising conformation for the purpose of X-ray crystallography, which is how the *Ppy* Fluc used here was derived. As such the natural oxidising conformation may be slightly different. Thus, the confidence of the position of K443 is uncertain.

The ACSM3 model created appears to closely resemble the *Ppy* Fluc oxidising conformation, based on the position of these two residues (ACSM K476 and R481; *Ppy* Fluc K443 and Q448), which – due to the 140° linker rotation – are radically reoriented in the adenylation conformation (Figure 3.42). Thus it appears that, in ACSM3, as these clashing residues do not form part of the active site in the adenylation conformation they permit the adenylation of D-LH₂ to occur. However, they disrupt the active site in the oxidising conformation and prevent oxidation of D-LH₂-AMP occurring. Further, these residues occur near the boundary of the D-LH₂ and AMP portions of D-LH₂-AMP, which must cleave during oxidation. These modelling observations fortify the findings of the LCMS work performed earlier within this chapter – that ACSM3 adenylates, but does not oxidise, D-LH₂.

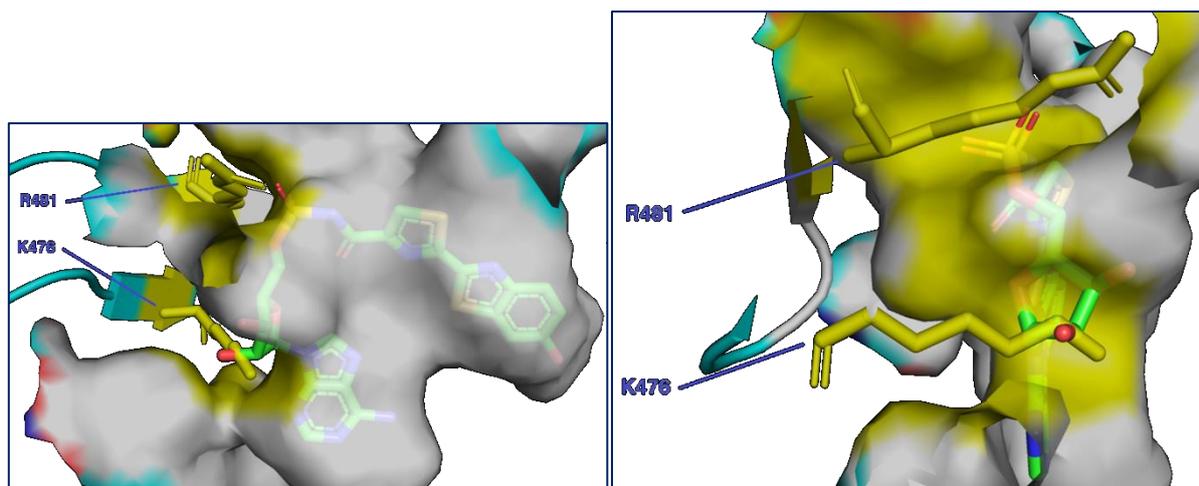


Figure 3.40. Influence of K476 and R481 residues (equivalent to K443 and Q448 in *Ppy Fluc*) on the “luciferyl adenylate binding pocket” of ACSM3. Two viewing angles are shown in the left and right Figures. The K476 and R481 residues are labelled and shown as sticks in yellow. The binding pocket is shown as surface view and coloured yellow where influenced by K476 and/or R481. Luciferyl adenylate is positioned via Autodock and shown as sticks, with carbons in green. Note that there is steric clash between part of the adenylate portion of luciferyl adenylate, and both K476 and R481.

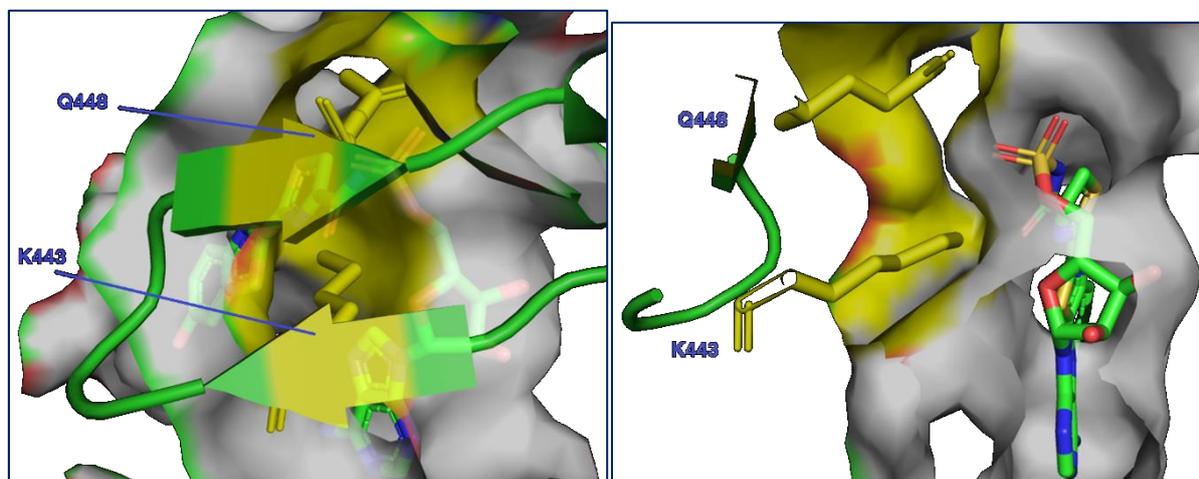


Figure 3.41. Influence of K443 and Q448 residues on the luciferyl adenylate binding pocket of *Ppy Fluc*. Two viewing angles are shown in the left and right Figures. *Ppy Fluc* is shown in the oxidising conformation. The K443 and Q448 residues are labelled and shown as sticks in yellow. The binding pocket is shown as surface view and coloured yellow where influenced by K443 and/or Q448. Luciferyl adenylate is positioned via Autodock and shown as sticks, with carbons in green.

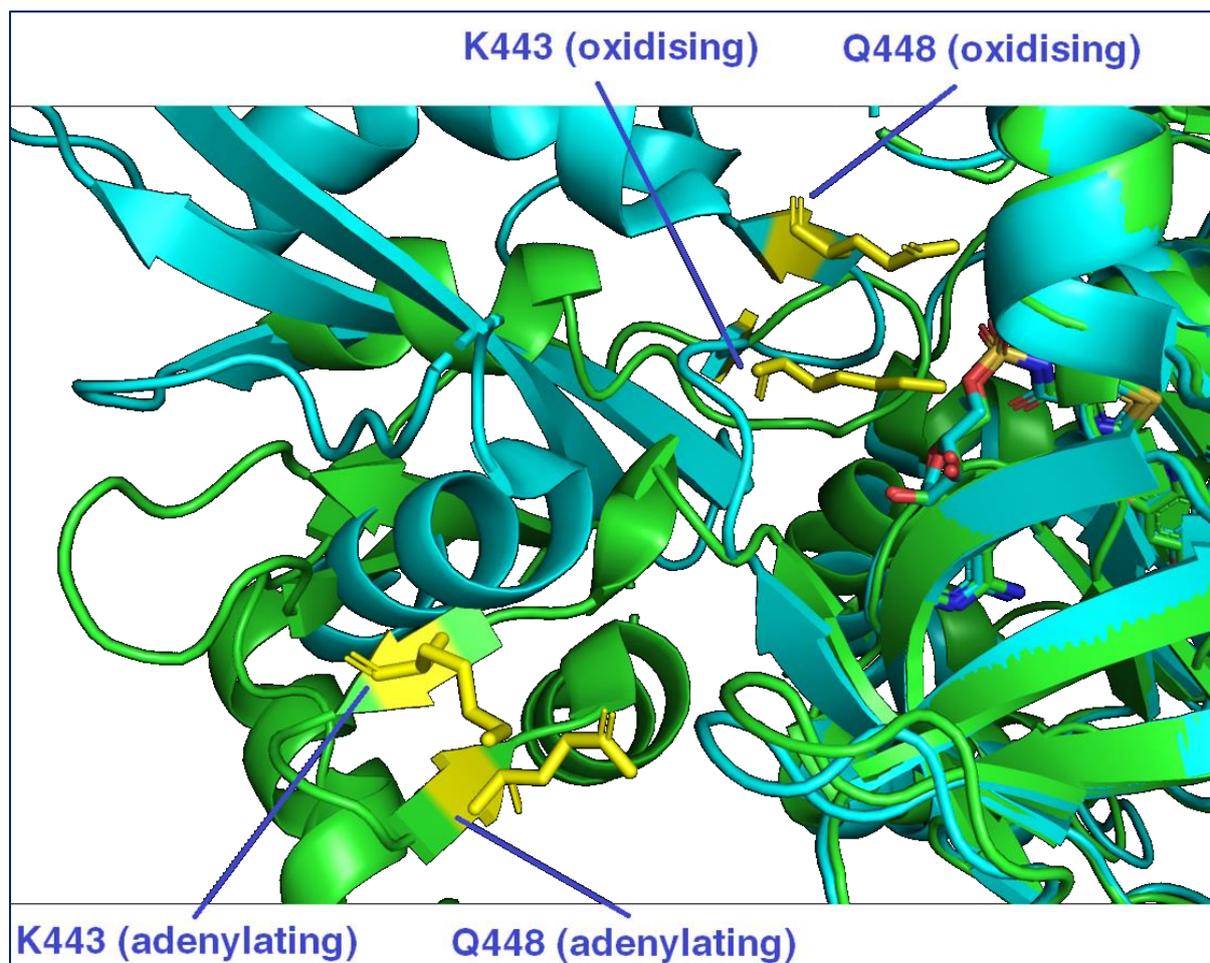


Figure 3.42. Alternating positions of K443 and Q448 residues between oxidising and adenylating conformations of *Ppy Fluc*. The adenylating conformation is displayed in green (cartoon). The oxidising conformation is displayed in cyan (cartoon).

3.4 Further Discussion

The work in this chapter sought to address whether a relatively small number of mutations could convert a murine acyl-CoA synthetase to have improved luciferin adenylation and the ability of luciferyl-adenylate oxidation, allowing bioluminescence. A directed evolution approach was first taken in the form of error-prone PCR to generate bioluminescent mutants of ACSM3. After this failed to generate any such mutants, a rational approach was taken. Analysis of the literature suggested ACSs like ACSM3 could be converted into functional Flucs through a small number of (in some cases single) substitutions. *Ppy Fluc* was chosen as the representative Fluc for substitution, and three components of it were separately substituted into ACSM3, namely: the active site consensus sequence, the substrate-determining and

interacting residues, and the R223 – T235 loop. However, these mutations failed to endow bioluminescence function when applied separately nor together.

3.4.1 Could a more optimal ACS been selected?

The ACS screening performed near the beginning of this chapter used the U.S. National Library of Medicine's protein BLAST tool. The WT *Ppy* Fluc protein sequence was used as the reference, and the most similar murine proteins found were medium-chain ACSs. Because of this, these proteins were selected as candidates for engineering. This was somewhat unexpected since Flucs are known to have evolved from long-chain ACSs. As was detailed earlier in the chapter there is high structural similarity between human ACSs and firefly luciferases (~85%), but low sequence similarity (~25%). BLAST tools rely on sequence similarity which is not necessarily indicative of structural similarity, which is more important for predicting protein function. As a result of this, a more advanced technique may be required for predicting the candidacy of ACSs, similar to the modelling performed which threaded a murine ACS sequence onto a human ACS 3D structure prior to 3D alignment to WT *Ppy* Fluc. Performing the candidate screen via this method would more likely ensure correct primary sequence alignment, and this may have identified long-chain murine ACSs as more appropriate than medium-chain ACSs for engineering. However this process relies on models based on other similar enzymes with known structure, ideally the crystal structure of the ACS of interest would be known, in case meaningful structural differences exist. Flucs appear to be functionally most similar to ACSs which catalyse the activation of fatty acids with a carbon chain length of C12 (Oba et al. 2006) which places it on the boundary between ACSMs and ACSLs. ACSM1 was experimentally identified as the best of the three ACSMs screened (namely ACSM1, ACSM2, and ACSM3) and this is perhaps to be expected since it performs optimally on fatty acids of carbon chain length C8 but is still ~50% as efficient with C12, whereas ACSM3 is optimal at C4 and loses almost all activity at C6 and longer (Fujino et al. 2001). There is no equivalent data for ACSM2 unfortunately.

3.4.2 Was epPCR adequate, and could it have been improved in throughput?

The failure of epPCR could be explained by the following: if it is assumed that there is one specific residue mutation which may endow bioluminescence function to murine ACSM3,

similar to the case of *Zophobus morio* ACS. For ACSs and Flucs which are ~550 residues in length, there are 11,000 unique single-substitution mutants possible (20 amino acids multiplied by 550 residues). If one screens 11,000 single-substitution mutants there is a 63.2% chance of finding the desired mutant [mathematical formula: probability of finding at least one specific construct = $1 - (\text{number of negative constructs} / \text{number of total constructs})^{\text{number of constructs screened}}$ = $1 - (10,999/11,000)^{11,000}$]. Approximately 33,000 mutants must be screened to have a 95% chance of finding the desired mutant. However, when using the specific MnCl₂ epPCR method (see 2.2.1.12) the number of colonies which are mutants (as opposed to WT) is likely ~25% (Jathoul, A., PhD thesis, University of Cambridge [not publically available]), therefore one may have to screen upwards of 132,000 colonies to have 95% probability of finding the desired mutant. It is estimated that the number of colonies per screened plate is ~500; five plates were screened therefore ~2,500 colonies were screened. This number is severely short of the >132,000 required for a >95% chance of success. Further, if endowment of bioluminescence function requires greater than one residue substitution, then the number of colonies one must screen for a >95% chance of success is orders of magnitude greater. Further, the MnCl₂ epPCR method is biased towards A→G and T→C transitions which may lower the diversity of the library (Lin-Goerke et al. 1997), and increase the proportion of duplicate mutants. This A→G and T→C bias can be countered (and the overall error rate of the polymerase increased) by instead using deuterium oxide (D₂O / ²H₂O) in the reaction mixture. The bias for transitions over transversions can be countered by also using H₂¹⁸O (Minamoto et al. 2012). From Sigma-Aldrich at the time of writing: D₂O is £2/mL and H₂¹⁸O is £264/mL, making the latter cost-prohibitive for this project. It may also be possible to alter the polymerase preference for certain bases by using dNTPs in unequal proportions. It is thus apparent that the epPCR implementation of high-throughput screening was not thorough enough, and would require either an extreme number of mutant library plates to be screened (>264 plates for single residue substitution), or a different approach entirely. One such approach involves the high-density printing of colonies onto plates with a flow cytometer which are then screened with D-LH₂ as usual. However, this method only marginally increases the efficiency (due to higher colony density) at the cost of increased complexity. Microfluidic and flow cytometric techniques have been developed for the directed evolution of fluorescent proteins and this is possible due to the ability to individually screen thousands of mutants per second, and each *E. coli* cell passes the laser/detector for

only a fraction of a second. However, bioluminescent proteins emit orders of magnitude fewer photons per second and thus require exposure to more sensitive detectors for a longer time, which precludes the use of this method. Several microfluidic technology company representatives were asked about options for bioluminescence screening, and they confirmed that it is not yet possible.

3.4.3 Unexplored techniques for identifying bioluminescent ACSM3 mutants

Elucidating why the other (non-epPCR) mutational strategies failed (*Ppy* Fluc: active site, SDRs, and R223 loop substitutions) is more difficult to predict definitively without running advanced molecular simulations which require supercomputer usage. Overall it can be inferred that since several engineering strategies failed to yield any functional luciferases from ACSM3-backbone constructs, it is likely that ACSM3 is too distant in function from D-LH₂ oxidation for such minor modifications to be successful (Figure 3.43). This is despite the LCMS data showing that ACSM3 is capable of D-LH₂ binding and adenylation. The domain exchange work (Chapter 4) suggests that ACSM1 would have been a better candidate for the mutations explored with ACSM3; in terms of sequence space ACSM1 may be closer to bioluminescence function than ACSM3. Further, it was revealed in Chapter 4 that *Ppy* Fluc with the CTD exchanged for those of ACSMs express and function more efficiently in HEK than *E. coli* (relative to WT *Ppy* Fluc), and this is likely true for ACSM backbone constructs too. Thus, these mutations should have been screened in HEK cells. Hybridisation mutagenesis techniques (most of which are PCR-based – sexual PCR) which sample a wider sequence space exist (Coco et al. 2001), and these may increase the possibility of discovering an ACSM mutant with bioluminescence function. There are many iterations of sexual PCR such as Random Chimeragenesis on Transient Templates (RACHITT), and these typically require ≥85% DNA sequence homology (Coco 2003); *Ppy* Fluc and ACSM3 were codon optimised in silico for maximal DNA homology but remained below 65%. Due to the low sequence homology between Flucs and murine ACSs it is not possible to perform the high-diversity techniques.

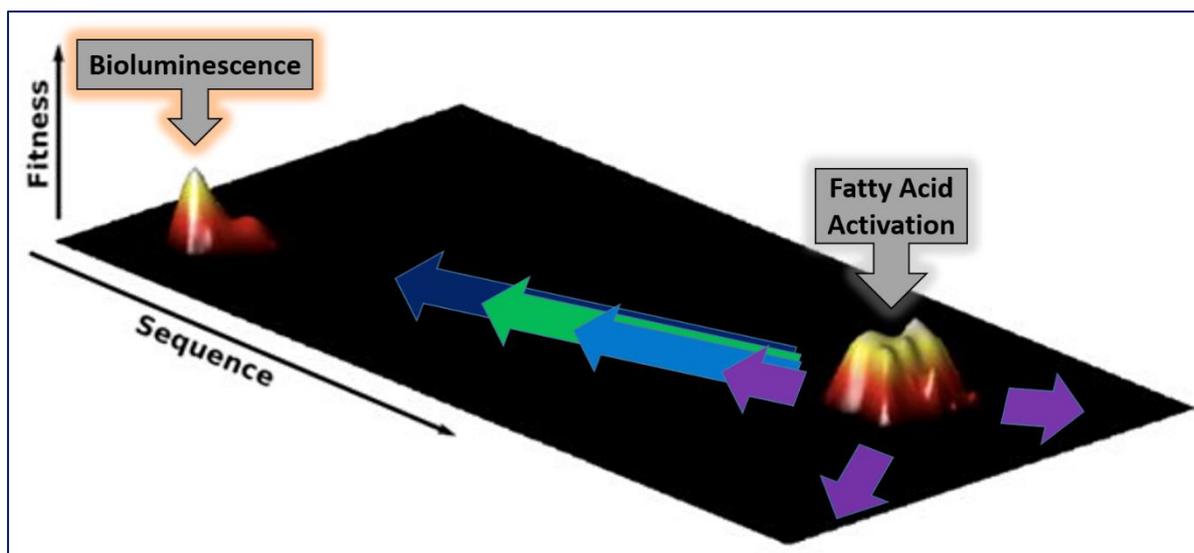


Figure 3.43. Adapted from (Romero and Arnold 2009). Presumptive visualisation of area of sequence space probed by the various protein engineering approaches. A different coloured arrow represents each approach attempted: ePCR random mutagenesis in purple; active site consensus sequence substitution in light blue; active site consensus sequence + substrate determining residue substitution in green; active site consensus sequence + substrate determining residue + R223 loop substitution in dark blue.

3.5 Conclusions

From the Ni-NTA purification assay, murine ACSM3 and ACSM1 appear to express poorly in *E. coli*, and this poor expression was not alleviated by removal of the MTS. Nevertheless, concentrated ACSM3 protein purifications did not produce bioluminescence with D -LH₂, leading to the conclusion that ACSM3 is non-bioluminescent. From the LCMS experiment it appears that ACSM3 does adenylate (but does not oxidise) D -LH₂. The random mutagenesis and rational mutation approaches undertaken within this chapter were unable to endow bioluminescence (and therefore oxygenase) function to ACSM3. Deeper modelling of ACSM3 revealed a region of steric clash with D -LH₂-AMP in the active site of the putative oxidising conformation. Overall, the strategy of creating a novel Fluc from the backbone of ACSM3 failed to generate any viable constructs. It was assumed that this approach had a low probability of creating bioluminescent mutants, but that any such mutants could be trivially optimised for mammalian applications. As such, a lower risk, more incremental approach was undertaken in Chapter 4.

4 Chapter 4 – Creation and In Vitro Testing of the *Photinus Pyralis* / Murine Acyl-CoA Synthetase 1, 2, and 3 Hybrid Constructs

4.1 Chapter Summary

In this work it was hypothesised that murine ACSM domains could be recombined with *Ppy* Fluc N- or C- terminal domains, in a modular fashion. This is based on the observation that it is possible to exchange Fluc terminal domains between different species, such as the reported chimeras between *Ppy* Fluc and *Luciola italic* Fluc (Branchini et al. 2014). As an alternative to the approach in Chapter 3 of adding small and potentially bioluminescence-endowing modifications to ACSM-backed constructs, I proposed domain exchange using the C-terminal domain (CTD) of a murine ACS in conjunction with the N-terminal domain (NTD) of *Ppy* Fluc. A truncated version of the NTD of *Ppy* Fluc is known to produce 0.03% of the bioluminescent specific activity of the complete WT *Ppy* Fluc (Zako et al. 2003), indicating that limited bioluminescent potential exists in this domain. Chimeras of the NTD of *Ppy* Fluc with murine ACSM1, 2, and 3 CTDs were constructed, as well as a range of chimeras between different ACSM NTDs and CTDs. Constructs combining the NTD of *Ppy* Fluc and the CTDs of ACSM1, 2, and 3 expressed in *E. coli* showed *Ppy* Fluc-relative specific activities of ~0.9%, ~0.004%, and ~0.2% respectively. In HEK cells, the chimera with the CTD of ACSM1 showed ~10% and CTD of ACSM3 ~2% of activity relative to *Ppy* Fluc. When increased expression of ACSM1 in HEK was accounted for the *Ppy* Fluc-relative brightness decreased to ~4%. The chimeras were red-shifting, and thus the effective light output when Hb filtered would likely be greatly improved relative to *Ppy* Fluc.

4.2 Overall Introduction

Ppy Fluc is composed of a relatively large (434 aa) N-terminal domain (NTD) and a smaller (104 aa) C-terminal domain (CTD) joined together by a flexible (8 aa) linker. These domains have been demonstrated to be modular as the domains of different species of luciferases can be interchanged while retaining function; indeed it was shown that a chimeric construct composed of the NTD of *Ppy* and CTD of *Luciola italic* (*Lit*) demonstrated a 1.4 fold greater bioluminescence quantum yield than WT *Ppy* Fluc, which is the brighter of the two parent

enzymes (Branchini et al. 2014). The CTD of *Ppy* Fluc is not essential for function but it greatly enhances it, as the NTD alone produces 0.03% the bioluminescent activity of the WT enzyme (Zako et al. 2003); i.e. the presence of the CTD allows the WT enzyme to be more than 3000 fold brighter. From an engineering perspective this is an ideal starting point as – since the CTD is redundant – the NTD with a modified CTD has a high probability of functioning. Further, luminometry or BLI is an easily quantifiable metric of luciferase fitness. As discussed in the general introduction it is believed that Flucs evolved from an ancestral non-bioluminescent ACS and these two proteins (in insects) have a high structural similarity due to their high protein sequence homologies (Prado et al. 2016). Mammalian and insect ACSs are also composed of a relatively large (436 aa) NTD and a smaller (103 aa) CTD joined together by a flexible (4 aa) linker – all sizes here are for murine ACSM1, but are similar across ACSs. It has previously been demonstrated that chimeric constructs composed of the NTD of *Ppy* Fluc and the CTD of ACS from *drosophila* (CG6178) retains 4% of the bioluminescent activity of the WT *Ppy* Fluc (Oba et al. 2006). The WT *drosophila* ACS used for hybridisation has greatest activity with fatty acids of chain-length C12 – at the boundary between short- and long-chain fatty acids.

For mammalian applications, thermal stability is of great importance for genetic reporters. x11Fluc is a *Ppy* Fluc mutant engineered to be more resistant to high temperatures and fluctuations in pH – at least 80% of maximum bioluminescence output over a pH range of 6.6-8.6. x11Fluc was created by combining and screening the mutations of three other stabilised *Ppy* Fluc mutants, namely x2Fluc, x4Fluc, and x5Fluc. The nomenclature for these four mutants is derived from the number of residue mutations they have compared to the WT enzyme, e.g. x11Fluc has eleven mutations. These mutations are all substitutions. The stabilising mutations identified for the creation of x11Fluc within the WT *Ppy* Fluc sequence are: F14R, L35Q, A105V, V182K, T214C, I232K, D234G, E354R, D357Y, S420T, and F465R (Jathoul et al. 2012). These substitutions occur disproportionately in the NTD with only F465R occurring in the CTD, since a proportionate distribution would suggest two in the CTD. From this it is implied that the WT NTD has a greater role in determining stability than the CTD. However, the F465R mutation increases the brightness of WT *Ppy* Fluc by ~1.5 fold, and is the second most potent thermal stabilising mutation of the five found in x5Fluc (4 fold brighter than WT *Ppy* Fluc after 42 °C incubation for 10 mins) (Law et al. 2006). However,

despite x5Fluc possessing a greater number of mutations it is much less thermo-stable than either x2Fluc or x4Fluc (20% activity remaining after 1hr at 40 °C for x5Fluc versus 75-80% for x2Fluc and x4Fluc) (Jathoul et al. 2012). Thus, mutagenesis of the WT *Ppy* Fluc CTD can improve both the thermal stability at 40-42 °C and brightness under ideal conditions (RT and pH 7.8) of the enzyme, but the greatest improvements are to be found in the NTD. Solely by modifying the CTD – such as with the replacement of a murine ACS CTD – it may be possible to increase the stability of the overall enzyme further than that given from F465R. This is important since in vivo physiological temperature in mammals where BLI experiments are performed is 37 °C.

4.3 Results and Discussion

4.3.1 Design, cloning, and screening of the *E. coli* chimera constructs

4.3.1.1 Design

The creation of the NTD *Ppy* Fluc / CTD *Lit* Fluc chimera (Branchini et al. 2014) exploited a silently introduced PacI site within the linker sequence of *Lit* Fluc (which occurs naturally in *Ppy* Fluc cDNA), and this resulted in the residues 1-439 of WT *Ppy* Fluc defining the *Ppy* Fluc NTD. The same splice site was used for designing the NTD *Ppy* Fluc / CTD *Mus* ACSM chimeras. Unlike the *Ppy* and *Lit* Flucs, PacI cannot be silently mutated into the equivalent position in any of the three ACSMs. Therefore the ACSM protein sequences must be mutated to allow PacI substitution, and table 4.1 lists the mutations required. As these residues were identified structurally for ACSM3 (Chapter 3), they could then be found by 2D alignment in ACSM2 and ACSM1 due to the high homology. Figure 4.1 shows the 3D sequence alignment of the residues encoded by PacI-containing cDNA, and the equivalent residues in ACSM3. The equivalent positions in ACSM1 and ACSM2 are shown in Figure 4.2.

<i>Ppy</i> Fluc	AKNIKKGP-APFYPLEDGTAGEQLHKAMKRYAL----VPGTIAFTDAHI-EVNITY-AEYFEMSVRLAEAMKRY-GLNTNHRIVVSSENSLOFFMPVLGALFIGVAVAPANDCYNERELLN	120
ACSM3	-----IE---IPEYFNFAKVDLDQWTNMEKAGKRLSNPAFWIDGNGEELRWSFEELGLLSRKFANILTEACSLQRGDRVMVILPKIPEWNLANVACLRTGTVLIPGTTQLTQKDILY	120
<i>Ppy</i> Fluc	SMNISQPTVVFVSKKGLQKILNVQKLPPIQKIIIMDSKTDYQGFQSMYTFVTSHLPPGFNEYDFVPESFDRDKTIALIMNSS----PKGVALPHRALAVRFSHARDPIFGNQIAPDTAI	240
ACSM3	RLQSSKAKCIITDDTLAPAVDAVAAKCENLHSLKLIVSQ-HSREGWGNLKEMMKY-A--SDS--HTCVDTKHD-EMMAIYFTSGTTGPPKMIGHTHSSFGLGLSVNGRF-WLDLIA-SDVM	240
<i>Ppy</i> Fluc	LSVVPFHGFGMFTTL-GYLISGFRVVLN-Y-RFEEELFLRSLQDYKIQSALLVPTLFSFLAKSTLIDKYDLSNLHEIASGGAPLSKEVGEAVAKRFHLPGRQYGLTETTSAILITPKG	360
ACSM3	WNTSDTGWAKSAWSSVFSWQACVFAHYLPRFESTSILQTLKFPITVFCAPTAYRMLVQN-DMSSYKFNSLKHCVSAGEPINPEVMEQWRKKTGL-DIYEGYQTET-VLICGNFKG	360
<i>Ppy</i> Fluc	D-DKPGAVGVVPPFEAKVVDLDTGKTLGVNQRGELSVR----GPMIMSGYVNNPEATNALIDKDWLHSGDIAYWDEDEHFFIVDRKSLIKYKGCQVAPAELESILLQHPNIFDAGVA	480
ACSM3	MKIKPGSMKPSAFDVKILD-ENGATLPPQEGDIALQVLPERPFLFTHYVDNPSKTASTLRGS-FYITGDRGYMDEGYFWFVARSDDIILSSGYRIGPFEVESALIEHPSIAESAVV	480
<i>Ppy</i> Fluc	GLPDDDAGELPAAVVLEHGKMT-EKEIVDYVASQVT---TAKKLRGGVVFVDEVPKGL-KLDARKIREILIKAK	554
ACSM3	SSPDPIRGEVVKAFIVLNPDYKSHDQEQLKKEIQEHVKKTTAPYKYPRKVEFIEELPKTVSGKVKRNLKKEWV--	554

Figure 4.1. Structure-derived protein sequence alignment of *Ppy* Fluc and ACSM3 (via TM-align). The residues containing the PacI site in *Ppy* Fluc (as DNA) and the equivalent residues in ACSM3 are marked by the red box.

ACSM1	mqwlksfqickvlqgfslsptqlhr-----rlfsrvgaprwndhdspeefnfasdvl	52
ACSM2	mhhllwkiprlftlwgneiscrtfhnikkliipi-----qwghqeapakfnfasdvi	51
ACSM3	mvmlrr-arcfqrllaipdpmrvlykdyrtatpqnfsnyesmkqdfkieipeyfnfakdvl	59
	* * . :: : . : * ****.***:	
ACSM1	dywaqmeeeegkrpsspafwvvnngggdeikwsfrklrdltcrtanvfeqicglqqgdhlal	112
ACSM2	dhwasvekagkrssgpalwvnmngsgkeikwsfrelseaskqtanvlsqacglhrgrdvav	111
ACSM3	dqwtmkekagkrlnspafwidnggeelrwsfeelgllsrkfanilteacslqrgdrvmv	119
	* *:::*** :***.***:*.*.***:***.* : : ***: *.***:***: :	
ACSM1	ilprvpewwlvvtvgcmrtgiifmpgttqlkakkdilyriqisrakaivttaslvpevesva	172
ACSM2	vlpripewwlmilgcmrtglvfmptiqmrssdilyrlqaskaraivagdevagevdava	171
ACSM3	ilpkipewwlanvaclrtgtvlipgttqltqkdilyrlqsskakciitddtlapvdava	179
	:***:***** :.***:*** :::*** * : .*****:* *:*:*:*: :. *::**	
ACSM1	secpdlkktlvvsdhshegwldfcslikaspdhtcikskmkdpmaiffstggtgypkma	232
ACSM2	pdcsflkikllvsensregwlnfkallkeastihqcvetesresaaiyftsgtsgppkma	231
ACSM3	akcenlhsklivsqhsregwgnlkemmyasdshtcvdtkhdenmaiiftsgtsgppkmi	239
	. * * : ***:***:***:*** : : : * * * * * : : : : * : ***:*****:* * **	
ACSM1	khnqglafrsyipscrkllkklktsdilwcmstdpgwilatvgcliepwtsgctvfihhlpq	292
ACSM2	ehshcslgikakmdaaswtglstsdiiwtisdtawimnilgaflepwwlgacifvhlhpk	291
ACSM3	ghthssfglglsvngfrwldliasdvmwntsdgtgwaksawssvfspwtqgacvfahylpr	299
	*.: . * :***:*** * * . * * . * : * * * * :	
ACSM1	fdpkvivevlfkypitqclaapgvyrmlvqqktsnlrfptlehcttggesllpeeyeqwk	352
ACSM2	fdsqtvklvlyssypintlvgapiiyrmllqqdlssykfphlhscfsggetllpetlenwk	351
ACSM3	festsilqtlskfpitvfcsaptayrmlvqndmssykfnslkhcvsagepinpevmeqwr	359
	* : : : * . * * * . * * * * * : * . * * . * * : * * : * * * * :	
ACSM1	qrtglsihevygqsetgissatlremkikrgsigkailpfdlqiidekgnilppntegy	412
ACSM2	aktgleireiygqtetglicrvsrtmkvkpgylgtafahydvqvideqgnvlpqgkegdi	411
ACSM3	kktgldiyegyqqtetvlicgnfkgnkikpgsmgkpspafdvkildengatlpqgqegdi	419
	:***.* * ***:*** : . : * * * * * * * . * : * : * : * * * * * * * * *	
ACSM1	girikptrplglfmeyenspestsevecgdfynsgdratideegyiwflrgddvinasg	472
ACSM2	airvkpiwpgmfsgyvdpkktqdnirgdfwlmgdrigkdpegyfhfigrsddiinssg	471
ACSM3	alqvlperpfglfthyvdnpsktastlrgsfyitgdrgymdedgyfwfvarsddiilssg	479
	: : : * * * * * * * * : * . * . * . * . * * : * * * * * * * * * * * * * * *	
ACSM1	yrigpaevenalaehpavaesavvsspdkdrgevvkafivlnpeflshdqeqlikelqhh	532
ACSM2	yrigpsevenalmehpavsetavisspdpsrgevvkafvvlapeflshdrdqtkvlqeh	531
ACSM3	yrigpfevesaliehpsiaesavvsspdpirgevvkafivlnpdykshdqeqlkkeiqeh	539
	***** ***.** ***:***:***:*** *****:*** * : : ***:*** * :*. *	
ACSM1	vksvtapykyprkvefvselpktvtgkikrkelrnkefgql---	573
ACSM2	vksvtapykyprkvefvldlpktvtgkieraklrakewktsgra	575
ACSM3	vkkttapykyprkvefieelpktvsgkvkrnelrkkewvtt---	580
	..***: :*****:***:* * * * * :	

Figure 4.2. 2D Protein sequence alignment of ACSM1, ACSM2, and ACSM3 (via Clustal Omega). The equivalent residues for the PacI-containing (as DNA) residues of *Ppy* Fluc are marked by the red box.

Full name (from NCBI)	acyl-CoA synthetase medium-chain family member 1	acyl-coenzyme A synthetase ACSM2, mitochondrial isoform 3	acyl-coenzyme A synthetase ACSM3, mitochondrial
Shorthand name	ACSM1	ACSM2	ACSM3
NCBI/GenBank reference code	EDL17174.1	NP_666309.1	NP_058566.3
Mutations required to introduce Pac1 site	V467L N469K	I466L	I474L L476K

Table 4.1. The protein sequence mutations required to introduce a PacI site into the linkers of ACSM1, ACSM2, and ACSM3 at the equivalent position as in *Ppy Fluc* cDNA linker.

Since the PacI site (and therefore the mutations required to introduce it) occurs in the linker sequence – and this region is a relatively large loop – it is thus assumed to be unlikely that mutations to this region will impact protein function (Halliwell et al. 2018). Further, the linker sequence of *Ppy Fluc* is flexible and thus should retain flexibility of the ACSM linker when partially substituted. The protein sequences for the three ACSMs were entered into Integrated DNA Technology’s codon optimisation tool and processed for optimal DNA codons in *E. coli*. The resulting DNA sequences were then altered to insert the PacI site in the linker, and to insert the flanking NdeI and BamHI sites required for subcloning into pET16b for expression with an N-terminal 10xHis-Tag. Once these changes were made the DNA sequences were ordered as DNA oligos of the NTD and CTD, or as (in the case of ACSM1) the complete gene. The ordered sequences for the three ACSMs are displayed as maps in Figures 4.3, 4.4, and 4.5. The domains of the ordered enzymes could be reassembled into a number of hybrid constructs (table 4.2). The NTD of *Ppy Fluc* – for pET16b expression as a truncated enzyme – was also ordered for bioluminescence comparison (Figure 4.6).



Figure 4.3. Map of gBlock DNA sequence encoding the entire ACSM1 with PacI mutation, for insertion into pET16b. An expanded version of this figure – with DNA and protein sequence shown – can be found in the appendix (Figure 4.3EX).

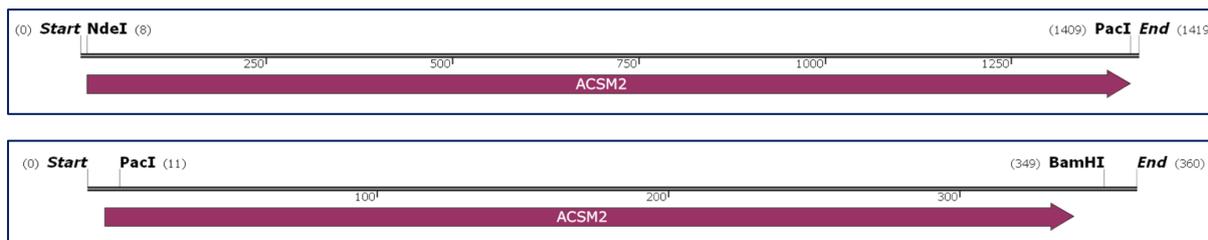


Figure 4.4. Maps of gBlock DNA sequences encoding the NTD and CTD of ACSM2 with PacI mutation, for insertion into pET16b. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figures 4.4EXa and 4.4EXb).

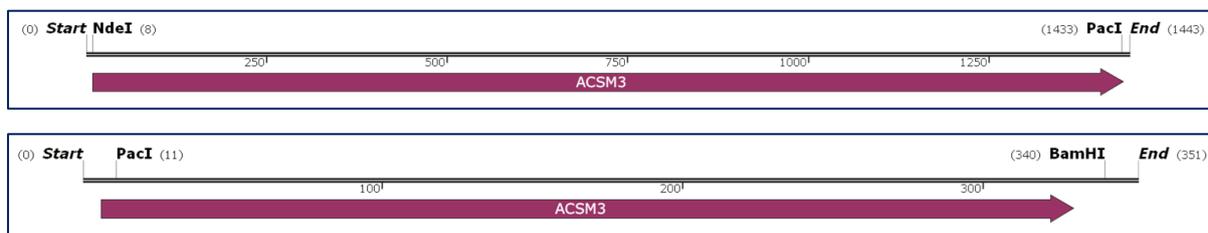


Figure 4.5. Maps of gBlock DNA sequences encoding the NTD and CTD of ACSM3 with PacI mutation, for insertion into pET16b. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figures 4.5EXa and 4.5EXb).

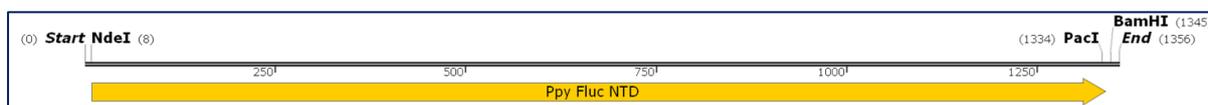


Figure 4.6. Map of gBlock DNA sequence encoding the NTD of *Ppy Fluc* (Pp00) (with PacI) with a stop codon and BamHI site for insertion into pET16b. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 4.6EX).

	<i>Ppy</i> Fluc NTD	ACSM1 NTD	ACSM2 NTD	ACSM3 NTD
<i>Ppy</i> Fluc CTD	<u>PpPp</u>	A1Pp	A2Pp	A3Pp
ACSM1 CTD	<u>PpA1</u>	A1A1	A2A1	A3A1
ACSM2 CTD	<u>PpA2</u>	A1A2	A2A2	A3A2
ACSM3 CTD	<u>PpA3</u>	A1A3	A2A3	A3A3
No CTD	<u>Pp00</u>	A100	A200	A300

Table 4.2. Constructs combining the *Ppy* Fluc and ACSM domains. Those in green were cloned and screened for bioluminescence function, those in red were not cloned due to time constraints or equivalent chimeras being non-bioluminescent. Those underlined were found to be bioluminescent with Δ -LH₂.

4.3.1.2 Cloning

The ACSM inserts were (along with *Ppy* Fluc in pET16b) terminally digested with NdeI, PacI, and BamHI (see 2.2.1.6) and purified via gel electrophoresis (see 2.2.1.4) for the NTDs and CTDs of each, giving 8 fragments. 12 constructs were to be created as displayed in table 4.2, and the nomenclature for these constructs is written as: NTD followed by CTD with two characters for each domain; *Ppy* Fluc is shortened to Pp, and ACSM1, 2, and 3 are shortened to A1, A2, and A3 respectively. For example, the chimera composed of the NTD of *Ppy* Fluc and the CTD of ACSM2 is written as PpA2. PpPp is the positive control and already existed in pET16b at this point but without the PacI site in the linker (due to codon optimisation), thus it was created alongside the other 10 constructs and acted as a positive cloning control. Using various combinations of the 8 purified (NTD and CTD) fragments and pre-digested pET16b vector, the 12 constructs were ligated (see 2.2.1.7). The ligation mixtures were then transformed into DH5 α cells (see 2.2.3.3), which were picked, partially DNA miniprep, and the highest-concentration sequence-correct cultures stored at -80 °C as glycerol stock.

4.3.1.3 Screening of the *E. coli* chimera constructs

4.3.1.3.1 Bioluminescence screening of colonies

The sequence-verified and highest-concentration DNA minipreps were retransformed into BL21 competent cells (see 2.2.3.3). ACSM3 (A3A3) was found in the previous chapter to be not detectably bioluminescent by colony IPTG induction and Δ -LH₂ spraying. Here it was

found that all constructs containing the NTD of ACSM1, ACSM2, and ACSM3 (A1A1, A2A2, A2A3, A2Pp, A3A3, A3A2, and A3Pp) were not detectably bioluminescent via the same method. The data for this are not shown. As single colonies on petri dishes the ACSM1, ACSM2, and ACSM3 CTD chimeras (PpA1, PpA2, and PpA3) along with Pp00 were below the detection threshold for the DSLR, however PpPp was clearly detectable (Figure 4.7). All constructs containing the NTD of *Ppy* Fluc (PpPp, PpA1, PpA2, PpA3, and Pp00) were found to be detectably bioluminescent as induced colonies on the PhotonIMAGER with the exception of Pp00 (Figure 4.8). The detectable constructs were also bright enough for a bandpass filter wheel spectrum to be acquired for each (Figure 4.9).

Given the success of the *Ppy* Fluc NTD-containing chimeras created and screened so far within this chapter, further work was anticipated to more finely exchange portions of the *Ppy* Fluc NTD with that of ACSM NTD; this work would render the ACSM NTD-containing constructs created and screened within this chapter somewhat redundant. This anticipated work was explored and is detailed within Chapter 6. Nevertheless, the *Ppy* Fluc NTD-containing chimeras were subjected to further screening within this chapter to determine both the optimal ACSM variant for *Ppy* hybridisation, and technique for expression and screening.

It is true that – as is the case with Pp00 – the ACSM NTD-containing constructs may be very dimly bioluminescent but below the detection limit for the PhotonIMAGER when screened as induced colonies. If this were the case there is no way to definitively exclude such a construct since its brightness is unknown, and even if the experiment were optimised to bring Pp00 above the detection threshold the ACSM NTD-containing constructs may still remain below.

Thus, the constructs with no detectable bioluminescence function as colonies were discarded at this stage in the project; the one exception was Pp00 as it is known to be very weakly bioluminescent (0.03% that of PpPp) from the literature (Zako et al. 2003). If Pp00 became detectable in later experiments within this Chapter (as purified protein from *E. coli*, or when expressed via HEK cells), then it could be used as a benchmark to quantify the brightness increases endowed by the presence of different ACSM CTDs for *Ppy* Fluc NTD.

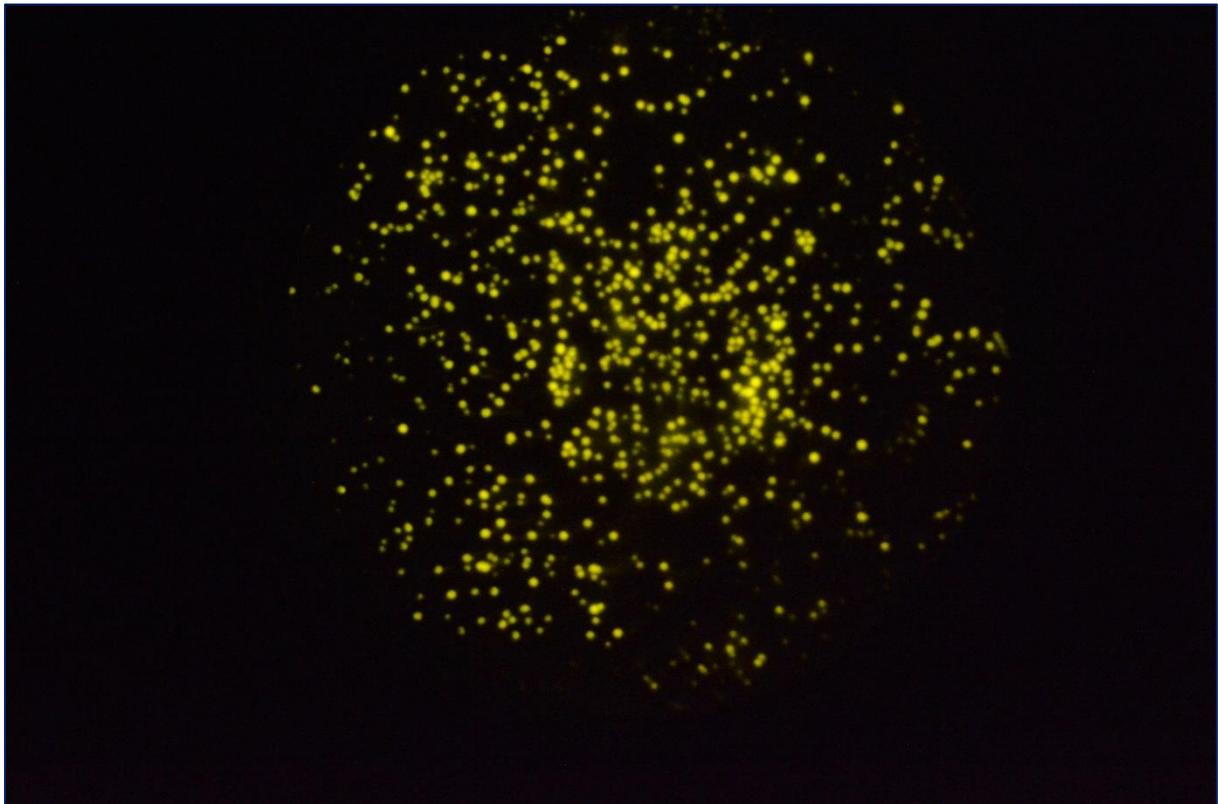


Figure 4.7. DSLR image of D-LH₂-sprayed PpPp *E. coli* colonies. The uniform colour indicates a monoclonal transformation.

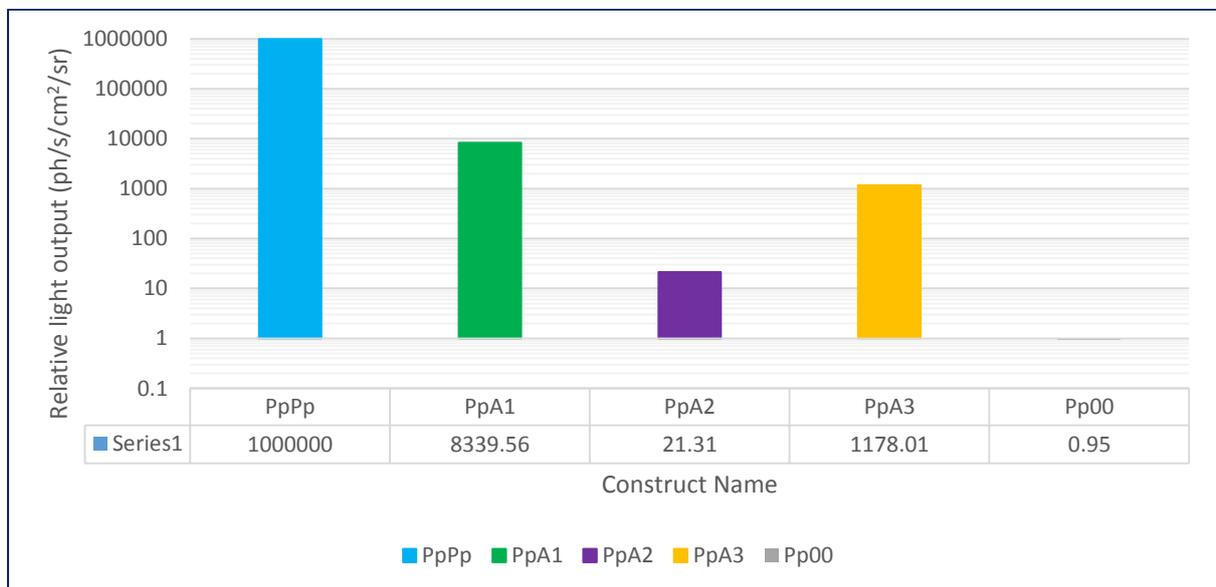


Figure 4.8. PpPp-relative brightness of PpA1, PpA2, PpA3, and Pp00. Note that the PpPp brightness bar has a value of 1,000,000 to allow the smaller bars to display correctly on a log scale.

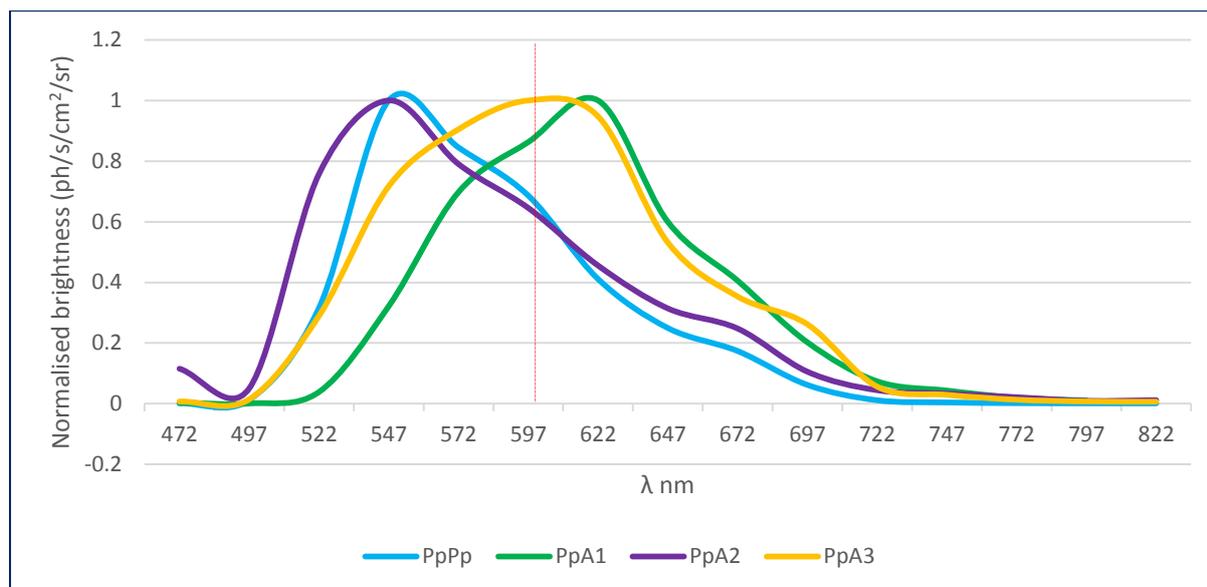


Figure 4.9. Normalised spectra of PpPp, PpA1, PpA2, and PpA3. The spectrum for PpA2 may be less reliable due to the severely reduced brightness (Figure 4.8). Emission to the right of the red dotted line (>600 nm) is considered to be within the bio-optical window.

4.3.1.3.2 Screening of purified proteins

4.3.1.3.2.1 Quantification of purified bioluminescent proteins

The bioluminescent constructs (PpPp, PpA1, PpA2, and PpA3) and Pp00 were expressed in bulk in *E. coli* and purified via nickel nitrilotriacetic acid (Ni NTA) column (see 2.2.3.7.2). Due to PpA1 and Pp00 being created at a later date, they were also purified at a later date alongside a repeat of PpPp. Therefore they represent the second of two batches of protein purifications, namely BatA and BatB.

4.3.1.3.2.1.1 SDS-PAGE gel analysis and quantification

Both batches of proteins were run on a SDS-PAGE gel (see 2.2.2.3). The gel was then Coomassie stained (see 2.2.2.4) (Figure 4.10). A degree of background staining was observed, particularly in the PpPp BatA lane. The stained gel was then scanned and the protein content of the bands quantified via ImageJ analysis (see 2.2.2.6) (Figure 4.11). As PpPp was present in both batches it was used as a control to allow comparison of constructs from different batches, and it was found that PpA1 and PpA2 express similarly to PpPp, whereas Pp00 and PpA3 are considerably less efficient (Figure 4.12).

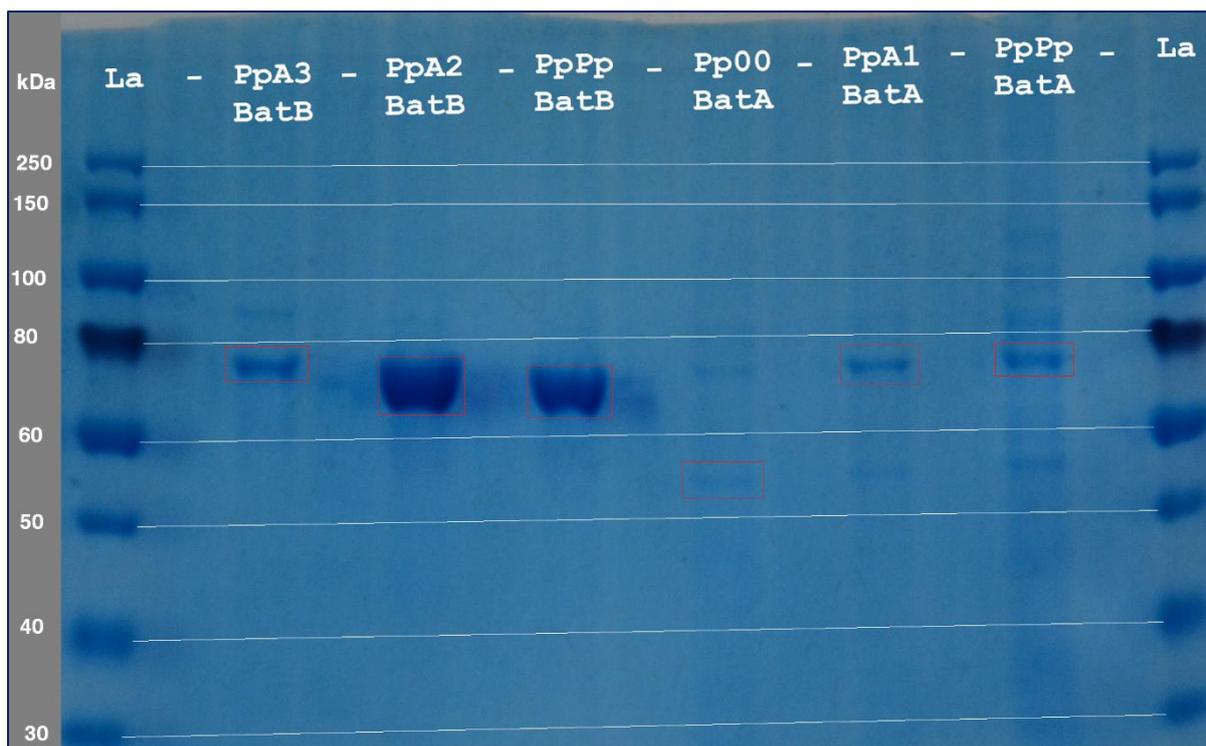


Figure 4.10. Coomassie-stained SDS-PAGE gel of PpPp, PpA1, PpA2, PpA3, and Pp00. “La” stands for ladder. Ladder bands of the same size have been bridged in white. Desired bands have been boxed in red.

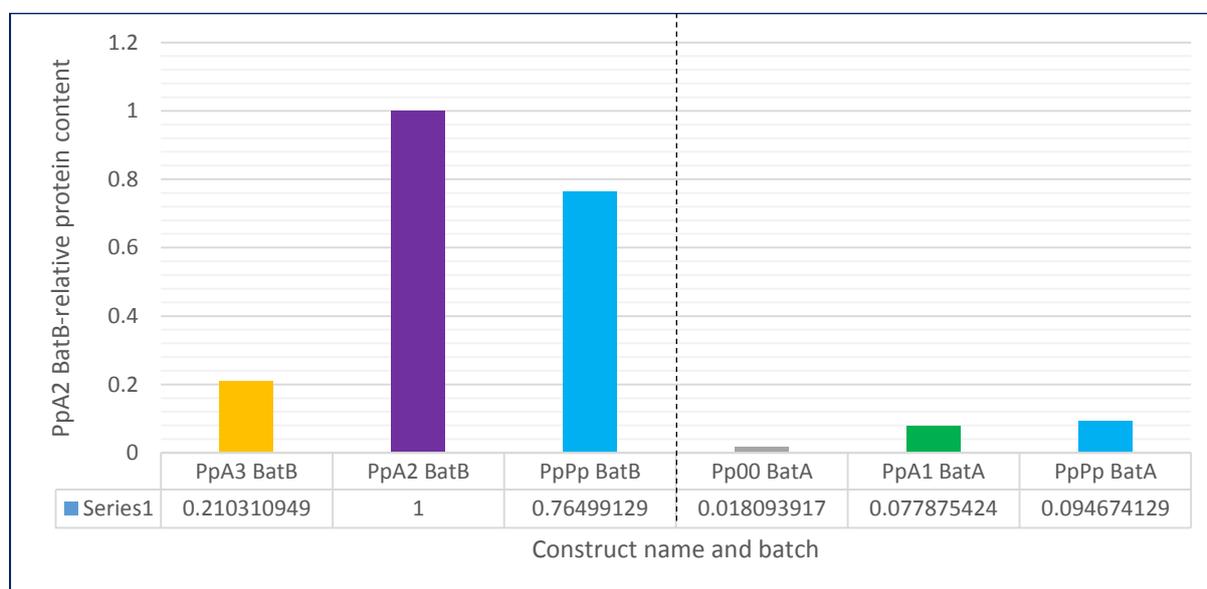


Figure 4.11. ImageJ analysis values of Coomassie-stained SDS-PAGE gel of PpPp, PpA1, PpA2, PpA3, and Pp00. All construct bands are relative to the highest concentration band (PpA2). Dotted line separates batch A from B.

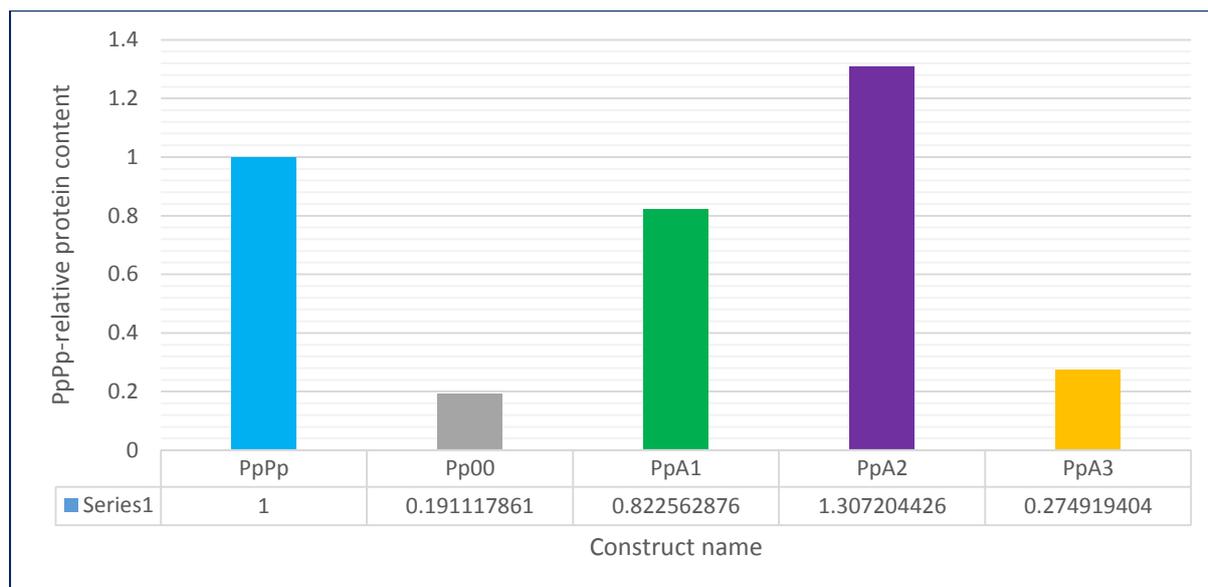


Figure 4.12. Combined ImageJ analysis values of Coomassie-stained SDS-PAGE gel of PpPp, PpA1, PpA2, PpA3, and Pp00. All construct expression efficiency is relative to PpPp.

4.3.1.3.2.1.2 Bradford assay

A Bradford assay was performed in triplicates (see 2.2.2.7) on the bioluminescent constructs (PpPp, PpA1, PpA2, and PpA3) and Pp00 using the same protein purification batches run on SDS-PAGE gel (Figure 4.13). As with the ImageJ analysis, the PpPp control allowed comparison between batches (Figure 4.14). The Bradford assay expression efficiency results did not match with the ImageJ analysis of the SDS-PAGE gel; this is likely due to high background protein present in the purification, evident from the SDS-PAGE gel (Figure 4.10).

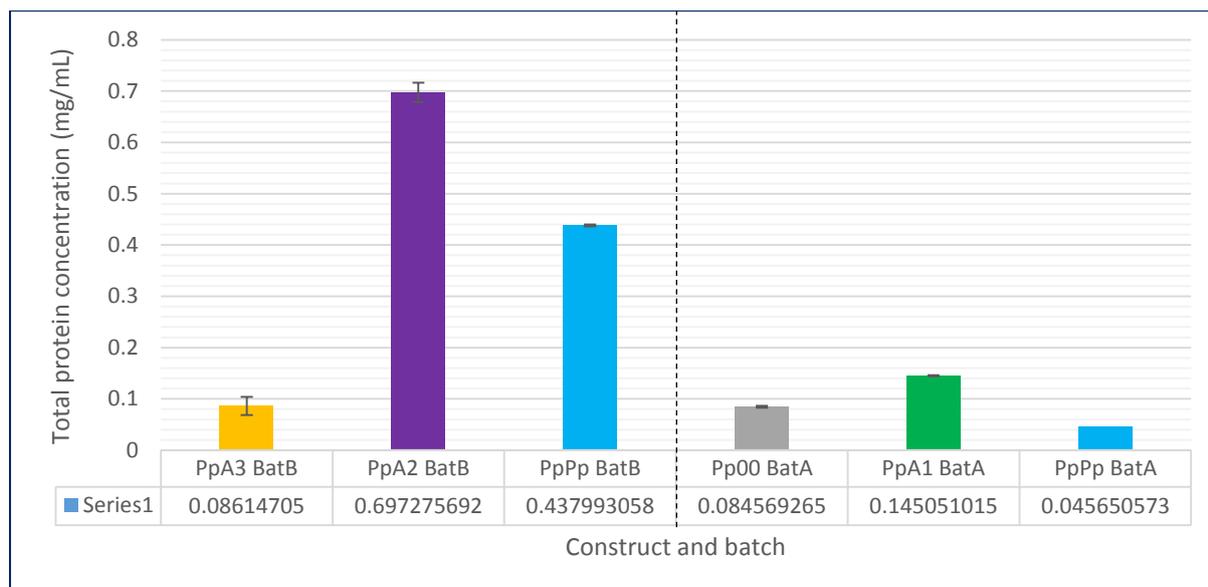


Figure 4.13. Bradford assay protein concentration values of construct protein purification batches of PpPp, PpA1, PpA2, PpA3, and Pp00. Dotted line separates batch A from B.

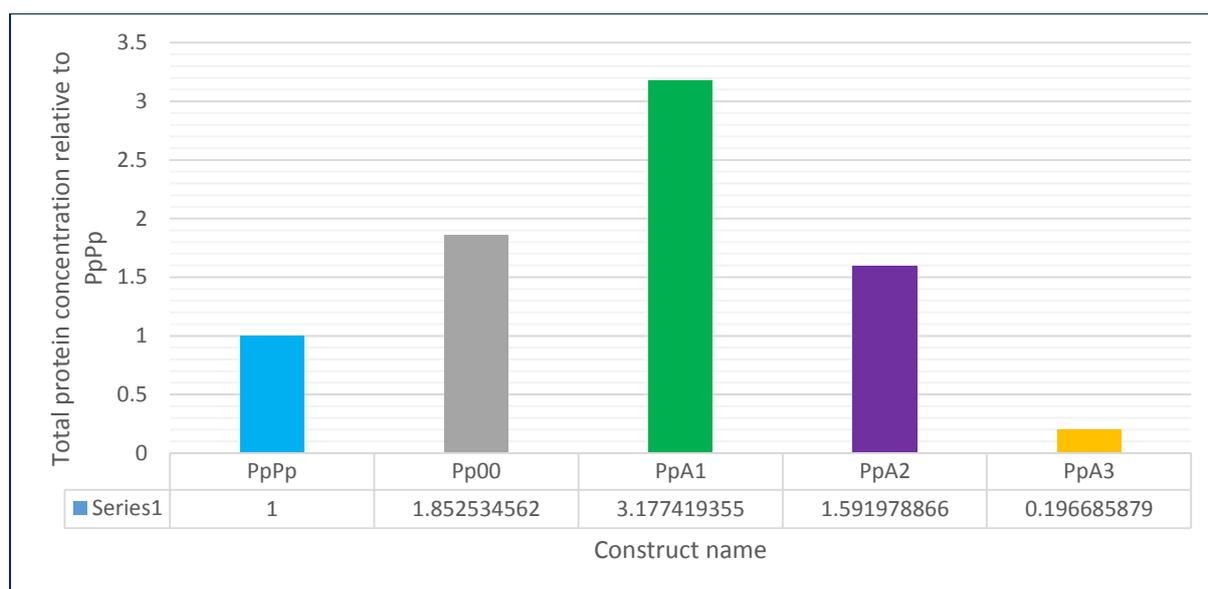


Figure 4.14. Bradford assay-derived expression efficiencies of PpPp, PpA1, PpA2, PpA3, and Pp00. All construct expression efficiency is relative to PpPp.

4.3.1.3.2.2 BLI screening with D-LH₂

Given that the SDS-PAGE gel revealed several non-specific bands (other than in PpA2 BatB and PpPp BatB), and these are included in the total protein for the value given by the Bradford assay, the Bradford assay values were not regarded as being an accurate reflection of the expressed construct protein. Thus, the ImageJ analysis of the SDS-PAGE gel was used to

estimate the PpPp-relative concentration of construct protein in each purification (figure 4.12), as it is more specific than the Bradford assay. With the concentration of construct protein known for each protein purification, bioluminescence screening could be performed with equal concentrations of construct protein per reaction (see 2.2.6.3). The PhotonIMAGER was used to acquire both total light output and spectra for the five construct purifications (Figure 4.15 and 4.16). These data revealed that the purified Pp00 fraction was indeed bioluminescent, but too dim for a spectrum to be acquired; however it is ~50 fold dimmer than would be expected from the literature (PpPp-relative brightness of 0.0006% rather than 0.03% (Zako et al. 2003)). The PpPp-relative specific activities and spectra of the chimeric constructs appear similar to the data from the colony screens (Figures 4.8 and 4.9); PpA1 was superior in terms of both overall brightness and degree of red-shifting, followed by PpA3 and PpA2.

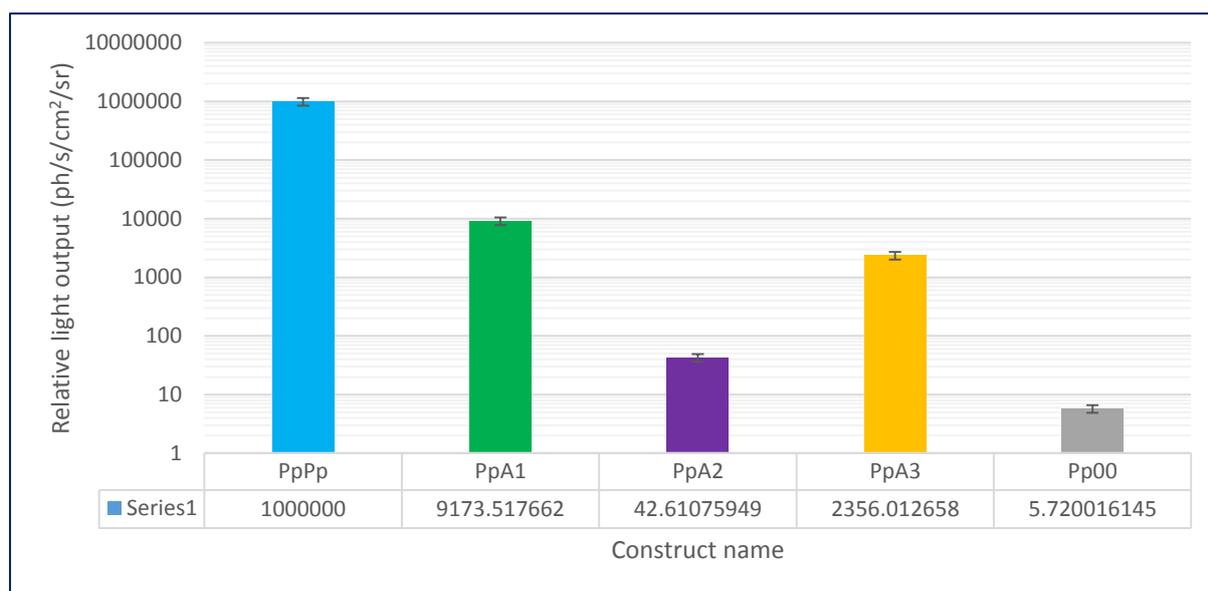


Figure 4.15. PpPp-relative specific activities of PpA1, PpA2, PpA3, and Pp00. Note that the PpPp brightness bar has a value of 1,000,000 to allow the smaller bars to display correctly on a log scale.

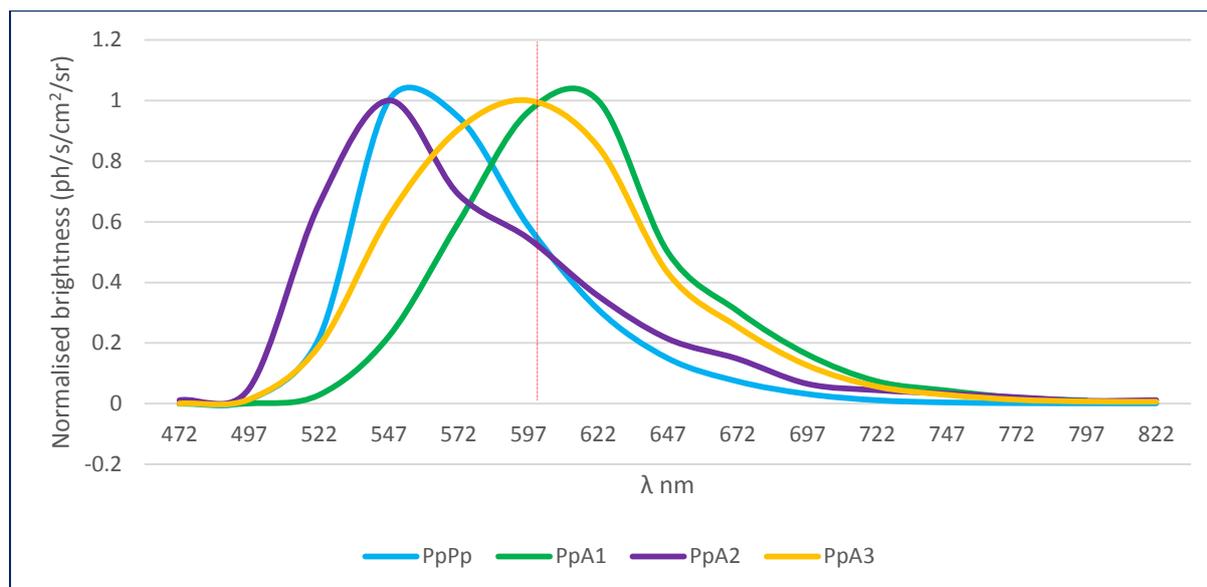


Figure 4.16. Normalised spectra of PpPp, PpA1, PpA2, and PpA3. The spectrum for PpA2 may be less reliable due to the severely reduced brightness (Figure 4.15). Emission to the right of the red dotted line (>600 nm) is considered to be within the bio-optical window.

4.3.2 Design, cloning, and screening of the mammalian chimera constructs

4.3.2.1 Design

From the *E. coli* work earlier in this chapter PpA1 followed by PpA3 was determined to be the superior ACSM-containing constructs tested thus far – they were the brightest and most red-shifted – as a result they were advanced to screening in mammalian cells. Dr Amit Jathoul kindly donated a mammalian expression vector known as “pCCL 305” (Dull et al. 1998). In pCCL 305 the EGFP has an internal ribosome entry site and thus is co-transcribed but independently translated. Although Flucs have previously been functionally expressed with C-terminal tags, the effect of tags on ACSM CTDs is unknown and as such this property of pCCL 305 was chosen as a precaution, and its map is shown in Figure 4.17.

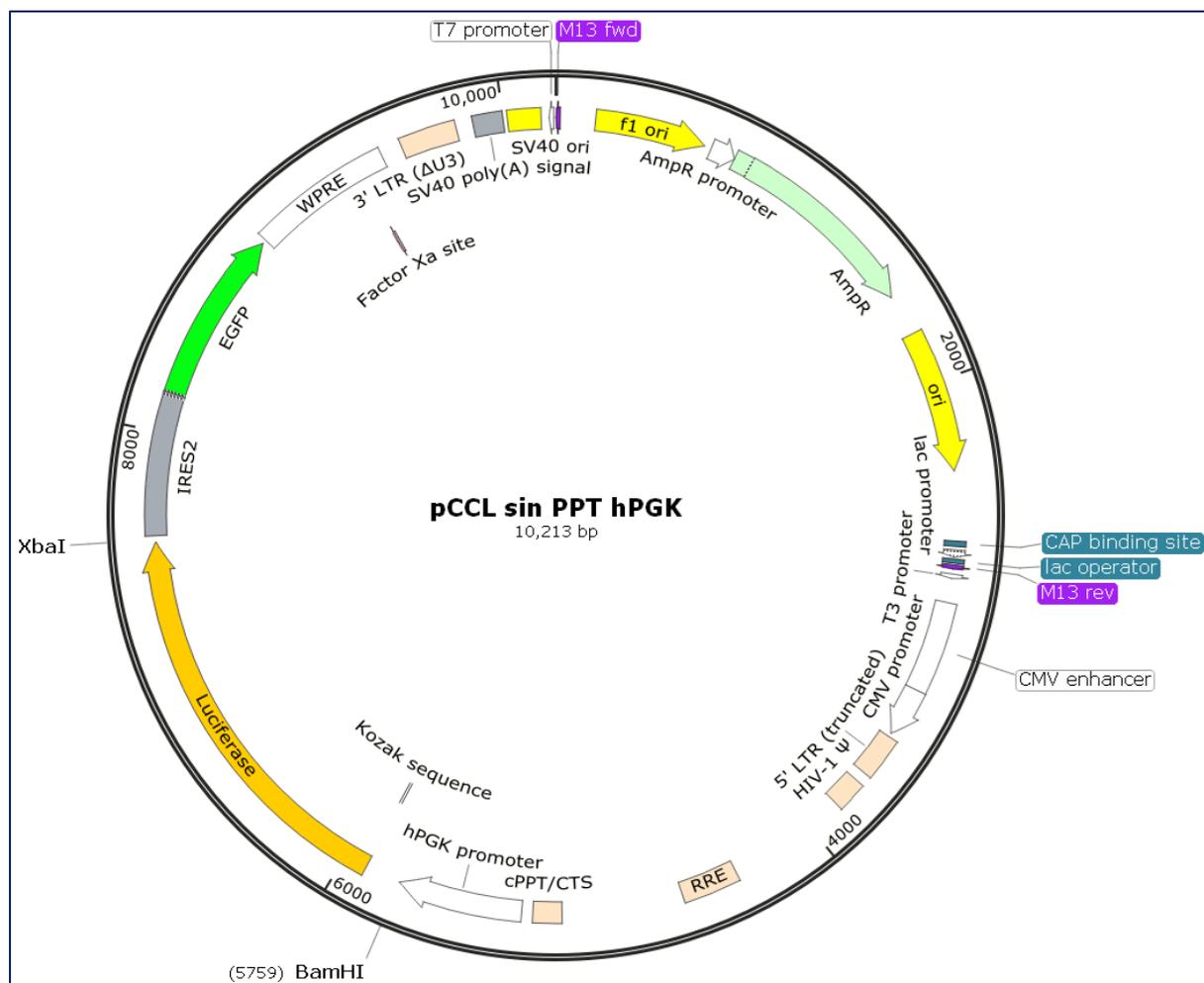


Figure 4.17. Plasmid map of pCCL 305 with a Fluc inserted. IRES2 stands for internal ribosome entry site (second generation), and allows independent translation of the EGFP.

pCCL 305 contains a MCS where inserts may contain a 5' BamHI and 3' XbaI restriction digestion site, as such the designed gBlock oligos needed to contain these. Despite only the CTDs of the ACSMs being required for this chapter, the NTDs and CTDs of ACSM1 and ACSM3 were ordered as continuous oligos, in anticipation of work explained in Chapter 6. The mitochondrial targeting sequences (MTSs) of the ACSMs were removed to facilitate cytoplasmic expression of the new reporters. The PacI restriction site was used for domain exchange of NTDs and CTDs in chapter 3 as this was established as a reliable method in the literature (Branchini et al. 2014). However, the placement of the PacI restriction site includes the first residue of the CTD in *Ppy* Fluc, and the first two residues of the CTD in ACSM1 and ACSM3, resulting in an incomplete domain exchange. Further, this likely means that the 8 aa flexible linker (only 4 aa in ACSM1 and ACSM3) is not being exploited effectively as the linker

is probably most tolerant to non-silent substitutions and splices in its centre – so that there is linker sequence up and down-stream of the splice – to grant the greatest protection against the splice interfering with the more highly-ordered functional domains flanking it. Thus, the *Ppy* Fluc protein sequence was re-examined for potential restriction sites using Resitefinder. HindIII was identified as a suitable candidate as it has a 6 bp recognition site, and is fully within the linker sequence. *Ppy* Fluc could be ordered containing both HindIII and PacI in the linker between the NTD and CTD, and since these restriction sites are introduced silently, their presence when not exploited will theoretically not affect protein expression. However, HindIII and PacI could only be introduced to ACSM1 and ACSM3 linker sequences non-silently (table 4.3), and as such these versions were ordered separately for comparison. Further, HindIII is more optimally positioned but required three (rather than two) mutations for incorporation.

The protein sequences for *Ppy* Fluc, ACSM1, and ACSM3 were entered into IDT's codon optimisation tool and codon optimised for murine expression. Then (for ACSM1 and ACSM3) the MTSs were removed, the HindIII and/or PacI sites introduced into the linker sequence, the deca-His tag and factor Xa site added upstream, and the BamHI and XbaI sites added at the 5' and 3' ends respectively. The inserts ordered are shown as maps in Figures 4.18 – 4.20.

Full name (from NCBI)	acyl-CoA synthetase medium-chain family member 1	acyl-coenzyme A synthetase ACSM3, mitochondrial
Shorthand name	ACSM1	ACSM3
NCBI/GenBank reference code	EDL17174.1	NP_058566.3
Mutations required to introduce PacI site	V467L N469K	I474L L476K
Mutations required to introduce HindIII site	D465K D466S V467L	D472K D473S I474L

Table 4.3. Residue mutations required to introduce PacI or HindIII into ACSM1 and ACSM3.

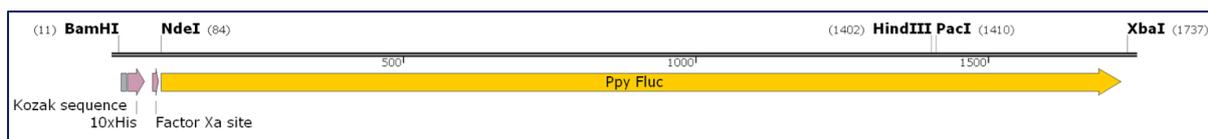


Figure 4.18. Map of gBlock DNA sequence encoding the entire WT *Ppy Fluc* with *PacI* and *HindIII*, for insertion into *pCCL 305*. An expanded version of this figure – with DNA and protein sequence shown – can be found in the appendix (figure 4.18EX).

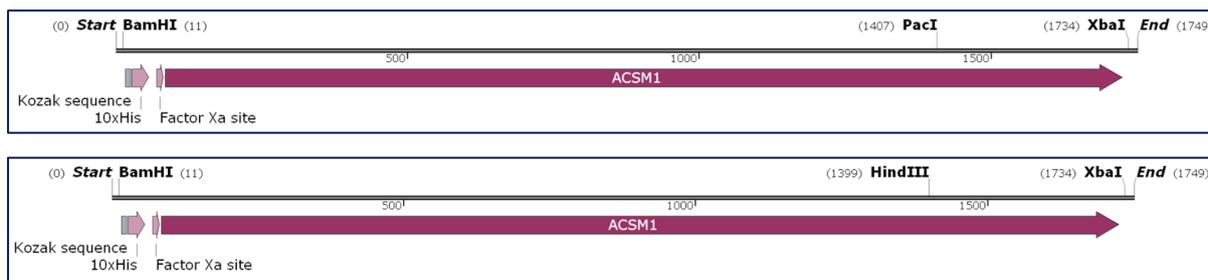


Figure 4.19. Maps of gBlock DNA sequences encoding the entire *ACSM1* with *PacI* or *HindIII*, for insertion into *pCCL 305*. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figures 4.19EXa and 4.19EXb).

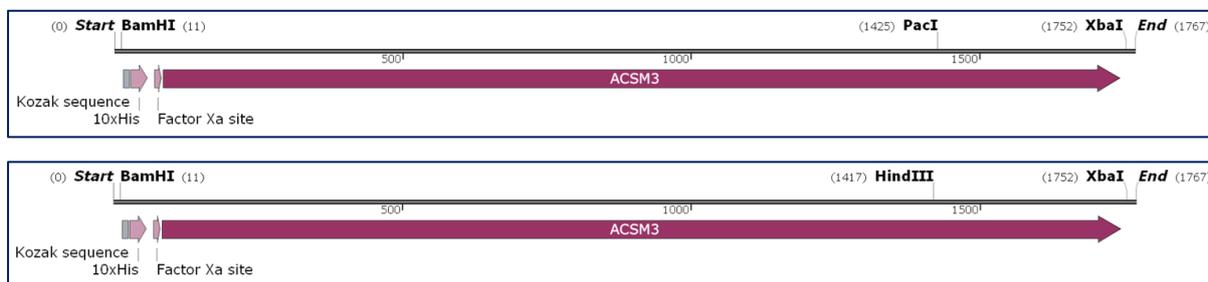


Figure 4.20. Maps of gBlock DNA sequences encoding the entire *ACSM3* with *PacI* or *HindIII*, for insertion into *pCCL 305*. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figures 4.20EXa and 4.20EXb).

4.3.2.2 Cloning

The constructs PpPp, PpA1 (*PacI*), PpA1 (*HindIII*), PpA3 (*PacI*), and PpA3 (*HindIII*) – each in *pCCL 305* – were created via T4 ligation and transformed into DH5 α *E. coli* (see 2.2.1.7 and 2.2.3.3). Colonies were picked, cultured, partly made into glycerol stocks, DNA miniprep, and sequenced (see 2.2.3.4, 2.2.3.5, 2.2.3.1, and 2.2.1.10).

4.3.2.3 BLI screening4.3.2.3.1 BLI screening with D-LH₂

The sequence-verified and highest concentration DNA minipreps for each of the five constructs were transformed in triplicate into HEK cells in 24-well plates (see 2.2.4.6), which were then screened for bioluminescence with D-LH₂ – and acquisitions taken for both light output and spectra (Figures 4.21 and 4.22). These data show that PpA1 and PpA3 are substantially (~10 fold each) brighter relative to PpPp when expressed in HEK rather than *E. coli* (Figure 4.21 vs. Figure 4.15). The spectrum of PpA1 was similar between *E. coli* and HEK, with a λ_{max} of ~620 nm; PpA3 λ_{max} was ~605 nm in *E. coli* and ~620 nm in HEK. However, in HEK, PpA1 was brighter than PpA3 at normalised bandpass filters redder than λ_{max} of ~620 nm. Thus, PpA1 appears superior to PpA3 for in vivo applications both in terms of brightness and colour. The λ_{max} of PpPp was ~560 nm in *E. coli* yet ~570 nm in HEK, perhaps indicating red-shift due to pH differences. The HindIII-spliced PpA1 and PpA3 variants were incrementally brighter than the PacI counterparts, but within the margin of error and have an effectively negligible difference in either brightness or spectra. For consistency with construct splicing anticipated for Chapter 6 (explained within that chapter) the HindIII variants of PpA1 and PpA3 were taken forward, whereas the PacI variants were discarded at this stage in the project.

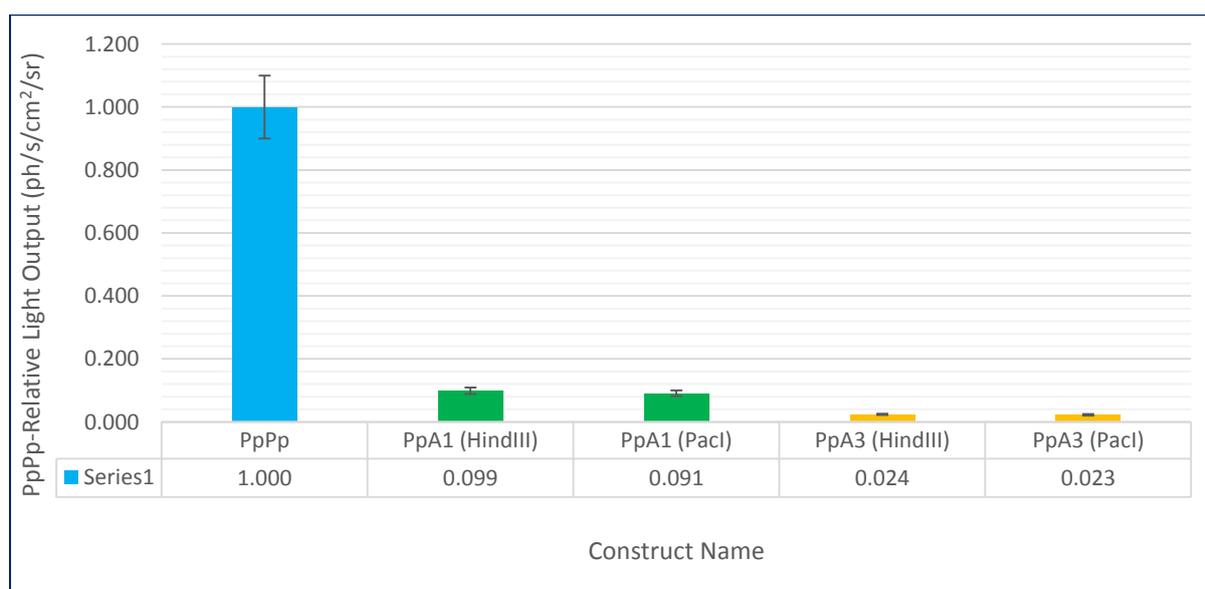


Figure 4.21. PpPp-relative light output of PpA1 (HindIII), PpA1 (PacI), PpA3 (HindIII), and PpA3 (PacI).

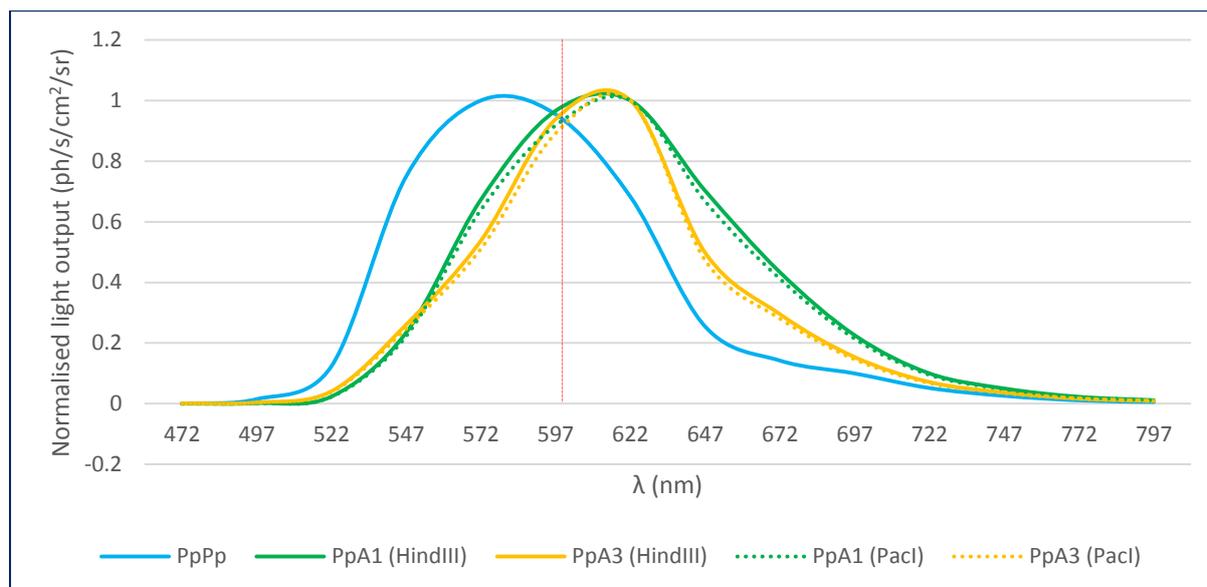


Figure 4.22. Normalised spectra of PpPp, PpA1 (HindIII), PpA1 (PacI), PpA3 (HindIII), and PpA3 (PacI). Emission to the right of the red dotted line (>600 nm) is considered to be within the bio-optical window.

4.3.2.3.2 BLI screening with infraluciferin (iLH₂) and 6-amino-D-luciferin (ALH₂)

PpA1 and PpA3 were screened with iLH₂ and ALH₂ alongside *Ppy* Fluc in the same manner as with D-LH₂ (see 2.2.6.5.2). Light output relative to *Ppy* Fluc was somewhat improved with ALH₂ and furthermore with iLH₂ (Figure 4.23), however absolute light output is greatly reduced relative to D-LH₂ these analogues. Spectra of PpA1 and PpA3 with ALH₂ and iLH₂ were very similar to that of *Ppy* Fluc (Figure 4.24). These results were expected as – unlike D-LH₂ – there are not alternative light emitting pathways for these analogues, precluding λ_{\max} shifting (Jathoul, A., PhD thesis, University of Cambridge [not publically available]).

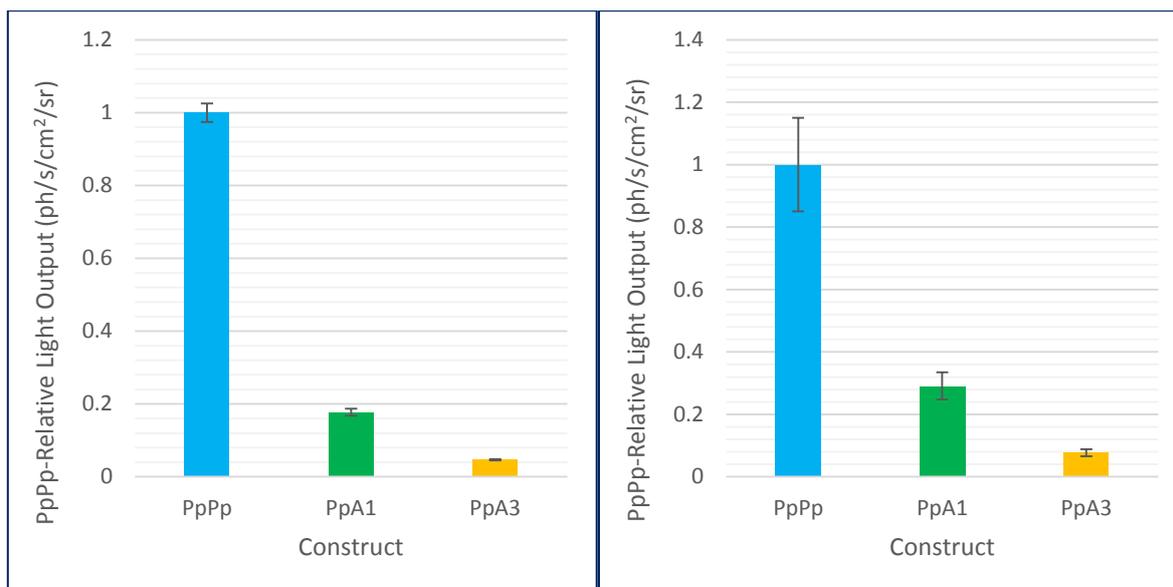


Figure 4.23. PpPp-relative light output of PpA1 (HindIII) and PpA3 (HindIII), with ALH₂ (left) and iLH₂ (right). The large error bars for iLH₂ are likely due to low brightness (low signal-to-noise ratio).

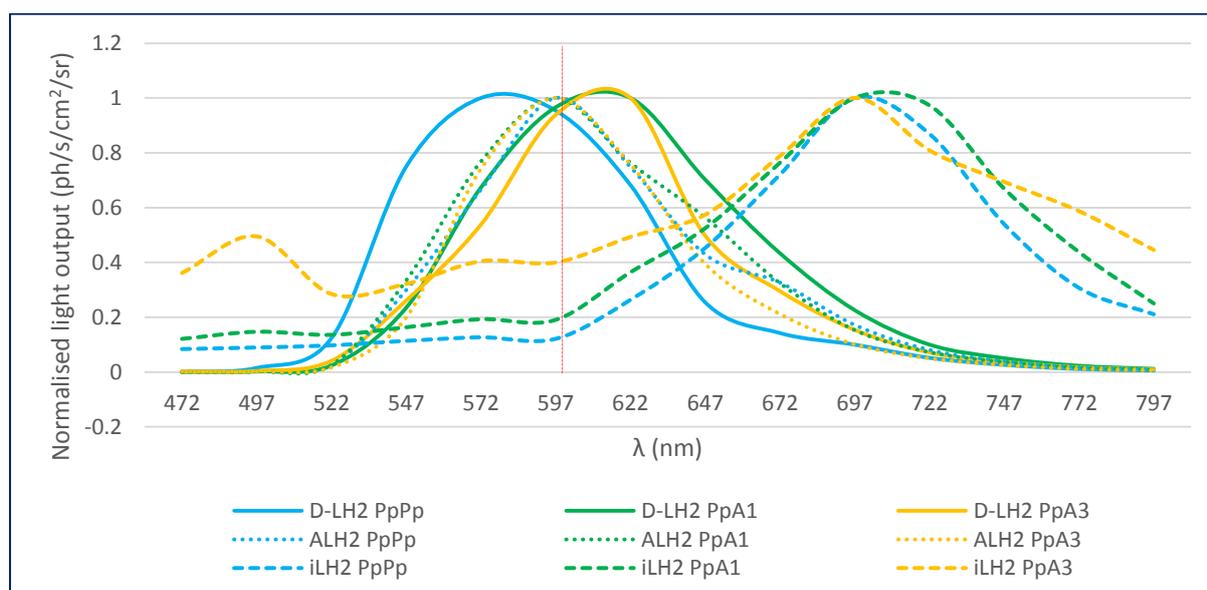


Figure 4.24. Normalised spectra of PpPp, PpA1 (HindIII), PpA3 (HindIII) with D-LH₂, ALH₂, and iLH₂. Emission to the right of the red dotted line (>600 nm) is considered to be within the bio-optical window.

4.3.3 Design, cloning, and screening of the N-terminal P2A-linked mCherry chimera constructs

4.3.3.1 Design

As previously explained, the pCCL 305 plasmid expresses EGFP independently of the construct and can therefore only quantify plasmid abundance and not construct expression. In mammalian cells, the latter is most readily quantified using co-translation of a fluorescent reporter. EGFP remains to fulfil its purpose of normalising for plasmid levels, and the second fluorescent protein for construct quantification must not overlap in terms of emission from excitation. Detection was to be carried out using a Becton, Dickinson and Company Biosciences LSRFortessa, or the Biospace Labs PhotonIMAGER for higher level expression. mCherry was chosen as the fluorescent protein due to its different excitation and emission wavelengths compared to EGFP as shown in Figures 4.25 and 4.26. mCherry, like EGFP, is monomeric and therefore efficiently expressed.

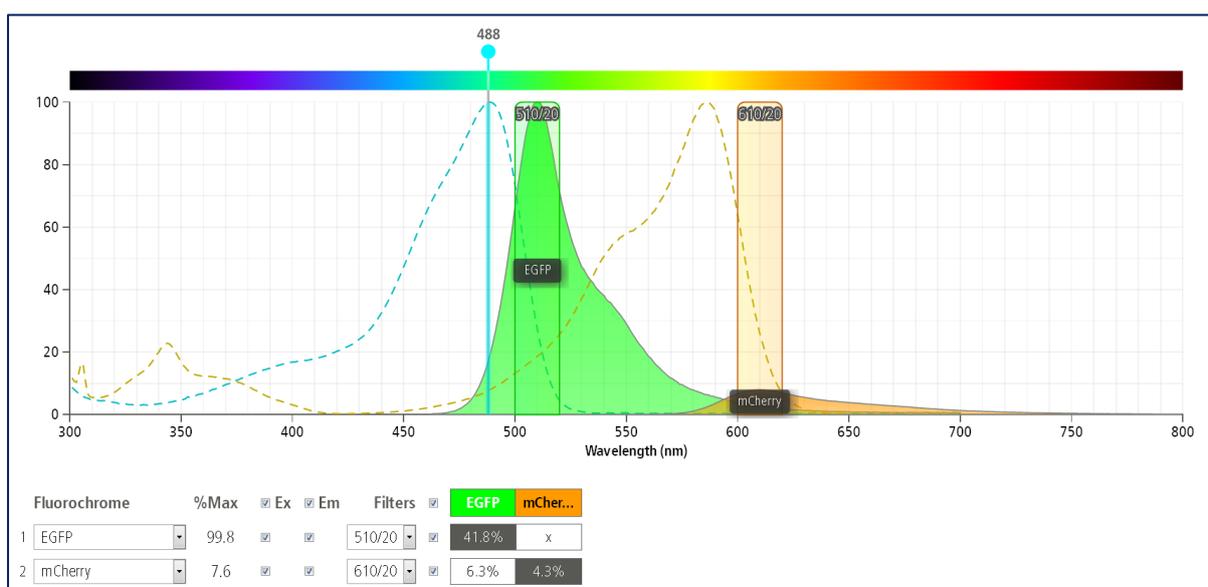


Figure 4.25. EGFP and mCherry absorption and emission spectra with 488 nm laser as available on the LSRFortessa. 488 nm is optimal for EGFP excitation.

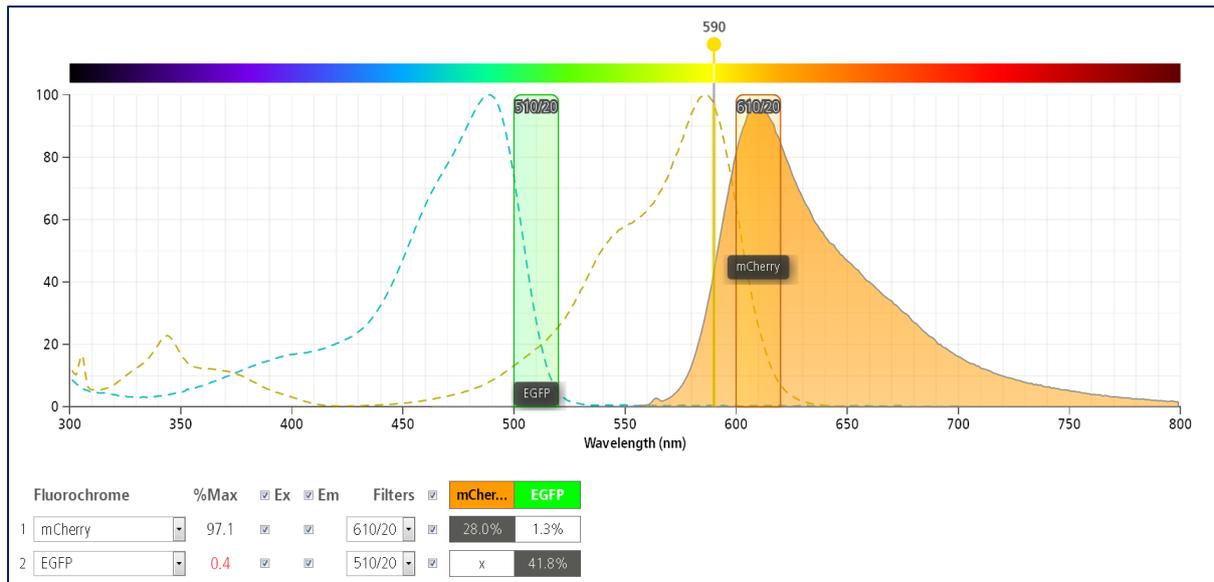


Figure 4.26. EGFP and mCherry absorption and emission spectra with 590 nm laser as available on the LSRFortessa. 590 nm is optimal for mCherry excitation.

The expression of mCherry must be directly linked to that of the construct for quantification and the “2A” self-cleaving peptide system provides a simple solution for 1:1 expression of two proteins. This short (18-22 aa) sequence is placed between the two constructs, and when translated by the ribosome one of the peptide bonds within the sequence fails and the two constructs cleave passively. There are four so-called “2A” peptide variants characterised, namely: porcine teschovirus-1 2A (P2A), *Thoseaasigna* virus 2A (T2A), equine rhinitis A virus (ERAV) 2A (E2A), and foot-and-mouth disease virus (FMDV) 2A (F2A). In a variety of cell and tissue types the P2A peptide consistently has the highest cleavage efficiency, followed by T2A, E2A, and F2A respectively (Kim et al. 2011). All 2A peptides contain the sequence GSG at their N-terminal end, and a NPGP at their C-terminal end; the cleavage site is directly upstream of the final proline. This is displayed in Figure 4.27.

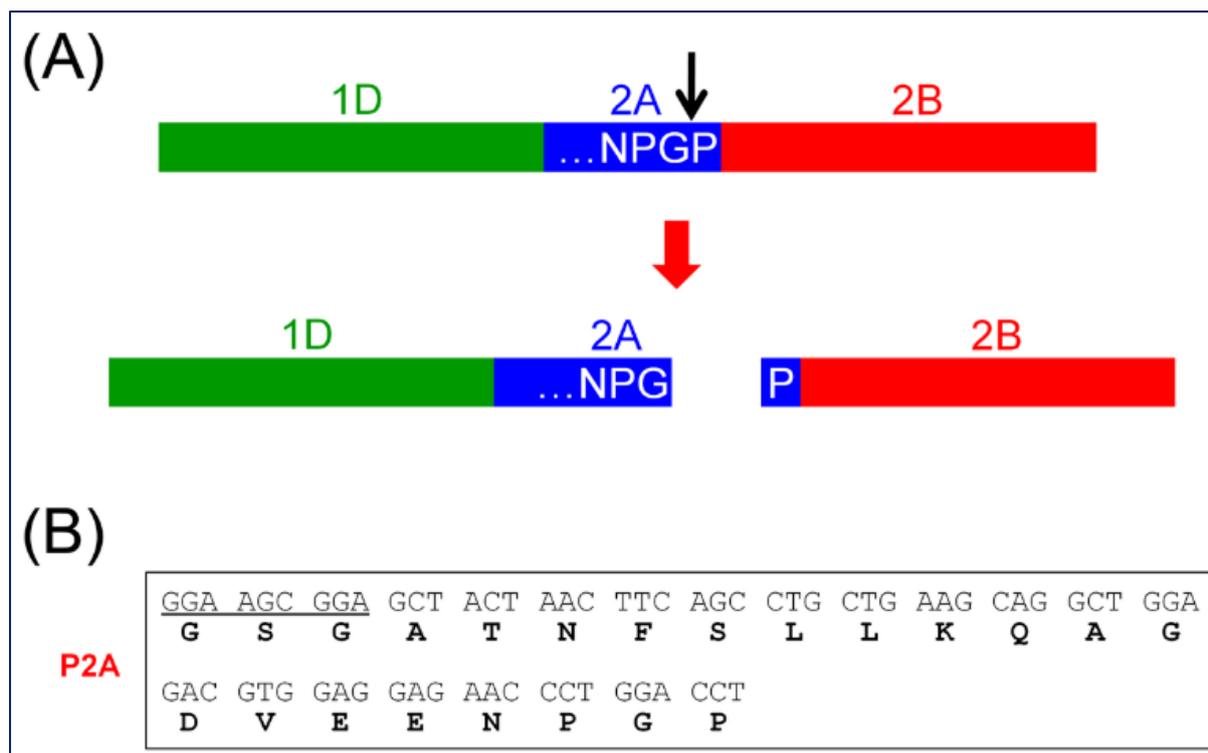


Figure 4.27. Taken from (Kim et al. 2011). Diagram of the 2A system (A) and sequence of P2A (B).

The 2A system, after cleavage, leaves a short peptide on the C-terminus of the upstream protein (21 residues for P2A) and a proline on the N-terminus of the downstream protein (Figure 4.27). Since mCherry is known to be extremely tolerant of terminal modifications it was decided that this protein should take the larger 21 aa peptide, therefore it would be upstream of P2A and construct. The luciferase would already have two N-terminal modifications, the deca-His tag and a Xa cleavage site. The deca-His tag is known to be highly flexible, and therefore the addition of a proline to the N-terminus is unlikely to affect construct function. Secondly, the Xa cleavage site means that the proline – along with the deca-His tag – can be post-translationally removed by the addition of Xa factor protease, leaving a single histidine (the smallest residue possible) on the N-terminus of the construct. The construction began by ordering a mCherry-encoding gBlock oligo which could be inserted using the BamHI site upstream of the Kozak sequence. The NdeI site between the deca-His tag plus Factor Xa site and the construct could not be easily used because it is not unique; a second NdeI site exists in the plasmid backbone. Therefore, since all constructs contain the NTD of *Ppy* Fluc, the same restriction site within this domain could be used for all three

without resorting to using the HindIII site between the two domains. A BstEII restriction site is present approximately halfway through the NTD and this was exploited for insertion of the mCherry sequence. The ordered sequence must itself contain the Kozak sequence upstream, and the deca-His tag plus Xa site downstream in order to retain these features. IDT's gBlock codon optimisation tool was used to codon optimise mCherry for mammalian expression, then the Kozak sequence, Xa factor site, deca-His tag, and *Ppy* Fluc (upstream of BstEII) features were added along with the N-terminal BamHI restriction site. Once assembled in silico the DNA was ordered as a gBlock oligo as shown in Figure 4.28.

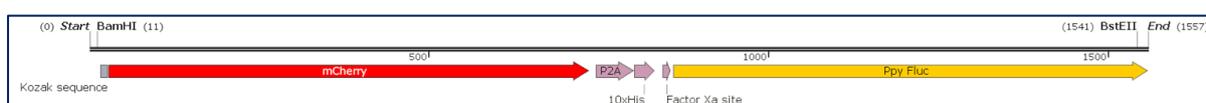


Figure 4.28. Map of gBlock DNA sequence encoding mCherry and P2A for splicing into *Ppy* Fluc in pCCL 305 via the flanking BamHI and BstEII sites. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 4.28EX).

4.3.3.2 Cloning

The constructs PpPp and PpA1 (HindIII) (both in pCCL 305) were created via T4 ligation with the previously cloned non-mCherry versions, and transformed into DH5 α *E. coli* (see 2.2.3.3). Colonies were picked, cultured, partly made into glycerol stocks, DNA miniprepmed, and sequenced (see 2.2.3.4, 2.2.3.5, 2.2.3.1, and 2.2.1.10).

4.3.3.3 Screening

4.3.3.3.1 Bioluminescence screening using the PhotonIMAGER

The sequence-verified and highest-concentration DNA minipreps were transformed into HEK cells and incubated to express the constructs (see 2.2.4.6). Bioluminescence screening was performed (by the addition of D-LH₂) as a precautionary check of enzyme function. These data (Figure 4.29) are very similar to the non-mCherry equivalents shown earlier (Figure 4.21).

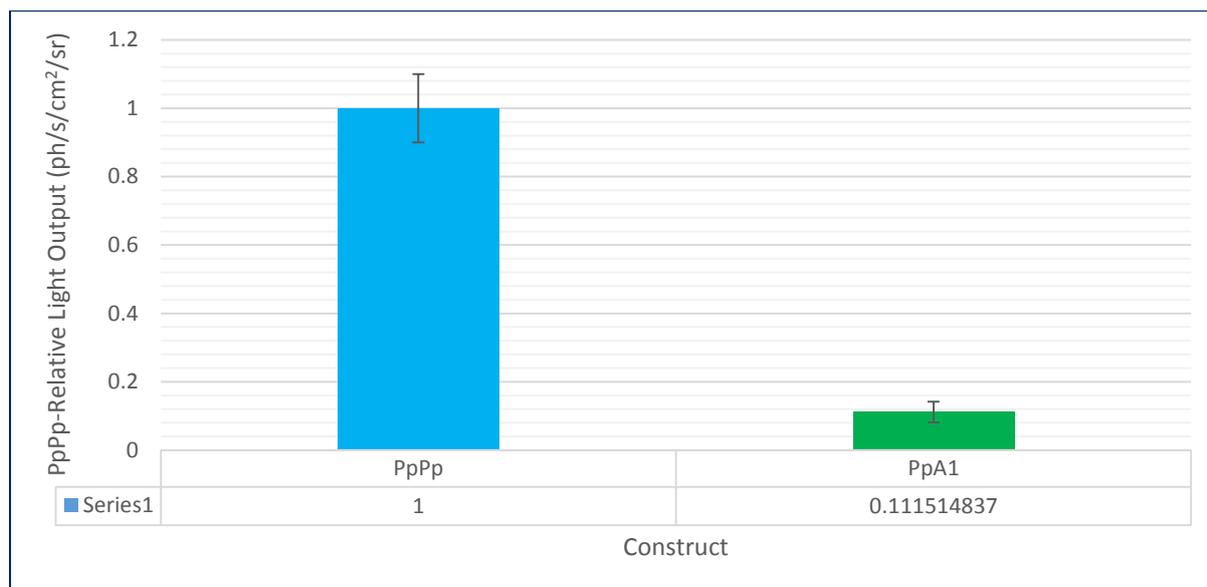


Figure 4.29. PpPp-relative light output of PpA1 (both with N-terminal-linked mCherry).

4.3.3.3.2 Fluorescence screening using the PhotonIMAGER

Prior to D-LH₂ addition, the transformed HEK cells were imaged for EGFP and mCherry fluorescence (see 2.2.6.5.1). As was seen without mCherry, the constructs were positive for EGFP signal, indicating plasmid expression and cell health. However, the mCherry signal was not above background fluorescence levels (data not shown, see Figure 9.1).

4.3.3.3.3 Fluorescence screening via flow cytometry

Due to the poor EGFP (and non-existent mCherry) fluorescence signal when acquired with the PhotonIMAGER, the flow cytometer was used for increased sensitivity. As the EGFP signal is stronger and more reliable, it can be used to differentiate positive from negative transformants independent of construct expression efficiency. Of those positive for EGFP (via FITC laser), the mCherry signal can be quantified. Although the LSRFortessa is capable of explicit mCherry quantification, the option for this in the acquisition software could not be identified even with the assistance of the available flow cytometry technician. As a substitute the settings for the PE-Texas Red[®] fluorophore were used, as this is commonly exploited by researchers using mCherry who do not have a flow cytometer explicitly capable of detecting it. The settings for PE-Texas Red[®] use a 568 nm laser rather than the 590 nm laser ideally used for excitation of mCherry. However, the emission filter of 610/20 nm is the same as mCherry. The 568 nm laser results in a 72.0% optimal excitation of mCherry, with the ideal

590 nm laser giving 97.1%. Since the emission filter is unchanged the proportion of the emitted light from mCherry which passes through the filter remains at 28%, and the leakage from EGFP remains at 1.3% (Figure 4.30).

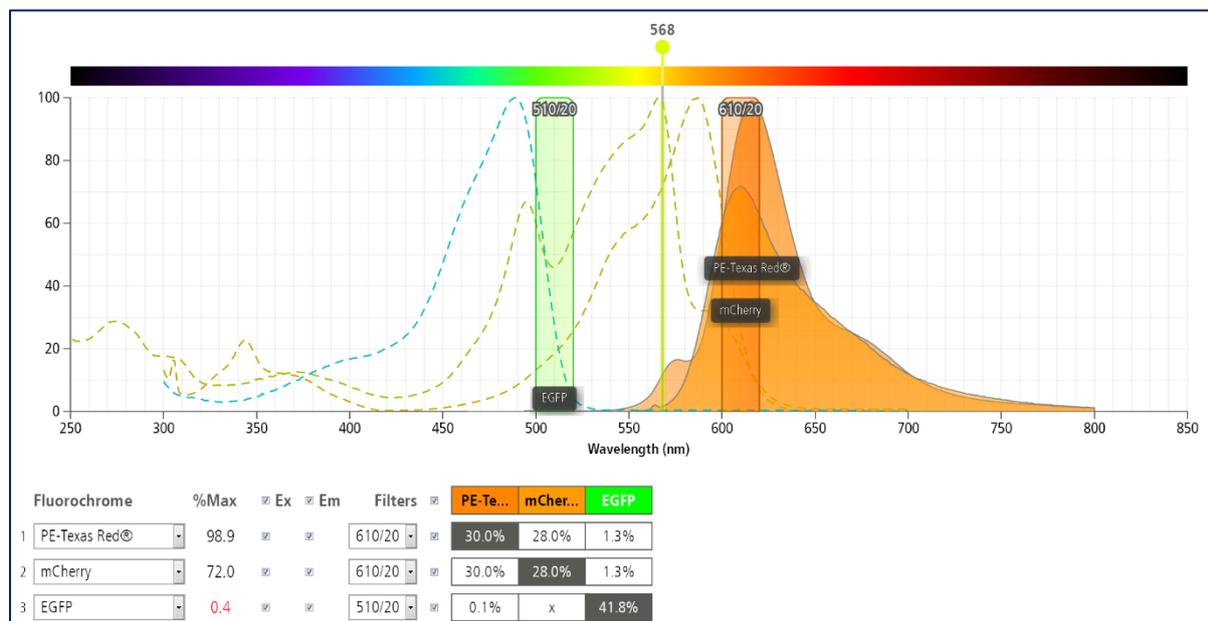


Figure 4.30. EGFP, mCherry, and PE-Texas Red® absorption and emission spectra with 568 nm laser as available on the LSRFortessa. 568 nm is optimal for PE-Texas Red® excitation.

The HEK cells were run through the LSRFortessa in triplicate (see 2.2.4.4.3) and the results analysed and exported using BD FACSDiva™ software. Specifically, live and single cells were plotted as a scatter plot with FITC-A on the y-axis and PE-Texas Red® on the y-axis. Then, the NT plot was used to set the quadrant gating threshold such that virtually all NT cells were in the lower left quadrant (Figure 4.31 i.). The cells within the top right quadrant were analysed for the average ratio between FITC and PE-Texas Red®, this was normalised for PpPp (Figure 4.32). These data show a substantial increase (~2.5 fold relative to Ppy Fluc) in protein expression in HEK for constructs containing the CTD of ACSMs (PpA1 and PpA3). Thus, the Ppy Fluc-relative specific activity of PpA1 is less than the ~10% implied from Figure 4.21 and 4.29, as controlling for expression reduces this number to ~4%.

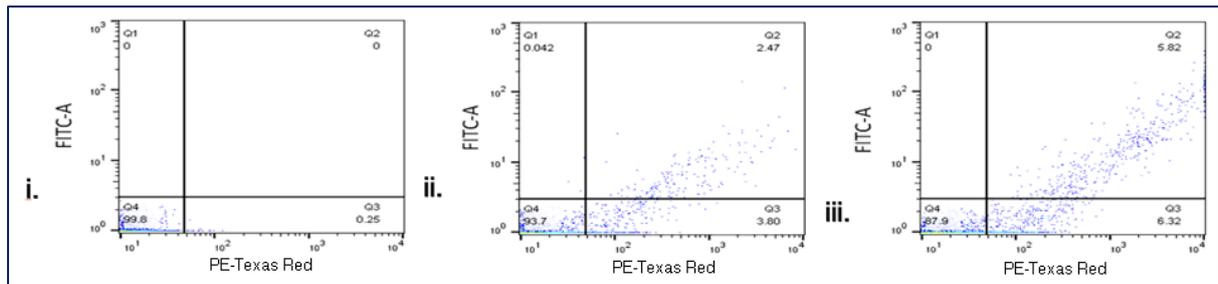


Figure 4.31. Quadrant-gated scatter plot of HEK cells for FITC and mCherry fluorescence. i. non-transformed, ii. *Ppy* Fluc (PpPp), iii. PpA1.

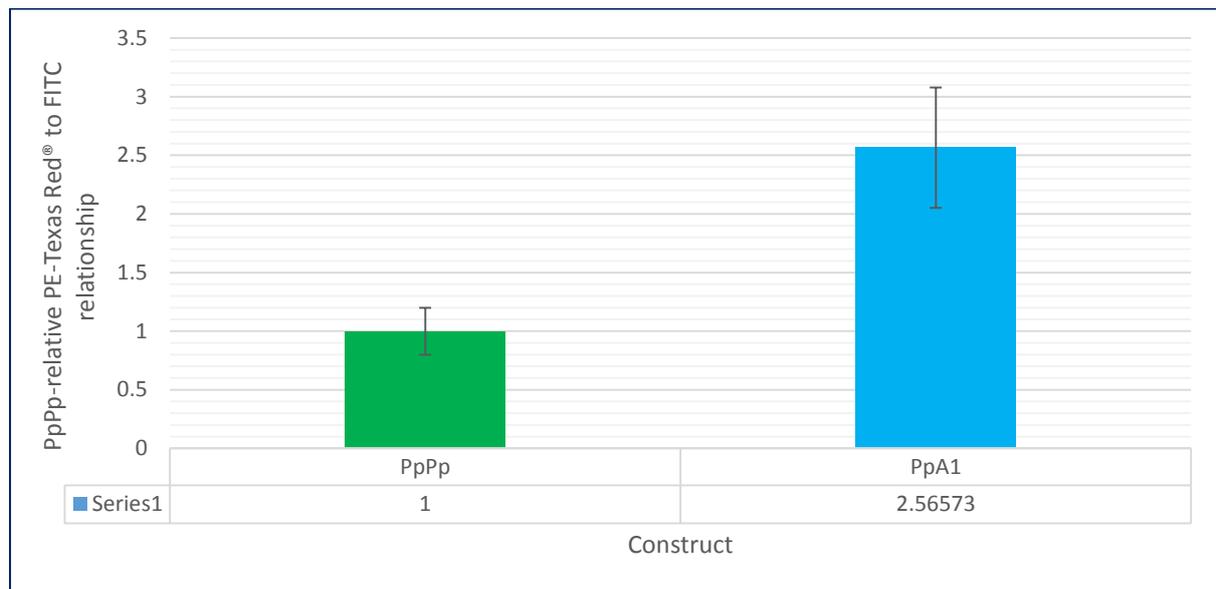


Figure 4.32. mCherry mean fluorescence indexes of FITC-positive: non-transformed, *Ppy* Fluc, and PpA1 transformed cells.

4.4 Further Discussion

The second section of work (Chapter 4) sought to address whether *Ppy* Fluc and murine ACS could be domain exchanged while retaining bioluminescence function, as was previously achieved in the literature with other insect ACS (Oba et al. 2006). Chimeric constructs were created using three murine ACSMs and *Ppy* Fluc. Those with the NTD of *Ppy* Fluc retained bioluminescence function, however to varying degrees – in HEK cells PpA1 was brighter than PpA3 (~10% vs ~2.5% respectively).

4.4.1 Would ACSLs be more suitable for hybridisation with Ppy Fluc?

Both ACSM1 and ACSM3 are technically *medium*-chain acyl-CoA synthetases, but based on optimal substrate size ACSM1 is closer to being an ACSL, and ACSM3 closer to being an ACSS. Given this information, it suggests that using a longer-chain ACS for the CTD of the chimera may improve brightness further. This would conform to the literature consensus that Ppy Fluc evolved from an ancestral ACSL. A possible mechanistic explanation for the superiority of longer-chain ACSs is that the binding sites on both the NTD and CTD are larger and more promiscuous, with greater tolerance for foreign substrates. Unfortunately there is limited information in the literature on murine ACSs, however their approximate chain length category is often known phylogenetically. Ideally, a mixture of ACSLs and ACSMs (and possibly ACSVLs) would have been screened to ascertain the optimal chain-length range for the CTD. This could be inferred for the NTD, and the optimal whole enzyme could then undergo the mutations detailed in Chapter 3.

4.4.2 Potential explanations for increased PpA1 expression relative to Ppy Fluc

The mCherry protein quantification of PpA1 and Ppy Fluc revealed a ~2.5 fold expression efficiency advantage of PpA1 over Ppy Fluc. The potential reasons for this are discussed here. Proteins can possess what are known as toxic folds which severely limit protein expression efficiency, such a feature was found in GFP (Crameri et al. 1996). Ppy Fluc has been subjected to intense mutagenesis by many researchers since its discovery, and thus such a toxic fold would likely have been discovered in the form of markedly brighter mutants.

However, the mCherry quantification work fails to fully account for the tenfold increase in brightness for PpA1 between *E. coli* and HEK expression. One explanation is simply that the overall structure of the CTD of ACSM1 is more suited to the mammalian cytoplasm, in terms of optimal pH and solubility. HEK cell cytosol has a pH of ~7.3, which is not optimal for Ppy Fluc which is optimal at pH 7.8. This is one of the advantages (and reasons for creating) one of the more pH stable mutants of Ppy Fluc. *E. coli* by contrast have less tightly controlled cytoplasm and can range from pH 7 – 8. ACSMs are mitochondrial enzymes, and mitochondrial pH is ~8.0, however ACSM1 pH stability is unknown outside of this. The pH dependence of PpA1 was not assayed, and such an experiment may elucidate this phenomenon. Another explanation is that PpA1 is more thermostable than Ppy Fluc, and

HEK-expressed constructs were expressed for a longer duration (24 – 48 hrs) pre-screening, whereas *E. coli* IPTG induction is only 3 hrs. Thus, the any thermostability advantage is magnified. Assays exist for quantifying the thermostability of luciferases and have been used in the development of mutants such as x11 Fluc, but only when expressed in *E. coli*, and it is uncertain how the *E. coli* physiological environment affects the stability of these novel constructs. Indeed, many characteristics of these novel chimeras are unexplored (pH dependence, k_m , k_{cat} , etc.) since as this was an engineering project with a defined end-goal, the details of the intermediate constructs were considered likely unhelpful in advancing the engineering.

4.4.3 Non-optimal ACSM1 used for hybridisation

The ACSM1 protein identified in the literature (Fujino et al. 2001) has a mitochondrial form available (NCBI Reference Sequence: NP_473435.1) which would be more in-line with the ACSM2 and ACSM3 variants used within Chapters 3 and 4, however this was unfortunately not noticed at the time of selection, only upon writing this thesis. For ACSM1, the difference between the non-mitochondrial and mitochondrial isotypes is the modification A478V; the effect of this modification is unknown and it occurs within the CTD which thus affects the sequence of the functional chimeras created from it (PpA1), and A1PpPpPpPpPpA1, and PpPsA1 for Chapter 6.

4.5 Conclusions

Although the ACSM NTD-containing constructs were not exhaustively screened for bioluminescence function dimmer than what is detectable as induced colonies; the potential for luciferases containing ACSM regions within the NTD was intended to be explored later (see Chapter 6) using the optimised expression via HEK cells identified within this chapter. The HEK cell work within this chapter demonstrated that chimeras containing the NTD of *Ppy* Fluc and the CTD of ACSM1, 2, or 3 were found to retain bioluminescence substantially brighter than *Ppy* Fluc with no CTD; this indicates that the ACSM CTSs serve a useful function. Brightness of PpA1 and PpA3 was increased tenfold relative to *Ppy* Fluc when expressed in HEK rather than *E. coli*. The brightest, PpA1, was found to be ~10% as bright as *Ppy* Fluc in HEK and was thus presumed to be viable for small animal imaging. Further, the λ_{max} of PpA1

was red-shifted with a greater proportion of emitted light within the bio-optical window. The increased *Ppy* Fluc-relative brightness of PpA1 in HEK is partly explained by the increased expression, 80% to 257% that of *Ppy* Fluc in *E. coli* and HEK respectively, this is not necessarily disadvantageous as increased expression efficiency in vivo is useful. What accounts for the remainder of the increased expression remains unexplored, but may be due to improved thermostability. These results are for when screened with D-LH₂; the low-brightness compatibility of PpA1 and PpA3 with iLH₂ and ALH₂ precludes in vivo use, but opens the possibility for further optimisation with these analogues. Further discussion can be found within Chapter 7. The relatively high *Ppy* Fluc-relative light output of PpA1 (~10%, but likely much high when Hb filtered due to red-shifted emission) met the threshold stated in the project Aims. Thus, PpA1 appeared viable as an in vivo genetic marker and was progressed into the in pilot in vivo phase of the project.

5 Chapter 5 – Creation and In Vivo Testing of Virally-Transduced HEK 293 Cell Lines of *Photinus Pyralis* / Murine Acyl-CoA Synthetase 1 Hybrid Construct

5.1 Chapter Summary

In this work, HEK cell lines were created by transduction with a lentivirus expressing the *Ppy* Fluc and PpA1 constructs (detailed in Chapter 4). These were sorted for expressing cells using flow cytometry prior to implantation in the rear flanks of NSG mice to test in vivo light emission. It was found over a ~20 min period post intraperitoneal injection of D-LH₂ that PpA1 emitted ~45% of the relative emission of *Ppy* Fluc, far greater than observed in vitro. Thus, PpA1 – like *Ppy* Fluc – is viable as a potential in vivo genetic marker, although potential immunogenic advantages remain unexplored.

5.2 Overall Introduction (animal ethical considerations and experimental design)

From the work of chapter 4, PpA1 was selected as the prime candidate construct for in vivo screening, followed by PpA3 and PpA2. In vivo screening is dramatically more expensive than mammalian cell culture work, and carries with it ethical concerns. Indeed, this PhD project is funded by the NC3Rs, a body whose sole purpose is to minimise animal suffering in research. Thus, two efforts were made to minimise the use of animals for this chapter. First, only one construct would be screened against *Ppy* Fluc, namely PpA1 as it was the best-performing construct created at this point in the project. Second, this chapter would represent a pilot experiment with reduced cohort sizes; this is precautionary against experimental design flaws and errors resulting in the needless use of large numbers of animals, and also allows the final in vivo experiment to be more finely tuned.

The NC3Rs has an online tool called the Experimental Design Assistant (EDA) (Percie du Sert et al. 2017), the purpose of which is to allow researchers to use minimal cohort sizes of animals in their experiments while maximising the statistical power. However, this tool is designed to plan experiments where the researcher desires to know whether there is a statistical difference between a control and an intervention treatment – for instance a drug.

Because the EDA appeared inappropriate for planning the experiment, I arranged a meeting with Dr Jessica Eddy (the NC3Rs regional programme manager for Cardiff and the South West of England) to discuss how to determine cohort sizes. It was confirmed that the EDA could not be used, and that we should fall-back onto the industry standard of a cohort size of $n = 8$. However, this would be increased to $n = 9$ to allow for unexpected mouse death, and/or removal of mice from the experiment for other reasons such as implant misinjection. Thus, for this pilot experiment I decided to use a reduced cohort size of $n = 9$ ($8 + 1$) for PpA1, and $n = 3$ ($2 + 1$) for PpPp – since this was positive control and the increased brightness would ensure better statistical uniformity.

The value of sorting cells prior to implantation was also to be roughly assessed by the inclusion of two mice, each to have either bulk-transduced PpPp or PpA1 implanted. This could inform how aggressively to sort cells for future experiments.

Although the NC3Rs' EDA could not be used to optimise the experiment, the ARRIVE guidelines could (Kilkenny et al. 2010). These are a set of practices for minimising animal suffering whilst maximising experimental reliability. These were followed in the planning and execution of the experiment, and the processing of data.

Transiently transfected cells carry plasmids that do not integrate into their genome. Since these are lost over time such cells are therefore unsuitable for many in vivo experiments for which cells may be imaged many days post-implant. Therefore, cell lines in which the construct is integrated into the genome of the cells are better suited as the cells can grow and divide theoretically indefinitely. Generally, transduced cells grow at half the rate of those not transduced. Therefore, transduced cells within a mixture of non-transduced cells will eventually be lost. This is why heterogeneous cell lines must be used relatively quickly once produced (on the order of days / weeks) depending on what percentage positive it is, and once implanted the experimental time is limited. Cell lines can be sorted on a flow cytometer to increase the percentage of positive cells by orders of magnitude. Sorting is often based on the use of a marker for expression, and in the case of pCCL 305 it would be the IRES EGFP.

The lentiviral particles integrate the pCCL 305 plasmid into the host cell's genome randomly, and as such some cells may have multiple copies of the construct gene, some only one, and some none. Depending on the loci of the inserted gene the expression can vary. Due to these factors, even cells which are 100% transduced will display a broad heterogeneity of construct gene expression. Multi-cell sorting can dramatically extend the usability of transplanted cells, however due to the aforementioned heterogeneity the low-expressing cells will grow faster than the high-expressing ones. One way of partially rectifying this issue is to only sort cells of a defined expression threshold, and another is to establish cell lines from individual sorted cells. The latter is somewhat analogous to picking *E. coli* colonies, as positive cells are each sorted into the wells of a plate and incubated, and then the cells which survive and grow can be further grown for implantation. Such a cell line is homogenous and can be grown/implanted indefinitely.

The process of creating cell lines with both systems is covered in Chapter 2 (see 2.2.4.7).

Although PpA1 appears to perform better with iLH₂ than D-LH₂ (in terms of red-shifted output and marginal Ppy-relative brightness), iLH₂ light output is already known to be dimmer than D-LH₂ when used in mammalian imaging (Jathoul et al. 2014), and for this reason only D-LH₂ was used for the in vivo experiment within this chapter. Further, D-LH₂ – unlike iLH₂ – is widely used within research, thus increasing the relevancy of this work for other researchers.

Mice were not shaved to be shaved prior to imaging, fur is known to increase light scatter. Red light is scattered to a lesser degree than blue, therefore the redder emission from PpA1 may have an advantage. The tissue of mice is also opaque due to the presence of collagen and other light scattering tissues. Researchers generally will shave mice as it allows for greater spatial resolution of the implant beneath. For the experiment within this thesis this step was to be omitted to minimise trauma to the mice, and because spatial resolution was not of interest – only light output and spectra.

5.3 Results and Discussion

5.3.1 Creation and preparation of virally-transduced cell lines for implantation

The PpPp and PpA1 pCCL 305 constructs (created within Chapter 4) were transduced into HEK cells via the third generation lentiviral system (see 2.2.4.7). The method was tripled in size for PpA1 as there were three fold the number of mice to implant. After this, cells were assessed and multi-cell sorted via FACS (see 2.2.4.4.4); the positively-sorted gate was set as the highest 2% of cells (by FITC signal) for PpPp. This PpPp-derived gate was then used to sort both the entirety of the transduced PpPp and PpA1 cells until only a small volume remained in the unsorted source tubes. Cells are in a weakened state after sorting and thus cannot be directly implanted, and it is unfeasible to sort tens of millions of cells necessary for implantation. Thus, sorted cells were reflasked and grown (with subculturing) for seven days. Samples of cells were assessed via the Luna cell counter for viability and the LSRFortessa for FITC (EGFP) signal (Figures 5.1 and 5.2)

Two days prior to this, NT HEK cells were lentivirally bulk-transduced (see 2.2.4.7) (without FACS) with PpPp and PpA1 in pCCL 305. Samples of NT and transduced cells were assessed via the Luna cell counter for viability (table 5.1) and the LSRFortessa (Figures 5.1 and 5.2) in parallel with the sorted cells above.

It was found – as was found for HEK transfections in Chapter 4 – that PpA1 has a greatly higher EGFP signal than PpPp when directly transfected. However, the sorting removed this advantage, and sorted PpPp had marginally increased EGFP signal than PpA1.

Cells were implanted into mice within 4 hrs of assessment.

Construct	Total number of live cells (millions)	Viability (live cells divided by total cells)
NT	5.41	98.4%
PpPp bulk-transduced	5.55	96.5%
PpPp multi cell sorted (FACS)	14.25	97.2%
PpA1 bulk-transduced	6.20	81.7%
PpA1 multi cell sorted (FACS)	45.5	89.2%

Table 5.1. State of HEK cells for implantation in terms of number and viability.

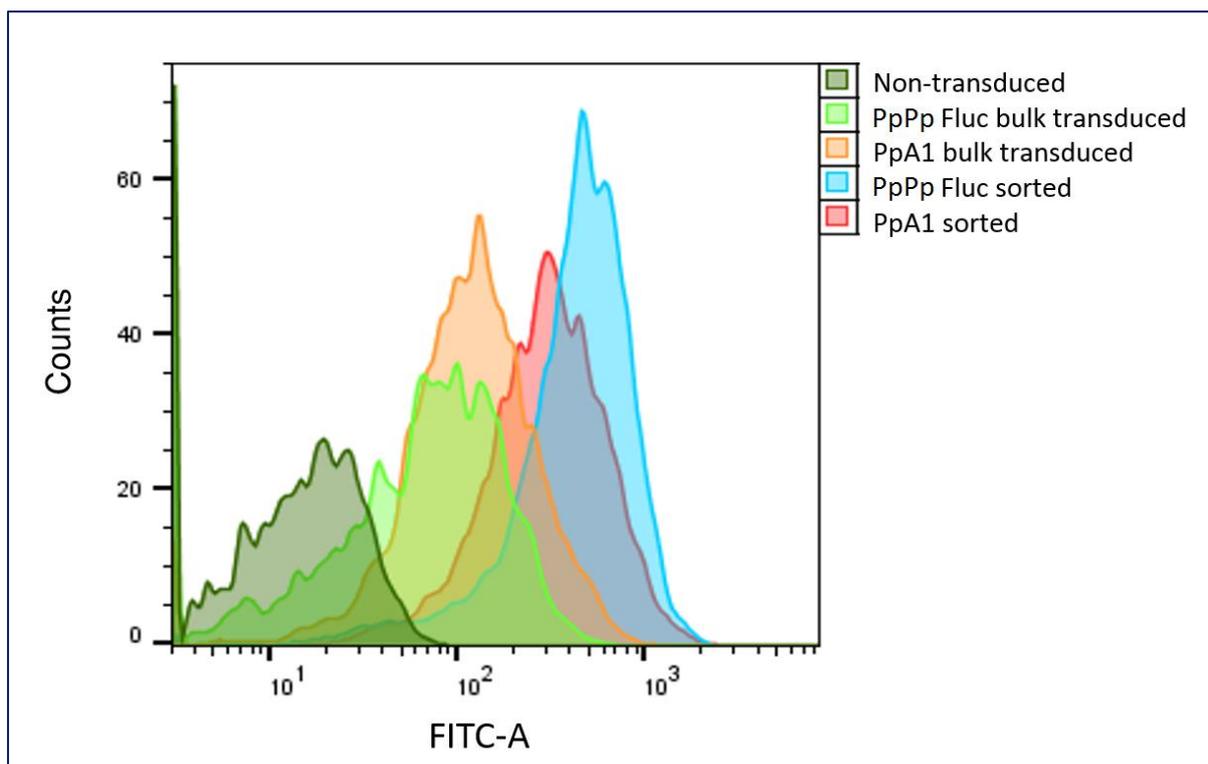


Figure 5.1. FITC histogram of HEK cells for implantation. Note that the vast majority of NT cells are zero FITC-A signal, those positive are likely due to PMT noise.

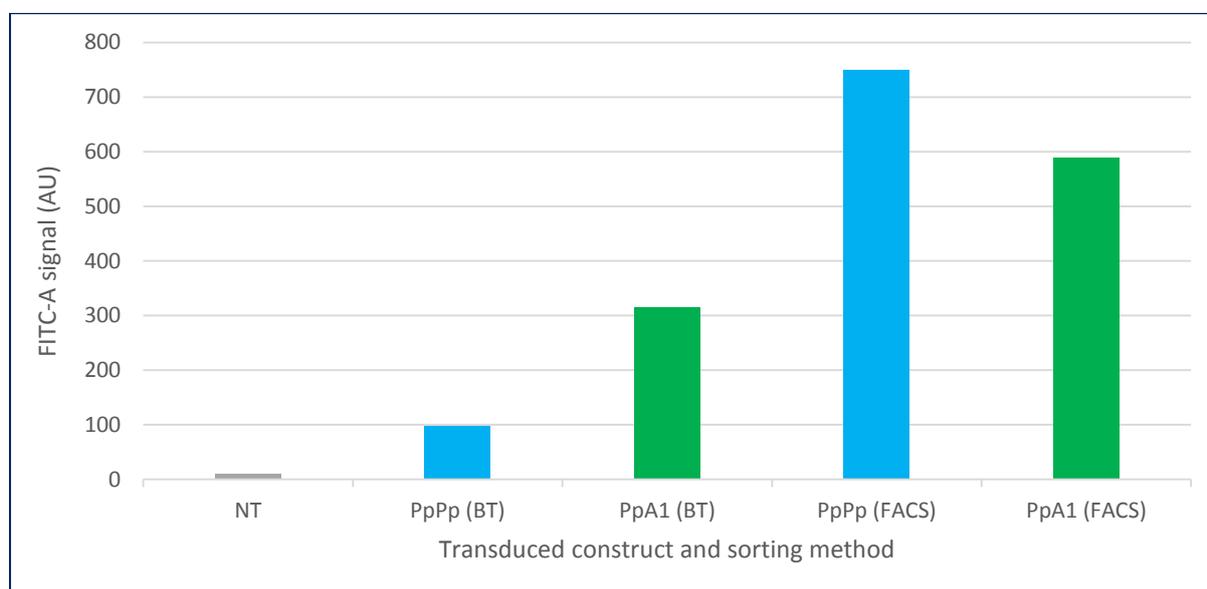


Figure 5.2. Mean fluorescent indexes of HEK cells for implantation.

5.3.2 Implantation of HEK into NSG mice

Matrigel™ is a gelatinous protein mixture produced by Corning, it is designed to represent the extracellular environment found in animal tissue. The high concentration variant is liquid at

4 °C, but rapidly polymerises and therefore solidifies at 37 °C. It is ideal for the implantation of cells into lab animals, since the cells can be injected while liquid but are prevented from seeping back out of the injection site by the rapidly-solidified Matrigel™.

The detailed Matrigel protocol can be found in Chapter 2 (see 2.2.5.2). Each implant was to contain 5×10^6 cells, however there were insufficient FACS PpA1 cells for 9 implants and thus all implants were reduced to 4.75×10^6 cells. Prior to mixing with Matrigel, the cells were resuspended in complete media to concentrations of 47.5×10^6 cells per mL and the Matrigel syringes prepared.

The mice were previously tail-marked (see 2.2.5.1) and these unique marks were now randomly associated with each implant via a written legend. This experiment was performed with my mentor (Dr Amit Jathoul, also Home Office licensed) which allowed double-blinding of the experiment, since he created the legend and kept it secret from me until after I completed the experiment (including data processing). The mice were then implanted (see 2.2.5.2). The mice were incubated in the positive isolator for 7 days until imaging.

5.3.3 FLI EGFP and BLI screening of implanted NSG mice with D-LH₂

The mice of cage A were injected with D-LH₂ and an acquisition performed for 20 mins (1,200 secs). The 20 min-long acquisition was chosen as it represents a time period which is approximately similar to that of many other in vivo researchers using the PhotonIMAGER. Further, it was assumed that a similar kinetic to that observed with PpA1 in vitro would be observed in vivo – with a brief peak light emission, followed by a sharp decline within 10 mins. The mice were imaged in groups of five. As there were three cages, each containing five mice, the cages of mice were screened in groups in the order of cage A, followed by cage B, followed by cage C. Once the acquisition of a group of five was completed the mice were killed whilst still unconscious via cervical dislocation and death was confirmed via palpation of the space between the skull and the spine. The aforementioned procedure was eventually repeated for the three sets of five mice. Although the identity of the implants in the mice was not known at this stage, it was noticed after briefly examining the acquired data for the mice of cage A that the light output of mouse number 3 was still increasing at the end of the acquisition, and that mice 4 and 5 had only begun to plateau shortly prior to the end of acquisition. It was not

possible to continue the acquisition as a meaningful amount of time had already passed since ending the acquisition, and the mice had already been sacrificed. Instead, the acquisition time was increased to 30 mins (1,800 secs) for the screening of the mice in the remaining cages B and C.

Once all mice had been screened the double-blinding legend was revealed to me (table 5.2). Bright-field images of the mice with BLI superimposed, with and without ROIs (and emission values) placed, can be seen in Figures 5.3 – 5.5

Cage letter	Mouse number	Construct	Sorting method	Misinjected implant?	Imaging time
A	1	PpA1	FACS	No	1200s
A	2	PpA1	FACS	No	1200s
A	3	PpPp	FACS	No	1200s
A	4	PpPp	FACS	No	1200s
A	5	PpPp	FACS	No	1200s
B	1	PpA1	FACS	No	1800s
B	2	PpA1	FACS	No	1800s
B	3	PpA1	FACS	No	1800s
B	4	PpA1	FACS	No	1800s
B	5	PpA1	FACS	YES	1800s
C	1	PpA1	Bulk-transduced	No	1800s
C	2	PpA1	FACS	YES	1800s
C	3	PpPp	Bulk-transduced	YES	1800s
C	4	Non-transduced	n/a	YES	1800s
C	5	PpA1	FACS	No	1800s

Table 5.2. Identity and implant status of each mouse, and the imaging acquisition time.

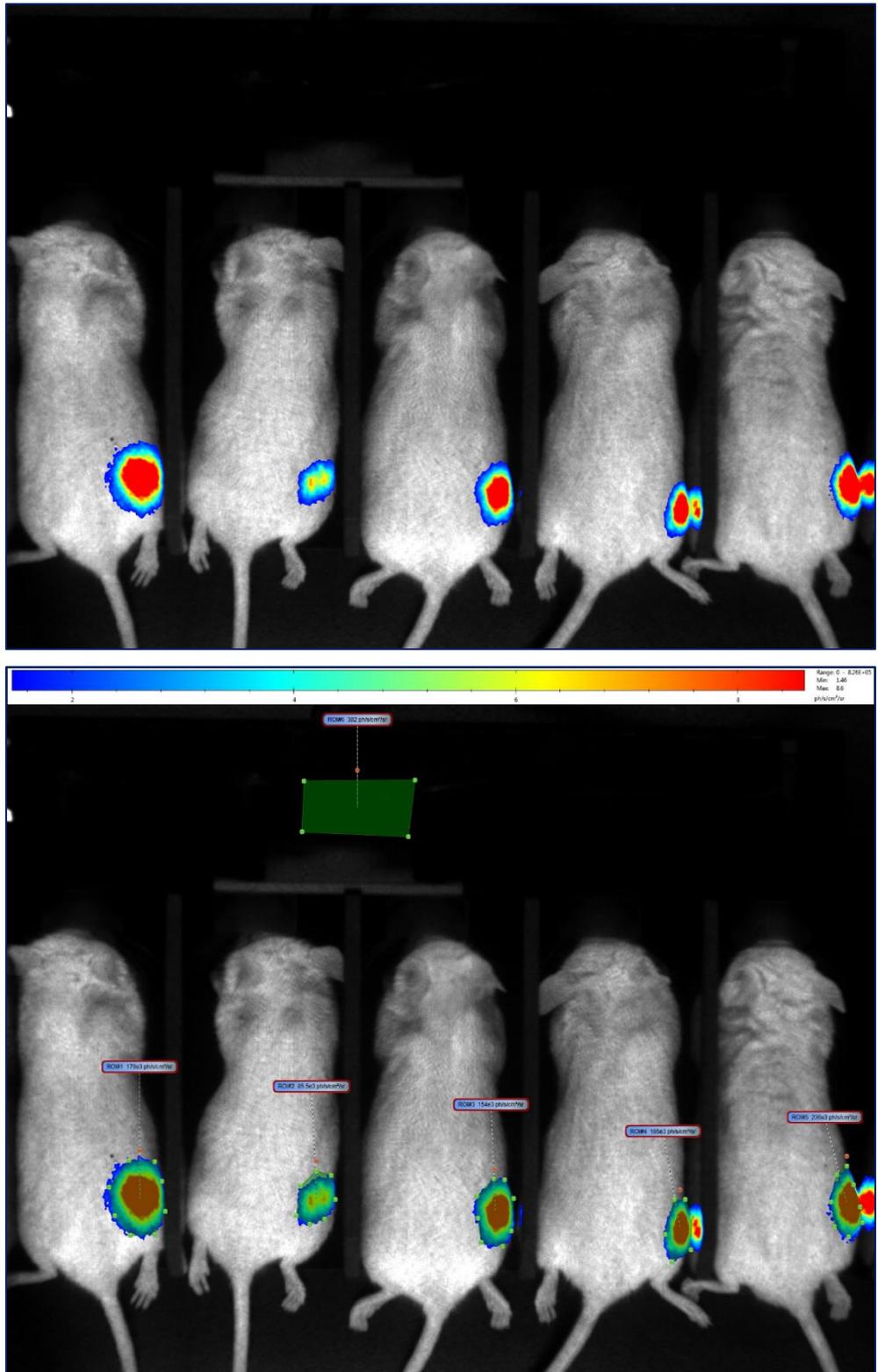


Figure 5.3. $d\text{-LH}_2$ acquisition overlaid on bright-field image of mice of Cage A, with ROIs. From left to right: FACS PpA1, FACS PpA1, FACS PpPp, FACS PpPp, FACS PpPp.

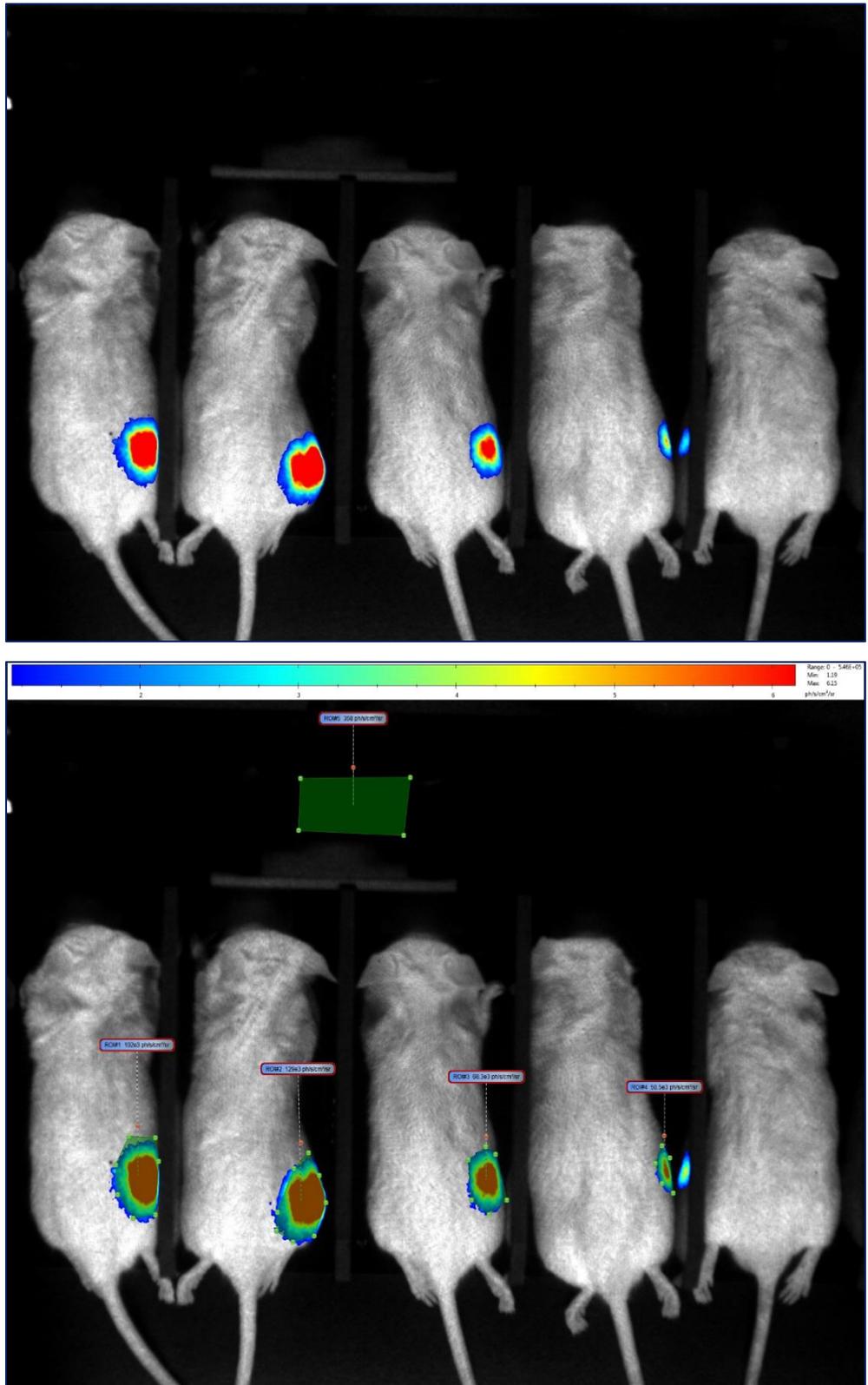


Figure 5.4. d -LH₂ acquisition overlaid on bright-field image of mice of Cage B, with ROIs. From left to right: FACS PpA1, FACS PpA1, FACS PpA1, FACS PpA1, FACS PpA1 (misinjected).

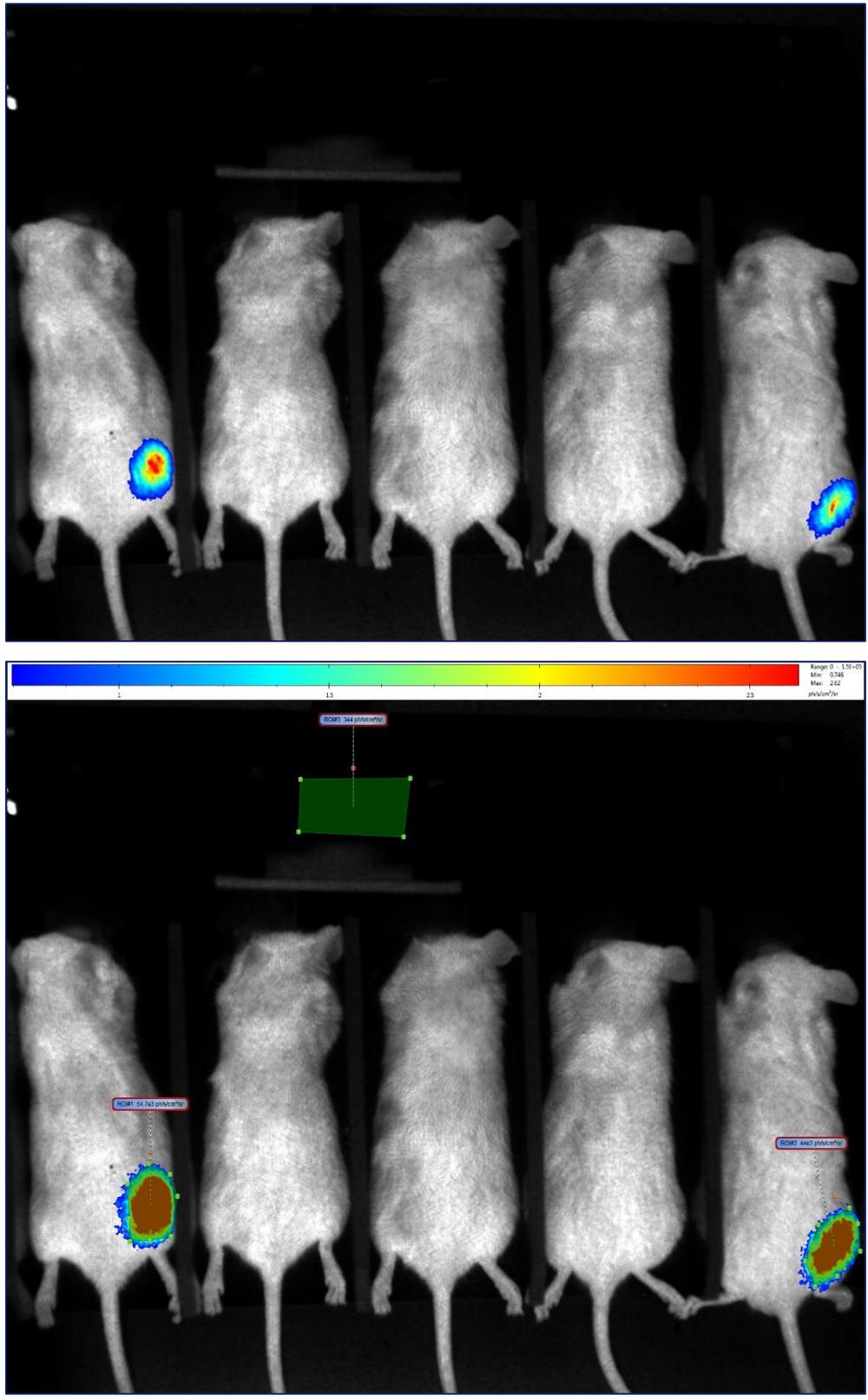


Figure 5.5. d-LH₂ acquisition overlaid on bright-field image of mice of Cage C, with ROIs. From left to right: bulk-transduced PpA1, FACS PpA1 (misinjected), bulk-transduced PpPp (misinjected), non-transduced (misinjected), FACS PpA1.

Unfortunately the following human errors occurred during the implantation part of the experiment as shown in Table 5.2: the implants for the sole NT, the sole PpPp bulk-transduced, and two of the nine sorted PpA1 were all misinjected. The NT was not strictly necessary due to the nature of BLI having virtually no background signal, and the presence of a non-emitting mouse negligibly affects said background; it was intended to be included as having a true negative control is considered good-practice. The loss of the PpPp bulk-transduced sample meant that the sample for the PpA1 bulk-transduced data were not processed, this is because a comparison to the sorted samples requires a PpPp-relative value. The loss of two of the nine sorted PpA1 samples results in a slightly lower-powered experiment from $n = 9$ to $n = 7$ (Figures 5.7 – 5.9); however $n = 8$ is the industry standard for similar in vivo experiments, and thus $n = 7$ is likely more than adequate given that this particular experiment was intended to be a pilot.

The light output for each mouse over the 20 or 30 mins is plotted in Figure 5.6. Unfortunately, all three PpPp FACS mice were in cage A (20 min acquisition), meaning that a direct comparison between PpA1 and PpPp for the final 10 (of 30) mins was not possible. However, the quality and consistency of the data for PpPp were high ($R^2 = 0.98$) that it could be reasonably extrapolated for this missing time (Figure 5.7). This analysis revealed that PpA1 light output peaks at $\sim 1,100$ secs post-injection, and PpPp at $\sim 1,300$ secs; this difference may be due to more rapid D -LH₂ depletion (dehydroluciferyl adenylate (L-AMP) inhibition) in PpPp. However, it is clear that even with this marginally longer duration, PpA1 total light output would still be markedly less than Ppy Fluc if the peaks are extrapolated to zero. Thus, PpA1 is likely using D -LH₂ inefficiently, with a higher proportion of non-bioluminescent oxidising reactions occurring.

Peak light output of PpA1 was $\sim 45\%$ that of PpPp (Figure 5.7), far higher than observed in vitro in HEK ($\sim 10\%$). Total light output of PpA1 was also $\sim 45\%$ that of PpPp (Figure 5.8). The spectral emission of PpPp and PpA1 was extremely similar (Figure 5.9), this may be due to non-optimal physiological conditions (e.g. pH), but is known to occur due to Hb filtration.

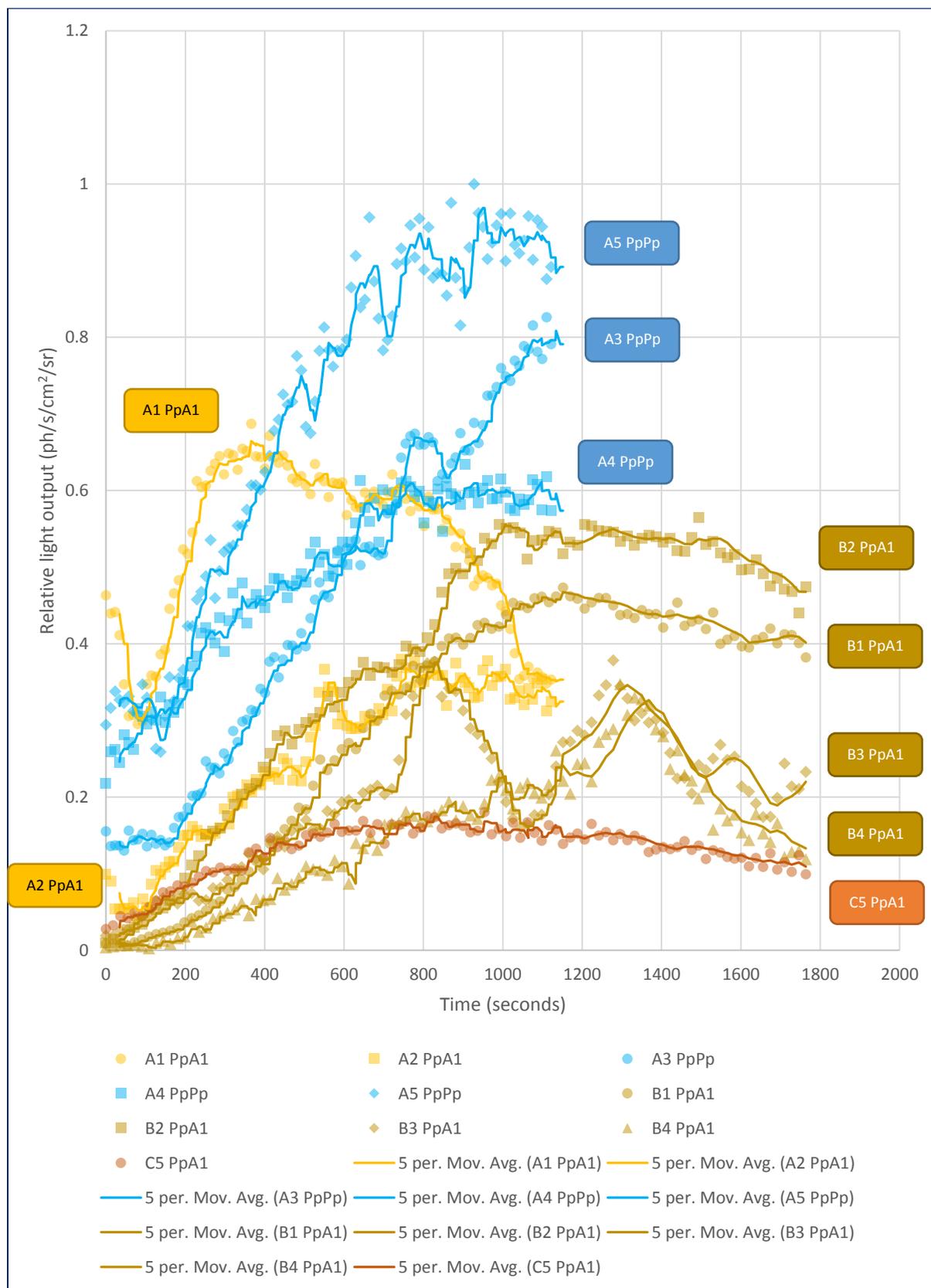


Figure 5.6. Longitudinally acquired relative (to brightest data point) light emission of all mice. Labels on chart area indicate cage and number of each mouse, followed by the implanted construct.

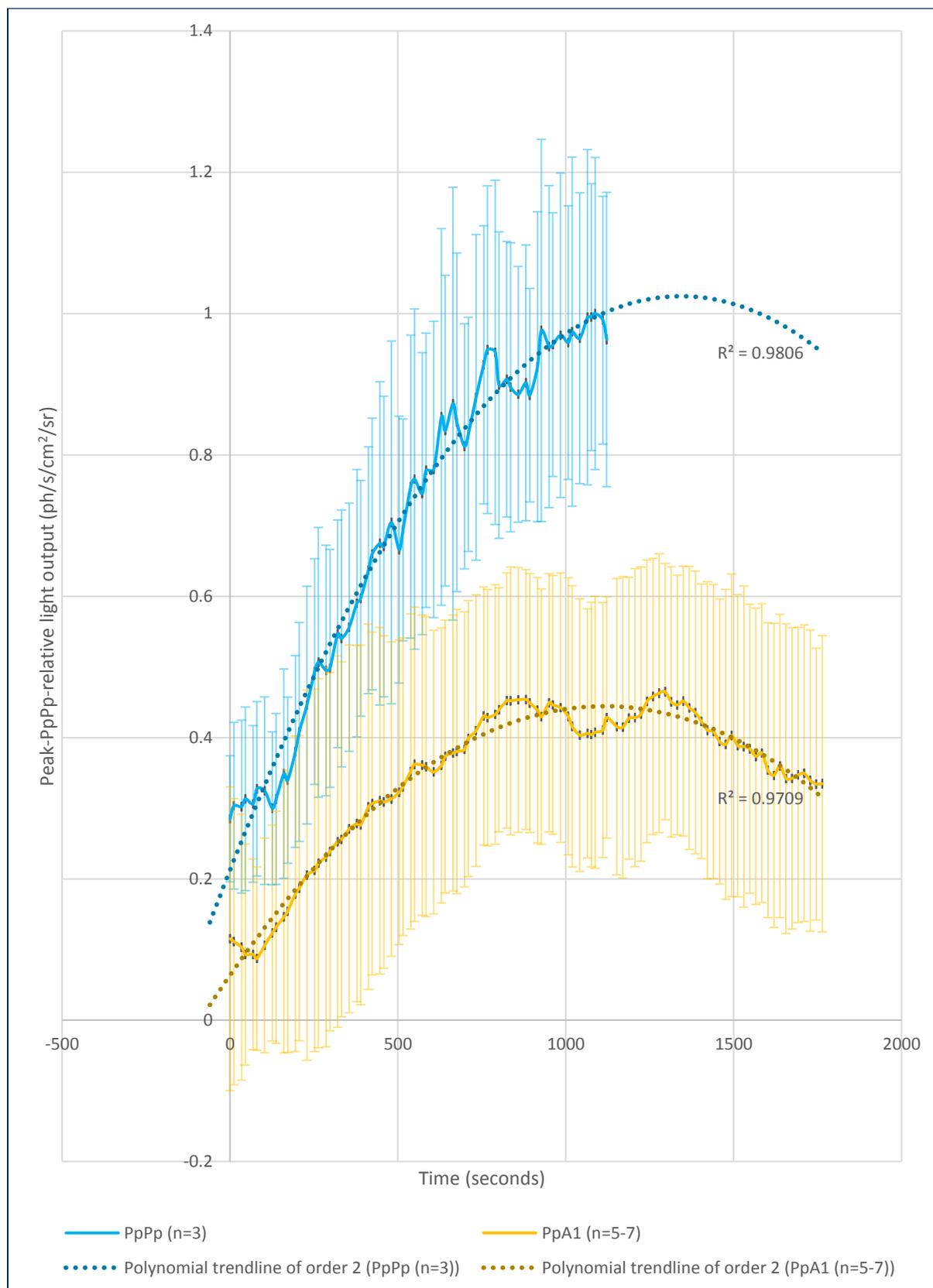


Figure 5.7. Longitudinally acquired PpPp-relative light emission of PpPp and PpA1 mice, averaged. The trendlines have been extrapolated 60 secs backwards as this was the approximate time between D-LH₂ injection and acquisition initiation.

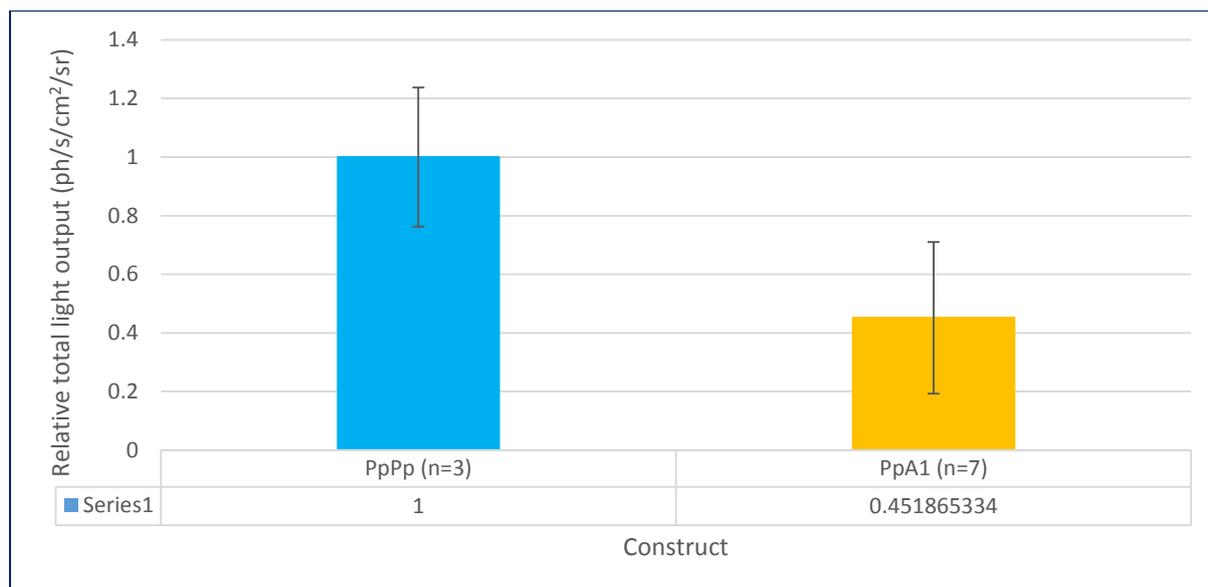


Figure 5.8. PpPp-relative total light emission of PpA1 mice. n = number of mice in cohort.

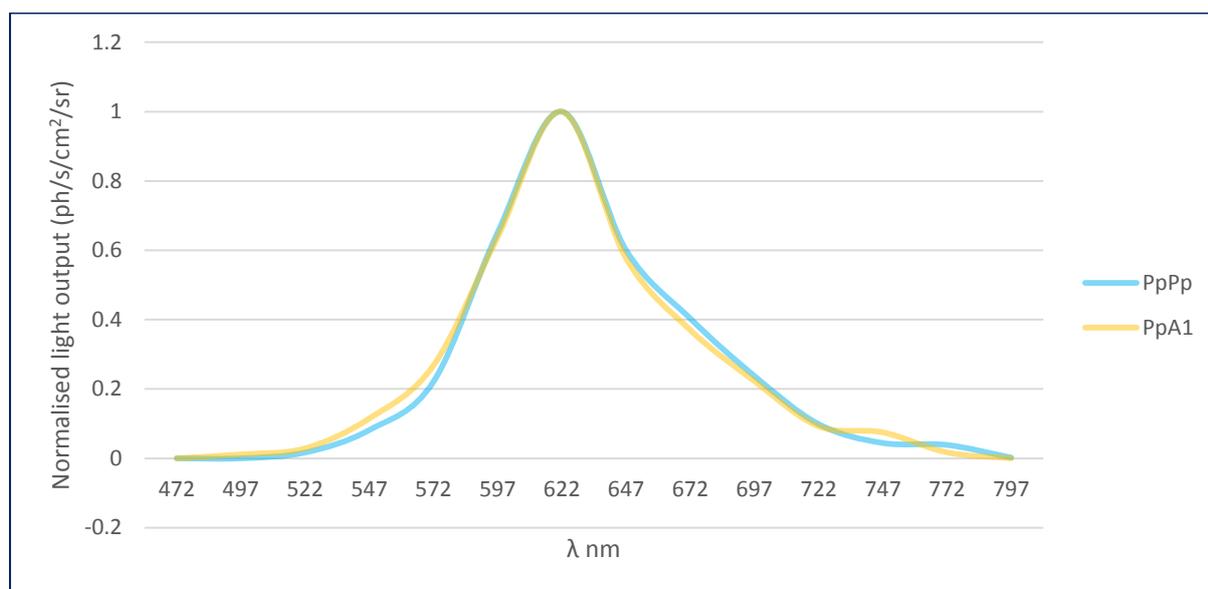


Figure 5.9. Normalised spectra of PpPp and PpA1 mice from Cage A. The emitted light from HEK cells passes through mouse tissue before reaching the sensor, both spectra are thus altered by Hb filtration and appear very similar.

5.4 Further Discussion

5.4.1 How does the increased expression of PpA1 relative to Ppy Fluc impact in vivo use?

As seen in Figure 5.1, the bulk transduced (unsorted) cells showed the same trend of the chimera expressing higher IRES EGFP than WT *Ppy* Fluc. This advantage was incidentally

removed for the sorted cells on which the conclusions for this chapter (PpA1 performing at ~45% that of PpPp) are drawn. Therefore, this indicates that the much better (~2.5 fold increase) expression of the chimeras in previous transduction experiments is not necessary for PpA1 to perform adequately and find use as a biomarker.

As mentioned, the randomly integrating lentiviral vectors leads to heterogeneous expression between individual cells. Thus, monoclonal cell populations are to be avoided for comparison of constructs as, if expressions are not equivalent, then the comparison is not fair. For this reason sorted polyclonal cells and unsorted (bulk transduced) cells were used for in vivo screening. Safe-harbours are loci in genomes which are “safe” to insert novel constructs into (e.g. insertion will not knockout a gene), but crucially whichever locus is chosen for insertion is identical between transformed cells. Thus, cells transformed this way are isogenic and healthier than their lentivirally transformed counterparts. There are further advantages to using such a system, these are outlined in table 5.3 below. Using such a system may have simplified the process of creating the transduced implants, opened the possibility for monoclonality, and improved the resulting data.

Key application point	Safe-harbour transformation	Lentiviral transformation	Safe-harbour benefit
Minimal time for useable integration	2 weeks	4 weeks	Faster
Polyclonal cell population	Isogenic	Highly variable	Cleaner
Integration site	1 targeted site	Multiple and random sites	Cleaner
Copy number	Limited	From low to high	More reliable
Monoclonal isolation	Not required	Recommended	Easier

Table 5.3 adapted from (<https://www.tebu-bio.com/blog/wp-content/uploads/2016/01/comparison-SF-vs-classic-method.jpg>). Advantages of using the Safe-harbour system over lentivirus for transduction of mammalian cells.

5.4.2 How were final data impacted by non-optimal mouse placement?

The right rear flanks of the mice were not angled optimally for BLI as much of the implant is occluded (Figures 5.3 – 5.5). This is especially true for the two right-most (PpPp) mice in Figure

5.3 where the light can be seen to be reflecting off of the plastic dividers between the mice; this may have given PpPp a disadvantage. This experiment could have been performed with the mice lying on their left sides, with their right flank fully exposed to the camera. This was not done primarily because the mice are less likely to fall out of the nosecone if placed on their stomachs. A mice awakening during acquisition will wander around the chamber, voiding its measurement, and may dislodge other mice in the process. However, given the aforementioned reflection issues, this may have been a risk worth taking.

Given that all mice were treated identically post-implant, the non-optimal angle should affect all constructs equally, with uniformly reduced light output. However, this assumes uniform implant placement, which is evidently not the case as some implants were occluded and others not. Thus, this non-optimal angle increased the noise in the processed data, ultimately contributing to the large error bars seen (figures 5.7 and 5.8).

5.4.3 How were final data impacted by premature acquisition termination?

For the Cage A mice (Figure 5.5), the mice were imaged for 20 mins each before stopping the acquisition and removing the mice. This was pre-planned based on the rapid peak and diminishment of light output observed for PpA1 in HEK cells in vitro. Only upon immediate analysis of these data was it realised how the light output of both PpPp and PpA1 had begun to plateau at the 20-min mark. Thus the imaging time was increased to 30 mins for the mice of Cages B and C (Figure 5.5). However, even with the 30 min acquisition PpA1 was still disadvantaged in relation to PpPp as its peak appears prolonged in vivo relative to PpPp (Figure 5.7). Therefore, had a longer acquisition been performed, PpA1 would have likely performed much better.

5.5 Conclusions

The Ppy Fluc-relative performance of PpA1 in vivo was greater than that in vitro in HEK cells. The error bars on the final quantification of brightness for PpPp and PpA1 were large, as expected for a pilot experiment with relatively small cohort sizes – especially considering the misinjection technicalities. Nevertheless, PpA1 is at least 15% as bright as PpPp (assuming worst-case error bar limits) which exceeds the minimum brightness requirement (10%) for

use as an in vivo reporter as set out in the project Aims. However, PpA1 is still likely immunogenic as the major BALB/c epitope remains, and presently PpA1 is objectively ultimately inferior to *Ppy* Fluc. Thus, PpA1 was progressed in the project for further engineering enhancements to resolve this.

6 Chapter 6 – Further engineering enhancements to the *Photinus Pyralis* / murine acyl-CoA synthetase 1 hybrid construct

6.1 Chapter Summary

The aim of the work in this chapter was to explore potential ways to reduce the presumed immunogenicity of the *Ppy* Fluc NTD of PpA1. Two different approaches were adopted.

One approach was to divide *Ppy* Fluc into eight sections called “microdomains” (μ Ds) which I hypothesised could be exchanged in a modular fashion with equivalent murine ACSM1 μ Ds without total loss of bioluminescence function, as was achieved in chapter 4. *Ppy* Fluc-backed constructs with one of each ACSM1 μ Ds (and one construct with two ACSM1 μ Ds) were created and screened in HEK cells with D-LH₂. Of these constructs only the N-terminal or C-terminal ACSM1 μ D substitutions retained bioluminescence function (0.06% and 9.69% that of *Ppy* Fluc respectively), and the latter construct is identical to PpA1 from Chapters 4 and 5.

The second approach taken was to substitute the region of *Ppy* Fluc known to be the strongest immunogenic epitope for that of another homologous luciferase to reduce the immunogenicity whilst retaining bioluminescence function. From the literature the epitope GFQSMYTFV was identified as a prime target (Limberis et al. 2009). Modelling revealed that this region would likely be functionally sensitive to mutations, therefore only similar luciferases were considered as candidates for substitution. Ten luciferases were considered, and using a combination of epitope prediction programs and BLAST alignments the equivalent region from *Photinus scintillans* was selected. Constructs of PpPp and PpA1 with the GFQSMYTFV region “seamlessly” substituted for that of *Psc* Fluc – PpPsPp and PpPsA1 respectively – were cloned. These mutants were screened in HEK cells and both were found to be brighter and to emit at a longer wavelength than their non-*Psc* counterparts; when filtered through haemoglobin PpPsPp was dimmer than PpPpPp, whilst PpPsA1 remained brighter than PpPpA1.

6.2 Overall Introduction

The results of Chapter 4 demonstrated that CTDs from murine ACSMs can greatly enhance the light output of the NTD of *Ppy Fluc*. This set a precedent for the potential of ACSM2, ACSM3, and especially ACSM1 (as the brightest hybrid) as potential substitutes for other regions of *Ppy Fluc*.

6.2.1 Substitution of the *Ppy Fluc* major BALB/c epitope

From speaking to fellow researchers who use mouse cancer models within Cardiff University, and from conversations with other researchers at various conferences such as The European Molecular Imaging Meeting, a preference was ascertained towards using the better understood and immunocompetent mouse models. This is because they more closely represent a human model (which is commonly immunocompetent) and allows them to understand the mouse model optimally. It also greatly simplifies maintenance and experimentation. BALB/c appears to be the most favoured strain and was the focus for creating a minimally immunogenic luciferase. Limberis *et al.* identified three epitopes within the *Ppy Fluc* sequence relevant for the BALB/c strain, with the epitope spanning G160 – V168 (protein sequence GFQSMYTFV) within the NTD being the most potent. This sequence is not highly conserved between luciferases (Figure 6.11) and as such its substitution with that of a different luciferase could reduce the immunogenicity while conserving bright bioluminescence. There are many well-characterised luciferases beyond *Ppy Fluc*, allowing many candidates.

6.2.2 Fragmented ACSM substitution of the *Ppy Fluc* N-terminal domain

It was found that replacing the NTD of *Ppy Fluc* with that of a murine ACSM removed bioluminescent function, however since it is possible to replace the CTD with that of an ACSM and retain some function it may also be possible to replace a limited portion of the NTD with an ACSM and also retain function. A logical strategy to limit the substitution is to divide the *Ppy Fluc* NTD into subdomains (Figure 6.1). However, unlike the large NTD and CTD, the first three structural subdomains do not appear sequentially in the primary sequence of the protein and are thus fragmented and each separated segment is treated here as a separate “microdomain” (μ D). Dividing both parent proteins into 8 fragments (as opposed to 2 in the

previous chimera work) provides 256 possible recombinant combinations rather than four. The mathematical formula for calculating these numbers is below.

$$\text{Number of hybrid constructs} = [\text{number of parent proteins}]^{\text{number of fragments}}$$

Thus, $2^8 = 256$. The nomenclature for the earlier chimera work was in order from N to C termini – [parentAparentB] “PaPb”, for instance PpA1. For the chimeras in this chapter, the same nomenclature was expanded to include the six additional NTD fragments, and the aforementioned chimera would be written as PpPpPpPpPpPpPpA1.

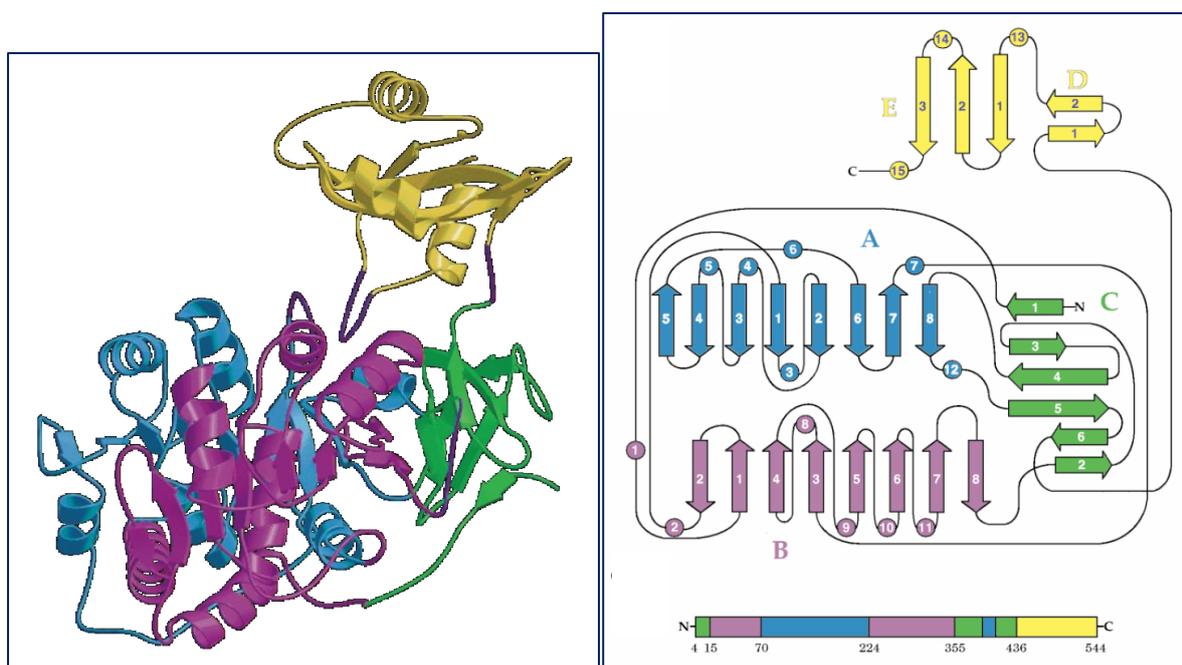


Figure 6.1 taken from (Conti et al. 1996). **Arrangement of the four structural subdomains derived from the eight sequential “microdomains” of *Ppy* Fluc.** (Left) 3D crystal structure of *Ppy* Fluc with subdomains highlighted in different colours. (Right) 2D schematic diagram of *Ppy* Fluc with subdomains highlighted in different colours, and primary sequence of *Ppy* Fluc with aforementioned subdomains highlighted. Note that in the primary sequence the subdomains do not all occur sequentially and the first three are fragmented giving rise to the 8 “microdomains” defined.

6.3 Results and Discussion

6.3.1 Design, cloning, and screening of microdomain-exchanged *Ppy* Fluc / ACSM1 constructs

6.3.1.1 Design

The MTSs of ACSM1 and ACSM3 do not appear to affect the expression nor function of proteins expressed in *E. coli*, however for mammalian expression these will result in constructs containing an ACSM NTD being localised to the mitochondria. This could cause complications, as the pH of mitochondria is different to that of the cytosol, and the permeability of this organelle to D-LH₂ and analogues is unknown. As such, the MTSs for the ACSM1 and ACSM3 were removed from gBlock oligos designed for mammalian expression.

The WT protein sequence of *Ppy* Fluc was uploaded to Integrated DNA Technologies' codon optimisation tool, and the protein optimised for human expression. The junction regions between the subdomains of the WT sequence of *Ppy* Fluc were examined and seven restriction enzymes were chosen for in-silico silent mutagenesis. From N to C termini these were: KpnI, BanII, BstEII, Aval, Avall, ClaI, and HindIII (Figures 6.2 and 6.3).

ACSM1 and ACSM3 were aligned to *Ppy* Fluc using TM-align and the Murine ACSM1 and ACSM3 were similarly optimised for mammalian codon usage, however as the peptide sequence is different in the junction regions to *Ppy* Fluc, it is not possible to create the chosen restriction sites without altering the encoded protein sequence. Therefore, the restriction sites were added non-silently meaning that ACSM1 and ACSM3 both contain a small number of residues of *Ppy* Fluc at each junction region. Although both ACSM1 and ACSM3 gBlocks were designed and ordered, due to time constraints ACSM3 was not used and its DNA/protein sequence is not shown here, however it can be found in the appendix (Figure 6i).

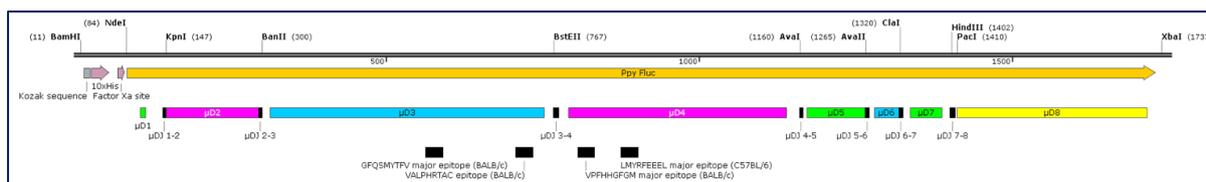


Figure 6.2. Map of gBlock DNA sequence encoding the 8 μ D *Ppy Fluc* parent. The position of the most potent epitope in BALB/c (GFQSMYTFV) is shown along with three less potent epitopes. The junction markers cover the residues containing the restriction sites. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 6.2EX).

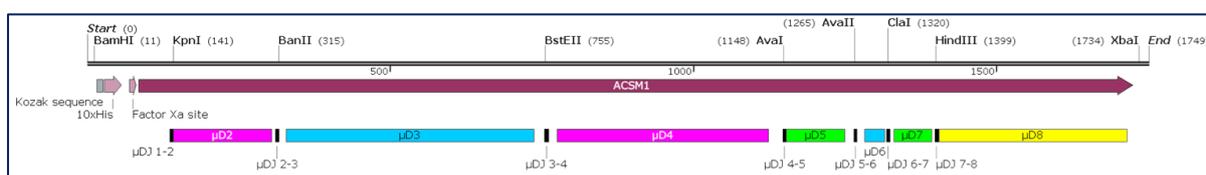


Figure 6.3. Map of gBlock DNA sequence encoding the 8 μ D *ACSM1* parent. The junction markers cover the residues containing the restriction sites, these junctions contain the necessary *Ppy Fluc* substitutions. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 6.3EX).

6.3.1.2 Cloning

In principle, any of the possible 256 combinations for *ACSM1* with *Ppy Fluc* could be created by digesting each parent into 8 fragments, purifying the necessary fragments and ligating together.

The parent fragments of *ACSM1*, and *Ppy Fluc* were obtained as double-stranded DNA from IDT. Each was terminally digested with *Bam*HI and *Xba*I and ligated into the pCCL 305 vector; this was then transformed into DH5 α cells and 5 mL of liquid culture grown overnight. Plasmid DNA was prepared via minipreps and 50 μ g digested with *Bam*HI and *Xba*I. The digest was run on an agarose gel and both the 305 vector backbone and insert extracted. Each insert was split into 8 individual digests to cut out a single μ D – *Kpn*I alone, *Kpn*I + *Ban*II, *Ban*II + *Bst*EII, etc. These were all run on an agarose gel and all fragments extracted. However, the smaller bands were not visible and could not be extracted.

To improve both the yield and purity of the purified fragments, each was PCR-amplified from the gBlock template prior to restriction digestion (Figures 6.4 and 6.5);(table 6.1). Extracting and purifying the smaller <100bp bands (and those with inefficient amplification) was difficult but possible as the bands themselves were barely visible to the eye (Figure 6.6).

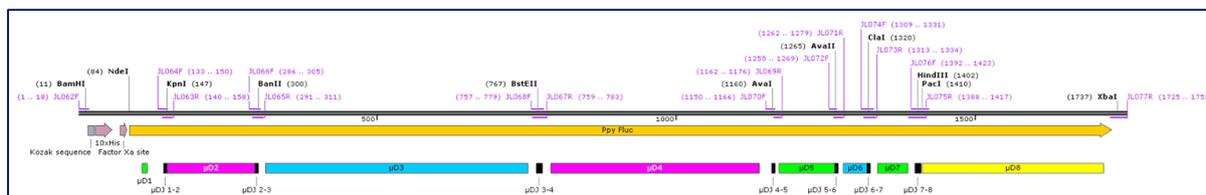


Figure 6.4. Primer binding locations for μ D amplification from *Ppy Fluc* gBlock oligo parent.

An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 6.4EX).

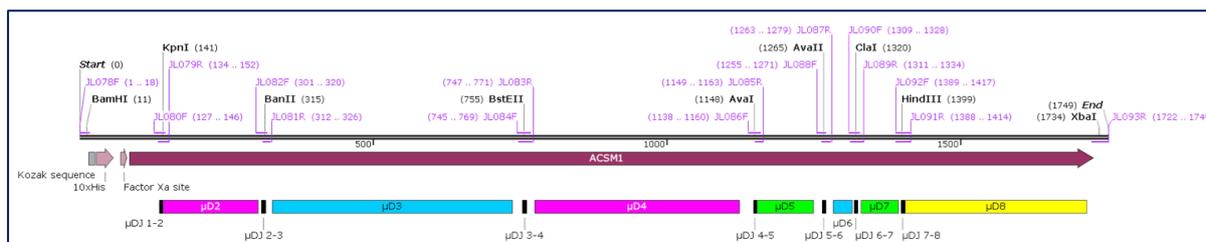


Figure 6.5. Primer binding locations for μ D amplification from *ACSM1* gBlock oligo parent.

An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 6.5EX).

Fragment	Forward primer – reverse primer	Length (bp)
<i>Ppy</i> Fluc μ D1	JL062F – JL063R	158
<i>Ppy</i> Fluc μ D2	JL064F – JL065R	179
<i>Ppy</i> Fluc μ D3	JL066F – JL067R	498
<i>Ppy</i> Fluc μ D4	JL068F – JL069R	420
<i>Ppy</i> Fluc μ D5	JL070F – JL071R	130
<i>Ppy</i> Fluc μ D6	JL072F – JL073R	80
<i>Ppy</i> Fluc μ D7	JL074F – JL075R	109
<i>Ppy</i> Fluc μ D8	JL076F – JL077R	361
ACSM1 μ D1	JL078F – JL079R	152
ACSM1 μ D2	JL080F – JL081R	200
ACSM1 μ D3	JL082F – JL083R	471
ACSM1 μ D4	JL084F – JL085R	419
ACSM1 μ D5	JL086F – JL087R	142
ACSM1 μ D6	JL088F – JL089R	80
ACSM1 μ D7	JL090F – JL091R	106
ACSM1 μ D8	JL092F – JL093R	361

Table 6.1. Desired PCR amplification products for *Ppy* Fluc and ACSM1 microdomains.

Primer sequences can be found in Chapter 2 (see 2.1.1.2).

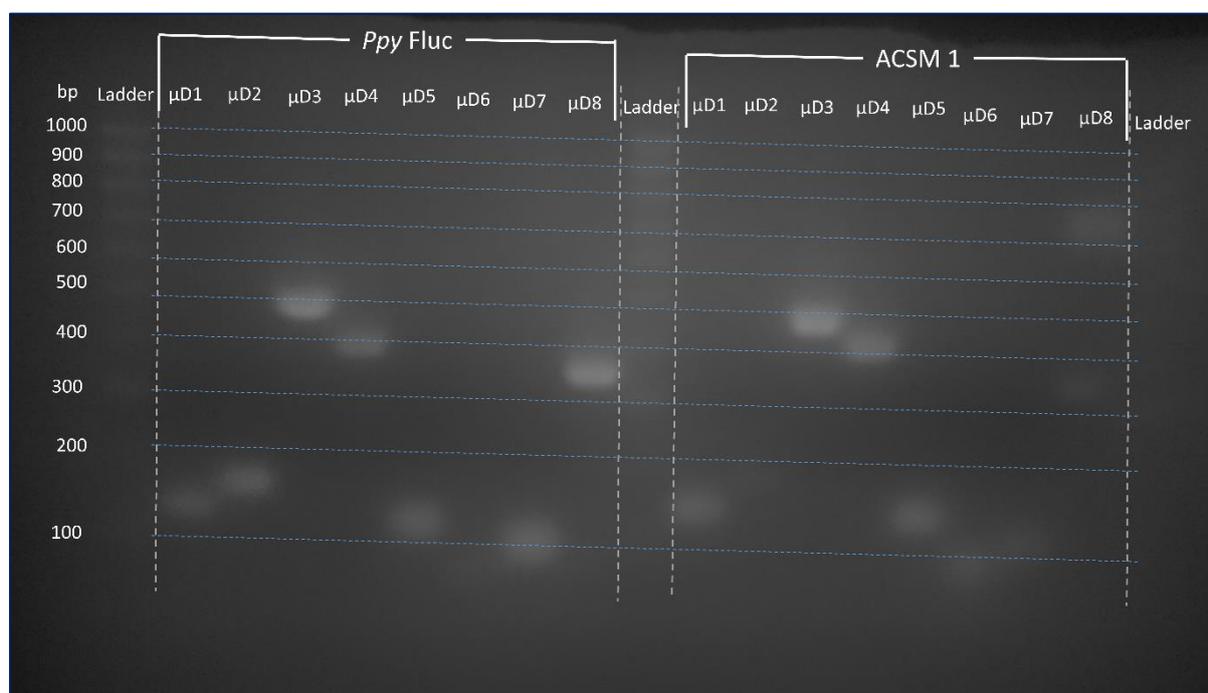


Figure 6.6. Gel electrophoresis of purified μ D PCR products of *Ppy* Fluc and ACSM1. All microdomain fragments were amplified successfully but in varying yields. The μ D8 lane of ACSM1 shows an extra band around 700 bp and is an undesired amplification product.

A pilot 9-fragment (8 μ Ds + pCCL 305 vector) ligation was attempted but this failed to produce any colonies of the correct sequence. Although theoretically possible, even with successful extraction such a method requires an 8-fragment ligation which is rate-limited by the slowest restriction-site end. Several attempts to ligate 8 *Ppy* Fluc and ACSM1 fragments together with the pCCL 305 vector failed. Therefore, it was decided that *Ppy* Fluc and ACSM1 should be terminally PCR-amplified prior to any digestions, but that for the first round of recombinants the number of fragments in the ligation should be minimised.

The first round of constructs to be created were those with a *Ppy* Fluc backbone and one ACSM1 μ D. Since there are 8 μ Ds this gives a total of 8 constructs, although one of these was the PpA1 chimera which has already been tested (Chapter 4 and 5), and thus 7 new constructs were to be created. All these constructs could be created through 3- or 4-fragment ligations (including the vector). For instance, PpPpA1PpPpPpPpPp could be created by digesting both ACSM1 and *Ppy* Fluc with BanII and BstEII and gel-extracting the three fragments: μ D1- μ D2 of *Ppy* Fluc, μ D3 of ACSM1, and μ D4- μ D8 of *Ppy* Fluc. These three fragments and vector backbone could then be ligated together. This approach allowed the creation of the 7 new aforementioned constructs and gave the potential for troubleshooting specific ligations

6.3.1.3 Screening

From the initial round of screening only the microdomain constructs A1PpPpPpPpPpPpPp – and the previously-screened PpPpPpPpPpPpPpA1 – emitted light with D-LH₂ at detectable levels, with the former being too dim to acquire a spectrum (Figure 6.7). The second round of cloning and screening was to combine the two bioluminescent ACSM1 μ D substitutions (the N and C terminal μ Ds) together into a construct with both, namely A1PpPpPpPpPpPpA1. This was cloned via restriction digestion and ligation. After DNA plasmid preparation from DH5 α *E. coli* this construct was transfected into HEK cells (see 2.2.4.6) and screened with D-LH₂, but failed to emit at detectable levels using the PhotonIMAGER (data not shown, see Figure 9.1).

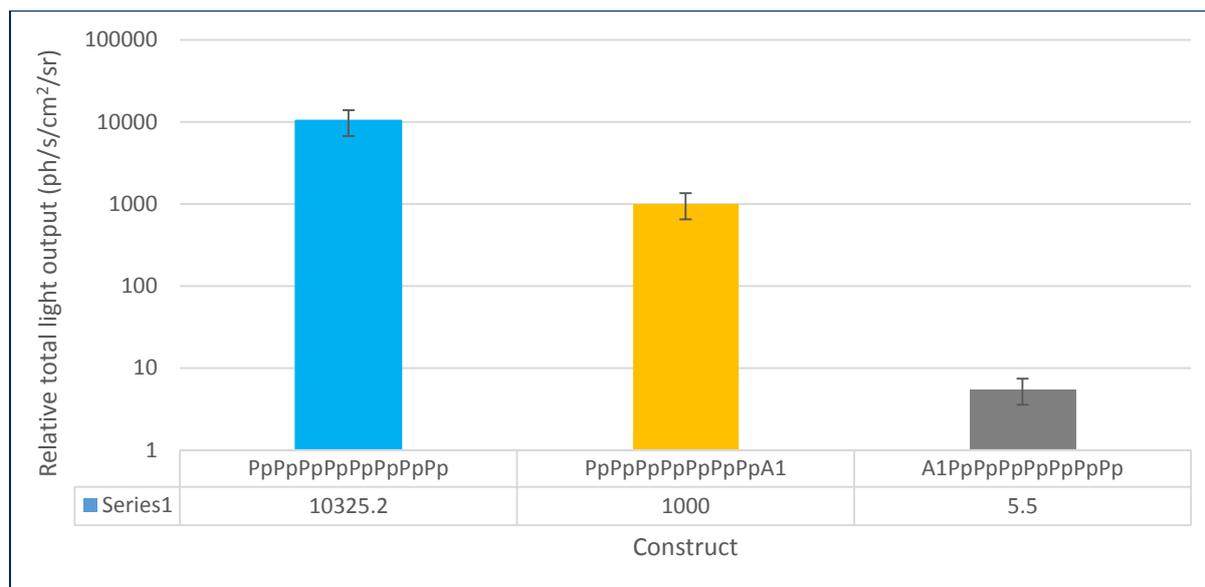


Figure 6.7. Light output of PpPpPpPpPpPpPpPpPp (*Ppy* Fluc), PpPpPpPpPpPpPpPpA1 (PpA1), and A1PpPpPpPpPpPpPpPp. Values are relative to PpPpPpPpPpPpPpPpA1 which has a value of 1000 to allow error bars to display more correctly for A1PpPpPpPpPpPpPp.

6.3.2 Design, cloning, and screening of epitope-substituted *Ppy* Fluc mutants

6.3.2.1 Design

The aforementioned paper (Limberis et al. 2009) reports using two epitope prediction programs called BIMAS and SYFPEIHTI (Figure 6.8). SYFPEIHTI is freely available to use, but a useable version of BIMAS could not be found. To follow the example the authors set (of using two programs) an equivalent program called Rankpep was used. The paper authors caution that their programs (along with others) are not highly reliable. For instance, if the entire *Ppy* Fluc sequence is entered into Rankpep and SYFPEIHTI they find that the GFQSMYTFV epitope is only the 11th and 4th most potent H2-K^d epitopes respectively – it is experimentally the most potent (Limberis et al. 2009). Secondly, SYFPEIHTI does not assess the likelihood of H2-D^d type binding to a given epitope – relevant for BALB/c immunogenicity and assessed by Limberis *et al.* For both these reasons the second program Rankpep was used as the primary guide for epitope potency.

In enzymes, terminals and loops are generally more tolerant to mutations than alpha helices and beta sheets. The GFQSMYTFV region was visualised in Visual Molecular Dynamics (VMD)

viewer to estimate its likely tolerance to mutation, and this revealed that approximately half of the sequence occupied an alpha helix, as displayed in Figure 6.9.

Predicted MHC class I peptide binders			
<i>Strain</i>	<i>Peptide sequence</i>	<i>MHC class I</i>	<i>Score</i>
C57BL/6	LMYRFEEEL	H2-K ^b	0.66 (B); 0 (S)
		H2-D ^b	16.6 (B); 17 (S)
BALB/c	GFQSMYTFV	H2-K ^d	345.6 (B); 22 (S)
		H2-D ^d	0.9 (B); ^a (S)
	VALPHRTAC	H2-K ^d	2.88 (B); 10 (S)
		H2-D ^d	0.03 (B); ^a (S)
	VPFHHGFGM	H2-K ^d	6.0 (B); 2 (S)
		H2-D ^d	0.3 (B); ^a (S)

Abbreviations: B, BIMAS; S, SYFPEITHI; MHC, major histocompatibility complex.
 The discrepancy of the scoring system as well as the nonidentification of an epitope by SYFPEITHI, which was scored by the BIMAS program can be attributed to the limitations of the epitope binding prediction programs in general.
^aNot identified.

Figure 6.8 taken from (Limberis et al. 2009). Three BALB/c (and one C57BL/6) epitopes identified within *Ppy* Fluc and their respective binding scores.

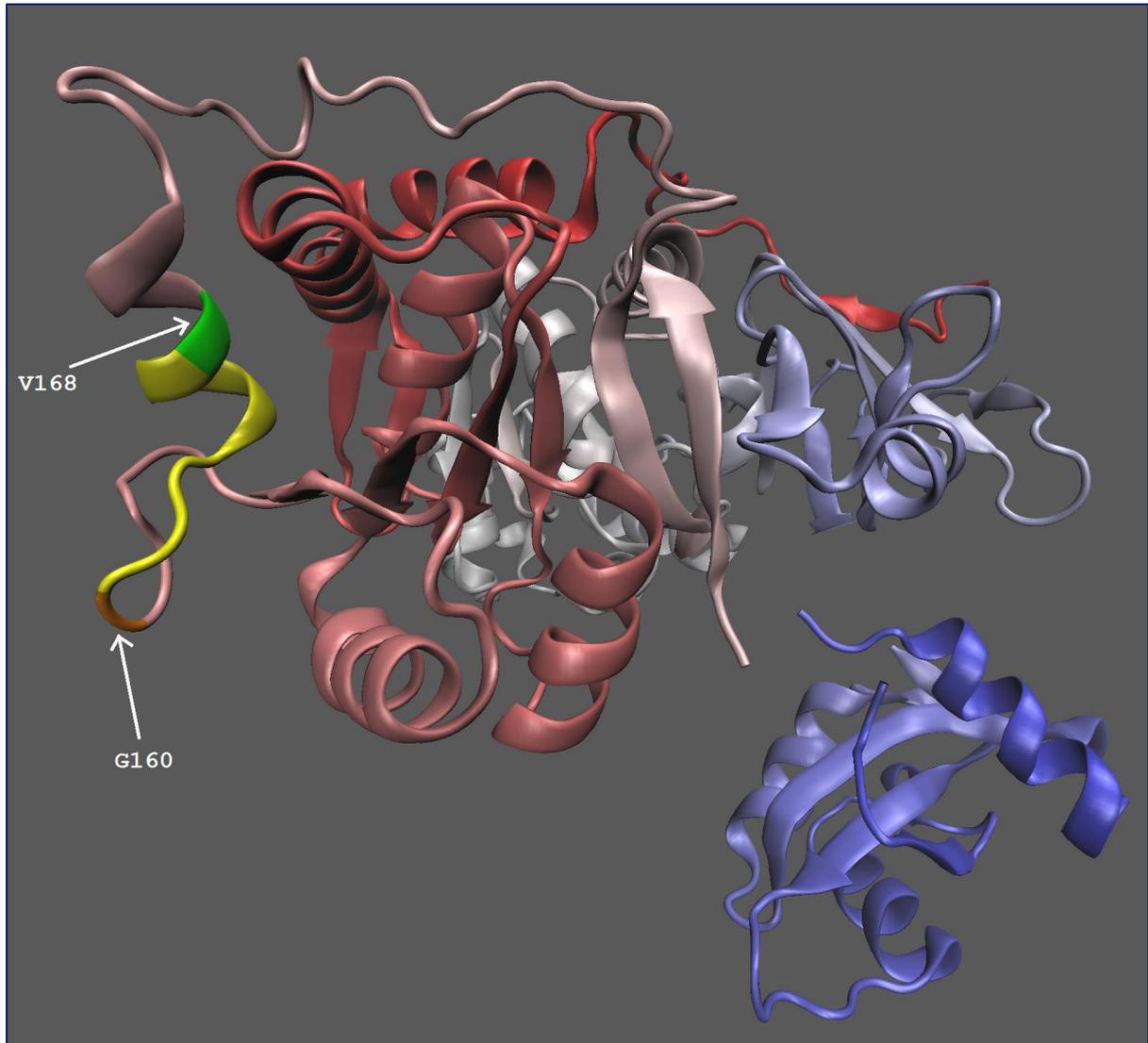


Figure 6.9. Major *Ppy* Fluc epitope (GFQSMYTFV) highlighted displayed in VMD viewer. The GFQSMYTFV epitope residues are displayed in yellow, the N-end of the epitope (V160) is displayed in orange and the C-end (G168) in green. The 3D protein structure used is 1lci.pdb.

Given the position of the GFQSMYTFV region, it was considered probable that any major change to this epitopic region could interrupt this helix and prevent proper folding of the protein. Therefore, only enzymes highly homologous to *Ppy* Fluc were considered as viable for epitope substitution. Ten candidate luciferases were selected primarily based on their known similarity to *Ppy* Fluc and their high degree of understanding in the literature. The list of these luciferases with their respective NCBI version codes and some key information about them are listed in table 6.2, as is *Photinus pyralis*.

The U.S. National Library of Medicine's protein BLAST tool was used (with *Ppy* Fluc as the reference luciferase) to identify luciferases with the greatest overall similarity as is shown in Figure 6.10.

The ten candidate luciferases along with *Ppy* Fluc were then aligned using CLC sequence viewer 8.0 (Figure 6.11). This software colour-codes residues based on their combination of neutral, acidic, basic, polar, and nonpolar properties. Therefore, the software allows the user to judge by eye how meaningfully different sequences are for a particular segment. The alignment showed that certain luciferases (*Luciola tsushimana*, *Luciola mingrelica*, *Luciola italica*, *Luciola cruciata*, *Abscondita terminalis*, and *Luciola parvula*) were highly divergent in this region. These were also the least homologous overall according to the BLAST (Figure 6.10).

Species	NCBI code	Geographical location
<i>Photinus pyralis</i>	P08659.1	North America
<i>Photinus scintillans (Psc)</i>	AOG75000.1	North America
<i>Lampyrus noctiluca (Lno)</i>	AAW72003.1	Europe
<i>Cratomorphus distinctus (Cdi)</i>	AAV32457.1	Brazil
<i>Photuris pensylvanica (Ppe)</i>	BAA05006.1	North America
<i>Luciola tsushimana (Lts)</i>	AAN40979.1	Japan
<i>Luciola mingrelica (Lmi)</i>	AAB26932.1	Eastern Europe
<i>Luciola italic (Lit)</i>	ABA03040.1	Italy
<i>Luciola cruciata (Lcr)</i>	BAW94790.1	Japan
<i>Abscondita terminalis (Ate)</i>	ABZ88151.1	South-east Asia
<i>Luciola parvula (Lpa)</i>	BAU71688.1	Japan

Table 6.2. Details of the ten chosen candidate luciferases. In *Photinus scintillans*, one residue is responsible for the orange colour difference to *Ppy* Fluc (Branchini et al. 2017). In *Luciola italic*, The N:C domain chimera (Ppy:Lui) is brighter than either parent luciferase (Branchini et al. 2014).

Descriptions		Graphic Summary	Alignments			
Sequences producing significant alignments						
		Download	Manage Columns			
		Show	100			
<input checked="" type="checkbox"/> select all 10 sequences selected						
GenPept Graphics Distance tree of results Multiple alignment						
Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> luciferase [Photinus scintillans]	1017	1017	98%	0.0	89.81%	AOG75000.1
<input checked="" type="checkbox"/> luciferase [Lampyrus noctiluca]	973	973	98%	0.0	84.38%	AAW72003.1
<input checked="" type="checkbox"/> luciferase [Cratomorphus distinctus]	957	957	98%	0.0	83.27%	AAV32457.1
<input checked="" type="checkbox"/> luciferase [Photuris pensylvanica]	794	794	99%	0.0	70.13%	BAA05006.1
<input checked="" type="checkbox"/> luciferase [Luciola cruciata]	781	781	98%	0.0	68.39%	BAW94790.1
<input checked="" type="checkbox"/> luciferase [Luciola mingrelica]	770	770	98%	0.0	67.65%	AAB26932.1
<input checked="" type="checkbox"/> luciferase [Abscondita terminalis]	769	769	98%	0.0	66.48%	ABZ88151.1
<input checked="" type="checkbox"/> luciferase [Luciola tsushimana]	765	765	98%	0.0	66.91%	AAN40979.1
<input checked="" type="checkbox"/> luciferase [Luciola italica]	759	759	98%	0.0	66.11%	ABA03040.1
<input checked="" type="checkbox"/> luciferase [Luciola parvula]	688	688	98%	0.0	59.63%	BAU71688.1

Figure 6.10. The homology of the ten chosen candidate luciferases to *Ppy* Fluc. The list from top to bottom is in order of overall homology (“Per Ident”).

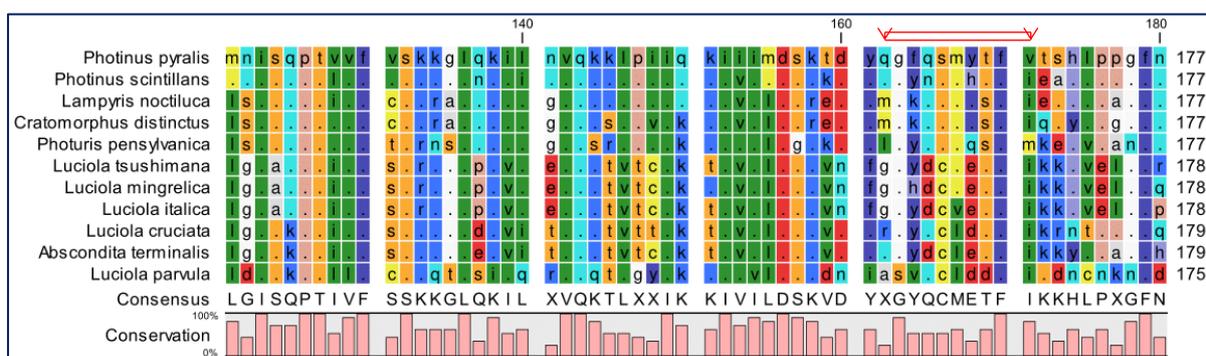


Figure 6.11. Alignment of the ten candidate luciferases with *Ppy* Fluc at the GFQSMYTFV region (marked in red). Note that the numbering is relative to the cumulative longest luciferases, not *Ppy* Fluc itself.

Due to the previously mentioned concern about the alpha helix, it was decided that the epitope should be substituted as seamlessly as possible. This means that the sequence to be substituted should be extended upstream and downstream of the GFQSMYTFV epitope until an identical residue to *Ppy* Fluc is reached. For instance, in the case of *Psc* Fluc the substituted segment (highlighted) must be extended downstream by two residues (underlined) until an identical residue is reached (green):

Ppy Fluc: ...dyq**gf**qsm~~yt~~fvts**h**lp...

Psc Fluc: ...dyq**gy**ns~~mt~~fiea**h**lp...

Substituted: ...dyq**gy**ns~~mt~~fiea**h**lp...

Evidently, because a greater number of residues may need to be substituted than the nine in the GFQSMYTFV region, the immunogenicity of the upstream and downstream sequences must be assessed. Further, novel potentially-epitopic nonamers are created at the upstream and downstream splices. To assess both of these effects the GFQSMYTFV-equivalent sequences were “seamlessly” substituted into *Ppy* Fluc, then the substituted region and 8 residues of *Ppy* Fluc upstream and downstream (33 residues total in length) assessed using Rankpep and SYFPEIHTI, the ten candidate luciferase were assessed (table 6.3). H2-k^d epitopes are generally more potent (Limberis et al. 2009), and as such these scores were given greater weight in choosing a substitution.

Species	Sequence (<i>Ppy</i> Fluc nonamer equivalent in bold, substituted sequence underlined, flanked by 8 mer <i>Ppy</i> Fluc)	H2-K ^d nonamer score		H2-D ^d nonamer score	
		Rankpep (pink if matches cleavage model)	SYFPEIHTI	Rankpep (pink if matches cleavage model)	SYFPEIHTI
<i>Photinus pyralis</i>	kiiimds ^k tdy qg <u>fqsm</u> yt <u>fv</u> ts ^h lppgfneyd	35.04 25.82 15.23	22 22 14	32.22 -3.56 -4.76	-
<i>Photinus scintillans</i>	kiiimds ^k tdy qg <u>ynsm</u> ht <u>fi</u> eahlppgfneyd	37.04 14.93 0.59	26 15 10	29.11 -1.92 -3.15	-
<i>Lampyrus noctiluca</i>	kiiimds ^k tdy qg <u>kqsm</u> ys <u>fi</u> eshlppgfneyd	30.85 14.42 2.61	20 15 15	32.57 -4.76 -5.97	-
<i>Cratomorphus distinctus</i>	kiiimds ^k tdy qg <u>kqsm</u> ys <u>fi</u> gshlppgfneyd	37.22 14.42 2.61	20 15 15	32.57 -4.76 -5.97	-
<i>Photuris pensylvanica</i>	kiiimds ^k tdy qg <u>ysmq</u> s <u>fm</u> kehlppgfneyd	31.42 13.93 4.73	15 15 12	13.05 0.42 -4.76	-
<i>Luciola tsushimana</i>	kiiimds ^k tdy qg <u>ydcm</u> et <u>fi</u> kkhlppgfneyd	46.03 13.12 0.00	24 14 13	23.23 -4.76 -5.97	-
<i>Luciola mingrelia</i>	kiiimds ^k tdy qg <u>hdcm</u> et <u>fi</u> kkhlppgfneyd	15.51 3.32 0.00	14 14 13	45.34 -4.76 -5.97	-
<i>Luciola italica</i>	kiiimds ^k tdy qg <u>ydcm</u> et <u>fi</u> kkhlppgfneyd	43.43 15.21 5.09	26 14 13	13.75 -4.76 -5.97	-
<i>Luciola cruciate</i>	kiiimds ^k tdy qg <u>ycld</u> t <u>fi</u> krntppgfneyd	50.81 24.17 14.80	27 15 13	5.24 3.78 -2.67	-
<i>Abscondita terminalis</i>	kiiimds ^k tdy qg <u>ydcl</u> et <u>fi</u> kkylppgfneyd	40.10 21.49 12.67	25 20 15	16.62 -4.76 -7.73	-
<i>Luciola parvula</i>	kiiimds ^k dnias sv <u>qcl</u> dd <u>fi</u> tshlppgfneyd	9.12 6.59 -1.28	16 16 15	-0.11 -4.76 -5.97	-
Consensus	kiiimds ^k tdy qg <u>ycm</u> et <u>fi</u> tshlppgfneyd	54.00 16.75 4.80	26 14 13	22.64 -4.76 -5.97	-

Table 6.3. Predicted equivalent epitopes of *Ppy* Fluc major BALB/c epitope from candidate luciferases. Both programs predict how epitopes (short chain peptides) will be generated by the proteasome, if the predicted epitope also matches this model it is coloured pink.

Due to severe time constraints of the PhD project only one epitope substitution was selected, and a low-risk approach was taken which placed likelihood of bioluminescence function above immunogenic reduction. This work thus represents a proof-of-concept for the tolerance of the GFQSMYTFV for substitution. Therefore, *Photinus scintillans* was selected primarily for its high overall and regional similarity to *Ppy* Fluc (Figures 6.10 and 6.11) despite its theoretically similar immunogenicity in the GFQSMYTFV region (Table 6.3). *Psc* Fluc is also well understood in the literature, for instance it is proposed that the orange light emission from *Psc* Fluc (as supposed to yellow-green from *Ppy* Fluc) is due entirely to a mutation at position 255 from

tyrosine to phenylalanine (Branchini et al. 2017). A red-shifting of emission is not undesirable, but red-shifts often result in substantial reductions in overall brightness. The GFQSMYTFV epitope spans the G160 to V168 positions, and so such a substitution avoids this particular mutation and the resulting hybrid was predicted to likely have similar colour to WT *Ppy* Fluc.

6.3.2.2 Cloning

The GFQSMYTFV epitope resides in μ D3 of *Ppy* Fluc, and the microdomain system was exploited to allow substitution of the epitope for that of *Psc* Fluc. This was achieved by designing and ordering a gBlock oligo of μ D3 of *Ppy* Fluc with the GFQSMYTFV epitope substituted for that of *Psc* Fluc, which was then restriction digested and ligated into pCCL 305 vectors containing PpPp and PpA1 respectively. The ordered gblock oligo containing *Ppy* Fluc μ D3 with the *Psc* Fluc substitution is shown in figure 6.12. To clarify, the entire μ D3 of *Psc* Fluc was *not* ordered. From this point on, PpPp with the GFQSMYTFV epitope substituted for *Psc* Fluc is referred to as PpPsPp, and PpA1 as PpPsA1; unmodified PpPp is referred to as PpPpPp, and PpA1 as PpPpA1.

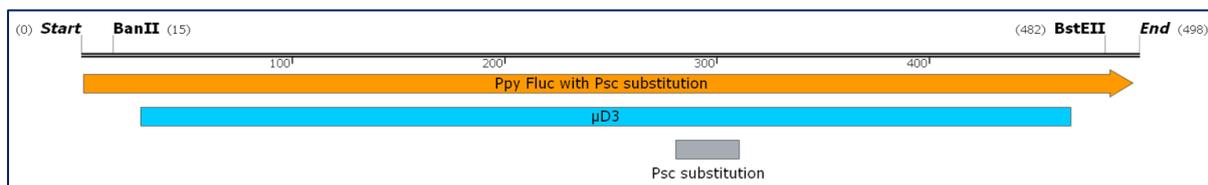


Figure 6.12. Map of gBlock oligo of μ D3 of *Ppy* Fluc with the *Psc* Fluc epitope replacement substituted. The grey box marks the region which has been substituted for *Psc* Fluc sequence; the sequence FQSMYTFVTS in WT *Ppy* Fluc is replaced with YNSMHTFIEA from *Psc* Fluc. An expanded version of this Figure – with DNA and protein sequence shown – can be found in the appendix (Figure 6.12EX).

6.3.2.3 Screening

Following plasmid DNA midiprep from DH5 α *E. coli* the constructs were transiently transformed into HEK cells (see 2.2.4.6). Following expression for 24 hrs the cells were screened with D-LH₂ for both light output, coarse spectral output – with and without haemoglobin filtration – and λ_{\max} emission. The λ_{\max} emission was measured using a BMG Labtech CLARIOstar[®] machine (see 2.2.6.2). For each construct a well containing the

transformed HEK cells was scraped so that the cells were concentrated in the centre of the well beneath the relatively small sensor. Then, one-by-one each well had D-LH₂ pipetted into it before being immediately inserted into the machine. Due to the low sensitivity of the sensor a full spectrum at maximum resolution (1nm divisions) would take greater than 10 mins, which would cause complications with the relative brightness diminishing throughout the measurement, causing the later and longer wavelengths to appear relatively dimmer. Secondly, the low sensitivity of the sensor causes difficulty detecting a signal for the very short and long wavelength photons (of which there are far fewer) above background, as such the spectrum appears clipped beyond a distance from the λ_{\max} , especially for the dimmer constructs. To resolve this a “coarse” wavelength screen of 10nm division resolution was performed. Once the λ_{\max} was identified at this resolution, a second high resolution wavelength screen of 1nm divisions was performed across a smaller bandwidth with the lower and upper wavelength limits set to ensure that the λ_{\max} was captured. The final λ_{\max} is shown within table 6.4 and the traces for the 10nm and 1nm screens shown in Figures 6.14 and 6.15.

Given that the main advantage of PpA1 in vivo was its red-shifted output and resulting better Hb penetration, further non-immunological improvements could be thus simulated in vitro. To simulate in vivo imaging, a screening of PpPpPp, PpPsPp, PpPpA1, and PpPsA1 transfected HEK cells was performed with and without a 1 cm thick haemoglobin phantom (see 2.1.5.4) between the cells and the camera sensor. Screening with D-LH₂ was performed 24 hrs post-transfection. The total light emission was measured and a full filter-wheel spectrum captured on the PhotonIMAGER (figures 6.16 – 6.19). These data are also included within table 6.4.

The Ppy Fluc-relative light output of PpPpA1 was 40% with the Hb phantom applied (Figure 6.17), indicating that the 45% Ppy Fluc-relative light output found in vivo was likely due almost entirely to Hb filtration. The spectra of all constructs was almost indistinguishable when filtered through Hb (Figure 6.19), as was observed in vivo (Figure 5.9), indicating the harsh filtration of light not in the bio-optical window.

Figure 6.13 and Table 6.5 culminate and average many experiments from throughout this thesis for PpPsPp, PpPpA1, and PpPsA1 light output relative to Ppy Fluc.

Construct	λ_{\max} (nm)	Relative brightness (Hb-) [up to 5EXP]	Relative brightness (Hb+) [1EXP]
PpPpPp	561	100%	100%
PpPsPp	571	~119%	~98%
PpPpA1	606	~6%	~40%
PpPsA1	615	~4%	~47%

Table 6.4. The impact of Hb filtration on PpPpPp, PpPsPp, PpPpA1, and PpPsA1 emission.

Hb- or Hb+ indicates whether a haemoglobin phantom was or was not used respectively.

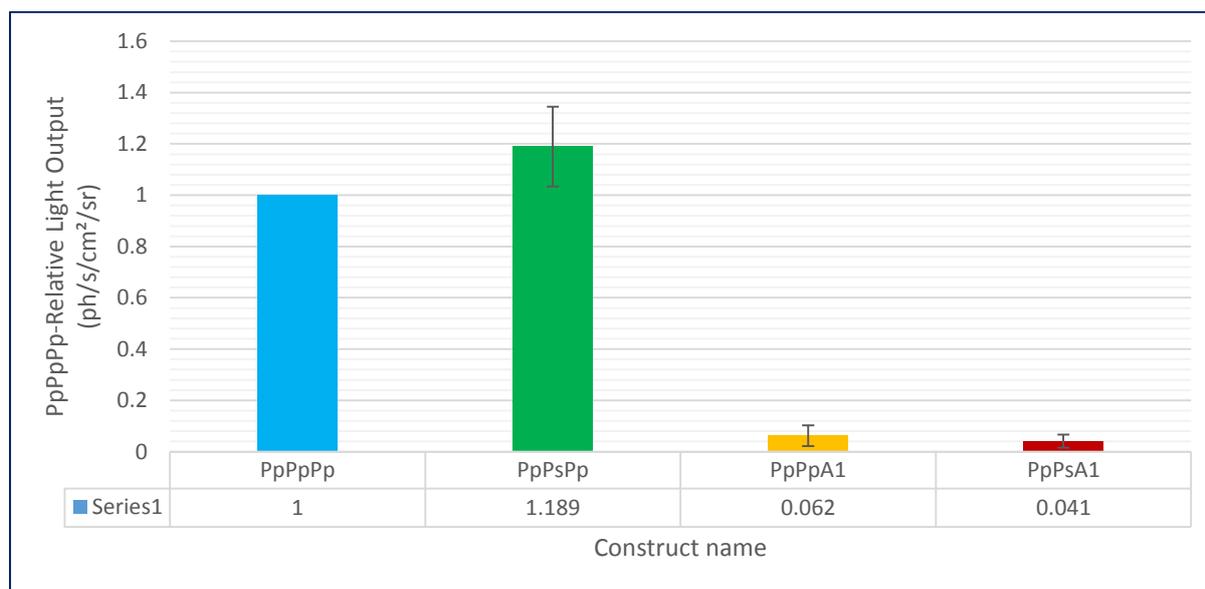


Figure 6.13. Ppy Fluc-relative light output of PpPsPp, PpPpA1, and PpPsA1. These data represent the average of up to five separate experiments in which these constructs were screened. These individual experiments are detailed in table 6.5.

Construct	PpPpPp-relative brightness of constructs for experiment dated:						Standard deviation between experiments
	03/12/2019	03/11/2019	14/10/2019	18/09/2018	15/08/2018	Average	
	Figure 6.16	n/a	n/a	Figure 4.29	Figure 4.21	Figure 6.13	
PpPpPp	1	1	1	1	1	1	n/a
PpPsPp	1.043	1.171	1.353	-	-	1.189	0.156
PpPpA1	0.051	0.033	0.018	0.112	0.097	0.062	0.041
PpPsA1	0.070	0.034	0.020	-	-	0.041	0.026

Table 6.5. Five experiments comparing the relative light output of novel luciferases to WT *Ppy Fluc*. Figure 6.13 is derived from the average column; Figure 6.16 is derived from the experiment dated 2019.12.03; Figure 4.21 (Chapter 4) is derived from the experiment dated 15.08.2019; Figure 4.29 (Chapter 4) is derived from the experiment dated 18.09.2018. None of these data has Hb filtration applied.

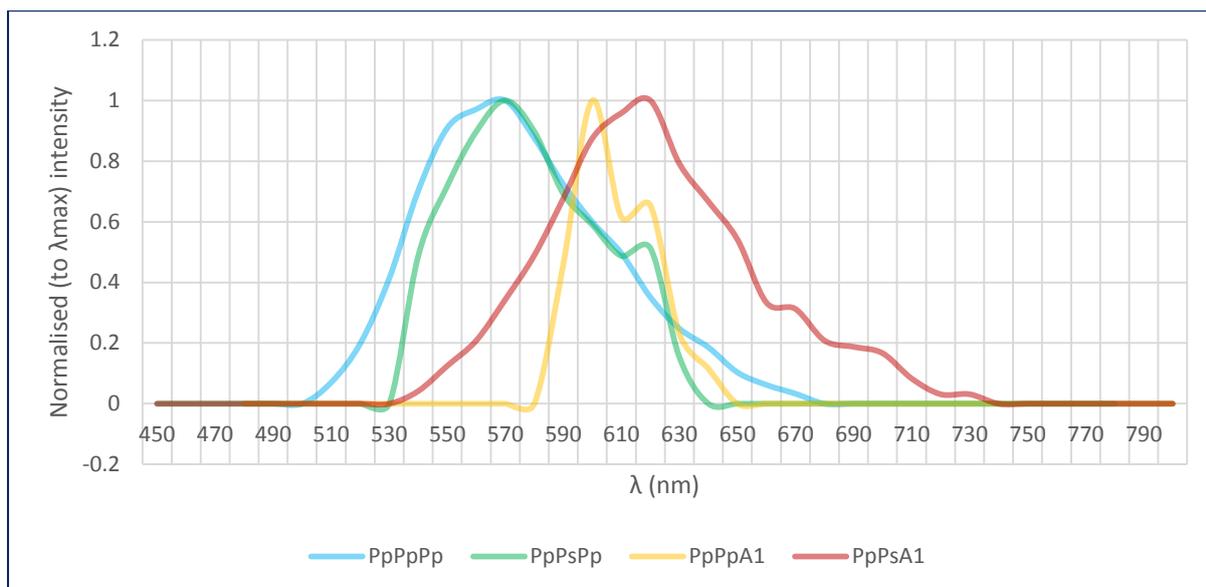


Figure 6.14. Low resolution (10 nm divisions) spectra screen of PpPpPp, PpPsPp, PpPpA1, and PpPsA1.

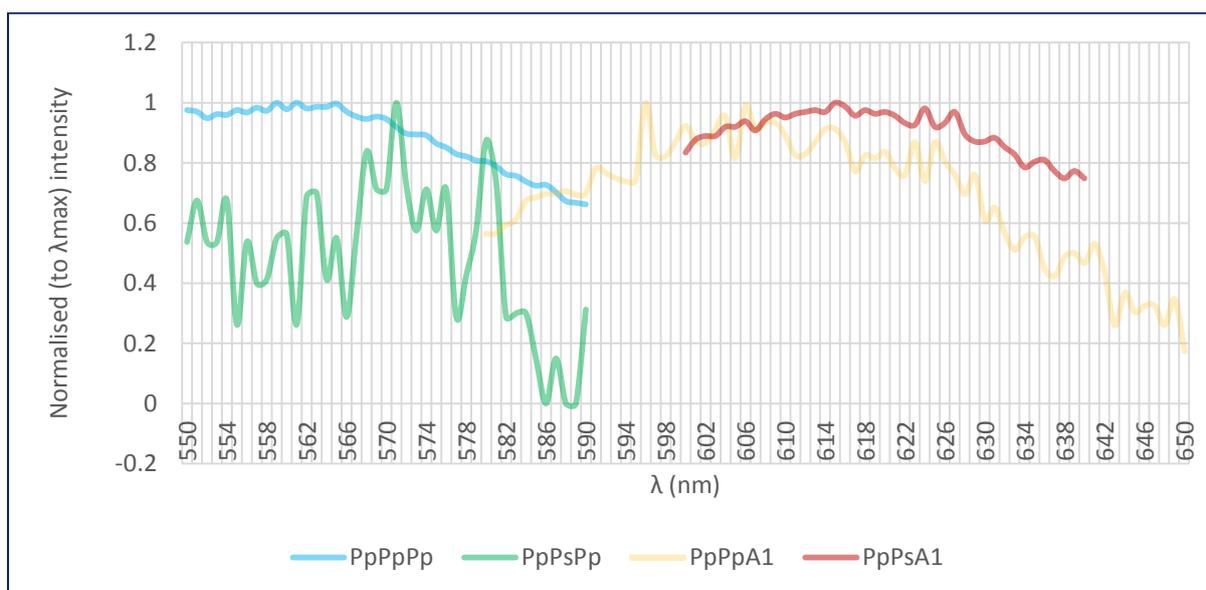


Figure 6.15. High resolution (1nm divisions) spectra screen of PpPpPp, PpPsPp, PpPpA1, and PpPsA1. The PpPsPp spectrum trace was somewhat erratic due to low signal, however the λ_{\max} was still ascertainable.

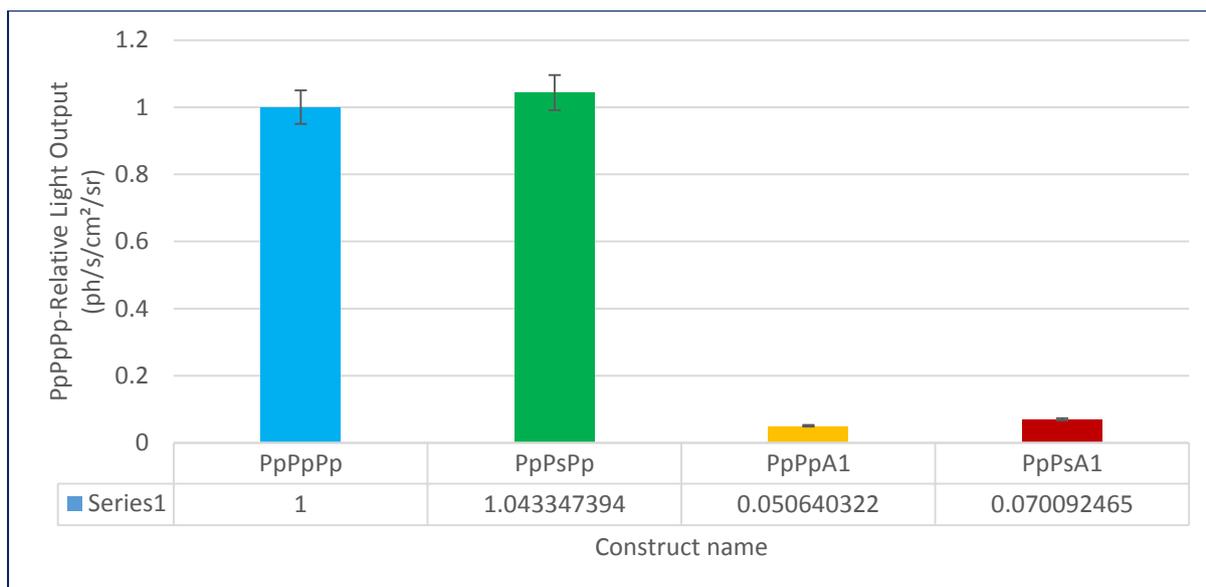


Figure 6.16. PpPpPp-relative light output of PpPsPp, PpPpA1, and PpPsA1; all *without* Hb phantom.

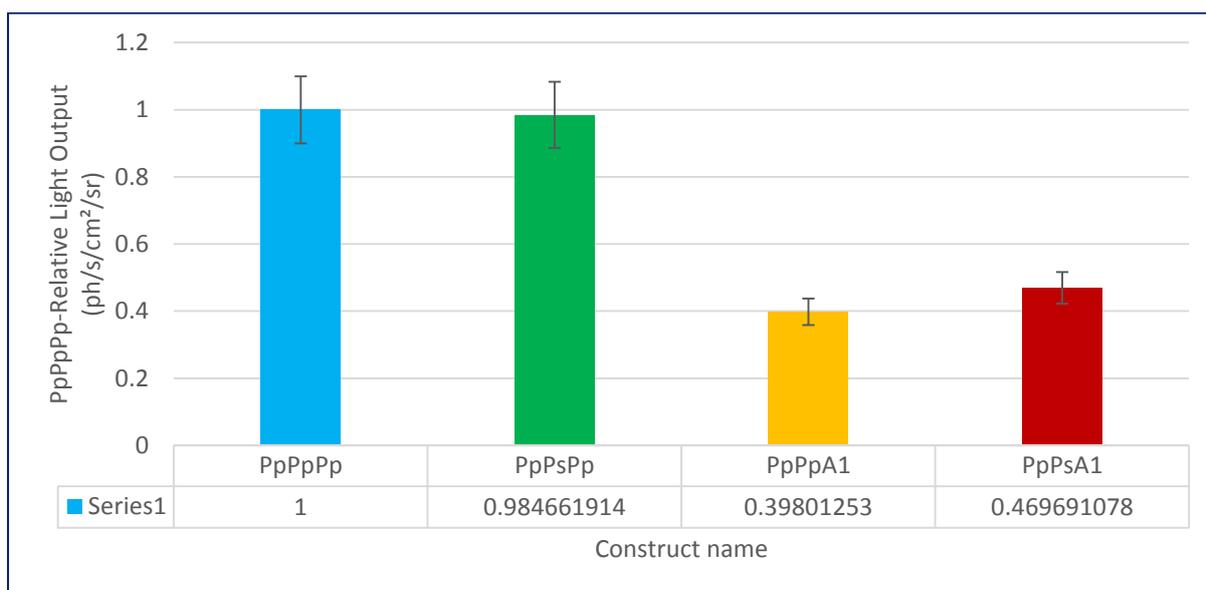


Figure 6.17. PpPpPp-relative light output of PpPsPp, PpPpA1, and PpPsA1; all *with* Hb phantom.

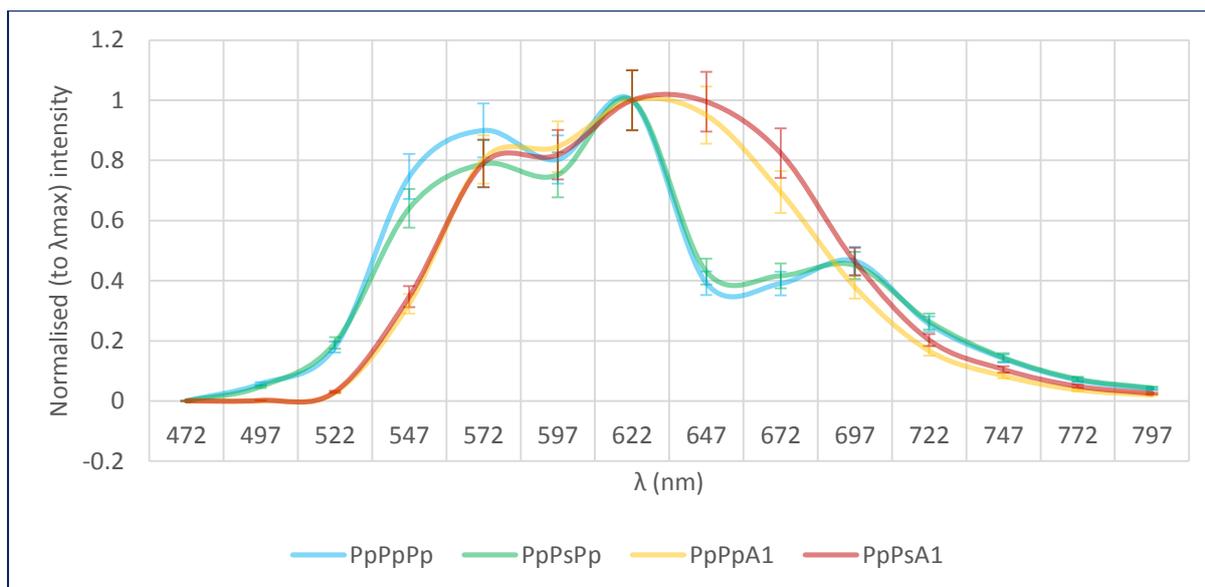


Figure 6.18. Normalised spectra of PpPpPp, PpPsPp, PpPpA1, and PpPsA1; all *without* Hb phantom.

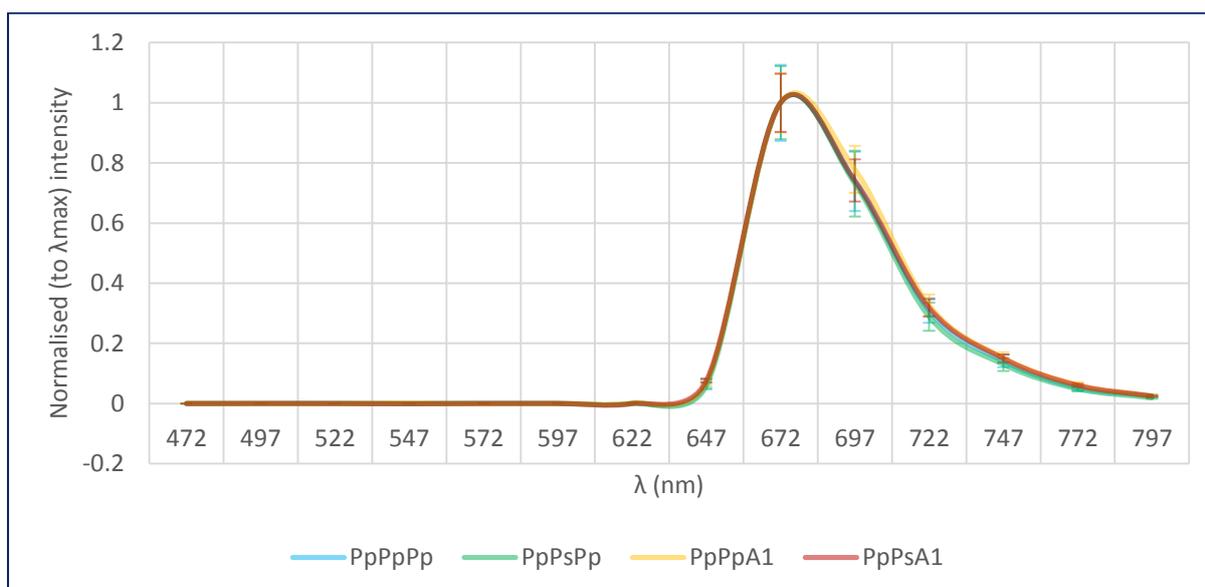


Figure 6.19. Normalised spectra of PpPpPp, PpPsPp, PpPpA1, and PpPsA1; all *with* Hb phantom.

6.4 Further Discussion

This work sought to address the remaining BALB/c major epitope in PpA1 via two strategies, μ D exchange, and epitope substitution.

6.4.1 Potential light output of the A1PpPpPpPpPpA1 construct

The construct A1PpPpPpPpPpA1 failed to bioluminesce at a detectable level when expressed in transiently transformed HEK cells. From a crude analysis, this is to be expected as the ACSM1 μ D8 causes a \sim 20 fold reduction in brightness relative to PpPpPp, and the ACSM1 μ D1 causes an almost 2000 fold reduction. Therefore, combining these μ Ds one might expect the construct to be almost 40,000 fold less bright than PpPpPp, which is seemingly beyond the sensitivity limit of the PhotonIMAGER when expressed in HEK cells. Protein purification from mammalian cells is much less efficient and more complicated than from *E. coli*. Had such a purification been performed, it may have been possible to quantify the light output of A1PpPpPpPpPpA1. Regardless, this construct would still have been discarded as too weakly bioluminescent to further engineer, especially given the relatively small increase in murine sequence over PpA1.

6.4.2 Impact of Hb filtration

In vitro in HEK cells the relative total brightness of PpA1 was consistently $<12\%$ that of PpPp when unfiltered (on average it was 5%). However, when filtered through a haemoglobin phantom this rose dramatically to $\sim 40\%$, demonstrating the importance of red-shifted emission for in vivo applications. Further, almost no light at wavelengths shorter than the bio-optical window (650 to 1200 nm) reached the camera (Figure 6.19). As previously stated, many red-shifted luciferin analogues exist, however these are dim when used by WT *Ppy* Fluc (confirmed with testing with PpA1). The Hb filtration work undertaken within this chapter confirmed its use as an accurate in vivo simulation.

6.5 Conclusions

For a *Ppy* Fluc-backbone construct, only μ D8 of *Ppy* Fluc (the CTD) can be exchanged with the ACSM1 equivalent μ D while retaining bioluminescence bright enough for potential in vivo use. Given that ACSM1 appears to be the optimal ACSM (Chapter 4), it is likely that this would hold true for μ Ds 1 – 7 of other ACSMs. The substitution of the G160 – V168 region of *Ppy* Fluc for that of *Psc* Fluc minimally affected the relative light output nor the spectra; this suggests that it is likely possible to substitute other (and less immunogenic) luciferases at this location while maintaining bioluminescence function, however further time and funding would be required

to explore this. The similar *Ppy* Fluc-relative light output of PpA1 when filtered through an Hb phantom versus through the living tissue of mice indicates that the advantage of PpA1 presently comes primarily from its red-shifted emission.

7 Chapter 7 – General Discussion, Conclusions, and Ideas for Future Work

7.1 Strategy Overview

The aim of this project was to use protein engineering techniques to improve the compatibility of firefly luciferase with mammalian systems and in particular to address the reported inherent immunogenicity of *Ppy* Fluc. The overall strategy was to make the luciferase sequence more similar to a mammalian protein and therefore less immunogenic. Two broad strategies were followed: the first was to mutate a mammalian enzyme with structural homology to luciferase to utilise luciferin as a substrate and emit light; and the second was to alter firefly luciferase, either to replace parts of the protein with amino acid sequence of mammalian origin or to remove predicted epitopic regions. The various constructs created, and the engineering progression of each is displayed in Figure 7.1.

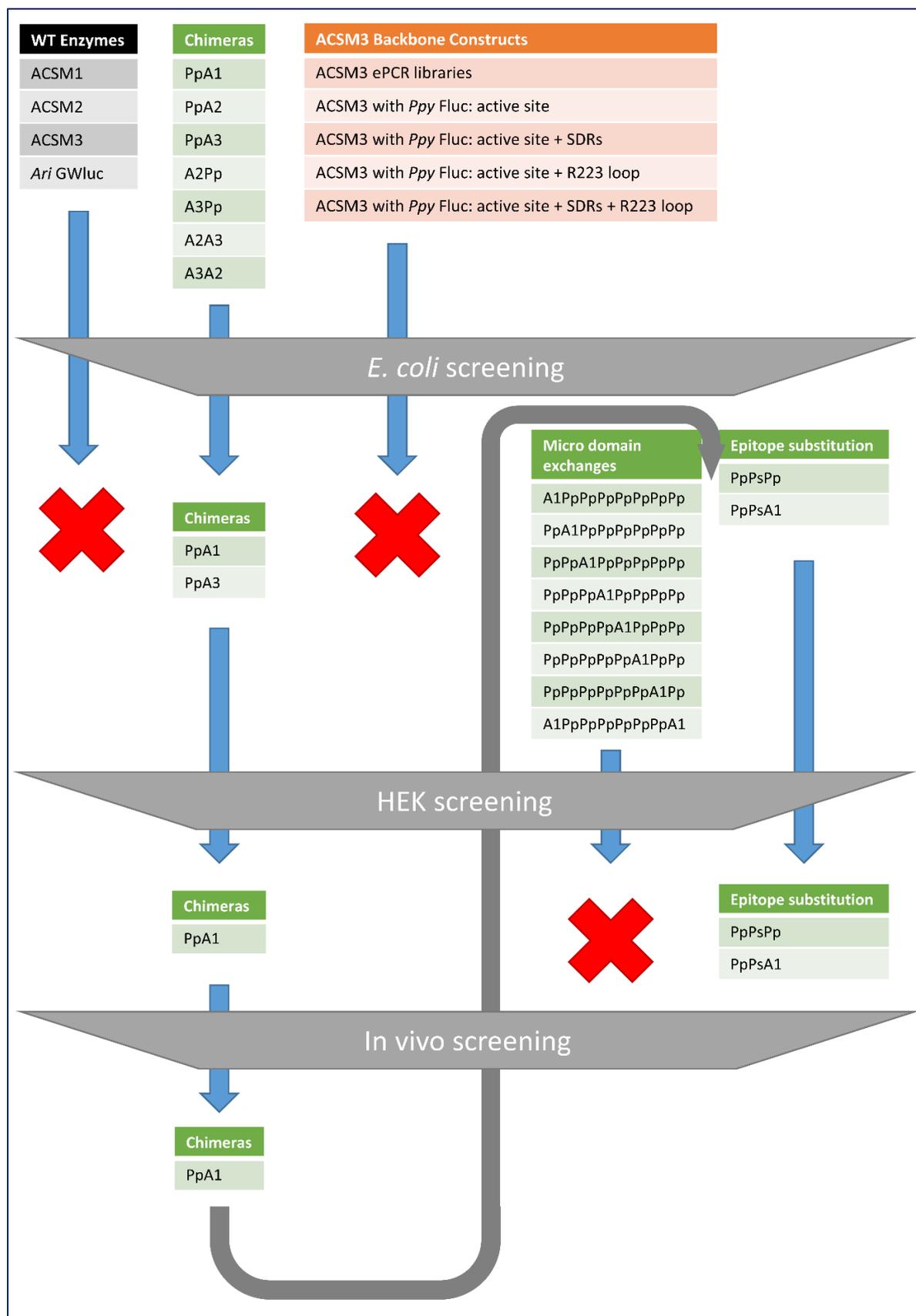


Figure 7.1. Flow chart of the progression and termination of constructs created within this project. A red X indicates that constructs were ultimately discarded.

Mammalian ACSs were identified as potential enzymes that might form a basis for both strategies. This was based on work previously published (Oba et al. 2009) showing that a single mutation of an insect ACS could produce a bioluminescent enzyme. The available literature initially reported that Fluc is analogous to a “true” long-chain ACS (optimal fatty acid length of C16) (Oba et al. 2003), however it was later revealed that it has greatest activity with fatty acids of length C12 (Oba et al. 2005);(Oba et al. 2006). Thus, the literature indicated that medium/long-chain ACSs most closely matched the activity of Fluc, and thus would be more amenable for conversion to a beetle luciferase. However, BLAST results indicated that for murine ACSs a refined medium/short-chain ACSM (ACSM3) was most homologous to *Ppy* Fluc. As such, it was decided to initially use ACSM3 for this project; this represents one of the only departures from the strategy that literature would have implied. ACSM1 (a promiscuous medium/long-chain ACSM that has 50% activity with C12 fatty acid) was introduced later in the project, and hybrids incorporating it dramatically outperformed those with ACSM3 in terms of brightness with D-LH₂. These results fortify the precedent set by the literature, and demonstrates the importance of adhering to it when possible. The mutants cloned from ACSM3 backbone all failed to endow bioluminescence function to ACSM3 (figure 7.1) (discussed separately and in greater detail later in this chapter). The domain-exchanged constructs created using the NTD of *Ppy* Fluc and CTD of ACSM1 and 3 retained bioluminescence bright enough to attempt improvements. The first improvement was to express these constructs in HEK cells rather than *E. coli* which improved light output to 10% for PpA1, partly due to increased protein expression efficiency over *Ppy* Fluc, which was bright enough to trial in live mice. The mouse experiment revealed that PpA1 was highly visible, and its *Ppy*-relative brightness increased further to 45% – which was likely entirely due to Hb filtration. The PpA1 construct had acceptable performance in terms of light emission, however its immunogenicity was unknown and likely unchanged as the major epitope of *Ppy* Fluc was still present. Further work was undertaken upon PpA1 to remove this potent epitope whilst retaining bioluminescence function, and a construct created which removed this epitope and incidentally improved Hb-filtered light output. This work is discussed in further detail within each chapter, general methodological difficulties are discussed here.

7.2 Technical and Methodological Challenges and Complications

Luciferases are known to both increase in brightness and red-shift at 37 °C, versus RT (Zhao et al. 2005). The included hot plate for the PhotonIMAGER was unstable and would fluctuate between ~34 °C and ~43 °C with approximately five mins between peaks. For in vivo experiments this is not an issue as the small animals are endotherms and also large enough that the fluctuations are negligible to their core temperature. However, for in vitro experiments – especially those with small volumes – the upper temperature will damage both *E. coli* and mammalian cells which could compromise longitudinal experiments. Further, the luciferases themselves would not encounter such high temperatures in vivo so it is not representative of the thermal stability required. A highly temperature-stable hot plate became available later in the project, but only after many in vitro experiments had been completed, although its relatively small size would have limited the number of plates imaged at once. As such, the accurate hot plate was not used whatsoever. The heating element of the included hot plate was not turned on (except for in vivo experiments), however the in vitro plates were still placed upon it during FLI and BLI acquisitions. The air temperature of the PhotonIMAGER room is relatively high and as such the stock hot plate would consistently report its temperature as ≥ 30 °C. A large and temperature-stable hot plate would have allowed a more meaningful screening of constructs in Chapter 4 prior to finally using mice. Nevertheless, the ACSM1 chimera was very clearly the superior murinised construct, even with the hot plate issues it was correct to advance it to the in vivo testing phase.

The IRES EGFP may not be truly independently translated. Transactivation is a phenomenon where the upstream mRNA recruits ribosomes and, when the ribosomes finish translating the upstream protein, they terminate in close physical proximity to the promoter for the downstream IRES gene. Therefore, efficiently translated upstream proteins may have a higher degree of transactivation of the downstream IRES EGFP. If this is the case, then it indicates that the chimeric constructs are more efficiently translated. This is supported by the data for the 1:1 expressed (with *Ppy* Fluc) mCherry detailed in Chapter 4 which found that PpA1 was expressed at ~2.6 that of PpPp. Another form of transactivation – known as “read through” is where the ribosome does not terminate at the stop codon of the upstream construct, and instead continues translation until the stop codon of the downstream construct is reached. In this case a continuous peptide is translated. It is for this reason that

many researchers choose to use a double stop codon, in case a ribosome skips the first. For this project a ribosome skipping the construct stop codon would lead to a construct linked at its C-terminus to the N-terminus of EGFP, with the IRES acting as an amorphous linker between the two. Flucs are known to tolerate C-terminal-linked tags (such as GST tags), and EGFP is known to tolerate N-terminal-linked constructs, so it is reasonable to assume that both proteins would remain somewhat functional. However functional, the exact efficiency loss – caused by a C-terminal tag – of the chimeric enzyme is unknown, and therefore it is possible that this phenomenon could cause the chimeric luciferases to appear less efficient than they would be if untagged. Given the wide use and understanding that this expression vector (pCCL 305), this phenomenon occurring to a degree high enough to meaningfully affect construct brightness is unlikely.

Bioluminescence screening of PpA1 in HEK cells was repeated within Chapter 6, however with reduced PpPp-relative brightness – PpA1 was ~10% that of PpPp in Chapter 4, but only ~4% in Chapter 6. Further, there were large discrepancies between experiments within the same chapters (table 6.5). There are a number of possible reasons for these discrepancies and are discussed here. The HEK cell screening required a large amount of plasmid DNA per transfection for screening, and as such multiple batches were harvested via DNA mini/midprep throughout the project. These harvested DNA preparations were expected to be of differing concentrations (yields), however this was accounted for by first quantifying with a NanoDrop spectrophotometer (see 2.2.1.3) and diluting appropriately for equimolar transfections. After dilution the samples were measured again and found to be all of the same concentration (within 5%). The NanoDrop is widely used within molecular biology and reveals both DNA concentration and approximate purity. The NanoDrop is not considered accurate for samples with DNA concentrations lower than 10ng/μL (according to manufacturer accuracy is between 2 – 15,000 ng/μL), but the PpA1 and PpPp preparations were always greater than 100 ng/μL and therefore this issue was assumed to be irrelevant. However, some researchers are sceptical of the results given by the NanoDrop even if above this threshold concentration, and instead opt to use other, more precise, machines such as the Agilent TapeStation. The TapeStation is not only more accurate and precise across a broader range on DNA concentrations, but can give more detailed information about the sample such as DNA lengths and contaminants. Regardless, given the simplicity and

established robustness of the DNA preparation technique and the careful steps taken to purify the sample, it was assumed that the final DNA concentration would be accurately quantified by the NanoDrop. Further, restriction digested plasmid DNA samples run on agarose gel did not show any exogenous DNA which could cause a sample to appear more concentrated; the DNA also digested into bands of the expected sizes, indicating plasmid integrity. Thus the TapeStation – which is more expensive and time consuming to use – was unfavoured. Further, as the plasmid preparations were consumed by transfection work, it is not possible to assess them via TapeStation. It seems unlikely that plasmid DNA concentration caused the differing *Ppy* Fluc-relative results of PpA1 between Chapters 4 and 6. An alternative explanation is HEK cell health which was a known variable, and it is not uncommon for HEK cells to degrade with each subculture. HEK cell *viability* was assessed prior to each transfection via Luna cell counter, however this simple technique cannot reveal the more subtle *health* differences which could affect protein expression efficiency. As two entirely different HEK call batches were used between Chapter 4 and 6 this explanation for differences in brightness is a distinct possibility. Further, different HEK subcultures were used throughout Chapter 4. In transiently-transfected HEK (and 4T1) cells both the IRES EGFP expression and construct brightness (relative to *Ppy* Fluc) varied substantially between individual experiments despite all factors remaining tightly controlled, further indicating a cell depreciation phenomenon. The purpose of the IRES EGFP is to allow the experimenter to compensate for minor differences in transfection efficiency and cell health between samples within the same experiment, and therefore achieve data with less variance. However, since the EGFP signal is consistently higher for the chimeric constructs this appears to be an inherent characteristic of these constructs, and EGFP compensation for BLI signal was ignored for all result conclusions.

In various experiments the pre-BLI EGFP signal from mammalian cells was not detectable above background levels despite producing a substantial bioluminescence signal, indicating that very high transfection efficiency is required for the PhotonIMAGER to detect FLI. This problem was encountered in Chapter 4 with failure to detect 1:1 expressed (with *Ppy* Fluc) mCherry, despite the mCherry fluorescent protein being clearly detectable by flow cytometry later. The PhotonIMAGER appears to have poor FLI capabilities in terms of sensitivity, it may

be advantageous to find a more sensitive yet still non-invasive machine for pre-BLI FLI screening.

7.3 Concluding Remarks

Bioluminescence with Δ -LH₂ is a highly refined enzymatic reaction which has evolved over millions of years. It is perhaps unsurprising that artificially recreating it with enzymes from organisms hundreds of millions of years apart phylogenetically was less fruitful than had optimistically been anticipated. Nevertheless, despite the high-risk high-reward strategy of ACSM-backbone mutants failing to produce bioluminescence, the chimera strategy created an enzyme confirmed as viable for in vivo imaging. With the limited time remaining in the project, this construct was theoretically improved further with negligible effect on brightness and colour. There are many further low-risk optimisations which could be incorporated, and unexplored strategies for identifying superior mutants based on PpA1 or PpPsA1.

7.4 Future Work

Due to the limited time remaining only one epitope substitution was tested in vitro for further testing in vivo. With adequate funding and lab time, many other candidate luciferases for epitope substitution could be screened in vitro. Further, the two minor BALB/c epitopes identified (Limberis et al. 2009) could also be substituted and screened. The final construct to continue into in vivo testing for bioluminescence and immunogenic properties would be the construct which provides the greatest theoretical immunogenic reduction while remaining functional. Immunogenicity can only be screened in vivo, as it is not yet possible to simulate the adaptive immune system accurately in vitro.

A large-scale in vivo experiment was planned, however this was not completed within the funding period for the PhD project, even with further extended funding it may not have been possible within the PhD project period. [The reason for delays was the repeated failure to create cell lines of the constructs, as the viral particles created via the third generation lentiviral system did not transduce cells. It is still unknown why this failed, as all components of the experiment were independently checked and deemed correct.] Therefore, future work for this project would be to complete this experiment. The experiment aim is to first create

virally-transduced HEK and 4T1 cell lines of PpPpPp, PpPsPp, PpPpA1, and PpPsA1 (although likely with an in vitro-screened, less immunogenic Fluc than *Psc*). The 4T1 cell lines are then to be implanted into immunocompetent BALB/c mice, and the HEK cell lines implanted into immunodeficient NSG mice. From that point on the HEK cell line half of this experiment is to be carried out identically to the previously completed in vivo HEK cell line experiment (Chapter 5), but with larger cohort sizes. As before, this half of the experiment would reveal the in vivo brightness, kinetics, and colour of the implanted constructs. The 4T1 side of the experiment would involve a period – of approximately three weeks – of post-implant incubation to allow the immune system to respond to the introduced foreign constructs, specifically, to allow the expression of anti-construct CD8+ T cells. The immunocompetent nature of the BALB/c mouse strain allows the implanted mice to be handled outside of a sterile environment. This allows the longitudinal monitoring of the implanted cells over the three week period, as the only difference between the cohorts is the construct; any statistically significant relative difference in implant size between cohorts can likely be attributed to the different immunogenic levels of the constructs. Finally, the BALB/c mice would be harvested and the splenocytes harvested. Then, a cytotoxicity assay would be performed similarly to Limberis *et al.* in which construct-expressing 4T1 cells are combined with splenocytes from the sacrificed mice. If anti-construct cytotoxic T cells are present it will lead to the killing of the construct-expressing 4T1 cells. There are a number of methods available for quantifying the level of killing such as chromium release or non-Fluc luciferase release, and revolve around detecting the lysis of cells as they are killed.

It is clear that the chimeric constructs are far better expressed (relative to WT *Ppy* Fluc) in mammalian cells than *E. coli*. This is unfortunate as directed evolution techniques are extremely difficult to perform in mammalian cells due to the fact that both single cell or “colony” screening and selection is microscopic in scale. “Colony” screening in such cells typically depends on the use of FAC-sorting into the wells of a plate, and screening the “colony forming units”; the positive wells can then be subcultured. Generally speaking, yeast is often a more capable organism for exogenous protein expression and is only moderately more difficult to work with than *E. coli*. Thus it, or a similar organism, may better lend itself to the high-throughput engineering of murine ACS-containing constructs.

After the major and minor BALB/c epitopes have been addressed, the engineering steps thereafter would be as follows. As previously stated, *Ppy* Fluc is known to be thermolabile and its emitted light heavily filtered by Hb. As a result, researchers rarely use the WT *Ppy* Fluc for in vivo work, but mutants which are pH and thermo stable, and which are optimised for use with luciferin analogues which emit within the bio-optical window. The mutations for pH and thermal stability are known (Jathoul et al. 2012), as are many which increase its brightness (Halliwell et al. 2018). Luciferin analogues such as iLH₂ are dim with WT *Ppy* Fluc (even with Hb filtration) but there are known mutations which can optimise the enzyme for use with this substrate. Thus, of all the aforementioned mutations, those which occur in the NTD of *Ppy* Fluc could be included with PpA1 and the benefits likely endowed.

The various chimeric constructs created and screened in Chapters 4, 5, and 6 were not subjected to any random mutagenesis. This was due to the assumption that broad rational mutations would give more rapid design progression. This could represent a further avenue for exploration.

8 References

Ando, Y. et al. 2008. Firefly bioluminescence quantum yield and colour change by pH-sensitive green emission. *Nature Photonics* 2(1), pp. 44–47. doi: 10.1038/nphoton.2007.251.

Ansari, A.M. et al. 2016. Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. *Stem Cell Reviews and Reports* 12(5), pp. 553–559. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27435468> [Accessed: 26 February 2020].

Auld, D.S. et al. 2009. A basis for reduced chemical library inhibition of firefly luciferase obtained from directed evolution. *Journal of Medicinal Chemistry* 52(5), pp. 1450–1458. Available at: <https://pubmed.ncbi.nlm.nih.gov/19215089/> [Accessed: 6 October 2020].

Baggett, B., Roy, R., Momen, S., Morgan, S., Tisi, L., Morse, D. and Gillies, R.J. 2004. Thermostability of firefly luciferases affects efficiency of detection by in vivo bioluminescence. *Molecular imaging* 3(4), pp. 324–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15802049> [Accessed: 26 February 2020].

Bechara, E.J.H. and Stevani, C. V. 2018. Brazilian bioluminescent beetles: Reflections on catching glimpses of light in the Atlantic Forest and Cerrado. *Anais da Academia Brasileira de Ciências* 90(1), pp. 663–679. doi: 10.1590/0001-3765201820170504.

Bradbury, M.S. et al. 2007. Optical bioluminescence imaging of human ES cell progeny in the rodent CNS. *Journal of neurochemistry* 102(6), pp. 2029–2039. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17555555> [Accessed: 26 February 2020].

Branchini, B.R. et al. 2017. Cloning of the Orange Light-Producing Luciferase from *Photinus scintillans*—A New Proposal on how Bioluminescence Color is Determined. *Photochemistry and Photobiology* 93(2), pp. 479–485. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27861940> [Accessed: 26 February 2020].

Branchini, B.R., Ablamsky, D.M., Murtiashaw, M.H., Uzasci, L., Fraga, H. and Southworth, T.L. 2007. Thermostable red and green light-producing firefly luciferase mutants for bioluminescent reporter applications. *Analytical Biochemistry* 361(2), pp. 253–262. doi:

References

10.1016/j.ab.2006.10.043.

Branchini, B.R., Magyar, R.A., Murtiashaw, M.H., Anderson, S.M., Helgerson, L.C. and Zimmer, M. 1999. Site-directed mutagenesis of firefly luciferase active site amino acids: A proposed model for bioluminescence color. *Biochemistry* 38(40), pp. 13223–13230. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10529195> [Accessed: 26 February 2020].

Branchini, B.R., Southworth, T.L., Fontaine, D.M., Davis, A.L., Behney, C.E. and Murtiashaw, M.H. 2014. A photinus pyralis and Luciola italica chimeric firefly luciferase produces enhanced bioluminescence. *Biochemistry* 53(40), pp. 6287–6289. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25264115> [Accessed: 26 February 2020].

Brutkiewicz, S. et al. 2007. The expression level of luciferase within tumour cells can alter tumour growth upon in vivo bioluminescence imaging. *Luminescence* 22(3), pp. 221–228. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17286245> [Accessed: 10 March 2020].

Chavelas-Adame, E., Hernández-Domínguez, E.E., Gaytán-Mondrangón, S., Rosales-León, L., Valencia-Turcotte, L. and Rodríguez-Sotres, R. 2011. *A Hitchhiker's Guide to the Modeling of the Three-Dimensional Structure of Proteins*. Available at: <http://salilab.org/modeller/examples/scoring/optimize.py> [Accessed: 26 February 2020].

Chou, T.C. and Moyle, R.L. 2014. Synthetic versions of firefly luciferase and Renilla luciferase reporter genes that resist transgene silencing in sugarcane. *BMC Plant Biology* 14(1), p. 92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24708613> [Accessed: 26 February 2020].

Choy, G., O'Connor, S., Diehn, F.E., Costouros, N., Alexander, H.R., Choyke, P. and Libutti, S.K. 2003. Comparison of noninvasive fluorescent and bioluminescent small animal optical imaging. *BioTechniques* 35(5), pp. 1022–1030. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14628676> [Accessed: 26 February 2020].

Christou, C. and Parks, R.J. 2011. Rational design of murine secreted alkaline phosphatase for enhanced performance as a reporter gene in mouse gene therapy preclinical studies. *Human Gene Therapy* 22(4), pp. 499–506. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21083426> [Accessed: 26 February 2020].

References

Coco, W.M. et al. 2001. DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nature Biotechnology* 19(4), pp. 354–359. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11283594> [Accessed: 26 February 2020].

Coco, W.M. 2003. RACHITT: Gene family shuffling by Random Chimeragenesis on Transient Templates. *Methods in molecular biology (Clifton, N.J.)* 231, pp. 111–27. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12824609> [Accessed: 26 February 2020].

Conti, E., Franks, N.P. and Brick, P. 1996. Crystal structure of firefly luciferase throws light on a super-family of adenylate-forming enzymes. *Structure* 4(3), pp. 287–298. doi: 10.1016/S0969-2126(96)00033-0.

Coulet, F., Nadaud, S., Agrapart, M. and Soubrier, F. 2003. Identification of Hypoxia-response Element in the Human Endothelial Nitric-oxide Synthase Gene Promoter. *Journal of Biological Chemistry* 278(47), pp. 46230–46240. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12963737> [Accessed: 10 March 2020].

Cramer, A., Whitehorn, E.A., Tate, E. and Stemmer, W.P.C. 1996. Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* 14(3), pp. 315–x1. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9630892> [Accessed: 26 February 2020].

Day, C.P. et al. 2014. Glowing head" mice: A genetic tool enabling reliable preclinical image-Based evaluation of cancers in immunocompetent allografts. *PLoS ONE* 9(11). doi: 10.1371/journal.pone.0109956.

Denburg, J.L., Lee, R.T. and McElroy, W.D. 1969. Substrate-binding properties of firefly luciferase. I. Luciferin-binding site. *Archives of Biochemistry and Biophysics* 134(2), pp. 381–394. doi: 10.1016/0003-9861(69)90297-5.

Dubey, P. 2012. Reporter gene imaging of immune responses to cancer: Progress and challenges. *Theranostics* 2(4), pp. 355–362. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22509199> [Accessed: 26 February 2020].

Duchatelet, L., Delroisse, J., Flammang, P., Mahillon, J. and Mallefet, J. 2019. Etmopterus spinax, the velvet belly lanternshark, does not use bacterial luminescence. *Acta Histochemica*

References

121(4), pp. 516–521. Available at: <https://pubmed.ncbi.nlm.nih.gov/31027729/> [Accessed: 24 June 2020].

Dull, T. et al. 1998. A third-generation lentivirus vector with a conditional packaging system. *JOURNAL OF VIROLOGY*, pp. 8463--8471. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.322.173> [Accessed: 1 September 2020].

Fleiss, A. and Sarkisyan, K.S. 2019. A brief review of bioluminescent systems (2019). *Current Genetics* 65(4), pp. 877–882. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/30850867> [Accessed: 17 April 2020].

Fraga, H., Esteves da Silva, J.C.G. and Fontes, R. 2004. Identification of luciferyl adenylate and luciferyl coenzyme a synthesized by firefly luciferase. *Chembiochem : a European journal of chemical biology* 5(1), pp. 110–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14695520> [Accessed: 26 February 2020].

Fraga, H., Fernandes, D., Fontes, R. and Esteves Da Silva, J.C.G. 2005. Coenzyme A affects firefly luciferase luminescence because it acts as a substrate and not as an allosteric effector. *FEBS Journal* 272(20), pp. 5206–5216. Available at: <https://pubmed.ncbi.nlm.nih.gov/16218952/> [Accessed: 1 November 2020].

Fujino, T. et al. 2001. Molecular Identification and Characterization of Two Medium-chain Acyl-CoA Synthetases, MACS1 and the Sa Gene Product. *Journal of Biological Chemistry* 276(38), pp. 35961–35966. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11470804> [Accessed: 26 February 2020].

Gandelman, O.A. et al. 2010. Novel bioluminescent quantitative detection of nucleic acid amplification in real-time. *PLoS ONE* 5(11), p. e14155. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21152399> [Accessed: 26 February 2020].

Halliwell, L.M., Jathoul, A.P., Bate, J.P., Worthy, H.L., Anderson, J.C., Jones, D.D. and Murray, J.A.H. 2018. ΔFlucs: Brighter *Photinus pyralis* firefly luciferases identified by surveying consecutive single amino acid deletion mutations in a thermostable variant. *Biotechnology and bioengineering* 115(1), pp. 50–59. Available at:

References

<http://www.ncbi.nlm.nih.gov/pubmed/28921549> [Accessed: 26 February 2020].

Heckman, K.L. and Pease, L.R. 2007. Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols* 2(4), pp. 924–932. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17446874> [Accessed: 26 February 2020].

Iwano, S. et al. 2018. Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science* 359(6378), pp. 935–939. Available at: <http://science.sciencemag.org/> [Accessed: 28 September 2020].

Jathoul, A., Law, E., Gandelman, O., Pule, M., Tisi, L. and Murray, J. 2012. Development of a pH-Tolerant Thermostable Photinus pyralis Luciferase for Brighter In Vivo Imaging. In: *Bioluminescence - Recent Advances in Oceanic Measurements and Laboratory Applications*. InTech. doi: 10.5772/37170.

Jathoul, A.P., Grounds, H., Anderson, J.C. and Pule, M.A. 2014. A dual-color far-red to near-infrared firefly luciferin analogue designed for multiparametric bioluminescence imaging. *Angewandte Chemie (International ed. in English)* 53(48), pp. 13059–63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25266918> [Accessed: 26 February 2020].

Khalil, A.A. et al. 2013. Subcutaneous Administration of D-Luciferin is an Effective Alternative to Intraperitoneal Injection in Bioluminescence Imaging of Xenograft Tumors in Nude Mice. *ISRN Molecular Imaging* 2013, pp. 1–7. Available at: [/pmc/articles/PMC4226276/?report=abstract](http://pmc/articles/PMC4226276/?report=abstract) [Accessed: 14 November 2020].

Khurana, P., Gokhale, R.S. and Mohanty, D. 2010. Genome scale prediction of substrate specificity for acyl adenylate superfamily of enzymes based on active site residue profiles. *BMC Bioinformatics* 11, p. 57. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20105319> [Accessed: 26 February 2020].

Kiddle, G. et al. 2012. GMO detection using a bioluminescent real time reporter (BART) of loop mediated isothermal amplification (LAMP) suitable for field use. *BMC biotechnology* 12, p. 15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22546148> [Accessed: 26 February 2020].

Kilkenny, C., Browne, W.J., Cuthill, I.C., Emerson, M. and Altman, D.G. 2010. Improving

References

bioscience research reporting: The arrive guidelines for reporting animal research. *PLoS Biology* 8(6). doi: 10.1371/journal.pbio.1000412.

Kim, J.H. et al. 2011. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS ONE* 6(4), p. e18556. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21602908> [Accessed: 5 March 2020].

Kitayama, A., Yoshizaki, H., Ohmiya, Y., Ueda, H. and Nagamune, T. 2003. Creation of a thermostable firefly luciferase with pH-insensitive luminescent color. *Photochemistry and photobiology* 77(3), pp. 333–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12685663> [Accessed: 26 February 2020].

Kobayashi, K.S. and Van Den Elsen, P.J. 2012. NLRC5: A key regulator of MHC class I-dependent immune responses. *Nature Reviews Immunology* 12(12), pp. 813–820. doi: 10.1038/nri3339.

Law, G.H.E., Gandelman, O.A., Tisi, L.C., Lowe, C.R. and Murray, J.A.H. 2006. Mutagenesis of solvent-exposed amino acids in *Photinus pyralis* luciferase improves thermostability and pH-tolerance. *The Biochemical journal* 397(2), pp. 305–12. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16551268> [Accessed: 3 March 2020].

Lee, A.S. et al. 2009. Effects of cell number on teratoma formation by human embryonic stem cells. *Cell Cycle* 8(16), pp. 2608–2612. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19597339> [Accessed: 26 February 2020].

Lehmann, S. et al. 2009. Longitudinal and multimodal in vivo imaging of tumor hypoxia and its downstream molecular events. *Proceedings of the National Academy of Sciences of the United States of America* 106(33), pp. 14004–14009. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19666490> [Accessed: 10 March 2020].

Limberis, M.P., Bell, C.L. and Wilson, J.M. 2009. Identification of the murine firefly luciferase-specific CD8 T-cell epitopes. *Gene therapy* 16(3), pp. 441–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19129859> [Accessed: 26 February 2020].

Lin-Goerke, J.L., Robbins, D.J. and Burczak, J.D. 1997. PCRr-based random mutagenesis using manganese and reduced DNTP concentration. *BioTechniques* 23(3), pp. 409–412. doi:

References

10.2144/97233bm12.

Mælandsmo, G.M. et al. 2005. Use of a murine secreted alkaline phosphatase as a non-immunogenic reporter gene in mice. *Journal of Gene Medicine* 7(3), pp. 307–315. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15515146> [Accessed: 26 February 2020].

Maghami, P., Ranjbar, B., Hosseinkhani, S., Ghasemi, A., Moradi, A. and Gill, P. 2010. Relationship between stability and bioluminescence color of firefly luciferase. *Photochemical and Photobiological Sciences* 9(3), pp. 376–383. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20221465> [Accessed: 26 February 2020].

McElroy, W.D. and Green, A.A. 1955. Enzymatic properties of bacterial luciferase. *Archives of Biochemistry and Biophysics* 56(1), pp. 240–255. Available at: <https://pubmed.ncbi.nlm.nih.gov/14377570/> [Accessed: 24 June 2020].

McELROY, W.D. and STREHLER, B.L. 1954. Bioluminescence. *Bacteriological reviews* 18(3), pp. 177–184. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC180800/> [Accessed: 24 June 2020].

Meister, G. and Tuschl, T. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature* 431(7006), pp. 343–349. doi: 10.1038/nature02873.

Minamoto, T., Wada, E. and Shimizu, I. 2012. A new method for random mutagenesis by error-prone polymerase chain reaction using heavy water. *Journal of Biotechnology* 157(1), pp. 71–74. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21963598> [Accessed: 26 February 2020].

Mofford, D.M., Liebmann, K.L., Sankaran, G.S., Reddy, G.S.K.K., Reddy, G.R. and Miller, S.C. 2017. Luciferase Activity of Insect Fatty Acyl-CoA Synthetases with Synthetic Luciferins. *ACS chemical biology* 12(12), pp. 2946–2951. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29073357> [Accessed: 26 February 2020].

Mofford, D.M., Reddy, G.R. and Miller, S.C. 2014a. Aminoluciferins extend firefly luciferase bioluminescence into the near-infrared and can be preferred substrates over d-luciferin. *Journal of the American Chemical Society* 136(38), pp. 13277–13282. doi: 10.1021/ja505795s.

References

Mofford, D.M., Reddy, G.R. and Miller, S.C. 2014b. Latent luciferase activity in the fruit fly revealed by a synthetic luciferin. *Proceedings of the National Academy of Sciences of the United States of America* 111(12), pp. 4443–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24616520> [Accessed: 26 February 2020].

Morise, H., Shimomura, O., Johnson, F.H. and Winant, J. 1974. Intermolecular energy transfer in the bioluminescent system of *aequorea*. *Biochemistry* 13(12), pp. 2656–2662. Available at: <https://pubs.acs.org/doi/abs/10.1021/bi00709a028> [Accessed: 24 June 2020].

Oba, Y., Iida, K. and Inouye, S. 2009. Functional conversion of fatty acyl-CoA synthetase to firefly luciferase by site-directed mutagenesis: a key substitution responsible for luminescence activity. *FEBS letters* 583(12), pp. 2004–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19450587> [Accessed: 26 February 2020].

Oba, Y., Ojika, M. and Inouye, S. 2003. Firefly luciferase is a bifunctional enzyme: ATP-dependent monooxygenase and a long chain fatty acyl-CoA synthetase. *FEBS Letters* 540(1–3), pp. 251–254. doi: 10.1016/S0014-5793(03)00272-2.

Oba, Y., Sato, M., Ojika, M. and Inouye, S. 2005. Enzymatic and genetic characterization of firefly luciferase and *Drosophila* CG6178 as a fatty acyl-CoA synthetase. *Bioscience, biotechnology, and biochemistry* 69(4), pp. 819–28. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15849423> [Accessed: 26 February 2020].

Oba, Y., Tanaka, K. and Inouye, S. 2006. Catalytic properties of domain-exchanged chimeric proteins between firefly luciferase and *Drosophila* fatty acyl-CoA synthetase CG6178. *Bioscience, Biotechnology and Biochemistry* 70(11), pp. 2739–2744. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17090919> [Accessed: 26 February 2020].

Pellegatti, P., Falzoni, S., Pinton, P., Rizzuto, R. and Di Virgilio, F. 2005. A novel recombinant plasma membrane-targeted luciferase reveals a new pathway for ATP secretion. *Molecular Biology of the Cell* 16(8), pp. 3659–3665. doi: 10.1091/mbc.E05-03-0222.

Percie du Sert, N. et al. 2017. The Experimental Design Assistant. *PLOS Biology* 15(9), p. e2003779. Available at: <https://dx.plos.org/10.1371/journal.pbio.2003779> [Accessed: 31 March 2020].

References

Podetz-Pedersen, K.M., Vezys, V., Somia, N. V, Russell, S.J. and McIvor, R.S. 2014. Cellular immune response against firefly luciferase after sleeping beauty-mediated gene transfer in vivo. *Human gene therapy* 25(11), pp. 955–65. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25093708> [Accessed: 26 February 2020].

Polyethylenimine (PEI), linear (1 mg/mL). 2008. *Cold Spring Harbor Protocols* 2008(4), p. pdb.rec11323-pdb.rec11323. doi: 10.1101/pdb.rec11323.

Prado, R.A., Santos, C.R., Kato, D.I., Murakami, M.T. and Viviani, V.R. 2016. The dark and bright sides of an enzyme: a three dimensional structure of the N-terminal domain of Zophobas morio luciferase-like enzyme, inferences on the biological function and origin of oxygenase/luciferase activity. *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* 15(5), pp. 654–65. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27101527> [Accessed: 26 February 2020].

Reed, S.E., Staley, E.M., Mayginnes, J.P., Pintel, D.J. and Tullis, G.E. 2006. Transfection of mammalian cells using linear polyethylenimine is a simple and effective means of producing recombinant adeno-associated virus vectors. *Journal of Virological Methods* 138(1–2), pp. 85–98. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16950522> [Accessed: 11 April 2020].

Regev-Rudzki, N., Yogev, O. and Pines, O. 2008. The mitochondrial targeting sequence tilts the balance between mitochondrial and cytosolic dual localization. *Journal of Cell Science* 121, pp. 2423–2431. Available at: <http://ihg.gsf.de/ihg/mitoprot.html>.

Romero, P.A. and Arnold, F.H. 2009. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* 10(12), pp. 866–876. doi: 10.1038/nrm2805.

Shapiro, E., Lu, C. and Baneyx, F. 2005. A set of multicolored Photinus pyralis luciferase mutants for in vivo bioluminescence applications. *Protein Engineering, Design and Selection* 18(12), pp. 581–587. Available at: <https://pubmed.ncbi.nlm.nih.gov/16243898/> [Accessed: 19 April 2021].

Shultz, L.D. et al. 2005. Human Lymphoid and Myeloid Cell Development in NOD/LtSz- scid IL2R γ null Mice Engrafted with Mobilized Human Hemopoietic Stem Cells . *The Journal of*

References

Immunology 174(10), pp. 6477–6489. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15879151> [Accessed: 24 April 2020].

Si, M., Xu, Q., Jiang, L. and Huang, H. 2016. Spytag/spycatcher cyclization enhances the thermostability of firefly luciferase. *PLoS ONE* 11(9). Available at: [/pmc/articles/PMC5033358/?report=abstract](http://pmc/articles/PMC5033358/?report=abstract) [Accessed: 28 September 2020].

Silva, J.R., Amaral, D.T., Hastings, J.W., Wilson, T. and Viviani, V.R. 2015. A transcriptional and proteomic survey of *Arachnocampa luminosa* (Diptera: Keroplatidae) lanterns gives insights into the origin of bioluminescence from the Malpighian tubules in Diptera. *Luminescence* 30(7), pp. 996–1003. Available at: <https://pubmed.ncbi.nlm.nih.gov/25676901/> [Accessed: 19 June 2020].

Strehler, B.L., Harvey, E.N., Chang, J.J. and Cormier, M.J. 1954. THE LUMINESCENT OXIDATION OF REDUCED RIBOFLAVIN OR REDUCED RIBOFLAVIN PHOSPHATE IN THE BACTERIAL LUCIFERIN-LUCIFERASE REACTION. *Proceedings of the National Academy of Sciences* 40(1), pp. 10–12. Available at: <https://europepmc.org/articles/PMC527928> [Accessed: 24 June 2020].

Sundlov, J.A., Fontaine, D.M., Southworth, T.L., Branchini, B.R. and Gulick, A.M. 2012. Crystal structure of firefly luciferase in a second catalytic conformation supports a domain alternation mechanism. *Biochemistry* 51(33), pp. 6493–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22852753> [Accessed: 26 February 2020].

Tiffen, J.C., Bailey, C.G., Ng, C., Rasko, J.E.J. and Holst, J. 2010. Luciferase expression and bioluminescence does not affect tumor cell growth in vitro or in vivo. *Molecular Cancer* 9, p. 299. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21092230> [Accessed: 26 February 2020].

TISI, L.C., LAW, G.H., GANDELMAN, O., LOWE, C.R. and MURRAY, J. 2002. THE BASIS OF THE BATHOCHROMIC SHIFT IN THE LUCIFERASE FROM PHOTINUS PYRALIS . World Scientific Pub Co Pte Lt, pp. 57–60. doi: 10.1142/9789812776624_0012.

Trowell, S.C., Dacres, H., Dumancic, M.M., Leitch, V. and Rickards, R.W. 2016. Molecular basis for the blue bioluminescence of the Australian glow-worm *Arachnocampa richardsae*

References

(Diptera: Keroplatidae). *Biochemical and Biophysical Research Communications* 478(2), pp. 533–539. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27457804> [Accessed: 26 February 2020].

Viviani, V.R. 2002. The origin, diversity, and structure function relationships of insect luciferases. *Cellular and Molecular Life Sciences* 59(11), pp. 1833–1850. doi: 10.1007/PL00012509.

Viviani, V.R., Amaral, D., Prado, R. and Arnoldi, F.G.C. 2011. A new blue-shifted luciferase from the Brazilian *Amydetes fanestratus* (Coleoptera: Lampyridae) firefly: Molecular evolution and structural/functional properties. *Photochemical and Photobiological Sciences* 10(12), pp. 1879–1886. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21983629> [Accessed: 6 May 2020].

Viviani, V.R. and Bechara, E.J.H. 1996. Larval *Tenebrio molitor* (Coleoptera: Tenebrionidae) Fat Body Extracts Catalyze Firefly D-Luciferin- and ATP-Dependent Chemiluminescence: A Luciferase-like Enzyme. *Photochemistry and Photobiology* 63(6), pp. 713–718. Available at: <http://doi.wiley.com/10.1111/j.1751-1097.1996.tb09620.x> [Accessed: 28 June 2020].

Viviani, V.R. and Ohmiya, Y. 2006. Bovine serum albumin displays luciferase-like activity in presence of luciferyl adenylate: Insights on the origin of protoluciferase activity and bioluminescence colours. *Luminescence* 21(4), pp. 262–267. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16791835> [Accessed: 17 April 2020].

Viviani, V.R., Prado, R.A., Arnoldi, F.C.G. and Abdalla, F.C. 2009. An ancestral luciferase in the Malpighi tubules of a non-bioluminescent beetle! *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology* 8(1), pp. 57–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19247530> [Accessed: 26 February 2020].

Viviani, V.R., Prado, R.A., Neves, D.R., Kato, D. and Barbosa, J.A. 2013. A route from darkness to light: emergence and evolution of luciferase activity in AMP-CoA-ligases inferred from a mealworm luciferase-like enzyme. *Biochemistry* 52(23), pp. 3963–73. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23705763> [Accessed: 26 February 2020].

References

Viviani, V.R., Silva Neto, A.J., Arnoldi, F.G.C., Barbosa, J.A.R.G. and Ohmiya, Y. 2008. The influence of the loop between residues 223-235 in beetle luciferase bioluminescence spectra: a solvent gate for the active site of pH-sensitive luciferases. *Photochemistry and photobiology* 84(1), pp. 138–44. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18173713> [Accessed: 26 February 2020].

Watkins, O.C., Sharpe, M.L., Perry, N.B. and Krause, K.L. 2018. New Zealand glowworm (*Arachnocampa luminosa*) bioluminescence is produced by a firefly-like luciferase but an entirely new luciferin. *Scientific reports* 8(1), p. 3278. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29459729> [Accessed: 26 February 2020].

Wood, K. V. 1995. THE CHEMICAL MECHANISM and EVOLUTIONARY DEVELOPMENT OF BEETLE BIOLUMINESCENCE. *Photochemistry and Photobiology* 62(4), pp. 662–673. Available at: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1751-1097.1995.tb08714.x> [Accessed: 28 June 2020].

Wood, K. V. and DeLuca, M. 1987. Photographic detection of luminescence in *Escherichia coli* containing the gene for firefly luciferase. *Analytical Biochemistry* 161(2), pp. 501–507. doi: 10.1016/0003-2697(87)90480-5.

Wood, K. V., Lam, Y.A., McElroy, W.D. and Seliger, H.H. 1989. Bioluminescent click beetles revisited. *Journal of Bioluminescence and Chemiluminescence* 4(1), pp. 31–39. Available at: <http://doi.wiley.com/10.1002/bio.1170040110> [Accessed: 14 June 2020].

Yamauchi, T. and Moroishi, T. 2019. Hippo Pathway in Mammalian Adaptive Immune System. *Cells* 8(5), p. 398. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/31052239> [Accessed: 22 April 2020].

Zako, T., Ayabe, K., Aburatani, T., Kamiya, N., Kitayama, A., Ueda, H. and Nagamune, T. 2003. Luminescent and substrate binding activities of firefly luciferase N-terminal domain. *Biochimica et Biophysica Acta - Proteins and Proteomics* 1649(2), pp. 183–189. doi: 10.1016/S1570-9639(03)00179-1.

Zhang, S.J. and Wu, J.C. 2007. Comparison of imaging techniques for tracking cardiac stem cell therapy. *Journal of Nuclear Medicine* 48(12), pp. 1916–1919. Available at:

References

<http://www.ncbi.nlm.nih.gov/pubmed/18056330> [Accessed: 26 February 2020].

Zhao, H., Doyle, T.C., Coquoz, O., Kalish, F., Rice, B.W. and Contag, C.H. 2005. Emission spectra of bioluminescent reporters and interaction with mammalian tissue determine the sensitivity of detection in vivo. *Journal of Biomedical Optics* 10(4), p. 041210. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16178634> [Accessed: 26 February 2020].

9 Appendix

9.1 Expanded Figures from Chapter 3



Figure 3.6EX.



Figure 3.7EXa.

Appendix

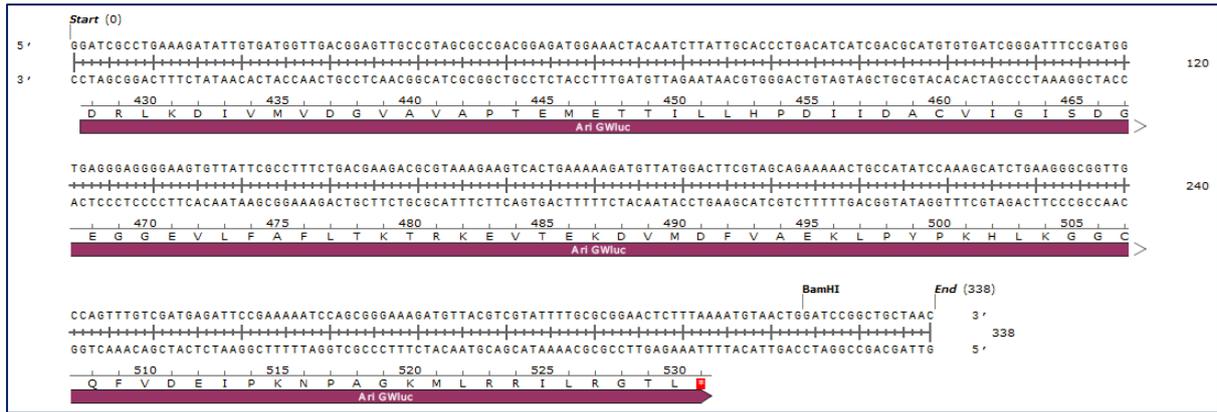


Figure 3.7EXb.

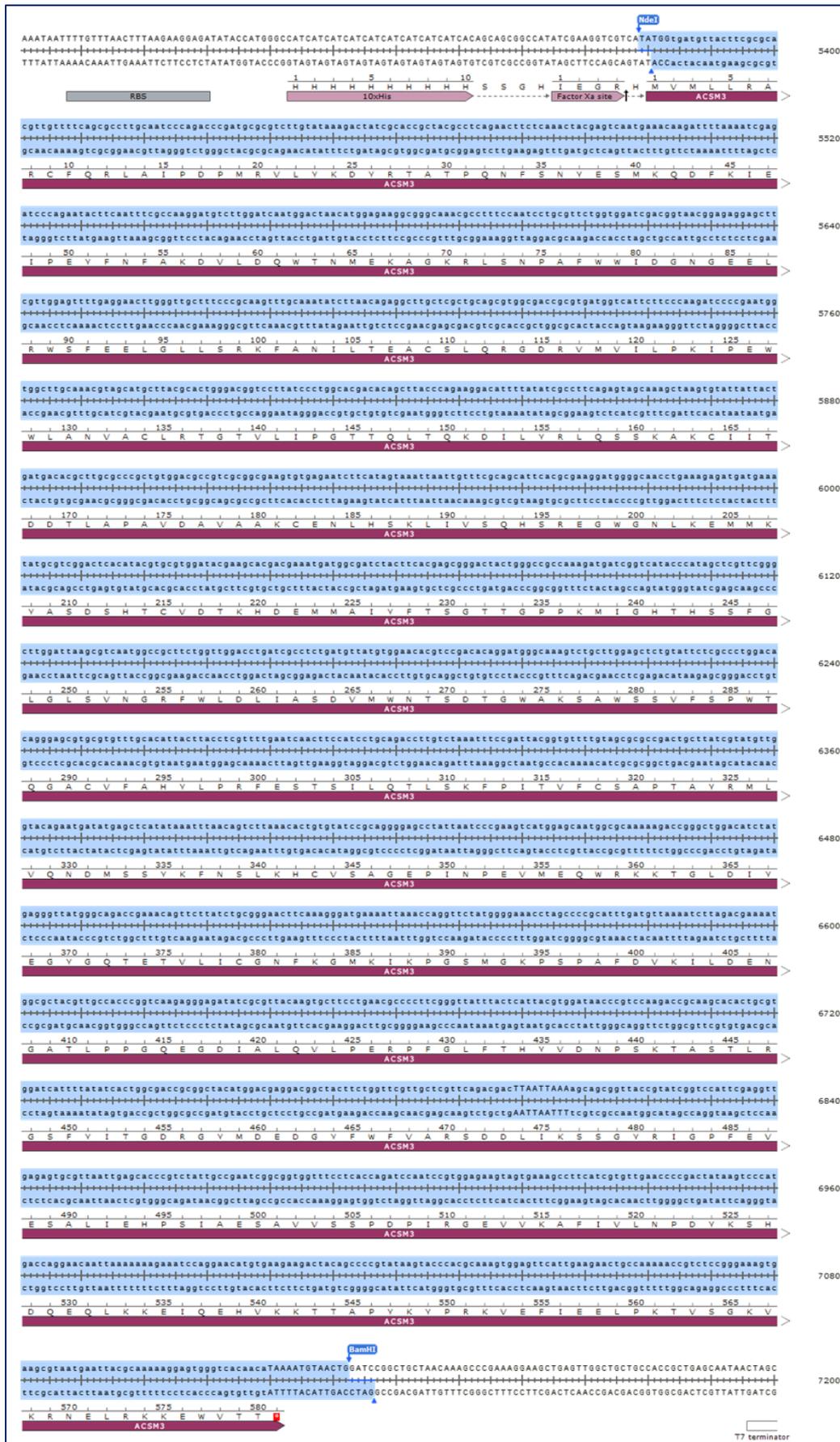


Figure 3.18EX.



Figure 3.26EXa.

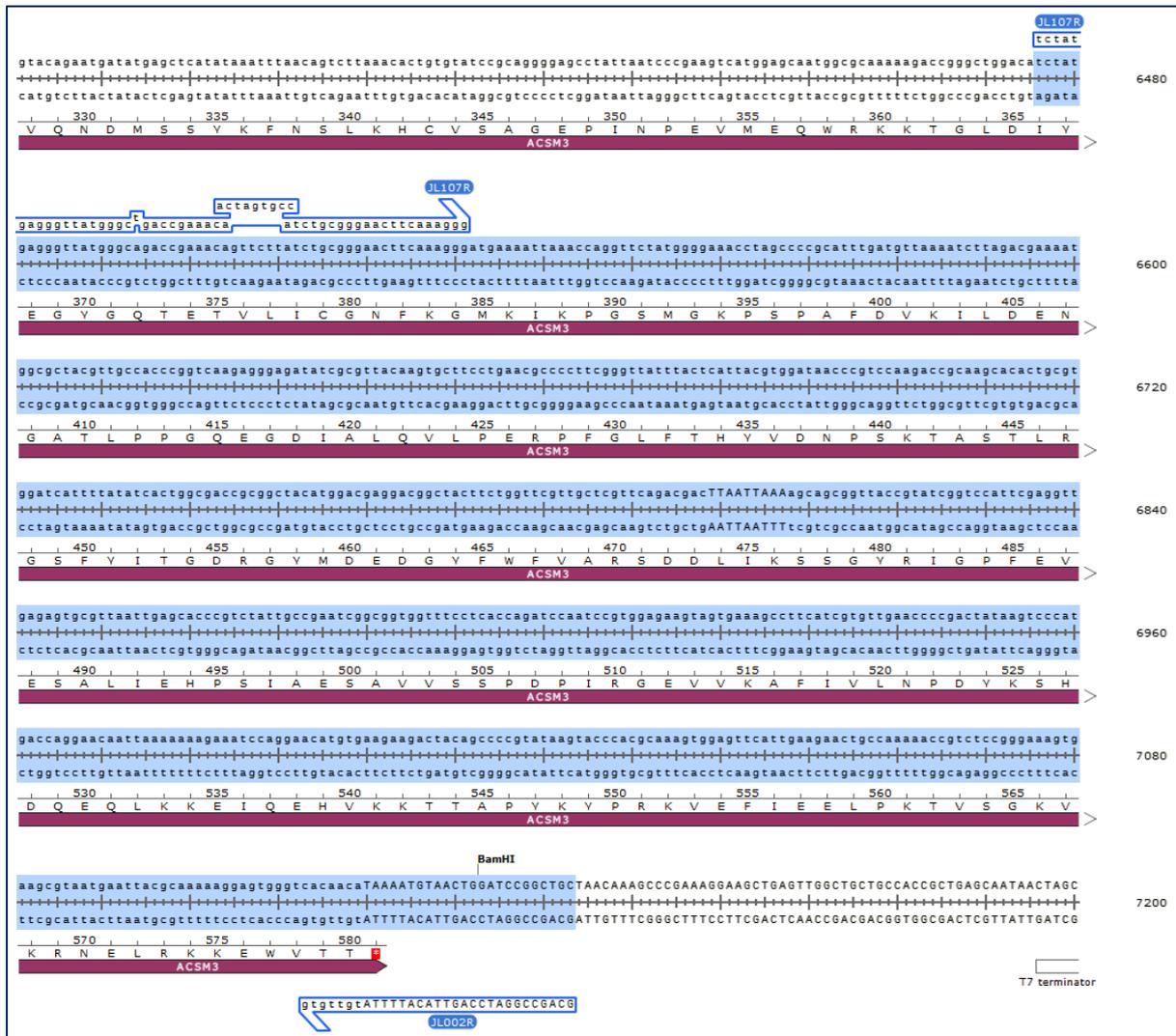


Figure 3.26EXb.

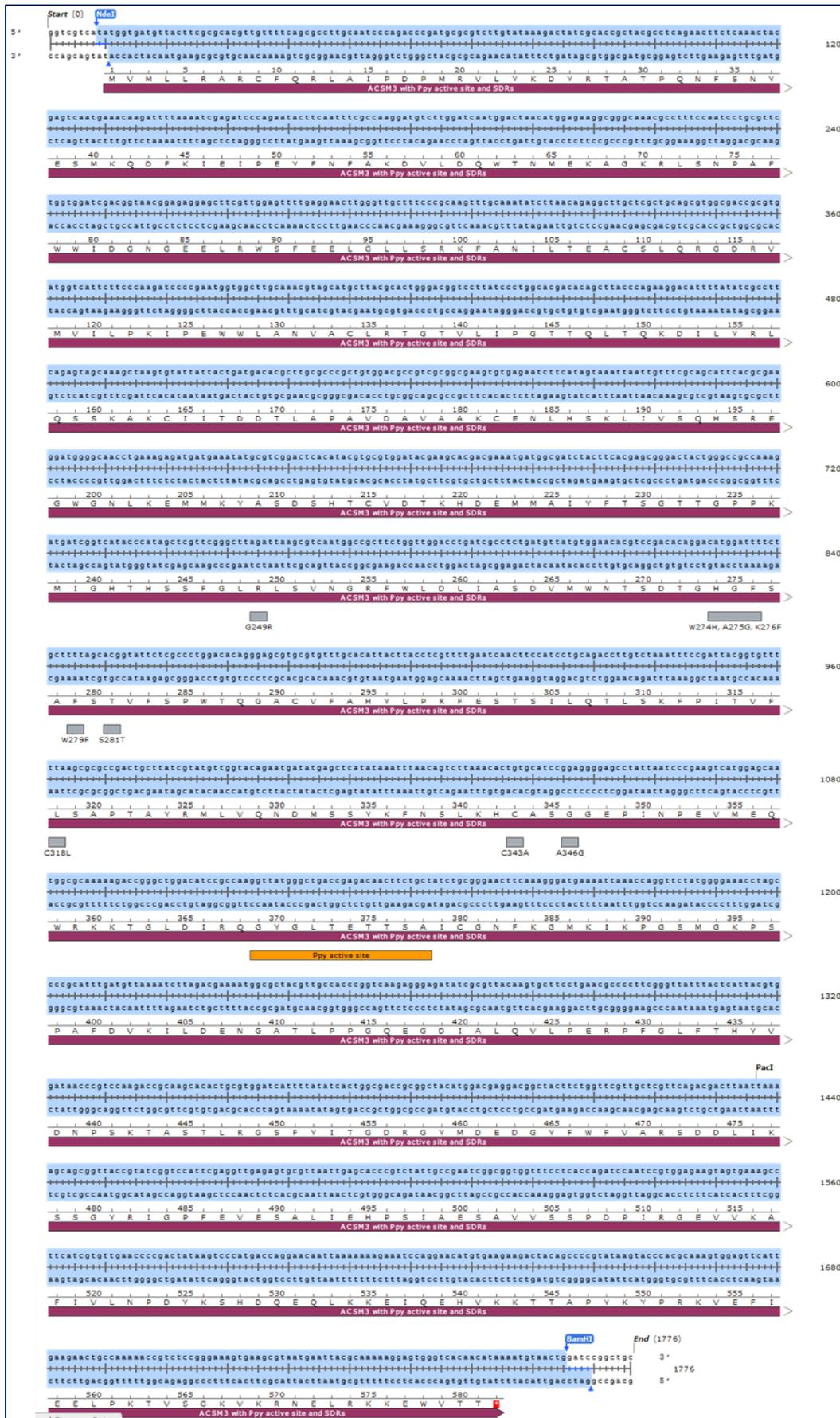


Figure 3.30EX.

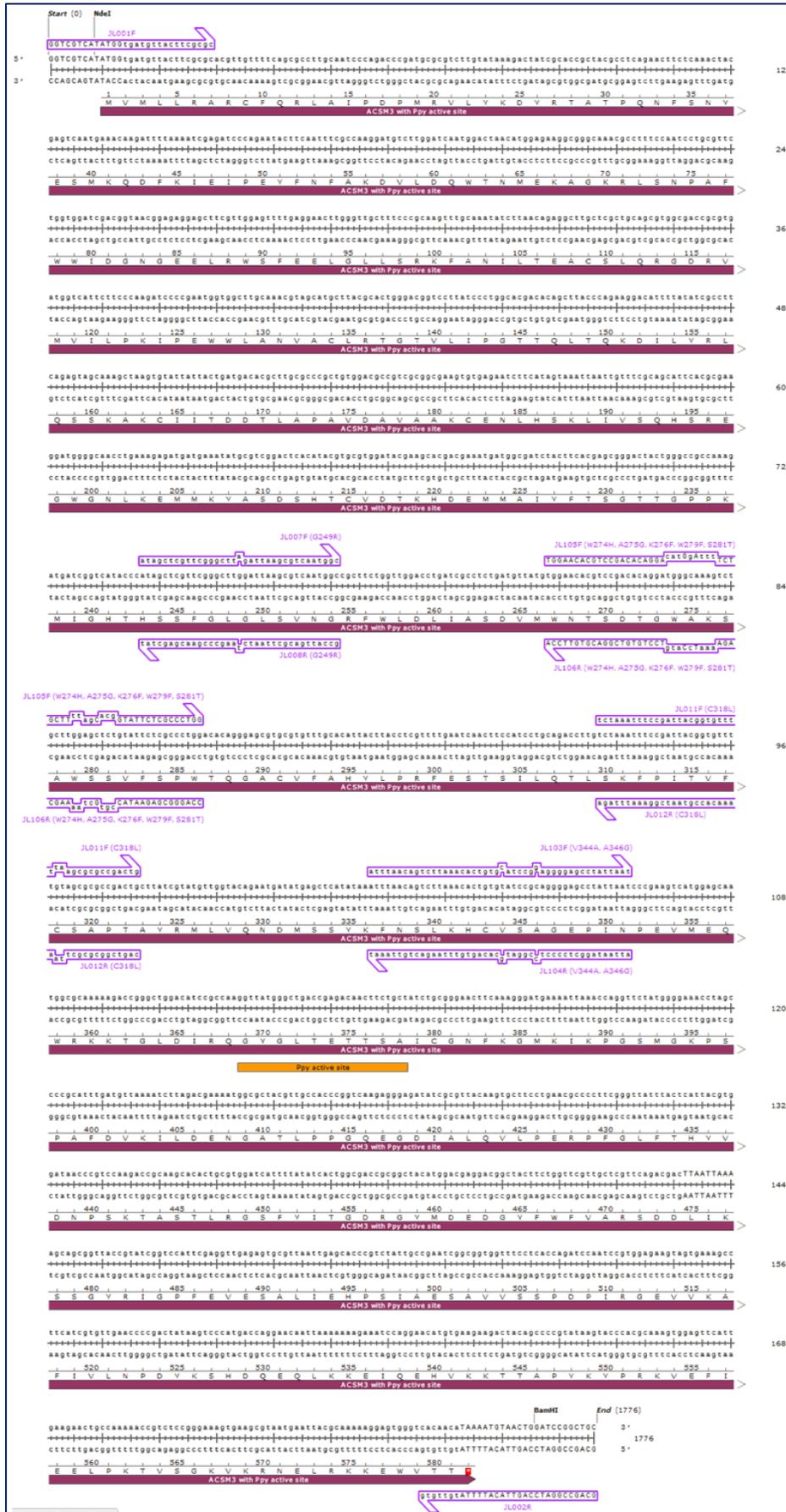


Figure 3.31EX.



Figure 3.36EX.

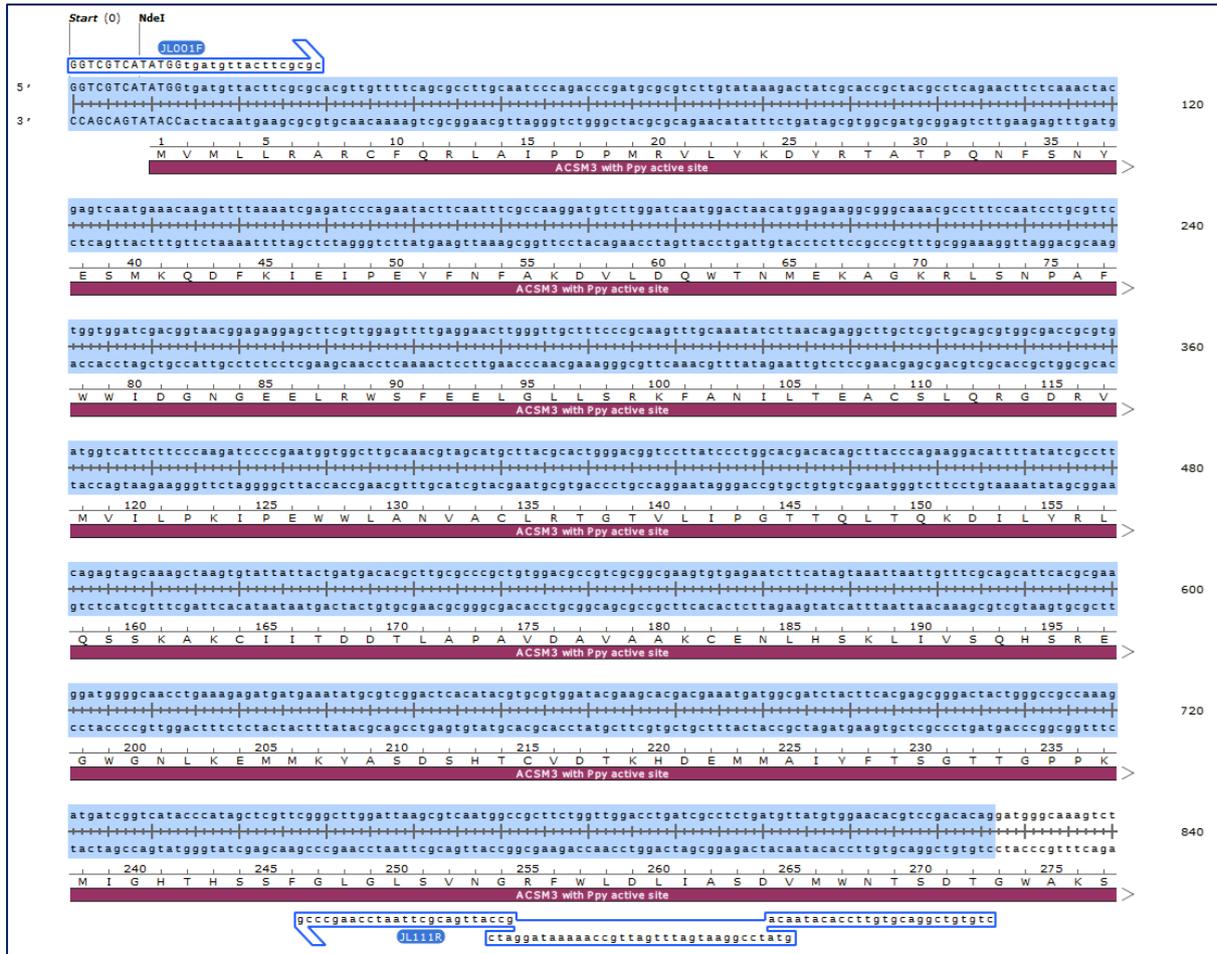


Figure 3.37EXa.

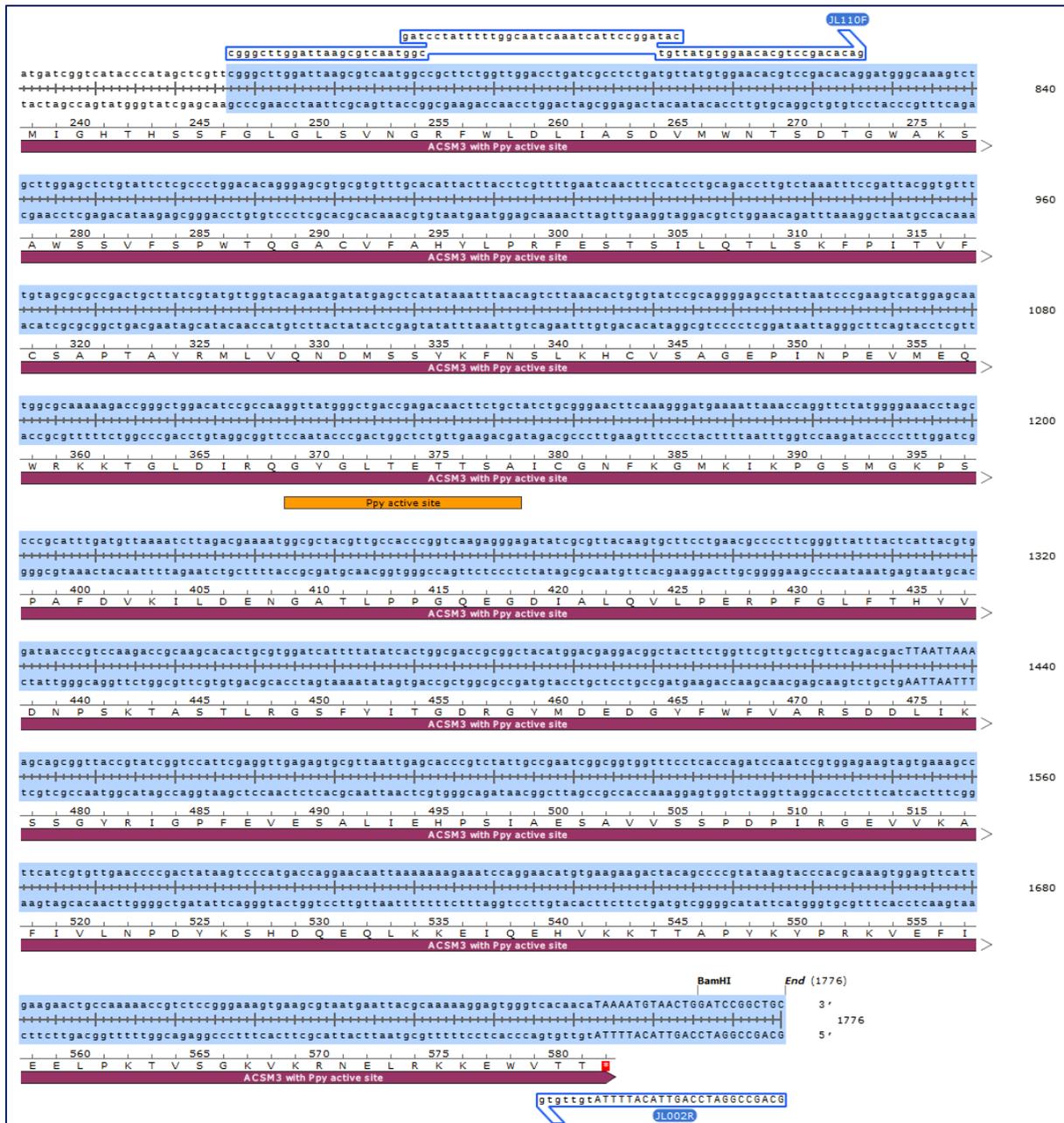


Figure 3.37EXb.

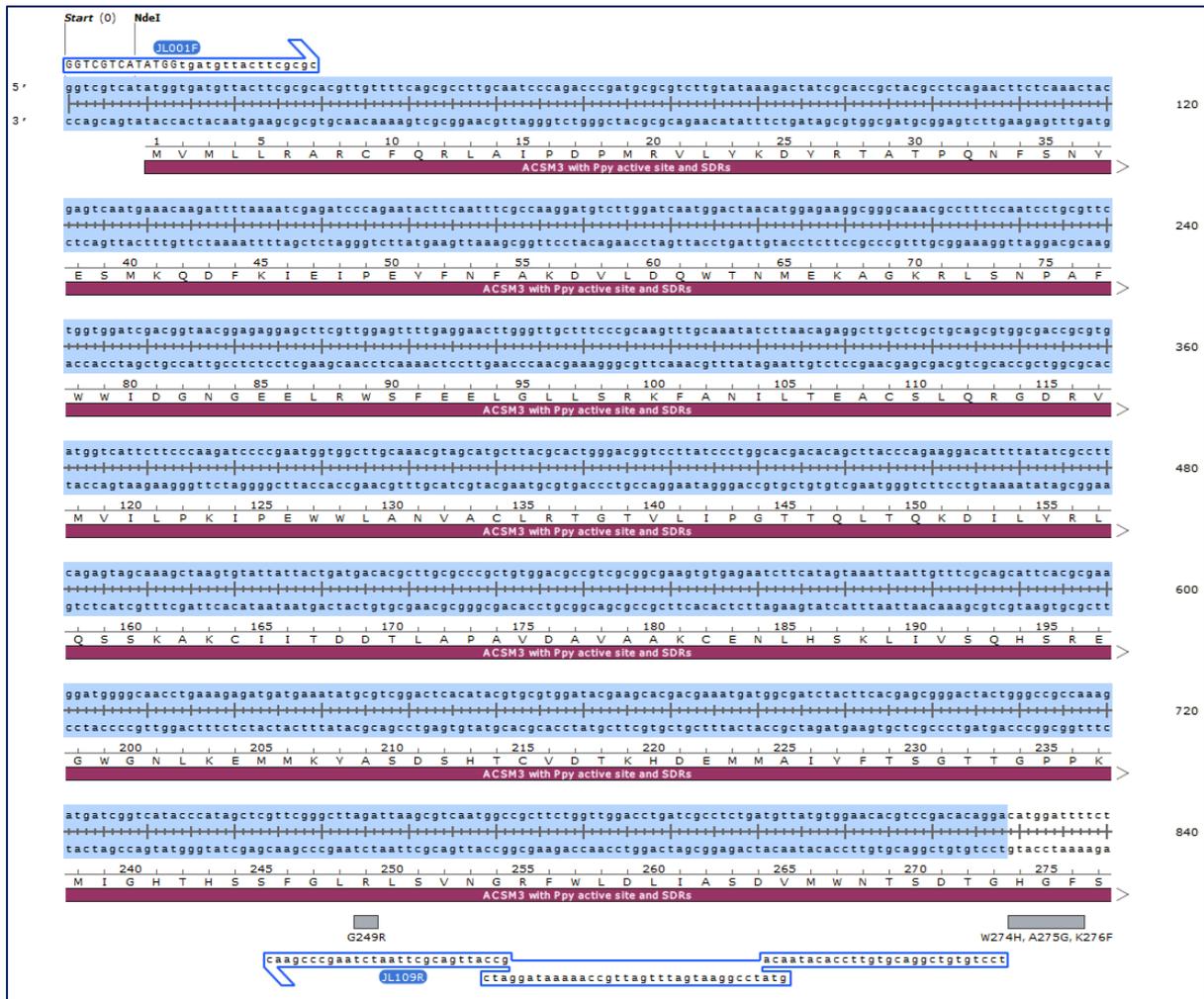


Figure 3.39EXa.

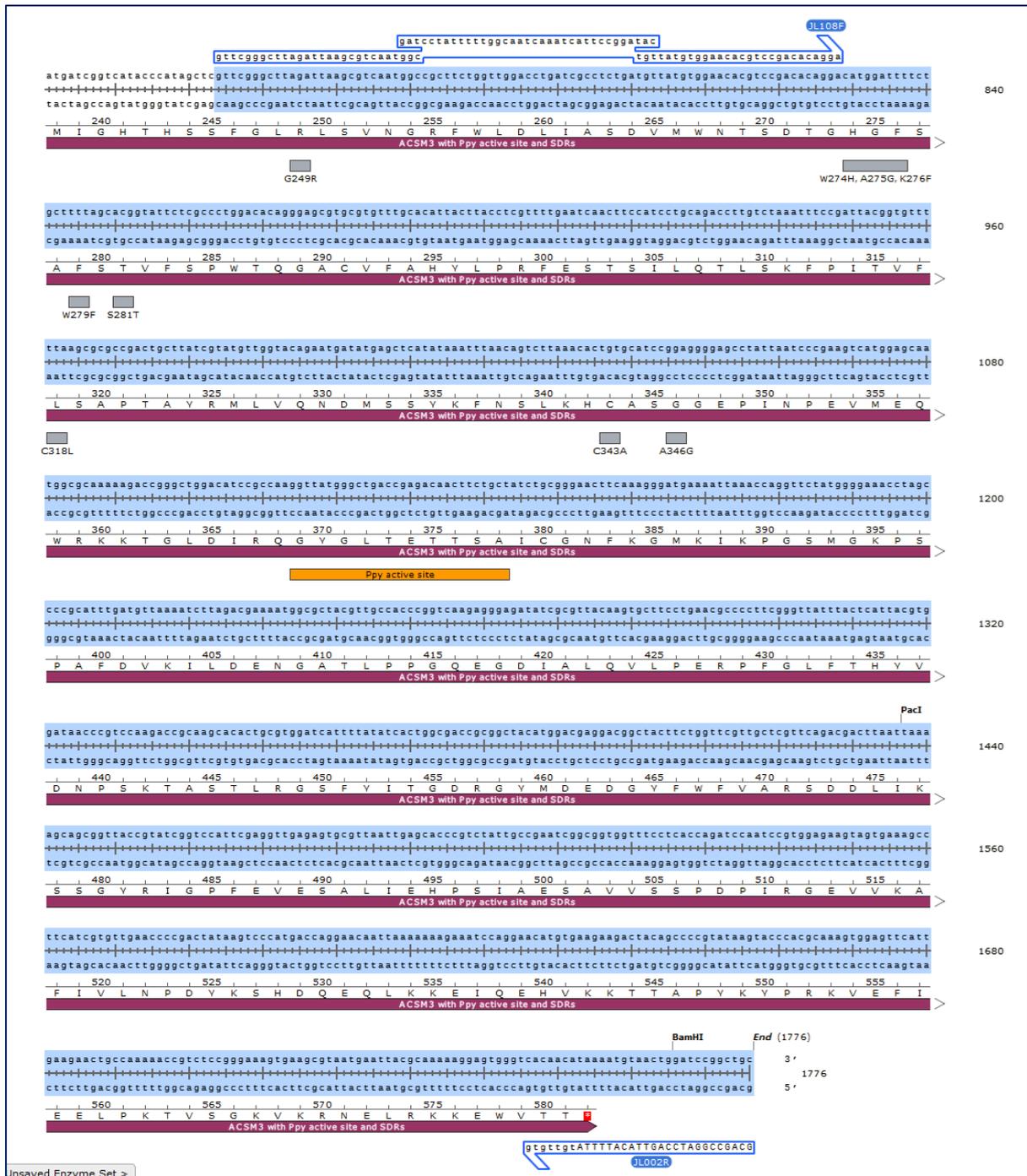


Figure 3.39EXb.

9.2 Expanded Figures from Chapter 4



Figure 4.3EX.

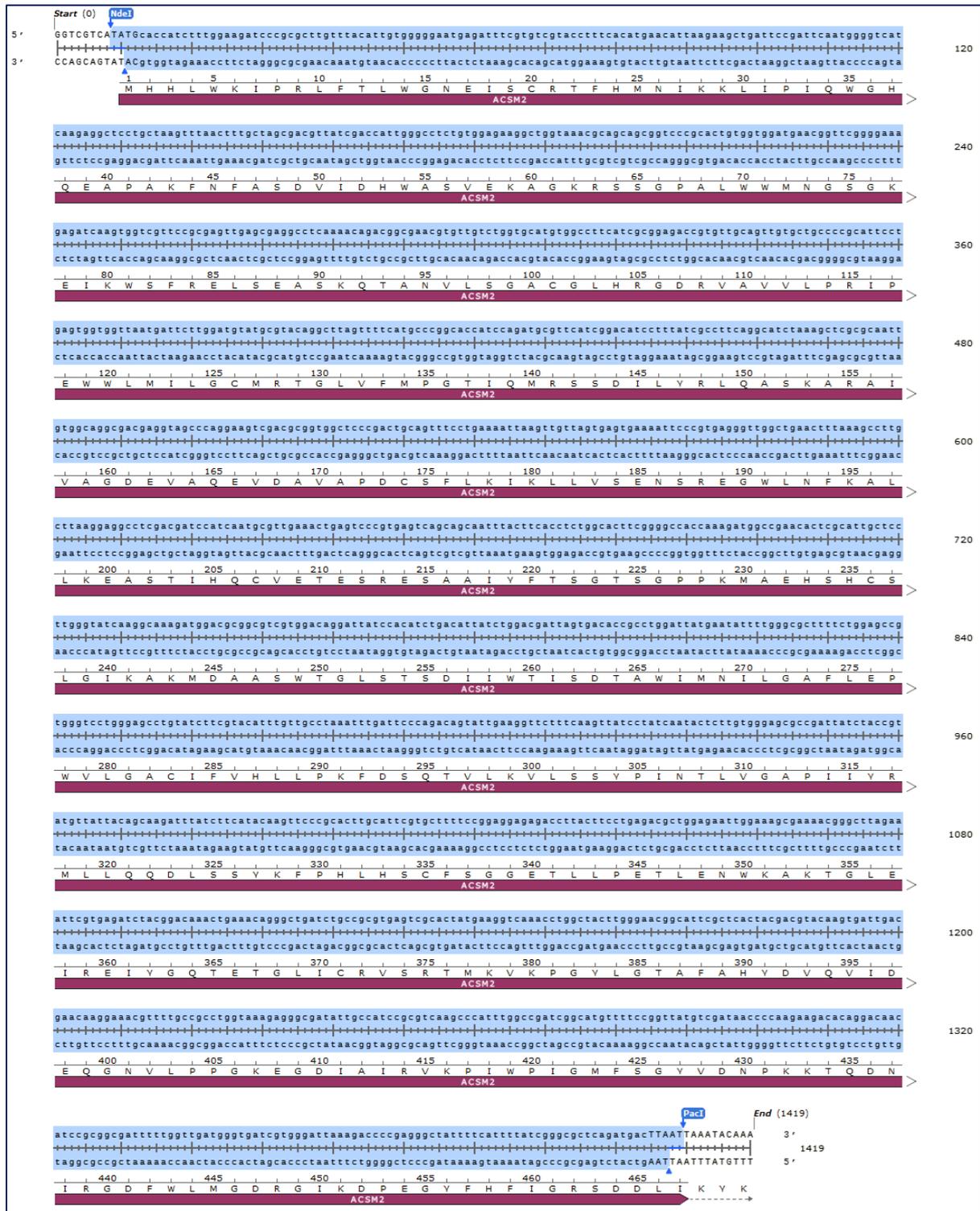


Figure 4.4EXa.

Appendix

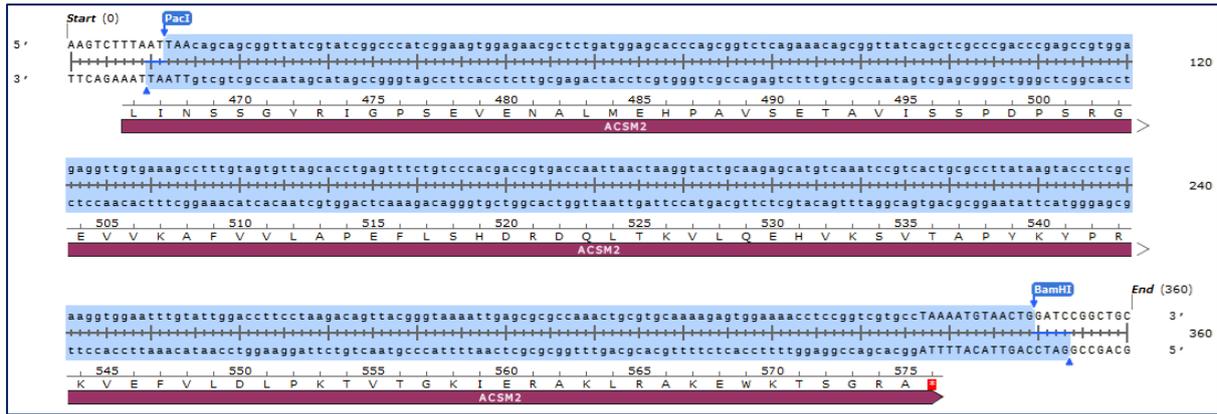


Figure 4.4EXb.



Figure 4.5EXa.

Appendix

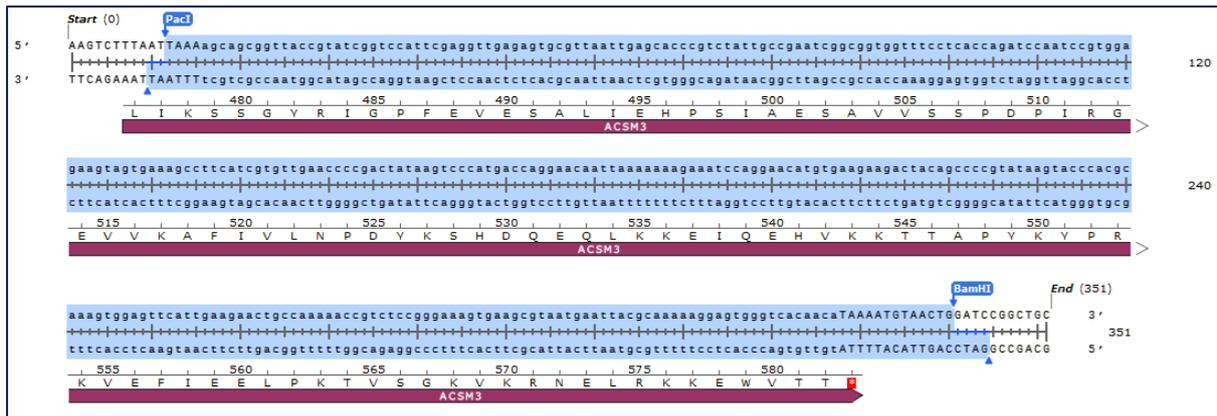


Figure 4.5EXb.

Appendix



Figure 4.19EXa.



Figure 4.19Exb.



Figure 4.20EXa.



Figure 4.20Xb.

9.3 Expanded Figures from Chapter 6

Appendix

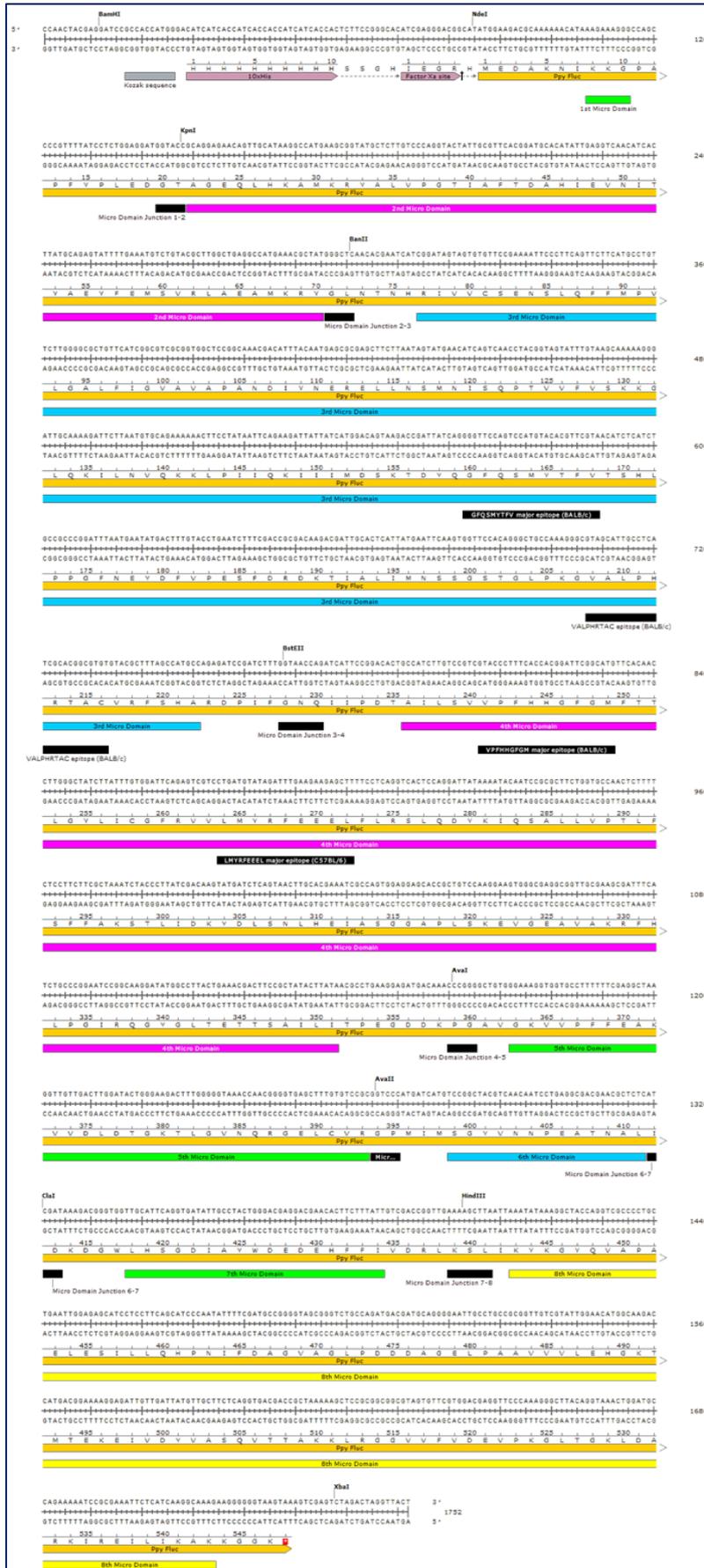


Figure 6.2EX.

Appendix

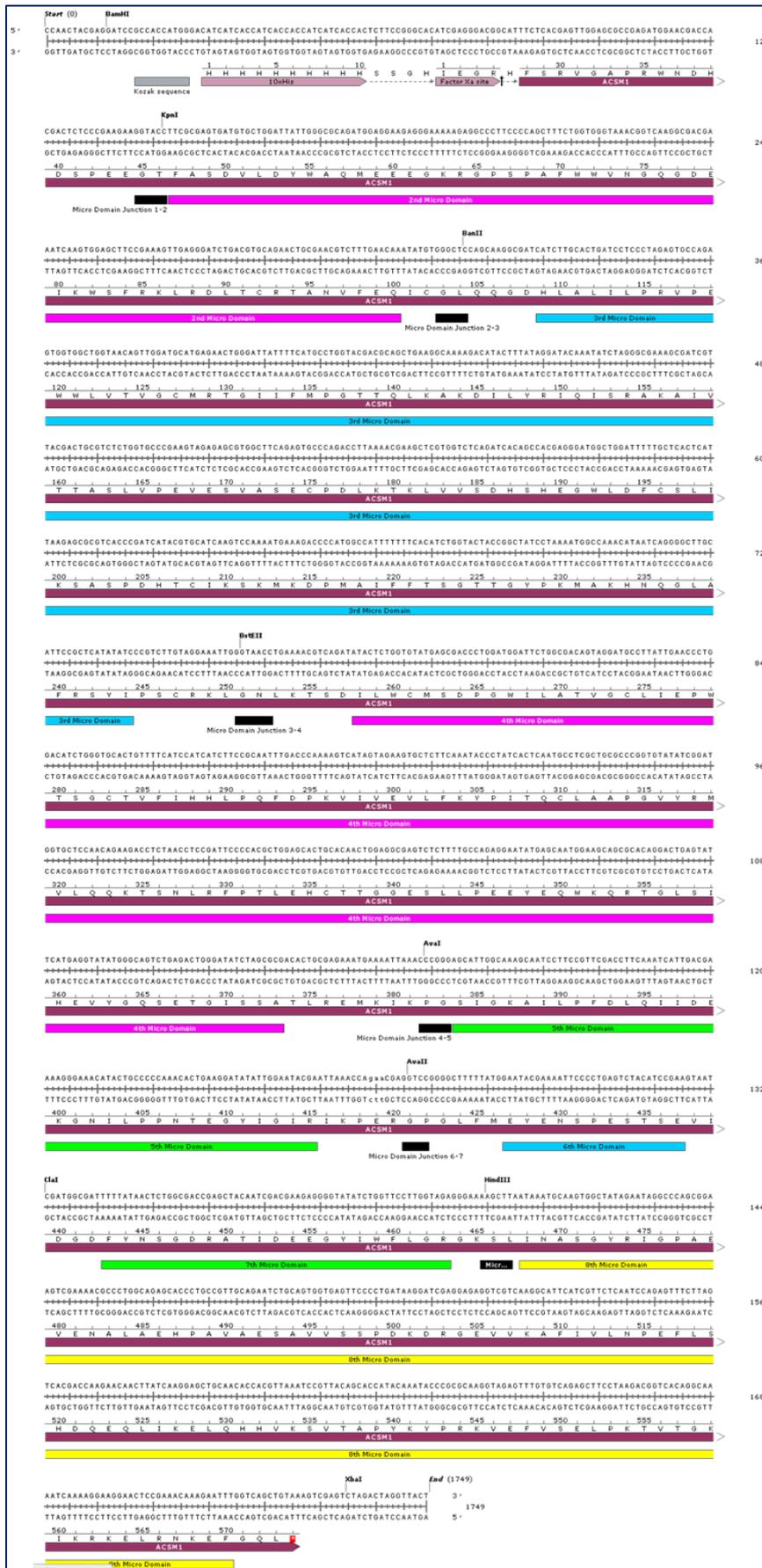


Figure 6.3EX.

Appendix

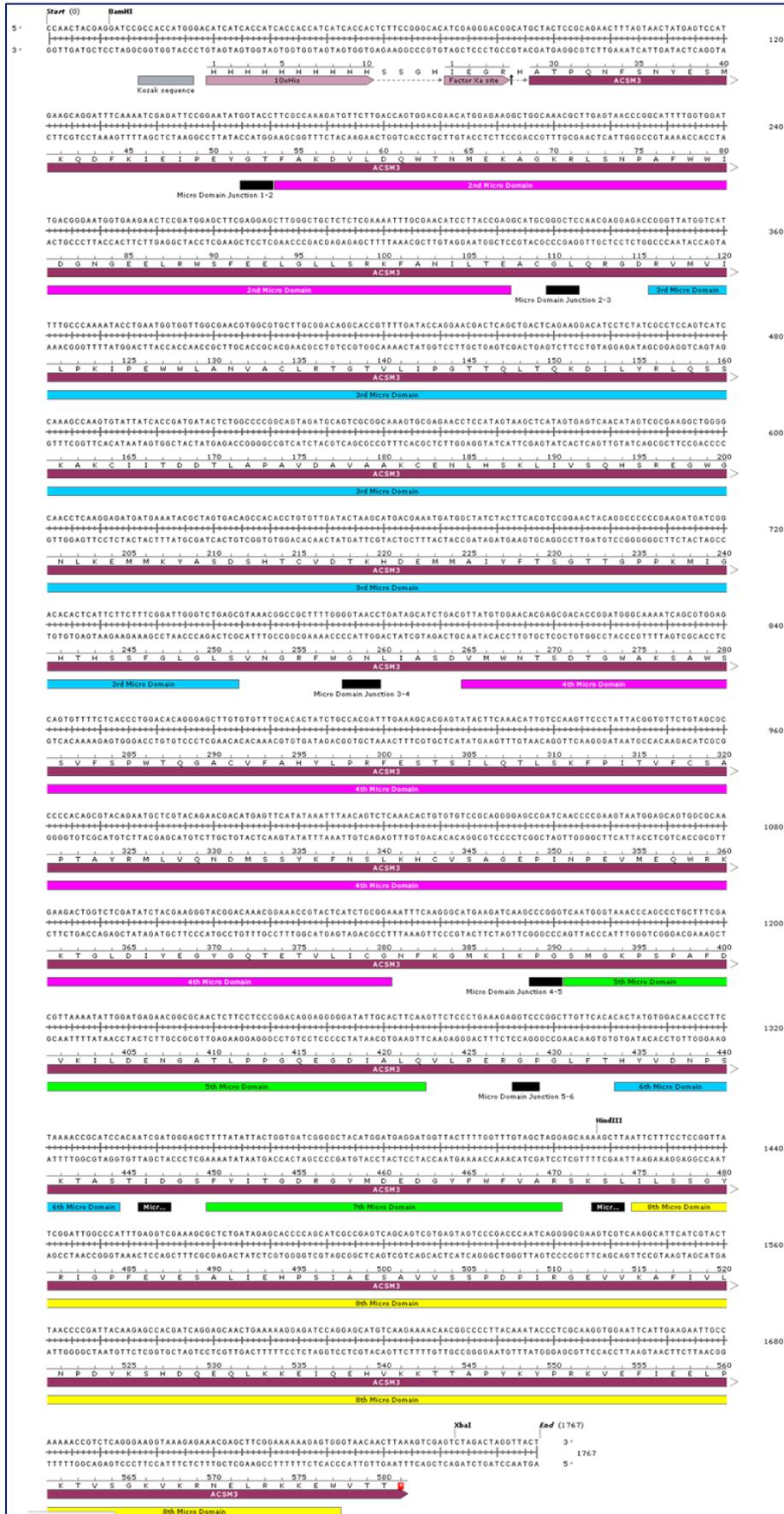


Figure 6i.

Appendix

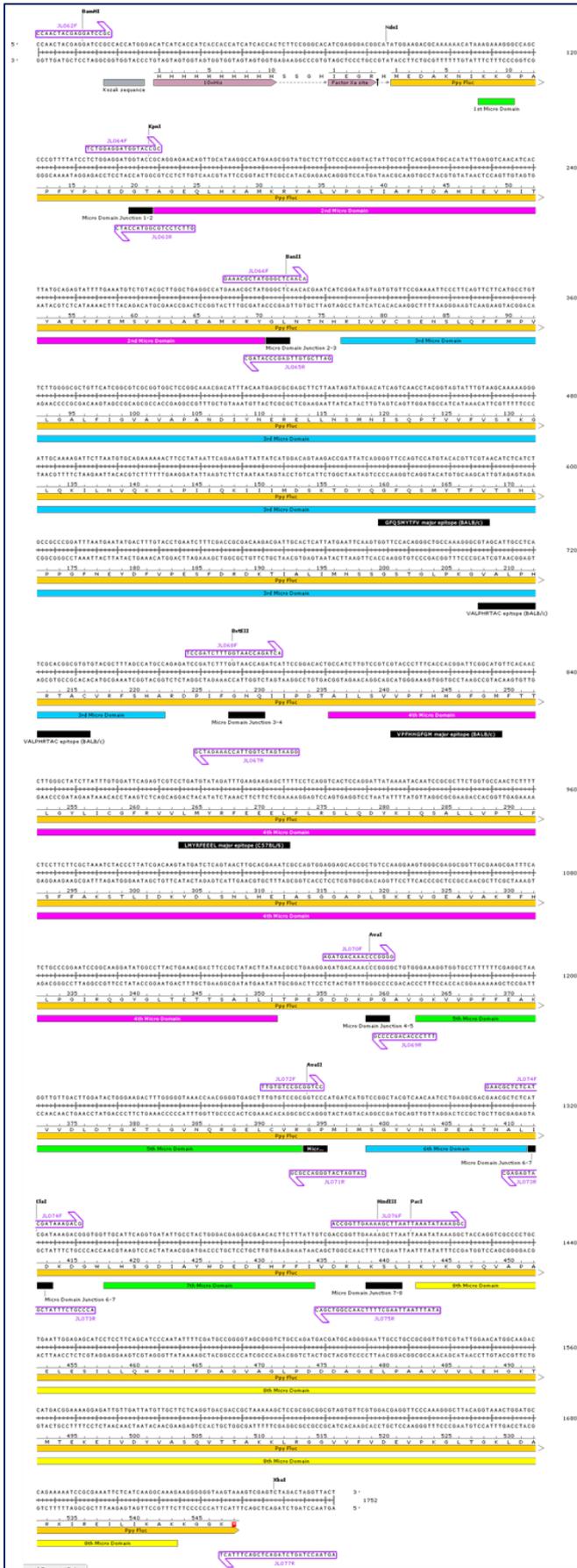


Figure 6.4EX.

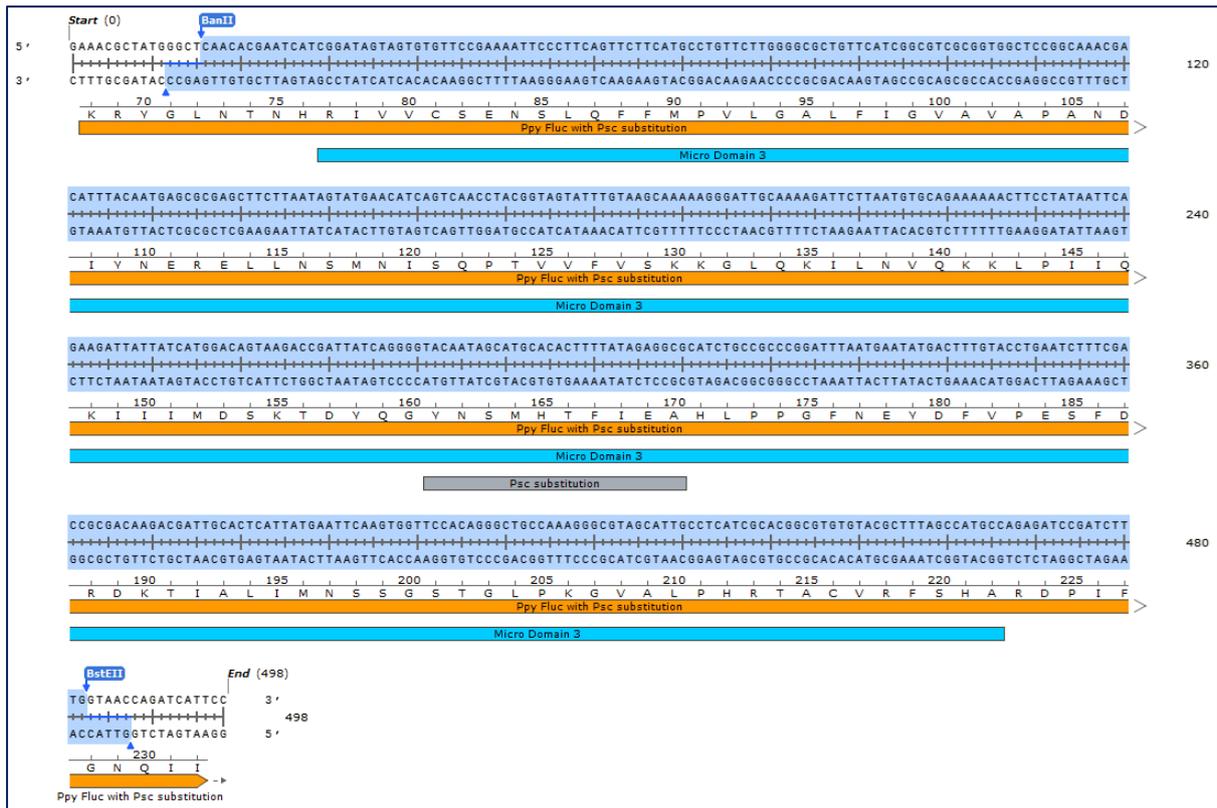


Figure 6.12EX.

9.4 Generic Figures from Various Chapters

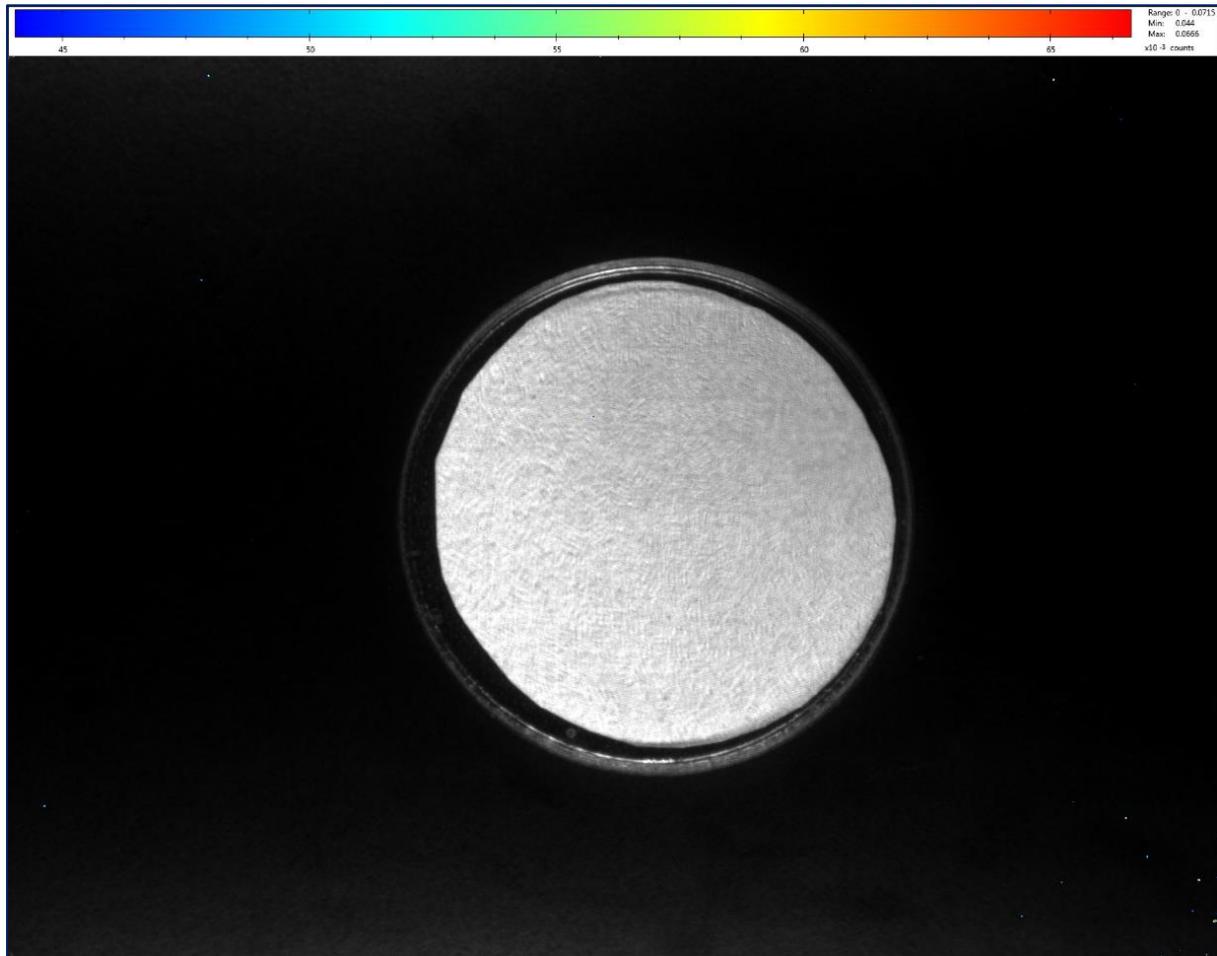


Figure 9.1. An example of a negative screen via the Biospace Lab PhotonIMAGER (bright-field photograph with photon acquisition superimposed). This particular example is of a nylon membrane lift of IPTG-induced colonies screened with $D^{-}LH_2$. The PhotonIMAGER automatically acquires at maximum aperture (highest sensitivity). There is some background noise, notably in the lower right of the image which is typical. The colonies are only visible on the membrane as a general haze in the bright-field image. The principle of determining a negative result from the PhotonIMAGER is the same for different formats (for instance using well-plates rather than petri dishes), and for fluorescence protein expression screening rather than bioluminescence.

Appendix

[This page is intentionally blank]