

**SYNTHETIC ALKALOID ENZYMOLOGY:
NEXT-GENERATION PRODUCTION OF ACTIVE
PHARMACEUTICAL INGREDIENTS AND
NUTRICHEMICALS**

WONG WEI JIE GARRETT

(B.Sc. (Hons.), NUS)

**A THESIS SUBMITTED FOR THE JOINT DEGREE OF
DOCTOR OF PHILOSOPHY**

**DEPARTMENT OF BIOCHEMISTRY
NATIONAL UNIVERSITY OF SINGAPORE**

AND

**DEPARTMENT OF MEDICINE
IMPERIAL COLLEGE LONDON**

2019

Supervisors:

Associate Professor Yew Wen Shan, Main Supervisor
Professor Paul Freemont, Co-Supervisor

Examiners:

Associate Professor Poh Chueh Loo, National University of Singapore
Dr. Karen Polizzi, Imperial College London
Associate Professor Liang Zhao-Xun, Nanyang Technological University

DECLARATION OF ORIGINALITY

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I duly acknowledge all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.



Wong Wei Jie Garrett
02 August 2019

COPYRIGHT DECLARATION

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a [Creative Commons Attribution-NonCommercial-No Derivatives 4.0 International Licence](#) (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text.

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

ACKNOWLEDGEMENTS

First and foremost, I am extremely grateful to my supervisors Prof. Paul Freemont and A.Prof. Yew Wen Shan for this excellent and rewarding opportunity, as well as their guidance and support over the past four years. I am also thankful for the generous financial support provided by the National University of Singapore Research Scholarship. Special thanks are also due to the members of my TAC, A.Prof Matthew Chang and A.Prof. Susanna Leong for the time taken out of their busy schedules to provide valuable suggestions that helped guide and improve my work.

I would also like to thank the people who have shared materials and expertise used in this project. Dr. David Bell and Dr. Chee Soo Mei, for their kind help and expertise in developing and improving the analytical methods used during the course of this project. Prof. Patrick Yizhi Cai and his group from the University of Manchester for kindly sharing the YeastFab suite of plasmids that formed an integral backbone to this project. Fellow lab members, Mr. Lim Li Rong for the *cloA* plasmids, and Mr. Tan Yong Quan for his help with creating the pathway acceptor and genome integration vectors.

To my research mentor Dr. Maybelle Go, whom I am wholly indebted to for her patience, guidance and friendship from the very beginning, I extend my greatest thanks. I would also like to extend my gratitude to the past and present members of the Freemont and Yew labs, SynCTI, and SynbiCITE who have all been excellent team-players and provided the brilliant discussion that helped shape this work.

Lastly, I would like to thank my family and friends for their unwavering support and understanding. The slowbros for all the timely and well-organised breaks. Most of all, Ms. Lydia Teo for always being my sunshine, pillar, and cheerleader.

Table of Contents

ACKNOWLEDGEMENTS	I
SUMMARY	VI
LIST OF PUBLICATIONS AND CONFERENCE PRESENTATIONS.....	VII
LIST OF TABLES.....	VIII
LIST OF FIGURES.....	IX
LIST OF ABBREVIATIONS	XIII
CHAPTER 1: INTRODUCTION	1
1.1 MICROBIAL CELL FACTORIES FOR THE PRODUCTION OF PRIVILEGED SCAFFOLDS.....	1
1.2 THE ERGOT ALKALOIDS	2
1.2.1 <i>A TUMULTUOUS JOURNEY.....</i>	2
1.2.2 <i>BIOSYNTHESIS AND GENETICS OF THE ERGOT ALKALOIDS</i>	4
1.2.3 <i>COMMERCIAL RELEVANCE AND CURRENT PRODUCTION METHODS.....</i>	12
1.3 A HETEROLOGOUS ERGOT PRODUCER: THE CURRENT STATE	14
1.4 SKETCHING THE BLUEPRINT FOR A MICROBIAL ASSEMBLY LINE..	15
1.4.1 <i>BREWING THE NEXT WAVE OF ERGOT ALKALOID PRODUCTION </i>	16
1.4.2 <i>YEASTFAB ASSEMBLY: A GENOME COMPILER.....</i>	17
1.4.3 <i>DESIGN → BUILD → TEST → LEARN</i>	20
CHAPTER 2: LAYING THE FOUNDATIONS FOR A MICROBIAL CELL FACTORY	22
2.1 INTRODUCTION	22
2.2 MATERIALS AND METHODS	25
2.2.1 <i>ADAPTATIONS TO THE YEASTFAB SYSTEM.....</i>	25
2.2.2 <i>IDENTIFICATION AND CLONING OF BIOLOGICAL PARTS.....</i>	26
2.2.2.1 <i>PROMOTERS AND TERMINATORS.....</i>	26
2.2.2.2 <i>IDENTIFICATION, CLONING, AND PURIFICATION OF dmaW AND easF</i>	29
2.2.3 <i>GOLDEN GATE ASSEMBLY</i>	30
2.2.4 <i>CHARACTERIZATION OF PROMOTER STRENGTHS</i>	31
2.2.4.1 <i>FLUORESCENT PROTEIN EXPRESSION</i>	31
2.2.4.2 <i>PRODUCTION OF DMAT.....</i>	31
2.2.5 <i>IN VITRO BIOSYNTHESIS OF DMAT AND 4DMA</i>	32
2.2.6 <i>GENOMIC INTEGRATION AND IN VIVO PRODUCTION OF 4DMA32</i>	
2.2.7 <i>LC-MS ANALYSIS OF SAMPLES.....</i>	33
2.3 RESULTS AND DISCUSSION	34
2.3.1 <i>ADAPTING YEASTFAB ASSEMBLY.....</i>	34
2.3.2 <i>CHARACTERIZATION OF PROMOTER STRENGTHS.....</i>	37
2.3.2.1 <i>FLUORESCENT PROTEIN EXPRESSION</i>	37

2.3.2.2	<i>PRODUCTION OF DMAT.....</i>	39
2.3.3	<i>TESTING GENOME INTEGRATION SITES WITH 4DMA PRODUCTION.....</i>	45
CHAPTER 3: BUILDING THE EARLY ERGOT PATHWAY TO CHANOCLAVINE-I		49
3.1	INTRODUCTION	50
3.2	MATERIALS AND METHODS	53
3.2.1	<i>CREATING AN ENTRY-LEVEL VECTOR FOR APPENDING SUMO- TAGS.....</i>	53
3.2.2	<i>IDENTIFICATION AND SCREENING OF easE ORTHOLOGS.....</i>	54
3.2.2.1	<i>GENERATING AN SSN FROM THE EFI-EST WEBTOOL.....</i>	54
3.2.2.2	<i>BUILDING THE SCREENING STRAIN YMC17.....</i>	56
3.2.2.3	<i>SCREENING SELECTED easE ORTHOLOGS FOR ACTIVITY</i>	59
3.2.2.4	<i>SEQUENCE ALIGNMENTS</i>	59
3.2.3	<i>DEVELOPMENT OF A CONSTITUTIVE CHANOCLAVINE-I PRODUCER.....</i>	60
3.2.3.1	<i>ASSEMBLY AND EVALUATION OF THE INITIAL STRAIN</i>	60
3.2.3.2	<i>IDENTIFICATION OF THE SHUNT PRODUCTS.....</i>	60
3.2.3.3	<i>PREPARATION OF BUFFERED YEAST CULTURES.....</i>	60
3.2.4	<i>DESIGN AND ASSEMBLY OF THE CHANOCLAVINE-I PROTOTYPE</i>	61
3.2.5	<i>VALIDATION OF CHANOCLAVINE-I PRODUCED.....</i>	61
3.2.5.1	<i>EXPRESSION AND PURIFICATION OF EASE_AJ.....</i>	61
3.2.5.2	<i>IN VITRO BIOSYNTHESIS OF CHANOCLAVINE-I.....</i>	62
3.2.6	<i>LC-MS ANALYSIS OF SAMPLES.....</i>	63
3.3	RESULTS AND DISCUSSION	64
3.3.1	<i>EXPANDING THE ENTRY-LEVEL VECTOR RANGE FOR FUSION TAGS.....</i>	64
3.3.2	<i>PROSPECTING FOR FUNCTIONAL PREDICTIONS.....</i>	65
3.3.2.1	<i>MINING WITH SEQUENCE SIMILARITY NETWORKS.....</i>	65
3.3.2.2	<i>A SCREENING METHOD FOR CHANOCLAVINE-I PRODUCTION.....</i>	67
3.3.3	<i>DEVELOPMENT OF A CONSTITUTIVE CHANOCLAVINE-I PRODUCER.....</i>	72
3.3.3.1	<i>AN INITIAL ASSESSMENT.....</i>	72
3.3.3.2	<i>DISCOVERY AND IDENTIFICATION OF SHUNT PRODUCTS</i>	73
3.3.3.3	<i>PREVENTING THE FORMATION OF SHUNT PRODUCTS</i>	76
3.3.4	<i>VALIDATION OF THE CHANOCLAVINE-I PRODUCED</i>	78
CHAPTER 4: DEVELOPING AND OPTIMIZING AN AGROCLAVINE PRODUCING STRAIN		82
4.1	INTRODUCTION	83
4.2	MATERIALS AND METHODS	87
4.2.1	<i>IDENTIFICATION AND CLONING OF easD AND easG.....</i>	87

4.2.2	BIOINFORMATICS FOR THE IDENTIFICATION OF ISOMERASE VARIANTS	87
4.2.3	SCREENING OF <i>easA</i> ORTHOLOGS	89
4.2.3.1	ASSEMBLY AND CONSTRUCTION OF PLASMID CONSTRUCTS....	89
4.2.3.2	PREPARATION OF AGROCLAVINE STANDARDS AND QUANTIFICATION.....	90
4.2.4	IMPROVING THE PRODUCTION OF AGROCLAVINE	90
4.2.4.1	ASSEMBLY AND CONSTRUCTION OF PLASMID CONSTRUCTS....	90
4.2.4.2	TESTING STRAIN STABILITY ACROSS MULTIPLE PROPAGATIONS	91
4.2.4.3	CONSTRUCTION AND TESTING OF AGCM33B AND AGCM32B....	92
4.2.4.4	LC-MS QUANTIFICATION OF AGROCLAVINE PRODUCED	94
4.2.5	VALIDATION OF AGROCLAVINE PRODUCTION USING ¹³ C-LABELLED TRYPTOPHAN.....	94
4.2.6	LC-MS ANALYSIS OF SAMPLES.....	94
4.3	RESULTS AND DISCUSSION	96
4.3.1	ISOLATING ISOMERASES FROM REDUCTASES	96
4.3.2	SCREENING FOR ISOMERASES	98
4.3.3	IMPROVING THE PRODUCTION OF AGROCLAVINE	103
4.3.3.1	WIDENING THE PATHWAY'S BOTTLENECKS.....	103
4.3.3.2	INSTABILITY IN THE INITIAL DESIGNS	105
4.3.4	VALIDATION OF AGROCLAVINE PRODUCTION WITH ¹³ C-LABELLED TRYPTOPHAN.....	109
CHAPTER 5: COMPLETING THE PATHWAY TO LYSERGIC ACID		112
5.1	INTRODUCTION	113
5.2	MATERIALS AND METHODS	118
5.2.1	GENERATING AN SSN OF THE <i>cloA</i> ORTHOLOGS	118
5.2.2	PROCUREMENT OF <i>cloA</i> ORTHOLOGS	118
5.2.2.1	TRANSFORMATION, CULTURE AND GROWTH CONDITIONS....	119
5.2.2.2	PREPARATION OF STANDARDS.....	119
5.2.2.3	QUANTIFICATION WITH A STANDARD CURVE.....	119
5.2.2.4	QUANTIFICATION BY STANDARD ADDITION	120
5.2.3	DEVELOPMENT OF A GENOMIC DLA PRODUCER.....	120
5.2.3.1	TRANSFORMATION, CULTURE AND GROWTH CONDITIONS....	120
5.2.3.2	QUANTIFICATION BY STANDARD ADDITION	121
5.2.4	CHARACTERIZATION OF DLAM33B.....	121
5.2.4.1	TRYPTOPHAN SUPPLEMENTATION EXPERIMENTS	121
5.2.4.2	¹³ C-TRYPTOPHAN FEEDING EXPERIMENTS	122
5.2.4.3	QUANTIFICATION OF ¹³ C-DLA WITH A STANDARD CURVE.....	122
5.2.4.4	CALCULATION OF PREDICTED LOGP VALUES	122
5.2.5	LC-MS ANALYSIS OF SAMPLES.....	122

5.2.6	STATISTICAL ANALYSIS.....	123
5.3	RESULTS AND DISCUSSION	124
5.3.1	BIOINFORMATICS GUIDED PREDICTION OF <i>cloA</i> ACTIVITY	124
5.3.2	DEMONSTRATING THE COMPLETE BIOSYNTHESIS OF D- LYSERGIC ACID.....	126
5.3.2.1	PLASMID-BASED TRIALS	126
5.3.2.2	QUANTIFYING PRODUCTION FROM PLASMID BASED STRAINS.....	129
5.3.2.3	QUANTIFICATION BY STANDARD ADDITION	132
5.3.2.4	TOWARDS A STABLE GENOMIC STRAIN	134
5.3.3	EVALUATING THE ASSEMBLED PATHWAY	137
5.3.3.1	EFFECT OF SUPPLEMENTING ADDITIONAL TRYPTOPHAN.....	137
5.3.3.2	STRAIN CHARACTERIZATION WITH LABELLED SUBSTRATES...	143
5.3.3.2.1	ASSESSING PATHWAY EFFICIENCY.....	143
5.3.3.2.2	VALIDATION OF PATHWAY TURNOVER AND FLUX TRACKING	144
	CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS	153
6.1	CONCLUSIONS	153
6.2	FUTURE DIRECTIONS	155
6.2.1	OPTIMIZATION, METABOLIC ENGINEERING, AND BIOPROCESS CONTROL.....	155
6.2.2	DIVERSIFICATION OF OXIDIZED TETRACYCLIC ERGOLINE PRODUCTS.....	157
6.2.3	BIOSYNTHESIS OF HALOGENATED ERGOT ALKALOIDS	159
6.3	LEARNING THROUGH CREATING.....	160
	APPENDIX	172

SUMMARY

One of the grand themes in the field of synthetic biology is the promise of the development of the tools that will help facilitate the global transition towards a renewable and sustainable bio-based economy. One such step towards that sustainable future is the use of microbial cell factories for the production of active pharmaceutical ingredients and nutraceuticals. In this line, the commercial production of the ergot alkaloids and their semi-synthetic derivatives is one such avenue that is appropriately poised to be improved on by being produced by microbial cell factories.

Ergot alkaloids have throughout history been known for their potent biological effects. These properties have been extensively exploited in the development of the semi-synthetic derivatives that are used in the treatments for Parkinsonism and dementia, amongst many others. Their production, however, is limited by the difficulty in the cultivation of the parasitic ergot fungi.

This thesis thus describes the progressive effort made to heterologously reconstitute the ergot alkaloid pathway to the main ergoline therapeutic precursor, *D*-lysergic acid, in *Saccharomyces cerevisiae*. Fundamental to this work is the adaptation of a Golden Gate-based cloning methodology, which would enable the easy and rapid assembly of customizable expression cassettes, and that is compatible with automation. Finally, this work presents for the first time, the development of a prototype yeast strain that before any extensive strain or bioprocess optimization, is capable of producing up to 71.5 $\mu\text{g/L}$ of *D*-lysergic acid. This demonstration effectively lays the groundwork for a more efficient commercial production method, as well as a platform for the discovery of novel ergot alkaloid therapeutics.

LIST OF PUBLICATIONS AND CONFERENCE PRESENTATIONS

Wong, G. W. J., Lim, L. R., Tan, Y. Q., Go, M. K., Bell, D. J., Freemont, P. S., Yew, W. S. (2020). Reconstituting the complete biosynthesis of the main ergoline-derived therapeutic precursor in *Saccharomyces cerevisiae*. (In preparation)

Wong, Garrett, Go, M. K., Freemont, P., Yew, W. S. (2019). Synthetic alkaloid enzymology: Next-generation production of active pharmaceutical ingredients and nutraceuticals. 26th Enzyme Mechanisms Conference. New Orleans, Louisiana, USA.

LIST OF TABLES

TABLE 2-1: LIST OF CONSTRUCTED GENOME INTEGRATION PLASMIDS	25
TABLE 2-2: LIST OF CONSTRUCTED PATHWAY ACCEPTOR PLASMIDS	26
TABLE 2-3: LIST OF PROMOTER SEQUENCES CLONED FROM <i>S.CEREVISIAE</i> S288C	27
TABLE 2-4: LIST OF TERMINATOR SEQUENCES CLONED FROM <i>S.CEREVISIAE</i> S288C. GENES IN RED REPRESENT THE GENES NOT INVOLVED DIRECTLY WITH THE ENZYMATIC PATHWAY.	28
TABLE 2-5: LIST OF GENES USED AND THEIR SOURCE	29
TABLE 2-6: EXACT MASS AND RETENTION TIME OF DMAT DETECTED.....	41
TABLE 2-7: EXACT MASS OF THE SIMPLEST FRAGMENT IONS FROM DMAT.....	45
TABLE 2-8: EXACT MASS AND RETENTION TIME OF DMAT AND 4DMA DETECTED.....	47
TABLE 2-9: EXACT MASS OF THE SIMPLEST FRAGMENT IONS FROM 4DMA	47
TABLE 3-1: LIST OF <i>EASE</i> ORTHOLOGS SELECTED FROM THE SSN GENERATED	56
TABLE 3-2: DETAILS OF THE SEQUENCE OF <i>EASC</i> SYNTHESIZED.....	57
TABLE 3-3: EXACT MASS AND RETENTION TIME OF CHANOCLOAVINE-I PRODUCED	70
TABLE 3-4: EXACT MASS AND RETENTION TIME OF MTCA AND DMA-MTCA PRODUCED	75
TABLE 3-5: EXACT MASS AND RETENTION TIME FOR CHANOCLOAVINE-I PRODUCED.....	80
TABLE 3-6: EXACT MASS OF THE SIMPLEST FRAGMENT IONS FROM CHANOCLOAVINE-I	80
TABLE 4-1: SEQUENCE DETAILS FOR THE GENES <i>EASD</i> AND <i>EASG</i> USED IN THE CONSTRUCTION OF AN AGROCLAVINE PRODUCER	87
TABLE 4-2: LIST OF HYPOTHETICAL ISOMERASE VARIANTS OF <i>EASA</i> SELECTED FROM THE SSN89	
TABLE 4-3: SUMMARIZED GENOTYPE OF THE REDESIGNED STRAINS.....	92
TABLE 4-4: LIST OF POTs ASSEMBLED FOR THE YMRWΔ15 INTEGRATION FRAGMENT	92
TABLE 4-5: LIST OF POTs ASSEMBLED FOR THE ARS308 INTEGRATION FRAGMENT.....	93
TABLE 4-6: LIST OF POTs ASSEMBLED FOR THE YPRCΔ17 INTEGRATION FRAGMENT	93
TABLE 4-7: EXACT MASS AND RETENTION TIME FOR THE AGROCLAVINE STANDARD AND AGROCLAVINE PRODUCED	100
TABLE 4-8: EXACT MASSES OF THE FRAGMENT IONS FROM AGROCLAVINE PRODUCED	100
TABLE 4-9: EXACT MASSES OF ¹³ C-2-AGROCLAVINE PRODUCED AND ITS FRAGMENT IONS	111
TABLE 5-1: SUMMARY OF <i>CLOA</i> ORTHOLOGS USED IN THIS STUDY	118
TABLE 5-2: SUMMARY TABLE OF THE <i>CLOA</i> ORTHOLOGS THAT WERE FOUND TO BE PRODUCING A COMPOUND WITH [M+H] ⁺ OF 269 M/Z	126
TABLE 5-3: EXACT MASS AND RETENTION TIME FOR THE COMPOUNDS DETECTED	127
TABLE 5-4: EXACT MASSES OF SELECT FRAGMENT IONS DETECTED FROM THE DLA STANDARD AND DLA DETECTED IN THE SAMPLES	128
TABLE 5-5: EXACT MASS AND RETENTION TIME FOR THE LYSERGOL STANDARD AND CLOA_PiPO SAMPLE, AS WELL AS THE MOST ABUNDANT FRAGMENT IONS	132
TABLE 5-6: SUMMARY OF THE CALCULATED DLA CONCENTRATION FROM THE CLOA_CPUR CONSTRUCT MEASURED BY STANDARD ADDITION	133
TABLE 5-7: EXACT MASS AND RETENTION TIME FOR THE DLA STANDARD AND DLAM33B SAMPLE, AS WELL AS THE MOST ABUNDANT FRAGMENT IONS	136
TABLE 5-8: SUMMARY OF THE CALCULATED DLA CONCENTRATION FROM THE CLOA_CPUR CONSTRUCT MEASURED BY STANDARD ADDITION	137
TABLE 5-9: EXACT MASSES OF ¹³ C-DLA PRODUCED AND ITS FRAGMENT IONS	146
TABLE 5-10: EXACT MASSES OF THE DETECTED 285, 286, 303, AND 304 M/Z COMPOUNDS	150
TABLE 5-11: COMPARISON OF THE PREDICTED LOGP VALUES AGAINST THEIR CORRESPONDING EXPERIMENTALLY DETERMINED RETENTION TIMES.....	152

LIST OF FIGURES

FIGURE 1-1: NATURAL AND SEMI-SYNTHETIC ERGOT ALKALOIDS WITH THEIR ESTABLISHED THERAPEUTIC USES. THE TETRACYCLIC ERGOLINE CORE HIGHLIGHTED IN PURPLE SERVES AS THE KEY PHARMACOPHORE, WHILE THE VARIOUS MODIFICATIONS TO THE RING HELP DETERMINE ITS TARGET RECEPTOR SPECIFICITY AND THEREFORE BIOLOGICAL ACTIVITY.	4
FIGURE 1-2: THE NUMBERING AND LETTERING CONVENTION FOR THE CORE ERGOLINE SCAFFOLD. ILLUSTRATION OF THE STRUCTURAL SIMILARITY BETWEEN SEROTONIN, DOPAMINE, AND ADRENALINE, ACCOUNTING FOR THE DIVERSE REPERTOIRE OF BIOLOGICAL ACTIVITIES OBSERVED FROM THE ERGOT ALKALOIDS. REPRESENTATIVE COMPOUNDS WITHIN THE MAJOR CLASSES OF ERGOT ALKALOIDS AND THEIR RESPECTIVE SUBDIVISIONS.	5
FIGURE 1-3: COMPLETE BIOSYNTHETIC PATHWAY OF THE ERGOT ALKALOIDS FROM TRYPTOPHAN TO THE ERGOPEPTIDES, INCLUDING THE DIVERGENT PATHWAY TO THE COMMUNESINS.	8
FIGURE 1-4: SCHEMATIC DIAGRAM OF THE YEASTFAB ASSEMBLY WORKFLOW, ILLUSTRATING THE HIERARCHICAL APPROACH TOWARDS ASSEMBLING COMPLEX MULTI-EXPRESSION CASSETTES FROM STANDARDIZED BIOLOGICAL PARTS.	19
FIGURE 1-5: (A) PATHWAY FOR THE BIOSYNTHESIS OF <i>D</i> -LYSERGIC ACID. (B) GENERAL STRATEGY TOWARDS RECONSTITUTING THE BIOSYNTHETIC PATHWAY TO <i>D</i> -LYSERGIC ACID IN YEAST. PART 1: <i>DMAW</i> AND <i>EASF</i> FOR THE PRODUCTION OF DMAT AND 4DMA. PART 2: <i>EASC</i> AND <i>EASE</i> FOR THE PRODUCTION OF CHANOC LAVINE-I. PART 3: <i>EASD</i> , <i>EASA</i> AND <i>EASG</i> FOR THE PRODUCTION OF AGROCLAVINE. PART 4: <i>CLOA</i> FOR THE PRODUCTION OF <i>D</i> -LYSERGIC ACID.	21
FIGURE 2-1: (A) ILLUSTRATION OF THE MODIFICATIONS MADE TO THE YEASTFAB ASSEMBLY WORKFLOW TO TAILOR THE NEEDS OF THIS PROJECT. (B) REPRESENTATIVE PLASMID MAP OF THE PATHWAY ACCEPTOR VECTORS CREATED. (C) REPRESENTATIVE PLASMID MAP OF THE GENOME INTEGRATION VECTORS CREATED. FULL IMAGE OF THE VECTOR MAPS OF (B) AND (C) ARE AVAILABLE IN APPENDIX A2.	36
FIGURE 2-2: CHARACTERIZATION OF YEAST PROMOTER STRENGTHS BASED ON EXPRESSION OF FLUORESCENT PROTEINS.	38
FIGURE 2-3: CHARACTERIZATION OF YEAST PROMOTER STRENGTHS BY PRODUCTION OF DMAT.	40
FIGURE 2-4: COMPARISON OF RETENTION TIME AND MASS SPECTRA OF DMAT PRODUCED <i>IN VIVO</i> (A) AND <i>IN VITRO</i> (B). MASS SPECTRA OF BOTH COMPOUNDS SHOW SIMILAR FRAGMENT IONS.	44
FIGURE 2-5: INTEGRATION OF <i>DMAW</i> AND <i>EASF</i> INTO THE YEAST GENOME FOR A 4DMA PRODUCING STRAIN.	46
FIGURE 3-1: (A) THE EARLY ERGOT ALKALOID PATHWAY LEADING TO CHANOC LAVINE-I FORMATION. THE FOCUS OF THIS CHAPTER IS THE EXTENSION OF THE PATHWAY FROM 4DMA TO CHANOC LAVINE-I. (B) HEME CATALASE ACTIVE SITE FROM THE PEROXISOMAL CATALASE OF <i>HANSENULA POLYMORPHA</i> , PDB ID: 2XQ1. (C) PREDICTED MODEL OF THE ACTIVE SITE IN <i>EASE_AJ</i> .	51
FIGURE 3-2: ILLUSTRATION OF METHOD DEvised TO CREATE A SUMO-TAG ENTRY-LEVEL VECTOR COMPATIBLE WITH YEASTFAB ASSEMBLY.	54
FIGURE 3-3: WORKFLOW FOR THE GENERATION OF SSN USING THE EFI-EST WEBTOOL.	55
FIGURE 3-4: ILLUSTRATION OF THE ORF RELEASED FROM pSGG-RFP. (A) FRAGMENT RELEASED FROM HCKAN_O. (B) FRAGMENT RELEASED FROM pSGG-RFP, SUMO-ORF FUSION. (C) VECTOR MAP OF pSGG-RFP.	64
FIGURE 3-5: SSN FOR THE EXPANDED ISOFUNCTIONAL CLUSTER OF <i>EASE</i> FROM THE REDEFINING OF THE ALIGNMENT SCORE FOR AN EDGE TO 170.	66
FIGURE 3-6: INCUBATION OF <i>EASC</i> WITH ECL SUBSTRATE PRODUCES CHEMILUMINESCENCE.	68

FIGURE 3-7: (A) ILLUSTRATION OF THE <i>EASE</i> SCREENING STRAIN (YMC17) AND SCREEN DESIGN. (B) REACTION PRODUCING CHANOCLOAVINE-I FROM 4DMA, CATALYSED BY <i>EASC</i> AND <i>EASE</i> . (C) EXTRACTED ION CHROMATOGRAMS OF <i>EASE</i> ORTHOLOGS TESTED FOR 257 M/Z. (D) RELATIVE AMOUNTS OF CHANOCLOAVINE-I PRODUCED FROM THE <i>A. JAPONICUS</i> AND <i>E. COENOPHILIA</i> ORTHOLOGS, ESTIMATED FROM THE AREA UNDER THE PEAK HIGHLIGHTED IN BLUE FROM C. (E, F) MASS SPECTRA FOR COMPOUND ELUTING AT 121 S.	69
FIGURE 3-8: SEQUENCE ALIGNMENTS FOR THE FAD-BINDING DOMAIN OF THE <i>EASE</i> ORTHOLOGS.	71
FIGURE 3-9: (A) CARTOON REPRESENTATION OF THE PMKU-C ₁₃ E ₇ STRAIN TESTED. (B) EXTRACTED ION CHROMATOGRAM OF THE STRAIN TESTED FOR 257 M/Z. (C) RELATIVE AMOUNTS OF DMAT, 4DMA, AND CHANOCLOAVINE-I PRODUCED, ESTIMATED FROM THE AREA UNDER THE PEAKS FOR THEIR RESPECTIVE EXTRACTED ION CHROMATOGRAMS.	72
FIGURE 3-10: PICTET-SPENGLER REACTION OF TRYPTOPHAN AND DMAT WITH FORMALDEHYDE PRODUCING THE SHUNT PRODUCTS MTCA AND ITS PRENYLATED FORM, DMA-MTCA.	73
FIGURE 3-11: (A) SHUNT PRODUCTS MTCA AND DMA-MTCA. (B) EXTRACTED ION CHROMATOGRAMS FOR 231 M/Z SHOWING THE PEAKS AT 102 S CORRESPONDING TO ONE STEREOISOMER OF MTCA. (C) EXTRACTED ION CHROMATOGRAMS FOR 299 M/Z SHOWING THE PEAK AT 181 S CORRESPONDING TO DMA-MTCA. (D, E) MASS SPECTRA FOR THE PEAKS AT 102 S FOR MTCA (F) MASS SPECTRA FOR THE PEAKS ELUTED AT 181 S FOR DMA-MTCA.	75
FIGURE 3-12: (A, B) BUFFERING THE CULTURE MEDIA AT pH 7 ERADICATES THE FORMATION OF MTCA AND DMA-MTCA. (C) PREVENTING THE FORMATION OF MTCA AND DMA-MTCA DOES NOT IMPROVE THE PRODUCTION OF CHANOCLOAVINE-I. (D) COMPARISON OF THE LEVELS OF INTERMEDIATES PRODUCED BETWEEN PMKU-C ₁₃ E ₇ AND PMKL-C _{GAL10} E _{GAL1} STRAINS.	76
FIGURE 3-13: (A) ILLUSTRATION OF STRAINS PMKL-C _{GAL10} E _{GAL1} AND THE GENOME INTEGRATED CONSTRUCT YOCE-C _{GAL10} E _{GAL1} . (B) EXTRACTED ION CHROMATOGRAMS FOR 257 M/Z, PEAK AT 123 S INDICATING THE PRODUCTION OF CHANOCLOAVINE-I FROM THE STRAINS CREATED. (C) EXTRACTED ION CHROMATOGRAMS FOR 257 M/Z OF THE <i>IN VITRO</i> BIOSYNTHETIC REACTIONS PERFORMED. THE PEAK ELUTING AT 146 S INDICATED THE PRODUCTION OF CHANOCLOAVINE-I. (D, E, F) COMPARISON OF THE MASS SPECTRA OBTAINED FROM THE PEAKS CORRESPONDING TO CHANOCLOAVINE-I.	79
FIGURE 4-1: REACTIONS CONVERTING CHANOCLOAVINE-I TO THE TETRACYCLIC ERGOCLAVINES.	84
FIGURE 4-2: PROPOSED REACTION MECHANISMS OF <i>EASA</i> FOR THE (A) ISOMERIZATION OF THE C8-C9 DOUBLE BOND IN THE FORMATION OF AGROCLAVINE; OR, (B) REDUCTION OF THE C8-C9 DOUBLE BOND IN THE FORMATION OF FESTUCLAVINE OR PYROCLAVINE.	85
FIGURE 4-3: WORKFLOW FOR THE GENERATION OF AN <i>EASA</i> SSN.	88
FIGURE 4-4: (A) INITIAL SSN GENERATED FOR <i>EASA</i> AT AN ALIGNMENT SCORE OF 85, CORRESPONDING TO 40% SEQUENCE IDENTITY; THIS WAS UNABLE TO DISTINGUISH ISOFUNCTIONAL CLUSTERS. (B) INCREASING THE ALIGNMENT SCORE TO 130, CORRESPONDING TO 60% SEQUENCE IDENTITY ALLOWED FOR THE IDENTIFICATION OF ISOFUNCTIONAL CLUSTERS.	96
FIGURE 4-5: EXPANDED SSN FOR THE ISOFUNCTIONAL CLUSTER OF <i>EASA</i> , OBTAINED BY DEFINING AN EDGE TO AN ALIGNMENT SCORE OF LARGER THAN 180.	97
FIGURE 4-6: SEQUENCE ALIGNMENT OF THE SELECTED <i>EASA</i> ORTHOLOGS WITH THE <i>N. LOLII</i> ORTHOLOG (M1S0Y0) AND <i>A. FUMIGATUS</i> ORTHOLOG (Q4WZ70) FOR THE CATALYTIC REGION CONTAINING THE IMPORTANT PHENYLALANINE/ TYROSINE RESIDE (HIGHLIGHTED IN RED).	98
FIGURE 4-7: SCREENING OF <i>EASA</i> ORTHOLOGS SELECTED.	99
FIGURE 4-8: QUANTIFICATION OF THE AGROCLAVINE PRODUCED BY THE <i>EASA</i> ORTHOLOGS.	102

FIGURE 4-9: SCREENING ALTERNATIVE SEGMENT DESIGNS.	104
FIGURE 4-10: (A) ILLUSTRATION OF THE POSSIBLE INVERSION OCCURRING ON THE <i>EAS</i> C/ <i>EAS</i> E CASSETTE AS A RESULT OF THE HOMOLOGY BETWEEN GAL1 AND GAL10 PROMOTER SEQUENCES. (B) ESTIMATION OF THE RELATIVE AMOUNTS OF INTERMEDIATES PRODUCED BY THE GA ^{E.COENOPHILIA} CD STRAIN FROM THE PEAK AREA RESPONSE OF THEIR RESPECTIVE EXTRACTED ION CHROMATOGRAMS AT 273, 287, 257, AND 239 M/Z.....	107
FIGURE 4-11: TESTING THE PRODUCTION YIELDS AND STABILITY OF THE NEW STRAINS AGCM33B AND AGCM32B.....	109
FIGURE 4-12: VALIDATION OF AGROCLAVINE PRODUCTION USING ¹³ C-LABELLED TRYPTOPHAN.	111
FIGURE 5-1: REACTIONS OXIDIZING THE CLAVINE ALKALOIDS IN THE LATE ERGOT ALKALOID PATHWAY.....	114
FIGURE 5-2: SEQUENCE SIMILARITY NETWORKS OF <i>CLOA</i>	116
FIGURE 5-3: EXPANDED SSN OF THE <i>CLOA</i> SUPERCLUSTER FROM FIGURE 5-2B, OBTAINED BY DEFINING AN EDGE AS A RELATIONSHIP CORRESPONDING TO A SEQUENCE IDENTITY GREATER THAN 60%.	117
FIGURE 5-4: EXPANDED SSN OF THE <i>CLOA</i> SUPERCLUSTER, OBTAINED BY DEFINING AN EDGE AS A RELATIONSHIP CORRESPONDING TO A SEQUENCE IDENTITY GREATER THAN 75%.	125
FIGURE 5-5: TESTING OF THE THREE <i>CLOA</i> ORTHOLOGS USING THE AGROCLAVINE PRODUCING STRAINS AGCM33B AND AGCM32B.....	127
FIGURE 5-6: MASS SPECTRA FROM THE MS-MS OF THE PEAKS ELUTING AT 125 S FROM THE (A) DLA STANDARD, (B) <i>C. PURPUREA</i> ORTHOLOG, (C) <i>E. COENOPHILIA</i> ORTHOLOG, AND (D) <i>P. IPOMOEAE</i> ORTHOLOG.	128
FIGURE 5-7: QUANTIFICATION OF THE DLA AND AGROCLAVINE PRODUCED IN THE PLASMID-BASED STRAINS BY STANDARD CURVE.....	129
FIGURE 5-8: (A) TOTAL ION CHROMATOGRAM FROM THE SELECTED ION MONITORING OF ALL FRAGMENTS PRODUCED FROM PRECURSORS WITH AN [M+H] ⁺ OF 255 M/Z. (B) MASS SPECTRA FROM THE MS-MS OF THE PEAK FROM THE LYSERGOL STANDARD. (C) MASS SPECTRA FROM THE MS-MS OF THE PEAK FROM THE <i>CLOA</i> _PIPO SAMPLE.	131
FIGURE 5-9: QUANTIFICATION OF DLA PRODUCED IN THE <i>CLOA</i> _CPUR SAMPLE BY STANDARD ADDITION.	133
FIGURE 5-10: CHARACTERIZATION OF DLAM33B.	135
FIGURE 5-11: QUANTIFICATION OF DLA PRODUCED BY DLAM33B BY STANDARD ADDITION.	136
FIGURE 5-12: QUANTIFICATION OF THE DLA AND AGROCLAVINE PRODUCED IN DLAM33B AND AGCM33B WHEN SUPPLEMENTED WITH ADDITIONAL TRYPTOPHAN.	139
FIGURE 5-13: (A) CALCULATED RELATIVE MOLARITY OF DLA/ AGROCLAVINE FROM SAMPLES FED WITH VARYING AMOUNTS OF TRYPTOPHAN SUPPLEMENTED. (B) THE FERRIC PEROXO INTERMEDIATE THAT IS GENERATED AS PART OF THE CYP450 MONOOXYGENASE CYCLE, AFTER PICKING UP MOLECULAR OXYGEN AND 2 ELECTRONS. (C) THE CATALYTIC CYCLE OF CYP450 MONOOXYGENASES PERFORMING THE ARCHETYPAL HYDROXYLATION REACTION.	140
FIGURE 5-14: CARTOON REPRESENTATION OF THE CLASS I (A) AND CLASS II (B) FUNGAL CYP450 SYSTEMS. (C) OVEREXPRESSION OF NCP1 FROM <i>S. CEREVISIAE</i> (NCP1_Sc) AND <i>A. THALIANA</i> (NCP1_At) ON A pESC-URA VECTOR TO SUPPLEMENT THE LEVELS OF CYPRs IN DLAM33B. (D) RELATIVE AMOUNTS OF DLA PRODUCED BY STRAINS OVEREXPRESSING NCP1_Sc AND NCP1_At DID NOT SHOW ANY SIGNIFICANT DIFFERENCE AGAINST THE EMPTY VECTOR NEGATIVE CONTROL OR DLAM33B.....	142
FIGURE 5-15: CHARACTERIZATION OF THE DLAM33B STRAIN BY FEEDING ¹³ C LABELLED TRYPTOPHAN.	143

FIGURE 5-16: PRODUCTION OF ^{13}C -DLA FROM ^{13}C -W. TOTAL ION CHROMATOGRAM FROM THE SELECTED ION MONITORING OF ALL FRAGMENTS PRODUCED FROM PRECURSORS WITH AN $[\text{M}+\text{H}]^+$ OF (A) 269 m/z, AND (B) 270 m/z. (C) OVERLAY OF THE MASS SPECTRA FOR THE PEAKS ELUTING AT 126 s. (D) MASS SPECTRA SHOWING THE DIFFERENCES BETWEEN CULTURES FED WITH TRYPTOPHAN OR ^{13}C -W.....	145
FIGURE 5-17: DETECTION OF ALL THE ^{13}C -LABELLED INTERMEDIATES ALONG THE RECONSTITUTED ERGOT BIOSYNTHESIS PATHWAY WHEN FED ^{13}C -LABELLED TRYPTOPHAN.	147
FIGURE 5-18: DETECTION OF MTCA AND DMA-MTCA FORMATION FROM ^{13}C -LABELLED TRYPTOPHAN.	148
FIGURE 5-19: DETECTION OF AN UNKNOWN PRODUCT DISPLAYING THE SAME +1 m/z SHIFT WHEN FED WITH ^{13}C -W.	149
FIGURE 5-20: REACTION MECHANISM PROPOSED BY KOZIKOWSKI ET AL., 1993, FOR THE FORMATION OF THE ERGOLINE C RING CATALYSED BY EASC AND EASE.	150
FIGURE 5-21: COMPARISON OF THE PREDICTED LOGP VALUES FOR ALL THE INTERMEDIATES OF THE ERGOT ALKALOID PATHWAY IN RELATION TO THE PROPOSED DETECTED INTERMEDIATES.	151

LIST OF ABBREVIATIONS

[M+H] ⁺	Molecular ion plus hydrogen
13C-AGC	¹³ C-2-Indole agroclavine
13C-DLA	¹³ C-2-Indole- <i>D</i> -lysergic acid
13C-W	¹³ C-2-Indole tryptophan
4DMA	4-Dimethylallyl- <i>L</i> -abrine
5FOA	5-Fluoroorotic acid
AUC	Area Under the Curve
bp	base pairs
BSA	Bovine Serum Albumin
CYP450	Cytochrome P450
DLA	<i>D</i> -Lysergic acid
DMA-MTCA	Dimethylallyl-MTCA
DMAPP	Dimethylallyl pyrophosphate
DMAT	4-Dimethylallyl- <i>L</i> -tryptophan
DNA	Deoxyribonucleic acid
ECL	Enhanced Chemiluminescence
EIC	Extracted Ion Chromatogram
EFI-EST	Enzyme Function Initiative – Enzyme Similarity Tool
ER	Endoplasmic Reticulum
ESI	Electrospray ionization
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
GFP	Green Fluorescent protein
GRAS	Generally regarded as safe
IPTG	Isopropyl β-thiogalactopyranoside
kDa	kilodalton

LB	Lysogeny Broth
LC-MS	Liquid Chromatography coupled Mass spectrometry
LSD	Lysergic acid diethylamide
m/z	mass-to-charge ratio
MCS	Multiple cloning sites
MOPS	3-(<i>N</i> -Morpholino)propanesulfonic acid
MS-MS	Tandem Mass spectrometry
MTCA	1-Methyl-1,2,3,4-tetrahydro- β -carboline-3-carboxylic acid
NAD ⁺	Nicotinamide adenine dinucleotide (oxidized)
NADH	Nicotinamide adenine dinucleotide (reduced)
NADP ⁺	Nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	Nicotinamide adenine dinucleotide phosphate (oxidized)
Ni ²⁺ -NTA	Nickel-Nitrilotriacetic acid
OD600	Optical density at 600 nm
PCR	Polymerase Chain Reaction
ppm	parts per million
PTFE	Polytetrafluoroethylene
rpm	revolutions per minute
SAM	S-adenosylmethionine
SC	Synthetic complete
SDS-PAGE	Sodium dodecyl sulphate - polyacrylamide gel electrophoresis
SSN	Sequence Similarity Network
TIC	Total Ion Chromatogram
TRIS	Tris(hydroxymethyl)aminomethane
v/v	volume per volume
w/v	weight per volume
YNB	Yeast Nitrogen bases

CHAPTER 1: INTRODUCTION

1.1 MICROBIAL CELL FACTORIES FOR THE PRODUCTION OF PRIVILEGED SCAFFOLDS

Synthetic biology is broadly defined as the design or redesign of biological systems towards the goal of imbuing some form of novel functionality (*Synthetic Biology: Applying Engineering to Biology : Report of a NEST High-Level Expert Group*, 2005). The field has been made accessible by the advancements made towards lowering the cost of DNA synthesis and sequencing (Hughes & Ellington, 2017). One of the premises from this perspective is that biology can be engineered in a manner similar to the design and engineering of electrical components (Endy, 2005).

An inherent application would be the development of microbial cell factories for the production of commercially important chemicals, as a “greener” and potentially more efficient substitute for current industrial practices based on synthetic chemistry. This is especially pertinent in our current era where global issues such as climate change and economic sustainability as a result of human industrial activity are becoming larger and more menacing concerns.

While the current level of biotechnology has not been able to provide a consistently economically competitive platform for the production of most bulk chemicals, high-value targets such as those for pharmaceutical products or flavour and fragrance compounds are ripe with opportunity (Erickson & Winters, 2012). Privileged structures or scaffolds, molecular structures that are capable of providing versatile and useful ligands for biological receptors, through the decoration of these scaffolds with different functional groups (Duarte, Barreiro, & Fraga, 2007), are an example of such high-value targets.

These structures are usually not easily accessible by total synthesis and are usually produced at low yields in nature.

The path towards improving the commercial viability of engineered microbial cell factories would be *via* targeting the production of these privileged scaffolds. This would enable the developed strains to compete across a broader market for all existing derivative products. Development efforts and costs would also be mitigated with the reduction of the degree of specialisation of the strain.

This work, therefore, seeks to demonstrate a proof of concept for the production of one such privileged scaffold – the ergoline ring, in the form of the common ergoline therapeutic precursor, *D*-lysergic acid.

1.2 THE ERGOT ALKALOIDS

1.2.1 A TUMULTUOUS JOURNEY

The history of the ergot alkaloids has been one which truly embodies the classical maxim in toxicology by Paracelsus, “*Alle Dinge sind Gift, und nichts ist ohne Gift, allein die Dosis macht das sein Ding kein Gift ist.*”, which translates, “All things are poison, and nothing is without poison, the dosage alone makes it so a thing is not a poison.” (Paracelsus, 1965). While the ergot alkaloids have been referenced throughout history for their therapeutic purposes, they were most notably remembered for their role in the mass poisonings they caused. This gave rise to the namesake “ergotism”, a description of the collective symptoms for ergot poisoning (de Groot, van Dongen, Vree, Hekster, & van Roosmalen, 1998).

Their use in modern medicine began in the early 1900s after Arthur Stoll had successfully isolated and identified ergotamine and ergometrine to be the key components from the sclerotia of the ergot fungus, *Claviceps purpurea*, that are responsible for the observed oxytocic effects (Minghetti & Crespi-Perellino, 1999). The use of ergotamine and ergometrine for the treatment of post-partum haemorrhage was short-lived. Their usage was soon found to be associated with an increase in the number of stillborn neonates. This prompted a search for ergot derivatives with fewer drastic side effects, that led to the eventual discovery of lysergic acid diethylamide (LSD) by Albert Hoffman (Hofmann, 2010).

As much as the reputation of the ergot alkaloids has been tarnished by their history, as both a poison and a psychedelic of abuse, their utility as a privileged scaffold cannot be discounted. LSD is one of the most potent drugs known, with binding affinities for several receptors reported in the low nanomolar range (Wacker et al., 2017). Several derivatives of the ergot alkaloids have also been found to be effective for numerous therapeutic applications in Parkinson's (Lieberman, Kupersmith, Estey, & Goldstein, 1976), dementia (Winblad et al., 2008), hypertension (Tandowsky, 1954), and possibly some forms of cancer (Mrusek, Seo, Greten, Simon, & Efferth, 2015) (Figure 1-1). Furthermore, LSD in recent times is being revisited for its application in treating depression and addiction (dos Santos et al., 2016; Liechti, 2017; Ly et al., 2018), despite the controversy amassed by its notorious use as a psychedelic of abuse.

For all the notoriety that has surrounded the history of the ergot alkaloids, there is still clearly a wealth of beneficial therapeutics yet to be tapped from the ergoline scaffold.

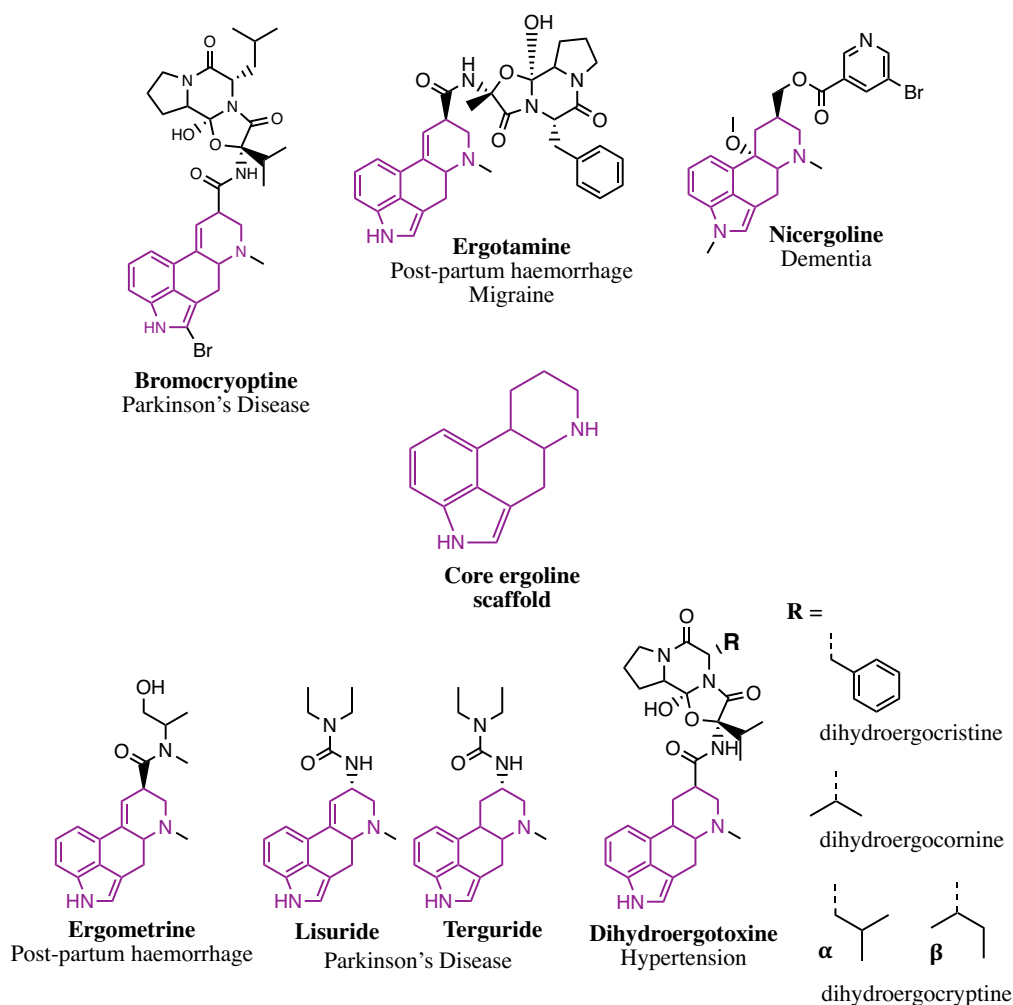


Figure 1-1: Natural and semi-synthetic ergot alkaloids with their established therapeutic uses. The tetracyclic ergoline core highlighted in purple serves as the key pharmacophore, while the various modifications to the ring help determine its target receptor specificity and therefore biological activity.

1.2.2 BIOSYNTHESIS AND GENETICS OF THE ERGOT ALKALOIDS

A major characteristic of most of the naturally occurring ergot alkaloids is the tetracyclic ergoline scaffold. This core scaffold, derived from *L*-tryptophan has been proposed to be the key structure responsible for the multitude of biological activities exhibited by the various ergot alkaloids and their derivatives (Pertz & Eich, 1999). This attributed to the way the ergoline ring mimics the positioning of the main functional groups in serotonin, dopamine, and adrenaline (Figure 1-2) (Mantegani, Brambilla, & Varasi, 1999).

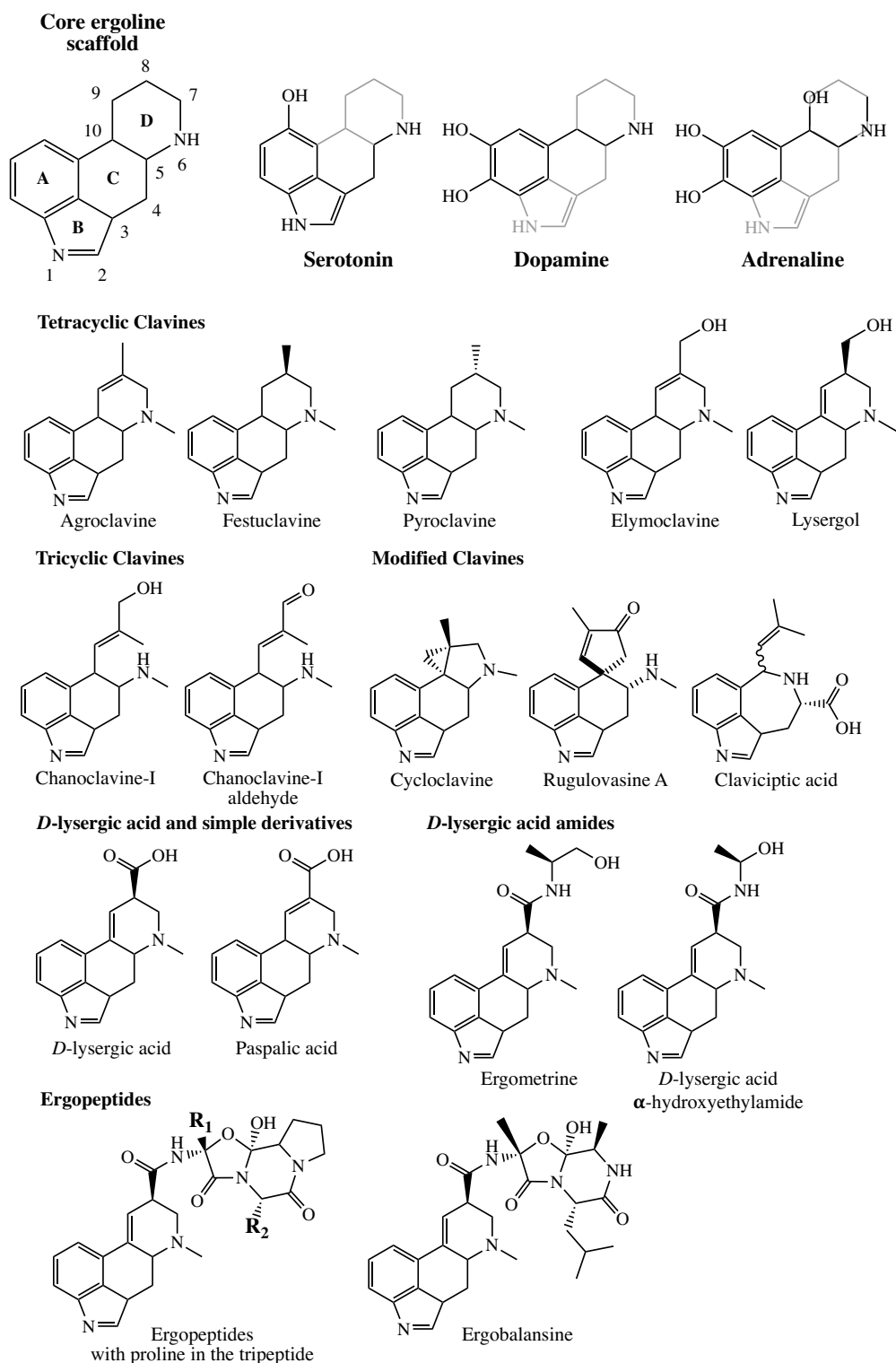


Figure 1-2: The numbering and lettering convention for the core ergoline scaffold. Illustration of the structural similarity between serotonin, dopamine, and adrenaline, accounting for the diverse repertoire of biological activities observed from the ergot alkaloids. Representative compounds within the major classes of ergot alkaloids and their respective subdivisions.

There are two main divisions of ergot alkaloids: the first are the derivatives of *D*-lysergic acid, which include the ergopeptides and other simple modifications to the carboxyl moiety. The second are the clavines, which encompasses the earlier intermediates along the pathway and their derivatives outside the main pathway to the tetracyclic ergot alkaloids, as well as the derivatives of 6,8-dimethylergoline (Figure 1-2).

The divergence is predominantly determined by the point at which they deviate from the pathway to the tetracyclic ergoline products. This has been attributed to the presence or absence of the respective late pathway enzymes that vary between the organisms that produce them. Our understanding of the biosynthesis of the ergot alkaloids has undoubtedly been advanced tremendously with the advent of Next-Generation Sequencing technologies and the ensuing flood of genomic data. Genome mining of the *Clavicipitaceae* and *Trichocomaceae* families of fungi, the main producers of these alkaloids, has identified several homologous gene clusters that could suitably account for the chemodiversity observed in the ergot alkaloids (Young et al., 2015).

Given the common starting point at *L*-tryptophan, the four enzymes involved in the early ergot pathway were found to be strongly conserved amongst all ergot-producing organisms. The pathway is initiated with the prenylation of tryptophan at the C-4 position of the indole ring with dimethylallyl pyrophosphate (DMAPP) catalysed by dimethylallyl prenyltransferase (*dmaW*) to produce 4-dimethylallyl tryptophan (DMAT). This step commits the substrate to the biosynthetic pathway (S.-L. Lee, Floss, & Heinstein, 1976). Next, a methyltransferase (*easF*/ FgaMT) catalyses the *N*-methylation of the amine at position 6 in DMAT, forming 4-dimethylallyl-1-

abrine (4DMA) (Otsuka, Quigley, Gröger, Anderson, & Floss, 1980; Rigbers & Li, 2008). The last detectable intermediate of the early pathway, chanoclavine-I, has been postulated to be a product of a multi-step reaction through the tandem action of genes with sequence similarity to a catalase (*easC*) and an FAD-dependent oxygenase (*easE*) (Goetz, Coyle, Cheng, O'Connor, & Panaccione, 2011; Lorenz, Olšovská, Šulc, & Tudzynski, 2010; K. Ryan, Moore, & Panaccione, 2013) (Figure 1-3, green arrows). In this reaction, the C10 position on the isoprene group is proposed to be first hydroxylated by *easC*, followed by a dehydration to produce a diene intermediate. This diene intermediate is then oxidized by *easE* to form an epoxide intermediate that cyclizes the C ring *via* the decarboxylation at the C5 position, which then leads to C5-C10 bond formation that opens the epoxide into chanoclavine-I (Kozikowski et al., 1993). However, the difficulty encountered with the heterologous expression and purification of *easE* has greatly limited any effort made to directly validate this hypothesis (Nielsen et al., 2014).

The next segment of the pathway is where the two main classes of ergot alkaloids diverge, most notably between the members of the *Clavicipitaceae* and *Trichocomaceae* families. Where the former tending to the formation of agroclavine and *D*-lysergic acid derivatives, the latter form the clavine type alkaloids that may lead to the formation of the fumigaclavines or the dihydroergot alkaloids.

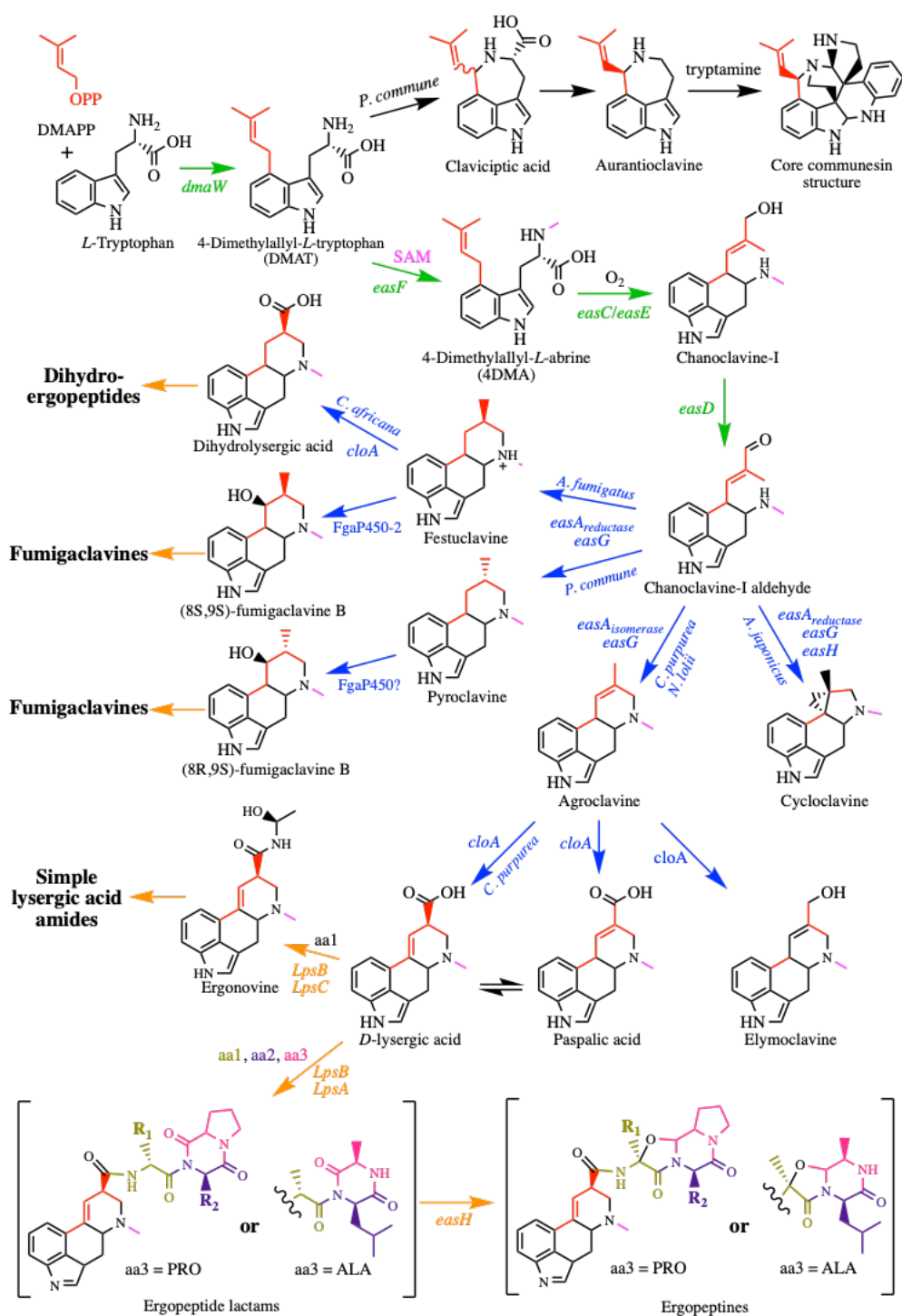


Figure 1-3: Complete biosynthetic pathway of the ergot alkaloids from tryptophan to the ergopeptides, including the divergent pathway to the communesins. Early ergot alkaloid pathway common to all ergot producing species are indicated by green arrows. Mid pathway diversification into the assortment of clavines found in the different lineages of ergot alkaloid producers are indicated by blue arrows. Late pathway derivatizations are indicated by orange arrows.

From chanoclavine-I, a short-chain dehydrogenase (*easD*) catalyses the oxidation of the -OH group using NAD^+ , to form chanoclavine-I-aldehyde (Wallwey, Matuschek, Xie, & Li, 2010). This step serves to activate the functional group for the cyclization of the ergoline D ring. The cyclization is catalysed by another enzyme pair, *easA* and *easG*. Sequence alignments of *easA* and its homologs show a high degree of similarity with the Old Yellow Enzyme (OYE) family. Previous work by Sarah O'Connor's group have demonstrated it to be responsible for reducing the C8-C9 double bond, which allows the carbonyl group to re-orientate and react with the secondary amine group to form the D ring and form an iminium intermediate (J. Z. Cheng, Coyle, Panaccione, & O'Connor, 2010b). This cyclization step is the first major structural branch point in the ergot alkaloid pathway, where the varying isoforms of *easA* and the presence of the downstream processing enzymes determine the ergot alkaloid profile. At least two isoforms of *easA* have been identified: a reductase form that completely reduces the C8-C9 double bond, and an isomerase form that re-orientates the carbonyl group and regenerates the C8-C9 double bond upon formation of the iminium intermediate (J. Z. Cheng, Coyle, Panaccione, & O'Connor, 2010a; Coyle, Cheng, O'Connor, & Panaccione, 2010). The enzyme *easG*, a NADPH oxidoreductase, catalyses the reduction of the iminium intermediates to form either festuclavine or agroclavine (Figure 1-3, blue arrows). Every branch from this point entails its unique lineage of enzymes, each tailoring the ergoline core further towards each individual end point.

For the production of *D*-lysergic acid, a cytochrome P450 enzyme (CYP450), clavine oxidase (*cloA*), catalyses the oxidation of the C8-methyl group (Haarmann, Ortel, Tudzynski, & Keller, 2006; Robinson & Panaccione,

2014). Early studies of *cloA* suggested that the oxidation of agroclavine is catalysed by the enzyme and possibly a second oxidase through a series of oxidized intermediates, elymoclavine and paspalic acid, leading to the formation of *D*-lysergic acid (Schardl, Panaccione, & Tudzynski, 2006). However, further studies have suggested that different isoforms of *cloA* might be capable of varying product profiles. Expression of the *C. purpurea cloA* in *C. fusiformis* produced *D*-lysergic acid, whereas the reverse did not, which corroborated with the ergot alkaloid profiles of *C. purpurea* and *C. fusiformis* (Lorenz, Wilson, Machado, Schardl, & Tudzynski, 2007). In the study conducted by the Panaccione group, expression of *cloA* from *Epichloe* sp. Lp1 in *A. fumigatus* produced only *D*-lysergic acid and no detectable paspalic acid or elymoclavine, suggesting that they may not be distinct intermediate products from disparate oxidation reactions (Robinson & Panaccione, 2014). More recently, a CYP450, sharing homology to *cloA* in certain festuclavine-producing species, was found to be responsible for catalysing a similar reaction on festuclavine to produce dihydrolysergol and dihydrolysergic acid (Arnold & Panaccione, 2017). These studies suggest that a variety of oxidised agroclavine/festuclavine products isolated from the different ergot producing species are the result of the varied product profiles from their respective *cloA* orthologs (Figure 1-3, blue arrows).

Further down the pathway to the ergopeptides, a pair of Non-Ribosomal Peptide Synthetases (NRPS) introduce another level of diversification. *LpsB* catalyses the activation *D*-lysergic acid or dihydrolysergic acid such that *LpsA* or *LpsC* may append either an assortment of tripeptide moieties or an *L*-alanine residue respectively, onto the acyl group to produce an ergopeptide (Walzel,

Riederer, & Keller, 1997). This is a manner similar to earlier branch points in the pathway, where different isoforms of *LpsI* present varying specificities to different amino acids substrates and add to the diversity of the ergopeptides observed (Ortel & Keller, 2009). Finally, an Fe^{2+} /2-ketoglutarate-dependent dioxygenase (*easH*), hydroxylates the *N*-(*D*-lysergyl-aminoacyl)-lactam which allows the spontaneous cyclolization into the terminal ergopeptine (Havemann, Vogel, Loll, & Keller, 2014) (Figure 1-3, orange arrows).

While the focus of this work is on the heterologous biosynthesis of *D*-lysergic acid and its accessible derivatives, the rich chemodiversity of the ergot alkaloids is not limited to this scope. In 2015, a group at UCLA were able to trace the biosynthesis of communesins from the *Penicillium* fungi to originate from the early ergot pathway intermediate, DMAT. DMAT was found to be shunted out into clavicipitic acid and subsequently aurantioclavine, *via* enzymes homologous to *easE* and *easC* (Lin et al., 2015); this was initially thought to be a derailment of the pathway (Figure 1-3, black arrows). Modified ergoline scaffolds, such as cycloclavine and rugulovasine can also be found from other species of *Aspergillum* and *Penicillium* fungi (Dorner, Cole, Hill, Wicklow, & Cox, 1980; Furuta, Koike, & Abe, 1982) (Figure 1-2).

Millions of years of evolution have inherently built an elaborate system for the diversification of ergot alkaloids. The development of a microbial cell factory for the production of a key ergoline precursor would set the stage for the next leap forward in the discovery of new natural or semisynthetic ergot-based therapeutics.

1.2.3 COMMERCIAL RELEVANCE AND CURRENT PRODUCTION METHODS

As it stands, the current demand for ergopeptines and lysergic acid for use in the manufacture of the semisynthetic ergot alkaloids-based therapeutics is estimated to be 8 and 10-15 tons respectively. Of this, approximately 60% of the supply is produced by the submerged fermentation of specially cultivated strains of *C. purpurea*, while the rest are obtained from field cultivation (Cvak, 1999).

Field cultivation of the ergot fungi as a commercial source of ergot alkaloids is the oldest large-scale method for its production. It entails the parasitic culture of the fungi on top of fields of susceptible cereal crops, most usually rye (Loveless, 1971). The main feature of field cultivation is the large variety of ergot alkaloids produced. This method, when adequately optimized, stands to be the most economically competitive method (Németh, 1999). It should be noted, however, that it is no trivial matter to maintain the required conditions for optimum yield. As with most agricultural methods, the entire process is affected by a whole suite of environmental factors that influence both the sclerotia yield as well as the ergot alkaloid yield per sclerotia (Németh, 1999). Furthermore, the parasitic nature of the ergot fungi adds another dimension of complexity, where additional factors such as infection rates and host effects on the variety of ergot alkaloids produced needs to be considered as well. One large concern surrounding field cultivation is the extensive care required to ensure the ergot infection does not spread beyond the desired cultivation area. Due to the potential risk of an outbreak of ergotism in both humans and livestock, ergot cultivation areas cannot be in the vicinity of any farms cultivating susceptible cereal strains.

The efficiency of ergot cultivation would therefore benefit greatly if they could be produced by fermentation, where the growth medium can be better controlled to reduce undesired side products and improve yields. Saprophytic culture of the ergot fungus is not a new concept; stationary surface cultures were originally used for the production of infecting material in field cultivation. These cultures involved the growth of the fungi either on the surface of a solid or liquid medium (Malinka, 1999). While these were useful for preparing infectious material, the surface area required makes it inadequate for industrial-scale manufacturing. Submerged cultivation is currently the method of choice for most microbial-based fermentation methods. Unfortunately, wild-type ergot producers are usually filamentous in nature and only produce the ergot alkaloids *in planta*. Special sclerotia-like strains adept for submerged cultures have been developed through extensive strain development, but these strains require constant selection to prevent any reversions or degeneration over long-term storage and successive generations in scale-up production (Malinka, 1999).

Over the years, a fair amount of effort has been geared towards strain engineering for reduced product variability and improved fermentation yields. The use of synthetic biology tools to deliver a microbial cell factory for ergot alkaloids presents an elegant solution to these persistent issues in ergot alkaloid manufacture. Through this approach, a single directed heterologous pathway could be grafted into each strain for the production of a single product, immediately resolving the problem of producing multiple types of ergot alkaloids that complicates the separation process. Developing such a strain

would also undoubtedly be crucial in understanding the extent of ergot alkaloid diversity.

1.3 A HETEROLOGOUS ERGOT PRODUCER: THE CURRENT STATE

The allure of a microbial cell factory capable of producing a singular ergot alkaloid product that is unencumbered by the natural regulatory elements in secondary metabolite biosynthesis has over the years attracted a fair amount of interest. Several groups have had exemplary amounts of success engineering certain parts of the pathway into a few heterologous systems, specifically in *A. nidulans*, *A. fumigatus*, and *S. cerevisiae* (Jakubczyk et al., 2015; Nielsen et al., 2014; Robinson & Panaccione, 2014; K. Ryan et al., 2013). These demonstrations have helped establish the necessary groundwork for our understanding of the pathway.

Daniel Panaccione's group first successfully demonstrated that the early ergot pathway could be grafted from *A. fumigatus* into *A. nidulans*, a model organism of the *Aspergillum* genera that does not contain the ergot biosynthesis gene cluster, for the production of chanoclavine-I. Their work tested out combinations of the early pathway to chanoclavine-I and validated earlier *easC* and *easE* knockout experiments from a bottom-up approach. This work confirmed that both enzymes are required for the turnover of 4DMA into chanoclavine-I (K. Ryan et al., 2013). They also identified the accumulation of ergotryptamine in strains lacking *easE*, suggesting that the decarboxylation of 4DMA is possibly independent of C-ring formation and that *easC* is the likely candidate for the C10 hydroxylation of 4DMA (K. L. Ryan, Akhmedov, & Panaccione, 2014).

The following year, they removed *easA* in *A. fumigatus* and replaced it with an ortholog from *Epichloe sp.* Lp1. This strain was shown to have switched from a festuclavine to an agroclavine producer, and the addition of *cloA* from *Epichloe sp.* Lp1 yielded an *A. fumigatus* strain capable of producing *D*-lysergic acid (Robinson & Panaccione, 2014).

A parallel series of experiments was performed by Sarah O'Connor's group in *S. cerevisiae*, using the orthologs of *dmaW*, *easF*, *easC*, and *easE* they had identified from *A. japonicus*. In their work, they identified signal peptide sequences in *easE* and *easC* and showed that they were crucial for *easE* function, presumably for its proper folding and generation of a holo-*easE* with FAD within the cell's ER, but not essential in *easC* (Nielsen et al., 2014). They also found that different orthologs of *easC* and *easE* from different sources were still be able to produce chanoclavine-I at similar levels. In a subsequent publication, they reported successfully extending the pathway to the formation of cycloclavine and festuclavine through the incorporation of *easD*, *easA*, *easG*, and *easH* (all sourced from *A. japonicus*) to their yeast host (Jakubczyk et al., 2015). Through some bioprocess optimization in a 1 L fermenter, they were able to produce cycloclavine from their yeast strain at a reasonably high titre of 529 mg L⁻¹.

1.4 SKETCHING THE BLUEPRINT FOR A MICROBIAL ASSEMBLY LINE

In spite of the success achieved in reconstituting parts of the ergot alkaloid pathway described earlier, it should be noted that they were performed using more traditional methods in molecular cloning designed for the purpose of the heterologous expression of one or two proteins. While excellent for directed

studies of a few gene targets, they are not particularly suited for the manipulation and optimization of large multi-component biosynthetic pathways. Even in the landmark works by the Keasling group for the semi-synthetic artemisinin project and the Smolke group's complete biosynthesis of the opioids, the methods employed to generate the ground-breaking constructs were tedious and less than congenial to automation (Galanie, Thodey, Trenchard, Interrante, & Smolke, 2015; Paddon et al., 2013).

A core tenet in synthetic biology is the ability to easily and rapidly iterate through the “Design→Build→Test→Learn” cycle, where biological data may be generated at a higher throughput to better match the pace of computational tools used for modelling, and design-of-experiment to facilitate the intelligent design and improvement of desired goals in a synergistic manner. At the heart of this issue is finding a suitable DNA assembly method that would adequately fulfil the requirements for this project. This would be ideally a system that allows ample flexibility for the variation of modular parts yet simple enough to be made compatible for automation and high-throughput construction.

1.4.1 BREWING THE NEXT WAVE OF ERGOT ALKALOID PRODUCTION

The first consideration towards creating a microbial cell factory is the choice of a host organism. From the engineering perspective, the host organism serves as the “motherboard” on which the additional “hardware” and associated “programming” are built on. Picking a DNA assembly method is then akin to choosing a programming language that would need to be compatible with the host.

For this project, *S.cerevisiae* or baker's yeast was selected as the host organism for a number of reasons. Firstly, the required post-translational modification essential for the functional expression of the enzyme *easE* (Nielsen et al., 2014) and the documented improved capability to express the cytochrome P450 family of enzymes, which *cloA* has been identified to belong to (Gonzalez & Korzekwa, 1995). Secondly, *S. cerevisiae* is not filamentous in nature, making it a more suitable host for the fermentative production of chemicals. Thirdly, as a model organism, there exists a wealth of knowledge that has been generated surrounding all aspects of its genetics, physiology, and metabolism (Walker, 1998). This repository of information would greatly facilitate the metabolic engineering and bioprocess optimization of the developed strain. Lastly, the ongoing Sc2.0 project promises an additional platform for further strain engineering and optimization through the use of SCRaMbLE (Jia et al., 2018). As the workhorse behind the millennia of producing alcoholic beverages, the associated fermentation technology would likely be easily translated to accommodate the new strains (Walker, 1998).

These points, in addition to its GRAS status and the comparative ease of introducing large DNA fragments into the yeast genome through homologous recombination, make it one of the most preferred hosts for the development of microbial cell factories (Yuan & Alper, 2019).

1.4.2 YEASTFAB ASSEMBLY: A GENOME COMPILER

Given the increasing complexity in the design of genetic networks and pathways that drive the field of synthetic biology, molecular biology toolkits are required to be versatile, efficient, and compatible with automation. A multitude of DNA assembly systems have since been developed to meet this

need (Casini, Storch, Baldwin, & Ellis, 2015). These methods are usually creative new ways of utilizing one or more of the classical methodologies such as restriction-ligation technology, homologous recombination, or long overlap PCR.

One system that most adequately meets the needs of this project is the Golden Gate method. It is a restriction-ligation based technique that capitalizes on the unique ability of type IIS restriction enzymes to cleave DNA a few bases down from its recognition site. This allows the user to define the overhanging sequences used and enables the concatenation of multiple DNA fragments in a predefined order all within a single reaction tube (Engler, Kandzia, & Marillonnet, 2008). Amongst the many Golden Gate systems available, the YeastFab assembly method provides a framework that would fit the desired workflow of this project with the fewest modifications (Guo et al., 2015).

Essentially, the system standardizes the core parts – promoters, ORFs, and terminators, on to entry-level vectors. These vectors can be hierarchically mixed and matched together to form assorted transcriptional units or “POTs” followed by pathways. The assembled pathways can be integrated *via* homologous recombination into the HO locus of their developed acceptor strain (JDY52) or into their designed plasmid (Figure 1-4).

To accommodate the workflow for this study, the pathway acceptor plasmids and the loci for genomic integration were modified to allow a direct assembly of the pathway into a pathway acceptor vector and integration onto sites within the genome of any *S. cerevisiae* S288C-derived strain. The details of these modifications are discussed in Chapter 2.

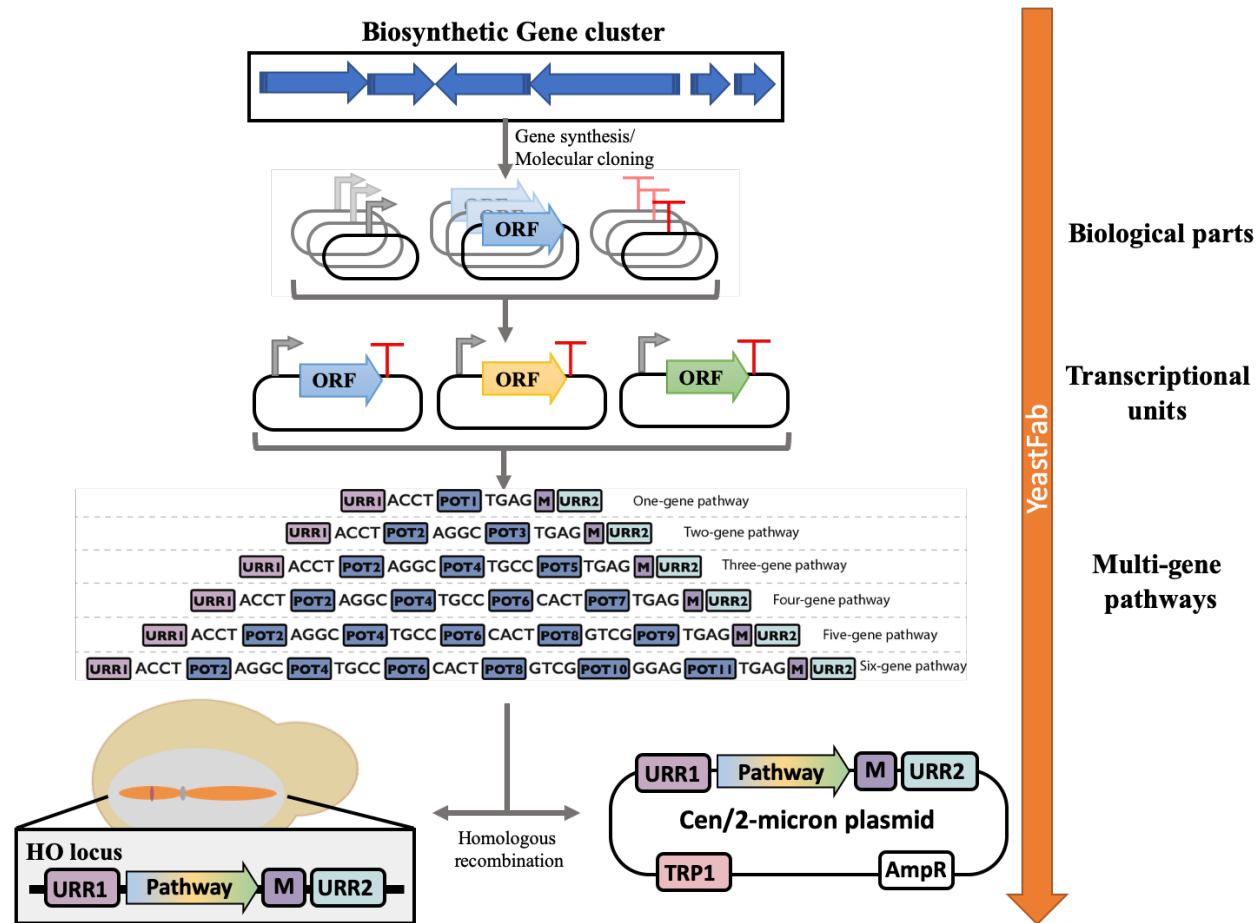


Figure 1-4: Schematic diagram of the YeastFab Assembly workflow, illustrating the hierarchical approach towards assembling complex multi-expression cassettes from standardized biological parts. Adapted from Guo et. al., 2015

1.4.3 DESIGN→ BUILD→ TEST→ LEARN

The reconstitution of the ergot alkaloid pathway to *D*-lysergic acid in yeast involves the heterologous expression of the eight enzymes that catalyse the six chemical transformations from *L*-tryptophan (Figure 1-5A). The approach taken for this project was to reconstruct the pathway in segments, partitioned at crucial points in the pathway whose products are easily detectable. The segments were built in *E. coli* and tested in yeast. They were initially screened as plasmids then subsequently integrated into the genome. This approach facilitated the troubleshooting process as well as enabled each prior strain to serve as a control for each subsequent strain.

Given the lack of commercially available standards for most of the pathway, a preparative strain serving as a negative control for each extension to the pathway was essential for demonstrating the creation of a new product (Figure 1-5B). This also confers the advantage of having strains at points along convenient branch points of the pathway to easily probe diverging directions down alternative branches in the ergot alkaloid pathway.

In the development of the next-generation of microbial cell factories for the production of ergoline-derived therapeutics, the aims of this study are:

- 1) The characterization of a set of yeast expression tools in the YeastFab assembly framework for rapid, modular pathway construction and optimization.
- 2) The complete reconstitution of the ergot alkaloid biosynthesis pathway to *D*-lysergic acid in *Saccharomyces cerevisiae*.

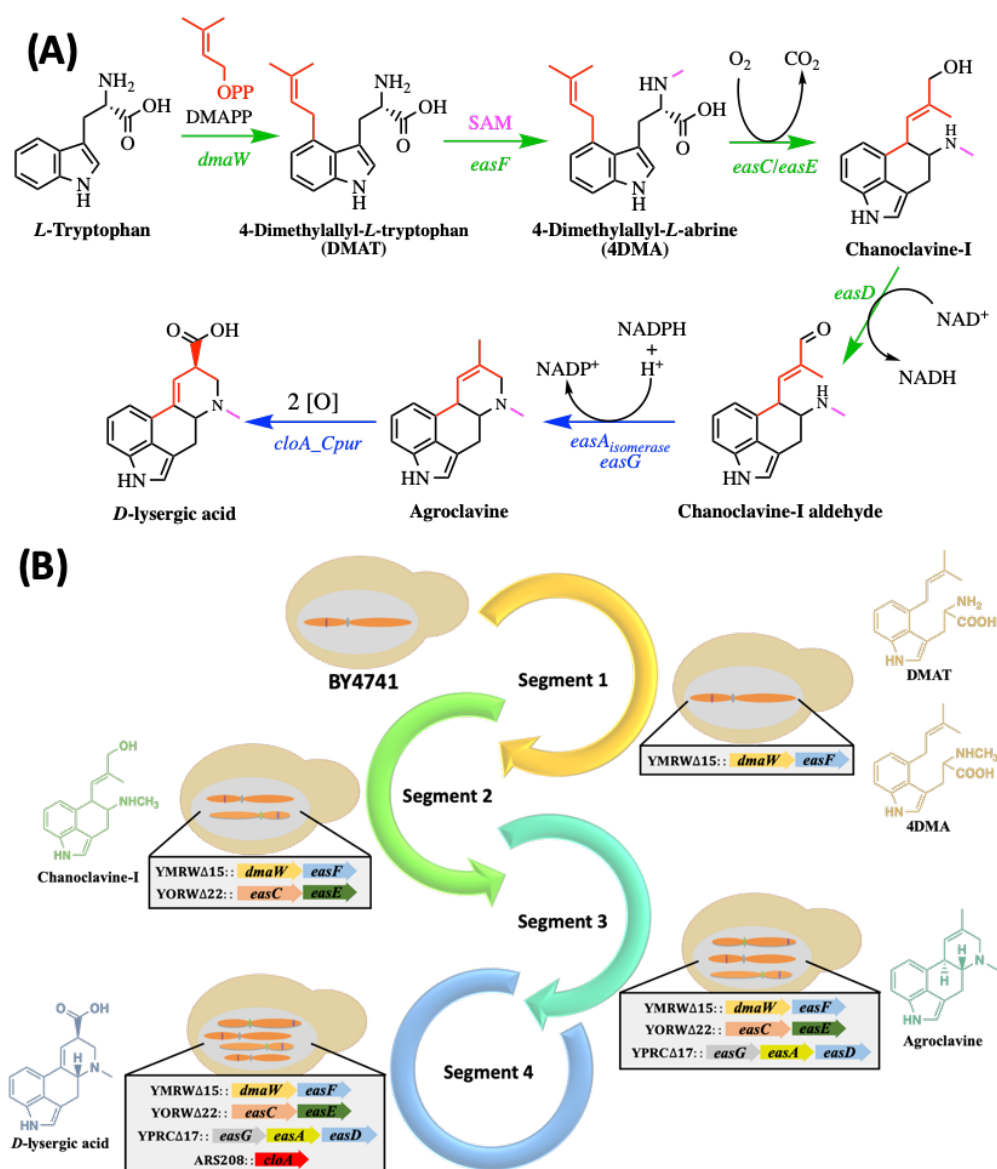


Figure 1-5: (A) Pathway for the biosynthesis of *D*-lysergic acid. (B) General strategy towards reconstituting the biosynthetic pathway to *D*-lysergic acid in yeast. Part 1: *dmaW* and *easF* for the production of DMAT and 4DMA. Part 2: *easC* and *easE* for the production of chanoclavine-I. Part 3: *easD*, *easA* and *easG* for the production of agroclavine. Part 4: *cloA* for the production of *D*-lysergic acid.

CHAPTER 2: LAYING THE FOUNDATIONS FOR A MICROBIAL CELL FACTORY

2.1 INTRODUCTION

Secondary metabolites such as the ergot alkaloids and numerous other biologically active compounds are typically biosynthesized through long and tenuous pathways. These pathways, more often than not, are comprised of several enzymes and are usually complexly regulated. The reconstitution of these pathways in more amenable heterologous hosts through traditional cloning methods are extremely tedious and time-consuming. The key challenge for the adaptation of a bio-based chemicals economy, built on microbial cell factories, is that first generation strains typically produce minute quantities of the target compound (Galanie et al., 2015). Extensive optimization is then required to try bring their yields and titres up to commercially competitive levels. Unfortunately, most of these prototypes are abandoned because the costs of further optimization are prohibitive and design optimization entails the reconstruction of the pathway painstakingly from scratch.

This project aims to design and build the ergot alkaloid pathway in yeast with the considerations of these issues in mind. The use of the Golden Gate methodology with the help of automation, should enable any further downstream pathway optimization and redesign to be performed rapidly, once the biological parts and major issues have been ironed out from the first iteration. To this end, the YeastFab assembly method was selected amongst the many available Golden Gate kits. The advantage of this system over the conventionally more popular kits such as the MoClo system (M. E. Lee, DeLoache, Cervantes, & Dueber, 2015), is the reduced number of parts for each

assembly reaction. In the many Golden Gate systems described in the literature, a common reported finding is that the assembly efficiency decreases as the number of assembly fragments increases (Engler et al., 2008). Given the large nature of this particular pathway, it would be prudent to limit the number of assembly fragments as much as possible - a trade-off in assembly flexibility for a higher degree of assembly efficiency.

The YeastFab assembly kit however, is not without its limitations. Firstly, it depends on the homologous recombination machinery in yeast to affix the assembled pathway onto either their pathway acceptor vector or onto the specially modified HO locus within their unique strain, JDY52. While *S. cerevisiae* has been shown to be an efficient system for the recombination of DNA, it is not the most suitable host to attempt to rapidly iterate through the design→build→test cycle. The doubling time for *E. coli* is half that of yeast (Plank & Harvey, 1979; Salari & Salari, 2017), which makes *E. coli* a much more attractive host to rapidly iterate through the cloning and assembly process. Furthermore the screening process of yeast transformants is complicated by the presence of a tough chitinous cell wall, this is not present in *E. coli* and hence simplifies and expedites the screening process.

Secondly, the YeastFab assembly kit offers only one integration locus for pathway construction. While it is theoretically possible to affix the entire pathway onto the site, having a single massive modification on one point in the genome would unlikely produce a very stable strain. It would be more prudent to distribute smaller parts of the pathway over multiple sites in the design of a stable production strain. Having multiple integration loci would also allow for

the incorporation of multiple copies of the same gene, thus enabling another method for expression level and metabolic flux tuning.

Lastly, the strain JDY52 used in the YeastFab system is a tryptophan auxotrophic strain. Building a pathway that utilizes tryptophan as the starting material in such a host would decouple the pathway from central metabolism and require to be constantly supplied with exogenous tryptophan for the production of *D*-lysergic acid as well as cell growth. A more efficient design would utilize a strain that allows for the pathway to tap into central metabolism and produce the desired products from the most economical feedstock available. It is hence imperative to adapt this system to fulfil these requirements prior to the assembly of the pathway..

2.2 MATERIALS AND METHODS

2.2.1 ADAPTATIONS TO THE YEASTFAB SYSTEM

The pathway acceptor plasmids and genome integration plasmids were created through Gibson Assembly using their published method (Gibson et al., 2009). For the genome integration plasmids, they were assembled from five fragments, each containing a separate vector element. The first fragment containing the URR2-URA3-URR2 was created *via* overlap PCR of the URR2 regions and the URA3 marker before being assembled with the remaining fragments. The second fragment containing the RFP expression cassette is flanked by the pathway insertion sites, which were amplified from the HCKan_O vector of the original YeastFab suite of plasmids. The third fragment containing the Amp resistance gene and the ColE1 origin of replication was amplified by PCR from the POT vectors of the original YeastFab suite of plasmids. The fragments containing the upstream and downstream genome integration homology regions were amplified from *S.cerevisiae* S288C genomic DNA by PCR. The integration sites chosen were based on reported integration efficiency and expression levels for those loci from the Nielson and Keasling groups (Table 2-1) (Apel et al., 2016; Bai Flagfeldt, Siewers, Huang, & Nielsen, 2009).

Table 2-1: List of constructed genome integration plasmids

Plasmid	Integration locus	Chromosome	Reference
pGAU-YMRW-RFP	YMRWΔ15	Chr. XIII	(Bai Flagfeldt et al., 2009)
pGAU-YPRC-RFP	YPRCΔ15	Chr. XVI	
pGAU-YORW-RFP	YORWΔ17	Chr. XV	
pGAU-A308-RFP	ARS308	Chr. III	(Apel et al., 2016)
pGAU-A208-RFP	ARS208	Chr. II	

The same strategy was used for the creation of the pathway acceptor plasmids (Table 2-2) and they were assembled from four fragments. The fragments containing the Kan resistance gene and ColE1 origin, and the RFP cassette flanked by the pathway insert sites were amplified by PCR from the HCKan_O plasmid. The fragments containing the yeast origin of replication (2-micron/ CEN-ARS) were amplified from pYES2/CT and the POT vectors respectively. The last fragment containing the yeast selection marker (URA3/ HIS3/ LEU2) were amplified by PCR from the POT vectors, pMV-HIS3, and pMV-LEU2 respectively. Both sets of pathway acceptor plasmids and genome integration plasmids were verified by sequencing and restriction digest.

Table 2-2: List of constructed pathway acceptor plasmids

Plasmid	Yeast origin of replication	Yeast selection marker	<i>E.coli</i> origin of replication	<i>E.coli</i> selection marker
pMKU	2-micron	URA3	ColE1	Kan ^R
pMKH	2-micron	HIS3	ColE1	Kan ^R
pMKL	2-micron	LEU2	ColE1	Kan ^R
pCKU	CEN-ARS	URA3	ColE1	Kan ^R
pCKH	CEN-ARS	HIS3	ColE1	Kan ^R
pCKL	CEN-ARS	LEU2	ColE1	Kan ^R

2.2.2 IDENTIFICATION AND CLONING OF BIOLOGICAL PARTS

2.2.2.1 PROMOTERS AND TERMINATORS

The promoters used in this project were chosen from the database published by the Promoter Database of *S.cerevisiae* (SCPD)(Zhu & Zhang, 1999). The selected promoters were chosen by the virtue of them having been characterized in other studies; this allowed for easy cross-reference when validating the promoter strength data produced in this study. Despite the extensive research done on baker's yeast, there is still a lack of understanding

surrounding its promoter sequence architecture. To simplify what constitutes a yeast promoter sequence and for modularity, each promoter is defined to be the sequence -500 base pairs to 0 upstream of its corresponding open reading frame. The selected promoters are thus named after the gene in the open reading frame they drive.

For this project, 19 such sequences were chosen and are listed in Table 2-3. They were amplified from the genomic DNA of *S.cerevisiae* S288C and cloned into the HCKan_P vectors and the pGY026 vector using the Golden Gate method and verified by sequencing.

Table 2-3: List of promoter sequences cloned from *S.cerevisiae* S288C

Promoter	Abbreviated Number	Promoter	Abbreviated Number	Promoter	Abbreviated Number
ENO2	1	GPM1	8	RPL3	-
PDC1	2	PDA1	9	HXT7	-
PGK1	3	PGI1	10	ACT1	-
PYK1	4	RPL8B	11	YEF3	-
TDH2	5	RPL15	12	TEF1	-
TEF2	6	TDH3	13		
TPI1	7	SSA1	14		

Terminator sequences used in this project were selected from a published database that were characterized for their effect on transcription and translation (Yamanishi et al., 2013). Similar to the promoter sequences, they were defined as the region beginning after the stop codon of an open reading frame to +500 base pairs downstream.

From the database, 17 sequences with the smallest difference in expression strength were chosen. To reduce the complexity and possible

number of promoter-gene-terminator combinations, each gene was arbitrarily assigned a terminator sequence in the course of developing the prototype strain (Table 2-4). The selected sequences were amplified from genomic DNA of *S.cerevisiae* S288C and cloned into the HCKan_T vectors using the same Golden Gate protocol and verified by sequencing.

Table 2-4: List of terminator sequences cloned from *S.cerevisiae* S288C. Genes in red represent the genes not involved directly with the enzymatic pathway.

Systematic Name	Gene Name	Relative expression strength (tPGK1)	Assigned gene
YHL015W	RPS20	1.953	<i>easE</i>
YDR165W	TRM82	1.959	ERO1
YKR001C	VPS1	1.961	PDI1
YGL161C	YIP5	1.967	<i>easC</i>
YDR368W	YPR1	1.968	<i>easD</i>
YLR206W	ENT2	1.971	<i>dmaW</i>
YER031C	YPT31	1.977	<i>easA</i>
YBL064C	PRX1	1.982	<i>easF</i>
YHL039W	EFM1	1.983	FAD1
YKL135C	APL2	2.022	<i>easG</i>
YDR206W	EBS1	2.104	-
YDL223C	HBT1	2.110	-
YLR029C	RPL15A	2.123	-
YDL133C-A	RPL41B	2.127	<i>cloA</i>
YDR382W	RPP2B	2.144	NCP1
YBR127C	VMA2	2.146	-
YIL043C	CBR1	2.156	mKOK

2.2.2.2 IDENTIFICATION, CLONING, AND PURIFICATION OF *dmaW* AND *easF*

The first two enzymes of the ergot alkaloid pathway, *dmaW* and *easF*, have been studied extensively and a structure previously obtained for *dmaW* (Metzger et al., 2009; Rigbers & Li, 2008; Stec et al., 2008; Unsöld & Li, 2005). The gene sequences were codon-optimized for yeast expression and synthesized into pET15b vectors by Genescript (Table 2-5). Using the synthesized genes as templates, they were amplified by PCR and cloned into the HCKan_O vectors *via* Golden Gate assembly for use in the YeastFab assembly system.

Table 2-5: List of genes used and their source

Gene	Source	PDB ID	Uniprot ID	Reference
<i>dmaW</i>	<i>Aspergillus fumigatus</i>	3I4X	Q50EL0	(Metzger et al., 2009)
<i>easF</i>	<i>Aspergillus fumigatus</i>	-	Q4WZ60	(Rigbers & Li, 2008)

The proteins were also recombinantly expressed in *E.coli* BL21(DE3) using the pET15b constructs. Cells were grown in 2YT media to an OD₆₀₀ of 0.7 at 37 °C, prior to induction with IPTG (2 mM) at 20 °C for 20 hrs, before they were harvested by centrifugation at 4 °C, 5000 rpm for 10 mins.

Harvested cells were resuspended in binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) at 4 °C and lysed *via* sonication. Cell debris was removed by centrifugation at 15000 rpm for 20 mins at 4 °C. All subsequent purification steps were done at 4 °C. The cell supernatant was added to 200 μ L of Ni²⁺-NTA chelating sepharose resin equilibrated in binding buffer. This was incubated for 30 mins with shaking at 140 rpm. The resin was washed three times with 2 mL wash buffer (60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) for 5 mins with shaking at 90 rpm for each washing step. The

bound protein was eluted twice from the resin with 500 μ L of His-Elute buffer (100mM L-histidine, 0.5M NaCl, 20 mM Tris-HCl, pH 7.9) for 10 mins with shaking at 70 rpm. The fractions were analysed using SDS-PAGE (Appendix A5-1). Fractions containing the protein were concentrated using 3 kDa molecular weight cut-off (MWCO) Ultra-0.5 spin filters (Amicon). The protein solution was dialysed against the storage buffer containing 50 mM Tris-HCl, pH 7.5, 5 mM CaCl₂, 50% glycerol and subsequently stored at -20 °C.

2.2.3 GOLDEN GATE ASSEMBLY

The golden Gate assembly mixture (10 μ L) contained 10 U of either BsaI-HF (NEB)(HCKan/ pMKU assembly) or Esp3I (NEB)(POT/ pGAU assembly) with 0.5 U of T4 DNA ligase (Thermo Scientific) and an equimolar mixture of part vectors (HCKan_P, HCKan_O, and HCKan_T) and acceptor plasmid in 1X T4 DNA ligase buffer (Thermo Scientific) and 10X BSA (NEB). The reactions were carried out at 37 °C for 10 mins, followed by 25 cycles of 37 °C for 5 mins, 25 °C for 10 mins, before finishing with 50 °C for 20 mins and 80 °C for 20 mins.

The reaction mixture was directly used for transformation in NEB10b chemically competent cells. The colonies were selected on either LB kanamycin (HCKan or pMKU) or LB ampicillin (POT/ pGAU) plates. The colonies were then screened for the presence of the gene by colony PCR and further analysed by restriction enzyme digestion or sequencing.

2.2.4 CHARACTERIZATION OF PROMOTER STRENGTHS

2.2.4.1 FLUORESCENT PROTEIN EXPRESSION

The sequence-verified pGY026 vectors containing the promoter inserts were transformed into chemically competent *S.cerevisiae* BY4741 cells (Gietz & Schiestl, 2007). Transformed cells were selected on SC-URA-LEU plates with 2% glucose.

Three separate colonies per promoter construct were picked and patched onto a fresh SC-URA-LEU plate. The patched colonies were then grown in 10 ml SC-URA with 2% glucose at 30 °C overnight; each construct was sub-cultured in 10 mL SC-URA 2% glucose at a 1:1000 dilution and grown at 30 °C. After 24 and 48 hrs, an aliquot of 100 μ L from each culture was removed and its fluorescence signal was measured at 530 nm for yeCitrine and 610 nm for mCherry using the BioTek Synergy H1 microplate reader with a gain value of 50. The fluorescence signal of yeCitrine was first normalized to the fluorescence signal of mCherry to account for plasmid copy number and cell density. The value of promoter strength was normalized against the value of the weakest promoter, PDC1.

2.2.4.2 PRODUCTION OF DMAT

The ensemble of assembled pMKU vectors containing *dmaW* driven by the various promoters were transformed in chemically competent *S.cerevisiae* BY4741 cells as described earlier and selected on SC-URA plates supplemented with 2% glucose.

Three independent colonies for each promoter construct were picked and patched onto a fresh SC-URA plate. Each clone was then grown in 10 mL SC-URA with 2% glucose at 30 °C for 48 hrs. A 1 mL aliquot of spent media

was filtered through a 0.2 μm PTFE syringe filter and analysed for the presence of DMAT using LC-MS. An estimate of the relative amounts of DMAT produced in each construct was obtained from the extracted ion chromatograms at m/z 257.1598 ± 20 ppm. These values were then normalized against the PDC1 driven construct.

2.2.5 *IN VITRO* BIOSYNTHESIS OF DMAT AND 4DMA

The *in vitro* biosynthesis of DMAT was prepared in a 60 μL reaction volume containing 50 mM Tris-HCl, pH 7.5, 5 mM CaCl_2 , 1 mM *L*-tryptophan, 1 mM DMAPP, and 10 μL purified *dmaW*. The reaction was incubated at 30 $^\circ\text{C}$ for 18 hrs. The reaction was stopped by filtering off the enzyme using a 3 kDa MWCO Ultra-0.5 spin filters (Amicon). The sample was either stored at -20 $^\circ\text{C}$ or immediately analysed using LC-MS. The *in vitro* biosynthesis of 4DMA was prepared similarly but with the addition of 1 mM SAM and 10 μL purified *easF*.

2.2.6 GENOMIC INTEGRATION AND *IN VIVO* PRODUCTION OF 4DMA

The linear DNA fragments to be integrated into the yeast genome were released from their pGAU vectors by restriction digest with BsaI-HF and subsequently gel purified. The linear DNA inserts were then transformed into chemically competent *S.cerevisiae* BY4741 cells. Transformed cells were selected on SC-URA plates with 2% glucose, and screened by performing a quick genomic DNA extraction followed by PCR (L  oke, Kristjuhan, & Kristjuhan, 2011). The verified clones were then grown in 10 mL SC-URA supplemented with 2% glucose at 30 $^\circ\text{C}$, for 48 hrs. A 1 mL aliquot of spent media was filtered through a 0.2 μm PTFE syringe filter analysed for the presence of 4DMA using LC-MS.

2.2.7 LC-MS ANALYSIS OF SAMPLES

All LC-MS analysis were performed using the Agilent 1290 Infinity LC System with an Agilent Zorbax Eclipse Plus C18 column, 2.1 x 100 mm dimensions, 1.8 μ m particle size. The samples were loaded into the column and eluted over a stepped gradient of Buffer A (water with 0.1% formic acid) and Buffer B (acetonitrile with 0.1% formic acid) at a constant flow rate of 0.5 mL min⁻¹. The column was washed with the following buffer programme – 95% A/ 5% B for 0.6 mins; 65% A/ 35% B for 2.6 mins; 1% A/ 99% B for 4.6 mins; 100% B for 2 mins. The column was re-equilibrated at 95% A/ 5% B for 1 min after every analysis.

The mass analysis was done using the Agilent Technologies 6550 iFunnel QTOF. The acquisition parameters were set to acquire MS data from an electrospray ionization source on positive mode for a mass range of m/z = 50 – 1000 m/z , at a scan rate of 3 spectra s⁻¹. The instrument parameters were set at gas temperature: 200 °C, gas flow: 12 L min⁻¹, nebulizer: 50 psig, sheath gas temperature: 350 °C, sheath gas flow: 11 L min⁻¹. The scan source parameters were set at VCap: 4500 V, Nozzle voltage: 0 V, Fragmentor: 110 V, Skimmer: 65 V, Octopole RF peak: 750 M.

2.3 RESULTS AND DISCUSSION

2.3.1 ADAPTING YEASTFAB ASSEMBLY

In the envisioned workflow for this project, the pathway segments were quickly designed, built, and screened first in *E. coli* before testing them for functionality in yeast. The YeastFab workflow was thus remodelled to accommodate this by the creation of new pathway acceptor vectors and a new suite of genome integration vectors (Figure 2-1A). Performing the molecular work using *E. coli* as the host would reduce the time taken for the assembly of pathway segments into testable plasmids from four days to one day. Additionally, the screening process would be greatly ameliorated because colony PCR can be performed easily and reliably in *E. coli*.

A series of pathway acceptor plasmids were created from the ensemble of vector elements available. A total of six new pathway acceptor plasmids were assembled, one for each combination of yeast origin of replication (2-micron/ CEN-ARS) and selection marker (URA3/ HIS3/ LEU2) (Figure 2-1B). This suite of vectors were designed to accommodate the 4-base overhang prefix of either POT1 or POT2 and the terminal 4-base overhang from the odd numbered POT-s.

Further in line with this project's workflow, a series of genome integration vectors were designed as well, to accept the pathway segments once they have been tested. These integration vectors were designed with a URA3 selection marker flanked by the 500 bp URR sites, which are uniquely homologous to each other but not to anywhere else within the yeast genome (Richardson, Nunley, Yarrington, Boeke, & Bader, 2010). These sites were originally used in the YeastFab assembly as recombination sites at the HO locus

or onto their pathway acceptor vector. In this design, the URR sequences allowed for the removal of the URA3 selection marker by homologous recombination upon counter-selection with 5-fluoroorotic acid (5FOA). This also enabled the recycling of the same selection marker for integration at alternative sites (Figure 2-1C).

The different regions on the yeast genome are known to affect the propensity for the acceptance of inserts, as well as the overall transcription and expression levels. The genome integration loci chosen and used for the creation of this suite of integration vectors were selected based on the loci characterization data from earlier studies (Apel et al., 2016; Bai Flagfeldt et al., 2009).

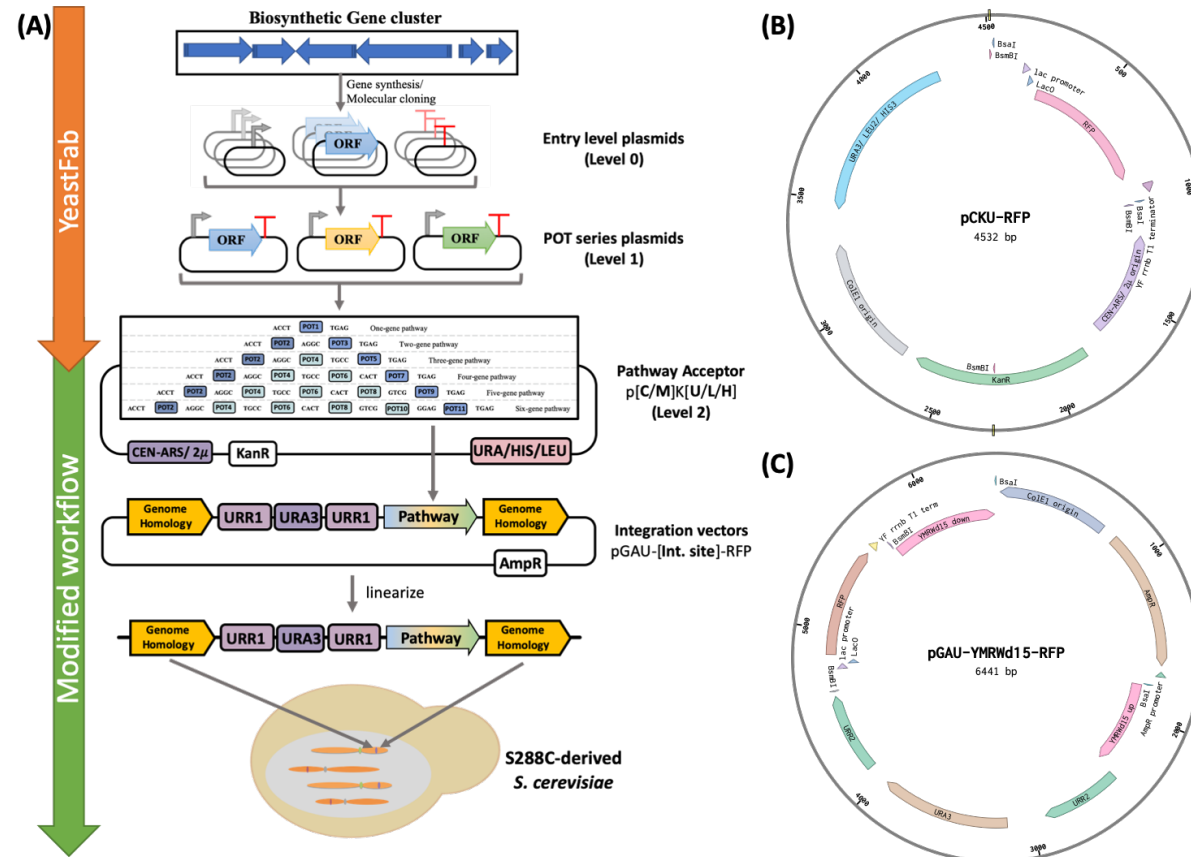


Figure 2-1: (A) Illustration of the modifications made to the YeastFab assembly workflow to tailor the needs of this project. (B) Representative plasmid map of the pathway acceptor vectors created. (C) Representative plasmid map of the genome integration vectors created. Full image of the vector maps of (B) and (C) are available in Appendix A2.

2.3.2 CHARACTERIZATION OF PROMOTER STRENGTHS

2.3.2.1 FLUORESCENT PROTEIN EXPRESSION

After adapting the YeastFab system, the next foundational requirement is a library of biological parts which the pathway will be built from. Biological parts such as promoters, open reading frames (ORFs), and terminators are required to be standardized and “domesticated” into the system. Different endogenous yeast promoters that have been previously characterized in other systems were tested for functionality in the context of the YeastFab expression cassette.

Eighteen promoters were cloned upstream of a yeCitrine reporter on the promoter characterisation plasmid (pGY026) used in the YeastFab study (Figure 2-2A). The characterization plasmid contained an mCherry expression cassette which acted as a more reliable measure for copy number and cell density. The expression of yeCitrine by the promoters was used to predict the expression of the target enzymes in the ergot alkaloid pathway.

The eighteen promoters tested were found to fall roughly into three groups of expression levels after 24 hrs: strong, medium, and weak (Figure 2-2B, C, D). The measured relative promoter strengths generally corroborated with existing published data at the level classifying them into strong, medium, or weak promoters (Partow, Siewers, Bjørn, Nielsen, & Maury, 2010). The specific orders of relative promoter strengths within each class however did not agree completely with the existing data. These were to be expected when considering the variations in experimental conditions used, the exact length and sequence used to define a “promoter” sequence, as well as the overall

architecture of the test construct and the different linker lengths and sequences used by the different studies.

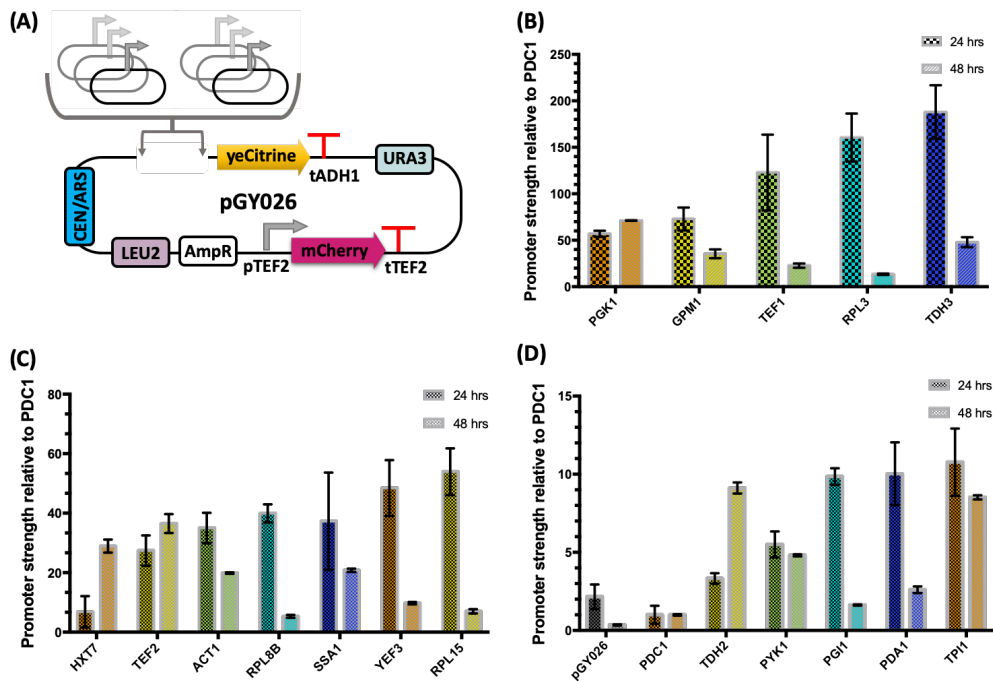


Figure 2-2: Characterization of yeast promoter strengths based on expression of fluorescent proteins. (A) Illustration of the characterization plasmid used for this experiment. (B) Strong promoters. (C) Medium promoters. (C) Weak promoters. Error bars represent standard deviation of calculated values from three biological replicates.

Interestingly, this double cassette system appeared to highlight some competition between the two cassettes. The 48 hr measurements for most promoters appear to drop whilst the reading for the TEF2 promoter on both yeCitrine and mCherry increased at the later phase of cell growth. It is useful to note such effects for pathway optimization; to capitalize on the effects of diauxic shifts to channel expression of late stage pathway enzymes. This experiment nevertheless suggested that the established methodology should adequately reflect the functionality of the procured promoter sequences in the context of the 4-base linker between promoter and ORF.

2.3.2.2 PRODUCTION OF DMAT

With a functional set of promoter sequences, the following step was the validation of the assembled expression cassettes. Most fortunately, the first enzyme of the ergot pathway, *dmaW*, has been extensively studied and shown to be very well-disposed to heterologous expression. It was hence elegantly situated to allow the validation of the relationship between expression strength and enzymatic turnover capacity, the adapted YeastFab pipeline, and the demonstration of the first reaction of the ergot pathway all at once.

A matrix of *dmaW* expression cassettes was constructed with 14 of the previous 18 promoters and was used to test the pathway acceptors (Figure 2-3A). In order to reduce the complexity and combinations to be tested throughout this project, each of the pathway enzymes were arbitrarily assigned a fixed terminator. The 14 *dmaW* expression cassettes were assembled with a GAL1 driven yEGFP on an odd numbered POT vector onto the pMKU-RFP pathway acceptor and screened for the production of DMAT. From the LC-MS analysis of the spent media from the tested constructs, a compound with an $[M+H]^+$ of $m/z = 273$ was consistently detected to elute at around 199-201 s (Figure 2-3C), suggesting the successful production of DMAT in all constructs (Table 2-6).

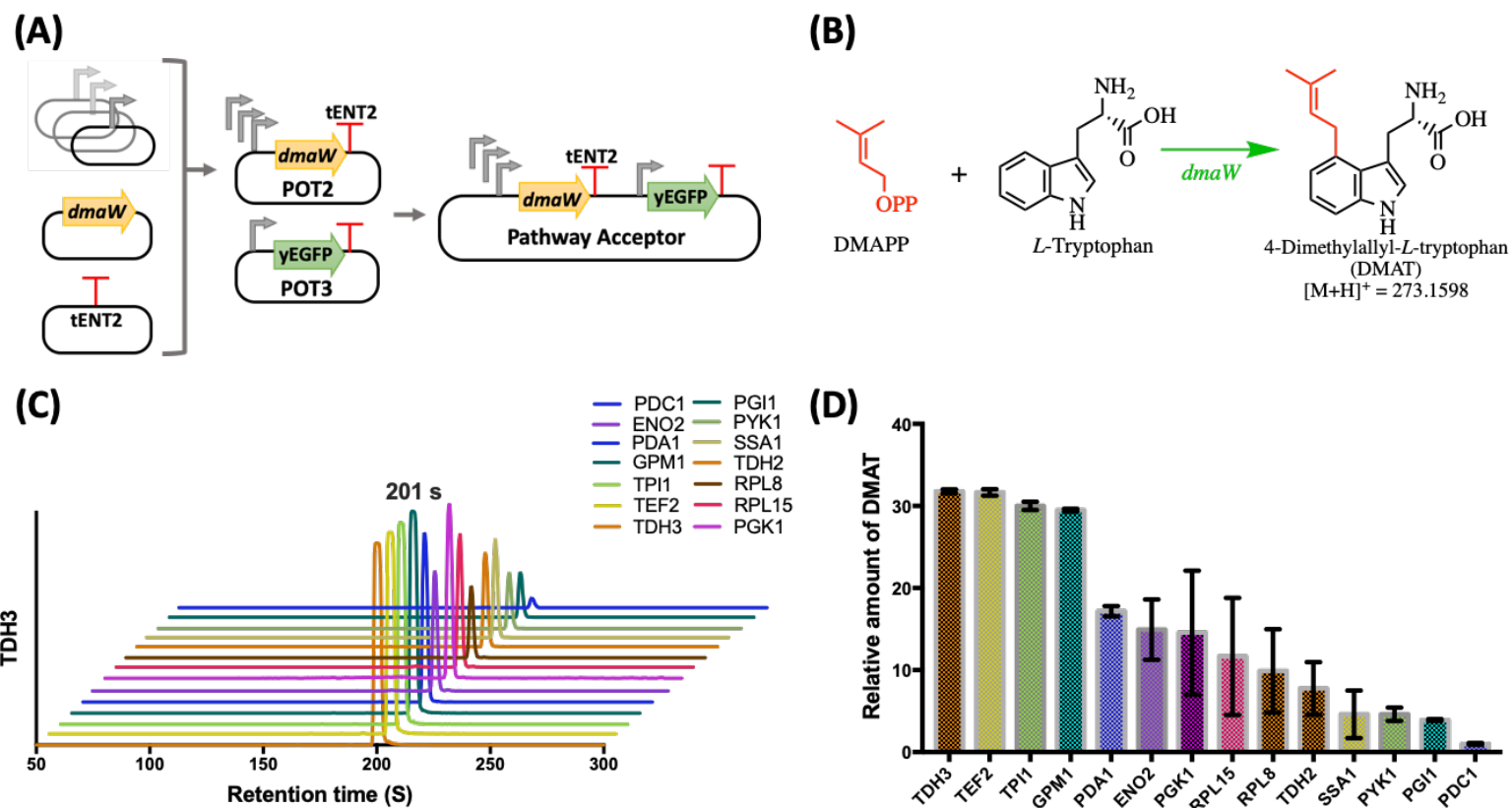


Figure 2-3: Characterization of yeast promoter strengths by production of DMAT. (A) Illustration of the assembly process to generate pMKU-dmaW plasmids. (B) Reaction catalysed by dmaW producing DMAT. (C) Overlaid extracted ion chromatograms for 273 m/z of the analysed samples showing peaks of varying sizes corresponding to the amount of DMAT produced. (D) Comparison of the relative amounts of DMAT produced, estimated from the AUC of the extracted ion chromatograms. Error bars represent the standard deviation calculated from three biological replicates

Table 2-6: Exact mass and retention time of DMAT detected

Promoter	Calculated m/z	Observed m/z	Mass error δ (ppm)	Retention time (s)
1	273.1598	273.1602	1.46	199.2
2	273.1598	273.1609	4.03	199.8
3	273.1598	273.1604	2.20	199.2
4	273.1598	273.1600	0.73	201.0
5	273.1598	273.1600	0.73	199.8
6	273.1598	273.1626*	10.25	201.0
7	273.1598	273.1619*	7.69	200.4
8	273.1598	273.1618*	7.32	200.4
9	273.1598	273.1603	1.83	199.8
10	273.1598	273.1599	0.37	200.4
11	273.1598	273.1602	1.46	199.2
12	273.1598	273.1605	2.56	198.6
13	273.1598	273.1624*	9.52	200.4
14	273.1598	273.1599	0.37	199.9

*: denotes saturating ions.

The estimation of the relative amounts of DMAT, produced by each construct largely corroborated with the earlier expression strength data measured from the expression of fluorescent proteins (Figure 2-3D), with two exceptions. The constructs driven by the TPI1 and PDA1 promoters performed very differently when compared with the expression of fluorescent proteins. The two promoters were found to be weak in the expression of fluorescent proteins, but were found to be among the strongest promoters when assessed for DMAT production. This observation suggested that when these promoters with shared transcriptional regulatory elements are placed in proximity to each other, they exert some form of cis-regulatory effects on one another that is likely due to some form of competitive binding to these regulatory elements or transcription factors. A review of the existing literature reporting identified transcriptional regulatory elements of the genes TPI1, PDA1, and TEF2 revealed that several

of these regulatory elements are common to these genes (Appendix A1-1) (Z. Hu, Killion, & Iyer, 2007; Lickwar, Mueller, Hanlon, McNally, & Lieb, 2012; Venters et al., 2011). The removal of these promoters in proximity to one another in the DMAT production experiment, releases this competitive suppression.

DMAT is not commercially available; thus, DMAPP and tryptophan were incubated with purified *dmaW* to catalyse the biosynthesis of DMAT. This was used for the verification and characterization of DMAT production. A comparison of the products from the *in vitro* and *in vivo* biosynthesis showed a similar compound being eluted at around 200-201 s with an $[M+H]^+$ that is within 20 ppm of the theoretical DMAT value of 273.1598 (Figure 2-4). Furthermore, the fragmentation patterns observed in both experiments show the similar production of two simple fragment ions from the cleavage of the amine group and the C10-C11 bond on the dimethylallyl moiety. Analysis of the fragments again showed that they are within 20 ppm of their respective theoretical values (Table 2-7).

Overall, the observed DMAT production levels from the *dmaW* promoter matrix would serve as a better indicator of the effects the different promoters would have on enzyme expression levels and therefore substrate turnover. This data served as a guide towards the selection of strong promoters as the prototype strain was developed. The successful use of the new pathway acceptors also demonstrated their functionality and successful remodel of the YeastFab workflow.

Additionally, this experiment has also highlighted the issue of potential promoter crosstalk that could occur when promoters sharing similar regulators

are in close proximity to one another. Whilst this would be an important and interesting aspect of such a system, it is not the focus of this work and was not pursued further.

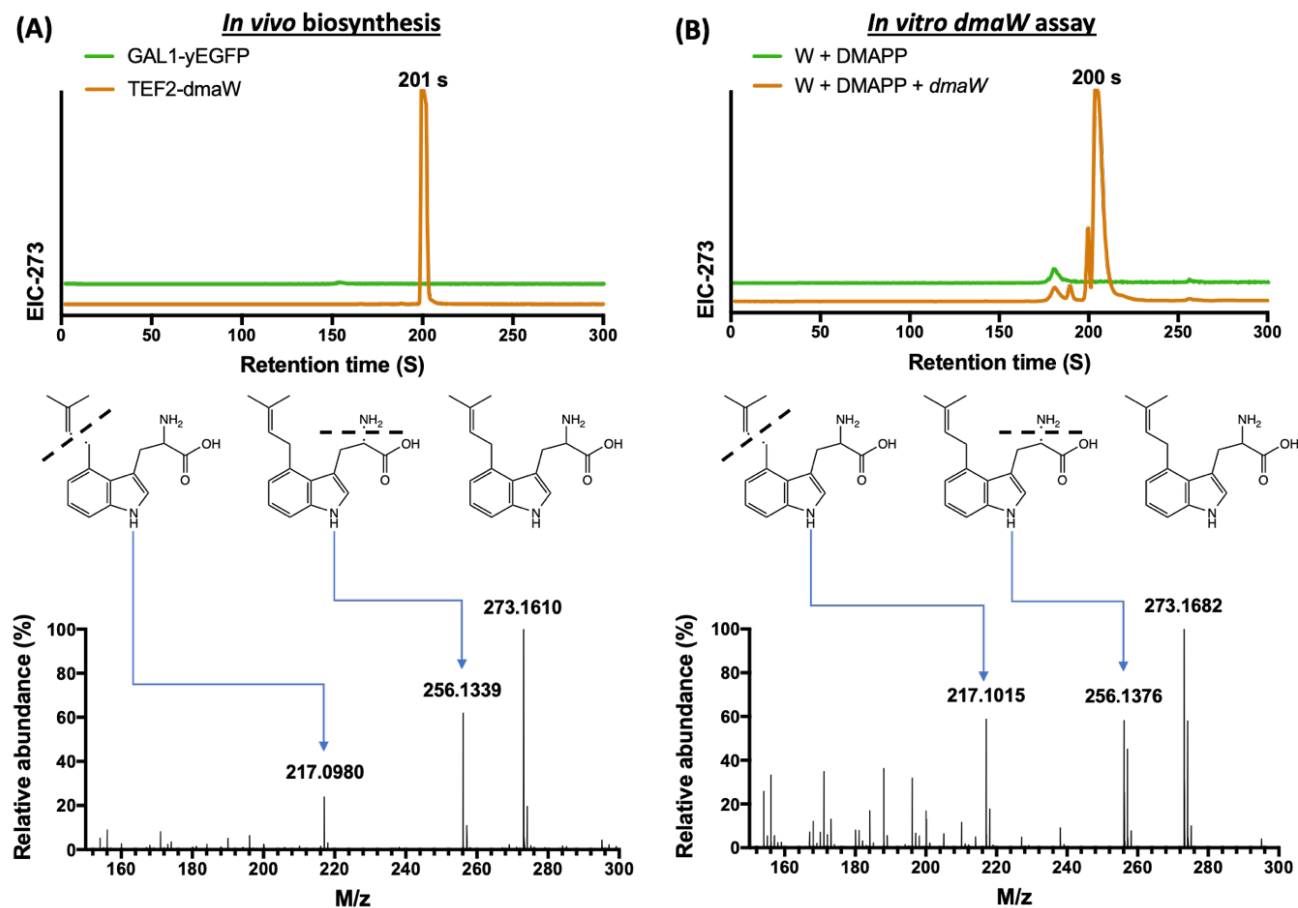


Figure 2-4: Comparison of retention time and mass spectra of DMAT produced *in vivo* (A) and *in vitro* (B). Mass spectra of both compounds show similar fragment ions. Structures of the simplest hypothesized fragment ions shown for comparison.

Table 2-7: Exact mass of the simplest fragment ions from DMAT

Fragment	Calculated m/z	Observed m/z		Mass Error δ (ppm)	
		TEF2- dmaW	<i>In vitro</i> assay	TEF2- dmaW	<i>In vitro</i> assay
- NH ₂	256.1338	256.1339	256.1376*	0.58	15.03
- C ₄ H ₇	217.0977	217.0980	217.1015*	1.38	17.50

*: denotes saturating ions

2.3.3 TESTING GENOME INTEGRATION SITES WITH 4DMA PRODUCTION

Following the successful demonstration of the production of DMAT and the functional validation of the pathway acceptor plasmids, this next experiment sought to test the genome integration plasmids created, as well as expand the pathway to the second reaction for the production of 4DMA (Figure 2-5B). Following the modified YeastFab workflow outlined earlier (Figure 2-1A), *dmaW* and *easF* were assembled into POTs driven by the TEF2 and GPM1 promoters respectively. These POTs were then condensed onto a pMKU plasmid and moved onto the genome integration vector, pGAU-YMRW, for integration at the YMRW Δ 15 loci to create the strain illustrated (Figure 2-5A). This strain was found to produce a compound with an $[M+H]^+$ of 273 m/z that eluted at around 201 s and a second compound with an $[M+H]^+$ of 287 m/z that eluted at around 206 s (Figure 2-5C), that alluded to the production of DMAT and 4DMA.

As with DMAT, there is no commercially available standard for 4DMA. The earlier *in vitro* DMAT biosynthesis catalysed by *dmaW* was expanded to include SAM and purified *easF*. The reaction mixture contained DMAPP, tryptophan, SAM, *dmaW*, and *easF* produced a compound with an $[M+H]^+$ of 287 m/z that eluted at around 207 s (Figure 2-5D). A comparison of the mass

spectra of both *in vivo* and *in vitro* products showed the accumulation of a compound with an $[M+H]^+$ within 20 ppm of the theoretical 4DMA value of 287.1598 m/z (Table 2-8). Analysis of the fragmentation patterns showed the production of similar fragment ions for the cleavage of the methylamino group and dimethylallyl moiety, all within 20 ppm of the theoretical values (Table 2-9).

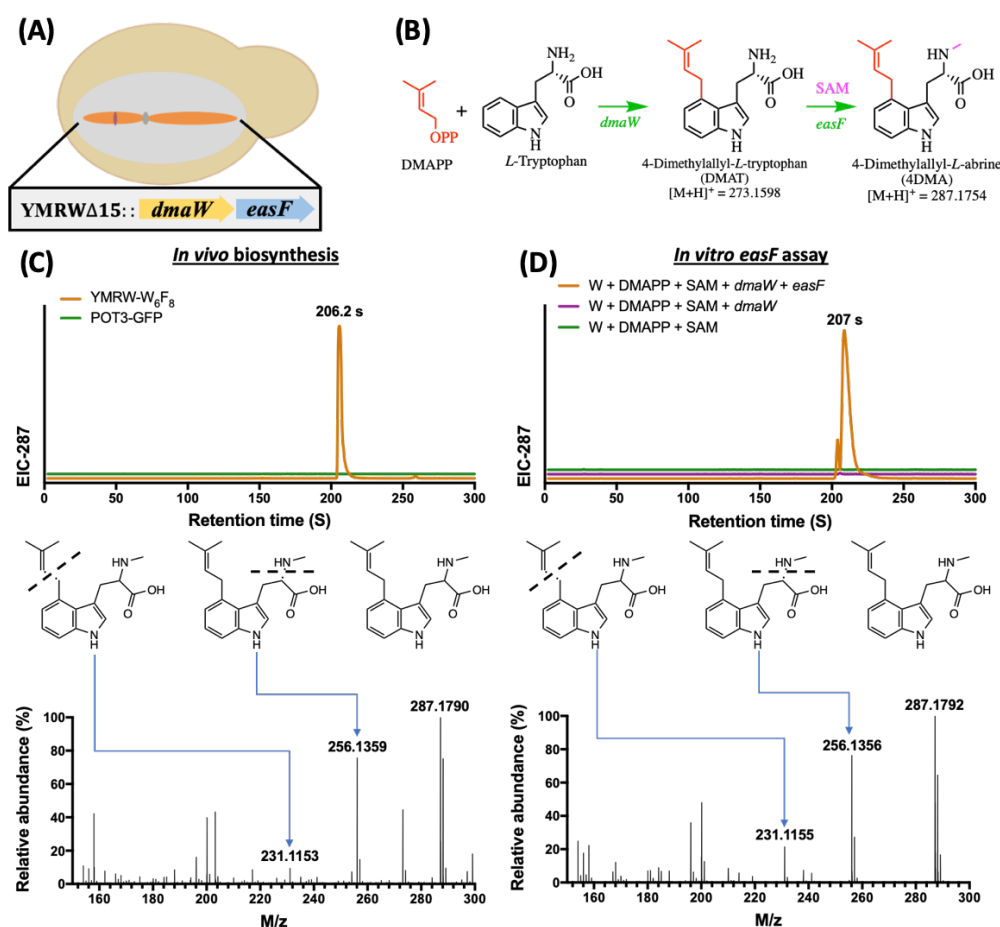


Figure 2-5: Integration of *dmaW* and *easF* into the yeast genome for a 4DMA producing strain. (A) Illustration of the strain created. (B) Reactions catalysed by the strain containing *dmaW* and *easF* converting tryptophan to DMAT and 4DMA. Comparison of the extracted ion chromatograms and mass spectra for *in vivo* (C) and *in vitro* (D) biosynthesized 4DMA.

Lending further support to the determination that these compounds observed are DMAT and 4DMA, is the nature of the major fragment ions were determined. The first fragment of 4DMA, obtained from the cleavage of the methylamino group produces the same fragment expected from the cleavage of

the amino group from DMAT. The second fragment of 4DMA, obtained from the cleavage of the dimethylallyl moiety, is exactly 14 m/z larger than the fragment from DMAT of the same cleavage site. This accounted for the appended methyl group in the reaction of DMAT to 4DMA from SAM.

Table 2-8: Exact mass and retention time of DMAT and 4DMA detected

Compound	Calculated M/z	Observed M/z		Mass Error δ (ppm)		Retention time (S)	
		YMRW-W ₆ F ₈	<i>In vitro</i> assay	YMRW-W ₆ F ₈	<i>In vitro</i> assay	YMRW-W ₆ F ₈	<i>In vitro</i> assay
DMAT	273.1598	273.1620	273.1648*	8.05	18.30	201	203
4DMA	287.1754	287.1790*	287.1792*	12.54	13.23	206	207

*: denotes saturating ions

Table 2-9: Exact mass of the simplest fragment ions from 4DMA

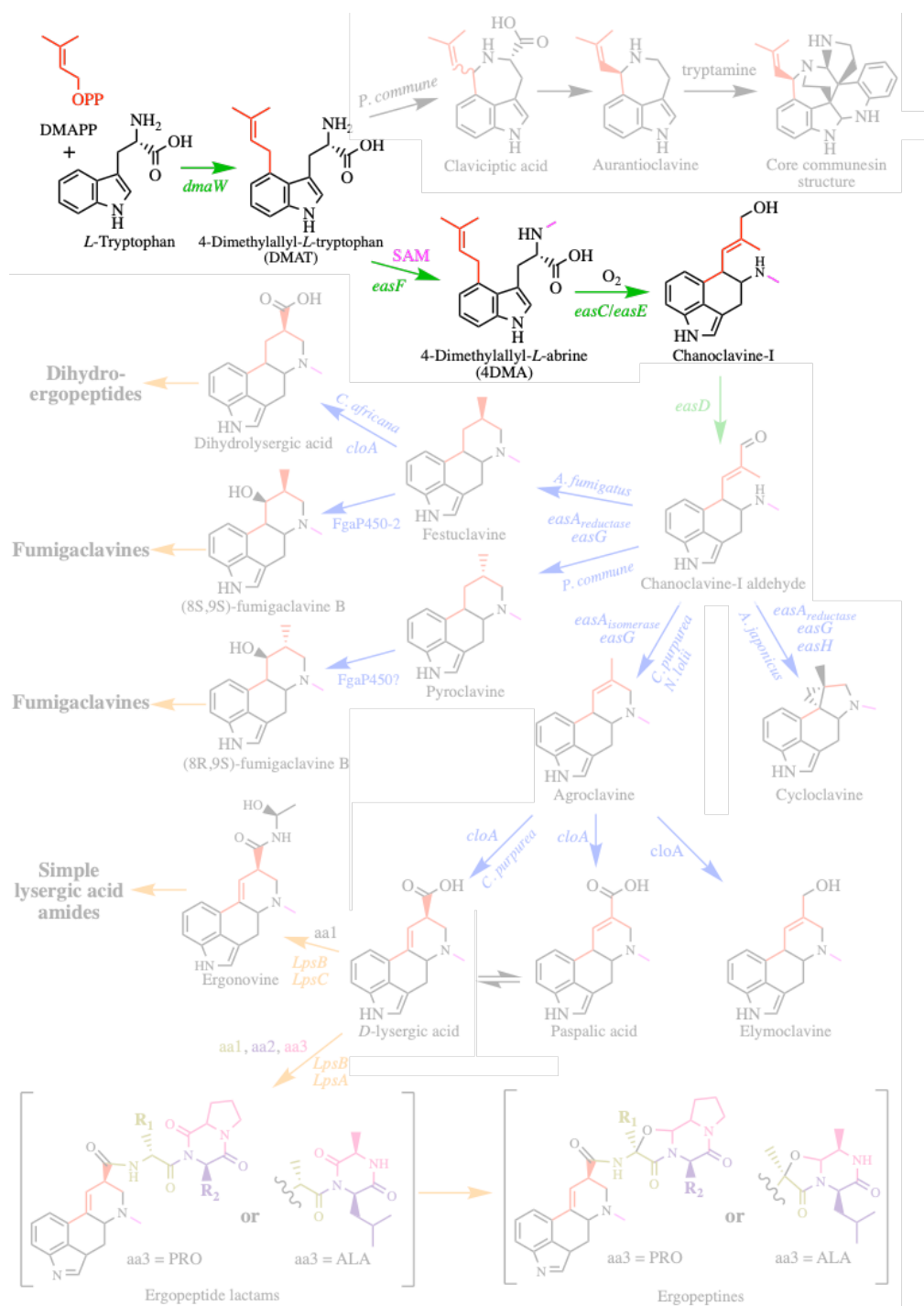
Fragment	Calculated m/z	Observed m/z		Mass Error δ (ppm)	
		YMRW-W ₆ F ₈	<i>In vitro</i> assay	YMRW-W ₆ F ₈	<i>In vitro</i> assay
- NHCH ₃	256.1338	256.1359	256.1356*	8.20	7.03
- C ₄ H ₇	231.1134	231.1153*	231.1155*	8.22	9.09

*: denotes saturating ions

While it might appear concerning that the mass error for several of the detected compounds are relatively large (>10 ppm) for high-resolution mass spectrometry, a much larger tolerance for error is acceptable in the accurate determination of small molecules (≤ 300 Da) because the permutations possible for a compound consisting of carbon, nitrogen, hydrogen, and oxygen are still relatively limited (Stoev, Xuan, Peycheva, & Scigelova, 2012). Furthermore, the samples with the larger errors observed are an artefact of peak broadening due to the saturation of the QTOF detector. This issue is easily resolved in subsequent iterations of LC-MS method development.

This experiment has suitably demonstrated the successful use of the newly expanded YeastFab system. This section showed the creation of a yeast strain harbouring the first two enzymes of the ergot alkaloid pathway. This served as the foundation on which the next few chapters were built upon.

CHAPTER 3: BUILDING THE EARLY ERGOT PATHWAY TO CHANOCLOAVINE-I



3.1 INTRODUCTION

The inclusion of two more enzymes is required to complete the early ergot pathway which is common to all ergot alkaloid-producing organisms. These two enzymes, *easC* and *easE*, have been previously identified to be responsible for the double oxidation event that circularizes the ergoline C ring to produce chanoclavine-I (Figure 3-1A) (Lorenz et al., 2010; Nielsen et al., 2014; K. Ryan et al., 2013). An early study on this reaction, using isotopic labelling, showed that the oxygen atom on the hydroxyl group in chanoclavine-I is from molecular oxygen (Kozikowski et al., 1993). This information, coupled with sequence similarity of *easC* to iron-heme catalases (Figure 3-1B) and *easE* to FAD-dependent berberine-bridge type oxidase enzymes (Figure 3-1C), has led to several hypotheses about the role each enzyme performs (Kozikowski et al., 1993; Kozikowski, Wu, Shibuya, & Floss, 1988).

Efforts to produce direct experimental validation for these hypotheses however, have not met with much success. This was largely because of the difficulty met by various groups in their attempts at the heterologous expression and purification of *easE* (Nielsen et al., 2014). Most studies have been in fungi that are very closely related to the source of the *easE* sequence used. The most successful “truly” heterologous expression of *easE* was performed by Sarah O’Connor’s group, using an ortholog of *easE* identified from *Aspergillus japonicus* (Jakubczyk et al., 2015; Nielsen et al., 2014). Even so, they were unable to purify the enzyme to adequately validate any of the proposed reaction mechanisms of the ergoline C ring formation.

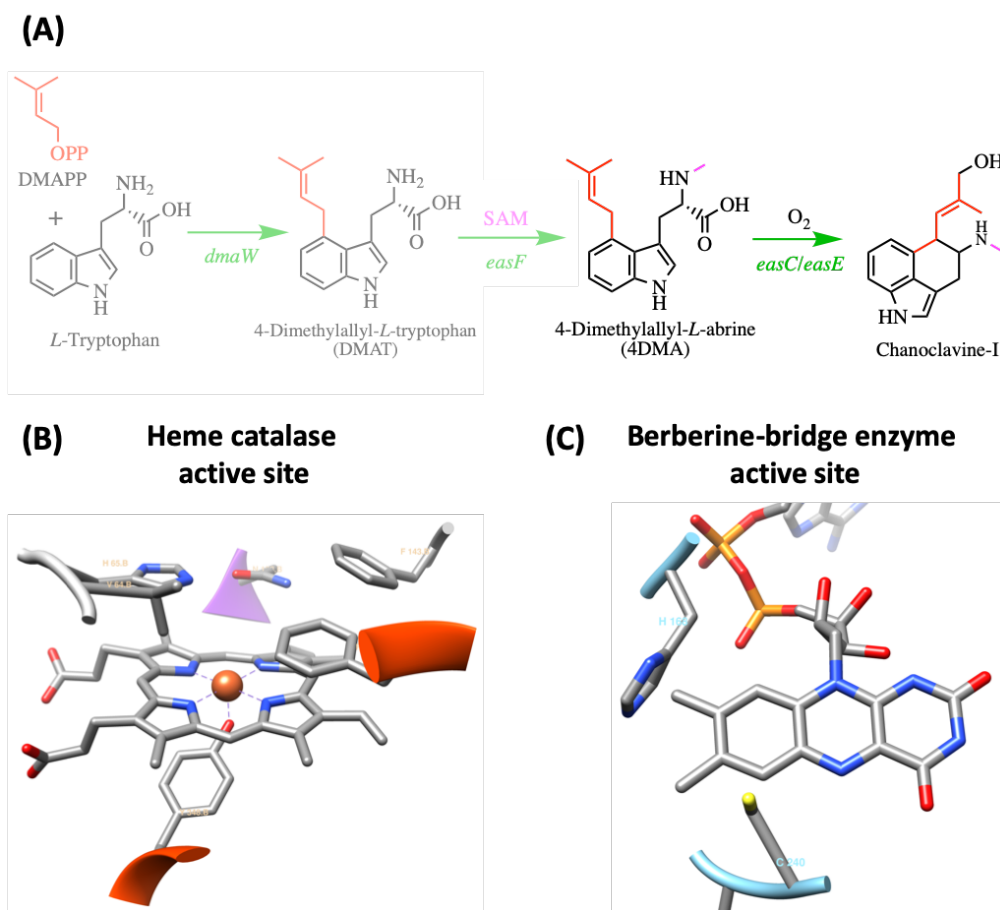


Figure 3-1: (A) The early ergot alkaloid pathway leading to chanoclavine-I formation. The focus of this chapter is the extension of the pathway from 4DMA to chanoclavine-I. (B) Heme catalase active site from the peroxisomal catalase of *Hansenula polymorpha*, PDB ID: 2XQ1. (C) Predicted model of the active site in *easE_Aj*. This model was created using the Phyre2 web portal (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015).

Bioinformatics analysis of the sequence of *easE* suggests its relation to bicovalently-bound FAD-dependent oxidases that require a certain degree of post-translational modification to be functionally expressed. This might account for the difficulty expressing it in prokaryotic hosts. Another common and plausible explanation is the erroneous prediction of the amino acid sequence. These errors are usually due to the difficulty in accurately identifying the introns and exons from DNA sequences and are not just limited to the ergot gene clusters, but for most eukaryotic gene structures (Brent & Guigo, 2004).

The approach taken for this part of the study was to find a diverse pool of *easE* orthologs in the hope that at least one would be successfully and functionally expressed. A Sequence Similarity Network (SSN) was generated using the sequence of *easE* from *A. japonicus* as the query. A set of orthologs was then selected from the predicted isofunctional cluster of *easE* and screened for expression and function in yeast. The best performing combination was incorporated into the prototype strain, completing the early pathway to the next stably accumulating intermediate, chanoclavine-I.

3.2 MATERIALS AND METHODS

3.2.1 CREATING AN ENTRY-LEVEL VECTOR FOR APPENDING SUMO-TAGS

The construction of an ORF entry-level vector that appends a SUMO-tag on the N-terminal site of the insert ORF was performed in two stages. The first stage involved the replacement of the MCS region within a pET28M-SUMO vector with the Golden Gate assembly insertion region from HCKan_O. This was done by linearizing pET28M-SUMO around the MCS region by PCR, using the primers to append the BsaI recognition sites and the 4-base prefix and suffix at its ends. The Golden Gate insert region from HCKan_O was then released by Esp3i digestion, gel purified, and ligated onto the linearized pET28M-SUMO fragment (Figure 3-2A). This produced a version of pET28M-SUMO compatible with Golden Gate cloning that was named p28SGG-RFP.

The lacI coding region and rop protein-coding region on p28SGG-RFP was removed and the Esp3i recognition sites were added in the second stage to create the simplified pSGG-RFP vector. This was done by the amplification of p28SGG-RFP into three fragments, that excluded these extraneous features (Figure 3-2B). The first fragment consisted of the T7 expression elements with the included Esp3i recognition site upstream of the 4-base prefix. The second was made up of the Golden Gate assembly insert region, with Esp3i recognition sites incorporated both upstream of the 4-base prefix and downstream of the 4-base suffix. The last fragment comprised of the *E.coli* origin of replication and Kan^R selection marker. These three fragments were then joined by Gibson Assembly as described earlier, to form the SUMO-tag entry-level vector, pSGG-RFP (Figure 3-2C).

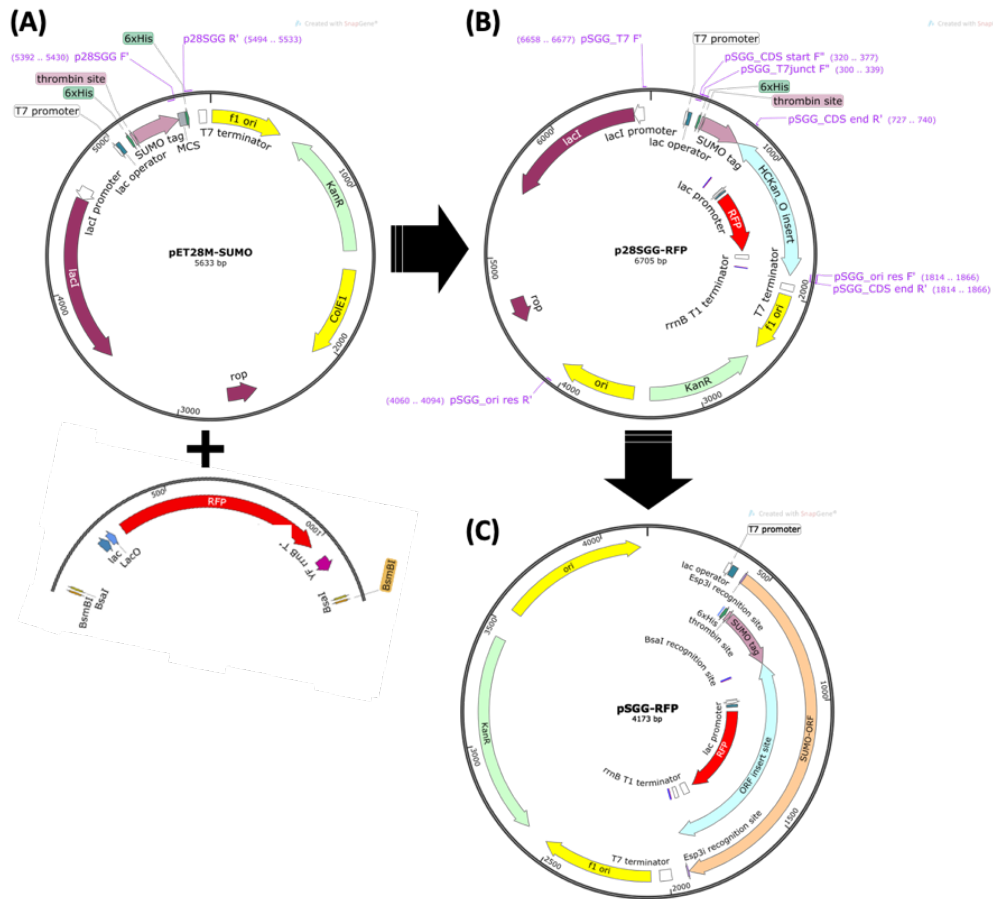


Figure 3-2: Illustration of method devised to create a SUMO-tag entry-level vector compatible with YeastFab assembly. (A) Commercially available SUMO-tag vector pET28M-SUMO was used as the starting material. (B) Modified pET28M-SUMO with the MCS site replaced with the YeastFab insert region. (C) Extraneous elements were removed to reduce vector size and complexity to create pSGG-RFP.

3.2.2 IDENTIFICATION AND SCREENING OF *easE* ORTHOLOGS

3.2.2.1 GENERATING AN SSN FROM THE EFI-EST WEBTOOL

To generate the SSN from the Enzyme Function Initiative – Enzyme Similarity Tool (EFI-EST) (Gerlt et al., 2015), the sequence of *easE* from *A. japonicus*, identified from the work published by the O'Connor group was used as the input (Option A). The initial BLAST parameters were set to pull a maximum of 9000 sequences with a minimum alignment E-value of 5 (Figure 3-3A). From the returned pool of sequences, an SSN was generated from the sequences with an alignment length ranging from 560 to 670 amino acids. The

edges were initially chosen with an E-value of at least 100, which corresponded to 40% sequence identity (Figure 3-3B). These initial parameters generated an output for the initial SSN representing the “known universe” of presumably FAD-dependent oxidases. The “*easE*” cluster was selected for by searching for nodes that were annotated, either manually or automatically, in the Uniprot database. This selection was expanded to include all the first neighbours and highlighted in orange (Figure 3-3C).

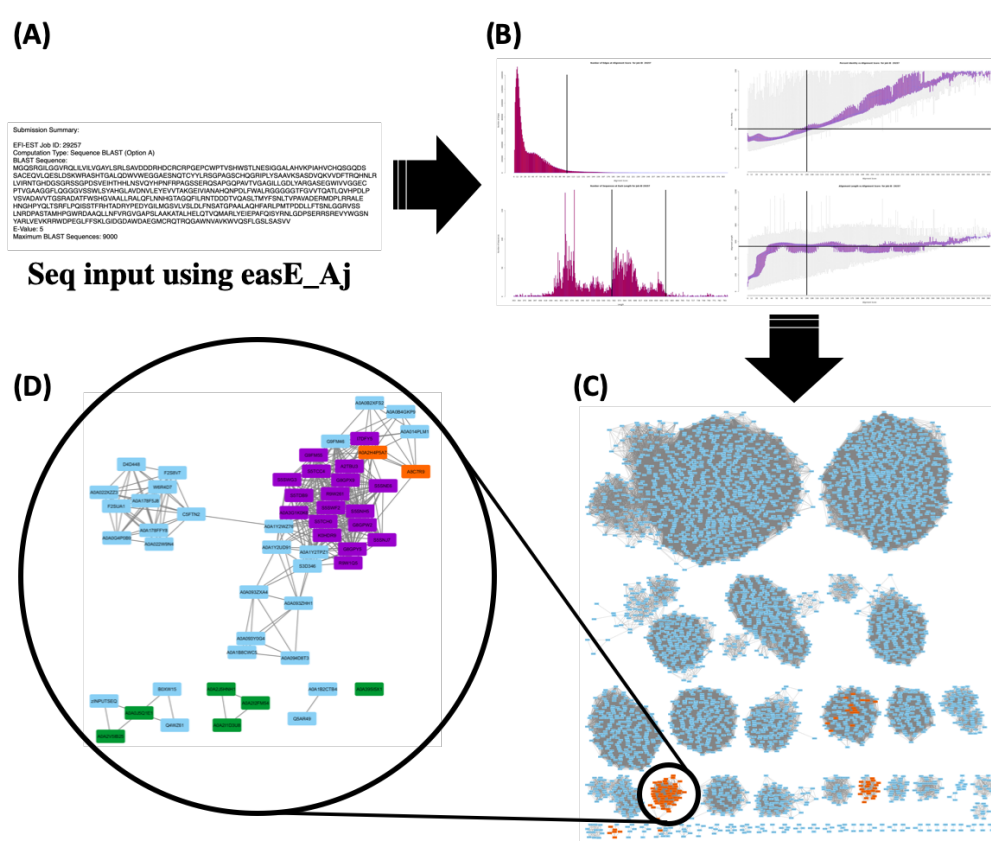


Figure 3-3: Workflow for the generation of SSN using the EFI-EST webtool. (A) Sequence input and initial BLAST parameters used. **(B)** Demographics of the sequences returned from the initial BLAST search. Selection of an alignment score corresponding to 40% sequence identity is used as a threshold for distinguishing isofunctional cluster. Selection of sequence length thresholds to limit the size of the network file. Full graphs are shown in appendix A4-1, 2, 3, 4. **(C)** Full SSN generated for returned “*easE*” hits, first neighbours of sequences annotated to be ergot alkaloid biosynthesis related are highlighted in orange. **(D)** Expansion of a single isofunctional cluster revealed the divergent lineages of *easE* enzymes.

From this pool of *easE*-related sequences, a hypothetical isofunctional cluster for *easE* was identified and further refined to display edges with an alignment score greater than 170. The cluster was found to then further diverge into sub-clusters that corresponded to the different lineages of *easE* enzymes from the different ergot alkaloid producing species (Figure 3-3D). Eight sequences on top of the sequence from *A. japonicus* were selected from the expanded isofunctional cluster of *easE* (Figure 3-3D). As none of the sequences have been annotated to have data demonstrating their existence at the protein level, two sequences from each of the four major sub-clusters were randomly chosen to represent each sub-cluster (Table 3-1). The selected sequences were codon-optimized for yeast expression, synthesized, and cloned onto pET15b and pSGG-RFP vectors.

Table 3-1: List of *easE* orthologs selected from the SSN generated

Abbreviated name	Organism	Uniprot ID
easE_Aj	<i>Aspergillus japonicus</i>	-
easE_Nl	<i>Neotyphodium lolii</i>	S5TDB9
easE_Ef	<i>Epichloe funkii</i>	R9W1Q5
easE_Pi	<i>Periglandular ipomoeae</i>	G9FM46
easE_Ee	<i>Epichloe elymi</i>	K0HDR9
easE_Ei	<i>Epichloe inebrians</i>	I7DFY5
easE_Cf	<i>Claviceps fusiformis</i>	A8C7R9
easE_Al	<i>Aspergillus lentulus</i>	A0A0S7DKV5
easE_Ec	<i>Epichloe coenophialia</i>	R9W261

3.2.2.2 BUILDING THE SCREENING STRAIN YMC17

In order to screen the identified *easE* orthologs, a functional variant of *easC* was required to build the envisioned screening strain. Earlier literature found that most orthologs of *easC* contained a classical PTS1 peroxisomal targeting sequence on the C-terminal end. The presence or absence of this

feature did not seem to negatively affect the production of chanoclavine-I (Goetz et al., 2011; Nielsen et al., 2014). The deletion of the sequence in the *A. japonicus* ortholog was even found to marginally improve chanoclavine-I production. The sequence from *Claviceps purpurea* was selected because it naturally does not contain the targeting sequence (Table 3-2). This sequence was synthesized into a pET15b vector by Genscript and cloned into the HCKan_O vector.

Table 3-2: Details of the sequence of *easC* synthesized

Gene	Source	Uniprot ID
<i>easC</i>	<i>Claviceps purpurea</i>	M1WA44

To test for the functional expression of *easC*, the pET15b construct was transformed into BL21(DE3) *E.coli* cells and selected on LB ampicillin plates. Cells were grown in 2YT to an OD₆₀₀ of 0.7 at 37 °C, before induction with IPTG (2 mM) at 20 °C for 20 hrs. They were subsequently harvested by centrifugation at 4 °C, 5000 rpm for 10 mins. Harvested cells were then resuspended in binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 7.9) at 4 °C and lysed *via* sonication. The cell lysate was then clarified by centrifugation at 15000 rpm for 20 mins at 4 °C. The protein was then purified from the cell supernatant in the same manner as previously described for *dmaW* and *easF*. Fractions were analysed on SDS-PAGE (Appendix A5-2) to detect the presence of *easC*. Fractions containing the protein were concentrated and dialysed in the same manner as previously described for *dmaW* and *easF*.

The catalase activity of *easC* was tested in two ways: these demonstrated the functional expression of *easC* in isolation of *easE*. A rough preliminary test

was first performed by adding 1 μ L of purified *easC* onto the edge of a 5 μ L drop of 1% H₂O₂ solution; this solution was observed for the catalytic release of O₂. After the formation of bubbles was observed, a simple assay to quantitatively test the activity of *easC* on ECL substrate was performed. Each reactions contained 60 μ L of pre-mixed ECL substrate and varying amounts of purified *easC* to a final volume of 100 μ L with 50 mM Tris-HCl, pH 7.5, 5 mM CaCl₂. The samples were incubated for 1 min at room temperature and measured for luminescence at 300 to 500 nm, using the BioTek Synergy H1 microplate reader.

The construction of the *easE* screening strain, YMC17, was done by first assembling an *easC* transcriptional unit with a GAL10 promoter and YIP5 terminator onto a POT5 acceptor vector, using the Golden Gate assembly protocol described in chapter 1. The created POT plasmid was put through the modified YeastFab pipeline with POTs containing *dmaW* and *easF*, as described in chapter 2 (Figure 2-1A), and integrated into a *S.cerevisiae* BY4741 host as previously described.

The successfully integrated strain was then cured of the URA3 marker by growing the strain in 5 mL YPD at 30 °C, followed by counter-selection on SC plates with 1 mg/L 5-fluoroorotic acid (5FOA) and 2% glucose at 30 °C. The colonies that formed were put through a second round of selection on a fresh SC plate with 5-FOA and grown overnight at 30 °C. The emergent colonies were then checked for the loss of the URA3 marker by patching onto a YPD plate and a SC-URA plate, as well as screened by PCR for the presence of the insert genes together with the URA3 marker.

3.2.2.3 SCREENING SELECTED *easE* ORTHOLOGS FOR ACTIVITY

The suite of selected *easE* orthologs were assembled into expression cassettes driven by a GAL1 promoter and a RPS20 terminator onto a POT3 acceptor vector. This ensemble were transformed into YMC17 chemically competent cells, prepared from the cured strains following the same protocol described previously. Successful transformants were selected for on SC-URA plates with 2% glucose. Three separate colonies per *easE* ortholog were picked and patched onto a fresh SC-URA plate. Each clone was then grown in 10 mL SC-URA supplemented with 0.1% glucose at 30 °C for 16 hrs.

Expression of *easC* and *easE* were induced by the addition of 500 μ L of filter-sterilized 40% galactose after the cells grew to an OD₆₀₀ of 0.8. The induction and production of chanoclavine-I was performed over 48 hrs at 24 °C, and the screw caps of the 50 mL falcon tubes were replaced with autoclaved aluminium foil covers to facilitate aeration. A 1 mL aliquot of spent media from each culture filtered through a 0.2 μ m PTFE syringe filter and analysed for the presence of chanoclavine-I using LC-MS. The relative amounts of chanoclavine-I biosynthesized, catalysed by *easE*_Aj and *easE*_Ec were estimated by the area under the peak from the extracted ion chromatogram at 257 m/z.

Statistical analysis of chanoclavine-I production levels was performed using the student's T-test.

3.2.2.4 SEQUENCE ALIGNMENTS

The sequence alignments of the selected *easE* orthologs with *easE*_Aj were performed using the Clustal Omega program, accessed through the Uniprot website (Pundir, Martin, O'Donovan, & Consortium, 2016).

3.2.3 DEVELOPMENT OF A CONSTITUTIVE CHANOC LAVINE-I PRODUCER

3.2.3.1 ASSEMBLY AND EVALUATION OF THE INITIAL STRAIN

The transcriptional unit for *easC* was assembled into the POT2 acceptor vector using the TDH3 promoter on HCKan_P, *easC* on HCKan_O, and the YIP5 terminator on HCKan_T. The transcriptional unit for *easE_Aj* was likewise assembled into the POT3 acceptor vector using the TPI1 promoter, *easE_Aj* on pSGG, and RPS20 terminator. These POTs were then condensed into the pMKU pathway acceptor plasmid and transformed into YMRW-W₆F₈ BY4741 chemically competent cells that were developed previously. Three colonies of the created construct were selected and tested using the same protocol described earlier. Samples were prepared for LC-MS analysis as mentioned previously.

The relative amounts of chanoclavine-I, 4DMA, and DMAT produced were estimated by the area under the peak from the extracted ion chromatogram at 257, 287, and 273 m/z respectively.

3.2.3.2 IDENTIFICATION OF THE SHUNT PRODUCTS

Commercial standards for MTCA (Sigma) were prepared by dissolving in fresh SC-URA media with 2% glucose to a final concentration of 10 μ M, and 1 mL of the solution used for LC-MS analysis.

3.2.3.3 PREPARATION OF BUFFERED YEAST CULTURES

Buffered SC-URA media (800 mL) was prepared by mixing 100 mL of 1 M MOPS pH 7, 100 mL 10X SC-URA dropout solution, 100 mL 10X Yeast Nitrogen Bases (YNB), 50 mL 40% glucose. The pH of the solution was

adjusted to pH 7 and the final volume was brought up to 1 L by adding deionized water and filter-sterilized. Cultures were grown and analysed using this buffered media and also regular SC-URA media previously described.

3.2.4 DESIGN AND ASSEMBLY OF THE CHANOCCLAVINE-I PROTOTYPE

The assembly of the pMKL- $C_{GAL10}E_{GAL1}$ construct was performed in a similar manner as pMKU- $C_{13}E_7$, but with the GAL10 and GAL1 promoter on HCKan_P and using the pMKL acceptor vector instead of pMKU. The pMKL- $C_{GAL10}E_{GAL1}$ construct was moved into the pGAU-YORW integration vector for insertion into the YMRW- W_6F_8 BY4741 genome. Three PCR verified colonies of the pMKL - $C_{GAL10}E_{GAL1}$ construct and YORW- $C_{GAL10}E_{GAL1}$ strain were cultured and the spent media analysed on the LC-MS as previously described.

The relative amounts of chanoclavine-I, 4DMA, and DMAT produced were measured by the area under the peak from the extracted ion chromatogram at 257, 287, and 273 m/z respectively.

3.2.5 VALIDATION OF CHANOCCLAVINE-I PRODUCED

3.2.5.1 EXPRESSION AND PURIFICATION OF EASE_AJ

The heterologous expression of easE_Aj cloned into the pET15 vector was attempted in BL21(DE3), Shuffle® T7 Express (NEB), and Origami™ B(DE3) *E.coli* cells (Novagen). The same expression protocol used previously with *dmaW*, *easF*, and *easC* was tested but no soluble protein was produced (Appendix A5-3). Extensive effort was undergone to find a successful expression condition, but to no avail.

The heterologous expression of the SUMO-easE_Aj fusion protein was then attempted in *S.cerevisiae* BY4741. Transformed cells containing the

POT3-pGAL1-SUMO-easE_Aj-tRPS20 vector were grown in SC-URA media supplemented with 0.1% glucose at 30 °C to an OD₆₀₀ of 0.8, prior to induction with galactose at a final concentration of 2%. Induction was carried out at 24 °C for 18 hrs. Cells were then harvested by centrifugation at 5000 rpm, 10 mins at 4 °C. The cells were then lysed by first resuspending them in 2 mL binding buffer (5 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl, pH 6) and treatment with lyticase from *Arthrobacter luteus* (Sigma) at 30 °C for 1 hr, followed by sonication. The cell lysate was then clarified by centrifugation at 15000 rpm for 20 mins at 4 °C. The purification of the protein was attempted using the Protomist® DTII with Ni²⁺-NTA chelating sepharose resin following the purification protocol described before, and fractions were analysed on SDS-PAGE (Appendix A5-4).

3.2.5.2 *IN VITRO* BIOSYNTHESIS OF CHANOC LAVINE-I

Purified *dmaW*, *easF*, and *easC* were prepared as described earlier. Cell lysates of easE_Aj and yEGFP were obtained following the expression protocol described previously, and used directly without further purification.

The *in vitro* biosynthesis of chanoclavine-I was prepared as 200 µL reactions. Each reaction consisted of 50 mM Tris-HCl, pH 7.5, 5 mM CaCl₂, 1 mM DMAPP, 1 mM *L*-tryptophan, 1 mM SAM, and 10 µL of each purified enzyme or cell lysate. Reactions were incubated at 24 °C for 48 hrs and were stopped by the removal of enzymes by filtration through a 3 kDa MWCO Ultra-0.5 spin filter (Amicon). The samples were either stored at -20 °C or directly analysed using LC-MS.

3.2.6 LC-MS ANALYSIS OF SAMPLES

All analyses using LC-MS were performed using the Agilent 1290 Infinity LC System. Screening of the *easE* orthologs and constructed strains were carried out using an Agilent Zorbax Eclipse Plus C18 column, 2.1 x 50 mm dimensions, 1.8 μ m particle size. Samples from the screen of *easE* orthologs were run using a shortened run method as previously described, that corresponded to the reduced column volume. Analysis of the *in vitro* reactions were carried out using an Agilent Zorbax Eclipse Plus C18 column, 2.1 x 100 mm dimensions, 1.8 μ m particle size and run on the same stepped gradient described in chapter 2. Mass analysis for all samples were performed as previously described.

3.3 RESULTS AND DISCUSSION

3.3.1 EXPANDING THE ENTRY-LEVEL VECTOR RANGE FOR FUSION TAGS

In the only successful heterologous expression of *easE*, the reported construct was expressed as a SUMO-fusion protein. Thus, the entry-level vector for ORFs was modified to similarly accept generically cloned ORFs but be released as a SUMO-ORF fusion when assembled into transcriptional units (Figure 3-4).

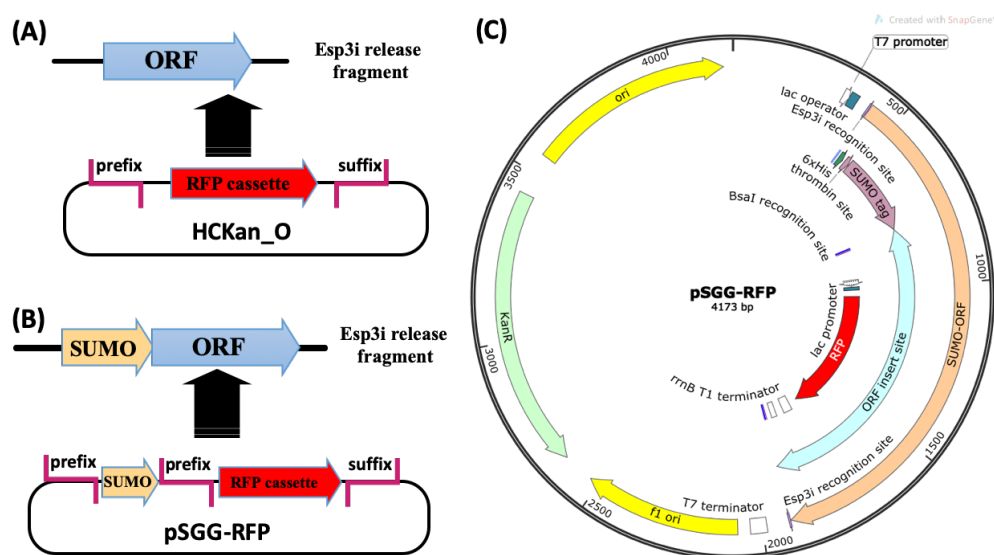


Figure 3-4: Illustration of the ORF released from pSGG-RFP. (A) Fragment released from HCKan_O. (B) Fragment released from pSGG-RFP, SUMO-ORF fusion. (C) Vector map of pSGG-RFP.

The modification of the entry-level vector to allow the assembled transcriptional units to express various common protein fusion constructs or tags expands the versatility of the assembly system, while maintaining the compatibility of the method with automation and ease of introducing new ORFs. With enough modified entry-level vectors for each of the different types of tags, this approach would facilitate the screening of an assortment of different tags and fusions that might help with the expression and solubility of new target

proteins, without needing to re-clone the target gene. This vector, pSGG-RFP, was henceforth used for the creation of SUMO-*easE* fusions that were screened for activity and subsequently built into the pathway.

3.3.2 PROSPECTING FOR FUNCTIONAL PREDICTIONS

3.3.2.1 MINING WITH SEQUENCE SIMILARITY NETWORKS

Like a double-edged sword, the exponentially increasing amount of genome sequencing data has facilitated the discovery and identification of the biosynthesis for numerous known and unknown secondary metabolites. It has, however, also made it unrealistic and unfeasible to manually curate and experimentally validate all the coding regions of the enzymes within these clusters. The only viable method to perform the herculean task of making sense of this colossal amount of data is to parse it with predictive algorithms computationally, and then analysing the summarized data.

When coupled with the extensive multitudes of ways in which different organisms might organize their introns and exons, the prediction of coding sequences from such data is more often than not prone to errors (Brent & Guigo, 2004). This is most evident in the common difficulty encountered in studies attempting to express and research proteins from higher eukaryotes with alternative splice variants (Mathé, Sagot, Schiex, & Rouzé, 2002; Zhang, 2002).

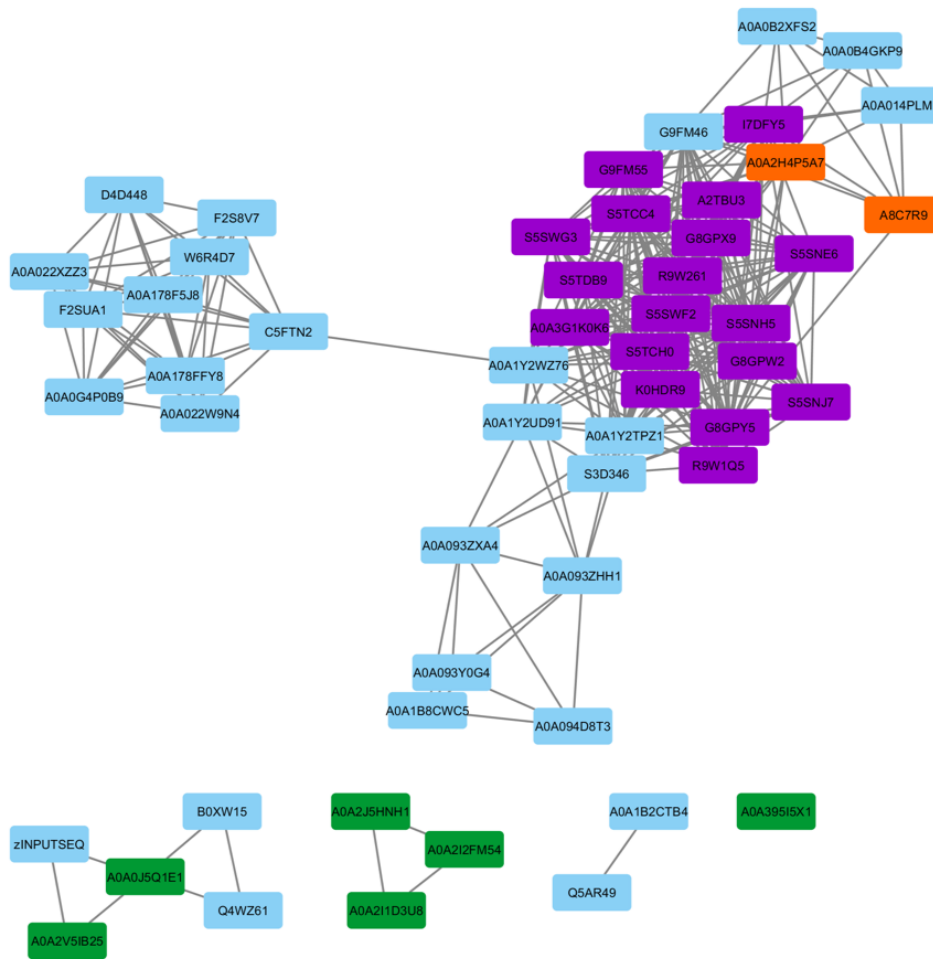


Figure 3-5: SSN for the expanded isofunctional cluster of *easE* from the redefining of the alignment score for an edge to 170. Nodes in **green represent sequences from the *Aspergillus* family, nodes in **purple** represent sequences from the *Epichloe* family, nodes in **orange** represents sequences from *Claviceps*, and nodes in **blue** represent sequences from the other collective families.**

To mitigate this, an SSN was generated using what presumably is a correctly annotated *easE* sequence. An SSN is one of several methods commonly used to visually represent sequence relationships and by proxy sequence function. In the representation of an SSN, each sequence surveyed is denoted by a node and is linked by an edge to other relevant nodes that show the degree of similarity equal to or greater than a predefined value. By using the EFI-EST webtool to generate an SSN, the program uses the input sequence and queries the Uniprot database to procure a collection of sequences that are of a specified degree of similarity to the query. An all-by-all BLAST is then

performed on this said collection to identify the degree of similarity between all sequences within this collection and used to draw the edges. The further refinement of the alignment score that defines an edge, will then split the pool of sequences into clusters. By setting an appropriate alignment score, these clusters will conceptually progress to delineate into isofunctional groups. A small library of orthologous proteins, selected from this pool might not only be properly annotated and functional, but also agreeable to heterologous expression and purification.

From the predicted isofunctional cluster of *easE*, the different orthologs from the various ergot-producing organisms could be further delineated by increasing the alignment score used as a cut-off for defining an edge (Figure 3-5). From this expanded isofunctional cluster, representative sequences for each of the major groups of organisms were chosen to be screened for *easE* activity.

3.3.2.2 A SCREENING METHOD FOR CHANOCLAVINE-I PRODUCTION

In the catalysis of 4DMA to chanoclavine-I, both *easC* and *easE* are essential for the reaction to take place. Fortuitously, *easC* has been shown to likely be able to function with any ortholog of *easE* regardless of the source, allowing the enzymes from different sources to be mixed and matched (Nielsen et al., 2014). Furthermore, the studies conducted by the Panaccione and O'Connor groups have demonstrated that functional expression of *easC* could be discerned in isolation of *easE*; either from the production of a side product, ergotryptamine, when co-expressed with *dmaW* and *easF* (K. L. Ryan et al., 2014) or by the presence of catalase activity (Nielsen et al., 2014).

With these considerations in mind, an *easE* screening yeast strain was designed to have *dmaW*, *easF*, and *easC* stably integrated on its genome. Plasmids containing expression cassettes of the various *easE* orthologs to be tested could be screened for the production of chanoclavine-I (Figure 3-7A). To create this strain, the ortholog of *easC* from *C. purpurea* was chosen and tested. *EasC* from *C. purpurea* was successfully expressed in *E. coli* and yeast, and the purified enzyme was found to exhibit chemiluminescence when incubated with ECL substrate (Figure 3-6); suggesting that this *easC* ortholog is functionally expressed.

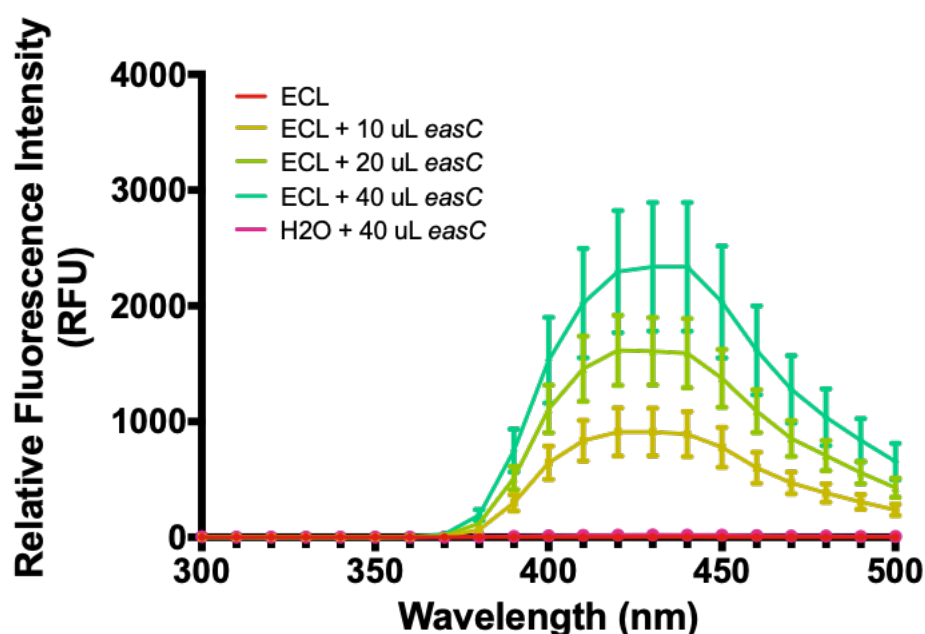


Figure 3-6: Incubation of *easC* with ECL substrate produces chemiluminescence. Demonstrating catalase activity from purified *easC*. Error bars were calculated from three technical replicates.

EasC was then put through the YeastFab pipeline to create the screening strain, YMC17, where *easC* is under the control of a GAL10 promoter as illustrated, and can be induced when required together with *easE* (Figure 3-10A). The eight selected orthologs of *easE*, together with the ortholog from *A. japonicus*, and a negative control (yEGFP) were all constructed to be under the

control of the GAL1 promoter on POT plasmids. These were transformed into YMC17 and screened for chanoclavine-I production. Of the eight *easE* orthologs, only one of them, from *Epichloe coenophialia*, was found to produce a compound with an $[M+H]^+$ of 257 m/z that corresponded to the compound produced by the *A. japonicus* ortholog (Figure 3-7C). Comparison of the mass spectra from both compounds and the theoretical mass values, suggested that they are both chanoclavine-I (Figure 3-7E, F) (Table 3-3).

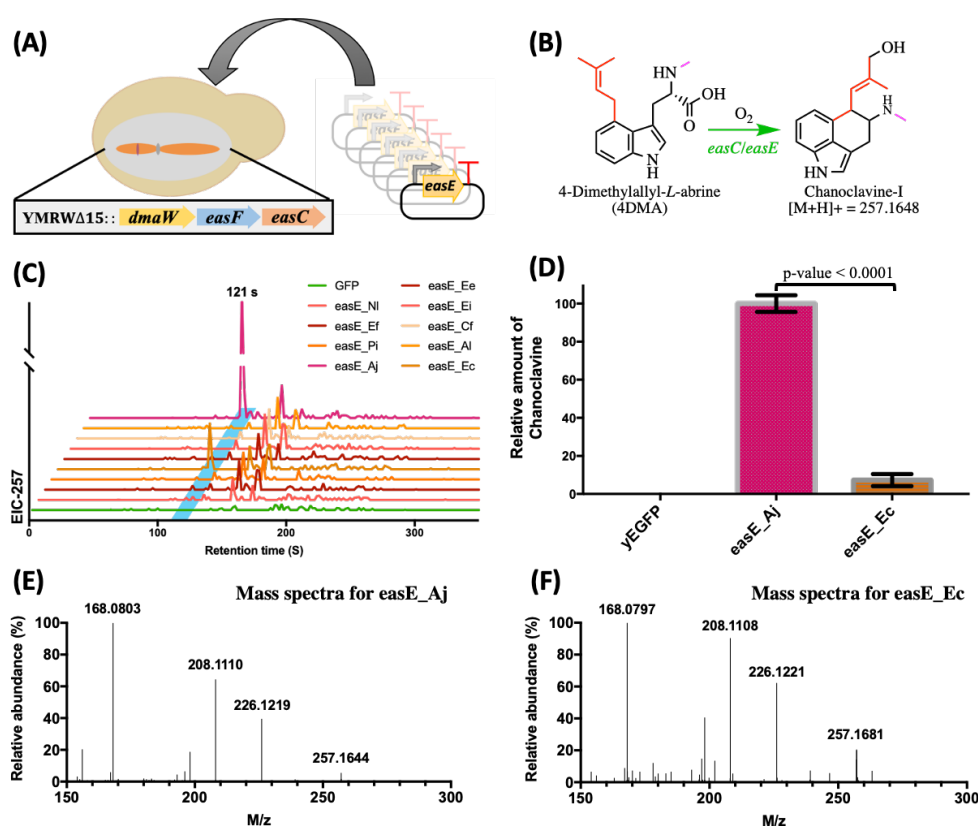


Figure 3-7: (A) Illustration of the *easE* screening strain (YMC17) and screen design. (B) Reaction producing chanoclavine-I from 4DMA, catalysed by *easC* and *easE*. (C) Extracted ion chromatograms of *easE* orthologs tested for 257 m/z. (D) Relative amounts of chanoclavine-I produced from the *A. japonicus* and *E. coenophialia* orthologs, estimated from the area under the peak highlighted in blue from C. Error bars represent standard deviation from three biological replicates. (E, F) Mass spectra for compound eluting at 121 s.

Table 3-3: Exact mass and retention time of Chanoclavine-I produced

<i>easE</i> source	Calculated m/z	Observed m/z	Mass error δ (ppm)	Retention time (s)
<i>A. japonicus</i>	257.1648	257.1644	-1.56	120.6
<i>E. coenophialia</i>	257.1648	257.1681	12.83	120.6

The *E. coenophialia* ortholog was, however, found to be producing approximately 10-fold less chanoclavine-I than the *A. japonicus* ortholog (Figure 3-7D). Further analysis of the sequence alignments between the *A. japonicus* ortholog with the selected orthologs revealed that the *E. coenophialia* ortholog shared more similarity in the first half of the predicted FAD-binding domain compared to the other seven (Figure 3-8).

This discrepancy in sequence similarity and the four amino acid gap in the catalytic domain is present in all dysfunctional *easE* sequences, hinting on the possibility that there is an exon that has been left out in the prediction of these seven sequences. It is possible that in the annotation of the *E. coenophialia* ortholog from its genomic DNA sequence, this particular exon was partially recognized and correctly translated. Therefore depositing a partially correct amino acid sequence that was sufficient to restore some semblance of activity but insufficient to match that of the *A. japonicus* ortholog.

The regulation of the production levels of secondary metabolites do not usually occur through the sequence variation of its enzymes, but rather at the transcriptional level. This suggests that the decreased production levels of chanoclavine-I from the *E. coenophialia* ortholog are more likely to be due to erroneous exon annotation than due to regulation of ergot alkaloid production.

(A)

S5TDB9	S5TDB9_EPINE	111	VSSS-ERPQEIPCHQGRRLPLYSAAVKSTAHVQGVIRFAKDHNRLRVIKNTGHDATGRSAA	169
R9W1Q5	R9W1Q5_9HYPO	111	VSSS-ERSQEIPCHQGRRLPLYSAAVKSTAHVQSVRFASDHNRLRVIKNTGHDATGRSAA	169
G9FM46	G9FM46_9HYPO	109	MTSPGEG-RGSACHQGRRLPLYSAVVQSTSDIQSSVRFASDHNRLRVIKNTGHDACGRSSS	167
K0HDR9	K0HDR9_EPIEL	111	VSSS-ERSQEIPCHQGRRLPLYSAAVKSTAHVQSVRFASDHNRLRVIKNTGHDATGRSAA	169
I7DFY5	I7DFY5_9HYPO	109	MTSPGEG-RRSVCHQGRRLPLYSAVVQSNISQSSVRFASDHNRLRVIKNTGHDACGRSSS	167
A8C7R9	EASE_CLAFS	109	TTWPGYHRTSDCHQGRRLPLYSAIVESTSDIQSCVKFANHNRLRVIKNSGHDATGRSSA	168
A0A0S7DKV5	A0A0S7DKV5_9EURO	110	H-SP---RGARCHQGRRLPLYSAAVESVDQVQAAVRFAQRHVRRLVVRNTGHDATGRSSA	164
R9W261	R9W261_EPICN	111	VSSS-ERPQEIPCHQGRRLPLYSAAVKSTAHVQGVIRFAKDHNRLRVIKNTGHDATGRSAA	169
easE_Aj		121	LRSG---PAGSCHQGRRLPLYSAAVKSASDVQKVDFTRQHNRLRVIRNTGHDGSGRSSG	176
*****:*****:*.:.:*.:.:*.:.:*****:*****:*****:				
S5TDB9	S5TDB9_EPINE	170	PDSLQIHTYFLKDIHYDDNFLVHGDA---GSGPAVTLGAGVHSEVYKHGIDHKYSVV	225
R9W1Q5	R9W1Q5_9HYPO	170	PDSLQIHTYFLKDIHYDDNFFAQGDAT---GSGPAVTLGAGVMHSEVYKHGIDHKYSVV	225
G9FM46	G9FM46_9HYPO	168	PNSFQIHTNLLKSI SFHNFVARGSTT---CSGPAVTLGAGVMHSEVYAHGVENGTYIV	223
K0HDR9	K0HDR9_EPIEL	170	PDSLQIHTYFLKDIHYDDNFFAQGDAT---GSGPAVTLGAGVMHSEVYKHGIDHKYSVV	225
I7DFY5	I7DFY5_9HYPO	168	PHSFQIHTNLLKSI SFHNFVARGSTT---GYGPAVTLGAGVMHSEVYAHGAEGYITIL	223
A8C7R9	EASE_CLAFS	169	PHSFQISTSSLKTI SLHNFVPRGSTT---GHGPAVTLGAGVMHSEVYAHGVKNAYTIL	224
A0A0S7DKV5	A0A0S7DKV5_9EURO	165	PDSFQIHTHRMKQIEYHDFRAVSSDI---DQGPVAVSVGAGVLLGEMVARGARDGVVVV	220
R9W261	R9W261_EPICN	170	PDSLQIHTYFLKDIHYDDNFLVHGDA---GSGPAVTLGAGVHSEVYKHGIDHKYSVV	225
easE_Aj		177	PDSVEIHTHLLNSVQYHPNFRPAGSSERQSAPGQPAVTVGAGILLGDLVARGASEGWIIV	236
..*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:				
S5TDB9	S5TDB9_EPINE	226	GGECPTVGI VGGFLQGGGVSSWSGFTRLGLAVDNVLEYQVVTANAELVIANEHQNDLFWA	285
R9W1Q5	R9W1Q5_9HYPO	226	GGECPTVGI VGGFLQGGGVSSWSGFTRLGLAVDNVLEYQVVTANAELVIANEHQNDLFWA	283
G9FM46	G9FM46_9HYPO	224	GGECPTVGA VGGFLQGGGVSSISHSFTKGLAVDNVLEYQVVTANADLVIANEHQNDLFWA	283
K0HDR9	K0HDR9_EPIEL	226	GGECPTVGI VGGFLQGGGVSSWSGFTRLGLAVDNVLEYQVVTANAELVIANEHQNDLFWA	283
I7DFY5	I7DFY5_9HYPO	224	GGECPTVGA VGGFLQGGGVSSISHSFTKGLAVDNVLEYQVVTANADLVIANEHQNDLFWA	283
A8C7R9	EASE_CLAFS	225	GGECPTVGA VGGFLQGGGVSSISHSFTKGLAVDNVLEYQVVTANADLVIANEHQNDLFWA	284
A0A0S7DKV5	A0A0S7DKV5_9EURO	221	GGECPTVGA TGGFLQGGGVSSPHSFINGLAVDNVLEYEVTAKEVIVVIANSSQNDLFWA	280
R9W261	R9W261_EPICN	226	GGECPTVGI VGGFLQGGGVSSWSGFTRLGLAVDNVLEYQVVTANAELVIANEHQNDLFWA	285
easE_Aj		237	GGECPTVGAAGGFLQGGGVSSWLSYAHGLAVDNVLEYEVVTAKGEIVIANAHQNDLFWA	296
*****:*****:*.:.:*****:*****:*****:*****:*****:*****:				
S5TDB9	S5TDB9_EPINE	286	LRGGGGGTGFGVVTQATVRAFPDDPTVSTVLVLSSTRADTSFWVKASIRLLSILRSCNQON	345
R9W1Q5	R9W1Q5_9HYPO	284	LRGGGGGTGFGVVTQATVRAFPDDPTVSTVLVVTSPRADTTLWEKAISIRLLSILRSCNQON	343
G9FM46	G9FM46_9HYPO	284	LKGGGGGTGFGIVTQATVRYVYDPPVVTATMIKTAFANTLFWTEGVRELLRLQLYFNQLH	343
K0HDR9	K0HDR9_EPIEL	284	LRGGGGGTGFGVVTQATVRAFPDDPTVSTVLVVTSPRADTTLWEKAISIRLLSILRSCNQON	343
I7DFY5	I7DFY5_9HYPO	284	LKGGGGGTGFGIVTQATVRYVYDPPVVTATMIKTAFANTLFWTEGVRELLRLQLYFNQLH	343
A8C7R9	EASE_CLAFS	285	LRGGGGGTGFGVQAQATIRVFPDDPTVVTATTIKAAVTNTMFWTEGVRELLRLQLYFNQLH	344
A0A0S7DKV5	A0A0S7DKV5_9EURO	281	LRGGGGGTGFGVTRATMRVHNDPVCISEVAVSGHRNNSLWTRGVAGLFSVLRSFNQOG	340
R9W261	R9W261_EPICN	286	LRGGGGGTGFGVVTQATVRAFPDDPTVSTVLVLSSTRADTSFWVKASIRLLSILRSCNQON	345
easE_Aj		297	LRGGGGGTGFGVVTQATLQVHPDLVPSVADAVTGSRADATFWSHGVAALLRALQFLNNHG	356
..*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:				

(B)

R9W261	R9W261_EPICN	110	LVSS-ERPQEIPCHQGRRLPLYSAAVKSTAHVQGVIRFAKDHNRLRVIKNTGHDATGRSAA	169
easE_Aj		120	YLRSGP---AGSCHQGRRLPLYSAAVKSASDVQKVDFTRQHNRLRVIRNTGHDGSGRSSG	176
*****:*****:*.:.:*****:*****:*****:*****:*****:*****:				
R9W261	R9W261_EPICN	170	PDSLQIHTYFLKDIHYDDNFLVHGDA---GSGPAVTLGAGVHSEVYKHGIDHKYSVV	225
easE_Aj		177	PDSVEIHTHLLNSVQYHPNFRPAGSSERQSAPGQPAVTVGAGILLGDLVARGASEGWIIV	236
..*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:				
R9W261	R9W261_EPICN	226	GGECPTVGI VGGFLQGGGVSSWSGFTRLGLAVDNVLEYQVVTANAELVIANEHQNDLFWA	285
easE_Aj		237	GGECPTVGAAGGFLQGGGVSSWLSYAHGLAVDNVLEYEVVTAKGEIVIANAHQNDLFWA	296
*****:*****:*.:.:*****:*****:*****:*****:*****:*****:				
R9W261	R9W261_EPICN	286	LRGGGGGTGFGVVTQATVRAFPDDPTVSTVLVLSSTRADTSFWVKASIRLLSILRSCNQON	345
easE_Aj		297	LRGGGGGTGFGVVTQATLQVHPDLVPSVADAVTGSRADATFWSHGVAALLRALQFLNNHG	356
..*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:.*.*.*.*:				

Figure 3-8: Sequence alignments for the FAD-binding domain of the *easE* orthologs. Alignments of the domain between orthologs of *E. coenophialia* and *A. japonicus* (B) showing a higher degree of similarity compared to an alignment between all selected orthologs (A). Red boxes in (A) and (B) highlighting region within the conserved FAD-binding domain that presents the most variation between the *easE* orthologs.

Nevertheless, these results have allowed the identification of a suitable combination of *easC* (*C. purpurea*) and *easE* (*A. japonicus*) sequences that can be used to complete the early ergot alkaloid pathway.

3.3.3 DEVELOPMENT OF A CONSTITUTIVE CHANOC LAVINE-I PRODUCER

3.3.3.1 AN INITIAL ASSESSMENT

Having identified a functional set of enzymes to complete the early ergot pathway, a construct with *easC* and *easE* driven by constitutive promoters was created to develop a more consistent producer of chanoclavine-I (Figure 3-9A). Going down the list of promoters characterized (Figure 2-3D), the next two strongest promoters TDH3 and TPI1 were selected and used to drive the expression of these two enzymes respectively. This test strain was found to be capable of producing the same 257 m/z compound as before (Figure 3-9B). It was noticed, however, that the amount of chanoclavine-I being produced was drastically less than previously observed, and approximately two orders of magnitude lower than the estimated levels of DMAT and 4DMA (Figure 3-9C).

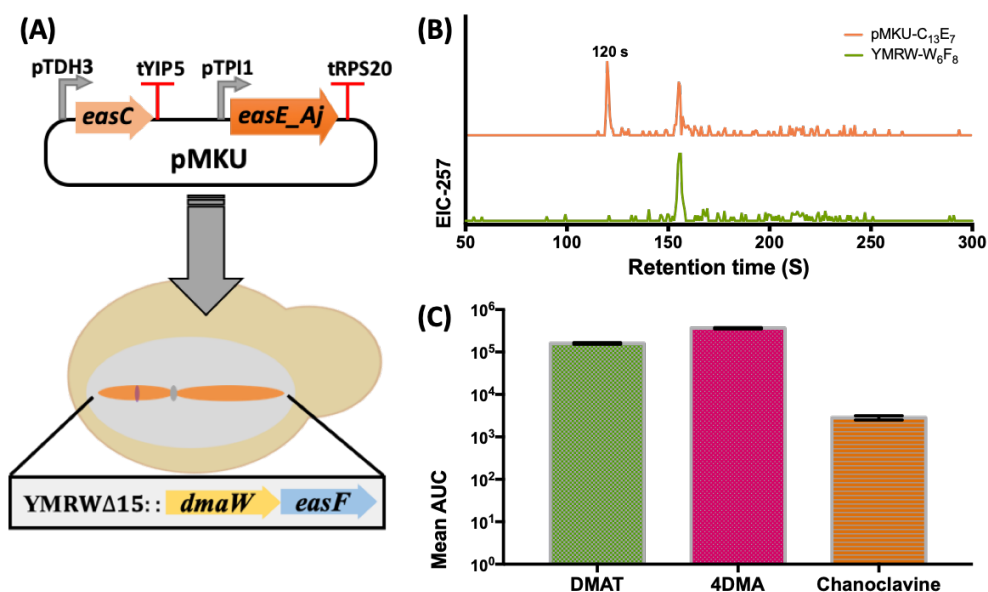


Figure 3-9: (A) Cartoon representation of the pMKU-C₁₃E₇ strain tested. (B) Extracted ion chromatogram of the strain tested for 257 m/z. (C) Relative amounts of DMAT, 4DMA, and chanoclavine-I produced, estimated from the area under the peaks for their respective extracted ion chromatograms. Error bars represent the standard deviation from three biological replicates.

3.3.3.2 DISCOVERY AND IDENTIFICATION OF SHUNT PRODUCTS

In addition to the accumulation of DMAT and 4DMA, this strain (pMKU-C₁₃E₇) as well as the 4DMA producing strain developed from the earlier chapter (YMRW-W₆F₈), was found to be producing two compounds with an [M+H]⁺ of 231 m/z and 299 m/z at levels that were abundantly observable from the Total Ion Chromatogram (TIC). A literature search suggested that these compounds might be MTCA and dimethylallyl-MTCA, shunt products formed from the reaction of tryptophan and DMAT with endogenous acetaldehyde non-enzymatically, at low pH (Figure 3-10) (Backhaus, Ludwig-Radtke, Xie, & Li, 2017).

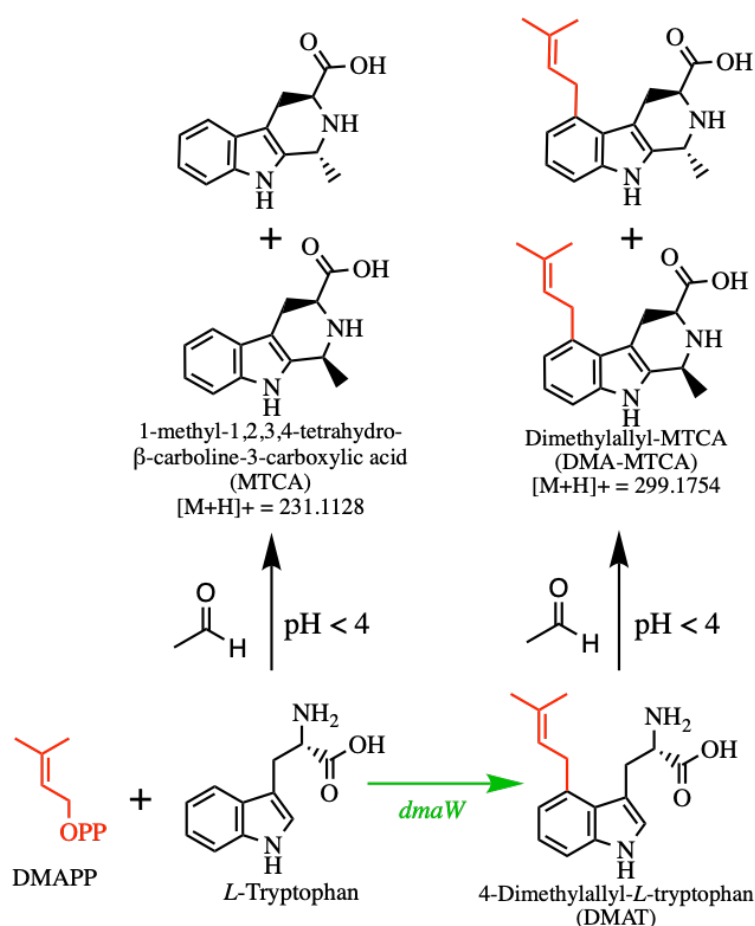


Figure 3-10: Pictet-Spengler reaction of tryptophan and DMAT with formaldehyde producing the shunt products MTCA and its prenylated form, DMA-MTCA. Media acidification during late log phase growth produces the conditions required for the reaction.

LC-MS analysis of commercially obtained MTCA dissolved in SC media verified the identity of the 231 m/z product to be MTCA, with both commercial standard and *in vivo* produced MTCA eluting at around 102 s (Figure 3-11B) and with similar mass spectral patterns (Figure 3-11D, E). Further analysis of the mass spectra of the 299 m/z compound and its most abundant fragment ions, consistently show a mass difference of 69 m/z from MTCA and its corresponding fragments that corresponds to the prenyl moiety (Figure 3-11F). This indicated DMAT also underwent the same Pictet–Spengler reaction as tryptophan at low pH with acetaldehyde; the 299 m/z product being the prenylated version of MTCA, dimethylallyl-MTCA (DMA-MTCA).

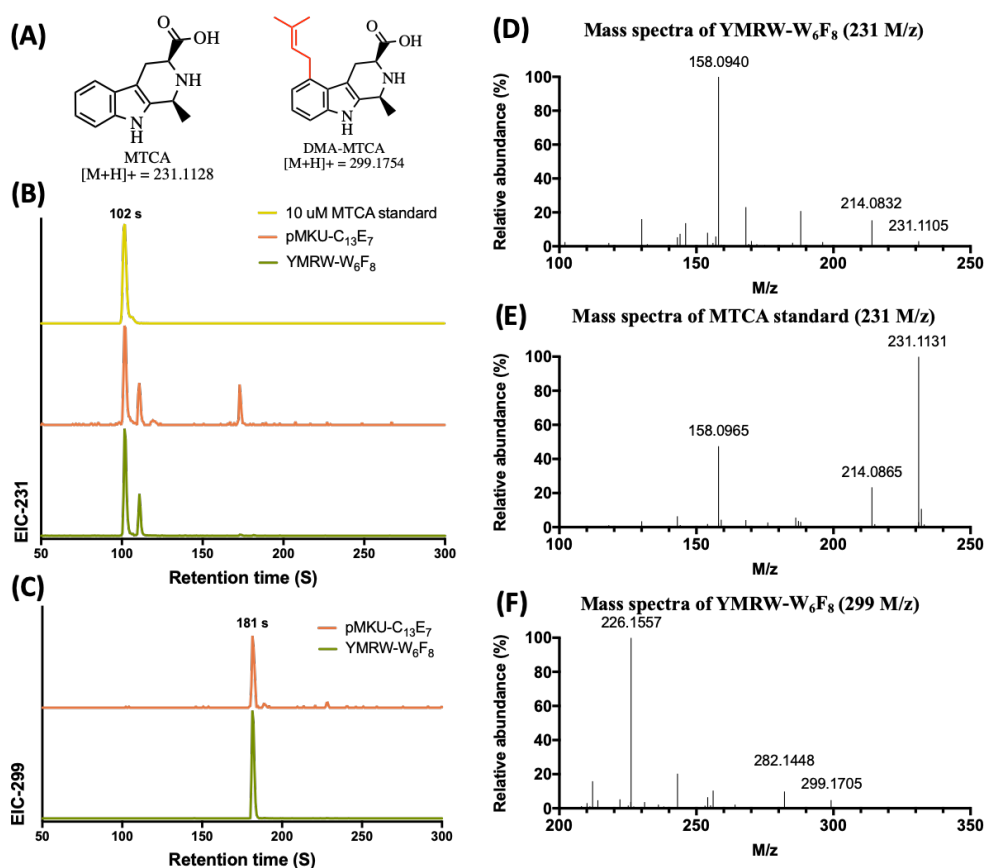


Figure 3-11: (A) Shunt products MTCA and DMA-MTCA. (B) Extracted ion chromatograms for 231 m/z showing the peaks at 102 s corresponding to one stereoisomer of MTCA. (C) Extracted ion chromatograms for 299 m/z showing the peak at 181 s corresponding to DMA-MTCA. (D, E) Mass spectra for the peaks at 102 s for MTCA (F) Mass spectra for the peaks eluted at 181 s for DMA-MTCA.

Table 3-4: Exact mass and retention time of MTCA and DMA-MTCA produced

Construct	MTCA			DMA-MTCA		
	Calc. m/z	Obs. m/z	Mass error δ (ppm)	Calc. m/z	Obs. m/z	Mass error δ (ppm)
YMRW-W ₆ F ₈	231.1128	231.1105	-9.95	299.1754	299.1705	-16.38
pMKU-C ₁₃ E ₇	231.1128	231.1100	-12.12	299.1754	299.1706	-16.04

3.3.3.3 PREVENTING THE FORMATION OF SHUNT PRODUCTS

Suspecting that the metabolic flux to chanoclavine-I was being significantly shunted away to produce these products at late log phase, pMKU-C₁₃E₇ and YMRW-W₆F₈ were grown in SC media buffered by 100 mM MOPS, pH 7. Maintaining the pH of the cultures above 4 managed to almost completely eradicate the production of MTCA and DMA-MTCA (Figure 3-12A, B). However, it did not improve the turnover of 4DMA to chanoclavine-I (Figure 3-12C).

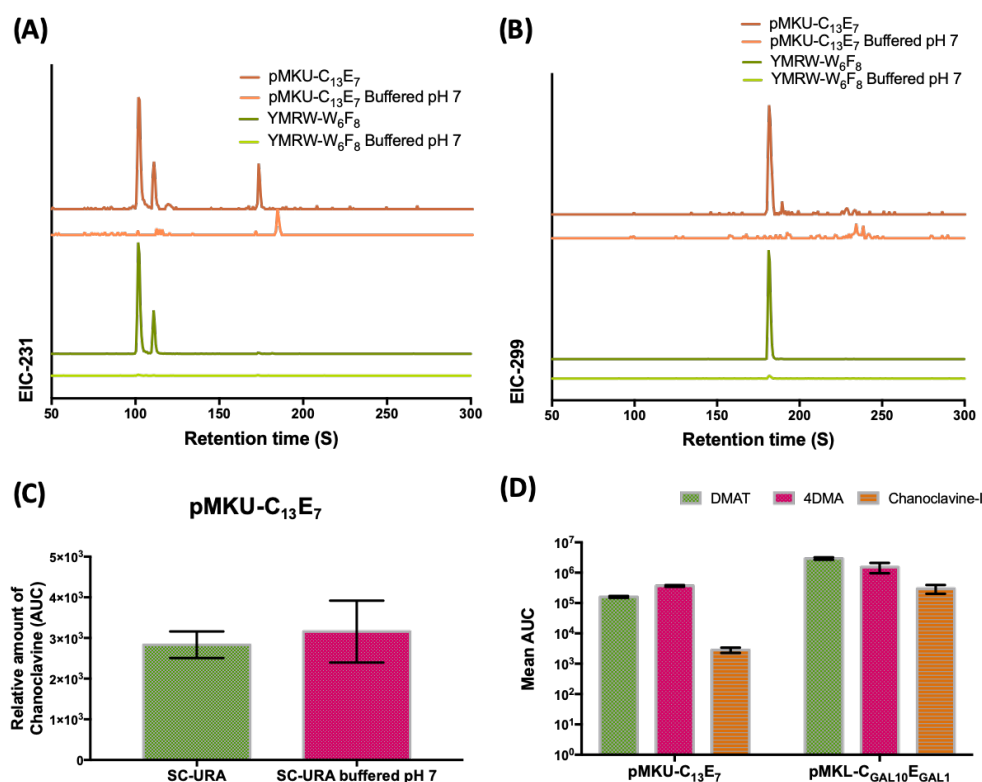


Figure 3-12: (A, B) Buffering the culture media at pH 7 eradicates the formation of MTCA and DMA-MTCA. (C) Preventing the formation of MTCA and DMA-MTCA does not improve the production of chanoclavine-I. Amounts of chanoclavine-I produced estimated from the area under the peak from the extracted ion chromatogram for 257 m/z. Error bars represent the standard deviation from three biological replicates. (D) Comparison of the levels of intermediates produced between pMKU-C₁₃E₇ and pMKL-C_{GAL10}E_{GAL1} strains. Relative amounts of the intermediates were estimated from the area under their respective peaks from their extracted ion chromatograms. Error bars represent standard deviation calculated from three biological replicates.

The poor production of chanoclavine-I may be due to the poor catalytic efficiency of *easC/easE*. This would also explain the required presence of four copies of *easC* in the strain developed by Sarah O'Connor's group that was shown to be capable of producing up to 529 mg/L of cycloclavine (Jakubczyk et al., 2015). Holding true to this hypothesis, when pMKU-C₁₃E₇ was compared against pMKL-C_{GAL10}E_{GAL1} it was immediately observed that detected levels of DMAT, 4DMA, and chanoclavine-I increased by approximately an order of magnitude (Figure 3-12D). More importantly, increasing the expression of *easC* and *easE* as much as possible, through the use of strong inducible promoters, narrows the difference in levels of DMAT, 4DMA and chanoclavine-I produced to within an order of magnitude. Through these series of experiments, it is postulated that the key limiting factor behind the turnover of 4DMA to chanoclavine-I was the catalytic efficiency of *easC/easE*. Unfortunately, the determination of kinetic parameters was not obtainable, due to the difficulty encountered in purifying *easE* from any source. Lastly, the determination of the shunt products MTCA and DMA-MTCA that formed at low pH, while not limiting at this juncture, would certainly be an important consideration during the eventual bioprocess development and optimization.

3.3.4 VALIDATION OF THE CHANOCLAVINE-I PRODUCED

Building on the earlier finding that pMKL-C_{GAL10}E_{GAL1} was the best performing construct, it was chosen to be moved into pGAU-YORW for integration into the genome. Both strains holding the construct on a plasmid and the genome integrated variant showed the same production profile (Figure 3-13B).

To validate the production of chanoclavine-I on all the strains created over the course of this phase, the same approach to create a biosynthesized standard was taken, similar to the production of DMAT and 4DMA. Despite the difficulty with purifying *easE*, it was noticed that the clarified cell lysate of the yeast strains solely expressing *easE_Aj* were capable of showing catalytic activity. Thus, this was used as a substitute for the purified enzyme in reproducing the production of chanoclavine-I observed *in vivo*. Following the procedure from the previous chapter, *in vitro* biosynthetic reactions were set up with purified *dmaW*, *easF*, *easC*, and yeast cell lysate either expressing *easE* alone or yEGFP (negative control).

From the *in vitro* biosynthetic reactions, a compound with the [M+H]⁺ of 257 m/z was found to elute at around 147 s. This was much later than the elution times of the corresponding compounds detected from the *in vivo* biosynthesis (Figure 3-13B, C). The mass spectra for these compounds, however, showed comparable masses that corresponded to the theoretical value of chanoclavine-I (Table 3-5). Analysis of the fragment ions produced from both sources of the compound further showed similar profiles (Figure 3-13D, E, F) (Table 3-6). This indicated that the detected compound from both reactions was chanoclavine-I.

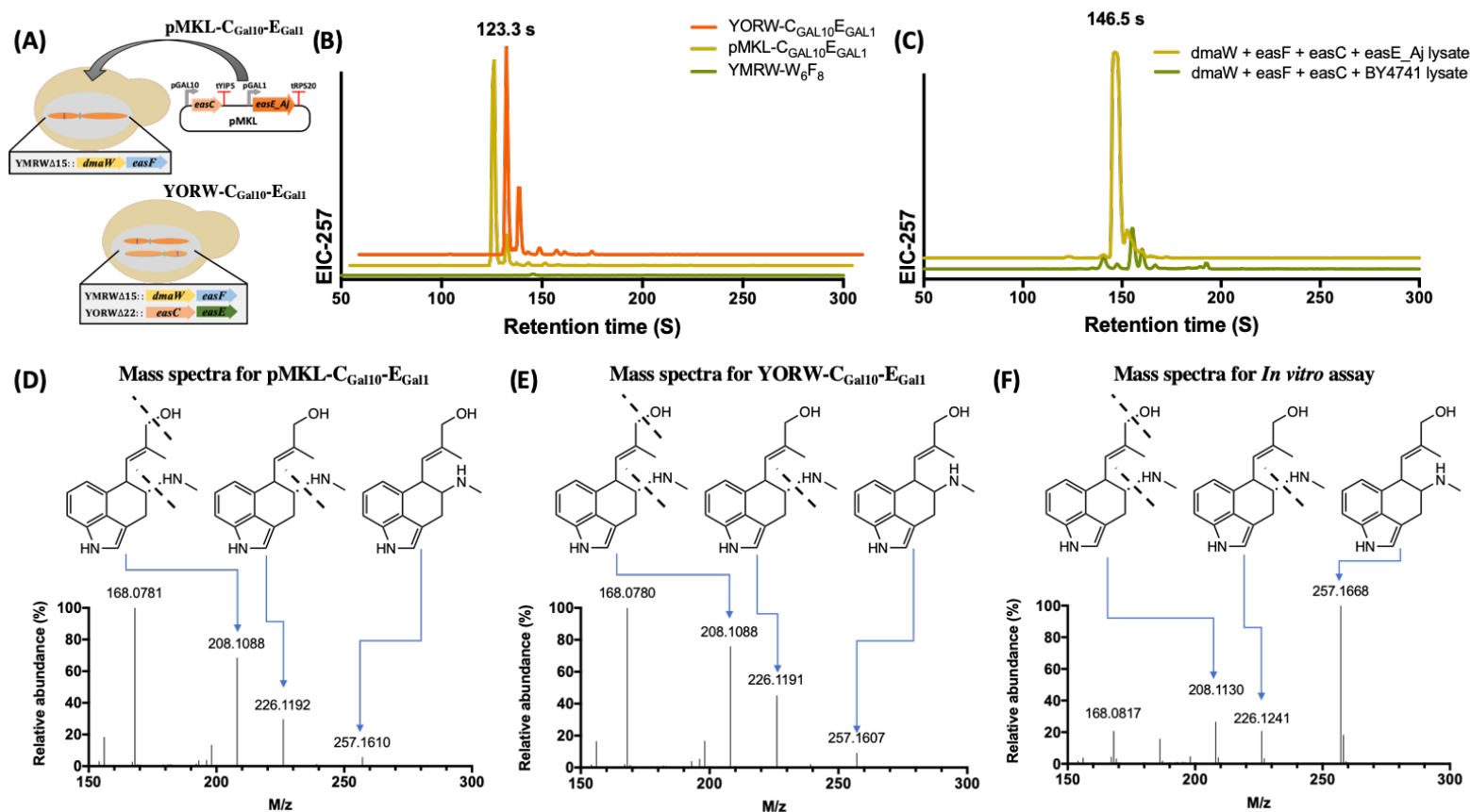


Figure 3-13: (A) Illustration of strains pMKL-C_{Gal10}^E_{Gal1} and the genome integrated construct YORW-C_{Gal10}^E_{Gal1}. (B) Extracted ion chromatograms for 257 m/z, peak at 123 s indicating the production of chanoclavine-I from the strains created. (C) Extracted ion chromatograms for 257 m/z of the *in vitro* biosynthetic reactions performed. The peak eluting at 146 s indicated the production of chanoclavine-I. (D, E, F) Comparison of the mass spectra obtained from the peaks corresponding to chanoclavine-I.

Table 3-5: Exact mass and retention time for chanoclavine-I produced

Construct	Calculated m/z	Observed m/z	Mass error δ (ppm)	Retention time (257 m/z)	Retention time (4DMA)	Relative retention to 4DMA
<i>In vitro</i> assay	257.1648	257.1668*	7.78	146.49	208.86	0.70
pMKL-C _{GAL10} E _{GAL1}	257.1648	257.1610	-14.78	121.81	183.29	0.66
YORW-C _{GAL10} E _{GAL1}	257.1648	257.1607	-15.94	123.27	183.44	0.67

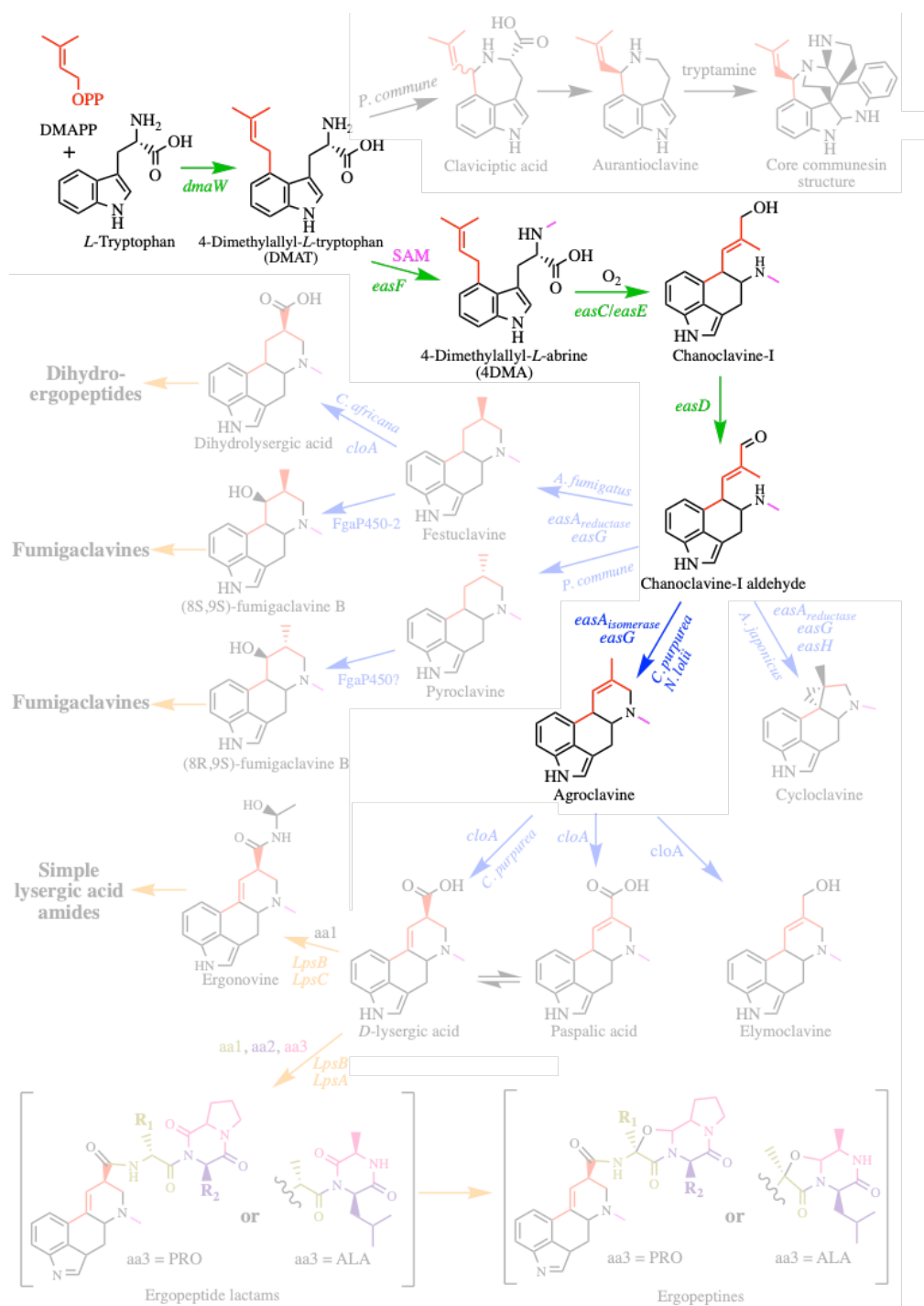
Table 3-6: Exact mass of the simplest fragment ions from chanoclavine-I

Fragment	Calculated m/z	Observed m/z			Mass Error δ (ppm)		
		pMKL -C _{GAL10} -E _{GAL1}	YORW -C _{GAL10} -E _{GAL1}	<i>In vitro</i> assay	pMKL -C _{GAL10} -E _{GAL1}	YORW -C _{GAL10} -E _{GAL1}	<i>In vitro</i> assay
- NHCH ₃	226.1226	226.1192	226.1191	226.1241	-15.04	-15.48	6.63
- OH, - NHCH ₃	208.1121	208.1088	208.1088	208.1130	-15.86	-15.86	4.32

Upon closer inspection of the *in vitro* biosynthetic reaction, the production of 4DMA was likewise shifted back by approximately 20 s compared to the 4DMA from the *in vivo* biosynthesis. The difference between the retention time for chanoclavine-I and 4DMA is the same in both experiments (Table 3-5). Thus, the discrepancy between the retention time of chanoclavine-I produced *in vivo* and *in vitro* may be attributed as an artefact stemming from the difference in column length used for LC-MS analysis. A 100 mm column was used for the analysis of the *in vitro* biosynthetic reactions, while a shorter 50 mm column was used for the analysis of the *in vivo* biosynthesis samples. The shorter column was used for the analysis of the *in vivo* samples to shorten the overall run time for each sample, in order to expedite the screening process.

This chapter has successfully replicated the work published by Sarah O'Connor's group for the heterologous production of chanoclavine-I in yeast (Nielsen et al., 2014); albeit using a different set of biological parts and from a different approach. The creation of this chanoclavine producing strain, YORW-C_{GAL10}E_{GAL1} (YOCE), solidified the platform where the pathway can be further expanded to agroclavine and *D*-lysergic acid.

CHAPTER 4: DEVELOPING AND OPTIMIZING AN AGROCLAVINE PRODUCING STRAIN



4.1 INTRODUCTION

The reaction producing agroclavine is a key structural branch point in the biosynthesis of the ergot alkaloids; the formation of the ergoline D ring from chanoclavine-I requires two chemical reactions. The first reaction is catalysed by *easD*, a short-chain alcohol dehydrogenase that oxidizes the hydroxyl group on chanoclavine-I to an aldehyde (Wallwey, Matuschek, & Li, 2010); which can react with the adjacent secondary amine (Figure 4-1). However, the *sp*²-*sp*² planarity (rigidity) of the C8-C9 double bond in chanoclavine-I-aldehyde prevents intramolecular cyclization, as the aldehyde functional group is held in an unproductive, distal orientation relative to the methylamino functional group.

The isoforms of *easA*, from the different lineages of ergot alkaloid producing organisms, catalyses either a reduction or a cis-trans isomerization across the C8-C9 alkene (J. Z. Cheng et al., 2010a, 2010b; Floss et al., 1968). This repositioning of the aldehyde group to a proximal and productive orientation, with respect to the methylamino group, facilitates the intramolecular cyclization. This reaction forms the ergoline D ring and creates an iminium intermediate. An NADPH-dependent oxidoreductase, *easG*, catalyses the reduction of the iminium intermediate that leads to the formation of a tetracyclic clavine (Matuschek, Wallwey, Xie, & Li, 2011). The reduction or retention of the C8-C9 double bond at the end of this process, depending on which isoform of *easA* catalysed the reaction, diverges the pathway towards the eventual formation of the fumigaclavines, dihydroergopeptides, the lysergic acid derivatives, or the more recently discovered cycloclavine (J. Z. Cheng et al., 2010a; Coyle et al., 2010; Jakubczyk et al., 2015; Wallwey, Matuschek, Xie, et al., 2010) (Figure 4-1).

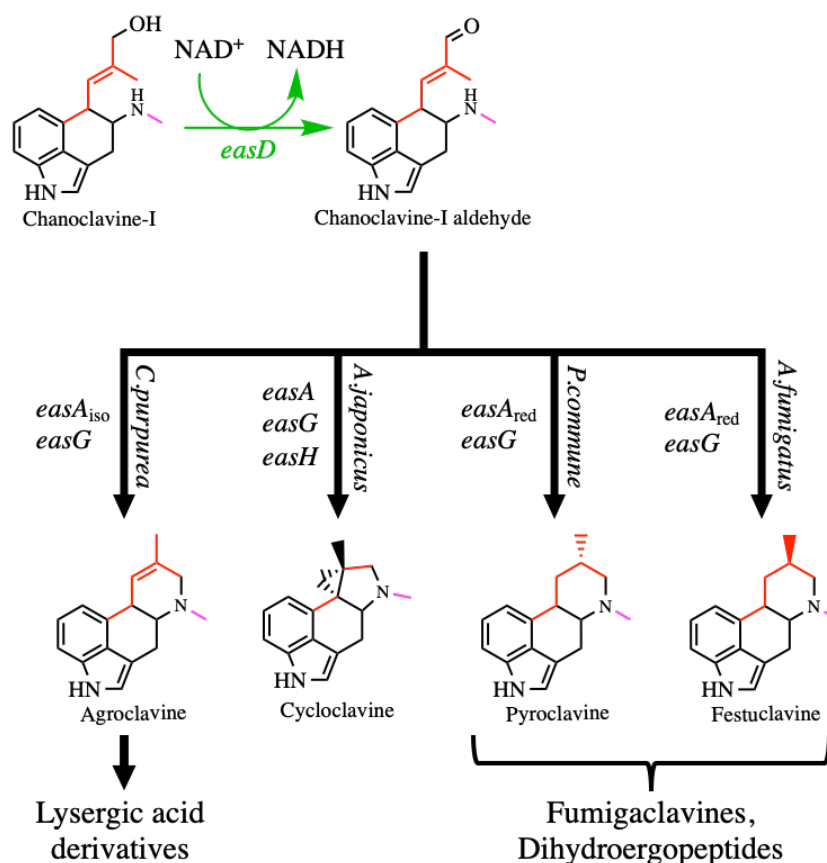


Figure 4-1: Reactions converting chanoclavine-I to the tetracyclic ergoclavines. Oxidation of the hydroxyl group on chanoclavine-I produces chanoclavine-I-aldehyde, which gets cyclized by different combinations of *easA* and *easG* from different ergot alkaloid producers to form the different clavine precursors of their respective ergot alkaloid profile.

The selection and incorporation of an isomerase variant for *easA* is therefore crucial to establish the successful heterologous production of *D*-lysergic acid in yeast. Earlier research into the mechanism of the ergoline D ring formation catalysed by *easA* has highlighted the importance of the tyrosine-178 residue in the *A. fumigatus* ortholog of *easA*, as well as its structurally corresponding phenylalanine-176 residue in the *N. lolii* ortholog (Figure 4-2) (J. Z. Cheng et al., 2010a). It was postulated that the hydroxyl group on tyrosine-178 provides the second proton that facilitates the complete reduction of the C8-C9 double bond. The lack of a hydroxyl group in phenylalanine prevents this from happening, allowing for rotation about C8-C9 bond before the hydride at

C9 is abstracted by FMN and the double bond reformed (Figure 4-2). This hypothesis was validated with the creation of the *N. lolii easA* F176Y mutant. This mutation switched its activity from being an isomerase to become a reductase. The *A. fumigatus easA* Y178F mutant, on the other hand, was unable to effect the corresponding switch to become an isomerase variant. This suggested that there were other unidentified factors that determined the precise nature of an *easA* ortholog. Nevertheless, this study pointed out an important feature that would help narrow the search for more isomerase variants that may be used to expand the pathway within the existing strain to produce agroclavine.

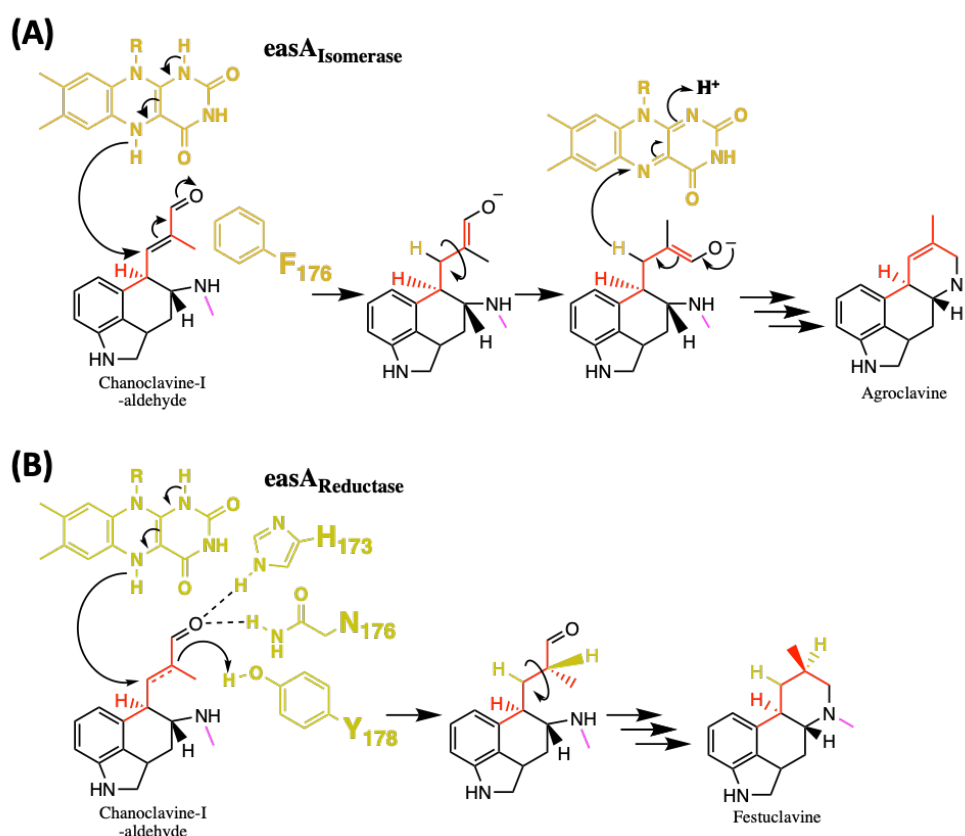


Figure 4-2: Proposed reaction mechanisms of *easA* for the (A) isomerization of the C8-C9 double bond in the formation of agroclavine; or, (B) reduction of the C8-C9 double bond in the formation of festuclavine or pyroclavine. The presence of the tyrosine at the 178 position is a key feature of the reductase variants of *easA*. Figure adapted from (J. Z. Cheng et al., 2010a, 2010b)

The sequence of a known isomerase variant of *easA* from *N. lolii* was used as the query to generate an SSN. While several *in vitro* biosynthetic studies have been performed on *easA*, they have predominantly been done using the reductase variant from *A. fumigatus* (Chilton, Ellis, & Lamb, 2014; Gerhards & Li, 2017; Matuschek, Wallwey, Wollinsky, Xie, & Li, 2012; Wallwey, Matuschek, Xie, et al., 2010). Furthermore, the isomerase variants characterized mostly within the source organism or a closely related species (Coyle et al., 2010; Floss et al., 1968; Robinson & Panaccione, 2014). This SSN served as the pool from which isomerase variants of *easA* orthologs were selected to be screened for functional expression in yeast. This chapter described the efforts made to identify an optimal variant of *easA* to extend the capabilities of the chanoclavine-I producing yeast to now produce agroclavine, and its subsequent improvement for enhanced stability.

4.2 MATERIALS AND METHODS

4.2.1 IDENTIFICATION AND CLONING OF *easD* AND *easG*

EasD from *A. fumigatus* and *easG* from *C. purpurea* have been reported to be successfully cloned, recombinantly expressed, and purified in *E. coli*; they were also demonstrated to be active in *in vitro* studies (Matuschek et al., 2011; Wallwey, Matuschek, & Li, 2010). These sequences were therefore selected, codon-optimized for yeast expression and synthesized onto pET15b vectors (Genescript). These were then cloned into the HCKan_O vector and assembled into an assortment of POTs, driven by various different promoters and at varying POT positions used in this study.

Table 4-1: Sequence details for the genes *easD* and *easG* used in the construction of an agroclavine producer

Gene	Source	Uniprot ID	Reference
<i>easD</i>	<i>Aspergillus fumigatus</i>	Q4WZ66	(Wallwey, Matuschek, & Li, 2010)
<i>easG</i>	<i>Claviceps purpurea</i>	M1WEN5	(Matuschek et al., 2011)

4.2.2 BIOINFORMATICS FOR THE IDENTIFICATION OF ISOMERASE VARIANTS

In the search for isomerase variants of *easA*, the sequence of *easA* from the *N. lolii* ortholog (J. Z. Cheng et al., 2010a; Robinson & Panaccione, 2014), was used as the input (Option A). Initial BLAST parameters were set to pull a maximum of 9000 sequences with a minimum alignment E-value of 5 (Figure 4-3A). The initial SSN generated from these sequences was limited to the sequences with an alignment length within 370 to 390 amino acids. Edges were initially specified as relationships with an E-value of 85, corresponding to 40% sequence identity (Figure 4-3B).

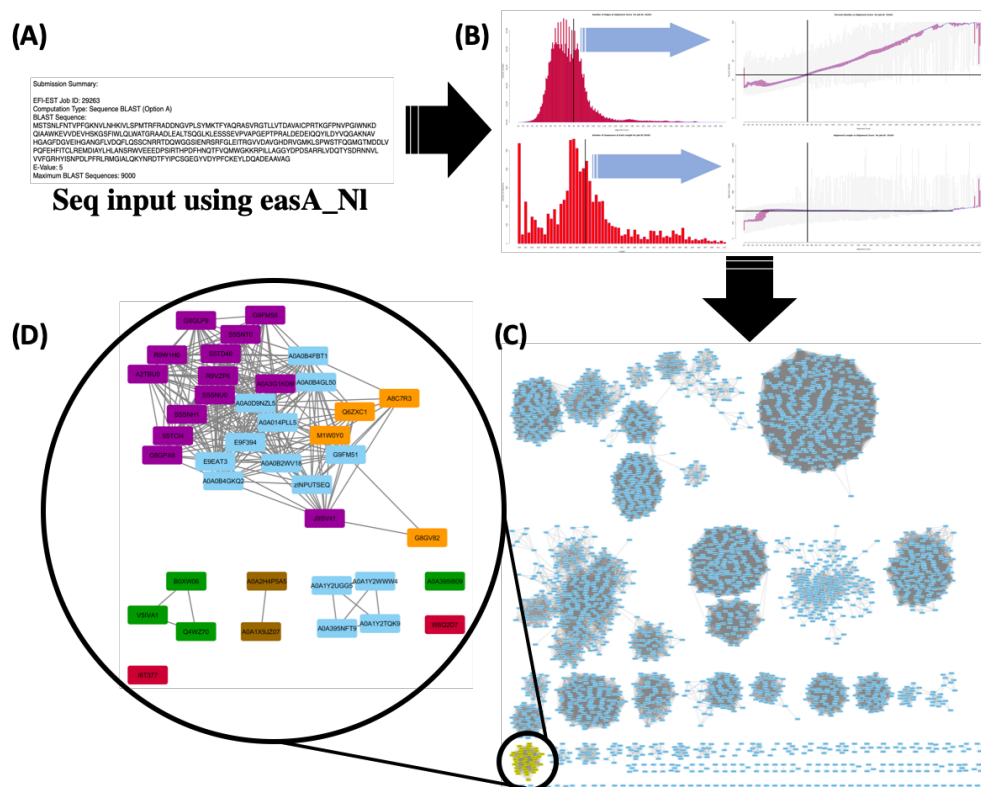


Figure 4-3: Workflow for the generation of an *easA* SSN. (A) Sequence input and initial BLAST parameters. (B) Demographics of the sequences returned from the initial BLAST. Selection of an alignment score corresponding to 40% sequence identity was used to define the initial threshold of an edge. Full graphs are shown in Appendix A4-5, 6, 7, 8 (C) SSN refined into isofunctional clusters at an alignment score of 130. (D) Expansion of the isofunctional cluster to delineate isomerase and reductase variants of *easA*.

This initial SSN returned a single massive cluster, which indicated that an edge threshold of 40% sequence identity was insufficient for isolating isofunctional clusters in this enzyme family. Increasing the E-value to 130, enabled the separation of the initial single cluster into more distinct clusters, from which a hypothetical *easA* isofunctional cluster was identified (Figure 4-3C). Further refinement on this cluster to an E-value of 180 allowed the delineation of clusters containing orthologs from known ergot alkaloid producers from the reductase and isomerase lineages (Figure 4-3D).

A sequence alignment of the orthologs found in the isomerase cluster was performed using the Clustal Omega program, accessed through the Uniprot

website. Alignments of the selected sequences and the *easA* ortholog from *A. fumigatus* are shown in Figure 4-6.

Three sequences from each representative family within the isomerase cluster were selected, and together with the *N. lolii* ortholog, were codon-optimized and synthesized into a pET15b vector (Table 4-2). These selected *easA* orthologs were then cloned into HCKan_O and used for the assembly of the POTs used in this study, following the same Golden Gate assembly protocol described in the earlier chapters.

Table 4-2: List of hypothetical isomerase variants of *easA* selected from the SSN

Abbreviated name	Organism	Uniprot ID
easA_Nl	<i>Neotyphodium lolii</i>	S5TCI4
easA_Pi	<i>Periglandular ipomoeae</i>	G9FM51
easE_Ec	<i>Epichloe coenophialia</i>	R9W1H0
easE_Cpur	<i>Claviceps purpurea</i>	M1W0Y0

4.2.3 SCREENING OF *easA* ORTHOLOGS

4.2.3.1 ASSEMBLY AND CONSTRUCTION OF PLASMID CONSTRUCTS

The pCKU-GAD constructs used to screen the selected *easA* orthologs were constructed by Golden Gate assembly as described in the earlier chapters. The three parts used were: *easG* on POT2, under the control of the ENO2 promoter and APL2 terminator; *easA* on POT4, under the control of the SSA1 promoter and YPT31 terminator; *easD* on POT5, under the control of the TDH3 promoter and YPR1 terminator. These were assembled into the pCKU-RFP acceptor vector, containing the low copy-number yeast origin of replication (CEN-ARS) and a URA3 selection marker.

The assembled and verified plasmid constructs were transformed into chemically competent YOCE competent cells, as described previously. Three

colonies per construct were selected and tested as described before. Spent media samples were prepared for LC-MS analysis as described previously.

4.2.3.2 PREPARATION OF AGROCLAVINE STANDARDS AND QUANTIFICATION

A stock solution of the agroclavine standard (Chiron) was prepared by dissolving 5 mg of the compound in 5 mL of LCMS grade ethanol and stored at -20 °C. A standard curve of agroclavine was generated by preparing various dilutions, ranging from 0.1 to 160 nM of agroclavine, using the spent media from the YOCE strain as the matrix. This covered the full linear range within the limit of detection. The media was prepared as described in chapter 3.

The amount of agroclavine produced by the tested strains was quantified by measuring the area under the curve for the peak response from the MS-MS ion transition of $[M+H]^+$ 239 $\rightarrow$ 208 m/z.

4.2.4 IMPROVING THE PRODUCTION OF AGROCLAVINE

4.2.4.1 ASSEMBLY AND CONSTRUCTION OF PLASMID CONSTRUCTS

The GACD constructs were assembled using the same *easA* and *easG* POTs as described previously with additional parts: *easC* on POT6, under the control of the TDH2 promoter and YIP5 terminator; *easD* on POT7, under the control of the TDH3 promoter and YPR1 terminator.

The GAFCD constructs were assembled in the same manner as the GACD constructs with the modifications: *easF* on POT6, under the control of the TEF2 promoter and PRX1 terminator; *easC* on POT8, under the control of the TDH2 promoter and YIP5 terminator; and *easD* on POT9, under the control of the TDH3 promoter and YIP5 terminator.

All constructs were assembled onto the pathway acceptor plasmid, pCKU-RFP, followed by the integration vector, pGAU-YPRC, using the modified YeastFab assembly pipeline described in chapter 2.

Genome integrated strains were created by linearizing the insert fragments from pGAU-YPRC and transformed into chemically competent YOCE cells as described earlier.

Three colonies from each verified construct were tested and prepared for LC-MS analysis as described earlier.

4.2.4.2 TESTING STRAIN STABILITY ACROSS MULTIPLE PROPAGATIONS

The YP-GAD/GACD/GAFCD strains were tested for stability across multiple propagations. The cells from the first round of biosynthesis were diluted in sterile deionised distilled water to an OD₆₀₀ of 2. From this suspension, 20 μ L was streaked out onto fresh SC-URA plates supplemented with 2% glucose and grown at 30 °C until single colonies emerged.

Three colonies per construct were then picked and patched onto a fresh SC-URA plate and further grown at 30 °C. These colonies were then grown as described earlier for the production of agroclavine, and similarly prepared for LC-MS analysis.

The relative proportions of the intermediates were estimated by the analysis of the peak area under the curve of their respective extracted ion chromatograms, and these values were normalized against the largest value for each construct.

4.2.4.3 CONSTRUCTION AND TESTING OF AGCM33B AND AGCM32B

AgcM33B and AgcM32B strains were envisioned to contain the genotype summarized in Table 4-3. The POTs required for the assembly of each constituent integration fragment were assembled from the existing library of parts, following the same Golden Gate assembly protocol described in chapter 2 (Table 4-4, 5, 6). A schematic representation of this assembly is illustrated in Appendix A4-1.

Table 4-3: Summarized genotype of the redesigned strains

Chromosome	Construct
Chr. XIII	YMRWΔ15:: <i>dmaW</i> ₅ / <i>easF</i> ₈ / <i>easE</i> ^{Aj} ₆ / <i>easC</i> ₃ / pCUP1-mKOK
Chr. XVI	YPRCΔ17:: <i>easG</i> ₁₁ / <i>easA</i> ^{Ec} ₁₄ / (<i>easF</i> ₆)/ <i>easC</i> ₅ / <i>easD</i> ₁₃
Chr. III	ARS308:: pRPL3-PDI1/ <i>easD</i> ₉ / pACT1-FAD1/ <i>easC</i> ₁₂ / <i>dmaW</i> ₇

The genes for PDI1 (chromosome III: 50221 – 48653) and FAD1 (chromosome IV: 373608 – 372688) were PCR-amplified from the genomic DNA of *S.cerevisiae* S288C and cloned into the HCKan_O entry-level vectors for “domestication” into the YeastFab pipeline as described in chapter 2.

Table 4-4: List of POTs assembled for the YMRWΔ15 integration fragment

POT number	Promoter	Gene	Terminator
2	TDH2	<i>dmaW</i>	ENT2
4	GPM1	<i>easF</i>	PRX1
6	TEF2	<i>easE</i> _Aj	RPS20
8	PGK1	<i>easC</i>	YIP5
9	CUP	mKOK	CBR1

Table 4-5: List of POTs assembled for the ARS308 integration fragment

POT number	Promoter	Gene	Terminator
2	RPL3	PDI1	VPS1
4	PDA1	<i>easD</i>	YPR1
6	ACT1	FAD1	EFM2
8	RPL15	<i>easC</i>	YIP5
9	TPI1	<i>dmaW</i>	ENT2

Table 4-6: List of POTs assembled for the YPRCΔ17 integration fragment

<u>AgcM33B</u>				<u>AgcM32B</u>			
POT number	Prom	Gene	Term	POT number	Prom	Gene	Term
2	RPL8B	<i>easG</i>	APL2	2	RPL8B	<i>easG</i>	APL2
4	SSA1	<i>easA_Ec</i>	YPT31	4	SSA1	<i>easA_Ec</i>	YPT31
6	TEF2	<i>easF</i>	PRX1	6	TDH2	<i>easC</i>	YIP5
8	TDH2	<i>easC</i>	YIP5	7	TDH3	<i>easD</i>	YPR1
9	TDH3	<i>easD</i>	YPR1				

Pathway segments were assembled first onto the pMKU-RFP acceptor vector and screened, after which they were transferred onto their respective integration vectors. Integration of the segments were performed sequentially, beginning with the YMRWΔ15 insert, followed by the ARS308 insert, and finally the YPRCΔ17 insert. At each iterative integration step, the transformants were screened by PCR and tested for functionality (Appendix A4-2), before they were cured of the URA3 marker. Thereafter, the strains were prepared as competent cells to move onto the next integration. These were all performed following the workflow and protocols described in chapters 2 and 3.

The completed and verified strains were tested by growing three biological replicates, using the same galactose induction protocol described earlier. A 1 mL aliquot of the spent media was filtered through a 0.2 μm PTFE syringe filter and analysed on the LC-MS.

The strains were also tested for stability and the relative amounts of intermediates accumulated were also performed as previously mentioned.

4.2.4.4 LC-MS QUANTIFICATION OF AGROCLAVINE PRODUCED

The agroclavine produced in the samples were quantified using a standard curve of agroclavine spiked into YOCE media as previously described.

4.2.5 VALIDATION OF AGROCLAVINE PRODUCTION USING ^{13}C -LABELLED TRYPTOPHAN

Stock solutions of tryptophan (Sigma) and ^{13}C -2-indole tryptophan (Sigma) were prepared by the dissolution of the powder in 20% (w/v) galactose to a final concentration of 10 mM, and filter-sterilized.

Tryptophan feeding was performed by culturing the strains as described earlier with a few differences. The prepared stock solutions containing 20% galactose, 10 mM tryptophan/ or ^{13}C -2-indole tryptophan were added to the cultures to a final concentration of 2% galactose, 1 mM tryptophan/ ^{13}C -2-indole tryptophan. Cultures were incubated and prepared for LC-MS analysis as previously mentioned.

4.2.6 LC-MS ANALYSIS OF SAMPLES

All analyses using LC-MS were performed using the Agilent 1290 Infinity LC System using an Agilent Zorbax Eclipse Plus C18 column, 2.1 x 50 mm dimensions, 1.8 μm particle size. Samples were run using the same method described earlier.

The acquisition parameters were set to acquire MS-MS data from an electrospray ionization source on positive mode. MS data acquisition was set to a mass range of 50 – 1000 m/z, at a scan rate of 3 spectra s^{-1} . MS-MS data

acquisition was set to targeted MS-MS mode for a mass range of 50 – 1000 m/z, MS-MS scan rate of 6 spectra s⁻¹, with a fixed collision energy of 16. Instrument parameters were set at; Gas temperature: 200 °C, Gas flow: 15 L min⁻¹, Nebulizer: 50 psig, Sheath Gas temperature: 350 °C, Sheath Gas flow: 10 L min⁻¹. Scan source parameters were set at VCap: 4000 V, Nozzle voltage: 0 V, Fragmentor: 110 V, Skimmer: 65 V, Octopole RF peak: 750 M.

Targeted masses for the samples screening the *easA* orthologs and testing strain designs were set to: 1) 239.1543 m/z, on a narrow isolation width (1.3 amu), collision energy of 15; 2) 257.1648 m/z, on a narrow isolation width (1.3 amu), collision energy of 15.

Targeted masses for the ¹³C-2-indole tryptophan feeding experiments were set to: 1) 239.1543 m/z, on a narrow isolation width (1.3 amu), collision energy of 15; 2) 240.1576 m/z, on a narrow isolation width (1.3 amu), collision energy of 15.

4.3 RESULTS AND DISCUSSION

4.3.1 ISOLATING ISOMERASES FROM REDUCTASES

Bioinformatics analyses on the protein sequence for *easA* showed it to be closely related to the Old Yellow Enzyme (OYE) family of FMN-binding oxidoreductases. This family of enzymes have been known to share a high degree of sequence similarity, with the largest source of sequence variation coming from the enzyme surface regions (Toogood, Gardiner, & Scrutton, 2010). Hence, it was unsurprising to find that the initial SSN generated, for an alignment score of 85, that corresponds to a sequence identity of 40%, could not differentiate the retrieved sequences (Figure 4-4A). Redefining an edge to an alignment score of 130, corresponding to a sequence identity of 60%, only then begins to fractionate the sequences into discernible clusters (Figure 4-4B).

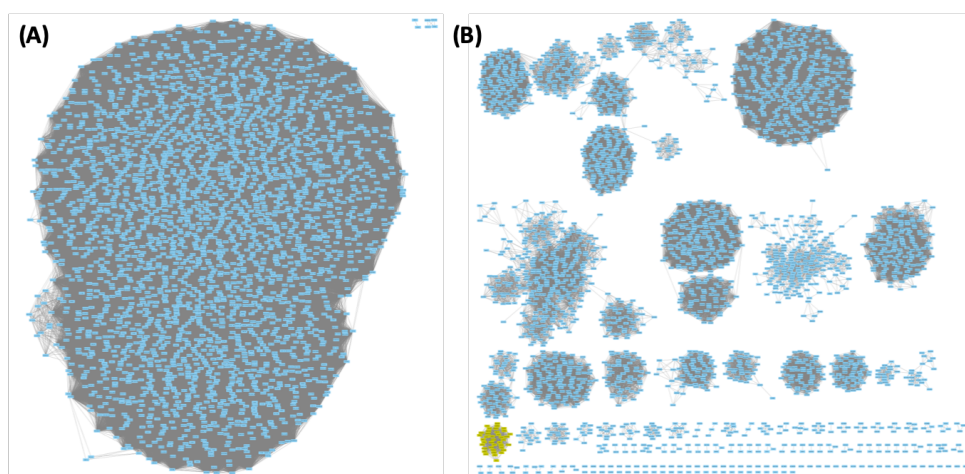


Figure 4-4: (A) Initial SSN generated for *easA* at an alignment score of 85, corresponding to 40% sequence identity; this was unable to distinguish isofunctional clusters. (B) Increasing the alignment score to 130, corresponding to 60% sequence identity allowed for the identification of isofunctional clusters. The hypothetical *easA* isofunctional cluster is highlighted in yellow.

From this SSN, the isofunctional cluster for *easA* enzymes could be identified from the sequences that were up to two degrees of separation. Further increasing the alignment score to 180, allowed the distinction of a tight cluster of sequences that contain several orthologs from known lysergic acid derivative

producers and numerous sub-clusters that were littered with sequences from fumigaclavine and dihydroergot alkaloid producers (Figure 4-5). While it would be possible to speculate the nature of the *easA* variant present within the gene cluster of an ergot alkaloid producing organism based on its ergot alkaloid product profile, this approach using an SSN, provided a more diverse pool of sequences from which a correctly predicted amino acid sequence may be found.

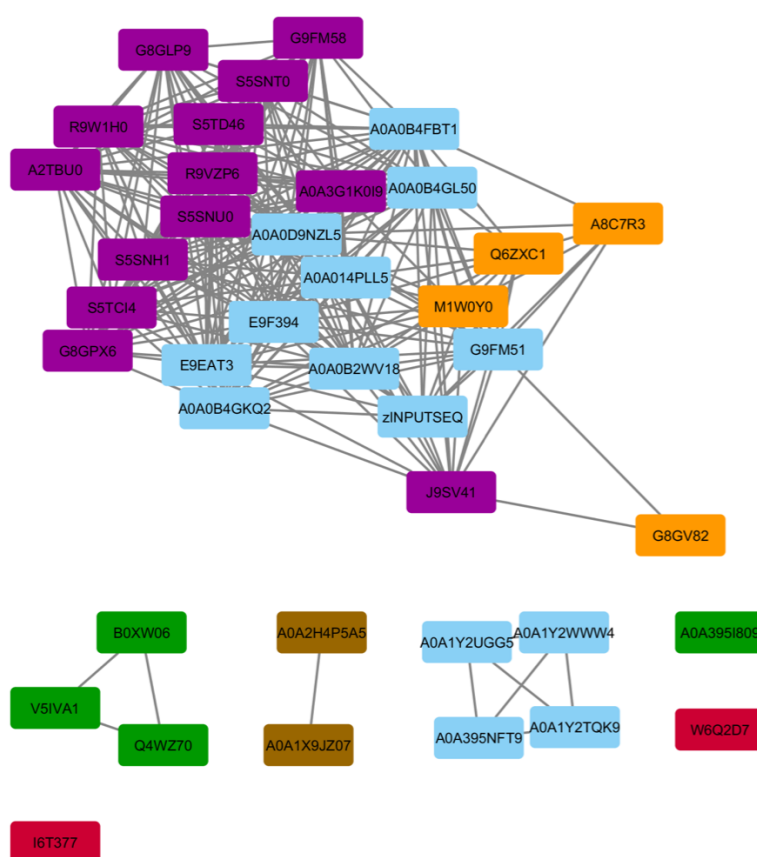


Figure 4-5: Expanded SSN for the isofunctional cluster of *easA*, obtained by defining an edge to an alignment score of larger than 180. Isomerase variants remain tightly clustered, while the reductase variants splinter off into several sub-clusters. Nodes in **purple denote sequences from the *Epichloe* family, nodes in **orange** denote sequences from the *Claviceps* family, nodes in **green** denote sequences from the *Aspergillus* family, nodes in **red** denote sequences from the *Penicillium* family, nodes in **brown** are for sequences from *Claviceps gigantea* and *Africana* (known producers of the dihydroergot alkaloids), and nodes in **blue** represent sequences from the other collective families not well characterized.**

This predicted isomerase cluster was aligned with the query sequence and the *A. fumigatus* reductase ortholog and they were confirmed to have the

characteristic phenylalanine residue at the 176 position (Figure 4-6). Four representative orthologs were then selected from the families of *easA* sequences present in the isomerase cluster to be screened.

```

S5TCI4 117  ANGFDLVSSSAVPVSPGEPTPRALSDDEINSYIGDFVQAAKNAVLEAGFDGVELHGANGF 176
G9FM51 117  SNGLELKSSSEVPVAPGEPTPRPLSEEEIQSYIFDYAQGAKNVHEAGFDGVEIHGANGF 176
R9W1H0 117  ANGFDLVSSSAVPVSPGEPTPRALSDDEINSYIGDFVQAAKNAVLEAGFDGVELHGANGF 176
M1W0Y0 117  SQGLKLESSSEVPVAPGEPTPRALDEDEIQQYILDYVQGAKNVHGAGFDGVEIHGANGF 176
Q4WZ70 121  D-MKDLISSAVPVEEKGPLRALTEDEIQQCIADFAQAARNAI-NAGFDGVEIHGANGY 178
          *  ***  ***      *  *  *  ::::  *  *  :  .  .  .  :  :  :  :  :  :  :  :  :  :  :  :

```

Figure 4-6: Sequence alignment of the selected *easA* orthologs with the *N. lolii* ortholog (M1S0Y0) and *A. fumigatus* ortholog (Q4WZ70) for the catalytic region containing the important phenylalanine/ tyrosine residue (highlighted in red).

4.3.2 SCREENING FOR ISOMERASES

The conversion of chanoclavine-I to agroclavine involves its progression first into chanoclavine-I-aldehyde catalysed by *easD*, before its cyclization into the iminium intermediate catalysed by *easA* and subsequent reduction to agroclavine catalysed by *easG*. As *easD* and *easG* have been previously shown to be very responsive to heterologous expression (J. Z. Cheng et al., 2010a; Wallwey, Matuschek, & Li, 2010), unlike the previous chapter, there was no pressing need to extensively test them in isolation from *easA*.

Test strains for each of the four *easA* orthologs (pCKU-GAD), with all three enzymes under the control of different constitutive promoters, were created to screen their ensuing product profiles for agroclavine and/or festuclavine (Figure 4-7A). The spent media samples from all the strains containing different *easA* orthologs eluted a compound with an $[M+H]^+$ of 239 m/z at around the same time as the standard of agroclavine spiked into the media of the YOCE strain (Figure 4-7B) (Table 4-7). The MS-MS fragmentation patterns between the agroclavine standard and the eluted compounds from the pCKU-GAD strains were found to be the same (Figure 4-7D, E). This confirmed the production of agroclavine (Table 4-8).

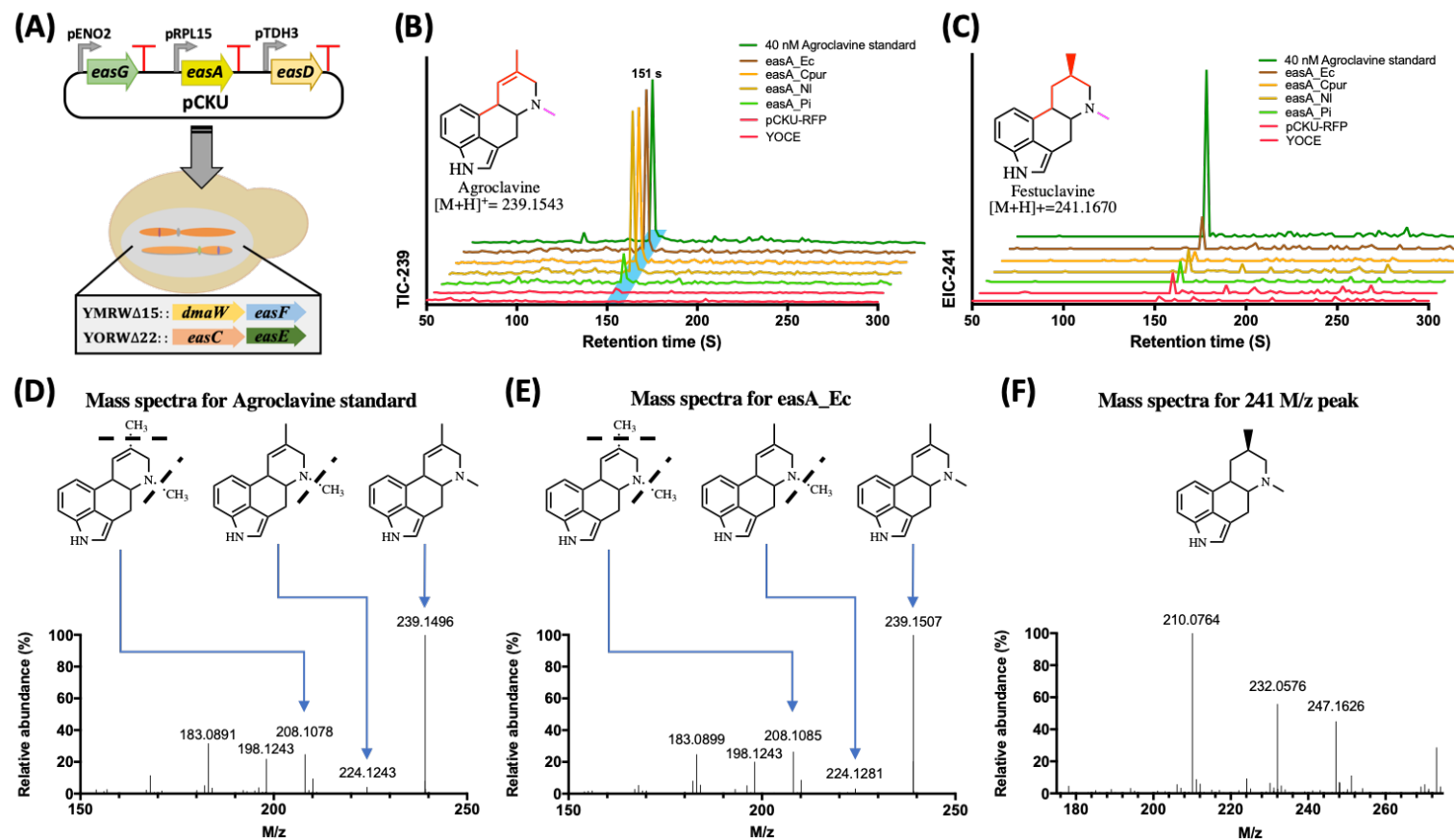


Figure 4-7: Screening of *easA* orthologs selected. (A) Illustration of the construct assembled for the screening experiments. (B) Total ion chromatogram from the selected ion monitoring of all fragments produced from precursors with an $[M+H]^+$ of 239 *m/z*. (C) Total ion chromatogram from the selected ion monitoring of all fragments produced from precursors with an $[M+H]^+$ of 241 *m/z*. (D, E) Mass spectra for the peaks eluting at 151 s from the agroclavine standard and sample of the *E. coenophialia* ortholog. (F) Mass spectra for the 241 *m/z* peak.

Table 4-7: Exact mass and retention time for the agroclavine standard and agroclavine produced

Construct	Calculated m/z	Observed m/z	Mass error δ (ppm)	Retention time (s)
<i>Agroclavine standard</i>	239.1543	239.1496	-19.65	150.8
easA_Pi	239.1543	239.1486	-23.83	151.3
easA_Nl	239.1543	239.1505	-15.89	151.4
easA_Cpur	239.1543	239.1502	-17.14	151.5
easA_Ec	239.1543	239.1507	-15.05	151.3

Table 4-8: Exact masses of the fragment ions from agroclavine produced

Construct	224 m/z fragment			208 m/z fragment			183 m/z fragment		
	Standard m/z	Observed m/z	Mass error δ (ppm)	Standard m/z	Observed m/z	Mass error δ (ppm)	Standard m/z	Observed m/z	Mass error δ (ppm)
easA_Pi	224.1243	224.1244	0.45	208.1078	208.1101	11.05	183.0891	183.0940	26.76
easA_Nl	224.1243	224.1272	12.94	208.1078	208.1090	5.77	183.0891	183.0893	1.09
easA_Cpur	224.1243	224.1260	7.59	208.1078	208.1084	2.88	183.0891	183.0890	-0.55
easA_Ec	224.1243	224.1281	16.96	208.1078	208.1085	3.36	183.0891	183.0899	4.37

Surprisingly, on the extracted ion chromatograms for the $[M+H]^+$ of 241 m/z, a small peak close to the retention time of agroclavine was found in all samples. This same peak was most pronounced in the agroclavine standard, and analysis of its mass spectra revealed that it likely be the fragments of agroclavine adducts. This indicated that the chosen orthologs of *easA* exclusively produced agroclavine and not festuclavine nor its (9S)-stereoisomer, pyroclavine.

A calibration curve of agroclavine was also created to quantify the amounts produced by these strains (Figure 4-8A). It was found that amongst the four tested orthologs, the *P. ipomoeae* ortholog was barely functional. The amount of agroclavine is almost at the lowest limit of detection by LC-MS. The other three orthologs were found to perform marginally better, producing agroclavine approximately in the range of 2.8-3.1 $\mu\text{g/L}$ (Figure 4-8B).

Putting it all together, despite the low levels of agroclavine being produced across all the orthologs tested, this is the first successful demonstration for the complete biosynthesis of agroclavine in an entirely heterologous host.

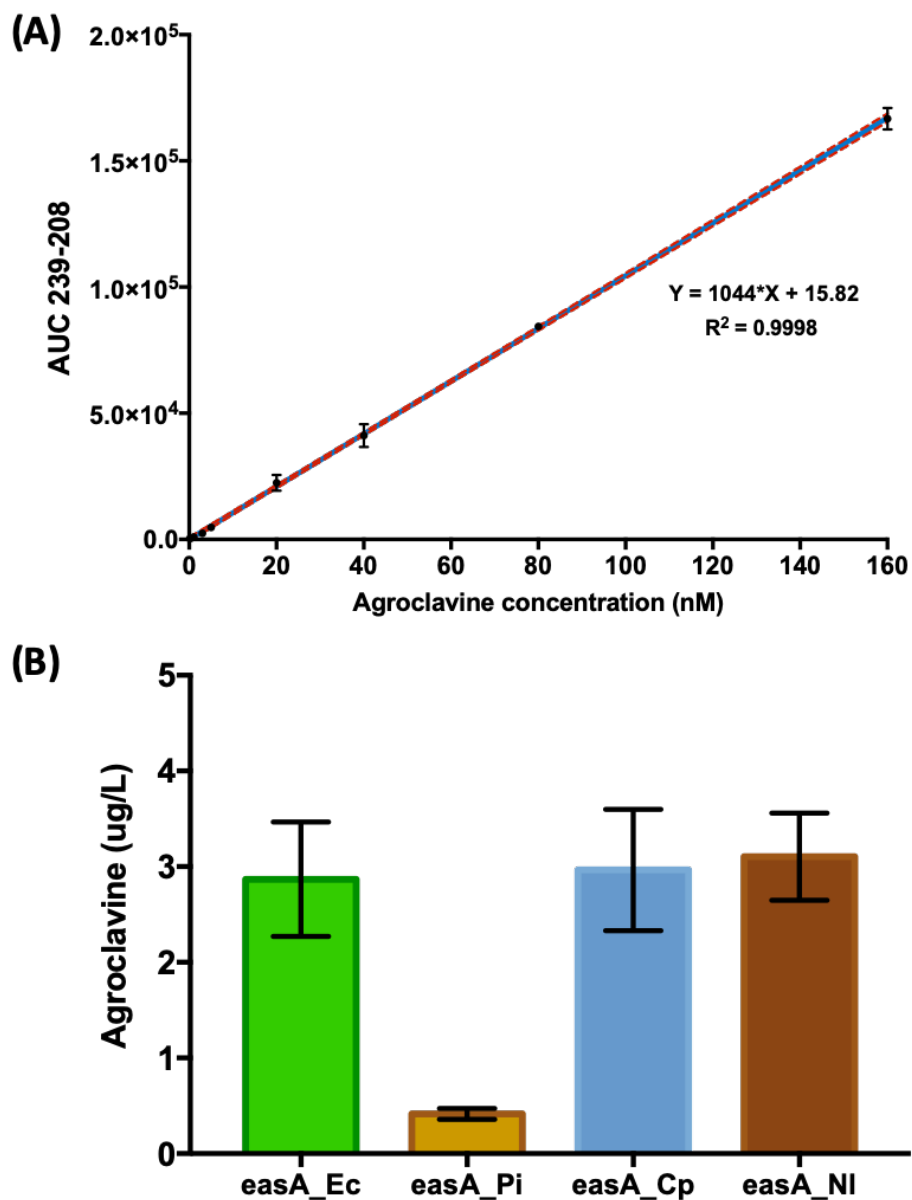


Figure 4-8: Quantification of the agroclavine produced by the *easA* orthologs. (A) Agroclavine standard curve established in YOCE media. Peak area response for the ion transition of 239 to 208 m/z plotted against spiked agroclavine concentrations, over a range of 0.1 – 160 nM. **(B)** Relative amounts of agroclavine produced in test strains to the standard curve, measured by the peak area response for the ion transition of 239 to 208 m/z. Error bars represent standard deviation from three biological replicates.

4.3.3 IMPROVING THE PRODUCTION OF AGROCLAVINE

4.3.3.1 WIDENING THE PATHWAY'S BOTTLENECKS

Armed with the knowledge that the selected sequences to complete the pathway to agroclavine are functional in yeast, the next goal was to improve the production yields sufficiently before completing the pathway to *D*-lysergic acid. This is exceptionally important because of the extremely low yields observed. Across all chemical transformations, whether biologically-catalysed or not, the efficiency of turnover at each reaction is rarely, if ever, a hundred percent. It is therefore not uncommon to observe the gradual decline of product over each successive intermediate. In terms of the yields measured, it is rather concerning that the detected amounts of agroclavine are so close to its limit of detection. Any further decline in signal over the subsequent reaction to produce *D*-lysergic acid would likely be within the range of the background noise and rendered undetectable.

Based on the information obtained from the previous chapter, the reaction catalysed by *easC/easE* is likely to be a choke point along the pathway. Furthermore, apart from buffering the media, the methylation of the amine group catalysed by *easF* is another way to prevent the diversion of metabolic flux of DMAT into DMA-MTCA. Two additional designs containing either an extra copy of *easC* (YP-GACD) or both *easC* and *easF* (YP-GAFCD) were built and integrated on the yeast genome, in an attempt to mitigate these identified limiting factors (Figure 4-9A).

Curiously, simply having the GAD constructs integrated onto the yeast genome improved agroclavine production around seven-fold to 21.7-28.1 $\mu\text{g/L}$

(Figure 4-9C), even though the pCKU vector used previously was a low copy number vector and should have been representative of a genomic construct.

Having an additional copy of *easC* driven by the TDH2 promoter on top of the GAL10 driven one produced the most improvement; especially for the GACD construct with the *easA* ortholog from *E. coenophialia*. This construct showed an increase of almost 20-fold, producing the equivalent of 55.2 $\mu\text{g/L}$ agroclavine (Figure 4-9B, C).

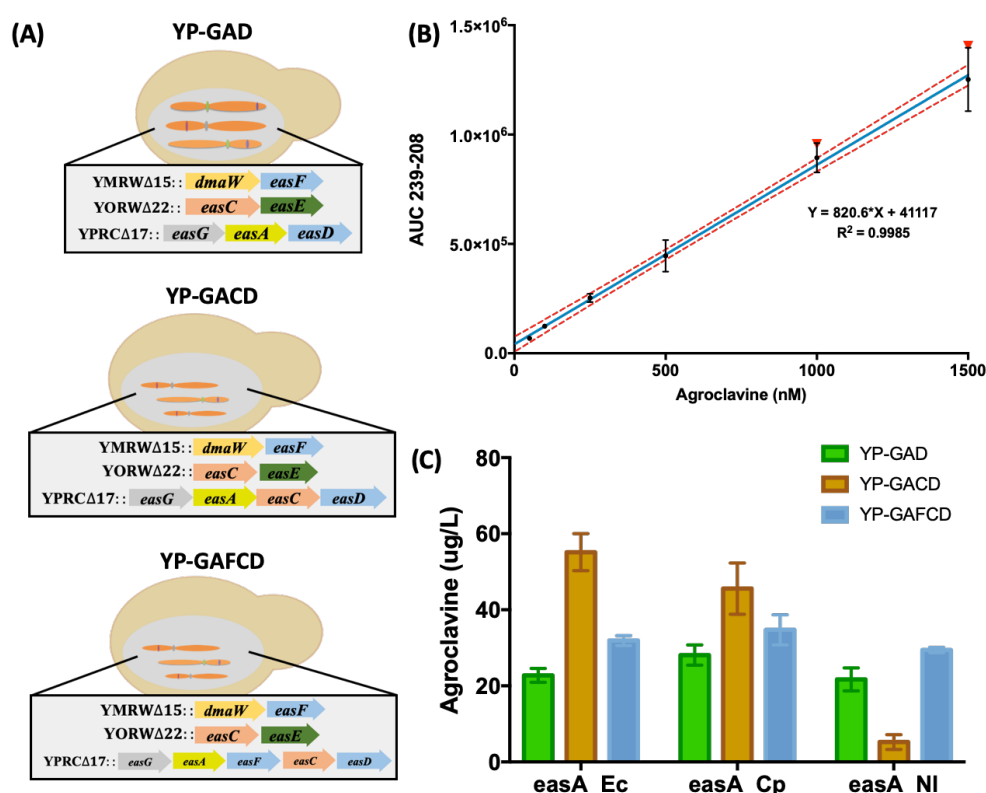


Figure 4-9: Screening alternative segment designs. (A) Illustration of GAD, GACD, and GAFCD constructs tested. **(B)** Agroclavine standard curve established in YOCE media. Peak area response for the ion transition of 239 to 208 m/z plotted against spiked agroclavine concentrations, over a range of 0.1 – 1.5 μM . **(C)** Relative amounts of agroclavine produced to the standard curve, measured by the peak area response for the ion transition of 239 to 208 m/z. Error bars represent standard deviation from three biological replicates.

The improvement was not observed in any construct containing the *N. lolii* ortholog. No improvement was also observed from the addition of a second copy of *easF* for all *easA* orthologs. Instead, it was noticed that while all other

strains grew to saturation, all GAFCD constructs as well as the GACD construct with the *N. lolii* ortholog appeared to stagnate at an OD₆₀₀ around 0.8-1, upon induction by galactose. This suggested that the poor performance observed was likely due to the lack of biomass accumulation. The decreased cell fitness observed in these strains was unusual, considering how they are more pronounced with the *N. lolii* orthologs.

While the new designs showed a very encouraging improvement in agroclavine yields, the peculiarities around the diminishing cell fitness raised another cause for concern. This prompted a series of repeats, to determine if further propagation of the other strains would show the same decline in cell fitness, as well as to re-introduce the GAD/GACD/GAFCD constructs into the YOCE strain to find out if it was a technical issue or something else altogether.

4.3.3.2 INSTABILITY IN THE INITIAL DESIGNS

From the repeats performed, it was noticed that all the constructs tested eventually displayed the same growth retardation as the *N. lolii* constructs. Furthermore, analysis of the AUC for each intermediate from their respective peaks on the extracted ion chromatograms showed the same relative decline in chanoclavine-I and agroclavine across multiple propagations. This indicated that all the constructs created were gradually losing the ability to produce chanoclavine-I and agroclavine, and started accumulating DMAT and 4DMA from the second propagation onwards (Figure 4-10B). The re-introduction of the *easC/easE* cassettes to these strains, or the re-creation of the YP-GAD/GACD/GAFCD strains from fresh YOCE competent cells, restored their ability to produce chanoclavine-I and agroclavine.

These findings suggested that the expression of the genes responsible for the production of chanoclavine-I and beyond are likely being selected against. While agroclavine has been noted for having cytotoxic effects to a few cancer cell lines and found to be mildly antimicrobial, it was also shown to not be toxic to another yeast strain, *Saccharomyces uvarum*. Therefore, it is unlikely to be inhibiting the growth of the closely related *S. cerevisiae* (Eich & Pertz, 1999; Mrusek et al., 2015). Instead, it is believed that the decline of cell fitness and therefore selection pressure experienced by the agroclavine producing strains are more likely a function of the metabolic burden exerted by the extensive heterologous pathway driven by the suite of strong promoters that were introduced (Carrier, Jones, & Keasling, 1998; Colletti et al., 2011; Z.-F. Hu et al., 2019).

Given the specific loss of production from chanoclavine-I, it was postulated that the selection pressure was likely targeting the *easC*_{GAL10}-*easE*_{GAL1} cassette for modification or deactivation. The GAL10 and GAL1 promoters are effectively the same sequence running in opposing directions; this was an oversight in the earlier design (Figure 4-10A). This allowed the endogenous yeast recombination machinery to easily flip the *easC* cassette or possibly form a loop to delete the region. With a sequence homology length of 332 bp, it is within the documented capability of the endogenous yeast recombination machinery (Hohnholz & Achstetter, 2019; Hua, Qiu, Chan, Zhu, & Luo, 1997).

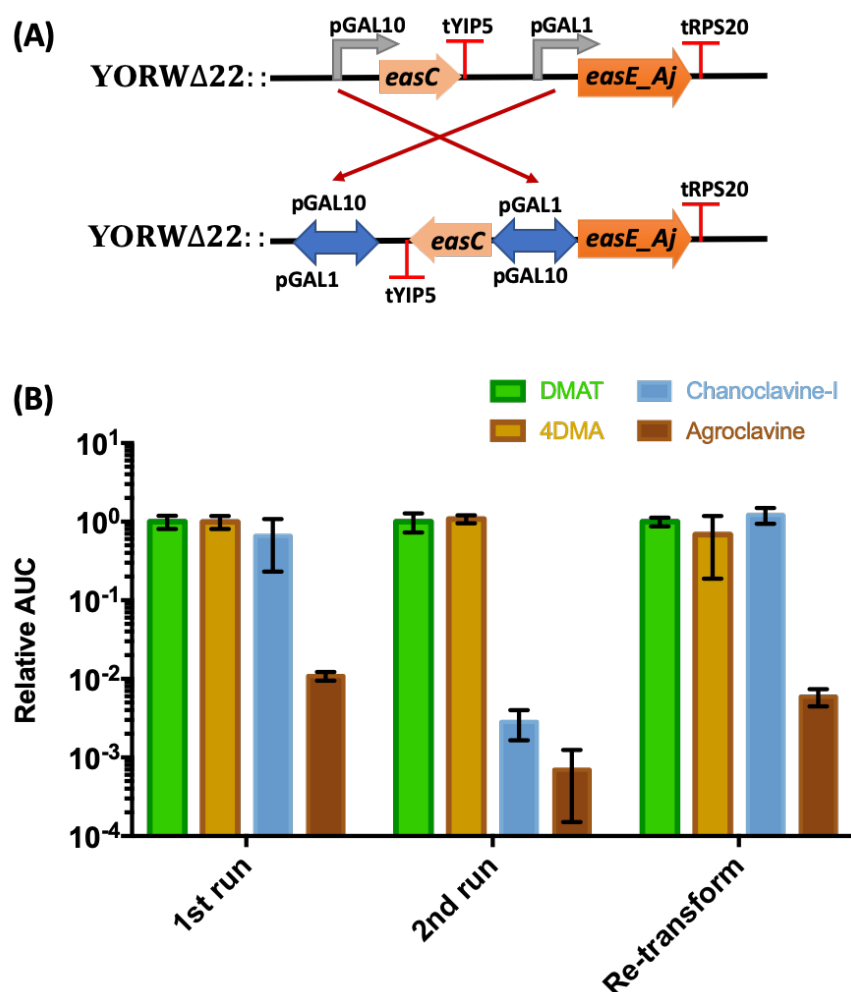


Figure 4-10: (A) Illustration of the possible inversion occurring on the *easC/easE* cassette as a result of the homology between GAL1 and GAL10 promoter sequences. (B) Estimation of the relative amounts of intermediates produced by the GA^{*E.coenophilia*}CD strain from the peak area response of their respective extracted ion chromatograms at 273, 287, 257, and 239 m/z. AUC values normalized to the largest mean value of each sample. Error bars represent the standard deviation calculated from three biological replicates.

In order to circumvent this issue as well as to prepare for further expansions to the pathway, an overhaul on the strain design was done. In this next design iteration, the *easA_Ec* ortholog was selected to trial a design similar to the one used in the cycloclavine producing yeast (Jakubczyk et al., 2015). This design sought to reallocate the stronger promoters to the difficult enzymes, such as *easE*, and maintain the expression levels of the other early pathway enzymes with multiple copies of expression cassettes driven by weaker promoters. At the same time, the re-design included extra copies of the genes

FAD1 and PDI1, which were cloned from the yeast genome to aid protein folding and the production of the co-factor, FAD, for *easE* (Nielsen et al., 2014) (Table 4-3).

When tested, these new strains (AgcM33B, and AgcM32B illustrated in Figure 4-11A) were found to produce the calibration curve equivalent of 35.7 and 40.2 $\mu\text{g/L}$ of agroclavine respectively (Figure 4-11B). These new strains, while found to produce marginally less agroclavine than the YP-GACD constructs, were more importantly no longer showing the same relative decline in chanoclavine-I and agroclavine production (Figure 4-11D). Instead, the relative peak area responses against 4DMA showed a shift in proportions of the detectable intermediates towards agroclavine downstream in the pathway. This phenomenon is likely because the host cells have accumulated larger quantities of each enzyme along the pathway on the second “passage” and more rapidly reaches this observed equilibrium of metabolic flux across the pathway. The equilibrium state is likely due to the accumulation of agroclavine to levels where its inhibitory effects on *dmaW* become apparent (L.-J. Cheng, Robbers, & Floss, 1980).

Having demonstrated the stability of the new designs, the two new strains (AgcM33B and AgcM32B), should now be more adequately to serve as the next stepping stone towards a heterologous *D*-lysergic acid producer.

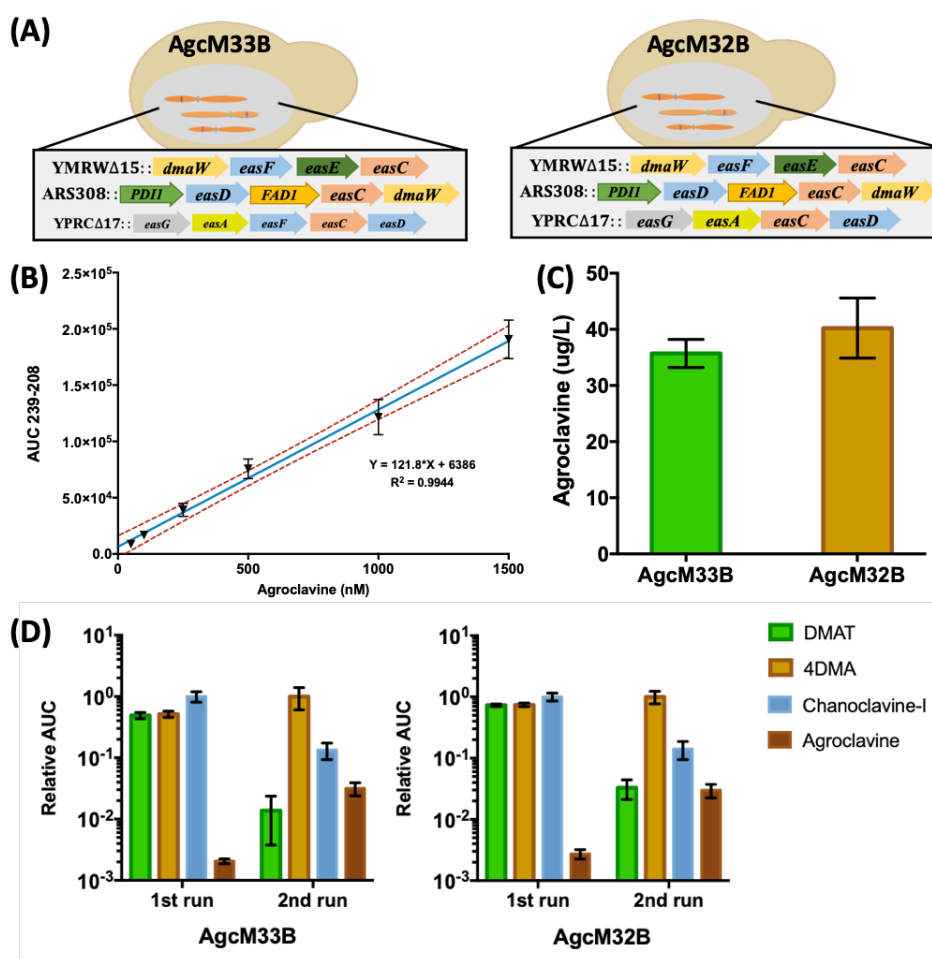


Figure 4-11: Testing the production yields and stability of the new strains AgcM33B and AgcM32B. (A) Diagrammatic representation of the strains created. (B) Agroclavine standard curve established in YOCE media. Peak area response for the ion transition of 239 to 208 m/z plotted against spiked agroclavine concentrations, over a range of 0.1 – 1.5 μ M. (C) Relative amounts of agroclavine produced to the standard curve, measured by the peak area response for the ion transition of 239 to 208 m/z. Error bars represent standard deviation from three biological replicates. (D) Estimation of the relative amounts of intermediates produced by the AgcM33B and AgcM32B strains from the peak area response of their respective extracted ion chromatograms at 273, 287, 257, and 239 m/z. AUC values normalized to the largest mean value of each sample. Error bars represent the standard deviation calculated from three biological replicates.

4.3.4 VALIDATION OF AGROCLAVINE PRODUCTION WITH ¹³C-LABELLED TRYPTOPHAN

In the earlier experiments, the production of agroclavine was validated by the comparison of its retention time, mass, and mass fragmentation patterns against a commercially-obtained standard. In the spirit of scientific rigour, the AgcM33B strain was fed with ¹³C-2-indole tryptophan (¹³C-W) to track the

production of ^{13}C -labelled intermediates along the reconstituted pathway, ending in ^{13}C -2-agroclavine (^{13}C -Agc). This served as an independent method for validating the production of agroclavine (Figure 4-12A).

The ion transitions for the precursors with an $[\text{M}+\text{H}]^+$ of 239 m/z replicated the earlier results, which demonstrated the elution of agroclavine at around 150 s, for both cultures fed with 1 mM tryptophan or ^{13}C -W (Figure 4-12B). This is as expected since the strain AgcM33B is capable of endogenously producing its own tryptophan. Furthermore, the SC media used for its culture is supplemented with unlabelled tryptophan. Analysis of the precursor ions with an $[\text{M}+\text{H}]^+$ of 240 m/z, showed the same peak eluting only from the culture fed with ^{13}C -W (Figure 4-12C). A comparison of the mass spectra from the two cultures illustrated a distinct shift of +1 m/z in the mass corresponding to agroclavine and all of its fragment ions, to levels well above the expected isotopic abundance (Figure 4-12D, E) (Table 4-9).

This experiment demonstrated the metabolism of ^{13}C -W into the pathway to produce ^{13}C -Agc; ergo the first successful, complete heterologous reconstitution of the pathway to agroclavine in yeast.

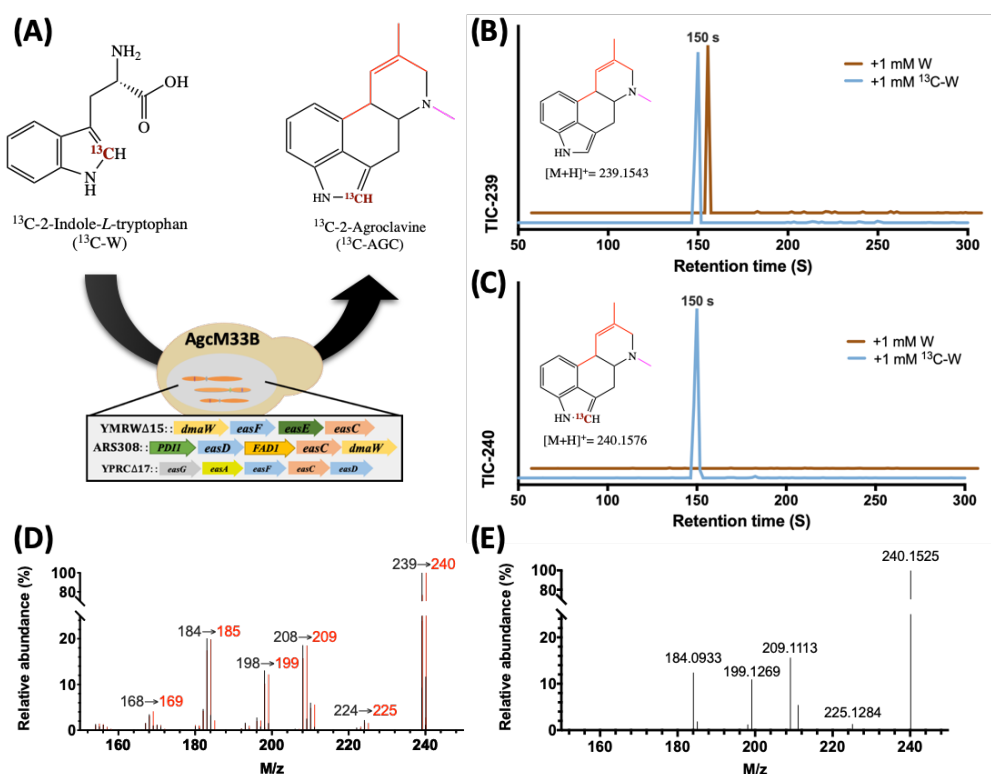


Figure 4-12: Validation of agroclavine production using ^{13}C -labelled tryptophan. (A) Cartoon representation of the experiment. (B) TIC from the selected ion monitoring of precursors with an $[M+H]^+$ of 239 m/z. (C) TIC from the selected ion monitoring of fragments precursors $[M+H]^+$ of 240 m/z. (D) Overlay of the mass spectra for the peaks at 150 s. Spectral lines in black represent the masses observed from tryptophan-fed samples. Spectral lines in red represent the masses observed from the culture fed with ^{13}C -W. (E) Difference in mass spectra between cultures fed with tryptophan or ^{13}C -W.

Table 4-9: Exact masses of ^{13}C -2-agroclavine produced and its fragment ions

Detected [^{13}C -M+H] $^+$ ions	Calculated m/z	Observed m/z	Mass error δ (ppm)
^{13}C -2-Agroclavine	240.1576	240.1525	-21.24
224→225	225.1276	225.1284	3.55
210→211	211.1272	211.1268	-1.89
208→209	209.1111	209.1113	0.96
198→199	199.1276	199.1269	-3.52
184→185	184.0924	184.0933	4.89

[illegible]

5.1 INTRODUCTION

The last step on the biosynthesis of *D*-lysergic acid (DLA) involves a two-step oxidation of the C8 methyl group in agroclavine and the migration of the double bond from the C8-C9 position to the C9-C10 position (Figure 5-1A). In 2006, Haarmann and colleagues identified the CYP450 monooxygenase, *cloA*, that catalysed this reaction (Haarmann et al., 2006). While the reactions required for the conversion of agroclavine to DLA can be deduced from their structural differences, not much is known about the sequence of events that lead to its formation, nor the precise mechanism by which this happens.

Previous studies showed that the reaction requires the presence of NADPH and molecular oxygen (Maier, Schumann, & Gröger, 1988; Quigley & Floss, 1981). Furthermore, corresponding homologous sequences have been identified in the gene clusters of all organisms producing oxidized tetracyclic clavine derivatives. From these organisms, different derivatives of agroclavine that correspond varying degrees of oxidation, have been isolated (Figure 5-1A) (Kozlovsky, 1999; Kren & Cvak, 1999). Elymoclavine and DLA have been isolated from numerous ergot fungi from the *Clavicipitaceae* family, where most of the research into the ergot alkaloid biosynthetic pathway was conducted on. Strains of *C. purpurea* with the proposed *cloA* gene knocked out were found to accumulate elymoclavine and agroclavine (Haarmann et al., 2006). These findings added credence to the school of thought that the conversion from agroclavine to DLA progressed stepwise through elymoclavine. Together with the finding that microsomal fractions of *C. purpurea* PCCE1 strains produced paspalic acid from elymoclavine, and that paspalic acid was predominantly produced in the closely related *Claviceps paspali* (Kobel, Schreier, &

Rutschmann, 1964), it was postulated that the reaction progresses from agroclavine to elymoclavine, followed by paspalic acid, before being spontaneously isomerized to DLA. Each individual oxidation step was proposed to be catalysed by a different oxidase, and that the isomerization of the C8-C9 double bond to the C9-C10 position happened spontaneously.

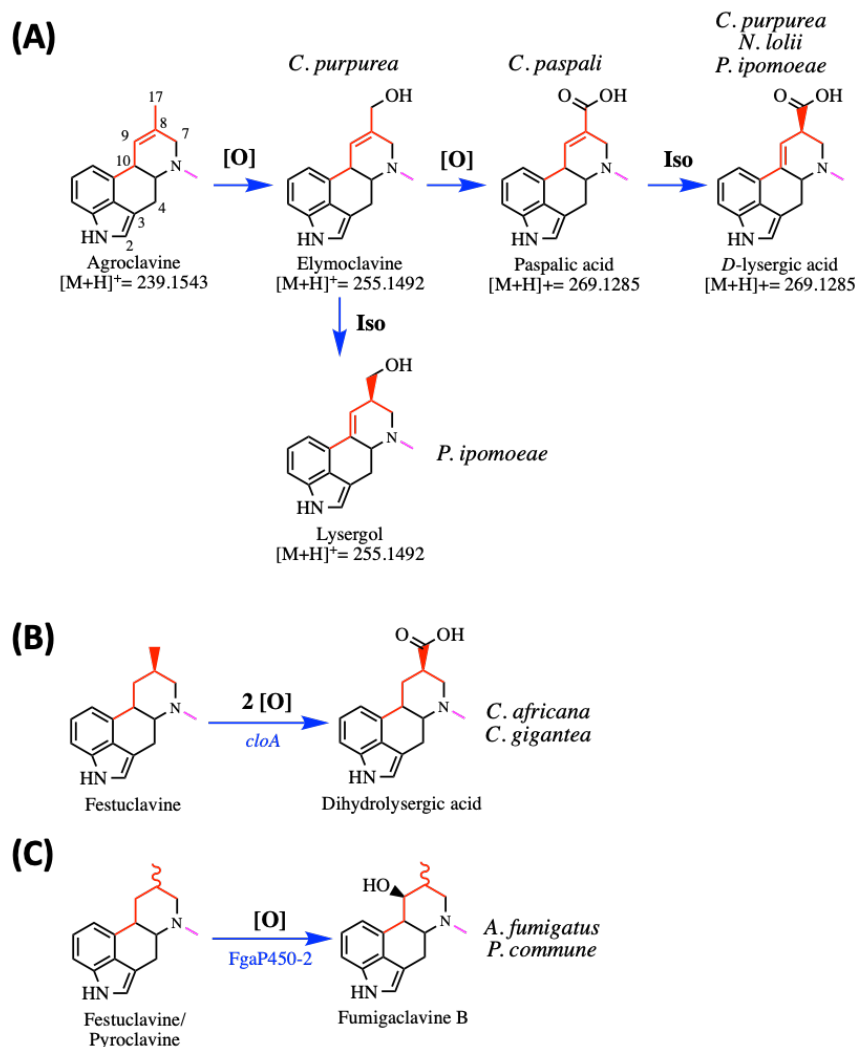


Figure 5-1: Reactions oxidizing the clavine alkaloids in the late ergot alkaloid pathway. (A) Hypothesized sequence of reactions catalysing the conversion of agroclavine to DLA and its derivatives in ergopeptine producing fungi. The isomerization of the C8-C9 double bond to produce DLA or lysergol could either occur catalytically or spontaneously in the presence of a base. (B) Reaction catalysed by the ortholog of *cloA* in the dihydroergot alkaloid producers. (C) Hydroxylation of festuclavine/ pyroclavine in the biosynthesis of the fumigaclavines.

Recent studies have found that when the *cloA* ortholog from *Epichloe* sp. Lp1 was heterologously expressed in *A. fumigatus* with the corresponding *easA*_{isomerase} gene, no production of elymoclavine or paspalic acid was detected (Robinson & Panaccione, 2014). Another study found that homologous sequences of *cloA* found in *Claviceps africana* and *Claviceps gigantea* are responsible for the double oxidation of festuclavine into dihydrolysergic acid (Figure 5-1B) (Arnold & Panaccione, 2017). Given these new findings, it is possible that sequence variations between the orthologs from different lineages or ergot alkaloid producers might be responsible for the varying extent of oxidation performed on the tetracyclic clavine products (Figure 5-1A, B, C). Certain variations between the *cloA* orthologs might favour the termination of the reaction at different degrees of oxidation, as well as the possibility for the migration of the C8-C9 double bond. The possible combinations of either a single or double oxidation, coupled with the option of the migration of the C8-C9 double bond to the C9-C10 position, would account for the production of elymoclavine (single oxidation, no isomerization), lysergol (single oxidation with isomerization), paspalic acid (double oxidation, no isomerization), and DLA (double oxidation with isomerization).

The characterization of the different *cloA* orthologs and their resulting product profiles is one aspect of a colleague's research. In his work, an SSN was generated to attempt to identify the different lineages of *cloA* that might account for the variability in the oxidized tetracyclic clavine profiles of the known ergot alkaloid-producing organism (Figure 5-2A). From the hypothetical clavine oxidase supercluster, the sequences could be further fractionated into several clusters that appeared to distinguish between the CYP450s responsible

for the oxidation of the different tetracyclic clavines (Figure 5-2B). Further refinement of the “C17-oxidase” cluster allowed for the separation of *cloA* orthologs from the *Epichloe*, *Claviceps*, *Metarhizium* families, the unusual variants from *C. africana* and *gigantea*, and lysergol-producing *P. ipomoeae*. This roughly corroborated with the findings for an exclusively DLA-producing group (*Epichloe*), elymoclavine and DLA-producing group (*C. purpurea*), the dihydrolysergic acid producing group (*C. Africana* and *gigantea*), and lysergol producing (*P. ipomoeae*) (Figure 5-2C). From four of the most dissimilar sub-clusters (Figure 5-3), 15 *cloA* orthologs were selected and characterized for their expression in yeast and production of elymoclavine, lysergol, paspalic acid, or DLA. Of these 15 sequences, five orthologs were found, when fed with agroclavine, to produce a compound with an $[M+H]^+$ of 269 m/z, corresponding to either paspalic acid or DLA.

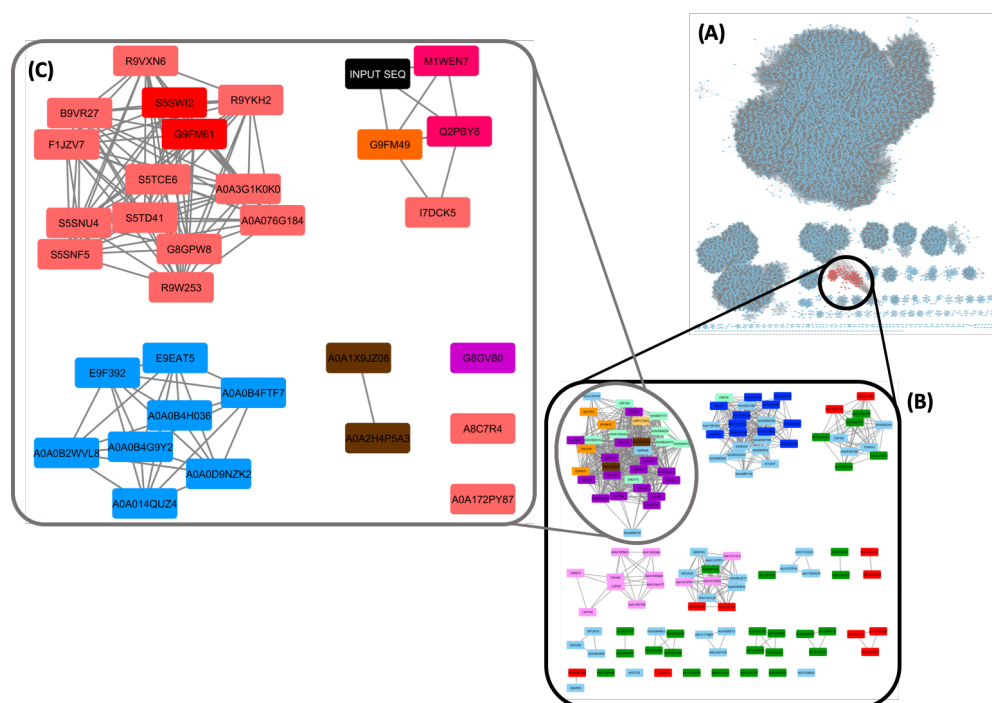


Figure 5-2: Sequence similarity networks of *cloA*. (A) Initial SSN covering the superfamily of cytochrome P450 monooxygenases. Hypothetical “*cloA*” supercluster coloured in orange. (B) Expansion of the “*cloA*” supercluster, showing the divergence of *cloA* sequences between the different families of ergot

alkaloid producers. (C) Further expansion of the cluster of *cloA* sequences from organisms identified to perform an ergoline C17 oxidation.

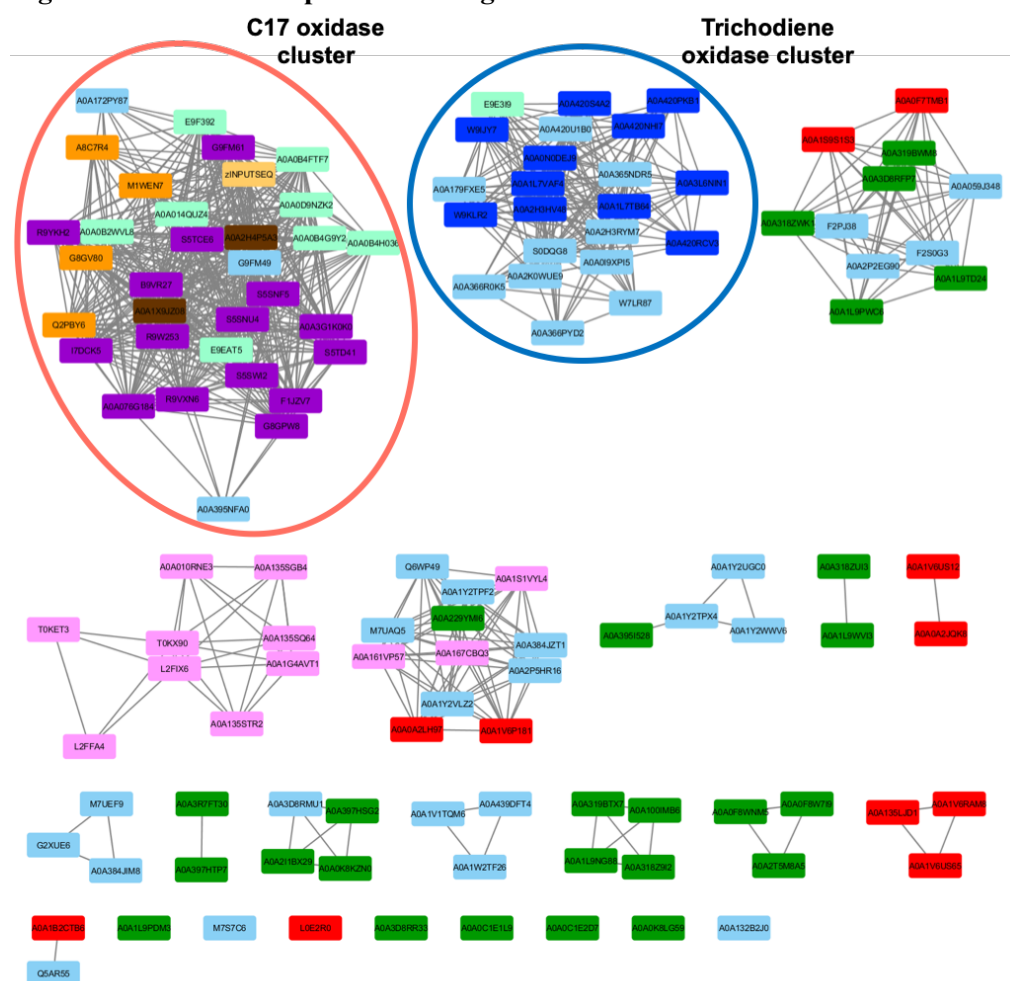


Figure 5-3: Expanded SSN of the *cloA* supercluster from Figure 5-2B, obtained by defining an edge as a relationship corresponding to a sequence identity greater than 60%. Nodes in **purple denote sequences from the *Epichloe* family, nodes in **orange** denote sequences from the *Claviceps* family, nodes in **green** denote sequences from the *Aspergillus* family, nodes in **red** denote sequences from the *Penicillium* family, nodes in **brown** are for sequences from *Claviceps gigantea* and *Africana* (known producers of the dihydroergot alkaloids), nodes in **pink** denote sequences from the *Colletotrichum* family, nodes in **teal** denote sequences from the *Metarhizium* family, nodes in **dark blue** denote sequences from the *Fusarium* family, and nodes in **light blue** represent sequences from the other collective families not well characterized.**

The completion of the ergot alkaloid pathway to DLA in yeast would require the incorporation of an ortholog of *cloA* that favours its production. This chapter describes the incorporation of three of the five orthologs, found to produce the 269 m/z compound, into the previously developed agroclavine producing strain (AgcM33B/ AgcM32B).

5.2 MATERIALS AND METHODS

5.2.1 GENERATING AN SSN OF THE *cloA* ORTHOLOGS

To generate the SSN of *cloA*, the sequence from *Epichloe* sp. Lp1 was used as the query for the initial BLAST search (Option A). The initial BLAST parameters were set to pull a maximum of 9000 sequences with a minimum alignment E-value of 5. An initial SSN was generated from all the returned sequences, and the edges were defined as relationships with an alignment score corresponding to greater than or equal to 40% sequence identity. These parameters return the SSN shown in Figure 5-2A. From this SSN, the predicted clavine oxidase supercluster was identified and coloured orange. Expansion of the supercluster by increasing the edge definition to greater than or equal to 60% sequence identity, produced the SSN in Figure 5-3. Further expansion of the “agroclavine” cluster by redefining an edge as a relationship greater than 77% produced the SSN in Figure 5-2C.

5.2.2 PROCUREMENT OF *cloA* ORTHOLOGS

Three of the five *cloA* orthologs identified to be producing a 269 m/z compound from agroclavine were obtained from a colleague, Lim. L. R. (Table 5-1). These sequences were cloned into a pYES2/CT vector, and are the constructs used by Lim, L. R.

Table 5-1: Summary of *cloA* orthologs used in this study

Abbreviated name	Organism	Uniprot ID
cloA_Cpur	<i>Claviceps purpurea</i>	M1WEN7
cloA_Ecoe	<i>Epichloe coenophiala</i>	R9VXN6
cloA_Pipo	<i>Periglandular ipomoeae</i>	G9FM49

5.2.2.1 TRANSFORMATION, CULTURE AND GROWTH CONDITIONS

The *cloA* constructs were transformed into chemically competent AgcM33B and AgcM32B cells that were prepared and transformed as described previously.

Three colonies per construct to be tested were grown as described before. Expression of the *cloA* orthologs were induced by the addition of 20% galactose to a final concentration of 2%, when the cultures reached an OD₆₀₀ of 0.8. Culture tubes (50 mL) were left uncapped and covered with autoclaved aluminium foil caps, and subsequently incubated further at 24 °C for 5 days. A 1 mL aliquot of spent media was prepared and analysed using LC-MS as described earlier.

5.2.2.2 PREPARATION OF STANDARDS

The stock solution of agroclavine (Chiron) was prepared as described in chapter 4. The stock solution of DLA (Chiron) was prepared by dissolving 10 mg of the compound in 10 mL 50% (v/v) LC-MS grade ethanol/ deionised water. The DLA standards were stored in the dark at -20 °C. A stock solution of lysergol standard (Toronto Research Chemicals) was prepared by dissolving the dry powder in DMSO to 100 mM and stored at -20 °C. The standards used for analysis were prepared by their appropriate dilutions in the filter-sterilized media of their corresponding negative controls.

5.2.2.3 QUANTIFICATION WITH A STANDARD CURVE

The standard curves for agroclavine and DLA were prepared by their appropriate dilutions in filter-sterilized media of their corresponding negative controls, over a range of concentrations from 10 nM to 1.5 µM.

The standard curve for agroclavine was generated by plotting the area under the curve for the peak response from the detection of the MS-MS ion transition of $[M+H]^+ 239 \rightarrow 208$ m/z. Likewise, the standard curve for DLA was generated by plotting the area under the curve for the peak response from the detection of the MS-MS ion transition of $[M+H]^+ 269 \rightarrow 223$ m/z.

5.2.2.4 QUANTIFICATION BY STANDARD ADDITION

The prepared DLA standard was diluted 1:1000 into 50% (v/v) LC-MS grade ethanol/ deionised water, and subsequently spiked into samples from each replicate of the pYES2/CT-*cloA*_Cpur construct. Three replicates were created for each concentration level of standard added.

To generate the calibration curve, a total of five concentration levels were used, corresponding to 0, 100, 200, 300, and 400 nM of DLA added. The calibration curve was obtained by plotting the area under the curve for the peak response from the detection of the MS-MS ion transition of $[M+H]^+ 269 \rightarrow 223$ m/z against the concentration of DLA added.

The concentration of DLA present in the original sample was calculated using the regression line obtained that corresponded to the added concentration producing double the peak response compared to the sample with no DLA added. Confidence intervals of the measured sample concentration were calculated from the derived confidence intervals of the regression line.

5.2.3 DEVELOPMENT OF A GENOMIC DLA PRODUCER

5.2.3.1 TRANSFORMATION, CULTURE AND GROWTH CONDITIONS

Genome integration plasmids for the *cloA* ortholog was obtained from Lim, L.R. The obtained plasmid construct was linearized by BsaI digest, and

gel purified. The purified linear integration fragment was transformed into chemically competent AgcM33B cells that were prepared and transformed as described previously. Successful transformants were selected for on SC-URA plates with 2% glucose, and screened for successful integration by PCR.

Three colonies per verified construct were selected and tested as described previously. Induction of *cloA* expression and DLA biosynthesis were performed as described earlier.

5.2.3.2 QUANTIFICATION BY STANDARD ADDITION

Quantification of the DLA produced by the DLAM33B strain was performed as described earlier.

5.2.4 CHARACTERIZATION OF DLAM33B

5.2.4.1 TRYPTOPHAN SUPPLEMENTATION EXPERIMENTS

Stock solutions of tryptophan (Sigma) were prepared in 20% (w/v) galactose to a concentration of 10 mM and 50 mM, and filter-sterilized.

Tryptophan supplementation was performed by culturing the strain as described in the earlier sections for the biosynthesis of DLA. Cultures were fed the additional tryptophan and induced for *cloA* expression concurrently by the addition of the prepared stock solution to a final concentration of either 1 mM tryptophan with 2% (w/v) galactose, or 5 mM tryptophan with 2% (w/v) galactose. Induction and biosynthesis was carried out as described previously.

Estimation of the quantities of agroclavine and DLA produced in the tryptophan supplementation experiments were performed using the standard curves generated.

5.2.4.2 ¹³C-TRYPTOPHAN FEEDING EXPERIMENTS

Stock solutions of tryptophan (Sigma) and ¹³C-2-indole tryptophan (CIL) were prepared as described in chapter 4. Tryptophan feeding was performed as described in chapter 4 as well, with the exception that cultures were induced for *cloA* expression and incubated at 24 °C, over 5 days. All spent media samples were prepared for LC-MS analysis as before.

5.2.4.3 QUANTIFICATION OF ¹³C-DLA WITH A STANDARD CURVE

Estimation of the amount of ¹³C-DLA produced from ¹³C-2-indole tryptophan was calculated using a standard curve of DLA. It is assumed that ¹³C-DLA and DLA are chemically identical and will produce the same peak area response for a given concentration and matrix when ionized using the electrospray ionization method.

5.2.4.4 CALCULATION OF PREDICTED LOGP VALUES

Predicted values of LogP for all the compounds discussed were obtained from the chemical properties function of ChemDraw® Version: 18.0.0.231 (4318).

5.2.5 LC-MS ANALYSIS OF SAMPLES

LC-MS analyses were performed using the Agilent 1290 Infinity LC System using an Agilent InfinityLab Poroshell 120 EC-C18 column, 2.1 x 100 mm dimensions, 1.9 µm particle size. Samples were run using the same LC method as previously described.

MS-MS analysis was performed as previously described. Targeted masses for the experiments testing the *cloA* orthologs and DLAM33B were set

to: 1) 239.1543 m/z, on a narrow isolation width (1.3 amu), collision energy of 20; 2) 255.1492 m/z, on a narrow isolation width (1.3 amu), collision energy of 20; 3) 269.1285 m/z, on a narrow isolation width (1.3 amu), collision energy of 20.

Targeted masses for the ^{13}C -labelled tryptophan experiments were set to: 1) 269.1285 m/z, on a narrow isolation width (1.3 amu), collision energy of 20; 2) 270.1318 m/z, on a narrow isolation width (1.3 amu), collision energy of 20.

5.2.6 STATISTICAL ANALYSIS

Comparisons between the levels of DLA produced by the constructs expressing the *cloA* orthologs episomally were performed using a One-way ANOVA. Analyses of the quantities of DLA and agroclavine produced from the supplementation of tryptophan were performed with a Two-way ANOVA.

5.3 RESULTS AND DISCUSSION

5.3.1 BIOINFORMATICS GUIDED PREDICTION OF *cloA* ACTIVITY

The utility of an SSN for the identification of a suite of enzymes for the catalysis of a desired reaction has been demonstrated extensively over the course of this project. This approach proved to be especially useful for delineating the slightly different reactions catalysed by otherwise highly similar enzymes, such as *easA*. Therefore, the same approach was therefore taken to probe the substrate and product specificities of the various *cloA* clades.

Within the predicted *cloA* supercluster, it was immediately apparent that the homologs from the *Aspergillum* and *Penicillium* families, that corresponded to the CYP450s that act to produce the fumigaclavines, were distinctly more divergent from the ones from the *Clavicipitaceae* family (Figure 5-3). This is rightly so, given the knowledge that they catalyse the hydroxylation of the C9 position on festuclavine or pyroclavine that would require the substrate to be oriented in a starkly different position relative to the catalytic iron-heme co-factor (Bilovol & Panaccione, 2016). Another distinct sub-cluster that was picked up in the *cloA* supercluster, were a group of sequences where around half were from the *Fusarium* family of fungi and annotated as trichodiene oxygenases. The other half of this group were annotated as *ent*-kaurene oxidases from the *Gibberella* family of plant parasitizing fungi. Both groups were likely to have been picked up because of the similar active site residues used for binding polycyclic substrates.

Once these extraneous clusters were excluded, the remaining pool of 49 predicted *cloA* sequences could be re-organised into four major groups (Figure 5-4). Of these, the most distinct 15 sequences between the four groups were

selected to be characterized as part of the study probing the features of *cloA* substrate and product specificity. In this separate work, five orthologs were found to catalyse the production of a 269 m/z product from agroclavine (Table 5-2), indicating the production of either paspalic acid or DLA (Lim, L.R., personal communication, 10 November 2017). Three of the most distinct sequences, *cloA*_Cpur, *cloA*_Ecoe, and *cloA*_Pipo were selected to be used for completing the pathway to DLA.

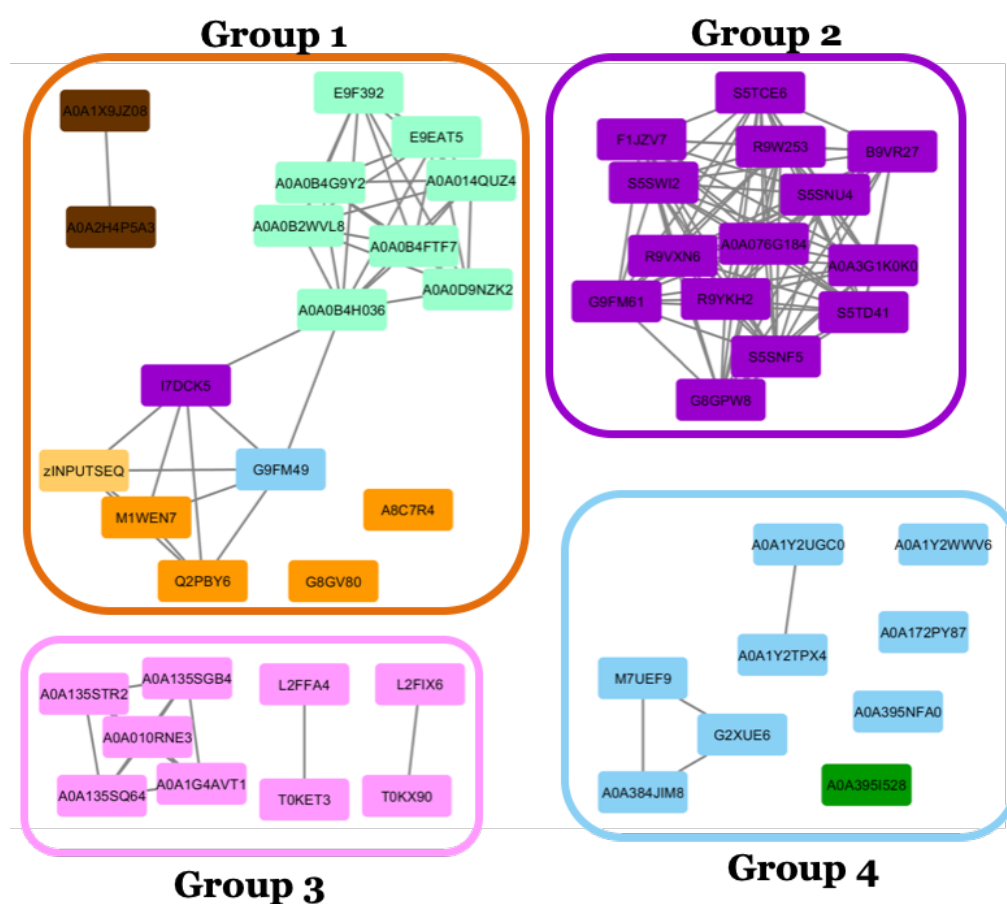


Figure 5-4: Expanded SSN of the *cloA* supercluster, obtained by defining an edge as a relationship corresponding to a sequence identity greater than 75%. Exclusion of the clusters containing the trichodiene oxidases and clusters with known C10-ergoline hydroxylases reduced the remaining sequences into four distinct groups. Fifteen of the most dissimilar sequences representing each of the four groups were selected and tested by Lim, L. R. for activity on agroclavine.

Table 5-2: Summary table of the *cloA* orthologs that were found to be producing a compound with [M+H]⁺ of 269 m/z

<i>CloA</i>	Organism	Uniprot ID
cloA_Cpur	<i>Claviceps purpurea</i>	M1WEN7
cloA_Ecoe	<i>Epichloe coenophiala</i>	R9VXN6
cloA_Pipo	<i>Periglandular ipomoeae</i>	G9FM49
cloA_Nlol	<i>Neotyphodium lolii</i>	S5SWI2
cloA_253	<i>Epichloe coenophialia</i>	R9W253

5.3.2 DEMONSTRATING THE COMPLETE BIOSYNTHESIS OF D-LYSERGIC ACID

5.3.2.1 PLASMID-BASED TRIALS

In the characterization of the *cloA* orthologs, the sequences were synthesized and cloned into pYES2/CT plasmids. In order to replicate the production of the 269 m/z compound observed, the previously developed agroclavine-producing strain (AgcM33B/ AgcM32B) was used as a platform for assaying these three orthologs (Figure 5-5A).

The presence of a compound with an [M+H]⁺ of 269 m/z was detected from the spent media samples from all three of the orthologs tested (Table 5-3). These results were consistent with the experiments where the strains, which only expressed a *cloA* ortholog, were fed with commercial agroclavine. Samples of spent media from the negative control (AgcM33B) spiked with commercial DLA, was found to co-elute the same 269 m/z compound (Figure 5-5B). This allowed for the determination of the compound to be DLA.

The fragmentation pattern obtained from the targeted MS-MS from these peaks showed that all orthologs presented a similar profile to the spiked standard (Figure 5-6A, B, C, D). This demonstrated the biosynthesis of DLA from agroclavine and the complete reconstitution of the ergot alkaloid biosynthesis pathway to DLA in yeast (Table 5-4).

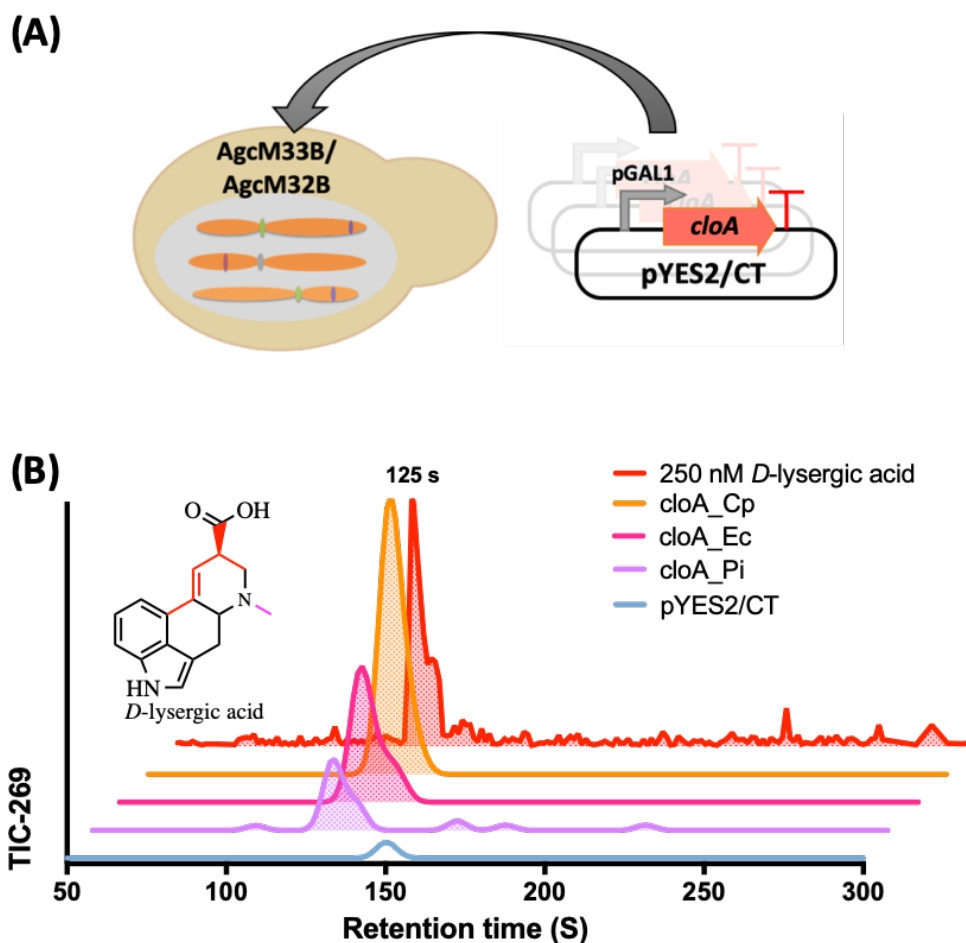


Figure 5-5: Testing of the three *cloA* orthologs using the agroclavine producing strains AgcM33B and AgcM32B. (A) Cartoon representation of the constructs used. (B) Total ion chromatogram from the selected ion monitoring of all fragments produced from precursors with an $[M+H]^+$ of 269 m/z. All orthologs tested eluted a compound with $[M+H]^+$ of 269 m/z at 125 s that corresponded to the DLA standard spiked into the media of the negative control.

Table 5-3: Exact mass and retention time for the compounds detected

Construct	Calculated m/z	Observed m/z	Mass error δ (ppm)	Retention time (s)
DLA standard	269.1285	269.1267	-6.69	125.99
cloA_C.pur	269.1285	269.1272	-4.83	126.12
cloA_E.coe	269.1285	269.1274	-4.09	126.00
cloA_P.ipo	269.1285	269.1275	-3.72	126.15

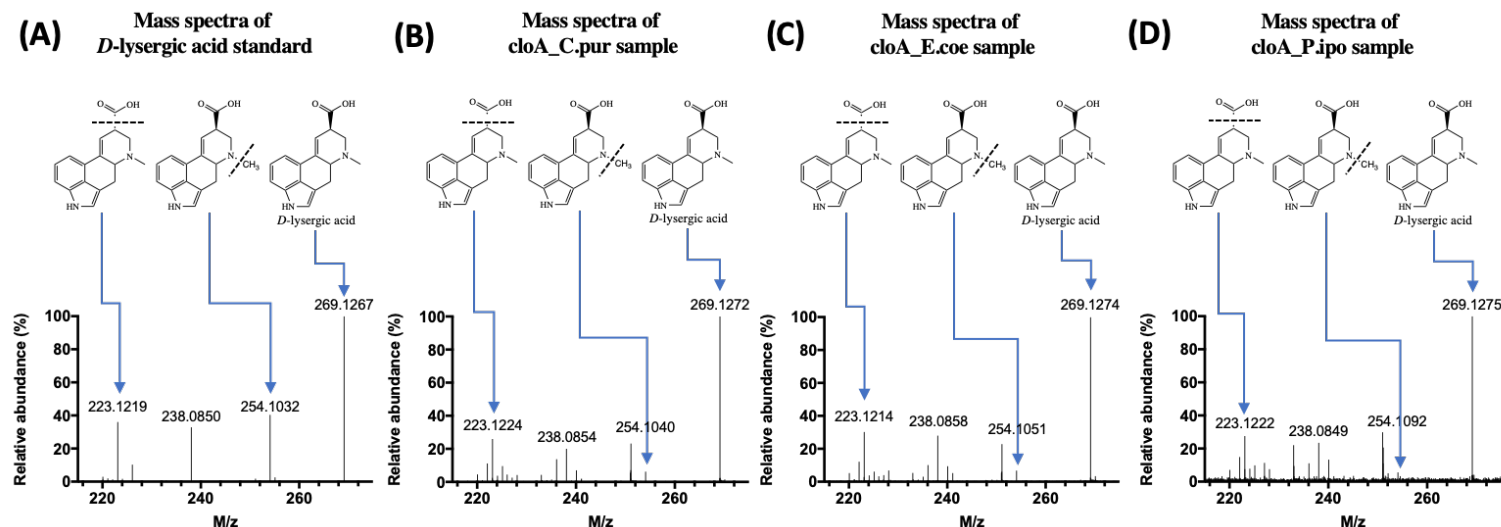


Figure 5-6: Mass spectra from the MS-MS of the peaks eluting at 125 s from the (A) DLA standard, (B) *C. purpurea* ortholog, (C) *E. coenophialia* ortholog, and (D) *P. ipomoeae* ortholog.

Table 5-4: Exact masses of select fragment ions detected from the DLA standard and DLA detected in the samples

Construct	254 m/z fragment			238 m/z fragment			223 m/z fragment		
	Standard m/z	Observed m/z	Mass error δ (ppm)	Standard m/z	Observed m/z	Mass error δ (ppm)	Standard m/z	Observed m/z	Mass error δ (ppm)
cloA_C.pur	254.1032	254.1040	3.15	238.0850	238.0854	1.68	223.1219	223.1224	2.24
cloA_E.coe	254.1032	254.1051	7.48	238.0850	238.0858	3.36	223.1219	223.1214	-2.24
cloA_P.ipo	254.1032	254.1092	23.61	238.0850	238.0849	-0.42	223.1219	223.1222	1.35

5.3.2.2 QUANTIFYING PRODUCTION FROM PLASMID BASED STRAINS

After establishing the production of DLA from the plasmid-based strains, a standard curve was generated to quantify the amounts of agroclavine and DLA produced (Figure 5-7A, C). The level of production of DLA varied across the three orthologs; while the *C. purpurea* and *E. coenophialia* orthologs produced similarly sized peaks, the peak from the *P. ipomoeae* ortholog suggested it was producing approximately three times less product than the other two orthologs (Figure 5-7B).

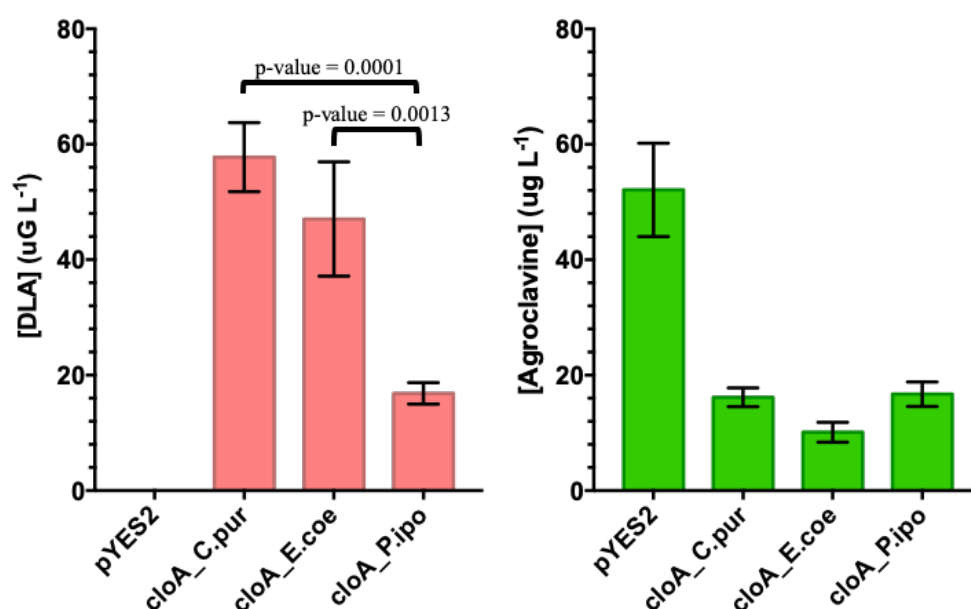


Figure 5-7: Quantification of the DLA and agroclavine produced in the plasmid-based strains by standard curve. (A) Relative amounts of DLA produced in test strains to the standard curve. Error bars represent standard deviation from three biological replicates. (B) Relative amounts of agroclavine produced in test strains to the standard curve. Error bars represent standard deviation from three biological replicates.

It could be possible that the *P. ipomoeae* ortholog is less active or was not expressed as well in the strain. However, all orthologs appeared to be consuming similar amounts of agroclavine (Figure 5-7D). This suggested that the *P. ipomoeae* ortholog was producing an alternative product from DLA. To search for this alternate product, the spent media samples of the three orthologs

were assessed for the presence of compounds with an $[M+H]^+$ of 255 m/z, which may indicate the presence of elymoclavine, lysergol, or chanoclavine-I-aldehyde. It was found that the *C. purpurea* and *E. coenophialia* orthologs did not contain any distinct product with this $[M+H]^+$. On the other hand, the *P. ipomoeae* ortholog was found to produce a compound with an $[M+H]^+$ of 255 m/z that eluted at around 127 s (Figure 5-8A, B). The close retention time with DLA suggested it was likely to be elymoclavine or lysergol rather than chanoclavine-I-aldehyde.

Fortunately, lysergol was commercially available. The elution time and mass fragmentation pattern of the 255 m/z compound produced by the *P. ipomoeae* ortholog was found to be similar to commercial lysergol (Table 5-5). This demonstrated that the *P. ipomoeae cloA* ortholog catalysed the production of both DLA and lysergol from agroclavine. This finding demonstrated the divergent capability inherent in the *cloA_P.ipo* ortholog, and corroborated earlier research on the isolation of lysergol from the seeds of the *P. ipomoeae* plant (Hofmann & Tscherter, 1960).

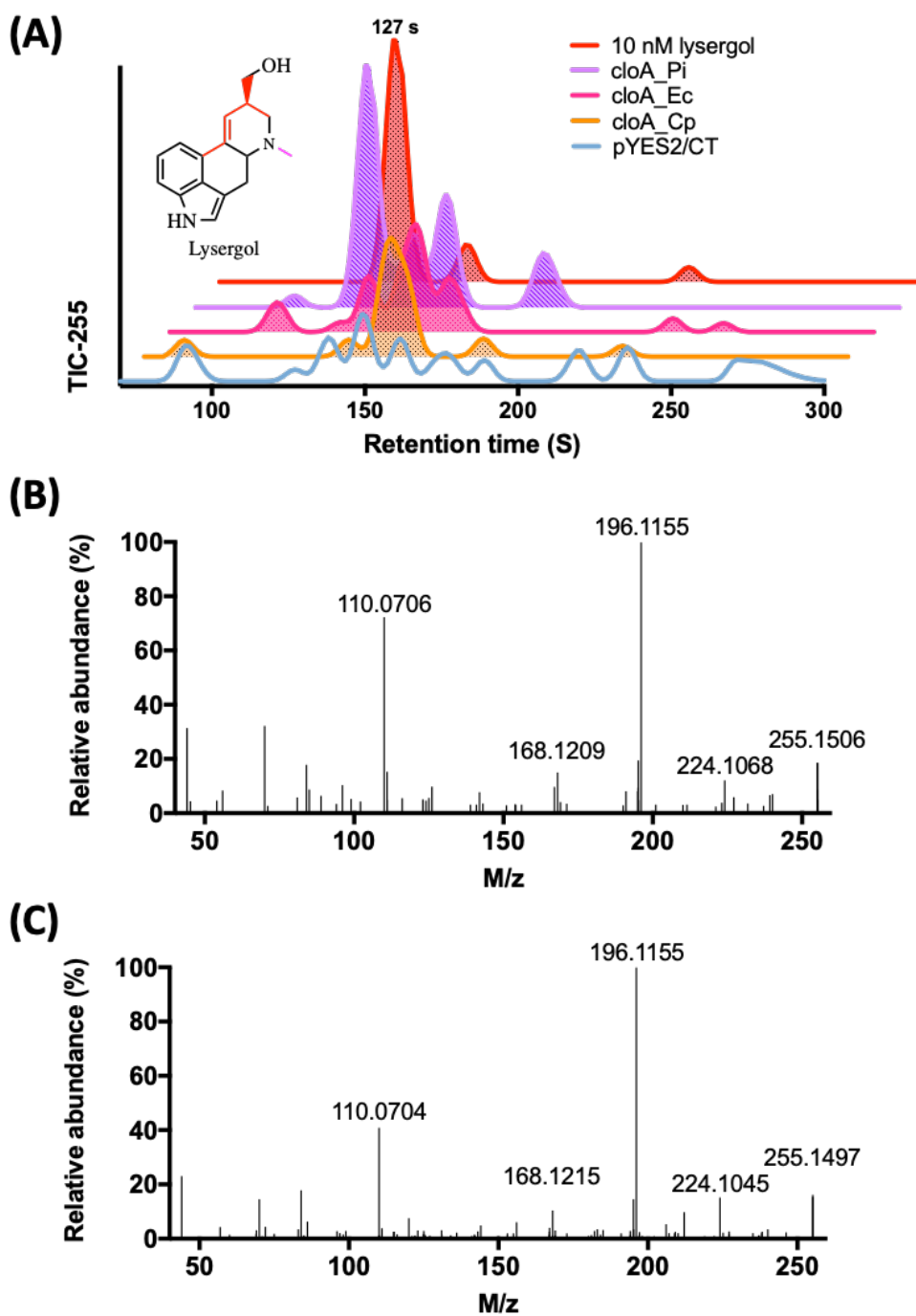


Figure 5-8: (A) Total ion chromatogram from the selected ion monitoring of all fragments produced from precursors with an $[M+H]^+$ of 255 m/z. (B) Mass spectra from the MS-MS of the peak from the lysergol standard. (C) Mass spectra from the MS-MS of the peak from the cloA_Pipo sample.

Table 5-5: Exact mass and retention time for the lysergol standard and cloA_Pipo sample, as well as the most abundant fragment ions

Construct	Calculated m/z	Observed m/z	Mass error δ (ppm)	Retention time (s)
<i>Lysergol standard</i>	255.1492	255.1506	5.49	127.0
cloA_P.ipo	255.1492	255.1497	1.96	127.2
Fragment Ions	Standard m/z			
224 m/z	224.1068	224.1045	-10.26	
196 m/z	196.1155	196.1155	0.00	
168 m/z	168.1209	168.1215	3.57	
110 m/z	110.0706	110.0704	-1.82	
44 m/z	44.0491	44.0487	-9.08	

These series of experiments are the first reported demonstration of the complete biosynthesis of both lysergol and DLA in an entirely heterologous host. It is also the first step towards supporting the hypothesis that the specific ortholog of *cloA* present in an ergot alkaloid producer is responsible for an additional dimension of diversification at the later segments of the ergot alkaloid pathway.

5.3.2.3 QUANTIFICATION BY STANDARD ADDITION

With a sense of the relative amounts of DLA the produced by the plasmid-based strains, it would be prudent to have a more absolute measure for future developments in both strain and bioprocess optimization.

The amount of DLA produced by the strain with the *C. purpurea* ortholog was quantitated using standard addition as a secondary measure to account for any potential confounding matrix effects (Figure 5-9). Using this

method, the concentration of DLA was determined to be $74.15 \pm 19.89 \mu\text{g/L}$ (Table 5-6). This calculated value was not significantly different from the value derived from the standard curve, and indicated that any possible matrix effects have already been accounted for.

While the overall yield was still considerably low for any direct commercial application, the intrinsic design allowed for an ample amount of space to perform further strain optimization.

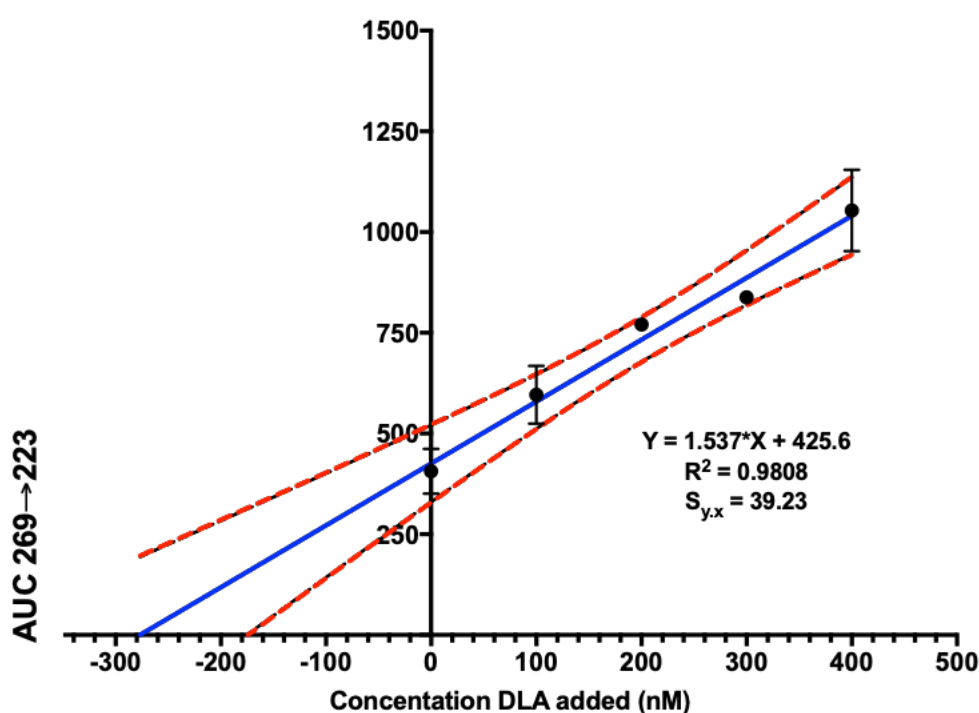


Figure 5-9: Quantification of DLA produced in the cloA_Cpur sample by standard addition. The calibration curve is obtained by plotting the area under the curve for the peak response from the detection of the MS-MS ion transition of $[M+H]^+ 269 \rightarrow 223 \text{ m/z}$ against the concentration of DLA added.

Table 5-6: Summary of the calculated DLA concentration from the cloA_Cpur construct measured by standard addition

pYES2/CT-cloA_C.pur	Concentration (nM)	Concentration ($\mu\text{g/L}$)
Mean concentration	<u>276.90</u>	<u>74.30</u>
95% Confidence Interval	± 74.15	± 19.89
Upper limit	351.05	94.19
Lower limit	202.76	54.40

5.3.2.4 TOWARDS A STABLE GENOMIC STRAIN

While the expression of the pathway in full or in parts from a plasmid would adequately serve to demonstrate its reconstitution in a heterologous host, housing the full pathway on the host genome would be a quintessential feature for a microbial cell factory of industrial relevance. Integrating the full pathway on the host genome confers the advantage of not requiring the perpetual maintenance of a selection pressure to retain the plasmid and ultimately pathway. Growth media catering to any form of selection, be it positive or negative, is generally more costly than generic non-selective media.

In line with the penultimate goal of a microbial cell factory for ergot alkaloid production, the expression cassette for *cloA_Cpur* was moved into the yeast genome. This new strain, housing the full pathway on its genome (DLAM33B) (Figure 5-10A), was checked for the production of DLA (Figure 5-10B, C, D) (Table 5-7), and quantitated by standard addition. This strain was found to be producing $71.48 \pm 29.81 \mu\text{g/L}$ of DLA, which is approximately the same amount produced by the plasmid-based strains (Figure 5-11) (Table 5-8).

This strain is a successful proof of concept in the utility of modern synthetic biology methods for the heterologous expression of complex biosynthetic pathways. With the existing characterized parts and assembly system based on the adaptation of the YeastFab assembly, it will be a trivial matter to re-design or re-build the existing strain.

Pushing forward to obtain commercially competitive yields will require a fair amount of optimization. The focus of the subsequent work therefore shifts towards the identification of the key factors that would improve the production yields.

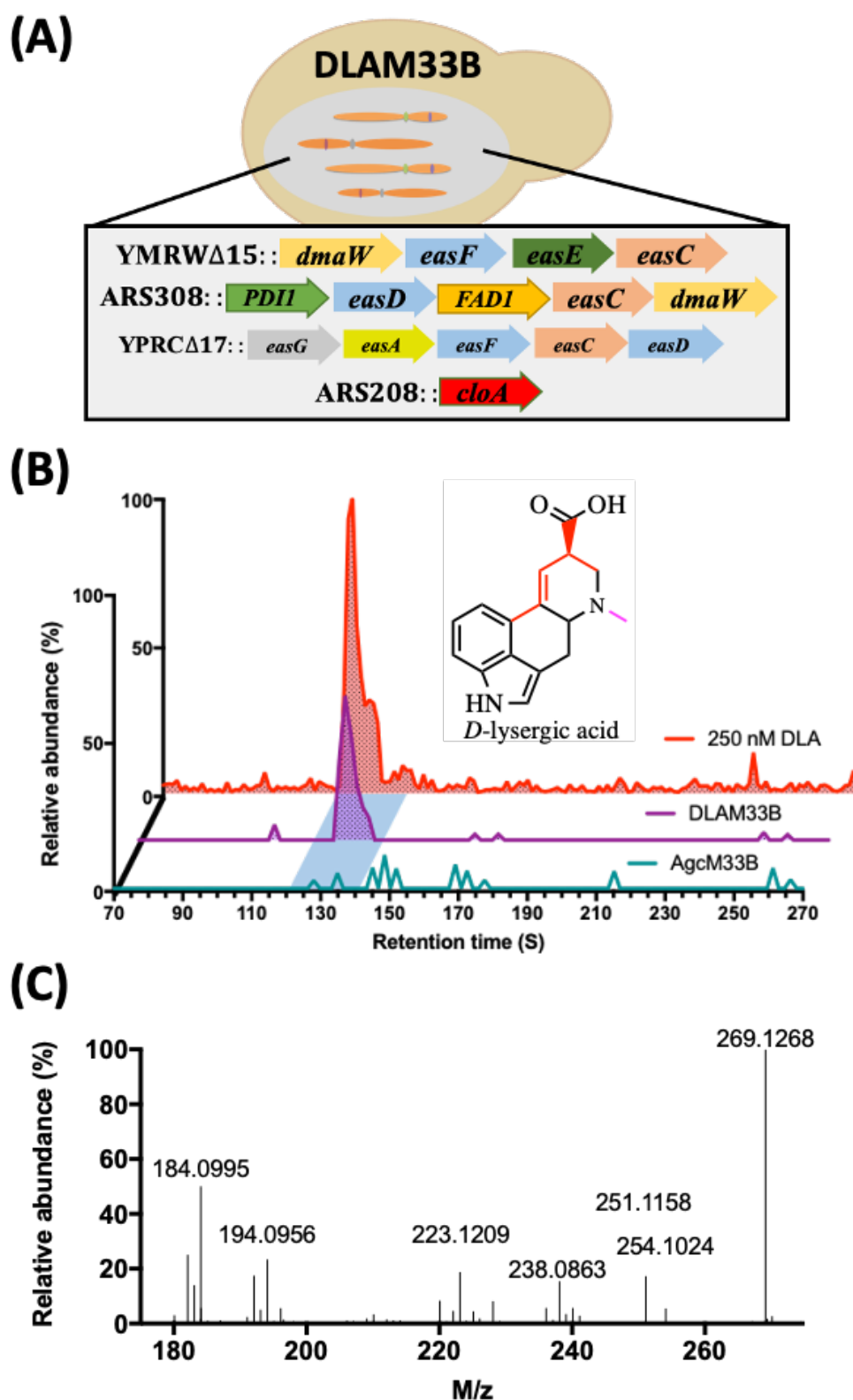


Figure 5-10: Characterization of DLAM33B. (A) Diagrammatic representation of the construct's genotype. (B) Total ion chromatogram from the selected ion monitoring of all fragments produced from precursors with an $[M+H]^+$ of 269 *m/z*. (C) Mass spectra from the MS-MS of the DLAM33B sample showing the same fragmentation pattern observed in the DLA standard and plasmid-based strains.

Table 5-7: Exact mass and retention time for the DLA standard and DLAM33B sample, as well as the most abundant fragment ions

Sample	Calculated m/z	Observed m/z	Mass error δ (ppm)	Retention time (s)
<i>DLA standard</i>	269.1285	269.1269	-5.95	125.0
DLAM33B	269.1285	269.1268	-6.32	124.9
Fragment Ions	Standard m/z			
254 m/z	254.1034	254.1024	-3.94	
238 m/z	238.0863	238.0863	0.00	
223 m/z	223.1205	223.1209	1.79	

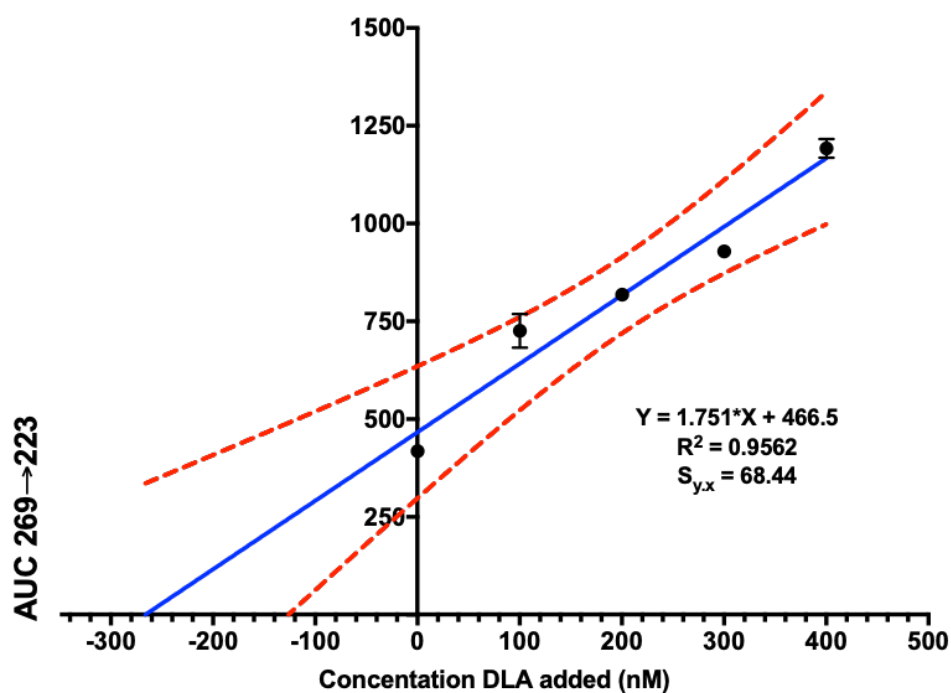


Figure 5-11: Quantification of DLA produced by DLAM33B by standard addition. The calibration curve was obtained by plotting the area under the curve for the peak response from the detection of the MS-MS ion transition of $[M+H]^+ 269 \rightarrow 223$ m/z against the concentration of DLA added.

Table 5-8: Summary of the calculated DLA concentration from the cloA_Cpur construct measured by standard addition

DLAM33B	Concentration (nM)	Concentration ($\mu\text{g/L}$)
Mean concentration	<u>266.42</u>	<u>71.48</u>
95% Confidence Interval	± 111.11	± 29.81
Upper limit	377.53	101.30
Lower limit	155.31	41.67

5.3.3 EVALUATING THE ASSEMBLED PATHWAY

5.3.3.1 EFFECT OF SUPPLEMENTING ADDITIONAL TRYPTOPHAN

With the knowledge gathered, we have arrived at the “Test→Learn” phase of this project. To achieve a commercially viable strain, the yield from DLAM33B needs to be improved by around five orders of magnitude, to the range of 1 g/L (Cvak, 1999). The approach taken to surmount this seemingly gargantuan obstacle is to identify the largest factors limiting the current yields as rationally as possible.

Several pertinent questions needed to be answered. Firstly, the efficiency of the pathway needs to be determined, and identify whether engineering improvements to the pathway or enhancing substrate production would most effectively bring up DLA yield. The simplest way to answer this is by supplementing the culture media with additional tryptophan. Within the generic SC-URA media used for yeast culture, there is already approximately 0.1 mM tryptophan. Supplementing 1 mM and 5 mM tryptophan to the media would increase the tryptophan concentration by 11 and 51 fold, respectively. This covered a comprehensive range within its solubility limit in water (55 mM).

Measuring the concentration of DLA produced from DLAM33B supplemented with additional tryptophan, found no corresponding increase in DLA (Figure 5-12B). The additional tryptophan did however, increase the amounts of agroclavine produced in both strains with or without *cloA* (Figure 5-12D). This demonstrated that at the current level of pathway efficiency, engineering any increase in substrate levels would have a minimal effect on yields and that any engineering effort should be directed towards improving the introduced pathway's turnover efficiency.

This result showed two main features. Firstly, the increase in agroclavine production with the increase in tryptophan suggests that the nature of inhibition agroclavine has been documented to exert on *dmaW* (L.-J. Cheng et al., 1980), is likely competitive in nature. The increase in tryptophan levels displaces the feedback inhibition from agroclavine towards a new equilibrium point.

Secondly, is the lack of change in DLA production levels, despite the increase in tryptophan and the ensuing increase of agroclavine levels. Comparing the molar ratios of DLA and agroclavine produced against the increase in tryptophan supplemented, illustrated the relationship between DLA and agroclavine produced resembles a reciprocal function (Figure 5-13A). A further comparison of the differences in agroclavine produced in the strain expressing *cloA* against the strain without *cloA*, suggested that an almost constant amount of agroclavine is being consumed by *cloA* across the varying levels of tryptophan supplementation. This would suggest that *cloA* is behaving as if it were another substrate in the reaction rather than a catalyst.

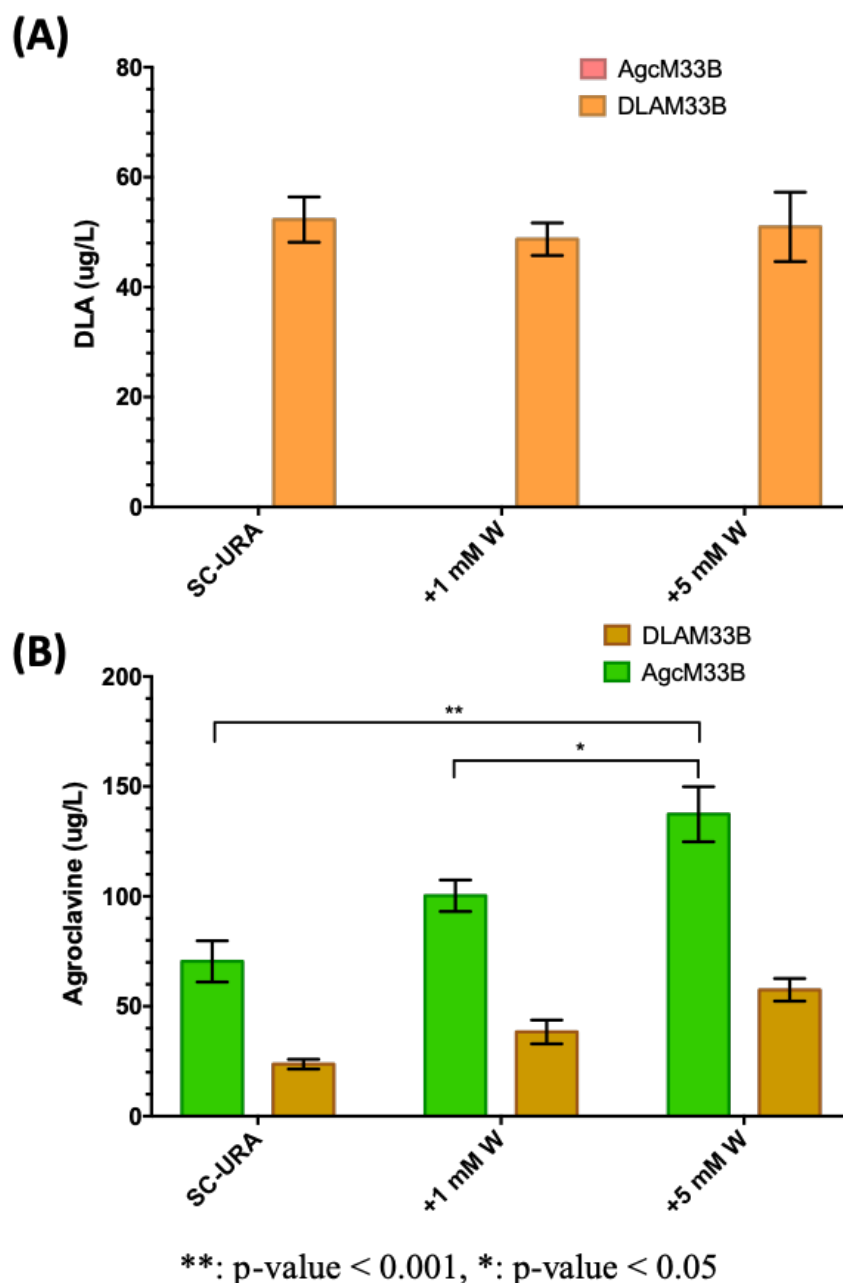


Figure 5-12: Quantification of the DLA and agroclavine produced in DLAM33B and AgcM33B when supplemented with additional tryptophan. (A) Relative amounts of DLA produced in test strains to the standard curve. Error bars represent standard deviation from three biological replicates. (B) Relative amounts of agroclavine produced in test strains to the standard curve. Error bars represent standard deviation from three biological replicates.

In a canonical CYP450 monooxygenase system, the capability to activate and oxidize an otherwise unreactive alkyl group stems from the generation of a ferric peroxo intermediate (Figure 5-13B), as part of its catalytic cycle (Figure 5-13C). This intermediate readily pulls protons to create, first a

ferric hydroperoxo intermediate, followed by the tetravalent Fe^{4+} -heme radical that performs the actual hydroxylation/oxidation reaction (Meunier, de Visser, & Shaik, 2004). Regeneration of this system requires some redox partner (usually a CYP450 reductase) to supply another pair of electrons to regenerate the reactive intermediates (Cook, Finnigan, Cook, Black, & Charnock, 2016).

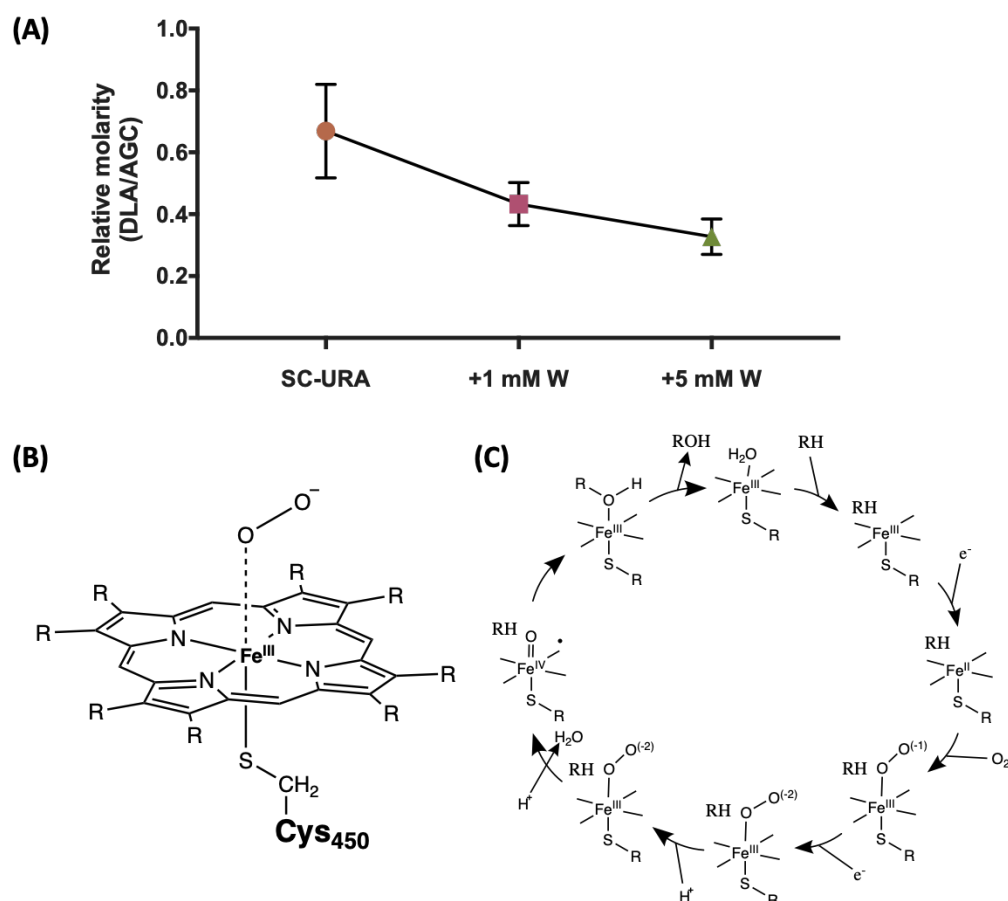


Figure 5-13: (A) Calculated relative molarity of DLA/ Agroclavine from samples fed with varying amounts of tryptophan supplemented. (B) The ferric peroxo intermediate that is generated as part of the CYP450 monooxygenase cycle, after picking up molecular oxygen and 2 electrons. (C) The catalytic cycle of CYP450 monooxygenases performing the archetypal hydroxylation reaction.

The lack of increase in turnover from agroclavine to DLA beyond what appears to be a fixed concentration, suggests that the endogenous CYP450 reductases (or alternative redox partners) in yeast are not adequately resetting *cloA* for subsequent reactions. This highlights the importance of engineering in additional support systems such as co-enzymes and co-factors.

Bioinformatics analysis of the *cloA* peptide sequence suggests that it is most likely either a class I or class II fungal CYP450 system. The class I systems are ER-anchored and depend on a CYPR for its supply of electrons (Figure 5-14A), while the class II systems are similarly ER-anchored but utilize an additional cytochrome B₅ (CYPB₅) partner to supply the second electron (Figure 5-14B) (Kumar, 2010). In an attempt to supplement the supporting redox system for *cloA*, two commonly used CYPRs from *S. cerevisiae* and *A. thaliana* were overexpressed in DLAM33B (Figure 5-14C) (Murakami, Yabusaki, Sakaki, Shibata, & Ohkawa, 1990).

Simply overexpressing the CYPRs alone however did not help to improve the DLA production titer (Figure 5-14D). This therefore suggests that *cloA* is likely a class II CYP450 system and would probably require the supplementation of CYPB₅ as well. Engineering any improvement in DLA yield will require the exploration of this hypothesis by complementing the system with a corresponding CYPB₅ and associated reductases.

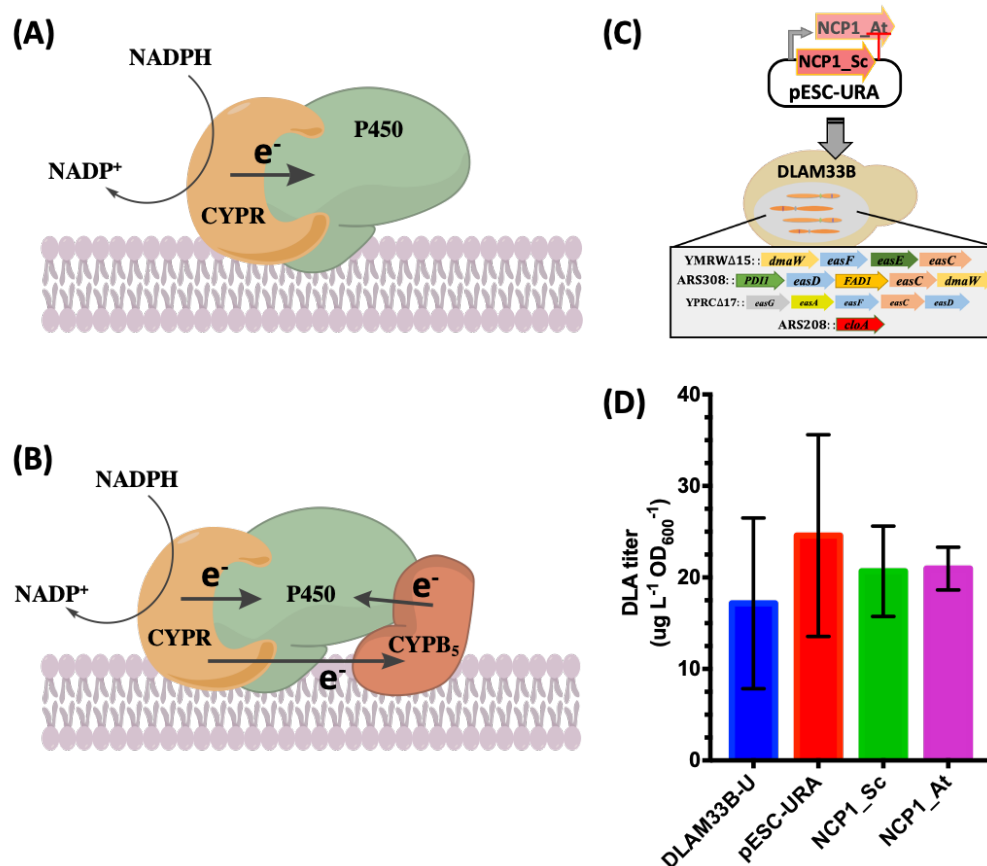


Figure 5-14: Cartoon representation of the class I (A) and class II (B) fungal CYP450 systems. (C) Overexpression of NCP1 from *S. cerevisiae* (NCP1_Sc) and *A. thaliana* (NCP1_At) on a pESC-URA vector to supplement the levels of CYPRs in DLAM33B. (D) Relative amounts of DLA produced by strains overexpressing NCP1_Sc and NCP1_At did not show any significant difference against the empty vector negative control or DLAM33B. Error bars were derived from three independent biological replicates.

5.3.3.2 STRAIN CHARACTERIZATION WITH LABELLED SUBSTRATES

5.3.3.2.1 ASSESSING PATHWAY EFFICIENCY

In the preceding section, pathway efficiency could be gauged from the tryptophan supplementation experiments. The genotype of the host yeast strain used, however, is not tryptophan auxotrophic. Therefore the actual amount of tryptophan available to the strain would have to include the amount it produces endogenously. Any measurement of pathway efficiency from the earlier section would have been confounded by this. In order to more accurately determine the efficiency of conversion of tryptophan to DLA, a similar experiment was conducted where ^{13}C -2-indole tryptophan was supplemented to the media.

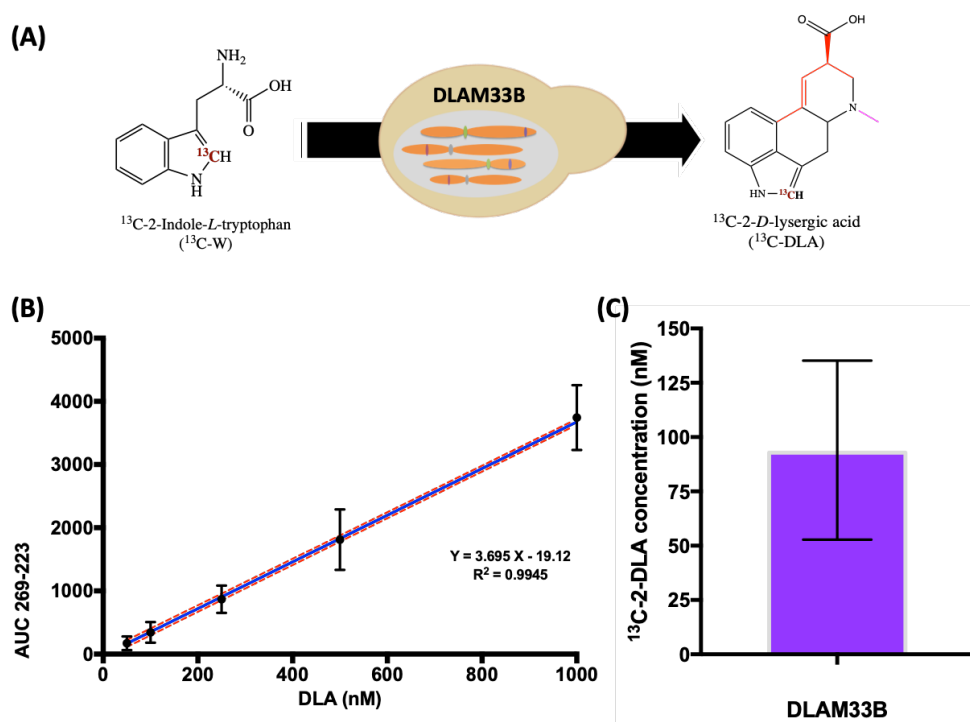


Figure 5-15: Characterization of the DLAM33B strain by feeding ^{13}C labelled tryptophan. (A) Cartoon representation of the experiment. (B) DLA standard curve established in AgcM33B media. Peak area response for the ion transition of 269 to 223 m/z plotted against spiked DLA concentrations, over a range of 50 – 1000 nM. (C) Estimation of ^{13}C -DLA produced by comparison to the standard curve of DLA.

^{13}C -labelled DLA produced is chemically identical to DLA, and should therefore be ionized to the same extent by electrospray ionization. The amount

of ^{13}C -DLA produced from ^{13}C -W was thus estimated using a standard curve of regular DLA (Figure 5-15). It was found that when fed with 1 mM ^{13}C -W, the DLAM33B strain produced approximately 92.8 nM of ^{13}C -DLA that accounts for around half (51.1%) the total amount of DLA produced. In general, this translates to an overall pathway efficiency level of $9.28 \times 10^{-3} \%$, and improving the turnover across the reaction catalysed by *cloA* could potentially double it.

5.3.3.2.2 VALIDATION OF PATHWAY TURNOVER AND FLUX TRACKING

The production of ^{13}C -DLA when ^{13}C -W is supplemented further validates the complete reconstitution of the ergot alkaloid pathway in yeast. As shown in chapter 4, feeding strains with labelled tryptophan produces a shift in the masses recorded for DLA by 1 m/z (Figure 5-16A, B, C, D). Analysis of the fragment ions produced further confirm the eluted compound at 126 s as the $[^{13}\text{C}\text{-M+H}]^+$ for DLA (Table 5-9).

Closer inspection of the data for all the detectable intermediates produced from the strains fed with tryptophan and ^{13}C -W, likewise show the same +1 m/z shift across all the previously detected intermediates in the pathway (Figure 5-17). This includes the shunt products MTCA and DMA-MTCA, formed from the reaction of tryptophan and DMAT with acetaldehyde after media acidification, that were discussed in chapter 3 (Figure 5-18).

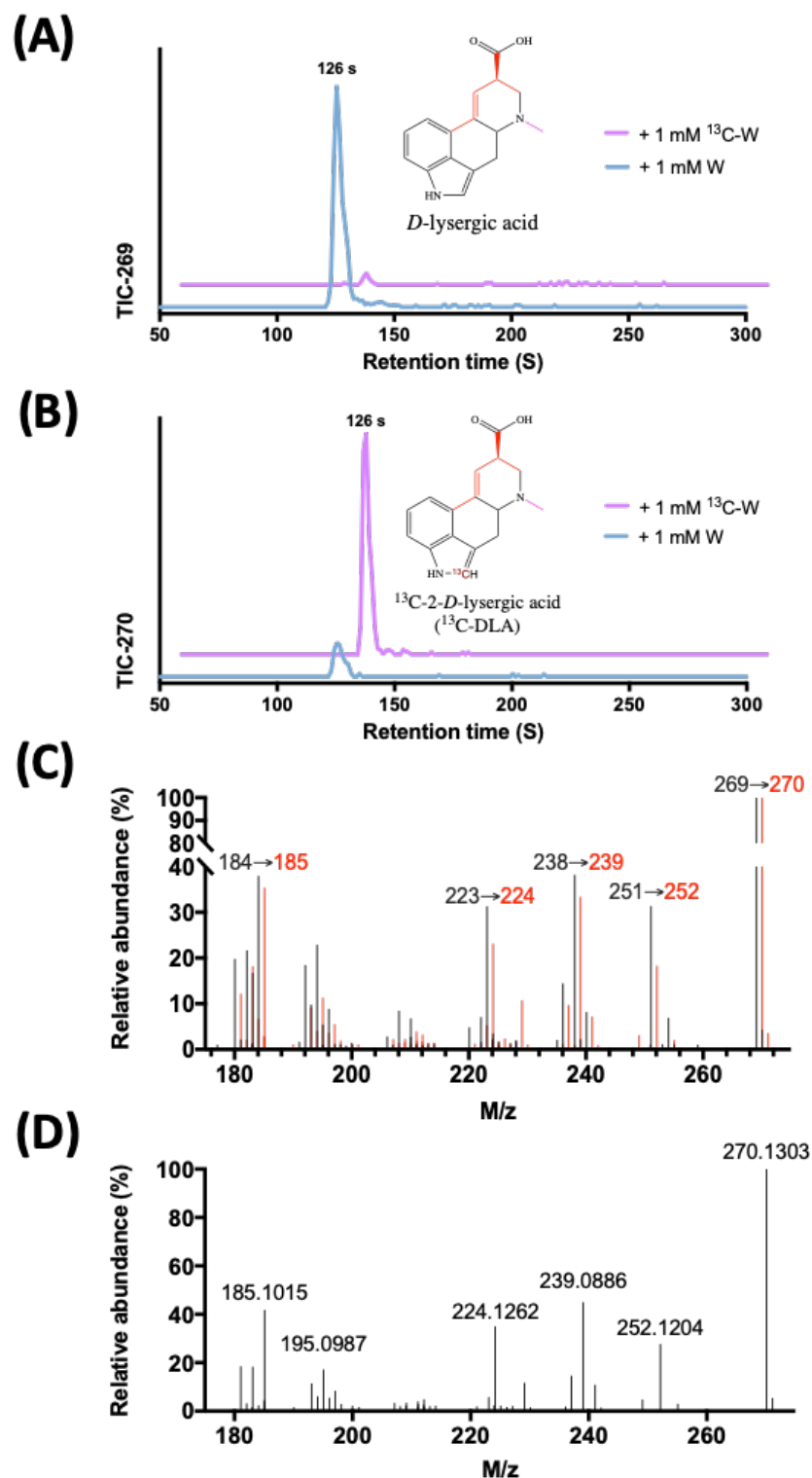


Figure 5-14: Production of ^{13}C -DLA from ^{13}C -W. Total ion chromatogram from the selected ion monitoring of all fragments produced from precursors with an $[\text{M}+\text{H}]^+$ of (A) 269 m/z , and (B) 270 m/z . (C) Overlay of the mass spectra for the peaks eluting at 126 s. Spectral lines in black represent the masses observed from the culture fed with tryptophan. Spectral lines in red represent the masses observed from the culture fed with ^{13}C -W. (D) Mass spectra showing the differences between cultures fed with tryptophan or ^{13}C -W.

Table 5-9: Exact masses of ^{13}C -DLA produced and its fragment ions

Detected [^{13}C-M+H]$^{+}$ ions	Calculated m/z	Observed m/z	Mass error δ (ppm)
^{13}C -2-DLA	270.1318	270.1303	-5.55
238 $\rightarrow$ 239	239.0896	239.0886	-4.18
223 $\rightarrow$ 224	224.1238	224.1262	10.71
184 $\rightarrow$ 185	185.0995	185.1015	10.80

In the process of tracking the different masses that were shifted by the incorporation of ^{13}C -W, a peculiar group of peaks were found to display the same +1 m/z shift. These peaks were found to elute at 170, 186, and 201 s and predominantly consisted of a compound with an $[\text{M}+\text{H}]^{+}$ of 285 m/z (Figure 5-19A, B). In this same group of peaks on the chromatogram, trace quantities of another mass of 303 m/z was spotted as well (Figure 5-19C).

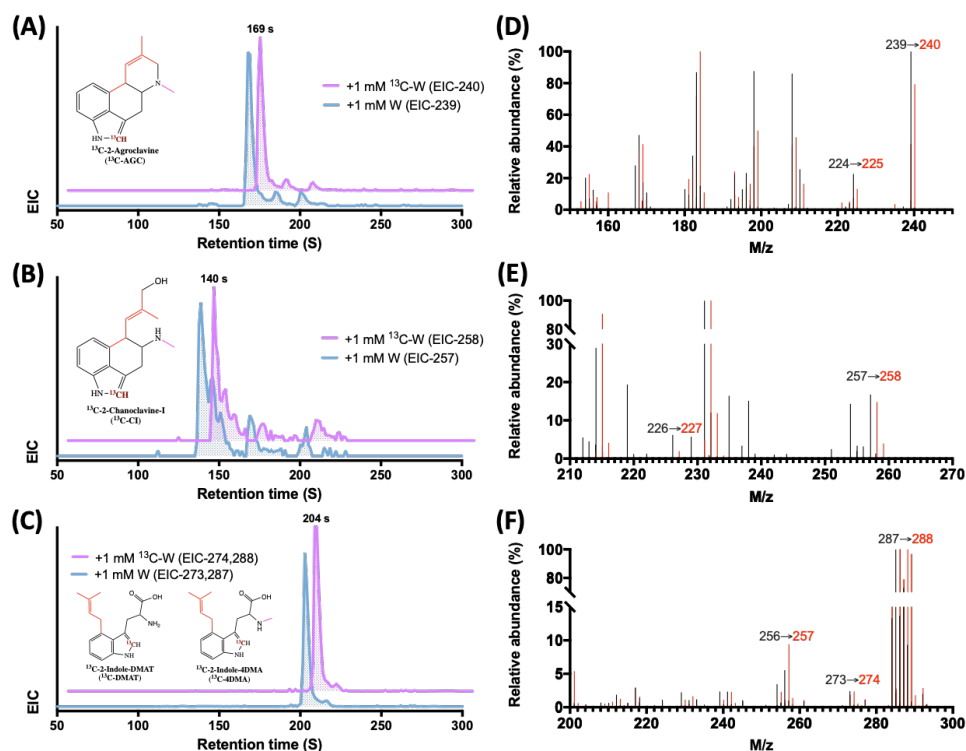


Figure 5-17: Detection of all the ^{13}C -labelled intermediates along the reconstituted ergot biosynthesis pathway when fed ^{13}C -labelled tryptophan. EIC for compounds with $[\text{M}+\text{H}]^+$ of (A) 239 and 240 m/z, corresponding to agroclavine and ^{13}C -agroclavine; (B) 257 and 258 m/z, corresponding to chanoclavine-I and ^{13}C -chanoclavine-I; (C) 273, 274, 287, and 288 m/z, for DMAT, 4DMA, and their ^{13}C -labelled counterparts. Overlay of the mass spectra for the peaks corresponding to (D) agroclavine, (E) chanoclavine-I, (D) DMAT and 4DMA. Spectral lines in black represent the masses observed from the culture fed with tryptophan. Spectral lines in red represent the masses observed from the culture fed with ^{13}C -W.

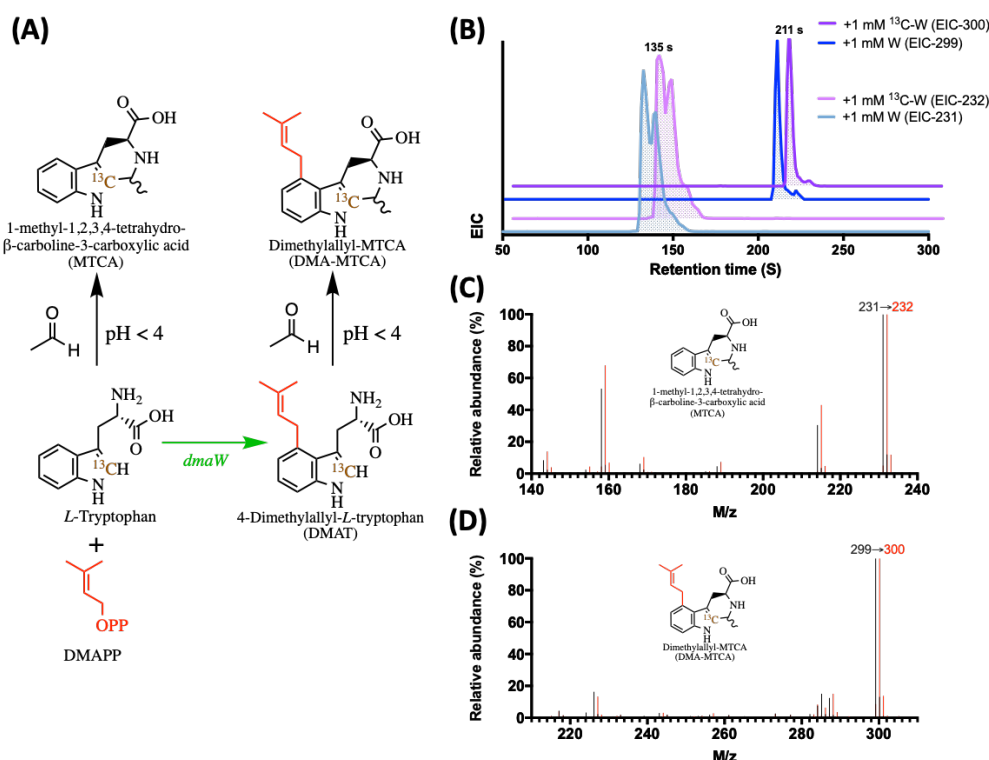


Figure 5-18: Detection of MTCA and DMA-MTCA formation from ^{13}C -labelled tryptophan. (A) Illustration of the reaction producing ^{13}C -labelled MTCA and DMA-MTCA. (B) Extracted ion chromatogram for compounds with $[\text{M}+\text{H}]^+$ of 231 (light blue) and 232 (pink) m/z , corresponding to MTCA and its ^{13}C -labelled version; 299 (blue) and 300 (purple) m/z , corresponding to DMA-MTCA and its ^{13}C -labelled version. Overlay of the mass spectra for the peaks corresponding to (C) MTCA, and (D) DMA-MTCA. Spectral lines in black represent the masses observed from the culture fed with tryptophan. Spectral lines in red represent the masses observed from the culture fed with ^{13}C -W.

In chapter 1 and 3, the proposed mechanisms for the cyclization of the ergoline C ring, by several groups were discussed (Kozikowski et al., 1993). Their work was based on labelling what would be C7 in the ergoline ring, postulated that the cyclization occurs in somewhat two distinct oxidations steps (Figure 5-20). The first oxidation step catalysed by one of the two enzymes, is proposed to hydroxylate 4DMA at the C10 position on the isoprene moiety (**1**). This product then gets dehydrated into a diene intermediate (**2**), which is then epoxidated by the second enzyme using molecular oxygen. Decarboxylation of the alpha carbon would then produce a carbanion on the C5, that would attack the C10 to push open the epoxide ring and in the process form the ergoline C

ring. While the deduction of this mechanism was based on extensive synthetic work, it has not been possible to be experimentally verified from reactions using purified enzymes.

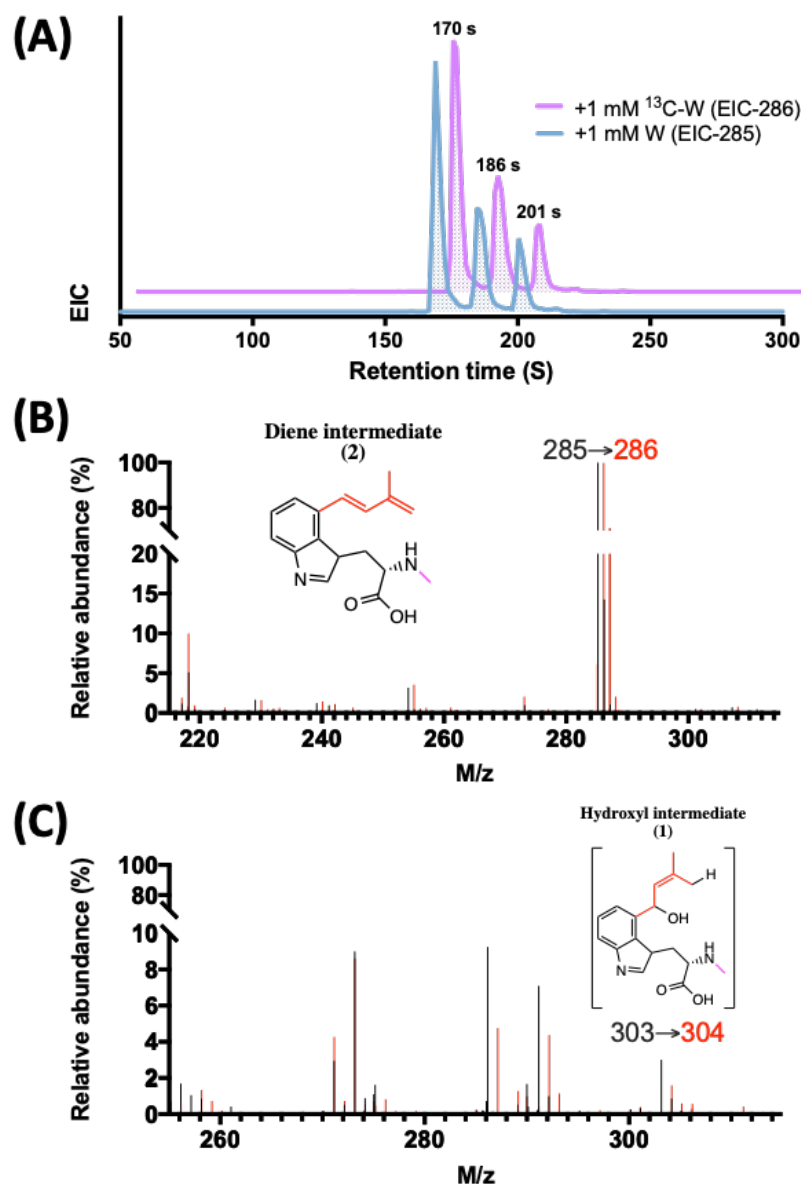


Figure 5-19: Detection of an unknown product displaying the same +1 m/z shift when fed with ^{13}C -W. (A) Extracted ion chromatogram for compounds with an $[M+H]^+$ of 285 and 286 m/z . (B) Overlay of the mass spectra from the peak eluting at 170 s. Structure of the diene intermediate shown for comparison. (C) Observation of trace quantities of an $[M+H]^+$ of 303 → 304 m/z , co-eluting with the 285 → 286 m/z compound. Structure of the hydroxyl intermediate shown for comparison.

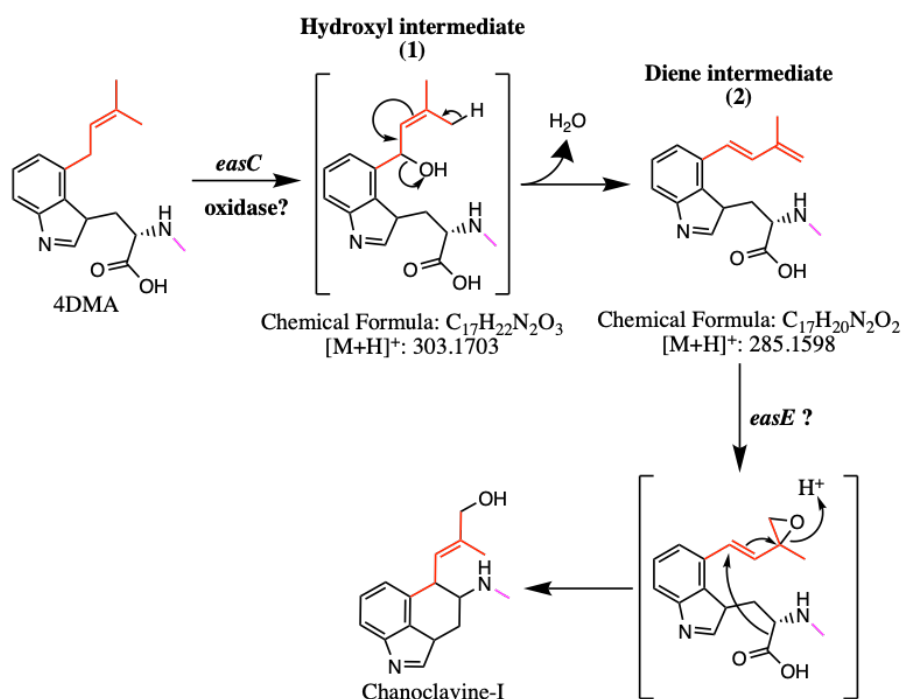


Figure 5-20: Reaction mechanism proposed by Kozikowski et al., 1993, for the formation of the ergoline C ring catalysed by easC and easE.

The minute difference between the theoretical $[M+H]^+$ of both intermediates (1) and (2) to the observed values, suggests that the eluted compounds are likely to be the proposed hydroxyl and diene intermediates, respectively (Table 5-10). Comparisons of the predicted $\text{LogP}_{(\text{octanol/water})}$ values for all the validated compounds and the hypothesized intermediates further support the observed retention time for the diene intermediate (Figure 5-21) (Table 5-11).

Table 5-10: Exact masses of the detected 285, 286, 303, and 304 m/z compounds

Detected masses	Molecular formula	Calculated m/z	Observed m/z	Mass error δ (ppm)
Intermediate (1)	$C_{17}H_{22}N_2O_3$	303.1703	303.1690	-4.29
Intermediate (2)	$C_{17}H_{20}N_2O_2$	285.1598	285.1599	0.35
^{13}C -(1)	$C_{16}^{13}\text{C}H_{22}N_2O_3$	304.1737	304.1710	-8.88
^{13}C -(2)	$C_{16}^{13}\text{C}H_{20}N_2O_2$	286.1631	286.1627	-1.40

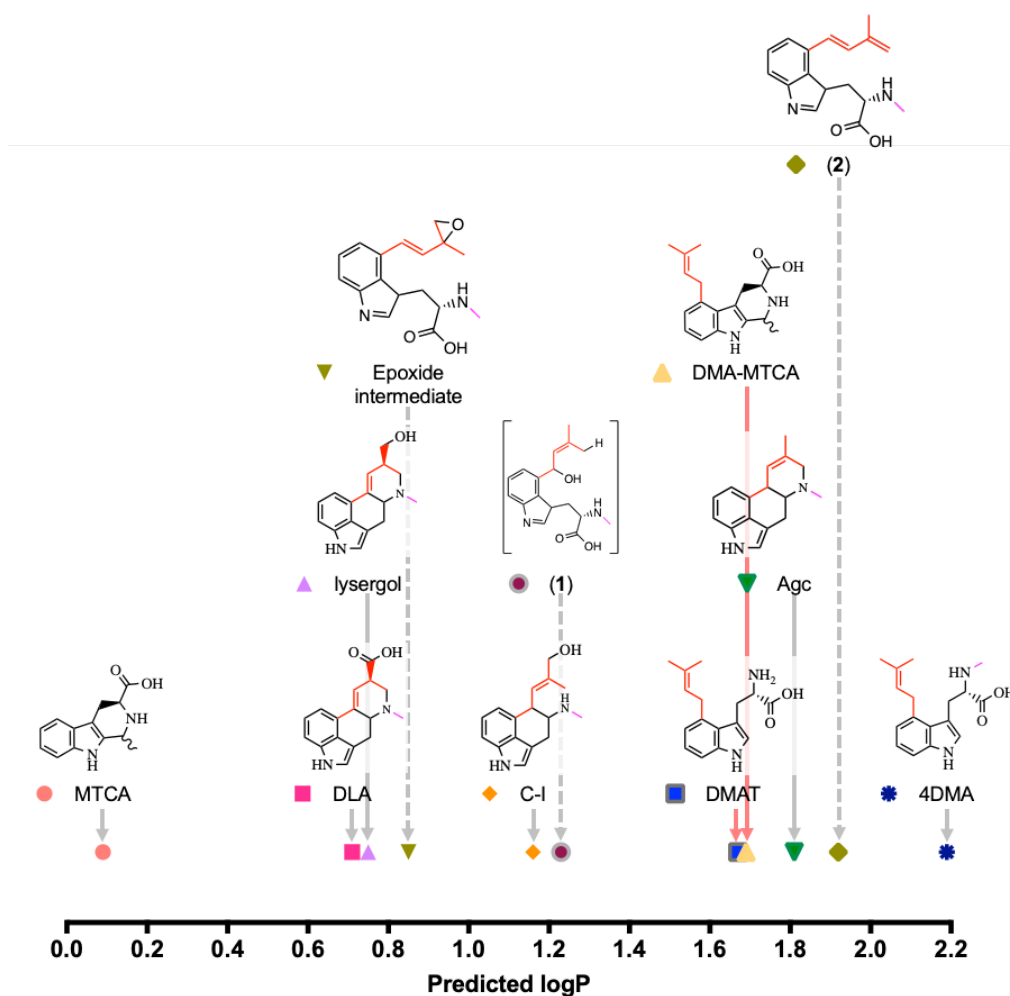


Figure 5-21: Comparison of the predicted LogP values for all the intermediates of the ergot alkaloid pathway in relation to the proposed detected intermediates. Values of LogP were calculated by Chemdraw®, with an expected standard deviation of 0.5 LogP units.

If the detected compounds are indeed the diene and hydroxyl intermediates, their detection from the *in vivo* biosynthesis would further solidify the validity of the proposed reaction mechanism. Further to this point, observing the accumulation of the diene intermediate in the background of an increased the copy number of *easC*, would support the hypothesis that *easC* is responsible for the hydroxylation reaction that leads to the formation of the diene intermediate. Therefore, alluding to *easE* being responsible for the epoxidation of the resulting diene.

Table 5-11: Comparison of the predicted LogP values against their corresponding experimentally determined retention times

Compound	Predicted LogP	Observed Retention time (s)
MTCA	0.09	135
DLA	0.71	125
Lysergol	0.75	127
Epoxide intermediate	0.85	-
chanoclavine-I	1.16	140
Intermediate (1)	1.23	-
DMAT	1.67*	202
DMA-MTCA	1.69*	211
Agroclavine	1.81	169
Intermediate (2)	1.92	-
4DMA	2.19	204

*: Instances where predicted LogP values did not corroborate with observed retention times.

Apart from shedding more light into the mechanism of the ergoline C ring formation, the detected accumulation of these intermediates in the culture media from *in vivo* biosynthesis has never been reported before. This finding would suggest that the *easC/easE* junction, despite the design overhaul, is still a major chokepoint and further reinforce the inference from chapter 3.

The characterization work discussed in this chapter has been focused on the metabolites that have been extruded into the culture media. A more in-depth analysis into the distribution of metabolic flux throughout the pathway *in vivo*, would bring more useful information to guide strain optimization work. Nevertheless, the knowledge generated here will indubitably be instrumental for guiding the way towards building further design iterations.

CHAPTER 6: CONCLUSIONS AND FUTURE DIRECTIONS

6.1 CONCLUSIONS

The rise of synthetic biology and its associated technologies brings along the hopes that it might do what the advent of synthetic chemistry did in the 19th century and perhaps go even further. Unlike the fundamentals of synthetic chemistry that are usually of petrochemical origin, the beauty in biological systems is that they are an inherently renewable resource and most definitely biodegradable. While the notion of pushing forward the transition into a sustainable bio-based economy is idealistically appealing, it is not currently an entirely feasible prospect. The work described in this thesis has thus sought to contribute to the collective effort of resolving this technological inhibition.

In line with the project goals, chapter two described the adaptations made to the YeastFab assembly methodology that allowed for the rapid and flexible construction of multi-gene pathways into the genome of any *S.cerevisiae* S288C-derived strain. In chapter four, the ease at which the strain design was completely revamped, demonstrated the utility of the adaptations made for rapidly creating new strains or making changes to the existing designs.

Over the course of chapters three, four, and five, the specific orthologs of *easE*, *easA*, and *cloA* required for the reconstitution of the pathway to DLA were progressively identified and incorporated into the prototype strain. This led to the demonstration of the first reported completely heterologous production of agroclavine, DLA, and lysergol in yeast.

From the characterization work on DLAM33B described in chapter five, two significant bottlenecks in the pathway that are severely limiting the production capability of the current design have been pointed out. The first point

is the reaction forming the ergoline C ring, catalysed by *easC* and *easE*. As demonstrated from the accumulation and detection of the diene intermediate of that reaction in chapter five, as well as the lack of metabolic flux through the junction even after the inhibition of the formation of the β -carboline shunt products in chapter three.

The second and more apparent point being the reaction catalysed by *cloA*, where supplementary redox support systems are required to restore the catalytic properties of *cloA*. Restoring the catalytic nature of this reaction would not only serve to improve DLA yield through the direct production of DLA, but by increasing the metabolic flux through the pathway *via* the removal of agroclavine accumulation and its associated inhibition on *dmaW* as well (L.-J. Cheng et al., 1980). These are all quintessential in the strain and bioprocess optimization process, to achieve the overarching goal of developing the next generation of microbial cell factories for the production medium to high value targets.

Slightly tangential to the original goals set out for this project are the implications for the detection of the diene intermediate from the *in vivo* biosynthesis reactions. This finding from chapter five, adds to the experimental validation for the reaction mechanism proposed for the formation of the ergoline C ring by *easC* and *easE* (Kozikowski et al., 1993). Additionally, the accumulation of this intermediate against the backdrop of a strain where the copy number of *easC* is disproportionately more than *easE*, might hint at a possible allocation of responsibilities of each enzyme for the hydroxylation and epoxidation reactions, respectively.

The work described in chapter three also identified an alternative functional ortholog of *easE*; collating more functional sequences of *easE* would help build a consensus sequence of the elements required for its functional expression and aid the analysis of how and where the amino acid sequence predictions for the rest of the orthologs could have gone wrong. These findings that might allow the enzyme to eventually be successfully purified for mechanistic and structural studies, to definitively validate the roles *easC* and *easE* perform.

Finally, the serendipitous demonstration of lysergol production from the *P. ipomoeae* ortholog of *cloA*, has provided compelling cause to rethink our understanding of the reactions catalysed by *cloA*. Another exciting avenue of discovery has opened up to be discussed further in future directions.

6.2 FUTURE DIRECTIONS

6.2.1 OPTIMIZATION, METABOLIC ENGINEERING, AND BIOPROCESS CONTROL

In this presented work, the development and characterization of DLAM33B have quite amply set the stage for an economically competitive microbial cell factory for the production of the ergoline scaffold. On top of the identified bottlenecks, several other variables could be further investigated to achieve the commercially competitive yields required.

The simplest and foremost would be to work out the bioprocessing controls that would help achieve this goal. It was found that pH control measures that restrict MTCA and DMA-MTCA formation did not improve chanoclavine-I production in the plasmid-based strain (pMKU-C₁₃E₇) (Chapter 3), but there ought to still be some merit to it. Sarah O'Connor's group used a

bioreactor, controlling for such variables in their demonstration for the production of cycloclavine to a yield of approximately 0.5 g/L (Jakubczyk et al., 2015). A follow-up on the optimization of these bioprocessing variables such as pH, temperature, aeration, and feed, upon the inclusion of adequate redox partners for *cloA*, could conceivably improve DLA yields to within the same order of magnitude.

Further strain optimization could also be performed by attempting to tune the expression levels of the enzymes within the pathway. This approach is usually an extremely laborious process, but by adopting the Design Of Experiment (DOE) methods that were traditionally used in bioprocess optimization (Telen, Logist, Van Derlinden, Tack, & Van Impe, 2012), and implemented with automation, would become a viable approach to improve product yields. The adaptations made to the YeastFab assembly methodology were expressly designed for this eventuality. On one end, the use of DOE software and its accompanying statistical inference methods would help limit the design space required to be sampled, while the use of the modified YeastFab assembly with automation on the other end would expedite and reduce the cost of the construct assembly process and its iterations.

Alternatively, a more randomized approach could be taken by replicating the strain (DLAM33B) in a synthetic yeast host, and thereafter genetic diversity could be artificially created through the use of SCRaMbLE (Jia et al., 2018; Jovicevic, Blount, & Ellis, 2014). Central to this method is the design of multiple loxP sites interspersed within the synthetic chromosomes of the synthetic yeast. Exposure of these sites to the Cre recombinase allows for the “shuffling” between these sites for the induction of genomic diversity.

Screening the libraries of the strain variants created has been successfully used by several groups to identify strains with alterations that enabled the improved production of their target products (Blount et al., 2018; Liu et al., 2018). Such a method would also likely help streamline the identification of modifications to tryptophan degradation pathways or regulatory mechanisms that might improve product yield.

Lastly, engineering key limiting enzymes, such as *easC*, *easE*, and *cloA*, for improved catalytic efficiency would also help to increase production levels. Given the existing knowledge of these enzymes, trials with *easC-easE* or *cloA*-CYP450 reductase fusions or other methods targeting the co-localization of these partner enzymes could be another viable avenue to explore.

Regardless of the approach, bringing the production yield of DLAM33B to 1 g/L would be the basic requirement for a strain useful for commercial applications. DLAM33B has quite adequately served its role as a proof of concept, actually translating it into something of industrial value would be the metaphorical cherry on top.

6.2.2 DIVERSIFICATION OF OXIDIZED TETRACYCLIC ERGOLINE PRODUCTS

The reaction oxidizing tetracyclic clavines, described in chapter five is one of the few points in the ergot alkaloid pathway that is still not very well understood. The existing literature on the reaction and the enzyme responsible for it, *cloA*, are angled towards the one-size-fits-all view that either *cloA* progressively oxidizes agroclavine into elymoclavine followed by a re-uptake and second oxidation to paspalic acid, or some other CYP450 performs the first oxidation and *cloA* performs the second.

The observed product profiles of the tested *cloA* orthologs, and the results presented by the Pannaccione group (Robinson & Panaccione, 2014), might however suggest that this is not entirely true. If *cloA* were to indeed perform a single two-electron oxidation at each pass, there should have been a substantial build-up of the intermediate product, elymoclavine. Especially within the DLAM33B system, which appeared to be limited in its capability to regenerate the redox capability of *cloA*. Instead, the products formed were either DLA alone or DLA and lysergol (in the *P. ipomoeae* ortholog).

We propose that the full linear progression of agroclavine to elymoclavine followed by paspalic acid and DLA might actually be catalysed entirely by a single enzyme, *cloA*. The varied product specificity observed is instead a result of some form of premature reaction termination within the different orthologs, where the sequence variation between orthologs control the extent of oxidation it is capable of performing.

The case for *cloA*'s capability to perform the double oxidations is partially supported by the finding that the *ent*-kaurene oxidases from the biosynthesis of gibberellins are considerably closely related to *cloA*. Several sequences of these *ent*-kaurene oxidases were picked up in the SSN of the *cloA* supercluster, within the cluster containing the CYP450s from the Fusarium family of trichodiene oxidases (Figure 5-2). The *ent*-kaurene oxidases have been previously shown to perform the complete oxidation of a methyl group to a carboxylic acid (Helliwell, Chandler, Poole, Dennis, & Peacock, 2001; Tudzynski, Hedden, Carrera, & Gaskin, 2001).

On the other hand, the divergent product specificities are supported by the apparent delineation of sub-clusters in the SSN by their clades and

associated divergent product profiles. The demonstration of lysergol production in the *P. ipomoeae* ortholog further solidifies this explanation. Furthermore, the presence of multiple *cloA* sequences picked up from the same organism in the SSN might account for the presence of the species presenting more diverse clavine and DLA derivative profiles.

Finally, given the high degree of selectivity the *cloA* enzymes show towards the C8-C9 double bond and the tertiary amino group (Maier et al., 1988), the isomerization of the C8-C9 double bond to the C9-C10 in lysergol and DLA might also serve as one mechanism for terminating the reaction and releasing the product. Further research into the *cloA* reaction mechanism and sequence diversity might show how these oxidation extents are determined, and possibly guide the development of novel DLA derivatives.

6.2.3 BIOSYNTHESIS OF HALOGENATED ERGOT ALKALOIDS

One attractive prospect of having a strain capable of heterologously producing DLA, is the ability to then use it as a platform for the *in vivo* derivatization of novel ergoline derivatives. This falls within the scope of expanding natural metabolism into the synthetic (Erb, Jones, & Bar-Even, 2017; Martin, Nielsen, Solomon, & Prather, 2009), and a particularly attractive target group are the development of halogenated ergoline derivatives.

Running through the semi-synthetic ergoline therapeutics available on the market and one would notice that of the six therapeutics being produced at more than 1000 kg per year (Cvak, 1999), two of them (Nicergoline and Bromocriptine) have a halogen appended to their structure (Figure 1-1). Diversification of the ergoline scaffold with halogens would likely yield more novel valuable products out of the pathway, or even enable further levels of

diversification through cross-coupling reactions utilizing the halogenated group (Maluenda & Navarro, 2015).

In 1998 and 2001, a Russian group reported the isolation and elucidation of the chemical structure for the compounds pibocin and pibocin B (Makarieva, Dmitrenok, Dmitrenok, Grebnev, & Stonik, 2001; Makarieva, Ilyin, Stonik, Lyssenko, & Denisenko, 1999). These were not only the first ergoline compounds isolated from marine organisms, but also the first natural halogenated ergot alkaloids isolated. Structural characterization of pibocin and pibocin B revealed the presence of a bromine substituted on the C2 position of the indole portion of the ergoline scaffold. Genome mining of the ascidian, *Eudistoma sp.* where these bromo-2-ergoline compounds were isolated from, could potentially unveil orthologous enzymes from the ergot alkaloid pathway that accepts halogenated tryptophan or a halogenase capable of performing this modification directly on the ergoline scaffold.

Finding and incorporating these undiscovered ergoline halogenases into the reconstituted pathway would certainly revolutionize the way the current semi-synthetic ergoline-derived therapeutics are manufactured.

6.3 LEARNING THROUGH CREATING

Through the course of this attempt at re-constructing the ergot alkaloid pathway, several findings relating to the fundamental understanding of this process were brought to light. Gaps in our knowledge were highlighted, and several new questions were raised. In this sense, through the process of creation or rather “re-creation”, we have pushed forward our understanding of the ergot alkaloid biosynthesis system. A core feature in the field of synthetic biology that strongly resonates with some of the famous last words from Richard P.

Feynman – “What I cannot create, I do not understand. Know to solve every problem that has been solved.”

REFERENCES

- Apel, A. R., d'Espaux, L., Wehrs, M., Sachs, D., Li, R. A., Tong, G. J., . . . Hillson, N. J. (2016). A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic acids research*, gkw1023.
- Arnold, S. L., & Panaccione, D. G. (2017). Biosynthesis of the pharmaceutically important fungal ergot alkaloid dihydrolysergic acid requires a specialized allele of *cloA*. *Appl. Environ. Microbiol.*, 83(14), e00805-00817.
- Backhaus, K., Ludwig-Radtke, L., Xie, X., & Li, S.-M. (2017). Manipulation of the precursor supply in yeast significantly enhances the accumulation of prenylated β -carboline. *ACS synthetic biology*, 6(6), 1056-1064.
- Bai Flagfeldt, D., Siewers, V., Huang, L., & Nielsen, J. (2009). Characterization of chromosomal integration sites for heterologous gene expression in *Saccharomyces cerevisiae*. *Yeast*, 26(10), 545-551.
- Bilovol, Y., & Panaccione, D. G. (2016). Functional analysis of the gene controlling hydroxylation of festuclavine in the ergot alkaloid pathway of *Neosartorya fumigata*. *Current genetics*, 62(4), 853-860.
- Blount, B., Gowers, G. F., Ho, J., Ledesma-Amaro, R., Jovicevic, D., McKiernan, R., . . . Ellis, T. (2018). Rapid host strain improvement by in vivo rearrangement of a synthetic yeast chromosome. *Nature communications*, 9(1), 1932.
- Brent, M. R., & Guigo, R. (2004). Recent advances in gene structure prediction. *Current opinion in structural biology*, 14(3), 264-272.
- Carrier, T., Jones, K. L., & Keasling, J. (1998). mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system. *Biotechnology and bioengineering*, 59(6), 666-672.
- Casini, A., Storch, M., Baldwin, G. S., & Ellis, T. (2015). Bricks and blueprints: methods and standards for DNA assembly. *Nature Reviews Molecular Cell Biology*, 16(9), 568.
- Cheng, J. Z., Coyle, C. M., Panaccione, D. G., & O'Connor, S. E. (2010a). Controlling a structural branch point in ergot alkaloid biosynthesis. *Journal of the American Chemical Society*, 132(37), 12835-12837.
- Cheng, J. Z., Coyle, C. M., Panaccione, D. G., & O'Connor, S. E. (2010b). A role for old yellow enzyme in ergot alkaloid biosynthesis. *Journal of the American Chemical Society*, 132(6), 1776-1777.
- Cheng, L.-J., Robbers, J. E., & Floss, H. G. (1980). End-product regulation of ergot alkaloid formation in intact cells and protoplasts of *Claviceps* species, strain SD 58. *Journal of Natural Products*, 43(3), 329-339.
- Chilton, A. S., Ellis, A. L., & Lamb, A. L. (2014). Structure of an *Aspergillus fumigatus* old yellow enzyme (EasA) involved in ergot alkaloid biosynthesis.

Acta Crystallographica Section F: Structural Biology Communications, 70(10), 1328-1332.

Colletti, P. F., Goyal, Y., Varman, A. M., Feng, X., Wu, B., & Tang, Y. J. (2011). Evaluating factors that influence microbial synthesis yields by linear regression with numerical and ordinal variables. *Biotechnology and bioengineering*, 108(4), 893-901.

Cook, D., Finnigan, J., Cook, K., Black, G., & Charnock, S. (2016). Cytochromes P450: history, classes, catalytic mechanism, and industrial application. In *Advances in protein chemistry and structural biology* (Vol. 105, pp. 105-126): Elsevier.

Coyle, C. M., Cheng, J. Z., O'Connor, S. E., & Panaccione, D. G. (2010). An old yellow enzyme gene controls the branch point between *Aspergillus fumigatus* and *Claviceps purpurea* ergot alkaloid pathways. *Appl. Environ. Microbiol.*, 76(12), 3898-3903.

Cvak, L. (1999). 13. INDUSTRIAL PRODUCTION OF ERGOT ALKALOIDS. In *Ergot: the genus Claviceps* (pp. 373): CRC Press.

de Groot, A. N., van Dongen, P. W., Vree, T. B., Hekster, Y. A., & van Roosmalen, J. (1998). Ergot alkaloids. *Drugs*, 56(4), 523-535.

Dorner, J., Cole, R., Hill, R., Wicklow, D., & Cox, R. (1980). *Penicillium rubrum* and *Penicillium biforme*, new sources of rugulovasines A and B. *Appl. Environ. Microbiol.*, 40(3), 685-687.

dos Santos, R. G., Osorio, F. L., Crippa, J. A. S., Riba, J., Zuardi, A. W., & Hallak, J. E. (2016). Antidepressive, anxiolytic, and antiaddictive effects of ayahuasca, psilocybin and lysergic acid diethylamide (LSD): a systematic review of clinical trials published in the last 25 years. *Therapeutic Advances in Psychopharmacology*, 6(3), 193-213.

Duarte, C. D., Barreiro, E. J., & Fraga, C. A. (2007). Privileged structures: a useful concept for the rational design of new lead drug candidates. *Mini reviews in medicinal chemistry*, 7(11), 1108-1119.

Eich, E., & Pertz, H. (1999). 15. ANTIMICROBIAL AND ANTITUMOR EFFECTS OF ERGOT ALKALOIDS AND THEIR DERIVATIVES. In *Ergot: the genus Claviceps* (pp. 441): CRC Press.

Endy, D. (2005). Foundations for engineering biology. *Nature*, 438(7067), 449.

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PloS one*, 3(11), e3647.

Erb, T. J., Jones, P. R., & Bar-Even, A. (2017). Synthetic metabolism: metabolic engineering meets enzyme design. *Current opinion in chemical biology*, 37, 56-62.

Erickson, B., & Winters, P. (2012). Perspective on opportunities in industrial biotechnology in renewable chemicals. *Biotechnology journal*, 7(2), 176-185.

Floss, H. G., Hornemann, U., Schilling, N., Kelley, K., Groeger, D., & Erge, D. (1968). Biosynthesis of ergot alkaloids. Evidence for two isomerizations in the isoprenoid moiety during the formation of tetracyclic ergolines. *Journal of the American Chemical Society*, 90(23), 6500-6507.

Furuta, T., Koike, M., & Abe, M. (1982). Isolation of Cycloclavine from the Culture Broth of *Aspergillus japonicus* SAITO. *Agricultural and Biological Chemistry*, 46(7), 1921-1922.

Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F., & Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science*, 349(6252), 1095-1100.

Gerhards, N., & Li, S.-M. (2017). A bifunctional old yellow enzyme from *Penicillium roqueforti* is involved in ergot alkaloid biosynthesis. *Organic & biomolecular chemistry*, 15(38), 8059-8071.

Gerlt, J. A., Bouvier, J. T., Davidson, D. B., Imker, H. J., Sadkhin, B., Slater, D. R., & Whalen, K. L. (2015). Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics*, 1854(8), 1019-1037.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, 6(5), 343-345.

Gietz, R. D., & Schiestl, R. H. (2007). Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nature protocols*, 2(1), 1.

Goetz, K. E., Coyle, C. M., Cheng, J. Z., O'Connor, S. E., & Panaccione, D. G. (2011). Ergot cluster-encoded catalase is required for synthesis of chanoclavine-I in *Aspergillus fumigatus*. *Current genetics*, 57(3), 201.

Gonzalez, F. J., & Korzekwa, K. R. (1995). Cytochromes P450 expression systems. *Annual review of pharmacology and toxicology*, 35(1), 369-390.

Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., . . . Zheng, Y. (2015). YeastFab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*. *Nucleic acids research*, 43(13), e88-e88.

Haarmann, T., Ortel, I., Tudzynski, P., & Keller, U. (2006). Identification of the cytochrome P450 monooxygenase that bridges the clavine and ergoline alkaloid pathways. *ChemBioChem*, 7(4), 645-652.

Havemann, J., Vogel, D., Loll, B., & Keller, U. (2014). Cyclolization of D-lysergic acid alkaloid peptides. *Chemistry & biology*, 21(1), 146-155.

- Helliwell, C. A., Chandler, P. M., Poole, A., Dennis, E. S., & Peacock, W. J. (2001). The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proceedings of the National Academy of Sciences*, 98(4), 2065-2070.
- Hofmann, A. (2010). *LSD-mein Sorgenkind: die Entdeckung einer "Wunderdroge"*: Klett-Cotta.
- Hofmann, A., & Tschertter, H. (1960). Isolierung von Lysergsäure-Alkaloiden aus der mexikanischen Zauberdroge Ololiuqui (*Rivea corymbosa* (L.) Hall. f.). *Experientia*, 16(9), 414-414.
- Hohnholz, R., & Achstetter, T. (2019). Recombination in yeast based on six base pairs of homologous sequences: structural instability in two sets of isomeric model expression plasmids. *Yeast*.
- Hu, Z., Killion, P. J., & Iyer, V. R. (2007). Genetic reconstruction of a functional transcriptional regulatory network. *Nature genetics*, 39(5), 683.
- Hu, Z.-F., Gu, A.-D., Liang, L., Li, Y., Gong, T., Chen, J.-J., . . . Zhu, P. (2019). Construction and optimization of microbial cell factories for sustainable production of bioactive dammarenediol-II glucosides. *Green Chemistry*.
- Hua, S.-b., Qiu, M., Chan, E., Zhu, L., & Luo, Y. (1997). Minimum Length of Sequence Homology Required for in Vivo Cloning by Homologous Recombination in Yeast. *Plasmid*, 38(2), 91-96.
- Hughes, R. A., & Ellington, A. D. (2017). Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harbor perspectives in biology*, 9(1), a023812.
- Jakubczyk, D., Caputi, L., Hatsch, A., Nielsen, C. A., Diefenbacher, M., Klein, J., . . . Naesby, M. (2015). Discovery and reconstitution of the cycloclavine biosynthetic pathway—enzymatic formation of a cyclopropyl group. *Angewandte Chemie International Edition*, 54(17), 5117-5121.
- Jia, B., Wu, Y., Li, B.-Z., Mitchell, L. A., Liu, H., Pan, S., . . . Li, B. (2018). Precise control of SCRaMbLE in synthetic haploid and diploid yeast. *Nature communications*, 9(1), 1933.
- Jovicevic, D., Blount, B. A., & Ellis, T. (2014). Total synthesis of a eukaryotic chromosome: Redesigning and SCRaMbLE-ing yeast. *Bioessays*, 36(9), 855-860.
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, 10, 845. doi:10.1038/nprot.2015.053
- Kobel, H., Schreier, E., & Rutschmann, J. (1964). 6-Methyl- Δ^8 , 9-ergolen-8-carbonsäure, ein neues Ergolinderivat aus Kulturen eines Stammes von *Claviceps paspali* STEVENSET HALL. 60. Mitteilung über Mutterkornalkaloide. *Helvetica Chimica Acta*, 47(4), 1052-1064.

- Kozikowski, A. P., Chen, C., Wu, J. P., Shibuya, M., Kim, C. G., & Floss, H. G. (1993). Probing ergot alkaloid biosynthesis: Intermediates in the formation of ring C. *Journal of the American Chemical Society*, 115(6), 2482-2488.
- Kozikowski, A. P., Wu, J. P., Shibuya, M., & Floss, H. G. (1988). Probing ergot alkaloid biosynthesis: identification of advanced intermediates along the biosynthetic pathway. *Journal of the American Chemical Society*, 110(6), 1970-1971.
- Kozlovsky, A. G. (1999). 18. PRODUCERS OF ERGOT ALKALOIDS OUT OF CLAVICEPS GENUS. In *Ergot: the genus Claviceps* (pp. 479): CRC Press.
- Kren, V., & Cvak, L. (1999). 7. Alkaloids and Other Metabolites of the Genus Claviceps. In *Ergot: the genus Claviceps* (pp. 188-216): CRC Press.
- Kumar, S. (2010). Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation. *Expert opinion on drug metabolism & toxicology*, 6(2), 115-131.
- Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. *ACS synthetic biology*, 4(9), 975-986.
- Lee, S.-L., Floss, H. G., & Heinstein, P. (1976). Purification and properties of dimethylallylpyrophosphate: tryptophan dimethylallyl transferase, the first enzyme of ergot alkaloid biosynthesis in Claviceps. sp. SD 58. *Archives of biochemistry and biophysics*, 177(1), 84-94.
- Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G., & Lieb, J. D. (2012). Genome-wide protein–DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature*, 484(7393), 251.
- Lieberman, A., Kupersmith, M., Estey, E., & Goldstein, M. (1976). Treatment of Parkinson's disease with bromocriptine. *New England Journal of Medicine*, 295(25), 1400-1404.
- Liechti, M. E. (2017). Modern clinical research on LSD. *Neuropsychopharmacology*, 42(11), 2114.
- Lin, H. C., Chiou, G., Chooi, Y. H., McMahon, T. C., Xu, W., Garg, N. K., & Tang, Y. (2015). Elucidation of the concise biosynthetic pathway of the communesin indole alkaloids. *Angewandte Chemie International Edition*, 54(10), 3004-3007.
- Liu, W., Luo, Z., Wang, Y., Pham, N. T., Tuck, L., Pérez-Pi, I., . . . Auer, M. (2018). Rapid pathway prototyping and engineering using in vitro and in vivo synthetic genome SCRaMbLE-in methods. *Nature communications*, 9(1), 1936.
- Löoke, M., Kristjuhan, K., & Kristjuhan, A. (2011). Extraction of genomic DNA from yeasts for PCR-based applications. *Biotechniques*, 50(5), 325-328.

- Lorenz, N., Olšovská, J., Šulc, M., & Tudzynski, P. (2010). Alkaloid cluster gene *ccsA* of the ergot fungus *Claviceps purpurea* encodes chanoclavine I synthase, a flavin adenine dinucleotide-containing oxidoreductase mediating the transformation of N-methyl-dimethylallyltryptophan to chanoclavine I. *Appl. Environ. Microbiol.*, 76(6), 1822-1830.
- Lorenz, N., Wilson, E. V., Machado, C., Schardl, C. L., & Tudzynski, P. (2007). Comparison of ergot alkaloid biosynthesis gene clusters in *Claviceps* species indicates loss of late pathway steps in evolution of *C. fusiformis*. *Appl. Environ. Microbiol.*, 73(22), 7185-7191.
- Loveless, A. (1971). Conidial evidence for host restriction in *Claviceps purpurea*. *Transactions of the British Mycological Society*, 56(3), 419-IN415.
- Ly, C., Greb, A. C., Cameron, L. P., Wong, J. M., Barragan, E. V., Wilson, P. C., . . . Paddy, M. R. (2018). Psychedelics promote structural and functional neural plasticity. *Cell reports*, 23(11), 3170-3182.
- Maier, W., Schumann, B., & Gröger, D. (1988). Microsomal oxygenases involved in ergoline alkaloid biosynthesis of various *Claviceps* strains. *Journal of basic microbiology*, 28(1-2), 83-93.
- Makarieva, T. N., Dmitrenok, A. S., Dmitrenok, P. S., Grebnev, B. B., & Stonik, V. A. (2001). Pibocin B, the first N-O-methylindole marine alkaloid, a metabolite from the Far-Eastern ascidian *Eudistoma* species. *Journal of Natural Products*, 64(12), 1559-1561.
- Makarieva, T. N., Ilyin, S. G., Stonik, V. A., Lyssenko, K. A., & Denisenko, V. A. (1999). Pibocin, the first ergoline marine alkaloid from the Far-Eastern ascidian *Eudistoma* sp. *Tetrahedron letters*, 40(8), 1591-1594.
- Malinka, Z. (1999). 12. SAPROPHYTIC CULTIVATION OF CLAVICHPS. In *Ergot: the genus Claviceps* (pp. 321): CRC Press.
- Maluenda, I., & Navarro, O. (2015). Recent developments in the Suzuki-Miyaura reaction: 2010–2014. *Molecules*, 20(5), 7528-7557.
- Mantegani, S., Brambilla, E., & Varasi, M. (1999). Ergoline derivatives: receptor affinity and selectivity. *Il Farmaco*, 54(5), 288-296.
- Martin, C. H., Nielsen, D. R., Solomon, K. V., & Prather, K. L. J. (2009). Synthetic metabolism: engineering biology at the protein and pathway scales. *Chemistry & biology*, 16(3), 277-286.
- Mathé, C., Sagot, M. F., Schiex, T., & Rouzé, P. (2002). Current methods of gene prediction, their strengths and weaknesses. *Nucleic acids research*, 30(19), 4103-4117.
- Matuschek, M., Wallwey, C., Wollinsky, B., Xie, X., & Li, S.-M. (2012). In vitro conversion of chanoclavine-I aldehyde to the stereoisomers festuclavine and pyroclavine controlled by the second reduction step. *RSC Advances*, 2(9), 3662-3669.

- Matuschek, M., Wallwey, C., Xie, X., & Li, S.-M. (2011). New insights into ergot alkaloid biosynthesis in *Claviceps purpurea*: an agroclavine synthase EasG catalyses, via a non-enzymatic adduct with reduced glutathione, the conversion of chanoclavine-I aldehyde to agroclavine. *Organic & biomolecular chemistry*, 9(11), 4328-4335.
- Metzger, U., Schall, C., Zocher, G., Unsöld, I., Stec, E., Li, S.-M., . . . Stehle, T. (2009). The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria. *Proceedings of the National Academy of Sciences*, 106(34), 14309-14314.
- Meunier, B., de Visser, S. P., & Shaik, S. (2004). Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chemical reviews*, 104(9), 3947-3980.
- Minghetti, A., & Crespi-Perellino, N. (1999). THE HISTORY OF ERGOT. In *Ergot* (pp. 16-39): CRC Press.
- Mrusek, M., Seo, E.-J., Greten, H. J., Simon, M., & Efferth, T. (2015). Identification of cellular and molecular factors determining the response of cancer cells to six ergot alkaloids. *Investigational new drugs*, 33(1), 32-44.
- Murakami, H., Yabusaki, Y., Sakaki, T., Shibata, M., & Ohkawa, H. (1990). Expression of cloned yeast NADPH-cytochrome P450 reductase gene in *Saccharomyces cerevisiae*. *The Journal of Biochemistry*, 108(5), 859-865.
- Németh, É. (1999). 11. PARASITIC PRODUCTION OF ERGOT ALKALOIDS. In *Ergot: the genus Claviceps*: CRC Press.
- Nielsen, C. A., Folly, C., Hatsch, A., Molt, A., Schröder, H., O'Connor, S. E., & Naesby, M. (2014). The important ergot alkaloid intermediate chanoclavine-I produced in the yeast *Saccharomyces cerevisiae* by the combined action of EasC and EasE from *Aspergillus japonicus*. *Microbial cell factories*, 13(1), 1.
- Ortel, I., & Keller, U. (2009). Combinatorial assembly of simple and complex D-lysergic acid alkaloid peptide classes in the ergot fungus *Claviceps purpurea*. *Journal of Biological Chemistry*, 284(11), 6650-6660.
- Otsuka, H., Quigley, F., Gröger, D., Anderson, J., & Floss, H. (1980). In Vivo and In Vitro Evidence for N-Methylation as the Second Pathway-Specific Step in Ergoline Biosynthesis. *Planta Medica*, 40(10), 109-119.
- Paddon, C. J., Westfall, P. J., Pitera, D., Benjamin, K., Fisher, K., McPhee, D., . . . Eng, D. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446), 528-532.
- Paracelsus, T. (1965). Die dritte Defension wegen des Schreibens der neuen Rezepte. *Septem Defensiones*, 1538, 510.
- Partow, S., Siewers, V., Bjørn, S., Nielsen, J., & Maury, J. (2010). Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast*, 27(11), 955-964.

- Pertz, H., & Eich, E. (1999). Ergot alkaloids and their derivatives as ligands for serotonergic, dopaminergic, and adrenergic receptors. *Ergot: The Genus Claviceps*. Amsterdam: Harwood Academic Publishers, 411-440.
- Plank, L., & Harvey, J. (1979). Generation time statistics of *Escherichia coli* B measured by synchronous culture techniques. *Microbiology*, 115(1), 69-77.
- Pundir, S., Martin, M. J., O'Donovan, C., & Consortium, U. (2016). UniProt tools. *Current protocols in bioinformatics*, 53(1), 1.29. 21-21.29. 15.
- Quigley, F. R., & Floss, H. G. (1981). Mechanism of amino acid. alpha-hydroxylation and formation of the lysergyl moiety in ergotamine biosynthesis. *The Journal of Organic Chemistry*, 46(2), 464-466.
- Richardson, S. M., Nunley, P. W., Yarrington, R. M., Boeke, J. D., & Bader, J. S. (2010). GeneDesign 3.0 is an updated synthetic biology toolkit. *Nucleic acids research*, 38(8), 2603-2606.
- Rigbers, O., & Li, S.-M. (2008). Ergot alkaloid biosynthesis in *Aspergillus fumigatus* overproduction and biochemical characterization of a 4-dimethylallyltryptophan N-methyltransferase. *Journal of Biological Chemistry*, 283(40), 26859-26868.
- Robinson, S. L., & Panaccione, D. G. (2014). Heterologous expression of lysergic acid and novel ergot alkaloids in *Aspergillus fumigatus*. *Appl. Environ. Microbiol.*, 80(20), 6465-6472.
- Ryan, K., Moore, C., & Panaccione, D. (2013). Partial reconstruction of the ergot alkaloid pathway by heterologous gene expression in *Aspergillus nidulans*. *Toxins*, 5(2), 445-455.
- Ryan, K. L., Akhmedov, N. G., & Panaccione, D. G. (2014). Identification and structural elucidation of ergotryptamine, a new ergot alkaloid produced by genetically modified *Aspergillus nidulans* and natural isolates of *Epichloë* species. *Journal of agricultural and food chemistry*, 63(1), 61-67.
- Salari, R., & Salari, R. (2017). Investigation of the best *Saccharomyces cerevisiae* growth condition. *Electronic physician*, 9(1), 3592.
- Schardl, C. L., Panaccione, D. G., & Tudzynski, P. (2006). Ergot alkaloids—biology and molecular biology. *The alkaloids: chemistry and biology*, 63, 45-86.
- Stec, E., Steffan, N., Kremer, A., Zou, H., Zheng, X., & Li, S. M. (2008). Two lysine residues are responsible for the enzymatic activities of indole prenyltransferases from fungi. *ChemBioChem*, 9(13), 2055-2058.
- Stoev, G., Xuan, Y., Peycheva, M., & Scigelova, M. (2012). Quantitative assessment of the contribution of high resolution mass spectrometric analysis to the reliability of compound confirmation. *Talanta*, 98, 19-27.

Synthetic Biology: Applying Engineering to Biology : Report of a NEST High-Level Expert Group. (2005). (E. C. D. G. f. Research Ed.): Office for Official Publications of the European Communities.

Tandowsky, R. M. (1954). Clinical Evaluation of Combined Hydrogenated Ergot Alkaloids (Hydergine) in Arterial Hypertension: With Special Reference to Their Action in Central Manifestations. *Circulation*, 9(1), 48-56.

Telen, D., Logist, F., Van Derlinden, E., Tack, I., & Van Impe, J. (2012). Optimal experiment design for dynamic bioprocesses: a multi-objective approach. *Chemical Engineering Science*, 78, 82-97.

Toogood, H. S., Gardiner, J. M., & Scrutton, N. S. (2010). Biocatalytic reductions and chemical versatility of the old yellow enzyme family of flavoprotein oxidoreductases. *ChemCatChem*, 2(8), 892-914.

Tudzynski, B., Hedden, P., Carrera, E., & Gaskin, P. (2001). The P450-4 gene of *Gibberella fujikuroi* encodes ent-kaurene oxidase in the gibberellin biosynthesis pathway. *Appl. Environ. Microbiol.*, 67(8), 3514-3522.

Unsöld, I. A., & Li, S.-M. (2005). Overproduction, purification and characterization of FgaPT2, a dimethylallyltryptophan synthase from *Aspergillus fumigatus*. *Microbiology*, 151(5), 1499-1505.

Venters, B. J., Wachi, S., Mavrich, T. N., Andersen, B. E., Jena, P., Sinnamon, A. J., . . . Hemeryck-Walsh, C. (2011). A comprehensive genomic binding map of gene and chromatin regulatory proteins in *Saccharomyces*. *Molecular cell*, 41(4), 480-492.

Wacker, D., Wang, S., McCorvy, J. D., Betz, R. M., Venkatakrishnan, A., Levit, A., . . . Nichols, D. E. (2017). Crystal structure of an LSD-bound human serotonin receptor. *Cell*, 168(3), 377-389. e312.

Walker, G. M. (1998). *Yeast physiology and biotechnology*: John Wiley & Sons.

Wallwey, C., Matuschek, M., & Li, S.-M. (2010). Ergot alkaloid biosynthesis in *Aspergillus fumigatus*: conversion of chanoclavine-I to chanoclavine-I aldehyde catalyzed by a short-chain alcohol dehydrogenase FgaDH. *Archives of microbiology*, 192(2), 127-134.

Wallwey, C., Matuschek, M., Xie, X.-L., & Li, S.-M. (2010). Ergot alkaloid biosynthesis in *Aspergillus fumigatus*: Conversion of chanoclavine-I aldehyde to festuclavine by the festuclavine synthase FgaFS in the presence of the old yellow enzyme FgaOx3. *Organic & biomolecular chemistry*, 8(15), 3500-3508.

Walzel, B., Riederer, B., & Keller, U. (1997). Mechanism of alkaloid cyclopeptide synthesis in the ergot fungus *Claviceps purpurea*. *Chemistry & biology*, 4(3), 223-230.

Winblad, B., Fioravanti, M., Dolezal, T., Logina, I., Milanov, I. G., Popescu, D. C., & Solomon, A. (2008). Therapeutic use of nicergoline. *Clinical drug investigation*, 28(9), 533-552.

- Yamanishi, M., Ito, Y., Kintaka, R., Imamura, C., Katahira, S., Ikeuchi, A., . . . Matsuyama, T. (2013). A genome-wide activity assessment of terminator regions in *Saccharomyces cerevisiae* provides a "terminatome" toolbox. *ACS synthetic biology*, 2(6), 337-347.
- Young, C. A., Schardl, C. L., Panaccione, D. G., Florea, S., Takach, J. E., Charlton, N. D., . . . Jaromczyk, J. (2015). Genetics, genomics and evolution of ergot alkaloid diversity. *Toxins*, 7(4), 1273-1302.
- Yuan, S.-F., & Alper, H. S. (2019). Metabolic engineering of microbial cell factories for production of nutraceuticals. *Microbial cell factories*, 18(1), 46.
- Zhang, M. Q. (2002). Computational prediction of eukaryotic protein-coding genes. *Nature Reviews Genetics*, 3(9), 698.
- Zhu, J., & Zhang, M. Q. (1999). SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics*, 15(7), 607-611.

APPENDIX

1. TPI1-PDA1-TEF2 REGULATION NETWORK

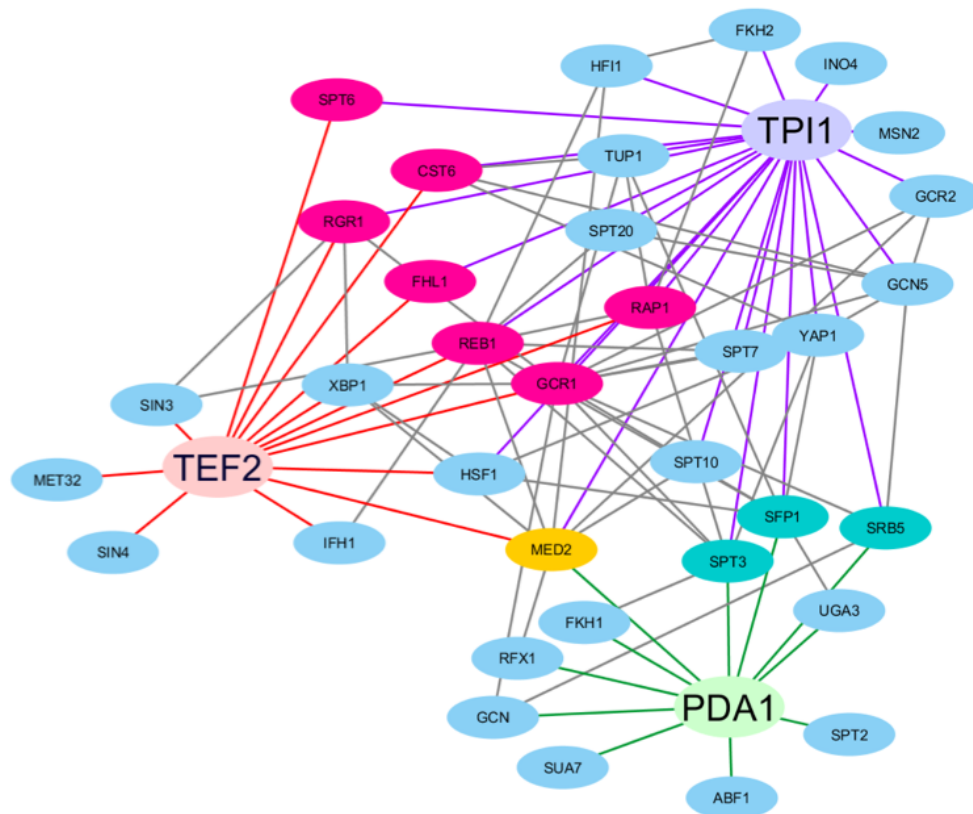


Figure A1-1: Transcriptional regulation network between the genes TPI1, PDA1, TEF2 in *Saccharomyces cerevisiae*. First degree common transcriptional regulators between TEF2- TPI1 are coloured in **red**, TPI1- PDA1 are coloured in **teal**, and TEF2-TPI1-PDA1 are coloured **orange**. Extensive interactivity between the transcriptional regulators of TEF2 and TPI1/PDA1 on both primary and secondary degrees of relationship could account for the discrepancy in observed promoter output as a result of proximal interactions on the plasmid-based reporter system.

2. PLASMID MAPS

2.1. Pathway acceptor vectors (p[M/C]K[U/L/H])

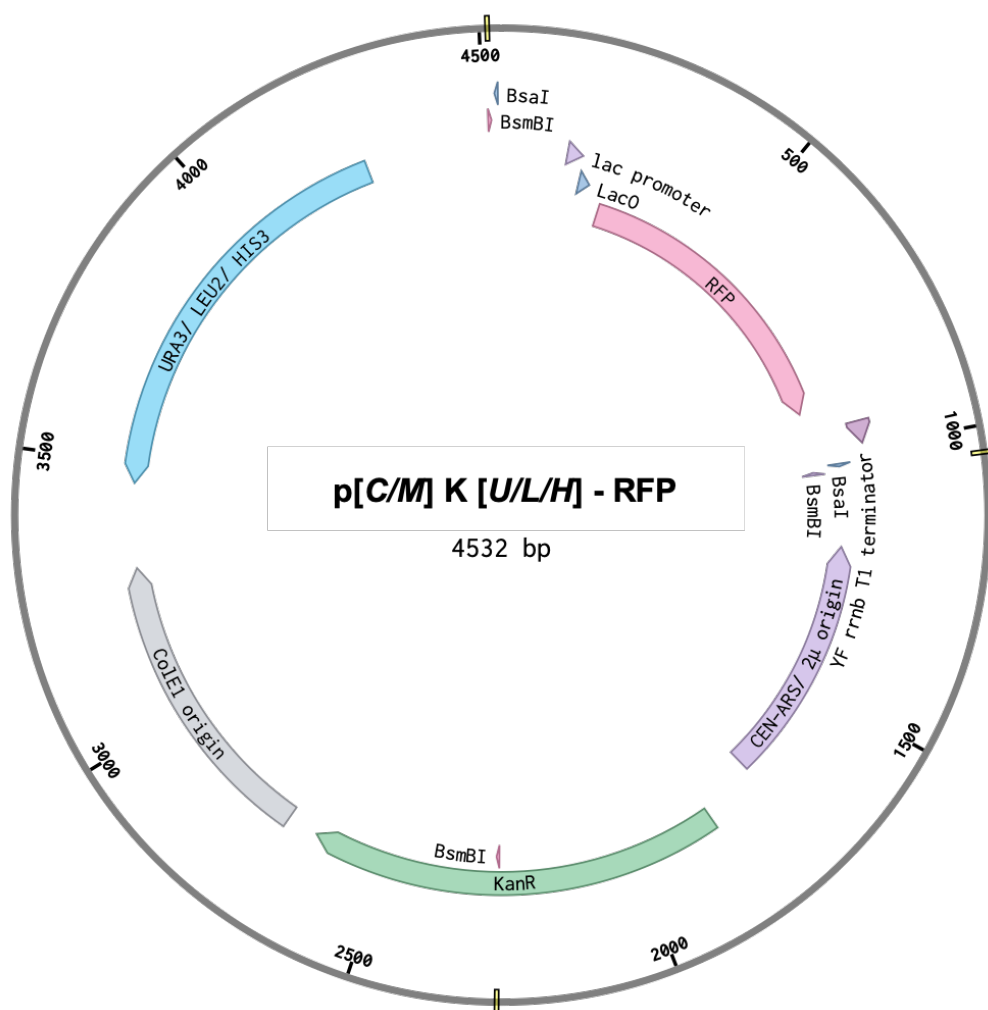


Figure A2-1: Enlarged representative plasmid map for the pathway acceptor vectors created (Chapter 2).

2.2. Genome integration vectors (pGAU-[int site]-RFP)

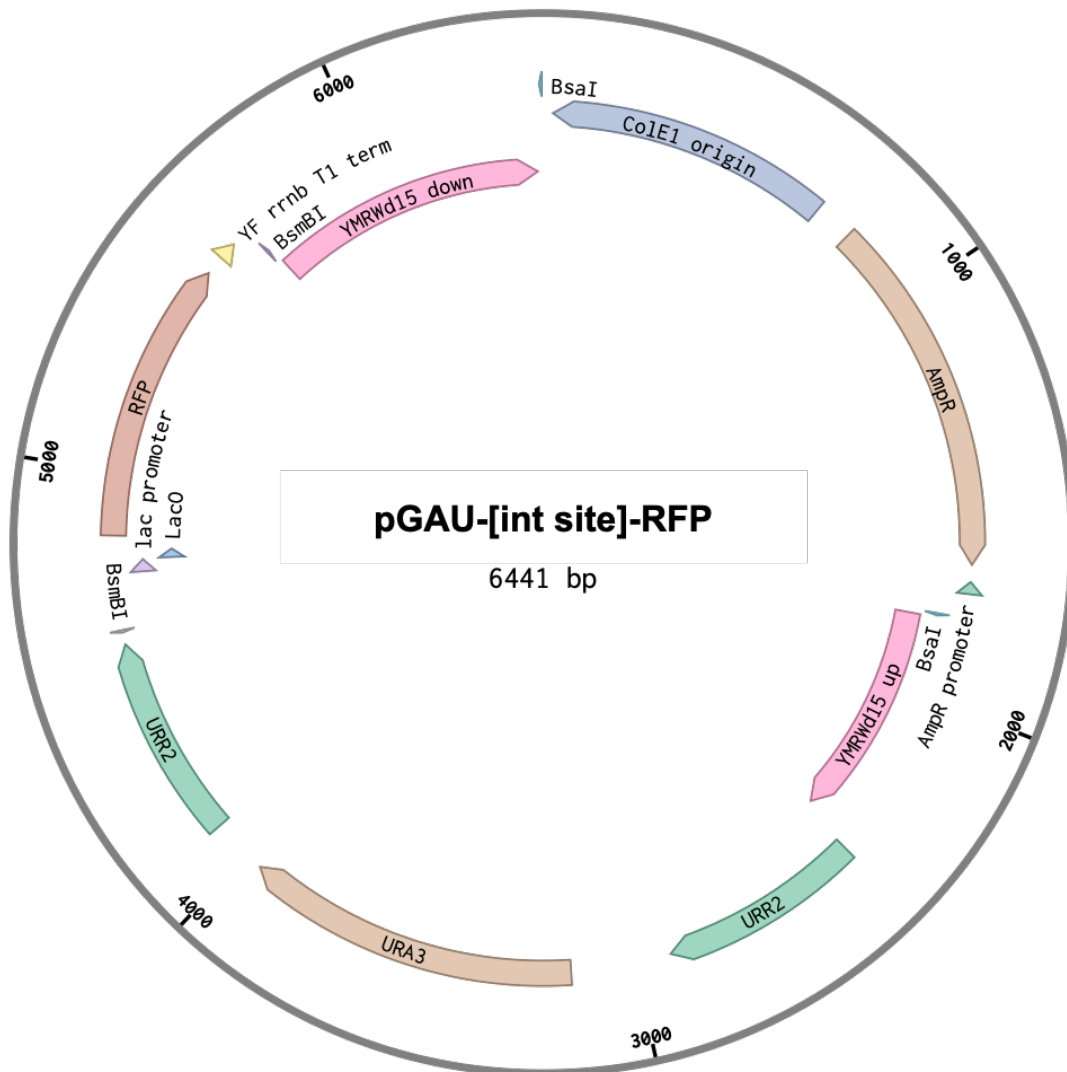


Figure A2-2: Enlarged representative plasmid map for the yeast genome integration vectors created (Chapter 2).

2.3. N-Terminal SUMO-Tag entry-level vector

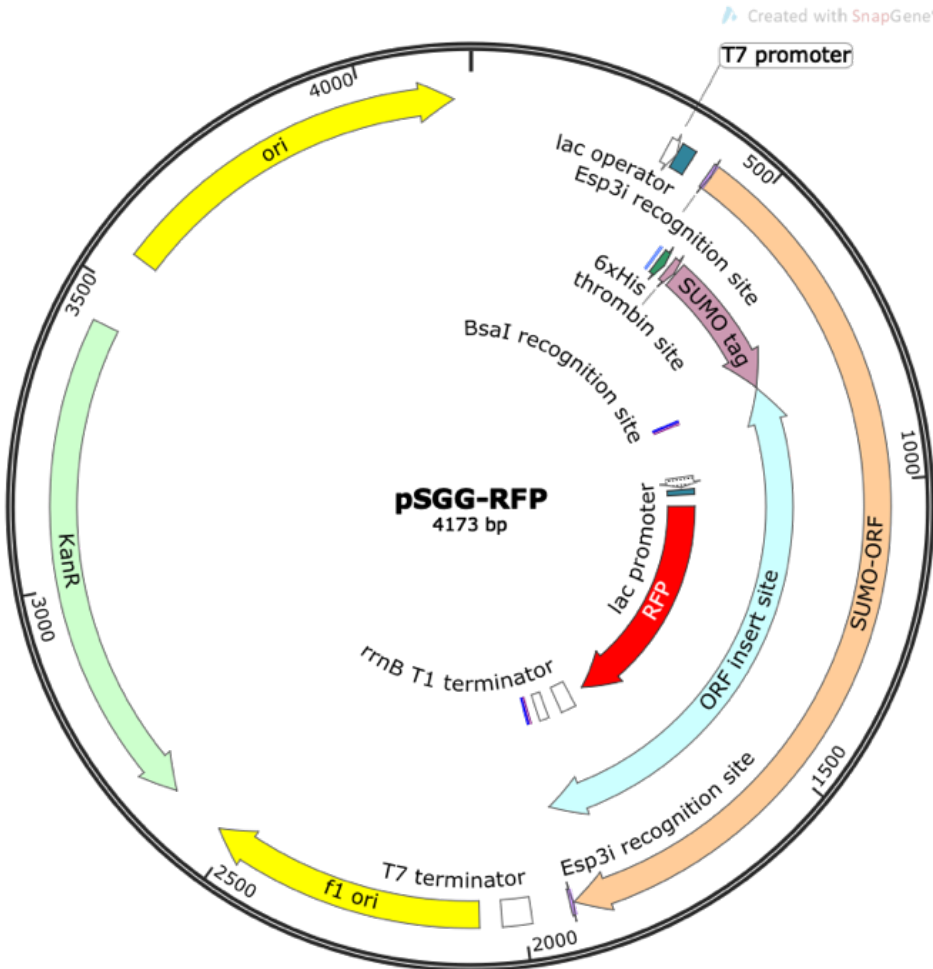


Figure A2-3: Enlarged vector map for the SUMO-ORF entry-level vector.

3. PRIMERS USED

3.1. Chapter 2

3.1.1. p [M/C] K [U/L/H]-RFP Assembly

Name	5'- Sequence - 3'
path-RFP F'	CCCTGACTAATGCCGACGTCTCGACCTCGAGACCGCAATACGCAAACCGCCTCTC
HIS3 F'	GAGAGGCGGTTTGCGTATTGCGGTCTCGAGGTCGAGACGTCGGCATTAGTCAGGG
HIS3 R'	CACGTATAGAATGATGCATTACCTTGTCTCTTCAGTATCATACTGTTTCGTATACA
Kan R'	TGTATACGAACAGTATGATACTGAAGATGACAAGGTAATGCATCATTCTATACGTG
Kan F'	GCACGTGATGAAAAGGACCCAGGTGGCATTGACTTGATCGGCACGTAAGAGGTTC
CEN R'	GAACCTCTTACGTGCCGATCAAGTCAATGCCACCTGGGTCCTTTTCATCACGTGC
Cen F'	CTACTAGAGTGGTCTCATGAGCGAGACGTCCGGCATCCGCTTACAGACAAGCTGT
path-RFP R'	ACAGCTTGTCTGTAAGCGGATGCCGGACGTCTCGCTCATGAGACCACTCTAGTAG

3.1.2. pGAU-[int site]-RFP Assembly

Name	5'- Sequence - 3'
pGAU-ORI F'	GCAAAAGGCCAGGAACCGT
pGAU-ORI R'	GTGGCACTTTTCGGGAAATG
YPRCd15up F'	ATTTCCCGAAAAGTGCCACGGTCTCTGTCAGATAACGCC
YPRCd15up R'	TGCTTCGAGCCAGGTAGAGATTGCGAAACCTATGCTCTGT
YMRWd15up F'	ATTTCCCGAAAAGTGCCACGGTCTCTGTCATCAATCAAAGCAACC
YMRWd15up R'	TGCTTCGAGCCAGGTAGAGAGCGGTGTAAGAAAATGACATAAAGTTTGA
YORWd17up F'	ATTTCCCGAAAAGTGCCACGGTCTCTGTCAGTGCACAAAGGCCATAATATTATGT
YORWd17up R'	TGCTTCGAGCCAGGTAGAGAATGGCATGAGTTATGGTTGCACAGT
ARS208up F'	ATTTCCCGAAAAGTGCCACGGTCTCTGTCAAAGGTAAGGGCTGTTACC
ARS208up R'	TGCTTCGAGCCAGGTAGAGAGTGTGTATGGTCCCTATTGGCACCG
ARS308up F'	ATTTCCCGAAAAGTGCCACGGTCTCTGTCATATTCAGAAAAATTATTC
ARS308up R'	TGCTTCGAGCCAGGTAGAGATTAGATAAAAAAGAAAAAATTCGAA
pGAU-URR2 F'	TCTCTACCTGGCTCGAAGCA
HcK_O-URA3 F'	CTGGTCTCGGATGCATACCACAGCTTTTCAATTCAATTCATCA
HcK_O-URA3 R'	CTGGTCTCGGCTACCTGATGCGGTATTTCTCCTTACG
pGAU-URR2 R'	TCTCTGCCTGAGGTGGTGC
pGAU-RFP F'	TGCAACACCTCAGGCAGAGAACCTAGAGACGGCAATACGCAAACCGCCTCT
pGAU-RFP R'	CTCATGAGACGGAGCCAGTGTGACTCTAGTAGAGAG
YPRCd15down F'	CACTGGCTCCGTCTCATGAGCGCTTGGAAGGTCGGGATG
YPRCd15down R'	TACGGTTCCTGGCCTTTTGCGGTCTCTCCGACATAAAGCAG
YMRWd15down F'	CACTGGCTCCGTCTCATGAGCGCTCATGGAAAATGCAACCG
YMRWd15down R'	TACGGTTCCTGGCCTTTTGTCATCGGTCTCTCCGAGCCG
YORWd17down F'	CACTGGCTCCGTCTCATGAGCGTCCCCAACAAAAGTGGGCTCTCA
YORWd17down R'	TACGGTTCCTGGCCTTTTGTCATCGGTCTCTCCGATCTTAAAGCTGGCTCCC
ARS208down F'	CACTGGCTCCGTCTCATGAGACTTGTAGTTTATGTGCTTTATAGT

ARS208down R'	TACGGTTCCTGGCCTTTTGCATCGGTCTCTCCGAATTAATTAATAAAATTTAAATAG
ARS308down F'	CACTGGCTCCGTCTCATGAGTCTTTGCTACATATTGCTACCACTT
ARS308down R'	TACGGTTCCTGGCCTTTTGCATCGGTCTCTCCGATGATAGAACGAGTACAAC

3.1.3. Cloning primers

3.1.3.1. ORFs

Name	5'- Sequence - 3'
dmaW_forGG	ATTAGGTCTCGGATGCACCATCACCATCACCACAT
EasF_forGG	ATTAGGTCTCGGATGACCATCTCTGCACCGCCGAT
dmaW_RevGGTag	ATTAGGTCTCGCAGCTCAATGTAAACCAGAG
EasF_RevGGTag	ATTAGGTCTCGCAGCGTTCAGACGCAGGCGCAG

3.1.3.2. Promoters and Terminators

Primers used to clone the promoters and terminators were designed by appending the following prefix and suffix sequences to 25 bps of the start and end of these sequences to be cloned.

	prefix	suffix
Promoters	GGTCTCGGGCT	GGTCTCGCATC
Terminators	GGTCTCGTAGC	GGTCTCGCCTC

3.2. Chapter 3

3.2.1. Assembly of pSGG-RFP

Name	5'- Sequence - 3'
pSGG_Ori Res F'	CTAGTAGCGGCCGCTGCAGGGTCTCTTAGCTGAGACGAGATCCGGCTGCTAACAAAGC
pSGG_Ori Res R'	CCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCATA
pSGG_CDS start F'	TTGTTTAACCTTAAGAAGGACGTCTCGGATGGGCAGCAGCCATCATCATCATCAC
pSGG_CDS end R'	GCTTTGTAGCAGCCGATCTCCGTCTCAGCTAAGAGACCCTGCAGCGGCCGCTACTAG
pSGG_T7 F'	CGCCAGCAACGCGGCCCTTTTACGGCGGGATCTCGACGCTCTCCC
pSGG_T7 R'	TCCTTCTTAAAGTTAAACAAAATTATTCTAGAGGGGAAT

3.3. Chapter 4

3.3.1. Cloning primers

Name	5'- Sequence - 3'
EasD_forGG	ATTAGGTCTCGGATGGCATCAGTGGAATCGCGTAT
EasD_RevGG	ATTAGGTCTCGGCTACGGCATACAGGCACCCA
EasG_forGG	ATTAGGTCTCGGATGACAGTATTATTAACAGGTGG
EasG_RevGG	ATTAGGTCTCGGCTACTTTCTTGCTCTCCATGCTGCC
easAPi_GGfwd	ATTAGGTCTCGGATGTCTACTTCCAACCTGTTCAACGAATTG
easAPi_GGrev	ATTAGGTCTCGCAGCAGCAACAGTAGTTTCTTGGTCACC
easANI_GGfwd	ATTAGGTCTCGGATGTCTACTTCTAACTTGTTCAACCCATTGCAA
easANI_GGrev	ATTAGGTCTCGCAGCAGCCAAAACAGCTTGCTTCTTGTTCAGT
easAEI_GGfwd	ATTAGGTCTCGGATGTCTACCTCTAACTTGTTCAACGCTT
easAEI_GGrev	ATTAGGTCTCGCAGCGTTCTTTCTCATAGAAGCAGTAGTAT
easACP_GGfwd	ATTAGGTCTCGGATGTCAACTTCTAATTGTTTAATAC
easACP_Ggrev	ATTAGGTCTCGGCTAACCTGCAACAGCAGCTTCATCAGC
PD11_forGG	ATTAGGTCTCGGATGAAGTTTCTGCTGGTGCCGTCC
FAD1_forGG	ATTAGGTCTCGGATGCAGTTGAGCAAGGCTGCTGAGATGTG
PD11_RevGG	ATTAGGTCTCGGCTACAATTCATCGTGAATGGCATCTTCTTCG
FAD1_RevGG	ATTAGGTCTCGGCTAATTCTTGATCCTGCCTGCTCTCTAAAG
FAD1_bsmBI F'	CGCATTAGAGATTTTATAAAGATATACCCTGAAACTGAAGCTATAGTGATAGGTATTAGACACACAG
FAD1_bsmBI R'	CTGTGTGTCTAATACCTATCACTATAGCTTCAGTTTCAGGGTATATCTTTATAAAATCTCTAAATGCG

4. Refinement of SSN parameters

4.1. SSN for *easE* orthologs

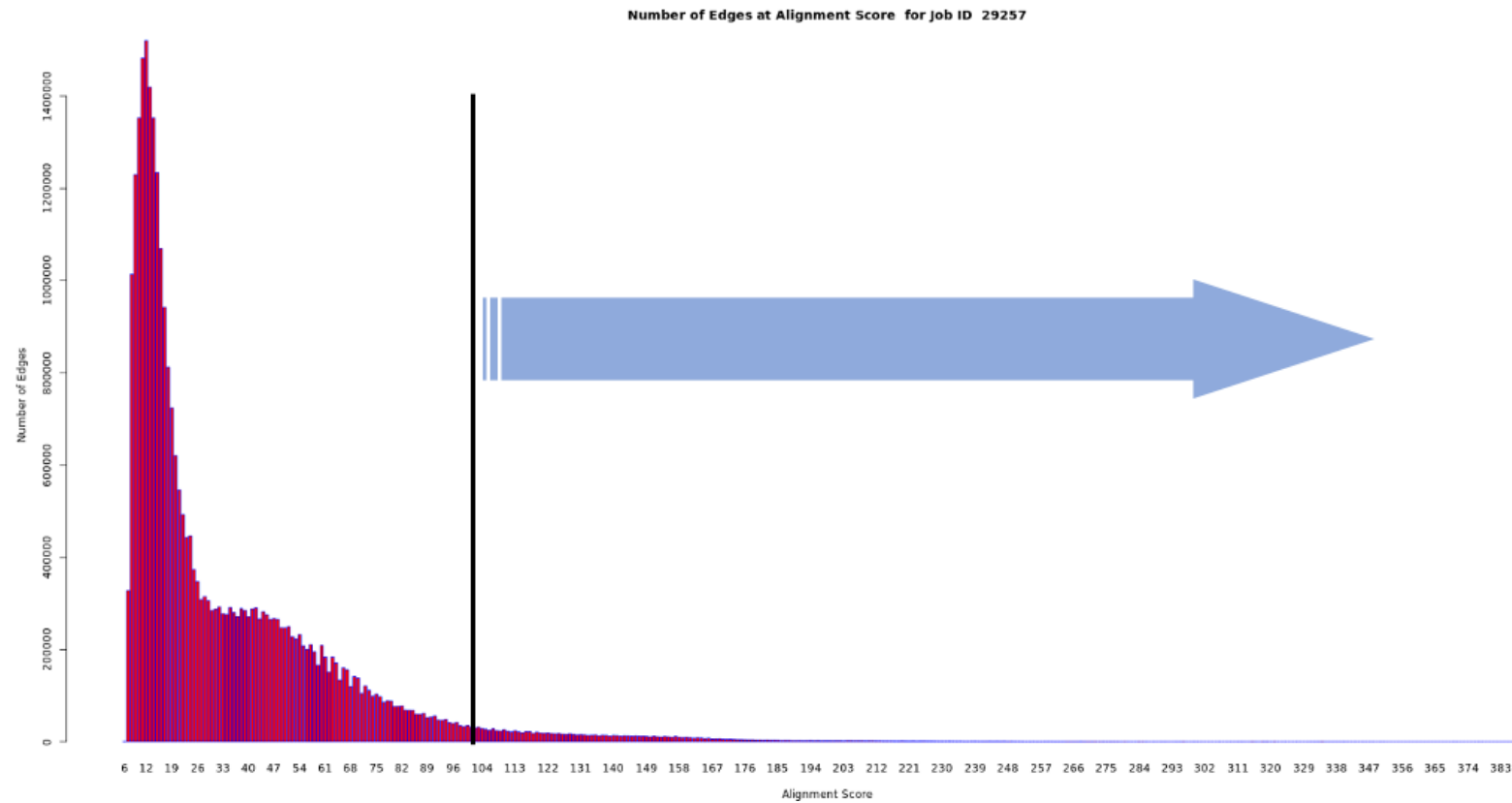


Figure A4-1: Graphical representation of the distribution of edges for each alignment score. Line in black represents the cut off used to generate the SSN used in this study.

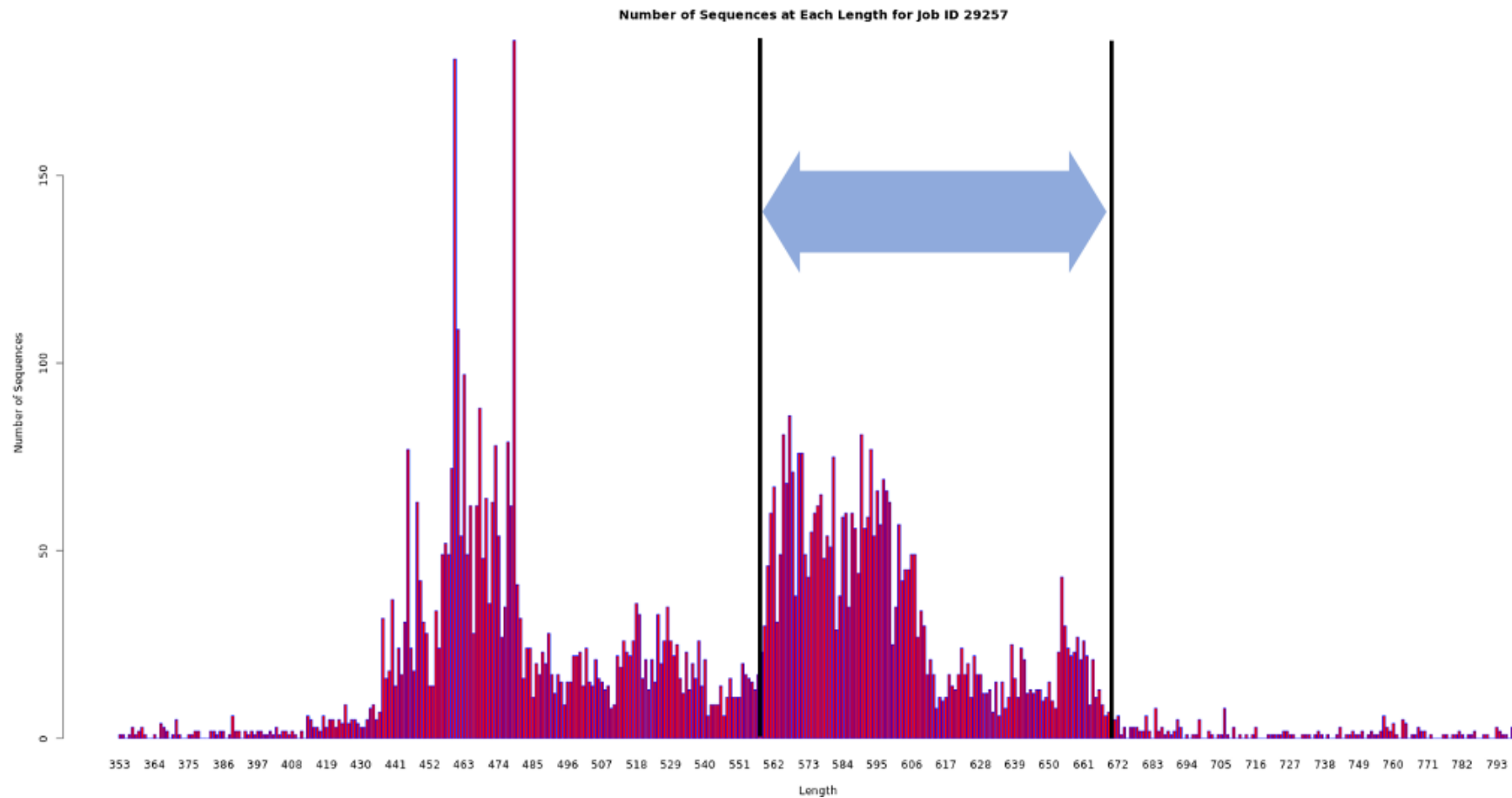


Figure A4-2: Graphical representation of the distribution of sequence lengths for the sequences retrieved by the initial BLAST search. Black lines indicate the selected sequence length window used for the generation of the SSN used in this study.

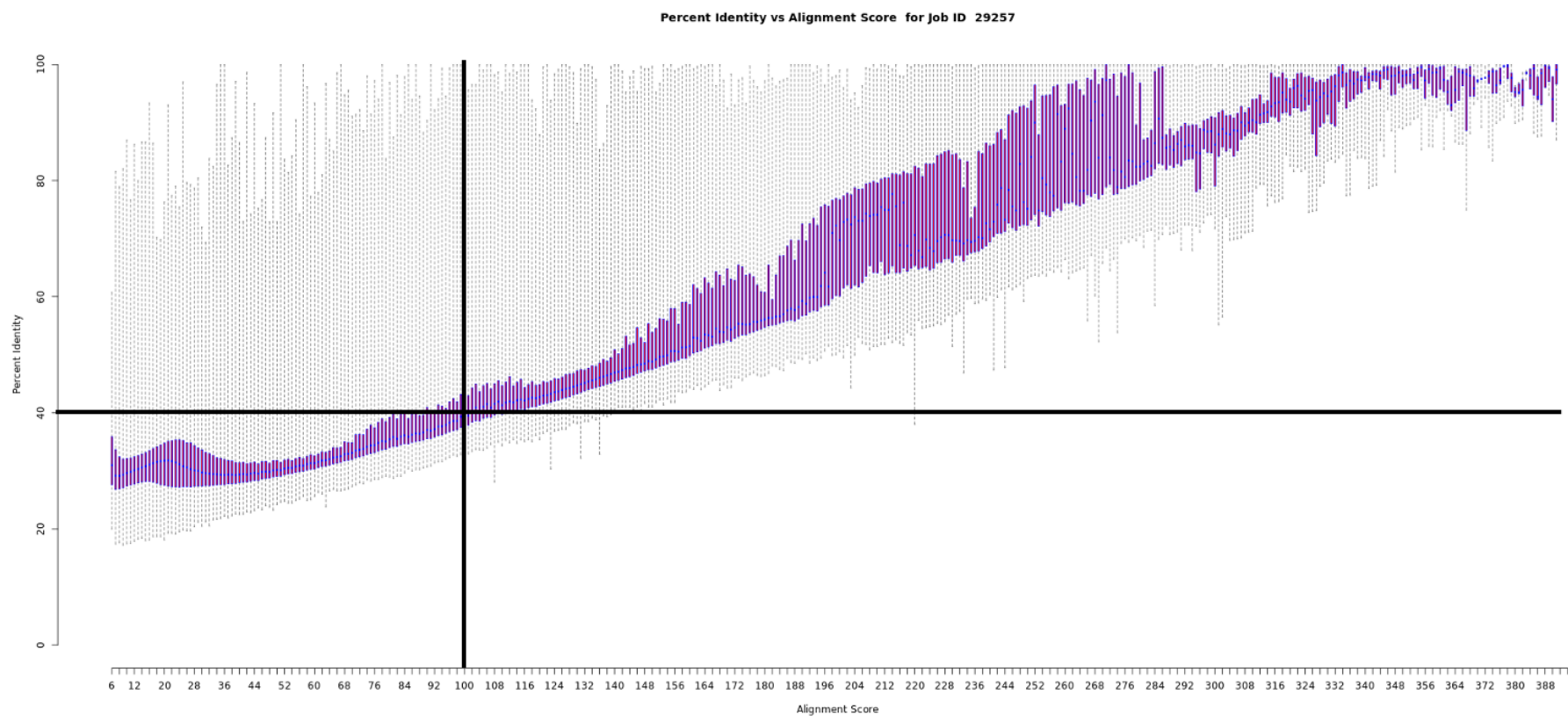


Figure A4-3: Correlation between sequence identity against their corresponding alignment scores. Selection of an alignment score of 100 approximates the sequences with 40% sequence identity.

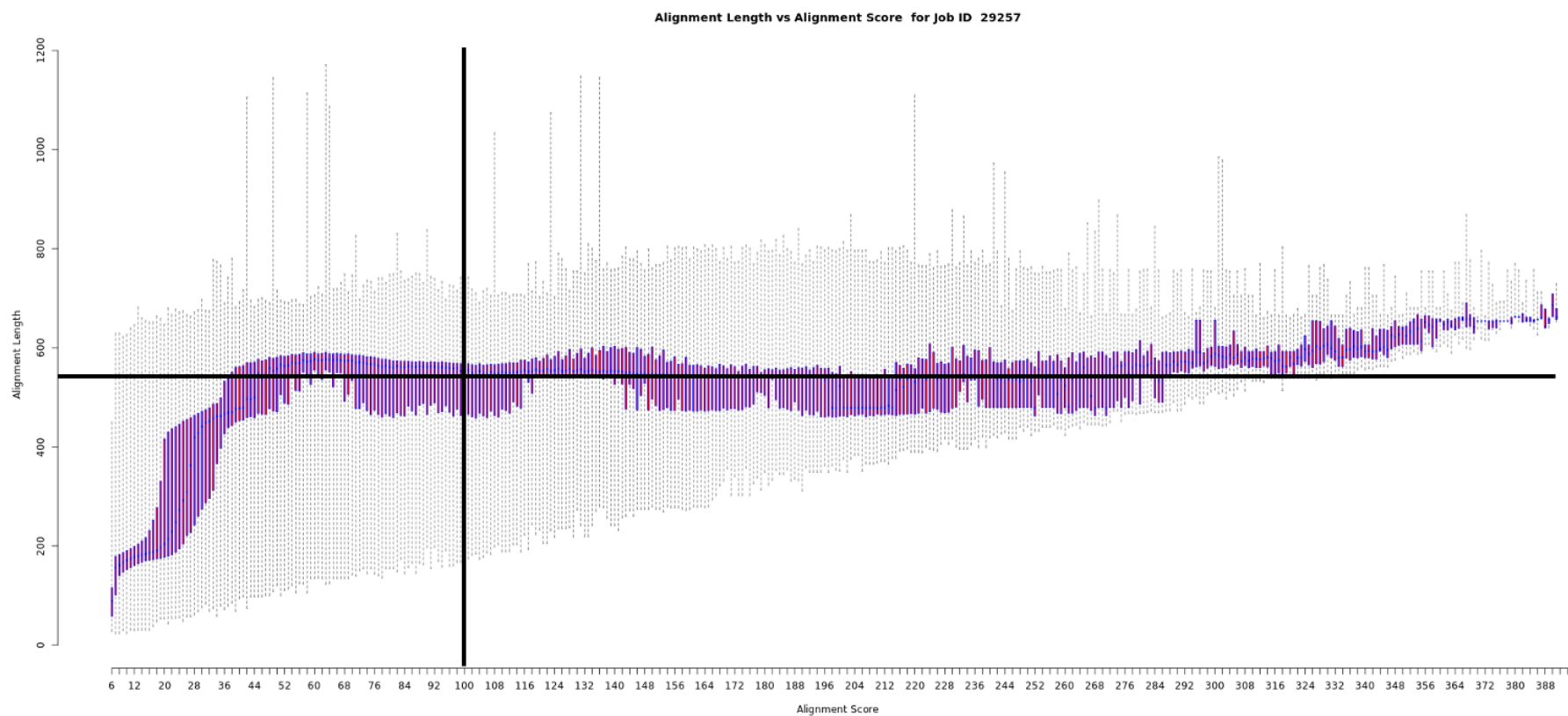


Figure A4-4: Correlation between alignment score and alignment length between the sequences retrieved. Selection of an alignment score of 100 ensures the alignment length between retrieved sequences do not fall below 500 aa and ensures the alignment covers the full query sequence.

4.2. SSN for *easA* orthologs

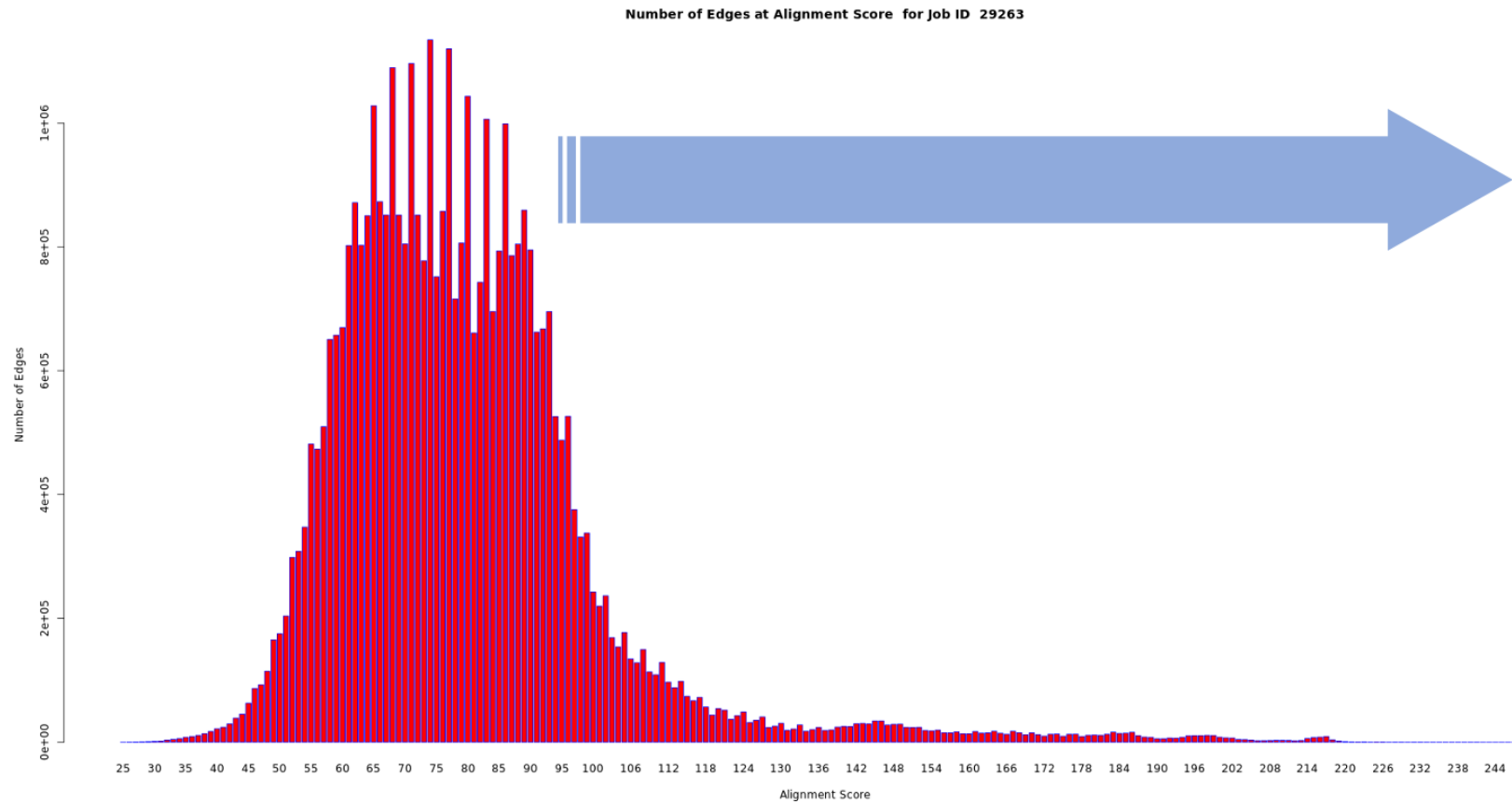


Figure A4-5: Graphical representation of the distribution of edges for each alignment score. Line in black represents the cut off used to generate the SSN used in this study

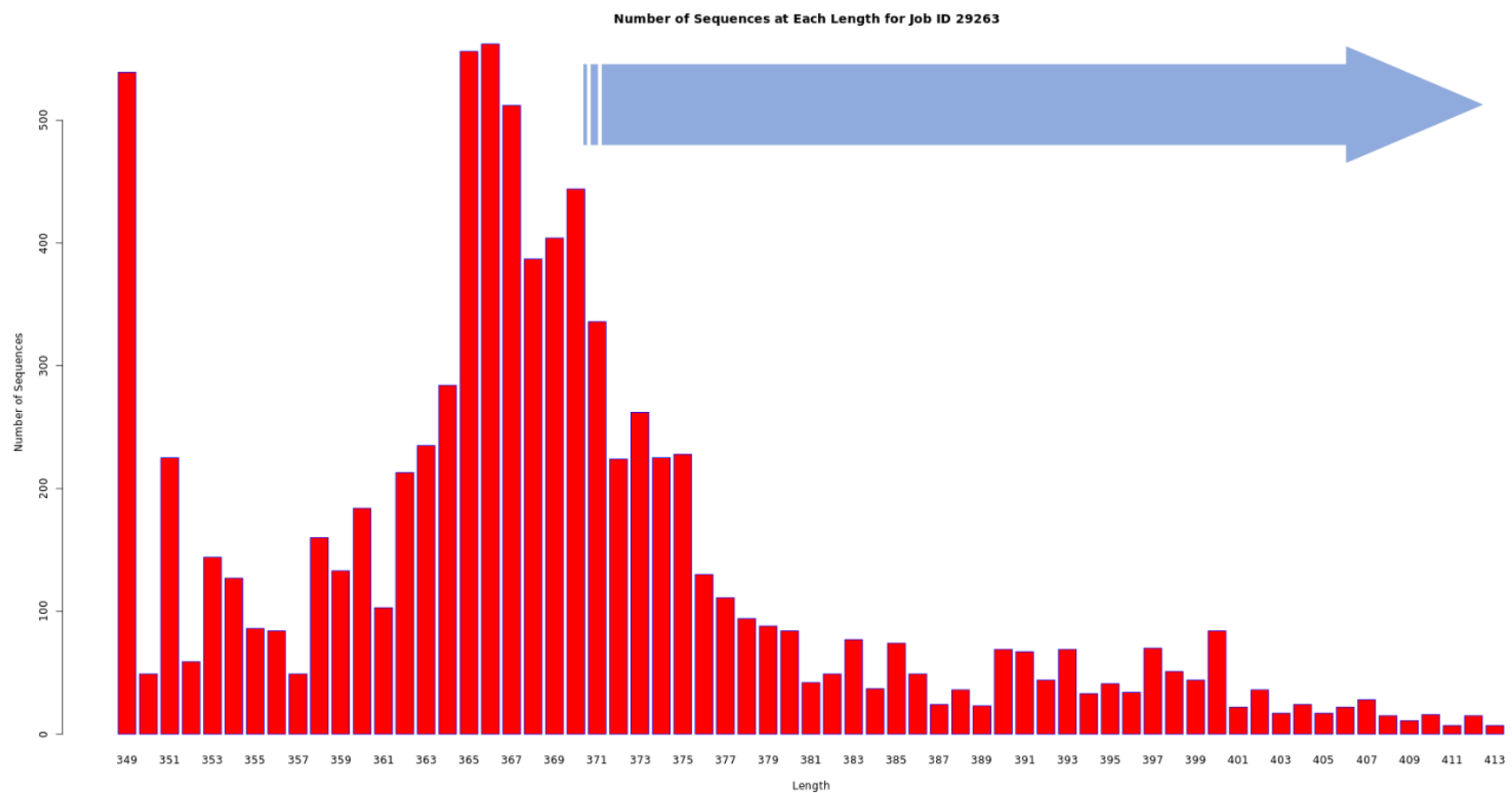


Figure A4-6: Graphical representation of the distribution of sequence lengths for the sequences retrieved by the initial BLAST search. Black lines indicate the selected sequence length window used for the generation of the SSN used in this study.

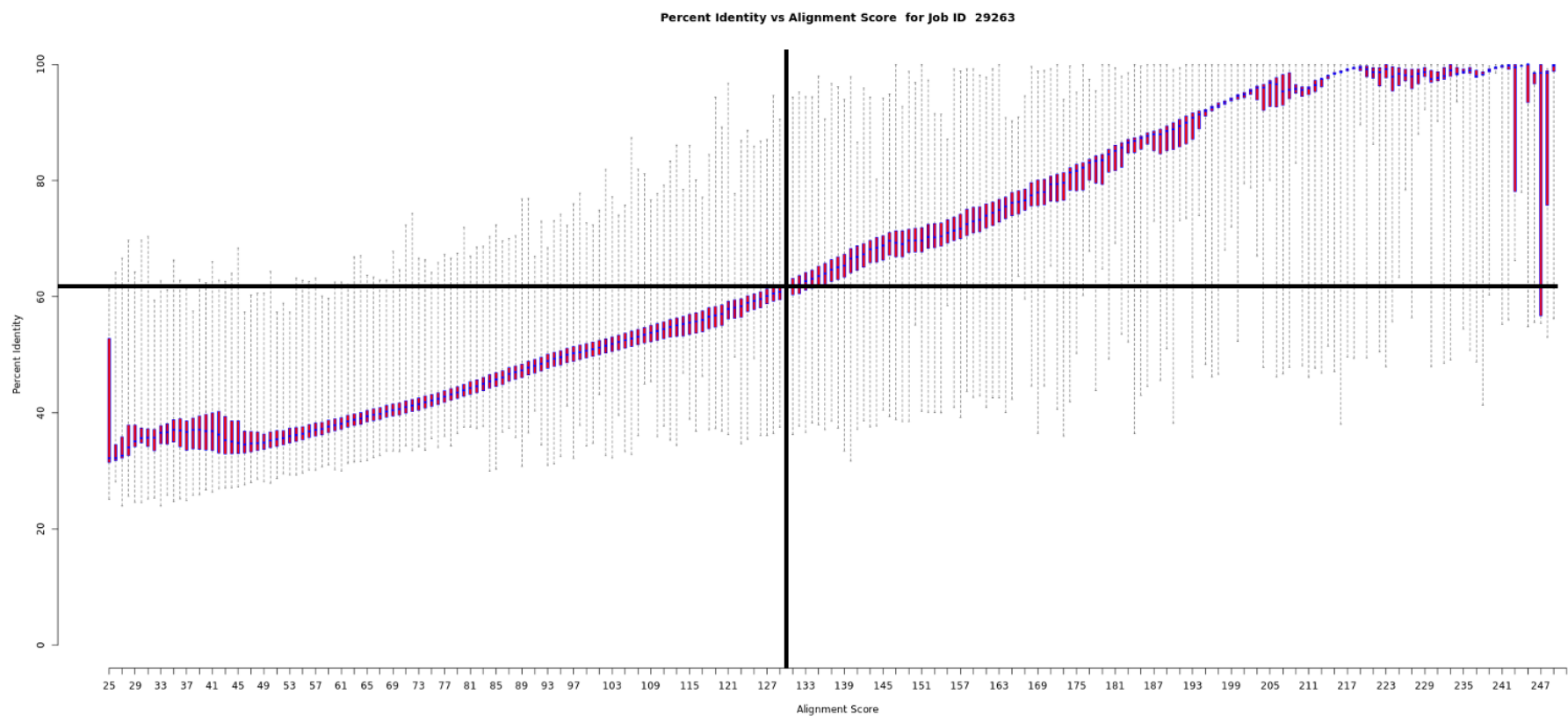


Figure A4-7: Correlation between sequence identity against their corresponding alignment scores. Selection of an alignment score of 130 approximates the sequences with 60% sequence identity.

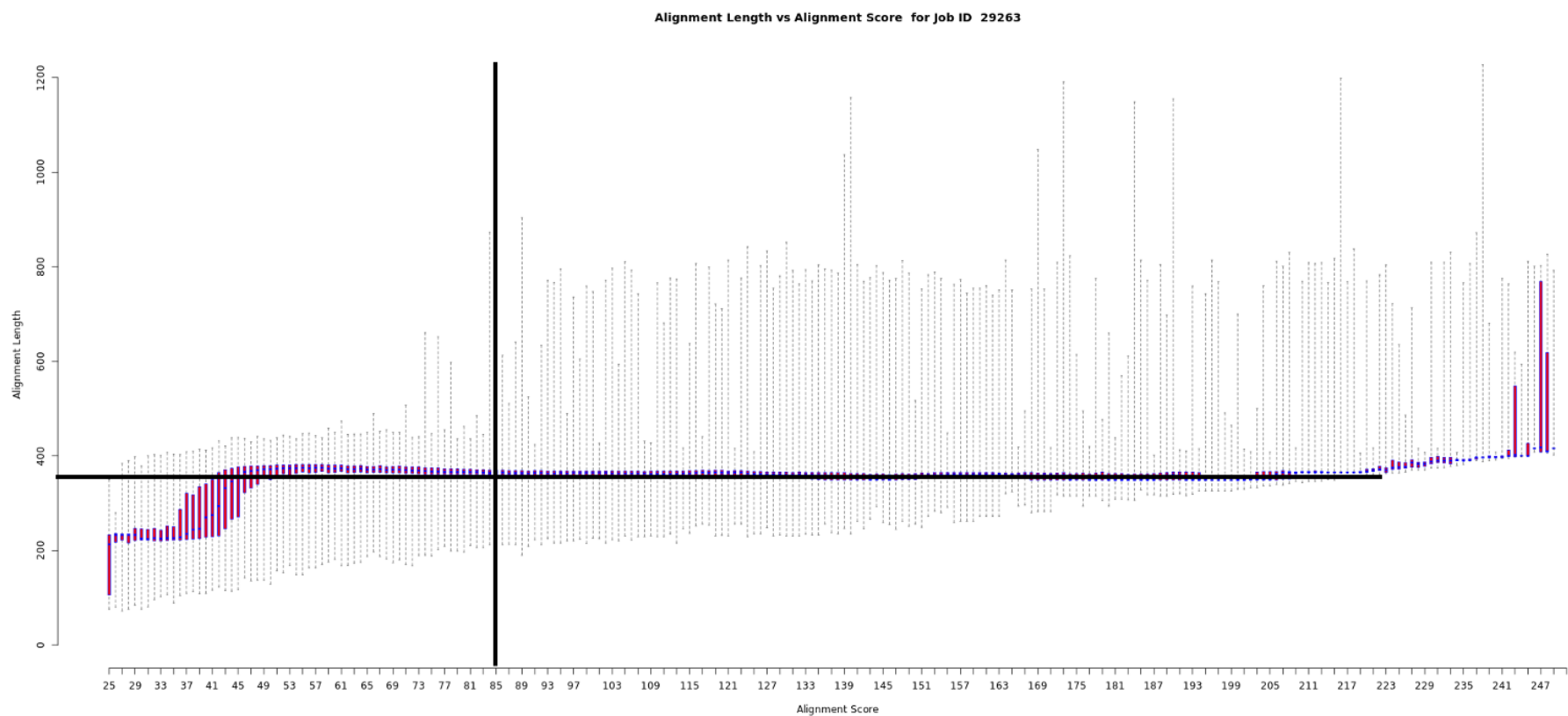


Figure A4-8: Correlation between alignment score and alignment length between the sequences retrieved. Selection of an alignment score of 130 ensures the alignment length between retrieved sequences do not fall below 300 aa and ensures the alignment covers the full query sequence.

5. EXPRESSION AND PURIFICATION OF ERGOT ALKALOID PATHWAY ENZYMES

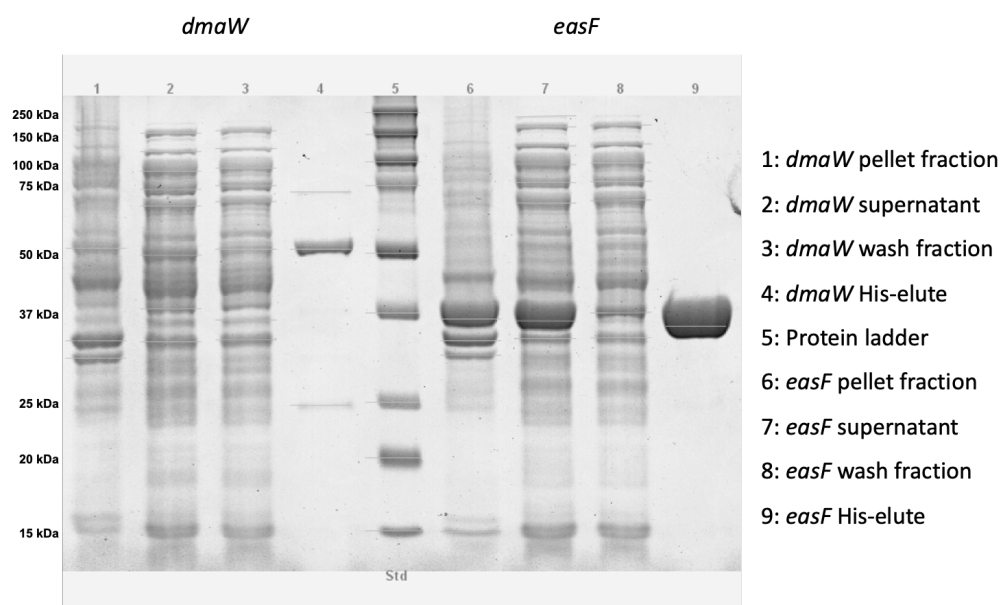


Figure A5-1: SDS-PAGE of purified *dmaW* (53 kDa) and *easF* (38 kDa) proteins from *A. fumigatus* recombinantly expressed in *E. coli*.

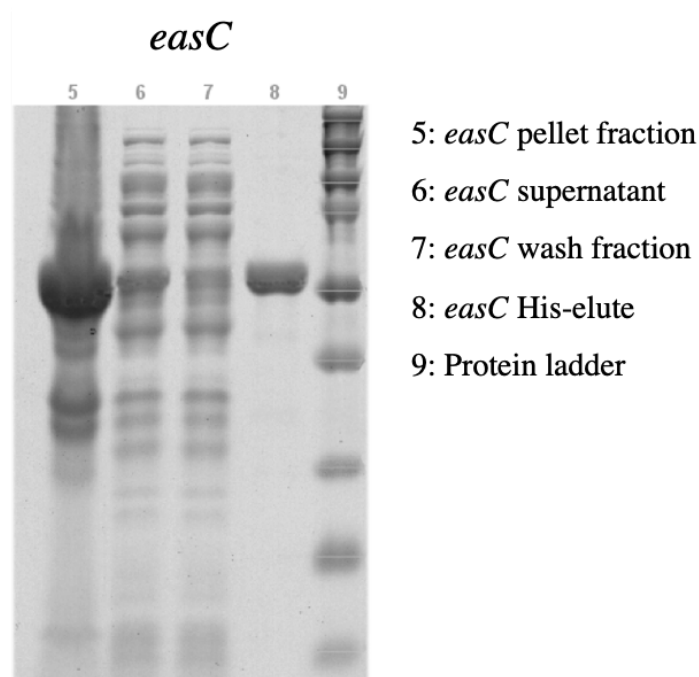


Figure A5-2: SDS-PAGE of *easC* (53 kDa) from *C. purpurea* expressed and purified from *E. coli*.

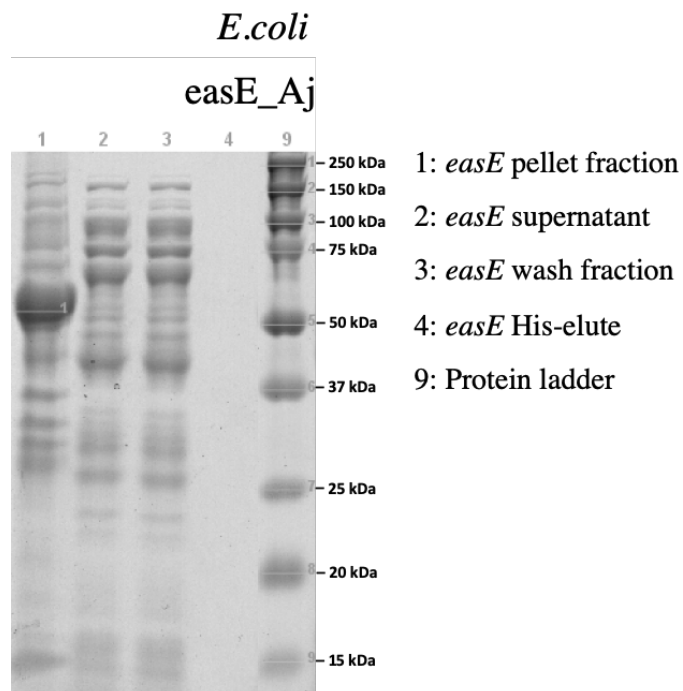


Figure A5-3: SDS-PAGE of *easE* (67 kDa) from *A. japonicus* expressed in *E. coli*.

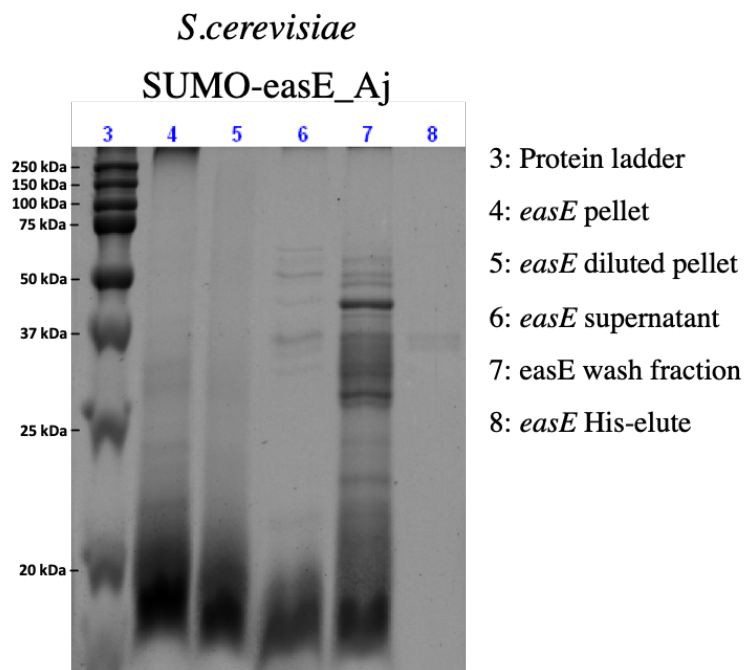


Figure A5-4: SDS-PAGE of SUMO-*easE* (79 kDa) from *A. japonicus* expressed in *S. cerevisiae*.

6. Assembly and testing of AgcM33B/ AgcM32B

6.1. Scheme for assembly

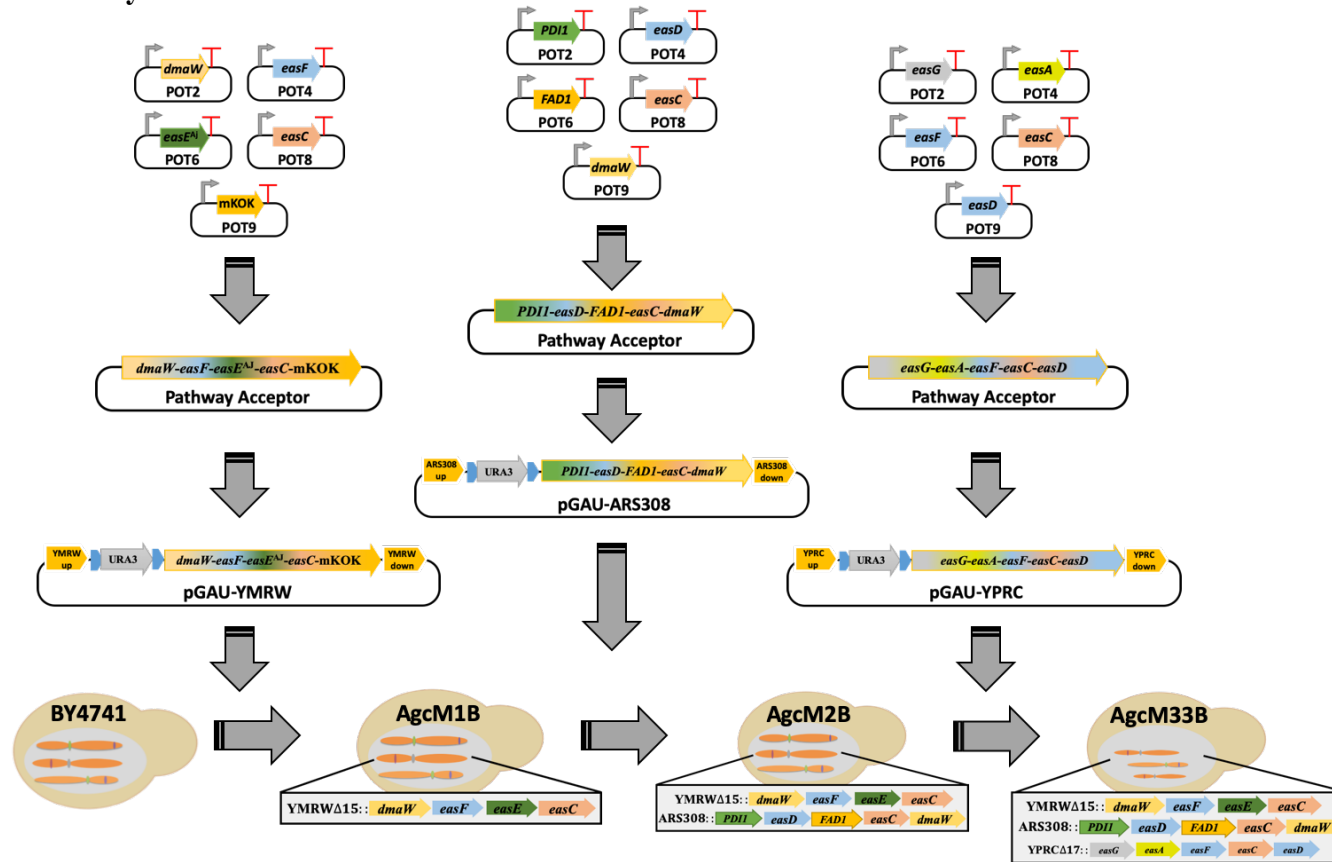
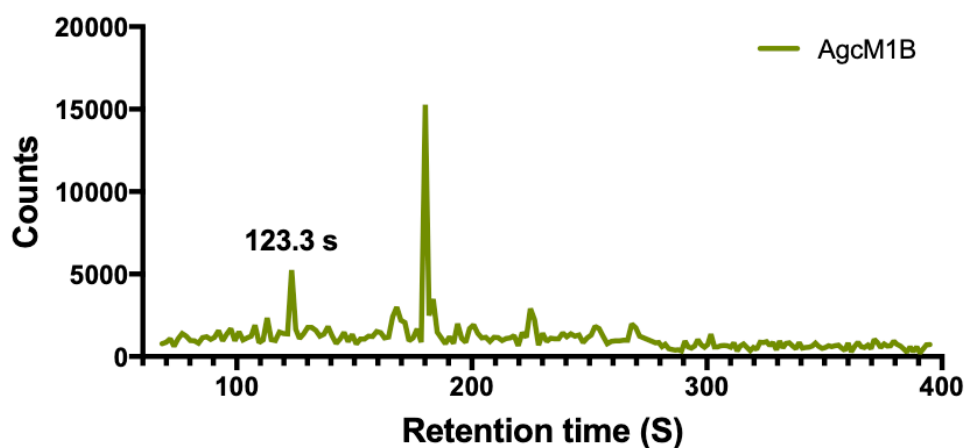


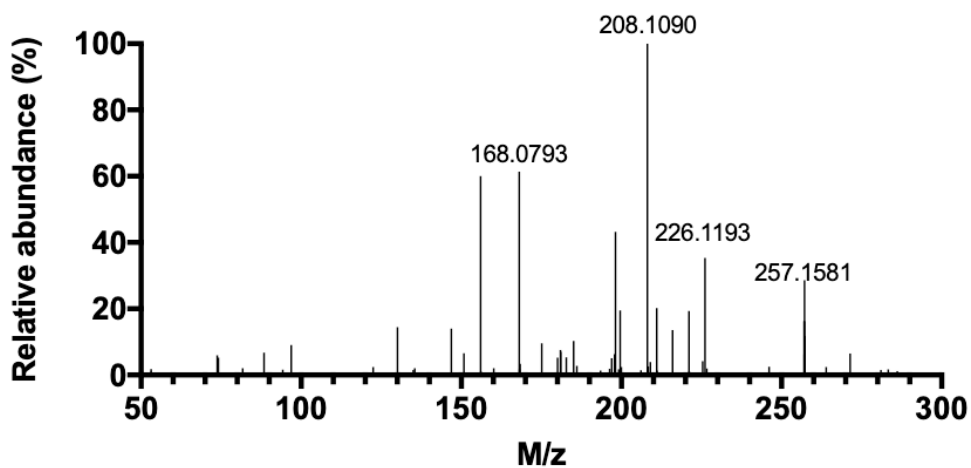
Figure A6-1: Schematic representation of the assembly process for AgcM33B/ AgcM32B.

6.2. Assembly checks for AgcM33B and AgcM32B

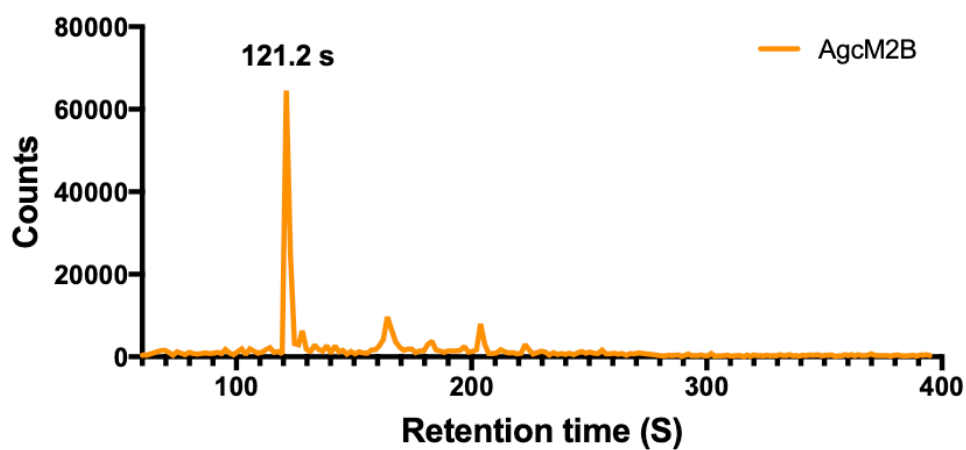
6.2.1. AgcM1B – Chromatograms (TIC: 257 m/z)



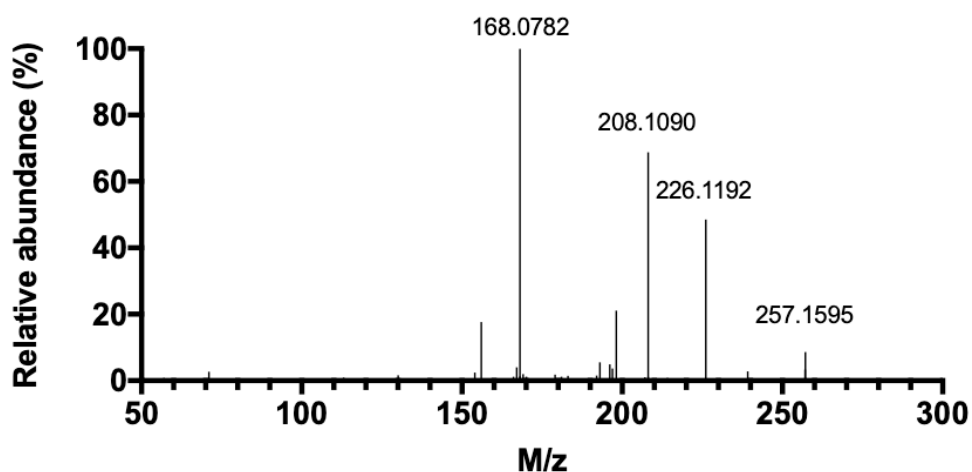
6.2.2. AgcM1B – Mass spectra



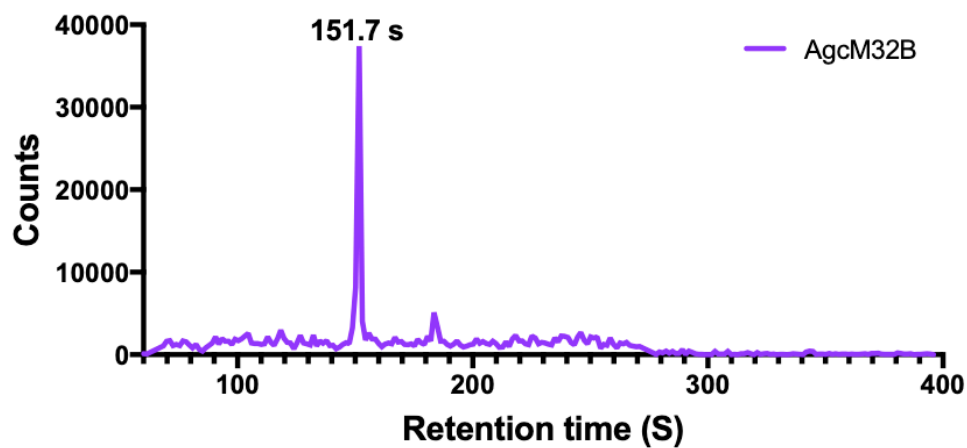
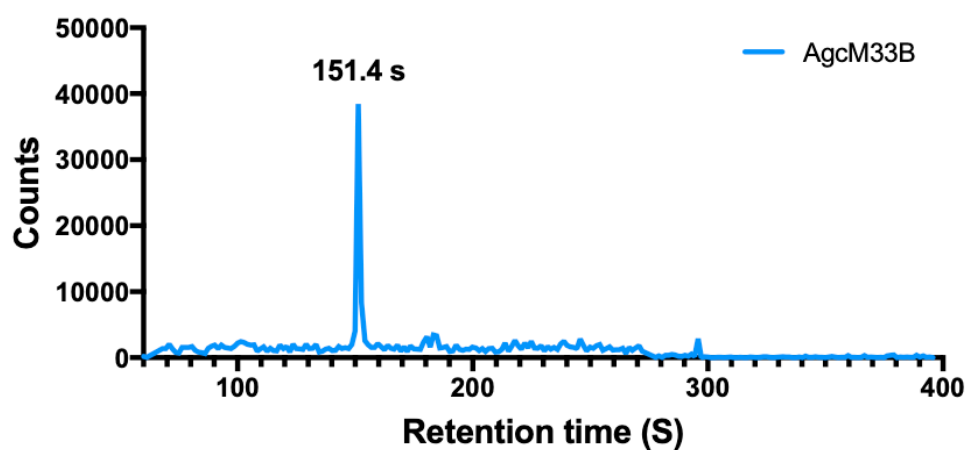
6.2.3. AgcM2B – Chromatograms (TIC: 257 m/z)



6.2.4. AgcM2B – Mass spectra

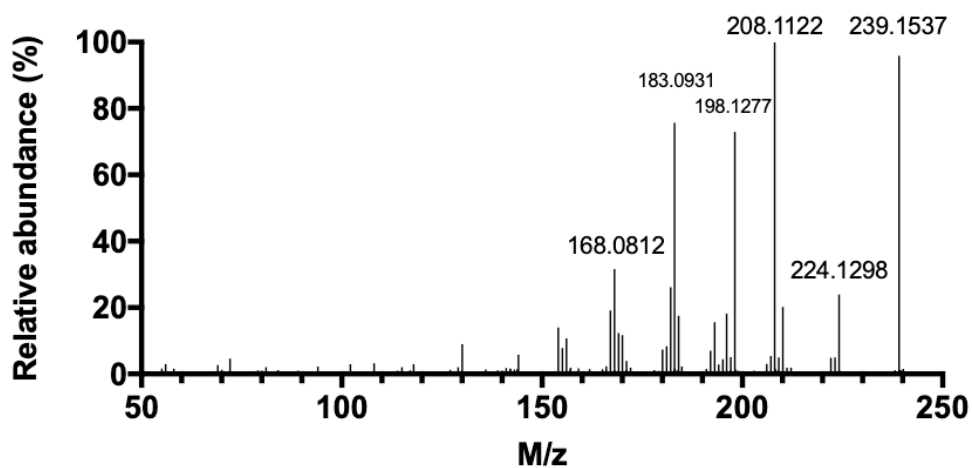


6.2.5. AgcM3XB – Chromatograms (TIC: 239 m/z)

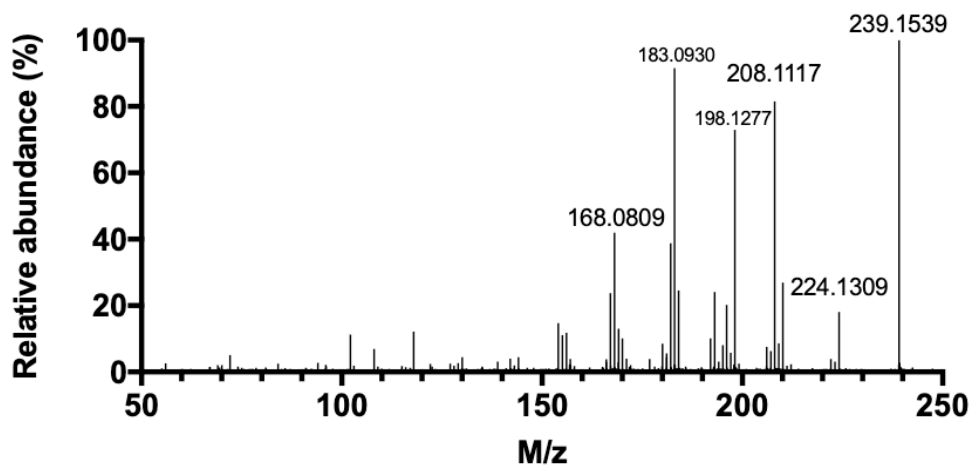


6.2.6. AgcM3XB – Mass spectra

6.2.6.1. AgcM33B



6.2.6.2. AgcM32B



7. Standard curves used for the quantification of DLA and Agroclavine
 7.1. Quantification of DLA and Agroclavine from the episomal expression of *cloA* in AgcM33/32B.

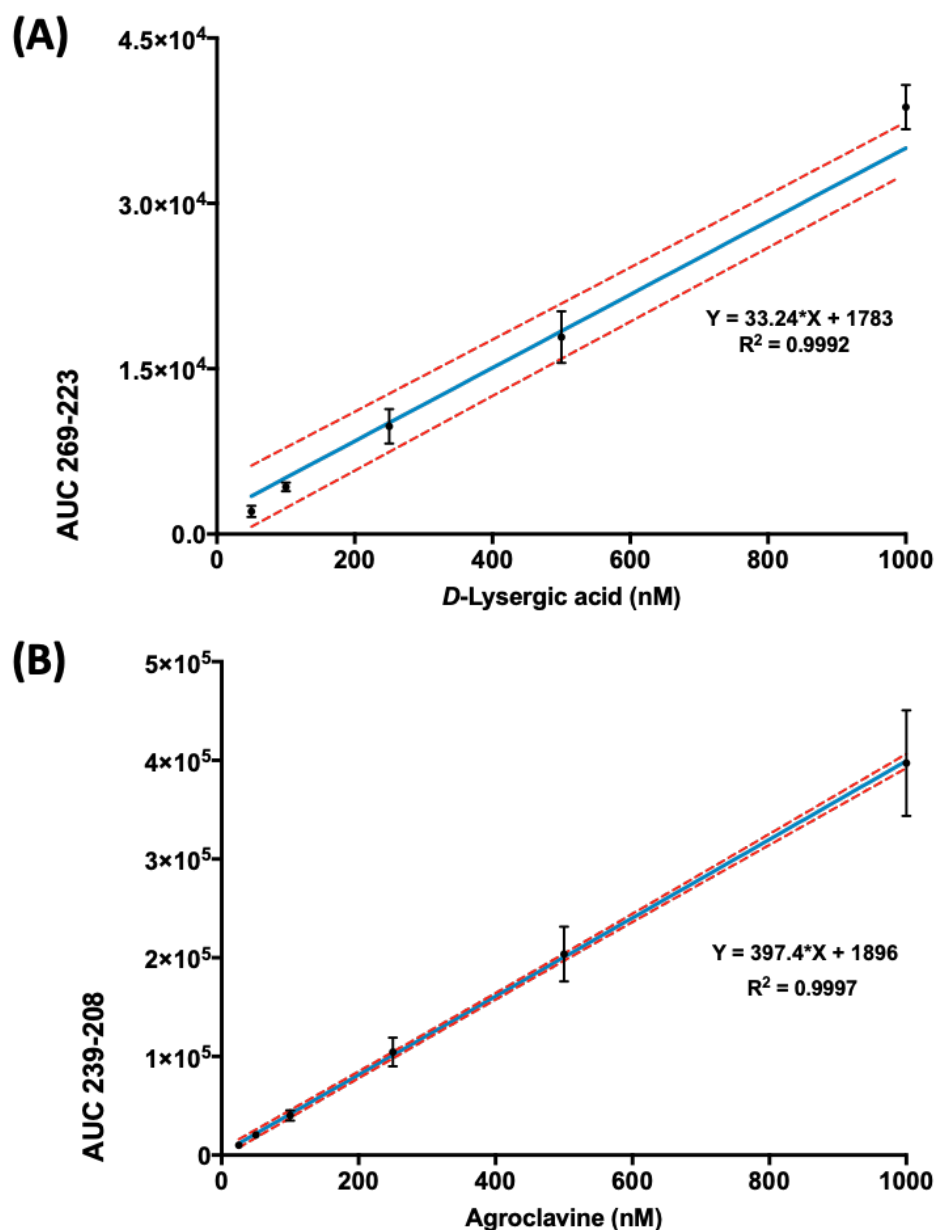


Figure A7-1: (A) DLA standard curve established in AgcM33B media. Peak area response for the ion transition of 269 to 223 m/z plotted against spiked DLA concentrations, over a range of 10 – 1000 nM. (C) Agroclavine standard curve established in YOCE media. Peak area response for the ion transition of 239 to 208 m/z plotted against spiked agroclavine concentrations, over a range of 10 – 1000 nM. Error bars were calculated from three independent replicates.

7.2. Standard curves used for the quantification of DLA and Agroclavine from the tryptophan supplementation experiments.

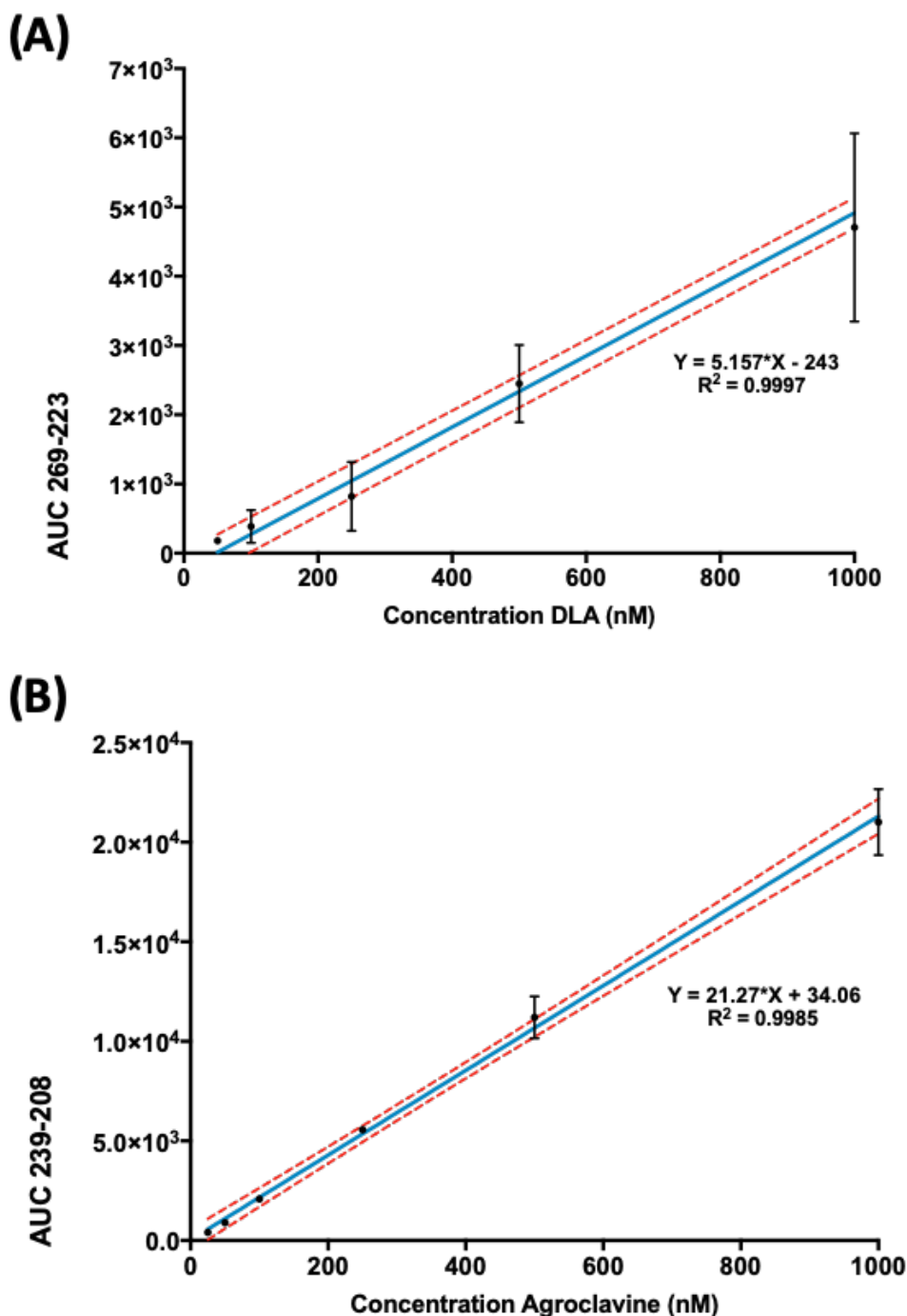


Figure A7-2: (A) DLA standard curve established in AgcM33B media. Peak area response for the ion transition of 269 to 223 m/z plotted against spiked DLA concentrations, over a range of 10 – 1000 nM. (B) Agroclavine standard curve established in YOCE media. Peak area response for the ion transition of 239 to 208 m/z plotted against spiked agroclavine concentrations, over a range of 10 – 1000 nM. Error bars were calculated from three independent replicates.

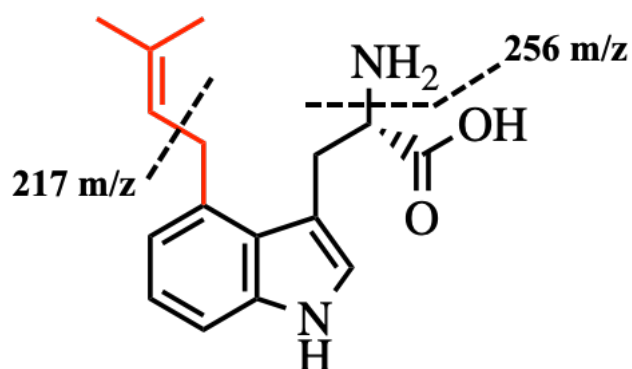
8. List of strains created.

Strain	Genotype		
	Chr.	Modifications	
YMWF	XIII	YMRWΔ15	P _{TEF2} -dmaW-T _{ENT2} P _{GPM1} -easF-T _{PRX1} P _{GAL1} -yEGFP-T _{CBR1}
YMC17	XIII	YMRWΔ15	P _{TEF2} -dmaW-T _{ENT2} P _{GPM1} -easF-T _{PRX1} P _{GAL10} -easC-T _{YIP5} P _{CUP1} -mKOk-T _{CBR1}
YOCE	XIII	YMRWΔ15	P _{TEF2} -dmaW-T _{ENT2} P _{GPM1} -easF-T _{PRX1} P _{GAL1} -yEGFP-T _{CBR1}
	XV	YORWΔ22	P _{GAL10} -easC-T _{YIP5} P _{GAL1} -easE_Aj-T _{RPS20}
AgcM1B	XIII	YMRWΔ15	P _{TDH2} -dmaW-T _{ENT2} P _{GPM1} -easF-T _{PRX1} P _{TEF2} -easE_Aj-T _{RPS20} P _{PGK1} -easC-T _{YIP5} P _{CUP1} -mKOk-T _{CBR1}
AgcM2B	XIII	YMRWΔ15	P _{TDH2} -dmaW-T _{ENT2} P _{GPM1} -easF-T _{PRX1} P _{TEF2} -easE_Aj-T _{RPS20} P _{PGK1} -easC-T _{YIP5} P _{CUP1} -mKOk-T _{CBR1}
	III	ARS308	P _{RPL3} -PDI1-T _{VPS1} P _{PDA1} -easD-T _{YPR1} P _{ACT1} -FAD1-T _{EFM2} P _{RPL15} -easC-T _{YIP5} P _{TPI1} -dmaW-T _{ENT2}

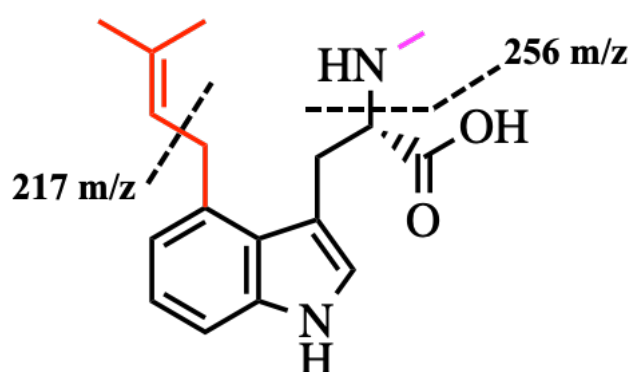
AgcM33B	XIII	YMRWΔ15	P_{TDH2}-dmaW-T_{ENT2} P_{GPM1}-easF-T_{PRX1} P_{TEF2}-easE_Aj-T_{RPS20} P_{PGK1}-easC-T_{YIP5} P_{CUP1}-mKOk-T_{CBR1}
	III	ARS308	P_{RPL3}-PDI1-T_{VPS1} P_{PDA1}-easD-T_{YPR1} P_{ACT1}-FAD1-T_{EFM2} P_{RPL15}-easC-T_{YIP5} P_{TPI1}-dmaW-T_{ENT2}
	XVI	YPRCΔ15	P_{RPL8B}-easG-T_{APL2} P_{SSA1}-easA_Ec-T_{YPT31} P_{TEF2}-easF-T_{PRX1} P_{TDH2}-easC-T_{YIP5} P_{TDH3}-easD-T_{YPR1}
DLAM33B	XIII	YMRWΔ15	P_{TDH2}-dmaW-T_{ENT2} P_{GPM1}-easF-T_{PRX1} P_{TEF2}-easE_Aj-T_{RPS20} P_{PGK1}-easC-T_{YIP5} P_{CUP1}-mKOk-T_{CBR1}
	III	ARS308	P_{RPL3}-PDI1-T_{VPS1} P_{PDA1}-easD-T_{YPR1} P_{ACT1}-FAD1-T_{EFM2} P_{RPL15}-easC-T_{YIP5} P_{TPI1}-dmaW-T_{ENT2}
	XVI	YPRCΔ15	P_{RPL8B}-easG-T_{APL2} P_{SSA1}-easA_Ec-T_{YPT31} P_{TEF2}-easF-T_{PRX1} P_{TDH2}-easC-T_{YIP5} P_{TDH3}-easD-T_{YPR1}
	II	ARS208	P_{GAL1}-cloA_Cpur-T_{RPL41B}

9. Proposed MSMS fragmentation ions

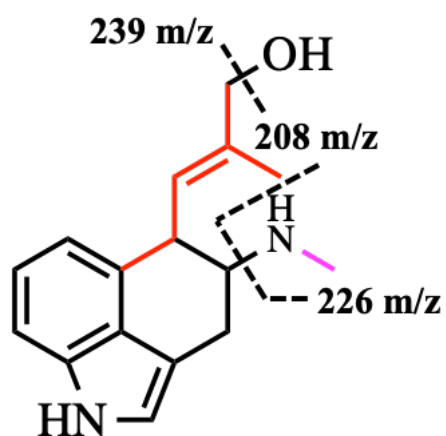
9.1. DMAT



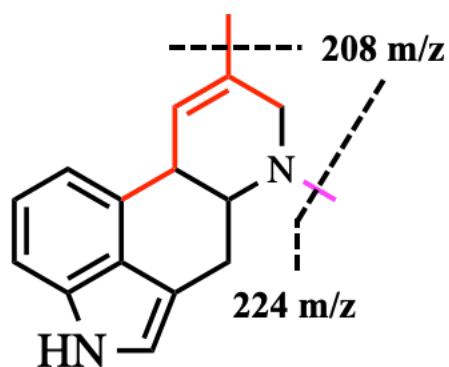
9.2. 4DMA



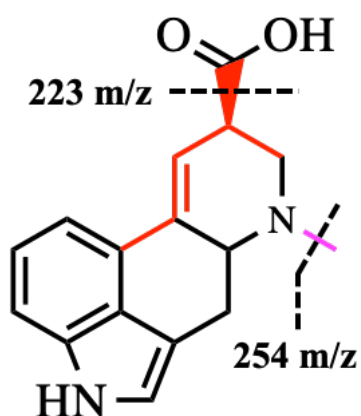
9.3. Chanoclavine-I



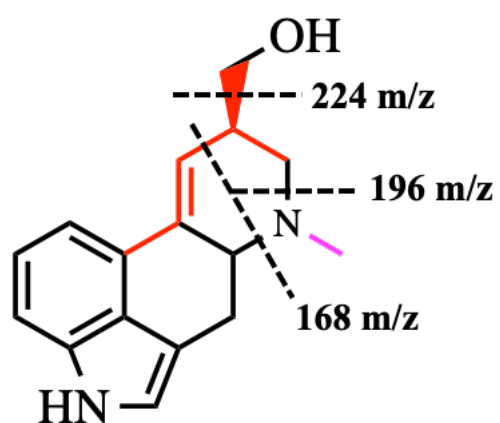
9.4. Agroclavine



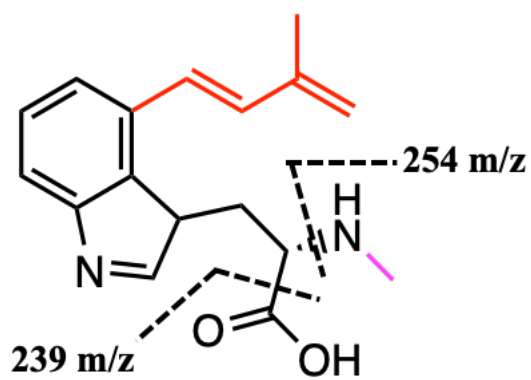
9.5. *D*-Lysergic acid



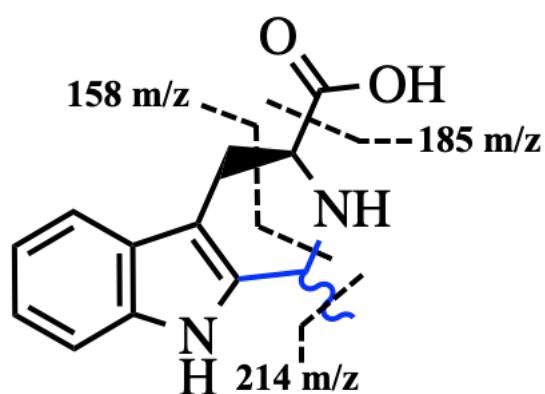
9.6. Lysergol



9.7. Diene intermediate



9.8. MTCA



9.9. DMA-MTCA

