

**The Role of Protein-Protein Interactions in Bisindole
Alkaloid Biosynthesis in *Catharanthus roseus***

Dissertation

zur Erlangung des akademischen Grades *doctor rerum naturalium*

(Dr. rer. nat)

vorgelegt dem Rat der Chemisch-Geowissenschaftlichen Fakultät
der Friedrich-Schiller-Universität Jena

von B.Sc Chloe Langley

geboren am 17.11.1995 in Gorleston-On-Sea, UK


Gutachter:

1. Prof. Dr. Arndt (Friedrich-Schiller-Universität, Jena)
2. Prof. Dr. Heintzmann (Friedrich-Schiller-Universität, Jena)
3. Prof. Dr. Hellmich (Friedrich-Schiller-Universität, Jena)

Tag der Verteidigung: 28.02.2024

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und unter Verwendung der angegebenen Hilfsmittel, persönlichen Mitteilungen und Quellen angefertigt habe.

Ort	Datum	Unterschrift der Verfasserin/des Verfassers
<u>Jena</u>	<u>28.02.2024</u>	<u>Chloe Langley</u> 

Zusammenfassung der Dissertation

Diese Arbeit befasst sich mit der mechanistischen und biophysikalischen Charakterisierung von Enzymen, die an der Biosynthese einer Gruppe von Verbindungen beteiligt sind, die als Monoterpen-Indolalkaloide (MIAs) bekannt sind. Die Biosynthese vieler MIAs ist komplex und erfordert viele biochemische Schritte, die in verschiedenen Zelltypen und subzellulären Kompartimenten organisiert sind. Angesichts der pharmakologischen Bedeutung dieser Verbindungen besteht ein großes Interesse daran, die Biochemie dieser Enzyme besser zu charakterisieren und zu verstehen, wie sie innerhalb der metabolisch komplexen Umgebung der Zelle organisiert sind.

Mehrere Alkoholdehydrogenase (ADH)-Enzyme führen atypische Reduktionsreaktionen in der MIA-Biosynthese durch, obwohl die biochemischen Grundlagen ihrer Katalyse bisher kaum verstanden wurden. In Kapitel 2 haben wir mithilfe von Röntgenkristallographie und biochemischen Studien einen Mechanismus für das ADH-Enzym Dihydroprecondylocarpinacetat-Synthase (DPAS) vorgeschlagen, das eine ungewöhnliche 1,4-Iminium-Reduktion katalysiert. Darüber hinaus berichten wir über den Mechanismus für das verwandte ADH-Enzym Geissoschizin-Synthase (GS), das eine 1,2-Iminium-Reduktion katalysiert. Mit Hilfe der vergleichenden Genomik und der phylogenetischen Analyse in Kapitel 3 decken wir die Expansion und Neofunktionalisierung der ADH-Enzymfamilie auf, was darauf hindeutet, dass sie eine Schlüsselrolle bei der Chemodiversität von MIA spielt.

Kapitel 4 untersucht die Rolle von Protein-Protein-Interaktionen zwischen MIA-Biosyntheseenzymen aus der Pflanze *Catharanthus roseus*. Besonderes Augenmerk liegt dabei auf dem ADH CrDPAS (charakterisiert in Kapitel 2) und den nachgeschalteten Cyclase-Enzymen Tabersonin-Synthase (CrTS) und Catharanthin-Synthase (CrCS), um die strukturelle Grundlage dieser Protein-Protein-Interaktion zu bestimmen. Darüber hinaus wurde das Interaktom der CrDPAS mit Hilfe von in vivo Proximity Tagging untersucht, um Interaktionen zwischen MIA und Enzymen der Phenylpropanoid-Biosynthese aufzudecken.

Die Charakterisierung dieser atypischen ADHs erweitert den chemischen Werkzeugkasten dieser Enzymfamilie und ist entscheidend für das Verständnis der Biosynthese einer breiten Palette pharmazeutisch wichtiger MIAs. Darüber hinaus hilft uns die Kartierung der Protein-

Protein-Interaktionen biosynthetischer Enzyme dabei, die physikalische Organisation von Proteinen innerhalb einer Zelle zu verstehen und kann die Grundlage für künftige Bioengineering-Anstrengungen zur Anpassung des Stoffwechselflusses bilden.

Thesis Abstract

This thesis details the mechanistic and biophysical characterisation of enzymes involved in the biosynthesis of the group of compounds known as monoterpene indole alkaloids (MIAs). The biosynthesis of many MIAs is complex, requiring many biochemical steps that are organised into different cell types and subcellular compartments. Given the pharmacological importance of these compounds, there is great interest to better characterise the biochemistry of these enzymes and to understand how they are organised within the metabolically complex environment of the cell.

Several alcohol dehydrogenase (ADH) enzymes perform atypical reduction reactions in MIA biosynthesis, though the biochemical basis of their catalysis was poorly understood. In Chapter 2, we used X-ray crystallography and biochemical studies to propose a mechanism for the ADH enzyme dihydroprecondylocarpine acetate synthase (DPAS), which catalyses an unusual 1,4-iminium reduction. Furthermore, we report the mechanism for the related ADH enzyme geissoschizine synthase (GS), which catalyses a 1,2-iminium reduction. Using comparative genomics and phylogenetic analysis in Chapter 3, we uncover the expansion and neofunctionalisation of the ADH family of enzymes, suggesting their role as key drivers in MIA chemodiversity.

Chapter 4 explores the role of protein-protein interactions between MIA biosynthetic enzymes from the plant *Catharanthus roseus*. Particular focus is given to the ADH *CrDPAS* (characterised in Chapter 2) and the downstream cyclase enzymes tabersonine synthase (*CrTS*) and catharanthine synthase (*CrCS*) to determine the structural basis of this protein-protein interaction. Furthermore, the interactions of *CrDPAS* was studied using in vivo proximity tagging, revealing inter-pathway interactions between MIA and phenylpropanoid biosynthetic enzymes.

The characterisation of these atypical ADHs expands the chemical toolbox of this enzyme family and is crucial for comprehending the biosynthesis of a wide range of pharmaceutically important MIAs. Additionally, mapping protein-protein interactions of biosynthetic enzymes helps us grasp the physical organisation of proteins within a cell and can provide the foundation of future bioengineering efforts to tailor metabolic flux.

List of Publications

Kunert, M., Langley, C., Ploss, K., Rodríguez López, C.E., Serna Guerrero, D.A., Rothe, E., Lucier, R., O'Connor, S.E., Sonawane, P.D. A promiscuous CYP87A enzyme activity initiates cardenolide biosynthesis in plants. *Nature Plants* doi: 10.1038/s41477-023-01515-9 (2023).

Stephan, B., Langley, C., Winkler, D., Basquin, J., Caputi, L., O'Connor, S.E., Kries, H. Directed Evolution of Piperazic Acid Incorporation by a Nonribosomal Peptide Synthetase. *Angewandte Chemie Int. Edition* e202304843 (2023).

Langley, C., Tatsis, E., Hong, B., Nakamura, Y., Paetz, C., Stevenson, C.E.M., Basquin, J., Lawson, D.M., Caputi, L., O'Connor, S.E. Expansion of the Catalytic Repertoire of Alcohol Dehydrogenases in Plant Metabolism. *Angewandte Chemie Int. Edition* 61, 48 (2022).

Hernández Lozada, N.J., Hong, B., Wood, J.C., Caputi, L., Basquin, J., Chuang, L., Kunert, M., Rodríguez López, C.E., Langley, C., Zhao, D., Buell, C.R., Lichman, B.R., O'Connor, S.E. Biocatalytic routes to stereo-divergent iridoids. *Nat Comms* 13, 4718 (2022).

Carqueijeiro, I., Langley, C., Grzech, D., Koudounas, K., Papon, N., O'Connor, S.E., Courdavault, V. Beyond the semi-synthetic artemisinin: metabolic engineering of plant-derived anti-cancer drugs, *Current Opinion in Biotechnology* 65 (17-24) (2020).

Dang, T., Franke, J., Carqueijeiro, I., Langley, C., Courdavault, V. and O'Connor, S. Sarpagan bridge enzyme has substrate-controlled cyclization and aromatization modes. *Nat Chem Biol* 14, 760–763 (2018).

Acknowledgements

This thesis is the culmination of 5 years of work across two countries and through a global pandemic. I would not have achieved this without the many people I have been lucky enough to have helped me throughout this time.

Firstly, I would like to thank my supervisor Sarah O'Connor who accepted me into her lab when I was an undergraduate student completely clueless to the world of academic research, and who has ever since given me her unwavering support and encouragement. I cannot thank you enough for giving me the freedom to follow my curiosities and for giving me the opportunity to be a member of the O'Connor lab.

I am grateful to all the members of the O'Connor lab – past and present. I cannot name you all but I would not be here without your help. Thank you to the members of the O'Commune, the Corona Crafting Crew, and the SOC Sockers. I've received help from all of the PhD students in the lab, but I especially want to thank Omar Kamileen for his friendship and support over the years, I am constantly amazed by the passion you have for your work and the respect and kindness you show to others. I am also very grateful to Matilde Florean for her positivity and for being the enabler in all my 'projects'.

I'd like to thank all the people who helped me achieve the work in this thesis, especially Lorenzo Caputi, Veit Gabe, Delia Serna Guerrero, Sarah Heinicke, Benke Hong, Manuela Inak, Maritta Kunert, Katrin Luck, Yoko Nakamura, Marlen Siegmund, and Kotaro Yamamoto. I especially thank Thu-Thuy Dang and Scott Farrow who both welcomed me into the lab and shared their passion for science with me as an undergraduate, without which I would not have pursued a PhD. I was also lucky to work with the extremely talented structural biologists Dave Lawson and Clare Stevenson at the JIC, and Jerome Basquin at the MPI Biochemistry. Thank you too to the institute's greenhouse team and haus technik – especially Eva Rothe, Andrea Schuenzel and Daniel Viet – without whom this work would not have been possible.

I'm grateful to my friends Elizabeth 'LizBiz' Jackson, Elin 'Swaggy' Isted, Ana 'Sloth Mother' Zadel, Joshua Daffydd Lewis, Matthew 'Ratthew' Parfitt, and the meme queens Mirha Wasim and Olive Taylor for keeping me going. Thank you to Mark and Tracey Garratley for your support and the endless supply of shortbread.

I would have not gotten through my PhD without my partner Ryan Garratley, who has rooted for me through at every step of this journey. You have believed in me when I could not see a way forward and I am so lucky to have you in my life.

I am forever grateful for the support of my family, especially my brothers Ross and Ellis, my Nanny and Grandad, and my dogs Charlie and Tilly. Lastly, I would like to thank my mum and dad for their constant love and for always being my biggest supporters. Thank you for giving me the courage to explore my passions. I would not have achieved any of this without your belief in me and I am so grateful to be your daughter.

Contents

Selbständigkeitserklärung	ii
Zusammenfassung der Dissertation	iv
Thesis Abstract.....	vi
List of Publications	viii
Acknowledgements.....	x
Contents.....	xii
List of Tables.. ..	xix
List of Figures	xxi
List of Abbreviations	xxix
Chapter 1. MIA Chemical Diversity and Biosynthesis	1
1.1. Plant Natural Products	1
1.2. Monoterpene Indole Alkaloids: Structure, Pharmacology and Biosynthetic Gene Discovery.....	3
1.3. MIA Enzyme Discovery in <i>Catharanthus roseus</i> : A Case Study of Vinblastine Biosynthesis	6
1.3.1. Corynanthe Scaffold Biosynthesis	7
1.3.2. Aspidosperma and Iboga Scaffold Biosynthesis	9
1.3.3. Vindoline Biosynthesis: Elaborating the Aspidosperma Scaffold.....	10
1.3.4. Vinblastine Biosynthesis: Dimerisation and Decoration of the Bisindole Scaffold 11	
1.4. Alcohol Dehydrogenase Enzymes: Overview and Evolutionary Origin.....	12
1.5. Protein-Protein Interactions in Plant-Specialised Metabolism.....	14
1.6. Scope of this Thesis	16
1.7. References.....	17
Chapter 2. The Structural and Mechanistic Study of CADs in MIA Biosynthesis	26

2.1	Introduction	26
2.1.1	Cinnamyl Alcohol Dehydrogenases	26
2.1.2	ADH Catalytic Mechanism	27
2.1.3	CAD-Catalysed Reductions in Monoterpene and MIA Biosynthesis	28
2.1.4	Chapter Overview	32
2.2	Results	33
2.2.1	DPAS-Catalysed Reduction of Precondylocarpine Acetate and Dehydrosecodine	33
2.2.2	Isotopic Labelling of the DPAS-Catalysed Reduction of Precondylocarpine Acetate	34
2.2.3	Stereoselectivity of the DPAS-Catalysed Production of Vincadifformine	35
2.2.4	DPAS-Catalysed Reduction of Strictosidine Aglycone	36
2.2.5	Solution of the <i>Cr</i> DPAS Structure	38
2.2.6	Solution of <i>Ti</i> DPAS2 Structure	39
2.2.7	Comparison of CAD Structures in Apocynaceae.....	43
2.2.8	<i>Cr</i> DPAS Mutagenesis Study on the Reduction of Precondylocarpine Acetate	49
2.2.9	Mechanism of <i>Cr</i> DPAS-Catalysed 1,4-Iminium Reduction of Precondylocarpine Acetate and Dehydrosecodine	52
2.2.10	<i>Cr</i> DPAS Mutagenesis Study on the Reduction of Strictosidine Aglycone..	53
2.2.11	Mechanism of <i>Cr</i> DPAS-Catalysed 1,4-Reduction of Vallesiachotamine	55
2.2.12	<i>Cr</i> GS and <i>Cr</i> THAS Mutagenesis Study on the Reduction of Strictosidine Aglycone.....	55
2.2.13	Mechanism of <i>Cr</i> GS-Catalysed 1,2-Iminium Reduction of 4,21-Dehydrogeissoschizine.....	59
2.3	Discussion.....	60

2.3.1	Elucidation of <i>CrDPAS</i> and <i>TiDPAS2</i> Structures Reveals Basis of 1,4-Reductions	60
2.3.2	Reduction of α,β -unsaturated Aldehydes in Plant Specialised Metabolism	61
2.3.3	Mechanism of <i>CrGS</i> -Catalysed 1,2-Iminium Reduction of 4,21-Dehydrogeissoschizine	63
2.4	Conclusions.....	64
2.5	Methods and Materials	65
2.5.1	Chemicals and molecular biology reagents	65
2.5.2	Cloning and mutagenesis	65
2.5.3	Protein Expression in <i>E. coli</i>	69
2.5.4	<i>CrPAS</i> Insect Cell Expression	70
2.5.5	Small-scale Protein Purification	70
2.5.6	<i>CrDPAS</i> , <i>TiDPAS2</i> , <i>CrGS</i> , <i>CrSGD</i> , <i>CrPAS</i> and <i>TbADH</i> Large-scale Protein Purification	71
2.5.7	Synthesis of NADPD	71
2.5.8	<i>In vitro</i> Enzyme Assays	72
2.5.9	UPLC-MS Analysis.....	72
2.5.10	Production and Isolation of <i>d</i> -angryline and <i>d</i> ₂ -vincadiformine.....	73
2.5.11	Production and Isolation of 19,20-dihydrovallesiachotamine.....	74
2.5.12	NMR of <i>d</i> -angryline, <i>d</i> -vincadiformine and 19,20-dihydrovallesiachotamine.....	75
2.5.13	ECD Measurement and Spectral Calculations of Vincadiformine.....	75
2.5.14	Protein Crystallisation	75
2.5.15	X-ray Data Collection, Processing and Structure Solution	76
2.5.16	Docking simulations	77
2.5.17	Phylogenetic analysis	77
2.6	References.....	79

Chapter 3. The Emergence of Atypical CADs as Drivers of MIA Chemical Diversity	83
3.1. Introduction	83
3.1.1. MIA Chemical Diversity in Gentianales	83
3.1.2. Gene Duplication and Neofunctionalisation in Plant-Specialised Metabolism	84
3.1.3. CADs in MIA Biosynthesis.....	85
3.1.4. Chapter Overview	86
3.2. Results	87
3.2.1. Genomic organisation of CADs in <i>C. roseus</i>	87
3.2.2. Phylogeny of CADs in <i>C. roseus</i>	89
3.2.3. Cell-Specific Expression Patterns of CADs in <i>C. roseus</i>	91
3.2.4. Phylogenetic Evolution of CADs in Gentianales.....	93
3.2.5. Genomic Synteny between <i>C. roseus</i> and <i>O. pumila</i>	95
3.3. Discussion.....	98
3.3.1. Emergence of CADs that Catalyse Atypical Reductions in <i>C. roseus</i>	98
3.3.2. Emergence of Atypical CADs in Gentianales Correlates with MIA Chemical Diversity.....	99
3.4. Conclusions	100
3.5. Methods and Materials.....	101
3.5.1. CAD Sequence Collection and Analysis	101
3.5.2. CAD Phylogenetic Analysis	103
3.5.3. Chemicals and molecular biology reagents	103
3.5.4. <i>CrDPAS</i> and <i>CrADH9</i> Cloning and Protein Expression in <i>E. coli</i>	103
3.5.5. Protein Purification	105
3.5.6. <i>In vitro</i> Enzyme Assays	106
3.5.7. UPLC-MS Analysis.....	106

3.5.8.	Gene Expression and Metabolite Level Analysis.....	106
3.5.9.	Synteny Analysis between <i>C. roseus</i> and <i>O. pumila</i> Genomes	107
3.6.	References.....	108
Chapter 4.	The Role of Protein-Protein Interactions in MIA Biosynthesis	113
4.1.	Introduction.....	113
4.1.1.	Protein-Protein Interactions in MIA Biosynthesis.....	113
4.1.2.	Methods Used to Detect Protein-protein Interactions.....	115
4.1.3.	Chapter Overview.....	118
4.2.	Results	119
4.2.1.	Interaction Network of <i>C. roseus</i> MIA Biosynthetic Proteins using Split-Luciferase	119
4.2.2.	Conservation of the ADH-Cyclase Protein-Protein Interaction.....	120
4.2.3.	Engineering the <i>CrDPAS</i> -Cyclase Interaction	122
4.2.4.	AP-MS Analysis of <i>CrDPAS</i> and <i>CrTS</i>	124
4.2.5.	Differential Scanning Fluorimetry Analysis of <i>CrDPAS</i> -Cyclase Complex.	124
4.2.6.	<i>In vivo</i> Proximity Tagging of <i>CrDPAS</i>	125
4.2.7.	Interactions between MIA and Phenylpropanoid Biosynthetic Enzymes	127
4.2.8.	Probing Protein-Protein Interactions in Pseudo-Scaffold MIA Biosynthesis	130
4.3.	Discussion.....	131
4.3.1.	Establishment of a <i>C. roseus</i> MIA Biosynthetic Protein-Protein Interaction Network	131
4.3.2.	Engineering MIA Enzyme-Enzyme Interactions	131
4.3.3.	Inter-pathway Interactions and Metabolic Crosstalk between MIA and Phenylpropanoid Biosynthesis in <i>C. roseus</i>	132
4.3.4.	Potential Role of Protein-Protein Interactions in Wider MIA Biosynthesis	132

4.4.	Conclusions	133
4.5.	Methods and Materials	134
4.5.1.	Chemicals and molecular biology reagents	134
4.5.2.	Cloning and mutagenesis	134
4.5.3.	Plant growth	149
4.5.4.	Transient Overexpression in <i>N. benthamiana</i>	149
4.5.5.	Split-luciferase Complementation Assays.....	150
4.5.6.	Transient Overexpression in <i>C. roseus</i>	150
4.5.7.	TurbolD Proximity-labelling in <i>C. roseus</i>	151
4.5.8.	Heterologous Expression and Protein Purification from <i>E. coli</i>	152
4.5.9.	<i>In vitro</i> Enzyme Assays	154
4.5.10.	UPLC-MS Analysis.....	154
4.5.11.	Complex Affinity Co-Purification.....	155
4.5.12.	Thermal Stability Assays.....	155
4.5.13.	Co-purification and TurbolD Sample Proteomics	156
4.5.14.	<i>C. roseus</i> Lignin Extraction and Quantification	156
4.5.15.	Phylogenetic Analysis.....	157
4.6.	References.....	159
Chapter 5.	Conclusions and Future Perspectives	163
5.1.	The Structural and Mechanistic Basis of Atypical CAD-Catalysed Reductions in MIA Biosynthesis	163
5.2.	Phylogenetic Study Reveals the Neofunctionalisation of CADs as Drivers of MIA Chemical Diversity	163
5.3.	Protein-Protein Interactions in MIA Biosynthesis and Beyond	164
5.4.	Future Directions: Biochemical and Biophysical Insights into MIA Biosynthesis and Beyond	165

5.5. References.....	167
----------------------	-----

Appendices 168

Appendix I. NMR Characterisation of <i>d</i> -angryline	168
Appendix II. NMR Characterisation of (-)-vincadifformine	170
Appendix III. NMR Characterisation of <i>d</i> ₂ -(+)-vincadifformine.....	173
Appendix IV. NMR Characterisation of 19,20-dihydrovallesiachotamine	176
Appendix V – X-Ray Data and Model Parameters used for Structure Solutions	196
Appendix VI. Identification of CADs in <i>C. roseus</i>	201
Appendix VII. Genomic Synteny between <i>C. roseus</i> and <i>O. pumila</i>	207
Appendix VIII. Subcellular Localisation of <i>C. roseus</i> MIA Biosynthetic Enzymes.....	212
Appendix IX. Split-Luciferase Assays of <i>C. roseus</i> MIA Biosynthetic Enzymes.....	213
Appendix X. Split-Luciferase Assays of CADs and α/β hydrolases	260
Appendix XI. Split-Luciferase Assays of Cyclase Enzyme Interaction Engineering	279
Appendix XII. AP-MS of CrDPAS and CrTS	282
Appendix XIII. DFS of CrDPAS	283
Appendix XIV. CrDPAS <i>In Vivo</i> Proximity Tagging	285
Appendix XV. Split-Luciferase Assays between <i>C. roseus</i> MIA and Phenylpropanoid Biosynthetic Enzymes	291
Appendix XVI. Split-Luciferase Assays of Ψ -Tabersonine Biosynthetic Enzymes	307

List of Tables

Table 1. Amino acid sequence identity (%) of the solved structures of CADs from Apocynaceae.....	44
Table 2. Primer sequences used for gene amplification and site-directed mutagenesis in Chapter 2.....	66
Table 3. Full length nucleotide sequences of genes used in Chapter 2.	67
Table 4. Genbank accession for sequences used to construct tree of maximum likelihood in Chapter 2.....	77
Table 5. Transcriptomes and genomes accessed for CAD sequence collection in Chapter 3.	101
Table 6. Full length nucleotide sequences of <i>CrDPAS</i> and <i>CrADH9</i> genes.....	104
Table 7. Primer sequences used in for gene amplification. In Chapter 3.	105
Table 8. Full-length nucleotide sequences of unpublished and synthetic genes used in Chapter 4.....	135
Table 9. GenBank accessions of previously deposited sequences used in Chapter 4.	142
Table 10. Primer sequences used for gene amplification and site-directed mutagenesis in Chapter 4.....	143
Table 11. Genbank accessions for sequences used to construct the tree of maximum likelihood in Chapter 4.....	157
Table 12. Comparison of ¹ H NMR data between angrinine and d-angrinine.	168
Table 13. ¹ H NMR data for (-)-vincadifformine in MeOH-d ₃	170
Table 14. ¹ H NMR data for d ₂ -(\pm)-vincadifformine in MeOH-d ₃	173
Table 15. ¹ H NMR spectra for 19,20-dihydrovallesiachotamine in MeOH-d ₃	177
Table 16. Summary of X-ray data and model parameters for <i>CrDPAS</i>	197
Table 17. Summary of X-ray data and model parameters for apo- <i>TiDPAS2</i>	198
Table 18. Summary of X-ray data and model parameters for precondylocarpine acetate-bound <i>TiDPAS2</i>	199
Table 19. Summary of X-ray data and model parameters for stemmadenine acetate-bound <i>TiDPAS2</i>	200
Table 20. Contigs encoding CADs retrieved from the <i>C. roseus</i> genome.....	201
Table 21. Contigs of CADs identified in the <i>C. roseus</i> leaf single-cell transcriptome.....	205

Table 22. Genes encoding enzymes involved in MIA biosynthesis or CADs in <i>C. roseus</i> that have genomic synteny with <i>O. pumila</i>	207
Table 23. Experimentally or predicted subcellular localisation of <i>C. roseus</i> MIA biosynthetic enzymes.	212
Table 24. Nucleic acid sequence identity of CADs tested for the α/β hydrolase interaction conservation.....	260
Table 25. Nucleic acid sequence identity of α/β hydrolases tested for ADH interaction conservation.....	260
Table 26. AP-MS results using <i>CrDPAS</i> or <i>CrTS</i> in <i>C. roseus</i> leaf tissue.	282
Table 27. Summary of <i>CrDPAS</i> $\Delta T^{\circ}m$ compared to <i>CrDPAS</i> control in various conditions investigating cyclase complex formation.....	283
Table 28. Contig matches with ≥ 1.5 log fold change in peptide hits between <i>CrDPAS</i> -TurboID protein pulldown compared to YFP-TurboID protein pulldown control.	285

List of Figures

Figure 1. Overview of the major classes of plant natural products.	1
Figure 2. MIA scaffolds generated from central precursor strictosidine.	3
Figure 3. Timeline of significant advances in MIA elucidation and biosynthetic pathway discovery	4
Figure 4. Structures of MIAs vinblastine and vincristine produced in <i>C. roseus</i>	6
Figure 5. Biosynthesis of corynanthe MIA scaffold in <i>C. roseus</i>	8
Figure 6. Biosynthesis of iboga and aspidosperma MIA scaffolds in <i>C. roseus</i>	9
Figure 7. Vindoline biosynthesis in <i>C. roseus</i>	11
Figure 8. Coupling of catharanthine and vindoline to form bisindole MIAs in <i>C. roseus</i>	12
Figure 9. Emergence and evolution of ADH enzyme family.....	13
Figure 10. Classifications of protein-protein complexes	14
Figure 11. Reversible reduction of aromatic cinnamyl aldehydes by CAD	26
Figure 12. Structural features of AtCAD5	27
Figure 13. Mechanism of the <i>Cr</i> 8HGO-catalysed reduction of 8-oxogeraniol.	27
Figure 14. CAD-catalysed reactions in vinblastine biosynthesis in <i>C. roseus</i>	28
Figure 15. <i>Cr</i> 8HGO catalysed oxidation of 8-hydroxygeraniol to 8-oxogeraniol.....	29
Figure 16. Characterised 1,2-iminium reduction reactions catalysed by CADs in MIA biosynthesis	30
Figure 17. DPAS-catalysed 1,4-iminium reduction of precondylocarpine acetate	31
Figure 18. UPLC-MS chromatograms of in vitro reactions of <i>Cr</i> DPAS and <i>Cr</i> GS with substrate precondylocarpine acetate and cofactor NADPH.....	33
Figure 19. Deuterium labelling of DPAS-catalysed 1,4-reductions of precondylocarpine acetate and dehydrosecodine	35
Figure 20. Stereoselectivity of vincadifformine cyclisation.....	36
Figure 21. DPAS-catalysed reduction of vallesiachotamine	37
Figure 22. Crystallisation of <i>Cr</i> DPAS	38
Figure 23. Apo- <i>Cr</i> DPAS crystal structure.	39
Figure 24. Apo- <i>Ti</i> DPAS2 crystal structure.....	40
Figure 25. Structure of <i>Ti</i> DPAS2 bound to precondylocarpine acetate	41
Figure 26. Structure of <i>Ti</i> DPAS2 bound to stemmadenine acetate.	42

Figure 27. Conformational changes between apo- and substrate-bound <i>TiDPAS</i>	43
Figure 28. MUSCLE amino acid alignment of solved structures of CADs from Apocynaceae and their respective PDB accessions.....	45
Figure 29. Catalytic zinc coordination in CADs from Apocynaceae.....	46
Figure 30. Structural zinc coordination in CADs from Apocynaceae.....	46
Figure 31. Cofactor binding in <i>Cr8HGO</i> and <i>CrGS</i>	47
Figure 32. Comparison of cofactor positioning between <i>Cr8HGO</i> , <i>CrGS</i> and <i>TiDPAS2</i>	48
Figure 33. Docking of NADPH cofactor in <i>TiDPAS2</i>	49
Figure 34. <i>CrDPAS</i> active site residues targeted by site-directed mutagenesis.	50
Figure 35. <i>In vitro</i> reactions of <i>CrDPAS</i> mutants with preconditionylcarpine acetate.....	51
Figure 36. Proposed mechanism of <i>CrDPAS</i> -catalysed reduction of preconditionylcarpine acetate	52
Figure 37. Computational docking of vallesiachotamine and NADPH in <i>CrDPAS</i>	53
Figure 38. Site-directed mutagenesis of <i>CrDPAS</i> on formation of 19,20-dihydrovallesiachotamine.....	54
Figure 39. Proposed mechanism of the <i>CrDPAS</i> -catalysed 1,4-reduction of vallesiachotamine.	55
Figure 40. Computational docking of 4,21-dehydrogeissoschizine in <i>CrGS</i> with cofactor NADP ⁺	56
Figure 41. LC-MS chromatograms of <i>in vitro</i> reactions of <i>CrGS</i> and mutants with substrate strictosidine, <i>CrSGD</i> and cofactor NADPH	57
Figure 42. UPLC-MS chromatograms of <i>CrGS</i> and <i>CrGS</i> Phe53Tyr mutant, and the corresponding <i>CrTHAS</i> and <i>CrTHAS</i> Tyr56Phe mutant <i>in vitro</i> reactions with substrate strictosidine, <i>CrSGD</i> and cofactor NADPH.	58
Figure 43. Proposed catalytic mechanism of <i>CrGS</i> catalysed 1,2-iminium reduction of 4,21-dehydrogeissoschizine to form geissoschizine.	59
Figure 44. Tree of maximum likelihood of previously characterised plant CADs.	61
Figure 45. Proposed mechanism of <i>Rauwolfia</i> VR2 catalysed 1,4-reduction of α,β -unsaturated aldehyde vomilenine.	62
Figure 46. Enzyme-catalysed reductions of α,β -unsaturated aldehydes in plant specialised metabolism.	62
Figure 47. Mechanism of <i>CrISY</i> catalysed reduction of 8-oxogeranial.....	63

Figure 48. CAD-catalysed reductions of strictosidine aglycone in <i>C. roseus</i>	83
Figure 49. Genomic organisation of CADs in <i>C. roseus</i>	87
Figure 50. 1,2-iminium reductions of strictosidine aglycone rearrangements catalysed by <i>CrGS</i> , <i>CrGS2</i> and <i>CrTHAS1-4</i>	89
Figure 51. Maximum likelihood phylogenetic tree of CADs in <i>C. roseus</i>	90
Figure 52. Maximum likelihood phylogenetic tree of 1,2-iminium reducing CADs in <i>C. roseus</i>	91
Figure 53. Hierarchical clustered heatmap of cell-type-specific expression patterns of CADs in <i>C. roseus</i> leaves.	92
Figure 54. Maximum likelihood phylogenetic tree of CADs in the order Gentianales	94
Figure 55. Maximum likelihood phylogenetic tree of species in the order Gentianales. ...	95
Figure 56. Biosynthetic pathway of the MIAs camptothecin and geissoschizine in <i>O. pumila</i> and <i>C. roseus</i> respectively, utilising the shared precursor strictosidine.	96
Figure 57. Syntenic analysis of <i>C. roseus</i> and <i>O. pumila</i> genomes.....	97
Figure 58. Cellular and subcellular localisation of elucidated steps of vinblastine biosynthesis in <i>C. roseus</i>	113
Figure 59. Previously studied enzyme-enzyme complexes in <i>C. roseus</i> MIA biosynthesis	114
Figure 60. Overview of methods used to detect protein-protein interactions discussed in this chapter	116
Figure 61. Pairwise testing of protein-protein interactions of <i>C. roseus</i> MIA biosynthetic enzymes enriched in epidermal cells using a split-luciferase assay	119
Figure 62. Proposed mechanism of tabersonine, (+)-catharanthine, and (-)-coronaridine formation from the intermediate dehydrosecodine.	120
Figure 63. Protein-protein interactions between ADHs and α/β hydrolases.....	121
Figure 64. Engineering <i>CrDPAS</i> -cyclase interaction	123
Figure 65. $T^{\circ}m$ of <i>CrDPAS</i> and <i>CrTS</i> individually and at stoichiometric ratios.	125
Figure 66. <i>In vivo</i> proximity labelling of <i>CrDPAS</i> in <i>C. roseus</i>	126
Figure 67. Inter-pathway protein-protein interactions between <i>C. roseus</i> MIA and phenylpropanoid biosynthetic enzymes.....	128
Figure 68. Metabolic crosstalk between phenylpropanoid and MIA biosynthesis in <i>C. roseus</i>	129

Figure 69. Biosynthesis of Ψ -tabersonine in <i>T. iboga</i>	130
Figure 70. Representative image of split-luciferase positive control <i>AtCHIL</i> and <i>AtCHS</i> in <i>N. benthamiana</i>	150
Figure 71. SDS-PAGE of TurboID streptavidin pull-down assays.....	152
Figure 72. Comparison of ^1H spectra of angryline and d-angryline. Loss of signal shown at H19, indicating deuterium incorporation.	169
Figure 73. ^1H NMR data for m/z 339, (-)-vincadifformine (standard).....	171
Figure 74. ^1H NMR data for m/z 339, (-)-vincadifformine (standard).....	172
Figure 75. Phase sensitive HSQC NMR data for m/z 341, d_2 -(\pm)-vincadifformine full range in $\text{MeOH-}d_3$	174
Figure 76. Phase sensitive HSQC NMR data for m/z 341, d_2 -(\pm)-vincadifformine, aliphatic range in $\text{MeOH-}d_3$	175
Figure 77. MS/MS spectra of 19,20-dihydrovallesiachotamine.	176
Figure 78. ^1H NMR data of 19,20-dihydrovallesiachotamine with water suppression, full range in $\text{MeOH-}d_3$	178
Figure 79. ^1H NMR data of 19,20-dihydrovallesiachotamine with water suppression, aldehyde range in $\text{MeOH-}d_3$	179
Figure 80. ^1H NMR data of 19,20-dihydrovallesiachotamine with water suppression, aromatic range in $\text{MeOH-}d_3$	180
Figure 81. ^1H NMR data of 19,20-dihydrovallesiachotamine with water suppression, aliphatic range in $\text{MeOH-}d_3$	181
Figure 82. NMR data of 19,20-dihydrovallesiachotamine, phase sensitive HSQC, full range in $\text{MeOH-}d_3$	182
Figure 83. NMR data of 19,20-dihydrovallesiachotamine, phase sensitive HSQC, aldehyde and aromatic range in $\text{MeOH-}d_3$	183
Figure 84. NMR data of 19,20-dihydrovallesiachotamine, phase sensitive HSQC, aliphatic range in $\text{MeOH-}d_3$	184
Figure 85. NMR data of 19,20-dihydrovallesiachotamine, HMBC, full range in $\text{MeOH-}d_3$	185
Figure 86. NMR data of 19,20-dihydrovallesiachotamine, HMBC, aldehyde and aromatic range in $\text{MeOH-}d_3$	186

Figure 87. NMR data of 19,20-dihydrovallesiachotamine, HMBC, aliphatic range in MeOH- d_3	187
Figure 88. NMR data of 19,20-dihydrovallesiachotamine, DEPTQ, power spectrum, full range in MeOH- d_3	188
Figure 89. NMR data of 19,20-dihydrovallesiachotamine, 1H - 1H DQF COSY with water suppression, magnitude mode processed, full range in MeOH- d_3	189
Figure 90. NMR data of 19,20-dihydrovallesiachotamine, 1H - 1H DQF COSY with water suppression, magnitude mode processed, aldehyde and aromatic range in MeOH- d_3 ..	190
Figure 91. NMR data of 19,20-dihydrovallesiachotamine, 1H - 1H DQF COSY with water suppression, magnitude mode processed, aliphatic range in MeOH- d_3	191
Figure 92. NMR data of 19,20-dihydrovallesiachotamine, 1H - 1H ROESY with water suppression, full range in MeOH- d_3	192
Figure 93. NMR data of 19,20-dihydrovallesiachotamine, 1H - 1H ROESY with water suppression, aldehyde and aromatic range in MeOH- d_3	193
Figure 94. NMR data of 19,20-dihydrovallesiachotamine, 1H - 1H ROESY with water suppression, aliphatic range in MeOH- d_3	194
Figure 95. Structure of 19,20-dihydrovallesiachotamine optimized using Gaussian 16 (DFT APFD/6-311G++(2d,p), solvent MeOH).	195
Figure 96. Gene structure of selected CADs from <i>C. roseus</i>	203
Figure 97 UPLC-MS chromatograms of in vitro reactions of CrADH9 and CrDPAS with substrate precondylocarpine acetate and cofactor NADPH	204
Figure 98. Amino acid sequence alignment of <i>OpADH1</i> and <i>OpADH2</i> with <i>C. roseus</i> CADs that catalyse a 1,2-reduction of an iminium moiety.	209
Figure 99. Hierarchical clustering of tissue-specific gene expression of CADs, and orthologues of <i>C. roseus</i> secoiridoid and MIA biosynthetic genes in <i>O. pumila</i>	210
Figure 100. Hierarchical clustering of tissue-specific accumulation of metabolites from <i>O. pumila</i>	211
Figure 101. Representative images of pairwise interactions of CrLAMT with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	215
Figure 102. Representative images of pairwise interactions of CrSLS with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	217

Figure 103. Representative images of pairwise interactions of <i>CrTDC</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	220
Figure 104. Representative images of pairwise interactions of <i>CrGS</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	223
Figure 105. Representative images of pairwise interactions of <i>CrGO</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	225
Figure 106. Representative images of pairwise interactions of <i>CrRedOx1</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	228
Figure 107. Representative images of pairwise interactions of <i>CrRedOx2</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	231
Figure 108. Representative images of pairwise interactions of <i>CrSAT</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	234
Figure 109. Representative images of pairwise interactions of <i>CrPAS</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	237
Figure 110. Representative images of pairwise interactions of <i>CrDPAS</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	240
Figure 111. Representative images of pairwise interactions of <i>CrTS</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	243
Figure 112. Representative images of pairwise interactions of <i>CrCS</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	246
Figure 113. Representative images of pairwise interactions of <i>CrCorS</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	249
Figure 114. Representative images of pairwise interactions of <i>CrT16H2</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	251
Figure 115. Representative images of pairwise interactions of <i>Cr16OMT</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	254
Figure 116. Representative images of pairwise interactions of <i>CrT3O</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	256
Figure 117. Representative images of pairwise interactions of <i>CrT3R</i> with <i>C. roseus</i> MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	259
Figure 118. Representative images of pairwise interactions of <i>AtCAD4</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	261

Figure 119. Representative images of pairwise interactions of <i>CrCAD</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	263
Figure 120. Representative images of pairwise interactions of <i>Cr2141</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	265
Figure 121. Representative images of pairwise interactions of <i>CrADH9</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	267
Figure 122. Representative images of pairwise interactions of <i>TiDPAS1</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	268
Figure 123. Representative images of pairwise interactions of <i>TiDPAS2</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	269
Figure 124. Representative images of pairwise interactions of <i>CrDPAS</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	271
Figure 125. Representative images of pairwise interactions of <i>CrRedOx1</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	273
Figure 126. Representative images of pairwise interactions of <i>CrGS</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	275
Figure 127. Representative images of pairwise interactions of <i>CrTHAS</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	276
Figure 128. Representative images of pairwise interactions of <i>CrT3R</i> with α/β -hydrolases tested by split-luciferase in <i>N. benthamiana</i>	278
Figure 129. Amino acid sequence alignment of cyclase enzymes from <i>C. roseus</i> and <i>T. iboga</i>	279
Figure 130. Representative images of pairwise interactions of <i>CrDPAS</i> interaction engineering with <i>CrCors</i> or <i>TiCorS</i> tested by split-luciferase in <i>N. benthamiana</i>	281
Figure 131. Effect of pH and [NaCl] on <i>CrDPAS</i> T ^m in the presence or absence of cyclase <i>CrTS</i> or <i>CrCS</i> added at equimolar concentrations.	284
Figure 132. Representative images of pairwise interactions of <i>CrC4H</i> with <i>C. roseus</i> lignin and MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	292
Figure 133. Representative images of pairwise interactions of <i>CrC3H</i> with <i>C. roseus</i> lignin and MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	294
Figure 134. Representative images of pairwise interactions of <i>CrCSE</i> with <i>C. roseus</i> lignin and MIA biosynthetic enzymes tested by split-luciferase in <i>N. benthamiana</i>	298

Figure 135. Representative images of pairwise interactions of *CrCCR* with *C. roseus* lignin and MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*302

Figure 136. Representative images of pairwise interactions of *CrCAD* with *C. roseus* lignin and MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*306

Figure 137. Representative images of pairwise interactions of *TiPAS1* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*308

Figure 138. Representative images of pairwise interactions of *TiPAS2* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*310

Figure 139. Representative images of pairwise interactions of *TiPAS3* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*312

Figure 140. Representative images of pairwise interactions of *CrPAS* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*313

List of Abbreviations

7DLH	7-deoxyloganic acid hydroxylase
8HGO	8-hydroxygeraniol oxidoreductase
16OMT	16-hydroxytabersonine- <i>O</i> -methyltransferase
βHAD	β-hydroxy acid dehydrogenase
ADH	Alcohol dehydrogenase
AKR	Aldo-keto reductase
AP-MS	Affinity purification mass spectrometry
BBE	Berberine bridge-like enzyme
BGC	Biosynthetic gene cluster
BiFC	Bimolecular fluorescence complementation
BirA	Biotin ligase
BP	Base pairs
C3'H	<i>p</i> -Coumaroyl ester 3-hydroxylase
C4H	Cinnamate-4-hydroxylase
CAD	Cinnamyl alcohol dehydrogenase
CCD	Charge-coupled device
CHIL	Chalcone isomerase-like
CHS	Chalcone synthase
CorS	Coronaridine synthase
COSY	Correlation Spectroscopy
CS	Catharanthine synthase
CSE	Caffeoyl shikimate esterase
CV	Column volumes
CXE	Carboxylesterase
D4H	Deacetoxyvindole 4'-hydroxylase
DAT	Deacetylvindoline 4- <i>O</i> -acetyl-transferase
DMAPP	Dimethylallyl diphosphate
DPAS	Dihydroprecondylocarpine acetate synthase
DSF	Differential scanning fluorimetry
ECD	Electronic circular dichroism

EDTA	Ethylenediaminetetraacetic acid
EIC	Extracted ion chromatogram
F5H	Ferulate-5-hydroxylase
FAD	Flavin adenine dinucleotide
G8H	Geraniol 8-hydroxylase
GES	Geraniol synthase
GGPPS	Geranylgeranyl diphosphate synthase
GroES	Common operon GroE Small protein
GS	Geissoschizine synthase
GO	Geissoschizine oxidase
HIDH	Hydroxyisoflavone dehydratase
HMBC	Heteronuclear multiple-bond correlation spectroscopy
HSQC	Heteronuclear single quantum correlation
HPLC	High-performance liquid chromatography
HYS	Heteroyohimbine synthase
IO	Iridoid oxidase
IPAP	Internal phloem-associated parenchyma
IPP	Isopentenyl pyrophosphate
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRED	Imine reductase
ISY	Iridoid synthase
LAMT	Loganic acid methyltransferase
MATE	Multidrug and toxic compound efflux
MDR	Medium-chain dehydrogenase/reductase
MEP	2C-methyl-D-erythritol-4-phosphate
MIA	Monoterpene indole alkaloid
MVA	Mevalonic acid
NADP(H)	Nicotinamide adenine dinucleotide phosphate (hydride)
NMR	Nuclear magnetic resonance
NMT	N-methyl transferase
NPF	Nitrate/peptide family

NS	Norfluorocurarine synthase
OMT	<i>O</i> -methyltransferase
P5βR	Progesterone 5- β -reductase
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
PAS	Precondylocarpine acetate synthase
PDB	Protein database
PNP	Plant natural product
PRX	Peroxidase
RedOx1	Reductase/oxidase 1
RedOx2	Reductase/oxidase 2
RMSD	Root mean square deviation
ROESY	Rotating frame nuclear Overhauser effect spectroscopy
SAT	Stemmadenine acetyltransferase
SDR	Short-chain dehydrogenase/reductase
SGD	Strictosidine deglucosidase
SLS	Secologanin synthase
SLTr	Secologanin transporter
STR	Strictosidine synthase
T3O	Tabersonine 3'-oxidase
T3R	Tabersonine-3-reductase
T11H	Tabersonine-11-hydroxylase
TCA	Tricarboxylic acid
TCEP	Tris(2-carboxyethyl)phosphine
TDC	Tryptophan decarboxylase
TE	Transposable elements
THAS	Tetrahydroalstonine synthase
TIC	Total ion chromatogram
T_m	Melting temperature
TS	Tabersonine synthase
UPLC-MS	Ultra-performance liquid chromatography mass spectrometry
VR	Vomilenine reductase

Chapter 1. MIA Chemical Diversity and Biosynthesis

1.1. Plant Natural Products

Plant natural products (PNPs), also known as specialised metabolites, comprise a diverse range of over 245,000 phytochemicals [1]. As sessile organisms, plants utilise these compounds for various biological functions including defence responses and as chemical signals [2] to overcome biotic and abiotic stresses. In addition to their ecological roles, PNPs have been harnessed by humans for thousands of years in traditional medicines to treat a range of diseases [3]. In 1805, the isolation of the first medicinal PNP morphine marked the beginning of the modern drug era. Today, 73% of pharmaceuticals are derived from natural products and PNPs continue to be a source of inspiration for novel drug design [4]. The diversity of PNPs can be largely divided into three major classes of compound: phenolics, terpenoids and alkaloids (Figure 1).

Phenolics are phytochemicals that contain one or more hydroxyl groups attached directly to an aromatic ring. These compounds are produced through the shikimate pathways from the precursor amino acids L-phenylalanine and L-tyrosine and include major groups such as flavonoids, stilbenes and lignins. Biological roles of phenolics include UV protection, structural integrity and nutrient uptake [5, 6].

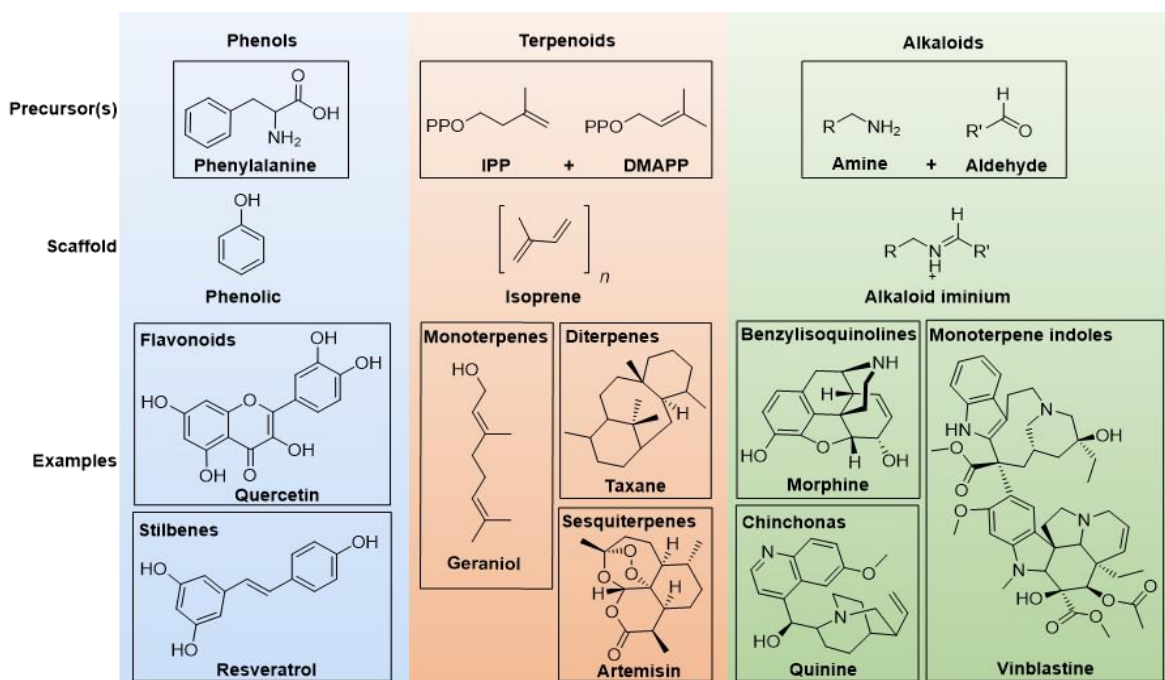


Figure 1. Overview of the major classes of plant natural products.

Terpenoids are produced from the condensation of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) to form an isoprene scaffold ($(C_5H_8)_n$). In plants, monoterpenoids ($n=2$) and diterpenoids ($n=4$) are typically synthesised via the plastid-localised 2C-methyl-D-erythritol-4-phosphate (MEP) pathway, whilst sesquiterpenoids ($n=3$) and triterpenoids ($n=6$) are synthesised from the cytosolic mevalonic acid (MVA) pathway, with some exceptions utilising both pathways ^[7]. Terpenoids exhibit diverse biological functions including as signalling molecules to attract pollinators and repel herbivores ^[8]. Additionally, some of these compounds have medicinal applications such as the anti-malarial artemisinin, the anti-cancer agent taxol and the anti-hyperglycemic stevioside ^[9].

Alkaloids are nitrogen-containing natural products that are categorised based on their biogenesis. True alkaloids derive their nitrogen atom(s) from an amino acid to form a heterocycle, whilst pseudo-alkaloids incorporate nitrogen enzymatically at a later stage of biosynthesis, and protoalkaloids do not form an amine heterocycle. The origin of the precursor molecules further classifies true alkaloids into families such as the benzyloquinolines, the monoterpene indoles and the chinchonas ^[10, 11]. In plants, alkaloids act as anti-herbivory and growth-regulating agents, however, these compounds have also garnered significant attention due to their pharmaceutical applications ^[12]. Currently, over 50 alkaloids are used in marketed medicines including the anti-malarial chinchona alkaloid quinine, the analgesic benzyloquinoline alkaloid morphine and the anti-cancer monoterpene indole alkaloid vinblastine ^[13].

Plants have evolved a diverse chemical arsenal to combat various biotic and abiotic stresses. Serendipitously, humans later harnessed this chemical repository for medicinal applications and continue to use nature as a source of inspiration for modern drug discovery and design. PNPs therefore remain a rich reservoir of nature's chemodiversity and understanding their biosynthesis is of significant biological and pharmaceutical importance.

1.2. Monoterpene Indole Alkaloids: Structure, Pharmacology and Biosynthetic Gene Discovery

Monoterpene indole alkaloids (MIAs) constitute a large class of compounds that have garnered considerable attention due to their structural diversity and pharmacological activities [14]. MIAs are exclusive to higher plants and are found in Apocynaceae, Gelsemiaceae, Loganiaceae and Rubiaceae families within the Gentianales order, and the Nyssaceae family from the Cornales order [15]. These alkaloids are composed of a monoterpene-derived unit (secologanin) and a tryptamine-derived indole unit, which undergo stereoselective condensation, forming the central precursor molecule strictosidine [16]. Subsequent formation of a central heterocycle from the secologanin-derived 9C unit forms the five major MIA scaffold types: corynanthe, aspidoasperma, iboga, yohimbine and strychnos (Figure 2). The structural diversity of MIAs enables their plethora

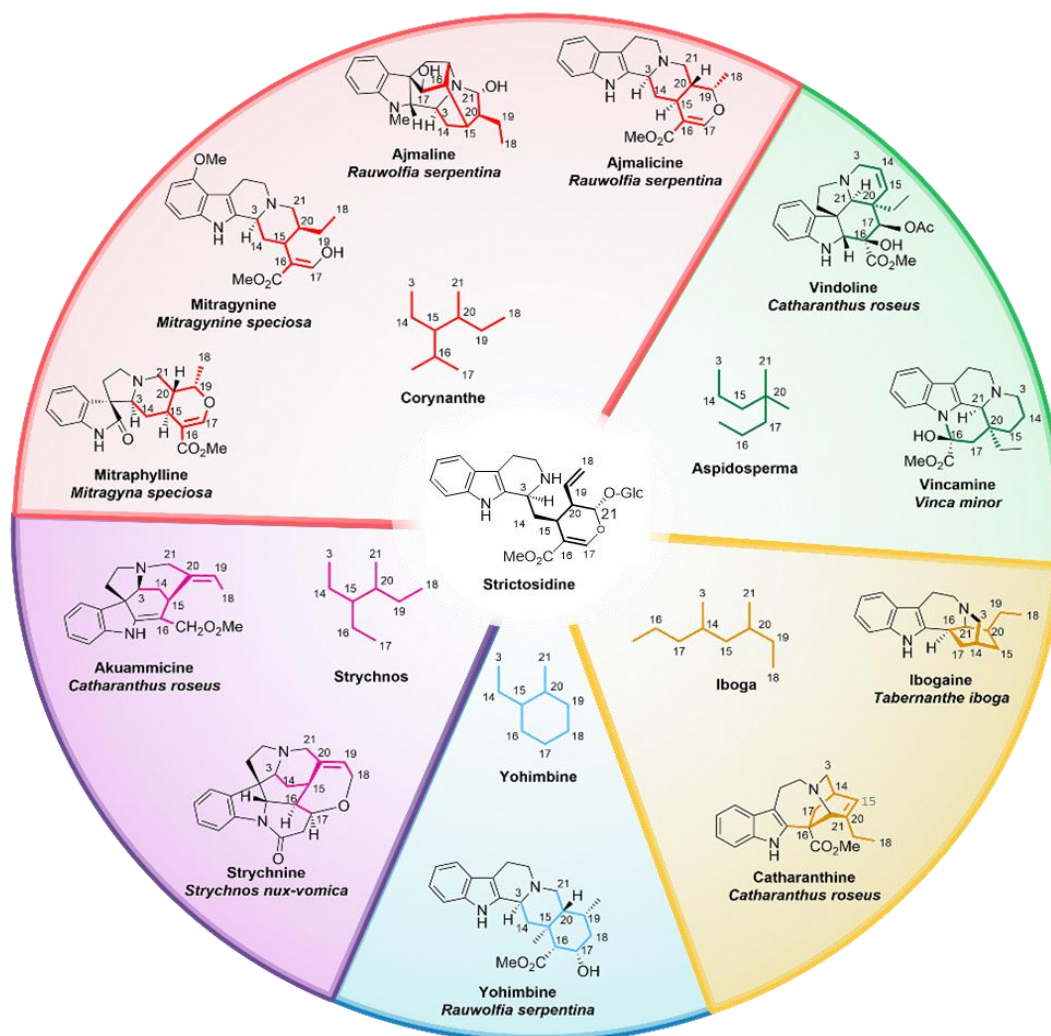


Figure 2. MIA scaffolds generated from central precursor strictosidine. Carbon numbering based on Scott *et al.*, [17]. Figure made using Biorender.

of pharmaceutical applications including the anti-cancer agent vinblastine, the anti-arrhythmic ajmaline and the convulsant strychnine. However, their often limited accumulation *in planta* has driven research to access these compounds using biological and chemical synthesis, as summarised in Figure 3.

Seminal research in the 1970s used radio-isotopic labelling of plant tissue and plant cell cultures to identify key pathway intermediates and form a chemical hypothesis of MIA biosynthesis [17–20]. These studies proposed the compound strictosidine as the key precursor of the MIA scaffold, which was later validated using molecular biology techniques to identify the corresponding biosynthetic enzyme strictosidine synthase (STR) [21–24]. However, the identification of MIA biosynthetic enzymes and the genes that encode them was hampered by a time-consuming “grind and find” approach, where enzymes were identified through purification from plant tissue or cell cultures. Additionally, the lack of accessible pathway intermediates for activity testing poses a challenge that remains in PNP pathway discovery.

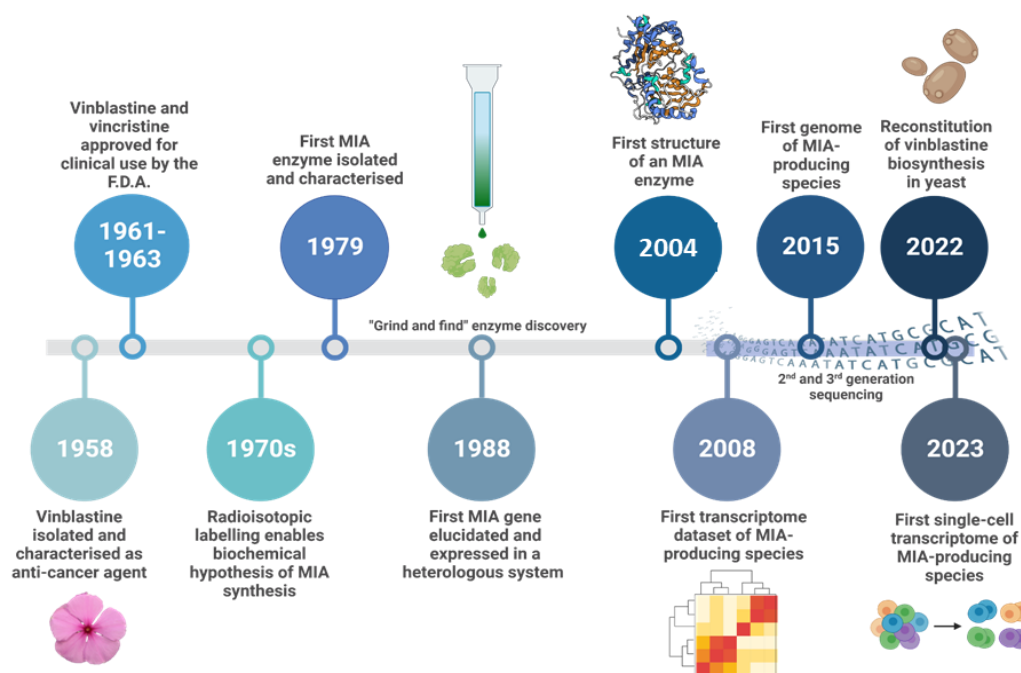


Figure 3. Timeline of significant advances in MIA elucidation and biosynthetic pathway discovery. Events include the isolation of vinblastine [53], the discovery of the first MIA biosynthetic enzyme [21] and gene [22], the first crystal structure of an MIA enzyme [81], the first transcriptome [149] and genome of a MIA-producing species [29], the reconstitution of vinblastine biosynthesis in yeast [46] and the generation of MIA-producing species single-cell transcriptomic and metabolomics [34]. Figure made using Biorender.

Advances in analytical and molecular biology techniques have accelerated the discovery of PNPs. The development of analytical instrumentation with improved sensitivity such as higher-field nuclear magnetic resonance (NMR) spectroscopy and ultra-performance liquid chromatography-mass spectrometry (UPLC-MS) has enabled the detection and characterisation of PNPs at increasingly smaller quantities, addressing the challenge posed by low-accumulating biosynthetic intermediates [25, 26]. These techniques are often used with computational metabolomics tools to identify PNPs within a complex sample [26].

The advent of second- and third-generation sequencing techniques in the last 15 years has facilitated the generation of transcriptomic and genomic data for numerous medicinal plants [27, 28]. Technologies such as PacBio and Illumina sequencing have enabled the assembly of numerous plant genomes [29–34], which were historically challenging due to their size and the presence of highly repetitive sequences [35]. This has enabled the identification of plant biosynthetic gene clusters (BGCs) [29, 31, 32, 36, 37], though these remain significantly less prevalent than those reported in bacteria. The availability of sequencing data has supported the discovery of MIA biosynthetic genes through comparative genomics, whereby genes are identified by encoding enzymes with homology to previously characterised enzymes, with similar protein structures often correlating with similar functions [38]. Co-expression analysis is also commonly used to identify genes that are expressed in a similar tissue- or cell type-specific pattern to known genes encoding other biosynthetic enzymes within a pathway. This technique has also been integrated with the co-localisation of biosynthetic intermediates identified by metabolomics to further improve the selection of candidate genes [34].

In addition to sequencing, advances in molecular biology techniques have accelerated the testing of candidate biosynthetic genes. Cloning techniques such as GoldenGate have streamlined the assembly of multiple genes within a single plasmid, enabling the efficient manipulation of gene constructs [39]. In conjunction, various heterologous hosts have been developed to express and test the activity of biosynthetic enzymes [40]. These activity screens often employ a combinatorial approach to increase throughput and identify enzymes that act on unstable or hard-to-access intermediates [41]. Moreover, virus-induced gene silencing (VIGS) has been developed in a range of plants to investigate the *in vivo* function of genes, providing valuable insights into their biological roles [42].

Collectively, these technological advances have substantially accelerated gene discovery in plants, thereby facilitating the elucidation of the biosynthetic pathways for several pharmaceutically important MIAs [43, 44], including the anti-cancer agent vinblastine from *Catharanthus roseus* as detailed in section 1.3. The discovery of PNP biosynthetic pathways provides insights into the biochemistry utilised by plants to synthesise these structurally complex compounds. Furthermore, it enables the production of MIAs [45–47] and non-natural MIA analogues through synthetic biology and metabolic engineering approaches [48–51]. The study of MIA biosynthesis thus contributes to a deeper understanding of nature’s chemical diversity and provides us greater access to this class of pharmacologically important compounds.

1.3. MIA Enzyme Discovery in *Catharanthus roseus*: A Case Study of Vinblastine Biosynthesis

C. roseus, commonly known as the Madagascar periwinkle or *Vinca rosea*, is a member of the Apocynaceae family of plants that has been extensively studied for its production of MIAs, most notably the anti-cancer agents vinblastine and vincristine (Figure 4). *C. roseus* extracts have been used in traditional medicines to treat a range of ailments, including cancers, diabetes and stomach disorders [52]. However, the biologically active alkaloids were serendipitously discovered in 1958 when Noble *et al.* observed a significant decrease in the white blood cell counts of mice following injection of *C. roseus* extracts, suggesting its potential use as a chemotherapy agent [53]. Vinblastine and the structurally related vincristine were isolated and shown to be effective cancer treatments [54, 55] by binding to microtubules and arresting cell mitosis [56], and subsequently approved for clinical use by the FDA [57]. Despite their low abundance *in planta* (0.003% based on dried plant material) [58], the small doses required for the pharmacological application of vinblastine and

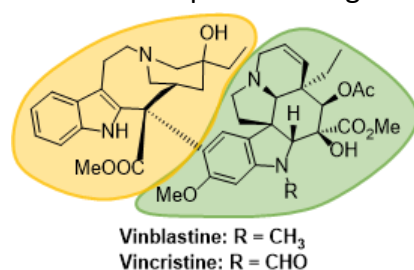


Figure 4. Structures of MIAs vinblastine and vincristine produced in *C. roseus*. Monomeric precursors catharanthine and vindoline coloured in yellow and green respectively.

vincristine enabled their clinical supply by isolation from plant material [53, 59]. Nevertheless, significant efforts have been made to improve the production of these MIAs using chemical synthesis [60–62] and synthetic biology [45–47] approaches.

Over the past 40 years, the biosynthesis of MIAs has been extensively studied in *C. roseus*, revealing a highly intricate and regulated process involving over 30 enzymatic steps. Biosynthesis occurs in 3 specialised cell types within the plant, with iridoid biosynthesis taking place in internal phloem-associated parenchyma (IPAP) cells, formation of corynanthe-, iboga- and aspidosperma-type alkaloids in epidermal cells, and vindoline biosynthesis and formation of the bisindole scaffold occurring in specialised idioblast cells [34, 63]. Vinblastine biosynthesis can broadly be split into four stages: formation of the corynanthe-type scaffold, formation of aspidosperma- and iboga-type scaffolds, elaboration of the aspidosperma-type scaffold to form vindoline, and the formation of the bisindole scaffold. The discovery and elucidation of vinblastine biosynthesis in *C. roseus* is discussed in detail below.

1.3.1. Corynanthe Scaffold Biosynthesis

The first committed step of MIA biosynthesis is the formation of the central precursor molecule strictosidine from the nitrogen-containing tryptamine and the monoterpene secologanin (Figure 5). Tryptamine is formed by the tryptophan decarboxylase (TDC) catalysed decarboxylation of the amino acid tryptophan [64] whilst secologanin originates from the MEP pathway [34, 65]. The Pictet-Spenglerase enzyme strictosidine synthase (STR) catalyses the condensation of tryptamine and secologanin to form strictosidine [18, 20, 21, 23, 66]. STR is localised within the cell vacuole [23, 67], necessitating the import of substrates into this sub-cellular compartment and the export of strictosidine to the cytosol. Whilst the tryptamine importer remains unknown, secologanin is imported into the vacuole by a multidrug and toxic compound efflux (MATE) transporter [34], and strictosidine is exported by a nitrate/peptide family transporter (NPF2.9) [68].

The glucose moiety of strictosidine is hydrolysed by strictosidine deglycosidase (SGD) in the cell nucleus to form the highly reactive aglycone [69–71]. Although a transporter to mediate the import of strictosidine into the nucleus has not been reported, examples of eukaryotic

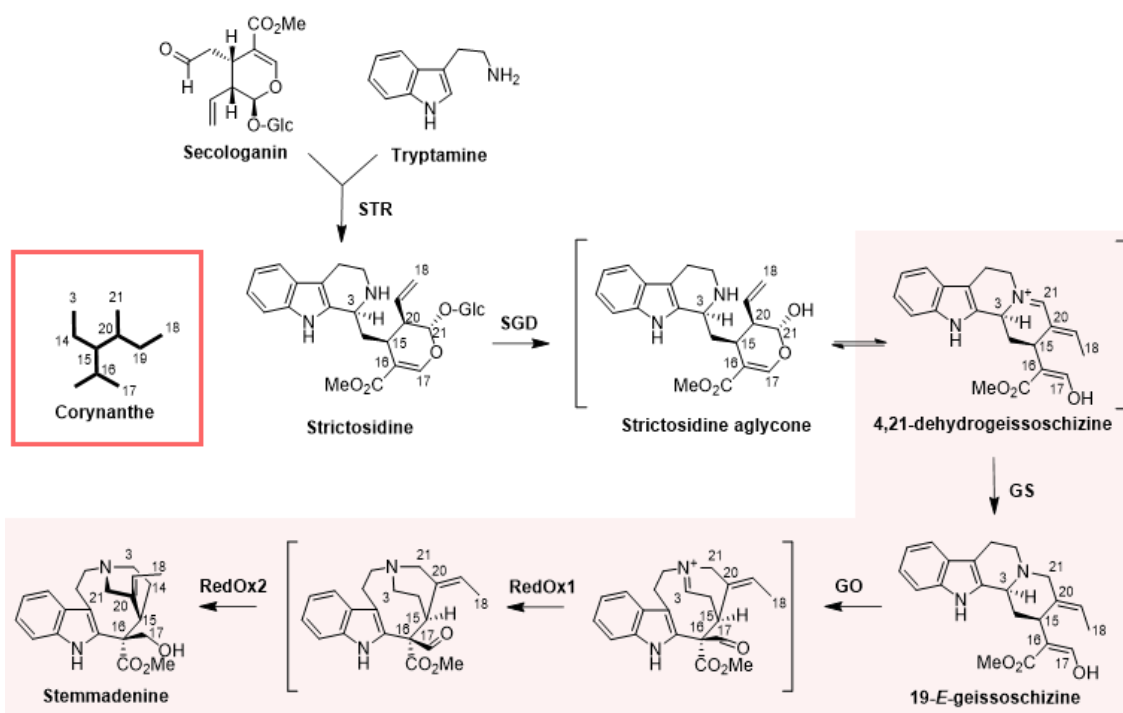


Figure 5. Biosynthesis of corynanthe MIA scaffold in *C. roseus* from the monoterpene secologanin and indole-derived tryptamine precursors.

nuclear pore complexes facilitating the movement of defence-related compounds suggest that the MIA biosynthetic intermediates could employ a similar mechanism [72, 73]. Strictosidine aglycone, which exists as a variety of structural isomers, can be reduced by a short-chain dehydrogenase/reductase (SDR) enzyme [74] or a variety of members of the cinnamyl alcohol dehydrogenase (CAD) subfamily of alcohol dehydrogenase (ADH) enzymes [75–77] to form different MIA scaffolds. Reduction of the strictosidine aglycone rearrangement 4,21-dehydrogeissoschizine by the CAD enzyme geissoschizine synthase (GS) produces the corynanthe-type alkaloid 19-*E*-geissoschizine [78]. This is subsequently oxidised by the cytochrome P450 (CYP) geissoschizine oxidase (GO) to form the highly unstable intermediate preakuammicine [17], although only the deformed product akuammicine was observed in *in vitro* assays [78, 79]. However, incubation of strictosidine with enzymes SGD, GO, an additional CAD named RedOx1 and an aldo-keto reductase (AKR) enzyme named RedOx2 resulted in the production of the known biosynthetic intermediate stemmadenine [79]. As initially proposed in early chemical hypotheses [17], preakuammicine spontaneously rearranges to an unstable iminium, which is reduced by RedOx1 to form a stable aldehyde intermediate. The aldehyde is then reduced by RedOx2 to generate stemmadenine, the precursor of aspidosperma- and iboga-type MIAs [80].

Orthologues of many of the enzymes involved in the formation of strictosidine and corynanthe-type MIAs have been reported in other species including *Strychnos nux-vomica* [43] and *Rauwolfia serpentina* [22, 81, 82], highlighting the shared biosynthetic pathways of these compounds as hypothesised in early chemical models [17].

1.3.2. Aspidosperma and Iboga Scaffold Biosynthesis

Further diversity of MIAs is achieved by the formation of the iboga- and aspidosperma-type scaffolds (Figure 6). The corynanthe-type MIA stemmadenine is acetylated by stemmadenine acetyltransferase (SAT) [79]. The resulting stemmadenine acetate is then oxidised by the berberine bridge-like enzyme (BBE) precondylocarpine acetate synthase (PAS) to form precondylocarpine acetate [83]. PAS likely catalyses stemmadenine acetate oxidation using a similar mechanism to previously characterised flavin adenine dinucleotide (FAD) dependent oxidases [84].

The reduction of precondylocarpine acetate catalysed by the CAD dihydroprecondylocarpine acetate synthase (DPAS) is discussed in Chapter 2 of this thesis. In brief, DPAS catalyses a 1,4-iminium reduction to form dihydroprecondylocarpine acetate, which following a spontaneous desacetoxylation, forms the highly reactive intermediate dehydrosecodine [83]. Early chemical hypotheses proposed dehydrosecodine as the key intermediate to aspidosperma- and iboga-type MIA scaffold formation [85–87].

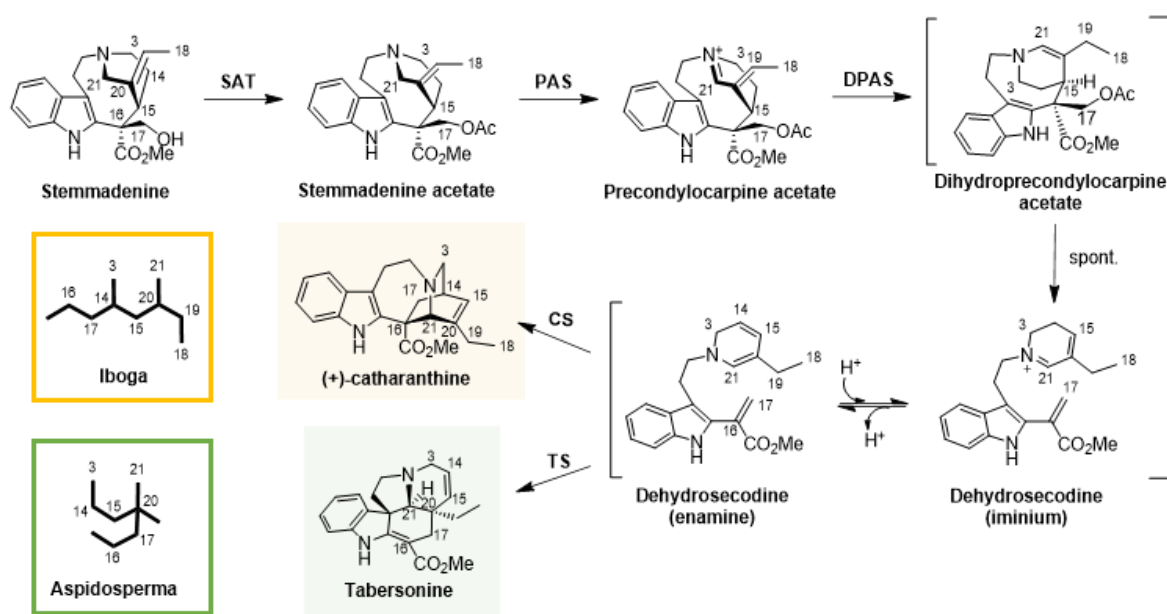


Figure 6. Biosynthesis of iboga and aspidosperma MIA scaffolds in *C. roseus*.

This was validated *in vivo* upon the discovery of the cyclase enzymes tabersonine synthase (TS) and catharanthine synthase (CS) [83, 88]. Incubation of the substrate precondylocarpine acetate with DPAS and either cyclase enzyme TS or CS led to the formation of tabersonine or the bisindole precursor catharanthine respectively (Figure 4). These enzymes were all found to be co-localised in the nucleocytoplasm, suggesting the formation of protein-protein complexes as discussed further in Chapter 4 of this thesis [83].

Since their initial discovery in *C. roseus*, orthologues of PAS, DPAS and the cyclase enzymes have also been reported in the closely related species *Tabernanthe iboga* [44, 89], revealing the shared biosynthetic pathway of aspidosperma- and iboga-type MIAs.

1.3.3. Vindoline Biosynthesis: Elaborating the Aspidosperma Scaffold

Tabersonine undergoes 7 further enzymatic steps to form the remaining precursor of bisindole MIAs - vindoline (Figure 7). Firstly, this compound is oxidised by the CYP enzyme tabersonine hydroxylase (T16H) - at C11 according to the carbon numbering used in this thesis - to form 11-hydroxytabersonine [90, 91]. Two isoforms of T16H have been reported which display different organ-dependent expression patterns suggesting divergence of their roles *in planta* [92]. The product of T16H is methylated by an *O*-methyltransferase (16OMT) to form 11-methoxytabersonine [93, 94]. A second isoform of 16OMT was identified to act on flavonoid substrates in addition to 11-hydroxytabersonine, suggesting a potential ancestral role of this enzyme in flavonoid biosynthesis [94]. The CYP enzyme tabersonine 3'-oxidase (T3O) forms the unstable 16-hydroxy-11-methoxy-tabersonine [95], which when incubated in the presence of the CAD enzyme tabersonine-3-reductase (T3R) forms the stable 16-hydroxy-11-methoxy-1,2-dihydroxytabersonine [95]. This product is subsequently methylated by an *N*-methyl transferase (NMT) to form deacetoxyvindoline [96]. NMT was localised to the thylakoids within the chloroplast [97, 98], though no transporter proteins have been identified which facilitate the movement of MIAs to this organelle.

The 2-oxoglutarate dependent-dioxygenase enzyme deacetoxyvindoline 4'-hydroxylase (D4H) catalyses the formation of deacetylvindoline [99-101], which is subsequently catalysed by deacetylvindoline 4-O-acetyl-transferase (DAT) to form vindoline [102, 103]. These penultimate steps of vindoline biosynthesis were found to be localised in the

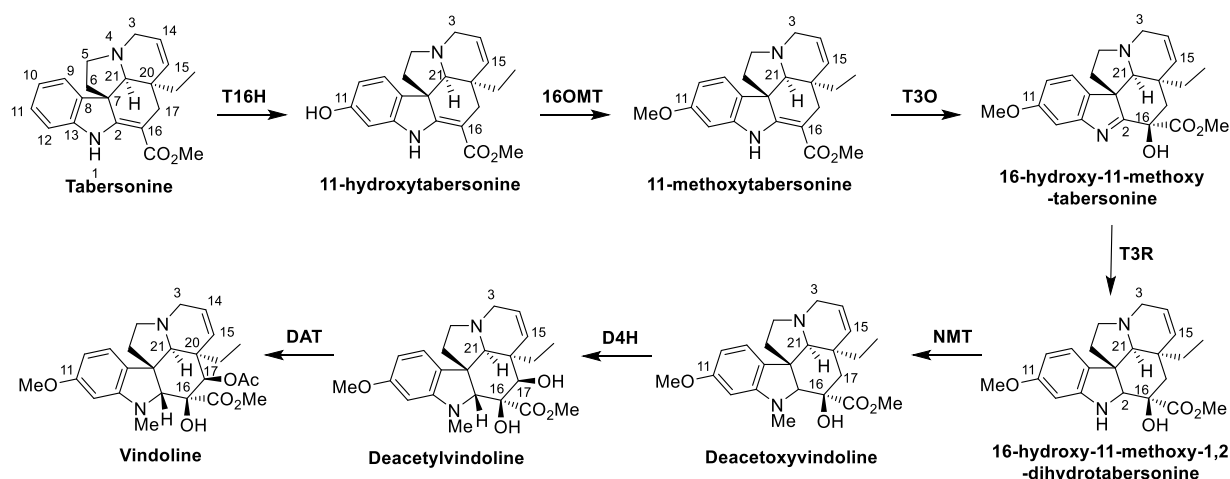


Figure 7. Vindoline biosynthesis from aspidosperma-type alkaloid tabersonine in *C. roseus*.

nucleocytoplasmic compartment within the cell^[98] and enriched in specialised idioblast cell types^[34]. This suggests the action of unknown transporter proteins to exchange deacetoxyvindoline between epidermal and idioblast cells. Elucidation of vindoline biosynthesis in conjunction with catharanthine biosynthesis thus enables access to the proposed precursors of bisindole MIAs.

1.3.4. Vinblastine Biosynthesis: Dimerisation and Decoration of the Bisindole Scaffold

The final steps of vinblastine biosynthesis require the dimerisation of the precursor's catharanthine and vindoline (Figure 8). These precursors were identified by *in vivo* isotopic label feeding experiments^[104], though the biosynthetic steps remain unknown. Catharanthine and vindoline are sequestered in the vacuole of idioblast cells through a proton antiport system^[105, 106]. Bis-indole MIA biosynthesis is then proposed to occur through the peroxidase-catalysed oxidation of catharanthine to form catharanthine iminium. This compound then spontaneously couples with vindoline to form the bisindole anhydrovinblastine iminium^[58, 107]. This oxidation was reported to be catalysed by the *C. roseus* peroxidase PRX1^[105, 108, 109], however, was also catalysed by the commercial enzyme horseradish peroxidase, suggesting that it is a nonselective reaction^[58, 110].

Anhydrovinblastine iminium is proposed to act as a central intermediate of bisindole MIAs observed in *C. roseus* including vinblastine^[111]. Isoforms of the CAD enzyme tetrahydroalstonine (THAS) catalyse the 1,2-iminium reduction of anhydrovinblastine iminium to form anhydrovinblastine^[34]. Vinblastine biosynthesis is proposed to go through

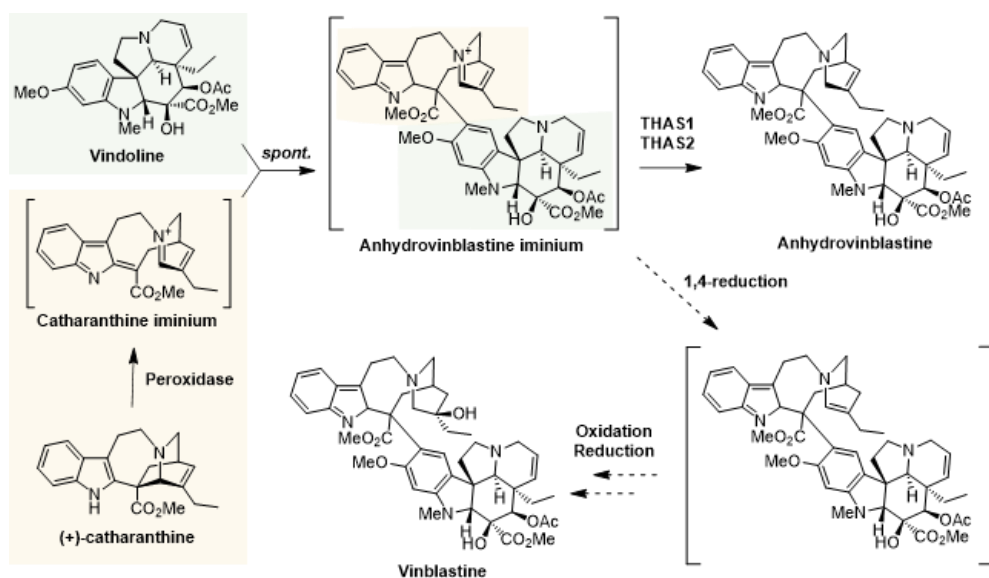


Figure 8. Coupling of catharanthine and vindoline to form bisindole MIAs in *C. roseus*. Catharanthine and vindoline precursors coloured in yellow and green respectively. Formation of vinblastine based on biochemical hypothesis by Langlois and Potier ^[111].

a 1,4-reduction of anhydrovinblastine iminium, reminiscent of the DPAS-catalysed reduction of dihydroprecondylocarpine acetate ^[83], before a final oxidation and reduction step (Figure 8). Members of the CAD family of enzymes have neofunctionalised to perform atypical reductions in MIA biosynthesis as discussed in Chapters 2 and 3 of this thesis. The final reductions to produce vinblastine may therefore be catalysed by one or more CAD enzymes in *C. roseus*.

Vinblastine biosynthesis is one of the longest and most complex PNP pathways studied to date and its elucidation is a culmination of over 40 years of study. Vinblastine biosynthesis in *C. roseus* not only acts as a model system for discovering MIA pathways in other species, but demonstrates the biochemical ingenuity of plants and the diversity of PNPs.

1.4. Alcohol Dehydrogenase Enzymes: Overview and Evolutionary Origin

ADHs are a large family of medium chain reductase (MDR) enzymes that catalyse the reversible oxidation of primary or secondary alcohols to their corresponding aldehyde or ketone using the cofactor nicotinamide adenine dinucleotide phosphate (hydride) (NADP(H)) ^[112–115]. ADHs are found in all kingdoms of life, though have phylogenetically and functionally diverged to act in a range of primary and specialised metabolic pathways ^[116],

¹¹⁷. They are proposed to have evolved from an ancestral SDR protein (Figure 9) ^[118, 119], which are structurally characterised by a single domain and a Rossmann-fold motif important for cofactor binding ^[120]. The merging of a SDR and a common operon GroE Small (GroES) chaperonin protein led to the formation of the MDR superfamily of enzymes ^[121]. MDRs later incorporated a variety of metal ions to form metalloenzyme MDRs, including the zinc-containing ADH family of enzymes.

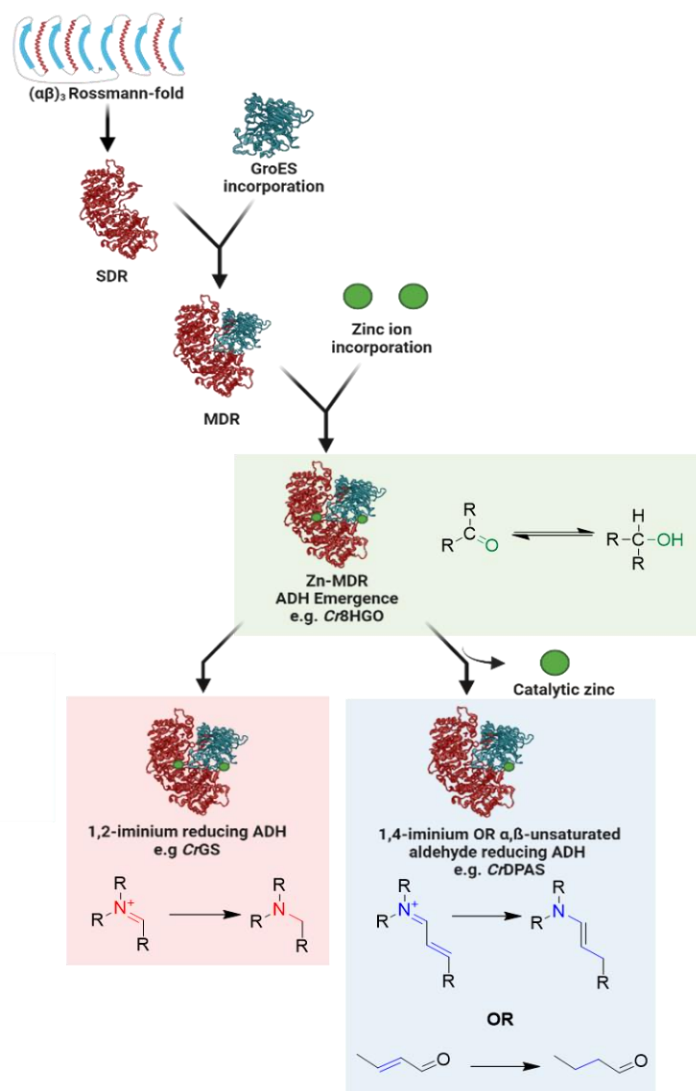


Figure 9. Emergence and evolution of ADH enzyme family. SDR superfamily emerged from dinucleotide binding Rossmann-fold which merged with GroES to form MDRs. Subsequent addition of two zinc ions formed the ADH family. ADHs further diverged in MIA biosynthesis to perform the 1,2-reduction of an iminium moiety (e.g. *CrGS*), or the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde (e.g. *CrDPAS*). Figure adapted from Jörnvall *et al.*, ^[150] and made using BioRender.

Members of the CAD subfamily of ADHs catalyse numerous reductions in MIA biosynthesis as outlined in section 1.3. Many CADs involved in MIA biosynthesis have functionally diverged from their typical catalytic activity of reducing a primary alcohol to an aldehyde to catalysing highly atypical reactions such as the 1,2- and 1,4- reduction of an iminium moiety (e.g. *CrGS* and *CrDPAS* respectively; Figure 9). Understanding the mechanism of these atypical CAD-catalysed reductions is therefore of biochemical and bioengineering interest and is the focus of Chapter 2 of this thesis. Furthermore, the expansion and neofunctionalisation of this enzyme family within Gentianales, as discussed in Chapter 3, sheds light on the evolution of MIA chemodiversity.

1.5. Protein-Protein Interactions in Plant-Specialised Metabolism

Plants produce over 350,000 specialised metabolites, with many being synthesised in parallel pathways utilising a common intermediate ^[122, 123]. Since seminal works in the 1960s ^[124, 125], the cell is now understood to be a metabolically complex and crowded space with protein-protein interactions mediating numerous biological processes. The dynamic assembly of protein complexes in plants is proposed to enable these sessile organisms to react to biotic and abiotic stresses in their environment. Complexes are formed by inter-chain non-covalent interactions between proteins and can be classified by their composition, structural and/or functional stability, and temporal stability (Figure 10) ^[126, 127]. Protein-protein complexes have a variety of biological functions in plants including co-

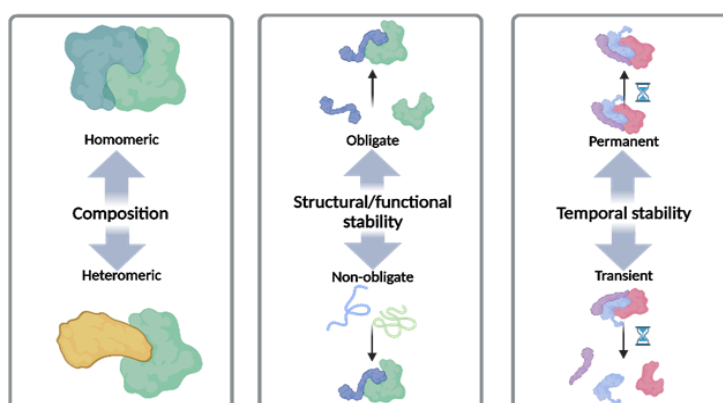


Figure 10. Classifications of protein-protein complexes. Complexes are classified by their composition (i.e. homo- (identical) or heteromeric (non-identical) protein chains), structural and/or functional stability as obligate (stable) or non-obligate (unstable) monomers, and temporal stability (i.e. permanent or transient). Figure made using Biorender.

localising organelles ^[128], regulating gene transcription ^[129, 130], and scaffolding biosynthetic enzymes ^[131, 132]. Furthermore, these complexes have been observed to form between proteins acting within different metabolic pathways to orchestrate defence responses ^[133].

Some enzymes physically interact to form supramolecular structures known as metabolons. These comprise two or more sequential enzymes in a biosynthetic pathway interacting to directly channel intermediates between their active sites. Metabolons can aid metabolic organisation by controlling flux at network branch points, and in some cases, improve metabolic efficiency ^[134]. Examples of metabolons have been found in highly conserved metabolic pathways across the kingdoms of life to produce compounds essential for cellular respiration such as purine ^[135] and ubiquinone biosynthesis ^[136].

The high number of enzymes and the generation of unstable intermediates in PNP biosynthetic pathways has led to speculation of the formation of protein complexes and/or metabolons. However, experimental and technical limitations including the limited access to often unstable intermediates make metabolons, particularly in PNP biosynthesis, challenging to study. Metabolons in plants have been most extensively studied in primary metabolism such as the tricarboxylic acid cycle, glycolysis and lignin biosynthesis ^[137–139]. Although less studied, there are metabolons and enzyme-enzyme complexes have been reported in a range of PNP pathways such as the biosynthesis of isoflavonoids ^[140], cyanogenic glucosides ^[141], iridoids ^[142], sporopollenin ^[143], camalexin ^[144], bitter acids ^[145] and flavonoids ^[146–148]. Identifying and characterising the protein-protein interactions within a biosynthetic pathway can enlighten our understanding of the metabolic organisation within a cell and the observed chemodiversity of PNPs.

As described in section 1.3, vinblastine biosynthesis is a long PNP pathway that contains several intermediates that act as metabolic branch points to generate chemical diversity (e.g. dehydrosecodine to form either aspidosperma- or iboga-type MIAs, Figure 6). In addition, several steps generate unstable intermediates that require the activity of further enzymes to generate a more stable compound (e.g. the formation of stemmadenine, Figure 5). Due to these observations, vinblastine biosynthesis is speculated to be facilitated or aided by the formation of protein-protein complexes and/or metabolons. Chapter 4 of this thesis explores the extent of protein-protein complexes in MIA biosynthesis and uses structural biology to elucidate the basis of an enzyme-enzyme interaction.

1.6. Scope of this Thesis

Great progress over the last 40 years have revealed the complex biosynthetic pathways of many pharmaceutically important MIAs including the anti-cancer agent vinblastine. This work has discovered that the CAD subfamily of ADH enzymes catalyse atypical reactions in MIA biosynthesis - namely, the 1,2- and the 1,4-reduction of an iminium moiety instead of the 1,2-reduction of an aldehyde. In Chapter 2 of this thesis, the activity of the MIA enzyme *CrDPAS* is confirmed as a bonafide 1,4-iminium reductase. Structural elucidation of DPAS orthologues from *C. roseus* and *T. iboga* reveal the highly unusual loss of residues involved in coordinating the catalytic zinc ion. In conjunction with mutational studies, we propose a mechanism for the *CrDPAS*-catalysed 1,4-iminium reduction of dihydroprecondylocarpine acetate, as well as the *CrGS*-catalysed 1,2-iminium reduction of 4,21-dehydrogeissoschizine to expand the catalytic repertoire of this enzyme family.

Building on the identification of key sequence motifs in CADs that catalyse atypical reductions, Chapter 3 of this thesis explores the expansion and neofunctionalisation of this enzyme family in MIA-producing species in Gentianales. Comparative genomics and phylogenetic analysis are used to propose the emergence of atypical CADs as requisite drivers in the evolution of MIA chemodiversity.

Protein-protein complexes have been proposed to facilitate the metabolic organisation of complex PNP pathways such as MIA biosynthesis. In Chapter 4 of this thesis, the pairwise interactions of 17 MIA enzymes from *C. roseus* are mapped using a split-luciferase assay to better understand the extent of protein-protein interactions in PNP biosynthesis. Furthermore, we use structural biology to elucidate the basis of interaction between the CAD *CrDPAS* and the downstream cyclase enzymes. Additionally, the interactions of *CrDPAS* are elucidated using *in vivo* proximity tagging, revealing inter-pathway interactions between MIA and phenylpropanoid biosynthetic enzymes.

The biochemical characterisation of CADs that catalyse atypical reductions expands the chemical toolbox of this enzyme family. Meanwhile, the protein-protein interactions of MIA biosynthetic enzymes can help us understand the organisation of PNP metabolism within a cell. The work in this thesis therefore explores how the neofunctionalisation of an enzyme family and how interactions between biosynthetic enzymes contribute to the biosynthesis and chemodiversity of this group of PNPs.

1.7. References

- [1] F. Ntie-Kang, D. Svozil, *Phys. Sci. Rev.* 2020, 5, 20180121.
- [2] L. Guo, H. Yao, W. Chen, X. Wang, P. Ye, Z. Xu, S. Zhang, H. Wu, *Hortic. Res.* 2022, 9, uhac223.
- [3] P. Wangchuk, *J. Biol. Act. Prod. Nat.* 2018, 8, 1–20.
- [4] H. Yuan, Q. Ma, L. Ye, G. Piao, *Molecules* 2016, 21, 559.
- [5] N. Kumar, N. Goel, *Biotechnol. Rep.* 2019, 24, e00370.
- [6] K. Li, S. Ji, W. Song, Y. Kuang, Y. Lin, S. Tang, Z. Cui, X. Qiao, S. Yu, M. Ye, *J. Nat. Prod.* 2017, 80, 334–346.
- [7] M. E. Bergman, B. Davis, M. A. Phillips, *Molecules* 2019, 24, 3961.
- [8] J. Gershenzon, N. Dudareva, *Nat. Chem. Biol.* 2007, 3, 408–414.
- [9] R. Jaeger, E. Cuny, *Nat. Prod. Commun.* 2016, 11, 9.
- [10] B. R. Lichman, *Nat Prod Rep* 2020, 38, 103–129.
- [11] P. Dey, A. Kundu, A. Kumar, M. Gupta, B. M. Lee, T. Bhakta, S. Dash, H. S. Kim, *Recent Adv. Nat. Prod. Anal.* 2020, 505–567.
- [12] B. Debnath, W. S. Singh, M. Das, S. Goswami, M. K. Singh, D. Maiti, K. Manna, *Mater. Today Chem.* 2018, 9, 56–72.
- [13] M. Heinrich, J. Mah, V. Amirkia, *Molecules* 2021, 26, 1836.
- [14] Q. Pan, N. R. Mustafa, K. Tang, Y. H. Choi, R. Verpoorte, *Phytochem. Rev.* 2016, 15, 221–250.
- [15] S. E. O'Connor, J. J. Maresh, *Nat Prod Rep* 2006, 23, 532.
- [16] A. E. Mohammed, Z. H. Abdul-Hameed, M. O. Alotaibi, N. O. Bawakid, T. R. Sobahi, A. Abdel-Lateff, W. M. Alarif, *Molecules* 2021, 26, 488.
- [17] A. I. Scott, *Accounts Chem Res* 1970, 3, 151–157.
- [18] J. Stöckigt, M. H. Zenk, *J. Chem. Soc., Chem. Commun.* 1977, 646–648.
- [19] A. R. Battersby, N. G. Lewis, J. M. Tippet, *Tetrahedron Lett.* 1978, 19, 4849–4852.
- [20] R. T. Brown, J. Leonard, S. K. Sleight, *Phytochemistry* 1978, 17, 899–900.

- [21] J. F. Treimer, M. H. Zenk, *Eur. J. Biochem.* 1979, *101*, 225–233.
- [22] T. M. Kutchan, N. Hampf, F. Lottspeich, K. Beyreuther, M. H. Zenk, *FEBS Lett.* 1988, *237*, 40–44.
- [23] T. D. McKnight, C. A. Roessner, R. Devagupta, A. I. Scott, C. L. Nessler, *Nucleic Acids Res.* 1990, *18*, 4939–4939.
- [24] T. M. Kutchan, *Phytochemistry* 1993, *32*, 493–506.
- [25] B. D. Hilton, G. E. Martin, *J. Nat. Prod.* 2010, *73*, 1465–1469.
- [26] J.L. Wolfender, J.-M. Nuzillard, J. J. J. van der Hoof, J.-H. Renault, S. Bertrand, *Anal. Chem.* 2019, *91*, 704–742.
- [27] J. H. Leebens-Mack, M. S. Barker, E. J. Carpenter, M. K. Deyholos, M. A. Gitzendanner, S. W. Graham, I. Grosse, Z. Li, M. Melkonian, S. Mirarab, M. Porsch, M. Quint, S. A. Rensing, D. E. Soltis, P. S. Soltis, D. W. Stevenson, K. K. Ullrich, N. J. Wickett, L. DeGironimo, P. P. Edger, I. E. Jordon-Thaden, S. Joya, T. Liu, B. Melkonian, N. W. Miles, L. Pokorny, C. Quigley, P. Thomas, J. C. Villarreal, M. M. Augustin, M. D. Barrett, R. S. Baucom, D. J. Beerling, R. M. Benstein, E. Biffin, S. F. Brockington, D. O. Burge, J. N. Burris, K. P. Burris, V. Burtet-Sarramegna, A. L. Caicedo, S. B. Cannon, Z. Çebi, Y. Chang, C. Chater, J. M. Cheeseman, T. Chen, N. D. Clarke, H. Clayton, S. Covshoff, B. J. Crandall-Stotler, H. Cross, C. W. dePamphilis, J. P. Der, R. Determann, R. C. Dickson, V. S. D. Stilio, S. Ellis, E. Fast, N. Feja, K. J. Field, D. A. Filatov, P. M. Finnegan, S. K. Floyd, B. Fogliani, N. García, G. Gâteblé, G. T. Godden, F. (Qi Y. Goh, S. Greiner, A. Harkess, J. M. Heaney, K. E. Helliwell, K. Heyduk, J. M. Hibberd, R. G. J. Hodel, P. M. Hollingsworth, M. T. J. Johnson, R. Jost, B. Joyce, M. V. Kapralov, E. Kazamia, E. A. Kellogg, M. A. Koch, M. V. Konrat, K. Könyves, T. M. Kutchan, V. Lam, A. Larsson, A. R. Leitch, R. Lentz, F.-W. Li, A. J. Lowe, M. Ludwig, P. S. Manos, E. Mavrodiev, M. K. McCormick, M. McKain, T. McLellan, J. R. McNeal, R. E. Miller, M. N. Nelson, Y. Peng, P. Ralph, D. Real, C. W. Riggins, M. Ruhsam, R. F. Sage, A. K. Sakai, M. Scascitella, E. E. Schilling, E.-M. Schlösser, H. Sederoff, S. Servick, E. B. Sessa, A. J. Shaw, S. W. Shaw, E. M. Sigel, C. Skema, A. G. Smith, A. Smithson, C. N. Stewart, J. R. Stinchcombe, P. Szövényi, J. A. Tate, H. Tiebel, D. Trapnell, M. Villegente, C.-N. Wang, S. G. Weller, M. Wenzel, S. Weststrand, J. H. Westwood, D. F. Whigham, S. Wu, A. S. Wulff, Y. Yang, D. Zhu, C. Zhuang, J. Zuidof, M. W. Chase, J. C. Pires, C. J. Rothfels, J. Yu, C. Chen, L. Chen, S. Cheng, J. Li, R. Li, X. Li, H. Lu, Y. Ou, X. Sun, X. Tan, J. Tang, Z. Tian, F. Wang, J. Wang, X. Wei, X. Xu, Z. Yan, F. Yang, X. Zhong, F. Zhou, Y. Zhu, Y. Zhang, S. Ayyampalayam, T. J. Barkman, N. Nguyen, N. Matasci, D. R. Nelson, E. Sayyari, E. K. Wafula, R. L. Walls, T. Warnow, H. An, N. Arrigo, A. E. Baniaga, S. Galuska, S. A. Jorgensen, T. I. Kidder, H. Kong, P. Lu-Irving, H. E. Marx, X. Qi, C. R. Reardon, B. L. Sutherland, G. P. Tiley, S. R. Welles, R. Yu, S. Zhan, L. Gramzow, G. Theißen, G. K.-S. Wong, *Nature* 2019, *574*, 679–685.
- [28] X. Su, L. Yang, D. Wang, Z. Shu, Y. Yang, S. Chen, C. Song, *Hortic. Res.* 2022, *9*, uhac075.
- [29] F. Kellner, J. Kim, B. J. Clavijo, J. P. Hamilton, K. L. Childs, B. Vaillancourt, J. Cepela, M. Habermann, B. Steuernagel, L. Clissold, K. Mclay, C. R. Buell, S. E. O'Connor, *Plant J* 2015, *82*, 680–692.

- [30] D. Zhao, J. P. Hamilton, G. M. Pham, E. Crisovan, K. Wiegert-Rininger, B. Vaillancourt, D. DellaPenna, C. R. Buell, *Gigascience* 2017, *6*, 1–7.
- [31] J. Franke, J. Kim, J. P. Hamilton, D. Zhao, G. M. Pham, K. Wiegert-Rininger, E. Crisovan, L. Newton, B. Vaillancourt, E. Tatsis, C. R. Buell, S. E. O'Connor, *Chembiochem* 2019, *20*, 83–87.
- [32] A. Raj, H. Hirakawa, R. Nakabayashi, S. Kikuchi, K. Hayashi, M. Rai, H. Tsugawa, T. Nakaya, T. Mori, H. Nagasaki, R. Fukushi, Y. Kusuya, H. Takahashi, H. Uchiyama, A. Toyoda, S. Hikosaka, E. Goto, K. Saito, M. Yamazaki, *Nat Commun* 2021, *12*, 405.
- [33] E. A. Stander, C. Cuello, C. Birer-Williams, N. Kulagina, H. J. Jansen, I. Carqueijeiro, L.-V. Méteignier, V. Vergès, A. Oudin, N. Papon, R. P. Dirks, M. K. Jensen, S. E. O'Connor, T. D. de Bernonville, S. Besseau, V. Courdavault, *G3: GenesGenomesGenet.* 2022, *12*, jkac268.
- [34] C. Li, J. C. Wood, A. H. Vu, J. P. Hamilton, C. E. R. Lopez, R. M. E. Payne, D. A. S. Guerrero, K. Gase, K. Yamamoto, B. Vaillancourt, L. Caputi, S. E. O'Connor, C. R. Buell, *Nat Chem Biol* 2023, 1–11.
- [35] W.-B. Jiao, K. Schneeberger, *Curr. Opin. Plant Biol.* 2017, *36*, 64–70.
- [36] S. K. Singh, B. Patra, P. Paul, Y. Liu, S. Pattanaik, L. Yuan, *Plant Sci* 2020, *293*, 110408.
- [37] J. Liang, T. An, J.-X. Zhu, S. Chen, J.-H. Zhu, R. J. Peters, R. Yu, J. Zi, *J Nat Prod* 2021, *84*, 2709–2716.
- [38] M. M. Alami, Z. Ouyang, Y. Zhang, S. Shu, G. Yang, Z. Mei, X. Wang, *Int. J. Mol. Sci.* 2022, *23*, 15932.
- [39] C. Engler, S. Marillonnet, *Methods Mol. Biol.* 2013, *1116*, 119–131.
- [40] I. Carqueijeiro, C. Langley, D. Grzech, K. Koudounas, N. Papon, S. E. O'Connor, V. Courdavault, *Curr Opin Biotech* 2020, *65*, 17–24.
- [41] K. Eljounaidi, B. R. Lichman, *Front. Chem.* 2020, *8*, 596479.
- [42] K. Yamamoto, D. Grzech, K. Koudounas, E. A. Stander, L. Caputi, T. Mimura, V. Courdavault, S. E. O'Connor, *Plant Physiol* 2021, *187*, 846–857.
- [43] B. Hong, D. Grzech, L. Caputi, P. Sonawane, C. E. R. López, M. O. Kamileen, N. J. H. Lozada, V. Grabe, S. E. O'Connor, *Nature* 2022, 1–6.
- [44] S. C. Farrow, M. O. Kamileen, L. Caputi, K. Bussey, J. E. A. Mundy, R. C. McAtee, C. R. J. Stephenson, S. E. O'Connor, *J Am Chem Soc* 2019, *141*, 12979–12983.
- [45] S. Brown, M. Clastre, V. Courdavault, S. E. O'Connor, *Proc. Natl. Acad. Sci.* 2015, *112*, 3205–3210.

- [46] J. Zhang, L. G. Hansen, O. Gudich, K. Viehrig, L. M. M. Lassen, L. Schrübbers, K. B. Adhikari, P. Rubaszka, E. Carrasquer-Alvarez, L. Chen, V. D'Ambrosio, B. Lehka, A. K. Haidar, S. Nallapareddy, K. Giannakou, M. Laloux, D. Arsovska, M. A. K. Jørgensen, L. J. G. Chan, M. Kristensen, H. B. Christensen, S. Sudarsan, E. A. Stander, E. Baidoo, C. J. Petzold, T. Wulff, S. E. O'Connor, V. Courdavault, M. K. Jensen, J. D. Keasling, *Nature* 2022, *609*, 341–347.
- [47] D. Grzech, B. Hong, L. Caputi, P. D. Sonawane, S. E. O'Connor, *ACS Synth. Biol.* 2023, *12*, 27–34.
- [48] W. Runguphan, J. J. Maresh, S. E. O'Connor, *Proc National Acad Sci* 2009, *106*, 13673–13678.
- [49] E. K. Leggans, K. K. Duncan, T. J. Barker, K. D. Schleicher, D. L. Boger, *J Med Chem* 2013, *56*, 628–639.
- [50] H. Gotoh, K. K. Duncan, W. M. Robertson, D. L. Boger, *Acs Med Chem Lett* 2011, *2*, 948–952.
- [51] A. A. Lopes, B. Chioca, B. Musquiari, E. J. Crevelin, S. de C. França, M. F. das G. F. da Silva, A. M. S. Pereira, *Sci Rep-uk* 2019, *9*, 11349.
- [52] S. Kumar, B. Singh, R. Singh, *J. Ethnopharmacol.* 2022, *284*, 114647.
- [53] R. L. Noble, C. T. Beer, J. H. Cutts, *Ann Ny Acad Sci* 1958, *76*, 882–894.
- [54] I. S. Johnson, H. F. Wright, G. H. Svoboda, J. Vlantis, *Cancer Res.* 1960, *20*, 1016–22.
- [55] O. H. Warwick, J. M. Darte, T. C. Brown, C. T. Beer, J. H. Cutts, R. L. Noble, *Cancer Res.* 1960, *20*, 1032–40.
- [56] M. A. Jordan, R. H. Himes, L. Wilson, *Cancer Res.* 1985, *45*, 2741–7.
- [57] P. Dhyan, C. Quispe, E. Sharma, A. Bahukhandi, P. Sati, D. C. Attri, A. Szopa, J. Sharifi-Rad, A. O. Docea, I. Mardare, D. Calina, W. C. Cho, *Cancer Cell Int.* 2022, *22*, 206.
- [58] J. P. Kutney, L. S. L. Choi, T. Honda, N. G. Lewis, T. Sato, K. L. Stuart, B. R. Worth, *Helvetica Chim. Acta* 1982, *65*, 2088–2101.
- [59] R. A. Rahim, N. H. Ahmad, K. M. A. Azzam, I. Mat, *Adv. Pharm. Bull.* 2017, *8*, 157–161.
- [60] S. Yokoshima, T. Ueda, S. Kobayashi, A. Sato, T. Kuboyama, H. Tokuyama, T. Fukuyama, *J Am Chem Soc* 2002, *124*, 2137–2139.
- [61] T. Kuboyama, S. Yokoshima, H. Tokuyama, T. Fukuyama, *P Natl Acad Sci Usa* 2004, *101*, 11966–11970.
- [62] H. Ishikawa, D. A. Colby, S. Seto, P. Va, A. Tam, H. Kakei, T. J. Rayl, I. Hwang, D. L. Boger, *J Am Chem Soc* 2009, *131*, 4904–4916.

- [63] K. Yamamoto, K. Takahashi, L. Caputi, H. Mizuno, C. E. Rodriguez-Lopez, T. Iwasaki, K. Ishizaki, H. Fukaki, M. Ohnishi, M. Yamazaki, T. Masujima, S. E. O'Connor, T. Mimura, *New Phytol* 2019, 224, 848-859.
- [64] V. D. Luca, C. Marineau, N. Brisson, *Proc. Natl. Acad. Sci.* 1989, 86, 2582–2586.
- [65] K. Miettinen, L. Dong, N. Navrot, T. Schneider, V. Burlat, J. Pollier, L. Woittiez, S. V. D. Krol, R. Lukan, T. Ilc, R. Verpoorte, K. M. Oksman-Caldentey, E. Martinoia, H. Bouwmeester, A. Goossens, J. Memelink, D. Werck-Reichhart, *Nat Commun* 2014, 5, 3606.
- [66] A. I. Scott, S. L. Lee, P. de Capite, M. G. Culver, C. R. Hutchinson, *Heterocycles* 1977, 7, 979.
- [67] T. D. McKnight, D. R. Bergey, R. J. Burnett, C. L. Nessler, *Planta* 1991, 185, 148–152.
- [68] R. M. E. Payne, D. Xu, E. Foureau, M. I. S. T. Carqueijeiro, A. Oudin, T. D. de Bernonville, V. Novak, M. Burow, C.-E. Olsen, D. M. Jones, E. C. Tatsis, A. Pendle, B. A. Halkier, F. Geu-Flores, V. Courdavault, H. H. Nour-Eldin, S. E. O'Connor, *Nat Plants* 2017, 3, 16208.
- [69] T. J. C. Luijendijk, L. H. Stevens, R. Verpoorte, *Plant Physiol. Biochem.* 1998, 36, 419–425.
- [70] A. Geerlings, M. M.-L. Ibañez, J. Memelink, R. van der Heijden, R. Verpoorte, *J. Biol. Chem.* 2000, 275, 3051–3056.
- [71] G. Guirimand, V. Courdavault, A. Lanoue, S. Mahroug, A. Guihur, N. Blanc, N. Giglioli-Guivarc'h, B. St-Pierre, V. Burlat, *Bmc Plant Biol* 2010, 10, 182.
- [72] C. E. Zimmerli, M. Allegretti, V. Rantos, S. K. Goetz, A. Obarska-Kosinska, I. Zagoriy, A. Halavatyi, G. Hummer, J. Mahamid, J. Kosinski, M. Beck, *Science* 2021, 374, 1341.
- [73] X. Wu, J. Han, C. Guo, *Int. J. Mol. Sci.* 2022, 23, 3031.
- [74] A. K. Stavrinides, E. C. Tatsis, T.-T. Dang, L. Caputi, C. E. M. Stevenson, D. M. Lawson, B. Schneider, S. E. O'Connor, *Chembiochem* 2018, 19, 940–948.
- [75] A. Stavrinides, E. C. Tatsis, E. Foureau, L. Caputi, F. Kellner, V. Courdavault, S. E. O'Connor, *Chem Biol* 2015, 22, 336–41.
- [76] A. Stavrinides, E. C. Tatsis, L. Caputi, E. Foureau, C. E. M. Stevenson, D. M. Lawson, V. Courdavault, S. E. O'Connor, *Nat Commun* 2016, 7, 12116.
- [77] C. Langley, E. Tatsis, B. Hong, Y. Nakamura, C. Paetz, C. E. M. Stevenson, J. Basquin, D. M. Lawson, L. Caputi, S. E. O'Connor, *Angew. Chem. Int. Ed.* 2022, 61, e202210934.
- [78] E. C. Tatsis, I. Carqueijeiro, T. D. D. Bernonville, J. Franke, T.-T. T. Dang, A. Oudin, A. Lanoue, F. Lafontaine, A. K. Stavrinides, M. Clastre, V. Courdavault, S. E. O'connor, *Nat Commun* 2017, 8, 316.

- [79] Y. Qu, M. E. A. M. Easson, R. Simionescu, J. Hajicek, A. M. K. Thamm, V. Salim, V. D. Luca, *Proc National Acad Sci* 2018, *115*, 3180–3185.
- [80] A. A. Qureshi, A. I. Scott, *Chem. Commun. (Lond.)* 1968, 945–946.
- [81] X. Ma, J. Koepke, G. Fritzsich, R. Diem, T. M. Kutchan, H. Michel, J. Stöckigt, *Biochim. Biophys. Acta (BBA) - Proteins Proteom.* 2004, *1702*, 121–124.
- [82] G. von Schumann, S. Gao, J. Stöckigt, *Bioorgan Med Chem* 2002, *10*, 1913–1918.
- [83] L. Caputi, J. Franke, S. C. Farrow, K. Chung, R. M. E. Payne, T.-D. Nguyen, T.-T. T. Dang, I. S. T. Carqueijeiro, K. Koudounas, T. D. de Bernonville, B. Ameyaw, D. M. Jones, I. J. C. Vieira, V. Courdavault, S. E. O'Connor, *Science* 2018, *360*, 1235–1239.
- [84] B. Daniel, B. Konrad, M. Toplak, M. Lahham, J. Messenlehner, A. Winkler, P. Macheroux, *Arch. Biochem. Biophys.* 2017, *632*, 88–103.
- [85] A. I. Scott, A. A. Qureshi, *Tetrahedron* 1974, *30*, 2993–3002.
- [86] A. I. Scott, C. C. Wei, *Tetrahedron* 1974, *30*, 3003–3011.
- [87] A. I. Scott, P. C. Cherry, C. C. Wei, *Tetrahedron* 1974, *30*, 3013–3019.
- [88] L. Caputi, J. Franke, K. Bussey, S. C. Farrow, I. J. C. Vieira, C. E. M. Stevenson, D. M. Lawson, S. E. O'Connor, *Nat Chem Biol* 2020, *16*, 383–386.
- [89] M. O. Kamileen, M. D. DeMars, B. Hong, Y. Nakamura, C. Paetz, B. R. Lichman, P. D. Sonawane, L. Caputi, S. E. O'Connor, *J Am Chem Soc* 2022, *144*, 19673–19679.
- [90] B. St-Pierre, V. D. Luca, *Plant Physiol* 1995, *109*, 131–139.
- [91] G. Schröder, E. Unterbusch, M. Kaltenbach, J. Schmidt, D. Strack, V. D. Luca, J. Schröder, *Febs Lett* 1999, *458*, 97–102.
- [92] S. Besseau, F. Kellner, A. Lanoue, A. M. K. Thamm, V. Salim, B. Schneider, F. Geu-Flores, R. Höfer, G. Guirimand, A. Guihur, A. Oudin, G. Glevarec, E. Foureau, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, B. St-Pierre, D. Werck-Reichhart, V. Burlat, V. D. Luca, S. E. O'Connor, V. Courdavault, *Plant Physiol* 2013, *163*, 1792–803.
- [93] D. Levac, J. Murata, W. S. Kim, V. D. Luca, *Plant J* 2007, *53*, 225–236.
- [94] P. L. Cruz, I. Carqueijeiro, K. Koudounas, D. P. Bomzan, E. A. Stander, C. Abdallah, N. Kulagina, A. Oudin, A. Lanoue, N. Giglioli-Guivarc'h, D. A. Nagegowda, N. Papon, S. Besseau, M. Clastre, V. Courdavault, *Protoplasma* 2023, *260*, 607–624.
- [95] Y. Qu, M. L. A. E. Easson, J. Froese, R. Simionescu, T. Hudlicky, V. DeLuca, *Proc National Acad Sci* 2015, *112*, 6224–6229.

- [96] V. Deluca, J. Balsevich, R. T. Tyler, W. G. Kurz, *Plant Cell Rep* 1987, 6, 458–61.
- [97] V. D. Luca, A. J. Cutler, *Plant Physiol* 1987, 85, 1099–102.
- [98] G. Guirimand, A. Guihur, P. Poutrain, F. Héricourt, S. Mahroug, B. St-Pierre, V. Burlat, V. Courdavault, *J Plant Physiol* 2011, 168, 549–557.
- [99] E. D. Carolis, F. Chan, J. Balsevich, V. D. Luca, *Plant Physiol*. 1990, 94, 1323–1329.
- [100] E. D. Carolis, V. D. Luca, *J. Biol. Chem.* 1993, 268, 5504–11.
- [101] F. Vazquez-Flota, E. D. Carolis, A. M. Alarco, V. D. Luca, *Plant Mol Biol* 1997, 34, 935–48.
- [102] R. Power, W. G. W. Kurz, V. D. Luca, *Arch. Biochem. Biophys.* 1990, 279, 370–376.
- [103] B. St-Pierre, P. Laflamme, A.-M. Alarco, V. D, e Luca, *Plant J* 1998, 14, 703–713.
- [104] A. I. Scott, F. Gueritte, S. L. Lee, *J. Am. Chem. Soc.* 1978, 100, 6253–6255.
- [105] M. Sottomayor, M. C. Pinto, R. Salema, F. DiCosmo, M. A. Pedreoo, A. R. Barcelo, *Plant Cell Environ* 1996, 19, 761–767.
- [106] I. Carqueijeiro, H. Noronha, P. Duarte, H. Gerós, M. Sottomayor, *Plant Physiol* 2013, 162, 1486–1496.
- [107] T. Endo, A. Goodbody, J. Vukovic, M. Misawa, *Phytochemistry* 1988, 27, 2147–2149.
- [108] M. Sottomayor, M. López-Serrano, F. DiCosmo, A. R. Barceló, *Febs Lett* 1998, 428, 299–303.
- [109] M. Sottomayor, A. R. Barceló, *Protoplasma* 2003, 222, 97–105.
- [110] A. Goodbody, T. Endo, J. Vukovic, J. Kutney, L. Choi, M. Misawa, *Planta Med.* 1988, 54, 136–140.
- [111] N. Langlois, P. Potier, *J. Chem. Soc., Chem. Commun.* 1979, 582–584.
- [112] J. Strommer, *Plant J* 2011, 66, 128–142.
- [113] B. V. Plapp, H. A. Charlier, S. Ramaswamy, *Arch Biochem Biophys* 2016, 591, 35–42.
- [114] B. V. Plapp, B. R. Savarimuthu, D. J. Ferraro, J. K. Rubach, E. N. Brown, S. Ramaswamy, *Biochemistry-us* 2017, 56, 3632–3646.
- [115] H. Jörnvall, T. Bergman, in *Encyclopedia of Metalloproteins*, Springer New York, 2013, 2349–2354.

- [116] C. E. Thompson, C. L. Fernandes, O. N. D. Souza, L. B. D. Freitas, F. M. Salzano, *J Mol Model* 2010, *16*, 919–928.
- [117] B. Persson, J. Hedlund, H. Jörnvall, *Cell Mol Life Sci* 2008, *65*, 3879–3894.
- [118] B. Persson, Y. Kallberg, J. E. Bray, E. Bruford, S. L. Dellaporta, A. D. Favia, R. G. Duarte, H. Jörnvall, K. L. Kavanagh, N. Kedishvili, M. Kisiela, E. Maser, R. Mindnich, S. Orchard, T. M. Penning, J. M. Thornton, J. Adamski, U. Oppermann, *Chem-biol Interact* 2009, *178*, 94–98.
- [119] H. Moummou, Y. Kallberg, L. B. Tonfack, B. Persson, B. van der Rest, *Bmc Plant Biol* 2012, *12*, 219.
- [120] K. L. Kavanagh, H. Jörnvall, B. Persson, U. Oppermann, *Cell Mol Life Sci* 2008, *65*, 3895.
- [121] H. Jörnvall, J. Hedlund, T. Bergman, U. Oppermann, B. Persson, *Biochem Bioph Res Co* 2010, *396*, 125–130.
- [122] G. Anarat-Cappillino, E. S. Sattely, *Curr. Opin. Plant Biol.* 2014, *19*, 51–58.
- [123] M.-L. Shih, J. A. Morgan, *Metab. Eng. Commun.* 2020, *10*, e00123.
- [124] M. Zalokar, *Exp. Cell Res.* 1960, *19*, 114–132.
- [125] E. S. Kempner, J. H. Miller, *Exp. Cell Res.* 1968, *51*, 150–156.
- [126] J. M. Matthews, 2013, 1965–1965.
- [127] Y. Ofran, B. Rost, *J. Mol. Biol.* 2003, *325*, 377–387.
- [128] Y. Zhang, A. Sampathkumar, S. M. L. Kerber, C. Swart, C. Hille, K. Seerangan, A. Graf, L. Sweetlove, A. R. Fernie, *Nat Commun* 2020, *11*, 1–15.
- [129] J. I. Watkinson, P. A. Bowerman, K. C. Crosby, S. B. Hildreth, R. F. Helm, B. S. J. Winkel, *Peerj* 2018, *2018*, e5598.
- [130] S. B. Hildreth, E. S. Littleton, L. C. Clark, G. C. Puller, S. Kojima, B. S. J. Winkel, *Plant J.* 2022, *110*, 932–945.
- [131] M. Gou, X. Ran, D. W. Martin, C.-J. Liu, *Nat Plants* 2018, *4*, 299–310.
- [132] Z. Ban, H. Qin, A. J. Mitchell, B. Liu, F. Zhang, J. K. Weng, R. A. Dixon, G. Wang, *Proc National Acad Sci* 2018, *115*, E5223–E5232.
- [133] S. J. Nintemann, D. Vik, J. Svozil, M. Bak, K. Baerenfaller, M. Burow, B. A. Halkier, *Front. Plant Sci.* 2017, *8*, 2028.
- [134] L. J. Sweetlove, A. R. Fernie, *Nat Commun* 2018, *9*, 2136.

- [135] V. Pareek, H. Tian, N. Winograd, S. J. Benkovic, *Science* 2020, *368*, 283–290.
- [136] M. H. Chehade, L. Pelosi, C. D. Fyfe, L. Loiseau, B. Rascalou, S. Brugière, K. Kazemzadeh, C.-D.-T. Vo, L. Ciccone, L. Aussel, Y. Cou te, M. Fontecave, F. Barras, M. Lombard, F. Pierrel, *Cell Chem Biol* 2019, *26*, 482-492.e7.
- [137] Y. Zhang, K. F. M. Beard, C. Swart, S. Bergmann, I. Krahnert, Z. Nikoloski, A. Graf, R. G. Ratcliffe, L. J. Sweetlove, A. R. Fernie, T. Obata, *Nat Commun* 2017, *8*, 15212.
- [138] J. W. A. Graham, T. C. R. Williams, M. Morgan, A. R. Fernie, R. G. Ratcliffe, L. J. Sweetlove, *Plant Cell* 2007, *19*, 3723–38.
- [139] C.-Y. Lin, Y. Sun, J. Song, H.-C. Chen, R. Shi, C. Yang, J. Liu, S. Tunlaya-Anukit, B. Liu, P. L. Loziuk, C. M. Williams, D. C. Muddiman, Y.-C. J. Lin, R. R. Sederoff, J. P. Wang, V. L. Chiang, *Front Plant Sci* 2021, *12*, 727932.
- [140] M. Dastmalchi, M. A. Bernards, S. Dhaubhadel, *Plant J* 2016, *85*, 689–706.
- [141] T. Laursen, J. Borch, C. Knudsen, K. Bavishi, F. Torta, H. J. Martens, D. Silvestro, N. S. Hatzakis, M. R. Wenk, T. R. Dafforn, C. E. Olsen, M. S. Motawia, B. Hamberger, B. L. M ller, J.-E. Bassard, *Science* 2016, *354*, 890–893.
- [142] J. Zhang, J. Shin, N. Tague, H. Lin, M. Zhang, X. Ge, W. Wong, M. J. Dunlop, J. Cheng, *Adv. Sci.* 2022, *9*, 2203887.
- [143] B. Lallemand, M. Erhardt, T. Heitz, M. Legrand, *Plant Physiol* 2013, *162*, 616–625.
- [144] S. Mucha, S. Heinzlmeir, V. Kriechbaumer, B. Strickland, C. Kirchhelle, M. Choudhary, N. Kowalski, R. Eichmann, R. H ckelhoven, E. Grill, B. Kuster, E. Glawischnig, *Plant Cell* 2019, *31*, 2697–2710.
- [145] H. Li, Z. Ban, H. Qin, L. Ma, A. J. King, G. Wang, *Plant Physiol.* 2015, *167*, 650–659.
- [146] T. Nakayama, S. Takahashi, T. Waki, *Front Plant Sci* 2019, *10*, 821.
- [147] J. Diharce, J. Golebiowski, S. Fiorucci, S. Antonczak, *Phys Chem Chem Phys* 2016, *18*, 10337–10345.
- [148] N. Fujino, N. Tenma, T. Waki, K. Ito, Y. Komatsuzaki, K. Sugiyama, T. Yamazaki, S. Yoshida, M. Hatayama, S. Yamashita, Y. Tanaka, R. Motohashi, K. Denessiouk, S. Takahashi, T. Nakayama, *Plant J* 2018, *94*, 372–392.
- [149] J. Murata, J. Roepke, H. Gordon, V. D. Luca, *Plant Cell* 2008, *20*, 524–542.
- [150] H. J rnvall, M. Landreh, L. J.  stberg, *Chem-biol Interact* 2015, *234*, 75–79.

Chapter 2. The Structural and Mechanistic Study of CADs in MIA Biosynthesis

2.1 Introduction

2.1.1 Cinnamyl Alcohol Dehydrogenases

CADs are a subfamily of ADHs that are named after their primary catalytic activity of reducing aromatic cinnamaldehydes to cinnamyl alcohols (Figure 11). Comparative phylogenetic and structural analyses suggest that both plant and bacterial CADs originated from a common glutathione-dependent formaldehyde dehydrogenase ancestor [1]. The prevalence of CADs in plants is strongly linked to their functional role in monolignol biosynthesis, though there are reports of members of this enzyme family acting on other defence-related substrates [2, 3]. In bacteria, the function of CADs is less known, though they are speculated to act in defence responses [4, 5].

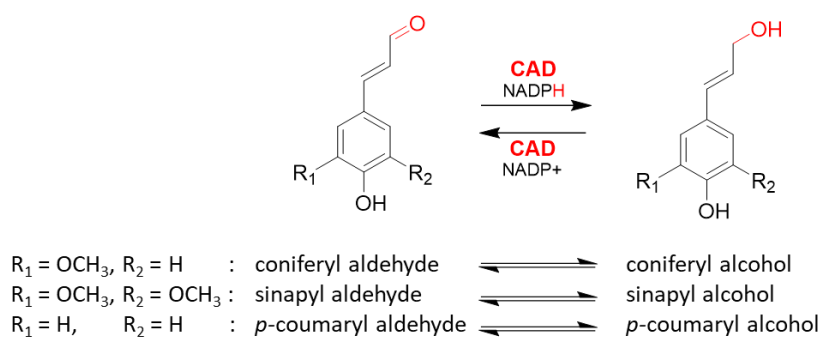


Figure 11. Reversible reduction of aromatic cinnamyl aldehydes to their respective primary alcohols catalysed by CAD in monolignol biosynthesis [8].

Structurally, CADs are comprised of two domains – a catalytic domain consisting of an ancestral GroES protein and a nucleotide-binding domain containing the Rossmann-fold motif that binds the adenine dinucleotide of the NADP(H) cofactor (Figure 12) [6]. The catalytic domain coordinates two zinc ions referred to as the catalytic and the structural zinc. The structural zinc is thought to maintain the tertiary structure of the protein and is tetrahedrally coordinated by the thiol groups of four highly conserved Cys residues. The catalytic zinc resides within the enzyme active site and acts as a Lewis acid during catalysis. It is typically coordinated by two thiol groups each from a Cys residue, an imidazole group from a His residue, and a variable fourth position. This is typically occupied by water molecule, though is substituted for a Glu residue in *Arabidopsis thaliana* CAD4 and CAD5 [7, 8].

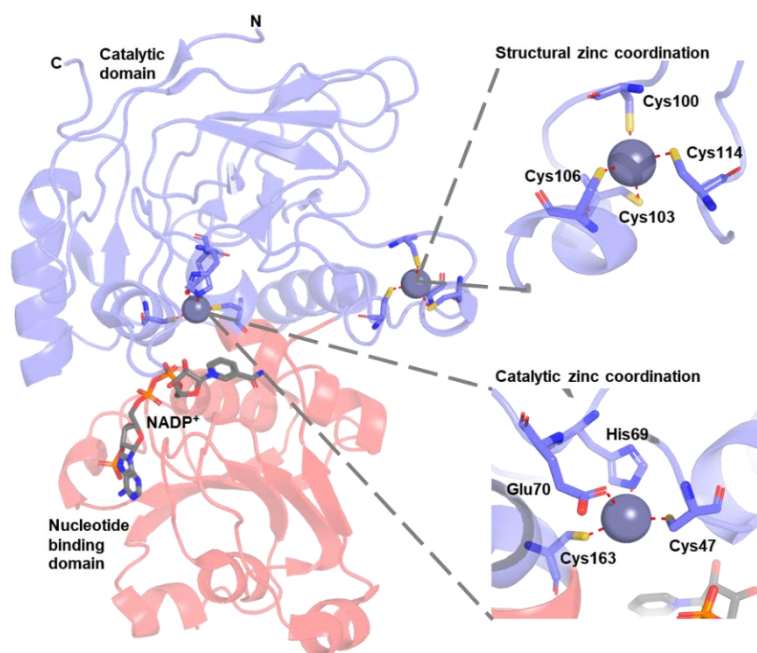


Figure 12. Structural features of AtCAD5 (PDB 2CF6). Catalytic domain is coloured blue and nucleotide binding domain is coloured red. Insets show residues involved in coordinating the structural (top) and the catalytic (bottom) zinc ions.

2.1.2 ADH Catalytic Mechanism

The catalytic mechanism of ADHs and their subfamilies (including CADs) is highly conserved with examples studied from mammals, yeast and plants [8–13]. During reduction, the aldehyde moiety of the substrate binds to the catalytic zinc in the fourth position, displacing the previously bound water molecule or residue (Figure 13). The binding of the NADP(H) cofactor relies on a conserved network of hydrogen bonds between residues within the Rossmann-fold. Notably, an imidazole group from a histidine residue (*Cr8HGO* His55) and

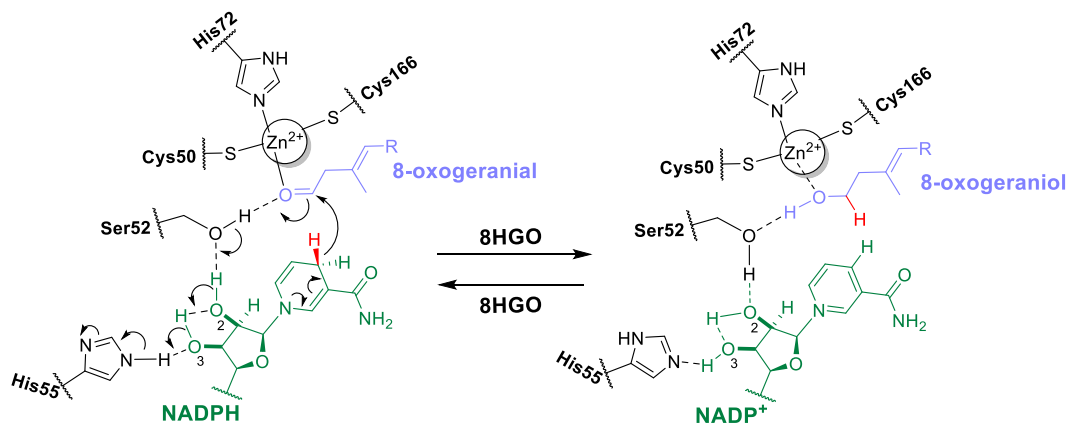


Figure 13. Mechanism of the *Cr8HGO*-catalysed aldehyde reduction of 8-oxogeranial. Figure adapted from Yuon *et al.*, [8].

a hydroxyl-containing residue (Cr8HGO Ser52) bind to the O3' and O2' groups of the cofactor ribose ring respectively. These residues collectively function as a proton relay during catalysis, facilitating the transfer of a hydride from the cofactor to the substrate's aldehyde moiety. The resulting primary alcohol product then leaves the active site and the water molecule or residue rebinds to the catalytic zinc in the fourth position. Typical CADs can also catalyse the oxidation of the alcohol product, performing the reaction in the reverse direction.

2.1.3 CAD-Catalysed Reductions in Monoterpene and MIA Biosynthesis

Members of the CAD subfamily catalyse numerous steps in MIA biosynthesis, including in vinblastine biosynthesis (Figure 14) [14–30]. CADs typically catalyse the reversible reduction of aldehydes of monolignol substrates. However, these enzymes have functionally diverged to act on non-aromatic substrates such as monoterpenes and MIAs. In addition, some CADs have been reported to catalyse the irreversible 1,2- or 1,4-reduction of an iminium moiety. The reduction chemistries catalysed by CADs within monoterpene and MIA biosynthesis is summarised below.

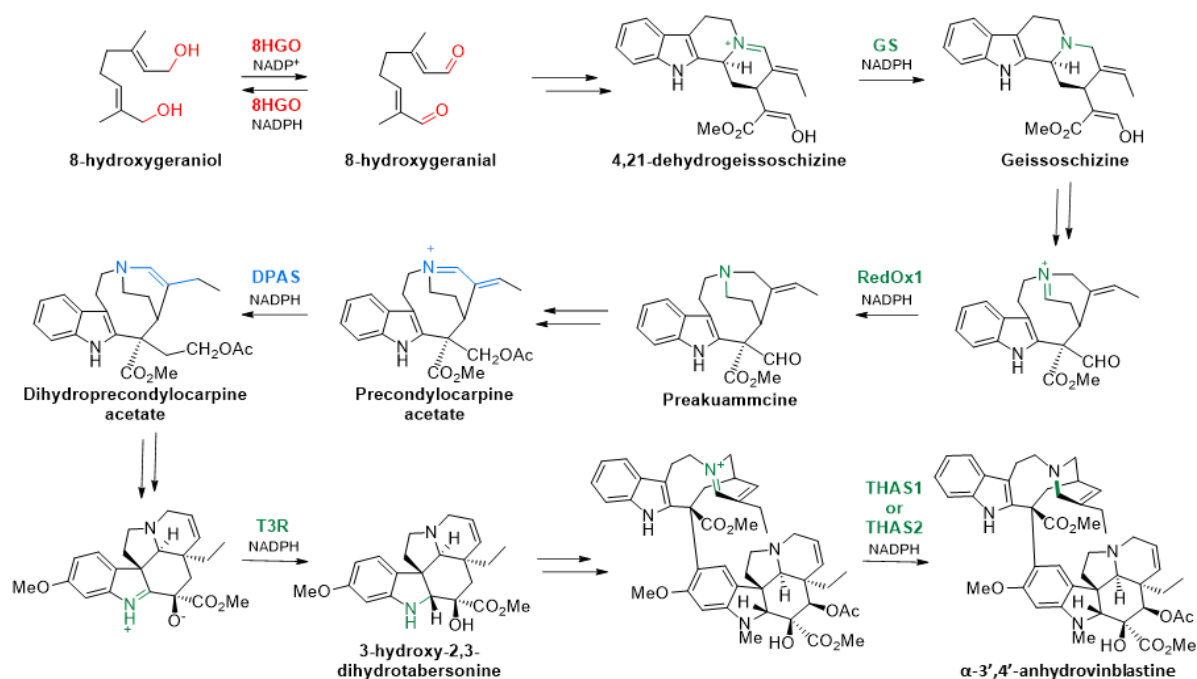


Figure 14. CAD-catalysed reactions in vinblastine biosynthesis in *C. roseus*. Typical oxidation of a primary alcohol to an aldehyde catalysed by 8HGO [31] coloured in red, 1,2-iminium reductions catalysed by GS [14], Redox1 [26], T3R [25], THAS1 and THAS2 [15, 16] coloured in green, and 1,4-iminium reduction catalysed by DPAS [22] coloured in blue.

2.1.3.1 CAD-Catalysed Aldehyde Reductions in Monoterpene Biosynthesis

Monoterpene biosynthesis forms the secologanin precursor subunit required for MIA scaffold biogenesis. The CAD *C. roseus* 8-hydroxygeraniol oxidase (*Cr8HGO*) catalyses two successive alcohol oxidations of the monoterpene 8-hydroxygeraniol to form 8-oxogeraniol or 8-hydroxygeraniol respectively. The remaining alcohol group of these intermediates then undergoes a further *Cr8HGO*-catalysed oxidation to form 8-oxogeraniol (Figure 15) [30]. While *Cr8HGO* is capable of catalysing the reverse reactions *in vitro*, it appears to perform a solely oxidative function *in vivo* [31]. The catalytic mechanism of *Cr8HGO* is similar to previously reported aldehyde-reducing CADs, whilst differences in the active site are proposed to enable the binding of the monoterpene substrates [31].

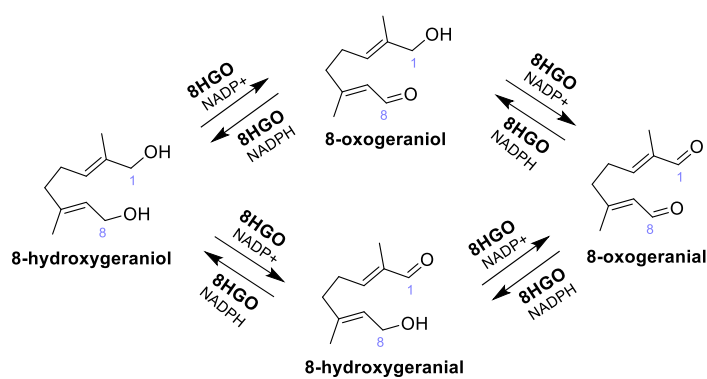


Figure 15. *Cr8HGO* catalysed oxidation of 8-hydroxygeraniol to 8-oxogeraniol. Figure adapted from Krithika *et al.*, [30].

2.1.3.2 CAD-Catalysed 1,2-Reduction of an Iminium Moiety in MIA Biosynthesis

In addition to their divergence to accept non-aromatic substrates, CADs have been reported to perform the 1,2-reduction of an iminium moiety in MIA biosynthesis (Figure 16). This atypical reduction chemistry was shown using isotopic labelling and subsequent structural characterisation of the enzymatic products of the CADs *C. roseus* tetrahydroalstonine synthase (*CrTHAS*; Figure 16C) and *C. roseus* heteroyohimbine synthase (*CrHYS*; Figure 16F). Furthermore, structural analyses of these enzymes revealed variations in otherwise highly conserved residues which typically form a proton relay between the substrate and cofactor during catalysis [15]. CADs have since been reported to catalyse the 1,2-reduction of an iminium moiety in MIA-producing species across Gentianales, as discussed in Chapter 3 of this thesis. This discovery highlights the catalytic divergence of CADs within MIA biosynthesis.

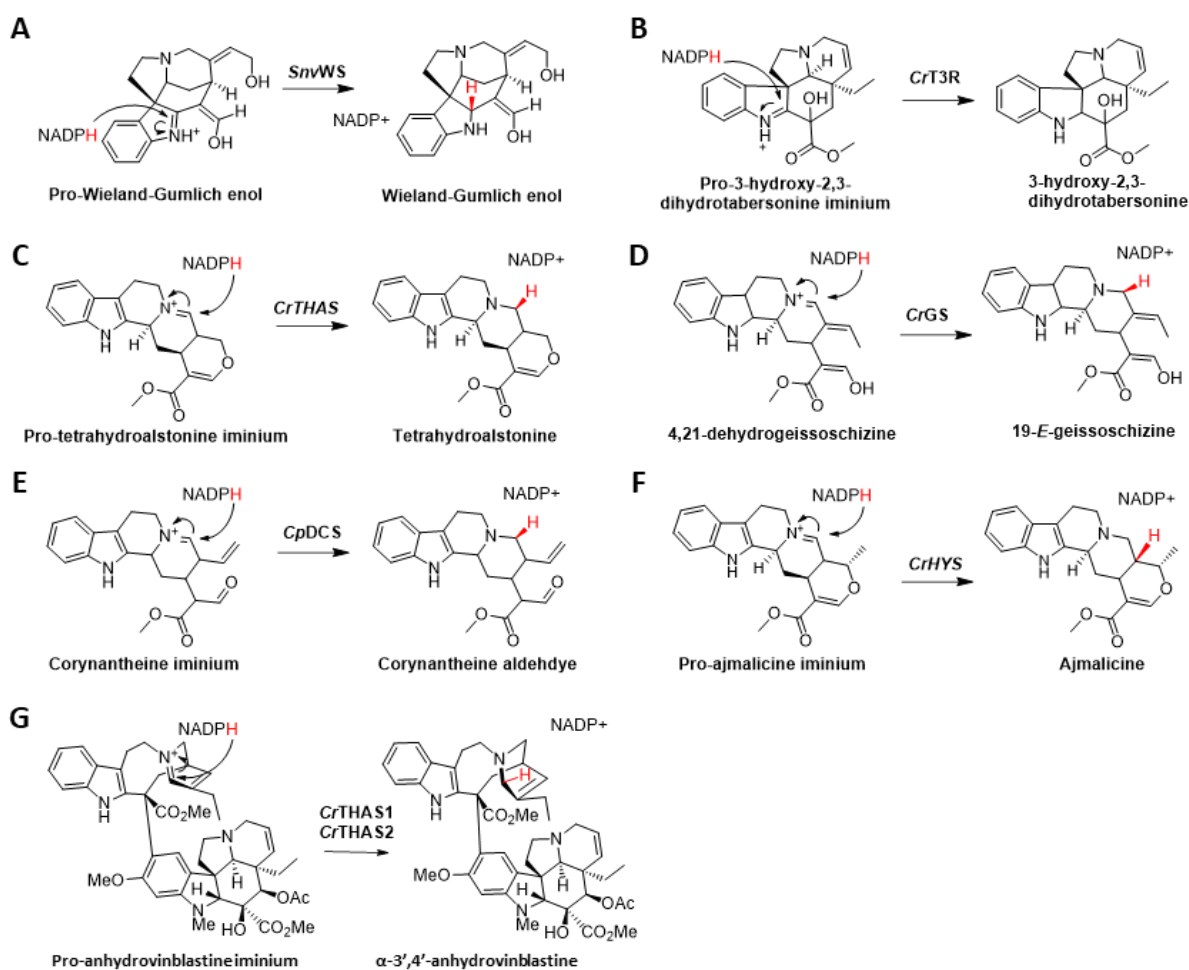


Figure 16. Characterised 1,2-iminium reduction reactions catalysed by CADs in MIA biosynthesis. **A.** *Strychnos nux-vomica* Wieland-Gumlich synthase (*SnvWS*)^[20]; **B.** *Catharanthus roseus* tabersonine-3-reductase (*CrT3R*)^[25]; **C.** *Catharanthus roseus* tetrahydroalstonine synthase (*CrTHAS*)^[16]; **D.** *Catharanthus roseus* geissoschizine synthase (*CrGS*)^[14]; **E.** *Chinchona pubescens* dihydrocorynantheine aldehyde synthase (*CpDCS*)^[29]; **F.** *Catharanthus roseus* heteroyohimbine synthase (*CrHYS*)^[15]; **G.** *CrTHAS1* or *CrTHAS2*^[28].

2.1.3.3 CAD-Catalysed 1,4-Reduction of an Iminium Moiety in MIA Biosynthesis

The CAD enzyme DPAS was recently proposed to catalyses the 1,4-iminium reduction of the MIA precondylocarpine acetate (Figure 17)^[22]. Although initially identified in *C. roseus* (*CrDPAS*)^[22], DPAS orthologues have since been reported in the related species *Tabernanthe iboga* (*TiDPAS1* and *TiDPAS2*)^[19]. This enzyme is proposed to catalyse a hydride addition at C19 of the substrate precondylocarpine acetate, resulting in the 1,4-reduction of the iminium moiety (Figure 17). The resulting 19,20-dihydroprecondylocarpine acetate is then postulated to undergo a spontaneous

desacetoxylation to form the highly unstable dehydrosecodine. Although this intermediate could not be isolated, the formation of the stable isomer angryline under acidic conditions enabled structural characterisation of this compound. Dehydrosecodine, in turn, can be cyclised by various α/β -hydrolase enzymes including *C. roseus* tabersonine synthase (*CrTS*) and *C. roseus* catharanthine synthase (*CrCS*), giving rise to the aspidosperma- or iboga-type MIAs (-)-tabersonine or (+)-catharanthine, respectively [22, 32]. Furthermore, DPAS catalyses the 1,4-iminium reduction of dehydrosecodine to form the unstable intermediate secodine, which is subsequently cyclised to yield vincadifformine [22].

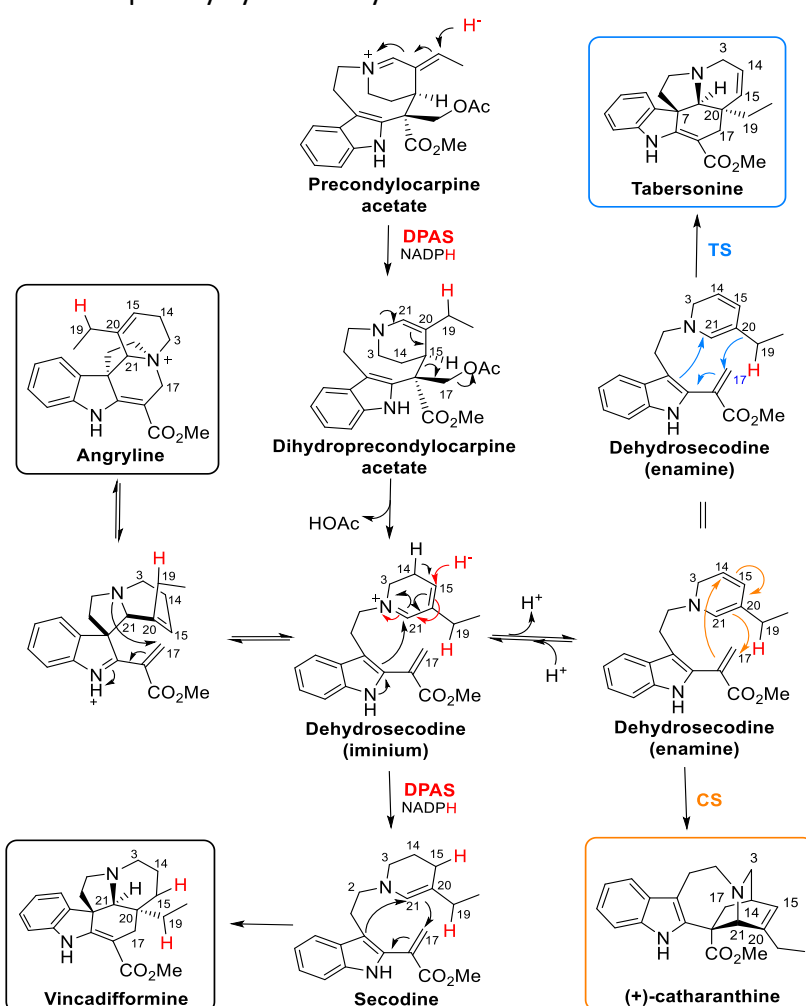


Figure 17. DPAS-catalysed 1,4-iminium reduction of precondylocarpine acetate. Reaction catalysed by *Catharanthus roseus* or *Tabernanthe iboga* dihydroprecondylocarpine acetate synthase orthologues (*CrDPAS*, *TiDPAS1*, or *TiDPAS2*). The resulting dehydrosecodine can rearrange in acidic conditions to form angryline, undergo a further DPAS-catalysed 1,4-reduction to form vincadifformine, or be cyclised by *Catharanthus roseus* tabersonine synthase (*CrTS*) or *Catharanthus roseus* catharanthine synthase (*CrCS*) enzymes to form tabersonine or catharanthine, respectively.

Contrary to these findings, the previously characterised 1,2-iminium reducing *CrGS* (Figure 16D) was also reported to catalyse the 1,4-iminium reduction of precondylocarpine acetate [26]. Nevertheless, the 1,4-reductions of precondylocarpine acetate and dehydrosecodine were previously unreported chemistries for an ADH enzyme. The chemical reduction of imines remains challenging and engineering of enzymatic imine reductases to accept non-native substrates has had limited success [33–35]. Therefore understanding the mechanism of the ADH-catalysed 1,2- and 1,4-reduction of an iminium moiety observed in MIA biosynthesis is of biochemical and bioengineering interest.

2.1.4 Chapter Overview

Members of the CAD subfamily perform atypical reductions in MIA biosynthesis, though the underlying mechanisms of these reactions are poorly understood. This chapter uses comparative structural analysis of CADs that act in monoterpene and MIA biosynthesis to identify key residues that enable the atypical 1,2- and 1,4-reduction of an iminium moiety. Notably, DPAS orthologues are found to have changes in residues that typically coordinate the catalytic zinc which result in the highly unusual loss of this ion. Additionally, we report the DPAS-catalysed 1,4-reduction of an α,β -unsaturated aldehyde and identify that the loss of residues coordinating the catalytic zinc underlies this hitherto unprecedented ADH chemistry. Furthermore, changes in otherwise conserved residues resulting in the atypical binding of the cofactor are found to form the mechanistic basis of the *CrGS*-catalysed 1,2-reduction of an iminium moiety. The findings discussed in this chapter expand the catalytic repertoire of the ADH family of enzymes and provide the mechanistic basis of these atypical reductions in MIA biosynthesis.

2.2 Results

2.2.1 DPAS-Catalysed Reduction of Precondylocarpine Acetate and Dehydrosecodine

Conflicting reports suggest that the enzymatic reduction of the substrate precondylocarpine acetate is catalysed by both *CrGS* and *CrDPAS* [22, 26]. To address this ambiguity, we conducted *in vitro* assays with the substrate precondylocarpine acetate and cofactor NADPH in the presence of either *CrDPAS* or *CrGS* and analysed the products using UPLC-MS (Figure 18). Given the instability of both the substrate and the product, determination of the steady-state kinetic constants for these reactions was not possible. Nevertheless, assays containing *CrDPAS* showed the formation of a product that co-eluted and had the same MS/MS fragmentation pattern as the chemically synthesised standard of angryline – a rearrangement of dehydrosecodine that forms under acidic conditions. In contrast, assays containing *CrGS* exhibited substantially reduced product levels, suggesting that *CrDPAS* is better able to catalyse the reduction of precondylocarpine acetate.

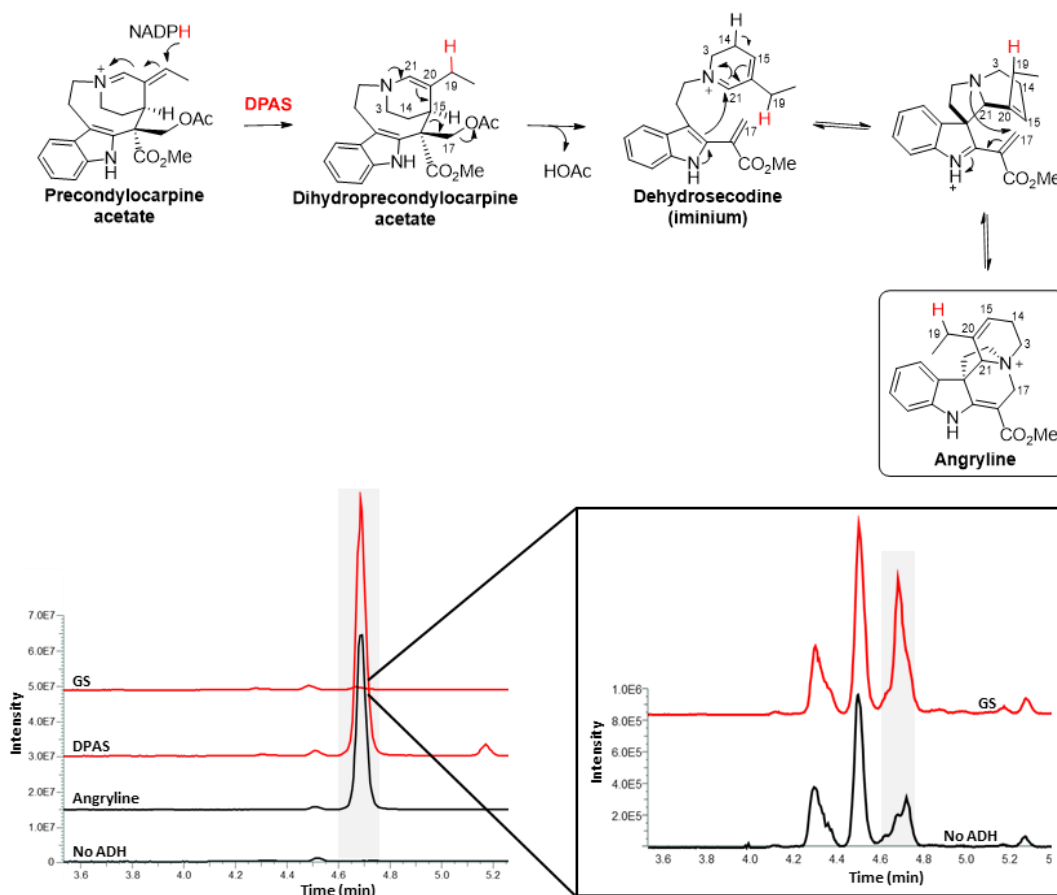


Figure 18. UPLC-MS chromatograms of *in vitro* reactions of *CrDPAS* and *CrGS* with substrate precondylocarpine acetate and cofactor NADPH. EIC m/z 337.180 ± 0.05 . Inset of *CrGS* and no ADH control reactions to show small amount of *CrGS*-specific product formation.

2.2.2 Isotopic Labelling of the DPAS-Catalysed Reduction of Precondylocarpine Acetate

DPAS orthologues from *C. roseus* (*CrDPAS*) and *T. iboga* (*TiDPAS1* and *TiDPAS2*) are proposed to catalyse the 1,4-reduction of the substrate precondylocarpine acetate (Figure 19A). Following a subsequent spontaneous desacetoxylation, this product forms the unstable intermediate dehydrosecodine which in turn can rearrange in acidic conditions to form the more stable compound, angryline. Alternatively, DPAS can catalyse the further reduction of dehydrosecodine to produce the unstable intermediate secodine, which subsequently cyclises to form the compound vincadifformine (Figure 19A). Catalysis of a 1,4-reduction by an ADH is unprecedented. Therefore, to investigate the mechanism of DPAS, we performed *in vitro* assays of precondylocarpine acetate and DPAS orthologues *CrDPAS*, *TiDPAS1* or *TiDPAS2*, using either cofactor (NADPH) or deuterated cofactor (pro-*R*-NADPD). Analysis using UPLC-MS showed that reactions containing DPAS and NADPH formed a product which co-eluted and had the same MS/MS spectra as the chemically synthesised standard of angryline. However, in analogous reactions using NADPD, the peak displayed a +1 *m/z* shift (Figure 19B and C). Through preparative-HPLC isolation and subsequent NMR analysis, we confirmed this product as *d*-angryline. Comparing the *d*-angryline and angryline ¹H NMR spectra revealed a signal loss at H19, indicating deuterium incorporation at C19 ^[22] (Figure 19A; Appendix I).

Furthermore, *in vitro* reactions with the substrate precondylocarpine acetate, a DPAS orthologue and the cofactor NADPH resulted in the generation of vincadifformine (Figure 19B). This product is proposed to form through the DPAS-catalysed 1,4-reduction of the intermediate dehydrosecodine (Figure 19A) ^[22]. Interestingly, repeating these assays using NADPD formed a product that co-eluted with the vincadifformine standard but had a +2 *m/z* shift, indicative of a doubly-reduced product (Figure 19B and C). Through preparative-HPLC isolation and NMR characterisation, we identified this peak as *d*₂-vincadifformine. Furthermore, comparing the *d*₂-vincadifformine and the vincadifformine ¹H NMR spectra showed a loss of signal at H15, indicative of deuterium incorporation at C15 (Figure 19A; Appendix II and III). These findings show that the hydride in the DPAS-catalysed reductions of both precondylocarpine acetate and dehydrosecodine is donated from the cofactor through a 1,4-iminium reduction mechanism at C19 and C15, respectively.

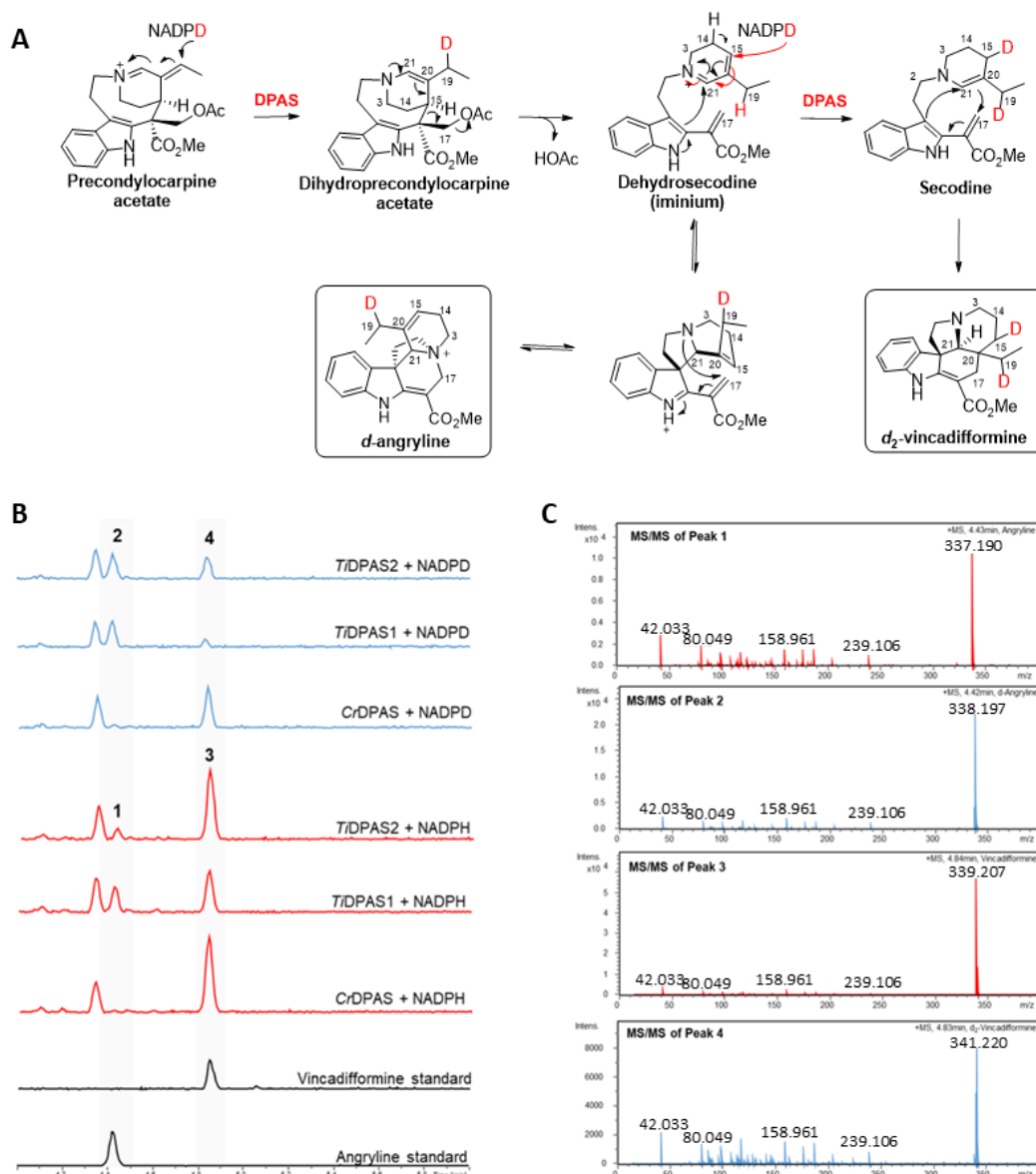


Figure 19. Deuterium labelling of DPAS-catalysed 1,4-reductions of precondylocarpine acetate and dehydrosecodine showing deuterium incorporation at C19 and C15, respectively. **A.** DPAS-catalysed reduction of precondylocarpine acetate and dehydrosecodine showing deuterium incorporation at C19 and C15, respectively. **B.** UPLC/MS TIC of *in vitro* reactions of DPAS orthologs with substrate precondylocarpine and either NADPH or NADPD. **C.** MS/MS² spectra of angryline (1), *d*-angryline (2), vincadifformine (3), and *d*₂-vincadifformine (4) products corresponding to peaks numbered in **B**.

2.2.3 Stereoselectivity of the DPAS-Catalysed Production of Vincadifformine

DPAS catalyses the reduction of the proposed substrate dehydrosecodine to form secodine, which subsequently cyclises to form vincadifformine (Figure 19A). However, it is not known whether DPAS catalyses the stereospecific cyclisation of secodine, or whether

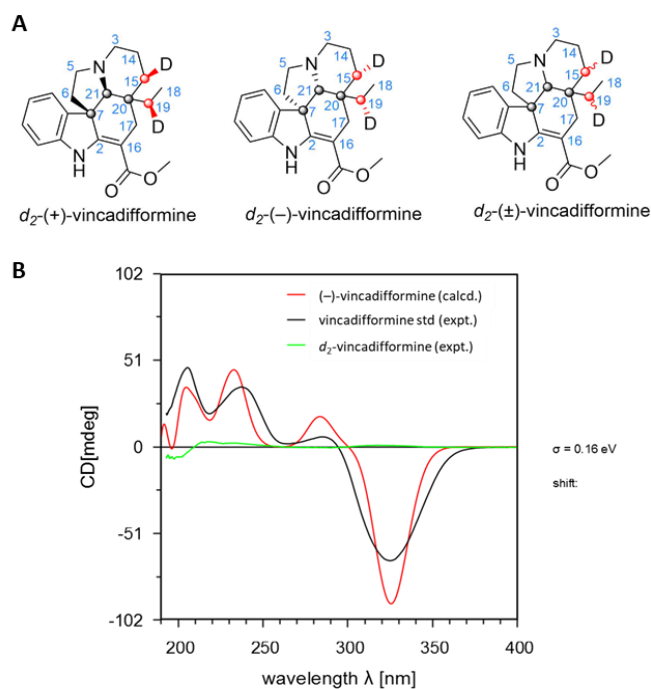


Figure 20. Stereoselectivity of vincadifformine cyclisation. **A.** Enantiomers of d_2 -vincadifformine. Chiral centres are highlighted. **B.** ECD spectra of d_2 -vincadifformine (green), vincadifformine standard (black) and calculated spectra of (-)-vincadifformine (red).

the cyclisation occurs spontaneously outside the enzyme active site, resulting in the observation of both enantiomers of vincadifformine (Figure 20A). To address this question, we used electronic circular dichroism (ECD) to compare the spectra of the enzymatically prepared d_2 -vincadifformine, the (-)-vincadifformine standard, and the calculated spectra of (-)-vincadifformine (Figure 20B). While the (-)-vincadifformine standard matched the calculated (-)-vincadifformine spectra, the enzymatically produced d_2 -vincadifformine exhibited a weak ECD signal. This observation is indicative of a racemic mixture of vincadifformine enantiomers, suggesting that the cyclisation of secodine occurs externally to the DPAS substrate pocket and is not catalysed by DPAS itself.

2.2.4 DPAS-Catalysed Reduction of Strictosidine Aglycone

Several CADs have been reported to catalyse the 1,2-iminium reduction of structural rearrangements of the substrate strictosidine aglycone [14–16]. To test whether DPAS catalyses a similar reduction, we conducted *in vitro* reactions using the substrate strictosidine and the enzyme CrSGD to generate the unstable aglycone in the presence of

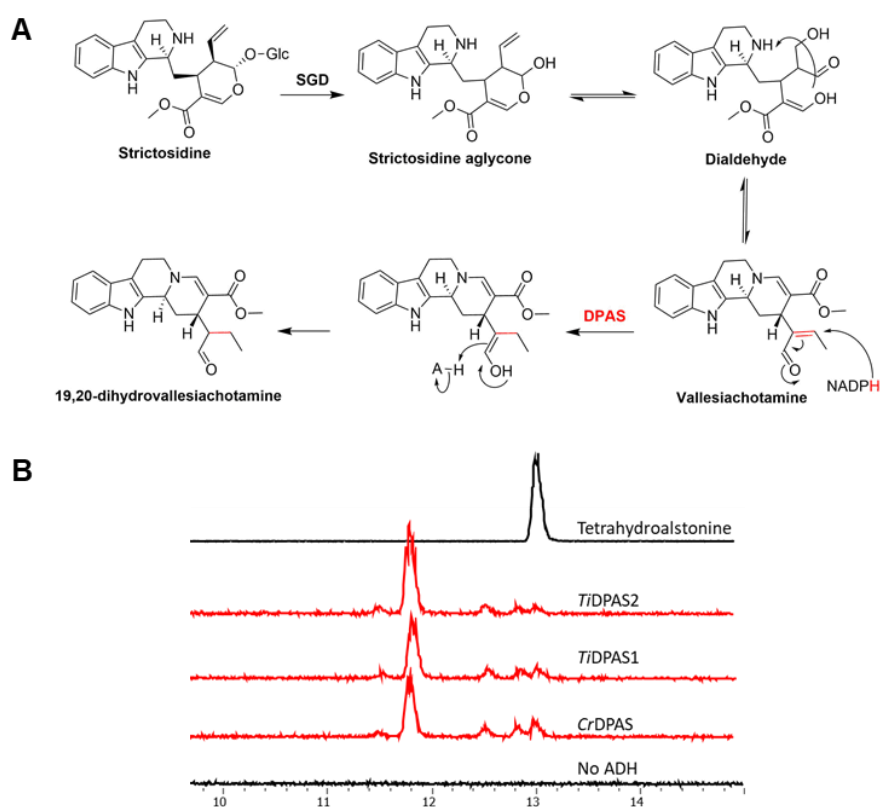


Figure 21. DPAS-catalysed reduction of vallesiachotamine. **A.** Proposed formation of 19,20-dihydrovallesiachotamine. **B.** UPLC/MS TIC of *in vitro* reactions of the substrate strictosidine, NADPH, *CrSGD* and either *CrDPAS*, *TiDPAS1* or *TiDPAS2*.

a DPAS orthologue (*CrDPAS*, *TiDPAS1* or *TiDPAS2*), and NADPH cofactor. Using UPLC-MS analysis we detected the formation of a peak that did not co-elute with previously characterised reduced products of strictosidine aglycone (Figure 21B). This was subsequently isolated by preparative HPLC and characterised by NMR as 19,20-dihydrovallesiachotamine (Figure 21A; Appendix IV). This compound was previously only partially characterised using chemical synthesis ^[36]. We propose that 19,20-dihydrovallesiachotamine is formed through a DPAS-catalysed 1,4-reduction of the structural rearrangement of strictosidine aglycone known as vallesiachotamine. A hydride transfer at C19 of the substrate reduces the α,β -unsaturated aldehyde, resulting in the spontaneous rearrangement of the C20 double-bond and ultimately giving rise to 19,20-dihydrovallesiachotamine (Figure 21A). This is the first report of a 1,4-reduction of an α,β -unsaturated aldehyde to be catalysed by an ADH, further expanding the chemical repertoire of this class of enzymes.

2.2.5 Solution of the CrDPAS Structure

DPAS catalyses the highly unusual 1,4-reduction of substrates precondylocarpine acetate, secodine and vallesiachotamine (Figure 19 and Figure 21). To understand the enzymatic mechanism of these reactions, we sought to solve the structure of CrDPAS using X-ray crystallography. To achieve this, CrDPAS was codon-optimised for expression in *E. coli* and the resulting protein purified using several rounds of affinity and size exclusion chromatography. The protein was then plated using the sitting-drop method with excess NADP⁺ and commercially available precipitant screens to yield crystals (Figure 22).

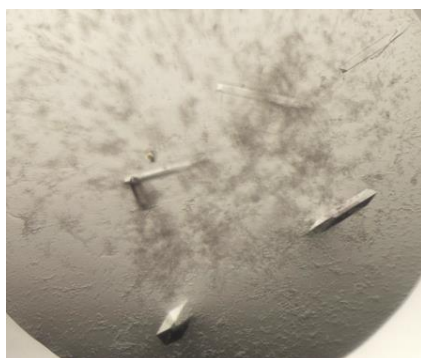


Figure 22. Crystallisation of CrDPAS. Crystals of 8 mg/mL CrDPAS protein with 1 mM NADP⁺ grown in 1.26 M ammonium sulfate precipitant and 200 mM lithium sulfate in 100 mM TRIS buffer pH 8.5.

The apo-CrDPAS structure was solved to 2.5 Å resolution as a homodimer using molecular replacement (Figure 23, Appendix V Table 16). However, a 34 amino acid region in each chain (between residues 102-136) of the structure lacked density within the 2Fo-Fc map, indicative of significant protein disorder. This missing region contained residues known to coordinate the structural zinc. Despite efforts to further optimise crystallisation conditions, we did not achieve an improved resolution or a complete structure of CrDPAS. However, analysis of the structure found that despite ADHs widely being reported to co-purify bound with the catalytic and structural zinc ions ^[8, 15], there was a lack of density in the 2Fo-Fc map at the expected site of the catalytic zinc (Figure 23C). Moreover, although NADP⁺ was present in both the crystallisation solution and cryoprotectant, no density correlating to the cofactor was observed in the CrDPAS structure.

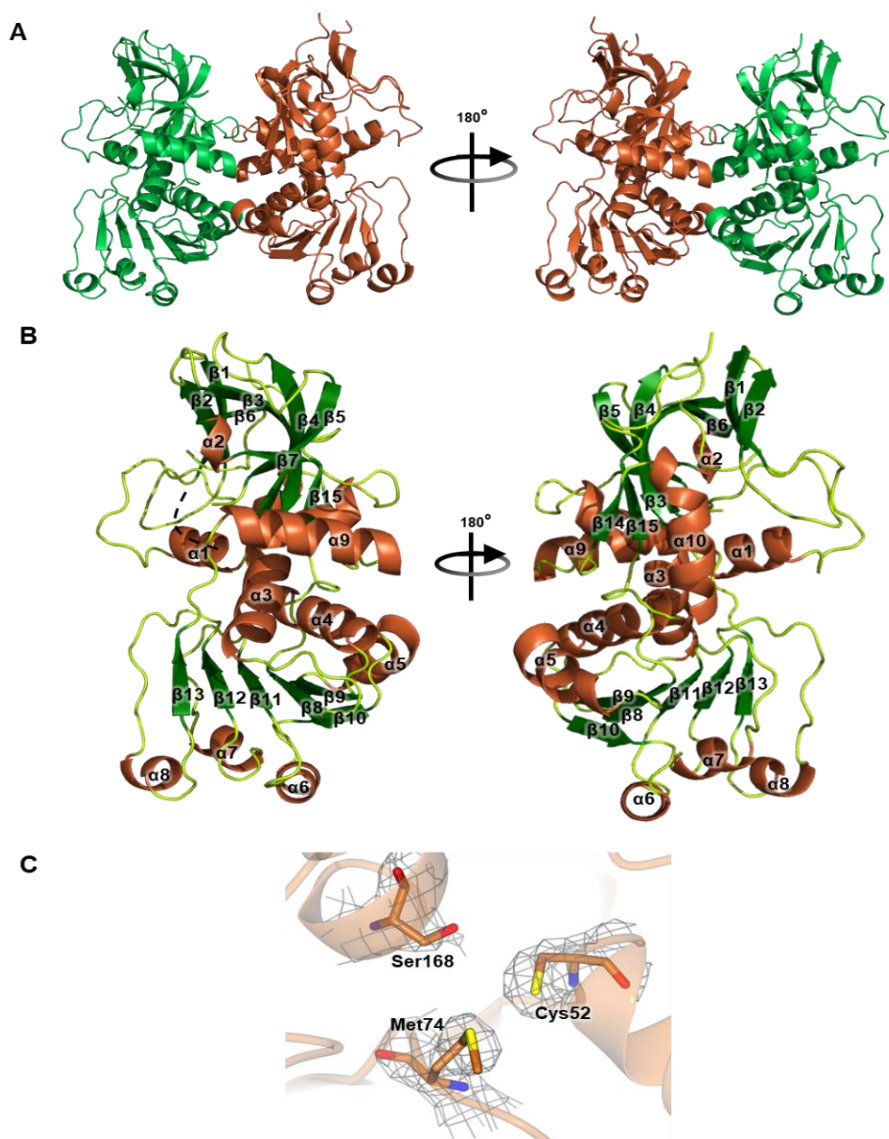


Figure 23. Apo-*CrDPAS* crystal structure. **A.** Partial structure of *CrDPAS* crystallised as a homodimer, coloured by chains. **B.** Monomer of *CrDPAS* coloured by secondary structure. 34 amino acid missing region represented by black dashed line. **C.** Loss of coordination of the catalytic zinc ion. Residues in positions that typically coordinate the ion shown as sticks. Density of 2Fo-Fc map shown as mesh contoured to 1.0 σ .

2.2.6 Solution of *TiDPAS2* Structure

The species *T. iboga* is closely related to *C. roseus* and also produces iboga- and aspidosperma-type alkaloids through the proposed intermediate dehydrosecodine^[37]. Two orthologues of *CrDPAS* have been identified in *T. iboga* (named *TiDPAS1* and *TiDPAS2*) that each catalyse the 1,4-reduction of precondylocarpine acetate, secodine and vallesiachotamine (Figure 19 and Figure 21)^[19]. Seeking to elucidate a complete structure

of DPAS, we solved the structure of apo-*Ti*DPAS2, achieving a resolution of 2.4 Å (Figure 24, Appendix V Table 17). *Ti*DPAS2 was crystallised as a homodimer and bound to the structural zinc ion by four Cys residues (Figure 24D). Remarkably, no density was observed in the 2Fo-Fc map at the expected sites of the catalytic zinc (Figure 24C) or the NADP⁺ cofactor. Despite extensive efforts, attempts to solve the structure of *Ti*DPAS1 were unsuccessful.

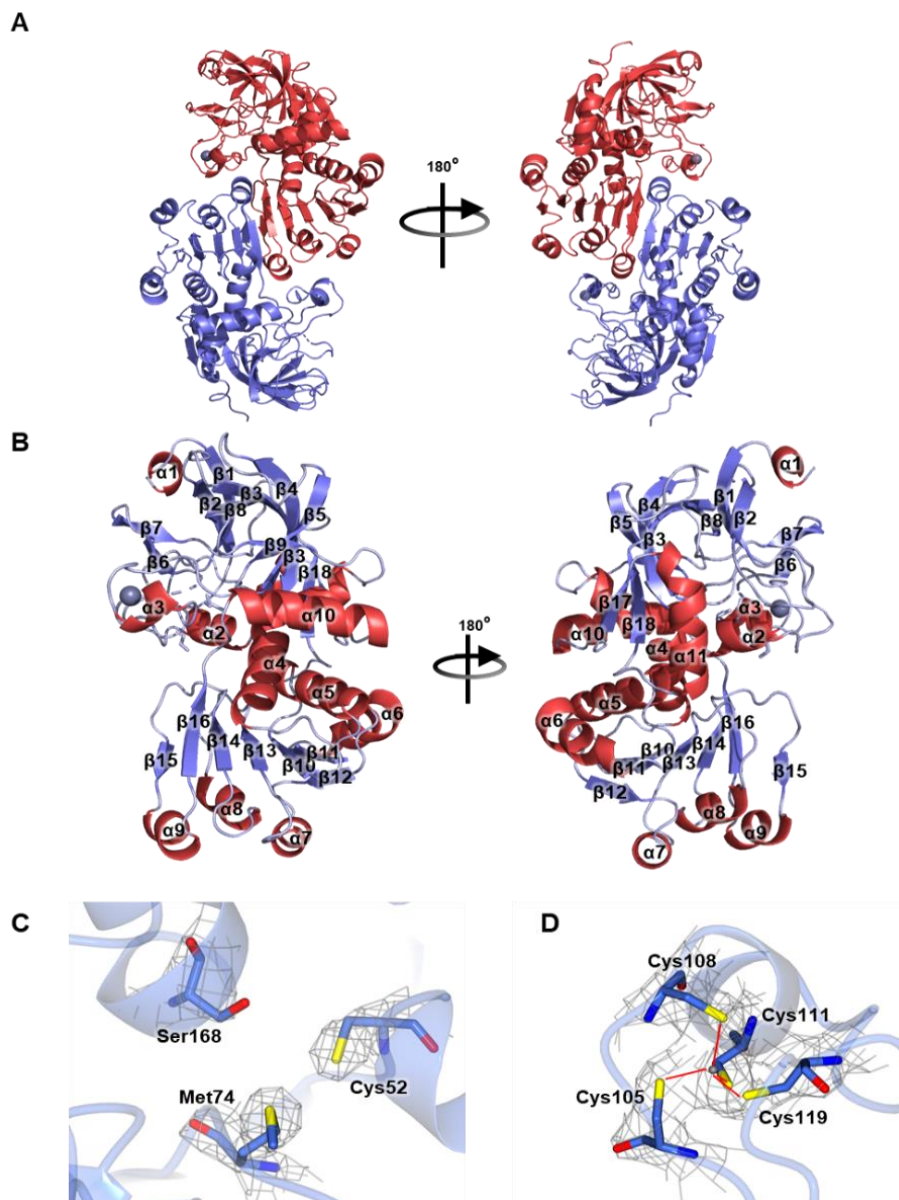


Figure 24. Apo-*Ti*DPAS2 crystal structure. **A.** Structure of *Ti*DPAS2 homodimer coloured by chains. **B.** Monomer of *Ti*DPAS2 coloured by secondary structure. **C.** Loss of coordination of the catalytic zinc. **D.** Coordination of the structural zinc. Residues in positions that typically coordinate metal ions shown as sticks. Density of 2Fo-Fc map shown as mesh contoured to 1.0 σ .

In addition, the structures of *Ti*DPAS2 bound to precondylocarpine acetate (Figure 25, Appendix V Table 18) and stemmadenine acetate (Figure 26, Appendix V Table 19) were solved to 1.8 Å and 2.2 Å, respectively. The substrate-bound structures both crystallised as homodimers, and as observed in the apo-structure, lacked density in the 2Fo-Fc map at the expected sites of the NADP⁺ cofactor and the catalytic zinc ion.

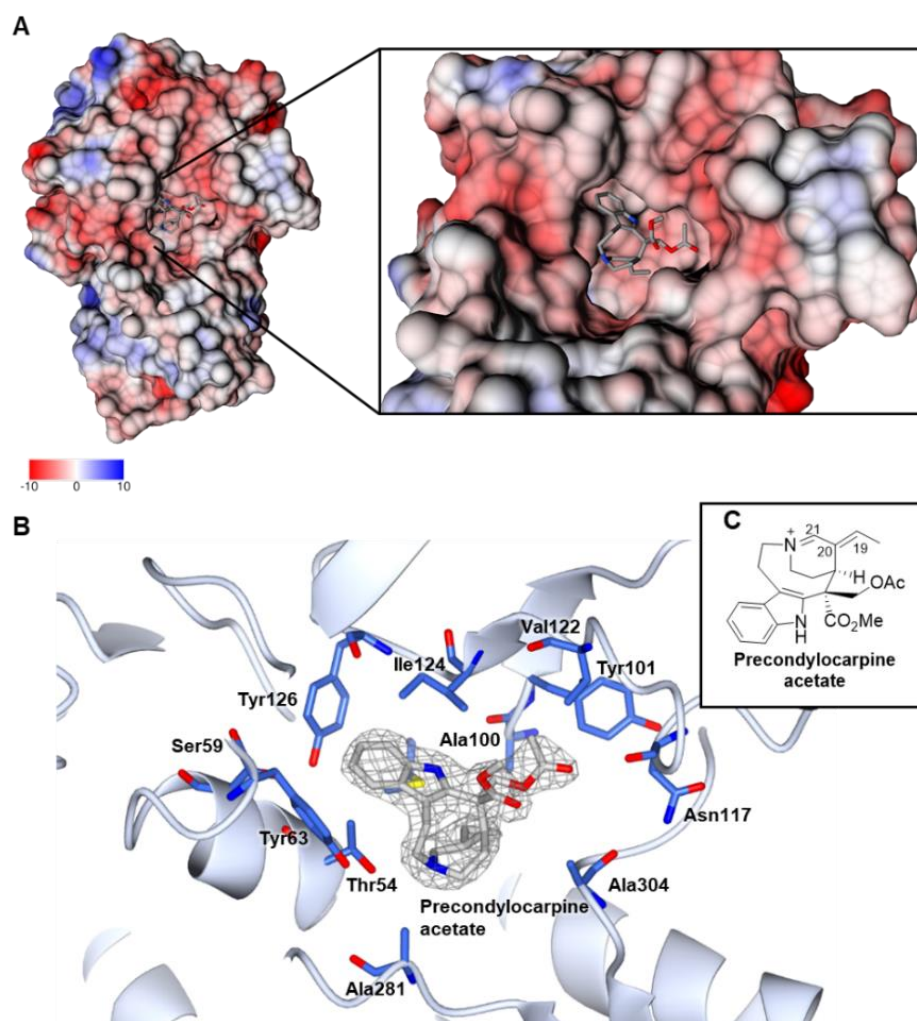


Figure 25. Structure of *Ti*DPAS2 bound to precondylocarpine acetate. **A.** Surface view of *Ti*DPAS2 bound to precondylocarpine acetate coloured by electrostatic charge. **B.** Substrate pocket of *Ti*DPAS2 bound to precondylocarpine acetate. Density of 2Fo-Fc map for substrate shown as mesh, contoured to 1.0 σ . **C.** Structure of precondylocarpine acetate.

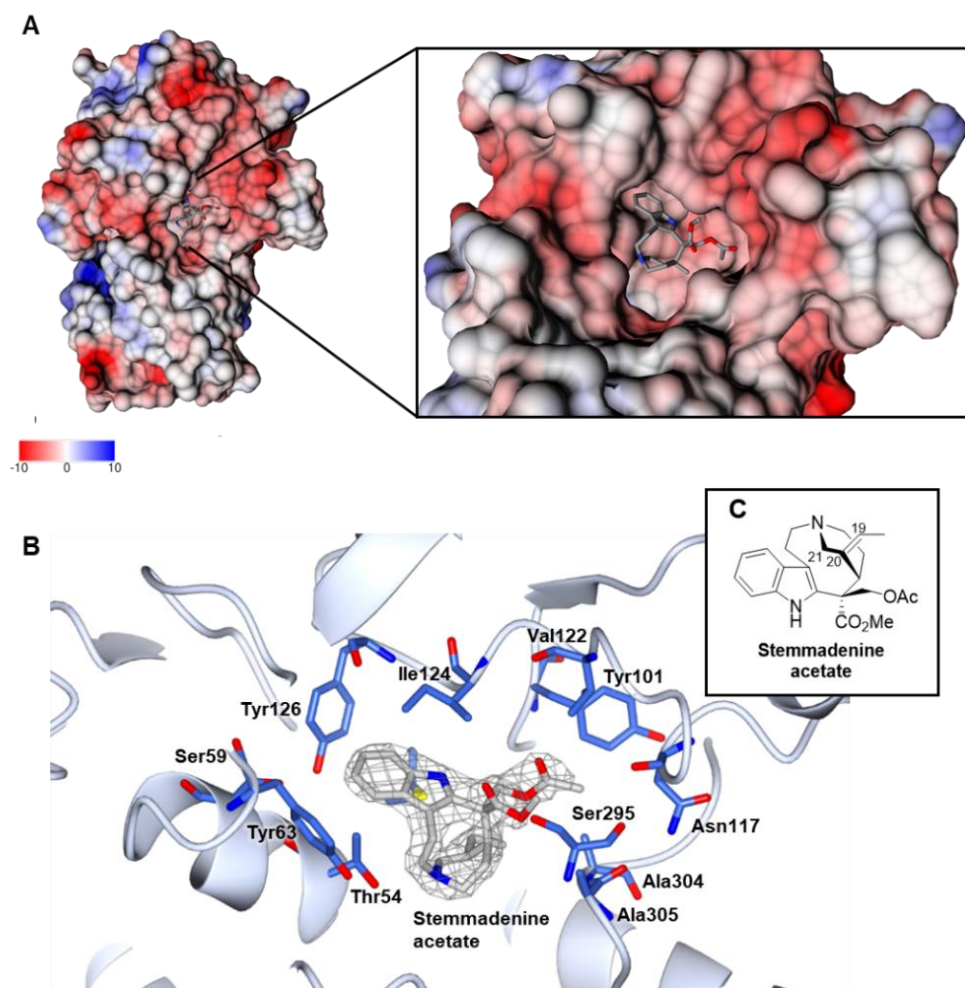


Figure 26. Structure of *TiDPAS2* bound to stemmadenine acetate. **A.** Surface view of *TiDPAS2* bound to stemmadenine acetate coloured by electrostatic charge. **B.** Substrate pocket of *TiDPAS2* bound to stemmadenine acetate. Density of 2Fo-Fc map for substrate shown as mesh, contoured to 1.0 σ . **C.** Structure of stemmadenine acetate.

The substrate pocket of *TiDPAS2* is framed by α -helix 2 and several flexible loop regions, suggesting there may be movement upon substrate binding. However, comparison of the apo- and substrate-bound *TiDPAS2* revealed only a 0.38 Å RMSD between the structures (Figure 27A and B). Conformational changes were detected in α -helix 2 upon precondylocarpine acetate binding, including a 1.0 Å shift of residue Thr54 to form a hydrogen bond with the substrate's iminium moiety (Figure 27C). Although this movement may be a resolution artefact from the structure solutions, it suggests a potential role of *TiDPAS2* Thr54 in binding the substrate within the active site.

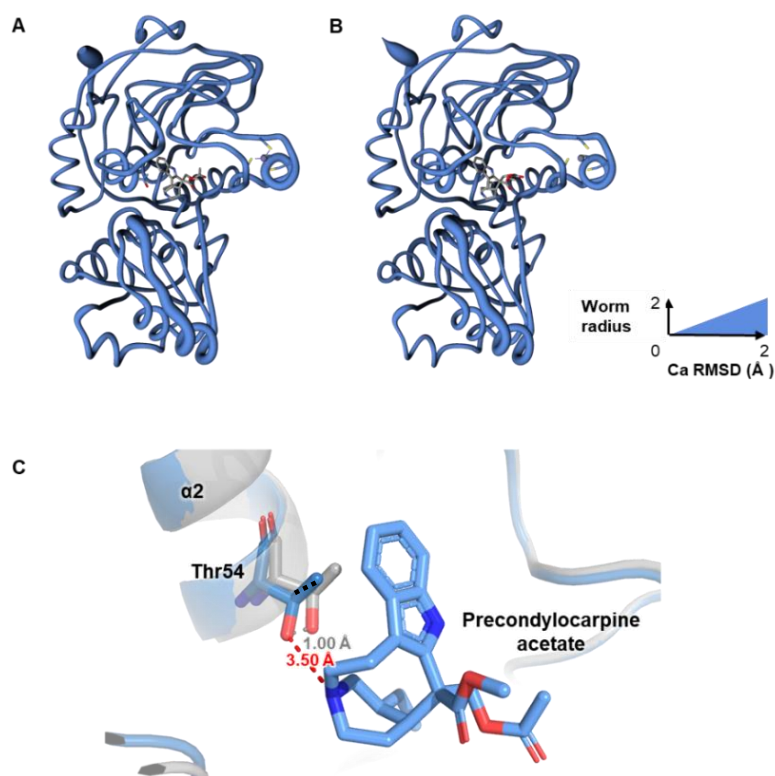


Figure 27. Conformational changes between apo- and precondylocarpine acetate bound (**A.**), and apo-*TiDPAS2* and stemmadenine acetate bound structures (**B.**). Ca RMSD rendered by worm radius. **C.** Movement of Thr54 between apo-*TiDPAS2* (grey) and precondylocarpine acetate-bound *TiDPAS2* (blue). Distance of Thr54 to the substrate's iminium moiety shown by red dashed line and movement between structures shown by black dashed line.

2.2.7 Comparison of CAD Structures in Apocynaceae

In addition to *CrDPAS* and *TiDPAS2*, the structures of four other CADs involved in monoterpene and MIA biosynthesis have previously been solved. Despite their high amino acid identity (Table 1), these enzymes catalyse different reduction reactions, namely the reduction of an aldehyde (*Cr8HGO*), the 1,2-reduction of an iminium moiety (*CrGS*, *CrTHAS1*, *CrTHAS2*) or the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde (*CrDPAS*, *TiDPAS2*). Therefore, to understand the mechanistic basis of these atypical reductions, we compared the structures of these CADs. Although their secondary structures were largely conserved, significant variability was observed in the region surrounding the substrate pocket previously reported to influence the enzyme's substrate specificity (namely *CrTHAS2* β -strands 5-9; Figure 28) ^[15].

Table 1. Amino acid sequence identity (%) of the solved structures of CADs from Apocynaceae. *Cr8HGO* (PDB 6KJ5), *CrGS* (PDB 8A3N), *CrTHAS1* (PDB 5FI3), *CrDPAS* (PDB 8B27), *TiDPAS2* (PDB 8B1V), *CrTHAS2* (PDB 5H81).

	<i>Cr8HGO</i>	<i>CrGS</i>	<i>CrTHAS1</i>	<i>CrDPAS</i>	<i>TiDPAS2</i>	<i>CrTHAS2</i>
<i>Cr8HGO</i>		56.63	62.88	60.67	61.50	59.03
<i>CrGS</i>	56.63		55.22	47.67	49.59	50.54
<i>CrTHAS1</i>	62.88	55.22		53.99	55.92	54.72
<i>CrDPAS</i>	60.67	47.67	53.99		86.30	51.34
<i>TiDPAS2</i>	61.50	49.59	55.92	86.30		51.88
<i>CrTHAS2</i>	59.03	50.34	54.71	51.34	51.88	

The structures of *CrDPAS* and *TiDPAS2* lacked density at the expected site of the catalytic zinc (Figure 23C and Figure 24C). Comparative structural and sequence analyses of CADs which co-purified with both these ions revealed two otherwise conserved residues typically involved in coordinating the catalytic zinc ion that differed in *CrDPAS* and *TiDPAS2*, (*CrDPAS* and *TiDPAS2* Met74 instead of His, and Ser168 instead of Cys; Figure 29) likely leading to the loss of this ion in these structures. Notably, the four Cys residues that coordinate the structural zinc ion were conserved between all structures (*TiDPAS2* Cys105, Cys108, Cys111 and Cys 119; Figure 30). This observation suggests that despite the lack of density in this region of the *CrDPAS* structure, the ion is likely conserved.

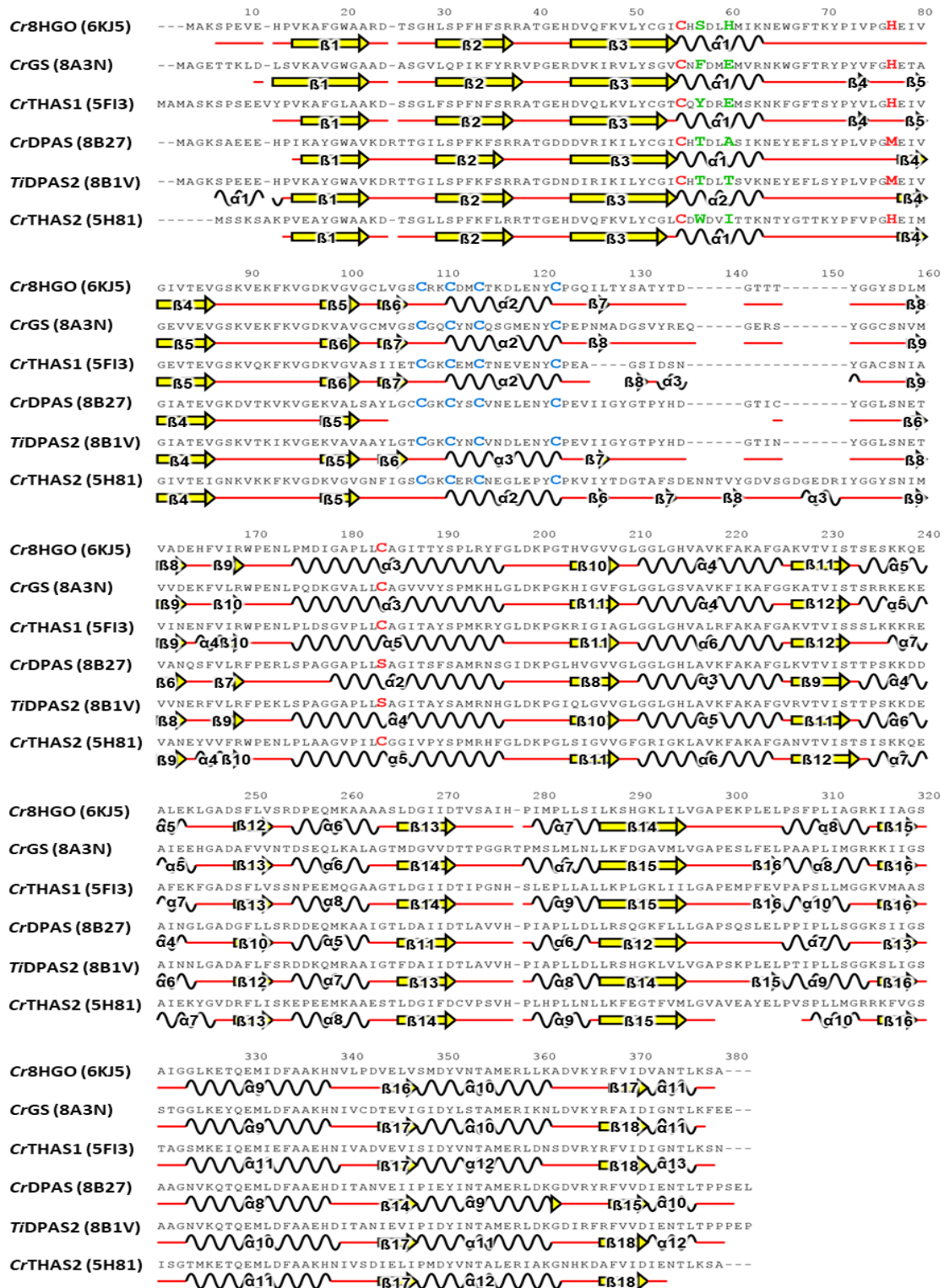


Figure 28. MUSCLE amino acid alignment of solved structures of CADs from Apocynaceae and their respective PDB accessions. Secondary structure features represented as follows: β strands as yellow arrows, α helices as black lines, and loops as red lines. Residues coordinating the catalytic zinc coloured in red, residues coordinating the structural zinc in blue and residues involved in the proton relay coloured in green. Figure made using 2dSS^[60].

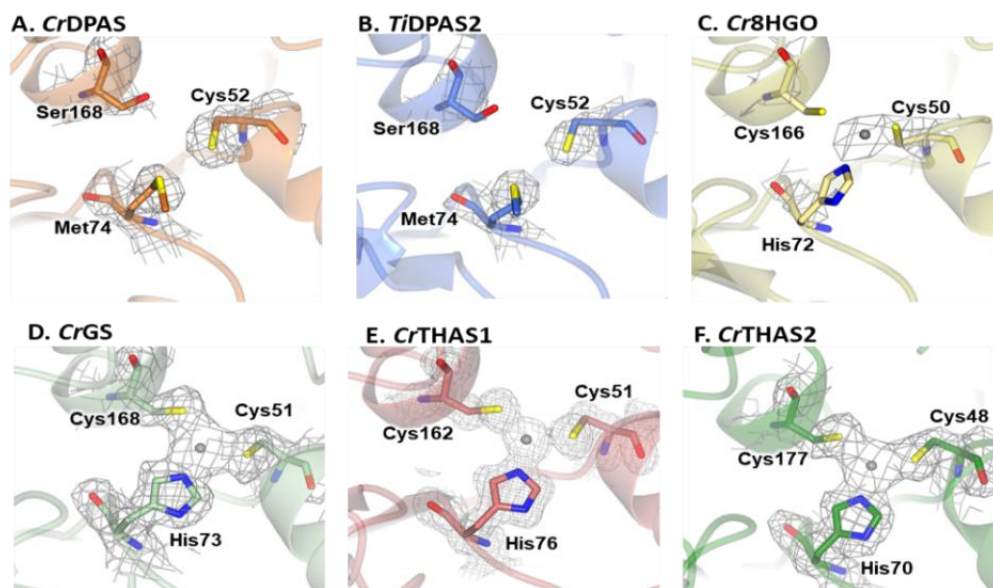


Figure 29. Catalytic zinc coordination in CADs from Apocynaceae. 2Fo-Fc density map contoured to 1.0σ and shown as mesh. **A.** *CrDPAS* (PDB 8B27), **B.** *TiDPAS2* (PDB 8B1V), **C.** *Cr8HGO* (PDB 6K3G), **D.** *CrGS* (PDB 8A3N), **E.** *CrTHAS1* (PDB 5FI3), **F.** *CrTHAS2* (PDB 5H81).

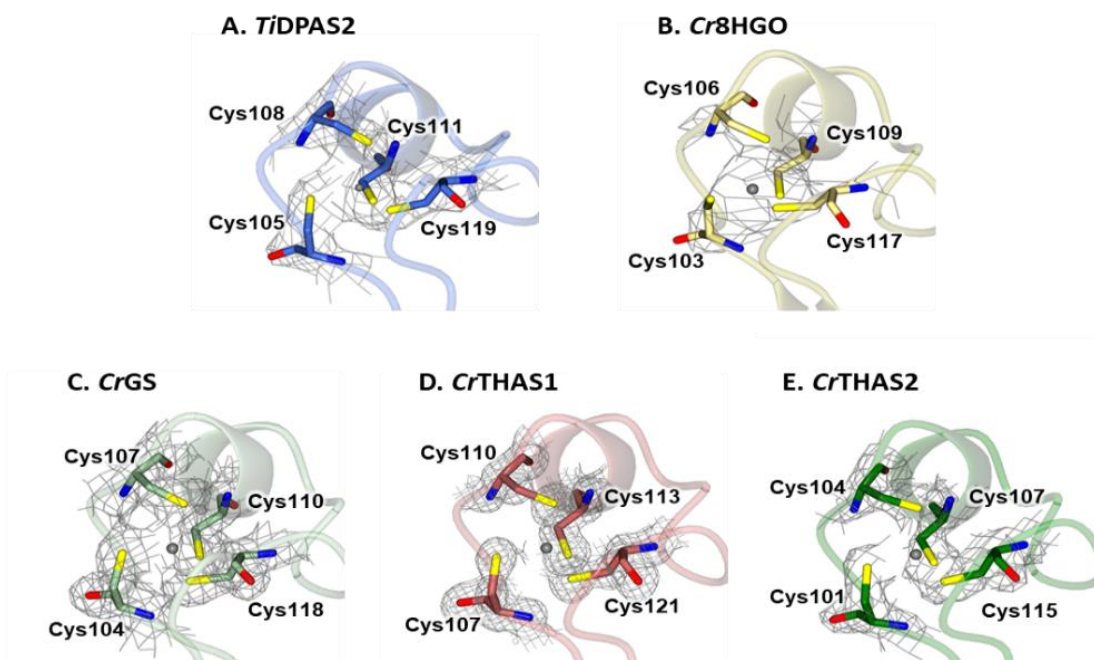


Figure 30. Structural zinc coordination in CADs from Apocynaceae. 2Fo-Fc density map contoured to 1.0σ and shown as mesh. **A.** *TiDPAS2* (PDB 8B1V), **B.** *Cr8HGO* (PDB 6K3G), **C.** *CrGS* (PDB 8A3N), **D.** *CrTHAS1* (PDB 5FI3), **E.** *CrTHAS2* (PDB 5H81).

CADs form a proton relay during catalysis comprised of a system of hydrogen bonds between highly conserved residues and the ribose ring of the cofactor and the substrate. This is observed in *Cr8HGO* His55 and Ser52 which binds to the O3' and the O2' of the cofactor, respectively (Figure 31A and B) [5, 11]. However, in *CrGS*, the inert Phe53 occupies the position that typically binds to O2' of the NADP⁺ (corresponding to *Cr8HGO* Ser52; Figure 31C and D), resulting in a 2.10 Å repositioning of the cofactor compared to the *Cr8HGO* structure (Figure 32A). This leads to *CrGS* Glu54 (corresponding to *Cr8HGO* His55) being too distal to form a hydrogen bond with O3'. Instead, this cofactor position binds to the carbonyl of *CrGS* Asn52 (corresponding to *Cr8HGO* His51). Similar changes in cofactor binding the cofactor were also observed in the sequences of the closely related 1,2-iminium catalysts *CrTHAS1* and *CrTHAS2* (Figure 28).

Despite the addition of excess NADP⁺ in both the crystallisation and cryoprotectant solutions, there was insufficient electron density to model the cofactor in the structures of *CrDPAS* and *TiDPAS2*. Therefore, NADPH was docked into the active site of *TiDPAS2* using

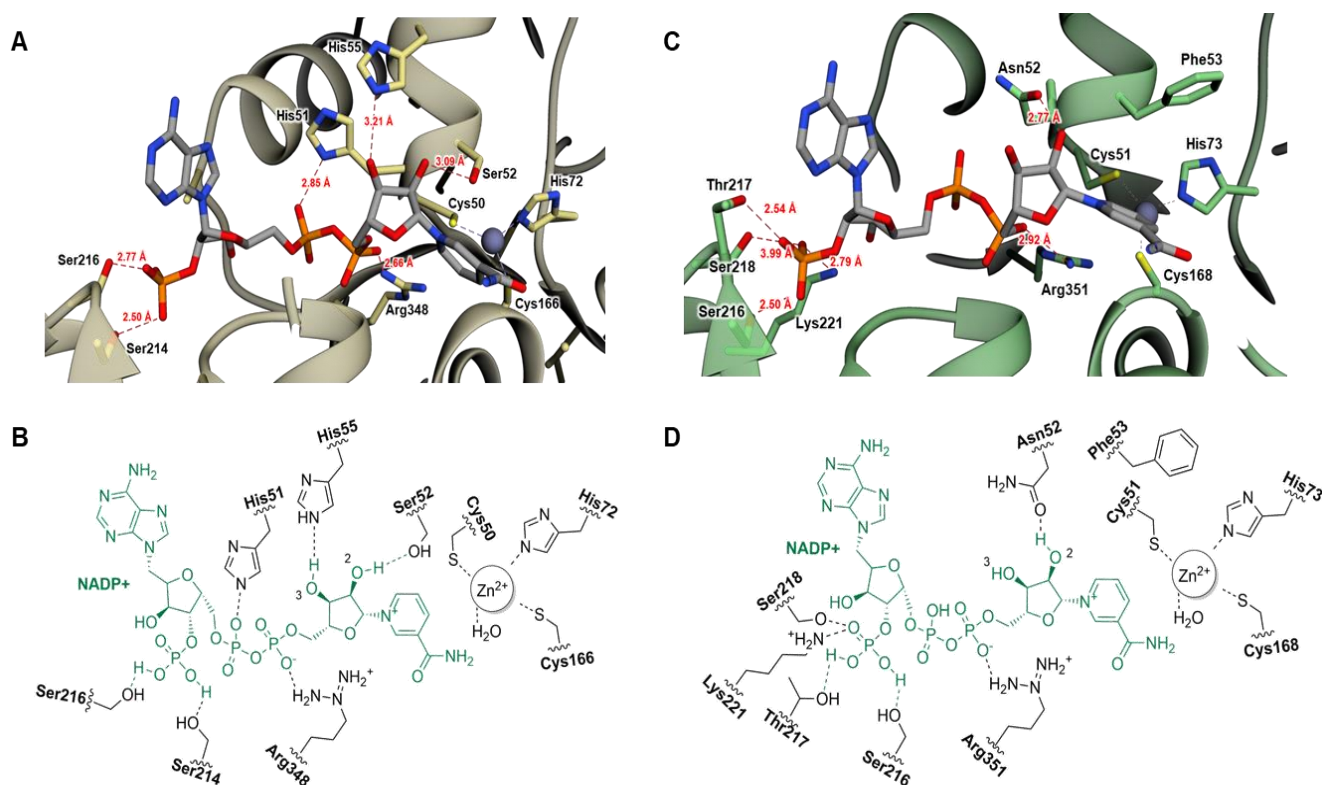


Figure 31. Cofactor binding in *Cr8HGO* and *CrGS*. NADP⁺ bound to *Cr8HGO* (PDB 6K3G; **A**) and *CrGS* (PDB 8A3N; **C**) structures. Hydrogen bonds represented by red dashed line. Cartoon representation of NADP⁺ and catalytic zinc binding in *Cr8HGO* (**B**) and *CrGS* (**D**).

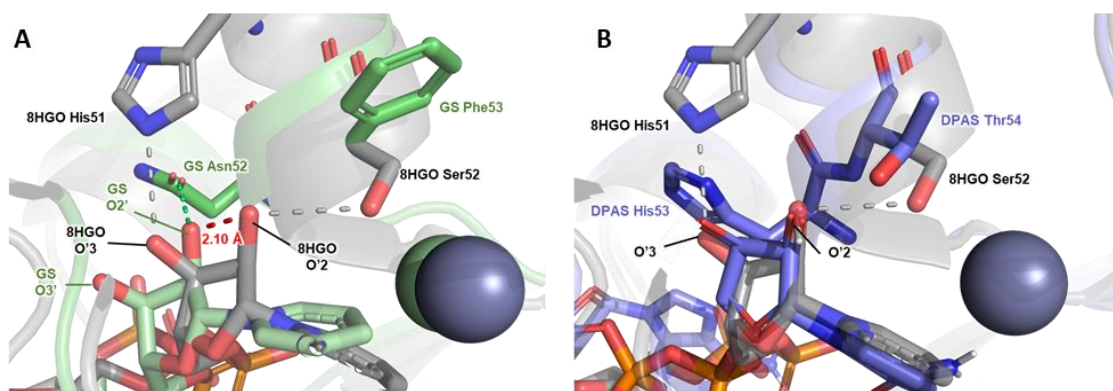


Figure 32. Comparison of cofactor positioning between *Cr8HGO* (PDB 6K3G) and either *CrGS* (A) or *TiDPAS2* (B). Grey dashed lines represent hydrogen bonds between cofactor and enzyme.

AutoDock Vina^[38] to better understand the position of the cofactor relative to the substrate (Figure 33). Subsequent comparison of the cofactor binding between *Cr8HGO* and *TiDPAS2* revealed notable differences in the structures, including the lack of binding between any *TiDPAS2* residue and the O3' of the cofactor. Notably, *TiDPAS2* Thr54 (corresponding to *Cr8HGO* Ser52) was observed to bind with O2' of the ribose ring and was situated 4.11 Å from the reduction site (C19) of the substrate precondylocarpine acetate, suggesting a possible role of this residue in catalysis. However, these changes only yielded a 0.22 Å movement of the cofactor between the *Cr8HGO* and *TiDPAS2* structures (Figure 32B).

The comparative structural and sequence analyses of characterised CADs from Apocynaceae reveals differences in otherwise highly conserved residues that are typically involved in catalysis. Namely, differences were observed between residues in positions that typically coordinate the catalytic zinc ion in the 1,4-iminium reducing *CrDPAS* and *TiDPAS2* and the aldehyde-reducing *Cr8HGO*. In addition, atypical residues were observed in positions that coordinate the cofactor in *CrDPAS*, *TiDPAS2* and *CrGS*, compared to the aldehyde-reducing *Cr8HGO*. These findings suggest that these residues may be involved in enabling the atypical reduction chemistries of *CrDPAS*, *TiDPAS2* and *CrGS*, and will form the basis of further biochemical study of these enzymes detailed in this chapter.

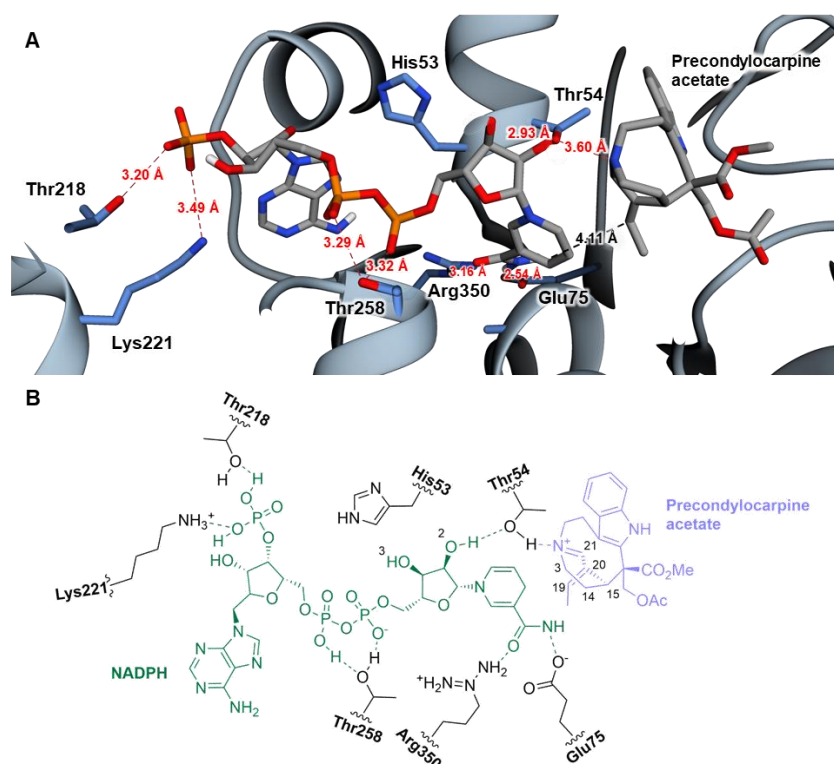


Figure 33. Docking of NADPH cofactor in *TiDPAS2*. **A.** Active site of precondylocarpine acetate-bound *TiDPAS2* docked with NADPH. Red dashed lines indicate hydrogen bonds, black line represents the distance of hydride transfer from cofactor to C19 of the substrate. **B.** Chemical representation of NADPH and precondylocarpine acetate binding in *TiDPAS2*.

2.2.8 *CrDPAS* Mutagenesis Study on the Reduction of Precondylocarpine Acetate

Comparative structural analyses of *CrDPAS* and *TiDPAS2* with other CADs from Apocynaceae revealed differences in otherwise highly conserved residues involved in catalysis. Specifically, residues involved in the proton relay (*CrDPAS* His53 and Thr54) and coordination of the catalytic zinc (*CrDPAS* Met74 and Ser168; Figure 34). To elucidate the functional implications of these differences, we performed site-directed mutagenesis on *CrDPAS* to introduce residues found in the corresponding position in either aldehyde-reducing (e.g. *Cr8HGO*) or 1,2-iminium reducing (e.g. *CrGS*) CADs. The *in vitro* activities of the resultant mutant proteins were tested with the substrate precondylocarpine acetate and cofactor NADPH and the products formed were detected using UPLC-MS (Figure 35).

The catalytic zinc ion typically acts as a Lewis acid during CAD-based catalysis though was absent in the structures of *CrDPAS* and *TiDPAS2* (Figure 23C and Figure 24C). Mutagenesis of the positions that typically coordinate this ion resulted in either the reduced (*CrDPAS*

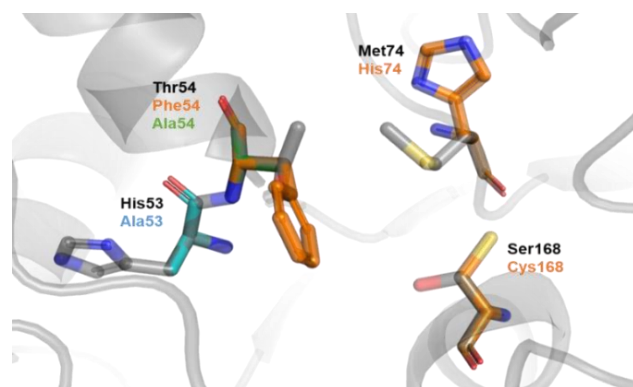


Figure 34. *CrDPAS* active site residues targeted by site-directed mutagenesis. The residues corresponding to residues in *Cr8HGO* or *CrGS* involved in the coordination of the catalytic zinc (Met74 and Ser168) and the proton relay (His53 and Thr54) shown as sticks.

Met74His) or increased (*CrDPAS* Ser168Cys) formation of the doubly-reduced product vincadifformine (Figure 35). Furthermore, the double mutant *CrDPAS* Met74His Ser168Cys displayed comparable activity to *CrDPAS* Met74His (Figure 35). These findings suggest that reintroducing residues that typically coordinate the catalytic zinc in *CrDPAS* does not influence the reduction of precondylocarpine acetate, though may affect the reduction of dehydrosecodine.

CAD-catalysed reductions typically require a proton relay system formed by a network of hydrogen bonds between the enzyme, cofactor, and substrate (Figure 13) [8]. As detailed in section 2.2.7, *TiDPAS2* Thr54 was found to form a hydrogen bond with the iminium moiety of the substrate precondylocarpine acetate and the O2' of NADPH (Figure 33), suggesting its role in catalysis. Additionally, it was hypothesised that *TiDPAS2* His53 may bind to the O3' of the cofactor and act as a base donor during catalysis. Therefore to assess the catalytic role of these residues, we performed site-directed mutagenesis on the corresponding positions in *CrDPAS* and tested the *in vitro* activity of the resulting proteins with precondylocarpine acetate and NADPH (Figure 35). These assays found that Ala replacement of either residue (*CrDPAS* Thr54Ala or His53Ala) resulted in comparable product profiles as wild-type enzyme, suggesting these residues do not have a catalytic role. However, the *CrDPAS* Thr54Phe mutant (the observed corresponding residue in *CrGS*) abolished enzyme activity, likely due to steric hindrance of the substrate.

These findings demonstrate that residues in positions that typically coordinate the catalytic zinc affect the *CrDPAS*-catalysed reduction of dehydrosecodine, though have no discernible

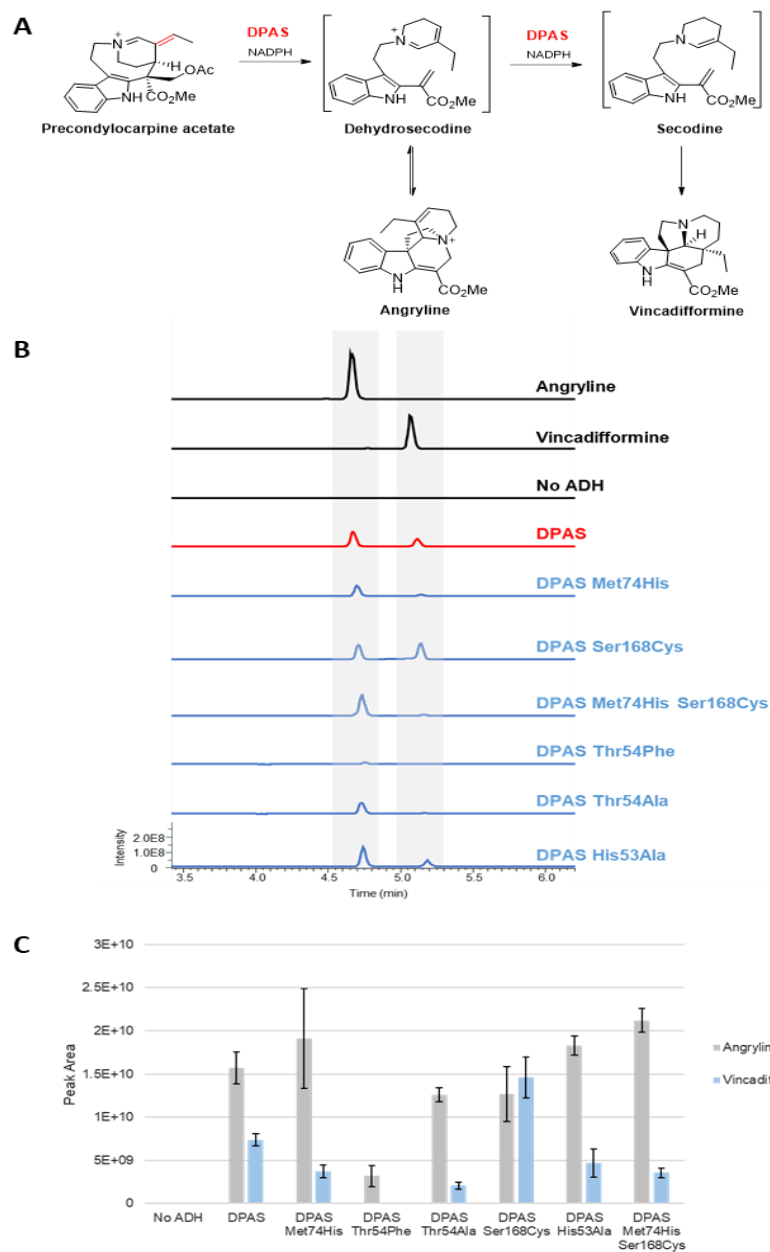


Figure 35. *In vitro* reactions of CrDPAS mutants with precondylocarpine acetate. **A.** DPAS-catalysed reduction of precondylocarpine acetate and dehydrosecodine. **B.** UPLC-MS chromatograms of CrDPAS activity assays. EIC m/z 337.05-340.05. **C.** Peak areas of angrlyne and vincadifformine products from CrDPAS mutant assays. $n=3$, bars show standard deviation.

effect on the reduction of precondylocarpine acetate. Furthermore, the reduction of precondylocarpine acetate was not affected by mutagenesis of residues in positions from the proton relay in aldehyde-reducing CADs (i.e. Cr8HGO). These results suggest that neither the loss of residues able to coordinate the catalytic zinc or those that typically form the proton relay are required for the CrDPAS-catalysed 1,4-reduction of precondylocarpine acetate.

2.2.9 Mechanism of CrDPAS-Catalysed 1,4-Iminium Reduction of Precondylocarpine Acetate and Dehydrosecodine

By combining our findings from isotopic labelling studies, structural analysis and site-directed mutagenesis, we propose the mechanism underlying the CrDPAS-catalysed 1,4-iminium reduction of precondylocarpine acetate and dehydrosecodine (Figure 36). The cofactor NADPH and the substrate bind in the active site aided by hydrogen bonding such as the observed interaction between the residue Thr54 and the iminium moiety. CrDPAS lacks the catalytic zinc which typically acts as a Lewis acid to stabilise the aldehyde of the substrate in ADH-catalysed reductions. However, we suggest the inherent reactivity of the unsaturated iminium of the substrate precondylocarpine acetate negates this requirement. Instead, a hydride transfer from the NADPH cofactor to C19 of the substrate results in a 1,4-reduction, forming dihydroprecondylocarpine acetate. We propose that either Thr54 or a water molecule within the active site act as a proton donor, resulting in the subsequent desacetoxylation and formation of the unstable intermediate dehydrosecodine. A second NADPH subsequently binds and transfers a hydride to C15 of the substrate, leading to a further 1,4-iminium reduction. The resulting secodine then leaves the enzyme active site and spontaneously cyclises to form vincadifformine.

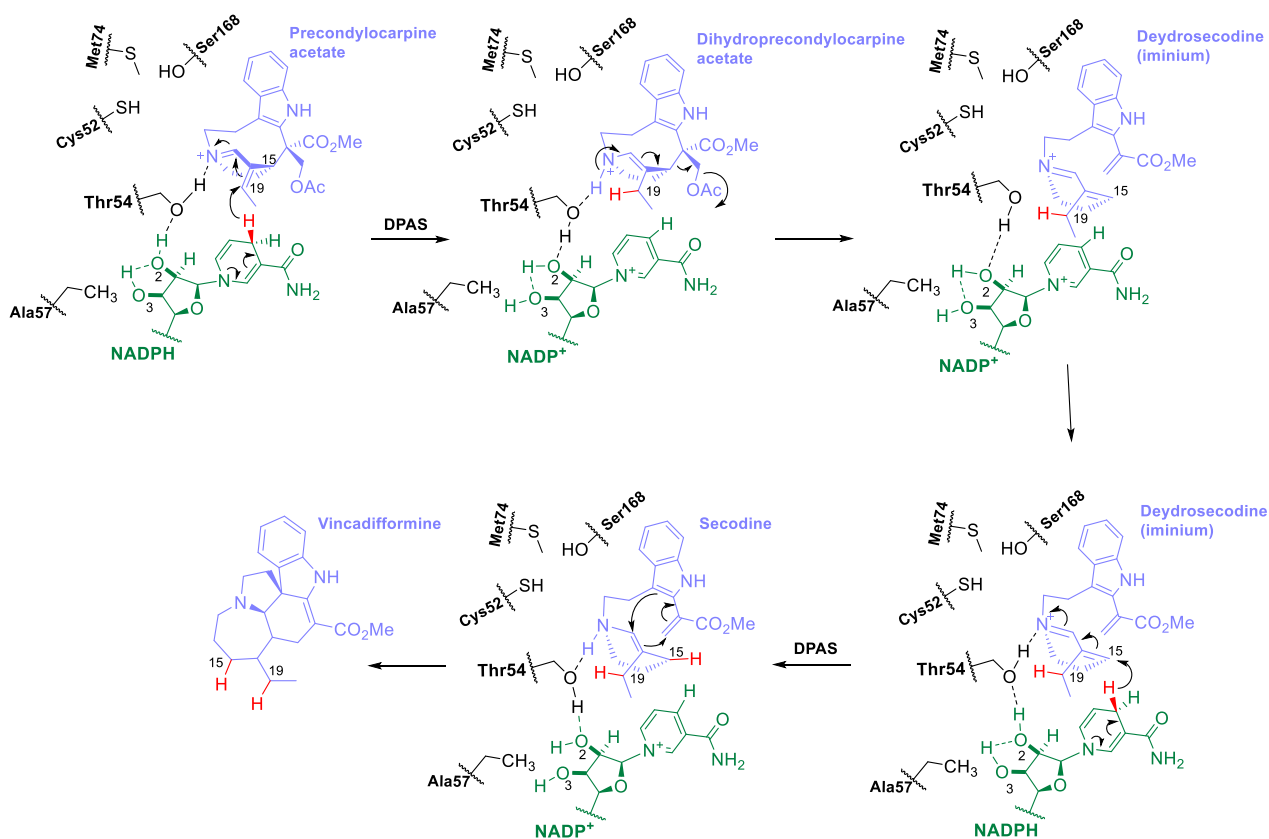


Figure 36. Proposed mechanism of CrDPAS-catalysed reduction of precondylocarpine acetate.

2.2.10 CrDPAS Mutagenesis Study on the Reduction of Strictosidine Aglycone

In addition to its activity with the substrate precondylocarpine acetate, DPAS orthologues from *C. roseus* and *T. iboga* were found to catalyse the 1,4-reduction of the α,β -unsaturated aldehyde vallesiachotamine to form 19,20-dihydrovallesiachotamine (Figure 38A). The reduction likely occurs by a hydride transfer from the NADPH cofactor to C19 of the substrate, mirroring the reduction of precondylocarpine acetate (Figure 36). To understand the structural basis of this reaction, vallesiachotamine was docked into the CrDPAS active site using AutoDock Vina^[38]. CrDPAS Thr54 was revealed to form a hydrogen bond with the substrate, suggesting its potential role in catalysis (Figure 37). Furthermore, the aldehyde of vallesiachotamine physically clashed with the position typically occupied by the catalytic zinc ion in ADHs, though is absent in CrDPAS (Figure 23C).

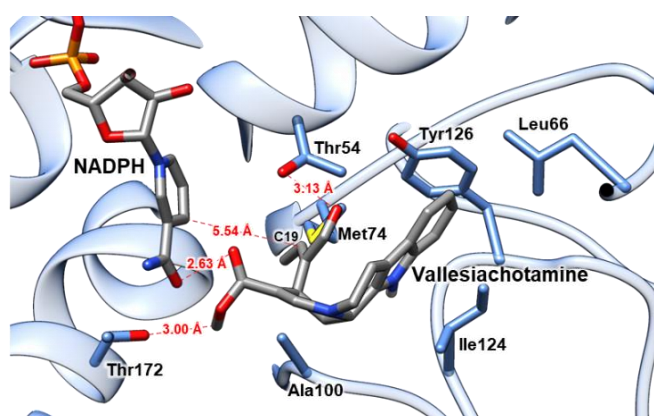


Figure 37. Computational docking of vallesiachotamine and NADPH in CrDPAS substrate pocket using AutoDock Vina^[38].

To probe the role of the two atypical CrDPAS residues in positions that usually coordinate the catalytic zinc, we performed site-directed mutagenesis and tested the activity of the resulting mutant proteins *in vitro* with substrate strictosidine aglycone (generated using the substrate strictosidine in a coupled assay with CrSGD) and cofactor NADPH. Restoring one of the two positions to residues that typically coordinate the catalytic zinc (i.e. CrDPAS Met74His or Ser168Cys) did not affect the formation of 19,20-dihydrovallesiachotamine detected using UPLC-MS (Figure 38B). However, the CrDPAS Met74His Ser168Cys double mutant abolished the production of 19,20-dihydrovallesiachotamine and instead formed tetrahydroalstonine, likely through the 1,2-iminium reduction of the pro-tetrahydroalstonine (Figure 38A)^[15, 16]. These findings suggest that the absence of the catalytic zinc ion is required for the CrDPAS-catalysed 1,4-reduction of vallesiachotamine

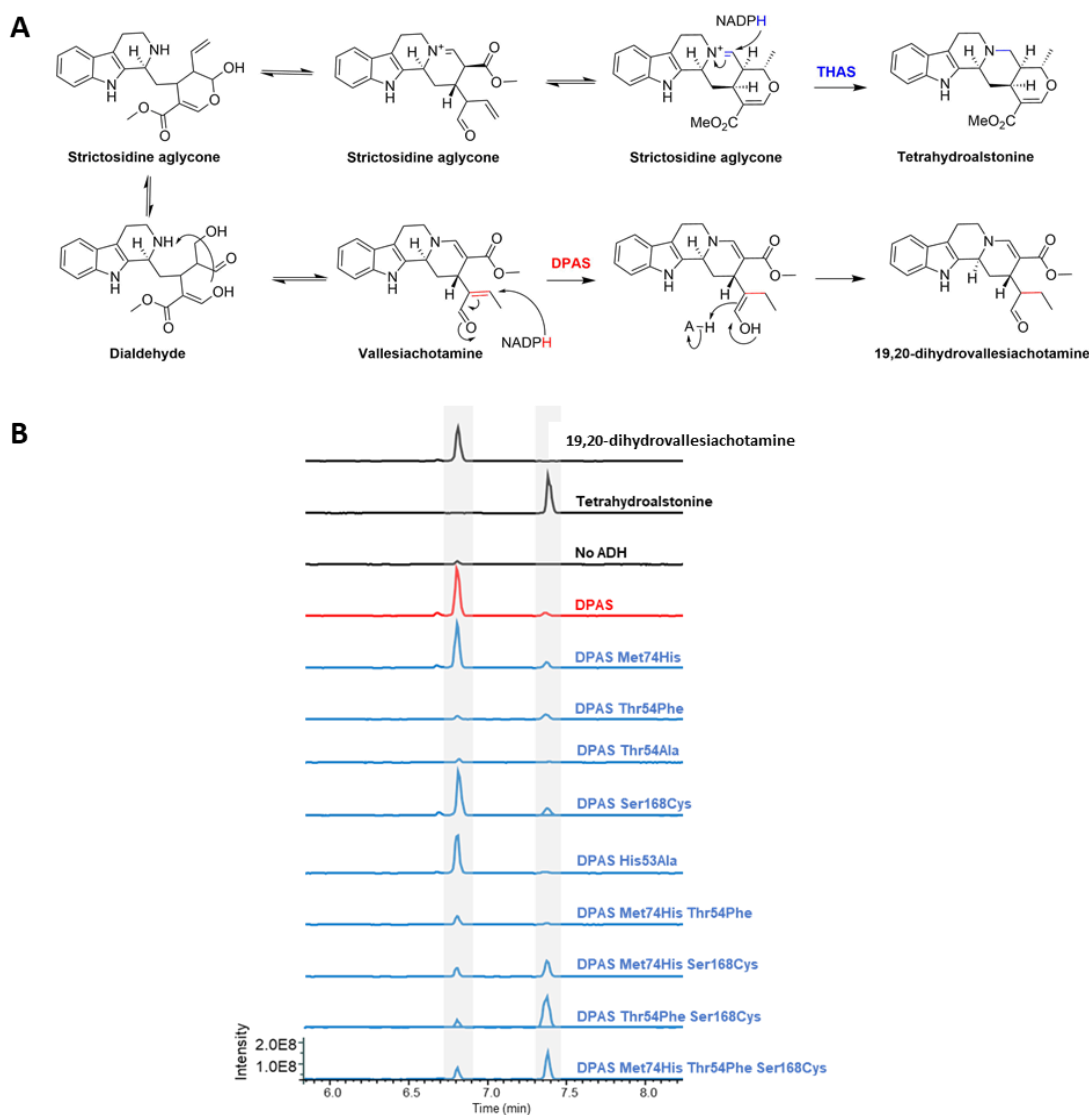


Figure 38. Site-directed mutagenesis of *CrDPAS* on formation of 19,20-dihydrovallesiachotamine. **A.** Proposed pathway of 19,20-dihydrovallesiachotamine and tetrahydroalstonine formation. **B.** UPLC-MS chromatograms of *in vitro* reactions of *CrDPAS* mutants, *CrSGD*, NADPH and substrate strictosidine. EIC m/z 353.185-353.225.

and reintroduction of residues that typically coordinate this ion introduced 1,2-iminium reduction activity. However, the restoration of the catalytic zinc ion could not be validated despite multiple attempts to crystallise *CrDPAS* Met74His Ser168Cys.

As detailed in section 2.2.7, *CrDPAS* and *TiDPAS2* had an atypical pattern of residues involved in binding the cofactor, including those which form a proton relay during catalysis in aldehyde-reducing CADs (i.e. *Cr8HGO*, Figure 33 and Figure 34). To explore the catalytic role of this atypical binding pattern, we conducted site-directed mutagenesis on *CrDPAS*

and observed that although His53Ala had comparable activity to the wild-type enzyme, Thr54Ala and Thr54Phe mutants resulted in reduced product formation (Figure 38B). These findings suggest that while His53 does not have a catalytic role in this reaction, Thr54 is involved in the reduction of vallesiachotamine, likely by the formation of a hydrogen bond with the substrate as observed in the substrate docking analysis (Figure 37).

2.2.11 Mechanism of CrDPAS-Catalysed 1,4-Reduction of Vallesiachotamine

Using the results of the structural and mutagenesis studies of CrDPAS we propose the mechanistic basis of the 1,4-reduction of the α/β -unsaturated aldehyde vallesiachotamine (Figure 39). Strictosidine aglycone can spontaneously rearrange to form various structural isomers, though it is not known whether this rearrangement occurs within the enzyme active site. However, vallesiachotamine binding within the CrDPAS active site is aided by the formation of hydrogen bonds with various residues such as Thr54 and the lack of steric hindrance as a result of the loss of the catalytic zinc ion (Figure 37). Hydride transfer from the NADPH cofactor to C19 of the substrate results in a 1,4-reduction of the unsaturated aldehyde. The hydrogen bond between CrDPAS Thr54 and the aldehyde stabilise the reaction intermediate, enabling the spontaneous rearrangement of the double bond to form 19,20-dihydrovallesiachotamine.

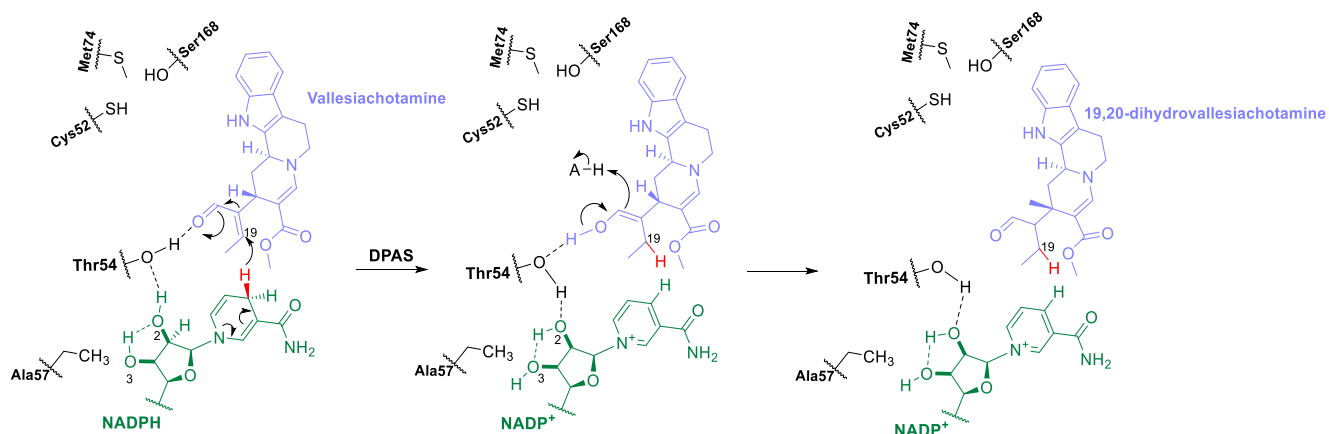


Figure 39. Proposed mechanism of the CrDPAS-catalysed 1,4-reduction of vallesiachotamine.

2.2.12 CrGS and CrTHAS Mutagenesis Study on the Reduction of Strictosidine Aglycone

CrGS catalyses the 1,2-iminium reduction of the structural rearrangement of substrate strictosidine aglycone named 4,21-dehydrogeissoschizine to form geissoschizine (Figure 41) [14]. Interestingly, tetrahydroalstonine was also observed to form as a minor product in

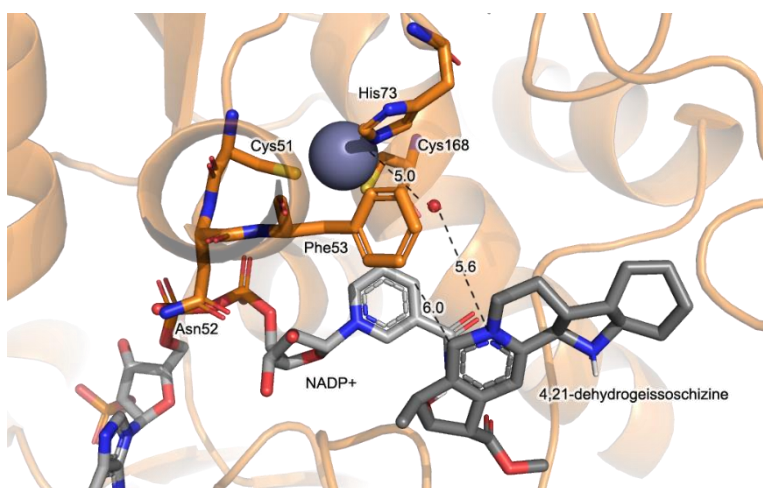


Figure 40. Computational docking of 4,21-dehydrogeissoschizine in CrGS with cofactor NADP⁺. The catalytic zinc ion is coloured in red, the co-crystallised cofactor NADP⁺ in white, and the docked substrate 4,21-dehydrogeissoschizine in grey.

CrGS *in vitro* reactions demonstrating enzyme promiscuity. To understand the structural basis of this reduction, we docked 4,21-dehydrogeissoschizine into the active site of CrGS using AutoDock Vina [38]. This resulted in a substrate pose in which the site of reduction (C21) was 6.0 Å from the hydride donor (Figure 40). Although no hydrogen bonds were observed between the enzyme and substrate, a water molecule was positioned 5.0 Å and 5.6 Å from the catalytic zinc ion and the charged iminium of the substrate, respectively. We therefore hypothesised that this water molecule may coordinate the catalytic zinc during catalysis and form a hydrogen bond with the substrate's iminium moiety to aid orientation within the enzyme pocket.

To understand the mechanism of CrGS-based catalysis, we compared its active site with the aldehyde-reducing Cr8HGO and the 1,4-iminium-reducing CrDPAS. As detailed in Section 2.2.7, we observed differences in residues that coordinate the catalytic zinc and form the proton relay between the substrate and cofactor during catalysis. We subsequently performed site-directed mutagenesis on these residues and tested the *in vitro* activity of these resulting mutants in the presence of CrSGD, cofactor NADPH, and the substrate strictosidine (Figure 41). The residue CrGS Phe53 was in a position that typically contributes to the proton relay during catalysis corresponding to Cr8HGO Ser52 or CrDPAS Thr54, though lacks the necessary hydroxyl-containing side-chain, (Figure 31C and D) [2]. Introducing a hydroxyl group at this position (Phe53Thr) thereby reinstating residues

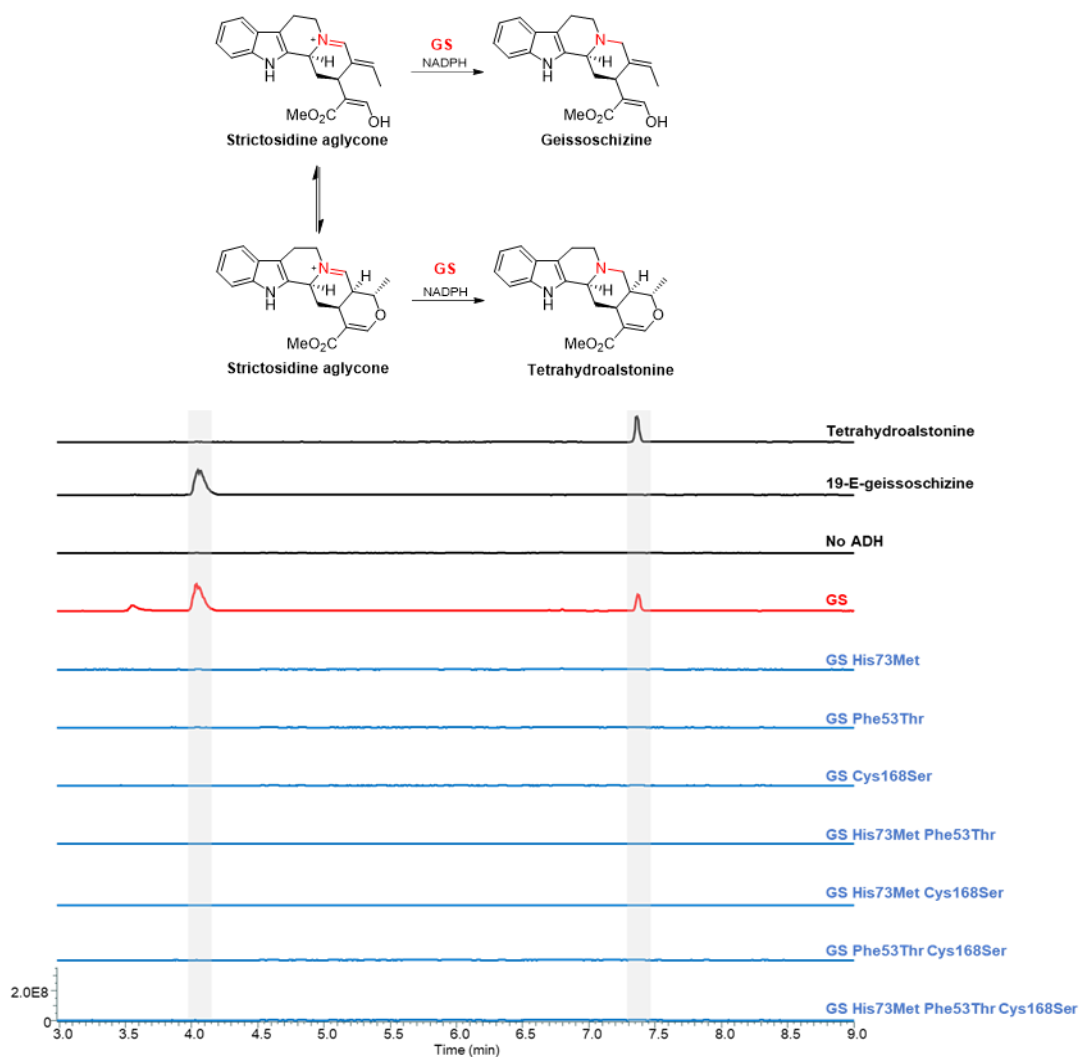


Figure 41. LC-MS chromatograms of *in vitro* coupled reactions of *CrGS* and mutants with substrate strictosidine, *CrSGD* and cofactor NADPH. These mutants probe the role of residues involved in coordination of the catalytic zinc and the proton relay. EIC m/z 353.185-353.225.

required to form the catalytic proton relay abolished *CrGS* activity. These results suggest that the loss of the proton relay system typically observed in aldehyde-reducing CADs is required for the *CrGS*-catalysed 1,2-iminium reduction of 4,21-dehydrogeissoschizine.

Mutagenesis study of the *CrDPAS*-catalysed reduction of strictosidine aglycone revealed that the reaction requires residues that are unable to coordinate the catalytic zinc ion (section 2.2.10, Figure 38). Therefore to probe the role of this ion in the *CrGS*-catalysed reduction of strictosidine aglycone, we mutated these positions to the corresponding residues observed in *CrDPAS* (*CrGS* His73Met and Cys168Ser). This abolished activity with the substrate, suggesting that the *CrGS* catalysed 1,2-iminium reduction of strictosidine aglycone requires the coordination of the catalytic zinc ion.

The mechanism of the closely-related 1,2-iminium-reducing CAD *CrTHAS1* was previously proposed to occur through a proton relay between the cofactor, the residue Tyr56 and the substrate pro-tetrahydroalstonine^[15]. However, *CrTHAS* Tyr56 corresponds to *CrGS* Phe53 which lacks the hydroxyl group required to form the proton relay. To address this mechanistic ambiguity, we generated *CrTHAS* Tyr56Phe and the corresponding *CrGS* Phe53Tyr mutants and tested their *in vitro* activity with the substrate strictosidine, NADPH and enzyme *CrSGD* (Figure 42). We detected comparable reaction products using UPLC-MS between the mutants and their wild-type counterparts. This suggests that this residue and thereby the formation of a proton relay is not involved in either the *CrGS*- or *CrTHAS*-catalysed 1,2-iminium reductions of strictosidine aglycone.

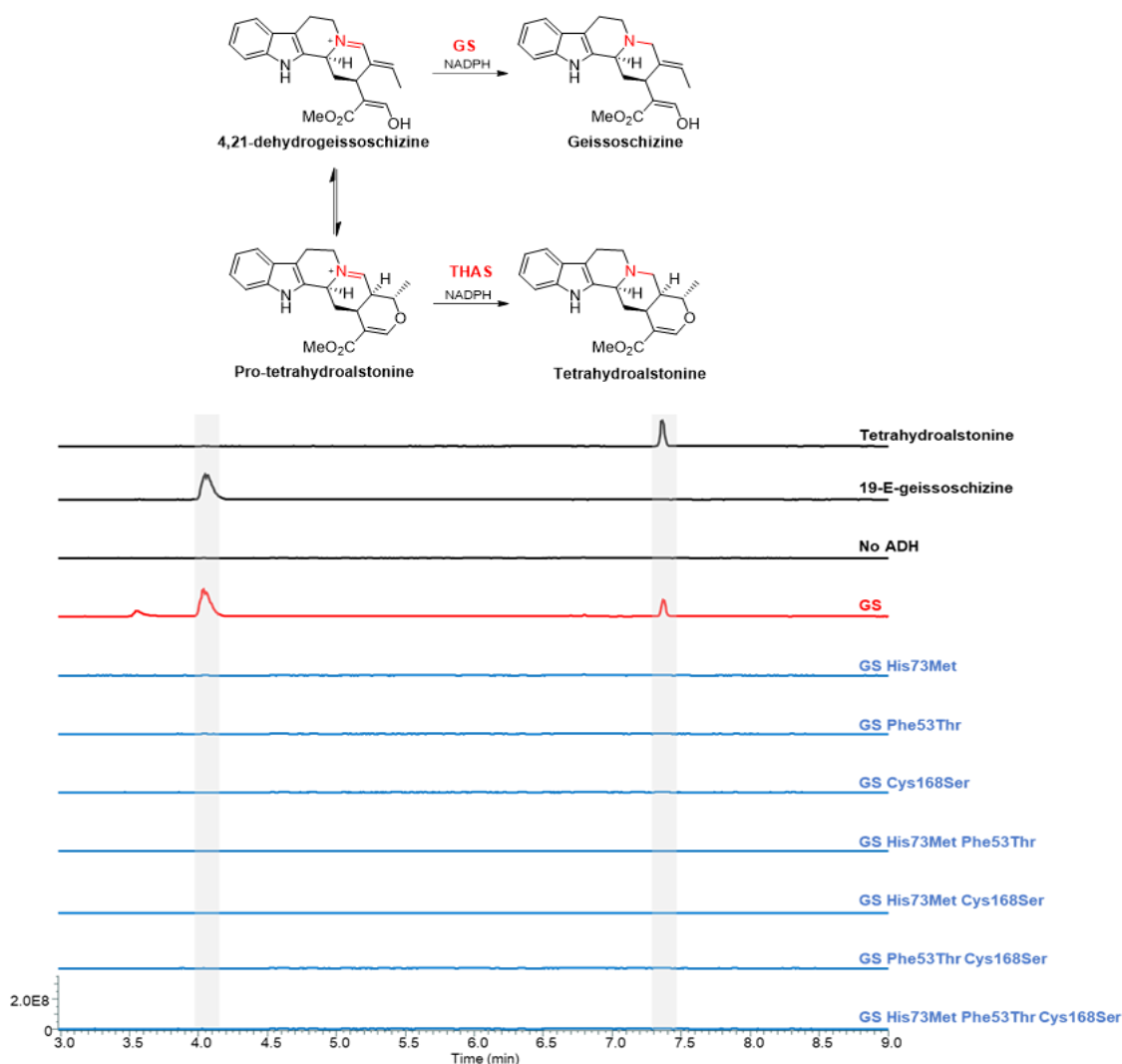


Figure 42. UPLC-MS chromatograms of *CrGS* and *CrGS* Phe53Tyr mutant, and the corresponding *CrTHAS* and *CrTHAS* Tyr56Phe mutant *in vitro* reactions with substrate strictosidine, *CrSGD* and cofactor NADPH. EIC m/z 353.185-353.225.

2.2.13 Mechanism of CrGS-Catalysed 1,2-Iminium Reduction of 4,21-Dehydrogeissoschizine

Findings from the comparative structural investigation and site-directed mutagenesis of CrGS enable us to propose the mechanism of the 1,2-iminium reduction of 4,21-dehydrogeissoschizine (Figure 43). The pattern of CrGS residues that bind the ribose ring of the cofactor is atypical to those observed in aldehyde-reducing CADs such as Cr8HGO, enabling the repositioning of NADPH within the active site. The resulting steric hindrance imposed by the cofactor in addition to residues such as CrGS Phe53 aid substrate orientation within the active site. Furthermore, we postulate that the binding of 4,21-dehydrogeissoschizine may be facilitated by a water molecule that coordinates the catalytic zinc in the fourth position. Instead of being displaced upon substrate binding as proposed in typical CAD-catalysed aldehyde reduction, this water forms a hydrogen bond with the iminium moiety of the substrate. These alterations in the enzyme active site enable 4,21-dehydrogeissoschizine to be orientated in a position conducive to hydride addition from the cofactor to C21 of the substrate. This results in a 1,2-iminium reduction to form geissoschizine.

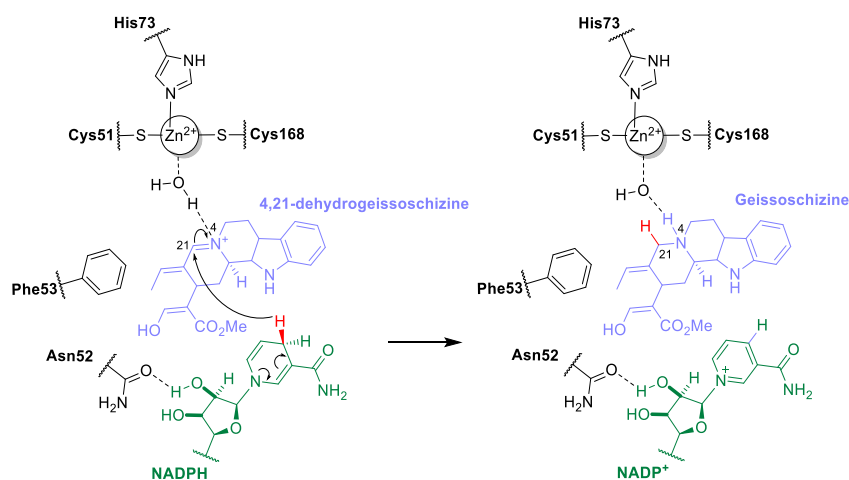


Figure 43. Proposed catalytic mechanism of CrGS catalysed 1,2-iminium reduction of 4,21-dehydrogeissoschizine to form geissoschizine.

2.3 Discussion

2.3.1 Elucidation of *CrDPAS* and *TiDPAS2* Structures Reveals Basis of 1,4-Reductions

DPAS orthologous from *C. roseus* and *T. iboga* are thus far the only ADHs reported to catalyse the 1,4-reduction of an iminium moiety or an α/β -unsaturated aldehyde. The results of isotopic labelling presented in this chapter demonstrate that these enzymes catalyse the 1,4-iminium reduction of the substrate precondylocarpine acetate and dehydrosecodine (Figure 19A). Though the cyclisation of the resulting secodine into vincadifformine was shown by ECD occur outside the enzyme active site (Figure 20). Furthermore, we report that DPAS orthologues from *C. roseus* and *T. iboga* catalyse the 1,4-reduction of the α,β -unsaturated aldehyde vallesiachotamine (Figure 21, Appendix IV) – the first report of this chemistry by an ADH.

To understand the structural basis of these reductions, the structures of *CrDPAS* and *TiDPAS2* were solved by X-ray crystallography. Surprisingly, there was a lack of density corresponding to catalytic zinc ion in both structures (Figure 23C and Figure 24C). The catalytic zinc is highly conserved in ADHs, with only one previous report of its loss in a prokaryote [39, 40]. Closer inspection revealed that both of these orthologues lack residues that are able to coordinate the ion (Figure 29). In addition to the elucidation of *apo-CrDPAS* and *apo-TiDPAS2* structures, the structures of *TiDPAS2* bound to substrates stemmadenine acetate and precondylocarpine acetate were solved (Figure 25 and Figure 26), which were used in conjunction with substrate and cofactor docking studies to identify residues that may be involved in catalysis (Figure 33 and Figure 37).

Findings from comparative structural analysis led us to identify differences in otherwise highly conserved active site residue and subsequently perform site-directed mutagenesis of to probe their catalytic role. These results suggest that the loss of residues that typically coordinate the catalytic zinc ion and the atypical pattern of cofactor binding residues observed in *CrDPAS* are not required for the 1,4-reduction of precondylocarpine acetate (Figure 35). In addition, mutagenesis of *CrDPAS* identified that the loss of coordination of the catalytic zinc ion and the residue Thr54 are required for the 1,4-reduction of the substrate vallesiachotamine, likely due to steric hindrance (Figure 38). These findings enable us to propose the mechanistic basis of these 1,4-reductions (Figure 36 and Figure 39) and expand the chemical repertoire of this class of enzymes.

2.3.2 Reduction of α,β -unsaturated Aldehydes in Plant Specialised Metabolism

The DPAS-catalysed formation of 19,20-dihydrovallesiachotamine is the first report of a 1,4-reduction of an α,β -unsaturated aldehyde by an ADH. Phylogenetic comparison of DPAS orthologues to other previously characterised CADs identified Vomilenine Reductase 2 (VR2) orthologues in *Rauwolfia tetraphylla* and *Rauwolfia serpentina* (70.9% and 71.6% nucleic acid sequence identity to *CrDPAS* respectively; Figure 44) [27]. VR2 reduces the substrate vomilenine through an unknown mechanism to form the MIA 19,20- $\alpha(S)$ -dihydrovomilenine. Sequence analysis of the VR2 orthologues identified atypical residues in positions that coordinate the catalytic zinc, mirroring observations in *CrDPAS* and *TiDPAS2* (Figure 23C and Figure 24C). These similarities in sequence identity and motifs enable us to propose suggest that VR2 catalyses the 1,4-reduction of the α,β -unsaturated aldehyde of the substrate vomilenine in an analogous mechanism as described for *CrDPAS* (Figure 46). If supported by future isotopic labelling experiments, this finding would suggest that this reduction mechanism arose in Apocynaceae and was maintained in *C. roseus*, *T. iboga*, *R. tetraphylla* and *R. serpentina*, giving insights into the evolution of these atypical CADs.

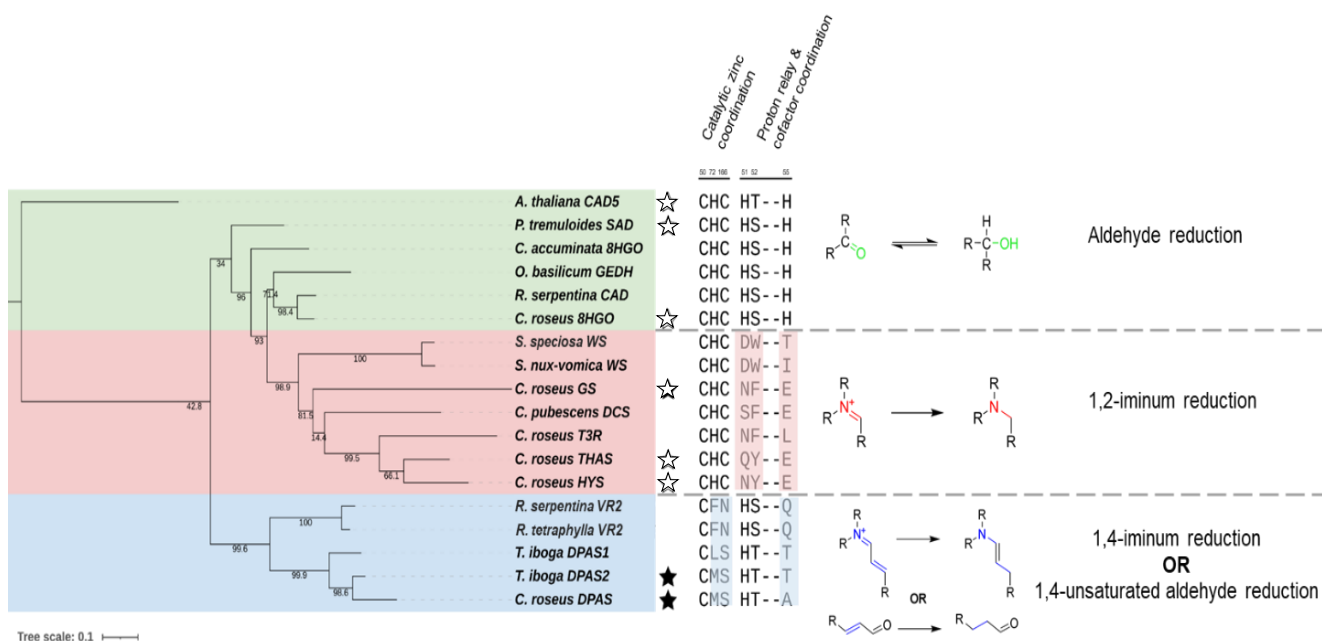


Figure 44. Tree of maximum likelihood of previously characterised plant CADs. Reduction chemistries and residues involved in the coordination of the catalytic zinc and forming the proton relay are shown. Residue numbering based on *Cr8HGO*, stars indicate proteins with structures solved in either previous work (line) or in this thesis (filled).

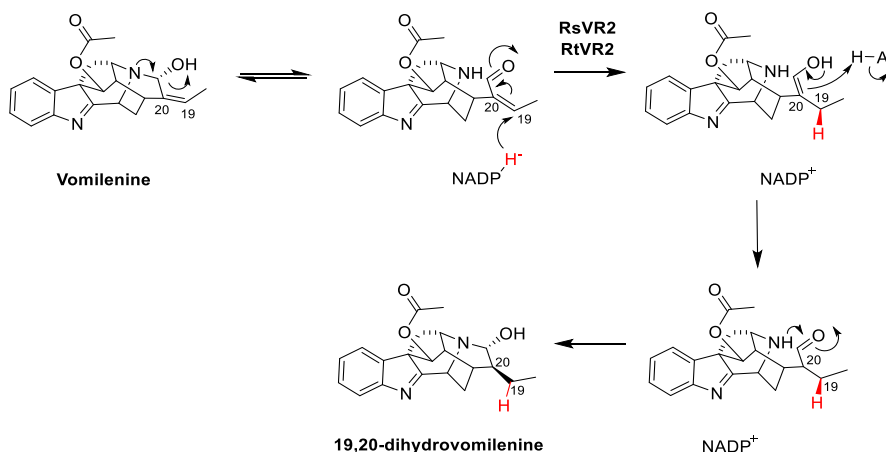


Figure 46. Proposed mechanism of *Rauwolfia* VR2 catalysed 1,4-reduction of α,β -unsaturated aldehyde vomilenine.

In addition to the ADHs DPAS and VR, the 1,4-reduction of an α,β -unsaturated aldehyde in plant specialised metabolism has been reported to be catalysed by the SDR *C. roseus* iridoid synthase (*CrISY*) in iridoid biosynthesis (Figure 45D) [41]. ADHs and SDRs are distantly related enzyme families resulting in the conservation of several structural features important for catalysis such as the cofactor binding Rossmann fold (Figure 45A and C) [42, 43]. Notably, the hydroxyl group of *CrISY* Tyr178 was found to provide the enolate hydrogen during the

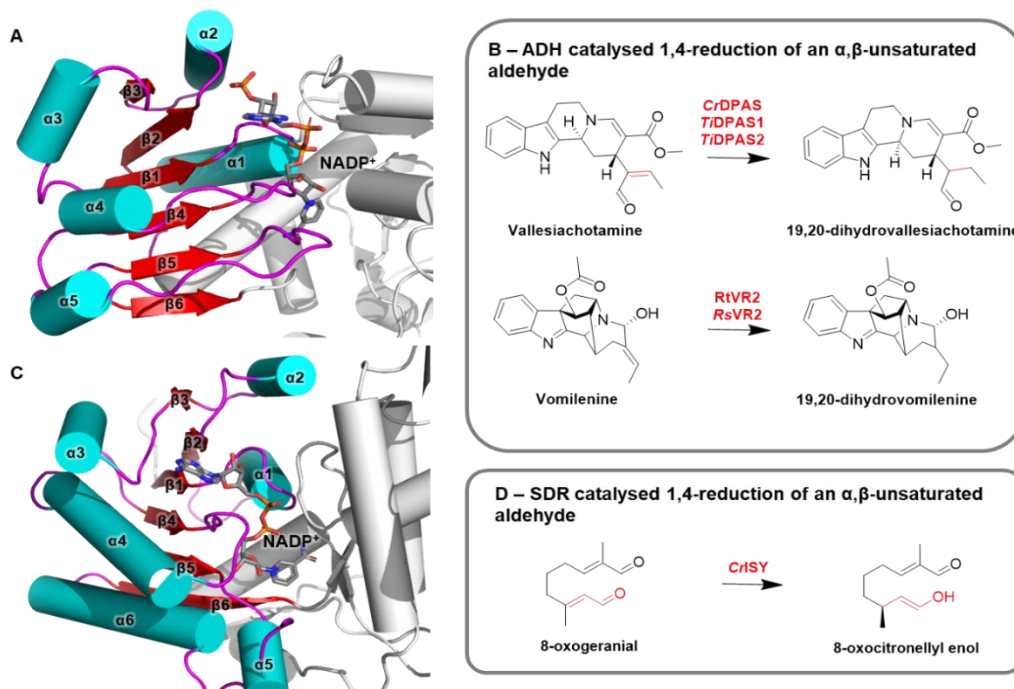


Figure 45. Enzyme-catalysed reductions of α,β -unsaturated aldehydes in plant specialised metabolism. **A.** Rossmann-fold of *TiDPAS2* bound to NADP^+ . **B.** DPAS and VR catalysed 1,4-reduction of vallesiachotamine and vomilenine. **C.** Rossmann-fold of *CrISY* bound to NADP^+ (PDB 5DCY). **D.** *CrISY* catalysed 1,4-reduction of 8-oxogeranial.

catalysis of 8-oxogeranial (Figure 47) ^[44], fulfilling a similar catalytic role to *CrDPAS* Thr54 in the reduction of vallesiachotamine (Figure 39). These findings suggest that the 1,4-reduction of α,β -unsaturated aldehydes have emerged in SDRs and ADHs by convergent evolution aided by the conserved cofactor-binding Rossmann fold.

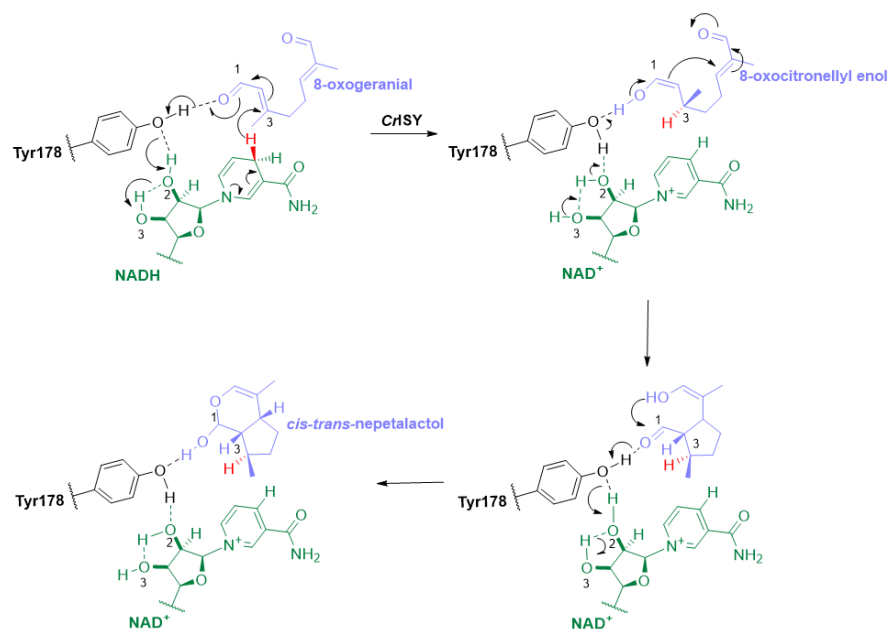


Figure 47. Mechanism of *CrISY* catalysed reduction of 8-oxogeranial to form 8-oxocitronellyl enol which spontaneously cyclises to form *cis-trans-nepetalactol*. Figure adapted from Hu *et al.*, ^[44].

2.3.3 Mechanism of *CrGS*-Catalysed 1,2-Iminium Reduction of 4,21-Dehydrogeissoschizine

Several CADs perform the 1,2-reduction of an iminium moiety in MIA biosynthesis, including the *CrGS* catalysed-reduction of 4,21-dehydrogeissoschizine (Figure 14). To understand the structural basis of this reduction chemistry, we *CrGS* and the aldehyde-reducing *Cr8HGO* and observed an atypical pattern of residues which were predicted to prevent the formation of the proton relay during catalysis and result in the repositioning of NADPH within the active site (Figure 31). Reintroduction of residues in *CrGS* to those able to form the proton relay abolished formation of geissoschizine, suggesting that the loss of this otherwise conserved catalytic feature is required for 1,2-iminium reduction (Figure 41). We additionally propose that a water molecule coordinates both the catalytic zinc and 4,21-dehydrogeissoschizine during catalysis to enable the 1,2-reduction of the iminium moiety (Figure 43).

2.4 Conclusions

CADs involved in MIA biosynthesis have neofunctionalised to accept non-aromatic substrates and to catalyse the atypical 1,2- or 1,4-reduction of an iminium moiety or the 1,4-reduction of an α,β -unsaturated aldehyde. This chapter investigates the structural and mechanistic basis of these reactions in this otherwise highly conserved family of MDRs.

CrDPAS is shown through isotopic labelling and the subsequent structural characterisation to catalyse the 1,4-iminium reduction of the substrates precondylocarpine acetate and dehydrosecodine. Furthermore, we report that DPAS orthologues from *C. roseus* and *T. iboga* catalyse the 1,4-reduction of the α,β -unsaturated aldehyde vallesiachotamine. These are the first reports of these chemistries to be catalysed by an ADH, expanding the chemical repertoire of this enzyme family. Structural elucidation of *CrDPAS* and *TiDPAS2* identified changes in otherwise highly conserved residues within the enzyme active site – namely, the loss of coordination of the catalytic zinc ion, and atypical residues involved in cofactor binding. In conjunction with site-directed mutagenesis, these findings enable us to propose the mechanisms of the DPAS-catalysed 1,4-reduction of precondylocarpine acetate, dehydrosecodine and vallesiachotamine.

Furthermore, a comparison of the structures of the 1,2-iminium-reducing *CrGS* with the aldehyde-reducing *Cr8HGO* reveals key differences between the enzyme's active sites. Findings of site-directed mutagenesis suggest that the formation of geissoschizine requires the repositioning of the cofactor and the resulting loss of the typical proton relay. These findings enable us to propose the mechanism for the *CrGS*-catalysed 1,2-iminium reduction of 4,21-dehydrogeissoschizine and provide the basis of 1,2-reductions catalysed by CADs in MIA biosynthesis.

The work presented in this chapter highlights the catalytic plasticity of this otherwise highly conserved class of enzyme within MIA biosynthesis and demonstrates their potential applications in bioengineering and enzyme discovery efforts.

The results in this chapter have been published in Langley *et al.*,^[21].

2.5 Methods and Materials

2.5.1 Chemicals and molecular biology reagents

All solvents used for extractions, chemical synthesis and preparative HPLC were HPLC grade, and solvents used for UPLC/MS were MS grade. All solvents were purchased from Sigma Aldrich. Carbenicillin, kanamycin sulfate, isopropyl β -D-thiogalactoside (IPTG) salts were purchased from Sigma. Synthetic genes were purchased from IDT. All gene amplifications and mutations were performed using Platinum II Superfi DNA Polymerase (Thermo Fisher). Constructs were transformed into vectors using In-Fusion kit (ClonTech Takara) and colony PCR was performed using Phire II mastermix (Thermo Fisher) according to manufacturer's instructions. PCR product purification was performed using Zymoclean Gel DNA Recovery kit (Zymo). Plasmid purification was performed using the Wizard Miniprep kit (Promega). Strictosidine, precondylocarpine acetate, stemmadenine acetate, angryline, vincadifformine, 19-*E*-geissoschizine and tetrahydroalstonine were enzymatically prepared and purified as previously described [16, 19, 22, 45].

2.5.2 Cloning and mutagenesis

Cloning of *CrDPAS*, *TiDPAS1*, *TiDPAS2*, *CrGS* and *CrTHAS* has been previously reported [14, 16, 19, 22]. Full-length *CrDPAS*, *TiDPAS2*, *GS* and *THAS* were amplified by PCR from the codon optimized synthetic genes listed in Table 3 using corresponding primers listed in Table 2. *DPAS*, *GS* and *THAS* mutants were generated by overlap extension PCR as previously reported [31]. PCR products were purified from 1% agarose gel and ligated into the BamHI and KPN1 restriction sites of pOPINK vectors for small-scale expression of *CrGS* and *CrGS* mutants. All other ADHs were cloned into the pOPINF vector. pOPINF and pOPINK were a gift from Ray Owens (Addgene plasmid #26042 and #41143 [46]). Constructs were ligated into vectors using the In-Fusion kit (Clontech Takara).

Table 2. Primer sequences used for gene amplification and site-directed mutagenesis. Cloning overhangs are underlined. Mutated codons are in bold.

Primers for full length gene amplification	
CrDPAS_Fwd	<u>AAGTTCTGTTTCAGGGCCCGATGGCAGGTAAAAGCGCAGAAGAAG</u>
CrDPAS_Rev	<u>ATGGTCTAGAAAGCTTTACAGTTCGCTAGGCGGTGTCAG</u>
TiDPAS1_Fwd	<u>AAGTTCTGTTTCAGGGCCCGATGGCAGTTAAGTCACCAGAAG</u>
TiDPAS1_Rev	<u>ATGGTCTAGAAAGCTTTACTCAGGGGGCGTAAGGGTGTTA</u>
TiDPAS2_Fwd	<u>AAGTTCTGTTTCAGGGCCCGATGGCGGGCAAATCCCCGAAG</u>
TiDPAS2_Rev	<u>ATGGTCTAGAAAGCTTTACGGTTCTGGAGGCGGAGTCAAAG</u>
CrGS_Fwd	<u>AAGTTCTGTTTCAGGGCCCGATGGCTGGTGAAACCACCAAAC</u>
CrGS_Rev	<u>ATGGTCTAGAAAGCTTTATTCTTCGAATTCAGGGTGTTAC</u>
CrTHAS_Fwd	<u>AAGTTCTGTTTCAGGGCCCGCAATGGCTTCAAAGTCACCTTCTG</u>
CrTHAS_Rev	<u>ATGGTCTAGAAAGCTTTAATTTGATTCAGAGTGTTT</u>
Primers for mutagenesis	
CrDPAS_M74H_Fwd	TATCCGCTGGTTCCTGGT CAT GAAATTGTTGGTATTGCAAC
CrDPAS_M74H_Rev	ACCAGGAACCAGCGGATAGCTCAG
CrDPAS_T54F_Fwd	GTATTGCGGCATTTGTCAT TC GATCTGGCAAGCATTAAAAAC
CrDPAS_T54F_Rev	ATGACAAATGCCGCAATACAGAATTTTG
CrDPAS_S168C_Fwd	GGTGCTCCGCTGCTGT GT GCAAGTATTACCAGCTTTAG
CrDPAS_S168C_Rev	CAGCAGCGGAGCACCGCCTGC
CrDPAS_T54A_Fwd	TATTGCGGCATTTGTCAT GCC GATCTGGCAAGCATTAAAAAC
CrDPAS_T54A_Rev	ATGACAAATGCCGCAATACAGAATTTTG
CrDPAS_H53A_Fwd	TGTATTGCGGCATTTGT GCT ACCGATCTGGCAAGCATT
CrDPAS_H53A_Rev	ACAAATGCCGCAATACAGAATTTTGA

CrGS_H73M_Fwd	TACCCGTACGTTTTCGGT ATGG AAACCGCTGGTGAAGTTGT
CrGS_H73M_Rev	ACCGAAAACGTACGGGTAACGGGT
CrGS_F53T_Fwd	GTA CT CTGGTGT TTG CAAC ACCG ACATGGAAATGGTTCGTAAC
CrGS_F53T_Rev	GTTGCAAACACCAGAGTACAGAACACGG
CrGS_C168S_Fwd	GGTGT TG CTCTGCTG AGCG CTGGTGT TG TTGTTACTC
CrGS_C168S_Rev	CAGCAGAGCAACACCTTTGTC
CrGS_F53Y_Fwd	TACTCTGGTGT TTG CAACT ACG ACATGGAAATGGTTCGT
CrGS_F53Y_Rev	GTTGCAAACACCAGAGTACAGAACACGG
CrTHAS_Y55F_Fwd	GTGGGACTTGCCA ATTTG ACAGGGAAATGAG
CrTHAS_Y55F_Rev	TTGGCAAGTCCACAGTATAATAC

Table 3. Full length nucleotide sequences of genes.

Codon optimised CrDPAS	ATGGCAGGTAAAAGCGCAGAAGAAGAACATCCGATTAAAGCATATGGTTGGG CAGTTAAAGATCGTACCACCGGTATTCTGAGCCCCTTTAAATTCAGCCGTCGTG CAACCGGTGATGATGATGTTCTGATCAAAATCTGTATTGCGGCATTTGTCATA CCGATCTGGCAAGCATTAAAAACGAATATGAGTTTCTGAGCTATCCGCTGGTTC CTGGTATGGAAATTGTTGGTATTGCAACCGAAGTTGGTAAAGATGTGACCAAA GTTAAAGTGGGTGAAAAAGTTGCACTGAGCGCATATCTGGGTTGTTGTGGTAA ATGTTATAGCTGCGTGAATGAGCTGGAAAATATTGTCCGGAAGTGATTATTG GTTATGGCACCCCGTATCATGATGGCACCATTTGTTATGGTGGTCTGAGCAATG AAACCGTTGCAAATCAGAGCTTTGTTCTGCGTTTTCCGGAACGTCTGAGTCCGG CAGGCGGTGCTCCGCTGCTGAGCGCAGGTATTACCAGCTTAGCGCAATGCGT AATAGCGGTATTGATAAACCGGGTCTGCATGTTGGTGTGTTGGTTTAGGTGGT CTGGGTCATCTGGCCGTTAAATTTGCAAAGCATTGGTCTGAAAGTGACCGTT ATTAGCACACACCGAGCAAAAAAGATGATGCAATTAATGGCCTGGGTGCAGA TGGTTTTCTGCTGAGCCGTGATGACGAGCAGATGAAAGCAGCAATTGGCACCC TGGATGCCATTATTGATACCCTGGCAGTTGTTTCATCCGATTGCACCGCTGCTGG ATCTGCTGCGTAGCCAGGGTAAATTTCTGCTGCTGGGTGCACCGAGCCAGAGC CTGGAAGTGCCTCCGATTCCTCTGCTGAGTGGTGGTAAAGCATTATTGGTAGC GCAGCAGGTAATGTTAAACAGACCCAAGAAATGCTGGATTTTGCAGCCGAACA TGATATTACCGCAACGTTGAAATTATCCCGATCGAATACATTAACACCGCAAT GGAACGCCTGGATAAAGGTGATGTGCGTTATCGTTTTGTGGTGGATATTGAAA ATACCCTGACACCGCCTAGCGAACTGTAA
------------------------	--

<p>Codon optimised <i>TiDPAS2</i></p>	<p>ATGGCGGGCAAATCCCCGAAGAAGAGCACCCGGTAAAAGCCTATGGCTGGG CAGTGAAAGACCGCACCCACGGGTATTCTGTCTCCATTCAAGTTCTCACGCCGGG CAACAGGTGATAATGACATTGCGATTAATAATTCTTTACTGCGGGATTTGCCATA CTGACTTGACATCAGTGAAAAACGAATACGAGTTCCTTTCATACCCCTTAGTACC GGGTATGGAGATCGTTGGTATCGCTACTGAAGTGGGAAGTAAGGTTACGAAA ATCAAGTTGGGGAGAAAGTGGCGGTGGCCGCGTACTTGGGGACTTGTGGTA AGTGTATAATTGCGTGAACGATCTTGAANAATACTGTCCCAGGTCATTATTG GATACGGTACTCCATATCACGACGGGACGATTAATACTATGGAGGACTTTCGAAC GAGACCGTAGTGAATGAGAGATTTGTCTTGCCTTTCCGGAAAAGCTTTCGCCG GCTGGCGGAGCTCCTCTGTTAAGCGCAGGGATTACCGCATACTCGGCTATGCG TAACCATGGACTTGACAAGCCAGGTATCCATCTTGGTGTGTCGGGCTTGGAG GACTGGGGCACCTTGCTGTTAAGTTTGTAAAGCGTTTGGCGTTCGTGTCACCG TCATCTACTACGCCTTCTAAAAAAGATGAGGCTATCAATAACCTGGGCGCGG ACGCCTTCTTATTCAGCCGGGACGATAAGCAGATGCGGGCTGCCATCGGTACC TTTGATGCCATAATCGATACGTTGGCGGTGGTCCATCCCATCGCACCATTATTA GATTTATTGCGTAGCCATGGTAAGCTGGTCTTGTGGCGCTCCGAGTAAGCCT TTAGAATTGCCAACAAATCCCACTACTGAGCGGGGGGAAGTCGCTTATAGGCTCC GCAGCAGGGAATGTCAAGCAAACCTCAGGAGATGTTGGACTTTGCTGCTGAACA TGATATCACCGCCAACATAGAGGTAATACCTATTGACTATATCAATACAGCCAT GGAACGTCTTGATAAGGGAGATATACGCTTTCGTTTCGTTGTAGACATCGAAAA CACTTTGACTCCGCCTCCAGAACCGTAA</p>
<p>Codon optimised <i>CrGS</i></p>	<p>ATGGCTGGTGAACCACCAAACCTGGACCTGTCTGTTAAAGCTGTTGGTTGGGG TGCTGCTGACGCTTCTGGTGTCTGCAGCCGATCAAATTCTACCGTCGTGTTCCG GGTGAACGTGACGTTAAATCCGTGTTCTGTAATCTGGTGTGTTGCAACTTCGAC ATGGAAATGGTTCGTAACAAATGGGGTTTACCCGTTACCCGTACGTTTTCGGT CACGAAACCGCTGGTGAAGTTGTTGAAGTTGGTTCTAAAGTTGAAAAATTCAA AGTTGGTGACAAAGTTGCTGTTGGTTGCATGGTTGGTCTTGCGGTCAGTGCTA CAACTGCCAGTCTGGTATGGAAAATACTGCCCAGAACCGAACATGGCTGACG GTTCTGTTTACCGTGAACAGGGTGAACGTTCTACGGTGGTTGCTCTAACGTTA TGGTTGTTGACGAAAATTCGTTCTGCGTTGGCCGAAAACCTGCCGCAGGAC AAAGGTGTTGCTCTGCTGTGCGCTGGTGTGTTGTTGTTACTCTCCGATGAAACAC CTGGGTCTGGACAAACCGGGTAAACACATCGGTGTTTTCGGTCTGGGTGGTCT GGGTTCTGTTGCTGTTAAATTCATCAAAGCTTTCGGTGGTAAAGCTACCGTTAT CTCTACCTCTCGTCGTAAGAAAAAGAAGCTATCGAAGAACACGGTGTGCTGACG CTTTCGTTGTTAACACCGACTCTGAACAGCTGAAAGCTCTGGCTGGTACCATGG ACGGTGTGTTGACACCACCCGGGTGGTCGTACCCCGATGTCTCTGATGCTGA ACCTGCTGAAATTCGACGGTGTGTTATGCTGGTTGGTGTCCGGAATCTCTGT TCGAACTGCCGGCTGCTCCGCTGATCATGGTTCGTAAAAAATCATCGTTCTT CTACCGTGGTCTGAAAGAATACCAGGAAATGCTGGACTTCGCTGCTAAACAC AACATCGTTTGCACACCGAAGTTATCGGTATCGACTACCTGTCTACCGCTATG GAACGTATCAAAAACCTGGACGTTAAATACCGTTTCGCTATCGACATCGGTAAC ACCCTGAAATTCGAAGAATAA</p>
<p>Codon optimised <i>CrTHAS</i></p>	<p>ATGGCAATGGCTTCAAAGTCACCTTCTGAAGAAGTATATCCAGTGAAGGCATTT GGTTTGGCTGCTAAGGATTCTTCTGGGCTTTTCTCTCCATTCAACTTCTCAAGAA GGGCCACAGGGGAACACGATGTGCAGCTCAAAGTATTATACTGTGGGACTTGC</p>

	<p>CAATATGACAGGGAAATGAGCAAAAACAAATTTGGATTTACAAGCTATCCTTAT GTTTTAGGGCATGAAATTGTGGGTGAGGTAAGTGAAGTTGGCAGCAAGGTGC AGAAATTCAAAGTCGGGGACAAAGTGGGCGTAGCAAGCATAATTGAACTTGT GGCAAATGTGAAATGTGTACAAATGAAGTTGAAAATTACTGTCCAGAAGCAGG ATCAATAGACAGCAATTACGGGGCATGTTCAAATATAGCAGTGATAAACGAGA ATTTTGCATCCGTTGGCCTGAAAATCTTCCTTTGGATTCTGGTGTTCCTCTTCTA TGTGCAGGAATCACGGCTTATAGTCCCATGAAACGTTATGGACTTGATAAACCT GGAAAACGTATCGGCATAGCCGGTCTAGGAGGACTTGGACATGTAGCTCTTAG ATTTGCCAAAGCTTTTGGGGCTAAGGTGACAGTGATTAGTTCCTCACTTAAGAA AAAACGTGAAGCCTTTGAGAAATTCGGAGCAGATTCTTTCTGGTCAGCAGTAA TCCAGAAGAAATGCAGGGTGCAGCAGGAACATTGGATGGGATCATAGACACT ATACCAGGGAATCACTCTCTTGAGCCACTCCTTGCTTTATTGAAGCCTCTGGGA AGCTTATCATTTTAGGTGCACCAGAAATGCCCTTTGAGTTCCCGCTCCTTCCCT GCTTATGGGTGGAAAAGTAATGGCTGCCAGTACTGCTGGGAGTATGAAGGAA ATACAAGAGATGATTGAATTTGCAGCAGAACAACATAGTAGCAGATGTGGA GGTTATCTCTATTGACTATGTGAACACTGCAATGGAGCGCCTTGATAACTCTGA TGTGAGATATCGTTTCGTGATTGATATAGGGAACACTCTGAAATCAAATTA</p>
<i>TbADH</i>	<p>AAGTTCTGTTTCAGGGCCCGAAAGGTTTTGCAATGCTCAGTATCGGTAAAGTTG GCTGGATTGAGAAGGAAAAGCCTGCTCCTGGCCATTTGATGCTATTGTAAGA CCTCTAGCTGTGGCCCTTGCACTTCGGACATTCATACCGTTTTTGAAGGCGCCA TTGGCGAAAGACATAACATGATACTCGGTCACGAAGCTGTAGGTGAAGTAGTT GAAGTAGGTAGTGAGGTAAGATTTTAAACCTGGTGATCGCGTTGTTGTGCC AGCTATTACCCCTGATTGGCGGACCTCTGAAGTACAAAGAGGATATCACCAGC ACTCCGGTGGAATGCTGGCAGGCTGGAATTTTGAATGTAAGATGGTGT TTTGGTGAATTTTTTCATGTGAATGATGCTGATATGAATTTAGCACATCTGCCTA AAGAAATTCATTGGAAGCTGCAGTTATGATTCCCGATATGATGACCACTGGTT TTCACGGAGCTGAACTGGCAGATATAGAATTAGGTGCGACGGTAGCAGTTTTG GGTATTGGCCCAGTAGGTCTTATGGCAGTCGCTGGTGCCAAATTGCGTGGAGC CGGAAGAATTATTGCCGTAGGCAGTAGACCAGTTTGTGTAGATGCTGCAAAT ACTATGGAGCTACTGATATTGTAACTATAAAGATGGTCCTATCGAAAGTCAGA TTATGAATCTAACTGAAGGCAAAGGTGTCGATGCTGCCATCATCGCTGGAGGA AATGCTGACATTATGGCTACAGCAGTTAAGATTGTTAAACCTGGTGGCACCATC GCTAATGTAAATTTTTGGCGAAGGAGAGTTTTGCCTGTTCTCGTCTTGAA TGGGGTTGCGGCATGGCTCATAAACTATAAAAGGCGGGCTATGCCCGGTGG ACGTCTAAGAATGGAAAGACTGATTGACCTTGTTTTTATAAGCGTGTGATCC TTCTAAGCTCGTCACTCACGTTTTCCGGGGATTTGACAATATTGAAAAGCCTTT ATGTTGATGAAAGACAAACCAAAGACCTAATCAAACCTGTTGTAATATTAGCA TAAAGCTTCTAGACCAT</p>

2.5.3 Protein Expression in *E. coli*

Constructs were transformed into chemically-competent *E. coli* Stellar cells (Clontech Takara) by heat shock at 42°C for 30 seconds and selected on LB agar containing 50µg/mL carbenicillin or kanamycin for pOPINF or pOPINK constructs respectively. Positive colonies were screened by colony PCR using primers listed in Table S1 and grown overnight at 37°C

shaking at 200 r.p.m. Plasmids were then isolated and constructs were sequence verified. Plasmids were transformed into chemically competent *E. coli* SoluBL21 cells by heat shock for 30 seconds at 42°C and selected on LB agar containing 50 µg/mL carbenicillin or kanamycin for pOPINF or pOPINK constructs respectively. For small scale protein purification, 10 mL starter cultures of LB with 50 µg/mL of the respective antibiotic and a colony of transformed construct in SoluBL21 cells were grown at 37°C 200 r.p.m. overnight. Media (100 mL 2xYT media) containing 50 µg/mL antibiotic was inoculated with 1 mL of the starter culture and grown until OD₆₀₀ of 0.6 was reached. For large scale purification, 20 mL starter cultures of LB with antibiotic and a colony of transformed construct in SoluBL21 cells were grown at 37°C 200 r.p.m. overnight. Media (1L 2xYT media) containing 50 µg/mL carbenicillin was inoculated with 10 mL of starter culture and grown until OD₆₀₀ of 0.6 was reached. Once cultures had reached the desired OD₆₀₀, cultures were transferred to 18°C 200 r.p.m shaking incubator for 30 minutes before protein expression was induced by addition of 300 µM IPTG, after which cultures were grown for an additional 16 hours.

2.5.4 CrPAS Insect Cell Expression

N-terminal His₆-tagged CrPAS was expressed in Sf9 insect cells as previously described ^[22]. Cells were harvested by centrifugation and the pellets frozen at –80°C until large-scale purification.

2.5.5 Small-scale Protein Purification

Cells were harvested by centrifugation at 4000 x *g* for 15 minutes and re-suspended in 10 mL buffer A1 (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) with addition of EDTA-free protease inhibitor cocktail (Roche Diagnostics Ltd.) and 10 mg lysozyme (Sigma). Cells were lysed at 4 °C using a sonicator (40% amplitude, 2 seconds on, 3 seconds off cycles for 2 minutes) and centrifuged at 35000 x *g* to remove insoluble cell debris. The supernatant was collected and filtered with 0.2 µm PES syringe filter (Sartorius) and purified by addition of 150 µL washed Ni-NTA agarose beads (QIAGEN). Samples were incubated on a rocking incubator at 4 °C for 1 hour. Beads were washed by centrifuging at 1000 x *g* for 1 minute to remove the supernatant, and then the beads were resuspended in 10 mL of A1 Buffer. This step was performed a total of three times. Protein was eluted by resuspending the beads in 600 µL of buffer B1 (50 mM Tris-

HCl pH 8.0, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole) before centrifuging for 1000 x *g* for 1 minute and then collecting the supernatant. This elution step was repeated to remove all Ni-NTA bound protein. Proteins were buffer exchanged into buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl) and concentrated using 10K Da molecular weight cut off centrifugal filter (Merck) and stored at –80 °C.

2.5.6 CrDPAS, TiDPAS2, CrGS, CrSGD, CrPAS and TbADH Large-scale Protein Purification

Cells were harvested by centrifugation at 3200 x *g* for 15 minutes and re-suspended in 50 mL buffer A1 (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) with addition of EDTA-free protease inhibitor cocktail (Roche Diagnostics Ltd.) and 10 mg lysozyme (Sigma). Dithiothreitol (Sigma) (final concentration of 0.05 mM) was additionally added to all buffers in purification of CrDPAS and TiDPAS2 for crystallisation to limit the formation of disulfide-bridges that may result in protein misfolding. Cells were lysed at 4 °C using a cell disruptor at 30 KPSI and centrifuged (35000 x *g*) to remove insoluble cell debris. The supernatant was collected and filtered with 0.2 µm PES syringe filter (Sartorius) and purified using an AKTA Pure FPLC (Cytiva). Sample was applied at 2 mL/min onto a His-Trap HP 5mL column (Cytiva) and washed with 5 column volumes (CV) of buffer A1 before being eluted with 5 CV of buffer B1. Protein was detected and collected using the UV 280 nm signal and then further purified on a Superdex Hiload 16/60 S200 gel filtration column (Cytiva) at a flow rate of 1 mL/min using buffer A4. Proteins were finally buffer exchanged into buffer A4 and concentrated using 10K Da molecular weight cut off centrifugal filter (Merck) before being snap frozen in liquid nitrogen and stored at -80 °C. For the crystallisation of CrDPAS and TiDPAS2, protein after gel filtration was incubated on a rocker overnight at 4°C with 3C protease to cleave the 6xHis-tag. Proteins were then passed through a 1mL HisTrap column (Cytiva) to remove the cleaved tag. Proteins were then buffer exchanged into buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl) containing 0.05 mM tris(2-carboxyethyl)phosphine (Sigma) and concentrated using 10K Da molecular weight cut off centrifugal filter (Merck) and stored at –80 °C.

2.5.7 Synthesis of NADPD

Deuterated pro-*R*-NADPD was produced *in vitro* as previously described ^[47] with minor modifications. A 20 mL reaction mixture containing 2 mM NADP⁺, 4 mM d₈-isopropanol, 1

mM semicarbazide and 5 μ M *TbADH* in 50 mM ammonium bicarbonate buffer at pH 7.5 was incubated at 30 °C. The progression of the reaction was monitored by a spectrophotometer at 340 nm. When no significant increase in absorbance was observed (approximately 3 hours), 300 μ L of Ni-NTA agarose beads (Qiagen) was added and the sample incubated rocking at room temperature for 30 minutes. The reaction was centrifuged to remove the Ni-NTA beads bound to *TbADH*, and the supernatant was filtered through a 45 μ m glass filter and lyophilized to remove the unreacted d_8 -isopropanol, the acetone that forms during the reaction and the buffer. The residue, containing primarily NADPD, was stored at -20 °C until use.

2.5.8 *In vitro* Enzyme Assays

Enzymatic assays with precondylocarpine acetate were performed in 50 mM HEPES buffer (pH 7.5) with 50 μ M precondylocarpine acetate in MeOH (not exceeding 5% of the reaction volume), 250 μ M NADPH cofactor (Sigma) and 150 nM enzyme to a final reaction volume of 100 μ L. Reactions were incubated for 30 minutes at 30 °C and shaking at 60 r.p.m. before being quenched with 1 volume of 70% MeOH with 0.1% HCO₂H.

Enzymatic assays with strictosidine aglycone were performed in 50 mM HEPES buffer (pH 7.5), 100 μ M strictosidine and 1 mM SGD to a final reaction volume of 100 μ L. Assays were incubated for 30 minutes at 30 °C and shaking at 60 r.p.m before 500 nM of ADH enzyme and 250 μ M NADPH was added. As control, the reactions were performed without the addition of ADH enzyme. Reactions were incubated for a further 30 minutes at 30 °C shaking at 60 r.p.m. before being quenched with 1 volume of 70% MeOH with 0.1% HCO₂H.

All enzymatic assays were centrifuged at 14000 x *g* for 15 minutes and the supernatant analysed by UPLC-MS.

2.5.9 UPLC-MS Analysis

All assays were analysed using a Thermo Scientific Vanquish UPLC coupled to a Thermo Q Exactive Plus orbitrap MS. For assays using precondylocarpine acetate, chromatographic separation was performed using a Phenomenex Kinetex C18 2.6 μ m (2.1 x 100 mm) column using water with 1% HCO₂H as mobile phase A and acetonitrile with 1% HCO₂H as mobile

phase B. Compounds were separated using a linear gradient of 10-30% B in 5 minutes followed by 1.5 minutes isocratic at 100% B. The column was then re-equilibrated at 10% B for 1.5 minutes. The column was heated to 40 °C and flow rate was set to 0.6 mL/min. For assays using strictosidine aglycone, separation was carried out using a Waters Acquity BEH C18 1.7 μm (2.1 x 50 mm) using 0.1% NH₄OH in water as mobile phase A and acetonitrile as mobile phase B. Compounds were separated using a linear gradient of 10-90% B in 9 minutes followed by 2 minutes isocratic at 90% B. The column was re-equilibrated at 10% B for 3 minutes. The column was heated to 50 °C and flow rate was set to 0.4 mL/min. MS detection was performed in positive ESI under the following conditions: spray voltage was set to 3.5 kV ~ 67.4 μA, capillary temperature set to 275 °C, vaporizer temperature 475 °C, sheath gas flow rate 65, sweep gas flow rate 3, aux gas flow rate 15, S-lens RF level to 55 V. Scan range was set to 200 - 1000 *m/z* and resolution at 17500.

2.5.10 Production and Isolation of *d*-angryline and *d*₂-vincadifformine

d-angryline was produced enzymatically from stemmadenine acetate using the same protocol previously described for the synthesis of angryline but replacing NADPH with NADPD [32]. Briefly, 0.25 mg of stemmadenine acetate, 40 μM flavin adenine dinucleotide (FAD) and 5 μg of *CrPAS* were combined in a total volume of 500 μL in 50 mM TRIS-HCl buffer pH 8.5 and incubated at 37 °C to form precondylocarpine acetate (reaction progress was monitored by LC-MS, *m/z* 395.19). After 2 hours, 1 mg of NADPD and 9 μg of *CrDPAS* were added to the reaction and incubated for 20 minutes at 37 °C to obtain *d*-angryline (*m/z* 338.19). Multiple reactions were prepared to obtain sufficient product for NMR characterization. After completion, the reactions were snap frozen in liquid nitrogen and stored at -80 °C.

*d*₂-vincadifformine was also produced enzymatically, but in this case NADPD was generated directly in the reaction mixture using an alcohol dehydrogenase from *E. coli* (Merck product 49854). Multiple 500 μL reactions were prepared to obtain sufficient product for NMR characterization. Each reaction contained 400 μM NADP⁺, 0.89 μg *d*₈-isopropanol, 1 μg of *TbADH*, 10 μg stemmadenine acetate, 0.8 μM *CrPAS* and 0.8 μM *TiDPAS1* in 50 mM HEPES buffer pH 7.5. The reactions were incubated at 30 °C for 1 hour, snap frozen in liquid nitrogen and stored at -80 °C until purification of the final product.

d-angryline and *d*₂-vincadifformine were purified by semi-preparative HPLC on an Agilent 1260 Infinity II HPLC system. The reactions were thawed and 500 μL of 90:9:1 MeOH:H₂O:HCO₂H was added to the deuterated samples. The samples were filtered through 0.2 μm PTFE disc filters (Sartorius) to remove the precipitated enzymes and injected onto a Phenomenex Kinetex XB-C18 5 μm (250 x 10 mm) column. Chromatographic separation was performed using 0.1% HCO₂H in water as mobile phase A and acetonitrile as mobile phase B. A linear gradient from 10% B to 40% B in 15 minutes was used for chromatographic separation of the compounds followed by a wash at 40% B for 5 minutes and a re-equilibration step to 10% B for 5 minutes. Flow rate was 6 mL/min. Elution of *d*-angryline and *d*₂-vincadifformine was monitored at two wavelengths, 330 and 254 nm. Fractions containing the compounds of interest were collected, dried under reduced pressure and stored at -80 °C until further analysis.

2.5.11 Production and Isolation of 19,20-dihydrovallesiachotamine

19,20-dihydrovallesiachotamine was produced enzymatically from 100 μM strictosidine reacted with 100 μM CrSGD in 50 mM HEPES buffer pH 7.5 in a 100 mL reaction at 30°C. After 90 minutes, 500 nM of CrDPAS and 250 μM NADPH was added and the reaction monitored. After 2 hours a further 500 nM CrDPAS was added to a final concentration of 1 μM and left for a further 3 hours until the reaction reached completion. The sample was snap frozen in liquid nitrogen and stored at -80 °C. For purification, the sample was thawed on ice and filtered through a 0.2 μm PTFE disc filter (Sartorius) to remove the precipitated enzymes and then passed through a Supelco DSC-18 column (MilliporeSigma) and eluted with methanol. Eluent was dried down in a rotovap and resuspended in 1.5 mL methanol. The product was purified on an Agilent 1290 Infinity II semi-preparative HPLC system using a Waters XBridge BEH C18 5 μm (10 x 250mm) column and using 0.1% NH₄OH in water as mobile phase A and acetonitrile as mobile phase B. Compounds were separated using a linear gradient of 10-65% B in 25 minutes followed by 10 minutes column re-equilibration at 10% B. Flow rate was set to 7mL/min. Compound was detected by measuring UV 290 nm and 254 nm signal. Fractions containing the compound of interest were collected and dried down using a rotovap and stored at -20 °C until NMR analysis.

2.5.12 NMR of *d*-angryline, *d*-vincadifformine and 19,20-dihydrovallesiachotamine

For *d*-angryline, NMR spectra were measured on a 400 MHz Bruker Advance III HD spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany). NMR spectra for 19,20-dihydrovallesiachotamine, (–)-vincadifformine and *d*₂-(±)-vincadifformine were measured on a 700 MHz Bruker Advance III HD spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany). For spectrometer control and data processing Bruker TopSpin ver. 3.6.1 was used. MeOH-*d*₃ was used as a solvent and all NMR spectra were referenced to the residual solvent signals at δ H 3.31 and δ C 49.0, respectively.

2.5.13 ECD Measurement and Spectral Calculations of Vincadifformine

ECD spectra were measured at 25 °C on a JASCO J-810 spectropolarimeter (JASCO cooperation, Tokyo, Japan) using a 350 μ L cell. Spectrometer control and data processing was accomplished using JASCO spectra manager II.

Based on the structure determined from NMR analysis a molecular model was created in GaussView ver.6 (Semichem Inc., Shawnee, Kansas, USA) and optimized using the semi-empirical method PM6 in Gaussian (Gaussian Inc., Wallingford, Connecticut, USA). The resulting structure was used for conformer variation with the GMMX processor of the Gaussian program package. Resulting structures were DFT-optimized with Gaussian ver.16 (APFD/6-31G(d)). A cut-off level of 4 kcal/mol was used to select conformers which were subjected to another DFT optimization on a higher level (APFD/6-311G+(2d,p)). All structures up to a deviation of 2.5 kcal/mol from the lowest energy conformer were used to determine the ECD-frequencies in a TD-SCF calculation on the same level as the former DFT optimization. The ECD curve was calculated from the Boltzmann-weighted contributions of all conformers with a cut-off level of two percent. Experimentally measured ECD data and calculated data were compared using SpecDis ver.1.71 ^[48].

2.5.14 Protein Crystallisation

Protein sequences were analysed for disordered regions using XstalPred web server ^[49]. Purified CrDPAS and TiDPAS2 were crystallised by sitting-drop vapour diffusion on MRC2 96-well crystallisation plates (SwissSci) with 0.3 μ L protein and 0.3 μ L precipitant solution drops dispensed by Oryx8 robot (Douglas Instruments).

CrDPAS was crystallised using JCSG screen (Jena Biosciences) with 1.26 M ammonium sulfate, 100 mM TRIS buffer pH 8.5 and 200mM lithium sulfate. Crystallisation condition with additional 1 mM NADP⁺ and 25% ethylene glycol was used as cryoprotectant.

TiDPAS2 was initially screened using PEG/Salt screen (Jena Biosciences) before condition optimization. Apo-*TiDPAS2* was crystallised in 17% w/v PEG 3350, 200 mM ammonium chloride and 0.75 mM angriline (no electron density corresponding to angriline was observed in the structure). 17% w/v PEG 3350, 220 mM ammonium chloride, 1 mM NADP⁺, 1 mM angriline and 25% ethylene glycol was used as cryoprotectant. Stemmadenine acetate-bound *TiDPAS2* was crystallised in 23% w/v PEG 3350, 250 mM sodium sulfate and 0.75 mM stemmadenine acetate, 23% w/v PEG 3350, 200 mM sodium sulfate, 1 mM NADP⁺, 1 mM stemmadenine acetate and 25% ethylene glycol was used as cryoprotectant. Precondylocarpine acetate-bound *TiDPAS2* was crystallised in 25% w/v PEG 3350, 180 mM sodium sulfate and 0.75 mM precondylocarpine acetate. 23% w/v PEG 3350, 200mM sodium sulfate, 1 mM NADP⁺, 1 mM precondylocarpine acetate and 25% ethylene glycol was used as cryoprotectant.

All crystals were soaked in the corresponding cryoprotectant before flash-cooling in liquid nitrogen.

2.5.15 X-ray Data Collection, Processing and Structure Solution

X-ray data sets for *CrDPAS* and *TiDPAS2* structures were recorded on the 10SA (PX II) beamline at the Paul Scherrer Institute (Villigen, Switzerland) at wavelength of 1.0 Å using a Dectris Eiger3 16M detector with the crystals maintained at 100K by a cryocooler. Diffraction data were integrated using XDS^[50] and scaled and merged using AIMLESS^[51]; data collection statistics are summarized in Appendix V. Structure's solution was automatically obtained by molecular replacement using the structure of tetrahydroalstonine synthase from *C. roseus* (PDB accession code 5FI3) as template with which *CrDPAS* and *TiDPAS2* share 54% and 56% amino acid identity respectively. In all cases the map was of sufficient quality to enable 90% of the residues expected for a homodimer to be automatically fitted using Phenix autobuild^[52, 53]. The models were finalized by manual rebuilding in COOT^[54] and refined using in Phenix refine.

All structures are in the PDB database under the following accessions: 8B27 (*CrDPAS*), 8B26 (*apo-TiDPAS2*), 8B1V (*precondylocarpine acetate-bound TiDPAS2*), 8B25 (*stemmadenine acetate-bound TiDPAS2*). The model statistics are reported in Appendix V.

2.5.16 Docking simulations

Ligands were docked into the active site of *TiDPAS* and *CrGS* using AutoDock Vina on the Webina webserver using default parameters [38, 55]. Coordinates of ligands were generated by PDBQTConvert. When assessing the results, we selected ligand orientations in which the 4-pro-*R* hydride of NADPH was in close proximity to the carbon being reduced; this orientation was not always the lowest possible energy solution. Results were visualised using PyMOL.

2.5.17 Phylogenetic analysis

Nucleic acid sequences of ADH genes were aligned using MUSCLE v5 [56]. A maximum likelihood phylogenetic tree was constructed using IQTree [57] using a best-fit substitution model followed by tree reconstruction using 1000 bootstrap alignments and the remaining parameters used default settings. Figures were made using iTOL version 6.5.2 [58].

Table 4. Genbank accession for sequences used to construct tree of maximum likelihood.

Gene Name	Genbank accession
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 1 (CAD1)	AT1G72680
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 2 (CAD2)	AT2G21730
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 3 (CAD3)	AT2G21890
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 4 (CAD4)	AT3G19450
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 5 (CAD5)	AT4G34230
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 6 (CAD6)	AT4G37970
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 7 (CAD7)	AT4G37980
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 8 (CAD8)	AT4G37990
<i>Arabidopsis thaliana</i> cinnamyl alcohol dehydrogenase 9 (CAD9)	AT4G39330
<i>Populus tremuloides</i> sinapyl alcohol dehydrogenase (SAD)	AF273256.1

<i>Camptotheca accuminata</i> 8-hydroxygeraniol oxidase (8HGO)	AY342355.1
<i>Ocimum basilicum</i> geraniol dehydrogenase (GEDH)	AY879284.1
<i>Rauwolfia serpentina</i> cinnamyl alcohol dehydrogenase (CAD)	KT369739.1
<i>Catharanthus roseus</i> 8-hydrogeraniol dehydrogenase (8HGO)	KF561458.1
<i>Strychnos speciosa</i> Wieland-Gumlich aldehyde synthase (WS)	OM304303.1
<i>Strychnos nux-vomica</i> Wieland-Gumlich aldehyde synthase (WS)	OM304294.1
<i>Catharanthus roseus</i> geissoschizine synthase (GS)	MF770507.1
<i>Cinchona pubescens</i> dihydrocorinantheine aldehyde synthase (DCS)	MW456554
<i>Catharanthus roseus</i> tabersonine 3- reductase (T3R)	KP122966.1
<i>Catharanthus roseus</i> tetrahydroalstonine synthase (THAS)	KM524258.1
<i>Catharanthus roseus</i> heteroyohimbine synthase (HYS)	KU865325.1
<i>Rauwolfia serpentina</i> vomilenine reductase 2 (VR2)	KT369740.1
<i>Rauwolfia tetraphylla</i> vomilenine reductase 2 (VR2)	KT369741.1
<i>Tabernanthe iboga</i> dihydroprecondylocarpine acetate synthase 1 (DPAS1)	MK840855.1
<i>Tabernanthe iboga</i> dihydroprecondylocarpine acetate synthase 2 (DPAS2)	MK840856.1
<i>Catharanthus roseus</i> dihydroprecondylocarpine acetate synthase (DPAS)	KU865331.1

2.6 References

- [1] O. Danielsson, H. Jörnvall, *Proc National Acad Sci* 1992, *89*, 9247–9251.
- [2] D.-M. Guo, J.-H. Ran, X.-Q. Wang, *J Mol Evol* 2010, *71*, 202–218.
- [3] S.-J. Kim, M.-R. Kim, D. L. Bedgar, S. G. A. Moinuddin, C. L. Cardenas, L. B. Davin, C. Kang, N. G. Lewis, *Proc National Acad Sci* 2004, *101*, 1455–1460.
- [4] J. Strommer, *Plant J* 2011, *66*, 128–142.
- [5] B. Mee, D. Kelleher, J. Frias, R. Malone, K. F. Tipton, G. T. M. Henehan, H. J. Windle, *Febs J* 2005, *272*, 1255–1264.
- [6] H. Jörnvall, M. Landreh, L. J. Östberg, *Chem-biol Interact* 2015, *234*, 75–79.
- [7] D. S. Auld, T. Bergman, *Cell Mol Life Sci* 2008, *65*, 3961.
- [8] B. Youn, R. Camacho, S. G. A. Moinuddin, C. Lee, L. B. Davin, N. G. Lewis, C. Kang, *Org Biomol Chem* 2006, *4*, 1687–1697.
- [9] T. D. Hurley, W. F. Bosron, J. A. Hamilton, L. M. Amzel, *Proc National Acad Sci* 1991, *88*, 8149–8153.
- [10] B. V. Plapp, S. Ramaswamy, *Biochemistry-us* 2012, *51*, 4035–4048.
- [11] B. V. Plapp, B. R. Savarimuthu, D. J. Ferraro, J. K. Rubach, E. N. Brown, S. Ramaswamy, *Biochemistry-us* 2017, *56*, 3632–3646.
- [12] S. R. Guntupalli, Z. Li, L. Chang, B. V. Plapp, R. Subramanian, *Biochemistry-us* 2021, *60*, 663–677.
- [13] B. V. Plapp, H. A. Charlier, S. Ramaswamy, *Arch Biochem Biophys* 2016, *591*, 35–42.
- [14] E. C. Tatsis, I. Carqueijeiro, T. D. D. Bernonville, J. Franke, T.-T. T. Dang, A. Oudin, A. Lanoue, F. Lafontaine, A. K. Stavrinides, M. Clastre, V. Courdavault, S. E. O’connor, *Nat Commun* 2017, *8*, 316.
- [15] A. Stavrinides, E. C. Tatsis, L. Caputi, E. Foureau, C. E. M. Stevenson, D. M. Lawson, V. Courdavault, S. E. O’Connor, *Nat Commun* 2016, *7*, 12116.
- [16] A. Stavrinides, E. C. Tatsis, E. Foureau, L. Caputi, F. Kellner, V. Courdavault, S. E. O’Connor, *Chem Biol* 2015, *22*, 336–41.
- [17] A. Awadasseid, W. Li, Z. Liu, C. Qiao, J. Pang, G. Zhang, Y. Luo, *Int J Biol Macromol* 2020, *162*, 1076–1085.

- [18] K. Miettinen, L. Dong, N. Navrot, T. Schneider, V. Burlat, J. Pollier, L. Woittiez, S. V. D. Krol, R. Lugan, T. Ilc, R. Verpoorte, K. M. Oksman-Caldentey, E. Martinoia, H. Bouwmeester, A. Goossens, J. Memelink, D. Werck-Reichhart, *Nat Commun* 2014, 5, 3606.
- [19] S. C. Farrow, M. O. Kamileen, L. Caputi, K. Bussey, J. E. A. Mundy, R. C. McAtee, C. R. J. Stephenson, S. E. O'Connor, *J Am Chem Soc* 2019, 141, 12979–12983.
- [20] B. Hong, D. Grzech, L. Caputi, P. Sonawane, C. E. R. López, M. O. Kamileen, N. J. H. Lozada, V. Grabe, S. E. O'Connor, *Nature* 2022, 1–6.
- [21] C. Langley, E. Tatsis, B. Hong, Y. Nakamura, C. Paetz, C. E. M. Stevenson, J. Basquin, D. M. Lawson, L. Caputi, S. E. O'Connor, *Angew. Chem. Int. Ed.* 2022, 61, e202210934.
- [22] L. Caputi, J. Franke, S. C. Farrow, K. Chung, R. M. E. Payne, T.-D. Nguyen, T.-T. T. Dang, I. S. T. Carqueijeiro, K. Koudounas, T. D. de Bernonville, B. Ameyaw, D. M. Jones, I. J. C. Vieira, V. Courdavault, S. E. O'Connor, *Science* 2018, 360, 1235–1239.
- [23] M. O. Kamileen, M. D. DeMars, B. Hong, Y. Nakamura, C. Paetz, B. R. Lichman, P. D. Sonawane, L. Caputi, S. E. O'Connor, *J Am Chem Soc* 2022, DOI 10.1021/jacs.2c08107.
- [24] A. Edge, Y. Qu, M. L. A. E. Easson, A. M. K. Thamm, K. H. Kim, V. D. Luca, *Planta* 2018, 247, 155–169.
- [25] Y. Qu, M. L. A. E. Easson, J. Froese, R. Simionescu, T. Hudlicky, V. DeLuca, *Proc National Acad Sci* 2015, 112, 6224–6229.
- [26] Y. Qu, M. E. A. M. Easson, R. Simionescu, J. Hajicek, A. M. K. Thamm, V. Salim, V. D. Luca, *Proc National Acad Sci* 2018, 115, 3180–3185.
- [27] M. Geissler, M. Burghard, J. Volk, A. Staniek, H. Warzecha, *Planta* 2016, 243, 813–824.
- [28] C. Li, J. C. Wood, A. H. Vu, J. P. Hamilton, C. E. R. Lopez, R. M. E. Payne, D. A. S. Guerrero, K. Gase, K. Yamamoto, B. Vaillancourt, L. Caputi, S. E. O'Connor, C. R. Buell, *Nat Chem Biol* 2023, 1–11.
- [29] F. Trenti, K. Yamamoto, B. Hong, C. Paetz, Y. Nakamura, S. E. O'Connor, *Org Lett* 2021, 23, 1793–1797.
- [30] R. Krithika, P. L. Srivastava, B. Rani, S. P. Kolet, M. Chopade, M. Soniya, H. V. Thulasiram, *Sci Rep-uk* 2015, 5, 1–6.
- [31] A. S. Sandholu, S. P. Mujawar, R. Krithika, H. V. Thulasiram, K. Kulkarni, *Proteins Struct Funct Bioinform* 2020, 88, prot.25891.
- [32] L. Caputi, J. Franke, K. Bussey, S. C. Farrow, I. J. C. Vieira, C. E. M. Stevenson, D. M. Lawson, S. E. O'Connor, *Nat Chem Biol* 2020, 16, 383–386.
- [33] P. Yao, Z. Xu, S. Yu, Q. Wu, D. Zhu, *Adv Synth Catal* 2019, 361, 556–561.

- [34] P. Stockinger, S. Roth, M. Müller, J. Pleiss, *Chembiochem* 2020, 21, 2689–2695.
- [35] S. L. Montgomery, A. Pushpanath, R. S. Heath, J. R. Marshall, U. Klemstein, J. L. Galman, D. Woodlock, S. Bisagni, C. J. Taylor, J. Mangas-Sanchez, J. I. Ramsden, B. Dominguez, N. J. Turner, *Sci. Adv.* 2020, 6, eaay9320.
- [36] K. T. D. D. Silva, G. N. Smith, K. E. H. Warren, *J Chem Soc D Chem Commun* 1971, 0, 905–907.
- [37] M. F. Bartlett, D. F. Dickel, W. I. Taylor, *J. Am. Chem. Soc.* 1958, 80, 126–136.
- [38] O. Trott, A. J. Olson, *J Comput Chem* 2010, 31, 455–461.
- [39] A. Vitale, F. Rosso, A. Barbarisi, T. Labella, S. D’Auria, *Gene* 2010, 461, 26–31.
- [40] A. Vitale, N. Thorne, S. Lovell, K. P. Battaile, X. Hu, M. Shen, S. D’Auria, D. S. Auld, *Plos One* 2013, 8, e63828.
- [41] T.-D. Nguyen, S. E. O’Connor, *Acs Chem Biol* 2020, 15, 1780–1787.
- [42] H. Jörnvall, J. Hedlund, T. Bergman, Y. Kallberg, E. Cederlund, B. Persson, in *Chemico-Biological Interactions*, Elsevier, 2013, pp. 91–96.
- [43] B. Persson, J. Hedlund, H. Jörnvall, *Cell Mol Life Sci* 2008, 65, 3879–3894.
- [44] Y. Hu, W. Liu, S. R. Malwal, Y. Zheng, X. Feng, T.-P. Ko, C.-C. Chen, Z. Xu, M. Liu, X. Han, J. Gao, E. Oldfield, R.-T. Guo, *Angewandte Chemie Int Ed* 2015, 54, 15478–15482.
- [45] M. Jarret, V. Turpin, A. Tap, J. Gallard, C. Kouklovsky, E. Poupon, G. Vincent, L. Evanno, *Angewandte Chemie Int Ed* 2019, 58, 9861–9865.
- [46] N. S. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. I. Stuart, R. J. Owens, *Nucleic Acids Res* 2007, 35, e45–e45.
- [47] S. S. Jeong, J. E. Gready, *Anal Biochem* 1994, 221, 273–277.
- [48] T. Bruhn, A. Schaumlöffel, Y. Hemberger, G. Bringmann, *Chirality* 2013, 25, 243–249.
- [49] L. Slabinski, L. Jaroszewski, L. Rychlewski, I. A. Wilson, S. A. Lesley, A. Godzik, *Bioinformatics* 2007, 23, 3403–3405.
- [50] W. Kabsch, *Acta Crystallogr Sect D Biological Crystallogr* 2010, 66, 125–132.
- [51] P. R. Evans, G. N. Murshudov, *Acta Crystallogr Sect D Biological Crystallogr* 2013, 69, 1204–1214.
- [52] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkóczi, V. B. Chen, T. I. Croll, B. Hintze, L.-W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R.

J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallogr Sect D* 2019, *75*, 861–877.

[53] N. Stein, *J Appl Crystallogr* 2008, *41*, 641–643.

[54] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Crystallogr Sect D Biological Crystallogr* 2010, *66*, 486–501.

[55] Y. Kochnev, E. Helleman, K. C. Cassidy, J. D. Durrant, *Bioinformatics* 2020, *36*, btaa579.

[56] R. C. Edgar, *Biorxiv* 2021, 2021.06.20.449169.

[57] J. Trifinopoulos, L.-T. Nguyen, A. von Haeseler, B. Q. Minh, *Nucleic Acids Res* 2016, *44*, W232–W235.

[58] I. Letunic, P. Bork, *Nucleic Acids Res* 2021, *49*, gkab301-.

[59] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, D. C. Richardson, *Nucleic Acids Res* 2007, *35*, W375–W383.

[60] D. P. Lotun, C. Cochard, F. R. J. Vieira, J. S. Bernardes, *Biorxiv* 2019, 649426.

[61] S. Zhao, R. B. Andrade, *J Org Chem* 2017, *82*, 521–531.

[62] M. E. Kuehne, U. K. Bandarage, A. Hammach, Y.-L. Li, T. Wang, *J Org Chem* 1998, *63*, 2172–2183.

Chapter 3. The Emergence of Atypical CADs as Drivers of MIA Chemical Diversity

3.1. Introduction

3.1.1. MIA Chemical Diversity in Gentianales

MIAs are a diverse class of compounds which derive from the central precursor molecule strictosidine [1]. These compounds are primarily found in the Gentianales order of plants including the Apocynaceae, Gelsemiaceae, Loganiaceae and Rubiaceae families, as well as the Nyssaceae family in the Cornales order [2]. Phylogenetic analysis has led to the emergence of STR, the enzyme that catalyses the condensation of tryptamine and secologanin, as being the evolutionary driver of MIA biosynthesis (Figure 48) [3]. In many

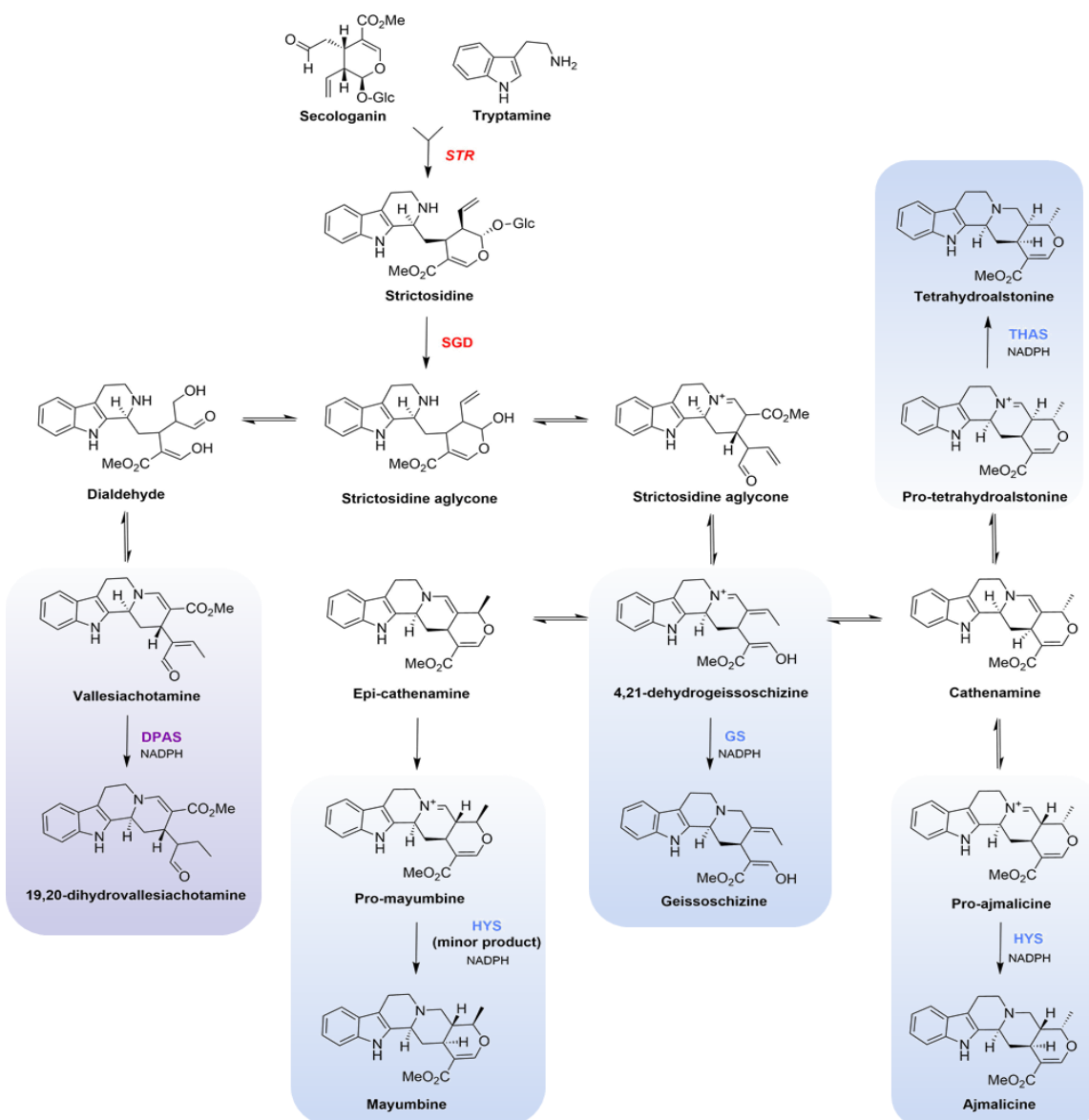


Figure 48. CAD-catalysed reductions of strictosidine aglycone in *C. roseus*. 1,4- or 1,2-reduction of an iminium moiety coloured in purple or blue, respectively.

MIA biosynthetic pathways, the resulting strictosidine is subsequently deglycosylated by the enzyme SGD to form strictosidine aglycone^[4-6]. This compound has the propensity to crosslink with proteins, suggesting its defensive role in plants, acting as a protective mechanism against herbivore attack^[7].

Beyond its defensive functions, strictosidine aglycone is a chemically versatile molecule that can reversibly rearrange to various structural isomers, each possessing a charged iminium moiety. However, to prevent excessive and detrimental protein crosslinking, a mechanism for neutralising the aglycone is imperative within the plant. This neutralisation is achieved through the reduction of the iminium moiety of strictosidine aglycone to form a less reactive compound - often catalysed by members of the CAD subfamily of enzymes (Figure 48)^[4, 8-10]. This atypical reduction by various CADs with different structural isomers of strictosidine aglycone gives rise a diverse range of MIA scaffold-types. Therefore, understanding the evolution of the CADs that catalyse reductions within MIA biosynthesis can reveal greater insights into the chemical diversity of these specialised metabolites within Gentianales.

3.1.2. Gene Duplication and Neofunctionalisation in Plant-Specialised Metabolism

The advancement of sequencing technologies over the past 15 years has significantly expanded access to genomic and transcriptomic data to an ever-increasing number of plant species, accelerating the discovery of biosynthetic genes^[11]. The extended read lengths facilitated by these technologies have resulted in the chromosome-length assembly of several plant genomes. This allows comparative genomic analyses and the identification of BGCs which in turn reveal insights into the evolution of biosynthetic genes involved in specialised metabolism.

A gene is understood to evolve by undergoing gradual mutagenesis. This can result in a change of expression, the partial or complete loss of function, or the gain of a new function in a process known as neofunctionalisation^[12, 13]. Such changes influence the selection pressure exerted on a gene, thereby guiding whether it is maintained or lost from the genome through the process of natural selection^[14]. Whilst most genes arise from whole genome or whole chromosome duplication events^[15], those involved in PNP biosynthesis

were found to often emerge from the lineage-specific expansion of their respective gene families. These smaller-scale duplication events are often driven by mobile sequences of DNA, known as transposable elements (TEs). These are capable of duplicating and/or relocating genes to form tandem duplication clusters ^[16], and are thought to mediate the formation of BGCs ^[17–19].

The genomic analyses of species which produce MIAs has identified examples of both BGCs and tandem duplication clusters involved in this specialised metabolic pathway ^[3, 6, 20–22]. Notably, the genome of *C. roseus* was previously assembled, however, technical limitations such as the shorter read lengths and a lack of closely related genomic comparisons prevented a more comprehensive analysis of genes encoding MIA biosynthesis including the CAD subfamily ^[22]. The growing repository and quality of genomic information available in recent years, including the recent chromosome-length assembly of the *C. roseus* genome ^[6] therefore provide the basis for understanding of the expansion and neofunctionalisation of this gene family and its contribution to the chemical diversity of MIAs in Gentianales.

3.1.3. CADs in MIA Biosynthesis

Exploring the evolution of a gene family that acts within MIA biosynthesis can provide insights into the divergence of these specialised metabolites. Much of the chemical diversity within MIAs is generated by the reduction of the various structural isomers of strictosidine aglycone. This reaction has been reported to be catalysed by several CAD enzymes ^[4, 8, 9, 23] and an SDR ^[24]. Notably, the CADs that reduce strictosidine aglycone catalyse either a highly unusual 1,2-reduction of an iminium moiety or the 1,4-reduction of an α,β -unsaturated aldehyde (Figure 48).

CADs are a subfamily of ADH enzymes that are prevalent in plants due to their functional role of reversibly reducing the aldehyde group of aromatic substrates in monolignol biosynthesis ^[25, 26]. In MIA biosynthesis, some CADs have been found to catalyse the atypical 1,2-reduction of an iminium moiety or the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde, the mechanistic basis of which is outlined in Chapter 2 of this thesis. These chemistries are attributed to specific alterations of critical residues within the catalytic pocket, enabling the sequence-based prediction of CADs likely to catalyse atypical

reductions. Consequently, an updated analysis of this enzyme family holds significant insights into the emergence of CADs that catalyse atypical reactions and whether this has contributed to the extensive chemical diversity of MIAs.

3.1.4. Chapter Overview

The CAD subfamily of genes encodes enzymes which catalyses atypical reduction reactions in MIA biosynthesis to generate much of the observed chemical diversity within this group of specialised metabolites. To understand the evolution of this gene family and its role in MIA biosynthesis, work in this chapter examines their genomic organisation, phylogenetic relationships, and expression patterns in *C. roseus*. Furthermore, we build on mechanistic findings detailed in Chapter 2 to explore the expansion and neofunctionalisation of CADs within Gentianales. We show that the presence of genes predicted to confer atypical reduction chemistries correlate with the species' production of MIAs. These findings highlight the neofunctionalisation of CADs as a crucial requisite for the diversification of MIAs in Gentianales and shed light on the evolution of these PNPs.

3.2. Results

3.2.1. Genomic organisation of CADs in *C. roseus*

The recent assembly of the chromosome-length *C. roseus* genome has opened up new avenues for investigating the genomic organisation of CADs within this species [6]. In this context, the CAD CrDPAS was subjected to a BLAST search against the *C. roseus* genome and the high sequence identity coding DNA sequences were retrieved. These were validated based on the identification of highly conserved residues involved in coordinating the structural zinc, resulting in the identification of 47 CADs (Appendix VI Table 20). These sequences ranged from 900-1377 coding base pairs (BP), though their gene sizes ranged from 1627-7273 BP. CADs were dispersed across 6 out of the 8 chromosomes of the *C. roseus* genome, with only one sequence unable to be scaffolded (Figure 49).

The reduction chemistry of each CAD was predicted based on residues involved in the coordination of the catalytic zinc and the proton relay as outlined in Chapter 2 of this thesis.

These were categorised as encoding enzymes that catalyse the typical reduction of an

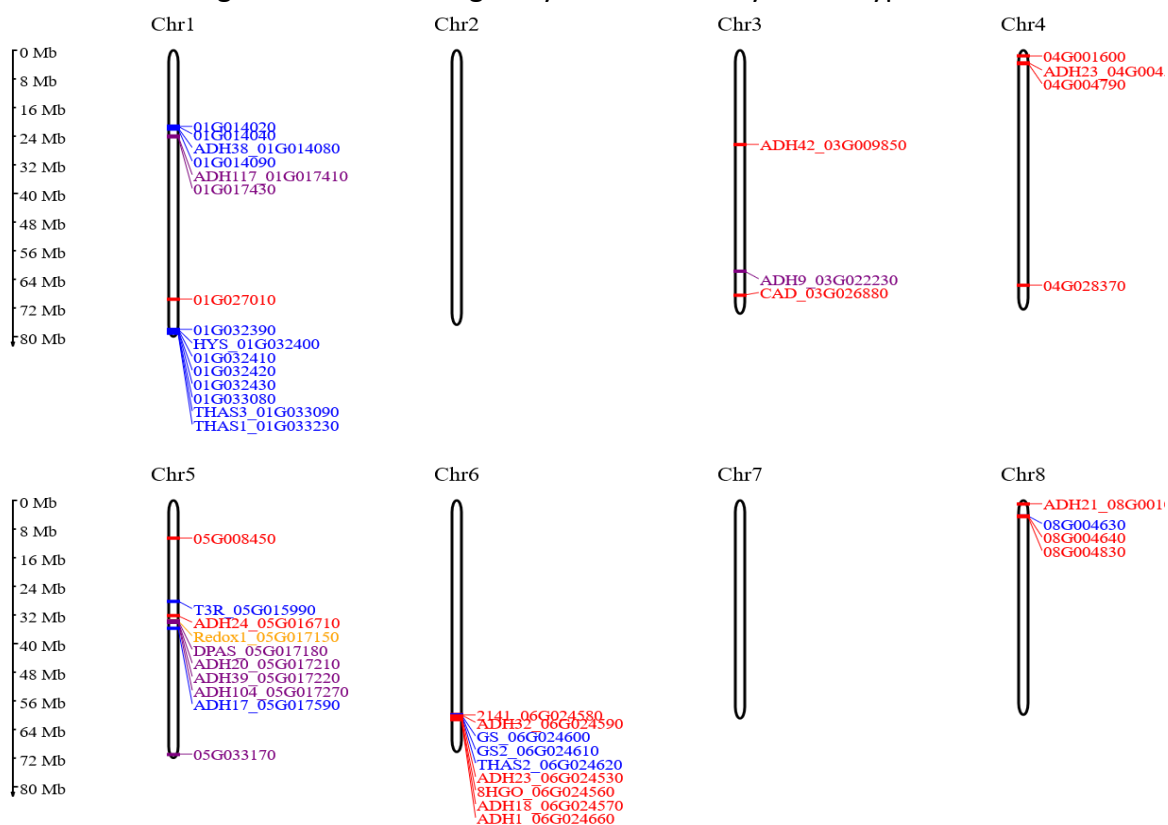


Figure 49. Genomic organisation of CADs in *C. roseus*. Genes coloured by predicted reduction types as typical aldehyde-reducing (red), 1,2-iminium reducing (blue), 1,4-iminium/ α,β -unsaturated aldehyde reducing (purple) and Redox1-like 1,2-iminium reducing (orange). Figure made using MG2C [41].

aldehyde (e.g. *Cr8HGO*), the 1,2-reduction of an iminium moiety reductions (e.g. *CrGS*), or the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde (e.g. *CrDPAS*). In addition, *CrRedOx1* is postulated to catalyse the 1,2-reduction of an iminium moiety [27], though unlike other CADs characterised to perform this chemistry, it was observed to have a Gly residue in the third position that typically coordinates the catalytic zinc ion instead of a Cys residue, so was classified as a distinct class. Remarkably, of the 47 CADs identified, 19 were predicted to catalyse typical reductions of an aldehyde, 19 to perform the 1,2-reduction of an iminium moiety, and 9 to catalyse the 1,4-reduction of iminium moiety or an α,β -unsaturated aldehyde (Appendix VI Table 20). It was noted that despite the gene structure of typical CADs varying between 4-10 exons, all the predicted atypical CADs had between 4-6 exons (Appendix VI Figure 96).

Analysis of the genomic positioning of CAD genes in *C. roseus* revealed several physical clusters, particularly of those predicted to encode enzymes which catalyse atypical reduction reactions (Figure 49). Further analysis of the cluster on chromosome 1 containing *CrGS* and *CrGS2* found that the exons of these genes shared a 50.7% nucleotide identity, suggesting their evolution by a recent tandem duplication event. However, the intron sequences of these genes were found to be similar to corresponding regions of *CrTHAS4* homologues (30.2-39.9% nucleotide similarity to *Cr01G032410*, *Cr01G032420*, *Cr01G032430* and *Cr01G033080*). Despite being located on chromosome 6 and 1, respectively, these genes encode CADs known to catalyse the reduction of different structural isomers of strictosidine aglycone (Figure 50). These observations therefore suggest that *CrGS* and *CrTHAS4* likely emerged by a TE-mediated chromosomal movement followed by subsequent neofunctionalisation to generate different MIA scaffolds.

In contrast, *CrDPAS* and *CrADH9* were found to differ in gene structure (Appendix VI Table 20) and chromosome location (Figure 49), despite their high exon nucleotide identity (82.5%). *CrDPAS* catalyses the 1,4-reduction of the substrate precondylocarpine acetate *in vitro*, whilst no product was observed in comparable reactions containing *CrADH9* (Appendix VI Figure 97). These results suggest that these genes may have emerged as dispersed duplicates and have since neofunctionalised to accept different substrates [15], further expanding the chemical diversity of MIAs through catalysing atypical reduction reactions.

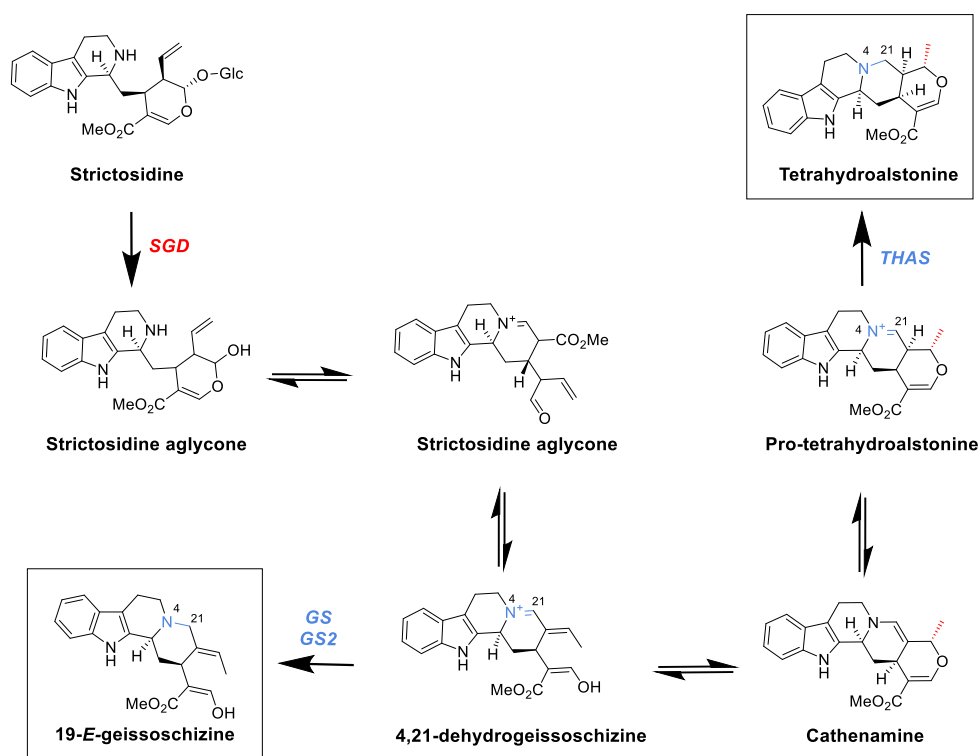


Figure 50. 1,2-iminium reductions of strictosidine aglycone rearrangements catalysed by *CrGS*, *CrGS2* and *CrTHAS1-4*.

3.2.2. Phylogeny of CADs in *C. roseus*

To further understand the expansion and the evolutionary relationships between members of the CAD gene family *C. roseus*, phylogenetic analysis was performed (Figure 51). Notably, each atypical class of CAD was found to have a monophyletic origin with a sister clade that contained typical CADs. This finding suggests that CADs underwent several divergence and subsequent neofunctionalisation events, each from an ancestral gene encoding an enzyme that catalysed the reduction of aldehyde. Some of these atypical classes of CADs were observed to have undergone further divergence to form distinct subclades.

The class of CADs identified or predicted to catalyse the 1,2-reduction of an iminium moiety could be further classified into three subclades (Figure 52). Subclade I and II contained genes known to encode CADs which reduce structural rearrangements of strictosidine aglycone. This included *CrGS* and *CrGS2*, both of which form geissoschizine in the first committed step of vinblastine biosynthesis [28], and *CrTHAS2* and *CrTHAS3* which form tetrahydroalstonine [4, 9]. The uncharacterised *Cr01G033080* was also identified to belong

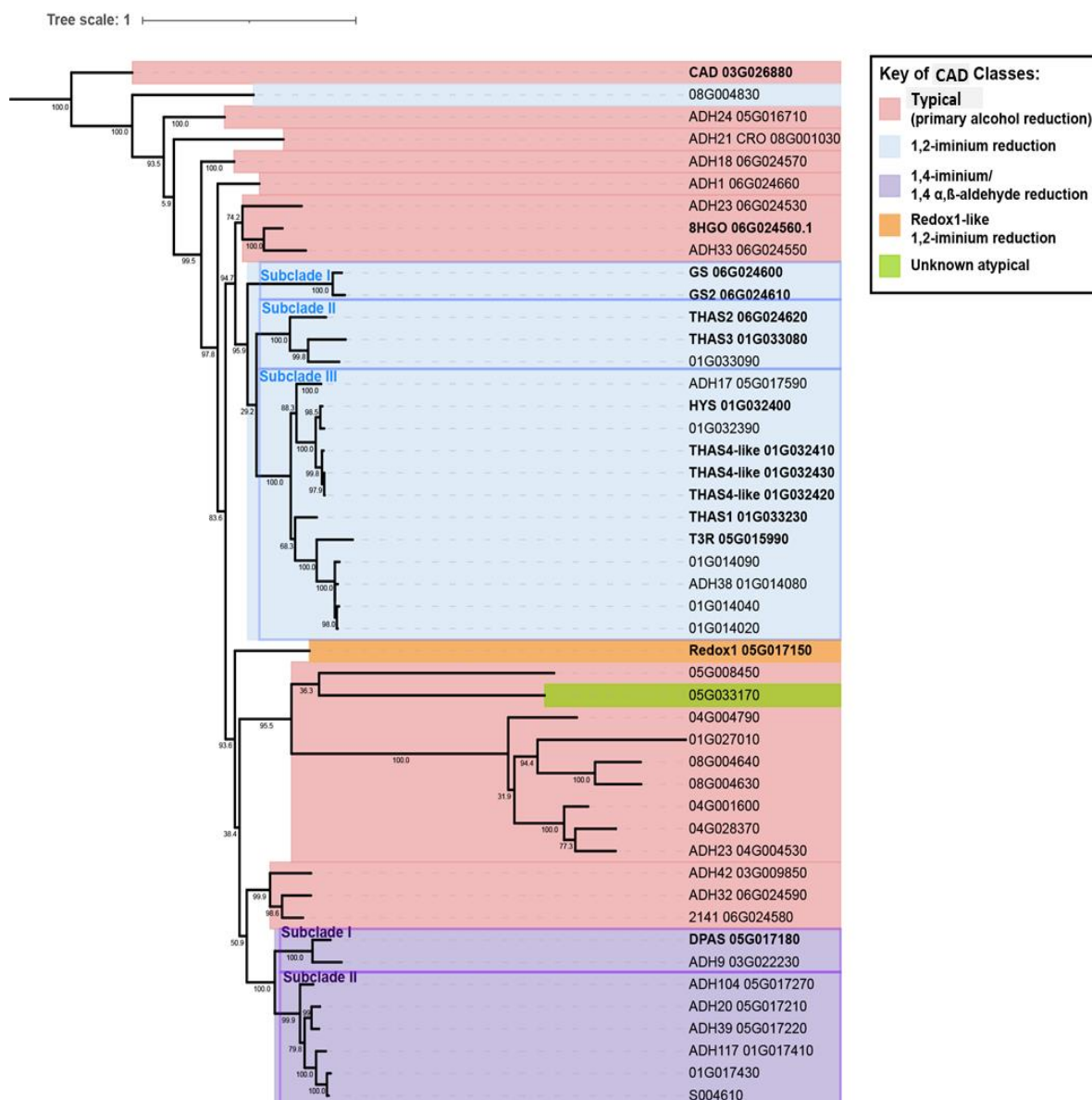


Figure 51. Maximum likelihood phylogenetic tree of CADs in *C. roseus*. Genes coloured by predicted reduction chemistries as either typical aldehyde (red), 1,2-iminium (blue), 1,4-iminium/ α,β -unsaturated aldehyde (purple), Redox1-like (orange) or unknown atypical (green) reductions based on residues involved in coordinating the catalytic zinc and the proton relay. Figure made using iTol [65].

to subclade II, suggesting it may also act on strictosidine aglycone. Subclade III contained genes which encode CADs that reduce strictosidine aglycone including *CrTHAS* orthologues and *CrHYS* which form tetrahydroalstonine and heteroyohimbine, respectively [9], as well as *CrT3R* which acts on aspidosperma-type MIAs in vindoline biosynthesis [28, 29]. The closely related uncharacterised *Cr01G014020*, *Cr01G014040*, *Cr01G014080*, and *Cr01G014090* may therefore also act on aspidosperma-type substrates.

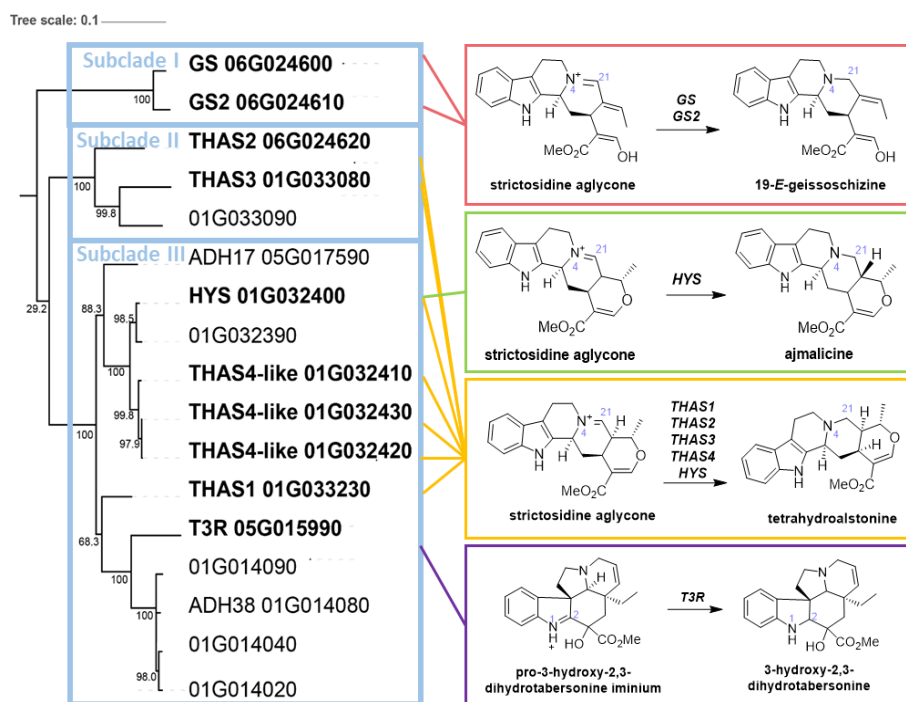


Figure 52. Maximum likelihood phylogenetic tree of 1,2-iminium reducing CADs in *C. roseus*. Characterised enzymes in bold with inset of catalysed reduction, tree visualised using iTol [65].

CrRedox1 and Cr05G033170 were observed to have an atypical residue in the third position involved in coordinating the catalytic zinc - namely, a Gly residue instead of the typical Cys. This observation is reflected in their large branch lengths (Figure 51). CrRedOx1 catalyses the 1,2-reduction of an iminium moiety, though is distinct from other CADs that perform this chemistry [27]. This finding suggests that this atypical chemistry has emerged through convergent evolution in *C. roseus*.

CADs with sequence patterns known to confer the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehydes (as observed in CrDPAS) formed a separate phylogenetic clade (Figure 51). Our analysis identified uncharacterised genes belonging to this clade such as CrADH104, CrADH39 and CrADH20. These findings suggests that these CADs may also encode enzymes capable of catalysing similar 1,4-reductions and are therefore interesting candidates for future gene discovery efforts.

3.2.3. Cell-Specific Expression Patterns of CADs in *C. roseus*

Recent advancements in single-cell transcriptomics have provided valuable insights into the cell-type-specific expression of genes within *C. roseus* leaf tissue [6]. Analysis of the genes

involved in MIA biosynthesis revealed they had enriched expression in 3 different cell types. Namely, genes encoding enzymes that act on iridoid and early secoiridoid biosynthesis were found to have enriched expressed in IPAP cells, those involved in early MIA biosynthesis were higher in epidermal cells, and late-stage MIA/bisindole biosynthesis genes were prominent in idioblast cells. This observation suggests that spatial localisation contributes to the generation of MIAs and may therefore facilitate the discovery of biosynthetic genes.

To understand the cell-specific expression pattern of CADs in *C. roseus*, we mapped the genes identified from the genome to the single-cell transcriptome, identifying 38 contigs (Appendix VI Table 21). In cases where several genome contigs mapped to the same single cell transcriptome contig, the highest identity hit was used. Among these sequences, 6 lacked expression data and were excluded from further analysis. The expression patterns of the remaining 32 CADs were analysed using hierarchical clustering (Figure 53).

Group VI of the hierarchical clustering contained genes enriched in epidermal cells based markers, including genes involved in MIA biosynthesis (Figure 53) [6]. This included CADs

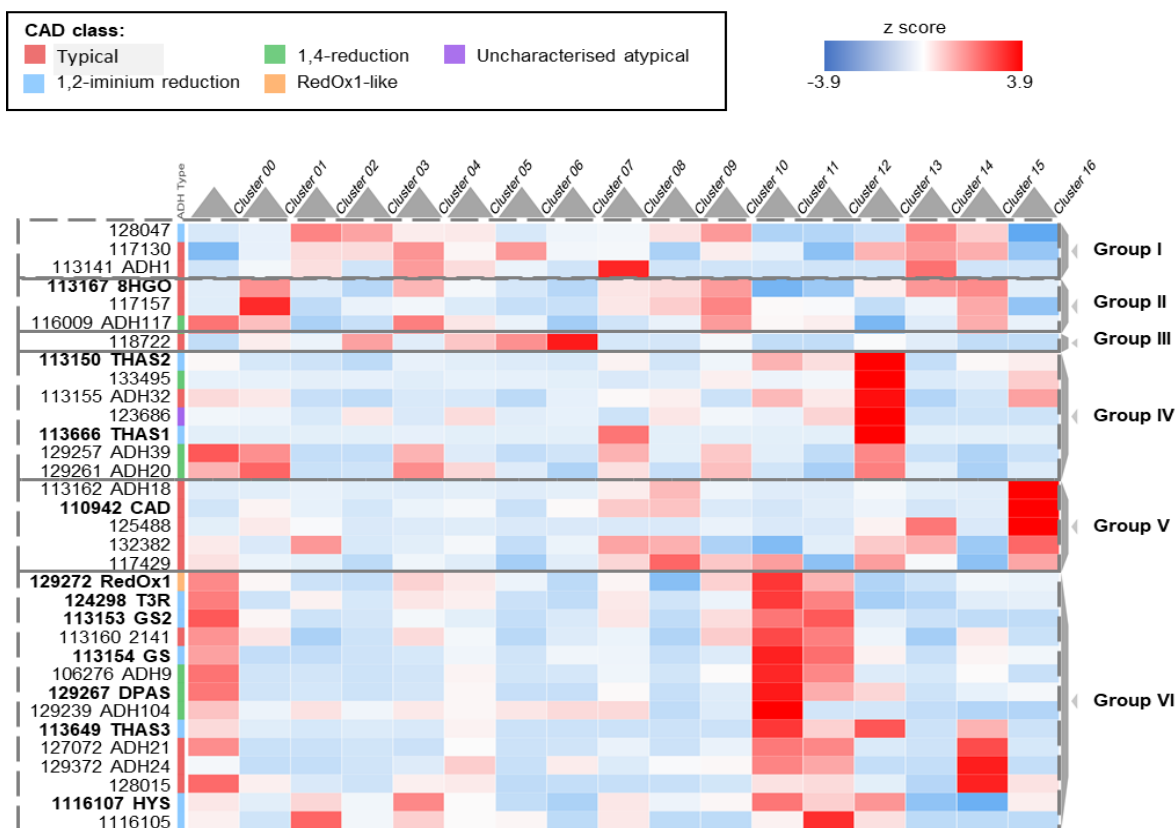


Figure 53. Hierarchical clustered heatmap of cell-type-specific expression patterns of CADs in *C. roseus* leaves. FPKM values normalised by z-score. Figure made using Clustergrammer [67].

that catalyse the 1,2-reduction of an iminium moiety such as *CrGS* [8] and *CrHYS* [9]. This group also contains genes known or predicted to encode enzymes which catalyse the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde such as *CrDPAS* [23] and the uncharacterised *CrADH9* and *CrADH104*. These findings suggest the potential role of these uncharacterised genes in late iridoid or early MIA biosynthesis.

Group IV of the hierarchical clustering comprised of genes with an enriched expression in a cell cluster determined by cell type markers to represent idioblast cells [6]. Idioblast cells were found to have enriched expression of genes involved in the final stages of MIA biosynthesis (Figure 53) [6, 30]. Namely, *CrTHAS1* and *CrTHAS2*, which were initially characterised to reduce the substrate strictosidine aglycone [4], though have since been shown to catalyse the production of the bis-indole MIA $\alpha,3'4'$ -anhydrovinblastine [6]. This observation highlights the role of cell-localised gene expression and substrate availability in metabolite production. Furthermore, our analysis identified 4 further CADs enriched in idioblast cells that have sequence motifs that confer the ability to catalyse a 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde (*CrADH39*, *CrADH20* and *Cr133495*), or the uncharacterised atypical *Cr123686*. Notably, it is hypothesised that one of the remaining unknown steps of vinblastine biosynthesis occurs through the 1,4-reduction of the anhydrovinblastine iminium intermediate (Figure 8) [31]. These findings therefore suggest the potential role of these uncharacterised CADs in late-stage MIA biosynthesis.

3.2.4. Phylogenetic Evolution of CADs in Gentianales

CADs have been reported to catalyse a range of atypical reduction reactions to generate MIA chemical diversity in an array of species across the order of Gentianales, as detailed in section 2.1.3. To explore the divergence and expansion of this enzyme family we therefore retrieved the sequences of 555 CADs from publicly available and in-house generated transcriptomes of 12 MIA-producing and 12 non-producing species within the Gentianales order, the closely orders Solanales and Lamiales, and the more distant Brassicales order. These sequences were validated as CADs based on the identification of highly conserved residues responsible for coordinating the structural zinc ion, and the reduction chemistry was subsequently predicted based on sequence motifs outlined in Chapter 2 of this thesis. This categorised each CAD as performing a typical aldehyde reduction, a 1,2-iminium

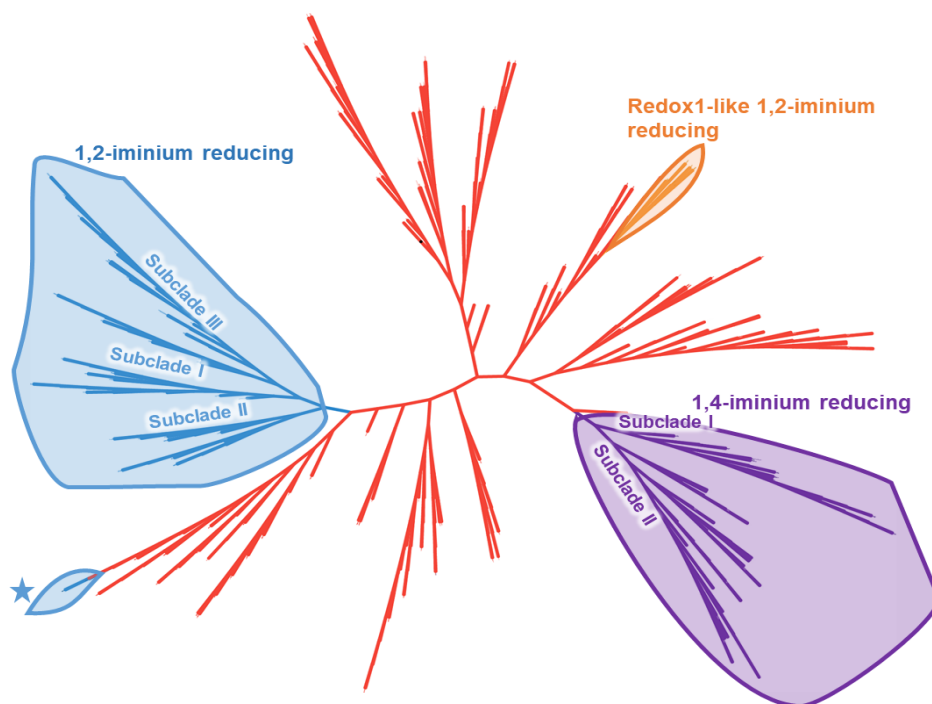


Figure 54. Maximum likelihood phylogenetic tree of CADs in the order Gentianales. Branches coloured by predicted activity as typical (red), 1,2-iminium reducing (blue), *CrRedox1*-like (orange) or 1,4-iminium or α,β -unsaturated aldehyde reducing (purple). 1,2-iminium reducing clade not observed in *C. roseus* denoted with a star. Figure made using iTol ^[65].

reduction or a 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde. Sequences which had a Gly residue in instead of a Cys in the typical third position that is responsible for coordinating the catalytic zinc ion were termed *CrRedOx1*-like.

Phylogenetic analysis of these sequences revealed that each atypical class of genes had a monophyletic origin within Gentianales with a sister clade containing typical CADs (Figure 54). This observation suggests that these classes of atypical CADs each emerged once in Gentianales from a typical ancestral CAD, as observed in *C. roseus* (section 3.2.2). Notably, the occurrence of CADs predicted to catalyse the 1,2-reduction of an iminium moiety within a given species correlated with its ability to produce MIAs (Figure 55). This suggests that the emergence of this atypical class occurred early in the divergence of Gentianales and is a necessary requisite for MIA biosynthesis. Examples of CADs predicted to catalyse *CrRedox1*-like reductions or 1,4-reductions of an iminium moiety or an α,β -unsaturated aldehyde appear solely in species within Apocynaceae, suggesting these emerged much later.

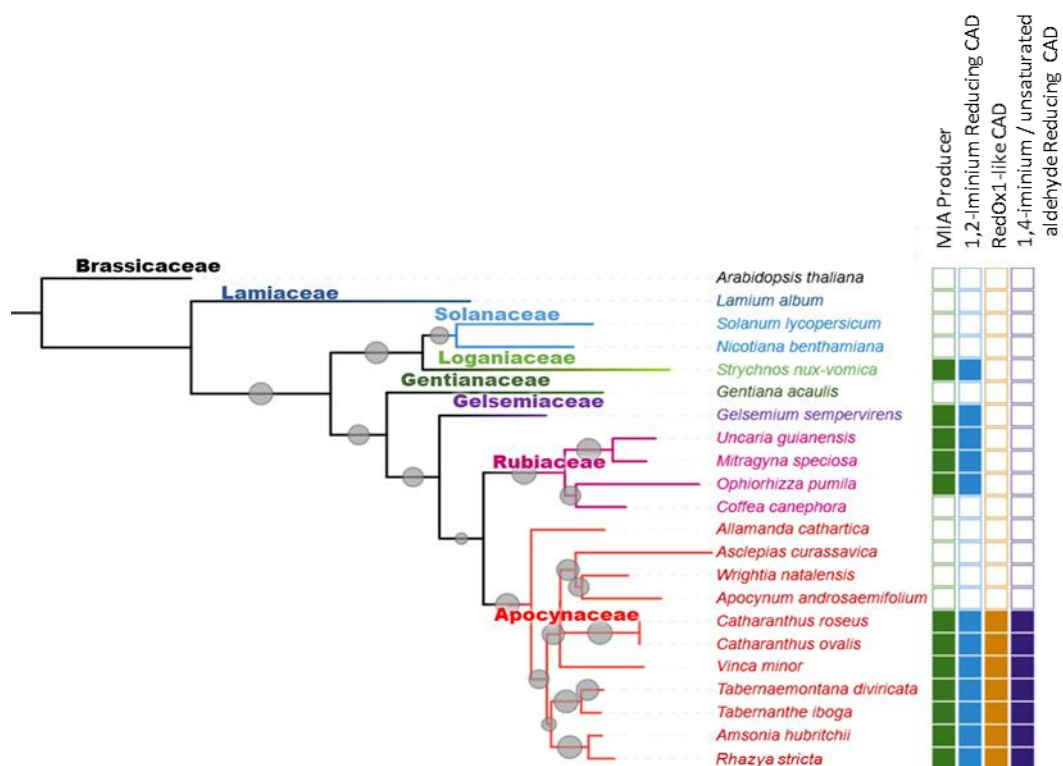


Figure 55. Maximum likelihood phylogenetic tree of species in the order Gentianales. Filled boxes indicate MIA production and the presence of predicted atypical reduction chemistries of CADs such as 1,2-iminium reducing, RedOx1-like, and 1,4-iminium/ α,β -unsaturated aldehyde reducing. Circle size represents bootstrap value. Figure made using iTol [65].

Our cross-species phylogenetic analysis further revealed genes from *Rhazya stricta* and *Amsonia hubritchii* which shared a pattern of active site residues observed in CADs known to catalyse the 1,2-reduction of an iminium moiety such as CrGS and detailed in section 2.2.13. However, these sequences were phylogenetically distinct and instead were more related to typical CADs (Figure 54). If upon further characterisation these sequences were found to catalyse a 1,2-iminium reduction, it would support that this atypical reduction has evolved multiple times in Apocynaceae through convergent evolution.

3.2.5. Genomic Synteny between *C. roseus* and *O. pumila*

Comparative genomics enables the identification of homologous regions of genomic organisation across species to facilitate an understanding of the evolutionary relationships of these genes, particularly in the context of biosynthetic pathways [3]. The genome

assemblies of *C. roseus* and *Ophiorrhiza pumila* have shed light on the evolution of MIA biosynthesis [3, 6]. These species both produce MIAs from the central precursor molecule strictosidine (Figure 56), though originate from the Apocynaceae and Rubiaceae families, respectively. Whilst MIA production in *C. roseus* goes through the deglycosylated strictosidine, *O. pumila* produces the MIA camptothecin through the hydrolysis and subsequent spontaneous coupling of the strictosidine methyl ester to form the proposed intermediate strictosamide. To understand the evolution of MIA biosynthetic pathways, we sought to compare the genomes of *C. roseus* and *O. pumila* genomes using syntenic analysis. As expected, we observed that genes involved in strictosidine biosynthesis shared genomic organisation between the species, though with the notable exceptions of the strictosidine and secologanin transporter proteins *CrNPF2.9* [5] and *CrSLTr* [6] respectively (Appendix VII Table 22, Figure 57) [22]. However, we did not observe synteny for orthologues of *CrSGD*, the first enzyme after the biosynthetic divergence between the species, or many genes that encode subsequent downstream enzymes (Appendix VII Table 22).

As the focus of our study pertains to the expansion and neofunctionalisation of the CAD subfamily of genes, we analysed these sequences (Appendix VII Table 22). We identified 6

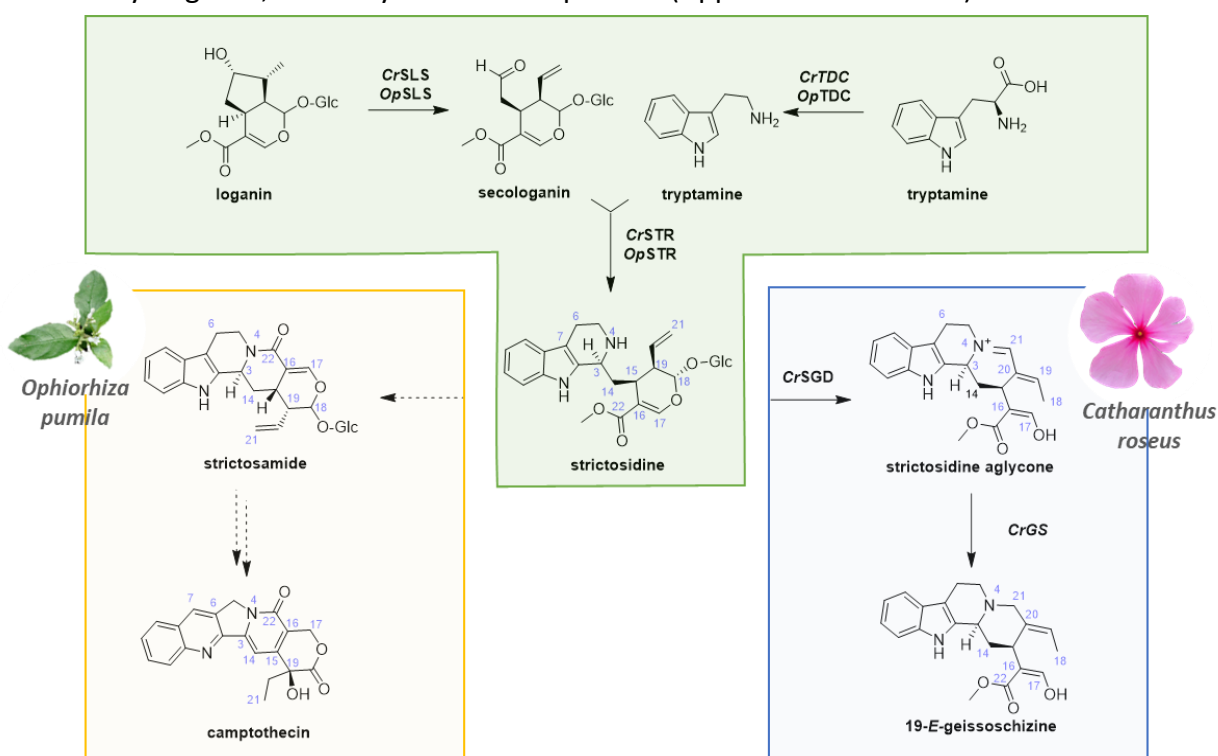


Figure 56. Biosynthetic pathway of the MIAs camptothecin and geissoschizine in *O. pumila* and *C. roseus* respectively, utilising the shared precursor strictosidine.

genes (*Cr04G004790*, *CrADH24*, *CrADH33*, *Cr08G004630* and *CrADH21*) which displayed synteny between the genomes and that were all predicted to catalyse the typical reduction of an aldehyde. This suggests their likely role in a conserved pathway such as lignin biosynthesis [26]. However, *CrGS* and *CrGS2* were observed to share genomic organisation with *Op0012230* and *Op0012250*, hereafter referred to as *OpADH1* and *OpADH2*, respectively. Sequence comparison found that *OpADH1* and *OpADH2* shared sequence motifs known to be crucial in catalysing the 1,2-reduction of an iminium moiety including atypical residues in positions important for cofactor binding as detailed in Chapter 2 of this thesis (Appendix VII Figure 98). The genomic synteny of CADs that likely catalyse the 1,2-reduction of an iminium moiety between *C. roseus* and *O. pumila* supports their monophyletic emergence and their subsequent conservation across plant families within Gentianales. Furthermore, *OpADH2* was found to co-express with other MIA biosynthetic genes in *O. pumila* (Appendix VII Figure 99) and was enriched in MIA-accumulating tissues (Appendix VII Figure 100). These findings suggest that *OpADH2* may function in MIA biosynthesis in *O. pumila* by likely catalysing the 1,2-reduction of an iminium moiety, and further highlights the emergence of CADs which catalyse atypical reductions as key drivers of the diversification of these PNPs.

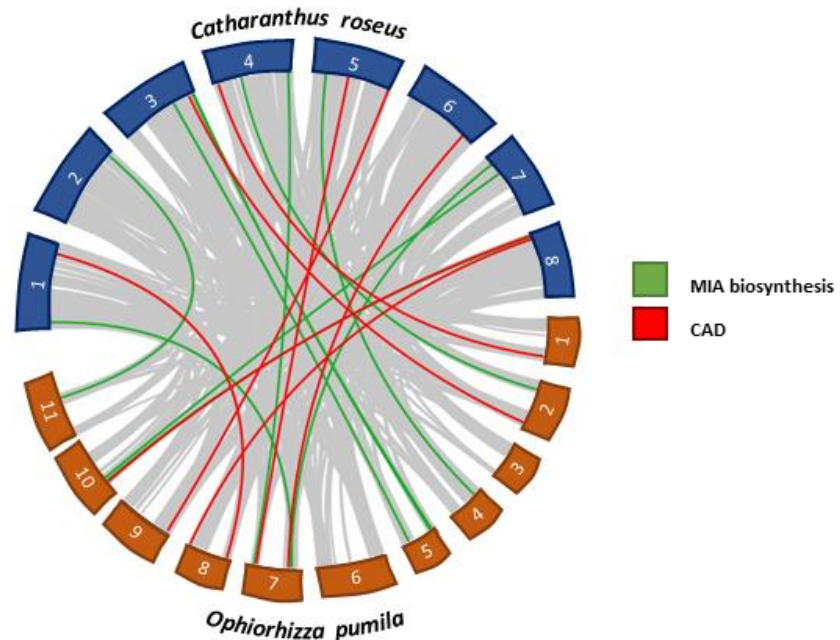


Figure 57. Syntenic analysis of *C. roseus* and *O. pumila* genomes. Genes involved in MIA biosynthesis (green) and CADs (red) are highlighted and mapped on to the corresponding chromosome of each species. Figure made using Circos [70].

3.3. Discussion

3.3.1. Emergence of CADs that Catalyse Atypical Reductions in *C. roseus*

The CAD lineage of genes has significantly expanded to 47 sequences *C. roseus* (Appendix VI Table 20) compared to the 9 observed in the model species *Arabidopsis thaliana* [32]. Using the mechanistic insights detailed in Chapter 2, we predicted that 19 of these genes encoded CADs that likely catalyse the 1,2-reduction of an iminium moiety, and 8 likely perform the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde. Notably, 1 CAD sequence was observed to have a novel pattern of catalytic residues, preventing the prediction of its catalytic activity, suggesting a potential further expansion of the catalytic repertoire of this enzyme family. The chromosome-length genome assembly of *C. roseus* enabled the identification of several physical clusters of CAD genes (Figure 49). Many of these clusters contained both CADs predicted to perform typical reduction, and those predicted to perform atypical reductions. These findings suggest that some of these atypical CADs in *C. roseus* emerged by tandem duplication from an ancestral aldehyde-reducing CAD, before being subsequently neofunctionalised to catalyse atypical reductions, resulting in the diversification of MIAs (Figure 52).

To further elucidate the potential biosynthetic roles of these uncharacterised CADs in *C. roseus*, we analysed their cell type specific expression patterns. Many characterised CADs such as *CrGS*, *CrRedOx1*, *CrDPAS* and *CrT3R* were found to be enriched in epidermal cells, alongside the uncharacterised *CrADH9* and *CrADH104* suggesting their potential role in MIA biosynthesis (Figure 53) [6, 30]. Genes encoding enzymes known to catalyse the late stages of MIA biosynthesis including *CrTHAS1* and *CrTHAS2* had enriched expression in idioblast cells in *C. roseus* leaves (Figure 53) [6, 30]. The uncharacterised *CrADH39* and *CrADH20* - both predicted to catalyse the 1,4-reduction of an iminium moiety – were also observed to have enriched expression in this cell type. Notably, vinblastine is suggested to form through the 1,4-reduction of the bisindole coupling iminium [31], suggesting the potential role of these CADs in the currently unknown steps of MIA biosynthesis.

3.3.2. Emergence of Atypical CADs in Gentianales Correlates with MIA Chemical Diversity

MIA biosynthesis is mostly limited to species within the Gentianales order of plants (Figure 55), with many of these pathways including steps catalysed by CADs performing atypical reduction reactions such as the 1,2- or 1,4-reduction of an iminium moiety^[10, 33, 34]. Yet, the expansion and neofunctionalisation of this gene family remained poorly understood. Therefore we performed a cross-species phylogenetic analysis and observed a monophyletic emergence of each atypical class of CAD, each likely diverging from a typical aldehyde-reducing ancestral enzyme (Figure 54). Furthermore, our findings reveal that CAD sequences predicted to encode enzymes which catalyse either *CrRedOx1*-like 1,2-iminium reduction or a 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde were only observed in MIA-producing species within the Apocynaceae family (Figure 54). This suggests that these classes of CADs that catalyse atypical reduction chemistries likely emerged after the family's divergence from Rubiaceae. Meanwhile CADs which were predicted to catalyse the 1,2-reduction activity of an iminium moiety were present in species across Gentianales (Figure 55).

To further explore the emergence CADs which perform atypical chemistries, we identified syntenic regions encoding CADs predicted to catalyse the 1,2-reduction of an iminium moiety between the genomes of *C. roseus* and *O. pumila* (Figure 98). These findings suggest that these genes emerged before the divergence of the Apocynaceae and Rubiaceae families approximately 96 million years ago^[35] and were subsequently conserved in these species. Furthermore, we identified the syntenic *OpADH2* was co-expressed with other MIA biosynthetic genes (Appendix VII Figure 99), suggesting its likely role in MIA biosynthesis. Remarkably, our phylogenetic analysis identified that the occurrence of CADs predicted encode enzymes which catalyse the 1,2-reduction of an iminium moiety to correlate with a species' ability to produce MIAs (Figure 55). Examples of enzymes from this class of CADs are known to act on early MIA biosynthetic intermediates such as strictosidine aglycone, thereby generating chemical diversity (Figure 48)^[4, 8, 9, 36]. These findings therefore suggest that the emergence and maintenance of these atypical CADs is a crucial requisite for MIA biosynthesis, driving the chemical diversity of these PNPs.

3.4. Conclusions

This chapter investigates the diversification of the catalytic repertoire of the CAD family of enzymes from performing the reduction of an aldehyde, to the 1,2-reduction of an iminium moiety or the 1,4-reduction of an iminium moiety or an α,β -unsaturated aldehyde in MIA biosynthesis. Building on the results of mechanistic investigations detailed in Chapter 2, we use phylogenetic and genomic analyses of this gene family in *C. roseus* to show that each class of atypical CAD likely independently evolved from an ancestral sequence encoding a typical aldehyde-reducing enzyme. The subsequent expansion of this gene lineage through tandem duplication and neofunctionalisation enabled the diversification of the reduction chemistries catalysed by CADs in MIA biosynthesis.

The findings described in *C. roseus* were mirrored in a cross-species phylogenetic analysis surveying the order of Gentianales. Furthermore, the identification of syntenic regions between the *C. roseus* and *O. pumila* genomes encoding CADs which catalyse atypical reductions further supports their monophyletic emergence. Additionally, we show that the presence of CADs that are predicted to catalyse the 1,2-reduction of an iminium moiety correlates to a species' ability to produce MIAs. These results therefore highlight the emergence and retention of CADs which catalyse atypical reduction chemistries as necessary evolutionary drivers of the chemical diversity of MIAs in Gentianales.

3.5. Methods and Materials

3.5.1. CAD Sequence Collection and Analysis

The sequences of CADs were retrieved from transcriptomes and genomes using a BLAST search of the full-length coding sequence of *CrDPAS*. Information about transcriptomes and genomes accessed is detailed in Table 5. Sequences were analysed using Geneious Prime and aligned using MUSCLE ^[37] using a maximum of 40 cycles. Putative CADs were verified by the presence of the four Cys residues involved in coordinating the structural zinc as detailed in Chapter 2 of this thesis. Their predicted reduction chemistry was determined by the analysis of residues in positions known to be important for the coordination of the catalytic zinc and the proton relay (see Chapter 2 for more details). Splign ^[38] was used to detect intron and exon regions of CADs in *C. roseus*. Intron and exon sequence similarity was determined by aligning sequences using MUSCLE ^[37]. The sequence alignment figure was made using ESript 3.0 ^[39].

The full-length sequence of each CAD in the *C. roseus* genome was blasted against the *C. roseus* leaf single-cell transcriptome (both reported in ^[6]) to find the corresponding contig. Contigs with a <90% sequence identity to their corresponding genome contig were used for further analysis. For single-cell transcriptome contigs matching more than one genome contig, the highest identity match was taken for further analysis. The genomic positioning of genes in *C. roseus* and *O pumila* was manually collected using the software JBrowse2 ^[40] and visualised using MG2C V2.1 ^[41].

Table 5. Transcriptomes and genomes accessed for CAD sequence collection.

Order	Family	Species	Transcriptome	MIA-producer
Brassicales	Brassicaceae	<i>Arabidopsis thaliana</i>	TAIR BLAST 2.9.0 ^[42]	✗
Lamales	Lamiaceae	<i>Lamium album</i>	In house	✗
Solanales	Solanaceae	<i>Nicotiana benthamiana</i>	NbenBase V1.0 ^[43]	✗
		<i>Solanum lycopersicum</i>	SolGenomics Tomato genome ITAG Release 2.4 ^[44]	✗

Gentianales	Loganiaceae	<i>Strychnos nux-vomica</i>	In house	✓ ^[10]
	Gentianaceae	<i>Exacum affine</i>	1000 Plants (KPUM) ^[45]	✗
		<i>Gentiana acaulis</i>	1000 Plants (ECTD) ^[45]	✗
	Gelsemiaceae	<i>Gelsemium sempervirens</i>	Published ^[21]	✓ ^[21]
	Rubiaceae	<i>Cinchona pubescens</i>	Published ^[46]	✓ ^[47]
		<i>Coffea canephora</i>	Coffee Genome Hub ^[48, 49]	✗
		<i>Galium boreale</i>	1000 Plants (WQRD) ^[45]	✗
		<i>Mitragyna speciosa</i>	In house	✓ ^[50]
		<i>Ophiorhizza pumila</i>	Published ^[3]	✓ ^[51]
		<i>Uncaria guianensis</i>	In house	✓ ^[52]
	Apocynaceae	<i>Allamanda cathartica</i>	1000 Plants (MGVU) ^[45]	✗
		<i>Amsonia hubrichtii</i>	PhytoMetaSyn ^[53]	✓ ^[54]
		<i>Apocynum androsaemifolium</i>	1000 Plants (JCLQ, UFQC) ^[45]	✗
		<i>Asclepia curassavica</i>	1000 Plants (DSUV) ^[45]	✗
		<i>Catharanthus ovalis</i>	PhytoMetaSyn ^[53]	✓ ^[55]
		<i>Catharanthus roseus</i>	Published ^[6]	✓ ^[28,56]
		<i>Rhazya stricta</i>	In house	✓ ^[57]
		<i>Rauwolfia serpentina</i>	PhytoMetaSyn ^[53]	✓ ^[58,59]
		<i>Tabernaemontana elegans</i>	In house	✓ ^[60]
		<i>Tabernanthe iboga</i>	In house	✓ ^[61, 62]
<i>Vinca minor</i>		In house	✓ ^[63]	
<i>Wrightia natalensis</i>	1000 Plants (EDEQ) ^[44]	✗		

3.5.2. CAD Phylogenetic Analysis

The nucleotide sequences for CADs extracted from *C. roseus* genome (Table 20), as well as transcriptomes and genomes of other plant species (Table 5) were aligned using MUSCLE v5^[37] using a maximum 40 cycle iterations. Maximum-likelihood phylogenetic trees of the resulting alignments were generated using the iQ-TREE web server^[64] using a best-fit substitution model. Bootstrap analysis was performed with 1000 iterations and a 0.99 minimum correlation coefficient. SH-aLRT was used to test for single branches (1000 replicates). Trees were searched using 0.5 perturbation strength and 100 tree stopping rule. All phylogenetic trees were visualised using iTol^[65].

3.5.3. Chemicals and molecular biology reagents

All solvents used for extractions, chemical synthesis and preparative HPLC were HPLC grade, and solvents used for UPLC/MS were MS grade. All solvents were purchased from Sigma Aldrich. Carbenicillin and isopropyl β -D-thiogalactoside (IPTG) salts were purchased from Sigma. All gene amplifications were performed using Platinum II Superfi DNA Polymerase (Thermo Fisher). Constructs were transformed into vectors using In-Fusion kit (ClonTech Takara) and colony PCR was performed using Phire II mastermix (Thermo Fisher) according to manufacturer's instructions. PCR product purification was performed using Zymoclean Gel DNA Recovery kit (Zymo). Plasmid purification was performed using the Wizard Miniprep kit (Promega). Precondylocarpine acetate and angryline were enzymatically prepared and purified as previously described^[56].

3.5.4. CrDPAS and CrADH9 Cloning and Protein Expression in *E. coli*

Cloning of CrDPAS has been previously reported^[57] and CrADH9 sequence was identified from the *C. roseus* transcriptome. Both sequences were codon optimised for expression in *E. coli* (Table 6) and amplified using corresponding primers listed in Table 7. PCR products were purified from 1% agarose gel and cloned into pOPINF vector (Addgene plasmid #26042^[66]) using the In-Fusion kit (Clontech Takara). Constructs were transformed into chemically-competent *E. coli* Stellar cells (Clontech Takara) by heat shock at 42°C for 30 seconds and selected on LB agar containing 50 μ g/mL carbenicillin. Positive colonies were screened by colony PCR using primers listed in Table 7 and grown overnight at 37°C shaking

at 200 r.p.m. Plasmids were then isolated and constructs were sequence verified. Plasmids were transformed into chemically competent *E. coli* SoluBL21 cells by heat shock for 30 seconds at 42°C and selected on LB agar containing 50 µg/mL carbenicillin. 10 mL starter cultures of LB with 50 µg/mL of the respective antibiotic and a colony of transformed construct in SoluBL21 cells were grown at 37°C 200 r.p.m. overnight. Media (100 mL 2xYT media) containing 50 µg/mL antibiotic was inoculated with 1 mL of the starter culture and grown until OD₆₀₀ of 0.6 was reached. Once cultures had reached the desired OD₆₀₀, cultures were transferred to 18°C 200 r.p.m shaking incubator for 30 minutes before protein expression was induced by addition of 300 µM IPTG, after which cultures were grown for an additional 16 hours.

Table 6. Full length nucleotide sequences of *CrDPAS* and *CrADH9* genes.

Codon optimised <i>CrDPAS</i>	<p>ATGGCAGGTAAAAGCGCAGAAGAAGAACATCCGATTAAAGCATATGGTTGGGC AGTTAAAGATCGTACCACCGGTATTCTGAGCCCGTTTAAATTCAGCCGTCGTGCA ACCGGTGATGATGATGTTTCGTATCAAATTCTGTATTGCGGCATTTGTCATACCG ATCTGGCAAGCATTAAAAACGAATATGAGTTTCTGAGCTATCCGCTGGTTCCTGG TATGGAAATTGTTGGTATTGCAACCGAAGTTGGTAAAGATGTGACCAAAGTTAA AGTGGGTGAAAAAGTTGCACTGAGCGCATATCTGGGTTGTTGTGGTAAATGTTA TAGCTGCGTGAATGAGCTGGAAAATATTGTCCGGAAGTGATTATTGGTTATGG CACCCCGTATCATGATGGCACCATTTGTTATGGTGGTCTGAGCAATGAAACCGTT GCAAATCAGAGCTTTGTTCTGCGTTTTCCGGAACGTCTGAGTCCGGCAGGCGGT GCTCCGCTGCTGAGCGCAGGTATTACCAGCTTAGCGCAATGCGTAATAGCGGT ATTGATAAACCGGGTCTGCATGTTGGTGTGTTGGTTTAGGTGGTCTGGGTCATC TGGCCGTTAAATTTGCAAAGCATTGGTCTGAAAGTGACCGTTATTAGCACCCAC ACCGAGCAAAAAGATGATGCAATTAATGGCCTGGGTGCAGATGGTTTTCTGCT GAGCCGTGATGACGAGCAGATGAAAGCAGCAATTGGCACCCCTGGATGCCATTAT TGATACCCTGGCAGTTGTTTCATCCGATTGCACCGCTGCTGGATCTGCTGCGTAGC CAGGGTAAATTTCTGCTGCTGGGTGCACCGAGCCAGAGCCTGGAAGTGCCTCCG ATTCCTCTGCTGAGTGGTGGTAAAAGCATTATTGGTAGCGCAGCAGGTAATGTTA AACAGACCCAAGAAATGCTGGATTTTGCAGCCGAACATGATATTACCGCCAACGT TGAAATTATCCCGATCGAATACATTAACACCGCAATGGAACGCCTGGATAAAGGT GATGTGCGTTATCGTTTTGTGGTGGATATTGAAAATACCCTGACACCGCCTAGCG AACTGTAA</p>
Codon optimised <i>CrADH9</i>	<p>ATGGCTCGTAAATCTCCGGAAGACGAACACCCGGTTAAAGCTTACGGTTGGGCT GTTAAAGACGGTACCACCGGTATCCTGTCTCCGTTCAAATTCTCTATCCGTGCTAC CGGTGACAACGACGTTTCGTATCAAATCCTGTAAGTGGTGGTTTCCGTTACCGAC CTGGCTGCTACCAAAAACGCTTTCCGTTTTCTGTCTTACCCGCTGGTTCCGGGTTT TCGTGAAATCGTTGGTATCGTTTCTGAAATCGGTAAAAACGTTAAAAAAGTTAAA GTTGGTAAAAAGTTGGTGTGCTCCGCACGTTGGTTCTTCCGGTAAATGCAAAT</p>

	<p>CTTGC GTT AACGAAGTTGAAA CTCTGCCCGAAACTGATCATCCCGTACGGTAC CCCGTACCACGACGGTACCATCTGCTACGGTGGTTTCTCTAACGAAACCGTTCGT GACGAACGTTTCGTTTTCCGTTTCCCGGAAAACCTGTCTCTGCCGGGTGGTGCTC CGCTGGTTTCTGCTGGTGTACCACCTACGGTGCTCTGCGTAACAACGGTCTGGA CAAACCGGGTCTGCACGTTGGTGTGGTGGTCTGGGTGGTCTGGGTACCTGGC TGTTAAATTCGCTAAAGCTCTGGGTGTTAAAGTTACCGTTATCTCTACCAACCCGT CTAAAGAACACGACGCTATCAACGGTTTCGGTGCTGACGCTTTCATCCTGACCCA CCACGAAGAAGAGATGAAAGCTGCTATGGGTACCCTGGACGGTATCCTGTACAC CGTTCGGTTGTTACGCTATCGCTCCGCTGCTGTCTCTGCTGGGTTCTCAGGGTA AATTCGTTCTGATCGGTGCTCCGTCTCAGCTGCTGGAAGTTCCGCCGATCCAGCT GCTGTTCCGGTGGTAAATCTATCATCGGTTCTGCTGCTGGTAACGTTAAACAGATC CAGGAAATGCTGGAATTCGCTGCTAAACACGACATCATCGCTAACGTTGAAATCA TCCAGATGGACTACATCAACACCGCTATGGAACGCTGGACAAAGGTGACGTTCT GTTACCGTTTCGTTATCGACATCGAAA CTCTGACCCTGCCGTCTGAAGTTTAA</p>
--	--

Table 7. Primer sequences used in for gene amplification. Cloning overhangs are underlined.

CrDPAS_Fwd	<u>AAGTTCTGTTTCAGGGCCCGATGGCAGGTA</u> AAAAGCGCAGAAGAAG
CrDPAS_Rev	ATGGTCTAGAAAGCTTTACAGTTCGCTAGGCGGTGTCAG
CrADH9_Fwd	<u>AAGTTCTGTTTCAGGGCCCGGCTCGTAAATCTCCGGAAG</u>
CrADH9_Rev	ATGGTCTAGAAAGCTTTAAACTTCAGACGGCAGGGTTCAG

3.5.5. Protein Purification

Cells were harvested by centrifugation at 4000 x *g* for 15 minutes and re-suspended in 10 mL buffer A1 (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) with addition of EDTA-free protease inhibitor cocktail (Roche Diagnostics Ltd.) and 10 mg lysozyme (Sigma). Cells were lysed at 4 °C using a sonicator (40% amplitude, 2 seconds on, 3 seconds off cycles for 2 minutes) and centrifuged at 35000 x *g* to remove insoluble cell debris. The supernatant was collected and filtered with 0.2 um PES syringe filter (Sartorius) and purified by addition of 150 µL washed Ni-NTA agarose beads (QIAGEN). Samples were incubated on a rocking incubator at 4 °C for 1 hour. Beads were washed by centrifuging at 1000 x *g* for 1 minute to remove the supernatant, and then the beads were resuspended in 10 mL of A1 Buffer. This step was performed a total of three times. Protein was eluted by resuspending the beads in 600 µL of buffer B1 (50 mM Tris-HCl pH 8.0, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole) before centrifuging for 1000 x *g* for 1 minute and then collecting the supernatant. This elution step

was repeated to remove all Ni-NTA bound protein. Proteins were buffer exchanged into buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl) and concentrated using 10K Da molecular weight cut off centrifugal filter (Merck) and stored at -80°C .

3.5.6. *In vitro* Enzyme Assays

Enzymatic assays with precondylocarpine acetate were performed in 50 mM HEPES buffer (pH 7.5) with 50 μM precondylocarpine acetate in MeOH (not exceeding 5% of the reaction volume), 250 μM NADPH cofactor (Sigma) and 150 nM enzyme to a final reaction volume of 100 μL . Reactions were incubated for 30 minutes at 30°C and shaking at 60 r.p.m. before being quenched with 1 volume of 70% MeOH with 0.1% HCO_2H . All enzymatic assays were centrifuged at $14000 \times g$ for 15 minutes and the supernatant analysed by UPLC-MS.

3.5.7. UPLC-MS Analysis

All assays were analysed using a Thermo Scientific Vanquish UPLC coupled to a Thermo Q Exactive Plus orbitrap MS. Chromatographic separation was performed using a Phenomenex Kinetex C18 2.6 μm (2.1 x 100 mm) column using water with 1% HCO_2H as mobile phase A and acetonitrile with 1% HCO_2H as mobile phase B. Compounds were separated using a linear gradient of 10-30% B in 5 minutes followed by 1.5 minutes isocratic at 100% B. The column was then re-equilibrated at 10% B for 1.5 minutes. The column was heated to 40°C and flow rate was set to 0.6 mL/min. MS detection was performed in positive ESI under the following conditions: spray voltage was set to 3.5 kV \sim 67.4 μA , capillary temperature set to 275°C , vaporizer temperature 475°C , sheath gas flow rate 65, sweep gas flow rate 3, aux gas flow rate 15, S-lens RF level to 55 V. Scan range was set to 200 - 1000 m/z and resolution at 17500.

3.5.8. Gene Expression and Metabolite Level Analysis

FPKMs from *C. roseus* ^[6] and *O. pumila* transcriptomes, as well as the *O. pumila* metabolite accumulation levels ^[3] were previously reported. Each dataset was normalised by z-score and clustered into a hierarchical heat map using Clustergrammer ^[67] using correlation distances and average linkage types.

3.5.9. Synteny Analysis between *C. roseus* and *O. pumila* Genomes

Synteny analysis between the genomes of *C. roseus* and *O. pumila* was performed using TBTOOLS software and MCScanX^[68, 69]. The peptide sequences of the coding regions from each genome were extracted and blast searched against one another using BlastXML. The number of hits and the number of alignments for the blast search was set to 5 and using an E-value cut-off of 1e-5. Text Merger for MCScanX was used to correctly format the genome .GFF file for MCScanX using GftGff2SmiGxt mode. MCScanX Wrapper was then run using the blast search result and the formatted .GFF file to identify collinearity between the genomes. The resulting collinearity files was used to identify gene pairs using Merger for MCScanX and the resulted plotted using Circos^[70].

3.6. References

- [1] A. E. Mohammed, Z. H. Abdul-Hameed, M. O. Alotaibi, N. O. Bawakid, T. R. Sobahi, A. Abdel-Lateff, W. M. Alarif, *Molecules* 2021, *26*, 488.
- [2] S. E. O'Connor, J. J. Maresh, *Nat Prod Rep* 2006, *23*, 532.
- [3] A. Rai, H. Hirakawa, R. Nakabayashi, S. Kikuchi, K. Hayashi, M. Rai, H. Tsugawa, T. Nakaya, T. Mori, H. Nagasaki, R. Fukushi, Y. Kusuya, H. Takahashi, H. Uchiyama, A. Toyoda, S. Hikosaka, E. Goto, K. Saito, M. Yamazaki, *Nat Commun* 2021, *12*, 405.
- [4] A. Stavrinides, E. C. Tatsis, E. Foureau, L. Caputi, F. Kellner, V. Courdavault, S. E. O'Connor, *Chem Biol* 2015, *22*, 336–41.
- [5] R. M. E. Payne, D. Xu, E. Foureau, M. I. S. T. Carqueijeiro, A. Oudin, T. D. de Bernonville, V. Novak, M. Burow, C.-E. Olsen, D. M. Jones, E. C. Tatsis, A. Pendle, B. A. Halkier, F. Geu-Flores, V. Courdavault, H. H. Nour-Eldin, S. E. O'Connor, *Nat Plants* 2017, *3*, 16208.
- [6] C. Li, J. C. Wood, A. H. Vu, J. P. Hamilton, C. E. R. Lopez, R. M. E. Payne, D. A. S. Guerrero, K. Gase, K. Yamamoto, B. Vaillancourt, L. Caputi, S. E. O'Connor, C. R. Buell, *Nat Chem Biol* 2023, 1–11.
- [7] G. Guirimand, V. Courdavault, A. Lanoue, S. Mahroug, A. Guihur, N. Blanc, N. Giglioli-Guivarc'h, B. St-Pierre, V. Burlat, *Bmc Plant Biol* 2010, *10*, 182.
- [8] E. C. Tatsis, I. Carqueijeiro, T. D. D. Bernonville, J. Franke, T.-T. T. Dang, A. Oudin, A. Lanoue, F. Lafontaine, A. K. Stavrinides, M. Clastre, V. Courdavault, S. E. O'connor, *Nat Commun* 2017, *8*, 316.
- [9] A. Stavrinides, E. C. Tatsis, L. Caputi, E. Foureau, C. E. M. Stevenson, D. M. Lawson, V. Courdavault, S. E. O'Connor, *Nat Commun* 2016, *7*, 12116.
- [10] B. Hong, D. Grzech, L. Caputi, P. Sonawane, C. E. R. López, M. O. Kamileen, N. J. H. Lozada, V. Grabe, S. E. O'Connor, *Nature* 2022, 1–6.
- [11] J. Kim, C. R. Buell, *Plant Physiol* 2015, *169*, 1532–1539.
- [12] B. R. Lichman, G. T. Godden, C. R. Buell, *Curr Opin Plant Biol* 2020, *55*, 74–83.
- [13] Z. Xu, X. Pu, R. Gao, O. C. Demurtas, S. J. Fleck, M. Richter, C. He, A. Ji, W. Sun, J. Kong, K. Hu, F. Ren, J. Song, Z. Wang, T. Gao, C. Xiong, H. Yu, T. Xin, V. A. Albert, G. Giuliano, S. Chen, J. Song, *Bmc Biol* 2020, *18*, 63.
- [14] J. A. Birchler, H. Yang, *Plant Cell* 2022, *34*, 2466–2474.
- [15] N. Panchy, M. Lehti-Shiu, S.-H. Shiu, *Plant Physiol* 2016, *171*, 2294–2316.

- [16] K. V. Krasileva, *Curr Opin Plant Biol* 2019, 48, 18–25.
- [17] S. J. Smit, B. R. Lichman, *Nat Prod Rep* 2022, 39, 1465–1482.
- [18] G. Polturak, Z. Liu, A. Osbourn, *Curr Opin Green Sustain Chem* 2021, 33, 100568.
- [19] P. Fan, P. Wang, Y.-R. Lou, B. J. Leong, B. M. Moore, C. A. Schenck, R. Combs, P. Cao, F. Brandizzi, S.-H. Shiu, R. L. Last, *Elife* 2020, 9, e56717.
- [20] D. Zhao, J. P. Hamilton, G. M. Pham, E. Crisovan, K. Wiegert-Rininger, B. Vaillancourt, D. DellaPenna, C. R. Buell, *Gigascience* 2017, 6, 1–7.
- [21] J. Franke, J. Kim, J. P. Hamilton, D. Zhao, G. M. Pham, K. Wiegert-Rininger, E. Crisovan, L. Newton, B. Vaillancourt, E. Tatsis, C. R. Buell, S. E. O'Connor, *Chembiochem* 2019, 20, 83–87.
- [22] F. Kellner, J. Kim, B. J. Clavijo, J. P. Hamilton, K. L. Childs, B. Vaillancourt, J. Cepela, M. Habermann, B. Steuernagel, L. Clissold, K. Mclay, C. R. Buell, S. E. O'Connor, *Plant J* 2015, 82, 680–692.
- [23] C. Langley, E. Tatsis, B. Hong, Y. Nakamura, C. Paetz, C. E. M. Stevenson, J. Basquin, D. M. Lawson, L. Caputi, S. E. O'Connor, *Angew. Chem. Int. Ed.* 2022, 61, e202210934.
- [24] A. K. Stavrinides, E. C. Tatsis, T.-T. Dang, L. Caputi, C. E. M. Stevenson, D. M. Lawson, B. Schneider, S. E. O'Connor, *Chembiochem* 2018, 19, 940–948.
- [25] S. de Vries, J. M. R. Fürst-Jansen, I. Irisarri, A. D. Ashok, T. Ischebeck, K. Feussner, I. N. Abreu, M. Petersen, I. Feussner, J. de Vries, *The Plant Journal* 2021, DOI 10.1111/tpj.15387.
- [26] D.-M. Guo, J.-H. Ran, X.-Q. Wang, *J Mol Evol* 2010, 71, 202–218.
- [27] Y. Qu, M. E. A. M. Easson, R. Simionescu, J. Hajicek, A. M. K. Thamm, V. Salim, V. D. Luca, *Proc National Acad Sci* 2018, 115, 3180–3185.
- [28] Y. Qu, M. L. A. E. Easson, J. Froese, R. Simionescu, T. Hudlicky, V. DeLuca, *Proc National Acad Sci* 2015, 112, 6224–6229.
- [29] A. Edge, Y. Qu, M. L. A. E. Easson, A. M. K. Thamm, K. H. Kim, V. D. Luca, *Planta* 2018, 247, 155–169.
- [30] K. Yamamoto, K. Takahashi, L. Caputi, H. Mizuno, C. E. Rodriguez-Lopez, T. Iwasaki, K. Ishizaki, H. Fukaki, M. Ohnishi, M. Yamazaki, T. Masujima, S. E. O'Connor, T. Mimura, *New Phytol* 2019, 224, np.16138.
- [31] N. Langlois, P. Potier, *J. Chem. Soc., Chem. Commun.* 1979, 0, 582–584.
- [32] S.-J. Kim, M.-R. Kim, D. L. Bedgar, S. G. A. Moinuddin, C. L. Cardenas, L. B. Davin, C. Kang, N. G. Lewis, *Proc National Acad Sci* 2004, 101, 1455–1460.

- [33] F. Trenti, K. Yamamoto, B. Hong, C. Paetz, Y. Nakamura, S. E. O'Connor, *Org Lett* 2021, 23, 1793–1797.
- [34] M. Geissler, M. Burghard, J. Volk, A. Staniek, H. Warzecha, *Planta* 2016, 243, 813–824.
- [35] N. Wikström, K. Kainulainen, S. G. Razafimandimbison, J. E. E. Smedmark, B. Bremer, *PLoS ONE* 2015, 10, e0126690.
- [36] Y. Qu, A. M. K. Thamm, M. Czerwinski, S. Masada, K. H. Kim, G. Jones, P. Liang, V. D. Luca, *Planta* 2018, 247, 625–634.
- [37] R. C. Edgar, *Biorxiv* 2021, 2021.06.20.449169.
- [38] Y. Kapustin, A. Souvorov, T. Tatusova, D. Lipman, *Biol Direct* 2008, 3, 20.
- [39] X. Robert, P. Gouet, *Nucleic Acids Res* 2014, 42, W320–W324.
- [40] C. Diesh, G. J. Stevens, P. Xie, T. D. J. Martinez, E. A. Hershberg, A. Leung, E. Guo, S. Dider, J. Zhang, C. Bridge, G. Hogue, A. Duncan, M. Morgan, T. Flores, B. N. Bimber, R. Haw, S. Cain, R. M. Buels, L. D. Stein, I. H. Holmes, *Biorxiv* 2022, 2022.07.28.501447.
- [41] J. Chao, Z. Li, Y. Sun, O. O. Aluko, X. Wu, Q. Wang, G. Liu, *Mol Horticulture* 2021, 1, 16.
- [42] T. Z. Berardini, L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, E. Huala, *Genesis* 2015, 53, 474–485.
- [43] K. Kurotani, H. Hirakawa, K. Shirasawa, Y. Tanizawa, Y. Nakamura, S. Isobe, M. Notaguchi, *Plant Cell Physiol* 2023, 64, 248–257.
- [44] N. Fernandez-Pozo, N. Menda, J. D. Edwards, S. Saha, I. Y. Tecle, S. R. Strickler, A. Bombarely, T. Fisher-York, A. Pujar, H. Foerster, A. Yan, L. A. Mueller, *Nucleic Acids Res* 2015, 43, D1036–D1041.
- [45] J. H. Leebens-Mack, M. S. Barker, E. J. Carpenter, M. K. Deyholos, M. A. Gitzendanner, S. W. Graham, I. Grosse, Z. Li, M. Melkonian, S. Mirarab, M. Porsch, M. Quint, S. A. Rensing, D. E. Soltis, P. S. Soltis, D. W. Stevenson, K. K. Ullrich, N. J. Wickett, L. DeGironimo, P. P. Edger, I. E. Jordon-Thaden, S. Joya, T. Liu, B. Melkonian, N. W. Miles, L. Pokorny, C. Quigley, P. Thomas, J. C. Villarreal, M. M. Augustin, M. D. Barrett, R. S. Baucom, D. J. Beerling, R. M. Benstein, E. Biffin, S. F. Brockington, D. O. Burge, J. N. Burris, K. P. Burris, V. Burtet-Sarramegna, A. L. Caicedo, S. B. Cannon, Z. Çebi, Y. Chang, C. Chater, J. M. Cheeseman, T. Chen, N. D. Clarke, H. Clayton, S. Covshoff, B. J. Crandall-Stotler, H. Cross, C. W. dePamphilis, J. P. Der, R. Determann, R. C. Dickson, V. S. D. Stilio, S. Ellis, E. Fast, N. Feja, K. J. Field, D. A. Filatov, P. M. Finnegan, S. K. Floyd, B. Fogliani, N. García, G. Gâteblé, G. T. Godden, F. (Qi Y. Goh, S. Greiner, A. Harkess, J. M. Heaney, K. E. Helliwell, K. Heyduk, J. M. Hibberd, R. G. J. Hodel, P. M. Hollingsworth, M. T. J. Johnson, R. Jost, B. Joyce, M. V. Kapralov, E. Kazamia, E. A. Kellogg, M. A. Koch, M. V. Konrat, K. Könyves, T. M. Kutchan, V. Lam, A. Larsson, A. R. Leitch, R. Lentz, F.-W. Li, A. J. Lowe, M. Ludwig, P. S. Manos, E. Mavrodiev, M. K. McCormick, M. McKain, T. McLellan, J. R. McNeal, R. E. Miller, M. N. Nelson, Y. Peng, P. Ralph, D. Real,

C. W. Riggins, M. Ruhsam, R. F. Sage, A. K. Sakai, M. Scascitella, E. E. Schilling, E.-M. Schlösser, H. Sederoff, S. Servick, E. B. Sessa, A. J. Shaw, S. W. Shaw, E. M. Sigel, C. Skema, A. G. Smith, A. Smithson, C. N. Stewart, J. R. Stinchcombe, P. Szövényi, J. A. Tate, H. Tiebel, D. Trapnell, M. Villegente, C.-N. Wang, S. G. Weller, M. Wenzel, S. Weststrand, J. H. Westwood, D. F. Whigham, S. Wu, A. S. Wulff, Y. Yang, D. Zhu, C. Zhuang, J. Zuidof, M. W. Chase, J. C. Pires, C. J. Rothfels, J. Yu, C. Chen, L. Chen, S. Cheng, J. Li, R. Li, X. Li, H. Lu, Y. Ou, X. Sun, X. Tan, J. Tang, Z. Tian, F. Wang, J. Wang, X. Wei, X. Xu, Z. Yan, F. Yang, X. Zhong, F. Zhou, Y. Zhu, Y. Zhang, S. Ayyampalayam, T. J. Barkman, N. Nguyen, N. Matasci, D. R. Nelson, E. Sayyari, E. K. Wafula, R. L. Walls, T. Warnow, H. An, N. Arrigo, A. E. Baniaga, S. Galuska, S. A. Jorgensen, T. I. Kidder, H. Kong, P. Lu-Irving, H. E. Marx, X. Qi, C. R. Reardon, B. L. Sutherland, G. P. Tiley, S. R. Welles, R. Yu, S. Zhan, L. Gramzow, G. Theißen, G. K.-S. Wong, *Nature* 2019, 574, 679–685.

[46] N. A. Canales, O. A. Pérez-Escobar, R. F. Powell, M. Töpel, C. Kidner, M. Nesbitt, C. Maldonado, C. J. Barnes, N. Rønsted, N. A. S. Przelomska, I. J. Leitch, A. Antonelli, *Gigabyte* 2022, 2022, 1–16.

[47] A. Murauer, M. Ganzera, *J Chromatogr A* 2018, 1554, 117–122.

[48] A. Dereeper, S. Bocs, M. Rouard, V. Guignon, S. Ravel, C. Tranchant-Dubreuil, V. Poncet, O. Garsmeur, P. Lashermes, G. Droc, *Nucleic Acids Res* 2015, 43, D1028–D1035.

[49] F. Denoeud, L. Carretero-Paulet, A. Dereeper, G. Droc, R. Guyot, M. Pietrella, C. Zheng, A. Alberti, F. Anthony, G. Aprea, J.-M. Aury, P. Bento, M. Bernard, S. Bocs, C. Campa, A. Cenci, M.-C. Combes, D. Crouzillat, C. D. Silva, L. Daddiego, F. D. Bellis, S. Dussert, O. Garsmeur, T. Gayraud, V. Guignon, K. Jahn, V. Jamilloux, T. Joët, K. Labadie, T. Lan, J. Leclercq, M. Lepelley, T. Leroy, L.-T. Li, P. Librado, L. Lopez, A. Muñoz, B. Noel, A. Pallavicini, G. Perrotta, V. Poncet, D. Pot, Priyono, M. Rigoreau, M. Rouard, J. Rozas, C. Tranchant-Dubreuil, R. VanBuren, Q. Zhang, A. C. Andrade, X. Argout, B. Bertrand, A. de Kochko, G. Graziosi, R. J. Henry, Jayarama, R. Ming, C. Nagai, S. Rounsley, D. Sankoff, G. Giuliano, V. A. Albert, P. Wincker, P. Lashermes, *Science* 2014, 345, 1181–1184.

[50] R. Veeramohan, A. I. Zamani, K. A. Azizan, H.-H. Goh, W. M. Aizat, M. F. A. Razak, N. S. M. Yusof, S. M. Mansor, S. N. Baharum, C. L. Ng, *Plos One* 2023, 18, e0283147.

[51] M. Yang, Q. Wang, Y. Liu, X. Hao, C. Wang, Y. Liang, J. Chen, Y. Xiao, G. Kai, *Bmc Biol* 2021, 19, 122.

[52] A. A. Lopes, B. Chioca, B. Musquiari, E. J. Crevelin, S. de C. França, M. F. das G. F. da Silva, A. M. S. Pereira, *Sci Rep-uk* 2019, 9, 11349.

[53] M. Xiao, Y. Zhang, X. Chen, E. J. Lee, C. J. S. Barber, R. Chakrabarty, I. Desgagné-Penix, T. M. Haslam, Y. B. Kim, E. Liu, G. MacNevin, S. Masada-Atsumi, D. W. Reed, J. M. Stout, P. Zerbe, Y. Zhang, J. Bohlmann, P. S. Covello, V. D. Luca, J. E. Page, D. K. Ro, V. J. J. Martin, P. J. Facchini, C. W. Sensen, *J Biotechnol* 2013, 166, 122–134.

[54] D. Williams, Y. Qu, R. Simionescu, V. D. Luca, *Plant J* 2019, 99, 626–636.

- [55] N. Langlois, R. Z. Andriamialisoa, N. Neuss, *Helv. Chim. Acta* 1980, *63*, 793–805.
- [56] L. Caputi, J. Franke, S. C. Farrow, K. Chung, R. M. E. Payne, T.-D. Nguyen, T.-T. T. Dang, I. S. T. Carqueijeiro, K. Koudounas, T. D. de Bernonville, B. Ameyaw, D. M. Jones, I. J. C. Vieira, V. Courdavault, S. E. O'Connor, *Science* 2018, *360*, 1235–1239.
- [57] A. Albeshri, N. A. Baeshen, T. A. Bouback, A. A. Aljaddawi, *Plants* 2021, *10*, 2508.
- [58] T.-T. T. Dang, J. Franke, I. Soares, T. Carqueijeiro, C. Langley, V. Courdavault, S. E. O'Connor, *Nat Chem Biol* 2018, *14*, 760–763.
- [59] B. Wiens, V. D. Luca, *Phytochemistry* 2016, *132*, 5–15.
- [60] A. Paterna, S. E. Gomes, P. M. Borralho, S. Mulhovo, C. M. P. Rodrigues, M.-J. U. Ferreira, *J Nat Prod* 2016, *79*, 2624–2634.
- [61] S. C. Farrow, M. O. Kamileen, L. Caputi, K. Bussey, J. E. A. Mundy, R. C. McAtee, C. R. J. Stephenson, S. E. O'Connor, *J Am Chem Soc* 2019, *141*, 12979–12983.
- [62] M. O. Kamileen, M. D. DeMars, B. Hong, Y. Nakamura, C. Paetz, B. R. Lichman, P. D. Sonawane, L. Caputi, S. E. O'Connor, *J Am Chem Soc* 2022, DOI 10.1021/jacs.2c08107.
- [63] S. Abouzeid, T. Hijazin, L. Lewerenz, R. Hänsch, D. Selmar, *Phytochemistry* 2019, *168*, 112110.
- [64] J. Trifinopoulos, L.-T. Nguyen, A. von Haeseler, B. Q. Minh, *Nucleic Acids Res* 2016, *44*, W232–W235.
- [65] I. Letunic, P. Bork, *Nucleic Acids Res* 2021, *49*, gkab301-.
- [66] N. S. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. I. Stuart, R. J. Owens, *Nucleic Acids Res* 2007, *35*, e45–e45.
- [67] N. F. Fernandez, G. W. Gundersen, A. Rahman, M. L. Grimes, K. Rikova, P. Hornbeck, A. Ma'ayan, *Sci Data* 2017, *4*, 170151.
- [68] Y. Wang, H. Tang, J. D. DeBarry, X. Tan, J. Li, X. Wang, T. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, *Nucleic Acids Res* 2012, *40*, e49–e49.
- [69] C. Chen, H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He, R. Xia, *Mol Plant* 2020, *13*, 1194–1202.
- [70] M. Krzywinski, J. Schein, Í. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, M. A. Marra, *Genome Res* 2009, *19*, 1639–1645.

Chapter 4. The Role of Protein-Protein Interactions in MIA Biosynthesis

4.1. Introduction

4.1.1. Protein-Protein Interactions in MIA Biosynthesis

Extensive research has sought to elucidate the biochemistry and physiology of vinblastine biosynthesis in *C. roseus* [1–10]. This complex pathway occurs over three different cell types [4, 11–14], with some steps localised within specific subcellular compartments [15–17] (Figure 58). Moreover, biosynthesis of this PNP involves the generation of unstable intermediates that can be catalysed by several competing enzymes to produce different products, thereby acting as metabolic branch points (Figure 59). Given the complexity and the presence of labile intermediates, protein-protein interactions have been speculated to help physically organise and/or facilitate MIA biosynthesis [2, 16]. Biosynthetic protein complexes have been previously identified at two metabolic branch points within *C. roseus* MIA biosynthesis: the formation of strictosidine aglycone and the generation of dehydrosecodine.

The *Cr*STR catalyses the first committed step of MIA biosynthesis production to generate the central precursory molecule strictosidine from which over 2000 MIAs derive [18]. The subsequent deglycosylation of strictosidine, catalysed by *Cr*SGD, generates the highly

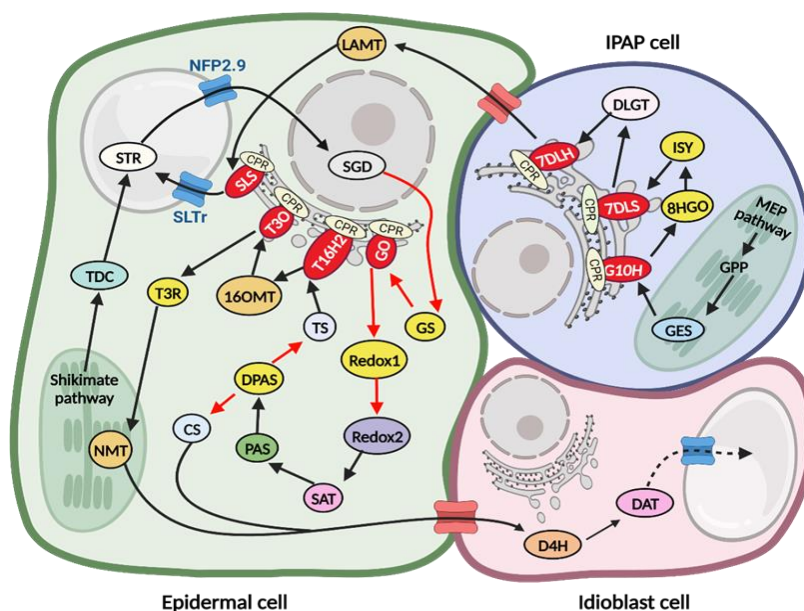


Figure 58. Cellular and subcellular localisation of elucidated steps of vinblastine biosynthesis in *C. roseus*. Iridoid and early secoiridoid biosynthesis predominantly occurs in internal phloem associated parenchyma (IPAP) cells, late secoiridoid and early MIA biosynthesis occurs in epidermal cells, and late MIA biosynthetic steps occur in idioblast cells. Steps with unstable intermediates indicated by red arrows. Figure made using BioRender.

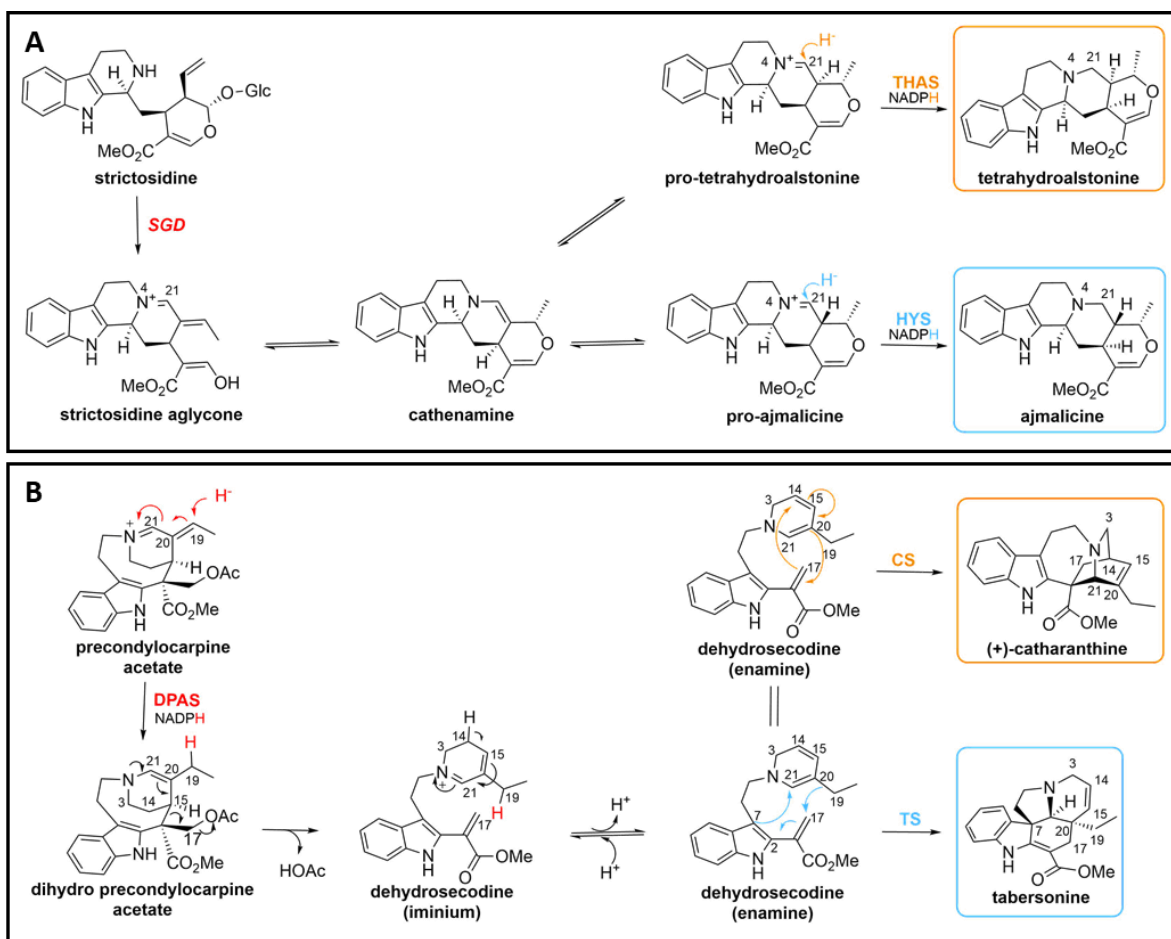


Figure 59. Previously studied enzyme-enzyme complexes in *C. roseus* MIA biosynthesis. **A.** *CrSGD*-catalysed deglycosylation of strictosidine and subsequent *CrTHAS*- or *CrHYS*-catalysed reduction of strictosidine aglycone. **B.** *CrDPAS* catalysed reduction of precondylocarpine acetate and subsequent cyclisation of dehydrosecodeine by cyclase enzymes *CrCS* or *CrTS* to form catharanthine or tabersonine, respectively.

reactive aglycone (Figure 59A). Reductases such as CADs and an SDR act on the structural rearrangements of strictosidine aglycone to produce a range of MIAs [9, 16, 19–21]. Previous studies found that *CrSGD* forms protein-protein complexes with the downstream enzymes *CrTHAS* and *CrHYS*, as evidenced by bimolecular fluorescence complementation (BiFC) imaging and affinity-pulldown studies [16, 19]. This suggests protein-protein complexes may have a role in controlling the metabolic flux at this branch point.

Protein-protein complexes have also been reported at the metabolic branch point of the generation of the highly unstable intermediate dehydrosecodeine. As detailed in Chapter 2 of this thesis, the CAD *CrDPAS* catalyses the reduction of substrate precondylocarpine acetate to form dehydrosecodeine, which is subsequently cyclised by the cyclase enzymes

(Figure 59B). Specifically, the enzyme *CrTS* generates the aspidosperma-type alkaloid tabersonine, which undergoes 7 additional enzymatic steps to form vindoline, and the enzyme *CrCS* generates the iboga-type alkaloid catharanthine. Vindoline and catharanthine are hypothesised to be the precursors for bis-indole alkaloids including vinblastine. Enzyme-enzyme complexes between *CrDPAS* and either *CrTS* or *CrCS* were observed using BiFC imaging ^[2], suggesting that these interactions may regulate the metabolic flux of dehydrosecodine and thereby the subsequent downstream MIAs. Interestingly, orthologues of DPAS and the cyclase enzymes have been identified in the closely related MIA-producing species *T. iboga*, though these were not tested for the formation of homologous protein-protein complexes ^[22, 23]. Studying this metabolic branch point could therefore provide valuable insights into the functional relevance and conservation of protein-protein interactions within MIA biosynthesis.

These previous reports prompted us to investigate the extent of protein-protein complexes in MIA biosynthesis in *C. roseus*. Our studies aimed to validate the interaction between *CrDPAS* and the cyclase enzymes and to understand the structural basis of this interaction. Furthermore, we utilised the orthologous enzymes from *T. iboga* to study whether these proteins complexes are conserved in closely related species, and to explore the formation of inter-pathway interactions between MIA and phenylpropanoid biosynthetic enzymes.

4.1.2. Methods Used to Detect Protein-protein Interactions

This chapter discusses various methods used to detect and study protein-protein interactions in MIA biosynthesis, taking into account the diversity of strength and temporal stability of these interactions. A summary of the methods used is provided below.

To identify novel interaction partners for a specific protein of interest (referred to as the bait protein), higher-throughput approaches are often employed. One commonly used technique is the affinity-purification coupled with mass-spectrometry (AP-MS). In this method, the bait protein is fused with a purification tag and exposed to a mixture of potential interacting proteins (Figure 60A). The bait protein is subsequently purified using affinity chromatography based on the specific tag along with any interacting proteins and the composition of the sample is identified using proteomic analysis ^[24]. Stringent washing

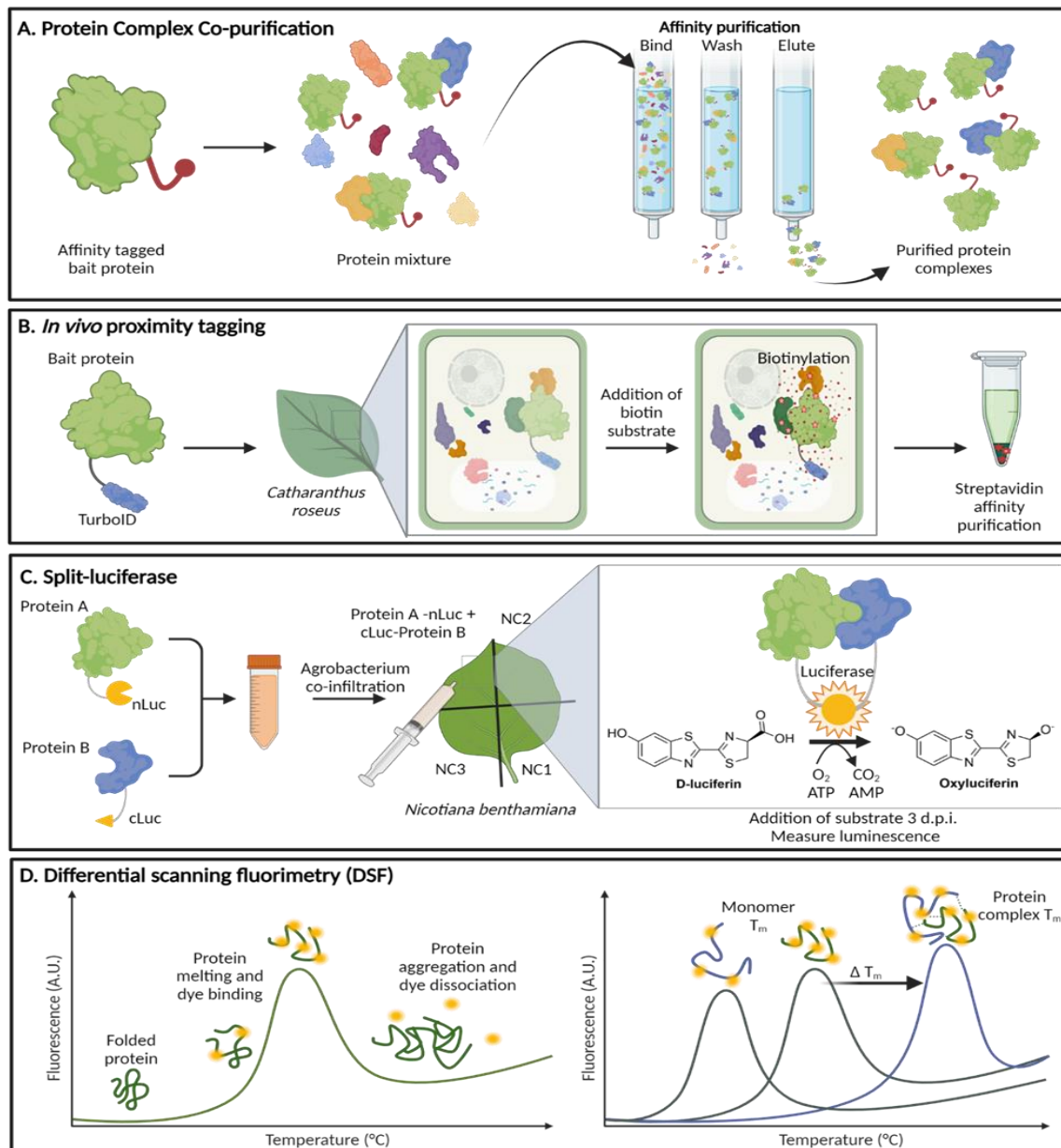


Figure 60. Overview of methods used to detect protein-protein interactions discussed in this chapter. **A.** Protein complex co-purification using affinity purification of tagged bait protein. **B.** *In vivo* proximity tagging using promiscuous biotin ligase TurboID and subsequent streptavidin-affinity purification of biotinylated proteins. **C.** Split-luciferase complementation assay to detect pairwise protein-protein interactions by luminescence. **D.** Differential scanning fluorimetry of protein thermal stability. Figure made using Biorender.

steps are typically applied during the purification process to reduce false-positive identifications of interacting proteins. However, these may result in the loss of weak or transient interactions, making this method more suitable for detecting stronger protein complexes.

Proximity tagging is an alternative method that can be particularly useful in detecting weak and or transient protein complexes *in vivo* (Figure 60B) [25, 26]. In this technique, the bait protein is fused with a promiscuous biotin ligase enzyme known as BirA, though other variants such as TurboID have since been developed [27, 28]. This fusion protein is then transiently or stably expressed in the plant. Upon addition of biotin, BirA catalyses the formation of biotinoyl-AMP, which covalently bonds to nearby primary amine groups such as lysine side chains. Due to the short half-life of biotinoyl-AMP, only primary amines within approximately 8 Å of the BirA protein are labelled. The biotin-labelled proteins in the sample are then extracted and enriched using streptavidin-affinity purification and subsequently identified by proteomic analysis [29]. This technique captures the *in vivo* interactions before further processing of the sample, enabling the detection of protein complexes that may not be observable using co-purification techniques.

After identification using high-throughput screening methods, protein complexes are validated using various biophysical techniques. Among these, fragment complementation assays are commonly employed to test pairwise protein-protein interactions due to their development for use in a variety of organisms and utilisation of various reporter proteins including fluorophores (e.g. BiFC), luciferases (e.g. split-luciferase), or antibiotic resistance (e.g. yeast two-hybrid). As detailed in section 4.1, protein complexes in MIA biosynthesis were previously detected using BiFC. This method requires each protein of interest to be fused to a fluorophore fragment. Upon formation of a protein complex, the fusion proteins irreversibly reconstitute to form a functional fluorophore reporter protein. However, the irreversible nature of the fluorophore reconstitution can lead to false-positive results, prompting the validation of protein-protein complexes using alternative techniques.

Split-luciferase is a fragment complementation assay, which like BiFC, requires each protein of interest to be fused with a non-functional luciferase protein fragment. However, unlike BiFC, the fragments can reversibly reconstitute to form a functional luciferase enzyme (Figure 60C) [30]. These fusion constructs are transiently expressed in the leaves of the host plant *N. benthamiana* to mitigate common plant-derived protein expression issues such as mislocalisation and misfolding. After several days, the substrate D-luciferin is added and oxidised by the reconstituted luciferase enzyme, forming light as a reaction by-product. The

formation of the protein-protein complex is therefore detected by measuring light emittance using a luminometer or a cooled charge-coupled device (CCD) camera.

The formation of protein-protein complexes is driven by the formation of hydrogen bonds, disulphide bridges and/or electrostatic forces between the partner proteins. These bonds contribute to increased thermal stability (T_m), enabling the measurement of protein complexes through differential scanning fluorimetry (DSF; Figure 60D) [31]. DSF uses a fluorescent dye reporter molecule, typically SYPRO Orange, which is quenched by water when initially incubated with the protein/proteins of interest. However, heating the sample causes protein secondary structure motifs to unfold, enabling the dye to bind to hydrophobic regions and thereby leading to fluorophore emission (T_m). Further increased temperatures induces protein aggregation, causing the dye to dissociate and return to its quenched state. Observing an increased T_m in a heterogeneous mixture of proteins compared to a homogeneous solution thus validates the formation of a protein complex and can provide insights into the stoichiometric ratio of the monomers.

4.1.3. Chapter Overview

In this chapter, we use a split-luciferase assay to investigate the pairwise interactions among 17 enzymes involved in vinblastine biosynthesis in *C. roseus*. Our findings revealed a complex network of interactions, shedding light on the intricate interplay among these enzymes. We validate the previously reported interaction between the enzyme *CrDPAS* and cyclase enzymes *CrTS* and *CrCS* and explore the conservation of this interaction across enzyme families. Furthermore, we use site-directed mutagenesis to engineer the interaction between *CrDPAS* and the cyclase enzymes. Additionally, we uncover inter-pathway interactions between MIA and phenylpropanoid biosynthetic pathways and provide evidence that suggest these protein-protein interactions may have a functional role in lignin biosynthesis in *C. roseus*.

4.2. Results

4.2.1. Interaction Network of *C. roseus* MIA Biosynthetic Proteins using Split-Luciferase

Single-cell transcriptomic analysis of *C. roseus* leaves previously demonstrated that a group of 19 consecutive MIA biosynthetic enzymes had enriched expression in epidermal cells. These enzymes included *CrLAMT*, *CrSLS*, *CrTDC*, *CrSTR*, *CrSGD*, *CrGS*, *CrGO*, *CrRedOx1*, *CrRedOx2*, *CrSAT*, *CrPAS*, *CrDPAS*, *CrTS*, *CrCS*, *CrCorS*, *CrT16H2*, *Cr16OMT*, *CrT3O* and *CrT3R* (Figure 61) [4]. With the exceptions of *CrSTR* and *CrSGD*, all of these enzymes were found to be localised in the cytosol or anchored to the cytosolic face of the endoplasmic reticulum membrane, implying that they are physically accessible to each other (Appendix VIII; Table 23). The co-localisation of these proteins in the same cell type is consistent with a model in which some or all of these biosynthetic enzymes interact with one another. Therefore, to gain insight into the extent of protein-protein interactions among these MIA biosynthetic enzymes, we tested the pairwise interactions of these 17 enzymes by transiently expressing proteins pairs in the heterologous host *N. benthamiana* using a split-luciferase system (Figure 61; Appendix IX). These assays revealed a complex network of protein-protein interactions, with certain enzymes (i.e. *CrTS*, *CrCS*, and *CrT3R*) interacting with many MIA

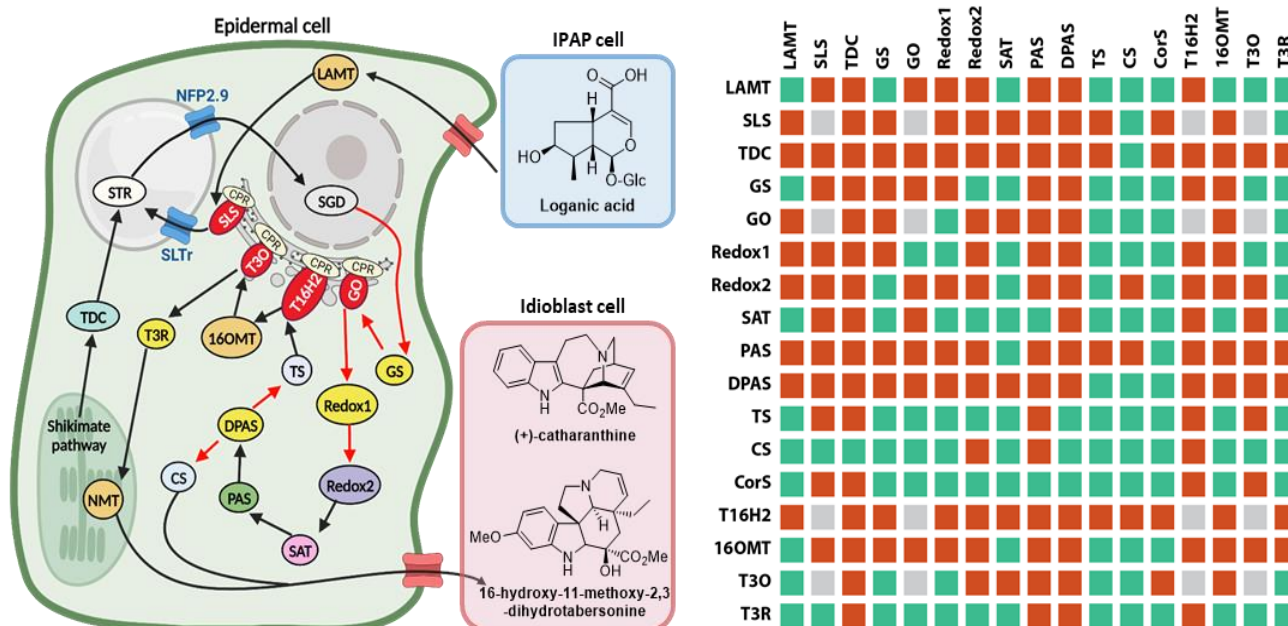


Figure 61. Pairwise testing of protein-protein interactions of *C. roseus* MIA biosynthetic enzymes enriched in epidermal cells using a split-luciferase assay. Green indicates interaction detected, red indicates no interaction detected, N/A refers to protein pairs not tested for interactions due to inaccessibility of luciferase fragment caused by each protein anchoring within the endoplasmic reticulum membrane.

enzymes are members of the α/β hydrolase superfamily, though these enzymes have undergone neofunctionalisation to act in MIA biosynthesis as detailed in Chapter 2.

The instability of dehydrosecodine raised speculation that DPAS and the cyclase enzymes formed protein-protein complexes, which were subsequently observed using BiFC imaging [2] and a split-luciferase assay (Figure 61). To gain insights into the conservation of this interaction, we expressed pairs of other ADHs and α/β hydrolase proteins in the heterologous host *N. benthamiana* and tested their pairwise interactions using a split-luciferase assay (Figure 63; Appendix X). Interestingly, despite their diverse functional roles, species of origin, and sequence identities (Appendix X Table 24 and Table 25), interactions

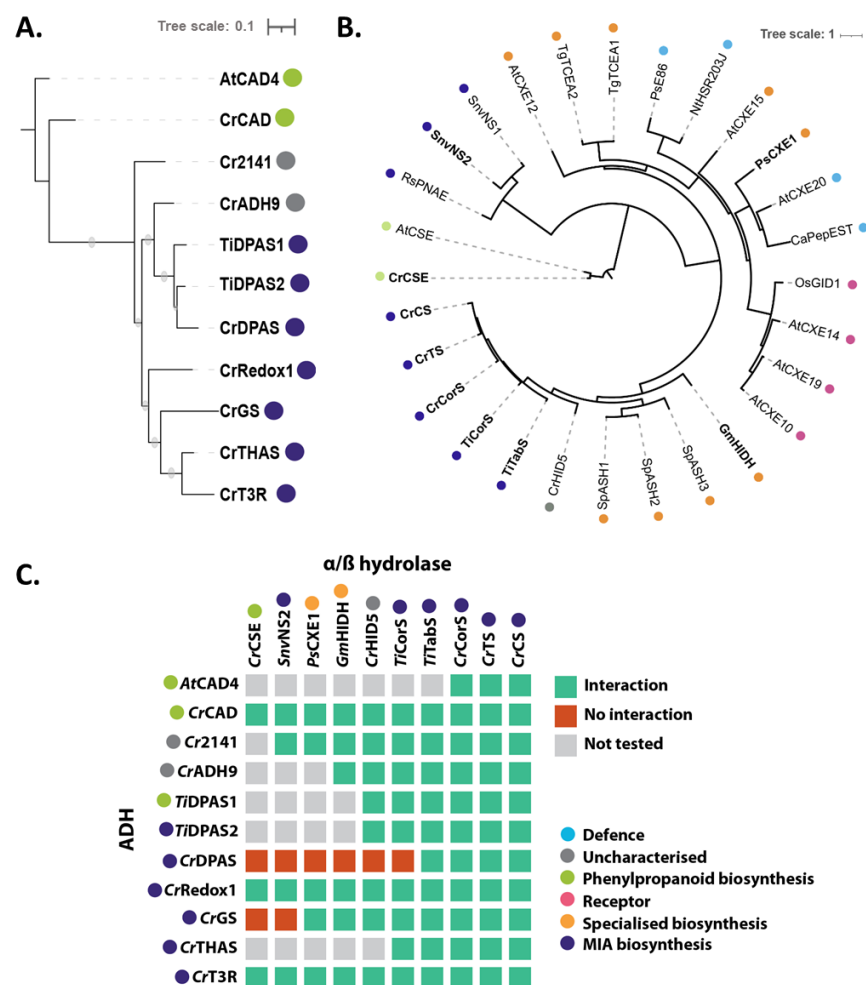


Figure 63. Protein-protein interactions between ADHs and α/β hydrolases. Coloured circles represent characterised enzyme function. Trees of maximum likelihood of CAD-like ADHs (**A**) and α/β hydrolases (**B**). Trees visualised using iTol [54], genes in bold tested for interactions. **C.** Pairwise interactions between CAD-like ADHs and α/β hydrolases by split-luciferase in heterologous host *N. benthamiana*.

between ADHs and α/β hydrolase enzymes were largely conserved. For instance, *Arabidopsis thaliana* CAD4, known for its role in lignin biosynthesis^[34], interacted with *C. roseus* cyclase enzymes that participate in MIA biosynthesis^[32]. The interactions between MIA and phenylpropanoid biosynthetic enzymes is expanded on in section 4.2.7. Furthermore, this interaction was conserved between *T. iboga* DPAS (*TiDPAS1* and *TiDPAS2*) and cyclase (*TiTabS* and *TiCorS*) orthologues, suggesting its functional role in other MIA producing species. However, some ADHs exhibited interaction specificity, such as *CrGS* which interacted with cyclase enzymes from both *C. roseus* and *T. iboga*, but not with more distantly related α/β hydrolases. Notably, *CrDPAS* interacted with the cyclase enzymes *CrTS*, *CrCS*, *CrCorS*, and *TiTabS*, but did not interact with *TiCorS*, as expanded on in section 4.2.3.

These findings highlight the conservation of protein-protein interactions between the ADH and α/β hydrolase families of enzymes. The conservation of these interactions across diverse plant species provides valuable insights into the evolution of protein complexes, and the potential metabolic interplay between plant biosynthetic pathways.

4.2.3. Engineering the *CrDPAS*-Cyclase Interaction

Our findings using a split-luciferase assay revealed that *CrDPAS* interacted with cyclase enzymes *CrCS*, *CrTS*, *CrCorS*, and *TiTabS*, but not with the closely related *TiCorS* (Figure 64A-D). We hypothesised that the loss of interaction between *CrDPAS* and *TiCorS* was attributed to residue changes on the surface of the cyclase enzymes. To this end, we utilised the previously solved structures of *CrCS*, *CrTS* and *TiCorS*^[32], and generated homology models of *CrCorS* and *TiTabS*. By comparison of the enzyme surfaces, we identified 5 residues conserved in all interacting cyclase enzymes but lost in *TiCorS*. However the corresponding *CrCorS* and *TiCorS* mutants did not confer the desired loss or introduction of interaction with *CrDPAS* (Appendix XI Figure 130E-H). Expanding our search, we identified all the differing surface residues between *CrCorS* and *TiCorS*, regardless of their conservation in the remaining cyclase enzymes. This led to the generation of 4-residue mutants of *CrCorS* and *TiCorS*. When we tested these mutants against *CrDPAS* using a split luciferase assay, we observed the engineered loss and introduction of a protein-protein interaction respectively (Appendix XI Figure 130I-L).

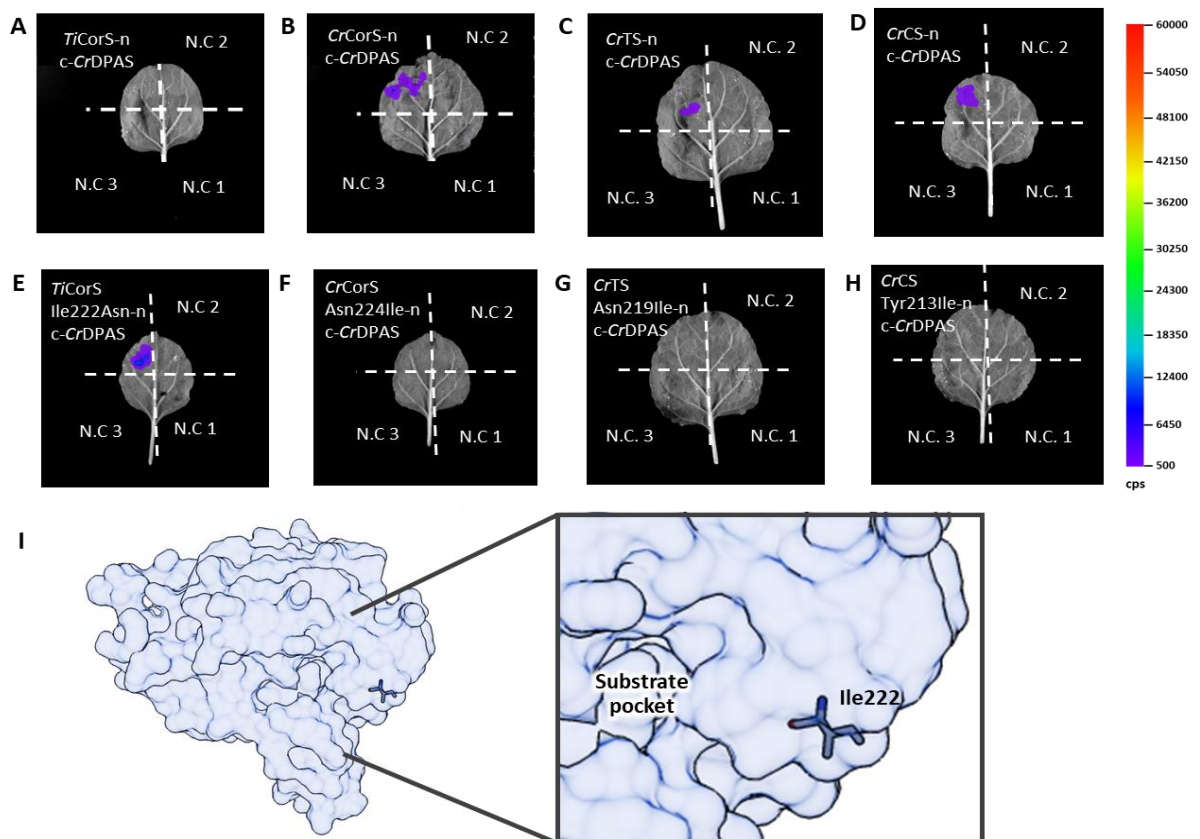


Figure 64. Engineering *CrDPAS*-cyclase interaction. Representative images of split-luciferase interaction between *CrDPAS* and wild-type *TiCorS* (A), *CrCorS* (B), *CrTS* (C), and *CrCS* (D), and corresponding mutants *TiCorS* Ile222Asn (E), *CrCorS* Asn224Ile (F), *CrTS* Asn219Ile (G), and *CrCS* Tyr213Ile (H). –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein and empty cLuc fragment, N.C. 2 represents cLuc-tagged protein and empty nLuc fragment, N.C. 3 represents nLuc and cLuc fragment negative controls. False colouring on images represents measured luminescence counts per second (cps). I. Surface view of *TiCorS* monomer structure (PDB 6RJ8^[32]) with inset showing Ile222 residue.

Subsequently, we generated single mutants based on this 4-residue mutant and identified *CrCorS* Asn224Ile and the corresponding *TiCorS* Ile222Asn as the crucial residue responsible for the loss or introduction of cyclase interaction with *CrDPAS* (Figure 64E and F; Appendix XI). Furthermore, introduction of the corresponding point mutation to *CrTS* or *CrCS* (Asn219Ile and Tyr213Ile respectively) abolished interaction with *CrDPAS* (Figure 64G and H). This suggests that the polar Asn, in contrast to the hydrophobic and less reactive Ile,

interacts with another polar or charged residue on the surface of *CrDPAS*, thereby facilitating the protein-protein interaction.

These findings provide the foundation for comprehending the structural basis of the interaction between *CrDPAS* and the cyclase enzymes. Future work to understand the metabolic effect of introducing or disrupting these enzyme-enzyme will contribute to our understanding of the metabolic role of protein-protein interactions in MIA biosynthesis.

4.2.4. AP-MS Analysis of *CrDPAS* and *CrTS*

AP-MS is a widely used technique to validate protein complexes observed using other methods and to identify novel interacting partners. In the context of our study, we sought to validate the complex observed between *CrDPAS* and the cyclase enzymes using BiFC^[2] and split-luciferase assays (Figure 61). To achieve this, we introduced 6X-His fusion tags to *CrDPAS* and *CrTS* proteins and added these to protein extracts from *C. roseus* leaves. Samples were subsequently purified using affinity chromatography and the resulting fractions were analysed by proteomics (Appendix XII Table 26). Surprisingly, our analysis did not reveal any MIA biosynthetic proteins that co-purified with *CrDPAS* or *CrTS*. This suggests that these interactions were too weak and/or transient to be detected, or require additional factors or conditions not present in this study. Despite not detecting direct interactions with MIA biosynthetic proteins, these findings provide valuable information about the stability and strength of the *CrDPAS*-cyclase complex.

4.2.5. Differential Scanning Fluorimetry Analysis of *CrDPAS*-Cyclase Complex

The increased T°_m of a mixture of proteins compared to their individual components has been established as an indirect measure of the formation of protein complexes. Building upon results of the protein complex between *CrDPAS* and the cyclase enzymes *CrTS* and *CrCS* using split-luciferase (Figure 61) and BiFC^[2] assays, we further investigated these interactions using DSF. To this end, we tested the T°_m using various stoichiometric ratios of partner proteins (Figure 65), as well as measuring the effect of cofactors and/or the substrate precondylocarpine acetate, and a range of buffer conditions (Appendix XIII, Table 27 and Figure 131). Despite these extensive efforts, the measured T°_m remained largely

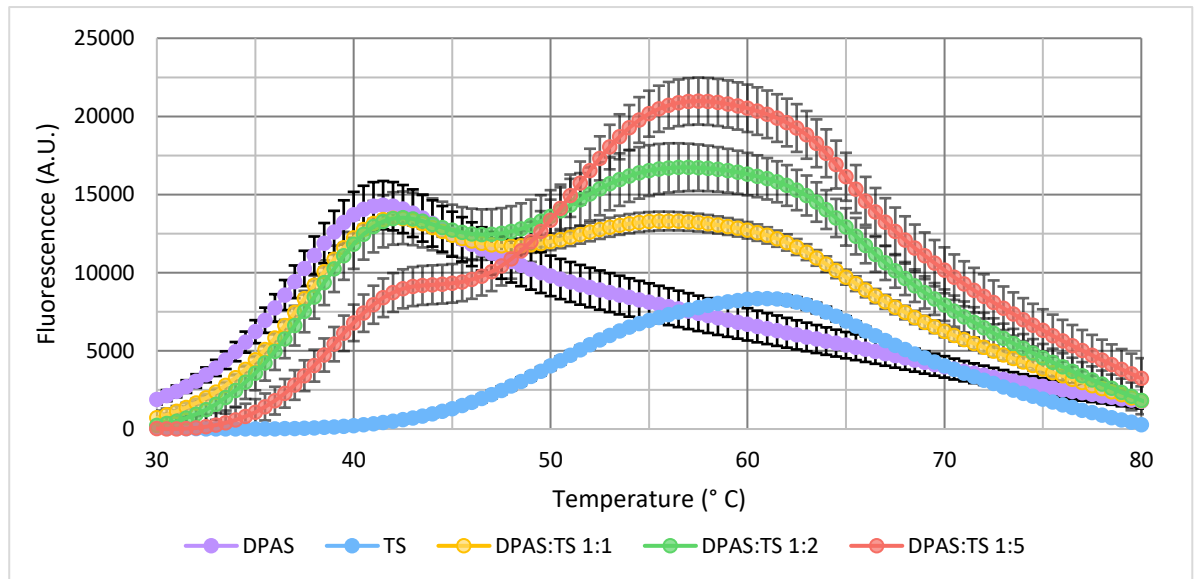


Figure 65. T^m of *CrDPAS* and *CrTS* individually and at 1:1, 1:2 and 1:5 stoichiometric ratios. Fluorescence measured at 490 nm/580 nm Ex/ Em. Bars depict standard deviation ($n = 3$).

unaffected in all tested conditions. These results suggest that either the protein complex between *CrDPAS* and the cyclase enzymes did not form under the conditions tested or was too weak to be detected using DSF. The absence of observable changes in T^m suggests that alternative factors or conditions such as the presence of chaperone-like proteins or plant-specific post-translational modifications may be critical for stabilising this protein complex. Therefore, while these assays did not provide evidence of a protein-protein complex between *CrDPAS* and the cyclase enzymes, these findings offer valuable insights into the dynamic nature of these interactions.

4.2.6. *In vivo* Proximity Tagging of *CrDPAS*

Proximity tagging is a technique used to identify proteins that are in the physical vicinity of a bait protein (Figure 66A) [26, 33]. It involves the transient or stable expression of the bait protein fused to a promiscuous biotin ligase in the plant of interest, followed by the application of biotin. Proteins physically close to the biotin ligase are biotinylated, facilitating their subsequent purification by streptavidin affinity chromatography and identification using proteomics. Thus, these tagged proteins are inferred to interact with the bait protein *in vivo*. Proximity tagging has proven particularly useful for identifying weak and/or temporal protein complexes that are often not detected using alternative co-purification techniques such as AP-MS.

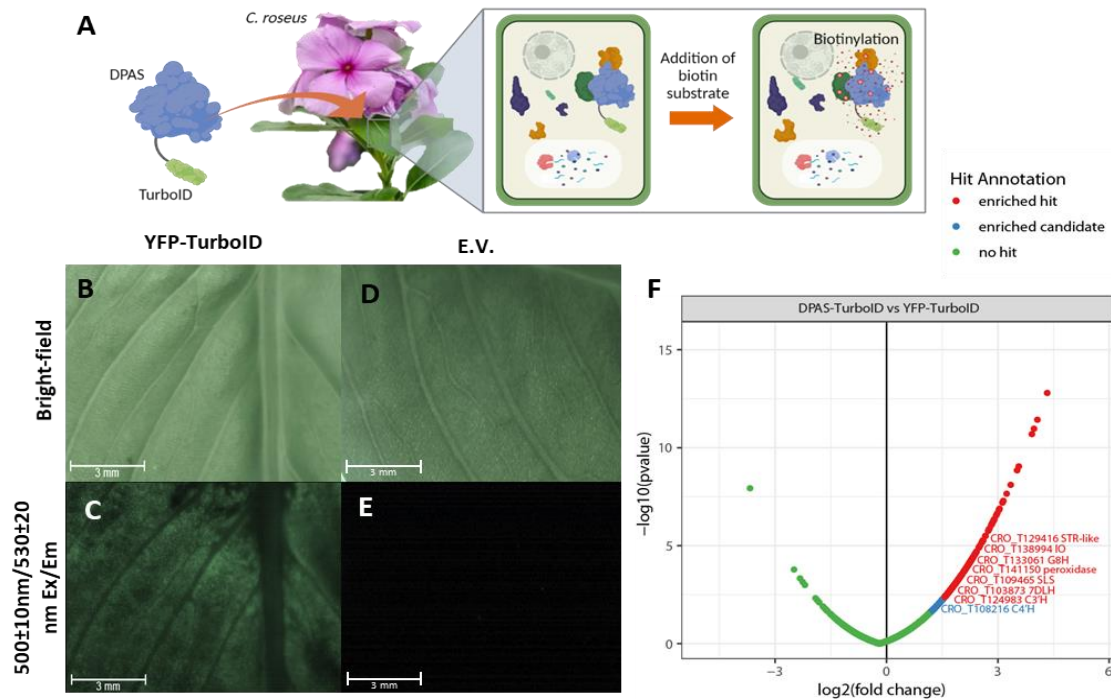


Figure 66. *In vivo* proximity labelling of *CrDPAS* in *C. roseus*. **A.** Schematic of *in vivo* proximity labelling by transient expression of *CrDPAS*-TurboID fusion protein in *C. roseus* followed by subsequent addition of biotin. Figure made using BioRender. Bright field (**B.** and **D.**) and 500±10 nm/530±20 nm Ex/Em filtered (**C.** and **E.**) images of *C. roseus* leaves transiently expressing YFP-TurboID or empty vector (E.V.) constructs respectively. **F.** Volcano plot of proteins enriched in DPAS-TurboID pull-down compared to YFP-TurboID control.

To utilise this technique, we developed an agrobacterium-mediated transformation protocol for *C. roseus* leaves to transiently overexpress the bait protein *CrDPAS* fused with the biotin ligase TurboID. As a control, we also overexpressed a YFP-TurboID fusion construct to account for any non-specific protein binding (Figure 66B and C). Following biotin feeding and subsequent streptavidin affinity purification, the results of proteomic analysis revealed the *CrDPAS*-TurboID specific enrichment of enzymes involved in MIA biosynthesis such as *CrSLS* (Figure 66F; Appendix XIV). However, an interaction between *CrDPAS* and *CrSLS* was not observed when these proteins were tested using a split-luciferase assay (Figure 61). Additionally, the cyclase enzymes *CrTS* or *CrCS* were not enriched in the *CrDPAS*-TurboID sample, contrasting previous findings from BiFC and split-luciferase assays (Figure 61) [2]. These results demonstrate the development and implementation of proximity tagging in *C. roseus* to identify proteins in close physical

proximity to a bait protein *in vivo*. However, complementary *in vitro* and *in vivo* biophysical methods are required to observe and validate the formation of protein-protein complexes between enzymes involved in MIA biosynthesis.

4.2.7. Interactions between MIA and Phenylpropanoid Biosynthetic Enzymes

In addition to the enrichment of MIA biosynthetic enzymes in *CrDPAS* proximity tagging studies (as detailed in 4.2.6), the technique also identified several enzymes involved in phenylpropanoid biosynthesis, namely ferulate-5-hydroxylase (F5H), *p*-coumaroyl ester 3-hydroxylase (C3'H) and cinnamate-4-hydroxylase (C4H; Appendix XIV). The phenylpropanoid pathway is highly conserved across land plants and produces various phenolics including lignin and flavonoids (Figure 67A) [34]. Notably, the final enzymatic step of lignin biosynthesis is catalysed by CAD, a member of the same enzyme family as *CrDPAS* as detailed in Chapter 3 of this thesis. Protein-protein interactions have been widely reported within the phenylpropanoid pathway, including interactions involving CAD [35–45]. Given the conserved interactions between ADHs and α/β -hydrolases (Figure 63), along with the results of *CrDPAS* proximity tagging, we speculated whether MIA biosynthetic enzymes may interact with phenylpropanoid enzymes. To investigate this hypothesis, we tested for interactions between *C. roseus* enzymes that are involved in phenylpropanoid or MIA biosynthesis using a split-luciferase assay (Figure 67B; Appendix XV). Our results revealed many inter-pathway interactions between MIA and phenylpropanoid biosynthesis, suggesting metabolic cross-talk between primary and specialised metabolic pathways.

Protein-protein interactions between consecutive enzymes in a biosynthetic pathway can improve metabolic flux through co-localisation and/or reducing the diffusion of reaction intermediates. For example, the formation of a protein complex between the phenylpropanoid biosynthetic enzymes in *Populus trichocarpa*, cinnamoyl-CoA reductase (*PtCCR*) and *PtCAD*, increased the metabolic flux towards lignin formation [39]. To explore whether orthologous interactions occur in *C. roseus*, we conducted pairwise testing of *CrCCR* using a split-luciferase assay. Surprisingly, we did not detect an interaction between *CrCCR* and either *CrCAD* (Figure 68A-B) or *CrDPAS* (Figure 68C-D), however we did observe interactions between *CrCCR* and the MIA biosynthetic enzyme *CrGS* (Figure 68E-F).

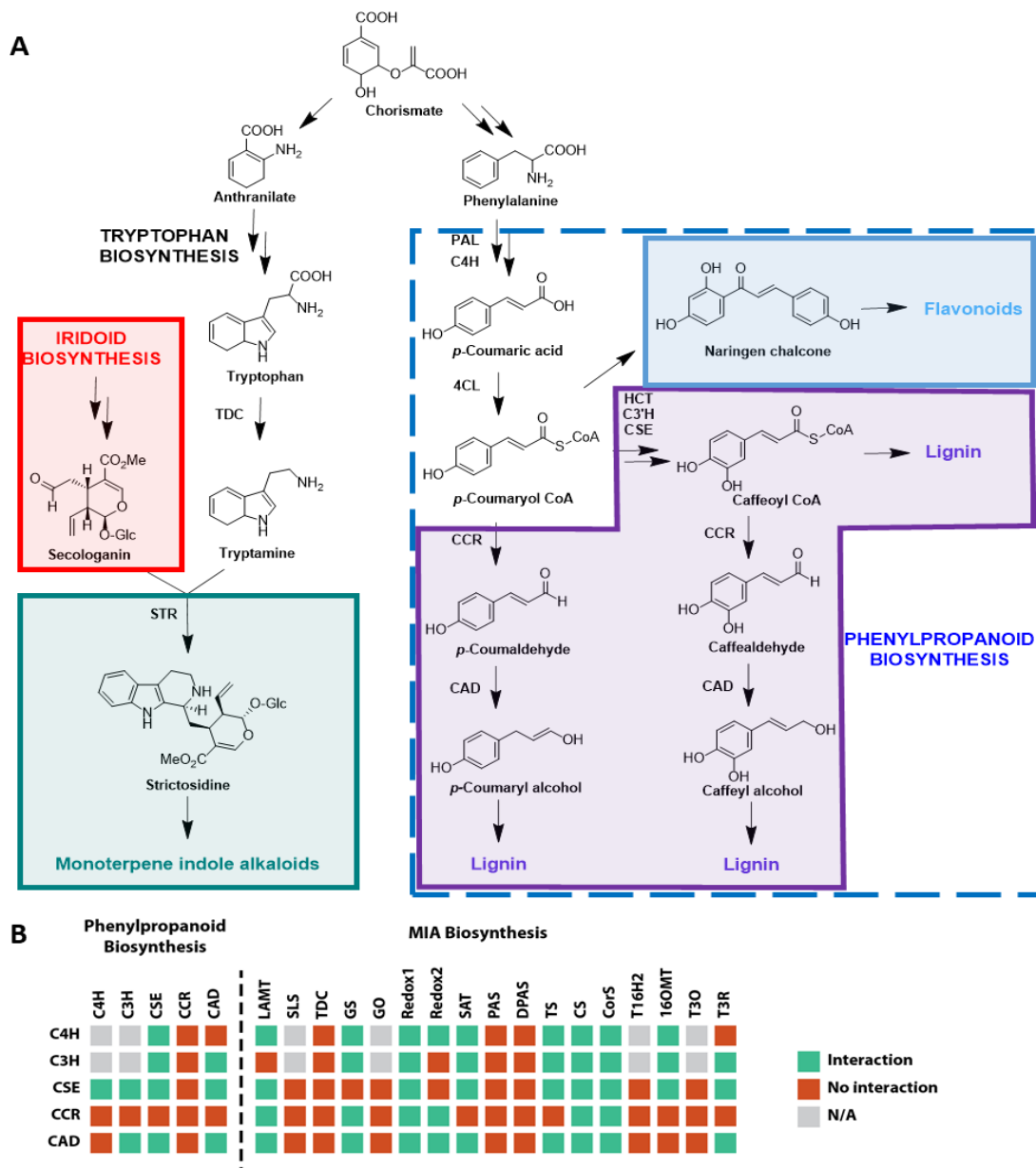


Figure 67. Inter-pathway protein-protein interactions between *C. roseus* MIA and phenylpropanoid biosynthetic enzymes. **A.** Pathway of MIA and phenylpropanoid biosynthesis from the central intermediate chorismate. **B.** Pairwise interactions between phenylpropanoid and MIA biosynthetic enzymes by split-luciferase in heterologous host *N. benthamiana*. N/A refers to protein pairs not tested for interactions due to inaccessibility of luciferase fragment due to membrane localisation.

To investigate the metabolic role of these inter-pathway interactions, we first determined that only *CrCAD* displayed activity against the product of *CrCCR*, cinnamaldehyde, *in vitro* (Figure 68G). This suggests that neither *CrGS* nor *CrDPAS* catalyse the final enzymatic step of lignin biosynthesis. We therefore hypothesised whether the formation of *CrGS-CrCCR*

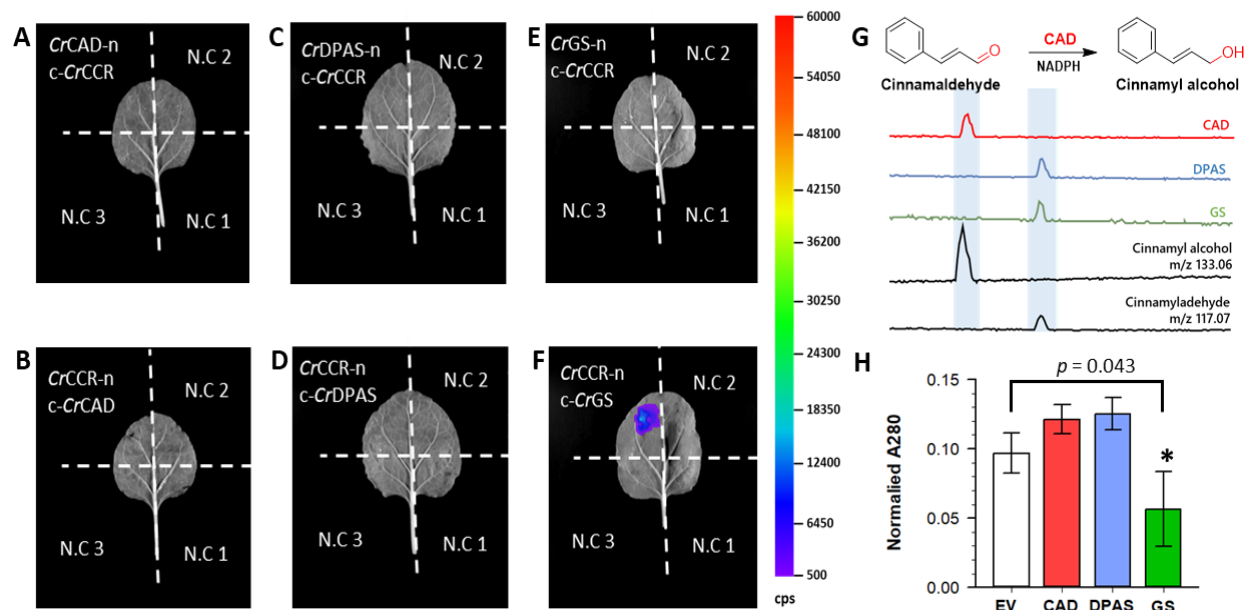


Figure 68. Metabolic crosstalk between phenylpropanoid and MIA biosynthesis in *C. roseus*.

Representative images of split luciferase interaction between *CrCCR* and either *CrCAD* (A-B), *CrDPAS* (C-D), and *CrGS* (E-F). –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents the N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein and cLuc fragment, N.C. 2 represents cLuc-tagged protein and nLuc fragment, N.C. 3 represents nLuc and cLuc fragment negative controls. False colouring represents measured luminescence counts per second (cps). **G.** LC-MS TIC of *in vitro* reactions of *CrCAD*, *CrDPAS* and *CrGS* reacted with substrate cinnamaldehyde and cofactor NADPH. **H.** Extracted lignin content of *C. roseus* leaves transiently overexpressing either empty vector (EV), CAD, DPAS or GS. A²⁸⁰ nm values normalised by sample fresh dry weight. *n* = 6 biological replicates, bars represent standard error, *p* value of a paired t-test.

protein complexes may disrupt lignin biosynthesis in *C. roseus*. To investigate this, we transiently overexpressed either *CrCAD*, *CrDPAS* or *CrGS* in *C. roseus* leaves and measured the extracted lignin content (Figure 68H). We observed that overexpression of *CrGS* significantly reduced the extracted lignin content, while plants overexpressing *CrCAD* or *CrDPAS* exhibited comparable results to the empty vector control. These results suggest that the formation of *CrGS-CrCCR* complexes hinders the metabolic flux to lignin formation through the formation of inter-pathway enzyme-enzyme interactions. Our results shed light on the physical coordination of primary and specialised metabolic biosynthetic enzymes in *C. roseus* and reveals the role of interactions between these proteins in modulating metabolic output.

4.2.8. Probing Protein-Protein Interactions in Pseudo-Scaffold MIA Biosynthesis

The elucidation of biosynthetic pathways in several MIA-producing species across Gentianales has identified some orthologous enzymes to those described in *C. roseus*. For instance, the pseudo-scaffold of tabersonine (Ψ -tabersonine) is generated in the closely related species *T. iboga* by recycling the upstream biosynthetic enzymes. Namely, *CrPAS* or the closely related *TiPAS1-3*, *TiDPAS1* or *TiDPAS2*, and *TiCorS* (Figure 69A). Interestingly, the combination of *CrPAS*, *TiDPAS1* and *TiCorS* formed less reaction side-products in *in vitro* reactions to produce Ψ -tabersonine [23], thus provoking speculation of the role of protein-protein interactions. Therefore, to investigate this hypothesis, we tested the pairwise interactions of these enzymes using a split luciferase assay. Our findings detected that only *CrPAS* interacted with *TiDPAS1* and *TiCorS* (Figure 69B; Appendix XVI), correlating with the results of *in vitro* reactions. These results suggest that protein-protein interactions play a significant role in Ψ -tabersonine biosynthesis in *T. iboga*. Furthermore, these findings reveal that orthologous protein complexes to those identified in *C. roseus* may be conserved across other MIA-producing species, raising questions about their evolution.

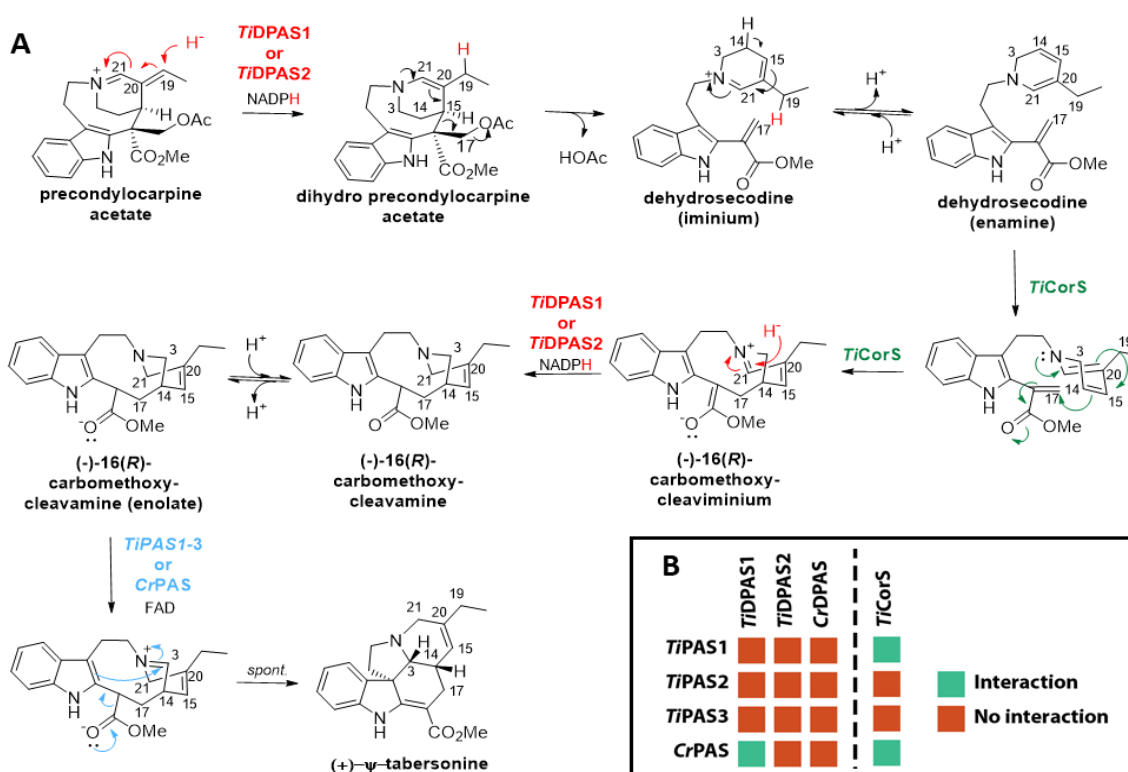


Figure 69. Biosynthesis of Ψ -tabersonine in *T. iboga*. **A.** Biosynthesis of Ψ -tabersonine. Figure adapted from Kamileen *et al.*, [23]. **B.** Pairwise interactions between Ψ -tabersonine biosynthetic enzymes tested by split-luciferase in heterologous host *N. benthamiana*.

4.3. Discussion

4.3.1. Establishment of a *C. roseus* MIA Biosynthetic Protein-Protein Interaction Network

The network of pairwise protein-protein interactions among 17 enzymes involved in vinblastine biosynthesis in *C. roseus*, all of which have enriched expression in epidermal cells^[4] and are accessible to the cytoplasm (Table 23), was mapped using a split-luciferase assay (Figure 61). Interestingly, interactions were observed between various MIA biosynthetic enzymes, including those acting in distal parts of the pathway. Notably, CrDPAS was observed to interact with the downstream cyclase enzymes CrTS and CrCS (Figure 62), thus corroborating earlier BiFC results^[2]. However, this interaction was not detected using other biophysical methods including AP-MS (Table 26) or DSF (Figure 65). These findings suggest that the interactions formed between these enzymes are either too weak and/or transient to be detected using these methods, or they may necessitate other conditions or factors such as post-translational modifications. Interestingly, the interactions between the CADs and α/β -hydrolases protein families (to which DPAS and the cyclase enzymes respectively belong) were revealed to be largely conserved despite their species origins and metabolic function (Figure 63). These results shed light on the evolution and conservation of a protein-protein interaction between two enzyme families.

4.3.2. Engineering MIA Enzyme-Enzyme Interactions

Protein-protein interactions between DPAS and the cyclase enzymes has been speculated to influence the metabolic flux of the intermediary molecule, dehydrosecodine. Building upon the observation that CrDPAS interacted with the cyclase enzymes CrCS, CrTS, CrCorS and TiTabS, but not with TiCorS (Figure 63C), we explored the basis of this interaction through comparative structural analysis. This led to the design of non-interacting cyclase TiCorS (Figure 64A), and interacting cyclase CrCorS (Figure 64B) mutants, culminating in the identification of a single surface residue responsible for the interaction with CrDPAS (Figure 64C-H). Future experiments will endeavour to determine the metabolic role of these mutant cyclase enzymes on the *in vivo* and *in vitro* production of MIAs.

4.3.3. Inter-pathway Interactions and Metabolic Crosstalk between MIA and Phenylpropanoid Biosynthesis in *C. roseus*

In vivo proximity tagging revealed that *CrDPAS* was physically close to both MIA and phenylpropanoid biosynthetic enzymes (Figure 66). Interactions between phenylpropanoid biosynthetic enzymes have been widely reported, including the lignin biosynthetic enzymes *PtCCR* and *PtCAD* (Figure 67A) [35–39, 46]. Given that many specialised metabolic pathways are proposed to have evolved from primary metabolism [47, 48], we speculated whether the interactions of *CrDPAS* had been conserved from *CrCAD* despite their functional divergence. To explore this, we tested the pairwise interactions between *C. roseus* MIA and phenylpropanoid biosynthetic enzymes (Figure 67B). Our findings revealed many inter-pathway interactions between proteins acting with these primary and specialised metabolic pathways, including the interaction between *CrCCR* and *CrGS* (Figure 67B). Furthermore, the reduction of lignin content in *C. roseus* leaves transiently overexpressing *CrGS* suggests a functional role of these inter-pathway interactions in altering metabolic output within a cell (Figure 68H). These results provide insight into how the formation of protein-protein complexes can physically organise primary and specialised biosynthetic enzymes within a cell. Furthermore, these interactions may provide a mechanism enabling plants to rapidly alter their metabolism to cope with various abiotic and biotic stresses.

4.3.4. Potential Role of Protein-Protein Interactions in Wider MIA Biosynthesis

Whilst the majority of work presented in this chapter focused on enzymes derived from *C. roseus*, we also present preliminary findings of enzyme-enzyme interactions in the closely related species *T. iboga*. Ψ -tabersonine is accessed by recycling the activities of upstream biosynthetic enzymes *CrPAS*, *TiDPAS1* and *TiCorS*, highlighting the role of metabolic plasticity in generating chemical diversity (Figure 69A) [23]. We observed that these enzymes interacted with one another when tested for pairwise interactions using a split-luciferase assay (Figure 69B). These results suggest the functional role of protein-protein interactions in Ψ -tabersonine biosynthesis in *T. iboga*. Furthermore, these findings support that some of the protein-protein interactions identified in *C. roseus* may be conserved in the MIA biosynthetic pathways of other closely related species.

4.4. Conclusions

The mapping of the protein-protein interactions of a selection of MIA biosynthetic enzymes from *C. roseus* provides valuable insights into the organisation of complex pathways within the cell. The pairwise testing of proteins from the ADH and the α/β -hydrolases families, suggests this interaction is largely conserved, though some exhibited more specificity, such as CrDPAS. However, the lack of detection using *in vitro* biophysical methods including AP-MS and DSF highlights the weak and/or transient nature of these interactions. By employing structural comparison and site-directed mutagenesis, a crucial surface residue is identified to be responsible for introducing or abolishing interaction between the cyclase enzymes and CrDPAS. These results suggest the potential for metabolic tuning at this bifurcation point.

Moreover, the discovery of inter-pathway interactions between enzymes from MIA and phenylpropanoid biosynthesis suggests the role of protein complexes in organising multiple metabolic pathways within the cell. Notably, the formation of CrGS-CrCCR protein complexes was associated with reduced lignin content in *C. roseus* leaves, indicating these interactions play a metabolic role within the plant. These results therefore not only hold promise for bioengineering applications in MIA biosynthesis, but are offer important insights into the broader organisation and evolution of metabolic pathways within a cell.

4.5. Methods and Materials

4.5.1. Chemicals and molecular biology reagents

All solvents used for extractions were HPLC grade and solvents used for UPLC/MS were MS grade. All solvents were purchased from Sigma Aldrich. Carbenicillin, kanamycin sulfate, gentamicin sulfate, rifampicin and isopropyl β -D-1-thiogalactopyranoside (IPTG) salts, cinnamaldehydes and cinnamyl alcohol were purchased from Sigma. D-luciferin was purchased from Promega. Synthetic genes were purchased from GeneWiz. All gene amplifications and mutations were performed using Platinum II Superfi DNA Polymerase (Thermo Fisher). Constructs were transformed into vectors using the In-Fusion kit (ClonTech Takara) and colony PCR was performed using Phire II mastermix (Thermo Fisher) according to the manufacturer's instructions. PCR product purification was performed using the Zymoclean Gel DNA Recovery kit (Zymo). Plasmid purification was performed using the Wizard Miniprep kit (Promega).

4.5.2. Cloning and mutagenesis

CrCorS, *CrHID5*, *CrCSE*, *CrCAD*, and *Cr2141* were previously identified from the *C. roseus* transcriptome and amplified from cDNA (Table 8). *CrC4H*, *CrC3H*, *CrCAD*, and *CrCCR* were identified from the *C. roseus* transcriptome based on sequence similarity based on previously characterised orthologs from *A. thaliana* and amplified from cDNA (Table 8). TurboID and linker sequence was based on work from Arora *et al.*,^[26]. TurboID, *PscXE1*, *GmHIDH*, *CrCorS* M1 and *TiCorS* M1 synthetic genes were ordered from GeneWiz (Table 8). The remaining MIA biosynthetic genes were amplified from the cDNA of their respective organism based on previously published sequences (Table 9).

Full-length genes were amplified using Platinum II Superfi DNA Polymerase (Thermo Fisher) and the corresponding primers to have the correct overhangs for pCambia, pCambia nLuc, pCambia cLuc, pHREAC, and pOPINF vectors (Table 10). cLuc Fwd and nLuc Rev primers were used to amplify constructs with the correct overhangs for the pCambia vector. *CrCorS*, *TiCorS*, *CrCS* and *CrTS* mutants were generated by overlap extension PCR as previously reported^[32] using corresponding primers in Table 10. For protein purification, *CrCAD* and *CrDPAS* were cloned into pOPINF and *CrGS* was cloned into

pOPINK vectors respectively. PCR products were purified from 1% agarose gel. pOPINF and pOPINK constructs were ligated using *HindIII* and *KpnI* restriction sites, pCambia nLuc using *KpnI* and *Sall* sites, pCambia cLuc using *KpnI* and *PstI* sites, pCambia using *KpnI* and *PstI* sites, and pHREAC using *BsaI* sites. Constructs were ligated into pCambia and pOPINF vectors using the In-Fusion kit (Clontech Takara). Constructs were ligated into pHREAC using the BsaI-HFv2 Golden Gate enzyme mix (New England BioLabs) as per the manufacturer's instructions. pOPINF was a gift from Ray Owens (Addgene plasmid # 26042^[49]) and pCambia nLuc and cLuc were both gifts from Richard Dixon.

In-Fusion and Golden Gate assembly products were transformed into *Escherichia coli* TOP10 cells (ThermoFisher) by heat shock at 42 °C for 30 seconds before incubating on ice for 2 minutes. Cells were then plated on LB agar containing the respective antibiotics (100 µg/mL carbenicillin for pOPINF transformants, 100 µg/mL kanamycin for pCambia and pHREAC transformants) and grown overnight at 37 °C. Colonies were screened by colony PCR using vector-specific sequencing primers (Table 10) and positive colonies were grown overnight in 10 mL liquid LB supplemented with the appropriate antibiotic at 37 °C shaking at 200 r.p.m. Plasmid DNA was isolated using the Wizard Miniprep kit (Promega) and sequenced.

Table 8. Full-length nucleotide sequences of unpublished and synthetic genes used in this chapter.

<i>C. roseus</i> CorS	ATGGCTTCCCAAACCTCCAACCTCAGATGAGACTCTTTTCGATCTTTCTCCATAC ATCAGAATCTTCAAAGATGGAAGAGTAGAAAGACTCCATAATACTCCTTATG TTCCCCATCACTTAATGATCCAGAAACCGGCTCTCTTGGAAAGACGTCCCA ATTTTCATCAAAGTTTCGGCTAGAATTTACCTTCCAAAAATCAGTGACCAGCA GGAAAATGAAGAAAACTCCAATTTTTGTTTATTTCCATGGGGCTGGCTTCT GTCTAGAATCTGCATTAGATCATTTCACACTTTTATCAAACACTTTGTAT CCGAAGCCAAAGCCATTGGGGTTTCGGTTGAATACAGACTCGCCCCGGAAC ACCCTTTACCCGCAGCTTATGAAGATTGCTGGGAAGCCCTCAATGGGTTCGC TTCTCACGTTTCGTCTCGACAATTCAAGCCTCAAGAGATCTATGGACAAGGAT CCATGGATAATCAACTATGGCGATTTTCGATAGACTCTATTTGGGGGGTGATA GTCCCGGTGGCAATATTGTTCAACGTAATTCTCAGAGCTGGAAAAGAGAA ATTGAATGGGGGAGTGAAAATTTGGGGGCAATTCAGTATTACCCATATTTT CTGATCCGGACGAGCTCGAAACAGAGTGATTATATGGAGAATGACTACAGG TGTTACTGGAAATTGGCTTATCCAAATGCTCCTGGTGGAACTGATAACCCAA TGATAAACCCACAGTTGAGAATGCTCCTGATTTGGCCGGATATGGTTGCTC
--------------------------	--

	CAGGCTGCTGATTTCAATGGTTGCTGATGAGACTAGAGATATAACTCTGCTT TTTATTGAGGCATTGAAGAAGAGCGGATGGAAAGGGCAATTGGATGTGGCT GATTTTGAAGCAGAGTTTTTTGACCTTTTCCAAACACAAACAGAGGTGGGCA AGAACATGATTAGACGCTTAACGTCTTTCATCAA
<i>C. roseus</i> HID5	ATGGCCTCCTCAGATGAGATTGCTATTGATATTTCTCCAGACATCATCCTCTA TAAATCCGGTAAGGTGGTAAGAGATTTTGTCCGACCATATGTTCCGCCATCA CTTGAAGATCCAACCACCGGTGTCTCTACTAAAGACGTCCCAATCTCAGCGG AAGTTTCTGCTAGAATCTACCTTCAAAGCTTGACACAGATGCACAAAAGTTC CCCATCTTGGTCTACTTCCACGGTGGAGGCTTCTGTTTGGTATCCGCCTTCGA TTCTTTATACAGCACTTACTTAAAATCCTTAGCCTCAGAAGCCAAAACAATTA TAATTTCACTCGAATTCCGCCTCGCTCCTGAGAACCCTTACCAGGTAGGTTAC GAAGATTGTTGGACTGCCCTCAATGGGTAGCTTACATGCCGTTGATAATT CCCTGTCTGTATTGATAGAGAACCCTGGCTAATCCACCACGGGAATCTTGA CAAAGTTTACATCGGAGGTGACAGTACCAGGGGTAACATGGTACATAACAT ATTAATGAAATCTGGTCGGGAAAAATTGAACGGTGACTTAAAAATCTCGGG AGGGATACTTTCTTACCCTTATTTCTTAATTAGTTCATGGGCTAAGAAAAGTG ATGAGGAATTATCAGATATGGTGAAAATGTATAAGAAATATTGGTTATTGTC TTGTCCTAGTGCTCCTGGTGGATTTGATAATCCGATGGTAAATCCGGTAGTT GAAGATGCTCCGAGCTTGGCCGGAATTGGGTGTTGGAGGTTACTTGTGATT ATGGCTATAGACGATTTAAGAGAAGCCCATCTTAGCTATGTGGAGGGATTG AAGAAAAGTGGGTGGAAAGGTGAATTGGAATTGGCTGATTTTGAAGGATAT GATCATTTCTTTGAGATCTTTAACCACCACTCAAAGGGCCAAGAATATTAT TCATCGAATAGCCTCTTTTATAAAGTAA
<i>C. roseus</i> CSE	ATGCCTTCAGAAGCAGCGCCCGGCTCAGGCAACAACGCCCGCGAATTTCT GGGGAGATATGCCGGAAGAAGAATACTATTCATCACAAGGAGTTCGAAACA AAAAATCCTATTTGAAACACCAGATGGAAAATCTTCACTCAATCATTCTT CCGTTAGATCCACAACAACCGATCAAAGGAACGGTATATATGAGCCATGGAT ACGGGTCCGATACAGGTTGGCTATTTCAAAGATCTGTATAAATTATGCGAA TTGGGGATACGCAGTGTTCGCGCGGATCTACTTGGGCATGGCCGATCAGA AGGGATCCGATGTTATCTTGGAGATATGAATAAAATTGCTGCTGCTTCTTGT ATTTTTCAAGAGTGTGAGGAATAGCGATGAATATAAGGAATTGCCGGCGTT TTTGTGGGAGAATCAATGGGTGGACTTGCTACTTTGCTCATGATTTTCAAT CGGAGCCAAATACTTGGACTGGATTGATTTTCTGCCCCTCTTTTGTCTTTC CTGAACCCATGAAGCCCTCAAAGGCAAGGCTATTCATGTACGGATTATTATT CGGGCTGGCAGATACATGGGCAGCAATGCCAGACAACAAAATGGTGGGAA AAGCCATAAAAGACCCTGAGAACTCAAGATAATTGCCAGCAATCCAAGAA GATACTGGTCTCCAAGGGTAGGAACCATGAGGGAACTTCTAAGGATGA CAGAGTATGTCCAGAACAATTTGACAAAGTTACGACACCGTTTTTAAACGGT CCACGGAACATCGGACGGCGTCACTTGTCCGACAGGATCAAAAATGTTGTAT GAAAAAGCAAGCAGTTCAGACAAGACATTGAAATTGTATGATGGAATGTAT CATTCTTTGATTCAAGGAGAGCCTGATGAAAATGCTAATCTTGTTTTGGCTGA TATGAGGGCTTGGATTGATGAAAGAGCTCAAAAATACGGTCCCAGAAAGTA A
<i>C. roseus</i> CAD	ATGGGGAGCTTGAAGAAGCAGAGAGAAAGACAATAATGGGATGGGCAGC AACTGATCCTTCAGGACAACCTTCCCCTACTCCTACTCCCTCAGAAACACAG

	<p>GTCCTGAAGATGTTTATATCAGGGTTATATGCTGCGGAGTTTGCCATACCGA TATTCATCAGACCAAGAATCACCTCGGCATGTCCAATTACCCCATGGTTCCTG GGCACGAAGTGGTAGGTGAAGTGGTGGAGGTGGGATCCAATGTGAGCAAG TTCAGAGTTGGTGAGCGTGTGGAGTAGGCATAATCGTTGGATCCTGCCAG AACTGCAGATCATGCGAAGCAGAGATAGAGCAATACTGCAACAAAAAGATT TGGACATACAATGATGTGTATACAGACGGCAATCCCACTCAAGGTGGATTTG CTAGTGCCATGGTCGTTGACCAGAAGTTTGTAGTGAAAATACCAGAGGGTA TGGATCCAGAGCAAGTAGCACCCCTTACTCTGTGCTGGGGTGACAGTGATA GTCCATTGAGCCATTTTGGGCTAAAGCAAAGTGGACTAAGAGGAGGCATAT TAGGACTTGGTGGTGTGGGCATATGGGAGTGAAAATAGCCAAAGCAATGG GGCATCATGTAACGGTCATAAGTTCTTCAGATAAGAAGAGAGAGGAAGCTT TGGACCACCTGGGCGCTGACGCATACTTGGTCAGCTCTGATGAGGGAAAGA TGCAGGAGGCTGCAGATTCACCTTGATTACATTATTGACACAGTTCCTGTTTT CATCCTCTGGAGCCATTTTATCATTGTTGAAAGTTGATGGAAAGTTGATTTT GATGGGAGTTATTAACCAGCCTTTGCAATTTACTCCAATGGTTATGCTAG GAAGGAAGTCAATAACAGGAAGCTTTATAGGTAGCATAAAAGAGACAGAA GAAGTACTTGAGTTCTGCAAGGAAAATAACCTAACTCCCAAATTGAAGTAG TGAAAATGGATTATATCAACAAGGCTTTTGAAGACTTGAAAAGAATGATGT CAGATATAGGTTTGTGTGGACGTTGCCGGCAGCAACCTTCTTGTGACCAC TAA</p>
<i>C. roseus</i> 2141	<p>ATGGCCGAAAATCACCAGAAGAGGAGCACCCAGTCAAGACCTATGGATTG GCTGCTCATGATTCATCTGGGGTTTTATCTCCGTTCAAATTCTCCAGGAGGGC AACTCTTGAGGATGATGTGAGGTTCAAGGTGCTATATTGTGGGATTTGTCAT ACTGACCTTCATTCGCTAAGAATGAGTGGGGTATTCGACCTATCCTCTTGT ACCAGGACATGAAATCGTAGGGGAAGTTACAGAGGTCCGGCGGCAAAGTTA CAAAGGTCAAGGTTGGAGATAAAGTTGGTGTGGCTGCTTGGTTGGTTCAT GCCGCACTTGATAATTGTCGTGCAGATCTTGAGAACTATTGTCCCAAAT GGTGCTAACCTATGCAAGTCAAACGTTGATGGAACGATTACCTATGGAGGC TATTCGAATGAGATGGTATGCAATGAACACTTTATTGTTGTTTTCCAGAGAA CCTACCACTTGATGGTGGGGCACCATTGCTTTGTGCCGGTATTACTGTGTAC AGTCCAATGAAATACTATGGCTTTGCCAAACCCGGGAGCCACATAGCTGTTA ATGGTCTTGGTGGACTTGGCCATGTGGCTGTTAAGTTTGCAAAGGCCATGG GAGCAAAAGTGACAGTTATAAGTACATCTGAGGGCAAGAAAGACGATGCC TCAATCGTTTGGGTGCAGATGCATTTTTGTTGAGCAGTAATCCAGAAGCACT GCAGGCTGCAACAGGCACATTTGATGGCATACTTAATACTATTTCTGCTAAG CACGCTATTATCCCATTGCTTGGTCTACTAAAGTCTCATGGCAAGCTTGTCT TCTTGGGGCACCCCGGAACCACTTGATCTTCACTCTGCTCCTTTGCTTATGG GGAGGAAGATGGTTGCTGGAAGTAGCATTGGAGGATTGAAGGAGACCCAA GAGATGCTTGATTTTGCCGAAAGCATAACATTAAGTGCAGATATAGAAGTCA TTTCCGCGGACAATATCAACACAGCTTTGGAGCGTCTGGCCAAGGGTGATGT TAGATATCGCTTTGTCCTTGACGTTGCAAAGACCTTGAAAGCTCCTTAA</p>
<i>C. roseus</i> CCR	<p>ATGCCGTCAGATTCCGGCAGAGTCGTCTGTGTTACCGGTGCCGCCGGTTACA TCGCTTCATGGATTGTCAAACCTTCTTCTGAAAAGGCTACACCGTCAGAGG AACTGTTAGAAATCCAGATGATCCAAAGAACAATCATTTAAGGGAATTAGAA GGAGCAAAGGAAAGATTAACACTGTGTAAAGCTGATCTACTTGATTATCAGA GTTTAAGACAAGCAATCGACGGCTGTGATGGAGTTTTCCACACTGCTTACC</p>

	<p>AGTTACCGATGACCCAGAACAAATGGTGGAGCCAGCAGTGATTGGGACAAA GAATGTAATCAATGCCGCTGCCGAAGCTAAGGTCCGCCGTGTGGTTTTACC TCGTCAATTGGTGCTGTTTCATATGGATCCAAACAGGCATCCTGATAAAGTCG TTGATGAGACTTGCTGGAGTGATCTTGATTTCTGCAAGAACTAAGAATTG GTATTGCTACGGGAAGGCAGTGGCGGAACAAACGGCATGGGAAGAAGCCA AAGCTAAAGGAGTGGACCTTGTTGATCACCCCTGTTTTGGTATTAGGGCC ATTGCTACAAAACACAGTGAATGCNAGNGTTCTTACATACTAAAGTATTTG ACTGGCTCGGCAAAGACATATGCCAATTCAGTACAAGCATATGTGCATGTTA AAGATGTTGCATTGGCACATATACTTCTATTTGAGACTCCTTCTGCTTCCGGA AGATACCTTTGTGCCGAAAGTGTGCTTCATCGAGGCGAAGTAGTTGAAATTC TGGCTAAATTTTTCCGGAGTATCCTATCCCCACAAAGTGTTGAGATGAGAC GAAGCCAAGAGCAAACCATACAAATTCTCAAACCAAAGCTGAAAGATTT GGGACTTGAATTTACACCAGTGAAGCAATGCCTTTATGAAACTGTCAAGAGT TTGCAGGAGAAAGGTCACCTTCCACTCCCTACTCAGGAGAACGATGAACCCC TCACAATAATCCGCTCTTAA</p>
<p><i>C. roseus</i> C3H</p>	<p>ATGAACATTTCTTCCCCACTACCGCCAACTCTCCACCACTTCTCCCTCCCTATG GCCCTTCTTTCTCTATTGCTCCTCACGTTTATTTTTCTCTTTCTAGCTTACTACC TTTACCAAAAATTCCGATTCAAACCTTCCACCCGGTCCCCGCCGTTACCCATC GTCGGAAACCTCTACGACGTTAAGCCGGTGAGGTTCCGATGTTTCTCCGAAT GGTCCGAACATTATGGACCGATTATATCGGTTTGGTTTGGCTCCACGCTAAA CGTTGTCGTTTCTAGCTCCGAATTAGCTAAGGAGGTTTTGAAAGAGAATGAT CAGCAATTGGCGGATCGGCACCGGAGTCGATCTGCCGCCAAGTTTAGTAGA GACGGACAGGACTTAATTTGGGCTGACTATGGACCTCACTATGTCAAGGTCA GAAAAGTGTGTACACTTGAATTGTTTTCTCCAAGAGGCTTGAAGCTCTGAG GCCATTAGAGAAGATGAGGTCACGGCCATGGTAGAGTCCATCTATAAAGA TTGCACCAATCCTGGCAATATAGGGAAAAGTCTGCTAGTGAAGAAGTACCTT GGAGCAGTGGCATTAAACAATATAACAAGACTTGCATTTGGAAAGCGTTTTG TGAAGTCTGAGGGAGTGATTGATGAGCAAGGTAAGGAGTTTAAAGAGATAG TTGCCAATGGATTGAAGCTGGGTGCATCTCTAGCCATGGCTGAGCACATCCC GTGGTTGCGTTGGCTGTTCCCTCTTGATGAAGCTGCATTTGCAAAGCACGGC GCTCGTAGGGACCGCTCACCCGTTCCATCATGGAAGAACACACTCTTGCTC GCCAGAAAAGTGGAGGAGCCAAGCAACACTTTGTTGATGCTTTGCTTACCCT CAAAGATCAATATGATCTTAGTGAAGACACCATCATTGGCCTTCTATGGGAT ATGATTACAGCGGGGATGGACACCACTGCCATTAGTGTTGAATGGGCTATG GCAGAGTTAATAAAGAATCCTAGGGTCCAACAAAAAGCCCAAGAGGAGTTG GACCGGGTAATCGGTTATGACCGGGTTATGACCGAACCAGACTTCTCAAACC TCCCTTACCTACAATGTGTAGCAAAGGAAGCACTAAGGTTGCACCCACCAAC ACCATTGATGCTTCTCACCGAGCCAATGCCAACGTGAAGATAGGCGGCTAC GACATCCCCAAGGGCTCAAACGTGCATGTAAACGTGTGGGCGGTTGCTCGT GATCCGGCCGTGTGGAAGAACCCTACAGAGTTCAGGCCGGAGAGGTTCTTG GAAGAGGATGTTGATATGAAGGGTCATGATTTTAGGCTACTTCCATTTGGTG CTGGTAGAAGAATATGCCAGGGGCACAATTGGGAATCAATCTAGTGGTAT CTATGTTGGGACACCTTTTGACCATTTTAATTGGGCTCCAGCTAATGGATTG AGCCCGGAAGAAATAGACATGGGGGAGAATCCGGGCCTGGTTACTTACATG AGGACACCACTTGAGGCAGTTCCTACACCAAGATTACCTGCAGAGTTATACA AACGTGTGCCTGTGGATATA</p>

<p><i>C. roseus</i> C4H</p>	<p>ATGGATCTTCTCCTCTTAGAGAAGACCCTTTTGGGTCTATTTGCGGCCATCAT TGTGGCCTCTGTAGTTTCAAAGCTACGAGGAAAGAAATTTAAGCTTCTCCA GGTCCTATCCCGGTACCGGTTTTGGAACTGGCTTCAAGTTGGGGATGACT TGAATCACAGAAATCTATCGGATTACGCTAAGAAATTTGGCGAAATTTCTTA CTTAGAATGGGCCAACGTAATCTGGTTGTGGTTTCATCTCCTGAACTGGCTA AAGAAGTTTTGCACACTCAGGGGGTTGAATTTGGCTCCCGTACTAGAAATGT TGTGTTTGATATCTTACAGGAAAAGGACAGGACATGGTTTTTACCGTTTAT GGTGAACATTGGAGGAAAATGAGAAGAATCATGACTGTCCCGTTTTTACTA ATAAAGTAGTTCAACAGTATAGATATGGATGGGAAGAAGAGGCAGCCCGTG TTGTTGAGGATGTGAAGAAAATCCTGAATCTGCAACTAATGGGATTGTATT GAGGAGAAGGTTACAACCTTATGATGTACAATAACATGTACAGGATTATGTTT GATAGAAGGTTTGGAGAGTGAGGATGATCCTTTTTTGTAACTTAAGGCCT TGAATGGTGAAGGAGTAGATTGGCCCAGAGCTTTGAGTACAATTATGGCG ATTTCAATCCAATTTTGGAGCCTTTCTTGGAGAGGTTATTTGAGGATCTGTAAG GAGGTTAAGGAGAGACGATTGCAGCTTTTCAAGGATTACTTCGTCGACGAA AGGAAGAAGTTGGGGAGTACAAAAGCATGGATAACAACAGCTTGAAATG TGCCATTGATCATATCCTAGAAGCTCAGCAAAGGGAGAGATCAACGAGGA TAATGTCCTTACATTGTTGAAAACATCAATGTTGCTGCCATCGAGACAACAC TATGGTCCATTGAGTGGGAATTGCAGAATTGGTGAACCACCCTGAAATCCA GAAGAAGCTACGAGACGAGCTTGATACTGTGCTAGGACCCGGCGTGCAGAT CACTGAACCGGATACTTACAAGTTACCATACTTCAGGCAGTGATCAAGGAG ACACTTCGTCTCAGAATGGCGATTCCCTTTTTGGTGCCTCACATGAACCTACA CGATGCCAAGCTTGGTGGCTATGACATTCCAGCGGAGAGCAAATACTGGT GAATGCCTGGTTTTAGCCAACAATCCGGAGCATTGGAAGAAGCCTGAAGA GTTGAGACCGGAAAGGTTCTTGGAAAGAGGAATCGAAAGTTGAGGCTAATGG CAATGACTTCAGATATCTACCATTTGGTGTGGTAGGAGAAGTTGCCCTGGT ATTATTCTAGCATTGCCAATTCTTGGCATTACTATAGGACGTTTGGTTCAGAA CTTTGAGCTTTTGCCTCCACCAGGACAATCTAAGATTGATACTAGTGAGAAA GGTGGACAATTCAGTTTGCACATTTTGAAGCACTCTACTATTGTAICTAAGCC CAGGACTTTTTAG</p>
<p><i>C. roseus</i> CorS Int M1</p>	<p>ATGGCTTCCCAAACCTCCAACCTCAGATGAGACTCTTTTCGATCTTCTCCATAC ATCAGAATCTTCAAAAACGGAAAGGTAGAAAGACTCCATAATACTCCTTATG TTCCCCATCACTTAATGATCCAGAAACCGGCGTCTCTTGGAAAGACGTCCCA ATTTTCATCAAAGTTTTGGCTAGAAATTTACCTTCAAAAATCAGTGACCAGCA GAAGAATGAAGAAAACTCCAATTTTTGTTATTTCCATGGGGCTGGCTTCT GTCTAGAATCTGCATTGAGATCATTTCACACTTTTATCAAACACTTTGTAT CCGAAGCCAAAGCCATTGGGGTTTTGGTTGAATACAGACTCGCCCCGGAAC ACCCTTACCCGACGCTTATGAAGATTGCTGGGAAGCCCTTCAATGGGTCGC TTCTACGTTTCGTCTCGACAATTCAAGCCTCAAGAGATCTATGGACAAGGAT CCATGGATAATCAACTATGGCGATCTCGATAGACTCTATTTGGGGGGTGATA GTCCCGGTGGCAATATTGTTTACAACGTACTTCTCAGAGCTGGAAAAGAGAA ATTGAATGGGGGAGTGAAAATTTGGGGGCAATTCAGTATTACCCATATTTT CTGATCCGGACGAGCTCGAAACAGAGTGATTATATGGAGAATGACTACAGG TGTTACTGGAAATTGGCTTATCCAATGCTCCTGGTGGAACTGATAACCCAA TGATAAACCCACAGTTGAGAATGCTCCTGATTTGGCCGGATATGGTTGCTC CAGGCTGCTGATTTCAATGGTTGCTGATGAGACTAGAGATATAACTCTGCTT</p>

	TTTCTTGAGGCATTGAAGAAGAGCGGATGGAAAGGGCAATTGGATGTGGCT GATTTTGAAGCAGAGTTTTTTGACCTTTTCCAAACACAAACAGAGGTGGGCA AGAACATGATTAGACGCTTAACGTCTTTCATCAA
<i>T. iboga</i> CorS Int M1	ATGGCTAATTCAACTGCAAACCTCTGATGAGATTGTTTTCGATCTTCATCCATA CATCAGAGTCTTTAAAGATGGCAGAGTAGAAAGACTTCACGACACCCCATAT GTTCCGCCATCACTTGAAGATCCAGCCACCGGTGTATCCTGGAAAGACGTCC CAATTTTCATCCGACGTTTCAGCTAGAGTCTACCTCCCGAAGATCAGCGAAGC GGAAGAAAAAAGCTCCCCATTTTCGTCTATTTCCATGGTGCAGGCTTCTGTC TGGAATCAGCCTTCAAATCATTTTTCCATACTTATGTTAAGCACGTTGTTGCC GAAACCAAAGCTGTCGGAGTTTCGGTTGAGTACAGACTCGCCCCGAGCAC CCTTTACCTGCGGCTTATGAAGATTGCTGGACTGCCCTTCAGTGGGTGGCTT CCCATGTTGGTCTTGACAACCTCCAGCCTCAAGAATGCTATTGATAAAGAGCC TTGGATAATCAACCATGGCGACTTCAATAAGCTTTACTTGGGTGGTGACAGT CCTGGTGGAAATATTGTGCACAACGTAAGTATTAGAGCTGGTAAGGAGAGC TTGCATGGCGGAGTGAAAATCCGGGGTGCAATCTTTATTACCCATATTTCTT GATCAGGACAAGCAAAAGACAGAGTGATTATATGGAGATTGACTATAGAGG CTACTGGAAGTTGGCTTATCCATCTGCTCCTGGCGGCACTGACAACCCAATG ATAAACCTGTAGCTAAGAATGCTCCTGATTTGGCCGGATATGGATGTTCGA GGCTGCTTGTTCATGGTTTCGGACGAGACCAGAGATATAACCTTCTCTAC ATTGAGGCATTGAAGAAGAGTGGGTGGAAAGGTGAATTGGAAGTGGGTGA CTACGAAGCACATTTCTTTGATTTGTTTCAGCCCTGAAAATGAAGTTGGCAAG ACTTGGATCAAACGTTCAAGCGATTTTCATCAACAAGGAGTAA
<i>C. roseus</i> DPAS- GGGS- TurboID- FLAG	ATGGCCGGAAAATCAGCAGAAGAAGAACATCCCATTAAGGCTTACGGATGG GCTGTAAAGATAGAACAACCTGGGATTCTTTCTCCCTTCAAATTTTCCAGAAG GGCAACAGGTGATGATGATGTCCGAATTAAGATACTCTACTGTGGAATTTGT CACACTGATCTTGCCTCAATCAAGAACGAATACGAGTTTCTTTCTTATCCTCTT GTGCCCCGGGATGGAGATCGTTGGAATAGCAACGGAGGTTGGAAAAGATGT CACAAAAGTGAAAGTTGGCGAAAAAGTAGCATTATCAGCCTATTTAGGATGT TGTGGCAAATGCTATAGTTGTGTAATGAACTCGAGAATTATTGTCCGGAAG TAATCATAGGTTATGGCACCCCATACCATGACGGAACAATTTGCTATGGGGG CCTTTCAAACGAAACTGTCGCAAATCAAAGTTTTGTTCTTCGTTTTCTGAAA GACTTTCTCCAGCTGGCGGAGCTCCTTTGCTTAGCGCCGGAATTACTTCGTTT AGTGCAATGAGAAATAGCGGCATCGACAAACCTGGATTACACGTGGGAGTC GTCGGTCTCGGCGGATTAGGTCATCTTGCTGTAAAATTTGCTAAGGCTTTTG GTCCTTAAAGTAACTGTTATTAGCACCCTCCAGCAAGAAGGATGATGCTAT AAATGGTCTTGGTGCTGATGGATTCTTACTCAGCCGCGATGATGAACAAATG AAGGCTGCTATTGGAACCTTGGATGCAATTATTGATACTGGCGGTTGTTT ATCCCATAGCACCATTGCTTGATCTCCTGAGAAGTCAAGGGAAATTTTTGTTA CTTGGGGCGCCATCTCAATCACTTGAGTTGCCACCTATTCCTTTATTATCAGG TGGGAAATCTATCATTGGAAGTGCAGCCGAAATGTGAAGCAAACCTCAAGA AATGCTTGATTTTGCAGCGGAGCATGATATAACTGCAAATGTTGAGATTATT CCAATAGAGTACATAAATACTGCAATGGAACGTTTAGACAAGGGCGATGTT AGATACCGATTTGTAGTTGACATCGAAAATACCTTGACTCCTCCGTCAGAGTT AGGAGGCGGTGGATCGAAAGACAATACTGTGCCTCTGAAGCTGATCGCTCT CCTGGCTAATGGCGAGTTCCATAGTGGCGAACAGCTGGGAGAAACCCTGGG CATGTCCAGGGCCGCTATCAACAAGCACATTGACTCTGCGCGACTGGGG

	<p>CGTGGACGTGTTACCGTGCCCCGAAAGGGCTACTCTCTGCCCGAGCCTATC CCGCTGCTGAACGCTAACAGATTCTGGGACAGCTGGACGGCGGGAGCGTG GCAGTCCTGCCTGTGGTCGACTCCACCAATCAGTACCTGCTGGATCGAATCG GCGAGCTGAAGAGTGGGGATGCTTGCATTGCAGAATATCAGCAGGCAGGG AGAGGAAGCAGAGGGAGGAAATGGTTCTCTCTTTTGGAGCTAACCTGTAC CTGAGTATGTTTTGGCGCCTGAAGCGGGGACCAGCAGCAATCGGCCTGGGC CCGGTCATCGGAATTGTCATGGCAGAAGCGCTGCGAAAGCTGGGAGCAGAC AAGGTGCGAGTCAAATGGCCAATGACCTGTATCTGCAGGATAGAAAGCTG GCAGGCATCCTGGTGGAGCTGGCCGGAATAACAGGCGATGCTGCACAGATC GTCATTGGCGCCGGGATTAACGTGGCTATGAGGCGCGTGGAGGAAAGCGT GGTCAATCAGGGCTGGATCACACTGCAGGAAGCAGGGATTAACCTGGACAG GAATACTCTGGCCGCTACGCTGATCCGAGAGCTGCGGGCAGCCCTGGA GTTTCGAGCAGGAAGGCCTGGCTCCATATCTGCCACGGTGGGAGAAGCTGGA TAACTTCATCAATAGACCCGTGAAGCTGATCATTGGGGACAAAGAGATTTTC GGGATTAGCCGGGGGATTGATAAACAGGGAGCCCTGCTGCTGGAACAGGA CGGAGTTATCAAACCCTGGATGGGCGGAGAAATCAGTCTGCGGTCTGCCGA AAAGGACTACAAAGACGATGACGATAAA</p>
<p>YFP- GGGS- TurboID- FLAG</p>	<p>ATGGTGAGCAAGGGCGAGGAGCTGTTACCCGGGGTGGTGCCCATCCTGGTC GAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGG CGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACC GGCAAGCTGCCCCGTGCCCTGGCCCACCCTCGTGACCACCTTCGGCTACGGCG TGCAGTGCTTCGCCCGCTACCCCGACCACATGCGCCAGCAGACTTCTTCAA GTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGAC GACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC CTGGGGCACAAGCTGGAGTACAACAGCCACAACGTCTATATCATG GCCGACAAGCAGAAGAACGGCATCAAGGTGAAGTTCAGATCCGCCACAAC ATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC ATCGGCGACGGCCCCGTGCTGCTGCCGACAACCACTACCTGAGCTACCACT CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGG AGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGG GAGGCGGTGGATCGAAAGACAATACTGTGCCTCTGAAGCTGATCGCTCTCT GGCTAATGGCGAGTTCATAGTGGCGAACAGCTGGGAGAAACCCTGGGCAT GTCCAGGGCCGCTATCAACAAGCACATTCAGACTCTGCGCGACTGGGGCGT GGACGTGTTACCGTGCCCGGAAAGGGCTACTCTCTGCCCGAGCCTATCCCG CTGCTGAACGCTAACAGATTCTGGGACAGCTGGACGGCGGGAGCGTGGC AGTCCTGCCTGTGGTCGACTCCACCAATCAGTACCTGCTGGATCGAATCGGC GAGCTGAAGAGTGGGGATGCTTGCATTGCAGAATATCAGCAGGCAGGGAG AGGAAGCAGAGGGAGGAAATGGTTCTCTCTTTTGGAGCTAACCTGTACCT GAGTATGTTTTGGCGCCTGAAGCGGGGACCAGCAGCAATCGGCCTGGGCCC GGTCATCGGAATTGTCATGGCAGAAGCGCTGCGAAAGCTGGGAGCAGACA AGGTGCGAGTCAAATGGCCAATGACCTGTATCTGCAGGATAGAAAGCTGG CAGGCATCCTGGTGGAGCTGGCCGGAATAACAGGCGATGCTGCACAGATCG TCATTGGCGCCGGGATTAACGTGGCTATGAGGCGCGTGGAGGAAAGCGTG GTCAATCAGGGCTGGATCACACTGCAGGAAGCAGGGATTAACCTGGACAGG AATACTCTGGCCGCTACGCTGATCCGAGAGCTGCGGGCAGCCCTGGA ACTG</p>

TTCGAGCAGGAAGGCCTGGCTCCATATCTGCCACGGTGGGAGAAGCTGGAT
AACTTCATCAATAGACCCGTGAAGCTGATCATTGGGGACAAAGAGATTTTCG
GGATTAGCCGGGGGATTGATAAACAGGGAGCCCTGCTGCTGGAACAGGAC
GGAGTTATCAAACCCTGGATGGGCGGAGAAATCAGTCTGCGGTCTGCCGAA
AAGGACTACAAAGACGATGACGATAAA

Table 9. GenBank accessions of previously deposited sequences used in this chapter.

Gene Name	Organism	GenBank Accession
LAMT	<i>Catharanthus roseus</i>	EU057974
SLS	<i>Catharanthus roseus</i>	KF415117
TDC	<i>Catharanthus roseus</i>	X67662
STR	<i>Catharanthus roseus</i>	X53602
SGD	<i>Catharanthus roseus</i>	AF112888
GS	<i>Catharanthus roseus</i>	MF770507
GO	<i>Catharanthus roseus</i>	MF770508
RedOx1	<i>Catharanthus roseus</i>	MF770509
RedOx2	<i>Catharanthus roseus</i>	MF770510
SAT	<i>Catharanthus roseus</i>	MF770511
PAS	<i>Catharanthus roseus</i>	MH213134
DPAS	<i>Catharanthus roseus</i>	KU865331
TS	<i>Catharanthus roseus</i>	MF770513
CS	<i>Catharanthus roseus</i>	MF770512
T16H2	<i>Catharanthus roseus</i>	JF742645
16OMT	<i>Catharanthus roseus</i>	EF444544
T3O	<i>Catharanthus roseus</i>	KP122967
T3R	<i>Catharanthus roseus</i>	KP122966
THAS	<i>Catharanthus roseus</i>	KM524258
PAS1	<i>Tabernanthe iboga</i>	MK840850
PAS2	<i>Tabernanthe iboga</i>	MK840851
PAS3	<i>Tabernanthe iboga</i>	MK840852
DPAS1	<i>Tabernanthe iboga</i>	MK840855

DPAS2	<i>Tabernanthe iboga</i>	MK840856
TabS	<i>Tabernanthe iboga</i>	MK840853
CorS	<i>Tabernanthe iboga</i>	MK840854
NS2	<i>Strychnos nux-vomica</i>	OM304292
CXE1	<i>Papaver somniferum</i>	JQ659006
HIDH	<i>Glycine max</i>	AB154415
CAD	<i>Arabidopsis thaliana</i>	AY302081

Table 10. Primer sequences used for gene amplification and site-directed mutagenesis. Cloning overhangs are underlined. Mutated codons are in bold.

Primers for pCambia vectors	
<i>CrLAMT_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTT</u> ATGGTTGCCACAATTGATT
<i>CrLAMT_nLuc_Rev</i>	ACGAGATCTGGTCGAAATTTCCCTTGC GTTTCAAGACAA
<i>CrLAMT_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTT</u> ATGGTTGCCACAATTGATT
<i>CrLAMT_cLuc_Rev</i>	TACGAACGAAAGCTCAATTTCCCTTGC GTTTCAAGACAA
<i>CrSLS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTT</u> ATGGAGATGGATATGGATA
<i>CrSLS_nLuc_Rev</i>	ACGAGATCTGGTCGAAAGCTCTCAAGCTTCTTGTAGATG
<i>CrTDC_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTT</u> ATGGGCAGCATTGATTCA
<i>CrTDC_nLuc_Rev</i>	ACGAGATCTGGTCGAAAGCTTCTTTGAGCAAATCATCG
<i>CrTDC_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTT</u> ATGGGCAGCATTGATTCA
<i>CrTDC_cLuc_Rev</i>	TACGAACGAAAGCTCAAGCTTCTTTGAGCAAATCATCGG
<i>CrGS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTT</u> ATGGCCGGAGAAACAACCAA
<i>CrGS_nLuc_Rev</i>	ACGAGATCTGGTCGAAATTCCTCAAATTTCAATGTATTT
<i>CrGS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTT</u> ATGGCCGGAGAAACAACC
<i>CrGS_cLuc_Rev</i>	TACGAACGAAAGCTCATTCTCAAATTTCAATGTATTT
<i>CrGO_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTT</u> ATGGAGTTTTCTTTCTCCTCA
<i>CrGO_nLuc_Rev</i>	ACGAGATCTGGTCGAAATCGTTAACAAGATGAGGAACCA
<i>CrRedOx1_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTT</u> ATGGCTGATCGCGTGAAGAC
<i>CrRedOx1_nLuc_Rev</i>	ACGAGATCTGGTCGAAAGACAGCTACTGTTGCATTCCC
<i>CrRedOx1_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTT</u> ATGGCTGATCGCGTGAAGAC
<i>CrRedOx1_cLuc_Rev</i>	TACGAACGAAAGCTCAGACAGCTACTGTTGCAT
<i>CrRedOx2_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTT</u> ATGGAAAAGCAAGTTGAGATCCC

<i>CrRedOx2_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACAAGTCTCCATCCCAAAGCTC</u>
<i>CrRedOx2_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGAAAAGCAAGTTGAGATCCC</u>
<i>CrRedOx2_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACAAGTCTCCATCCCAAAGCT</u>
<i>CrSAT_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCACCCCAGATGCA</u>
<i>CrSAT_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAATTGCTAAAATCAGTGTCCAGAA</u>
<i>CrSAT_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCACCCCAGATGCA</u>
<i>CrSAT_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAATTGCTAAAATCAGTGTCCAGA</u>
<i>CrPAS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGATAAAAAAAGTCCCAATAG</u>
<i>CrPAS_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAAAGTTCGACTTGTAATGGAGAG</u>
<i>CrPAS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGATAAAAAAAGTCCCAATA</u>
<i>CrPAS_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAAAGTTCGACTTGTAATGGAGA</u>
<i>CrDPAS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCCGGAAAATCAGCAGA</u>
<i>CrDPAS_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAATAACTCTGACGGAGGAGTCAAG</u>
<i>CrDPAS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCCGGAAAATCAGCAGA</u>
<i>CrDPAS_cLuc_Rev</i>	<u>TACGAACGAAAGCTCATAACTCTGACGGAGGAGTCAA</u>
<i>CrTS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGGTTCTCAGATGAGACTA</u>
<i>CrTS_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACTTGATGAAAGAAGCTAAACGTC</u>
<i>CrTS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGGTTCTCAGATGAG</u>
<i>CrTS_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACTTGATGAAAGAAGCTAAACGTC</u>
<i>CrCS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGATGAGACTATTTGGGAT</u>
<i>CrCS_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAATTTGATGAAAGATGCTAAACGTC</u>
<i>CrCS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGATGAGACTATTTGGGAT</u>
<i>CrCS_cLuc_Rev</i>	<u>TACGAACGAAAGCTCATTGATGAAAGATGCTAAACGTC</u>
<i>CrCorS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCTTCCCAAACCTCCAA</u>
<i>CrCorS_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAATTTGATGAAAGACGTTAAGCGTC</u>
<i>CrCorS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCTTCCCAAACCTCCAA</u>
<i>CrCorS_cLuc_Rev</i>	<u>TACGAACGAAAGCTCATTGATGAAAGACGTTAAGCGTC</u>
<i>CrT16H2_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGAGTTGTATTATTTTCCACCTTT GCCTTCC</u>
<i>CrT16H2_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAATATTTACCTTTGAGAGAAGAAGCAGA ATAAGGAAATG</u>
<i>Cr16OMT_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGATGTTCAATCTGAGG</u>
<i>Cr16OMT_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAAGGATAAACCTCAATGAGACTCC</u>

<i>Cr16OMT_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGATGTTCAATCTGAG</u>
<i>Cr16OMT_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAAGGATAAACCTCAATGAGACTC</u>
<i>CrT30_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGAGTTTCATGAATCT</u>
<i>CrT30_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAATGCATAGGACGTAGCGATT</u>
<i>CrT3R_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGTCTAGTGAAATGGCT</u>
<i>CrT3R_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAGGGTGATTTGAAAGTGTTTCCA</u>
<i>CrT3R_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGTCTAGTGAAATGGCT</u>
<i>CrT3R_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAGGGTGATTTGAAAGTGTTTCCAA</u>
<i>CrCAD_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGGGAGCTTGGAAGAAGCA</u>
<i>CrCAD_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAGTGGTCAACAAGAAGGTTGCT</u>
<i>CrCAD_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGGGAGCTTGGAAGAA</u>
<i>CrCAD_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAGTGGTCAACAAGAAGGTTGCT</u>
<i>Cr2141_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCCGAAAATCACCAGAA</u>
<i>Cr2141_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAAGGAGCTTTCAAGGTCTTTGCA</u>
<i>Cr2141_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCCGAAAATCACCAGA</u>
<i>Cr2141_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAAGGAGCTTTCAAGGTCTTTGCA</u>
<i>CrADH9_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCTGGAAAATCACCAGA</u>
<i>CrADH9_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAAGGAGTTAGAGTGTTCCCAATAT</u>
<i>CrADH9_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCTGGAAAATCACCAGAA</u>
<i>CrADH9_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAAGGAGTTAGAGTGTTCCCAATAT</u>
<i>CrTHAS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCAATGGCTTCAAAGTCA</u>
<i>CrTHAS_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAATTTGATTTAGAGTGTTCCCTA</u>
<i>CrTHAS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCAATGGCTTCAAAGT</u>
<i>CrTHAS_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAATTTGATTTAGAGTGTTCCCTA</u>
<i>CrHID5_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCCTCCTCAGATGAGATT</u>
<i>CrHID5_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACTTTATAAAAGAGGCTATTTCGAT</u>
<i>CrHID5_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCCTCCTCAGATGAGATT</u>
<i>CrHID5_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACTTTATAAAAGAGGCTATTTCGA</u>
<i>CrCSE_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGCCTTCAGAAGCAGCGC</u>
<i>CrCSE_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACTTTCTGGGACCGTATTTTTGA</u>
<i>CrCSE_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGCCTTCAGAAGCAGC</u>
<i>CrCSE_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACTTTCTGGGACCGTATTTTTGA</u>

<i>CrCCR_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGCCGTCAGATTCCGGCA</u>
<i>CrCCR_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAAGAGCGGATTATTGTGAGGGGT</u>
<i>CrCCR_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGCCGTCAGATTCCGGCA</u>
<i>CrCCR_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAAGAGCGGATTATTGTGAGGGGT</u>
<i>CrC4H_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGATCTTCTCCTCTTAGAGA</u>
<i>CrC4H_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAAAAAGTCCTGGGCTTGAGTACAA</u>
<i>CrC3H_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGAACATTTCTTCCCCACTA</u>
<i>CrC3H_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAATATATCCACAGGCACACGTTTG</u>
<i>TiPAS1_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGTATACTACTGAAGTTC</u>
<i>TiPAS1_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAAAAGTTCGTCTTTGGAAGCAAGAG</u>
<i>TiPAS1_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGTATACTACTGAAGTTC</u>
<i>TiPAS1_cLuc_Rev</i>	<u>TACGAACGAAAGCTCAAAGTTCGTCTTTGGAAGCAAGAG</u>
<i>TiPAS2_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGTTGAAGTCTCTAAAGTTCT</u>
<i>TiPAS2_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACGATGATTCGTCTTGTGAAGTGA</u>
<i>TiPAS2_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTGTTGAAGTCTCTAAAGTTCT</u>
<i>TiPAS2_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACGATGATTCGTCTTGTGAAGTGA</u>
<i>TiPAS3_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGTTAGCAGAAGTCTCC</u>
<i>TiPAS3_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACAATTCATCATGTAAAGTTAGAG</u>
<i>TiPAS3_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGTTAGCAGAAGTCTCC</u>
<i>TiPAS3_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACAATTCATCATGTAAAGTTAGAG</u>
<i>TiDPAS1_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCTGTAAAATCACCTGAA</u>
<i>TiDPAS1_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAATTCGGTGGAGTTAGTGTGTT</u>
<i>TiDPAS1_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCTGTAAAATCACCTGA</u>
<i>TiDPAS1_cLuc_Rev</i>	<u>TACGAACGAAAGCTCATTCCGGTGGAGTTAGTGTGTT</u>
<i>TiDPAS2_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCAGGAAAATCACCAGAA</u>
<i>TiDPAS2_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACGGTTCTGGCGGAGGAGTTAA</u>
<i>TiDPAS2_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCAGGAAAATCACCAGAA</u>
<i>TiDPAS2_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACGGTTCTGGCGGAGGAGTTAA</u>
<i>TiTabS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCTTCTTCAACTGAAAG</u>
<i>TiTabS_nLuc_Rev</i>	<u>ACGAGATCTGGTCGAACTCCTTGTTGATGAAAGACGTTA</u>
<i>TiTabS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCTTCTTCAACTGAAAGCT</u>
<i>TiTabS_cLuc_Rev</i>	<u>TACGAACGAAAGCTCACTCCTTGTTGATGAAAGACGTTA</u>

<i>TiCorS_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCTAATTCAACTGCAA</u> ACT
<i>TiCorS_nLuc_Rev</i>	ACGAGATCTGGTCGAACTCCTTGTTGATGAAATCGCTT
<i>TiCorS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCTAATTCAACTGCAA</u> A
<i>TiCorS_cLuc_Rev</i>	TACGAACGAAAGCTCACTCCTTGTTGATGAAATCGCT
<i>AtCAD4_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGGAAGTGTAGAAGCAG</u>
<i>AtCAD4_nLuc_Rev</i>	ACGAGATCTGGTCGAAGTTTGTAGTTGTTGCAGC
<i>AtCAD4_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGGAAGTGTAGAAGCAG</u>
<i>AtCAD4_cLuc_Rev</i>	TACGAACGAAAGCTCAGTTTGTAGTTGTTGCAGC
<i>AtCHIL_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGGAACAGAGATGGTCATG</u>
<i>AtCHIL_nLuc_Rev</i>	ACGAGATCTGGTCGAAGTTAAACTGCGGAGATTGAATC
<i>AtCHS_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGTGATGGCTGGTGCTTC</u>
<i>AtCHS_cLuc_Rev</i>	TACGAACGAAAGCTCAGAGAGGAACGCTGTGCAAG
<i>SnvNS2_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGAAGTTGCGAATGCG</u>
<i>SnvNS2_nLuc_Rev</i>	ACGAGATCTGGTCGAAAACCACTTTCTCAGCTATCTCTAC
<i>SnvNS2_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGAAGTTGCGAATGCG</u>
<i>SnvNS2_cLuc_Rev</i>	TACGAACGAAAGCTCAAACCACTTTCTCAGCTATCTCTAC
<i>PsCXE1_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCAGATCCTTATGAATTCCTAAT</u>
<i>PsCXE1_nLuc_Rev</i>	ACGAGATCTGGTCGAAGTATATAAATTCGTCTGTTAAAATAAAA TGT
<i>PsCXE1_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCAGATCCTTATGAATTCCTAAT</u>
<i>PsCXE1_cLuc_Rev</i>	TACGAACGAAAGCTCAGTATATAAATTCGTCTGTTAAAATAAAA TGT
<i>GmHIDH_nLuc_Fwd</i>	<u>CGGGGGACGAGCTCGTTATGGCGAAGGAGATAGTGA</u>
<i>GmHIDH_nLuc_Rev</i>	ACGAGATCTGGTCGAAAACCAAGAAAAGAAGCCAAGCGT
<i>GmHIDH_cLuc_Fwd</i>	<u>ACGCGTCCCGGGGCGTTATGGCGAAGGAGATAGTGA</u>
<i>GmHIDH_cLuc_Rev</i>	TACGAACGAAAGCTCAAACCAAGAAAAGAAGCCAAGCGT

Primers for pOPIN vectors

<i>CrDPAS_pOPINF_Fwd</i>	<u>AAGTTCTGTTTCAGGGCCCGATGGCAGGTAAAAGCGCAGAAG</u> AAG
<i>CrDPAS_pOPINF_Rev</i>	<u>ATGGTCTAGAAAGCTTTACAGTTCGCTAGGCGGTGTCAG</u>
<i>CrCAD_pOPINF_Fwd</i>	<u>AAGTTCTGTTTCAGGGCCCGATGGGGAGCTTGGAAGAAG</u>
<i>CrCAD_pOPINF_Rev</i>	<u>ATGGTCTAGAAAGCTTTAGTGGTCAACAAGAAGGTTG</u>
<i>CrTS_pOPINF_Fwd</i>	<u>AAGTTCTGTTTCAGGGCCCGATGGGTTCTCAGATGAGACTATT</u> TTTG

<i>CrTS_pOPINF_Rev</i>	<u>ATGGTCTAGAAAGCTTTACTTGATGAAAGAAGCTAAACGTCTG</u> AG
<i>CrCS_pOPINF_Fwd</i>	<u>AAGTTCTGTTTCAGGGCCCGGCTCCCAAACCTCAACCTCAG</u> ATGA
<i>CrCS_pOPINF_Rev</i>	<u>ATGGTCTAGAAAGCTTTACTCATGTTTGATGAAAGATGCTAA</u> ACG
<i>CrGS_pOPINK_Fwd</i>	<u>AAGTTCTGTTTCAGGGCCCGATGGCTGGTGAAACCACCAAAC</u>
<i>CrGS_pOPINK_Rev</i>	<u>ATGGTCTAGAAAGCTTTATTCTTCGAATTCAGGGTGTTAC</u>

Primers for pHREAC vector

<i>CrDPAS_TurboID_Fwd</i>	<u>GGCTACGGTCTCTAAAAATGGCCGGAAAATCAGCAGAAGAAG</u>
<i>CrDPAS_TurboID_Rev</i>	<u>GGCTACGGTCTCGAGCGCTATTTATCGTCATCGTCTTTG</u>
<i>YFP_TurboID_Fwd</i>	<u>GGCTACGGTCTCTAAAAATGGTGAGCAAG</u>
<i>YFP_TurboID_Rev</i>	<u>GGCTACGGTCTCGAGCGCTATTTATCGTCATC</u>

Primers for mutagenesis

<i>CrCorS_Asn32Asp_Fwd</i>	GTAGAAAGACTCCAT G ATACTCCTTATGTTCC
<i>CrCorS_Asn32Asp_Rev</i>	ATGGAGTCTTTCTACTCTTCCATCTTTG
<i>CrCorS_Ser216Lys_Fwd</i>	AAACAGAGTGATTATATGGAGAATGACTAC
<i>CrCorS_Ser216Lys_Rev</i>	ATAATCACTCTGTTTTTGTCTCGTCCGGATCAGGA
<i>CrCorS_Asn224Ile_Fwd</i>	AAACAGAGTGATTATATGGAG A TAGACTACAGGTGT
<i>CrCorS_Asn224Ile_Rev</i>	CTCCATATAATCACTCTGTTTCGAGC
<i>CrCorS_Glu301His_Fwd</i>	TGGCTGATTTTGAAGC A CACTTTTTTGACCTTTTCCA
<i>CrCorS_Glu301His_Rev</i>	TGCTTCAAATCAGCCACATCCAATTGCCCT
<i>TiCorS_Asp32Asn_Fwd</i>	GTAGAAAGACTTCACA A TACCCCATATGTT
<i>TiCorS_Asp32Asn_Rev</i>	GTGAAGTCTTTCTACCTTGCCG
<i>TiCorS_Lys214Ser_Fwd</i>	AGACAGAGTGATTATATGGAGATTGAC
<i>TiCorS_Lys214Ser_Rev</i>	ATAATCACTCTGTCT G GAGCTTGTCTGA
<i>TiCorS_Ile222Asn_Fwd</i>	AGACAGAGTGATTATATGGAG A ACGACTATAGAGGCT
<i>TiCorS_Ile222Asn_Rev</i>	CTCCATATAATCACTCTGTCTTTTGCTTG
<i>TiCorS_His299Glu_Fwd</i>	GGTGACTACGAAGC A GAATTCTTTGATTTGT
<i>TiCorS_His299Glu_Rev</i>	TGCTTCGTAGTCACCCACTTCCA
<i>CrTS_Asn219Ile_Fwd</i>	ACAGAGTGATTATATGGAG A TAGAGTATAGATCTTACT
<i>CrTS_Asn219Ile_Rev</i>	CCATATAATCACTCTGTTTTGAACTCGTCCTGATTA
<i>CrCS_Tyr213Ile_Fwd</i>	AAACTTAGTGATGATTTTGAG A TAAACTACACATGTTACTGGA
<i>CrTS_Tyr213Ile_Rev</i>	AATCATCACTAAGTTTCGTGCTGGTTGGGATAATGAA

4.5.3. Plant growth

Nicotiana benthamiana plants used in transient overexpression experiments were grown for 3-4 weeks in a greenhouse with a 16-hour light/ 8-hour dark light cycle at 22 °C and 60% relative humidity. Plants were transferred to a York chamber with the same light and humidity conditions 24 hours before *Agrobacterium* infiltration. *Catharanthus roseus* cultivar “Sunstorm Apricot” seeds were obtained from Syngenta (2012). *C. roseus* plants used in transient overexpression experiments were grown for 4-5 weeks in a York chamber with a 16-hour light/ 8-hour dark light cycle at 21-28 °C and 60% relative humidity until agrobacterium-mediated transformation. *C. roseus* plants were grown in the lab using the same light source and under the same light cycle after infiltration. Plants used in all experiments were watered periodically as required.

4.5.4. Transient Overexpression in *N. benthamiana*

Constructs cloned into pCambia were transformed into *Agrobacterium tumefaciens* GV3101 cells (GoldBio) by electroporation and recovered in 1mL LB for 3 hours at 28 °C. Cells were plated on LB agar containing 30 µg/mL kanamycin, 30 µg/mL gentamicin and 20 µg/mL rifampicin and grown at 28 °C for 2 days. A single colony was grown in 10 mL LB containing 30 µg/mL kanamycin, 30 µg/mL gentamicin and 20 µg/mL rifampicin at 28 °C overnight shaking at 200 r.p.m. Cells were harvested by centrifugation at 4000 x g for 10 minutes and the pellet was gently resuspended in infiltration buffer (10 mM MES, 10 mM MgCl₂, pH 5.6, 200 µM acetosyringone) to OD₆₀₀ = 0.6. Resuspended cultures were incubated gently shaking in the dark for 3 hours before being mixed 1:1 with the remaining strain so each strain was at a final concentration of OD₆₀₀ = 0.3. Strains were infiltration in each quadrant of the abaxial side of the 3-4 week-old *N. benthamiana* leaf. Leaves from the 2nd fully expanded leaf pair were infiltrated. Four leaves from four different plants were infiltrated for each combination, and two independent repeats were carried out for each combination to reduce plant-to-plant and batch-to-batch effects. Four leaves were infiltrated with the *A. thaliana* chalcone isomerase-like *AtCHIL*-nLuc and *A. thaliana* chalcone synthase *AtCHS*-cLuc constructs in each batch as a positive control.

4.5.5. Split-luciferase Complementation Assays

N. benthamiana leaves overexpressing nLuc or cLuc constructs were imaged 3 days post infiltration in a NightShade LB 985 (Berthold Technologies). Leaves were sprayed with 0.5 mM solution of d-luciferin and incubated in the dark for 5 minutes before being imaged on their abaxial side. Images were exposed for 0.1 seconds and luminescence emission was exposed for 20 seconds with 8 x 8 pixel binning. Up to four leaves of the same combination were imaged per experiment until three leaves gave the same result, and each experiment was repeated an independent time. The previously published *AtCHIL*-nLuc and *AtCHS*-cLuc combination was used as a positive control in each experimental batch of plants (Figure 70) [50].

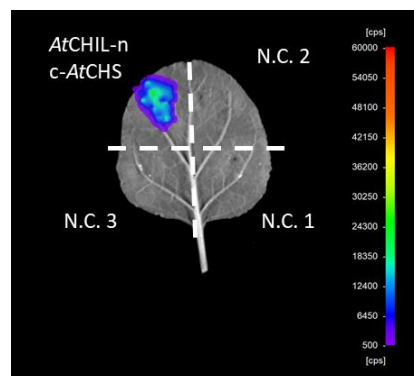


Figure 70. Representative image of split-luciferase positive control *AtCHIL* and *AtCHS* in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents *AtCHIL*-nLuc construct and cLuc fragment, N.C. 2 represents *AtCHS*-cLuc construct and nLuc fragment, N.C. 3 represents nLuc and cLuc fragment negative controls. Luminescence measured by counts per second (cps) and represented by false colour.

4.5.6. Transient Overexpression in *C. roseus*

Constructs cloned into pCambia were transformed into *Agrobacterium tumefaciens* GV3101 cells (GoldBio) by electroporation and recovered in 1mL LB for 3 hours at 28 °C. Cells were plated on LB agar containing 30 µg/mL kanamycin, 30 µg/mL gentamicin and 20 µg/mL rifampicin and grown at 28 °C for 2 days. A single colony was grown in 10 mL LB containing 30 µg/mL kanamycin, 30 µg/mL gentamicin and 20 µg/mL rifampicin at 28 °C overnight shaking at 200 r.p.m. 50 µL of overnight culture was transferred to 100 mL LB containing 30 µg/mL kanamycin, 30 µg/mL gentamicin and 20 µg/mL rifampicin and grown

at 28 °C overnight shaking at 200 r.p.m. Cells were harvested by centrifugation at 4000 x g for 10 minutes and the pellet was gently resuspended in infiltration buffer (10 mM MES, 10 mM MgCl₂, pH 5.6, 200 μM acetosyringone) to OD₆₀₀ = 0.8. Resuspended cultures were incubated gently shaking in the dark for 3 hours. 50 mL of resuspended culture was placed in a glass beaker with 0.002% v/v Silwet L-77 (UPL Germany). 4 1 cm holes were made into each corner of both leaves of the first fully expanded leaf pair of 4-5 week *C. roseus* plants using a hole puncher. A plastic cover placed over the plant pots to prevent soil fall out. Plants were submerged in resuspended cultures and placed in a desiccator. Plants were vacuum infiltrated at 30 mBar for 5 minutes before the vacuum was rapidly released and returned to room pressure. This was repeated for a total of five rounds. Plants were removed from the resuspension culture and left in the dark overnight before being returned to the normal growth light cycle. Maximal fluorophore signal was observed in the first fully expanded leaf pair 5 days post infiltration.

4.5.7. TurboID Proximity-labelling in *C. roseus*

C. roseus plants 5 d.p.i expressing TurboID constructs were submerged in a glass beaker containing a 50 mL solution of 50 μM biotin. Biotin was fed by vacuum infiltration of plants at 30 mBar for 5 minutes before the vacuum was rapidly released and returned to room pressure. This was repeated for a total of five rounds. Plants were returned to normal growth conditions and harvested at various time points. Excessive labelling time may result in aspecific tagging. Sufficient biotin labelling was observed after 3 hours, therefore tissue from the first fully expanded leaf pair was collected after 3 hours and snap-frozen in liquid nitrogen. Tissue was stored at -80 °C until protein extraction.

Harvested tissue was crushed to a fine powder in a liquid nitrogen-cooled mortar and pestle and then resuspended in protein buffer (25 mM Tris-HCl and 20 mM NaCl pH 8.0) at 1g fresh tissue weight per 1 mL protein buffer. Samples were incubated gently rocking for an hour at 4 °C. Samples were centrifuged at 3500 x g for 15 minutes and the supernatant was removed. The supernatant was then centrifuged at 17000 x g for 20 minutes. Protein was quantified using Pierce Rapid Gold BCA Protein Assay Kit (Thermo Fisher) according to the manufacturer's instructions on 96-well clear flat-bottomed plates (Sigma-Aldrich) and absorbance was measured using a CLARIOstar plate reader (BMG Labtech). Protein extracts

were normalised to 12 mg and made up to a final volume of 1 mL in protein buffer. 100 μ L of Streptavidin Dynabeads (Thermo Fisher) per sample were washed in 1 mL PBS buffer gently by pipetting. Samples were then placed on the DynaMag SampleRack (Thermo Fisher) for 3 minutes to separate the beads and the supernatant removed. This washing process was repeated a further 2 times. 1 mL of the normalised sample was added to 100 μ L of washed beads and samples were incubated gently rocking at room temperature for 30 minutes. Samples were placed on the DynaMag SampleRack for 3 minutes to separate the beads and the supernatant removed. Beads were resuspended in 1 mL of PBS. This washing step was repeated a further 2 times. Proteins were eluted from DynaBeads with 50 μ L 2X Lamelli buffer (Thermo Fisher) containing 25 mM biotin and placed on the DynaMag SampleRack for 3 minutes and the supernatant was collected. The supernatant was boiled at 95 $^{\circ}$ C for 5 minutes to denature the proteins. Protein samples were then frozen at -20 $^{\circ}$ C before being analysed by proteomics. Protein quantity was estimated by running samples on SDS-PAGE (Thermo Fisher) and staining the gel using the Pierce Silver Stain kit (Thermo Fisher) according to the manufacturer's instructions (Figure 71).

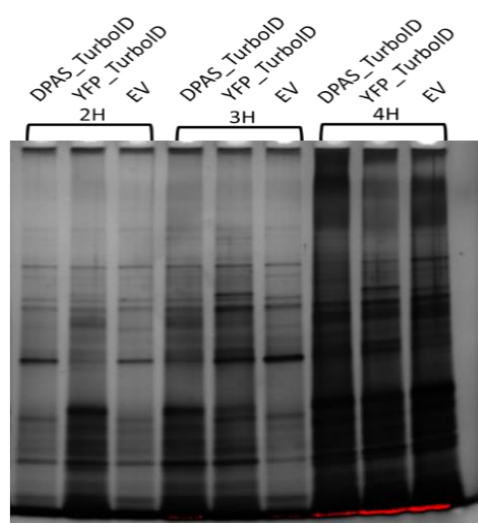


Figure 71. SDS-PAGE of DPAS-TurboID, YFP-TurboID and pHREAC empty vector (EV) streptavidin pull-down assays. Samples taken at 2, 3 and 4 hour post biotin feeding time points. Gel developed using ProteoSilver staining.

4.5.8. Heterologous Expression and Protein Purification from *E. coli*

CrDPAS, *CrCAD* *CrTS* and *CrCS* constructs cloned into pOPINF plasmid and *CrGS* construct cloned into pOPINK plasmid. Plasmids were transformed into *E. coli* BL21 (DE3) (ThermoFisher) cells by heat shock at 42 $^{\circ}$ C for 30 seconds. Cells were plated on LB agar

containing 100 µg/mL carbenicillin plates and grown overnight at 37 °C. A single colony was then picked and grown in 10 mL 2x YT media containing 100 µg/mL carbenicillin at 37 °C shaking at 200 r.p.m. overnight.

For *CrDPAS*, *CrTS* and *CrCS* protein expression, 1 mL of the overnight culture was added to 1L 2xYT media containing 100 µg/mL carbenicillin and grown at 37 °C shaking at 200 r.p.m. until $OD_{600} = 0.6-0.8$. Cultures were then transferred to an 18 °C incubator shaking at 200 r.p.m for 30 minutes before protein production was induced with 0.2 mM IPTG and incubated overnight (16-18h). Cells were harvested by centrifugation at 3200 x g for 15 minutes and re-suspended in 50 mL buffer A1 (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) with the addition of EDTA-free protease inhibitor cocktail (Roche Diagnostics Ltd.) and 10 mg lysozyme (Sigma). Cells were lysed on ice using a sonicator (40% amplitude, 2 seconds on, 3 seconds off cycles for 3 minutes) and centrifuged (35000 x g) to remove insoluble cell debris. The supernatant was collected and filtered with a 0.2 µm PES syringe filter (Sartorius) and purified using an AKTA Pure FPLC (Cytiva). The sample was applied at 2 mL/min onto a His-Trap High-Performance 5mL column (Cytiva) and washed with 5 column volumes (CV) of buffer A1 before being eluted with 5 CV of buffer B1 (50 mM Tris-HCl pH 8.0, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole). Protein was detected and collected using the UV 280 nm signal and then further purified on a Superdex Hiload 16/60 S200 gel filtration column (Cytiva) at a flow rate of 1 mL/min using buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl). Proteins were finally buffer exchanged into buffer A4 and concentrated using a 10K Da molecular weight cut-off centrifugal filter (Merck) before being snap frozen in liquid nitrogen and stored at -80 °C.

For *CrCAD* and *CrGS* protein expression, 1 mL of the overnight culture was added to 100 mL 2xYT media containing 100 µg/mL carbenicillin and grown at 37 °C shaking at 200 r.p.m. until $OD_{600} = 0.6-0.8$. Cultures were then transferred to an 18 °C incubator shaking at 200 r.p.m for 30 minutes before protein production was induced with 0.2 mM IPTG and incubated overnight (16-18h). Cells were harvested by centrifugation at 4000 x g for 15 minutes and re-suspended in 10 mL buffer A1 (50 mM Tris-HCl pH 8, 50 mM glycine, 500 mM NaCl, 5% glycerol, 20 mM imidazole) with the addition of EDTA-free protease inhibitor cocktail (Roche Diagnostics Ltd.) and 10 mg lysozyme (Sigma). Cells were lysed at 4 °C using

a sonicator (40% amplitude, 2 seconds on, 3 seconds off cycles for 1.5 minutes) and centrifuged at 35000 x g to remove insoluble cell debris. The supernatant was collected and filtered with a 0.2 µm PES syringe filter (Sartorius) and purified by the addition of 150 µL washed Ni-NTA agarose beads (QIAGEN). Samples were incubated on a rocking incubator at 4 °C for 1 hour. Beads were washed by centrifuging at 1000 x g for 1 minute to remove the supernatant, and then the beads were resuspended in 10 mL of A1 Buffer. This step was performed a total of three times. Protein was eluted by resuspending the beads in 600 µL of buffer B1 (50 mM Tris-HCl pH 8.0, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole) before centrifuging for 1000 x g for 1 minute and then collecting the supernatant. This elution step was repeated to remove all Ni-NTA-bound proteins. Proteins were buffer exchanged into buffer A4 (20 mM HEPES pH 7.5, 150 mM NaCl) and concentrated using a 10K Da molecular weight cut-off centrifugal filter (Merck) and stored at -80 °C.

4.5.9. *In vitro* Enzyme Assays

Enzymatic assays with cinnamyl aldehyde were performed in 50 mM HEPES buffer (pH 7.5) with 50 µM substrate, 250 µM NADPH cofactor (Sigma) and 500 nM enzyme to a final reaction volume of 100 µL. Reactions were incubated for 30 minutes at 30°C and shaken at 60 r.p.m. before being quenched with 1 volume of 70% MeOH with 0.1% HCO₂H.

4.5.10. UPLC-MS Analysis

All assays were analysed using a Thermo Scientific Vanquish UPLC coupled to a Thermo Q Exactive Plus orbitrap MS. For assays using precondylocarpine acetate, chromatographic separation was performed using a Phenomenex Kinetex C18 2.6 µm (2.1 x 100 mm) column using water with 1% HCO₂H as mobile phase A and acetonitrile with 1% HCO₂H as mobile phase B. Compounds were separated using a linear gradient of 10-30% B in 5 minutes followed by 1.5 minutes isocratic at 100% B. The column was then re-equilibrated at 10% B for 1.5 minutes. The column was heated to 40°C and the flow rate was set to 0.6 mL/min. For assays using strictosidine aglycone, separation was carried out using a Waters Acquity BEH C18 1.7 µm (2.1 x 50 mm) using 0.1% NH₄OH in water as mobile phase A and acetonitrile as mobile phase B. Compounds were separated using a

linear gradient of 10-90% B in 9 minutes followed by 2 minutes isocratic at 90% B. The column was re-equilibrated at 10% B for 3 minutes. The column was heated to 50°C and the flow rate was set to 0.4 mL/min. MS detection was performed in positive ESI under the following conditions: spray voltage was set to 3.5 kV ~ 67.4 μ A, capillary temperature set to 275°C, vaporizer temperature 475°C, sheath gas flow rate 65, sweep gas flow rate 3, aux gas flow rate 15, S-lens RF level to 55 V. Scan range was set to 200 - 1000 m/z and resolution at 17500.

4.5.11. Complex Affinity Co-Purification

5g of *C. roseus* leaf tissue was snap-frozen in liquid nitrogen and ground using a mortar and pestle until it formed a fine powder. The powder was added to 10 mL A1 buffer with the addition of EDTA-free protease inhibitor cocktail (Roche Diagnostics Ltd.) and 10 mg lysozyme (Sigma) and incubated gently rocking at 4 °C for 2 hours. Protein extracts were centrifuged at 1000 x g for 1 minute to remove insoluble cell debris and 1mg of 6X-His-tagged purified CrDPAS or CrTS was added to the supernatant. The samples were incubated gently rocking at 4 °C for 2 hours. Ni-NTA beads (QIAGEN) were washed by centrifuging at 1000 x g for 1 minute to remove the supernatant, and then the beads were resuspended in 10 mL of A1 Buffer. This step was performed a total of three times. 200 μ L washed Ni-NTA agarose beads were added to each sample and incubated on a rocking incubator at 4 °C for 1 hour. Samples were centrifuged at 1000 x g for 1 minute to remove the supernatant, and then the beads were resuspended in 10 mL of A1 Buffer. This step was performed a total of three times. Protein was eluted by resuspending the beads in 600 μ L of buffer B1 (50 mM Tris-HCl pH 8.0, 50 mM glycine, 500 mM NaCl, 5% glycerol, 500 mM imidazole) before centrifuging for 1000 x g for 1 minute and then collecting the supernatant. This elution step was repeated to remove all Ni-NTA-bound proteins. Protein samples were snap frozen and stored at -20 °C until analysed by proteomics.

4.5.12. Thermal Stability Assays

Protein was resuspended to the desired concentrations in DI water and stored on ice until use. A 50X stock solution SYPRO Orange Dye (Thermo Fisher) was prepared by diluting 2.5 μ L into 250 μ L of DI water and kept in the dark until use. 12.5 μ L of each buffer solution

was added to the desired wells of a clear 96-well PCR plate (Thermo Fisher). JBScreen Thermofluor FUNDAMENT HTS buffer solutions (Jena Bioscience) were used to analyse the optimal pH and salt concentration of buffers for protein melt curves. For analysis of the effect of substrate and cofactor, 50 mM HEPES (pH 8.0) buffer was used and samples were carried out in triplicate. 10 μ L of resuspended protein and 2.5 μ L of the 50X stock solution of SYPRO Orange Dye was added to each well to a final volume of 25 μ L. Plates were sealed using a PCR Plate Sealing Film (BioRad) and centrifuged at 300 x g for 1 minute. Melt temperature analysis was performed using a CFX96 Touch Real-Time PCR system (BioRad) measuring 490 nm/ 580 nm Ex/Em wavelengths. Starting temperature was set to 10 °C to 95 °C in 0.5 °C increments for 10 seconds. Melt curves analysis was performed using CFX Maestro software (BioRad).

4.5.13. Co-purification and TurboID Sample Proteomics

For proteomics, samples were dissolved in 6 M guanidine HCl, 50 mM Tris-HCl (pH 8.0), 3 mM DTT and heated at 37 °C for 1 hour. 50 mM NH_4HCO_3 and 1 mM CaCl_2 was added until guanidine HCl concentration was reduced below 1 M. Trypsin protease was added and incubated at 37 °C overnight and the reaction stopped by freezing at -20 °C. Peptides were identified based on the *C. roseus* transcriptome.

4.5.14. *C. roseus* Lignin Extraction and Quantification

5 mg of leaf tissue was harvested from the first fully expanded leaf pair of *C. roseus* plants 5 d.p.i. Tissue was sliced into 1 mm strips using a blade and placed in 2 mL glass UPLC vials. The lignin extraction and quantification method was based on the acetyl-bromide-based protocol previously published by Chang *et al.*,^[51]. Briefly, to extract the cell wall residue, 1 mL DI water was added to each sample and heated to 98 °C for 30 minutes in a water bath. The water was removed from each sample and discarded, and 1 mL of 100% ethanol was added, and samples were heated to 76 °C for 30 minutes. Ethanol was removed and 1 mL chloroform was added and heated to 59 °C for 30 minutes. Chloroform was removed and 1 mL acetone was added and heated to 54 °C for 30 minutes. Acetone was removed and samples were dried in a GeneVac at 40 °C under 30 mBar for approximately 4 hours until

completely dry. Cell wall residue was dissolved in 0.1 mL 25% acetyl bromide in acetic acid and 4 μ L 60% perchloric acid was added. Samples were incubated at 70 °C for 30 minutes shaking at 850 r.p.m. until completely dissolved. Samples were then centrifuged at 21300 x g for 15 minutes and the supernatant separated from the pellet. 0.2 mL 2M NaOH and 0.5 mL acetic acid was added to the supernatant. The pellet was washed with 0.5 mL acetic acid. The pellet wash was combined with the supernatant and a further 1.2 mL acetic acid was added to a final sample volume of 2 mL. Samples were incubated for 20 minutes at room temperature. Extracted lignin was measured by A_{280nm} value using a Nanodrop and values normalised by the sample fresh weight.

4.5.15. Phylogenetic Analysis

Nucleic acid sequences of ADH and α/β -hydrolase genes were aligned using MUSCLE5^[52]. Sequences used for ADH alignment are shown in Figure 63A described in Table 8 and Table 9, and sequences used for α/β -hydrolases alignment are shown in Figure 63B described in Table 8 and Table 11. Maximum likelihood phylogenetic trees were constructed using IQTree using a best-fit substitution model followed by tree reconstruction using 1000 bootstrap alignments and the remaining parameters used default settings^[53]. Tree visualisation and figures were made using iTOL version 6.5.2^[54].

Table 11. Genbank accessions for sequences used to construct the tree of maximum likelihood in Figure 6B.

Gene Name	Genbank accession
<i>Arabidopsis thaliana</i> Caffeoyl shikimate esterase (CSE)	NP_175685
<i>Arabidopsis thaliana</i> Carboxylesterase 10 (CXE10)	AT3G05120
<i>Arabidopsis thaliana</i> Carboxylesterase 12 (CXE12)	AT3G48690
<i>Arabidopsis thaliana</i> Carboxylesterase 14 (CXE14)	AT3G63010
<i>Arabidopsis thaliana</i> Carboxylesterase 15 (CXE15)	AT5G06570
<i>Arabidopsis thaliana</i> Carboxylesterase 19 (CXE19)	AT5G27320
<i>Arabidopsis thaliana</i> Carboxylesterase 20 (CXE20)	AT5G62180
<i>Capsicum annuum</i> Esterase (PepEST)	AAF77578
<i>Catharanthus roseus</i> Tabersonine synthase (TS)	MF770513
<i>Catharanthus roseus</i> Catharanthine synthase (CS)	MF770512

<i>Glycine max</i> 2-hydroxyisoflavanone dehydratase (HIDH)	NP_001389 539
<i>Nicotiana tabacum</i> Carboxylesterase 15 (HSR203J)	XP_016474 715
<i>Oryza sativa</i> Gibberellin insensitive dwarf 1 (GID1)	BAE45340
<i>Papaver somniferum</i> Carboxylesterase 1 (CXE1)	AFB74618
<i>Pisum sativum</i> HSR203J homolog (E86)	BAA85654
<i>Rauwolfia serpentina</i> Polyneuridine aldehyde esterase (PNAE)	AF178576
<i>Solanum pennellii</i> Acylsugar acyl hydrolase 1 (ASH1)	KT282359
<i>Solanum pennellii</i> Acylsugar acyl hydrolase 2 (ASH2)	KT282360
<i>Solanum pennellii</i> Acylsugar acyl hydrolase 3 (ASH3)	KT282361
<i>Strychnos nux-vomica</i> Norfluorocurarine synthase 1 (NS1)	OM304291
<i>Strychnos nux-vomica</i> Norfluorocurarine synthase 2 (NS2)	OM304292
<i>Tabernanthe iboga</i> Tabersonine synthase (TabS)	MK840853
<i>Tabernanthe iboga</i> Coronaridine synthase (CorS)	MK840854
<i>Tulipa gesneriana</i> Tuliposide A-converting enzyme 1 (TCEA1)	I4DST8
<i>Tulipa gesneriana</i> Tuliposide A-converting enzyme 2 (TCEA2)	I4DST9

4.6. References

- [1] Y. Qu, M. L. A. E. Easson, J. Froese, R. Simionescu, T. Hudlicky, V. DeLuca, *Proc National Acad Sci* 2015, *112*, 6224–6229.
- [2] L. Caputi, J. Franke, S. C. Farrow, K. Chung, R. M. E. Payne, T.-D. Nguyen, T.-T. T. Dang, I. S. T. Carqueijeiro, K. Koudounas, T. D. de Bernonville, B. Ameyaw, D. M. Jones, I. J. C. Vieira, V. Courdavault, S. E. O'Connor, *Science* 2018, *360*, 1235–1239.
- [3] Y. Qu, M. E. A. M. Easson, R. Simionescu, J. Hajicek, A. M. K. Thamm, V. Salim, V. D. Luca, *Proc National Acad Sci* 2018, *115*, 3180–3185.
- [4] C. Li, J. C. Wood, A. H. Vu, J. P. Hamilton, C. E. R. Lopez, R. M. E. Payne, D. A. S. Guerrero, K. Gase, K. Yamamoto, B. Vaillancourt, L. Caputi, S. E. O'Connor, C. R. Buell, *Nat Chem Biol* 2023, 1–11.
- [5] B. St-Pierre, V. D. Luca, *Plant Physiol* 1995, *109*, 131–139.
- [6] S. Besseau, F. Kellner, A. Lanoue, A. M. K. Thamm, V. Salim, B. Schneider, F. Geu-Flores, R. Höfer, G. Guirimand, A. Guihur, A. Oudin, G. Glevarec, E. Foureau, N. Papon, M. Clastre, N. Giglioli-Guivarc'h, B. St-Pierre, D. Werck-Reichhart, V. Burlat, V. D. Luca, S. E. O'Connor, V. Courdavault, *Plant Physiol* 2013, *163*, 1792–803.
- [7] J. G. M. Eng, M. Shahsavarani, D. P. Smith, J. Hájíček, V. D. Luca, Y. Qu, *Nat Commun* 2022, *13*, 3335.
- [8] A. Edge, Y. Qu, M. L. A. E. Easson, A. M. K. Thamm, K. H. Kim, V. D. Luca, *Planta* 2018, *247*, 155–169.
- [9] E. C. Tatsis, I. Carqueijeiro, T. D. D. Bernonville, J. Franke, T.-T. T. Dang, A. Oudin, A. Lanoue, F. Lafontaine, A. K. Stavrinos, M. Clastre, V. Courdavault, S. E. O'connor, *Nat Commun* 2017, *8*, 316.
- [10] V. Deluca, J. Balsevich, R. T. Tyler, W. G. Kurz, *Plant Cell Rep* 1987, *6*, 458–61.
- [11] K. Yamamoto, K. Takahashi, L. Caputi, H. Mizuno, C. E. Rodriguez-Lopez, T. Iwasaki, K. Ishizaki, H. Fukaki, M. Ohnishi, M. Yamazaki, T. Masujima, S. E. O'Connor, T. Mimura, *New Phytol* 2019, *224*, nph.16138.
- [12] G. Guirimand, A. Guihur, P. Poutrain, F. Héricourt, S. Mahroug, B. St-Pierre, V. Burlat, V. Courdavault, *J Plant Physiol* 2011, *168*, 549–557.
- [13] G. Guirimand, A. Guihur, O. Ginis, P. Poutrain, F. Héricourt, A. Oudin, A. Lanoue, B. St-Pierre, V. Burlat, V. Courdavault, *Febs J* 2011, *278*, 749–763.
- [14] S. Mahroug, V. Courdavault, M. Thiersault, B. St-Pierre, V. Burlat, *Planta* 2006, *223*, 1191–1200.

- [15] G. Guirimand, V. Courdavault, A. Lanoue, S. Mahroug, A. Guihur, N. Blanc, N. Giglioli-Guivarc'h, B. St-Pierre, V. Burlat, *Bmc Plant Biol* 2010, *10*, 182.
- [16] A. Stavriniades, E. C. Tatsis, E. Foureau, L. Caputi, F. Kellner, V. Courdavault, S. E. O'Connor, *Chem Biol* 2015, *22*, 336–41.
- [17] R. M. E. Payne, D. Xu, E. Foureau, M. I. S. T. Carqueijeiro, A. Oudin, T. D. de Bernonville, V. Novak, M. Burow, C.-E. Olsen, D. M. Jones, E. C. Tatsis, A. Pendle, B. A. Halkier, F. Geu-Flores, V. Courdavault, H. H. Nour-Eldin, S. E. O'Connor, *Nat Plants* 2017, *3*, 16208.
- [18] A. Rai, H. Hirakawa, R. Nakabayashi, S. Kikuchi, K. Hayashi, M. Rai, H. Tsugawa, T. Nakaya, T. Mori, H. Nagasaki, R. Fukushima, Y. Kusuya, H. Takahashi, H. Uchiyama, A. Toyoda, S. Hikosaka, E. Goto, K. Saito, M. Yamazaki, *Nat Commun* 2021, *12*, 405.
- [19] A. Stavriniades, E. C. Tatsis, L. Caputi, E. Foureau, C. E. M. Stevenson, D. M. Lawson, V. Courdavault, S. E. O'Connor, *Nat Commun* 2016, *7*, 12116.
- [20] A. K. Stavriniades, E. C. Tatsis, T.-T. Dang, L. Caputi, C. E. M. Stevenson, D. M. Lawson, B. Schneider, S. E. O'Connor, *Chembiochem* 2018, *19*, 940–948.
- [21] C. Langley, E. Tatsis, B. Hong, Y. Nakamura, C. Paetz, C. E. M. Stevenson, J. Basquin, D. M. Lawson, L. Caputi, S. E. O'Connor, *Angew. Chem. Int. Ed.* 2022, *61*, e202210934.
- [22] S. C. Farrow, M. O. Kamileen, L. Caputi, K. Bussey, J. E. A. Mundy, R. C. McAtee, C. R. J. Stephenson, S. E. O'Connor, *J Am Chem Soc* 2019, *141*, 12979–12983.
- [23] M. O. Kamileen, M. D. DeMars, B. Hong, Y. Nakamura, C. Paetz, B. R. Lichman, P. D. Sonawane, L. Caputi, S. E. O'Connor, *J Am Chem Soc* 2022, DOI 10.1021/jacs.2c08107.
- [24] J. H. Morris, G. M. Knudsen, E. Verschueren, J. R. Johnson, P. Cimermancic, A. L. Greninger, A. R. Pico, *Nat. Protoc.* 2014, *9*, 2539–2554.
- [25] A. Mair, S. L. Xu, T. C. Branon, A. Y. Ting, D. C. Bergmann, *Elife* 2019, *8*, e47864.
- [26] D. Arora, N. B. Abel, C. Liu, P. van Damme, K. Yperman, D. Eeckhout, L. D. Vu, J. Wang, A. Tornkvist, F. Impens, B. Korbei, J. van Leene, A. Goossens, G. de Jaeger, T. Ott, P. N. Moschou, D. van Damme, *Plant Cell* 2020, *32*, 3388–3407.
- [27] K. Kido, S. Yamanaka, S. Nakano, K. Motani, S. Shinohara, A. Nozawa, H. Kosako, S. Ito, T. Sawasaki, *Elife* 2020, *9*, e54983.
- [28] K. F. Cho, T. C. Branon, S. Rajeev, T. Svinkina, N. D. Udeshi, T. Thoudam, C. Kwak, H. W. Rhee, I. K. Lee, S. A. Carr, A. Y. Ting, *Proc National Acad Sci* 2020, *117*, 12143–12154.
- [29] M. Fairhead, M. Howarth, *Methods Mol. Biol.* 2014, *1266*, 171–184.
- [30] T. Azad, A. Tashakor, S. Hosseinkhani, *Anal. Bioanal. Chem.* 2014, *406*, 5541–5560.

- [31] K. Gao, R. Oerlemans, M. R. Groves, *Biophys. Rev.* 2020, *12*, 85–104.
- [32] L. Caputi, J. Franke, K. Bussey, S. C. Farrow, I. J. C. Vieira, C. E. M. Stevenson, D. M. Lawson, S. E. O'Connor, *Nat Chem Biol* 2020, *16*, 383–386.
- [33] T. C. Branon, J. A. Bosch, A. D. Sanchez, N. D. Udeshi, T. Svinkina, S. A. Carr, J. L. Feldman, N. Perrimon, A. Y. Ting, *Nat Biotechnol* 2018, *36*, 880–887.
- [34] S. de Vries, J. M. R. Fürst-Jansen, I. Irisarri, A. D. Ashok, T. Ischebeck, K. Feussner, I. N. Abreu, M. Petersen, I. Feussner, J. de Vries, *The Plant Journal* 2021, DOI 10.1111/tpj.15387.
- [35] C.-Y. Lin, Y. Sun, J. Song, H.-C. Chen, R. Shi, C. Yang, J. Liu, S. Tunlaya-Anukit, B. Liu, P. L. Loziuk, C. M. Williams, D. C. Muddiman, Y.-C. J. Lin, R. R. Sederoff, J. P. Wang, V. L. Chiang, *Front Plant Sci* 2021, *12*, 727932.
- [36] J. P. Wang, B. Liu, Y. Sun, V. L. Chiang, R. R. Sederoff, *Front Plant Sci* 2019, *9*, 1942.
- [37] H.-C. Chen, Q. Li, C. M. Shuford, J. Liu, D. C. Muddiman, R. R. Sederoff, V. L. Chiang, *Proc National Acad Sci* 2011, *108*, 21253–21258.
- [38] M. Gou, X. Ran, D. W. Martin, C.-J. Liu, *Nat Plants* 2018, *4*, 299–310.
- [39] X. Yan, J. Liu, H. Kim, B. Liu, X. Huang, Z. Yang, Y. J. Lin, H. Chen, C. Yang, J. P. Wang, D. C. Muddiman, J. Ralph, R. R. Sederoff, Q. Li, V. L. Chiang, *New Phytol* 2019, *222*, 244–260.
- [40] M. Dastmalchi, *The Plant Journal* 2021, DOI 10.1111/tpj.15446.
- [41] T. Nakayama, S. Takahashi, T. Waki, n.d.
- [42] T. Nakayama, S. Takahashi, T. Waki, *Front Plant Sci* 2019, *10*, 821.
- [43] J. I. Watkinson, P. A. Bowerman, K. C. Crosby, S. B. Hildreth, R. F. Helm, B. S. J. Winkel, *PeerJ* 2018, *2018*, e5598.
- [44] T. Waki, D. C. Yoo, N. Fujino, R. Mameda, K. Denessiouk, S. Yamashita, R. Motohashi, T. Akashi, T. Aoki, S. I. Ayabe, S. Takahashi, T. Nakayama, *Biochem Bioph Res Co* 2016, *469*, 546–551.
- [45] N. Fujino, N. Tenma, T. Waki, K. Ito, Y. Komatsuzaki, K. Sugiyama, T. Yamazaki, S. Yoshida, M. Hatayama, S. Yamashita, Y. Tanaka, R. Motohashi, K. Denessiouk, S. Takahashi, T. Nakayama, *Plant J* 2018, *94*, 372–392.
- [46] J.-E. Bassard, L. Richert, J. Geerinck, H. Renault, F. Duval, P. Ullmann, M. Schmitt, E. Meyer, J. Mutterer, W. Boerjan, G. D. Jaeger, Y. Mely, A. Goossens, D. Werck-Reichhart, *Plant Cell* 2012, *24*, 4465–4482.
- [47] H. A. Maeda, *Front Plant Sci* 2019, *10*, 881.

- [48] P. L. Cruz, I. Carqueijeiro, K. Koudounas, D. P. Bomzan, E. A. Stander, C. Abdallah, N. Kulagina, A. Oudin, A. Lanoue, N. Giglioli-Guivarc'h, D. A. Nagegowda, N. Papon, S. Besseau, M. Clastre, V. Courdavault, *Protoplasma* 2023, 260, 607–624.
- [49] N. S. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. I. Stuart, R. J. Owens, *Nucleic Acids Res* 2007, 35, e45–e45.
- [50] Z. Ban, H. Qin, A. J. Mitchell, B. Liu, F. Zhang, J. K. Weng, R. A. Dixon, G. Wang, *Proc National Acad Sci* 2018, 115, E5223–E5232.
- [51] X. F. Chang, R. Chandra, T. Berleth, R. P. Beatson, *J Agr Food Chem* 2008, 56, 6825–6834.
- [52] R. C. Edgar, *Biorxiv* 2021, 2021.06.20.449169.
- [53] J. Trifinopoulos, L.-T. Nguyen, A. von Haeseler, B. Q. Minh, *Nucleic Acids Res* 2016, 44, W232–W235.
- [54] I. Letunic, P. Bork, *Nucleic Acids Res* 2021, 49, gkab301-.
- [55] S. S. Sahu, C. D. Loaiza, R. Kaundal, *Aob Plants* 2019, 12, plz068.
- [56] X. Robert, P. Gouet, *Nucleic Acids Res* 2014, 42, W320–W324.

Chapter 5. Conclusions and Future Perspectives

5.1. The Structural and Mechanistic Basis of Atypical CAD-Catalysed Reductions in MIA Biosynthesis

Members of the CAD family of enzymes catalyse several atypical reduction reactions in MIA biosynthesis such as the 1,2-reduction of the iminium moiety of the substrate strictosidine aglycone ^[1]. In addition, findings in Chapter 2 demonstrate that the CAD *CrDPAS* catalyses the 1,4-iminium reduction of the substrates precondylocarpine acetate and dehydrosecodine, and the 1,4-reduction of an α,β -unsaturated aldehyde vallesiachotamine - the first reports of these chemistries to be catalysed by an ADH enzyme.

To uncover the mechanistic basis of these reductions, we elucidated the structures of the 1,4-iminium reducing *CrDPAS* and *TiDPAS2* and compared their active sites with other closely related CADs known to act in MIA biosynthesis including the 1,2-iminium reducing *CrGS* and aldehyde reducing *Cr8HGO*. Our study identified key differences known to be involved in catalysis. Namely, *CrDPAS* and *TiDPAS2* had atypical residues in positions that typically coordinate the catalytic zinc, resulting in the loss of this ion, whilst *CrGS* had residues that resulted in atypical binding and positioning of the cofactor. The findings from structural analyses in conjunction with mutational studies of *CrDPAS* and *CrGS* enable us to propose the mechanistic basis of these reduction reactions. Current engineering efforts to improve the substrate promiscuity of enzymatic imine reductases have had limited success ^[2]. It is therefore hoped that the discovery of ADHs capable of catalysing the reduction of an iminium moiety ^[3-5] may aid future bioengineering efforts.

5.2. Phylogenetic Study Reveals the Neofunctionalisation of CADs as Drivers of MIA Chemical Diversity

Building on our understanding of CAD sequence motifs important in catalysing atypical reductions in MIA biosynthesis, Chapter 3 explores the phylogenetic relationship of this gene family in *C. roseus*. Genomic analysis reveals that the CAD gene family has undergone more recent expansion in *C. roseus*, with several instances of physical clusters of these genes in the *C. roseus* genome that likely arose through several tandem duplication events. In addition, phylogenetic and sequence analyses suggest that each class of atypical

reduction chemistry had a monophyletic branch of corresponding CAD genes and a sister clade of aldehyde-reducing CADs. These findings suggest that each atypical class of this gene family emerged once from a typical-CAD ancestral genes, before being neofunctionalised to catalyse either the 1,2- or 1,4-reduction of an iminium moiety.

Furthermore, we expanded our analysis across the order of Gentianales to reveal that CADs with a similar pattern of atypical residues required for the 1,2-reduction of an iminium moiety were exclusively found in species known to produce MIAs. CADs that catalyse this atypical reduction often act on different structural isomers of the early biosynthetic intermediate strictosidine aglycone to form a diverse range of MIA scaffolds^[1, 6, 7]. The role of atypical CADs in the generation of chemical diversity therefore supports the neofunctionalisation and maintenance of these genes as key evolutionary drivers of MIA biosynthesis.

5.3. Protein-Protein Interactions in MIA Biosynthesis and Beyond

The large number of biosynthetic enzymes and the presence of unstable intermediates have led to speculation regarding the role of protein-protein interactions in MIA biosynthesis. Several protein complexes between MIA biosynthetic enzymes have been previously reported^[7, 8], however, a more comprehensive study was required to understand the extent of these interactions. To this end, Chapter 4 reports the pairwise testing of 17 biosynthetic enzymes involved in vinblastine biosynthesis which are colocalised in the same cell-type in *C. roseus* using a split-luciferase assay^[9]. These findings revealed an extensive network of protein-protein interactions between these enzymes, suggesting their potential metabolic role in MIA biosynthesis.

We focussed our study on characterising the interaction between *CrDPAS* and the downstream cyclase enzymes *CrTS* and *CrCS* due to the potential metabolic role of this complex on the flux of the unstable intermediary compound, dehydrosecodine^[8]. However, whilst interactions between these proteins was observed using a split-luciferase assay, it was not detected using other *in vitro* biophysical methods such as AP-MS or DSF, highlighting the weak and/or dynamic nature of these interactions. To probe the structural basis of this interaction, we identified a single surface residue on the cyclase enzymes

through mutational studies which selectively engineered the interaction between these proteins. These results form the basis for future engineering efforts to understand the metabolic effect of protein-protein interaction between these MIA biosynthetic enzymes.

In addition, we present findings that support the formation of inter-pathway protein complexes between enzymes that act within specialised metabolism (MIA biosynthesis) and primary metabolism (phenylpropanoid biosynthesis). Phenylpropanoid biosynthesis is an ancient pathway ^[10] comprising of enzymes that in many cases, have subsequently undergone neofunctionalisation and been co-opted to act in specialised metabolism ^[11], including MIA biosynthesis ^[12]. The observation of these inter-pathway interactions may be a result of conservation from their ancestral primary metabolic predecessors. Furthermore, we postulate that these interactions have a functional role in reducing lignin biosynthesis in *C. roseus* leaves. These preliminary findings suggest that inter-pathway protein-protein interactions may act as a mechanism to modify the metabolic output of a cell, enabling cross-talk between primary and specialised metabolism within the plant.

5.4. Future Directions: Biochemical and Biophysical Insights into MIA Biosynthesis and Beyond

MIAs are a structurally diverse group of PNPs of great biological and pharmaceutical importance, including the anti-cancer compound vinblastine from *C. roseus* ^[13]. This doctoral thesis explores the biochemical and biophysical characterisation of enzymes participating in MIA biosynthesis, with the aim of further unravelling the intricacies of this pathway.

Work in this thesis explores the remarkable divergence of reduction chemistries catalysed by members of the CAD enzyme subfamily in MIA biosynthesis. These atypical reduction reactions, which are unreported outside the context of MIA biosynthesis, were found to have arisen from deviations in otherwise highly conserved residues within the enzyme's active site. These alterations facilitated their neofunctionalisation to perform either the 1,2-reduction of an iminium moiety, or the 1,4-reduction of either an iminium moiety or an α,β -unsaturated aldehyde. This discovery highlights the remarkable catalytic plasticity of

these enzymes and underscores their potential utility in bioengineering efforts of enzymatic catalysts of challenging imine reductions [2].

Furthermore, comprehensive phylogenetic and genomic analysis of the CAD gene family revealed their role as key evolutionary drivers of MIA chemical diversity in Gentianales. The identification of sequence motifs important in catalysis enabled us to identify the emergence of atypical class of CADs after the divergence of Gentianales and correlate our findings with the presence of MIAs in these species. Additionally, this work identified uncharacterised CADs with the potential to catalyse other atypical reductions, offering promising leads for the discovery of additional MIA biosynthetic genes.

Lastly, this thesis delves into characterising the biophysical interactions among biosynthetic proteins involved in vinblastine biosynthesis within *C. roseus*. In this context, the structural basis of the interaction between *CrDPAS* and the downstream cyclase enzymes *CrTS* and *CrCS* was elucidated. This finding suggests enzyme-enzyme interactions may exert an effect on the metabolic flux of their biosynthetic intermediary molecule, although further *in vivo* and *in vitro* experimentation is required to test this hypothesis. Additionally, we reveal inter-pathway interactions between enzymes within MIA and phenylpropanoid biosynthesis and provide evidence of their potential metabolic role in regulating lignin biosynthesis in *C. roseus*. The full extent of inter-pathway interactions between MIA biosynthetic enzymes and proteins acting in other pathways and their potential metabolic roles requires further exploration. However, these initial results provide insight into the evolution of these protein complexes and reveal the potential organisational principles governing multiple metabolic pathways within the cell.

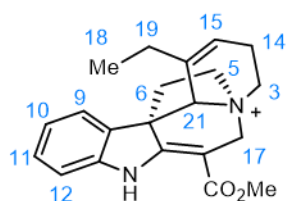
5.5. References

- [1] A. Stavriniades, E. C. Tatsis, L. Caputi, E. Foureau, C. E. M. Stevenson, D. M. Lawson, V. Courdavault, S. E. O'Connor, *Nat Commun* 2016, 7, 12116.
- [2] P. Stockinger, S. Roth, M. Müller, J. Pleiss, *Chembiochem* 2020, 21, 2689–2695.
- [3] S.-J. Kim, M.-R. Kim, D. L. Bedgar, S. G. A. Moinuddin, C. L. Cardenas, L. B. Davin, C. Kang, N. G. Lewis, *Proc National Acad Sci* 2004, 101, 1455–1460.
- [4] B. Youn, R. Camacho, S. G. A. Moinuddin, C. Lee, L. B. Davin, N. G. Lewis, C. Kang, *Org Biomol Chem* 2006, 4, 1687–1697.
- [5] C. Langley, E. Tatsis, B. Hong, Y. Nakamura, C. Paetz, C. E. M. Stevenson, J. Basquin, D. M. Lawson, L. Caputi, S. E. O'Connor, *Angew. Chem. Int. Ed.* 2022, 61, e202210934.
- [6] E. C. Tatsis, I. Carqueijeiro, T. D. D. Bernonville, J. Franke, T.-T. T. Dang, A. Oudin, A. Lanoue, F. Lafontaine, A. K. Stavriniades, M. Clastre, V. Courdavault, S. E. O'Connor, *Nat Commun* 2017, 8, 316.
- [7] A. Stavriniades, E. C. Tatsis, E. Foureau, L. Caputi, F. Kellner, V. Courdavault, S. E. O'Connor, *Chem Biol* 2015, 22, 336–41.
- [8] L. Caputi, J. Franke, S. C. Farrow, K. Chung, R. M. E. Payne, T.-D. Nguyen, T.-T. T. Dang, I. S. T. Carqueijeiro, K. Koudounas, T. D. de Bernonville, B. Ameyaw, D. M. Jones, I. J. C. Vieira, V. Courdavault, S. E. O'Connor, *Science* 2018, 360, 1235–1239.
- [9] C. Li, J. C. Wood, A. H. Vu, J. P. Hamilton, C. E. R. Lopez, R. M. E. Payne, D. A. S. Guerrero, K. Gase, K. Yamamoto, B. Vaillancourt, L. Caputi, S. E. O'Connor, C. R. Buell, *Nat Chem Biol* 2023, 1–11.
- [10] S. de Vries, J. M. R. Fürst-Jansen, I. Irisarri, A. D. Ashok, T. Ischebeck, K. Feussner, I. N. Abreu, M. Petersen, I. Feussner, J. de Vries, *The Plant Journal* 2021, DOI 10.1111/tpj.15387.
- [11] H. A. Maeda, *Front Plant Sci* 2019, 10, 881.
- [12] P. L. Cruz, I. Carqueijeiro, K. Koudounas, D. P. Bomzan, E. A. Stander, C. Abdallah, N. Kulagina, A. Oudin, A. Lanoue, N. Giglioli-Guivarc'h, D. A. Nagegowda, N. Papon, S. Besseau, M. Clastre, V. Courdavault, *Protoplasma* 2023, 260, 607–624.
- [13] S. E. O'Connor, J. J. Maresh, *Nat Prod Rep* 2006, 23, 532.

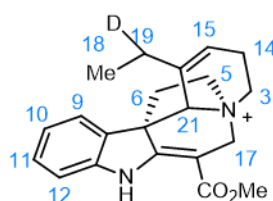
Appendices

Appendix I. NMR Characterisation of *d*-angryline

Table 12. Comparison of ^1H NMR data between angryline and *d*-angryline. Multiplicity abbreviations: s, singlet; d, doublet; m, multiplet; bs, broad singlet; bd, broad doublet; dd, doublet of doublets; bdd, broad doublet of doublets; ddd, doublet of doublet of doublets.



angryline
 ^1H 600 MHz,
 MeOH- d_3 , 25 °C



d-angryline
 ^1H 400MHz,
 MeOH- d_3 , 25 °C

Position	angryline ^[6] ^1H (600 MHz, MeOH- d_3)	<i>d</i> -angryline ^1H (400 MHz, MeOH- d_3)
3	4.05 (m, 1H) 3.76 (m, 1H)	4.04 (m, 1H) 3.82 (m, 1H)
5	4.06 (m, 1H) 3.88 (m, 1H)	4.07 (m, 1H) 3.85 (m, 1H)
6	2.66 (m, 1H) 2.51 (m, 1H)	2.69 (m, 1H) 2.54 (m, 1H)
9	7.49 (d, 7.5, 1H)	7.50 (d, 7.5, 1H)
10	7.06 (dd, 7.6, 7.6, 1H)	7.07 (td, 7.5, 0.9, 1H).
11	7.33 (dd, 7.8, 7.8, 1H)	7.35 (td, 7.8, 1.1, 1H)
12	7.10 (d, 7.6, 1H)	7.12 (d, 7.9, 1H),
14	2.82 (m, 1H) 2.66 (m, 1H)	2.85 (m, 1H) 2.69 (m, 1H)
15	5.66 (s, 1H)	5.68 (s, 1H)
17	4.41 (dd, 15.0, 2.0, 1H) 4.06 (m, 1H)	4.44 (dd, 15.0, 2.2, 1H) 4.07 (m, 1H)
18	0.78 (t, 7.3, 3H)	0.78 (t, 7.4, 3H).
19	1.43 (m, 2H)	1.46 (m, 1.36H)
21	5.11 (s, 1H)	5.12 (s, 1H)
CO ₂ Me	3.78 (s, 3H)	3.81 (s, 3H)

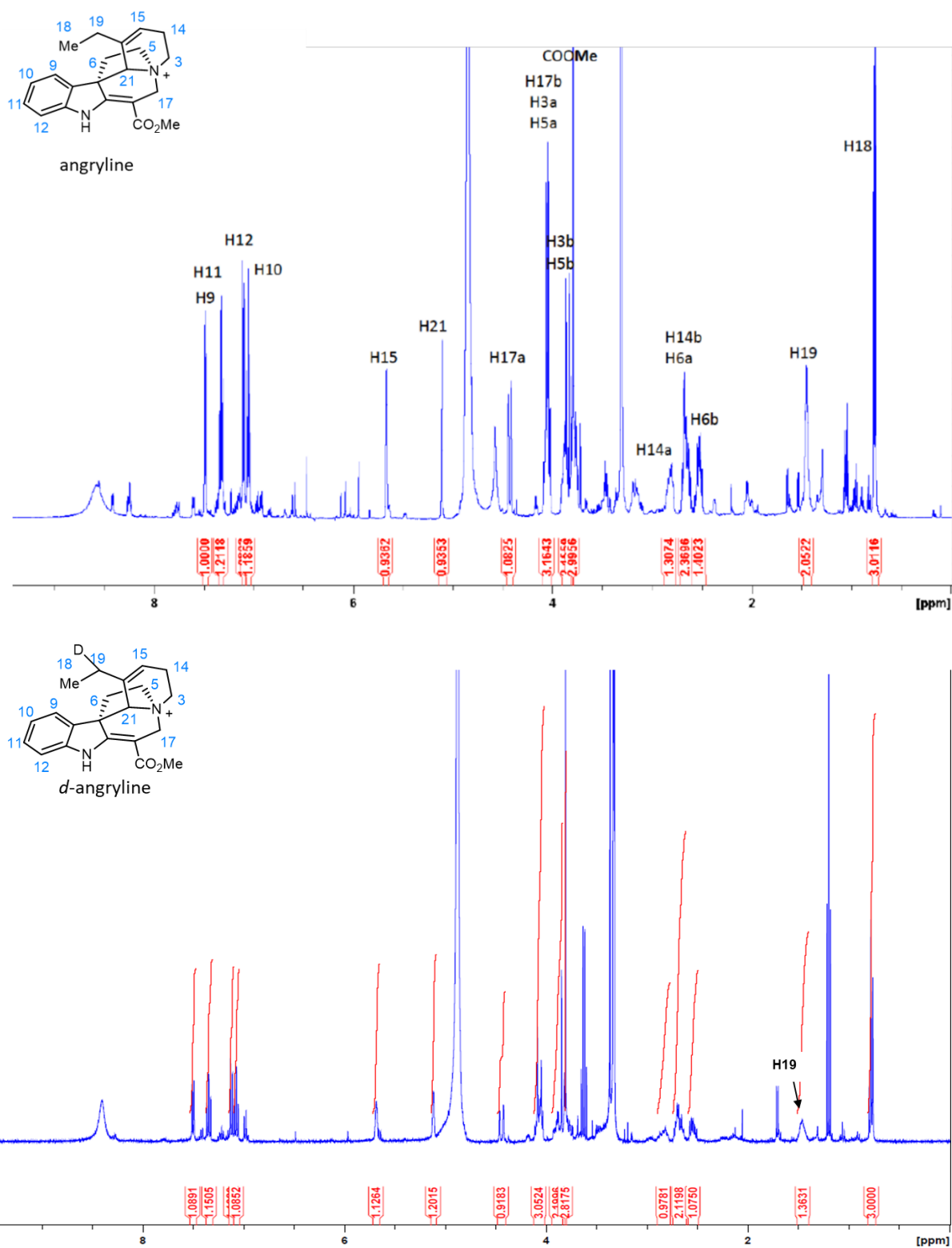
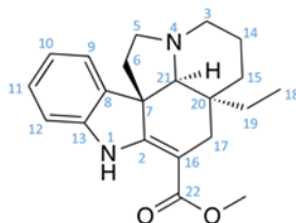


Figure 72. Comparison of ^1H spectra of angryline and *d*-angryline. Loss of signal shown at H19, indicating deuterium incorporation.

Appendix II. NMR Characterisation of (-)-vincadifformine

Table 13. ^1H NMR data for (-)-vincadifformine in $\text{MeOH-}d_3$. Multiplicity abbreviations: s, singlet; d, doublet; m, multiplet; bs, broad singlet; bd, broad doublet; dd, doublet of doublets; bdd, broad doublet of doublets; ddd, doublet of doublet of doublets.



(-)-vincadifformine
 ^1H 700 MHz, $\text{MeOH-}d_3$, 25 °C

pos.	δ_{H}	mult.	J_{HH}	δ_{C}
1	9.16	bs	-	-
2	-	-	-	169.1
3a	2.45	ddd	10.9/10.9/3.0	51.4
3b	3.11	m	-	51.4
4	-	-	-	-
5a	2.59	ddd	11.3/8.5/4.8	52.5
5b	2.91	m	-	52.5
6a	1.64	ddd	11.3/4.8/1.2	46.6
6b	1.99	ddd	11.3/11.3/6.6	46.6
7	-	-	-	57.0
8	-	-	-	138.9
9	7.20	bd	7.2	121.8
10	6.84	bdd	7.7/7.2	121.5
11	7.09	ddd	7.7/7.7/0.9	128.5
12	6.89	bd	7.7	110.5
13	-	-	-	144.8
14a	1.54	m	-	22.8
14b	1.84	m	-	22.8
15a	1.29	ddd	13.1/13.1/4.9	33.9
15b	1.80	m	-	33.9
16	-	-	-	92.8
17a	2.28	dd	15.2/1.8	26.9
17b	2.70	d	15.2	26.9
18	0.57	t	7.3	7.3
19a	0.95	m	-	30.4
19b	0.63	m	-	30.4
20	-	-	-	39.0
21	2.51	bs	-	73.8
22	-	-	-	170.2
OMe	3.74	s	-	51.3

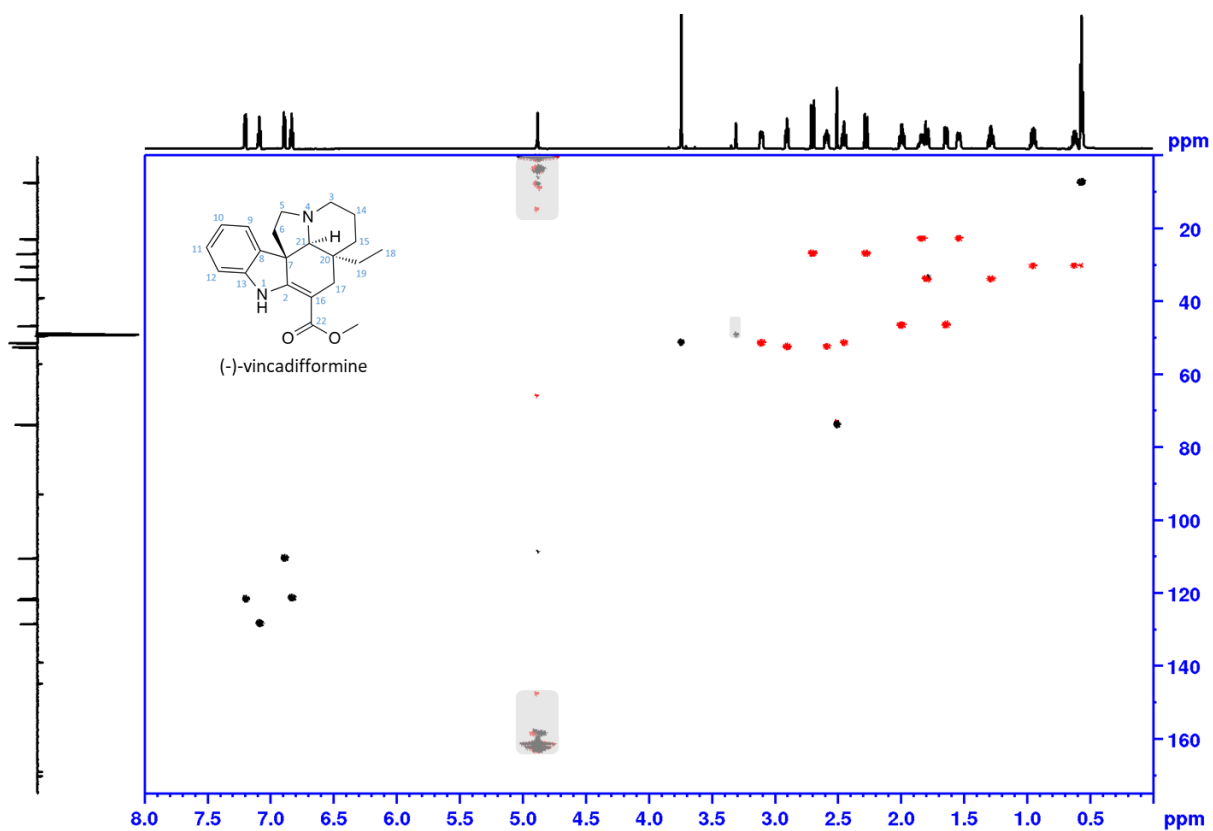


Figure 73. ¹H NMR data for *m/z* 339, (-)-vincadifformine (standard). Phase sensitive HSQC, full range in MeOH-*d*₃. Shaded areas mark impurity and solvent, red: CH₂, black: CH, CH₃. NMR data of (-)-vincadifformine in chloroform-*d* has been previously reported [61, 62].

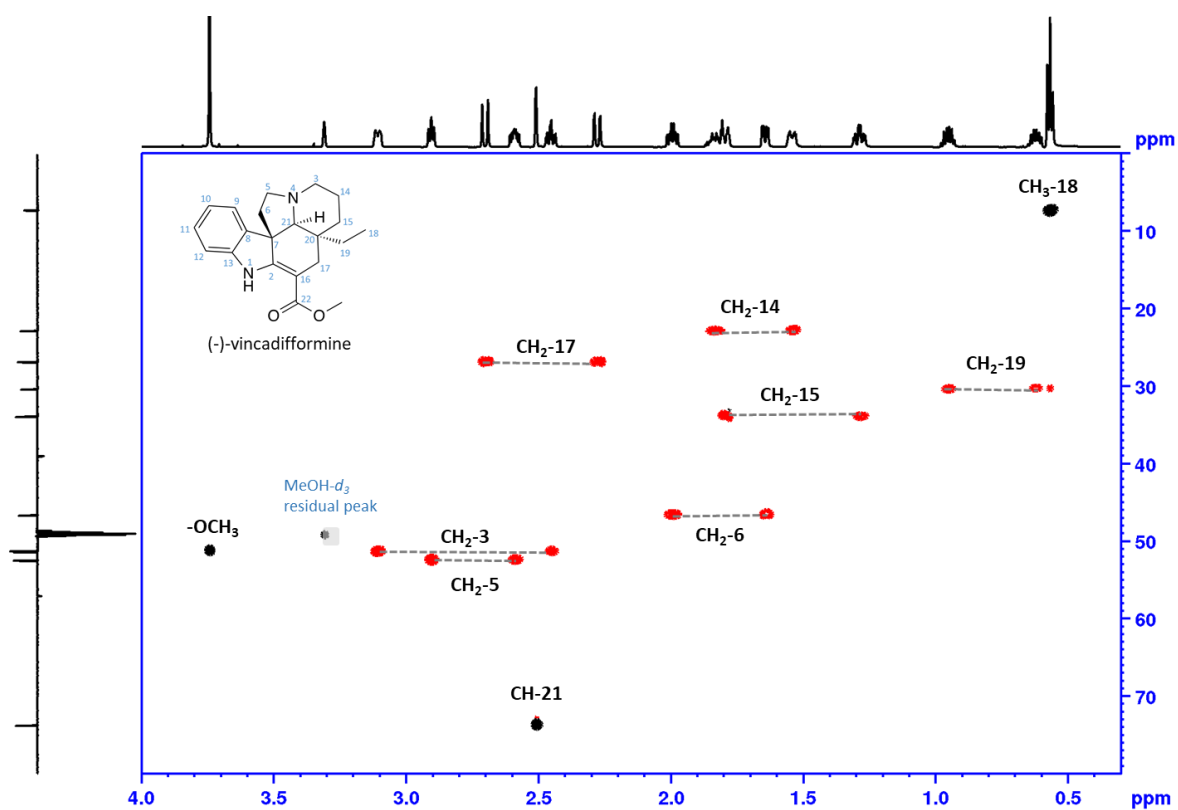
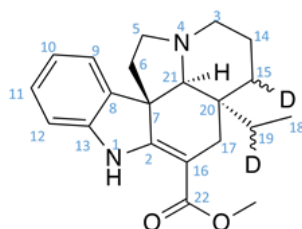


Figure 74. ^1H NMR data for m/z 339, (-)-vincadifformine (standard). Phase sensitive HSQC, aliphatic range in $\text{MeOH-}d_3$. Shaded areas mark impurity and solvent, red: CH_2 , black: CH , CH_3

Appendix III. NMR Characterisation of *d*₂-(+)-vincadifformine

Table 14. ¹H NMR data for *d*₂-(±)-vincadifformine in MeOH-*d*₃. Multiplicity abbreviations: s, singlet; d, doublet; m, multiplet; bs, broad singlet; bd, broad doublet; dd, doublet of doublets; bdd, broad doublet of doublets; ddd, doublet of doublet of doublets.



*d*₂-(±)-vincadifformine
¹H 700 MHz, MeOH-*d*₃,
 25 °C

pos.	δ_{H}	mult.	J_{HH}	δ_{C}
1	9.18	bs	-	-
2	-	-	-	169.1
3a	2.51	m**	-	51.3
3b	3.14	m	-	51.3
4	-	-	-	-
5a	2.65	ddd	11.3/9.3/4.8	52.5
5b	2.95	m	-	52.5
6a	1.69	ddd	11.3/4.8/1.2	46.6
6b	2.03	ddd	11.3/11.3/6.4	46.6
7	-	-	-	56.9
8	-	-	-	138.8
9	7.23	bd	7.3	121.8
10	6.85	bdd	7.7/7.3	121.6
11	7.1	ddd	7.7/7.7/0.8	128.5
12	6.91	bd	7.7	110.6
13	-	-	-	144.8
14a	1.56	m	-	22.4
14b	1.85	m	-	22.4
15a*	1.30	m**	-	33.2
15b*	1.79	m	-	33.2
16	-	-	-	92.8
17a	2.30	dd	15.1	26.9
17b	2.69	dd	15.2/2.9	26.9
18	0.57	d	7.3	7.1
19a*	0.94	m**	-	30.1
19b*	0.63	m**	-	30.1
20	-	-	-	38.7
21	2.59	bs	-	73.6
22	-	-	-	170.2
OMe	3.75	s	-	51.4
*as CH signal				
** overlapped signals <i>J</i> unresolved				

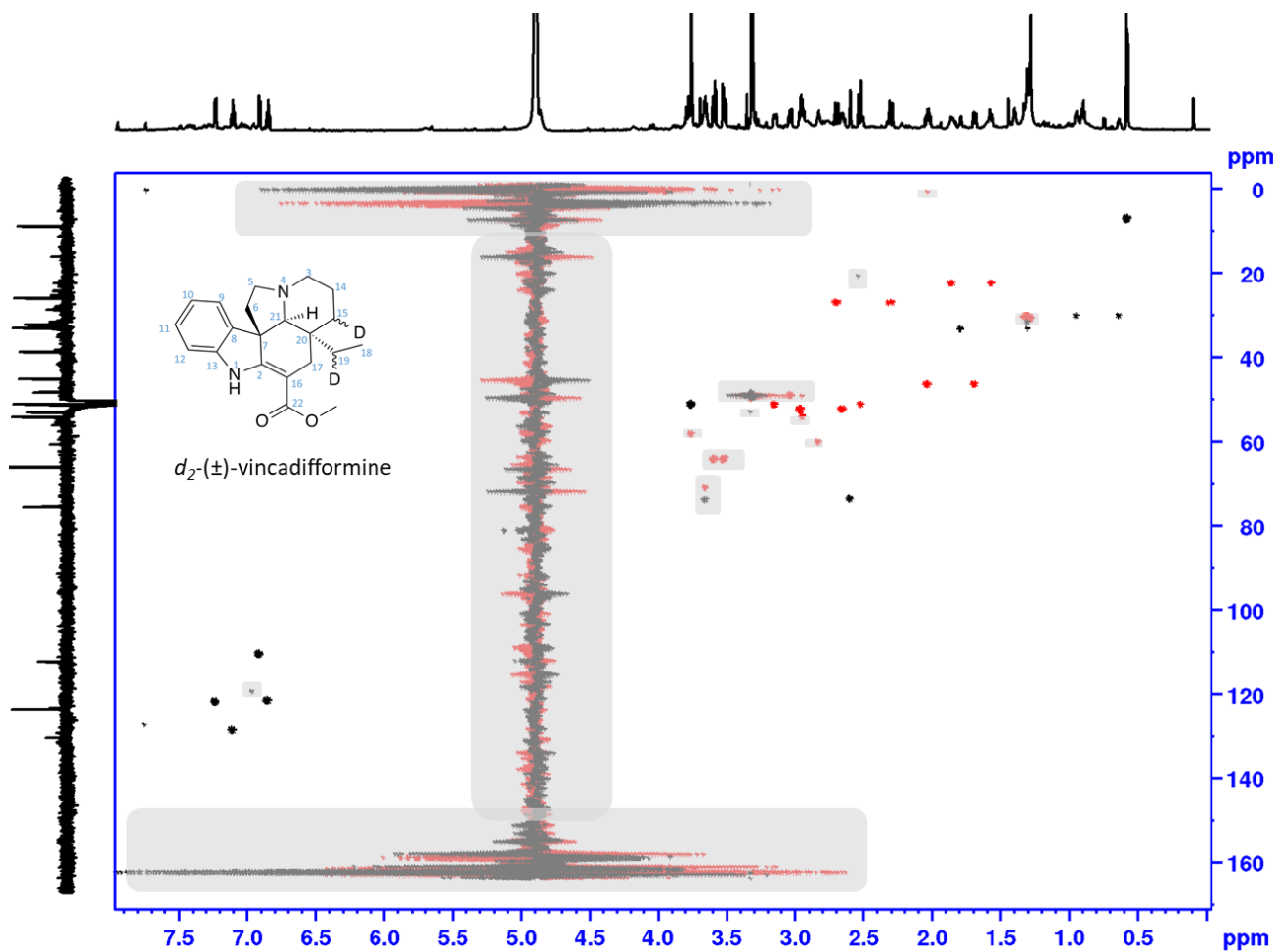


Figure 75. Phase sensitive HSQC NMR data for m/z 341, d_2 -(±)-vincadifformine full range in $\text{MeOH-}d_3$. Shaded areas mark impurity and solvent, red: CH_2 , black: CH , CH_3

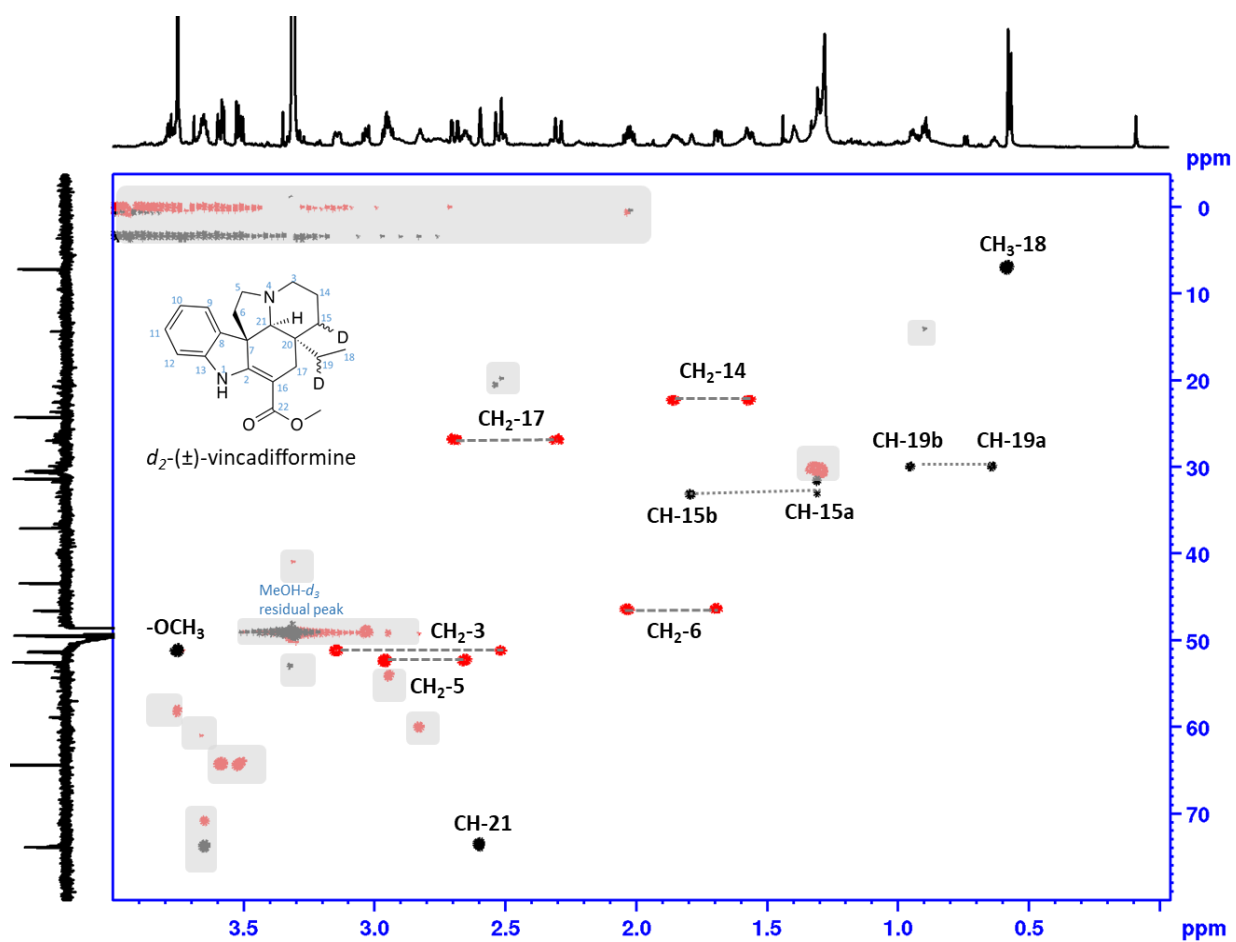


Figure 76. Phase sensitive HSQC NMR data for m/z 341, d_2 -(\pm)-vincadifformine, aliphatic range in MeOH- d_3 . Shaded areas mark impurity and solvent, red: CH_2 , black: CH , CH_3

Appendix IV. NMR Characterisation of 19,20-dihydrovallesiachotamine

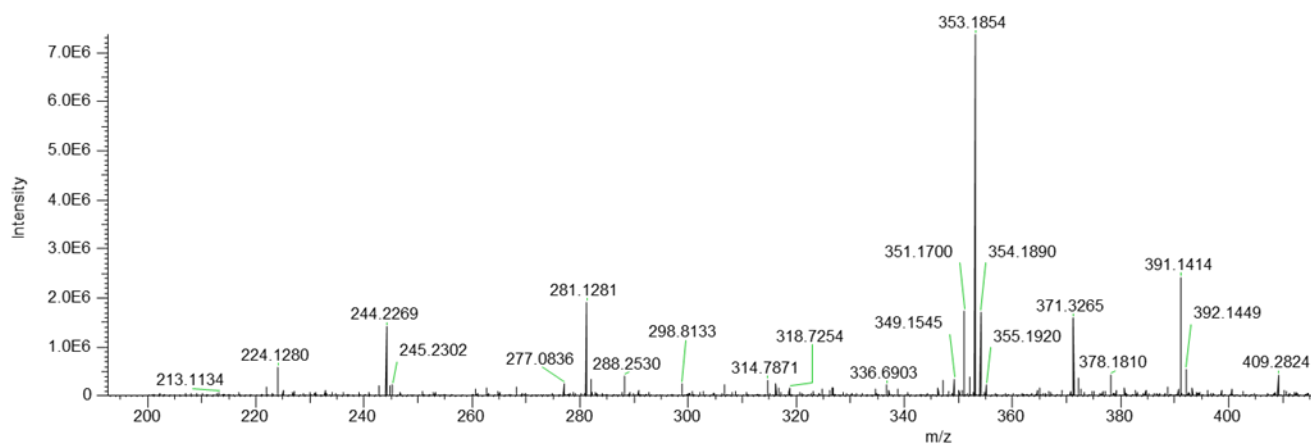
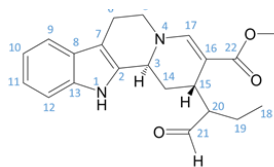


Figure 77. MS/MS spectra of 19,20-dihydrovallesiachotamine. Formula: $C_{21}H_{24}N_2O_3$; observed mass: 353.1854; theoretical mass: 353.1860; error 1.6988 p.p.m.

Table 15. ^1H NMR spectra for 19,20-dihydrovallesiachotamine in $\text{MeOH-}d_3$.



19,20-dihydrovallesiachotamine
 ^1H 700 MHz, $\text{MeOH-}d_3$, 25 °C

pos.	δ_{H}	mult.	J_{HH}	δ_{C}
1	10.4	<i>bs</i>	-	-
2	-	-	-	133.9
3 α	4.55	<i>bd</i>	12.0	49.9
4	-	-	-	-
5 α	3.61	<i>ddd</i>	13.0/12.4/3.9	52.2
5 β	3.81	<i>dd</i>	13.0/5.4	52.2
6 α	2.76	<i>m</i>	-	22.9
6 β	2.86	<i>m</i>	-	22.9
7	-	-	-	108.2
8	-	-	-	127.9
9	7.39	<i>bd</i>	7.8	118.6
10	6.98	<i>dd</i>	7.8/7.2	119.8
11	7.06	<i>dd</i>	8.0/7.2	122.3
12	7.28	<i>bd</i>	8.0	112.0
13	-	-	-	138.3
14 α	2.50	<i>ddd</i>	13.8/4.0/2.0	32.5
14 β	1.60	<i>ddd</i>	13.8/12.0/5.0	32.5
15 β	3.18	<i>ddd</i>	7.5/5.0/2.0	31.6
16	-	-	-	94.1
17	7.74	<i>s</i>	-	148.9
18	0.93	<i>t</i>	7.4	12.5
19a	1.82	<i>ddt</i>	14.0/9.6/7.4	20.8
19b	1.54	<i>m</i>	-	20.8
20	2.42	<i>m</i>	-	60.2
21	9.72	<i>d</i>	3.4	208.0
22	-	-	-	170.9
OMe	3.67	<i>s</i>	-	51.1

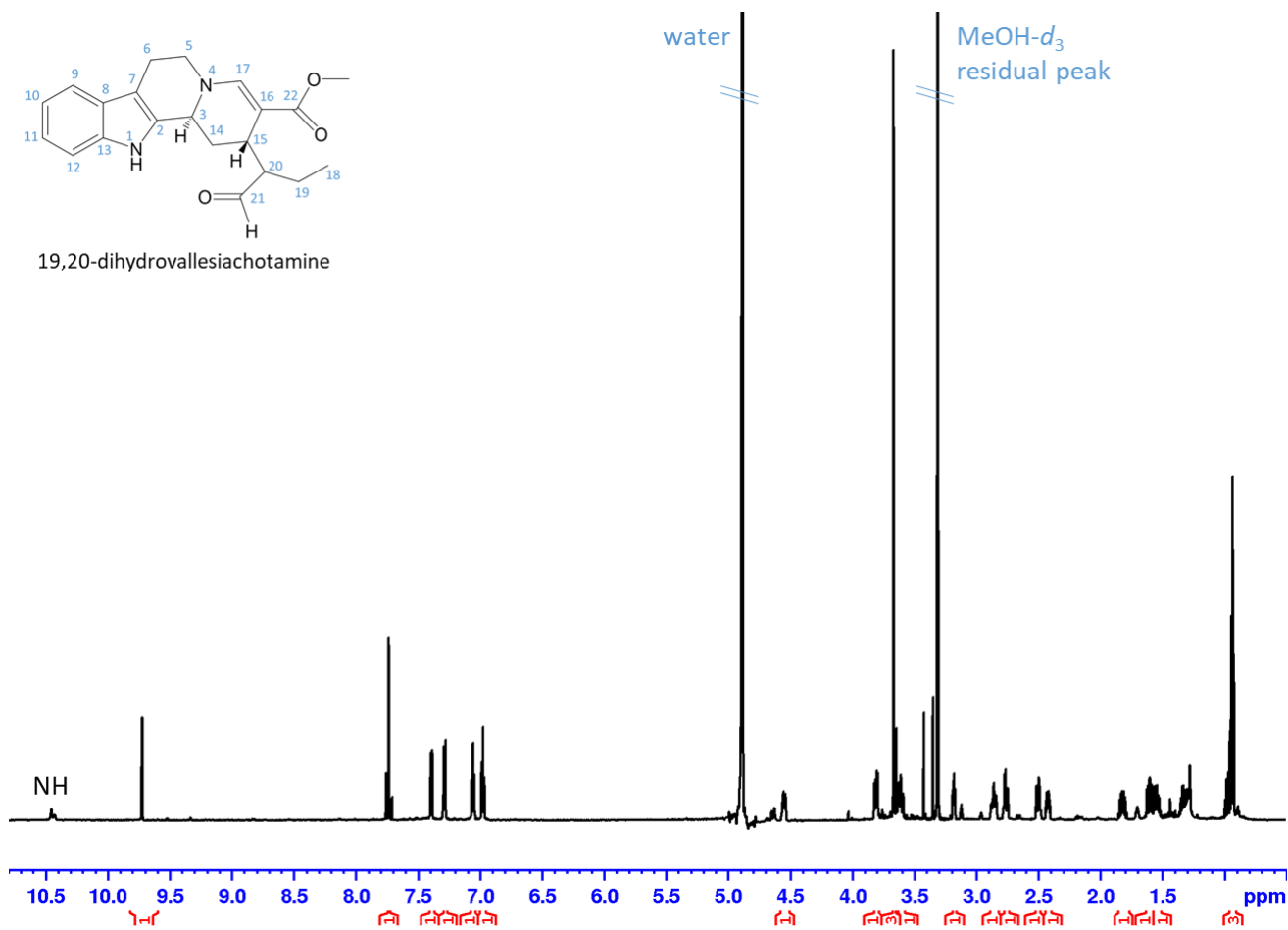


Figure 78. ^1H NMR data of 19,20-dihydrovallesiachotamine with water suppression, full range in $\text{MeOH-}d_3$

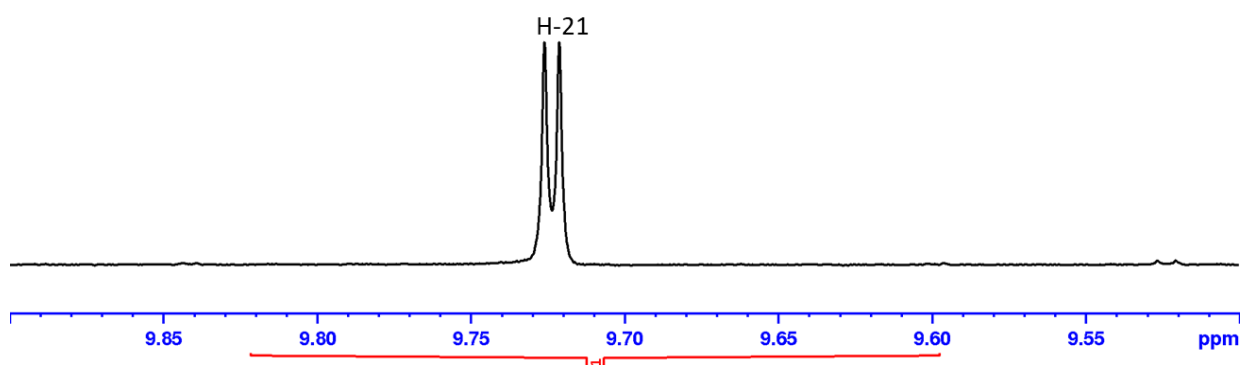
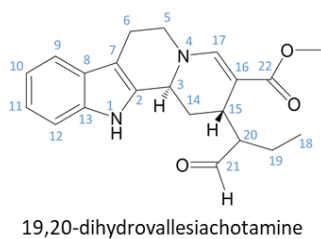


Figure 79. ^1H NMR data of 19,20-dihydrovallesiachotamine with water suppression, aldehyde range in $\text{MeOH-}d_3$

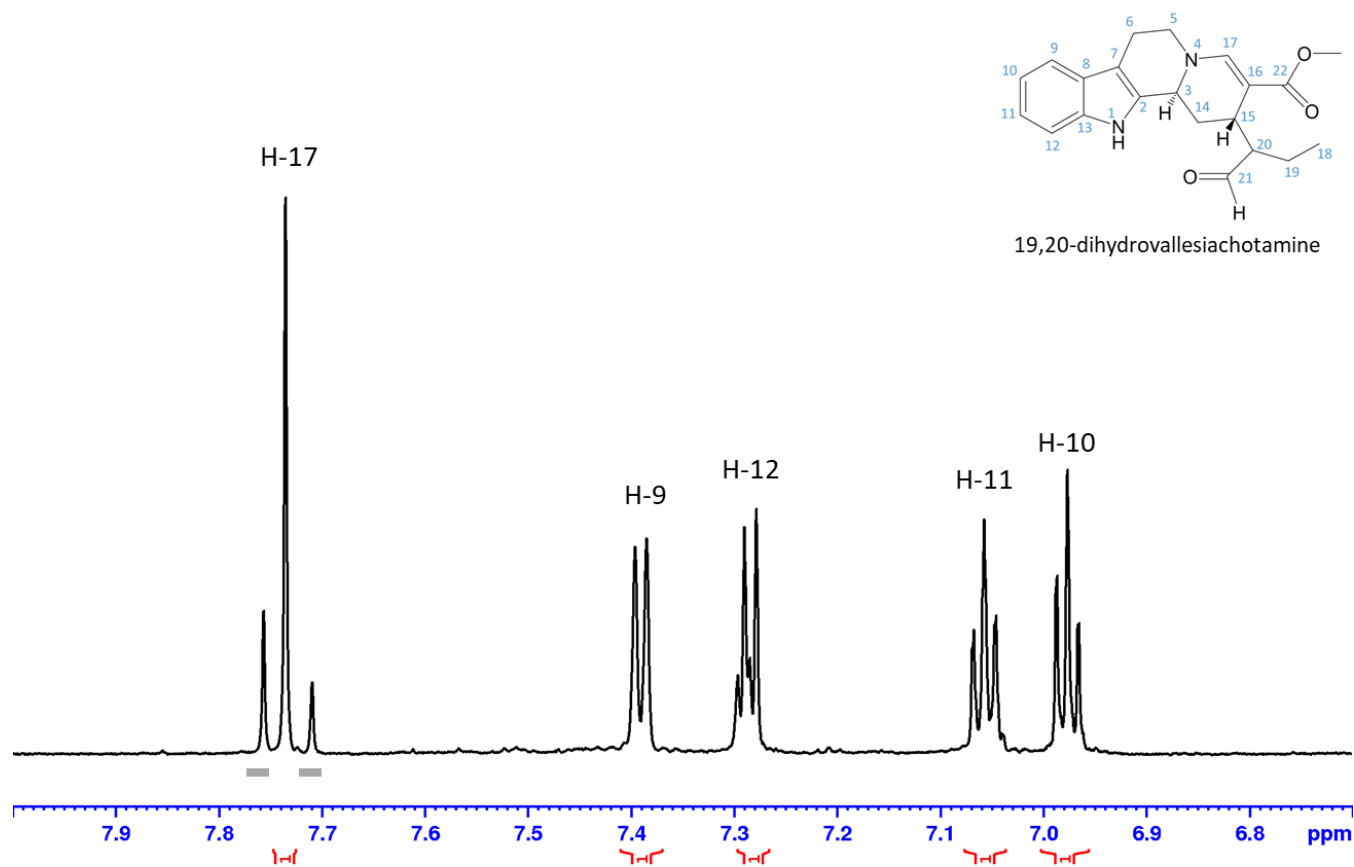


Figure 80. ¹H NMR data of 19,20-dihydrovallesiachotamine with water suppression, aromatic range in MeOH-*d*₃. Grey bars indicate impurities.

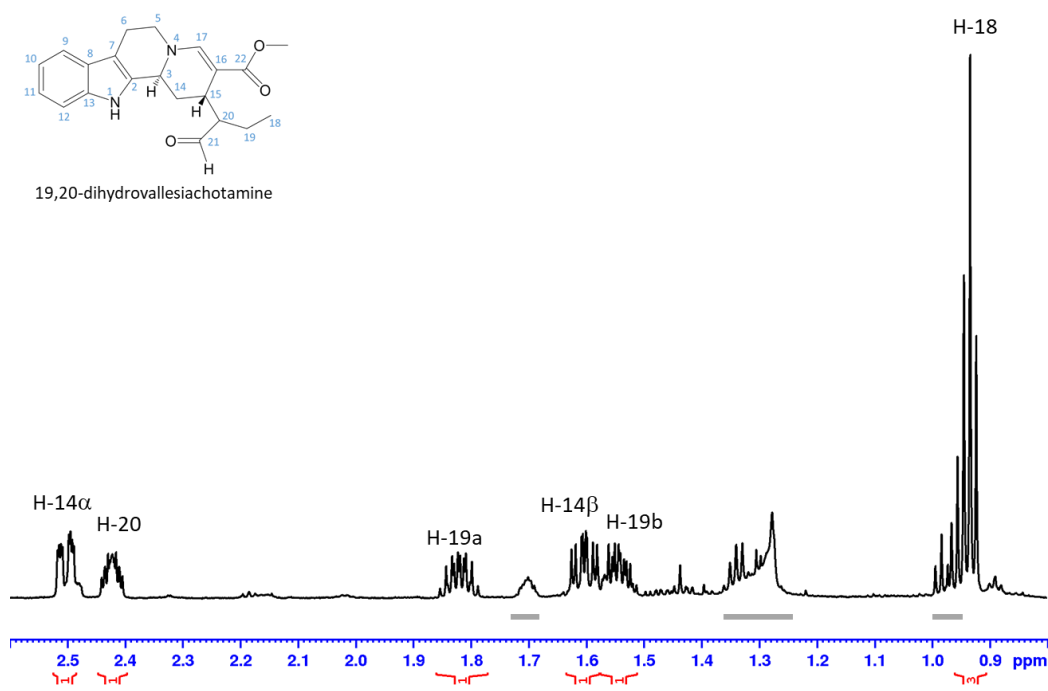
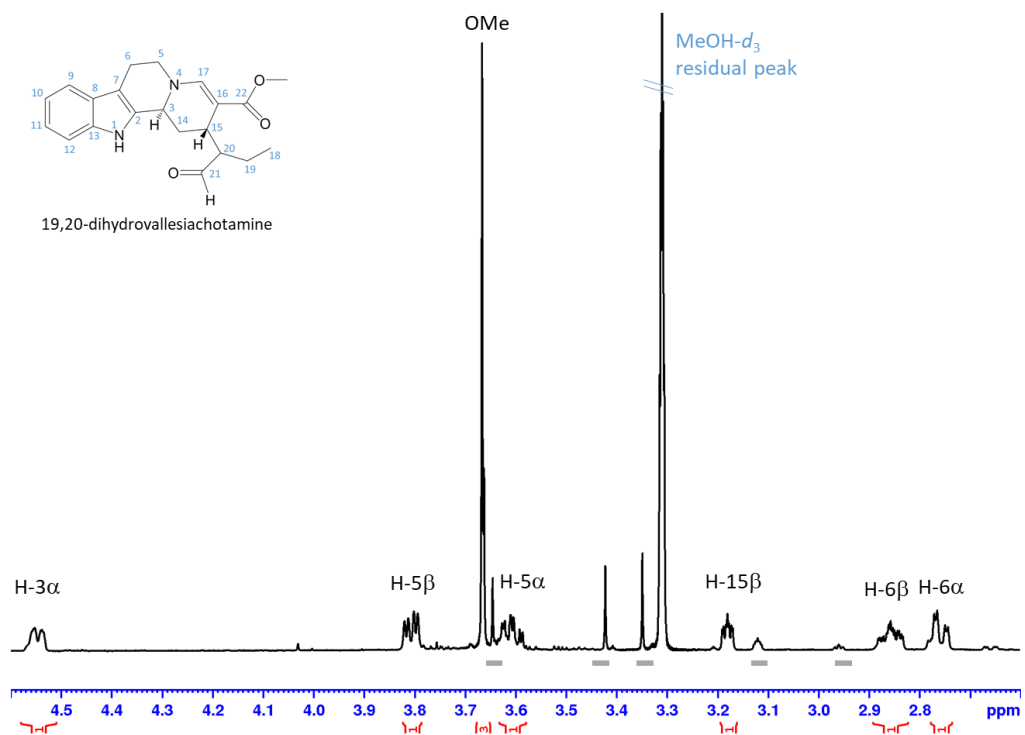


Figure 81. ¹H NMR data of 19,20-dihydrovallesiachotamine with water suppression, aliphatic range in MeOH-*d*₃. Grey bars indicate impurities.

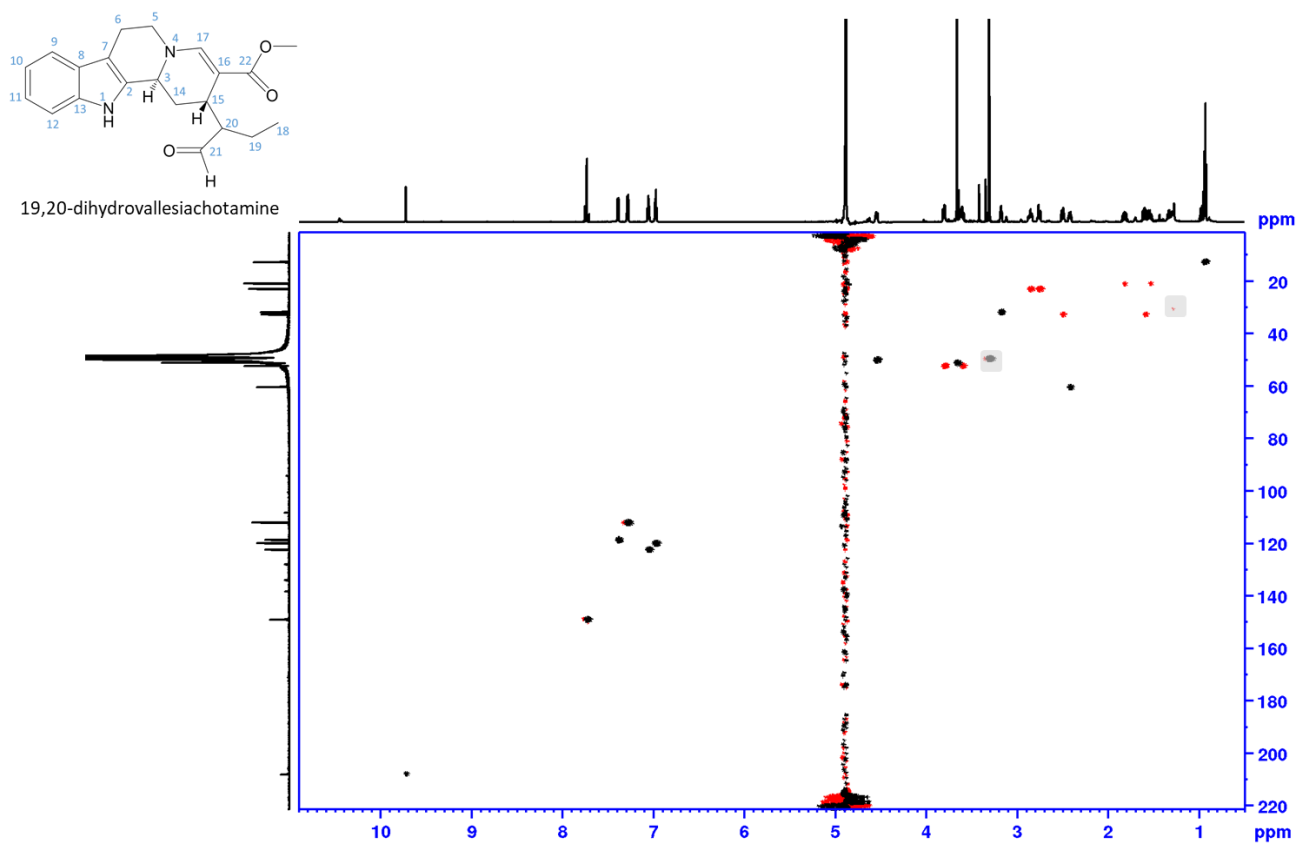


Figure 82. NMR data of 19,20-dihydrovallesiachotamine, phase sensitive HSQC, full range in MeOH- d_3 . Shaded areas mark impurity and solvent, red: CH₂, black: CH, CH₃

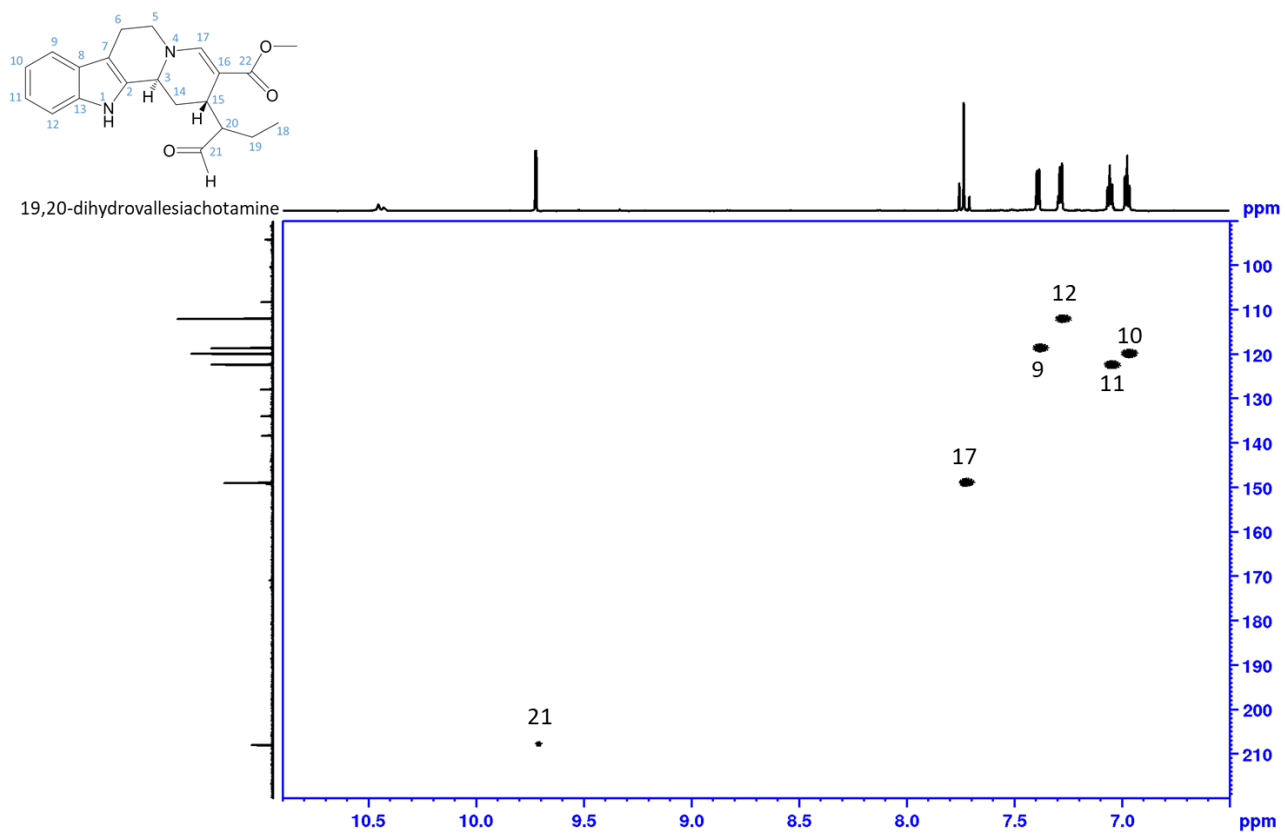


Figure 83. NMR data of 19,20-dihydrovallesiachotamine, phase sensitive HSQC, aldehyde and aromatic range in MeOH-*d*₃. Shaded areas mark impurity and solvent, red: CH₂, black: CH

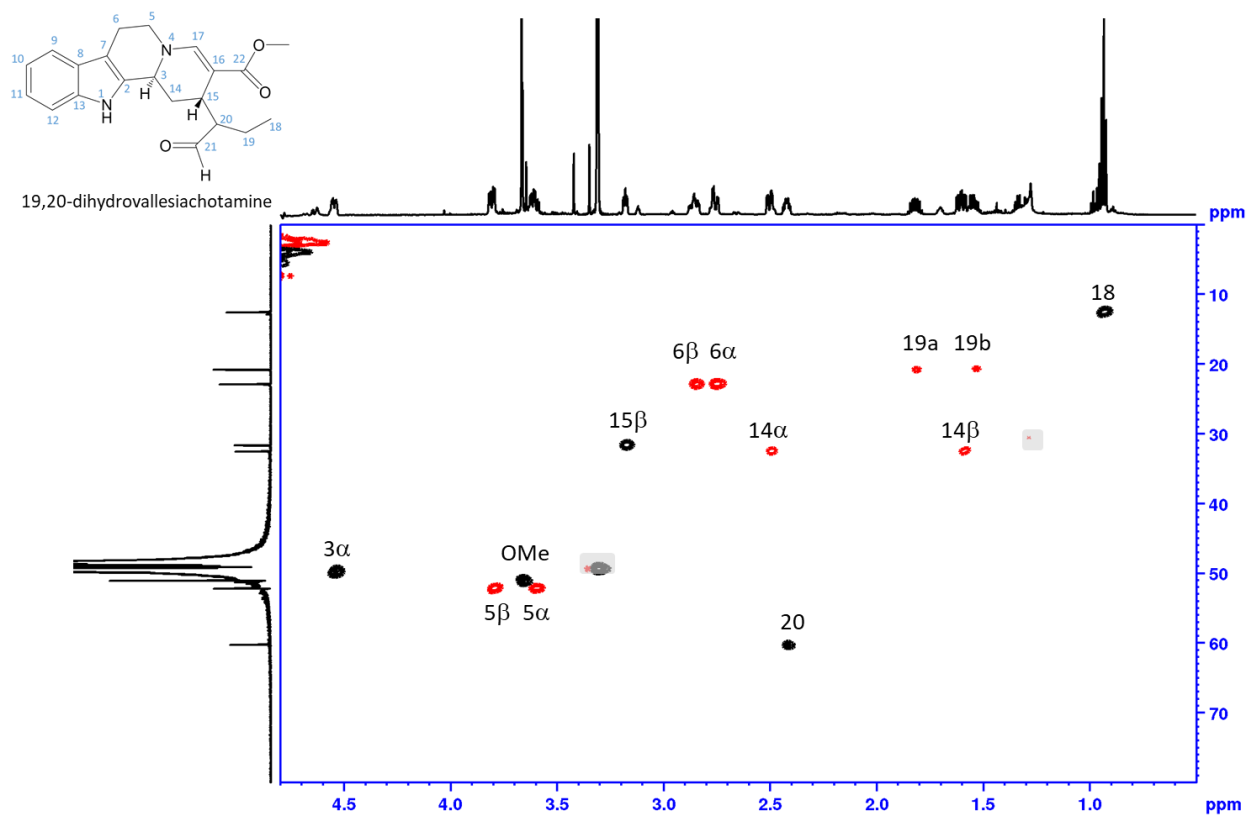


Figure 84. NMR data of 19,20-dihydrovallesiachotamine, phase sensitive HSQC, aliphatic range in MeOH- d_3 . Shaded areas mark impurity and solvent, red: CH₂, black: CH

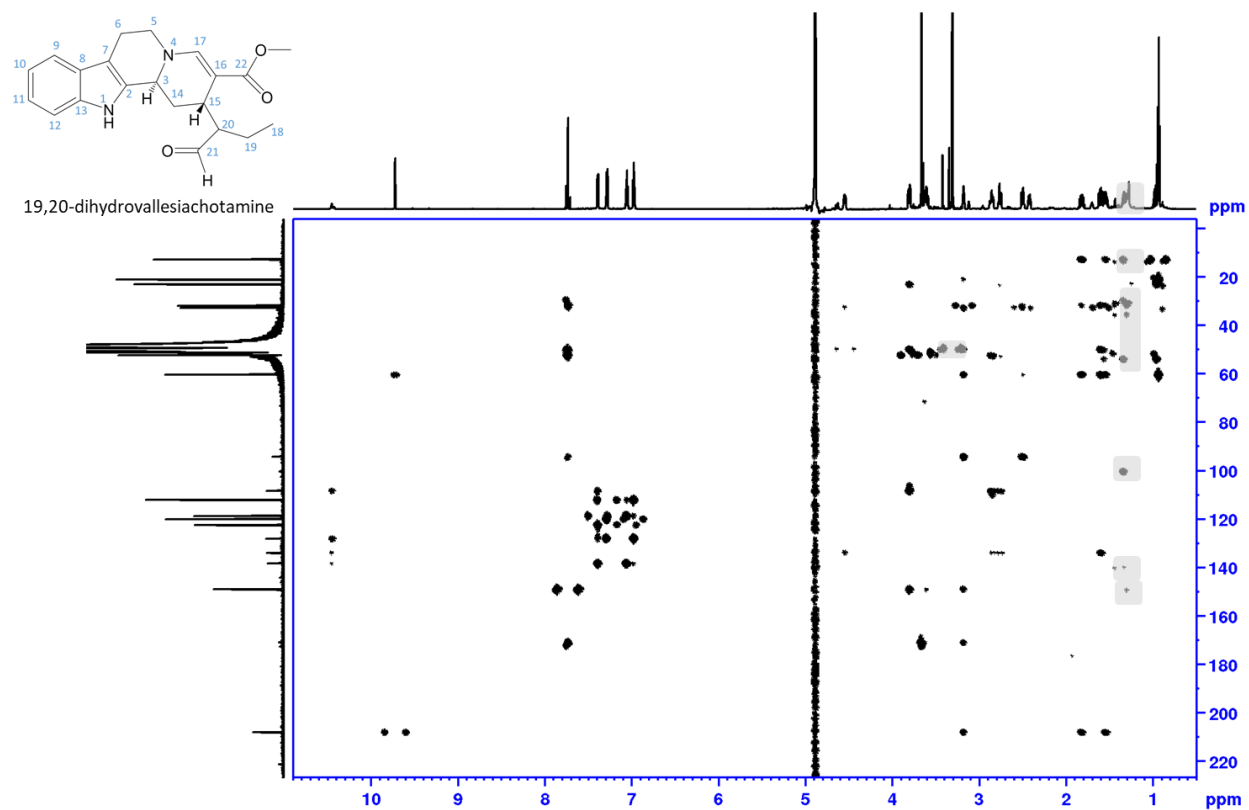


Figure 85. NMR data of 19,20-dihydrovallesiachotamine, HMBC, full range in MeOH- d_3 . Shaded areas mark impurity and solvent.

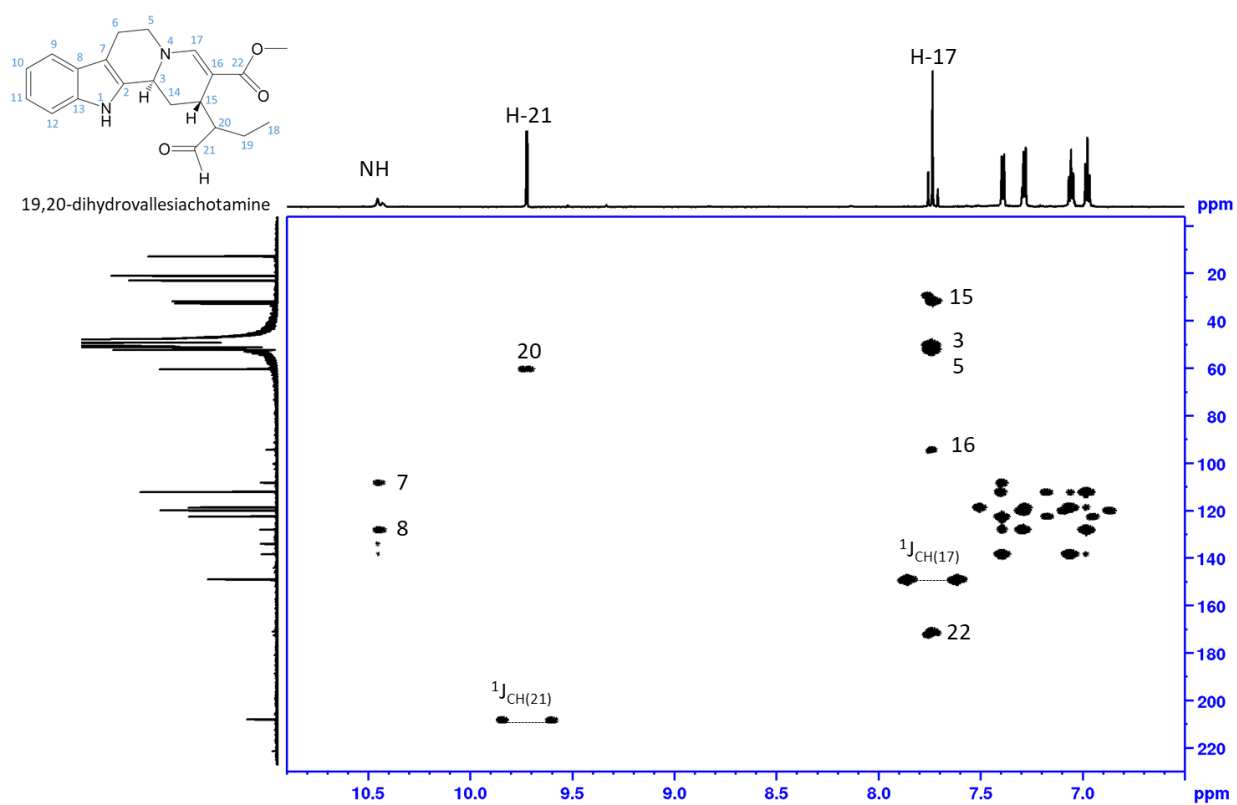


Figure 86. NMR data of 19,20-dihydrovallesiachotamine, HMBC, aldehyde and aromatic range in $\text{MeOH-}d_3$. Shaded areas mark impurity and solvent.

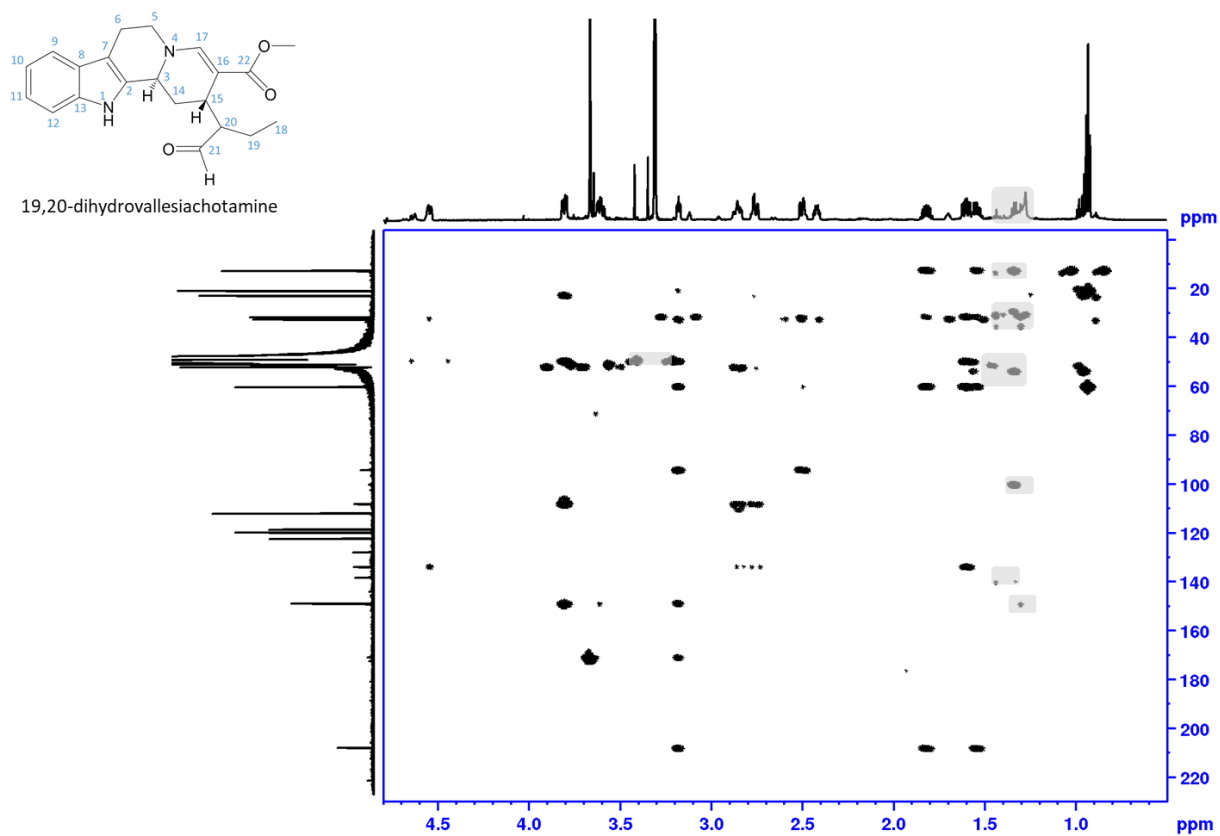


Figure 87. NMR data of 19,20-dihydrovallesiachotamine, HMBC, aliphatic range in MeOH-*d*₃. Shaded areas mark impurity and solvent.

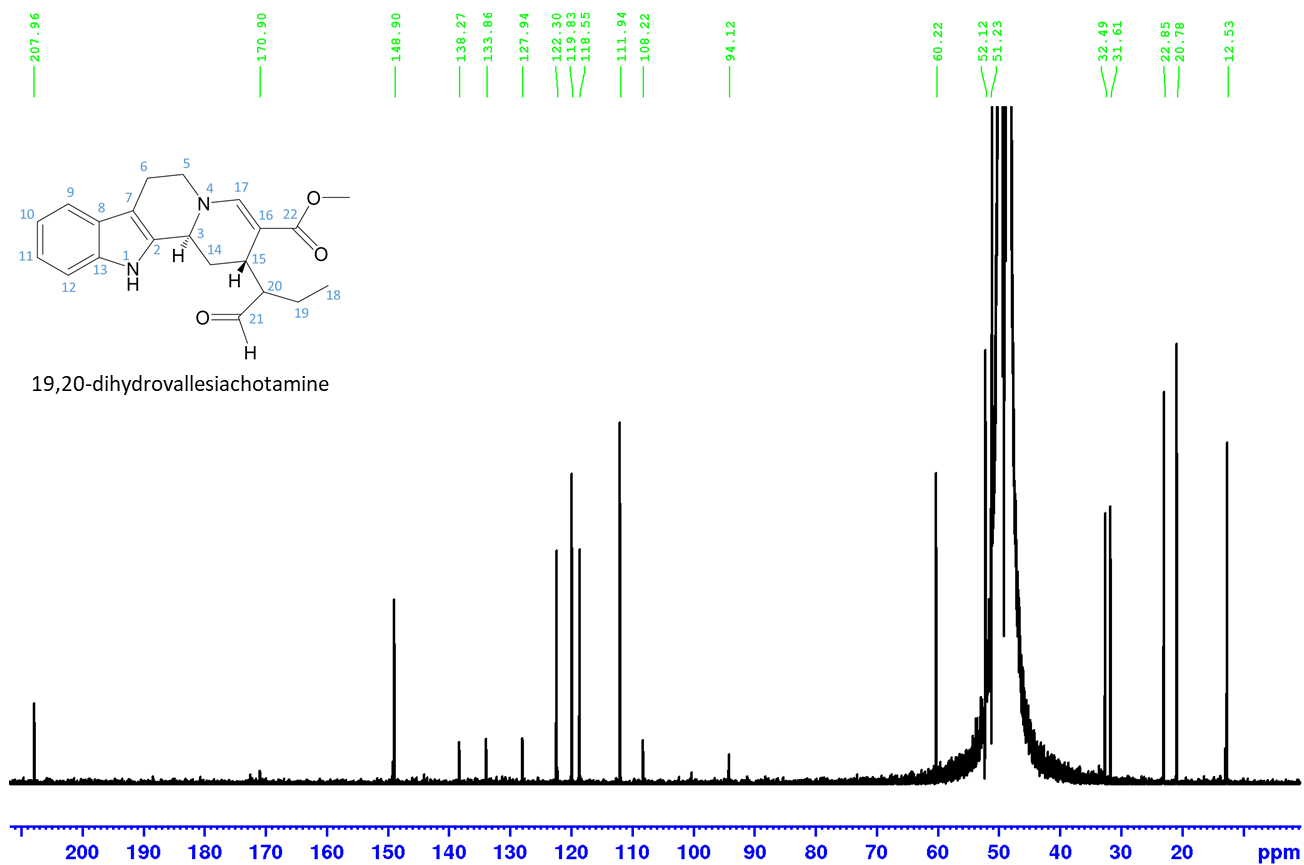


Figure 88. NMR data of 19,20-dihydrovallesiachotamine, DEPTQ, power spectrum, full range in MeOH-*d*₃.

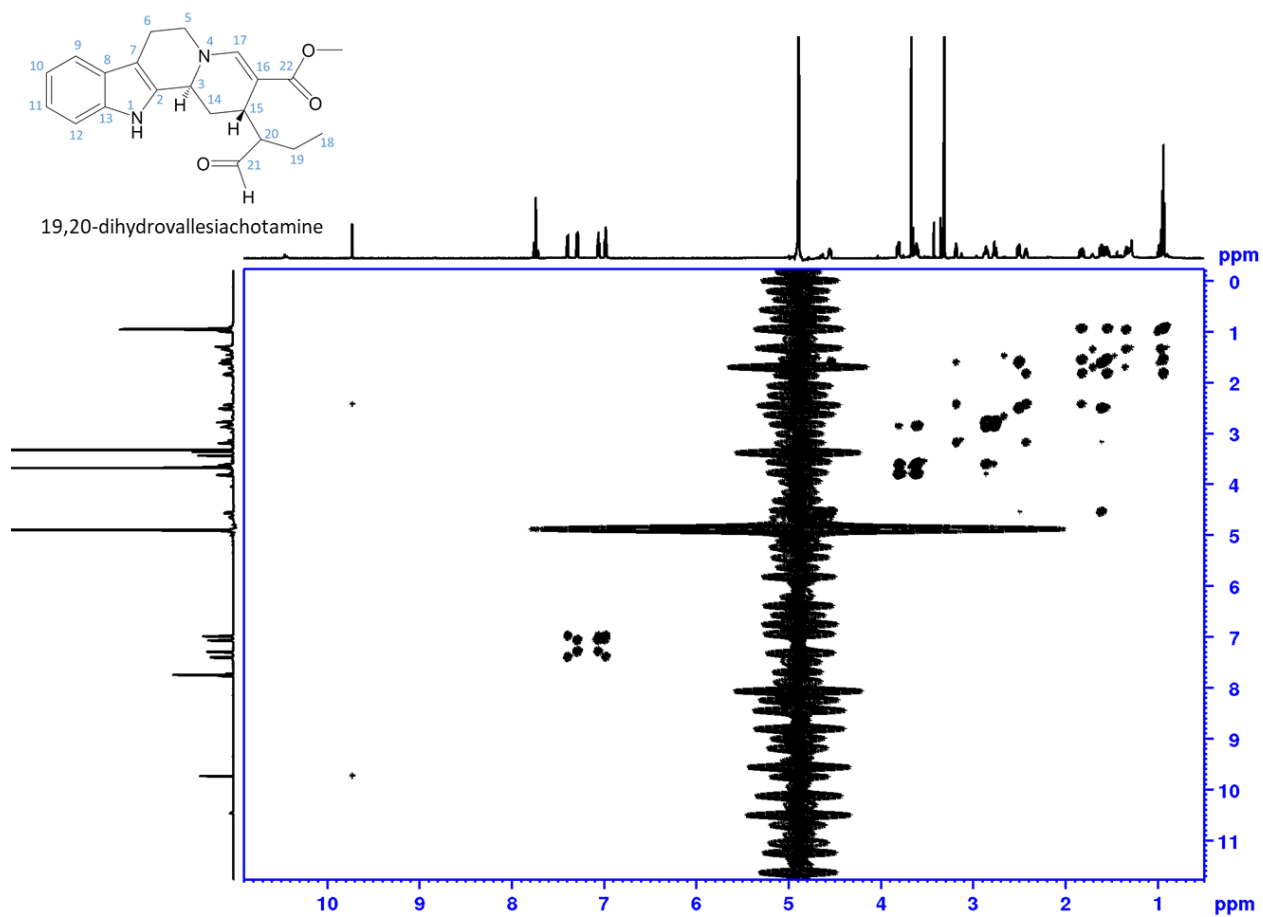


Figure 89. NMR data of 19,20-dihydrovallesiachotamine, ^1H - ^1H DQF COSY with water suppression, magnitude mode processed, full range in $\text{MeOH-}d_3$.

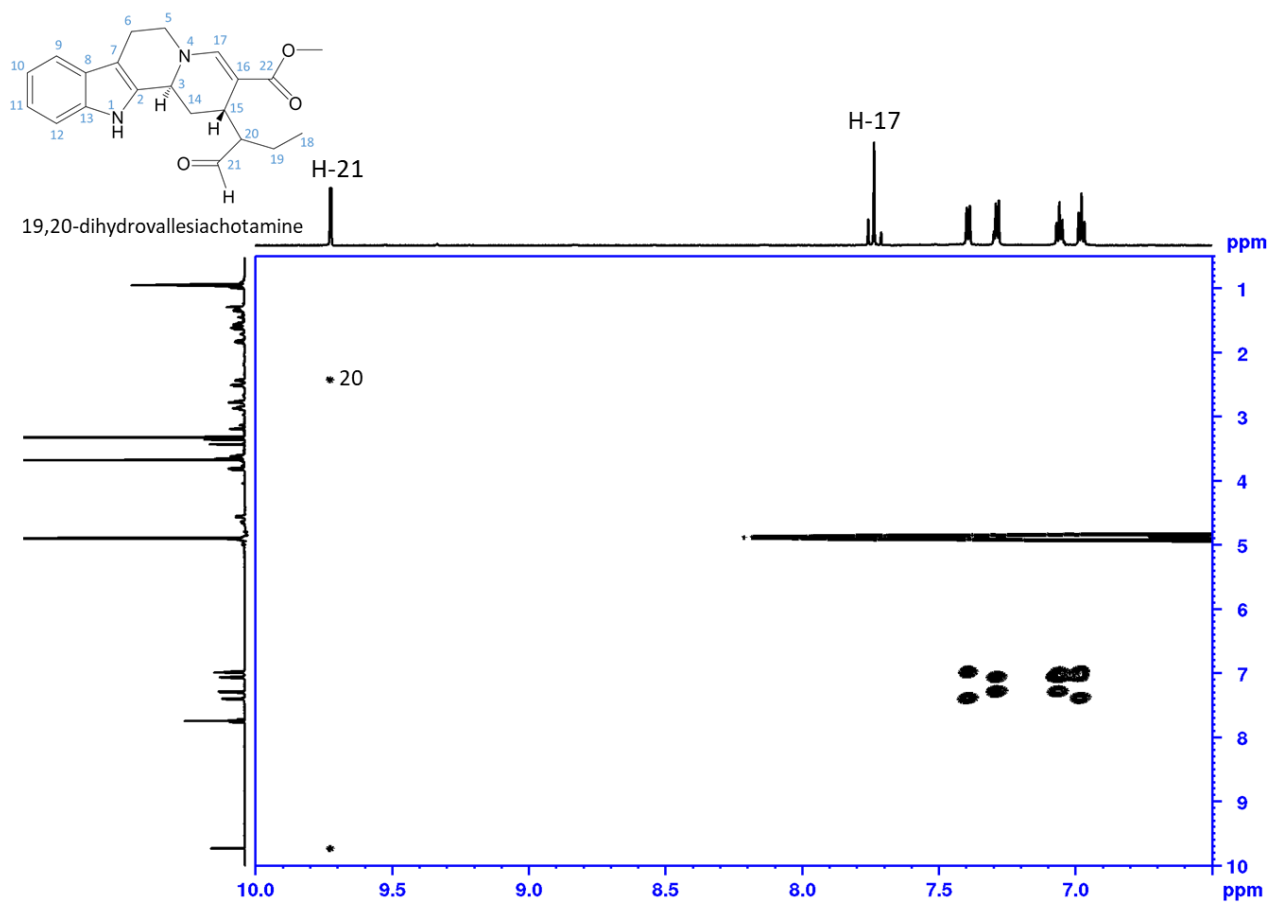


Figure 90. NMR data of 19,20-dihydrovallesiachotamine, ^1H - ^1H DQF COSY with water suppression, magnitude mode processed, aldehyde and aromatic range in $\text{MeOH-}d_3$.

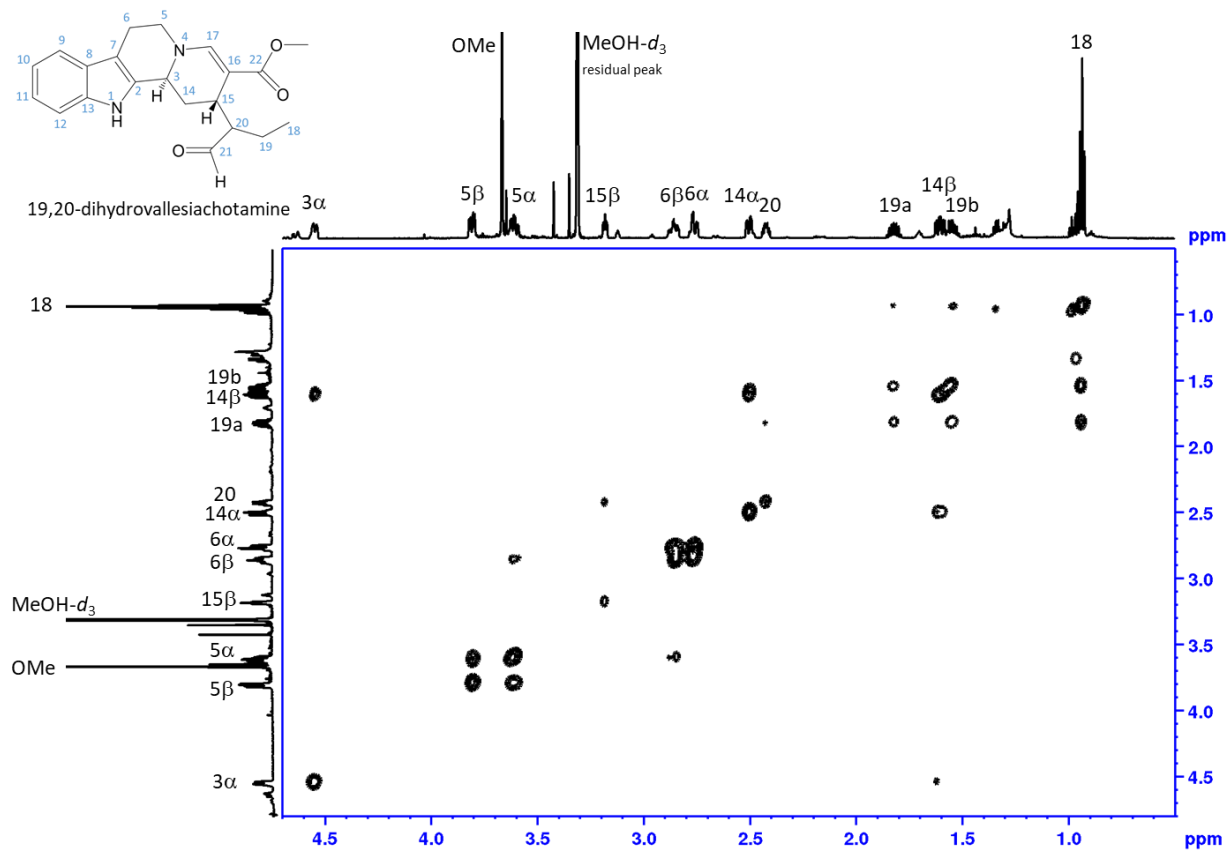


Figure 91. NMR data of 19,20-dihydrovallesiachotamine, ^1H - ^1H DQF COSY with water suppression, magnitude mode processed, aliphatic range in MeOH- d_3 .

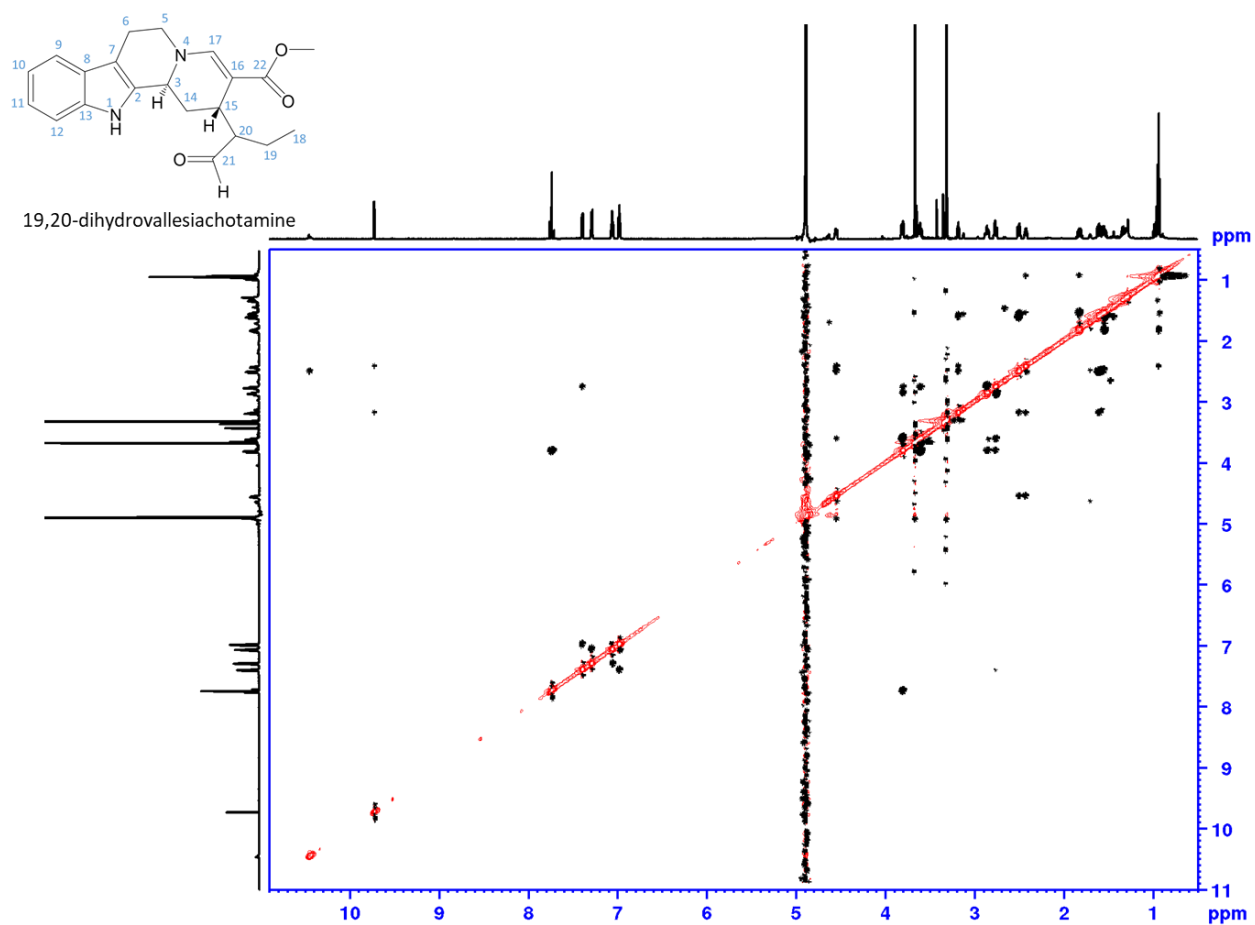


Figure 92. NMR data of 19,20-dihydrovallesiachotamine, ^1H - ^1H ROESY with water suppression, full range in $\text{MeOH-}d_3$

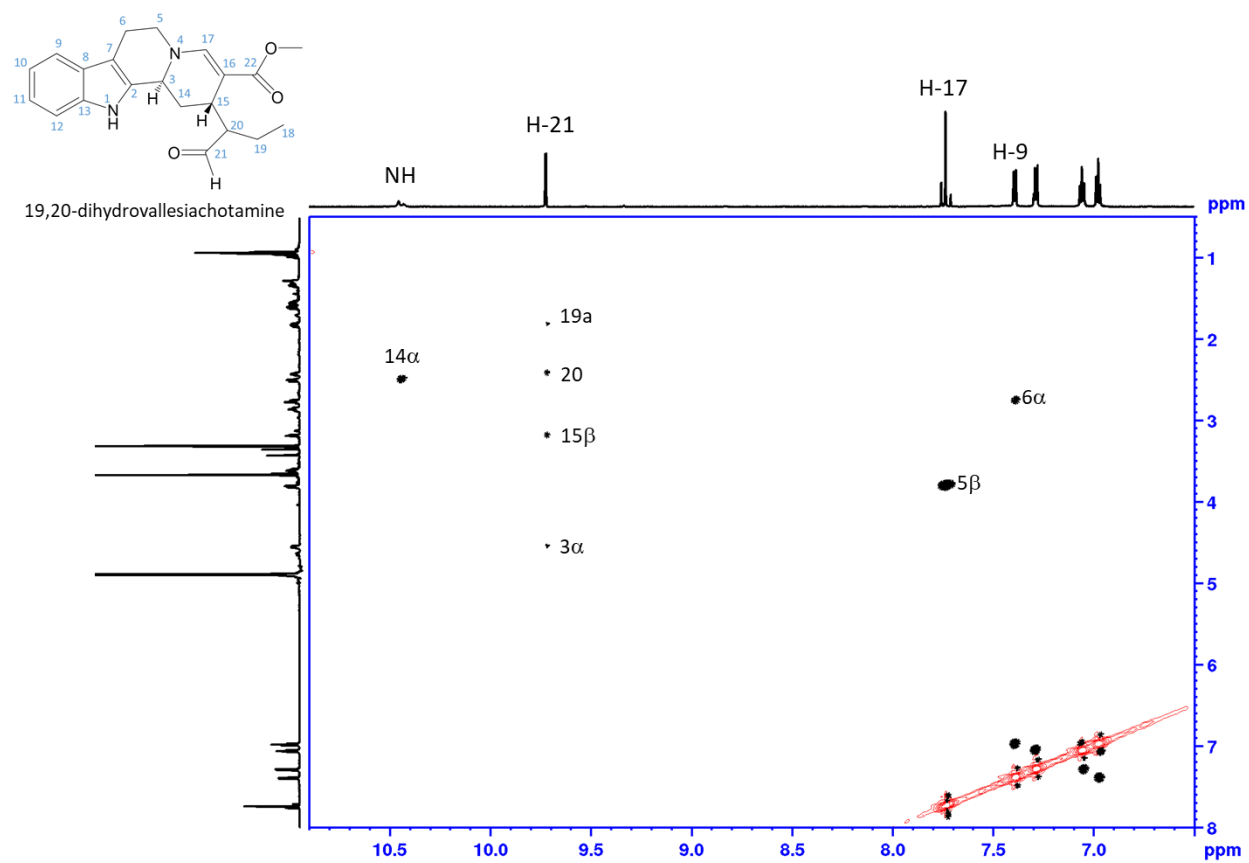


Figure 93. NMR data of 19,20-dihydrovallesiachotamine, ^1H - ^1H ROESY with water suppression, aldehyde and aromatic range in $\text{MeOH-}d_3$

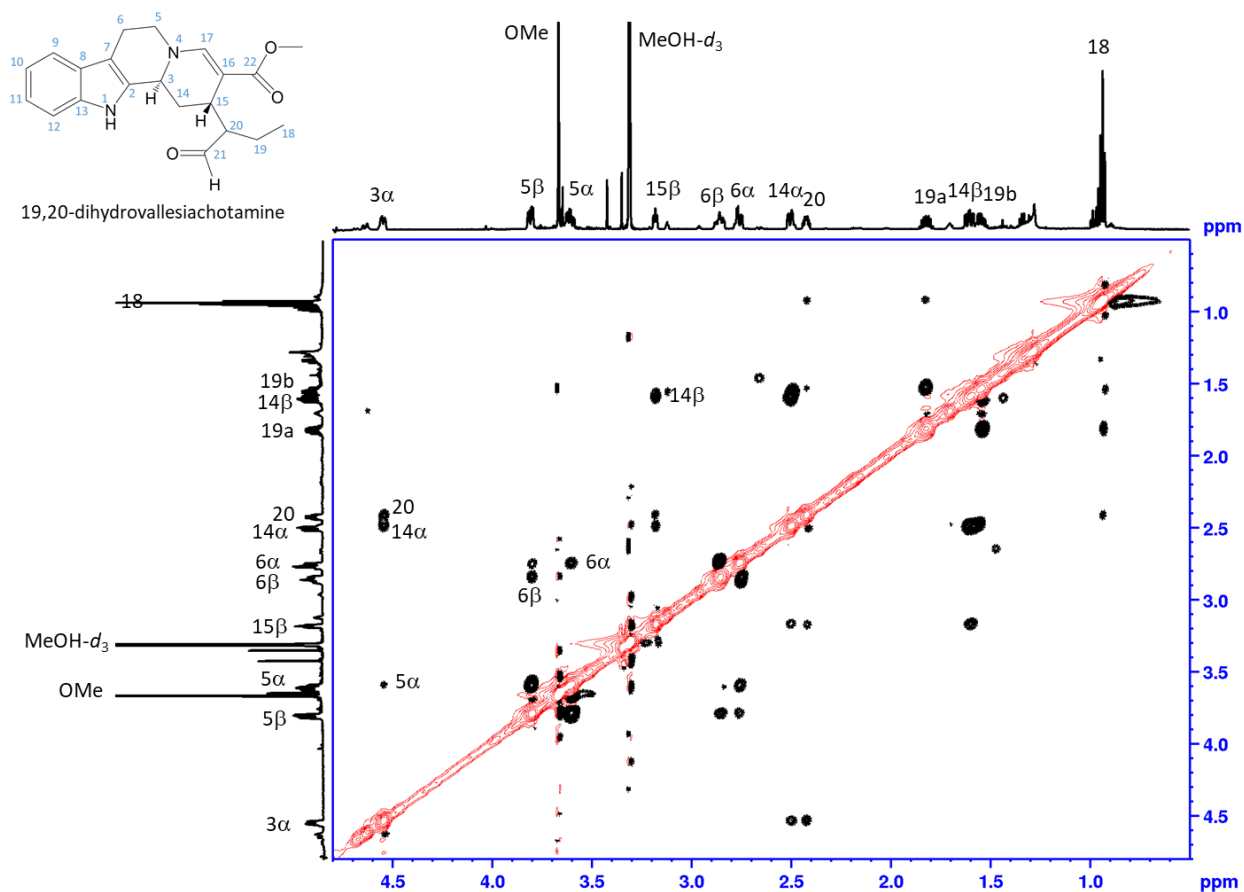


Figure 94. NMR data of 19,20-dihydrovallesiachotamine, ^1H - ^1H ROESY with water suppression, aliphatic range in MeOH- d_3

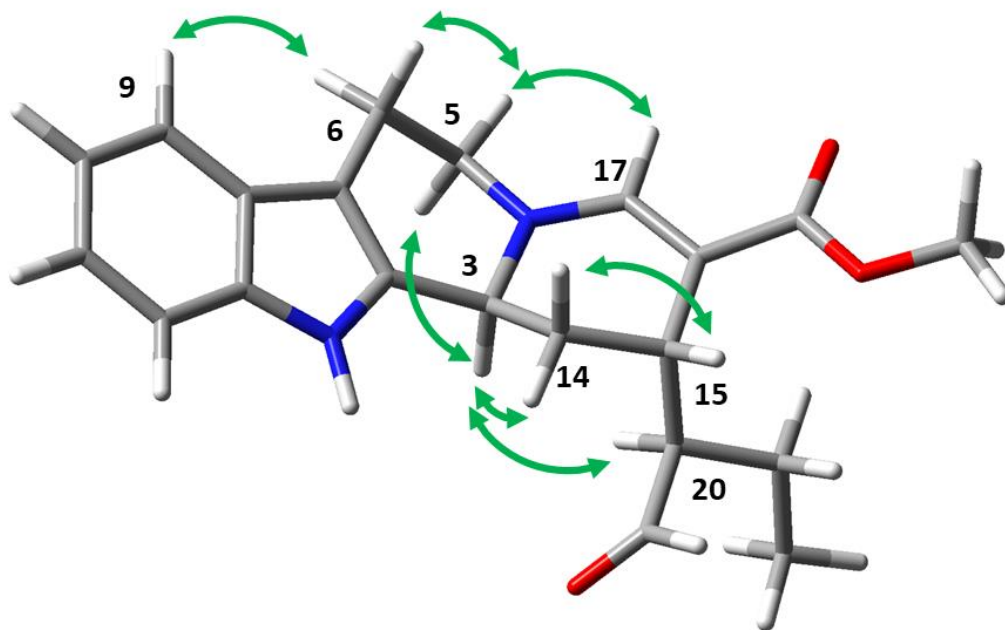


Figure 95. Structure of 19,20-dihydrovallesiachotamine optimized using Gaussian 16 (DFT APFD/6-311G++(2d,p), solvent MeOH). Important ROESY correlations extracted from NMR data are depicted in green.

Appendix V – X-Ray Data and Model Parameters used for Structure Solutions

Key for Solution Tables

^a $R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$.

^b $R_{\text{meas}} = \sum_{hkl} [N/(N - 1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th observation of reflection hkl , $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl .

^c $CC_{1/2}$ is the correlation coefficient between symmetry equivalent intensities from random halves of the dataset.

^d The data set was split into "working" and "free" sets consisting of 95 and 5% of the data respectively. The free set was not used for refinement.

^e The R-factors R_{work} and R_{free} are calculated as follows: $R = \sum(|F_{\text{obs}} - F_{\text{calc}}|) / \sum |F_{\text{obs}}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

^f As calculated using MolProbity ^[59].

Table 16. Summary of X-ray data and model parameters for CrDPAS.

Data collection	
Paul Scherrer Institute	10SA (PX II)
Wavelength (Å)	1
Resolution range (Å)	44.62 - 2.45 (2.548 - 2.45)
Space Group	<i>P</i> 21 21 21
Cell parameters (Å)	<i>a</i> = 61.019, <i>b</i> = 114.015, <i>c</i> = 143.357, β = 90°
Total no. of measured reflections	494135 (51201)
Unique reflections	37564 (3719)
Multiplicity	13.2 (13.8)
Mean <i>I</i> / σ (<i>I</i>)	21.46 (3.10)
Completeness (%)	98.7 (96.8)
<i>R</i> _{merge} ^a	0.2154 (1.406)
<i>R</i> _{meas} ^b	0.2242 (1.46)
<i>CC</i> _{1/2} ^c	0.999 (0.879)
Wilson <i>B</i> value (Å ²)	53.18
Refinement	
Reflections used in refinement	37560 (3719)
Reflections used for R-free	1877 (186)
<i>R</i> _{work}	0.2217 (0.2772)
<i>R</i> _{free}	0.2501 (0.3143)
<i>CC</i> _{work}	0.942 (0.845)
<i>CC</i> _{free}	0.958 (0.704)
Protein residues	640
Number of non-hydrogen atoms	4668
macromolecules	4566
ligands	43
solvent	69
Ramachandran plot:	98.1/1.58/0.32
favoured/allowed/disallowed ^f (%)	
Rotamer outliers (%)	3.97
R.m.s. bond distance deviation (Å)	0.007
R.m.s. bond angle deviation (°)	0.98
Clashscore	19.95
Mean <i>B</i> factors:	62.93/54.82/76.91/62.91
protein/waters/ligands/overall (Å ²)	
PDB accession code	8B27

Statistics for the highest-resolution shell are shown in parentheses.

Table 17. Summary of X-ray data and model parameters for apo-TiDPAS2.

Data collection	
Paul Scherrer Institute	10SA (PX II)
Wavelength (Å)	1
Resolution range (Å)	41.64 - 2.421 (2.508 - 2.421)
Space Group	<i>P</i> 21 21 21
Cell parameters (Å)	$a = 74.422$, $b = 78.124$, $c = 131.207$, $\beta = 90^\circ$
Total no. of measured reflections	341639 (17075)
Unique reflections	29562 (2702)
Multiplicity	11.6 (6.3)
Mean $I/\sigma(I)$	16.49 (1.32)
Completeness (%)	98.98 (91.90)
R_{merge}^a	0.08578 (0.8446)
R_{meas}^b	0.0897 (0.9206)
$CC_{1/2}^c$	0.999 (0.785)
Wilson B value (Å ²)	64.90
Refinement	
Reflections used in refinement	29531 (2700)
Reflections used for R-free	1477 (135)
R_{work}	0.2082 (0.3777)
R_{free}	0.2552 (0.4209)
CC_{work}	0.965 (0.812)
CC_{free}	0.901 (0.716)
Protein residues	716
Number of non-hydrogen atoms	5305
macromolecules	5269
ligands	2
solvent	34
Ramachandran	plot: 98.87/1.13/0.00
favoured/allowed/disallowed ^f (%)	
Rotamer outliers (%)	2.43
R.m.s. bond distance deviation (Å)	0.006
R.m.s. bond angle deviation (°)	0.95
Clashscore	5.27
Mean B factors:	71.16/57.66/62.53/71.07
protein/waters/ligands/overall (Å ²)	
PDB accession code	8B26

Statistics for the highest-resolution shell are shown in parentheses.

Table 18. Summary of X-ray data and model parameters for precondylocarpine acetate-bound *TiDPAS2*.

Data collection	
Paul Scherrer Institute	10SA (PX II)
Wavelength (Å)	1
Resolution range (Å)	39.81 – 1.882 (1.949 – 1.882)
Space Group	<i>P</i> 21 21 21
Cell parameters (Å)	<i>a</i> = 72.888, <i>b</i> = 79.624, <i>c</i> = 130.801, β = 90°
Total no. of measured reflections	809479 (78567)
Unique reflections	62174 (5895)
Multiplicity	13.0 (13.3)
Mean <i>I</i> / σ (<i>I</i>)	14.05 (0.88)
Completeness (%)	99.49 (95.74)
<i>R</i> _{merge} ^a	0.1082 (3.23)
<i>R</i> _{meas} ^b	0.1128 (3.357)
<i>CC</i> _½ ^c	0.999 (0.463)
Wilson <i>B</i> value (Å ²)	40.94
Refinement	
Reflections used in refinement	62152 (5895)
Reflections used for R-free	3104 (295)
<i>R</i> _{work}	0.1927 (0.4735)
<i>R</i> _{free}	0.2216 (0.5240)
<i>CC</i> _{work}	0.972 (0.696)
<i>CC</i> _{free}	0.966 (0.671)
Protein residues	716
Number of non-hydrogen atoms	5601
macromolecules	5272
ligands	97
solvent	242
Ramachandran plot:	97.33/2.67/0.00
favoured/allowed/disallowed ^f (%)	
Rotamer outliers (%)	0.93
R.m.s. bond distance deviation (Å)	0.004
R.m.s. bond angle deviation (°)	0.71
Clashscore	3.89
Mean <i>B</i> factors:	44.88/47.01/46.74/45.00
protein/waters/ligands/overall (Å ²)	
PDB accession code	8B1V

Statistics for the highest-resolution shell are shown in parentheses.

Table 19. Summary of X-ray data and model parameters for stemmadenine acetate-bound *TiDPAS2*.

Data collection	
Paul Scherrer Institute	10SA (PX II)
Wavelength (Å)	1
Resolution range (Å)	39.92 – 2.24 (2.32 – 2.24)
Space Group	<i>P</i> 21 21 21
Cell parameters (Å)	<i>a</i> = 73.186, <i>b</i> = 79.845, <i>c</i> = 130.922, β = 90°
Total no. of measured reflections	432608 (21387)
Unique reflections	35719 (2561)
Multiplicity	12.1 (8.4)
Mean <i>I</i> / σ (<i>I</i>)	17.69 (1.77)
Completeness (%)	94.79 (68.96)
$R_{\text{merge}}^{\text{a}}$	0.1239 (1.273)
$R_{\text{meas}}^{\text{b}}$	0.1294 (1.358)
$CC_{\frac{1}{2}}^{\text{c}}$	0.999 (0.586)
Wilson <i>B</i> value (Å ²)	44.54
Refinement	
Reflections used in refinement	35691 (2561)
Reflections used for R-free	1786 (128)
R_{work}	0.1737 (0.3245)
R_{free}	0.2199 (0.3957)
CC_{work}	0.972 (0.790)
CC_{free}	0.957 (0.700)
Protein residues	717
Number of non-hydrogen atoms	5530
macromolecules	5272
ligands	114
solvent	168
Ramachandran plot:	96.49/3.51/0.00
favoured/allowed/disallowed ^f (%)	
Rotamer outliers (%)	2.79
R.m.s. bond distance deviation (Å)	0.148
R.m.s. bond angle deviation (°)	4.02
Clashscore	5.86
Mean <i>B</i> factors:	45.79/46.44/45.13/45.78
protein/waters/ligands/overall (Å ²)	
PDB accession code	8B25

Statistics for the highest-resolution shell are shown in parentheses.

Appendix VI. Identification of CADs in *C. roseus*

Table 20. Contigs encoding CADs retrieved from the *C. roseus* genome.

Name	Contig	CAD type	Chromosome	Direction	Start position	End position	Gene length (bp)	Number of exons	Coding sequence length (bp)
	S004610	1,4-iminium	Scaffold 09	Fwd	37214	39828	2614	5	1095
	01G014020	1,2-iminium	1	Fwd	20847005	20849541	2536	4	1077
	01G014040	1,2-iminium	1	Fwd	20982703	20985249	2546	4	1077
ADH38	01G014080	1,2-iminium	1	Fwd	21409862	21412414	2552	4	1077
	01G014090	1,2-iminium	1	Fwd	21660916	21663437	2521	5	900
ADH117	01G017410	1,4-iminium	1	Fwd	32476926	32479384	2458	5	1095
	01G017430	1,4-iminium	1	Rev	32526597	32524231	2366	5	1047
	01G027010	Typical	1	Fwd	68444118	68445956	1838	6	1236
	01G032390	1,2-iminium	1	Fwd	76798229	76803026	4797	5	1074
HYS	01G032400	1,2-iminium	1	Fwd	76817574	76820091	2517	5	1089
	01G032410	1,2-iminium	1	Fwd	76830463	76833921	3458	5	1089
	01G032420	1,2-iminium	1	Fwd	76847844	76851152	3308	5	1089
	01G032430	1,2-iminium	1	Fwd	76847844	76851152	3308	5	1089
	01G033080	1,2-iminium	1	Rev	77630006	77627023	2983	5	1134
THAS3	01G033090	1,2-iminium	1	Rev	77640569	77633747	6822	5	1152
THAS1	01G033230	1,2-iminium	1	Rev	77777894	77774074	3820	5	1071
	03G009850	Typical	3	Fwd	25914923	25922196	7273	7	1299
ADH9	03G022230	1,4-iminium	3	Rev	60738581	60733835	4746	5	1098
CAD	03G026880	Typical	3	Fwd	67341888	67344419	2531	5	1086
	04G001600	Typical	4	Rev	1533999	1531930	2069	10	1143
	04G004530	Typical	4	Fwd	3464070	3467040	2970	10	1143
	04G004790	Typical	4	Fwd	3680824	3684319	3495	9	1140
	04G028370	Typical	4	Fwd	64596046	64598786	2740	10	1143
	05G008450	Typical	5	Rev	10372377	10366145	6232	8	1377
T3R	05G015990	1,2-iminium	5	Rev	28291913	28288757	3156	5	1056
ADH24	05G016710	Typical	5	Fwd	31655252	31659758	4506	5	1086
RedOx1	05G017150	RedOx1-like	5	Fwd	33046696	33050140	3444	5	1065
DPAS	05G017180	1,4-iminium	5	Fwd	33141553	33146215	4662	5	1098
ADH20	05G017210	1,4-iminium	5	Fwd	33221022	33223944	2922	5	1095
ADH39	05G017220	1,4-iminium	5	Rev	33262382	33259651	2731	5	1095
ADH104	05G017270	1,4-iminium	5	Fwd	33514985	33518674	3689	5	1083
ADH17	05G017590	1,2-iminium	5	Rev	35153908	35150434	3474	5	1074
	05G033170	Uncharacterised atypical ADH	5	Rev	69851825	69849971	1854	7	1170

ADH23	06G024530	Typical	6	Fwd	58741266	58743372	2106	5	1086
ADH33	06G024550	Typical	6	Fwd	58764154	58766850	2696	5	1089
8HGO	06G024560	Typical	6	Fwd	58779322	58780949	1627	4	1191
ADH18	06G024570	Typical	6	Fwd	58814190	58816678	2488	5	1092
Cr2141	06G024580	Typical	6	Fwd	58861701	58864080	2379	5	1086
ADH32	06G024590	Typical	6	Fwd	58911502	58916377	4875	5	1086
GS	06G024600	1,2-iminium	6	Rev	58929151	58927274	1877	5	1095
GS2	06G024610	1,2-iminium	6	Rev	58941541	58938474	3067	5	1089
THAS2	06G024620	1,2-iminium	6	Rev	58962346	58960348	1998	5	1116
ADH1	06G024660	Typical	6	Fwd	59100734	59104326	3592	5	1086
ADH21	08G001030	Typical	8	Rev	984953	981610	3343	6	1089
	08G004630	Typical	8	Fwd	4239805	4243545	3740	9	1194
	08G004640	Typical	8	Fwd	4256033	4258739	2706	10	1215
	08G004830	1,2-iminium	8	Rev	4461823	4458412	3411	6	1095

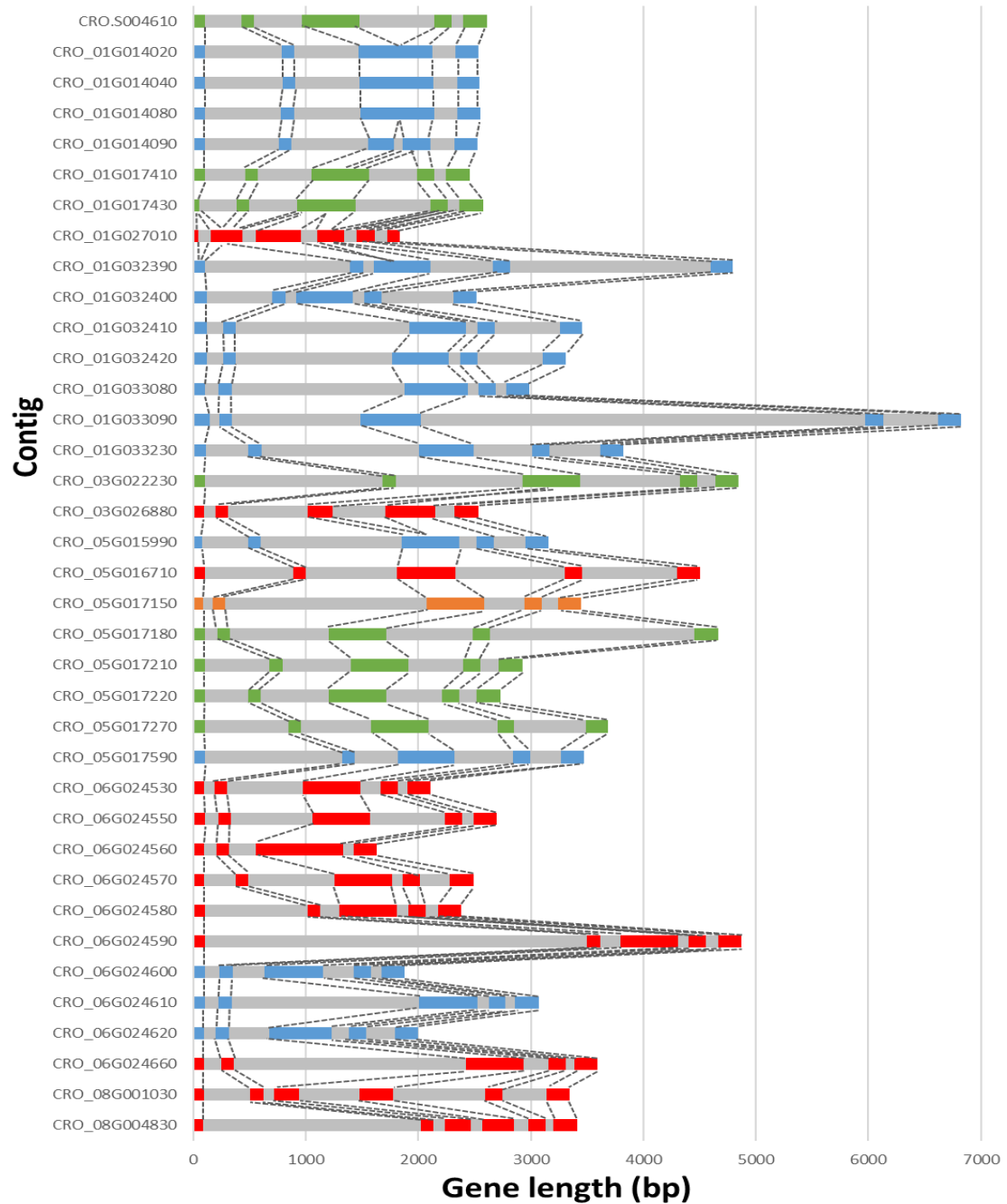


Figure 96. Gene structure of selected CADs from *C. roseus*. Exons coloured by predicted reduction chemistry as either typical aldehyde reducing (red), 1,2-iminium reducing (blue), 1,4-iminium/ α,β -unsaturated aldehyde reducing (green) or RedOx1-like 1,2-iminium reducing (orange), introns coloured in grey. Homologous exons shown as dashed lines.

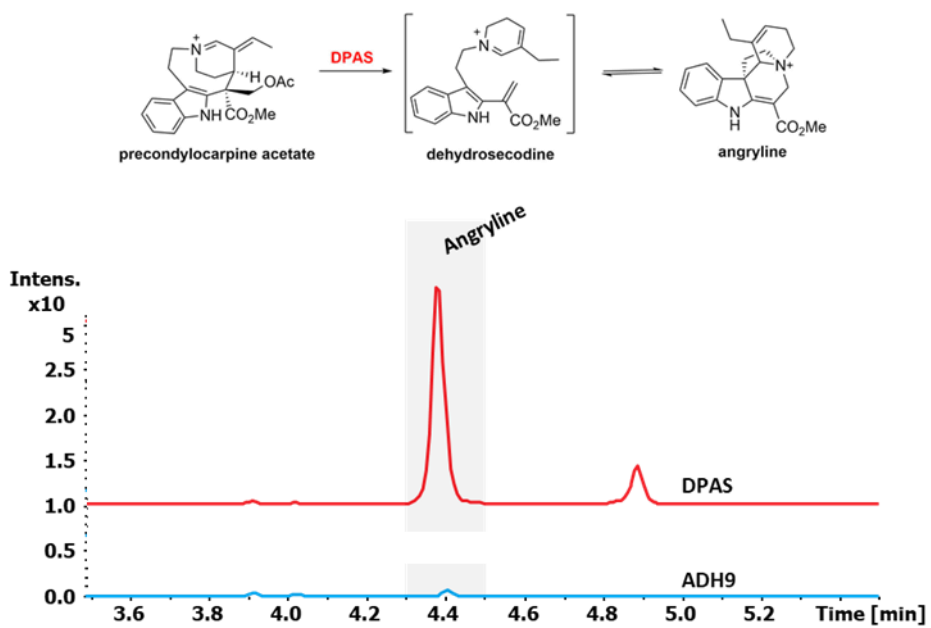


Figure 97 UPLC-MS chromatograms of *in vitro* reactions of CrADH9 and CrDPAS with substrate precondylocarpine acetate and cofactor NADPH. EIC m/z 337.1908 \pm 0.05

Table 21. Contigs of CADs identified in the *C. roseus* leaf single-cell transcriptome. Genes with low quality matches (<90% sequence similarity) are highlighted, sequences with duplicate single cell contig matches are marked by *. + denotes CADs with no recorded expression in the dataset.

Name	Genome Contig	CAD type	Chromosome	Single cell Contig	Identity (%)
	S004610	1,4-iminium	Scaffold 09	T133495	96.8
	01G014020	1,2-iminium	1	T120551**	97.8
	01G014040	1,2-iminium	1	T120551**	99.2
ADH38	01G014080	1,2-iminium	1	T117489 ⁺	99.5
	01G014090	1,2-iminium	1	T117489	78.0
ADH117	01G017410	1,4-iminium	1	T116009	98.5
	01G017430	1,4-iminium	1	T133495	87.1
	01G027010	Typical	1	T107879	84.9
	01G032390	1,2-iminium	1	T116105	99.5
HYS	01G032400	1,2-iminium	1	T116107*	99.2
	01G032410	1,2-iminium	1	T116107*	90.3
	01G032420	1,2-iminium	1	T116107*	89.0
	01G032430	1,2-iminium	1	T116107*	88.7
	01G033080	1,2-iminium	1	T113649	84.4
THAS3	01G033090	1,2-iminium	1	T113649	100.0
THAS1	01G033230	1,2-iminium	1	T113666	100.0
	03G009850	Typical	3	T125488	99.4
ADH9	03G022230	1,4-iminium	3	T106276	99.7
CAD	03G026880	Typical	3	T110942	100.0
	04G001600	Typical	4	T117429	100.0
	04G004530	Typical	4	T117157	100.0
	04G004790	Typical	4	T117130	100.0
	04G028370	Typical	4	T118722	100.0
	05G008450	Typical	5	T132382	93.9
T3R	05G015990	1,2-iminium	5	T124298	100.0
ADH24	05G016710	Typical	5	T129372	93.0
RedOx1	05G017150	RedOx1-like	5	T129272	100.0
DPAS	05G017180	1,4-iminium	5	T129267	100.0
ADH20	05G017210	1,4-iminium	5	T129261	100.0
ADH39	05G017220	1,4-iminium	5	T129257	100.0

ADH104	05G017270	1,4-iminium	5	T129239	91.1
ADH17	05G017590	1,2-iminium	5	T130593 ⁺	100.0
	05G033170	Uncharacterised atypical ADH	5	T123686	100.0
ADH23	06G024530	Typical	6	T113170 ⁺	100.0
ADH33	06G024550	Typical	6	T113168 ⁺	100.0
8HGO	06G024560	Typical	6	T113167	90.9
ADH18	06G024570	Typical	6	T113162	97.3
Cr2141	06G024580	Typical	6	T113160	100.0
ADH32	06G024590	Typical	6	T113155	95.0
GS	06G024600	1,2-iminium	6	T113154	100.0
GS2	06G024610	1,2-iminium	6	T113153	94.5
THAS2	06G024620	1,2-iminium	6	T113150	100.0
ADH1	06G024660	Typical	6	T113141	100.0
ADH21	08G001030	Typical	8	T127072	100.0
	08G004630	Typical	8	T128015	100.0
	08G004640	Typical	8	T128017 ⁺	100.0
	08G004830	1,2-iminium	8	T128047	93.3

Appendix VII. Genomic Synteny between *C. roseus* and *O. pumila*

Table 22. Genes encoding enzymes involved in MIA biosynthesis or CADs in *C. roseus* that have genomic synteny with *O. pumila*.

Gene name	<i>C. roseus</i> chromosome no.	<i>C. roseus</i> contig no.	<i>C. roseus</i> chromosome position	<i>O. pumila</i> chromosome no.	<i>O. pumila</i> contig no.	<i>O. pumila</i> chromosome position
G8H	Chr1	01G006750	5625533: 5628506	Chr7	07_g000725	3988470: 3990699
NPF2.1	Chr2	02G025780	71581858: 71589059	Chr1 1	11_g0082800	36905489: 36910805
MATE2	Chr3	03G016570	46991099: 46996540	Chr5	05_g0060800	28196827: 28204828
CAD	Chr3	03G026880	67341687: 67344673	Chr2	02_g0073180	39415719: 39418861
GES	Chr3	03G031560	70762775: 70768219	Chr5	05_g0006110	3359253: 3362648
STR	Chr3	03G032330	71668285: 71672478	Chr5	05_g0008300	4385001: 4387208
TDC	Chr3	03G032340	71677828: 71679666	Chr5	05_g0008350	4411007: 4423844
7DLH	Chr3	03G018360	51553563: 51556569	Chr5	05_g0058250	26867950: 26870659
LAMT	Chr3	03G019430	54209120: 54212389	Chr5	05_g0056110	25759985: 25763218
ADH	Chr4	04G004790	3680690: 3684754	Chr1	01_g0069870	33487543: 33490695
SLS2	Chr4	04G013510	24039159: 24042147	Chr2	02_g001299	7570958: 7575077
GGPPS	Chr4	04G032090	67982251: 67984041	Chr7	07_g0083430	38735179: 38741401
RedOx2	Chr5	05G008760	11150554: 11154711	Chr4	04_g0012040	6938637: 6945461

ADH24	Chr5	05G016710	31655095: 31660003	Chr7	07_g0077810	35942412: 35945028
ADH	Chr5	05G033170	69849971: 69851825	Chr9	09_g0004900	3072734: 3075082
GS	Chr6	06G024600	58926982: 58929303	Chr7	07_g0012250	6936485: 6941694
GS2	Chr6	06G024610	58929151: 58927274	Chr7	07_g0012220	6906865: 6909134
ISY	Chr7	07G007660	6958033: 6960384	Chr7	07_g0008900	4971893: 4973566
PRX1	Chr7	07G012950	17312716: 17318106	Chr1 0	10_g0010700	6067283: 6090265
ADH	Chr8	08G004630	4239717: 4243977	Chr8	08_g0085440	38692328: 38697017
ADH21	Chr8	08G001030	981336: 985217	Chr1 0	10_g0001910	938397: 940652

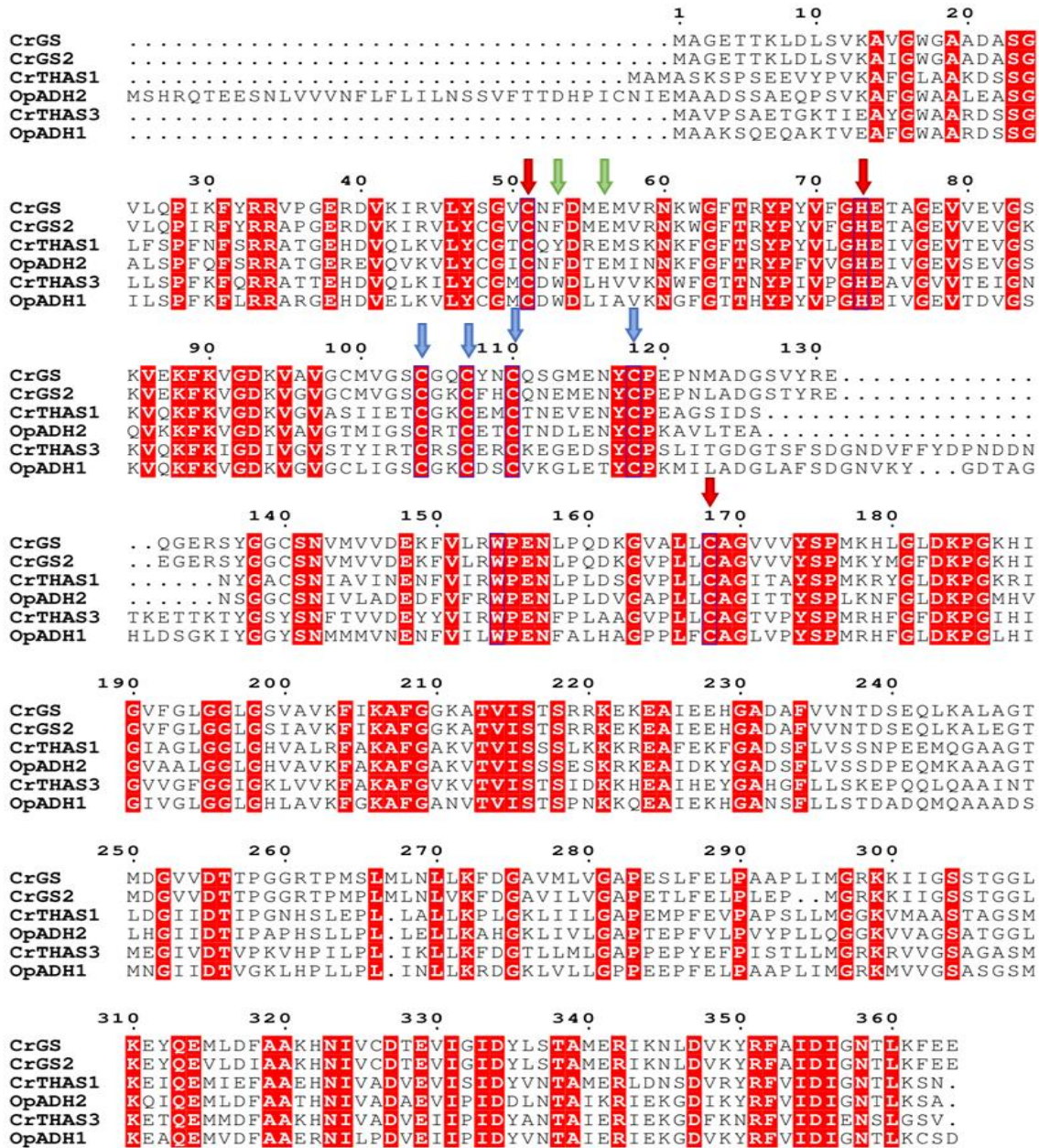


Figure 98. Amino acid sequence alignment of *OpADH1* and *OpADH2* with *C. roseus* CADs that catalyse a 1,2-reduction of an iminium moiety. Residues involved in coordinating the structural zinc ion (blue arrows), residues that coordinate the catalytic zinc ion (red arrows), and residues in positions involved in cofactor binding (green arrows) are highlighted. Figure made using ESPript 3.0 [38].

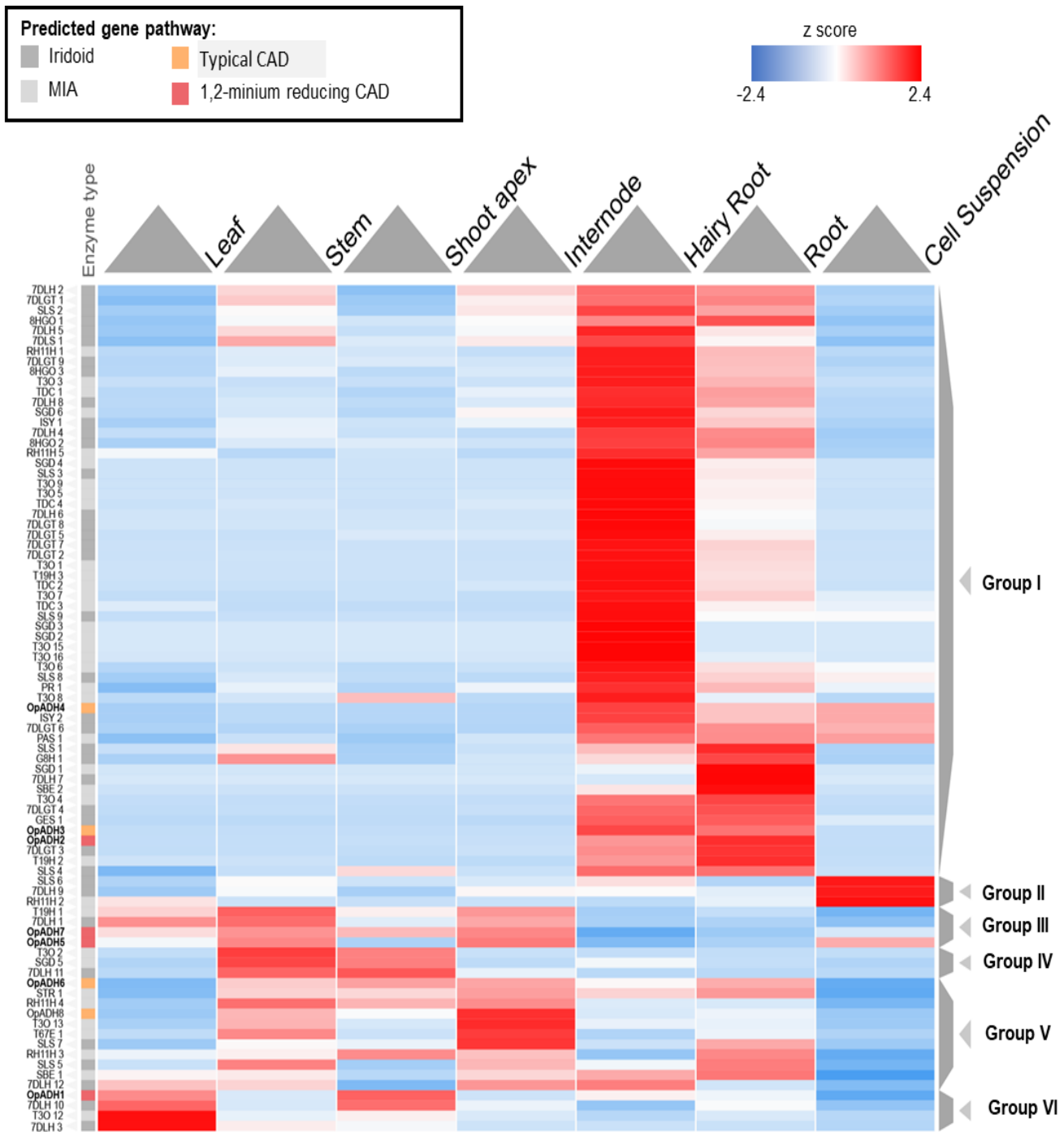


Figure 99. Hierarchical clustering of tissue-specific gene expression of CADs, and orthologues of *C. roseus* secoiridoid and MIA biosynthetic genes in *O. pumila*. CADs shown in bold. Figure adapted from data published by Rai *et al.*, [3] and made using Clustergrammer [66].

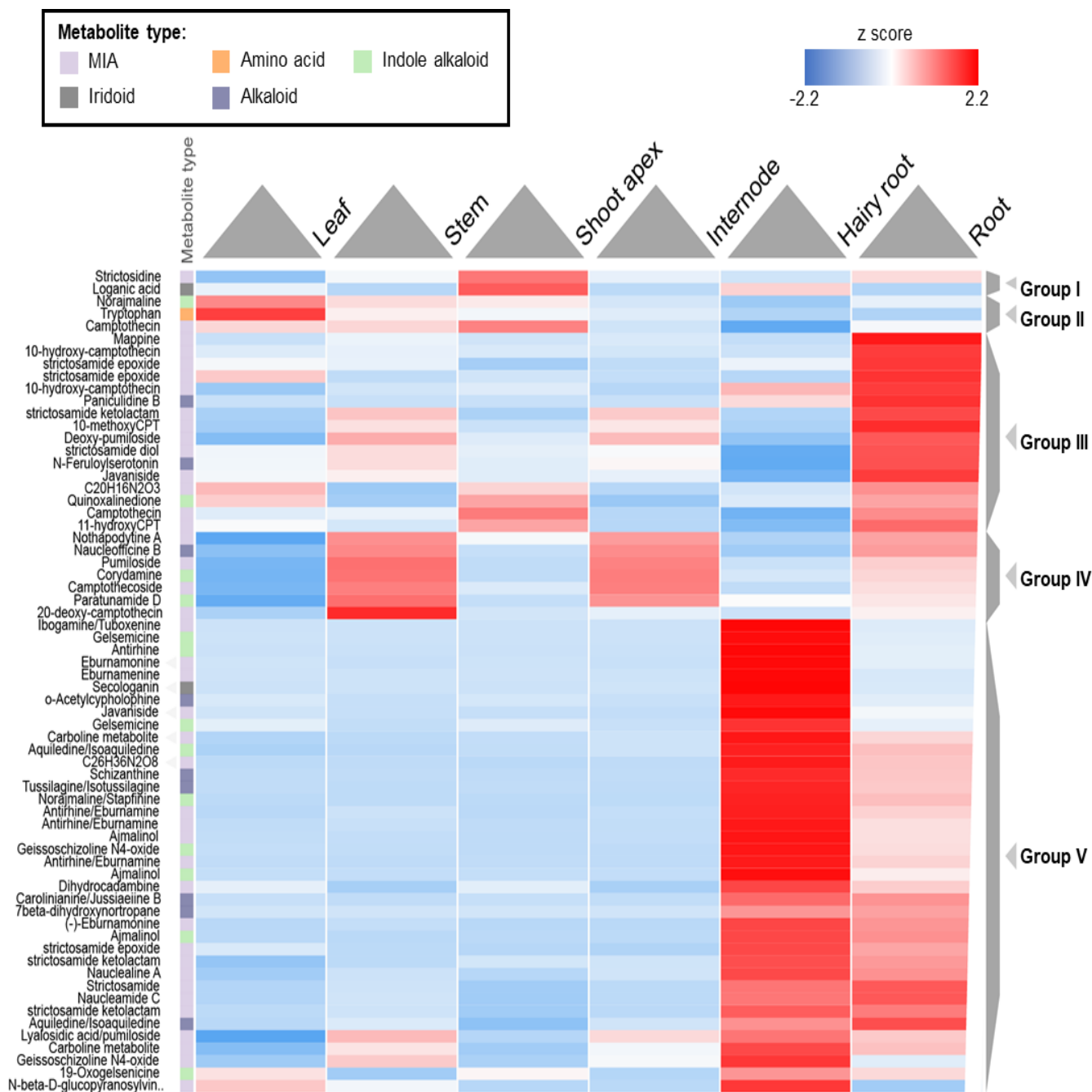


Figure 100. Hierarchical clustering of tissue-specific accumulation of metabolites from *O. pumila*. Metabolites predicted based on MS/MS spectral library. Figure adapted from Rai *et al.*, [3] and made using Clustergrammer [66]. Alkaloids are mostly enriched in root tissue, (clustered in groups III and V), though camptothecin accumulates in the stem tissue (clustered in group IV).

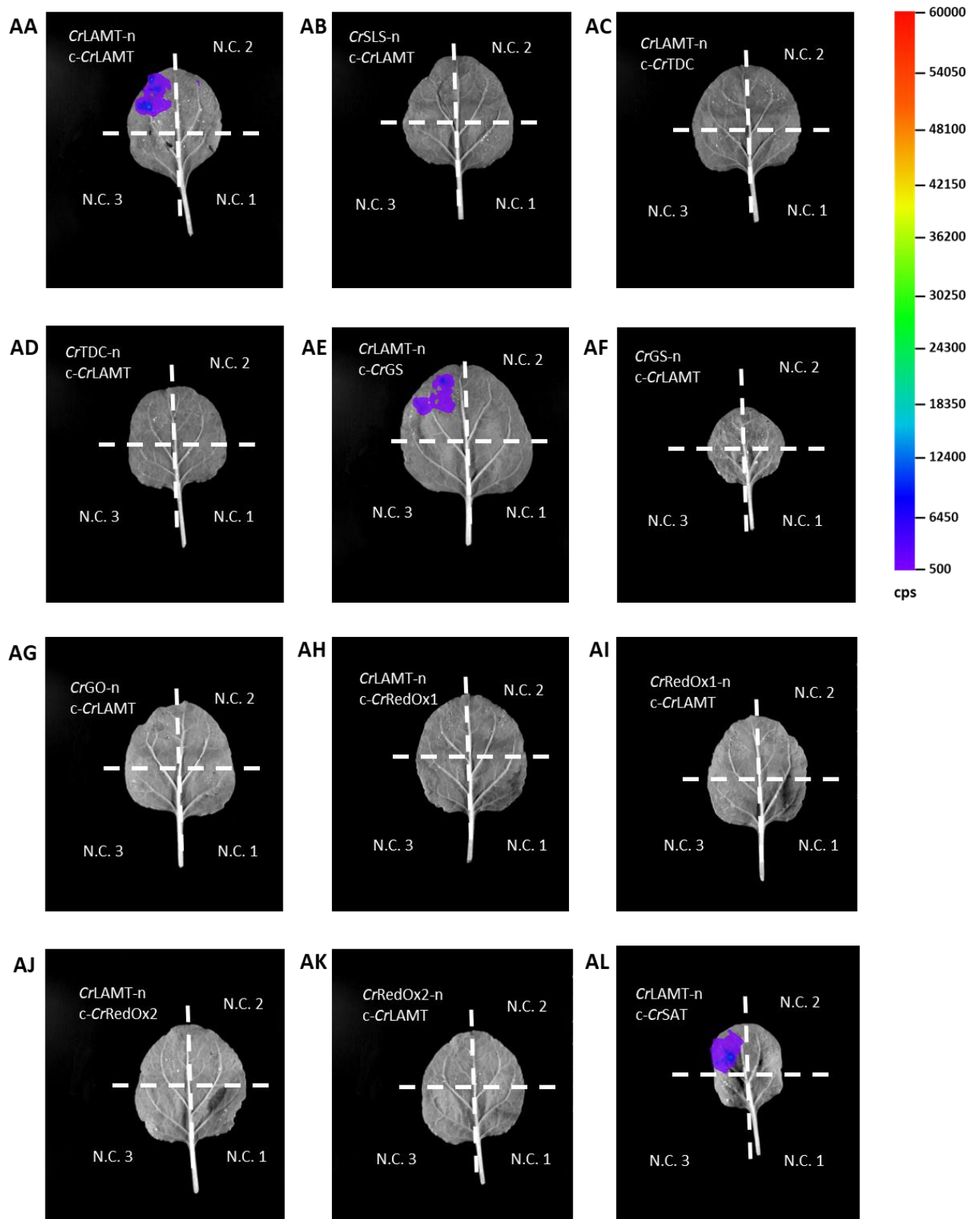
Appendix VIII. Subcellular Localisation of *C. roseus* MIA Biosynthetic Enzymes

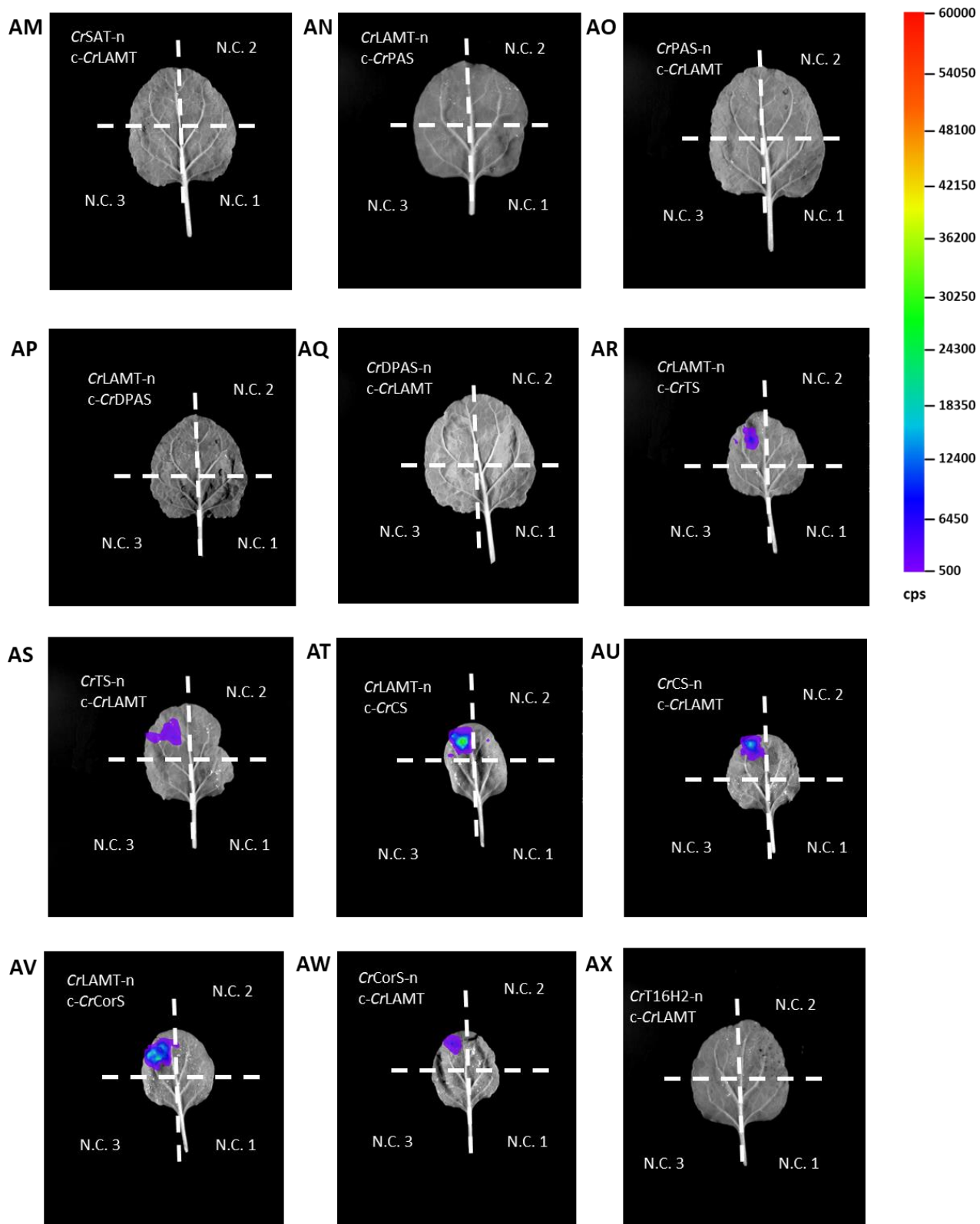
Table 23. Experimentally or predicted subcellular localisation of *C. roseus* MIA biosynthetic enzymes.

Enzyme Name	Enzyme Class (BRENDA number)	Experimental localisation	Plant-mSubP predicted localisation ^[55]
LAMT	Methyl transferase		Cytosolic
SLS	Cytochrome P450 type 72		Endoplasmic reticulum
TDC	Decarboxylase		Cytosolic
STR	Amine lyase	Vacuolar ^[15]	
SGD	Deglucosidase	Nuclear ^[15,16]	
GS	CAD-like ADH	Nucleocytosolic ^[9]	
GO	Cytochrome P450 type 71	Endoplasmic reticulum ^[9]	
RedOx1	CAD-like ADH		Cytosolic
RedOx2	Aldo-keto reductase		Cytosolic
SAT	Acetyltransferase		Cell membrane
PAS	FAD-dependent berberine bridge enzyme	Endoplasmic reticulum, moving to vesicles ^[2]	
DPAS	CAD-like ADH	Nucleocytosolic ^[2]	
TS	α/β hydrolase	Nucleocytosolic ^[2]	
CS	α/β hydrolase	Nucleocytosolic ^[2]	
CorS	α/β hydrolase		Cytosolic
T16H2	Cytochrome P450 type 71		Endoplasmic reticulum
16OMT	Methyl transferase		Cytosolic
T3O	Cytochrome P450 type 71		Endoplasmic reticulum
T3R	CAD-like ADH		Plastid

Appendix IX. Split-Luciferase Assays of *C. roseus* MIA Biosynthetic Enzymes

Representative split-luciferase result of testing protein-protein interactions of *C. roseus* MIA biosynthetic enzymes in heterologous host *N. benthamiana*.





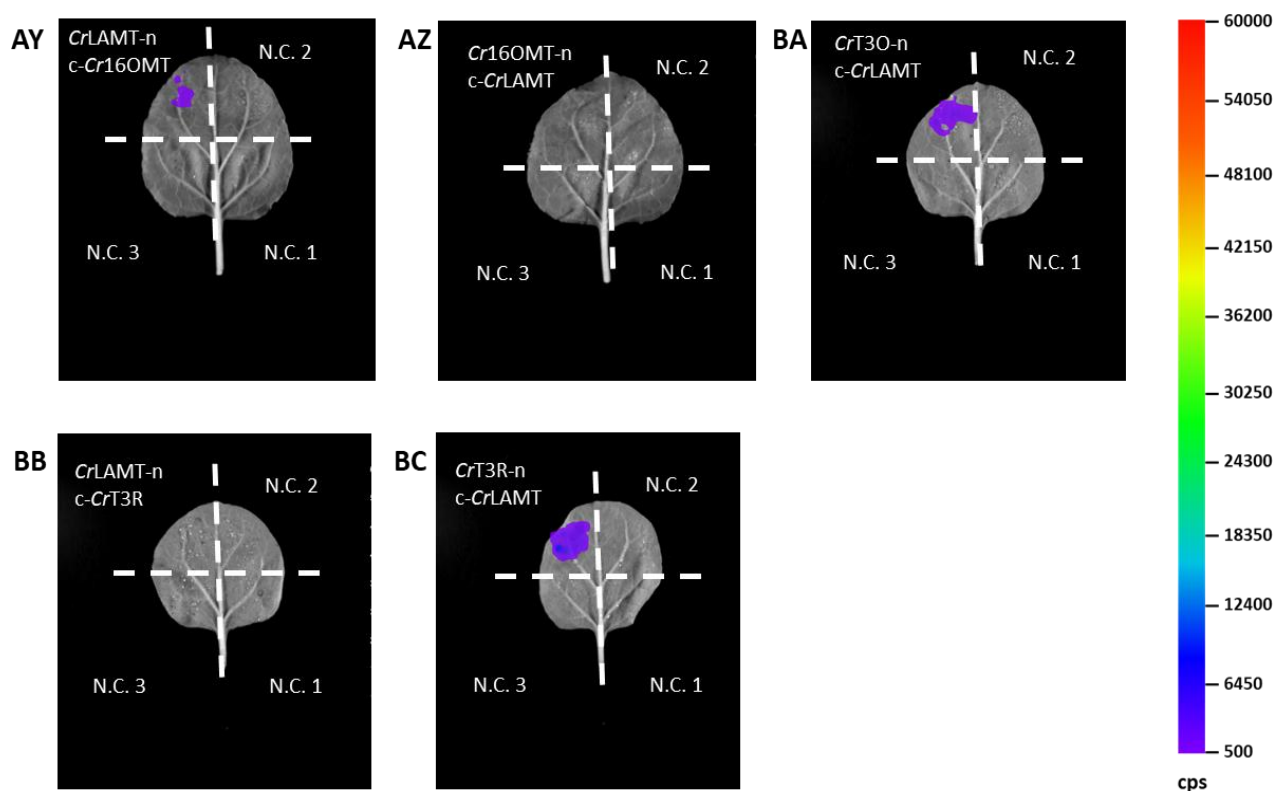
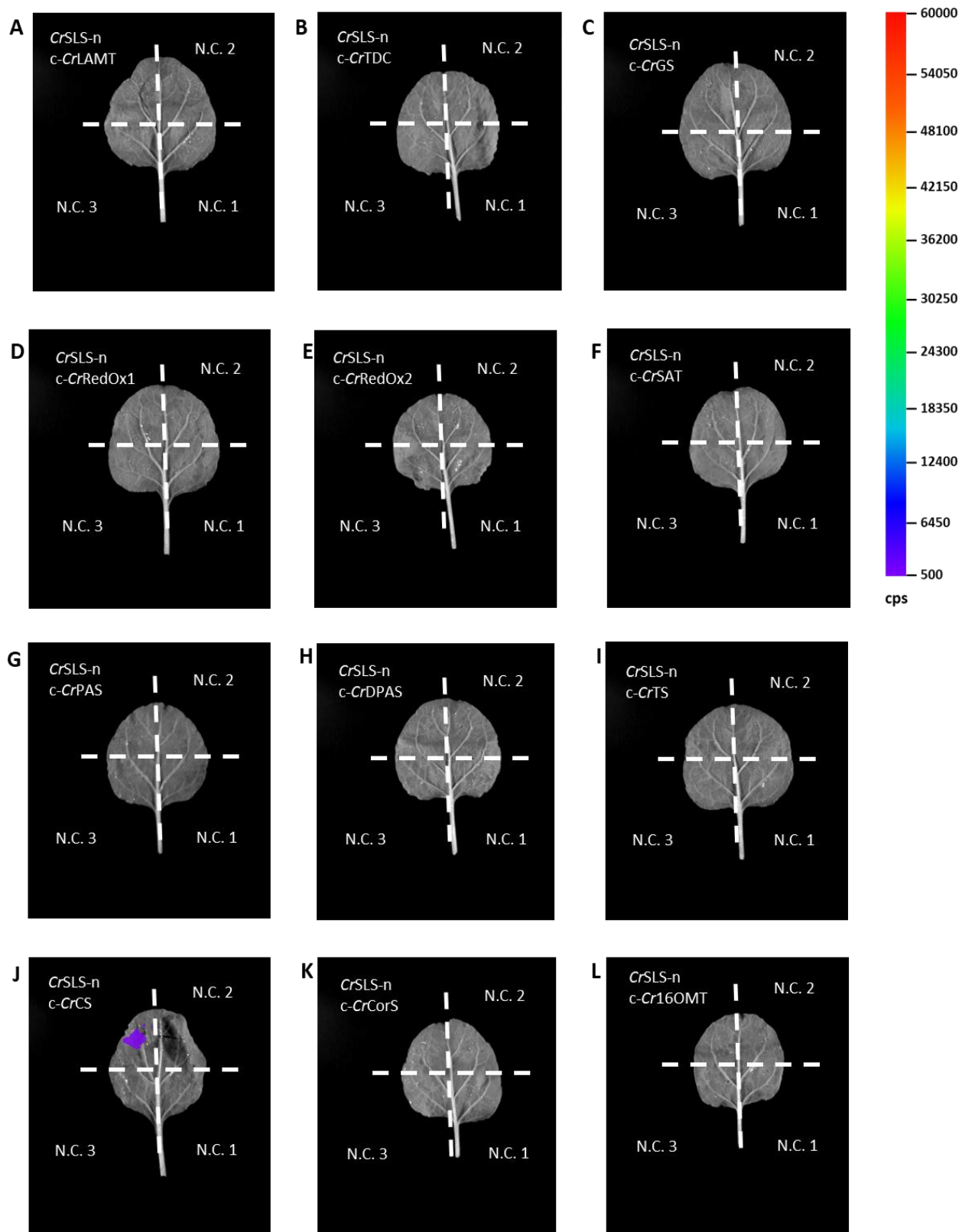


Figure 101. Representative images of pairwise interactions of *CrLAMT* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA.** LAMT-LAMT; **AB.** SLS-LAMT; **AC-D.** SLS-TDC; **AE-F.** SLS-GS; **AG.** GO-LAMT; **AH-I.** LAMT-RedOx1; **AJ-K.** LAMT-RedOx2; **AL-M.** LAMT-SAT; **AN-O.** LAMT-PAS; **AP-Q.** LAMT-DPAS; **AR-S.** LAMT-TS; **AT-U.** LAMT-CS; **AV-W.** LAMT-CorS; **AX.** T16H2-LAMT; **AY-Z.** LAMT-16OMT; **BA.** T3O-LAMT; **BB-C.** LAMT-T3R.



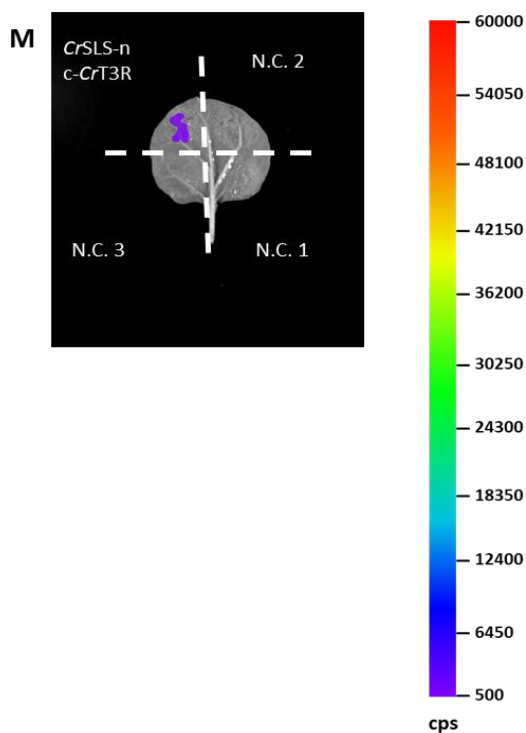
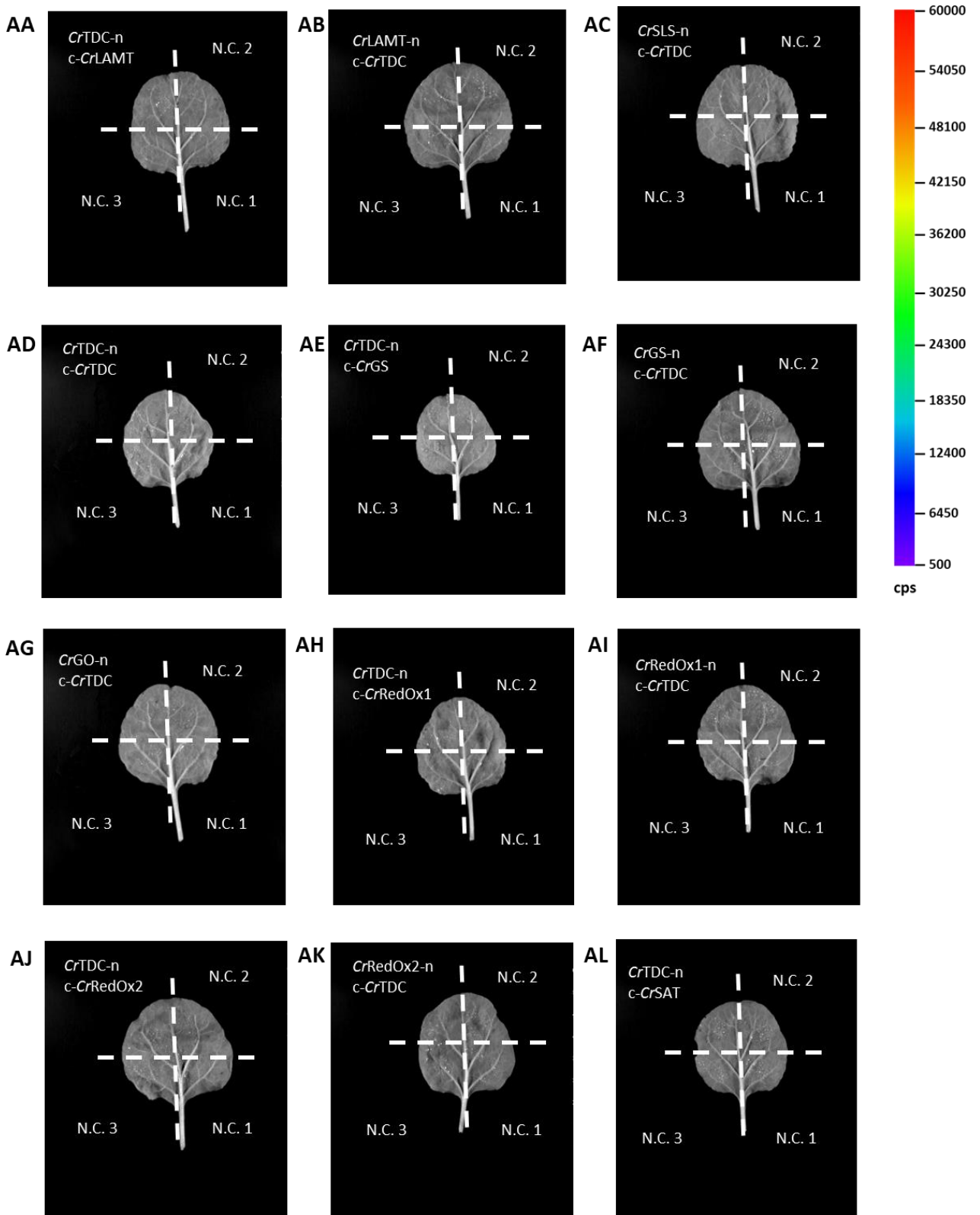
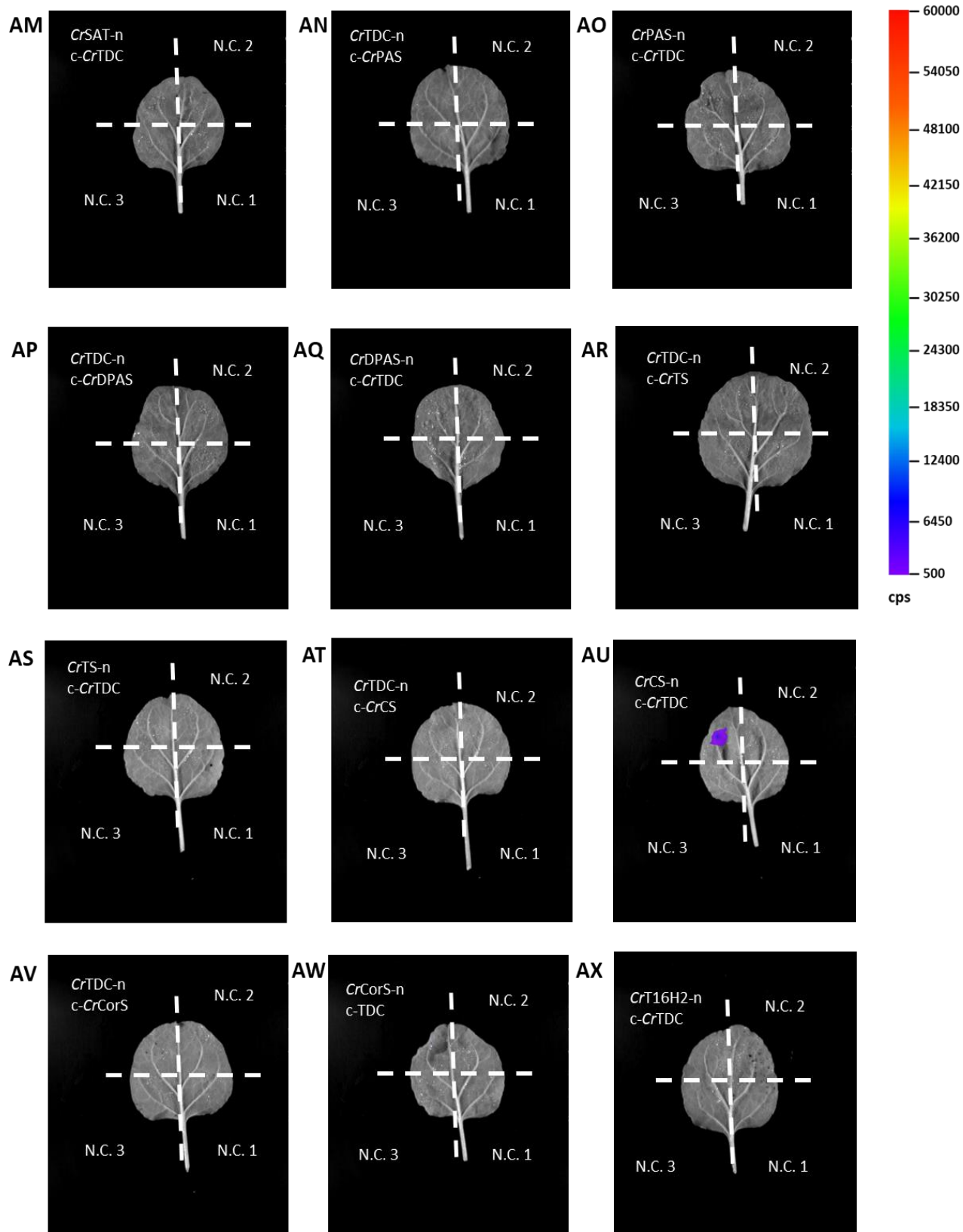


Figure 102. Representative images of pairwise interactions of *CrSLS* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A.** SLS-LAMT; **B.** SLS-TDC; **C.** SLS-GS; **D.** SLS-RedOx1; **E.** SLS-RedOx2; **F.** SLS-SAT; **G.** SLS-PAS; **H.** SLS-DPAS; **I.** SLS-TS; **J.** SLS-CS; **K.** SLS-CorS; **L.** SLS-16OMT; **M.** SLS-T3R.





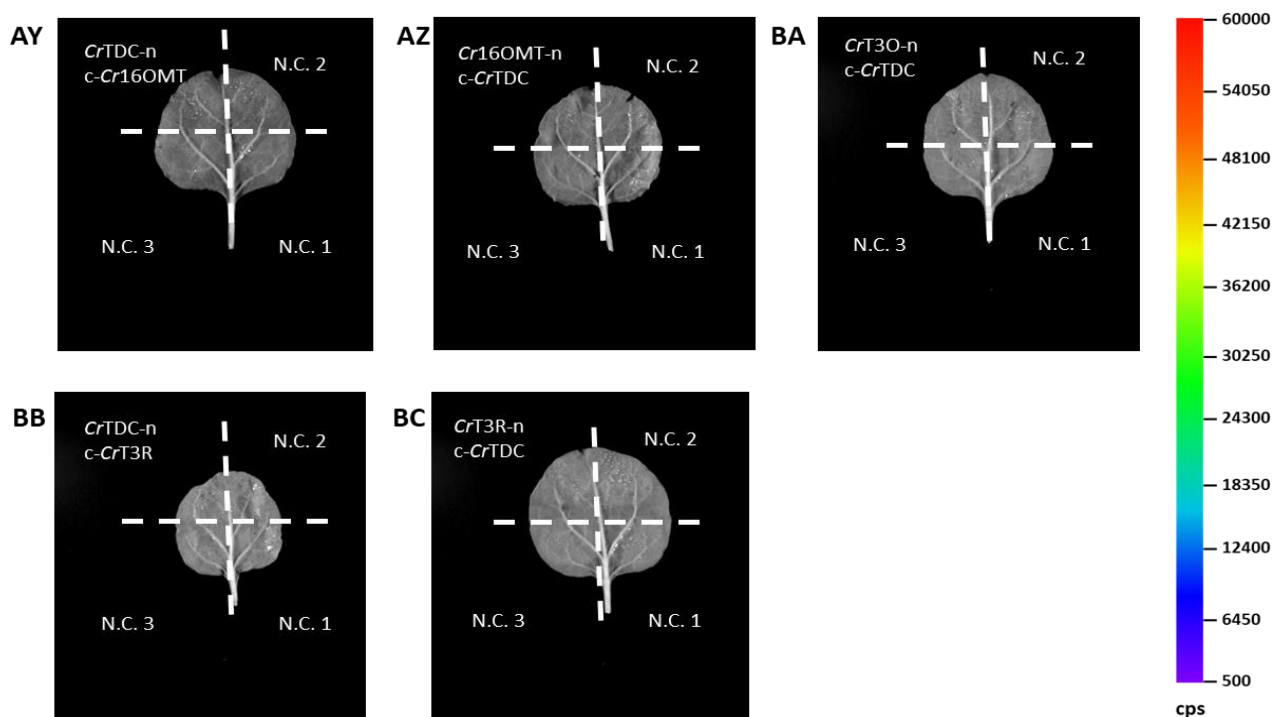
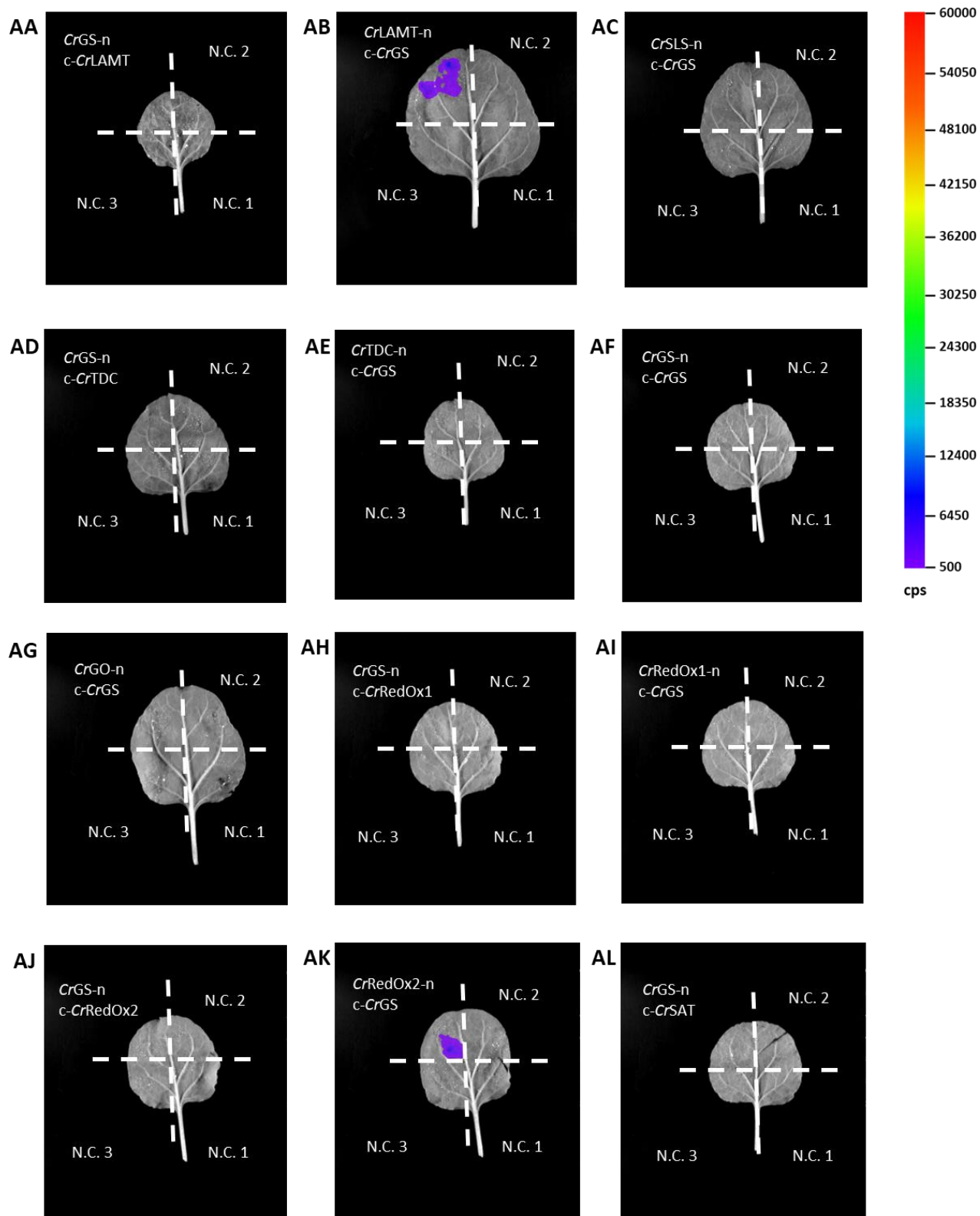
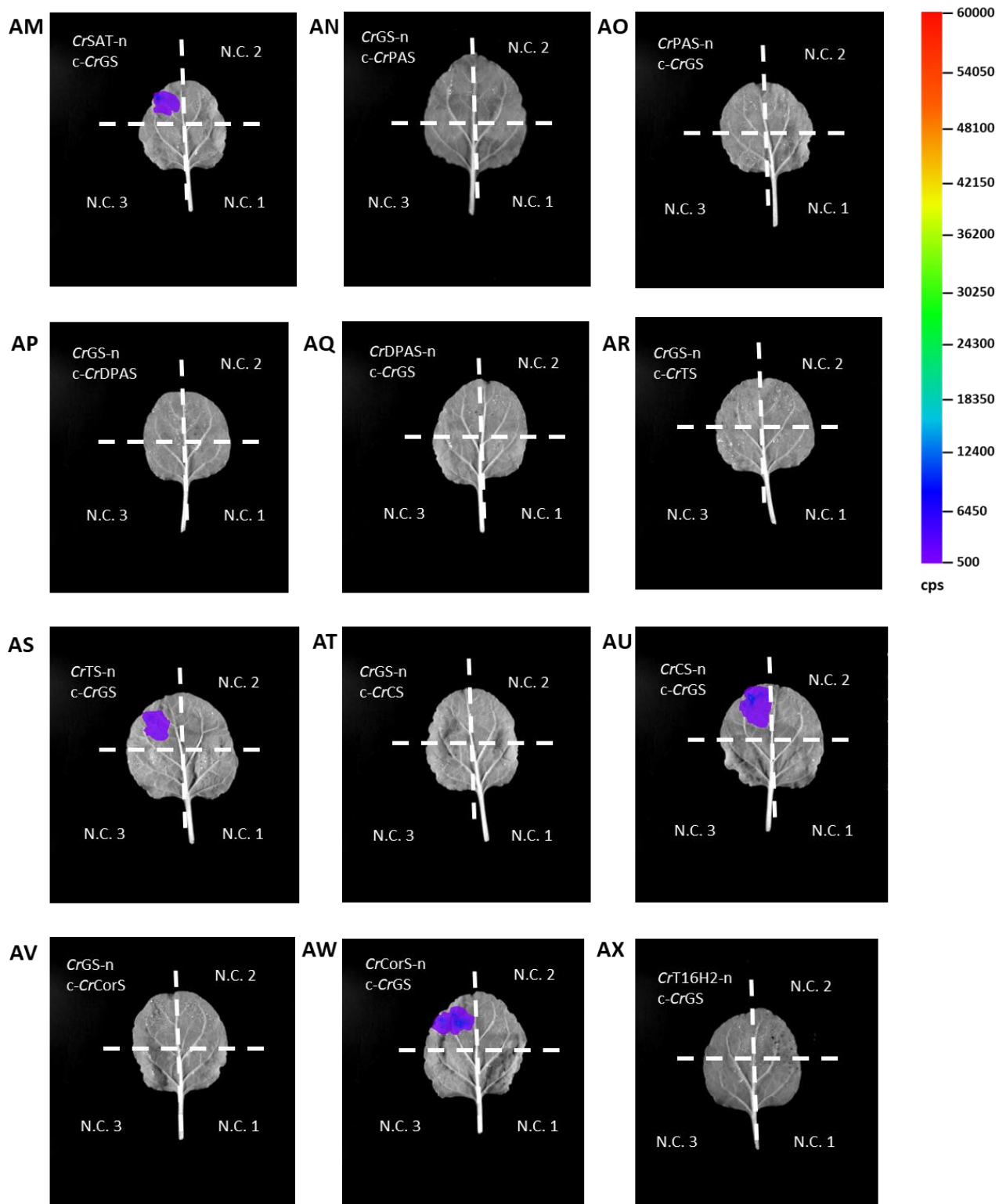


Figure 103. Representative images of pairwise interactions of *CrTDC* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** TDC-LAMT; **AC.** SLS-LAMT; **AD.** TDC-TDC; **AE-F.** TDC-GS; **AG.** GO-TDC; **AH-I.** TDC-RedOx1; **AJ-K.** TDC-RedOx2; **AL-M.** TDC-SAT; **AN-O.** TDC-PAS; **AP-Q.** TDC-DPAS; **AR-S.** TDC-TS; **AT-U.** TDC-CS; **AV-W.** TDC-CorS; **AX.** T16H2-TDC; **AY-Z.** TDC-16OMT; **BA.** T30-TDC; **BB-C.** TDC-T3R.





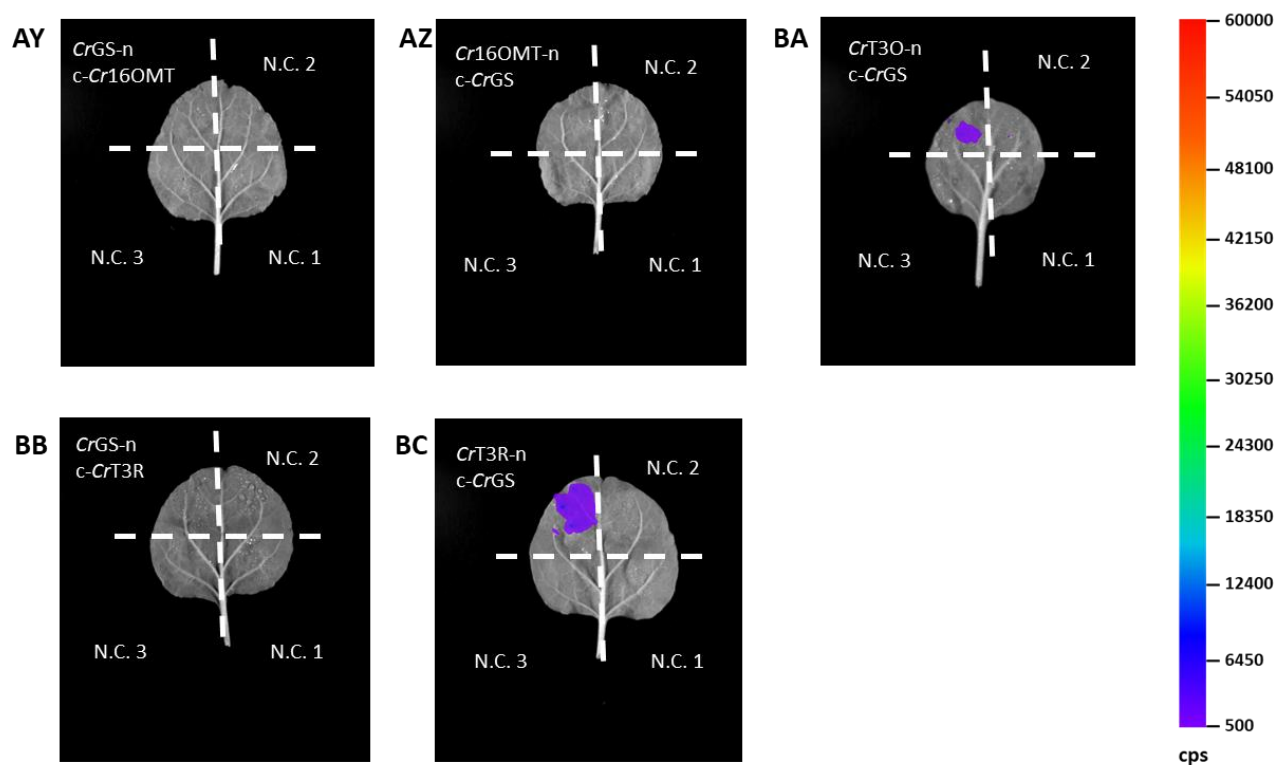
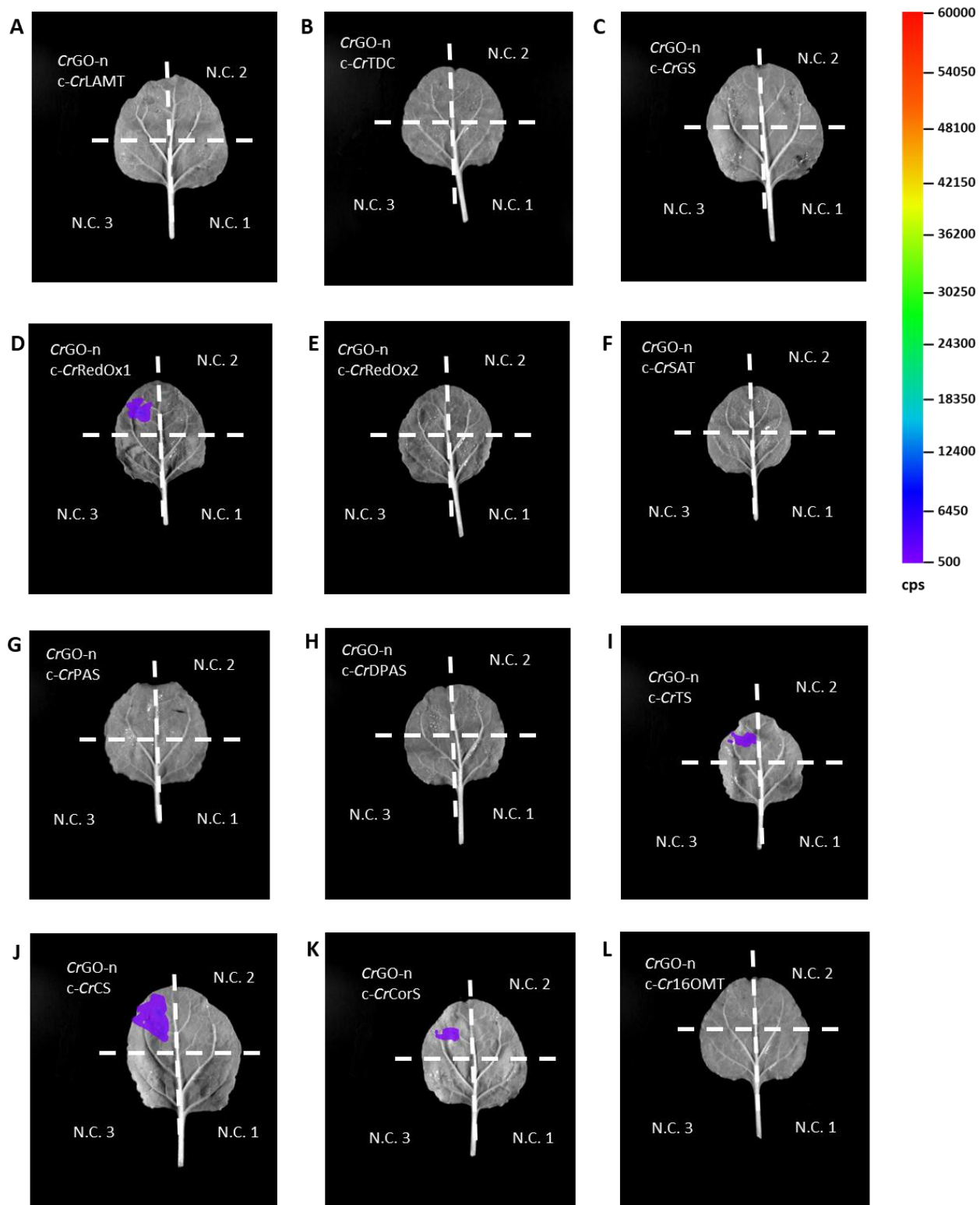


Figure 104. Representative images of pairwise interactions of *CrGS* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** GS-LAMT; **AC.** SLS-GS; **AD-E.** GS-TDC; **AF.** GS-GS; **AG.** GO-GS; **AH-I.** GS-RedOx1; **AJ-K.** GS-RedOx2; **AL-M.** GS-SAT; **AN-O.** GS-PAS; **AP-Q.** GS-DPAS; **AR-S.** GS-TS; **AT-U.** GS-CS; **AV-W.** GS-CorS; **AX.** T16H2-GS; **AY-Z.** GS-16OMT; **BA.** T30-GS; **BB-C.** GS-T3R.



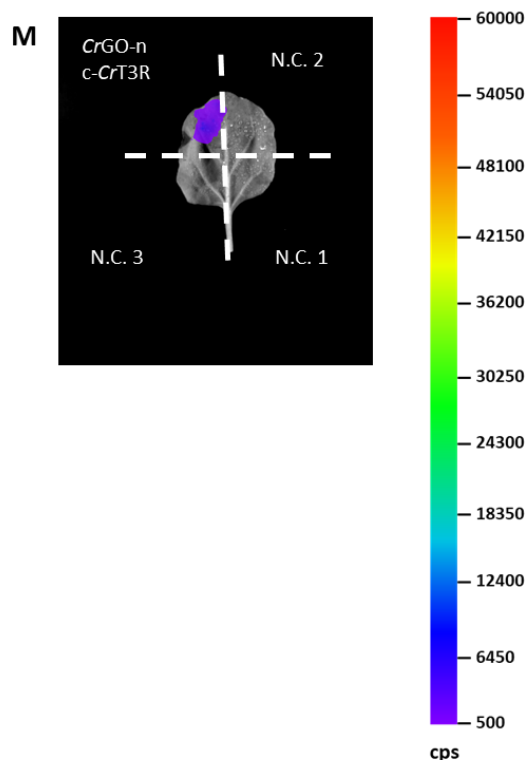
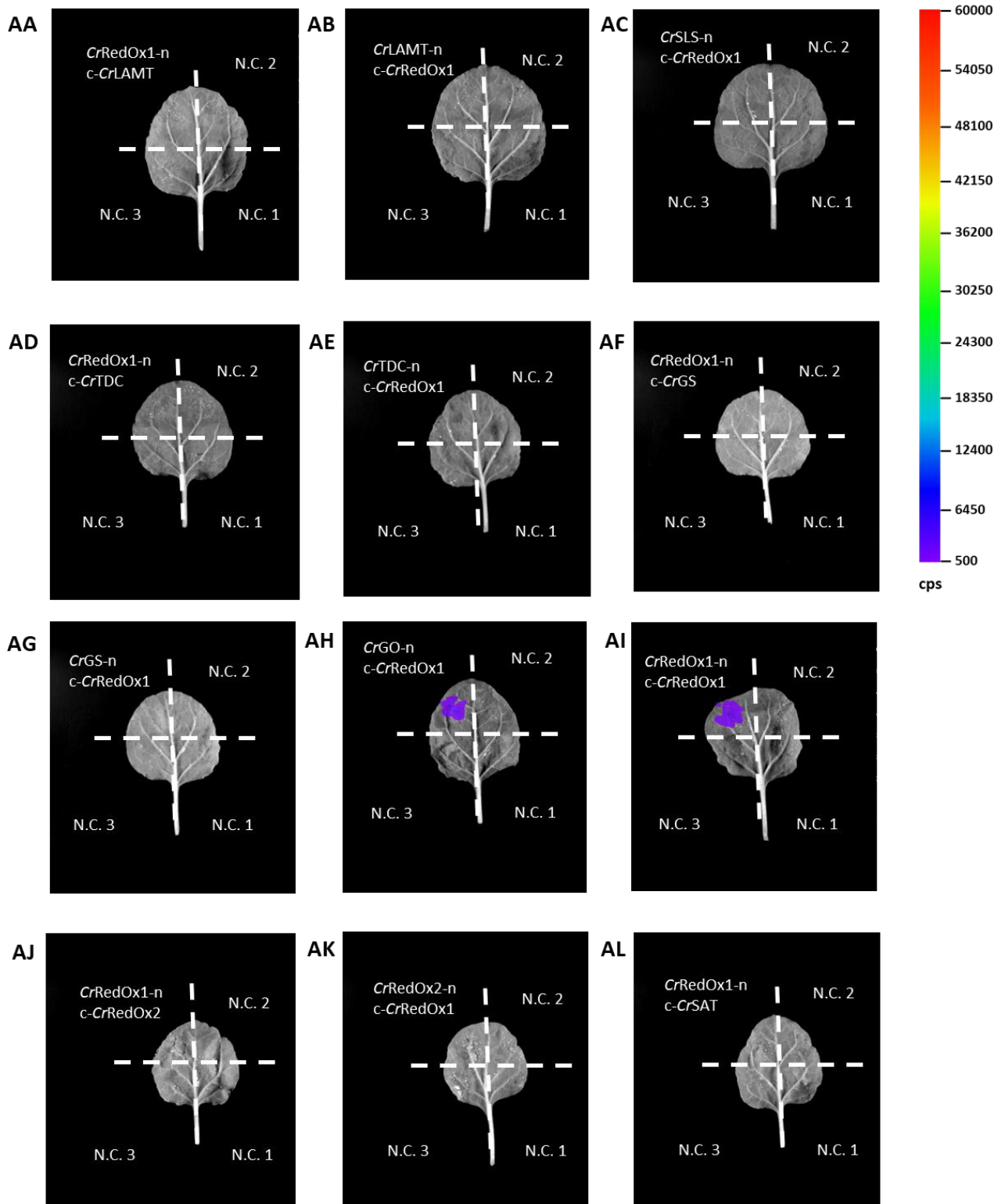
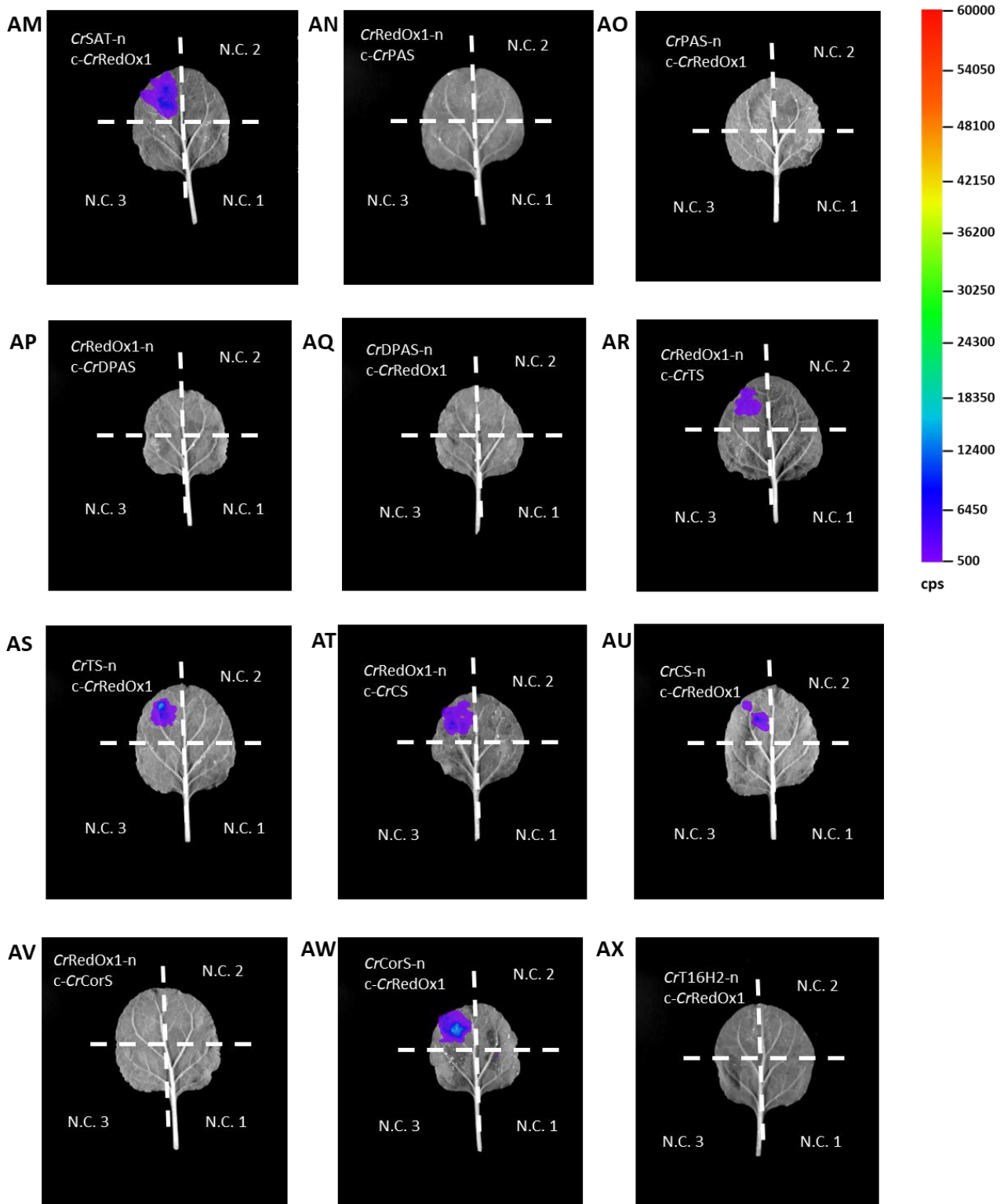


Figure 105. Representative images of pairwise interactions of *CrGO* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A.** GO-LAMT; **B.** GO-TDC; **C.** GO-GS; **D.** GO-RedOx1; **E.** GO-RedOx2; **F.** GO-SAT; **G.** GO-PAS; **H.** GO-DPAS; **I.** GO-TS; **J.** GO-CS; **K.** GO-CorS; **L.** GO-16OMT; **M.** GO-T3R.





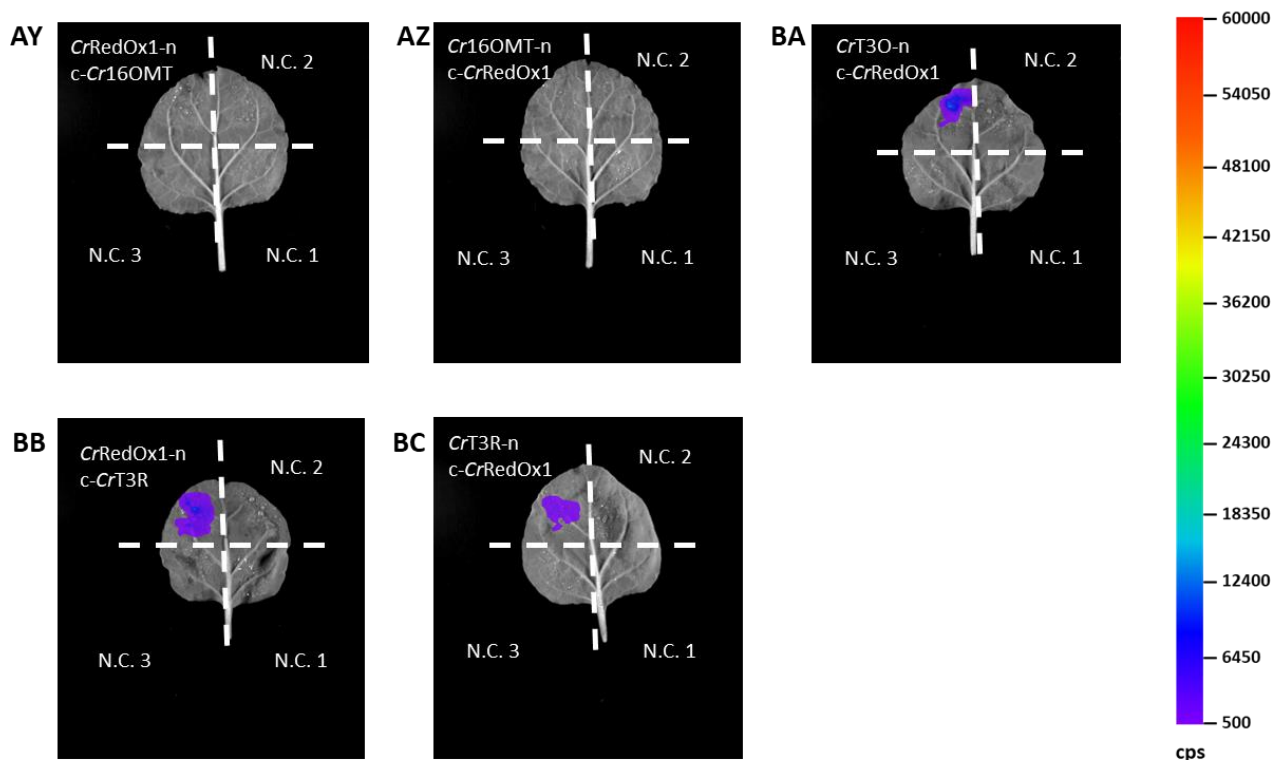
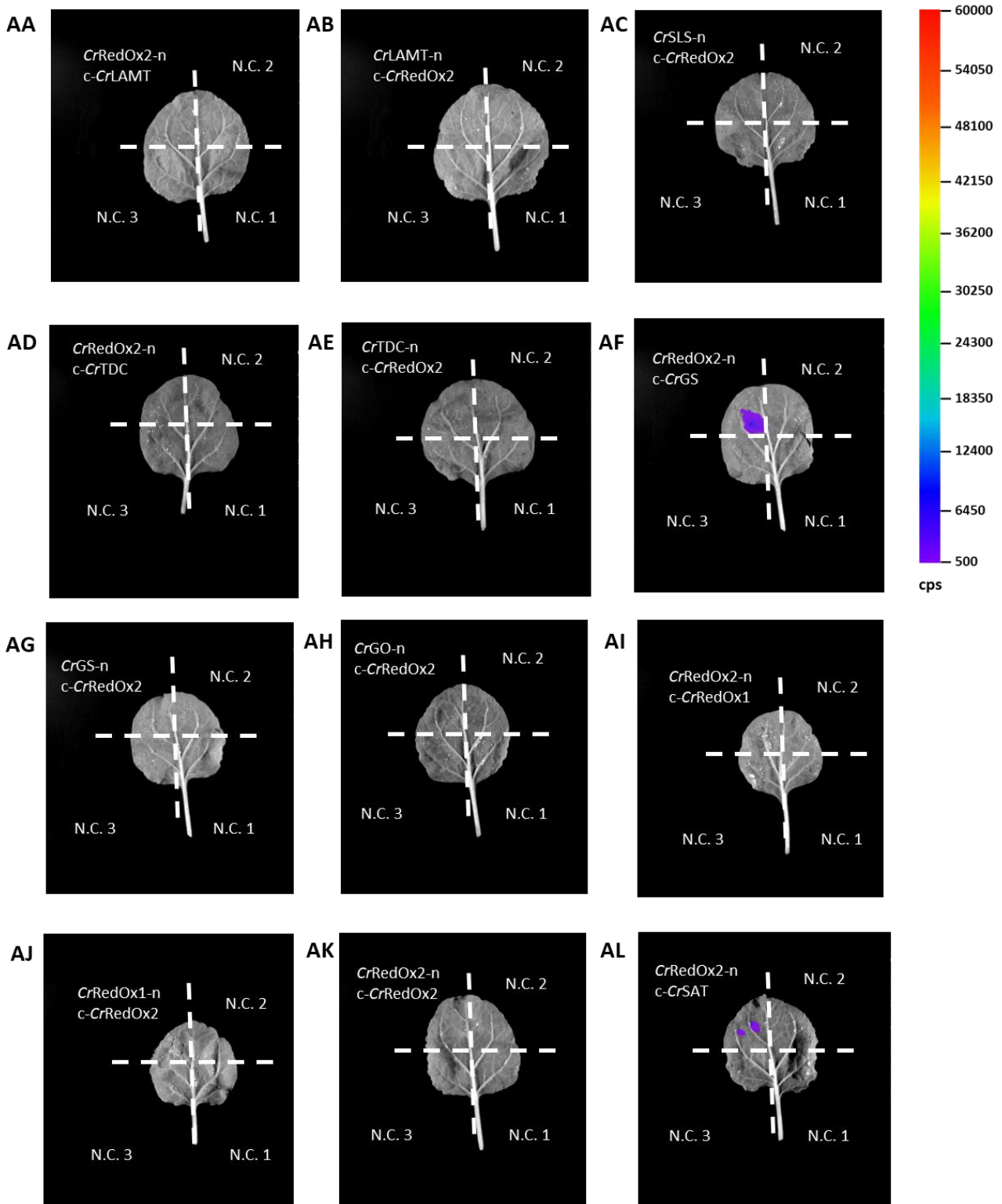
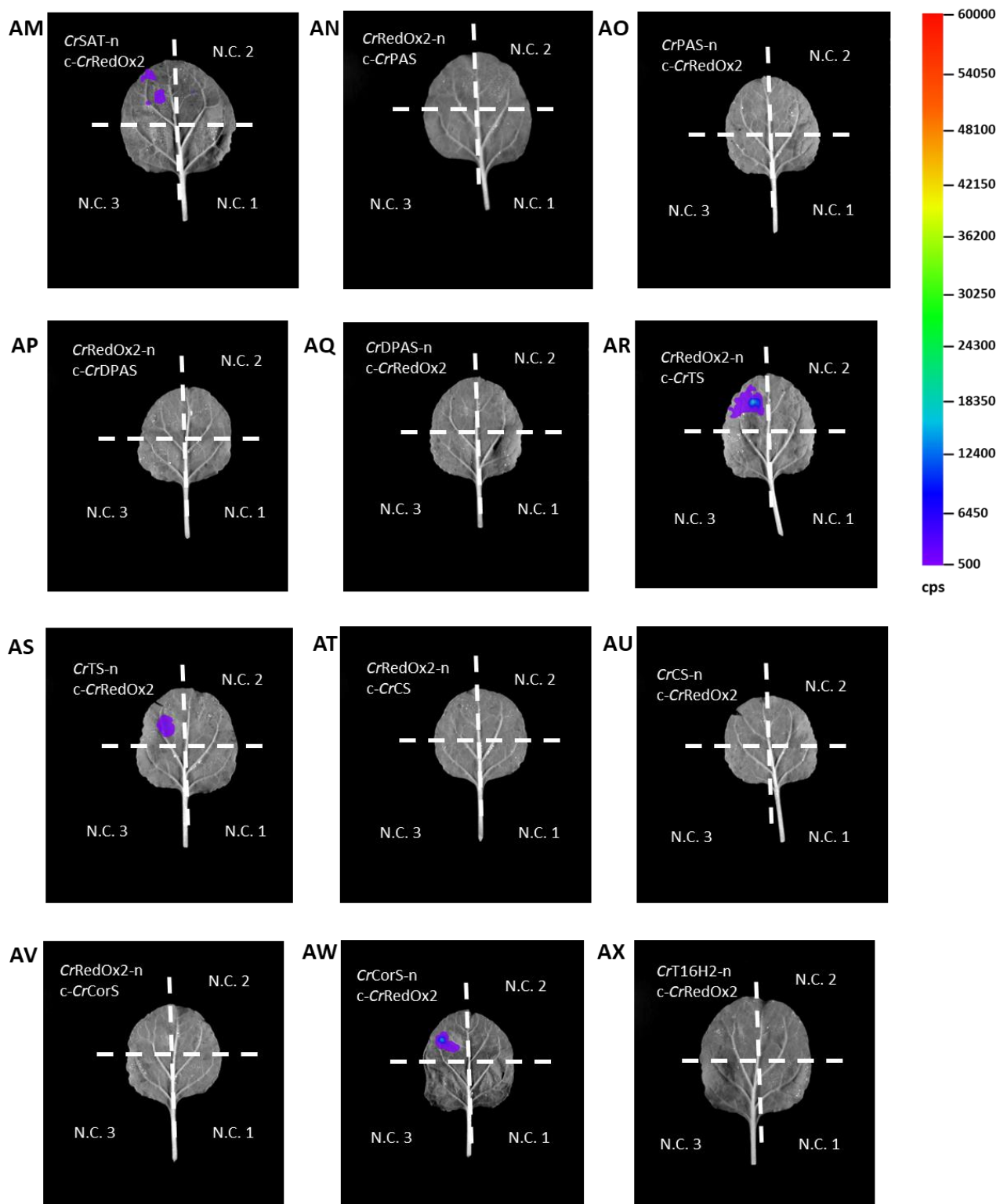


Figure 106. Representative images of pairwise interactions of *CrRedOx1* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** RedOx1-LAMT; **AC.** SLS-RedOx1; **AD-E.** RedOx1-TDC; **AF-G.** RedOx1-GS; **AH.** GO-RedOx1; **AI.** RedOx1-RedOx1; **AJ-K.** RedOx1-RedOx2; **AL-M.** RedOx1-SAT; **AN-O.** RedOx1-PAS; **AP-Q.** RedOx1-DPAS; **AR-S.** RedOx1-TS; **AT-U.** RedOx1-CS; **AV-W.** RedOx1-CorS; **AX.** T16H2-RedOx1; **AY-Z.** RedOx1-16OMT; **BA.** T3O-RedOx1; **BB-C.** RedOx1-T3R.





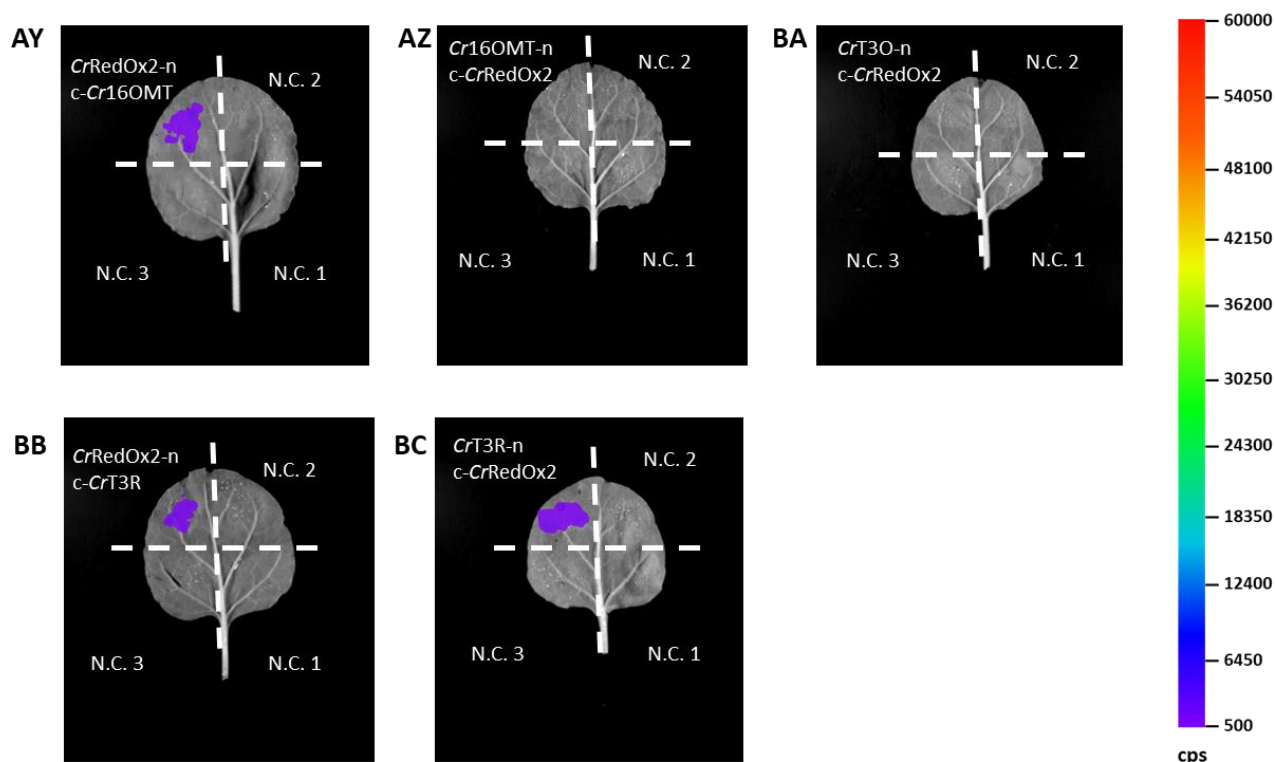
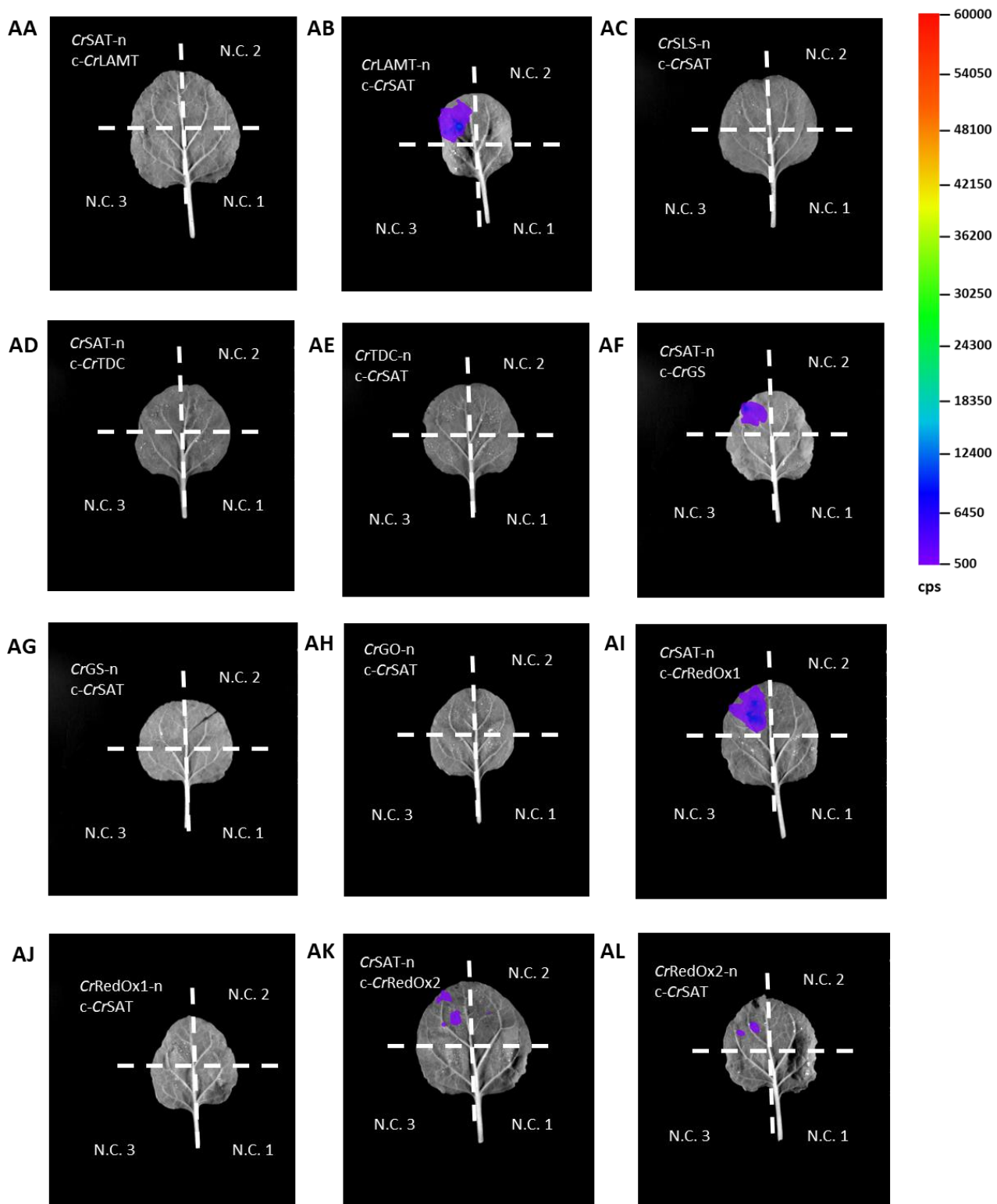
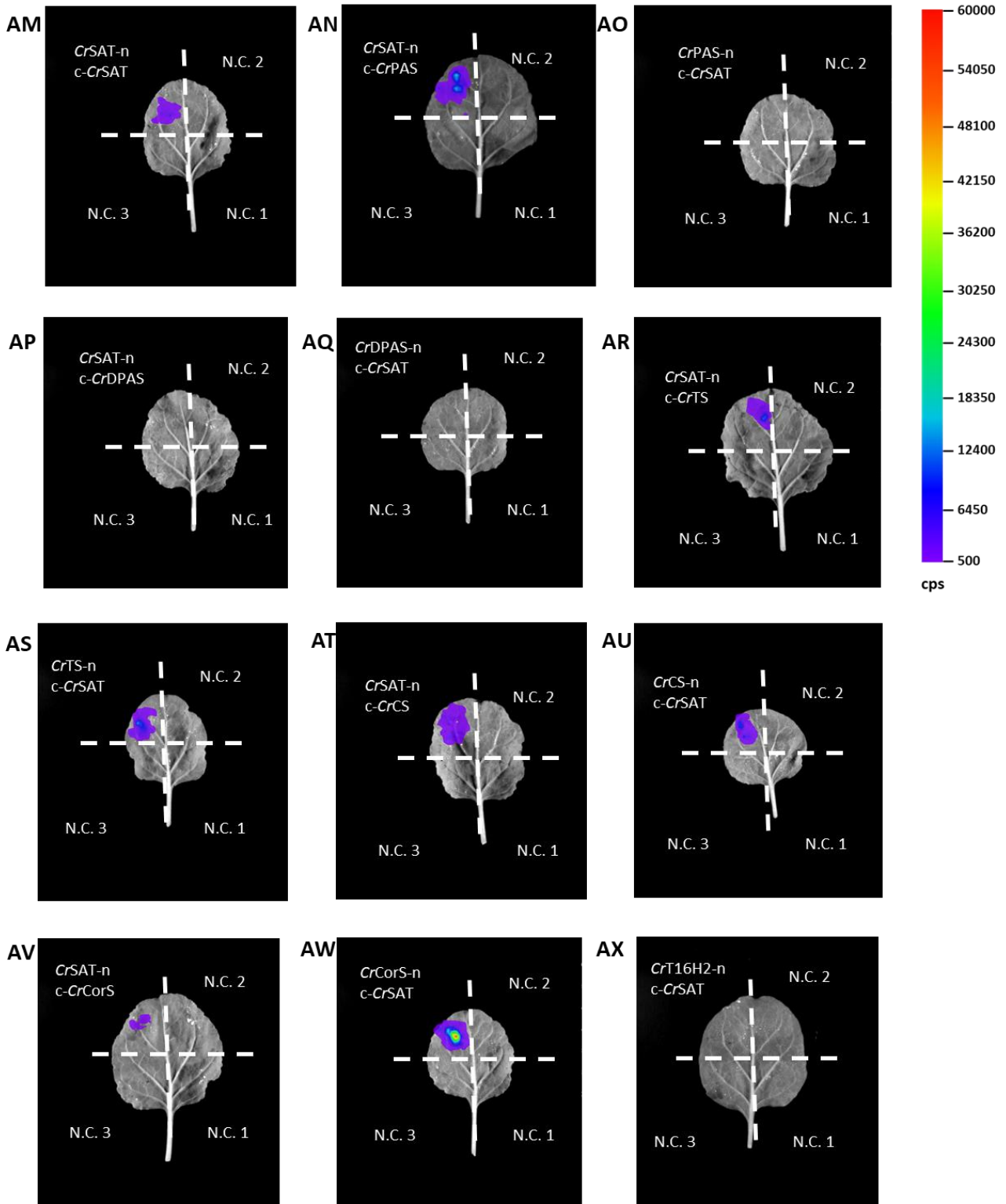


Figure 107. Representative images of pairwise interactions of *CrRedOx2* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** RedOx2-LAMT; **AC.** SLS-RedOx2; **AD-E.** RedOx2-TDC; **AF-G.** RedOx2-GS; **AH.** GO-RedOx2; **AI-J.** RedOx1-RedOx2; **AK.** RedOx2-RedOx2; **AL-M.** RedOx2-SAT; **AN-O.** RedOx2-PAS; **AP-Q.** RedOx2-DPAS; **AR-S.** RedOx2-TS; **AT-U.** RedOx2-CS; **AV-W.** RedOx2-CorS; **AX.** T16H2-RedOx2; **AY-Z.** RedOx2-16OMT; **BA.** T3O-RedOx2; **BB-C.** RedOx2-T3R.





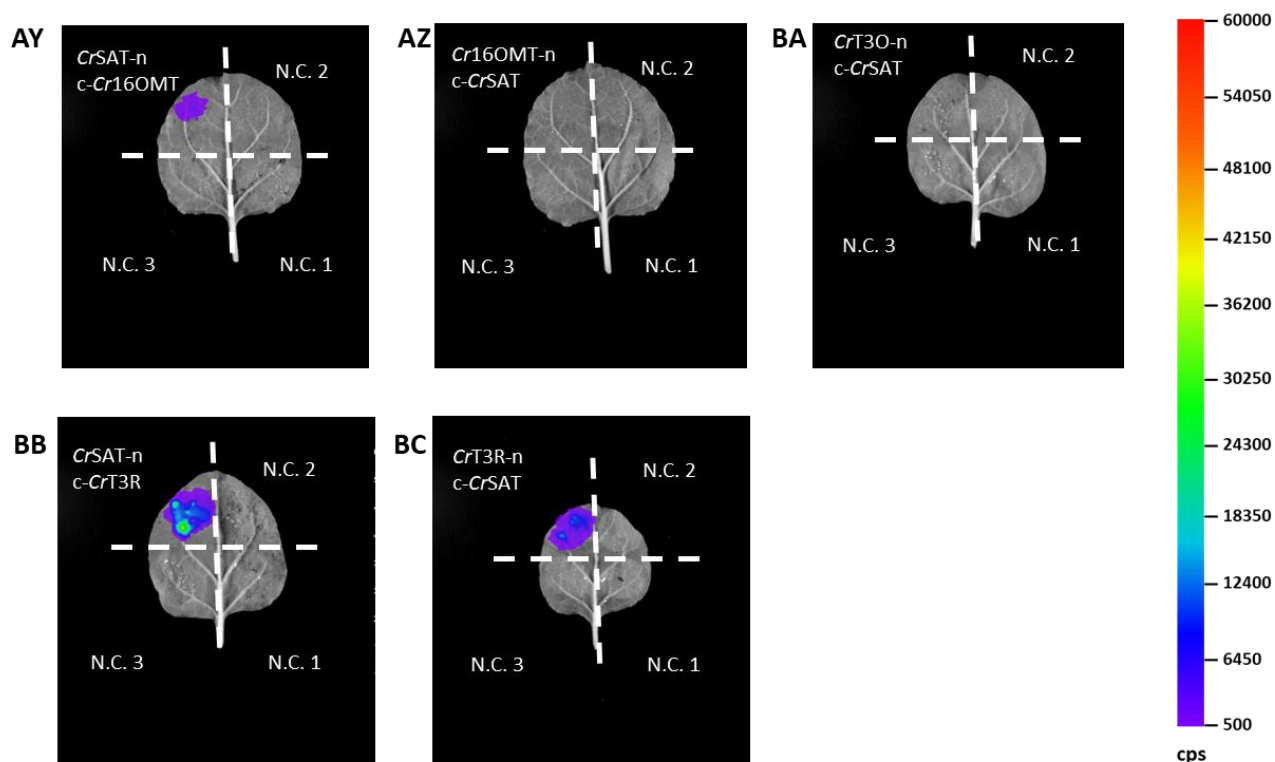
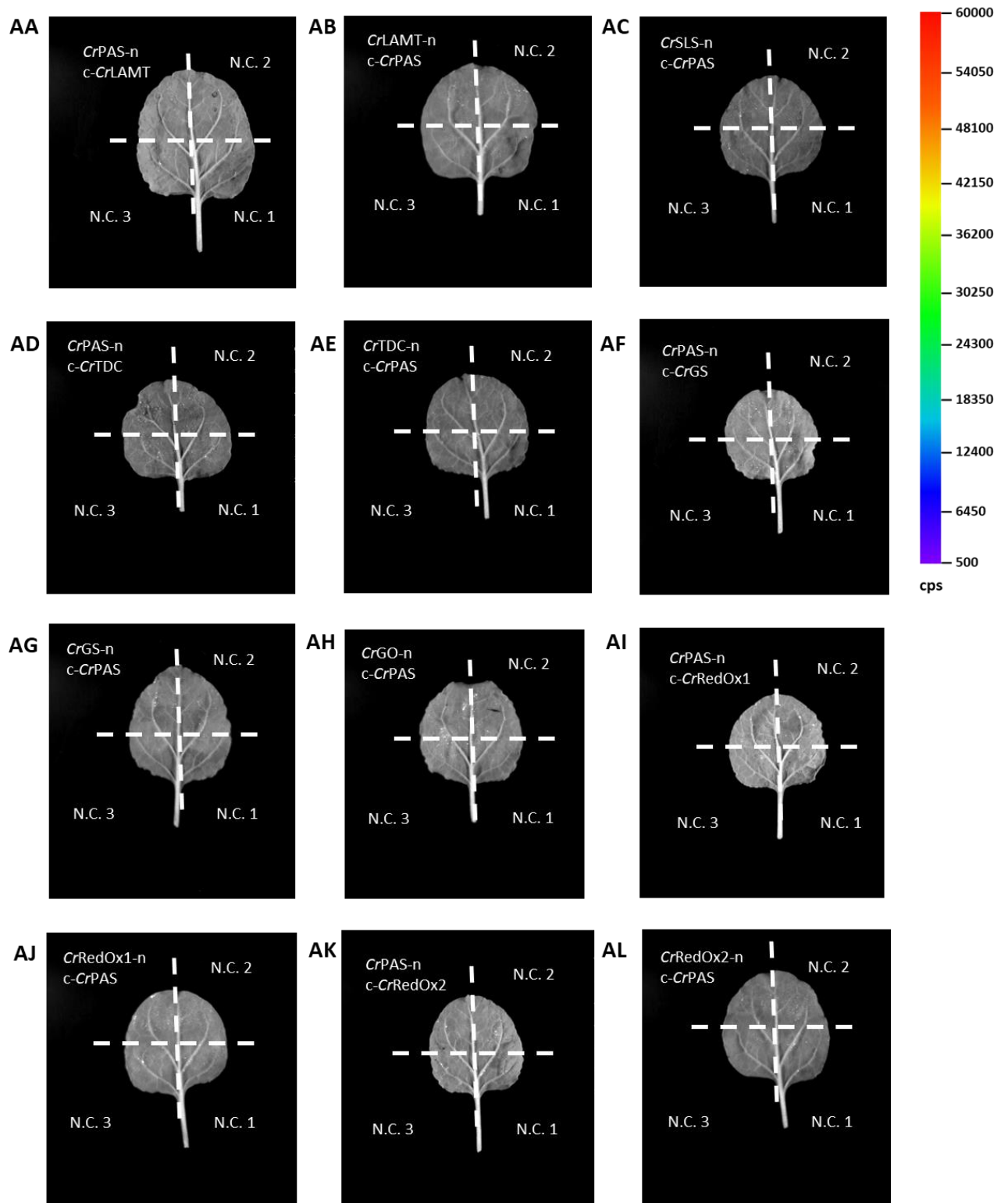
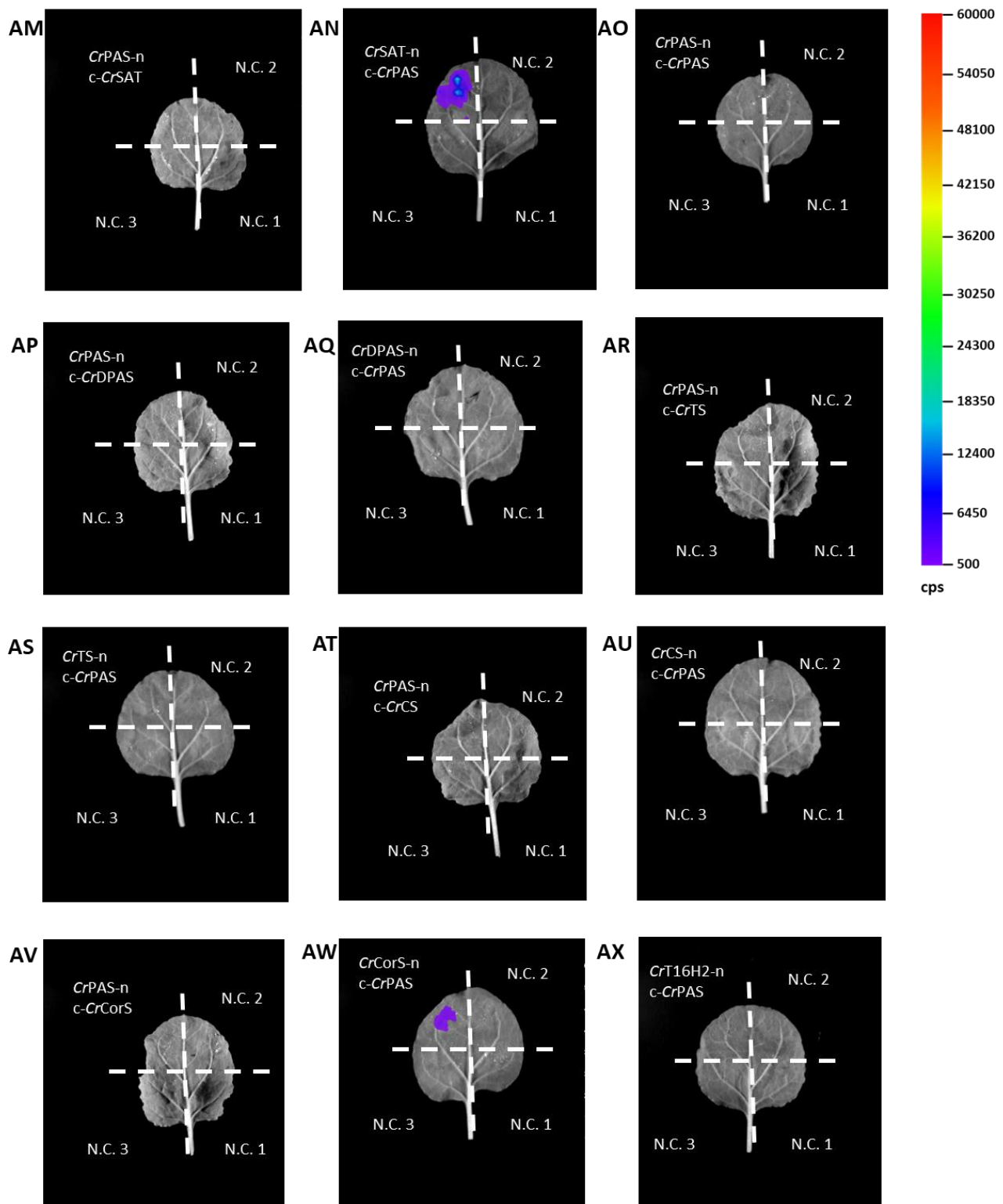


Figure 108. Representative images of pairwise interactions of *CrSAT* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** SAT-LAMT; **AC.** SLS-SAT; **AD-E.** SAT-TDC; **AF-G.** SAT-GS; **AH.** GO-SAT; **AI-J.** SAT-RedOx1; **AK-L.** SAT-RedOx2; **AM.** SAT-SAT; **AN-O.** SAT-PAS; **AP-Q.** SAT-DPAS; **AR-S.** SAT-TS; **AT-U.** SAT-CS; **AV-W.** SAT-CorS; **AX.** T16H2-SAT; **AY-Z.** SAT-16OMT; **BA.** T3O-SAT; **BB-C.** SAT-T3R.





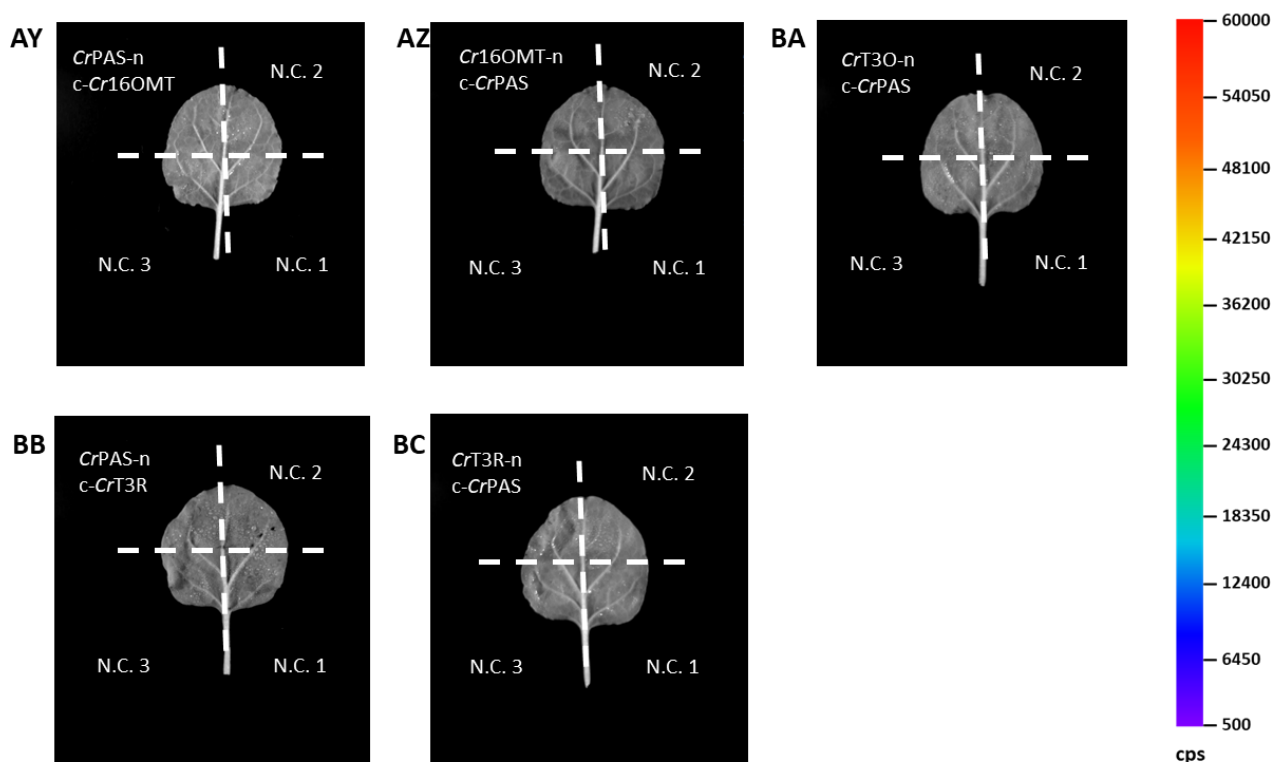
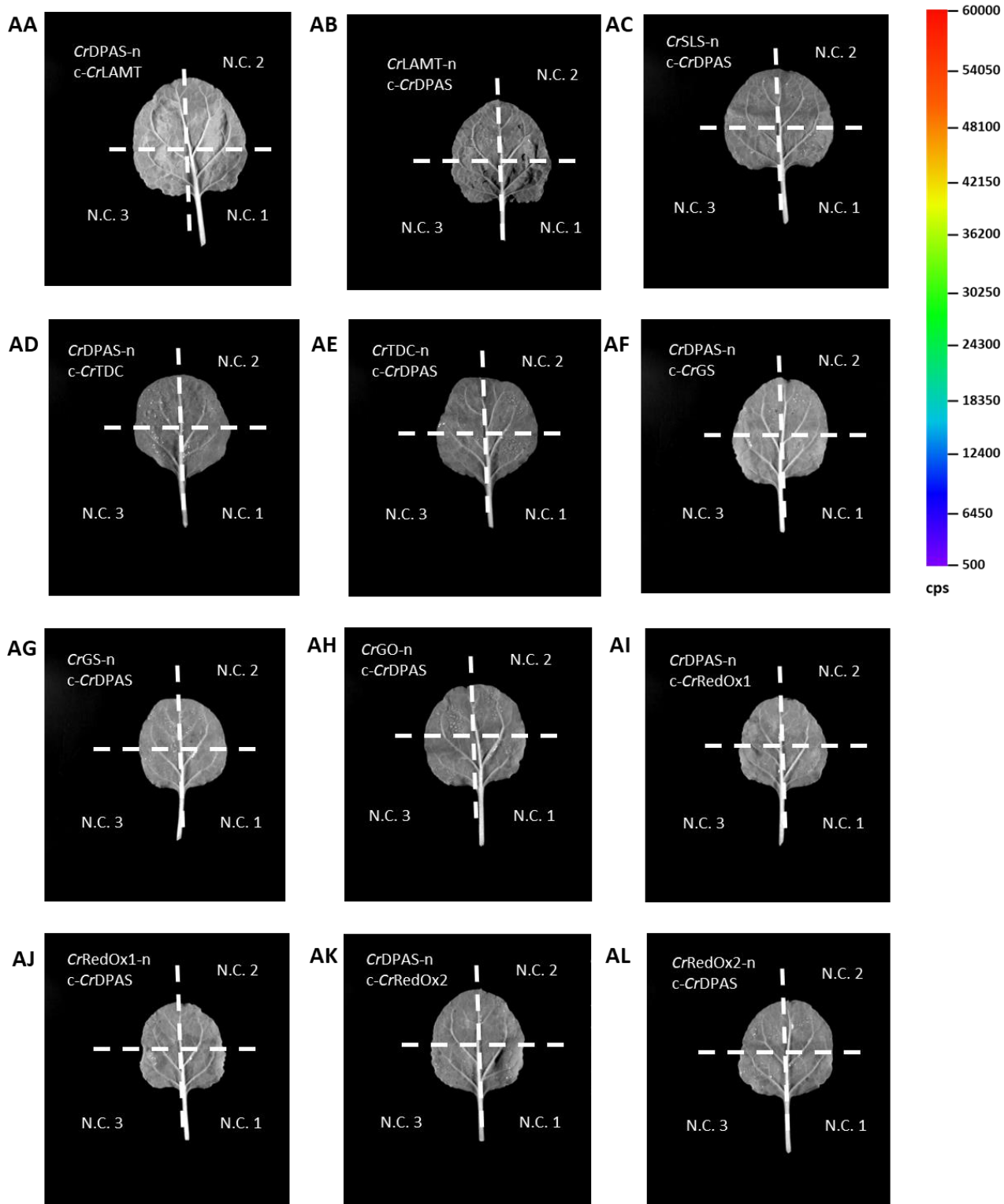
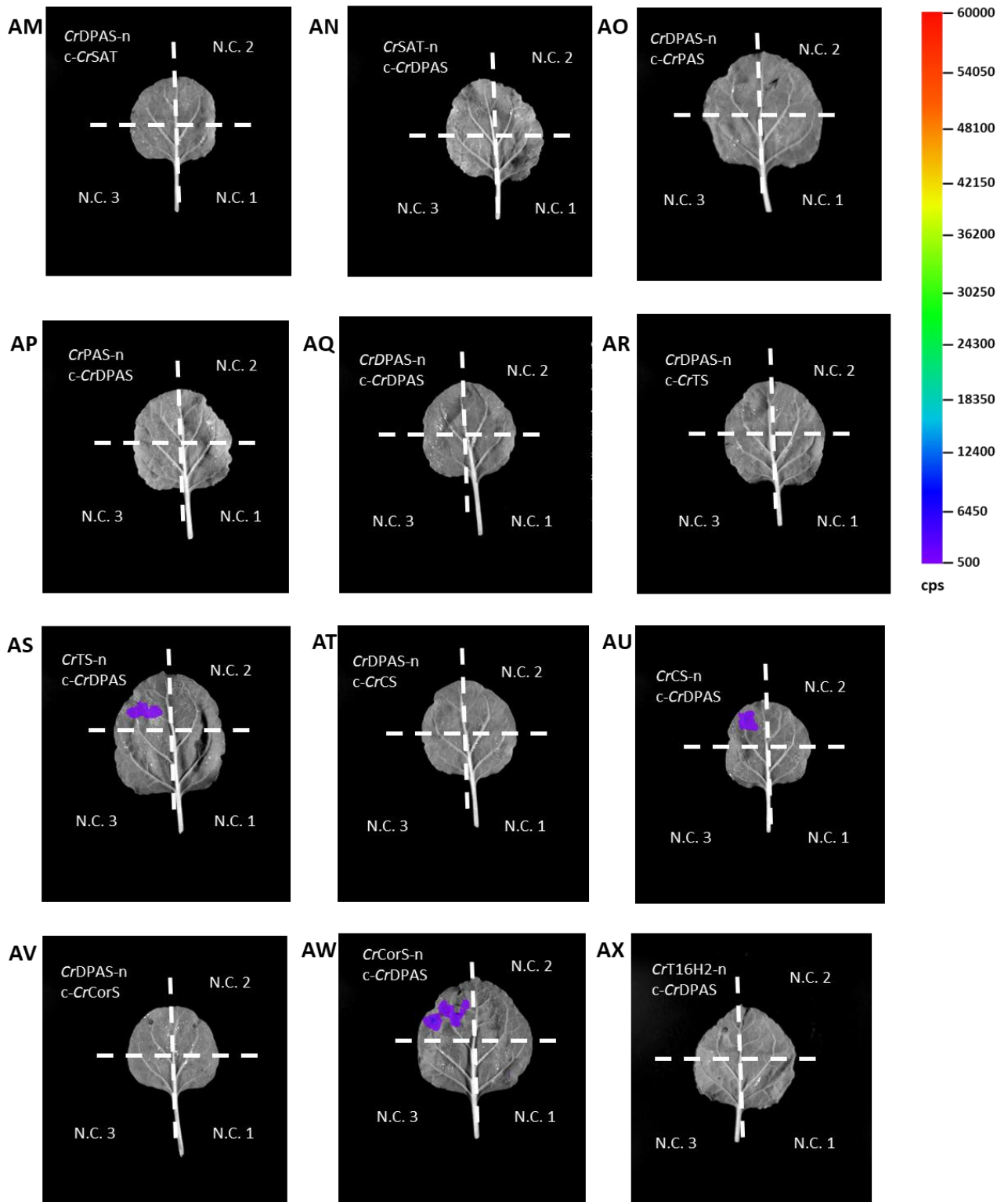


Figure 109. Representative images of pairwise interactions of *CrPAS* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** PAS-LAMT; **AC.** SLS-PAS; **AD-E.** PAS-TDC; **AF-G.** PAS-GS; **AH.** GO-PAS; **AI-J.** PAS-RedOx1; **AK-L.** PAS-RedOx2; **AM-N.** PAS-SAT; **AO.** PAS-PAS; **AP-Q.** PAS-DPAS; **AR-S.** PAS-TS; **AT-U.** PAS-CS; **AV-W.** PAS-CorS; **AX.** T16H2-PAS; **AY-Z.** PAS-16OMT; **BA.** T3O-PAS; **BB-C.** PAS-T3R.





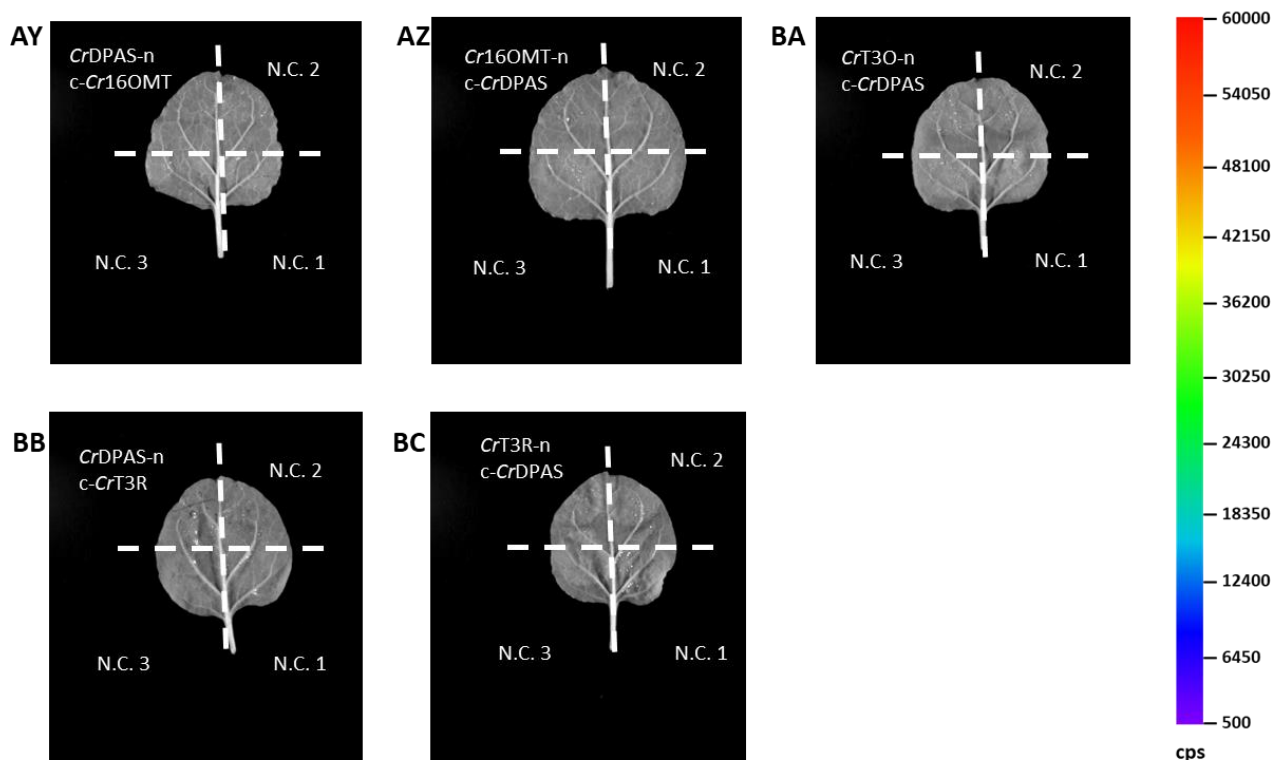
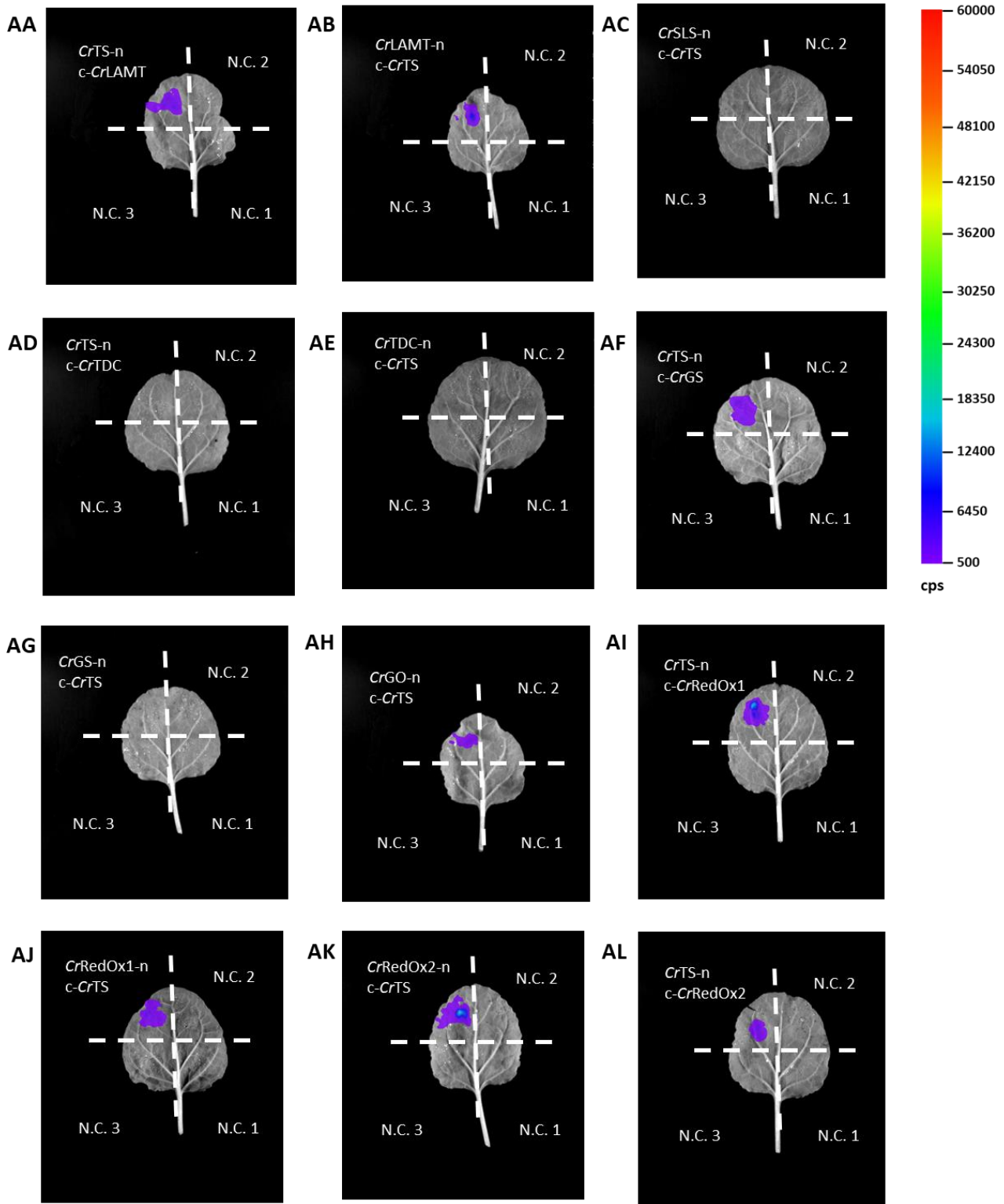
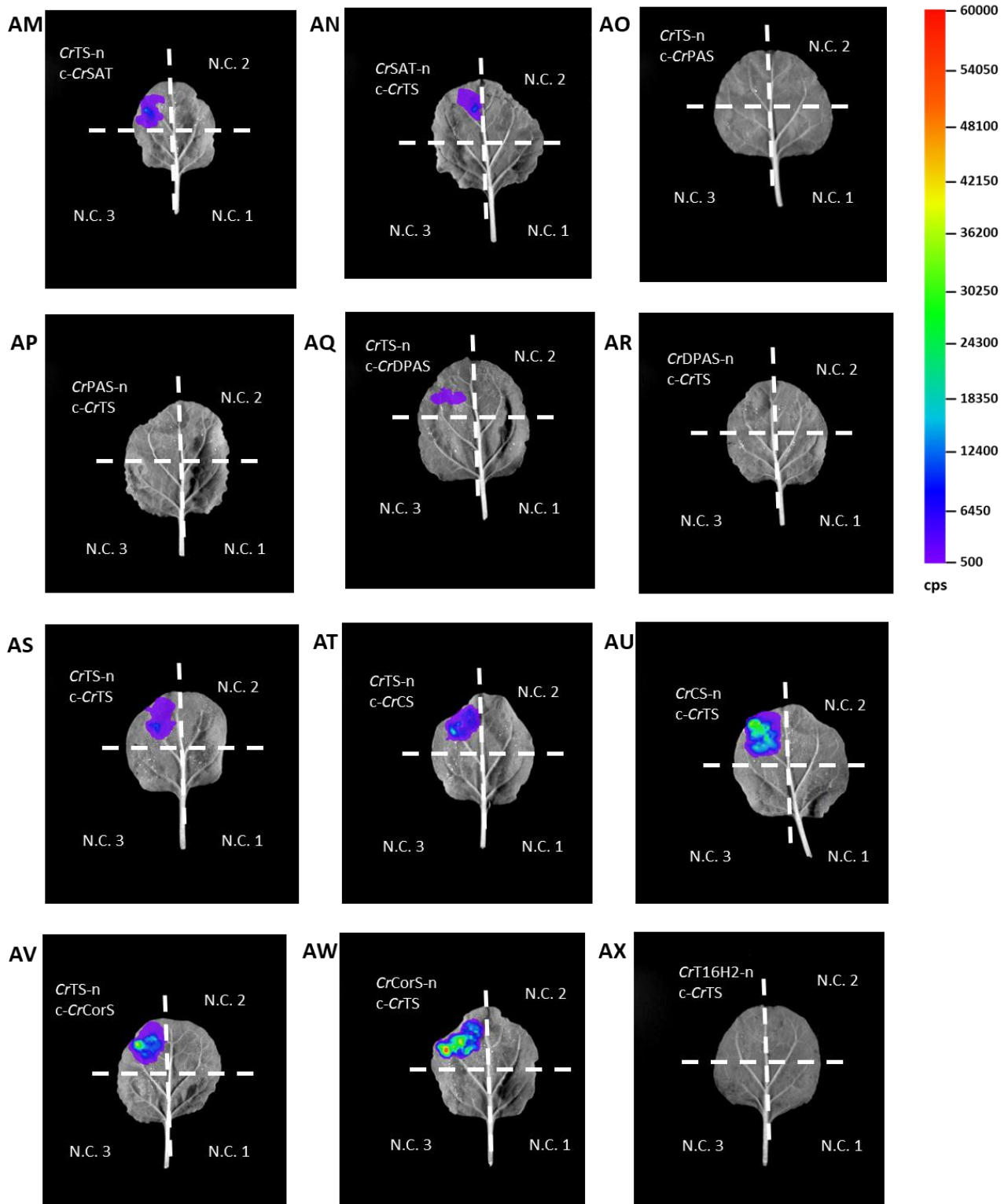


Figure 110. Representative images of pairwise interactions of *CrDPAS* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** DPAS-LAMT; **AC.** SLS-DPAS; **AD-E.** DPAS-TDC; **AF-G.** DPAS-GS; **AH.** GO-DPAS; **AI-J.** DPAS-RedOx1; **AK-L.** DPAS-RedOx2; **AM-N.** DPAS-SAT; **AO-P.** DPAS-PAS; **AQ.** DPAS-DPAS; **AR-S.** DPAS-TS; **AT-U.** DPAS-CS; **AV-W.** DPAS-CorS; **AX.** T16H2-DPAS; **AY-Z.** DPAS-16OMT; **BA.** T3O-DPAS; **BB-C.** DPAS-T3R.





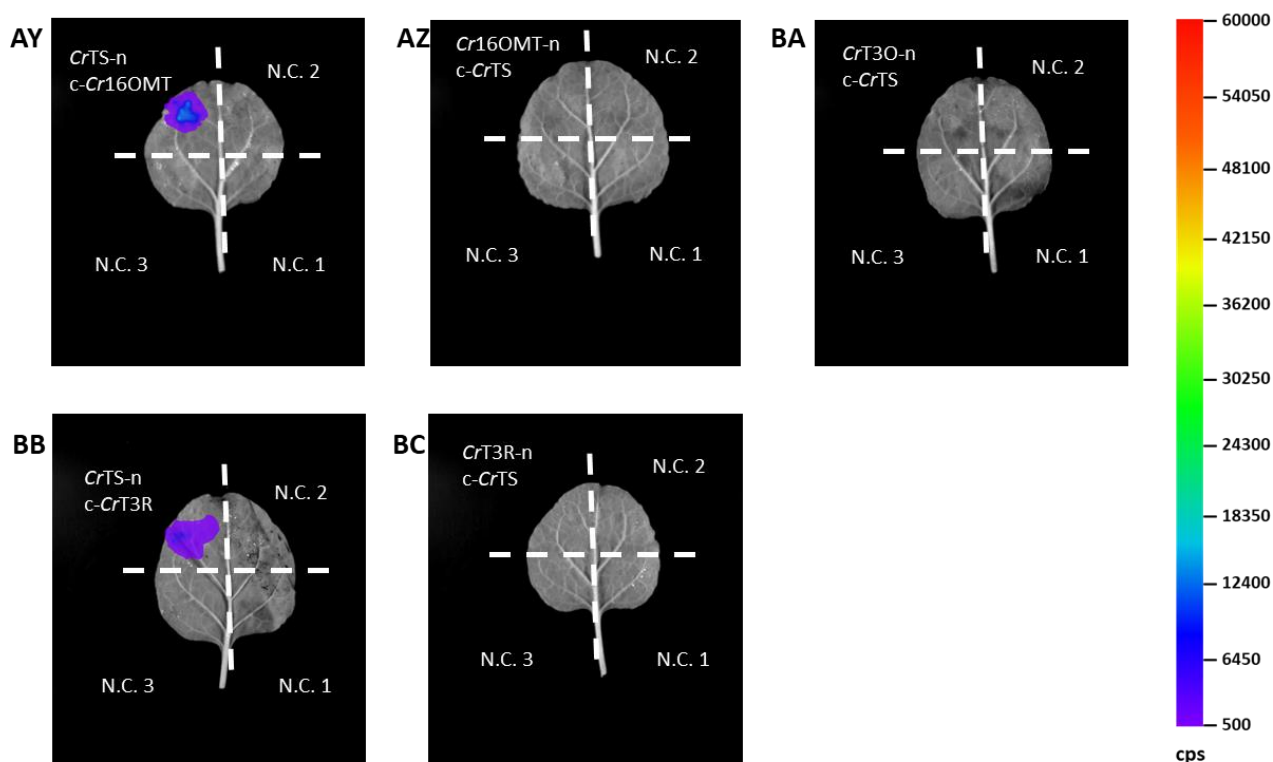
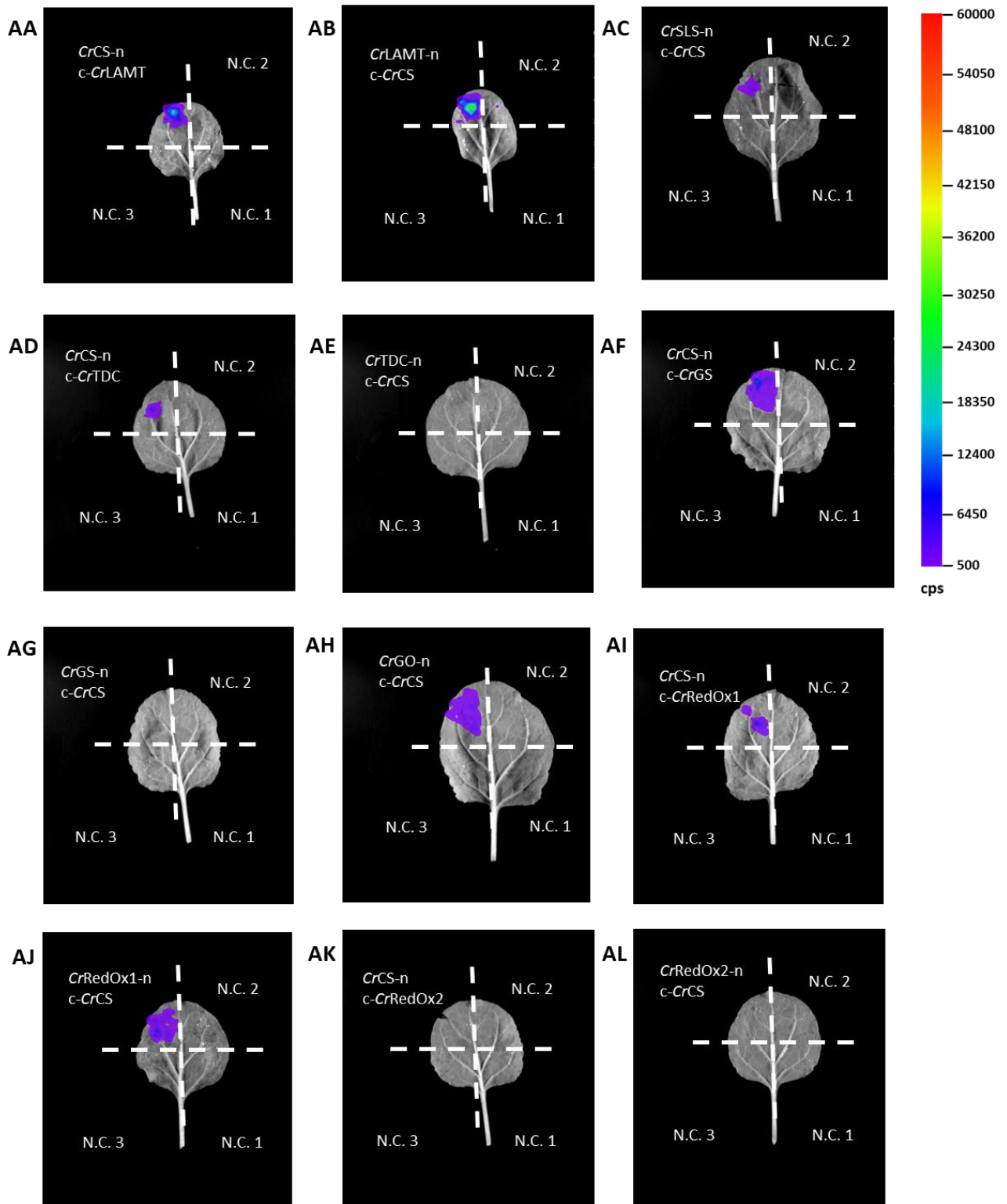
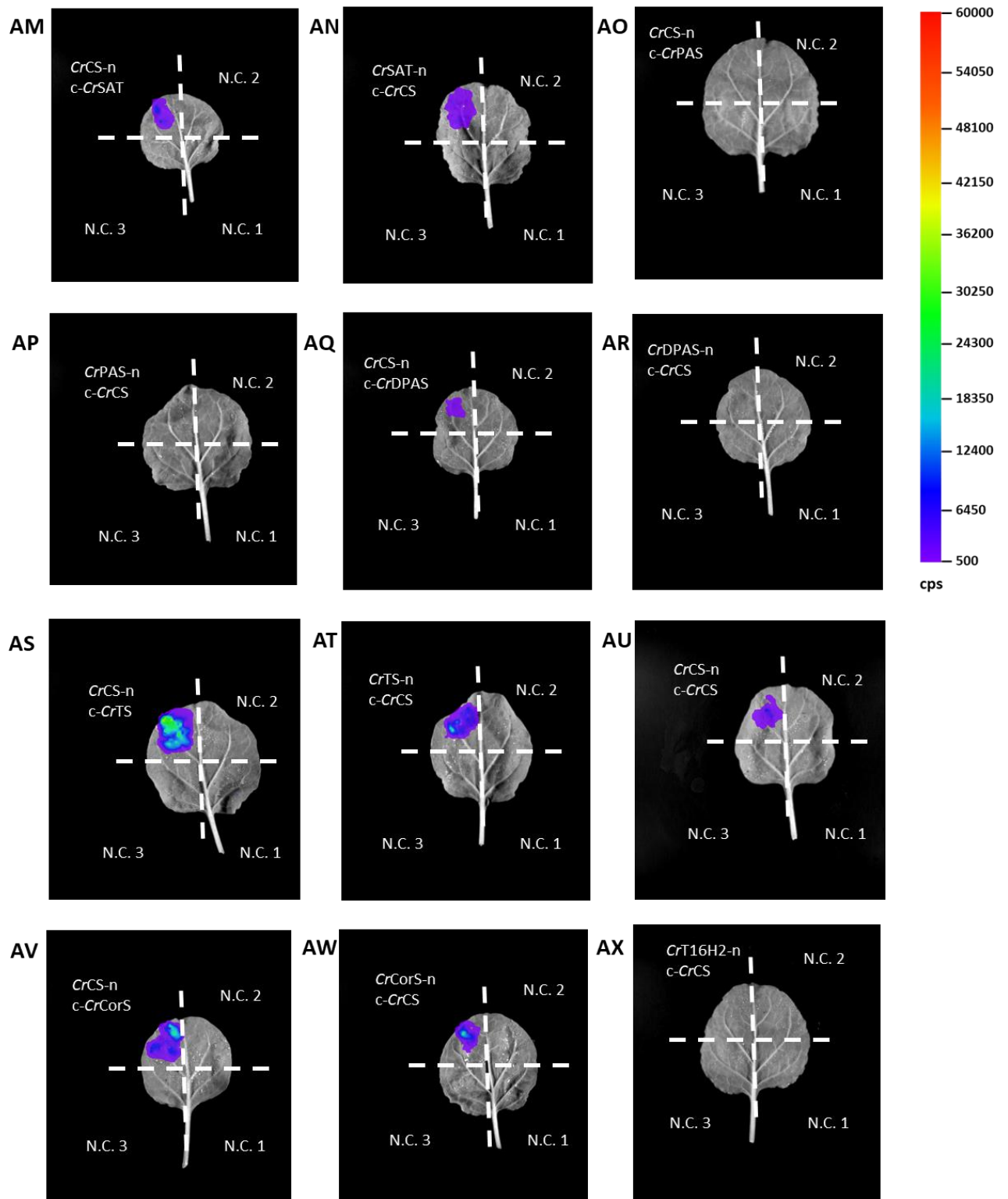


Figure 111. Representative images of pairwise interactions of *CrTS* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** TS-LAMT; **AC.** SLS-TS; **AD-E.** TS-TDC; **AF-G.** TS-GS; **AH.** GO-TS; **AI-J.** TS-RedOx1; **AK-L.** TS-RedOx2; **AM-N.** TS-SAT; **AO-P.** TS-PAS; **AQ-R.** TS-DPAS; **AS.** TS-TS; **AT-U.** TS-CS; **AV-W.** TS-CorS; **AX.** T16H2-TS; **AY-Z.** TS-16OMT; **BA.** T3O-TS; **BB-C.** TS-T3R.





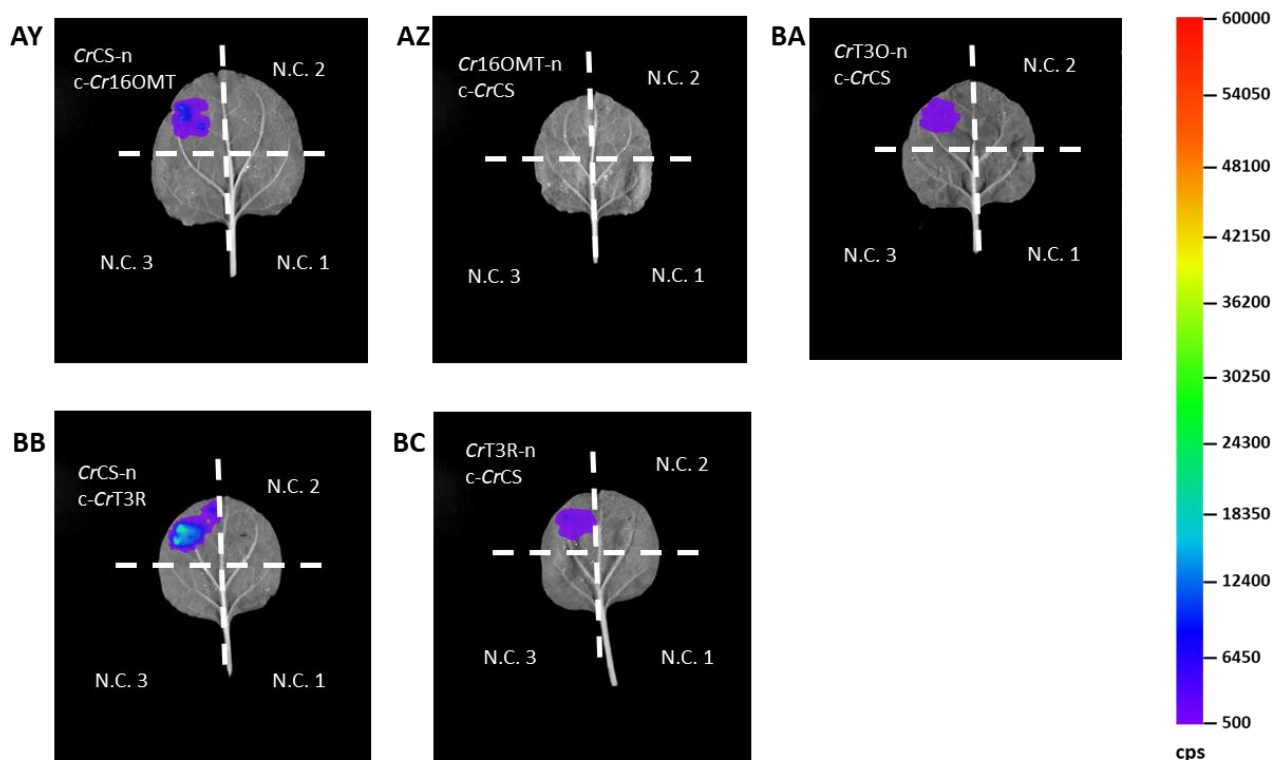
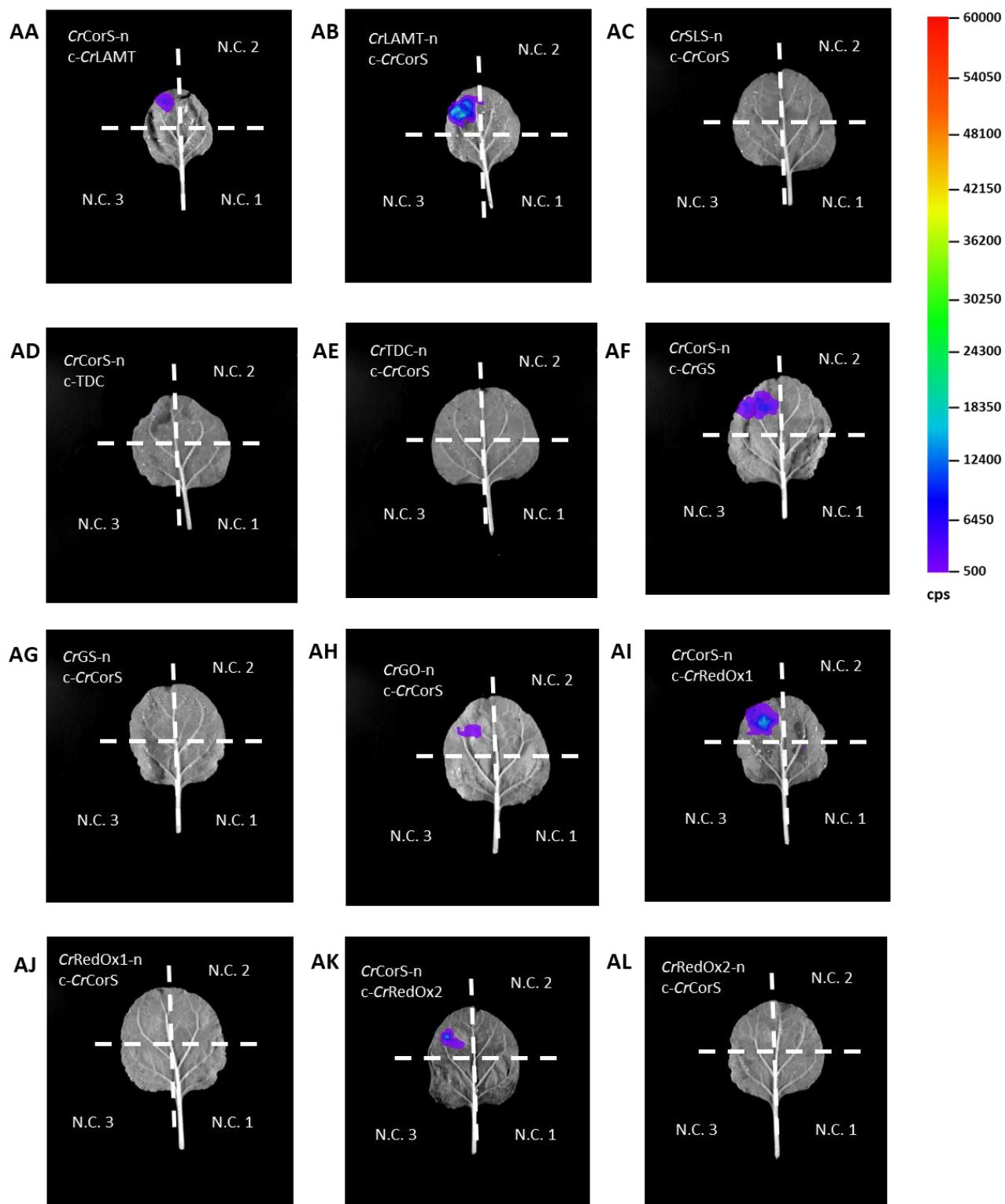
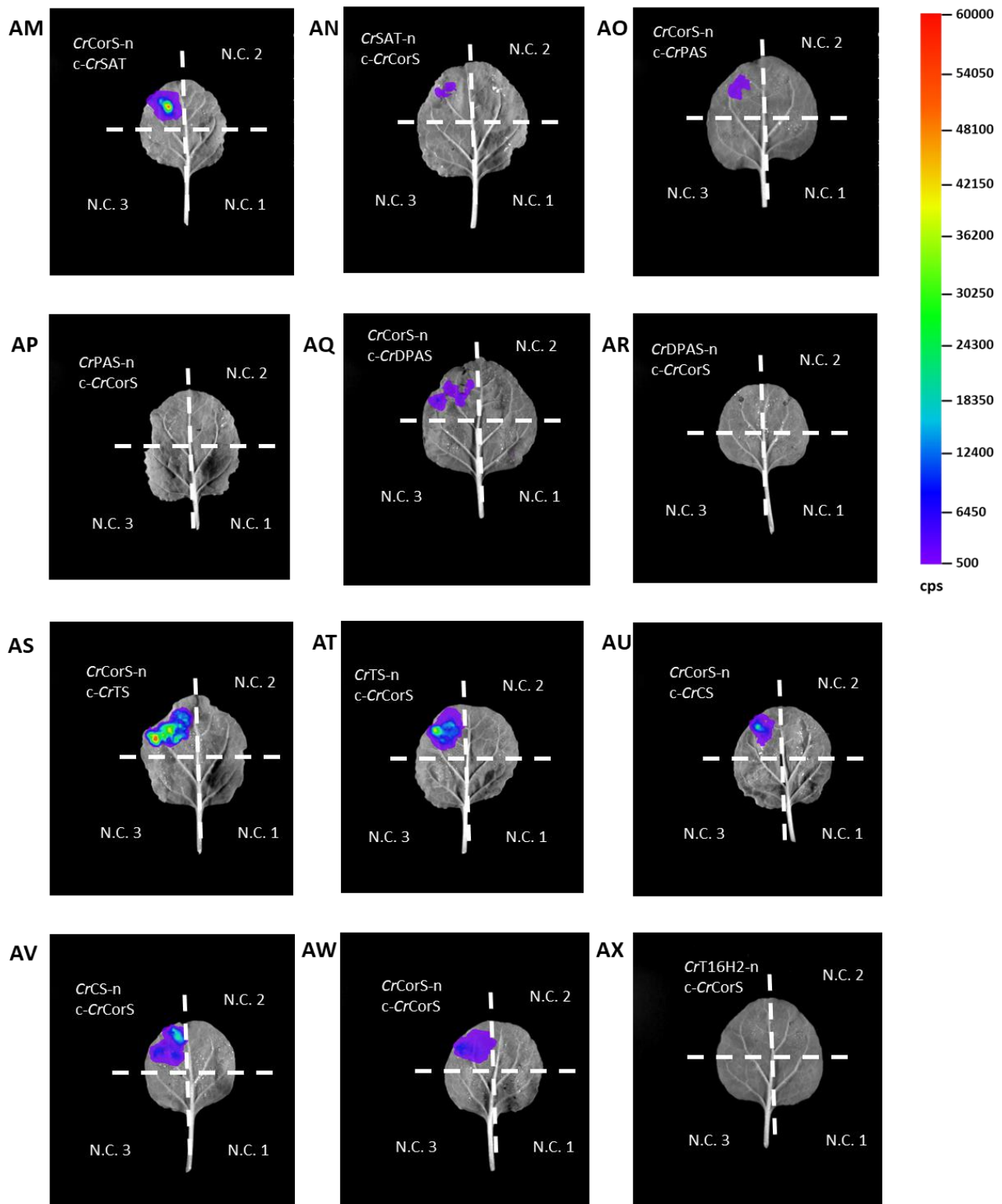


Figure 112. Representative images of pairwise interactions of *CrCS* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** CS-LAMT; **AC.** SLS-CS; **AD-E.** CS-TDC; **AF-G.** CS-GS; **AH.** GO-CS; **AI-J.** CS-RedOx1; **AK-L.** CS-RedOx2; **AM-N.** CS-SAT; **AO-P.** CS-PAS; **AQ-R.** CS-DPAS; **AS-T.** CS-TS; **AU.** CS-CS; **AV-W.** CS-CorS; **AX.** T16H2-CS; **AY-Z.** CS-16OMT; **BA.** T3O-CS; **BB-C.** CS-T3R.





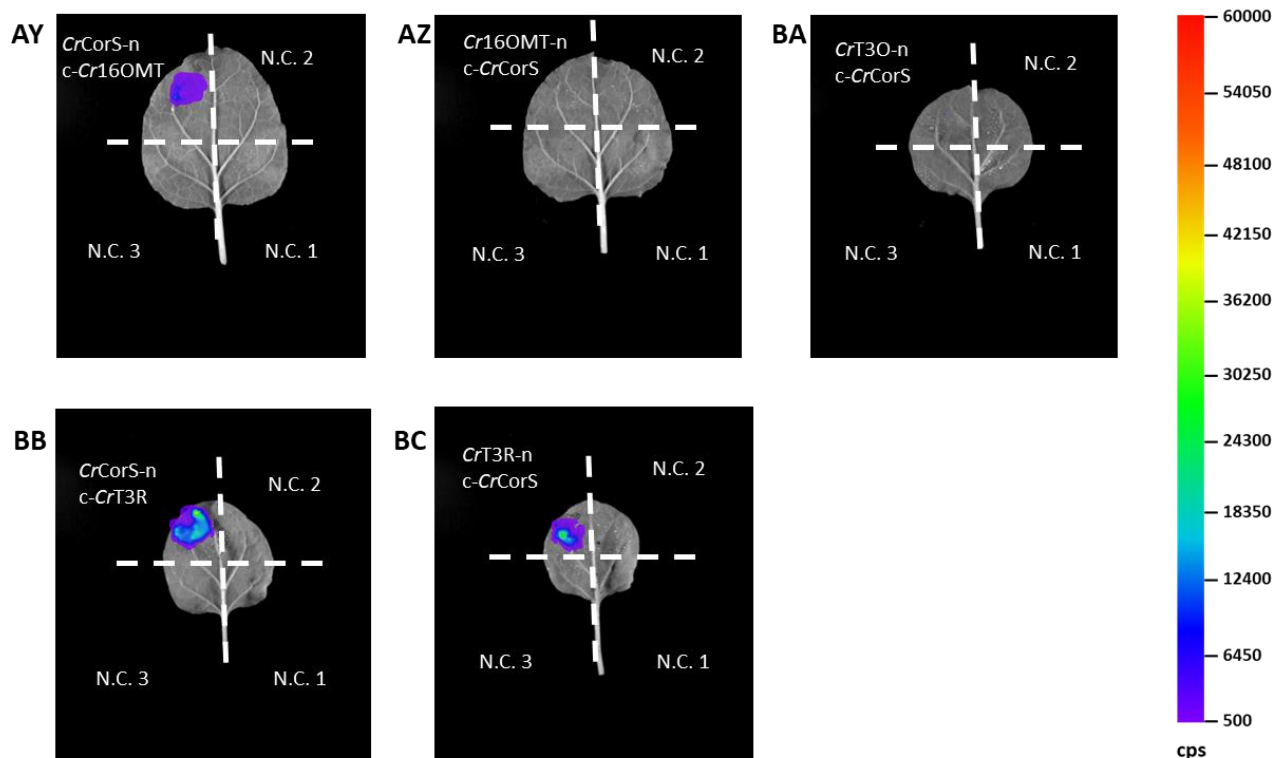
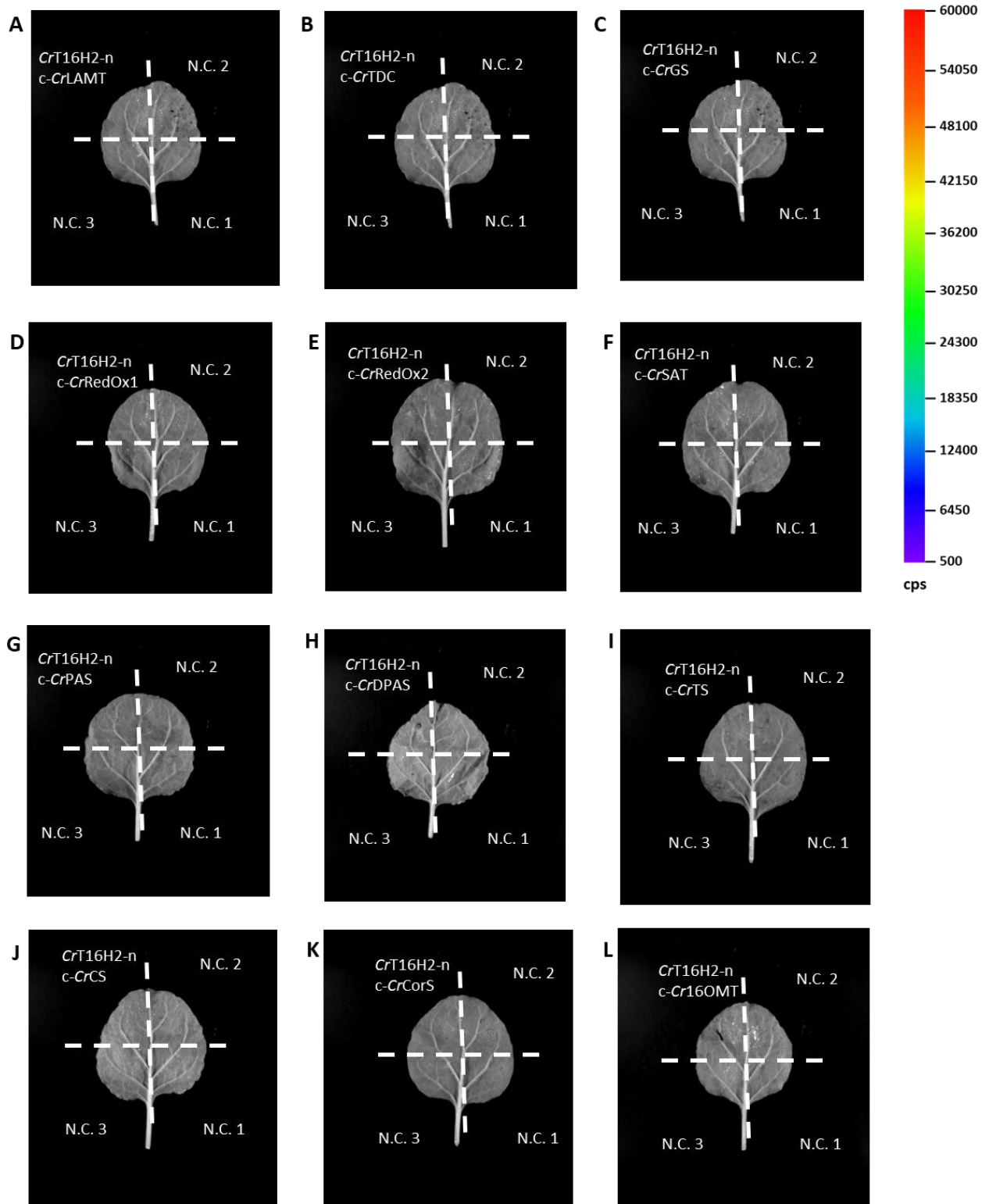


Figure 113. Representative images of pairwise interactions of *CrCorS* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** CorS-LAMT; **AC.** SLS-CorS; **AD-E.** CorS-TDC; **AF-G.** CorS-GS; **AH.** GO-CorS; **AI-J.** CorS-RedOx1; **AK-L.** CorS-RedOx2; **AM-N.** CorS-SAT; **AO-P.** CorS-PAS; **AQ-R.** CorS-DPAS; **AS-T.** CorS-TS; **AU-V.** CorS-CS; **AW.** CorS-CorS; **AX.** T16H2-CorS; **AY-Z.** CorS-16OMT; **BA.** T30-CorS; **BB-C.** CorS-T3R.



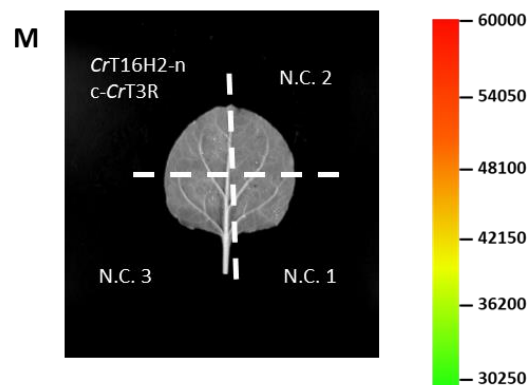
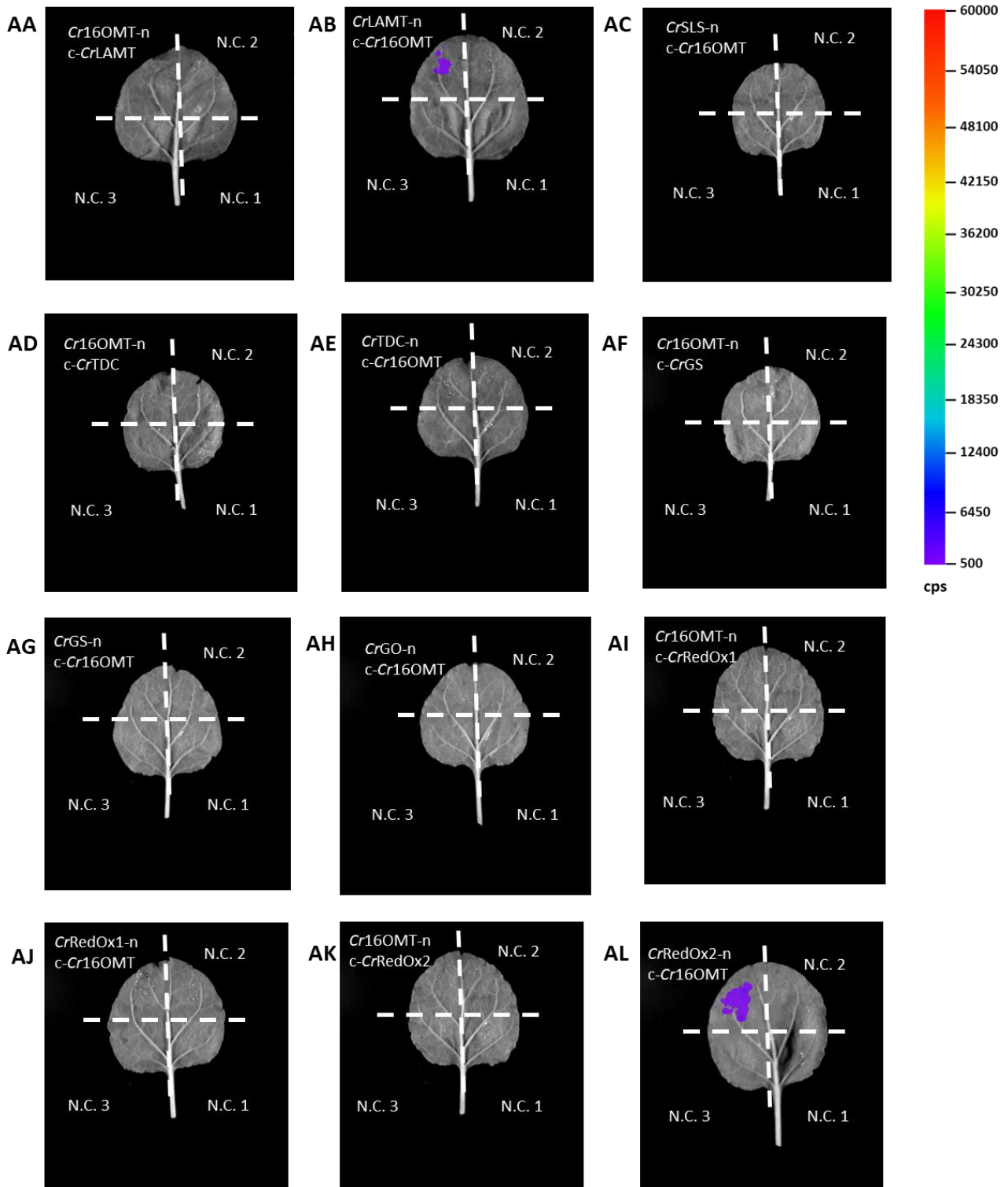
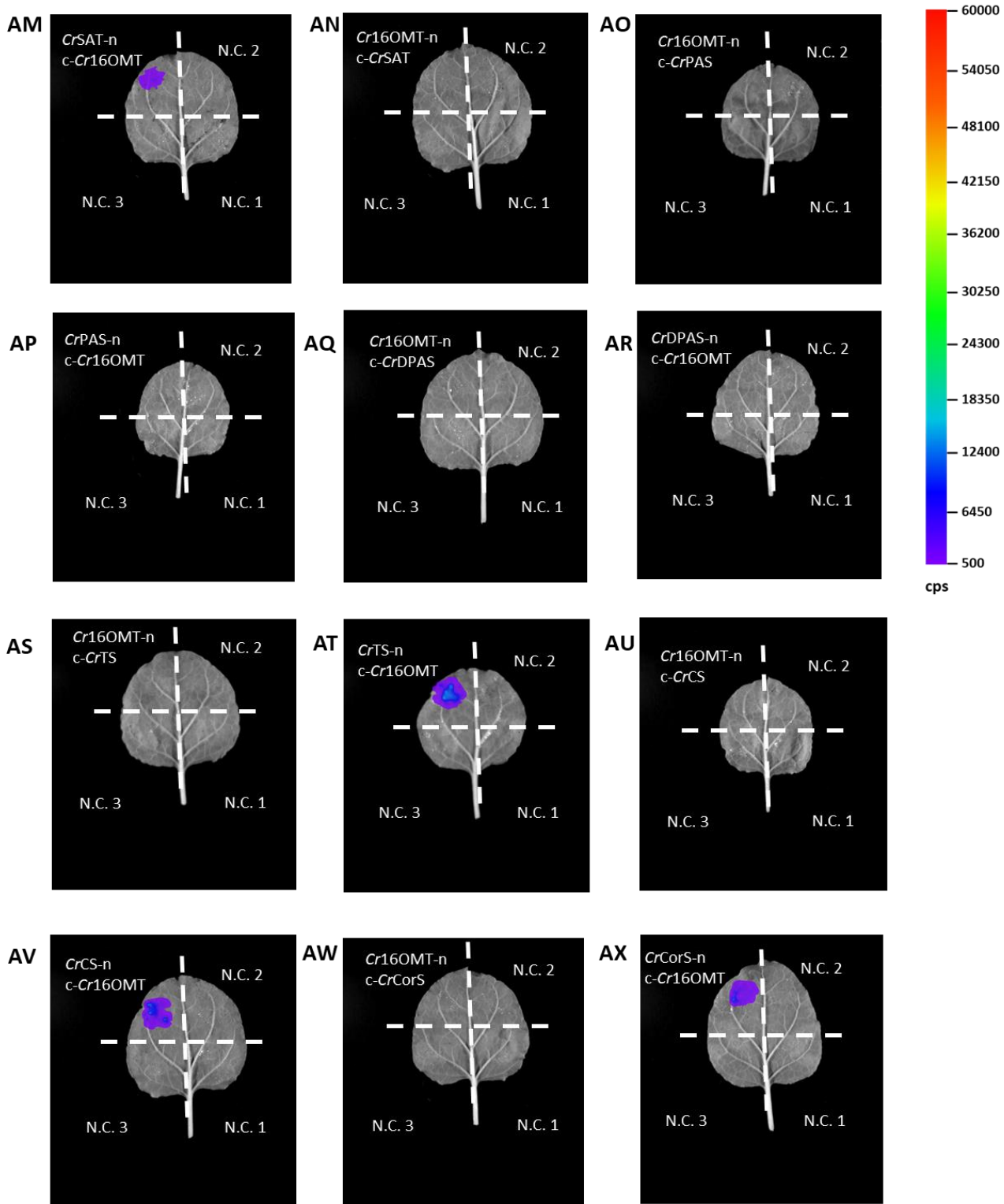


Figure 114. Representative images of pairwise interactions of CrT16H2 with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A.** T16H2-LAMT; **B.** T16H2-TDC; **C.** T16H2-GS; **D.** T16H2-RedOx1; **E.** T16H2-RedOx2; **F.** T16H2-SAT; **G.** T16H2-PAS; **H.** T16H2-DPAS; **I.** T16H2-TS; **J.** T16H2-CS; **K.** T16H2-CorS; **L.** T16H2-16OMT; **M.** T16H2-T3R.





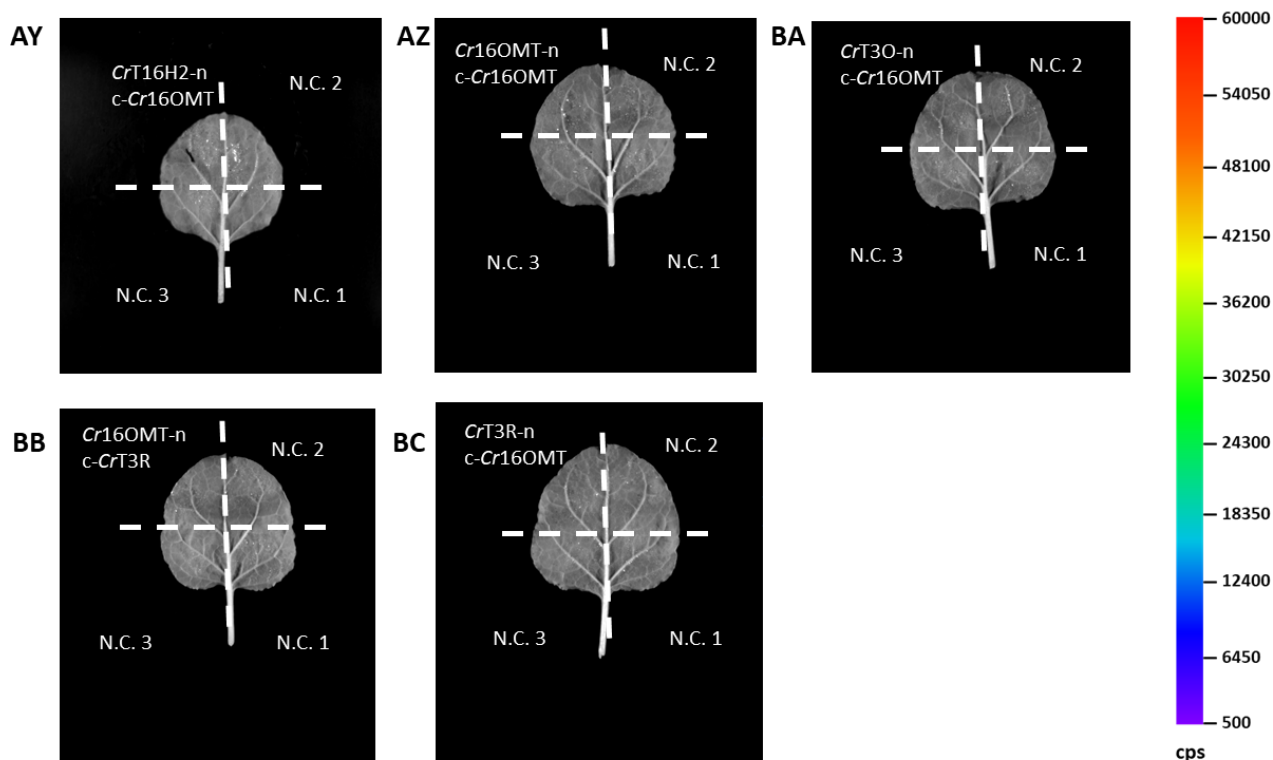
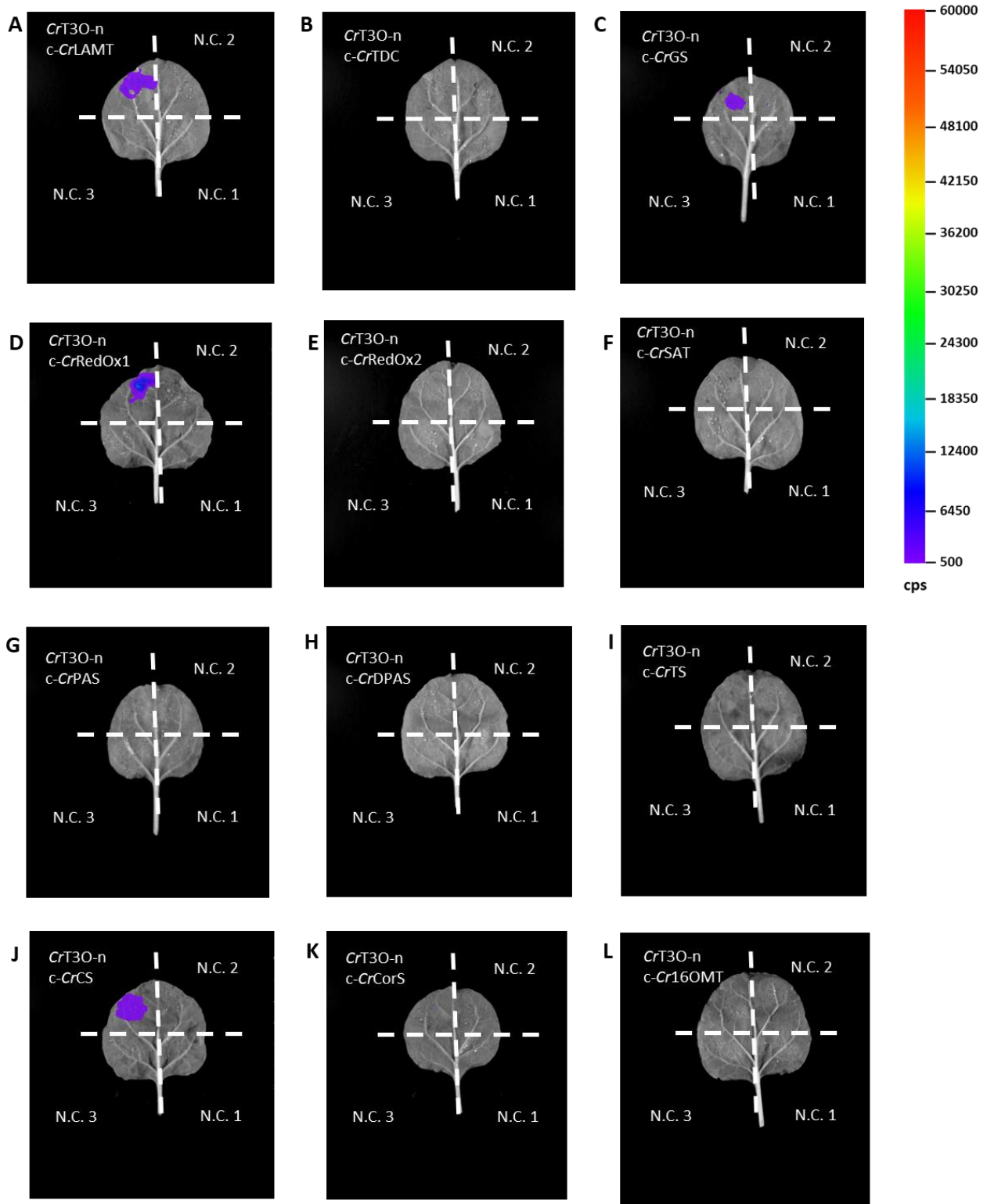


Figure 115. Representative images of pairwise interactions of *Cr16OMT* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** 16OMT-LAMT; **AC.** SLS-16OMT; **AD-E.** 16OMT-TDC; **AF-G.** 16OMT-GS; **AH.** GO-16OMT; **AI-J.** 16OMT-RedOx1; **AK-L.** 16OMT-RedOx2; **AM-N.** 16OMT-SAT; **AO-P.** 16OMT-PAS; **AQ-R.** 16OMT-DPAS; **AS-T.** 16OMT-TS; **AU-V.** 16OMT-CS; **AW-X.** 16OMT-CorS; **AY.** T16H2-16OMT; **AZ.** 16OMT-16OMT; **BA.** T30-16OMT; **BB-C.** 16OMT-T3R.



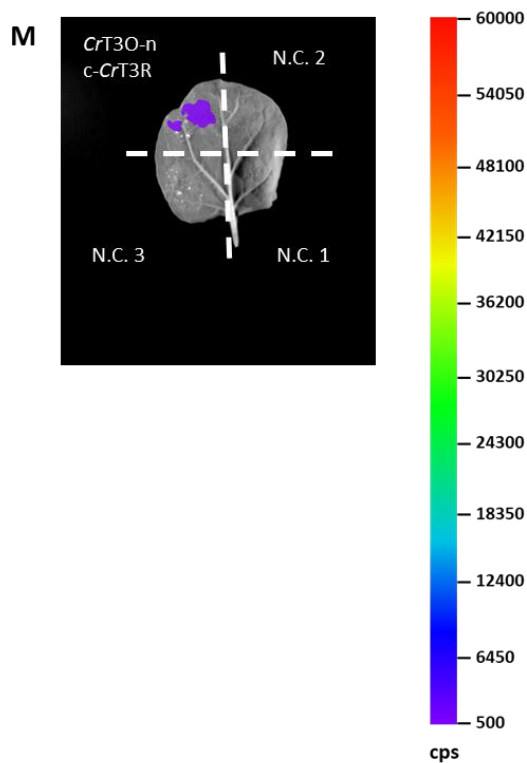
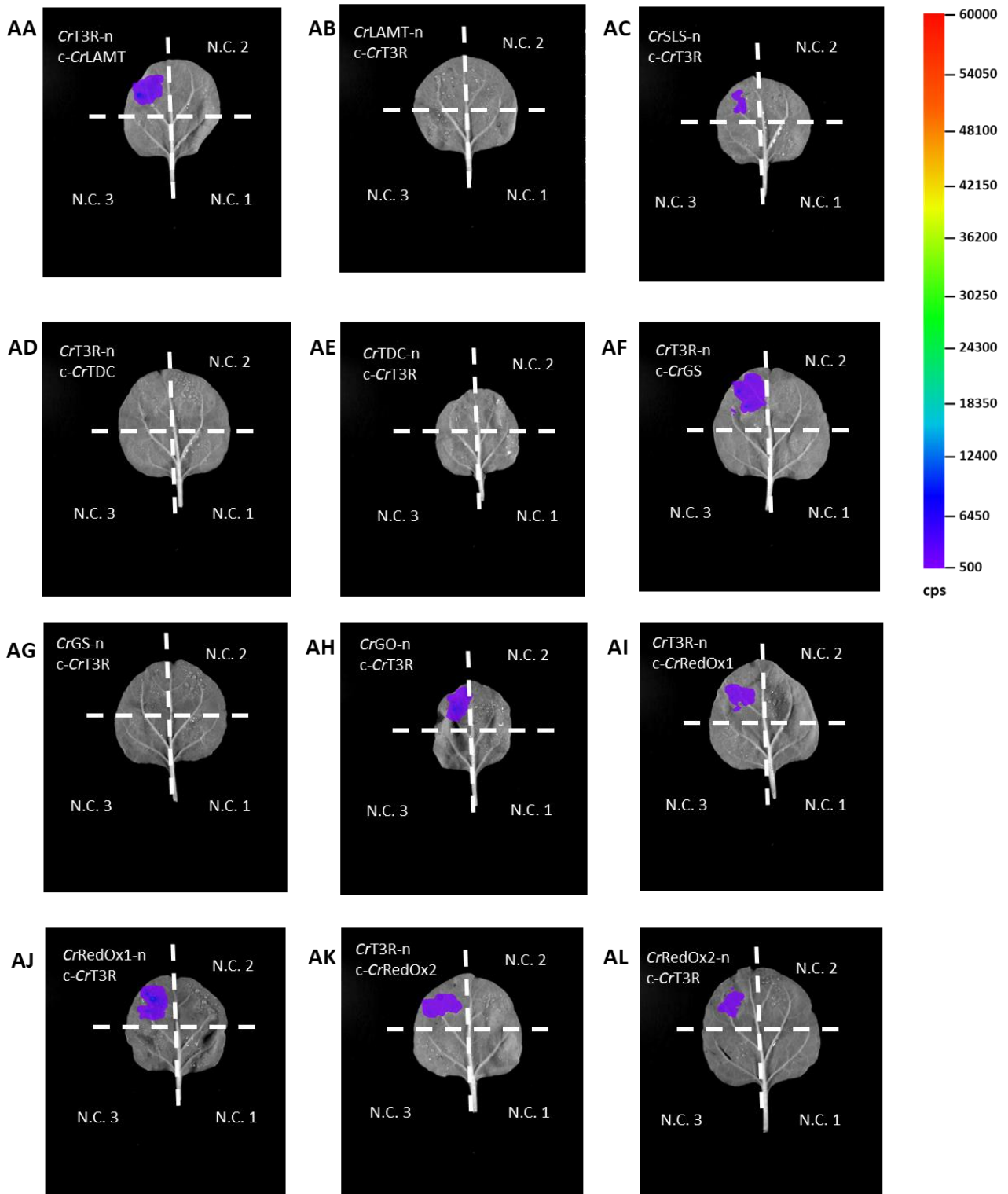
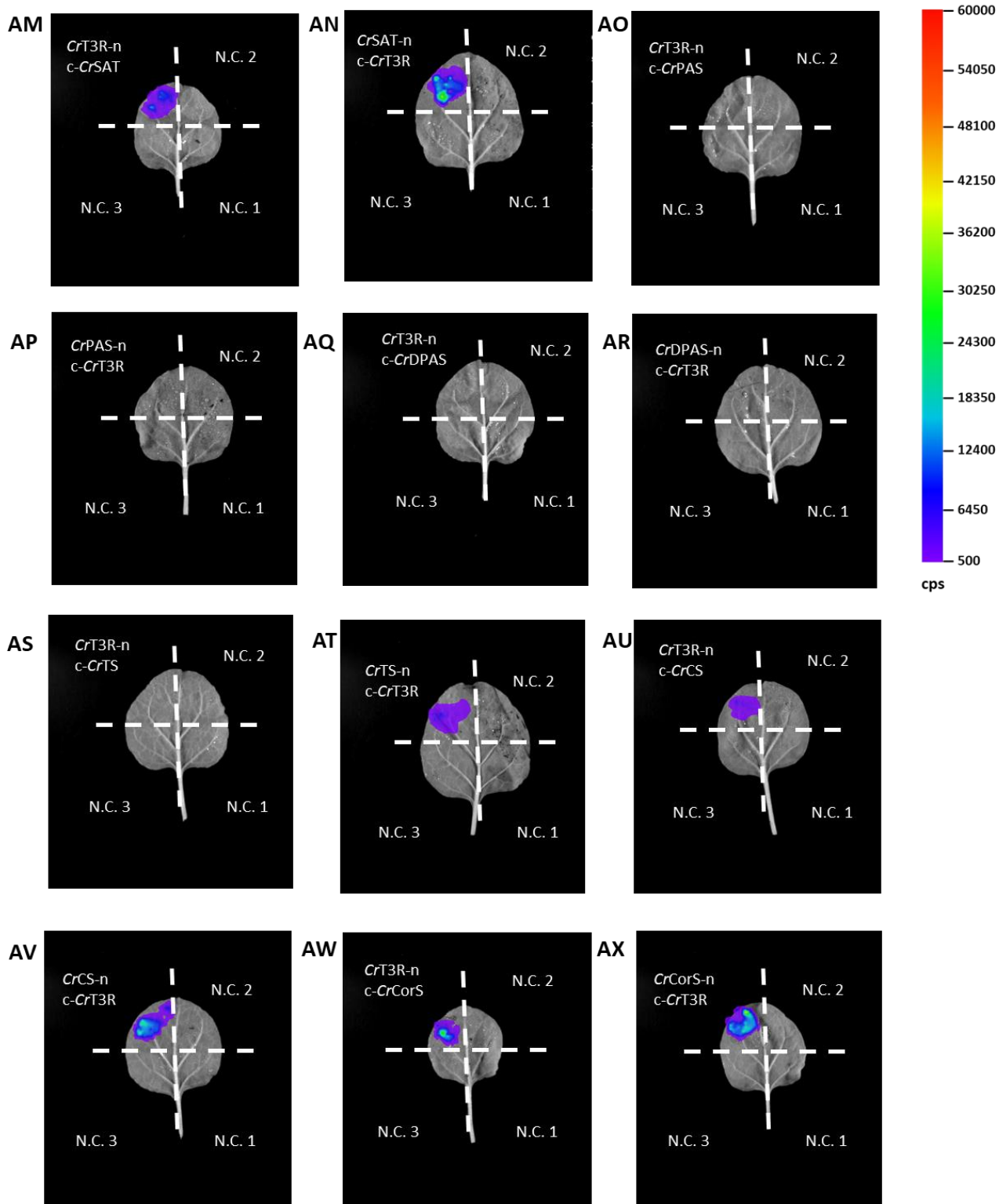


Figure 116. Representative images of pairwise interactions of *CrT30* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A.** T3O-LAMT; **B.** T3O-TDC; **C.** T3O-GS; **D.** T3O-RedOx1; **E.** T3O-RedOx2; **F.** T3O-SAT; **G.** T3O-PAS; **H.** T3O-DPAS; **I.** T3O-TS; **J.** T3O-CS; **K.** T3O-CorS; **L.** T3O-16OMT; **M.** T3O-T3R.





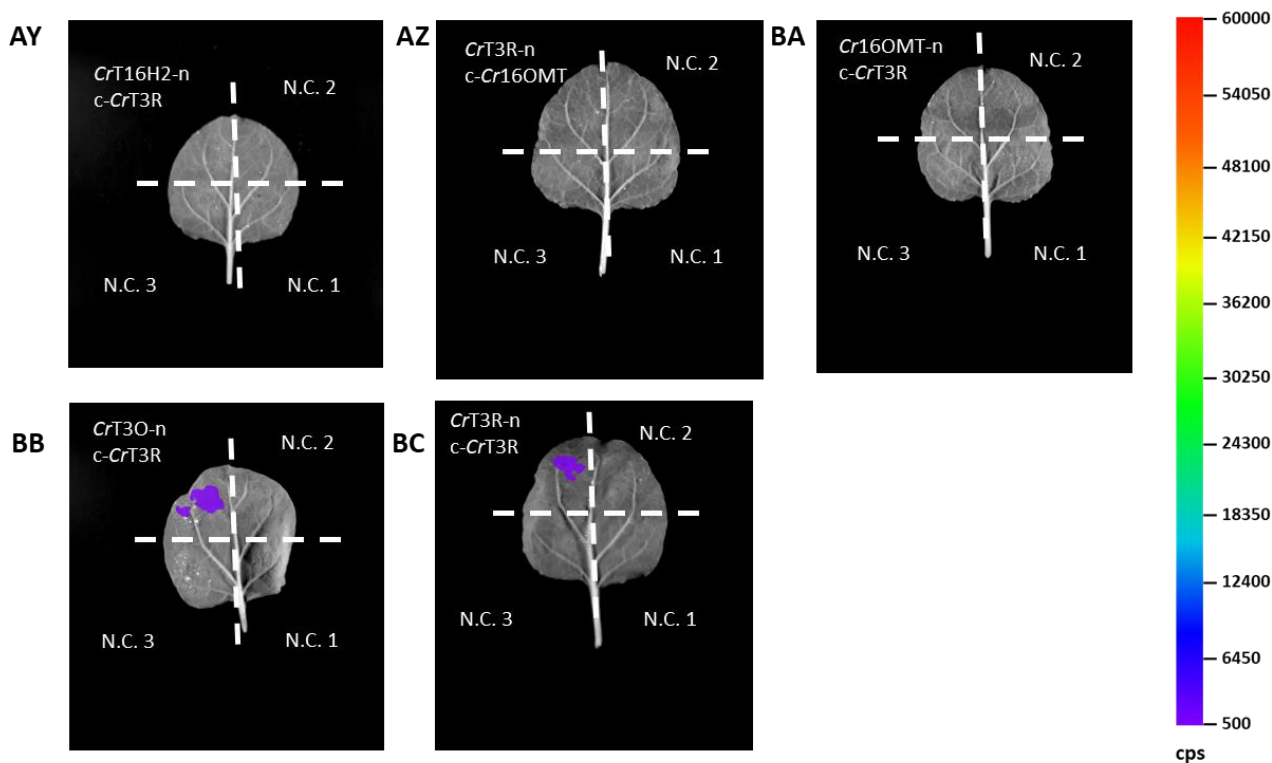


Figure 117. Representative images of pairwise interactions of *CrT3R* with *C. roseus* MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA-B.** T3R-LAMT; **AC.** SLS-T3R; **AD-E.** T3R-TDC; **AF-G.** T3R-GS; **AH.** GO-T3R; **AI-J.** T3R-RedOx1; **AK-L.** T3R-RedOx2; **AM-N.** T3R-SAT; **AO-P.** T3R-PAS; **AQ-R.** T3R-DPAS; **AS-T.** T3R-TS; **AU-V.** T3R-CS; **AW-X.** T3R-CorS; **AY.** T16H2-T3R; **AZ-BA.** T3R-16OMT; **BB.** T3O-16OMT; **BC.** T3R-T3R.

Appendix X. Split-Luciferase Assays of CADs and α/β hydrolases

Representative split-luciferase result of testing protein-protein interactions between CADs and α/β -hydrolases in heterologous host *N. benthamiana*.

Table 24. Nucleic acid sequence identity of CADs tested for the α/β hydrolase interaction conservation.

AtCAD4		70.97	49.73	51.76	52.15	53.66	54.18	53.59	51.55	49.11	50.73
CrCAD4	70.97		50.36	53.41	52.24	54.70	54.27	54.97	51.28	50.98	51.37
Cr2141	49.73	50.36		59.44	62.13	58.59	60.15	60.04	57.36	55.43	57.55
CrADH9	51.76	53.41	59.44		74.68	60.00	58.31	59.39	57.97	57.18	58.24
TiDPAS1	52.15	52.24	62.13	74.68		61.19	62.28	63.83	59.29	60.55	61.20
TiDPAS2	53.66	54.70	58.59	60.00	61.19		64.83	64.91	61.25	60.00	61.71
CrDPAS	54.18	54.27	60.15	58.31	62.28	64.83		67.87	64.56	63.73	65.84
CrRedOx1	53.59	54.97	60.04	59.39	63.83	64.91	67.87		71.36	70.14	73.39
CrGS	51.55	51.28	57.36	57.97	59.29	61.25	64.56	71.36		76.09	81.19
CrTHAS	49.11	50.98	55.43	57.18	60.55	60.00	63.73	70.14	76.09		82.04
CrT3R	50.73	51.37	57.55	58.24	61.20	61.71	65.84	73.39	81.19	82.04	

Table 25. Nucleic acid sequence identity of α/β hydrolases tested for ADH interaction conservation.

	CrCSE	SnvNS2	PsCXE1	GmHIDH	CrHID5	CrCS	CrCorS	CrTS	TiCorS	TiTabS
CrCSE		22.32	24.94	26.16	33.47	33.45	37.14	35.09	34.72	34.81
SnvNS2	22.32		28.60	28.78	30.97	33.27	33.27	32.91	32.48	31.16
PsCXE1	24.94	28.60		39.63	43.49	43.92	44.84	45.54	43.79	44.17
GmHIDH	26.16	28.78	39.63		54.16	52.92	53.73	55.12	54.41	54.91
CrHID5	33.47	30.97	43.49	54.16		60.64	63.62	64.96	65.83	65.53
CrCS	33.45	33.27	43.92	52.92	60.64		81.07	80.50	72.80	72.70
CrCorS	37.14	33.27	44.84	53.73	63.62	81.07		84.23	78.56	78.36
CrTS	35.09	32.91	45.54	55.12	64.96	80.50	84.23		76.28	78.33
TiCorS	34.72	32.48	43.79	54.41	65.83	72.80	78.56	76.28		85.38
TiTabS	34.81	31.16	44.17	54.91	65.53	72.70	78.36	78.33	85.38	

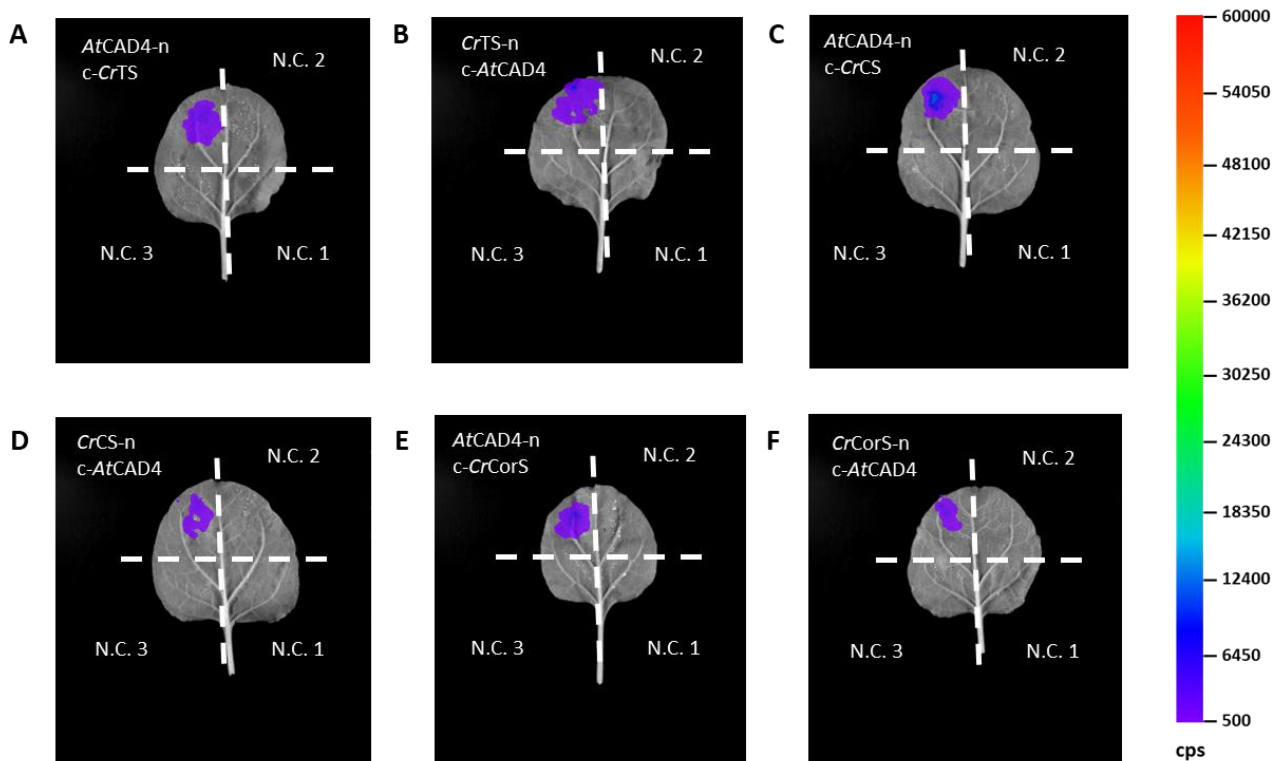
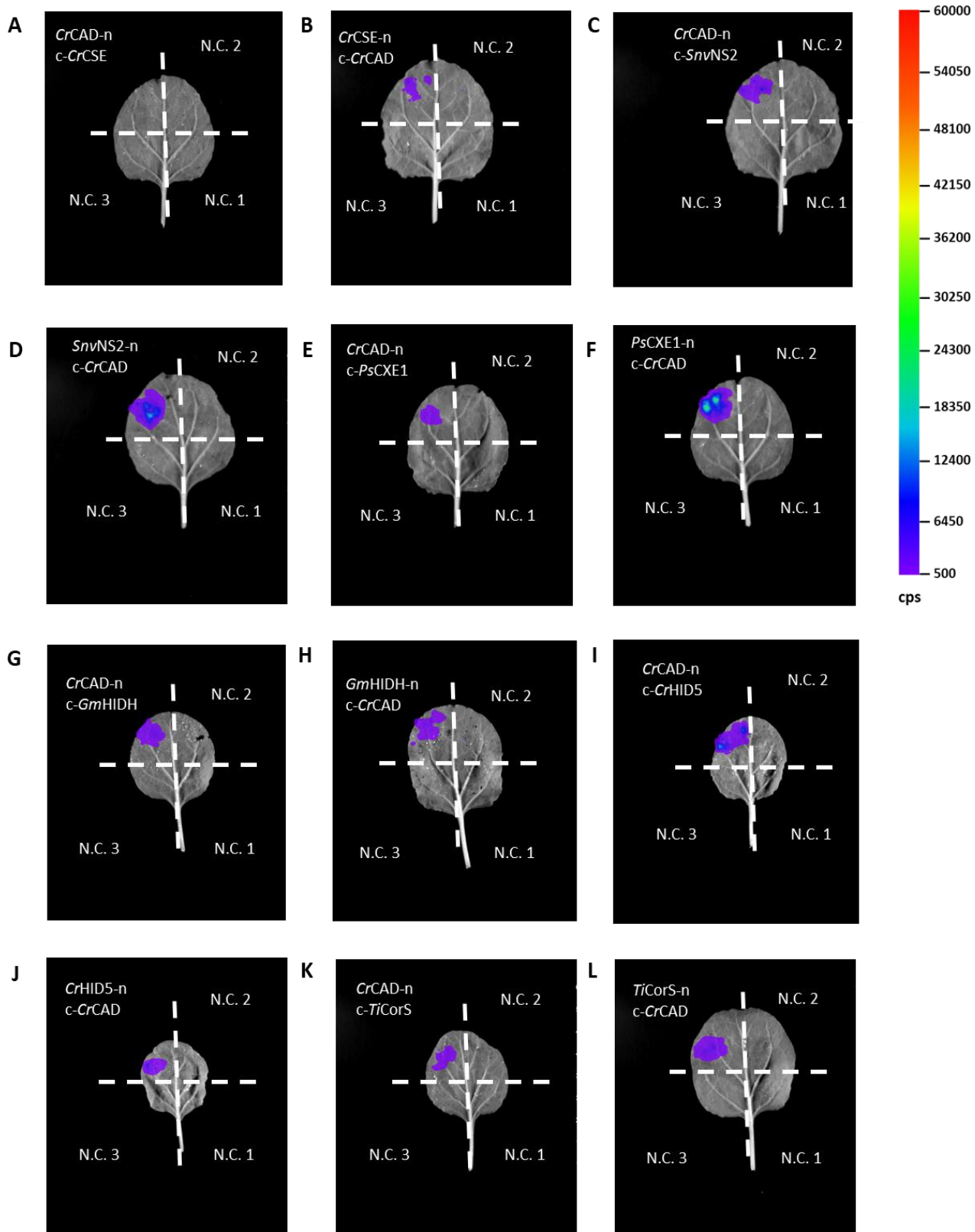


Figure 118. Representative images of pairwise interactions of *AtCAD4* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *AtCAD4*-*CrTS*; **C-D.** *AtCAD4*-*CrCS*; **E-F.** *AtCAD4*-*CrCorS*.



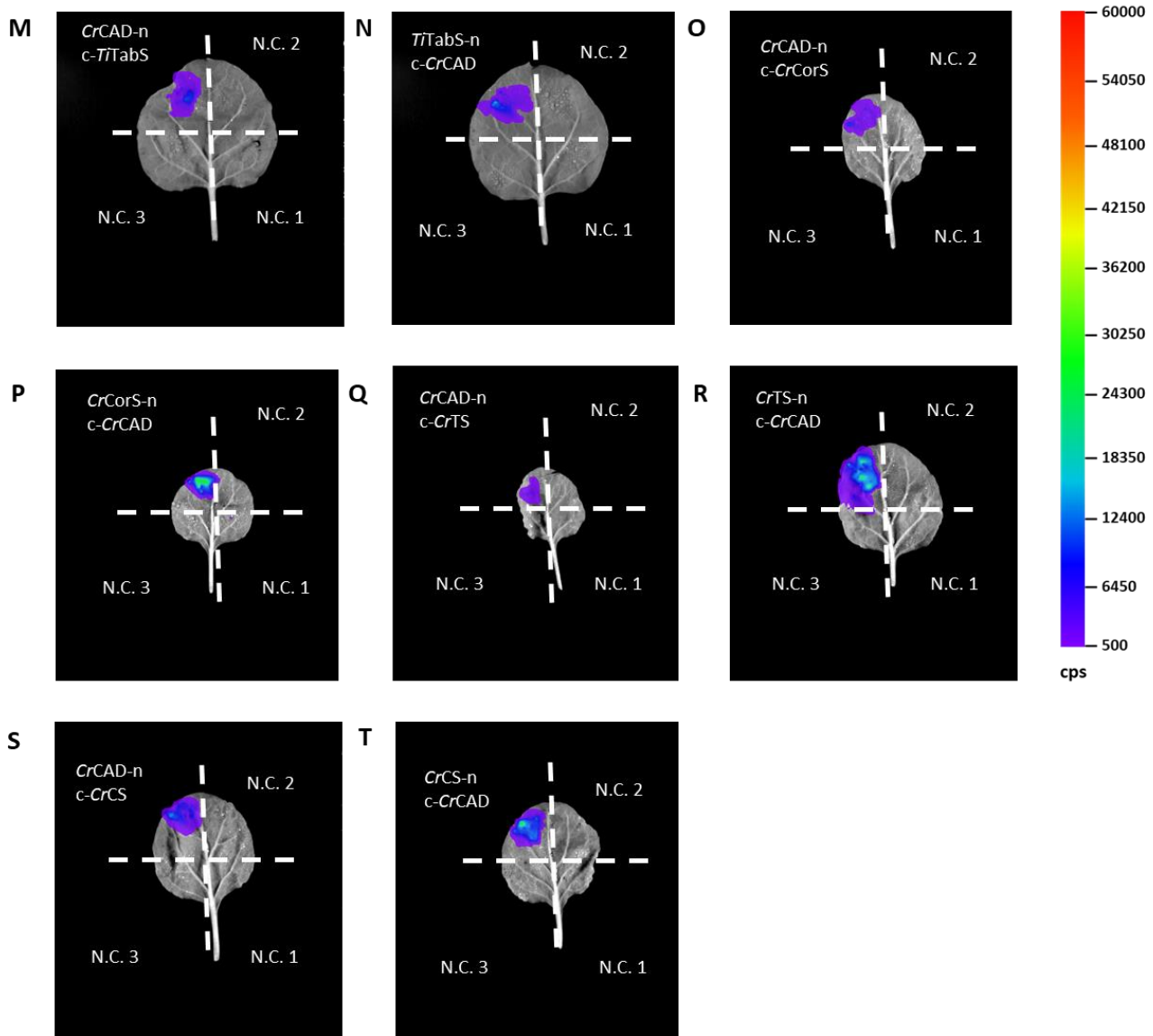
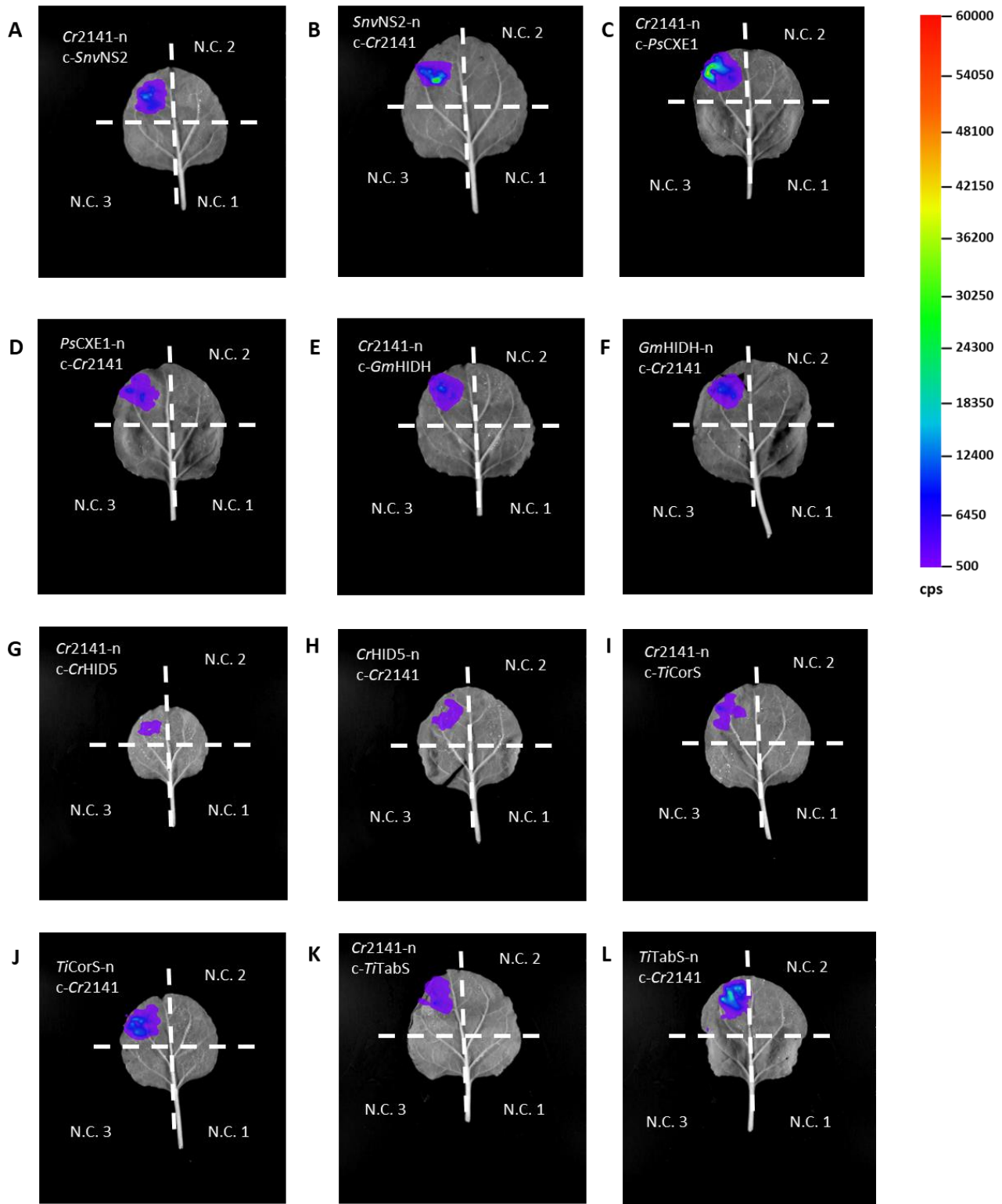


Figure 119. Representative images of pairwise interactions of *CrCAD* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrCAD-CrCSE*; **C-D.** *CrCAD-SnvNS2*; **E-F.** *CrCAD-PsCXE1*; **G-H.** *CrCAD-GmHIDH*; **I-J.** *CrCAD-CrHID5*; **K-L.** *CrCAD-TiCorS*; **M-N.** *CrCAD-TiTabS*; **O-P.** *CrCAD-CrCorS*; **Q-R.** *CrCAD-CrTS*; **S-T.** *CrCAD-CrCS*.



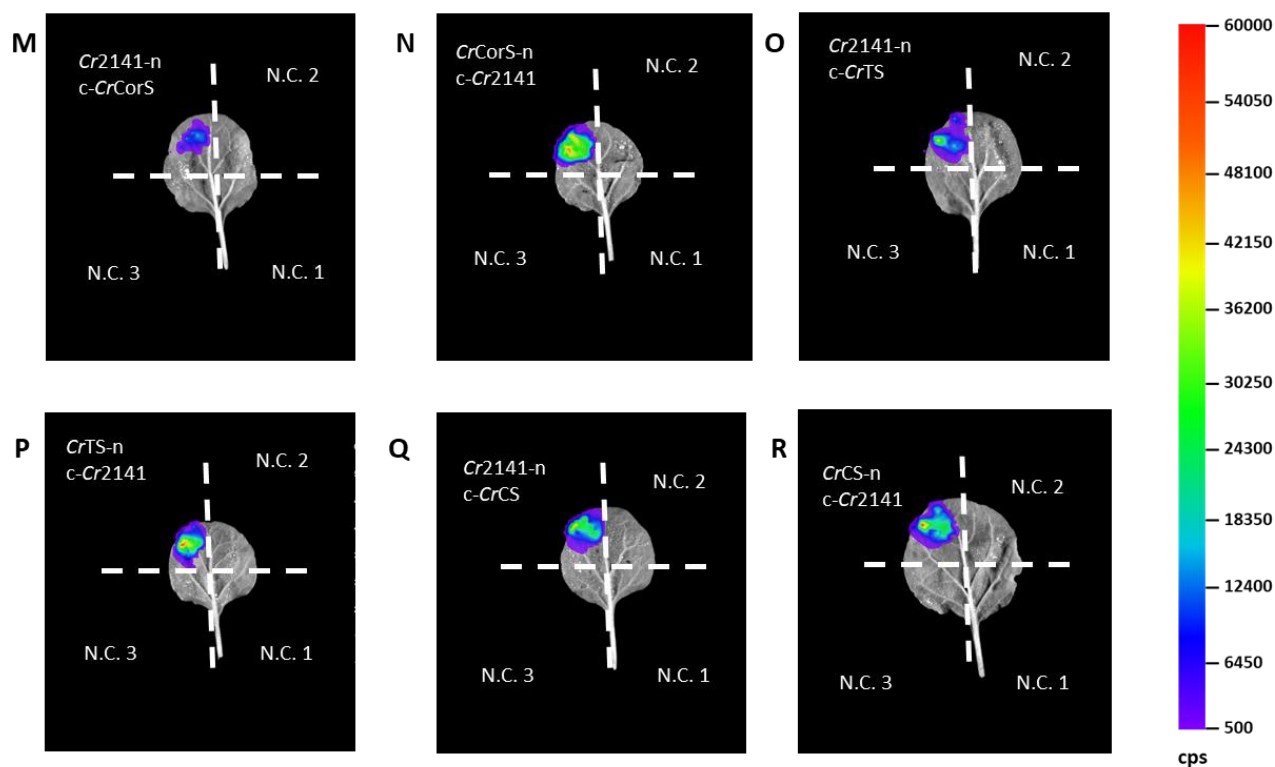
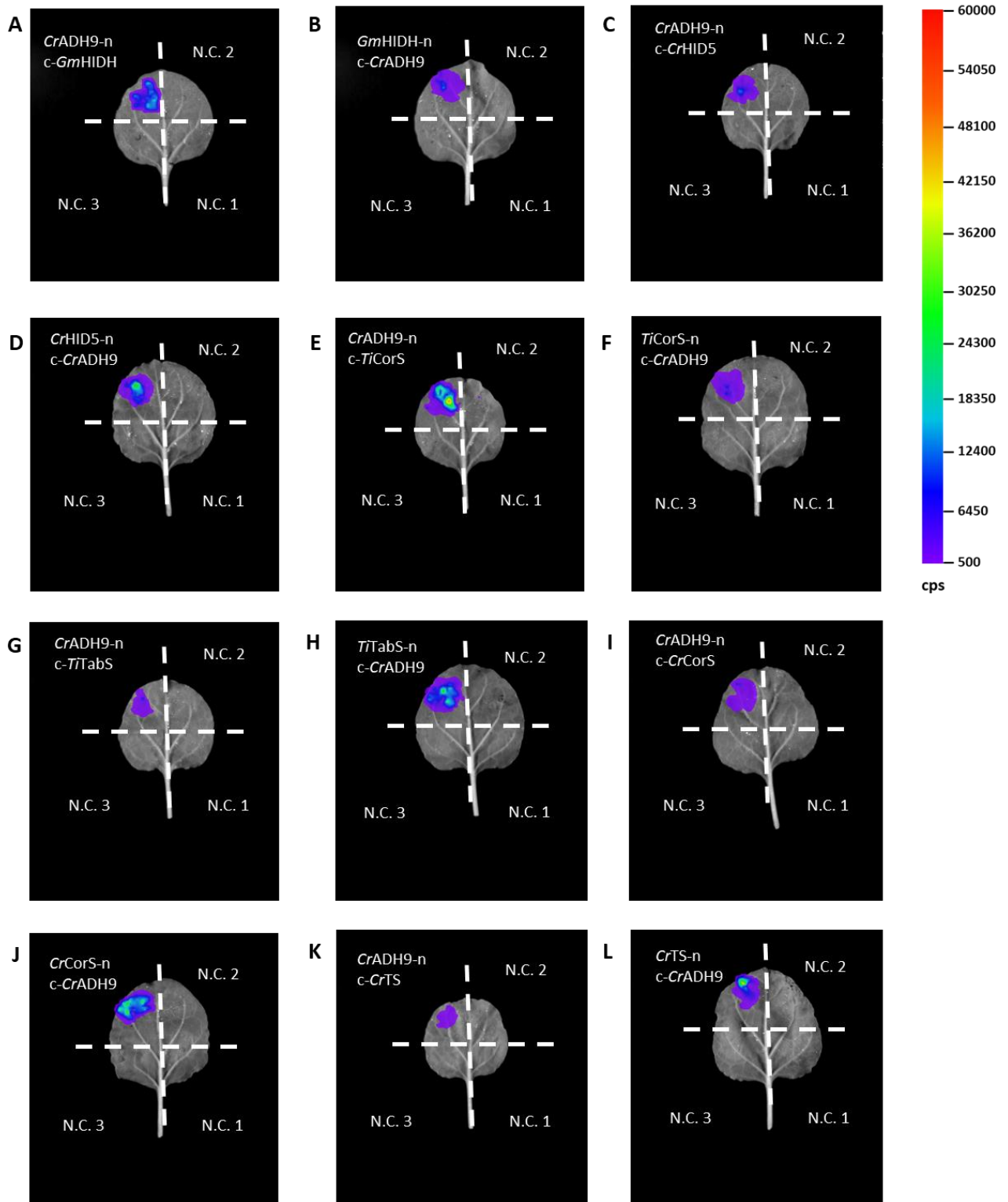


Figure 120. Representative images of pairwise interactions of *Cr2141* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *Cr2141-SnvNS2*; **C-D.** *Cr2141-PsCXE1*; **E-F.** *Cr2141-GmHIDH*; **G-H.** *Cr2141-CrHID5*; **I-J.** *Cr2141-TiCorS*; **K-L.** *Cr2141-TITabS*; **M-N.** *Cr2141-CrCorS*; **O-P.** *Cr2141-CrTS*; **Q-R.** *Cr2141-CrCS*.



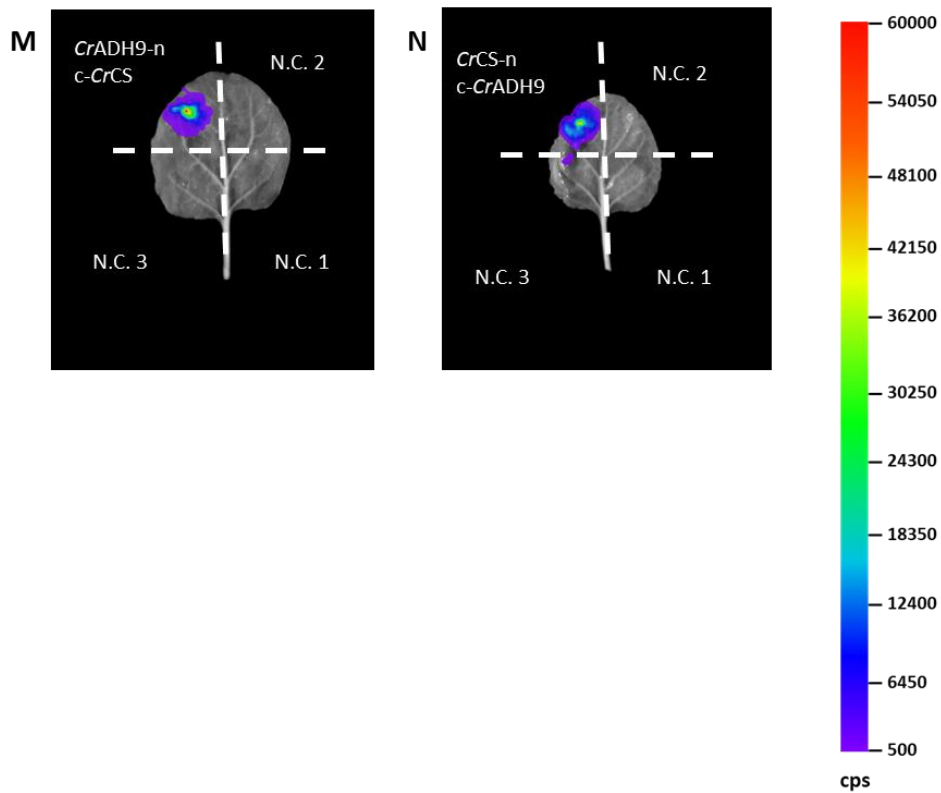


Figure 121. Representative images of pairwise interactions of *CrADH9* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrADH9-GmHIDH*; **C-D.** *CrADH9-CrHID5*; **E-F.** *CrADH9-TiCorS*; **G-H.** *CrADH9-TiTabS*; **I-J.** *CrADH9-CrCorS*; **K-L.** *CrADH9-CrTS*; **M-N.** *CrADH9-CrCS*.

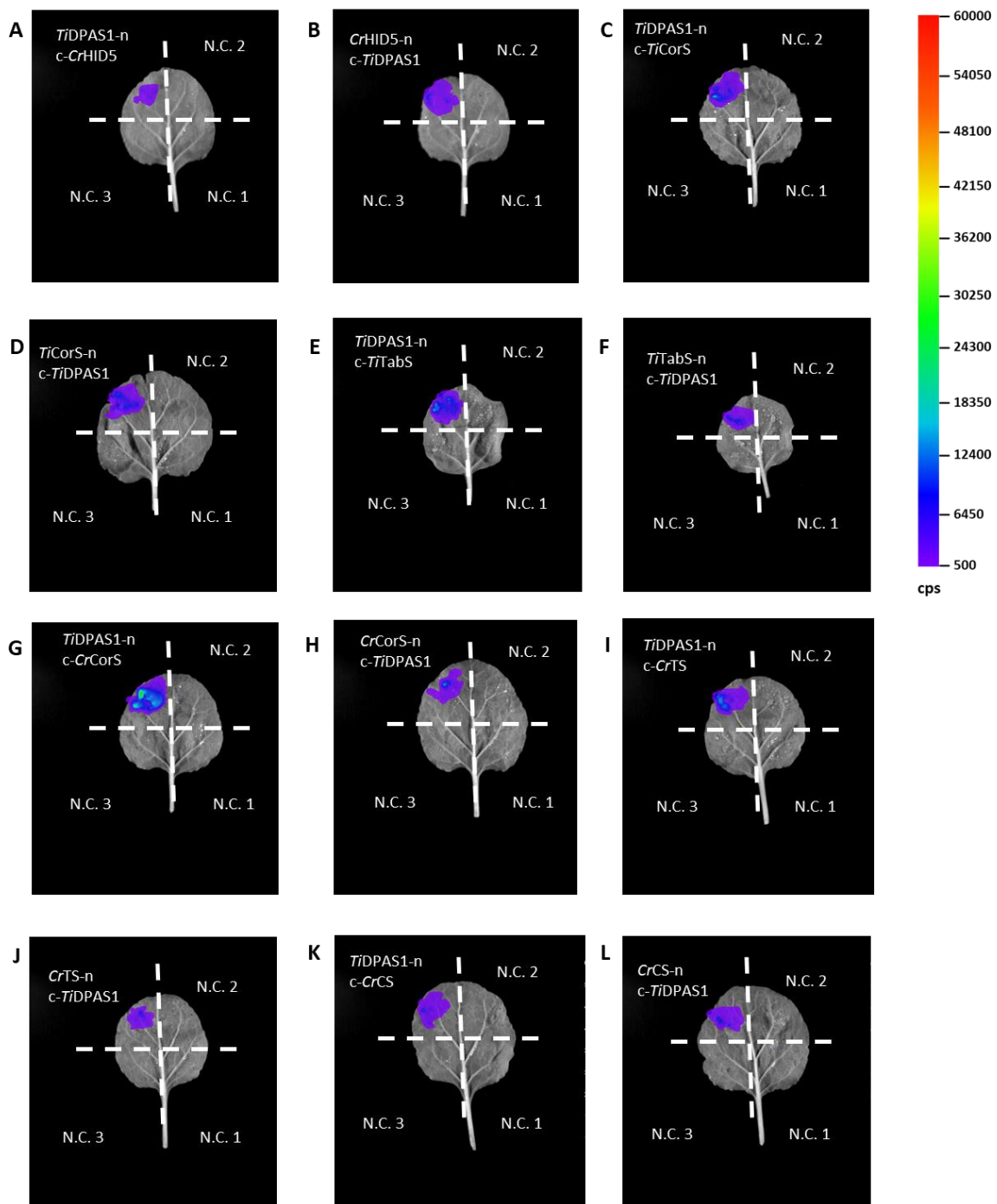


Figure 122. Representative images of pairwise interactions of *TiDPAS1* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *TiDPAS1*-*CrHID5*; **C-D.** *TiDPAS1*-*TiCorS*; **E-F.** *TiDPAS1*-*TITabS*; **G-H.** *TiDPAS1*-*CrCorS*; **I-J.** *TiDPAS1*-*CrTS*; **K-L.** *TiDPAS1*-*CrCS*.

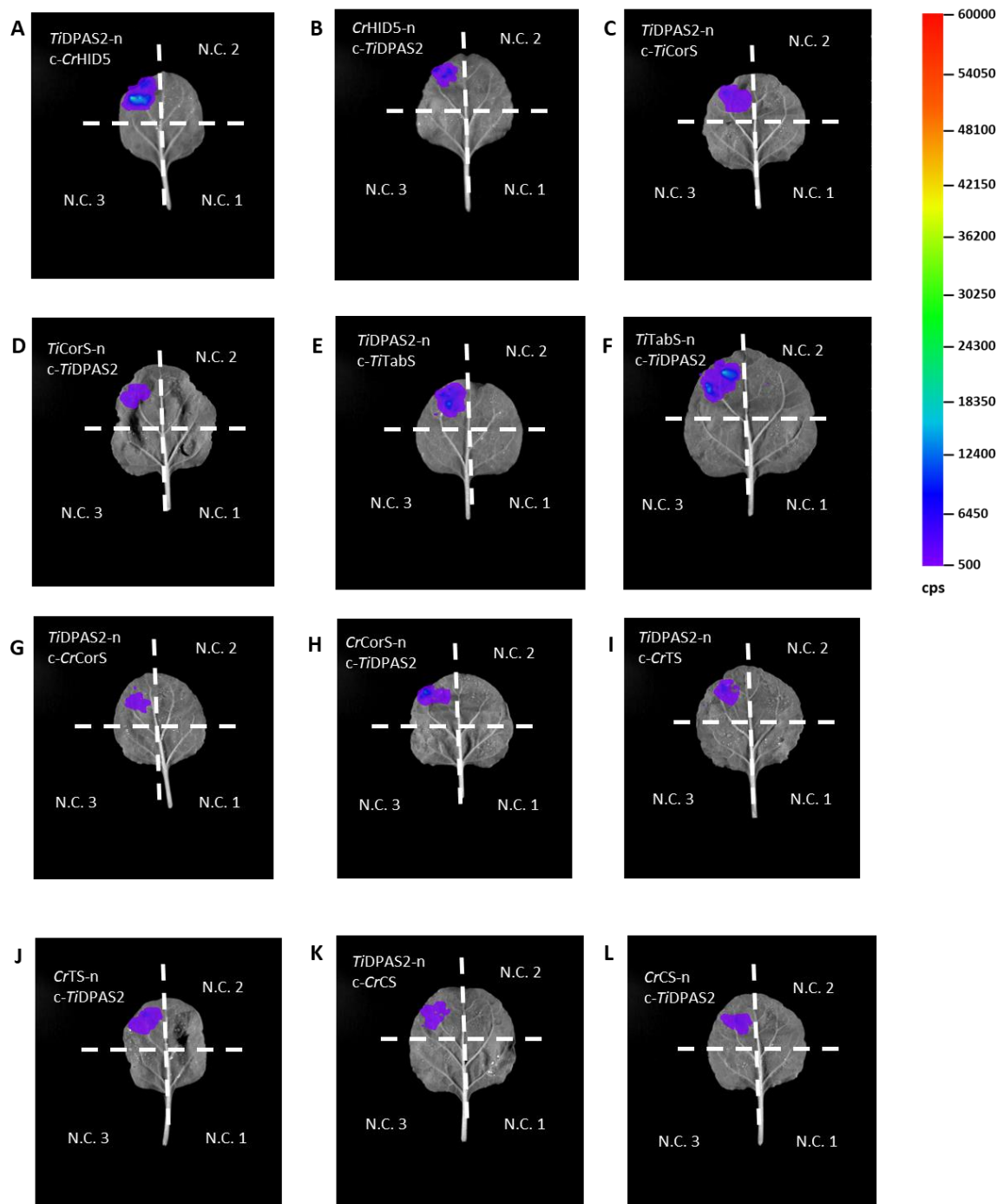
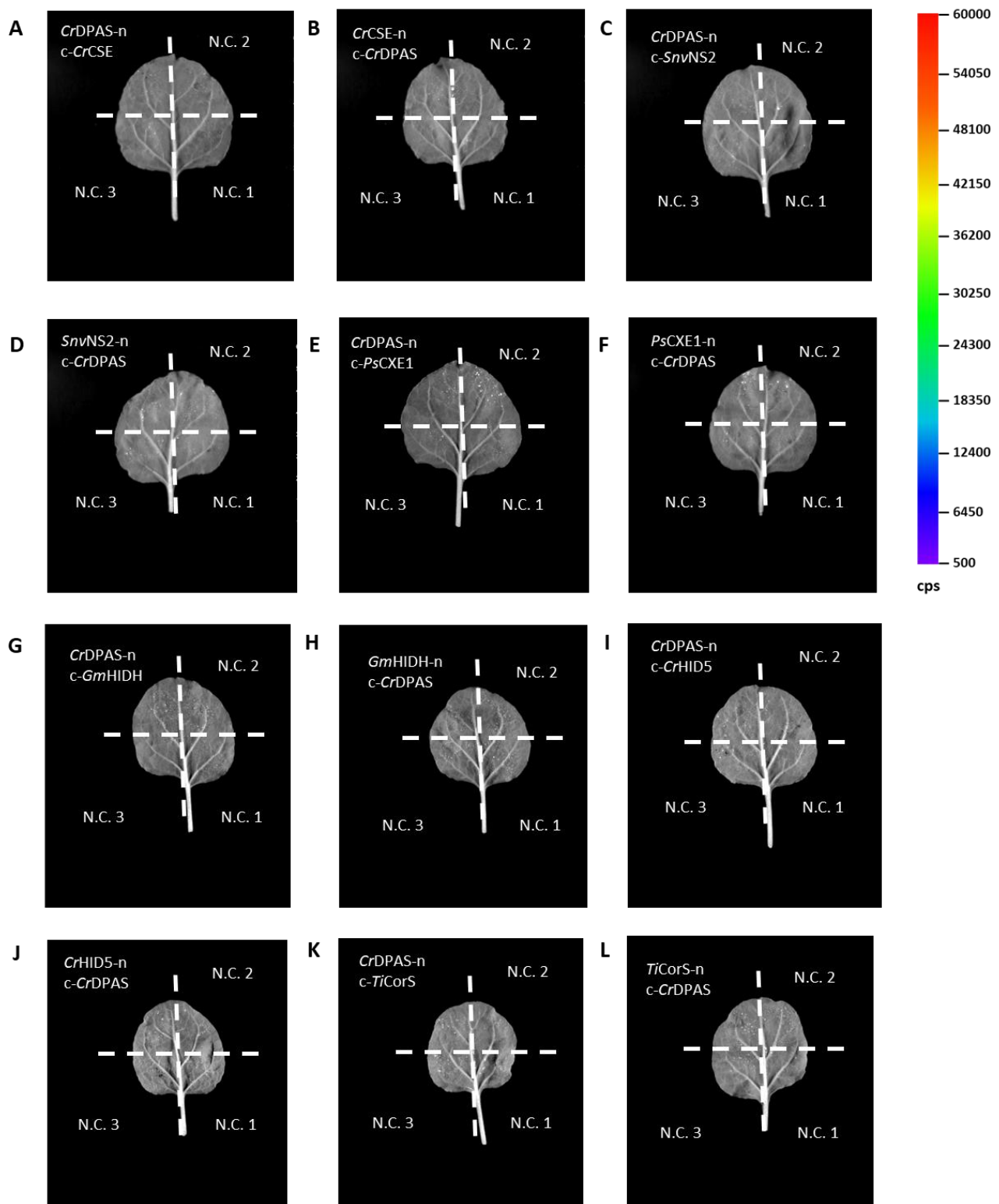


Figure 123. Representative images of pairwise interactions of *TiDPAS2* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *TiDPAS2-CrHID5*; **C-D.** *TiDPAS2-TiCorS*; **E-F.** *TiDPAS2-TiTabS*; **G-H.** *TiDPAS2-CrCorS*; **I-J.** *TiDPAS2-CrTS*; **K-L.** *TiDPAS2-CrCS*.



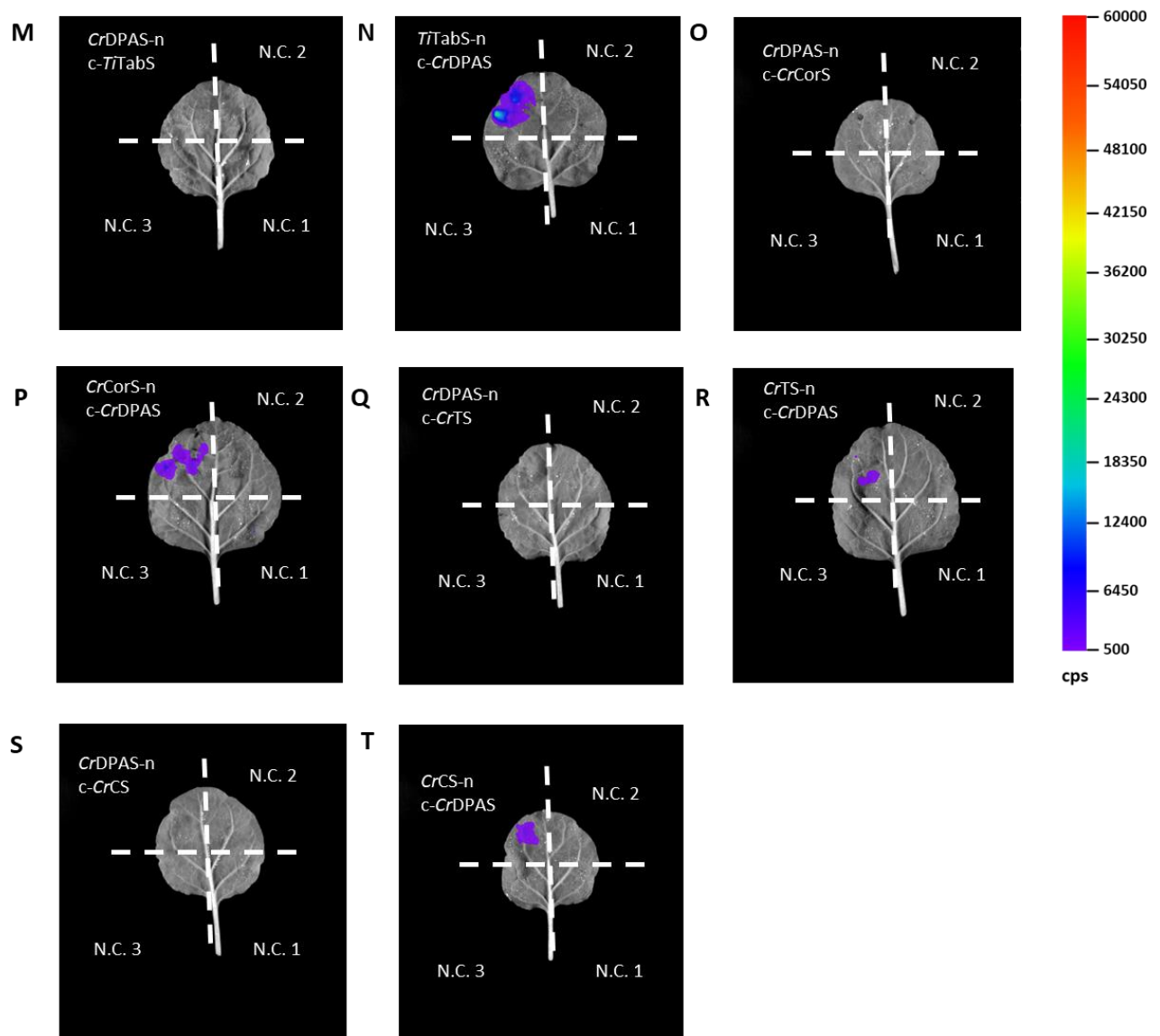
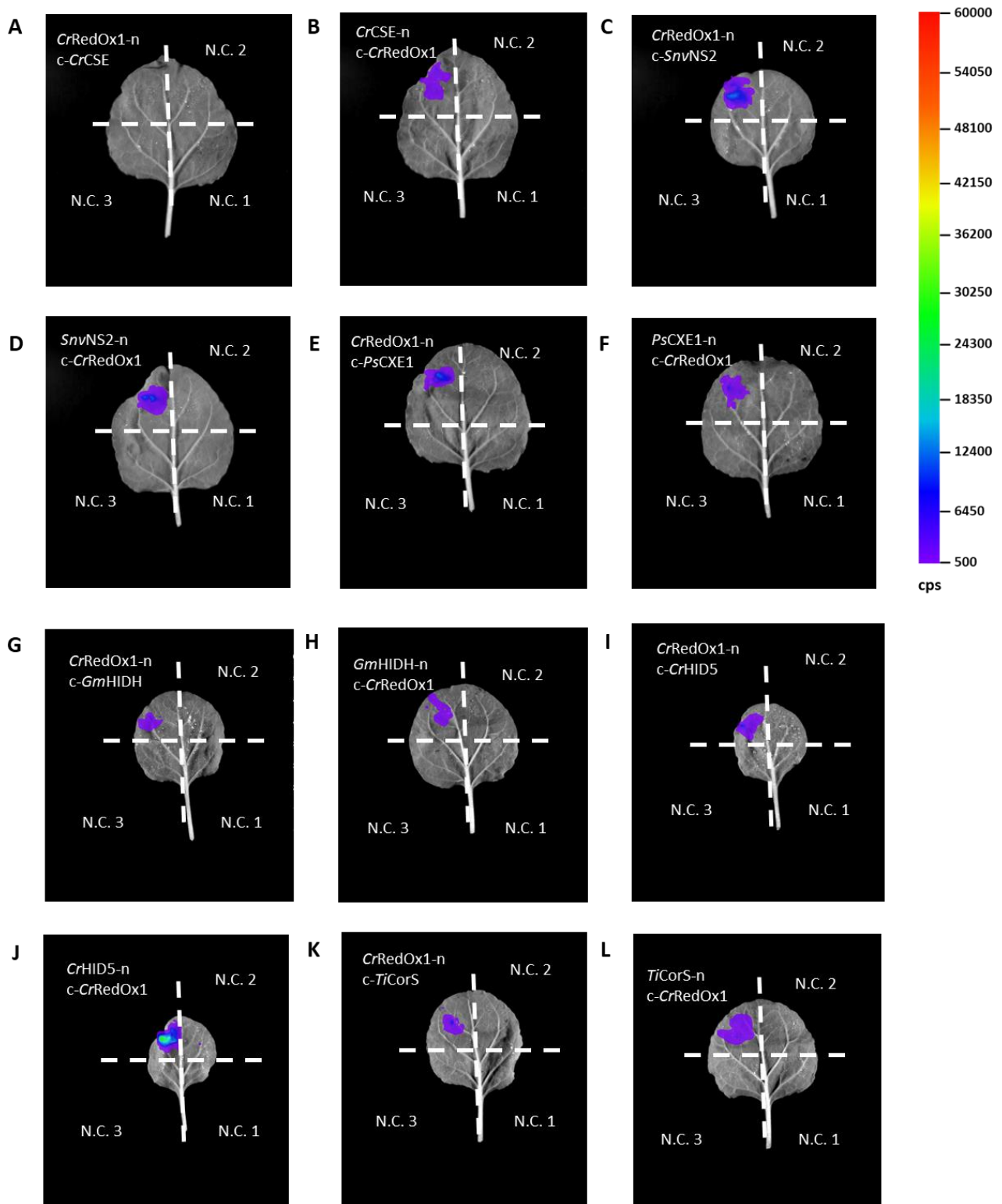


Figure 124. Representative images of pairwise interactions of *CrDPAS* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrDPAS-CrCSE*; **C-D.** *CrDPAS-SnvNS2*; **E-F.** *CrDPAS-PsCXE1*; **G-H.** *CrDPAS-GmHIDH*; **I-J.** *CrDPAS-CrHID5*; **K-L.** *CrDPAS-TiCorS*; **M-N.** *CrDPAS-TTabS*; **O-P.** *CrDPAS-CrCorS*; **Q-R.** *CrDPAS-CrTS*; **S-T.** *CrDPAS-CrCS*.



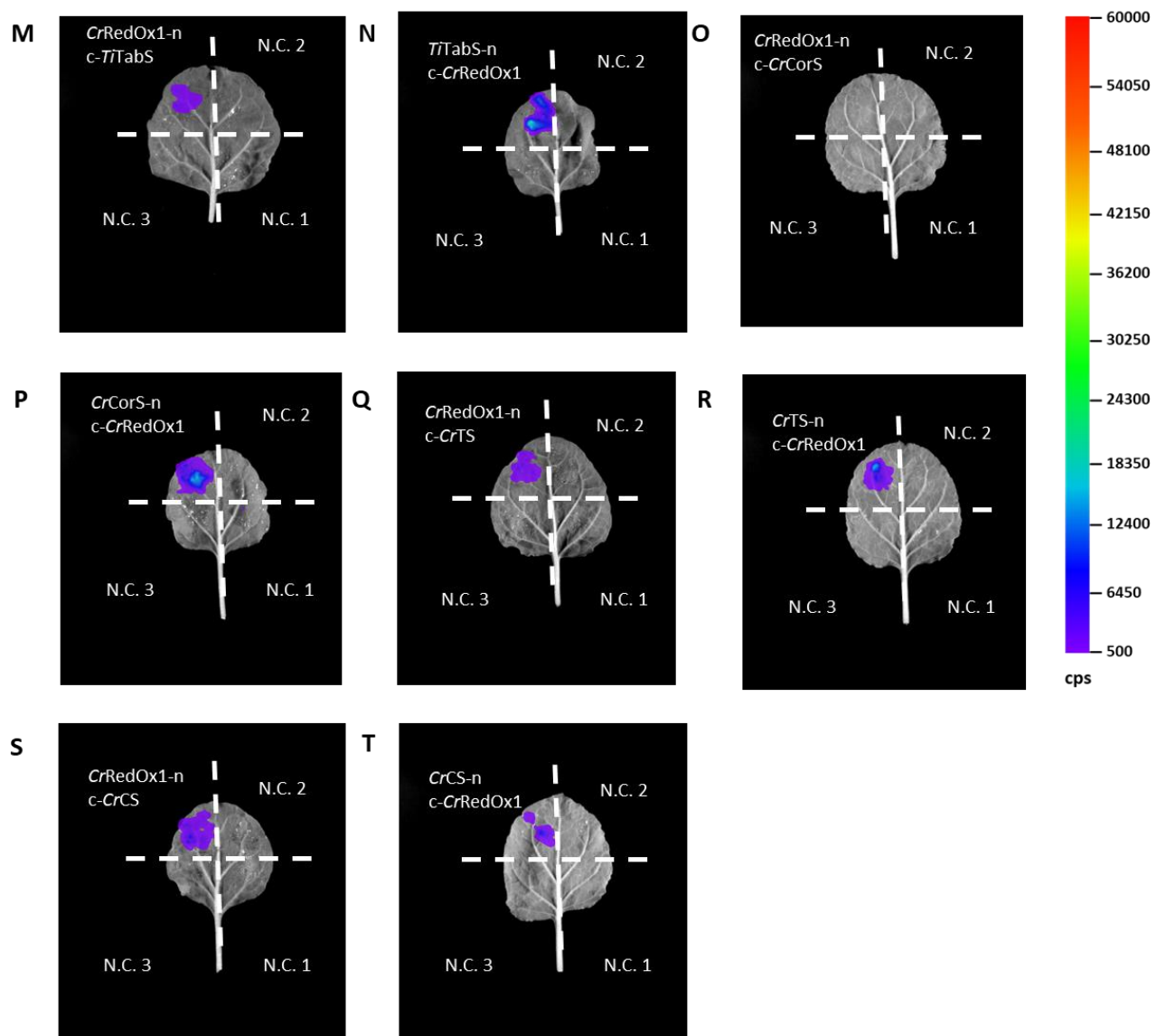
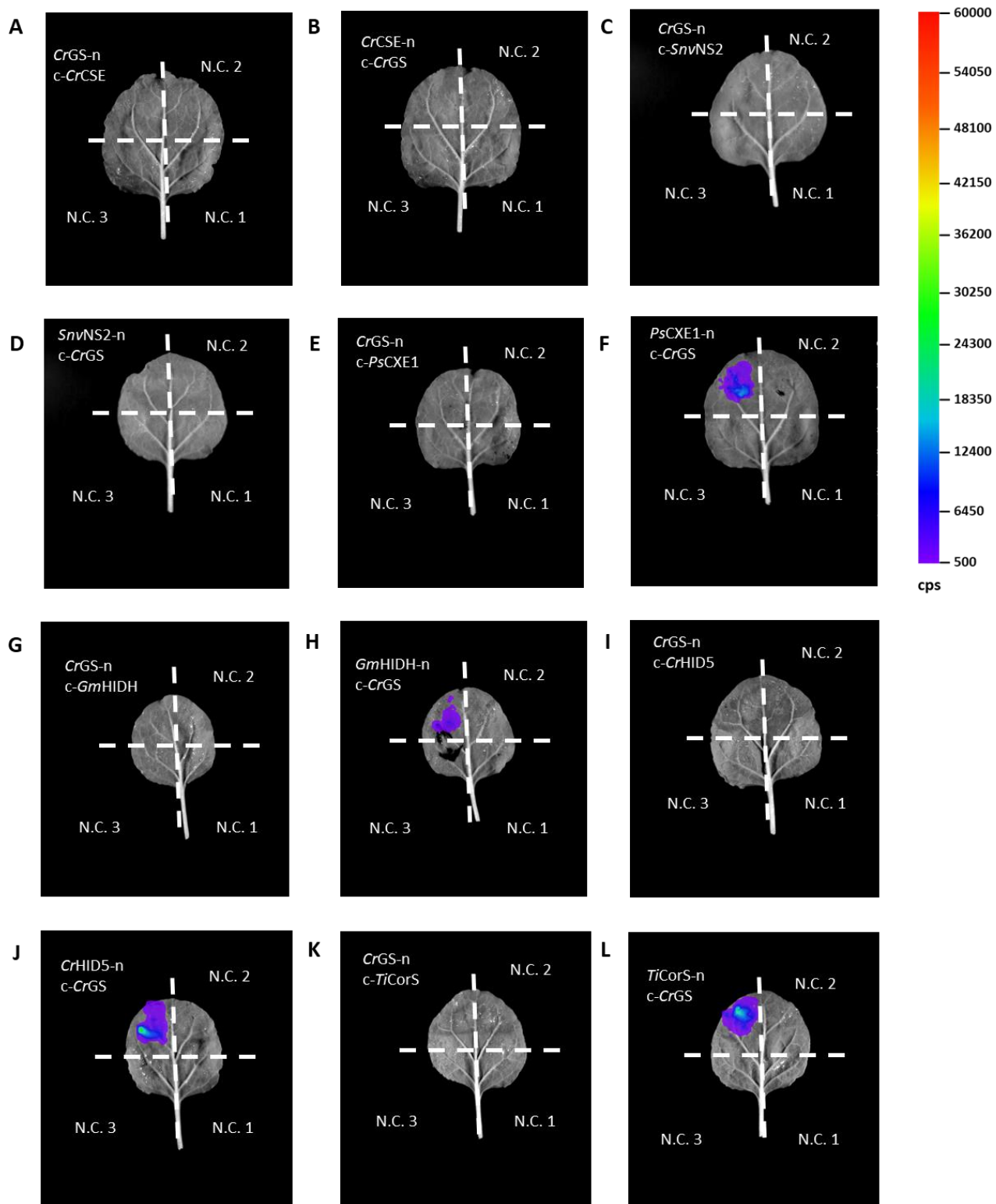


Figure 125. Representative images of pairwise interactions of *CrRedOx1* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrRedOx1-CrCSE*; **C-D.** *CrRedOx1-SnvNS2*; **E-F.** *CrRedOx1-PsCXE1*; **G-H.** *CrRedOx1-GmHIDH*; **I-J.** *CrRedOx1-CrHID5*; **K-L.** *CrRedOx1-TiCorS*; **M-N.** *CrRedOx1-TiTabS*; **O-P.** *CrRedOx1-CrCorS*; **Q-R.** *CrRedOx1-CrTS*; **S-T.** *CrRedOx1-CrCS*.



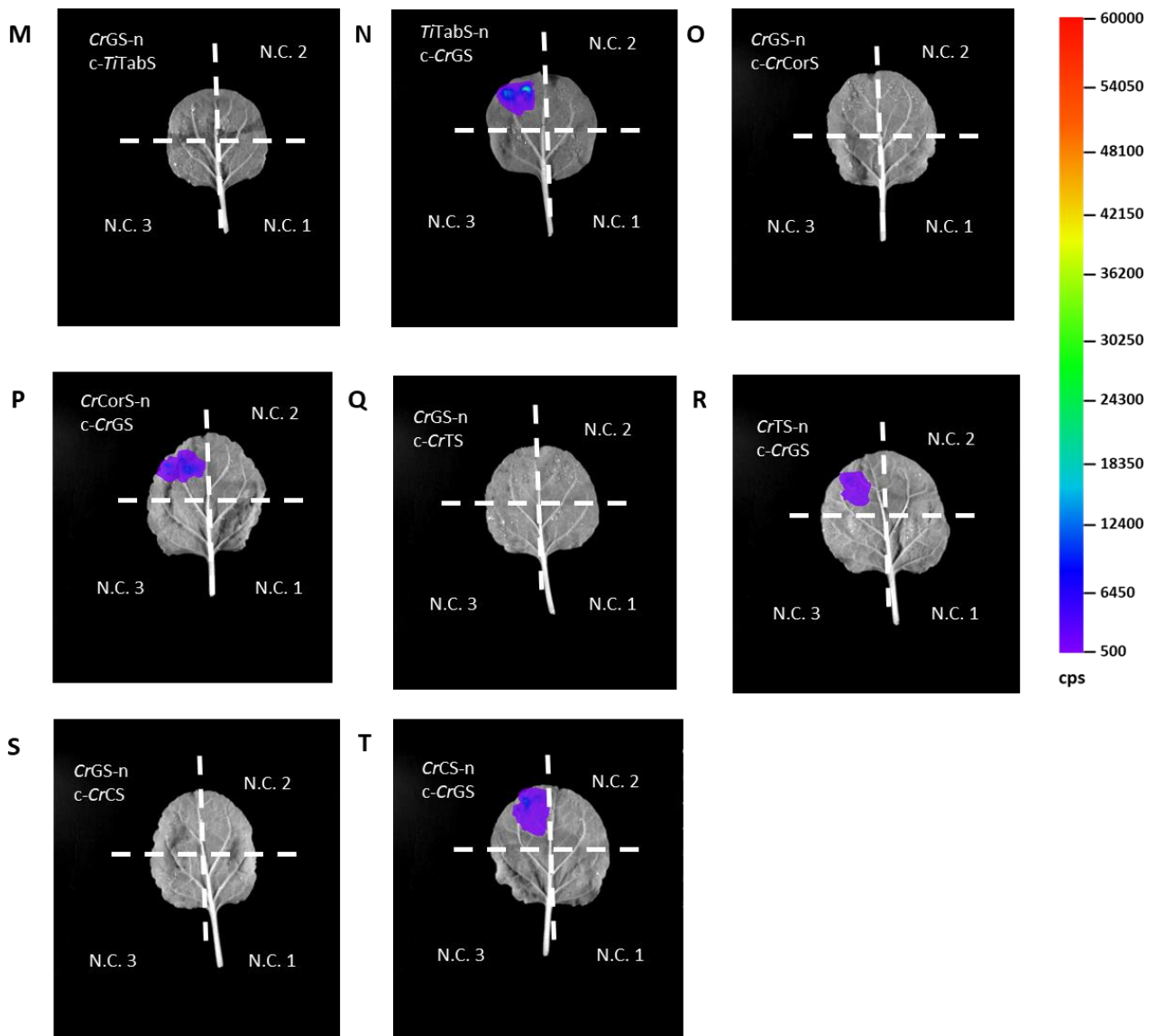


Figure 126. Representative images of pairwise interactions of *CrGS* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrGS-CrCSE*; **C-D.** *CrGS-SnvNS2*; **E-F.** *CrGS-PsCXE1*; **G-H.** *CrGS-GmHIDH*; **I-J.** *CrGS-CrHID5*; **K-L.** *CrGS-TiCorS*; **M-N.** *CrGS-TITabS*; **O-P.** *CrGS-CrCorS*; **Q-R.** *CrGS-CrTS*; **S-T.** *CrGS-CrCS*.

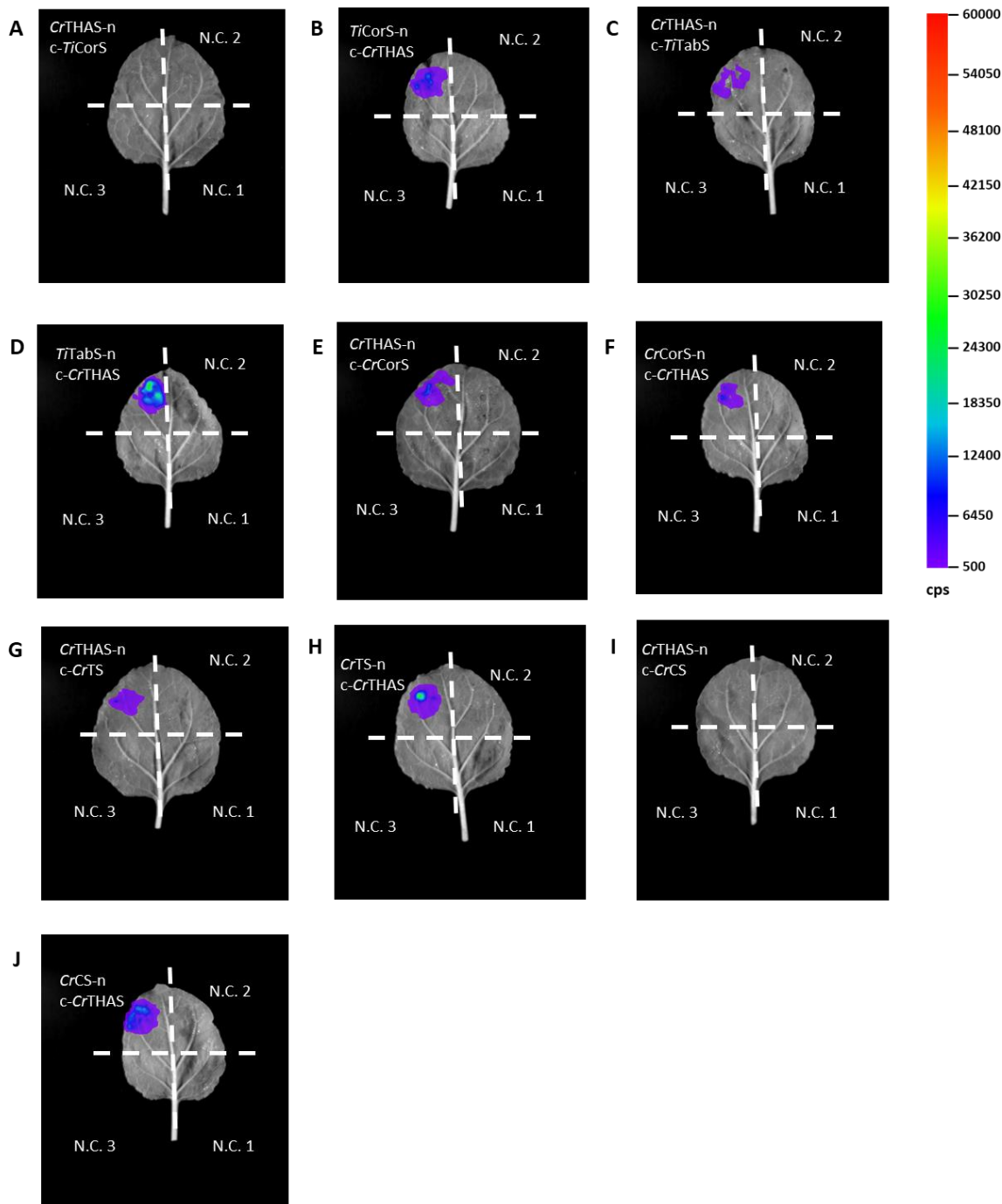
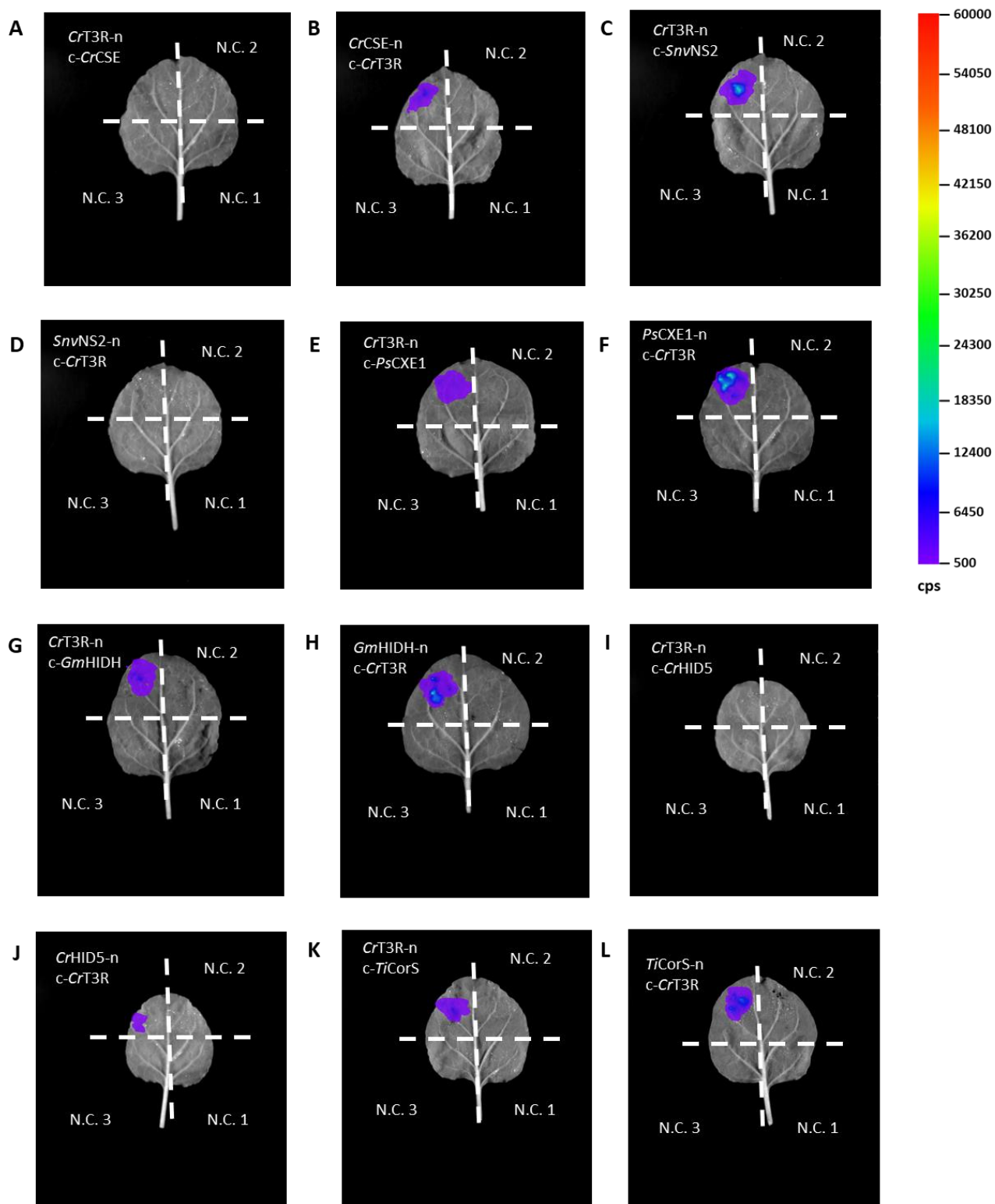


Figure 127. Representative images of pairwise interactions of *CrTHAS* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrTHAS-TiCorS*; **C-D.** *CrTHAS-TTTabS*; **E-F.** *CrTHAS-CrCorS*; **G-H.** *CrTHAS-CrTS*; **I-J.** *CrTHAS-CrCS*.



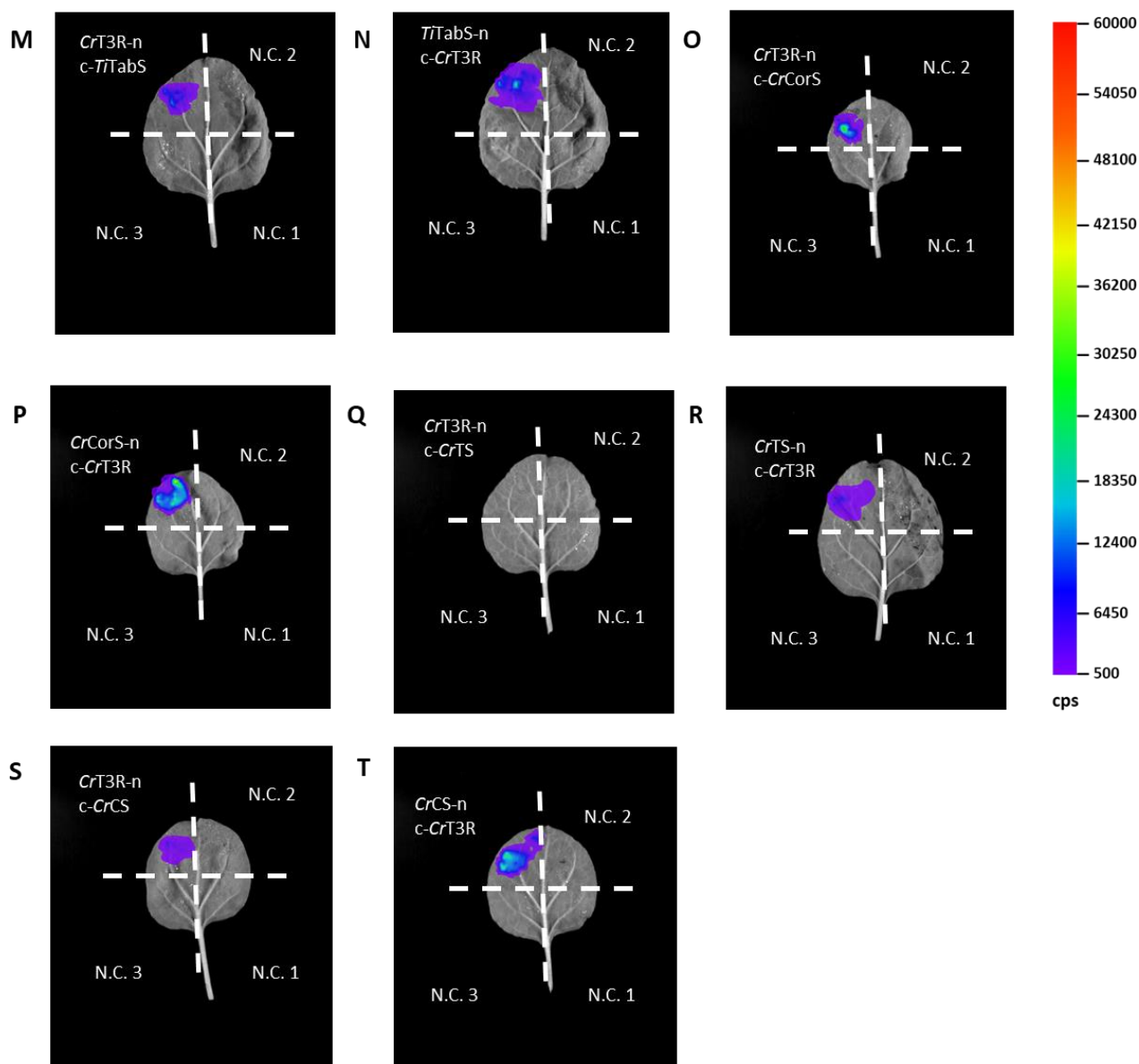


Figure 128. Representative images of pairwise interactions of *CrT3R* with α/β -hydrolases tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrT3R-CrCSE*; **C-D.** *CrT3R-SnvNS2*; **E-F.** *CrT3R-PsCXE1*; **G-H.** *CrT3R-GmHIDH*; **I-J.** *CrT3R-CrHID5*; **K-L.** *CrT3R-TiCorS*; **M-N.** *CrT3R-TTabS*; **O-P.** *CrT3R-CrCorS*; **Q-R.** *CrT3R-CrTS*; **S-T.** *CrT3R-CrCS*.

Appendix XI. Split-Luciferase Assays of Cyclase Enzyme Interaction Engineering

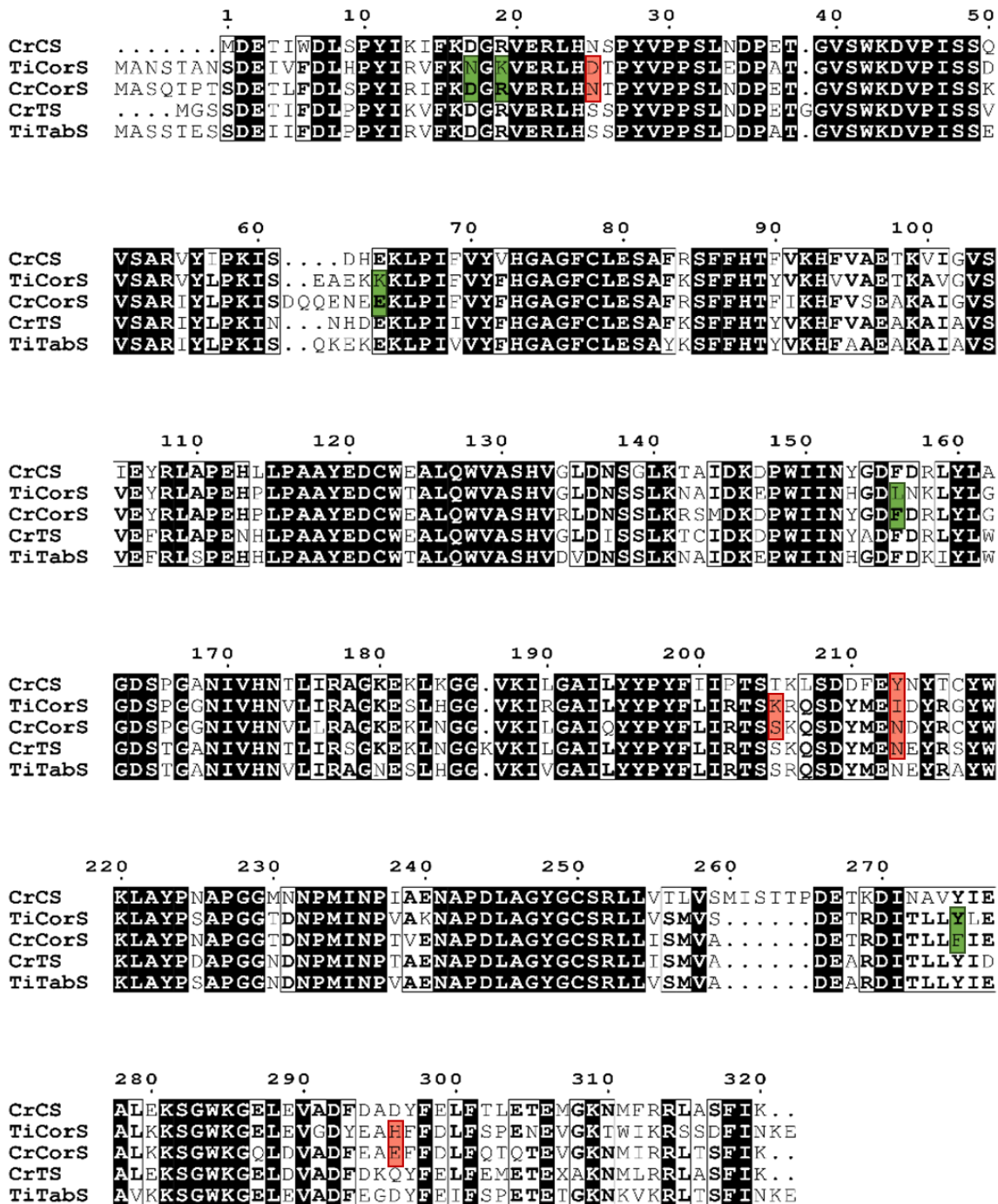
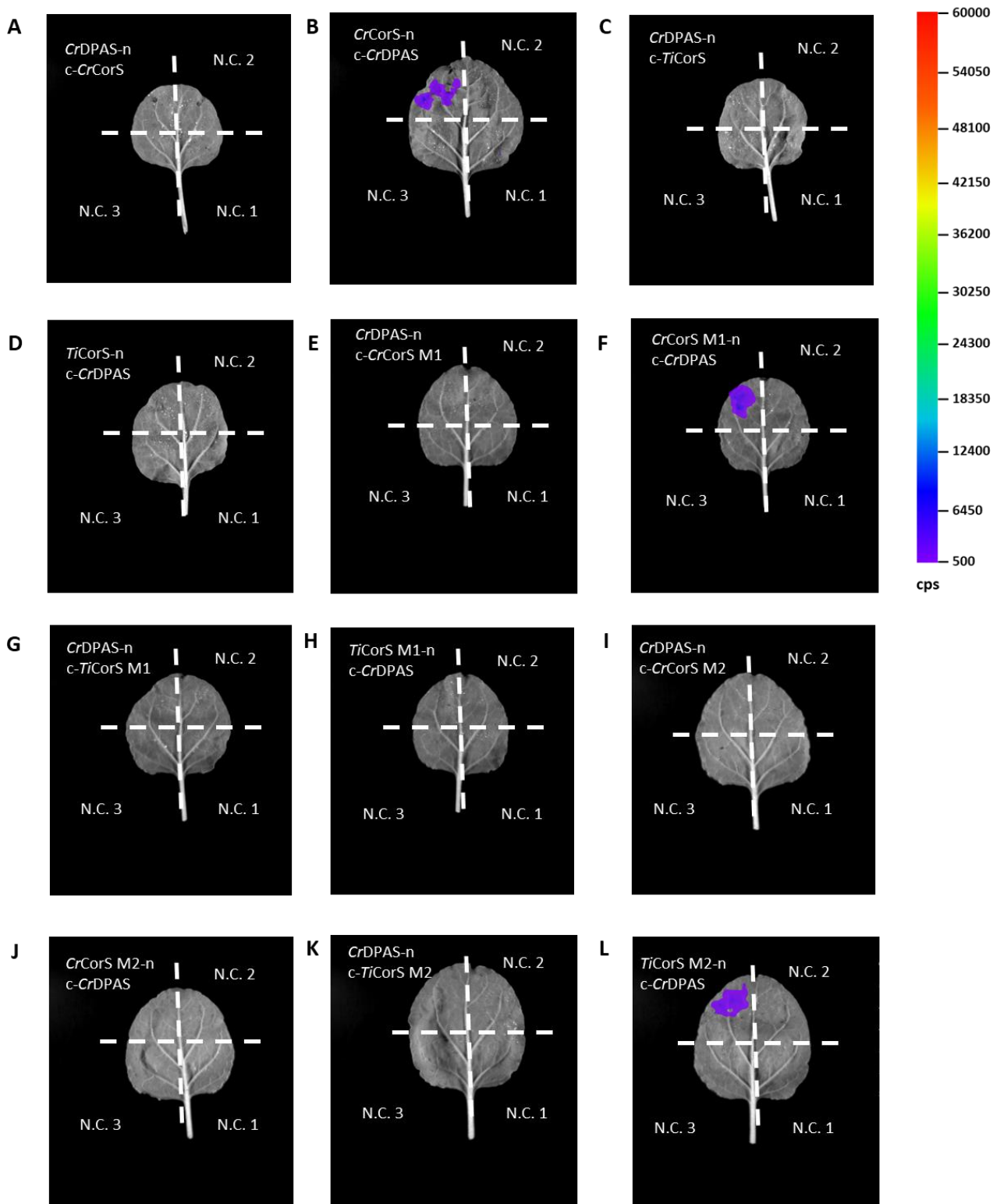


Figure 129. Amino acid sequence alignment of cyclase enzymes from *C. roseus* and *T. iboga*. Surface residues mutated in *CrCorS* and *TiCorS* in M1 highlighted in green, residues mutated in *CrCorS* and *TiCorS* in M2 highlighted in red. Figure made using ESPrnt 3.0 [56].



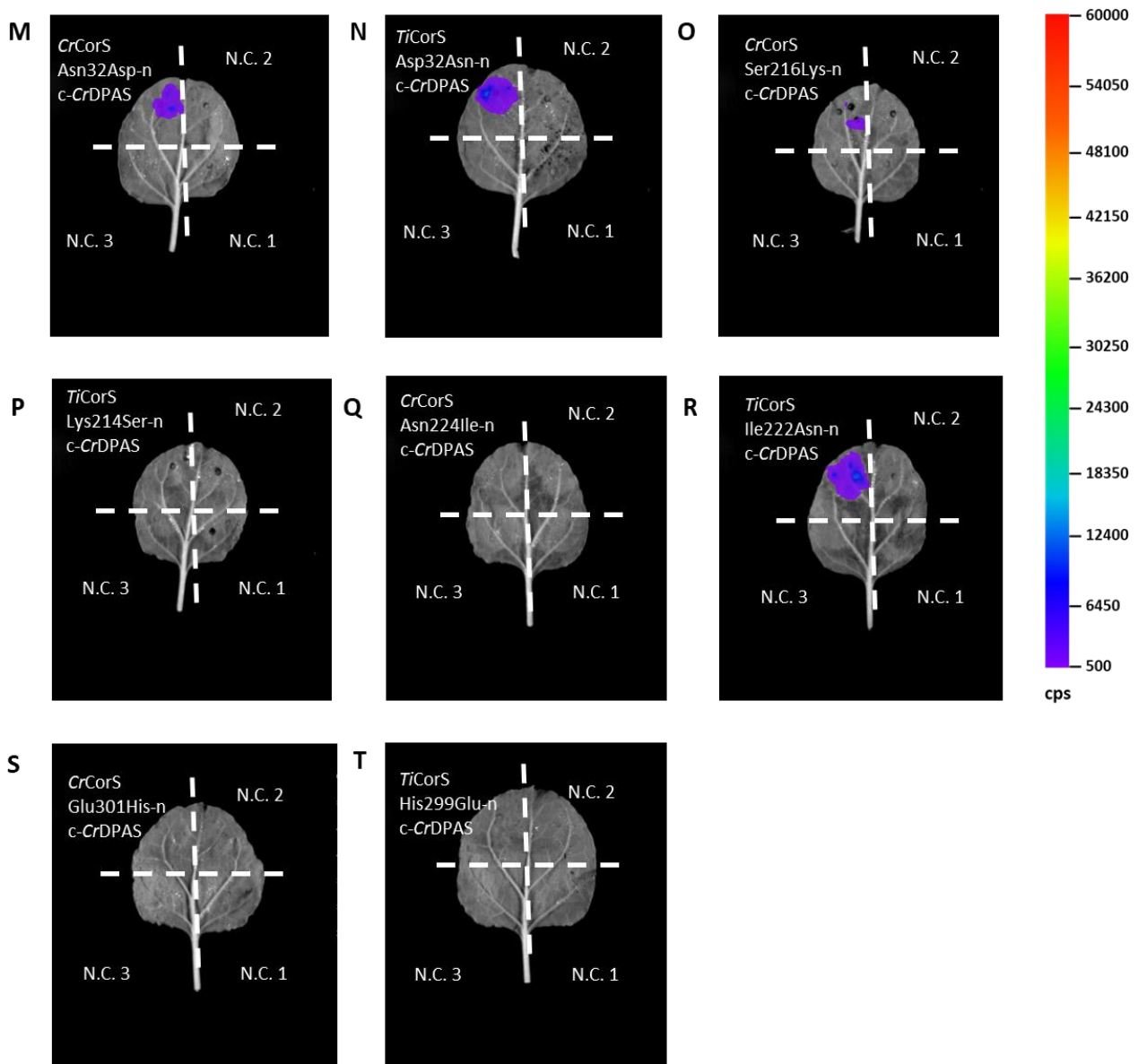


Figure 130. Representative images of pairwise interactions of CrDPAS interaction engineering with CrCors or TiCorS tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** CrDPAS-CrCorS; **C-D.** CrDPAS-TiCorS; **E-F.** CrDPAS-CrCorS M1; **G-H.** CrDPAS-TiCorS M1; **I-J.** CrDPAS-CrCorS M2; **K-L.** CrDPAS-TiCorS M2; **M.** CrCorS Asn32Asp-CrDPAS; **N.** TiCorS Asp32Asn-CrDPAS; **O.** CrCorS Ser216Lys-CrDPAS; **P.** TiCorS Lys214Ser-CrDPAS; **Q.** CrCorS Asn224Ile-CrDPAS; **R.** TiCorS Ile222Asn-CrDPAS; **S.** CrCorS Glu301His-CrDPAS; **T.** TiCorS His299Glu-CrDPAS.

Appendix XII. AP-MS of CrDPAS and CrTS

Table 26. AP-MS results using 6X-His CrDPAS or 6X-His CrTS as bait proteins in protein extract from *C. roseus* leaf tissue.

Accession number	Annotated name	Spectrum Counts		
		DPAS bait	TS bait	Blank bait
CRO_T011212	TS	6	1108	26
CRO_T033537	DPAS	303	0	0
CRO_T020363	MAR-binding filament-like protein	72	35	28
CRO_T005426	DNA-binding enhancer protein-related	63	50	12
CRO_T005174	hydroxyproline-rich glycoprotein family protein	59	32	14
CRO_T026844	zinc finger (CCCH-type) family protein / RNA recognition motif (RRM)-containing protein	26	12	0
CRO_T008025	plastid transcriptionally active	22	0	0
CRO_T011810	thioredoxin M-type	18	3	0
CRO_T026310	GDSL-like Lipase/Acylhydrolase superfamily protein	17	3	3
CRO_T024551	GDSL-like Lipase/Acylhydrolase superfamily protein	16	6	2
CRO_T015760	ZIM-like	13	8	0
CRO_T030110	nodulin-related protein	13	0	0
CRO_T018424	Lipase/lipoxygenase, PLAT/LH2 family protein	11	0	0
CRO_T001292	glyceraldehyde 3-phosphate dehydrogenase A subunit	10	0	0
CRO_T026604	conserved hypothetical protein	9	0	0
CRO_T020754	DUF1499 domain containing protein	9	0	0
CRO_T013082	Dehydrin family protein	7	0	0
CRO_T005451	uridylyltransferase-related	6	0	0
CRO_T018160	hypothetical protein	6	0	0
CRO_T003870	Pentatricopeptide repeat (PPR) superfamily protein	6	0	0
CRO_T009576	glyceraldehyde-3-phosphate dehydrogenase B subunit	6	0	0
CRO_T024124	Lipase/lipoxygenase, PLAT/LH2 family protein	5	0	0

Appendix XIII. DFS of CrDPAS

Table 27. Summary of CrDPAS $\Delta T^{\circ}m$ compared to CrDPAS control in various conditions investigating cyclase complex formation.

Complex condition	CrDPAS $\Delta T^{\circ}m$ ($^{\circ}C$)
2 μ M CrDPAS + 2 μ M CrTS + 10 μ M NADPH	0
2 μ M CrDPAS + 3 μ M CrTS + 10 μ M NADPH	1
2 μ M CrDPAS + 4 μ M CrTS + 10 μ M NADPH	1
2 μ M CrDPAS + 2 μ M CrCS + 10 μ M NADPH	0
2 μ M CrDPAS + 3 μ M CrCS + 10 μ M NADPH	1
2 μ M CrDPAS + 4 μ M CrCS + 10 μ M NADPH	1
2 μ M CrDPAS + 2 μ M CrTS + 10 μ M NADP ⁺	1
2 μ M CrDPAS + 3 μ M CrTS + 10 μ M NADP ⁺	1
2 μ M CrDPAS + 4 μ M CrTS + 10 μ M NADP ⁺	1
2 μ M CrDPAS + 2 μ M CrCS + 10 μ M NADP ⁺	1
2 μ M CrDPAS + 3 μ M CrCS + 10 μ M NADP ⁺	1
2 μ M CrDPAS + 4 μ M CrCS + 10 μ M NADP ⁺	0
2 μ M CrDPAS + 2 μ M CrTS + 10 μ M NADP ⁺ + 10 μ M precondylocarpine acetate	2
2 μ M CrDPAS + 3 μ M CrTS + 10 μ M NADP ⁺ + 10 μ M precondylocarpine acetate	1
2 μ M CrDPAS + 4 μ M CrTS + 10 μ M NADP ⁺ + 10 μ M precondylocarpine acetate	1
2 μ M CrDPAS + 2 μ M CrCS + 10 μ M NADP ⁺ + 10 μ M precondylocarpine acetate	0
2 μ M CrDPAS + 3 μ M CrCS + 10 μ M NADP ⁺ + 10 μ M precondylocarpine acetate	1
2 μ M CrDPAS + 4 μ M CrCS + 10 μ M NADP ⁺ + 10 μ M precondylocarpine acetate	0

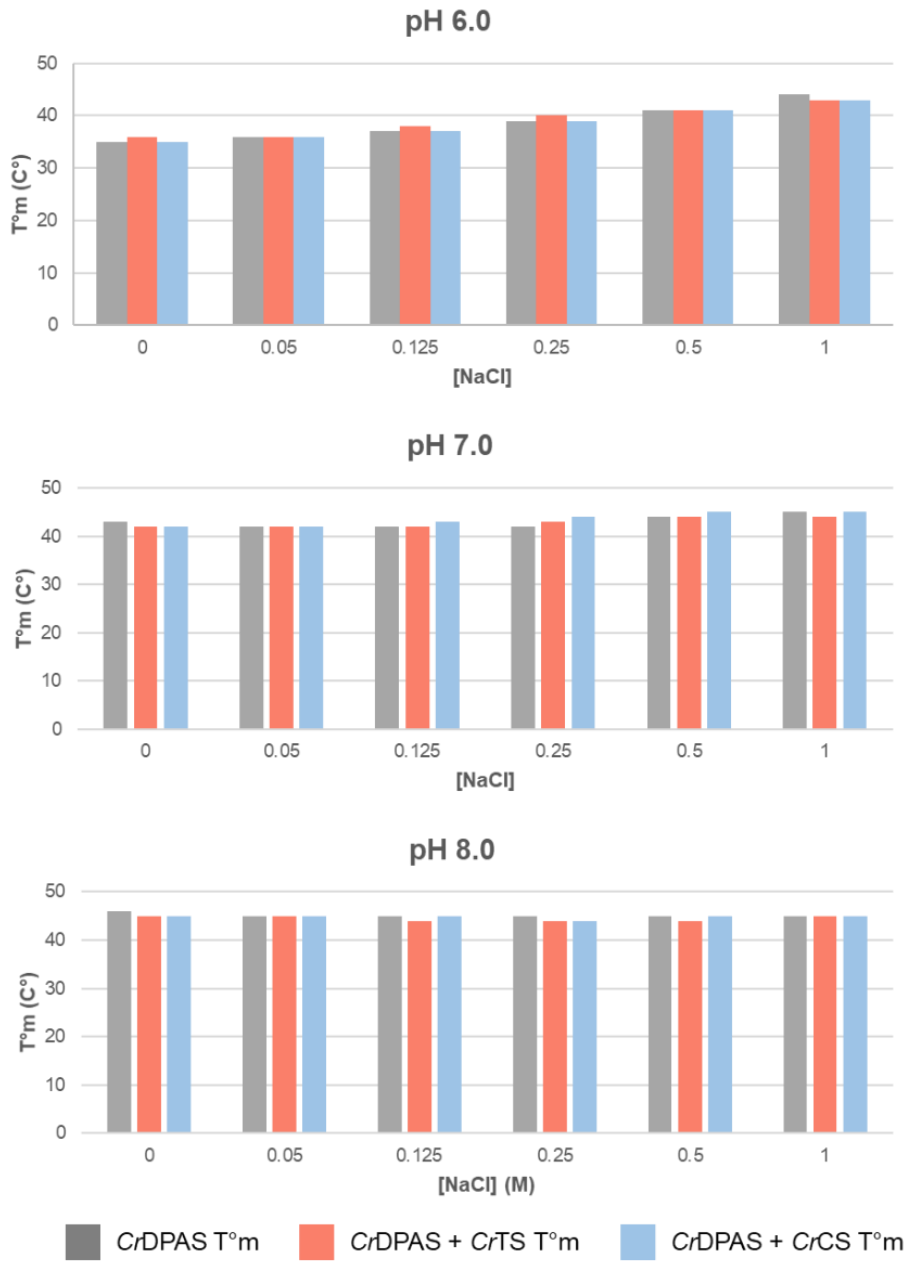


Figure 131. Effect of pH and $[NaCl]$ on $CrDPAS T^{\circ}m$ in the presence or absence of cyclase $CrTS$ or $CrCS$ added at equimolar concentrations.

Appendix XIV. CrDPAS *In Vivo* Proximity Tagging

Table 28. Contig matches with ≥ 1.5 log fold change in peptide hits between CrDPAS-TurboID protein pulldown compared to YFP-TurboID protein pulldown control. Contigs encoding known proteins involved in iridoid, MIA or phenylpropanoid biosynthesis are highlighted.

Contig	Description	logFC DPAS_TurboID vs. YFP_TurboID
CRO_T112140 CRO_T139361	photosystem II reaction center protein D	4.33
CRO_T110594	photosystem II family protein	4.07
CRO_T101917	photosystem II light harvesting complex gene 2.1	3.97
CRO_T125675	Chlorophyll A-B binding family protein	3.92
CRO_T117139	photosystem I subunit E-2	3.56
CRO_T137836	voltage dependent anion channel	3.52
CRO_T130227	Photosystem I, PsaA/PsaB protein	3.35
CRO_T118278	CAAD domains of cyanobacterial aminoacyl-tRNA synthetase domain containing protein	3.24
CRO_T133555	photosystem I subunit F	3.15
CRO_T125574	light harvesting complex of photosystem II	3.12
CRO_T103006	calcium sensing receptor	3.05
CRO_T133025	photosystem II subunit O-2	3.03
CRO_T103910	FtsH extracellular protease family	3.03
CRO_T106640	Tetratricopeptide repeat (TPR)-like superfamily protein	2.98
CRO_T133139	Major facilitator superfamily protein	2.95
CRO_T132223	Photosystem I, PsaA/PsaB protein	2.90
CRO_T112598	peroxin 11c	2.90
CRO_T105753	thylakoid lumen 18.3 kDa protein	2.88
CRO_T131591	3-oxo-5-alpha-steroid 4-dehydrogenase family protein	2.88
CRO_T120926	Inorganic H pyrophosphatase family protein	2.86
CRO_T122644	sulfate transmembrane transporters	2.85
CRO_T131186	ATP synthase subunit beta	2.83
CRO_T108207	NAD(P)-binding Rossmann-fold superfamily protein	2.82
CRO_T121366	Outer envelope pore protein 24, chloroplastic	2.77
CRO_T138933	BCL-2-associated athanogene	2.76
CRO_T109159	ATPase, F0 complex, subunit B/B', bacterial/chloroplast	2.74
CRO_T127711	MAR binding filament-like protein	2.67
CRO_T140657	light harvesting complex photosystem II	2.67
CRO_T110634	Rhodanese/Cell cycle control phosphatase superfamily protein	2.60
CRO_T130034	cytochrome P450, family 81, subfamily D, polypeptide	2.60

CRO_T108889	Rhodanese/Cell cycle control phosphatase superfamily protein	2.60
CRO_T131207	cytochrome P450, family 71, subfamily B, polypeptide CYP71BT1	2.58
CRO_T117548	Protein of unknown function (DUF3411)	2.57
CRO_T131660	ATPase, F1 complex, gamma subunit protein	2.57
CRO_T111279	Rhodanese/Cell cycle control phosphatase superfamily protein	2.52
CRO_T124321	Protein of unknown function DUF106, transmembrane	2.50
CRO_T110723	FtsH extracellular protease family	2.49
CRO_T113334	sugar transporter	2.49
CRO_T109968	STT7 homolog STN7	2.49
CRO_T133479	sodium/calcium exchanger family protein / calcium-binding EF hand family protein	2.43
CRO_T133143	translocon at the outer envelope membrane of chloroplasts 75-III	2.41
CRO_T121413	3-ketoacyl-CoA synthase	2.41
CRO_T101960	Auxin-responsive family protein	2.39
CRO_T104118 CRO_T121814	prohibitin	2.39
CRO_T110421	Protein TRIGALACTOSYLDIACYLGLYCEROL 4, chloroplastic	2.37
CRO_T104055	sodium/calcium exchanger family protein / calcium-binding EF hand family protein	2.37
CRO_T105084	synaptotagmin A	2.36
CRO_T140744	Mitochondrial substrate carrier family protein	2.35
CRO_T114330 CRO_T114331	chlorophyll A/B binding protein	2.33
CRO_T135151	membrane-associated progesterone binding protein MSBP	2.33
CRO_T138610	photosystem II subunit R	2.32
CRO_T115362	strictosidine synthase-like STR-like	2.32
CRO_T116728	conserved hypothetical protein	2.31
CRO_T131097	H(+)-ATPase	2.31
CRO_T105683	endomembrane-type CA-ATPase	2.30
CRO_T130288	plasma membrane intrinsic protein 1C	2.30
CRO_T124780	Pheophorbide a oxygenase family protein with Rieske [2Fe-2S] domain	2.30
CRO_T101459	thylakoid-associated phosphatase	2.29
CRO_T127515	vacuolar proton ATPase A3	2.29
CRO_T139710	photosystem I subunit D-2	2.28
CRO_T137837	SPFH/Band 7/PHB domain-containing membrane-associated protein family	2.28
CRO_T106319	Protein of unknown function (DUF1682)	2.28
CRO_T138504	cellulose synthase like E1	2.28

CRO_T124886	MORN (Membrane Occupation and Recognition Nexus) repeat-containing protein	2.27
CRO_T112921	NADH:cytochrome B5 reductase	2.27
CRO_T126592	Protein kinase superfamily protein	2.27
CRO_T132921	Mitochondrial substrate carrier family protein	2.26
CRO_T110069	conserved hypothetical protein	2.26
CRO_T103367	allantoinase	2.25
CRO_T116935	peptide transporter	2.24
CRO_T117189	cytochrome B5 isoform B	2.24
CRO_T124539	Mitochondrial substrate carrier family protein	2.23
CRO_T114993	conserved hypothetical protein	2.23
CRO_T103026	signal peptide peptidase	2.23
CRO_T133576	Kunitz family trypsin and protease inhibitor protein	2.22
CRO_T135117	Remorin family protein	2.22
CRO_T134088	prohibitin	2.22
CRO_T140630	thylakoid ATP/ADP carrier	2.21
CRO_T120082	DnaJ / Sec63 Brl domains-containing protein	2.21
CRO_T124194	alpha-mannosidase	2.20
CRO_T134480	Mitochondrial ATP synthase subunit G protein	2.18
CRO_T101224	rubredoxin family protein	2.18
CRO_T118762	H(+)-ATPase	2.17
CRO_T122302	gamma carbonic anhydrase	2.17
CRO_T113675	conserved hypothetical protein	2.15
CRO_T124573	Carbohydrate-binding-like fold	2.14
CRO_T138396	trigalactosyldiacylglycerol2	2.14
CRO_T124261	copper ion binding;cobalt ion binding;zinc ion binding	2.13
CRO_T124454	DNA repair ATPase-related	2.13
CRO_T110277	calnexin	2.13
CRO_T139024	catalytics	2.12
CRO_T140929	pleiotropic drug resistance	2.12
CRO_T138994	Iridiod oxidase IO	2.11
CRO_T139096	Protein TIC 56, chloroplastic	2.10
CRO_T121266	ferulic acid 5-hydroxylase F5H	2.10
CRO_T121397	acclimation of photosynthesis to environment	2.09
CRO_T129147		2.09
CRO_T129149	Major facilitator superfamily protein	2.09
CRO_T110600	cytochrome P450, family 71, subfamily B, polypeptide	2.08
CRO_T116078	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	2.08
CRO_T127029	cytochrome P450, family 82, subfamily C, polypeptide	2.08
CRO_T124635	chloroplast outer envelope protein	2.07
CRO_T138614	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	2.06
CRO_T137443	wall-associated kinase	2.03
CRO_T129416	strictosidine synthase-like STR-like	2.03

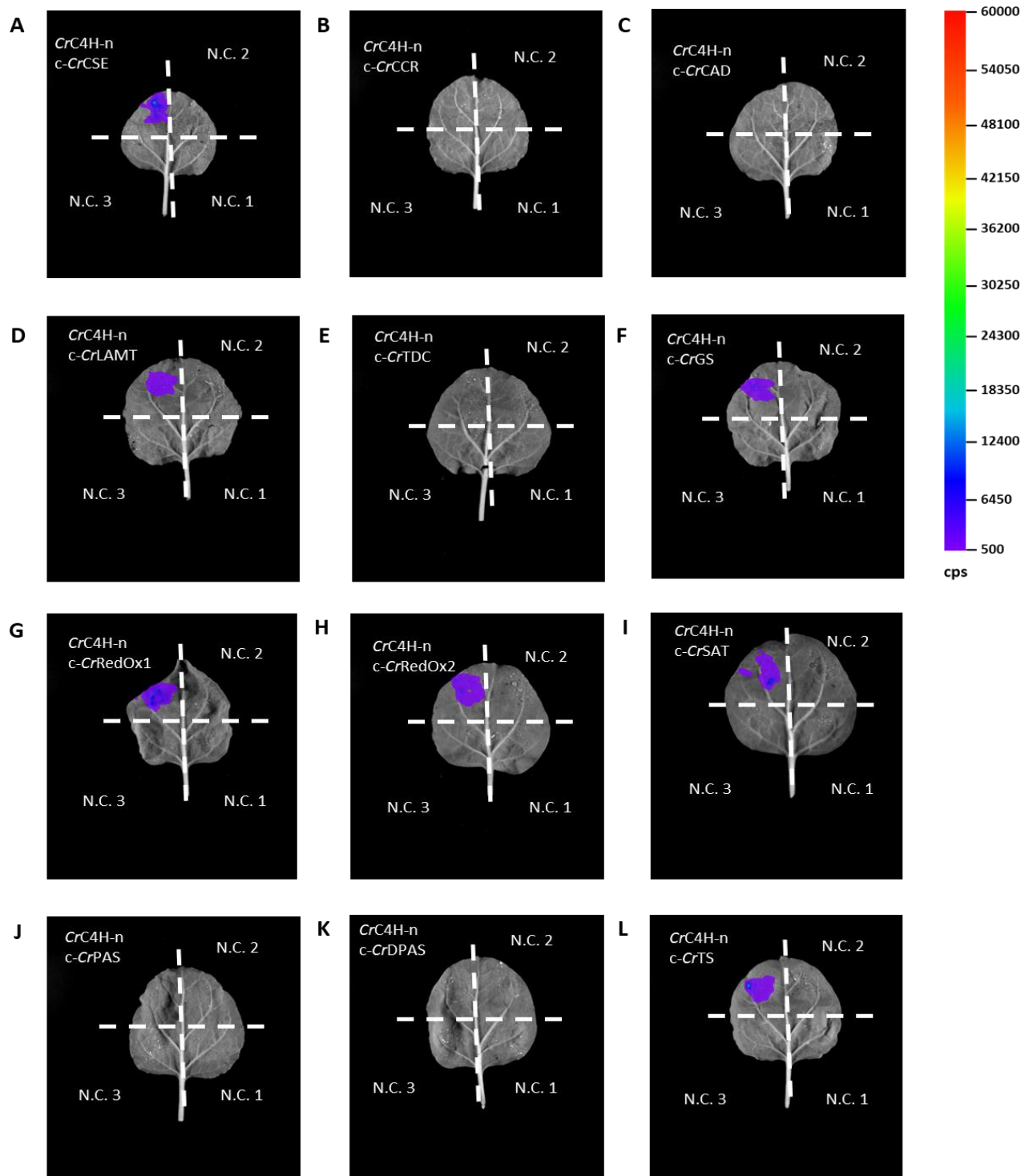
CRO_T127776	Outer envelope pore protein 24A, chloroplastic	2.03
CRO_T117246	FK506-binding protein 15 kD-1	2.02
CRO_T121327	AAA-type ATPase family protein	2.02
CRO_T137859	Leucine-rich repeat protein kinase family protein	2.00
CRO_T140376	ATP synthase D chain, mitochondrial	2.00
CRO_T117694	hexokinase	2.00
CRO_T100981	plant uncoupling mitochondrial protein	1.99
CRO_T131978	cytochrome P450, family 706, subfamily A, polypeptide	1.98
CRO_T131128	SPFH/Band 7/PHB domain-containing membrane-associated protein family	1.97
CRO_T111252	nitrate transmembrane transporters	1.97
CRO_T108532	cytochrome P450, family 77, subfamily A, polypeptide	1.96
CRO_T128640	voltage dependent anion channel	1.95
CRO_T125663	Rieske (2Fe-2S) domain-containing protein	1.95
CRO_T133061	Geraniol 8-hydroxylase G8H	1.94
CRO_T113510	multidrug resistance-associated protein	1.94
CRO_T128882	Plastid-lipid associated protein PAP / fibrillin family protein	1.94
CRO_T108517	pleiotropic drug resistance	1.93
CRO_T113501	cell elongation protein / DWARF1 / DIMINUTO (DIM)	1.93
CRO_T123749	Glycosyl hydrolase superfamily protein	1.93
CRO_T120479	B-cell receptor-associated 31-like	1.92
CRO_T105449	PDI-like 5-2	1.92
CRO_T113642	ABC2 homolog	1.91
CRO_T100235	delta subunit of Mt ATP synthase	1.90
CRO_T133734	calreticulin 1a	1.89
CRO_T129457	germin	1.89
CRO_T124651	alpha/beta-Hydrolases superfamily protein	1.89
CRO_T105640	Transmembrane proteins 14C	1.88
CRO_T124913	adenine nucleotide transporter	1.87
CRO_T123688	non-intrinsic ABC protein	1.87
CRO_T128469	transporter associated with antigen processing protein	1.87
CRO_T123657	PDI-like 1-1	1.87
CRO_T105618		1.86
CRO_T105619		
CRO_T105621		
CRO_T105684	Histone superfamily protein	
CRO_T127975	Subtilisin-like serine endopeptidase family protein	1.86
CRO_T100278	ADP/ATP carrier	1.85
CRO_T110680	Ribophorin I	1.84
CRO_T108938	Leucine-rich repeat (LRR) family protein	1.83
CRO_T129178	conserved hypothetical protein	1.83
CRO_T122157	SEC12P-like 2 protein	1.83
CRO_T113597	peptide transporter	1.83
CRO_T105087	ascorbate peroxidase	1.82
CRO_T137738	Insulinase (Peptidase family M16) protein	1.81

CRO_T140881	cytochrome P450, family 81, subfamily K, polypeptide	1.80
CRO_T122648	nucleotide transporter	1.80
CRO_T126869	Ribophorin I	1.80
CRO_T127016	Subtilase family protein	1.80
CRO_T124766	gamma tonoplast intrinsic protein	1.80
CRO_T134427	Insulinase (Peptidase family M16) protein	1.79
CRO_T138838	translocon at the outer envelope membrane of chloroplasts	1.79
CRO_T105023	sterol methyltransferase	1.78
CRO_T133407	Curculin-like (mannose-binding) lectin family protein	1.78
CRO_T117787	conserved hypothetical protein	1.77
CRO_T118263	VIRB2-interacting protein	1.77
CRO_T113918	alpha/beta-Hydrolases superfamily protein	1.77
CRO_T141150	Peroxidase superfamily protein	1.77
CRO_T111277	cytochrome P450, family 71, subfamily A, polypeptide	1.76
CRO_T126954	plasma membrane intrinsic protein 2;5	1.76
CRO_T124369	Ubiquitin-specific protease family C19-related protein	1.76
CRO_T135052	conserved hypothetical protein	1.76
CRO_T109465 CRO_T109472 CRO_T113655	Secologanin synthase SLS	1.75
CRO_T119647	GRIM-19 protein	1.75
CRO_T131784	phospholipase C	1.75
CRO_T110794	NAD(P)-binding Rossmann-fold superfamily protein	1.74
CRO_T112631	Remorin family protein	1.74
CRO_T123311	MATE efflux family protein	1.74
CRO_T121601	Single hybrid motif superfamily protein	1.74
CRO_T134187	Protein of unknown function (DUF3754)	1.74
CRO_T101292	SPFH/Band 7/PHB domain-containing membrane-associated protein family	1.73
CRO_T106860	annexin	1.72
CRO_T138528	Outward rectifying potassium channel protein	1.71
CRO_T103125	voltage dependent anion channel	1.71
CRO_T137206	hexokinase	1.71
CRO_T102025	Major facilitator superfamily protein	1.71
CRO_T116310	autoinhibited Ca ²⁺ -ATPase	1.71
CRO_T103881	cytochrome P450, family 72, subfamily A, polypeptide	1.69
CRO_T117079	ribophorin II (RPN2) family protein	1.69
CRO_T113525	amino acid transporter	1.68
CRO_T141060	copper ion binding;cobalt ion binding;zinc ion binding	1.68
CRO_T132186	ATP synthase alpha/beta family protein	1.68
CRO_T126144	DUF1517 domain containing protein	1.67
CRO_T134124	translocase of outer membrane 20 kDa subunit	1.67
CRO_T130743	PDI-like 1-4	1.67
CRO_T134091	lysine histidine transporter	1.66

CRO_T109497	sterol methyltransferase	1.66
CRO_T103873	7-deoxyloganic acid hydroxylase 7DLH	1.66
CRO_T120100	Zn-dependent exopeptidases superfamily protein	1.65
CRO_T110779	polyol/monosaccharide transporter	1.65
CRO_T139926	receptor like protein	1.65
CRO_T132383	outer envelope protein of 80 kDa	1.64
CRO_T109875	ATP synthase subunit alpha	1.64
CRO_T130538	AAA-type ATPase family protein	1.64
CRO_T115920	Histone superfamily protein	1.62
CRO_T111337	cytochrome P450, family 71, subfamily B, polypeptide	1.62
CRO_T101250	TMPIT-like protein	1.62
CRO_T124983	cytochrome P450, family 98, subfamily A, polypeptide	1.61
CRO_T118238	Protein kinase superfamily protein	1.61
CRO_T101194	AAA-type ATPase family protein	1.61
CRO_T131467	Oligosaccharyltransferase complex/magnesium transporter family protein	1.61
CRO_T138641	Leucine-rich repeat transmembrane protein kinase	1.60
CRO_T111384	ATP binding cassette subfamily B1	1.60
CRO_T130116	Heat shock protein 70 (Hsp 70) family protein	1.60
CRO_T107192	pectin methylesterase	1.60
CRO_T101280	Mitochondrial import inner membrane translocase subunit Tim17/Tim22/Tim23 family protein	1.60
CRO_T113477	Early-responsive to dehydration stress protein (ERD4)	1.59
CRO_T118877	calcium ATPase	1.59
CRO_T133593	PDI-like 1-2	1.58
CRO_T116686	conserved hypothetical protein	1.58
CRO_T122630	LETM1-like protein	1.58
CRO_T122503	Cytochrome P450 superfamily protein	1.58
CRO_T117258	Auxin-responsive family protein	1.58
CRO_T127722	autoinhibited H(+)-ATPase isoform	1.57
CRO_T139269	cytochrome BC1 synthesis	1.57
CRO_T113237	NADH-ubiquinone dehydrogenase, mitochondrial, putative	1.56
CRO_T119412	tobamovirus multiplication 2A	1.55
CRO_T133290	Translation initiation factor 3 protein	1.55
CRO_T117377	translocon at the inner envelope membrane of chloroplasts	1.54
CRO_T117321	translocase of the outer mitochondrial membrane	1.53
CRO_T139711	PLAT/LH2 domain-containing lipoxygenase family protein	1.52
CRO_T124973	syntaxin of plants	1.51
CRO_T101649	Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	1.50
CRO_T119276	hypothetical protein	1.50

Appendix XV. Split-Luciferase Assays between *C. roseus* MIA and Phenylpropanoid Biosynthetic Enzymes

Representative split-luciferase result of testing protein-protein interactions of *C. roseus* phenylpropanoid and MIA biosynthetic enzymes in heterologous host *N. benthamiana*.



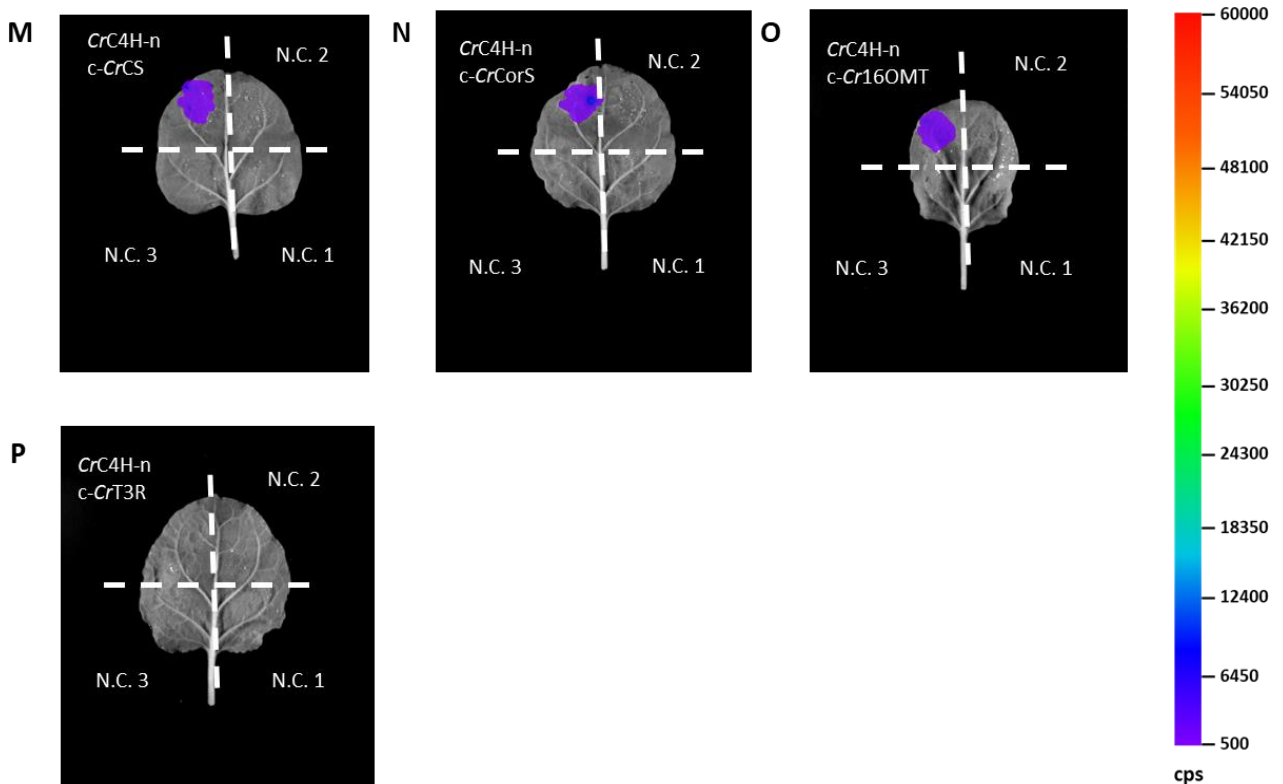
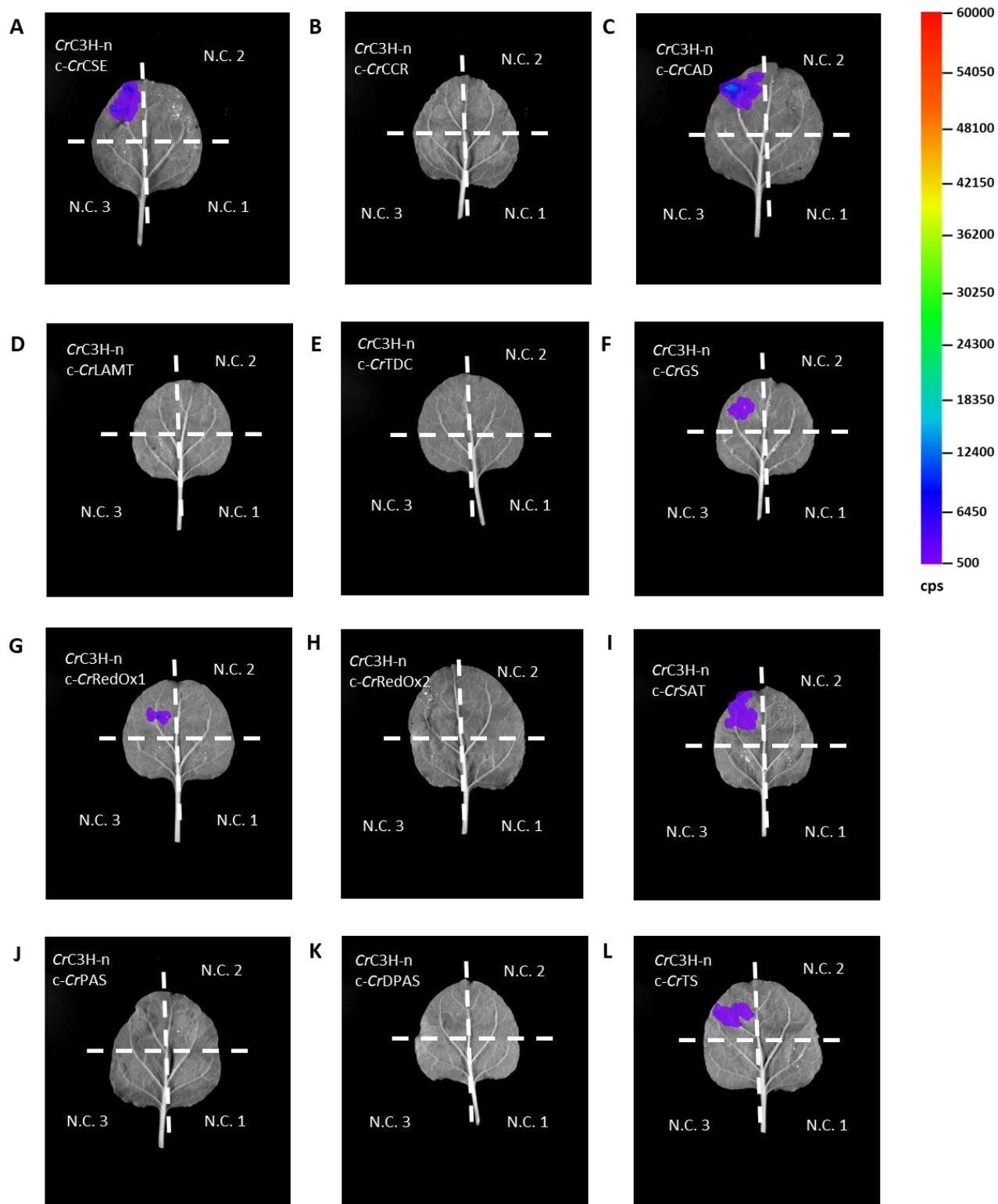


Figure 132. Representative images of pairwise interactions of *CrC4H* with *C. roseus* lignin and MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A.** C4H-CSE; **B.** C4H-CCR; **C.** C4H-CAD; **D.** C4H-LAMT; **E.** C4H-TDC; **F.** C4H-GS; **G.** C4H-RedOx1; **H.** C4H-RedOx2; **I.** C4H-SAT; **J.** C4H-PAS; **K.** C4H-DPAS; **L.** C4H-TS; **M.** C4H-CS; **N.** C4H-CorS; **O.** C4H-16OMT; **P.** C4H-T3R.



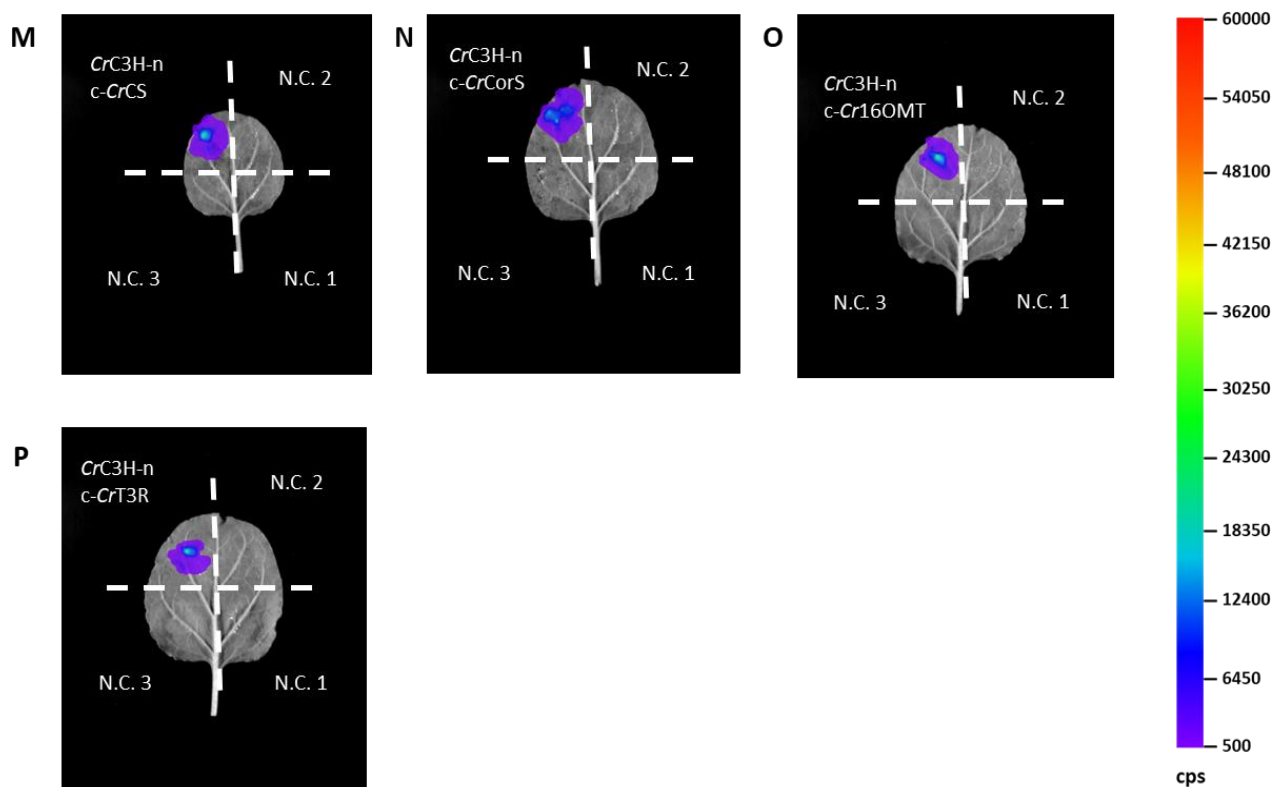
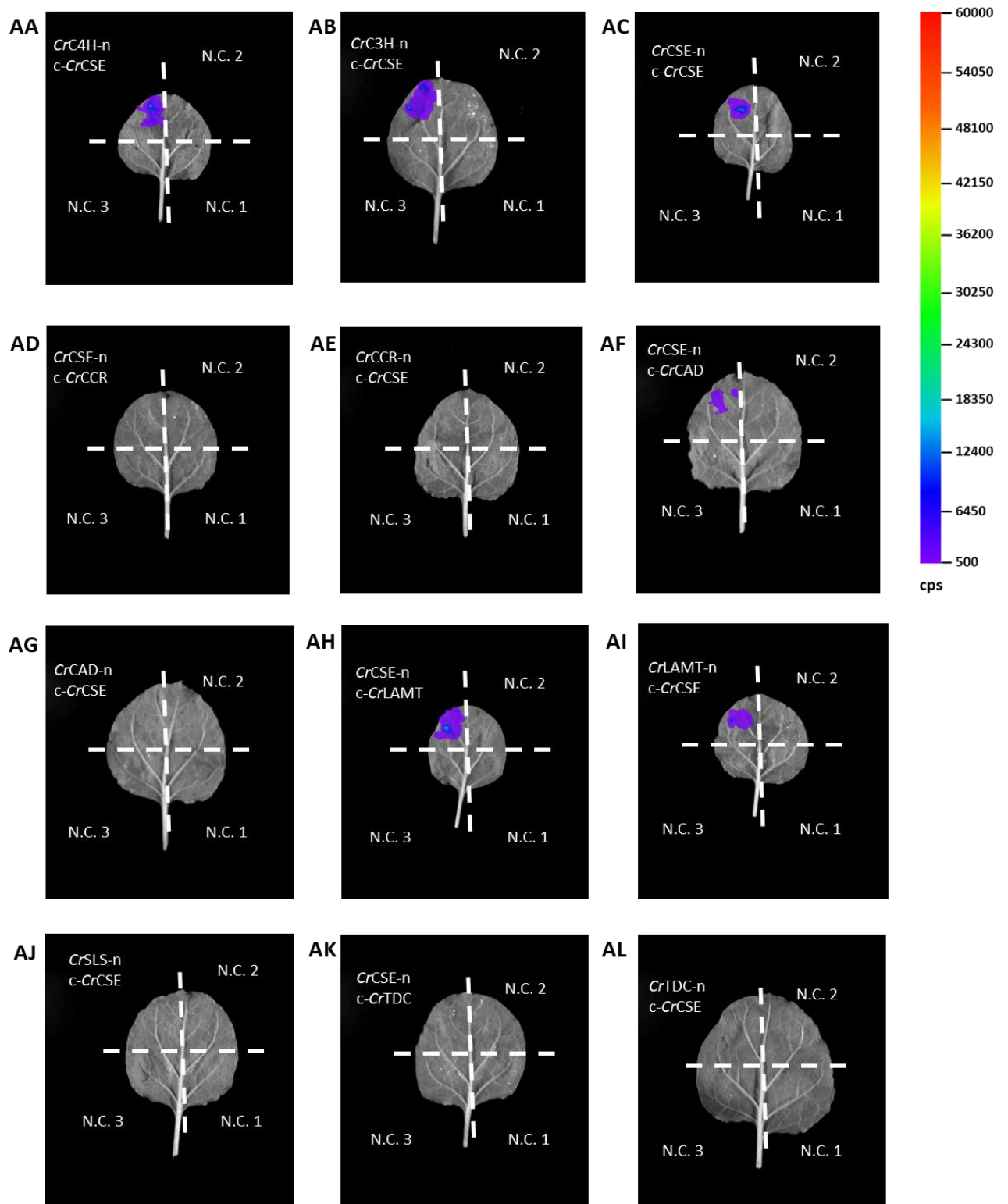
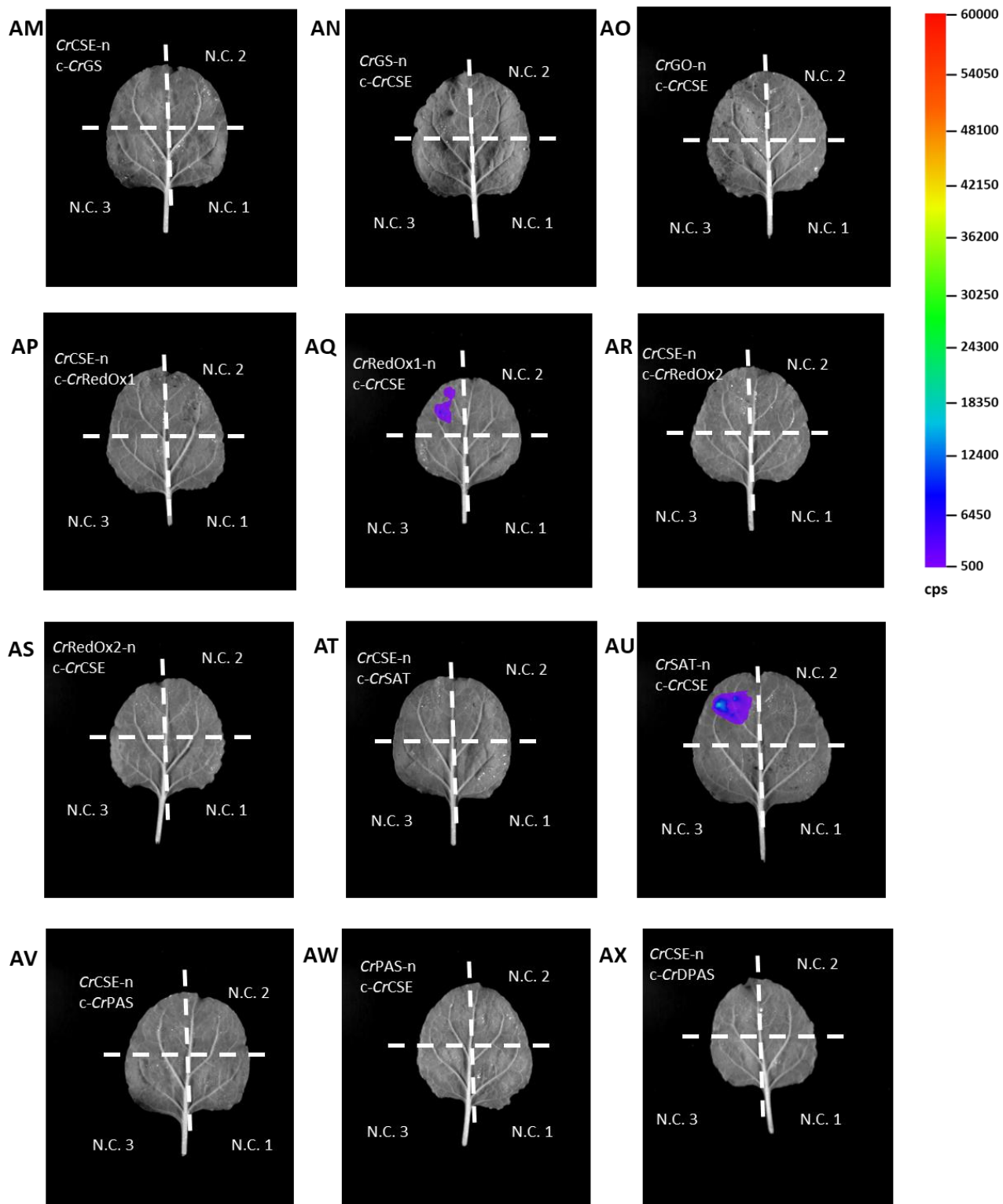
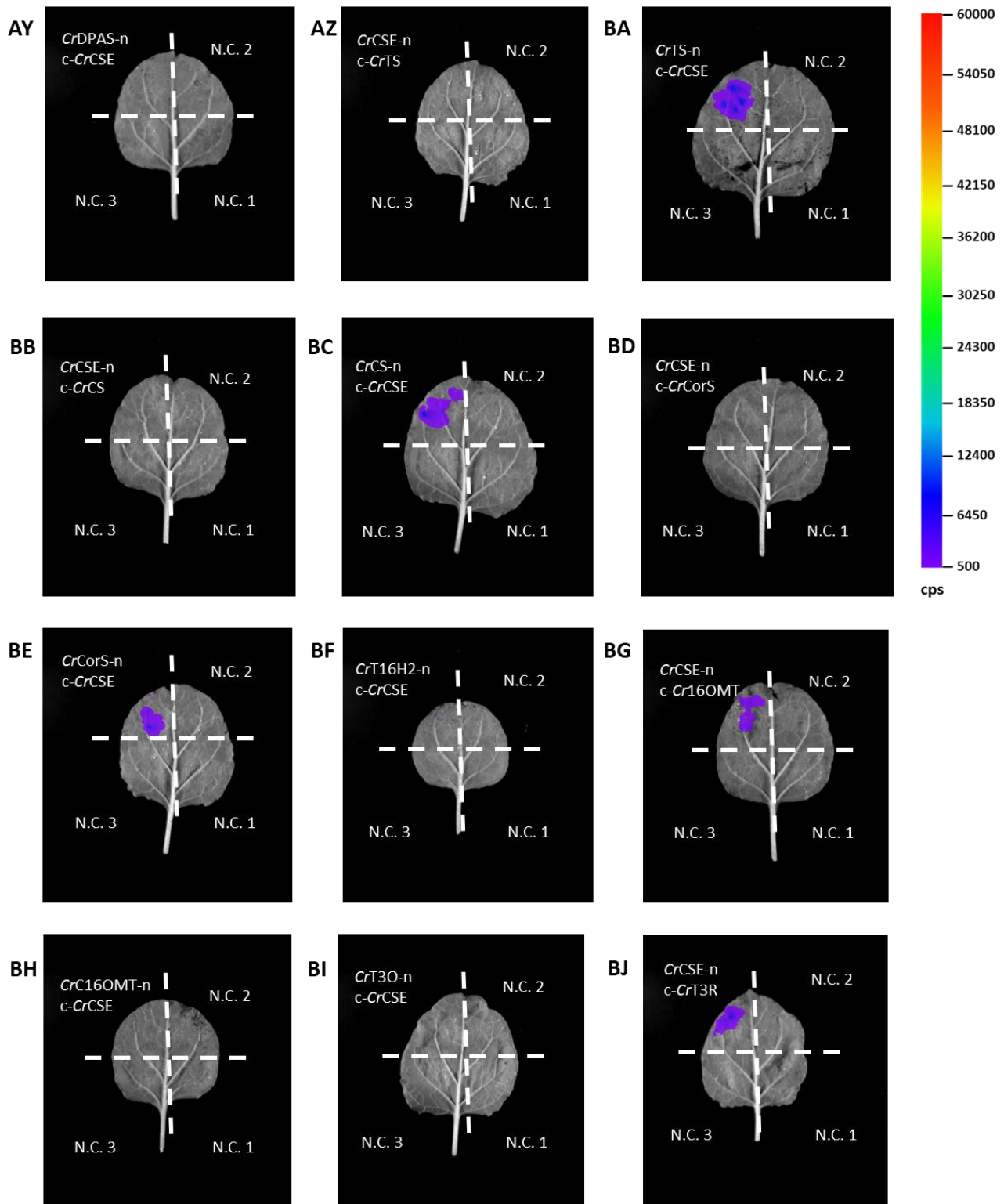


Figure 133. Representative images of pairwise interactions of CrC3H with *C. roseus* lignin and MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A.** C3H-CSE; **B.** C3H-CCR; **C.** C3H-CAD; **D.** C3H-LAMT; **E.** C3H-TDC; **F.** C3H-GS; **G.** C3H-RedOx1; **H.** C3H-RedOx2; **I.** C3H-SAT; **J.** C3H-PAS; **K.** C3H-DPAS; **L.** C3H-TS; **M.** C3H-CS; **N.** C3H-CorS; **O.** C3H-16OMT; **P.** C3H-T3R.







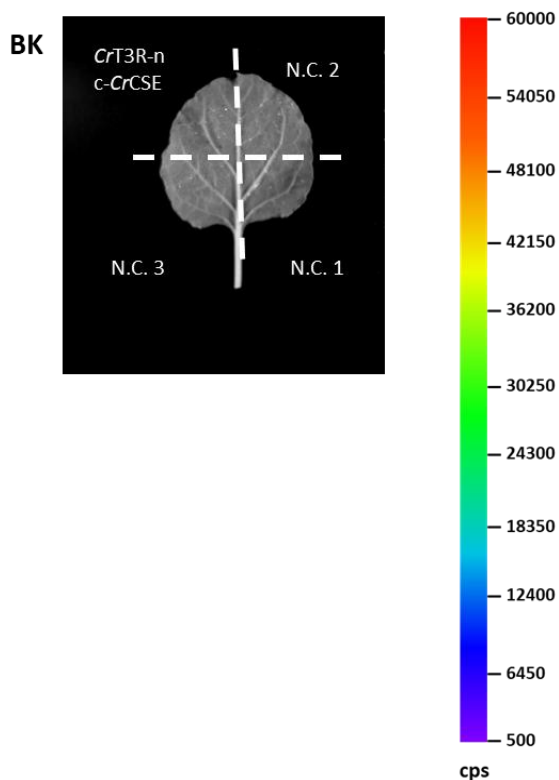
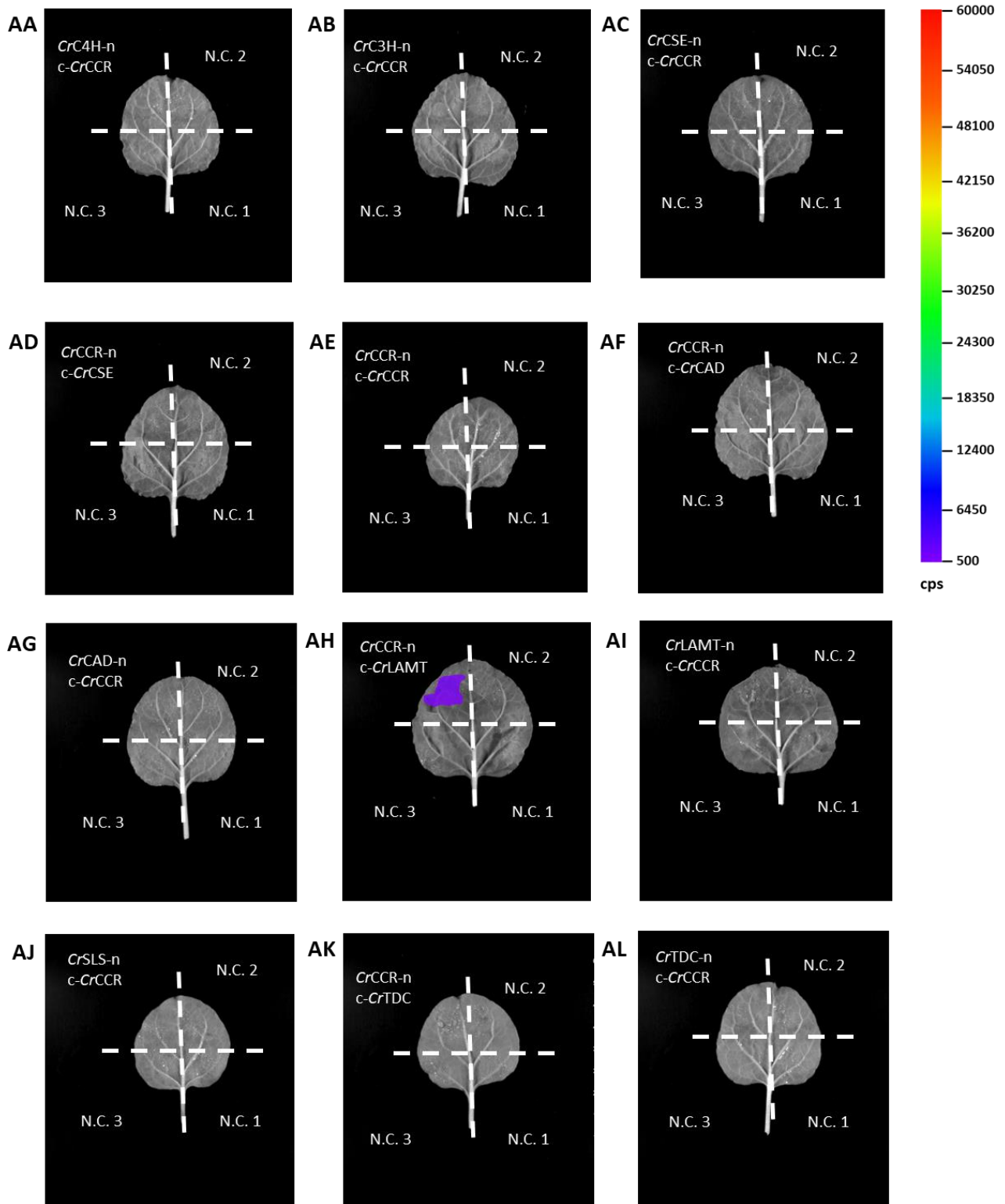
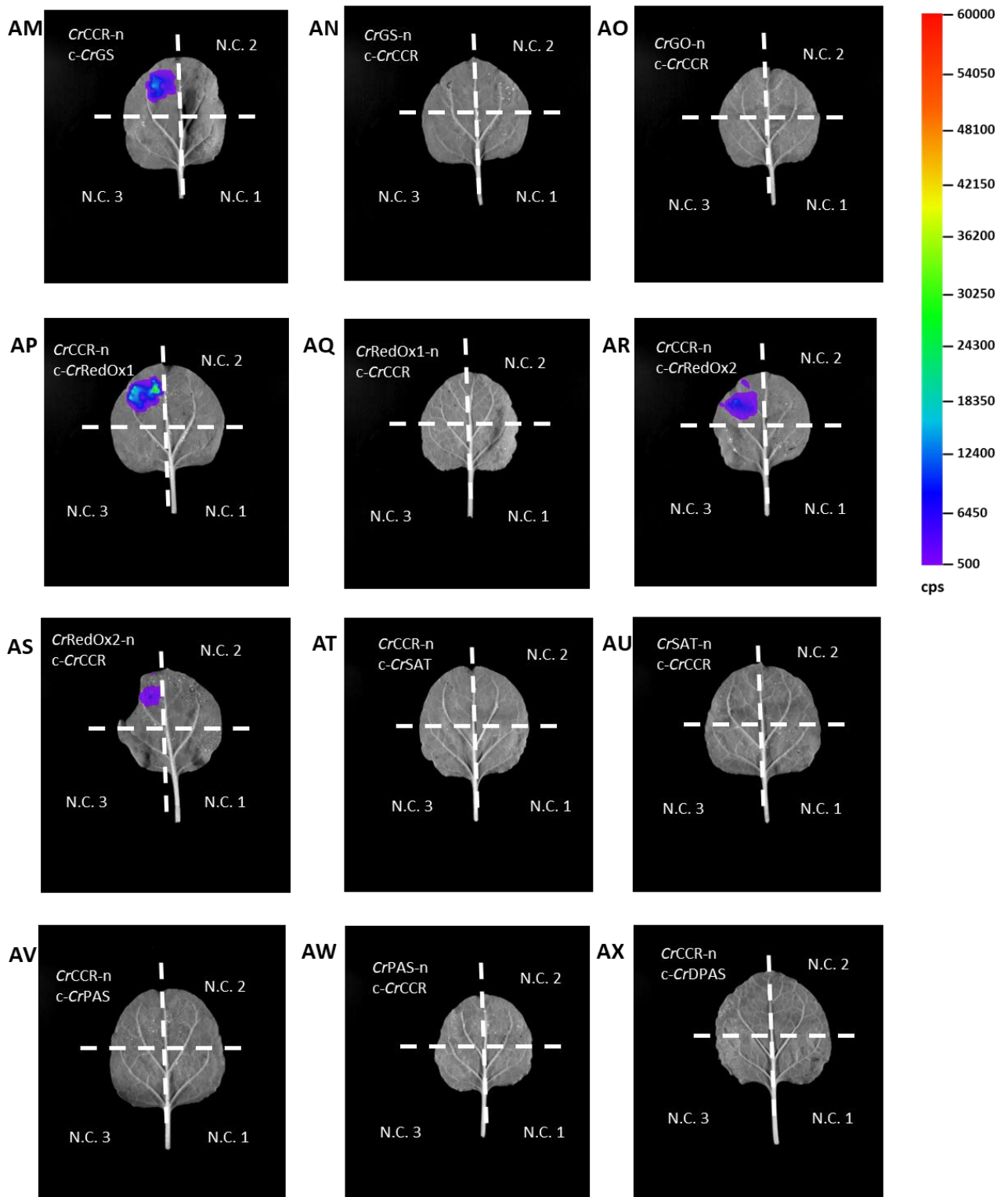
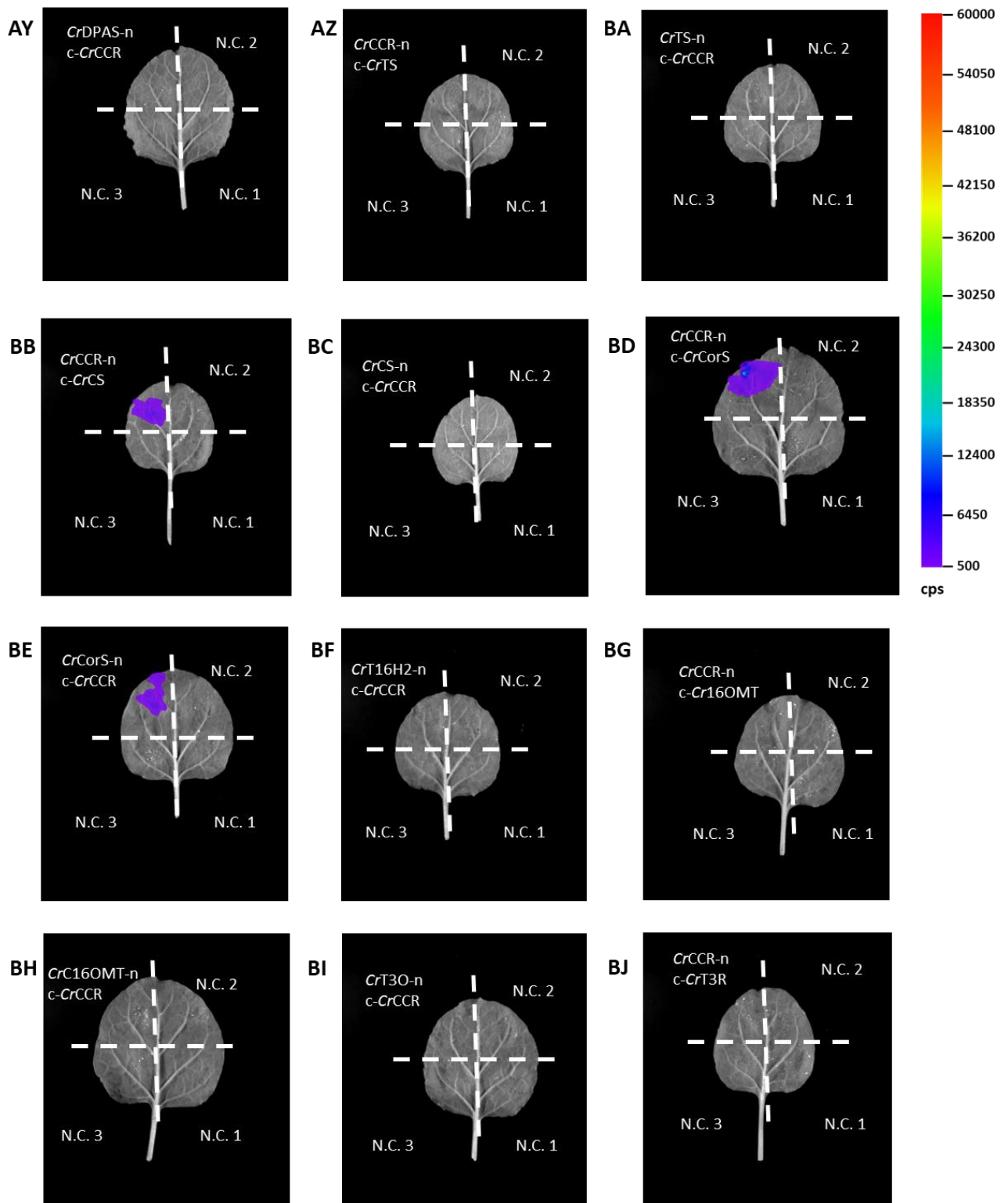


Figure 134. Representative images of pairwise interactions of *CrCSE* with *C. roseus* lignin and MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA.** C4H-CSE; **AB.** C3H-CSE; **AC.** CSE-CSE; **AD-E.** CSE-CCR; **AF-G.** CSE-CAD; **AH-I.** CSE-LAMT; **AJ.** SLS-CSE; **AK-L.** CSE-TDC; **AM-N.** CSE-GS; **AO.** GO-CSE; **AP-Q.** CSE-RedOx1; **AR-S.** CSE-RedOx2; **AT-U.** CSE-SAT; **AV-W.** CSE-PAS; **AX-Y.** CSE-DPAS; **AZ-BA.** CSE-TS; **BB-C.** CSE-CS; **BD-E.** CSE-CorS; **BF.** T16H2-CSE; **BG-H.** CSE-16OMT; **BI.** T3O-CSE; **BJ-K.** CSE-T3R.







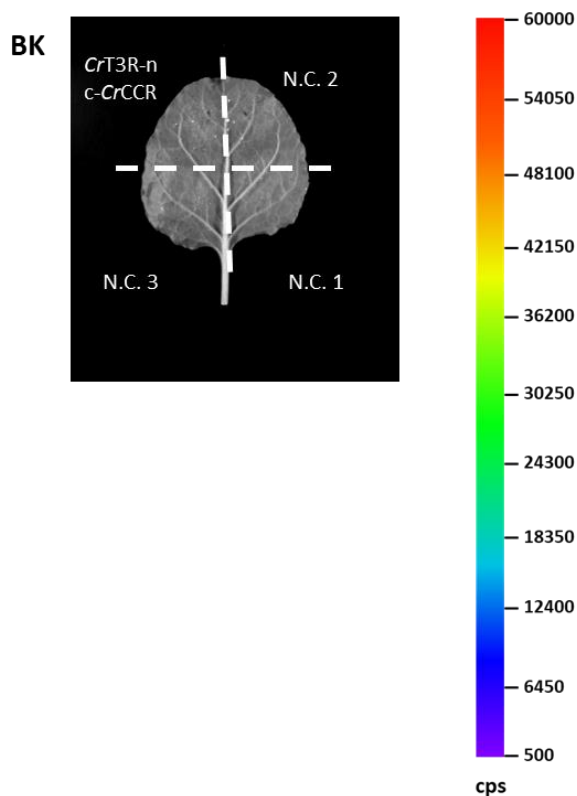
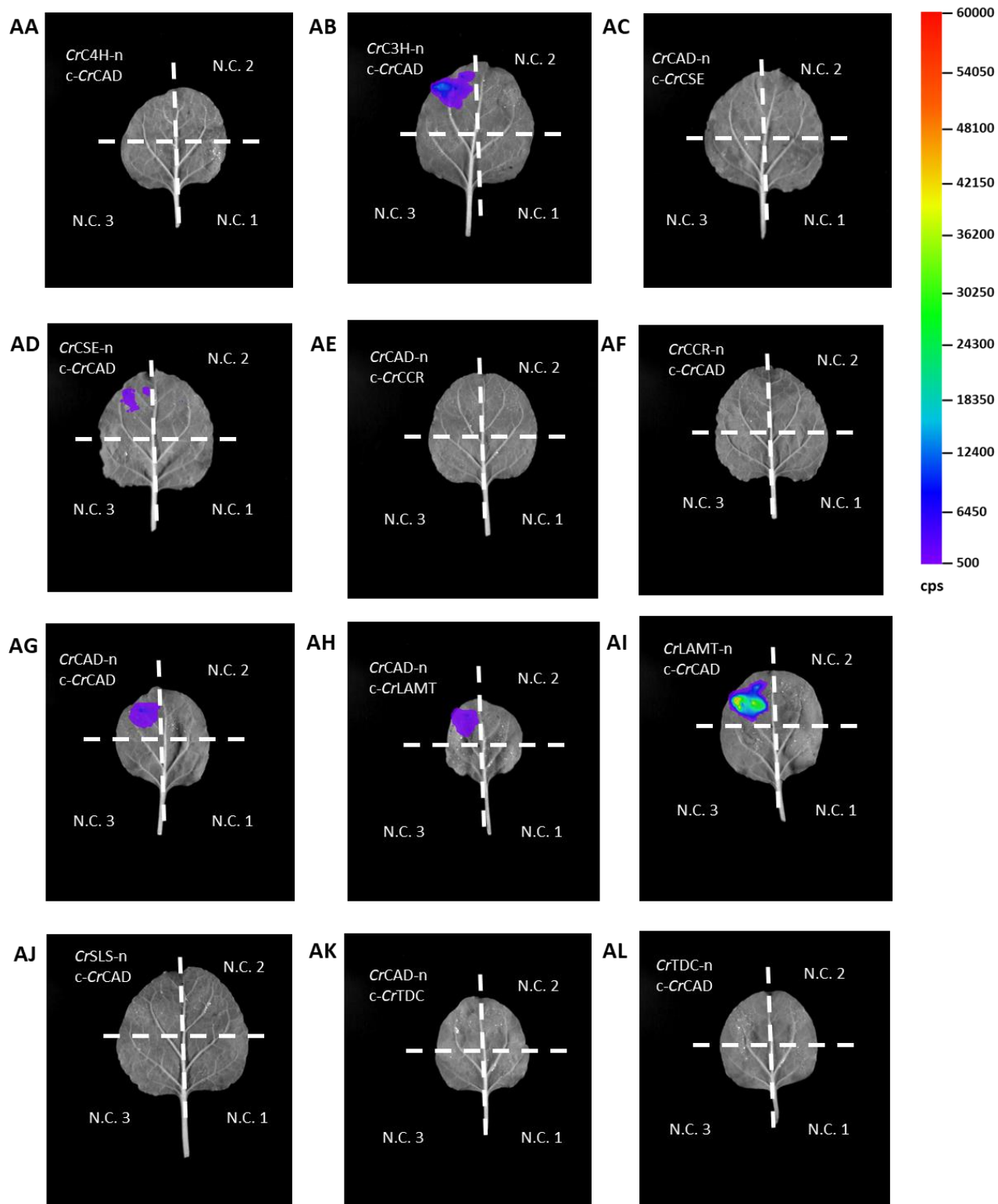
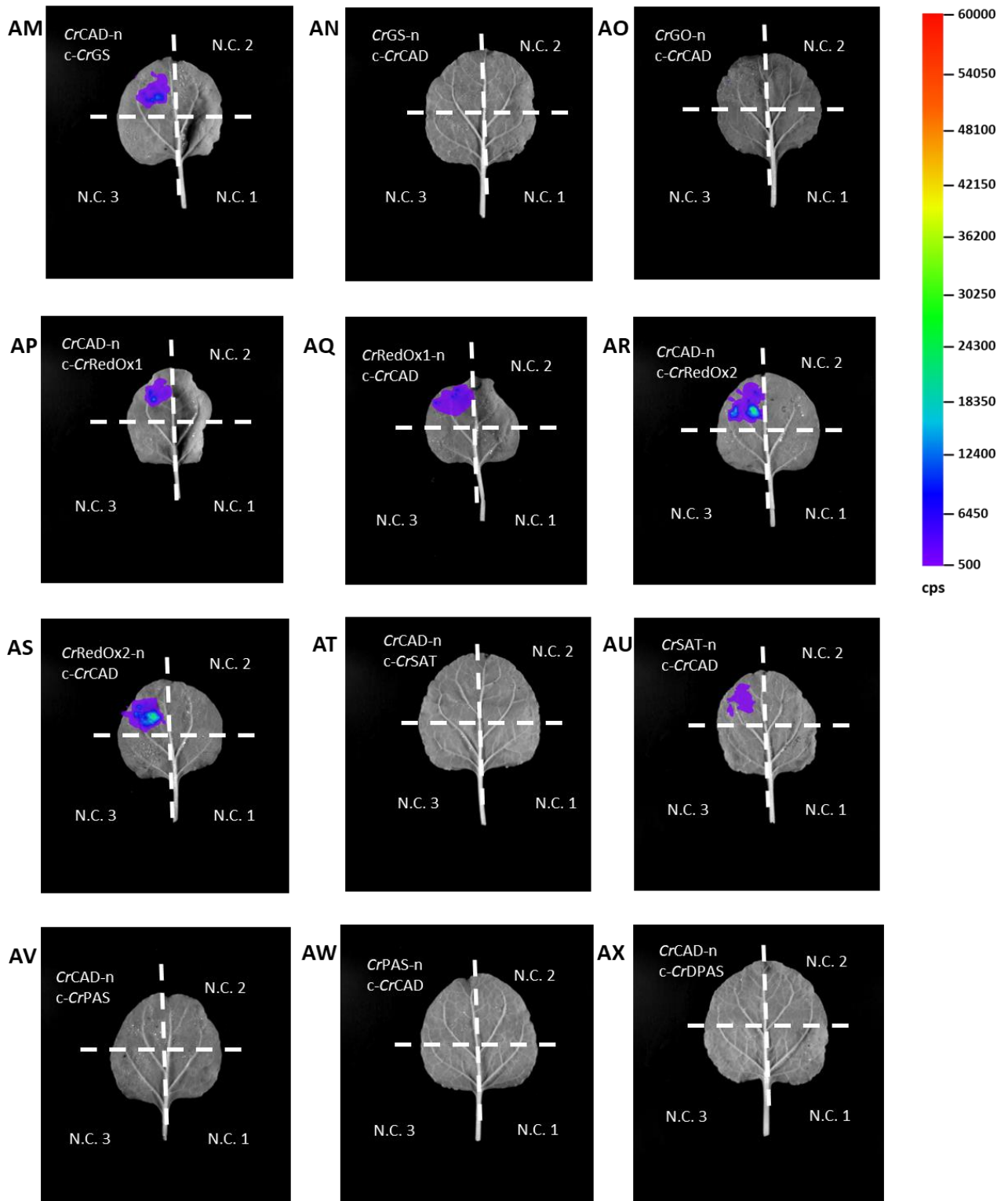
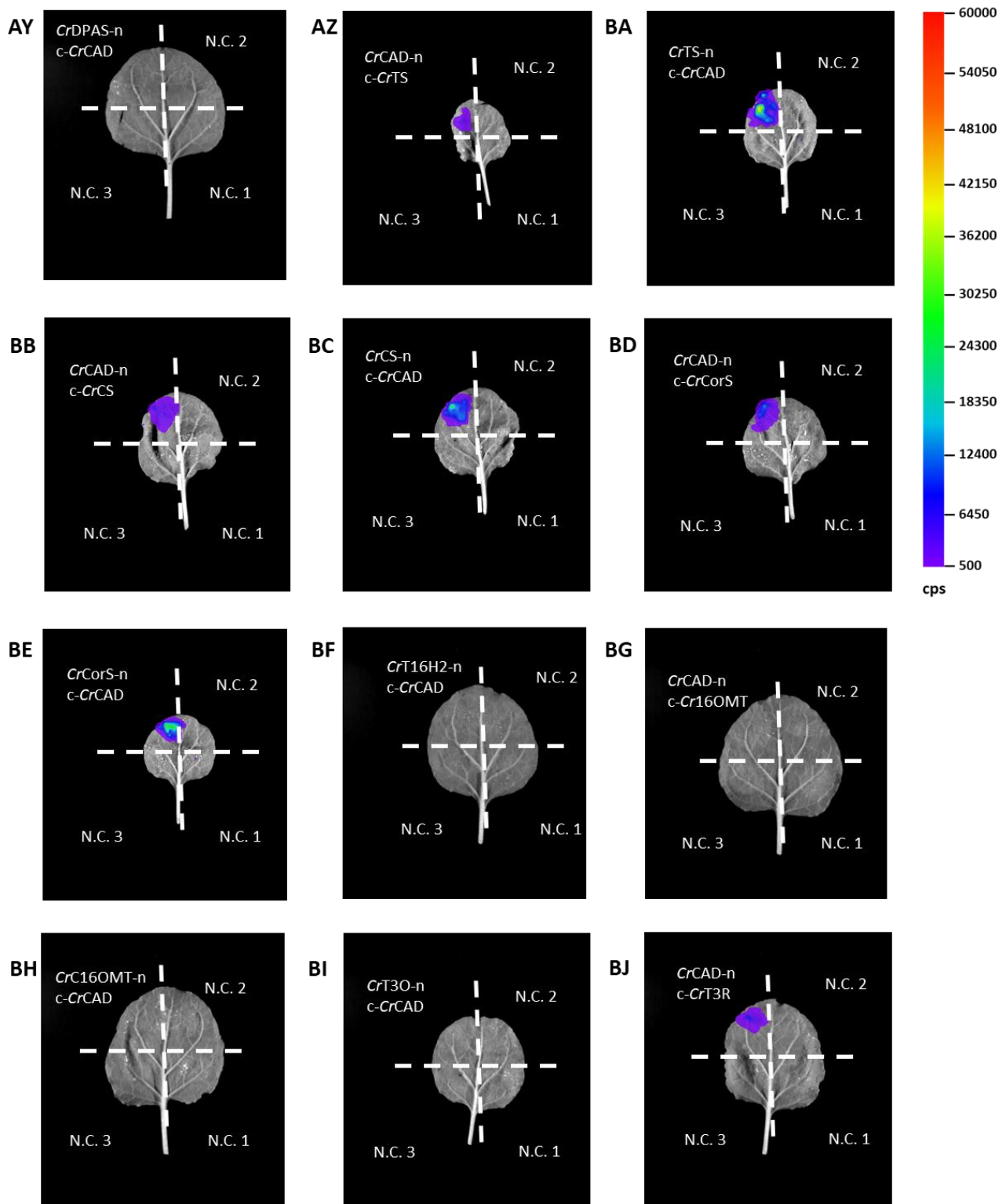


Figure 135. Representative images of pairwise interactions of *CrCCR* with *C. roseus* lignin and MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA.** C4H-CCR; **AB.** C3H-CCR; **AC-D.** CCR-CSE; **AE.** CCR-CCR; **AF-G.** CCR-CAD; **AH-I.** CCR-LAMT; **AJ.** SLS-CCR; **AK-L.** CCR-TDC; **AM-N.** CCR-GS; **AO.** GO-CCR; **AP-Q.** CCR-RedOx1; **AR-S.** CCR-RedOx2; **AT-U.** CCR-SAT; **AV-W.** CCR-PAS; **AX-Y.** CCR-DPAS; **AZ-BA.** CCR-TS; **BB-C.** CCR-CS; **BD-E.** CCR-CorS; **BF.** T16H2-CCR; **BG-H.** CCR-16OMT; **BI.** T3O-CCR; **BJ-K.** CCR-T3R.







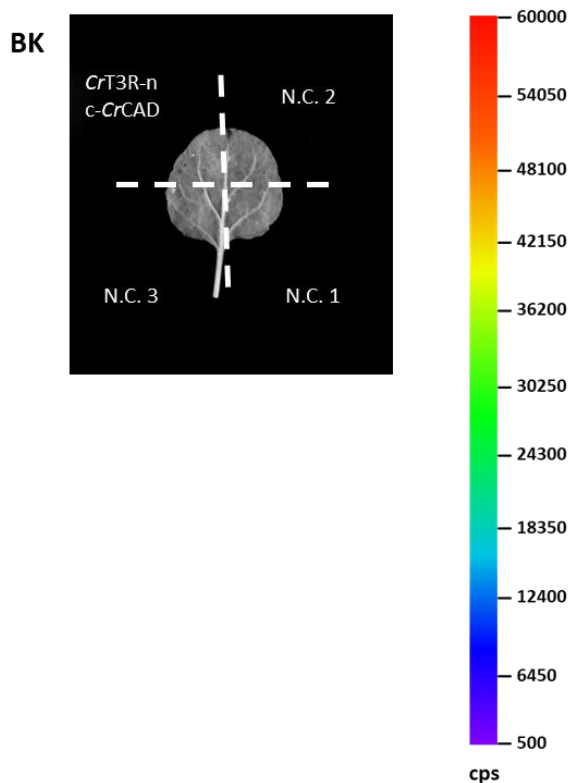
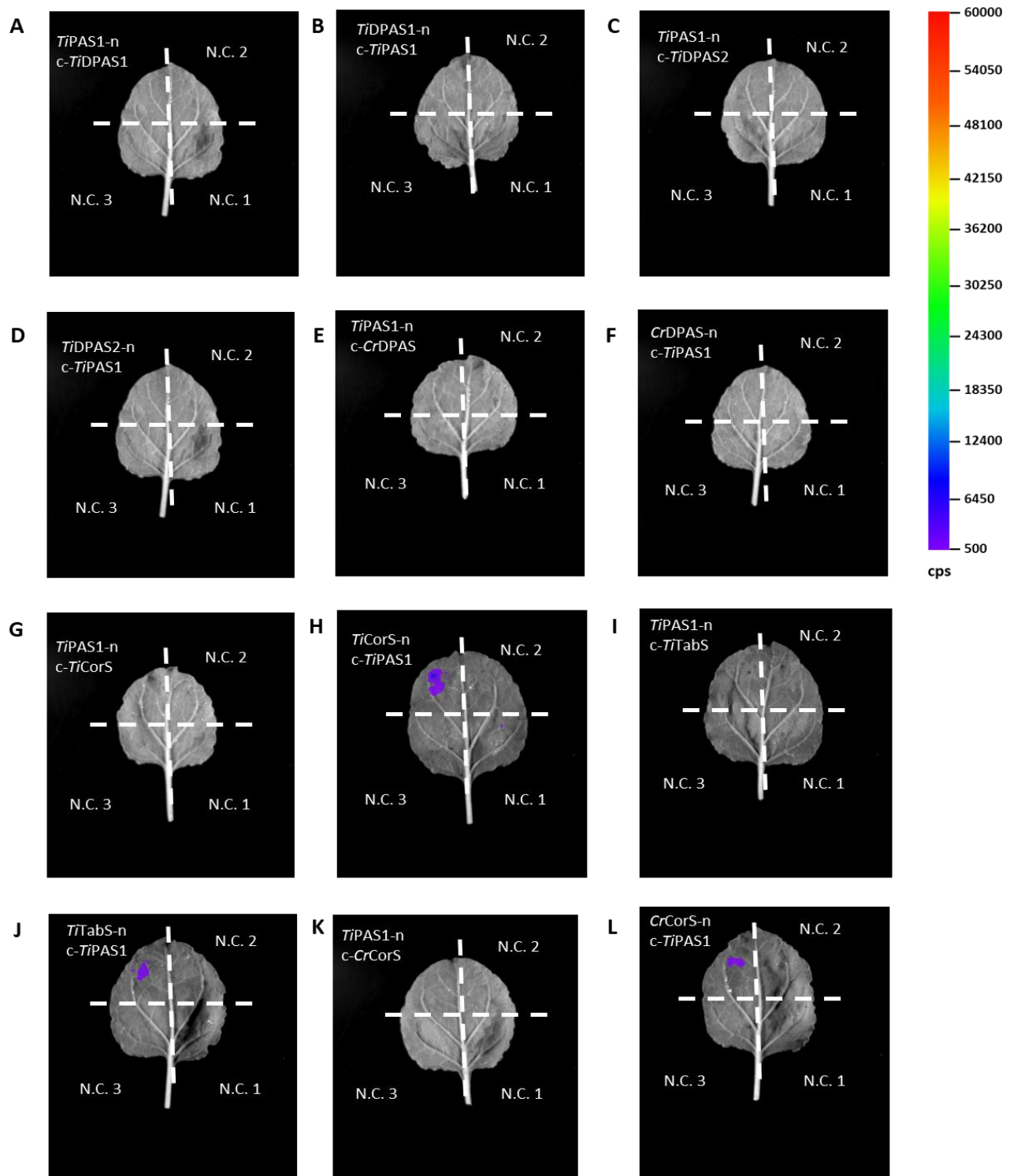


Figure 136. Representative images of pairwise interactions of CrCAD with *C. roseus* lignin and MIA biosynthetic enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **AA.** C4H-CAD; **AB.** C3H-CAD; **AC-D.** CAD-CSE; **AE-F.** CAD-CCR; **AG.** CAD-CAD; **AH-I.** CAD-LAMT; **AJ.** SLS-CAD; **AK-L.** CAD-TDC; **AM-N.** CAD-GS; **AO.** GO-CAD; **AP-Q.** CAD-RedOx1; **AR-S.** CAD-RedOx2; **AT-U.** CAD-SAT; **AV-W.** CAD-PAS; **AX-Y.** CAD-DPAS; **AZ-BA.** CAD-TS; **BB-C.** CAD-CS; **BD-E.** CAD-CorS; **BF.** T16H2-CAD; **BG-H.** CAD-16OMT; **BI.** T3O-CAD; **BJ-K.** CAD-T3R.

Appendix XVI. Split-Luciferase Assays of Ψ -Tabersonine Biosynthetic Enzymes

Representative split-luciferase result of testing protein-protein interactions of enzymes involved in Ψ -tabersonine biosynthesis.



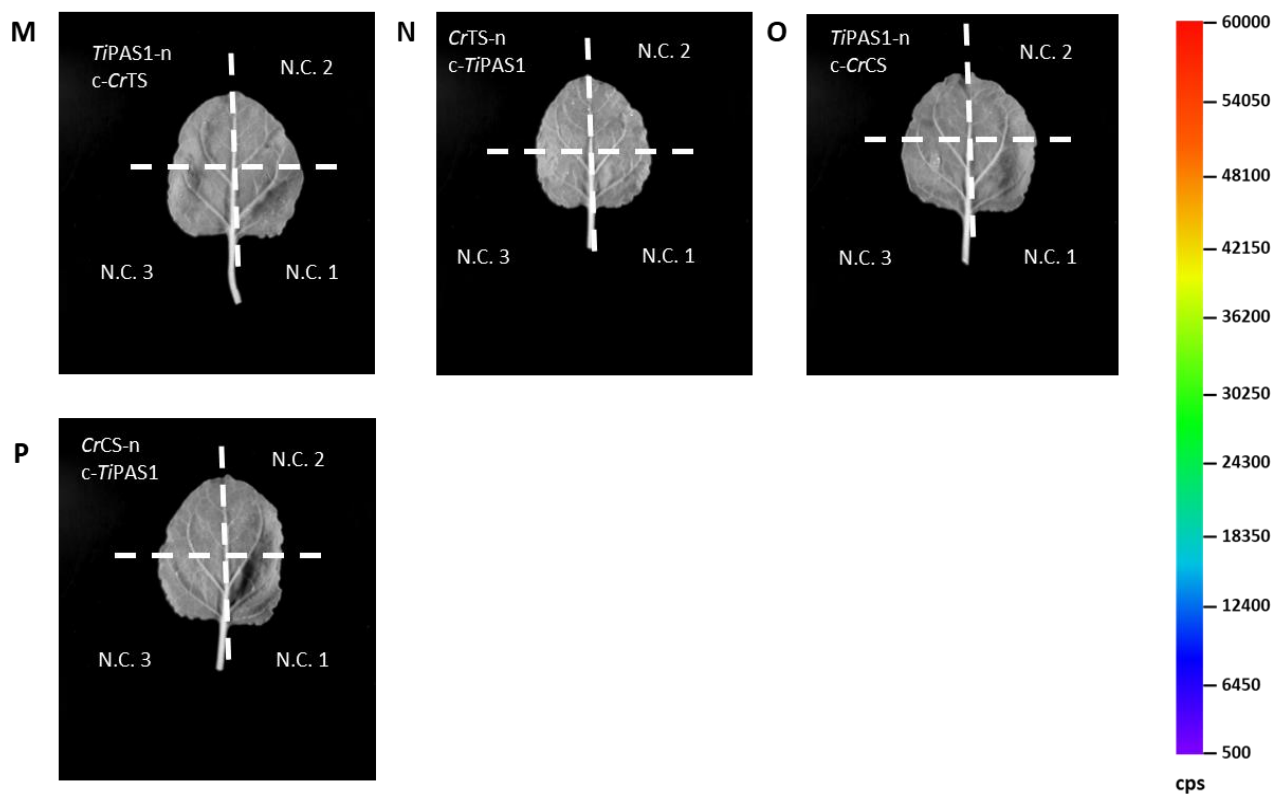
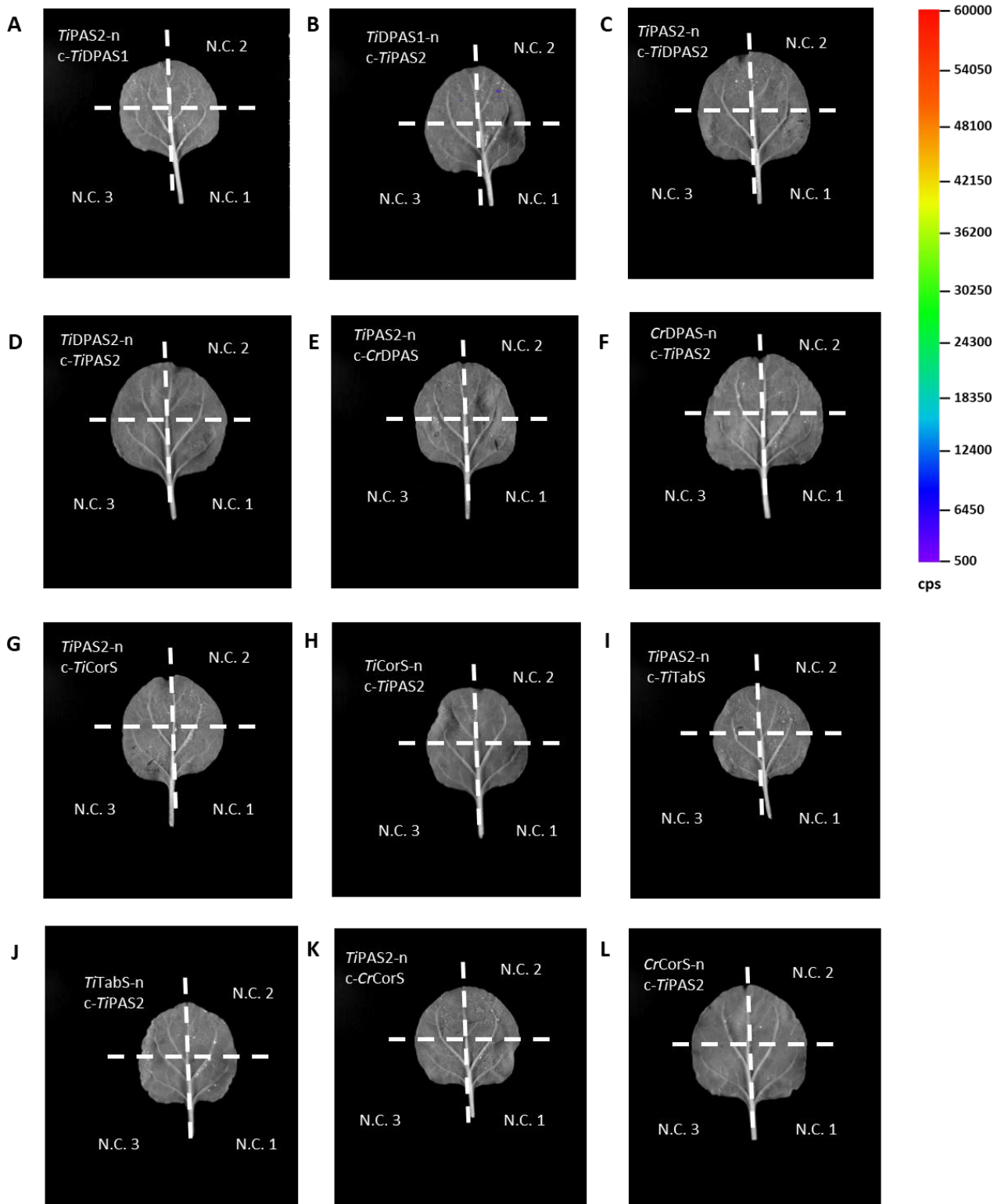


Figure 137. Representative images of pairwise interactions of *TiPAS1* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *TiPAS1-TiDPAS1*; **C-D.** *TiPAS1-TiDPAS2*; **E-F.** *TiPAS1-CrDPAS2*; **G-H.** *TiPAS1-TiCorS*; **I-J.** *TiPAS1-TiTabS*; **K-L.** *TiPAS1-CrCorS*; **M-N.** *TiPAS1-CrTS*; **O-P.** *TiPAS1-CrCS*.



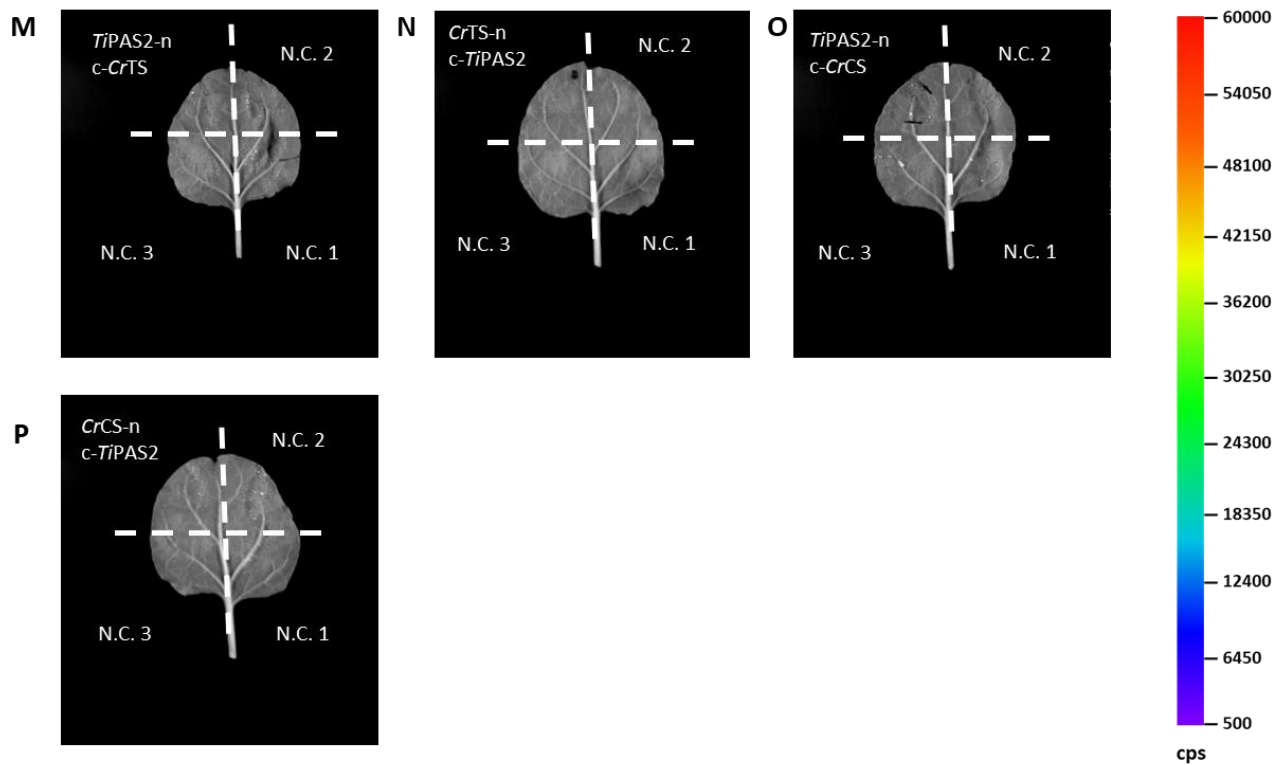
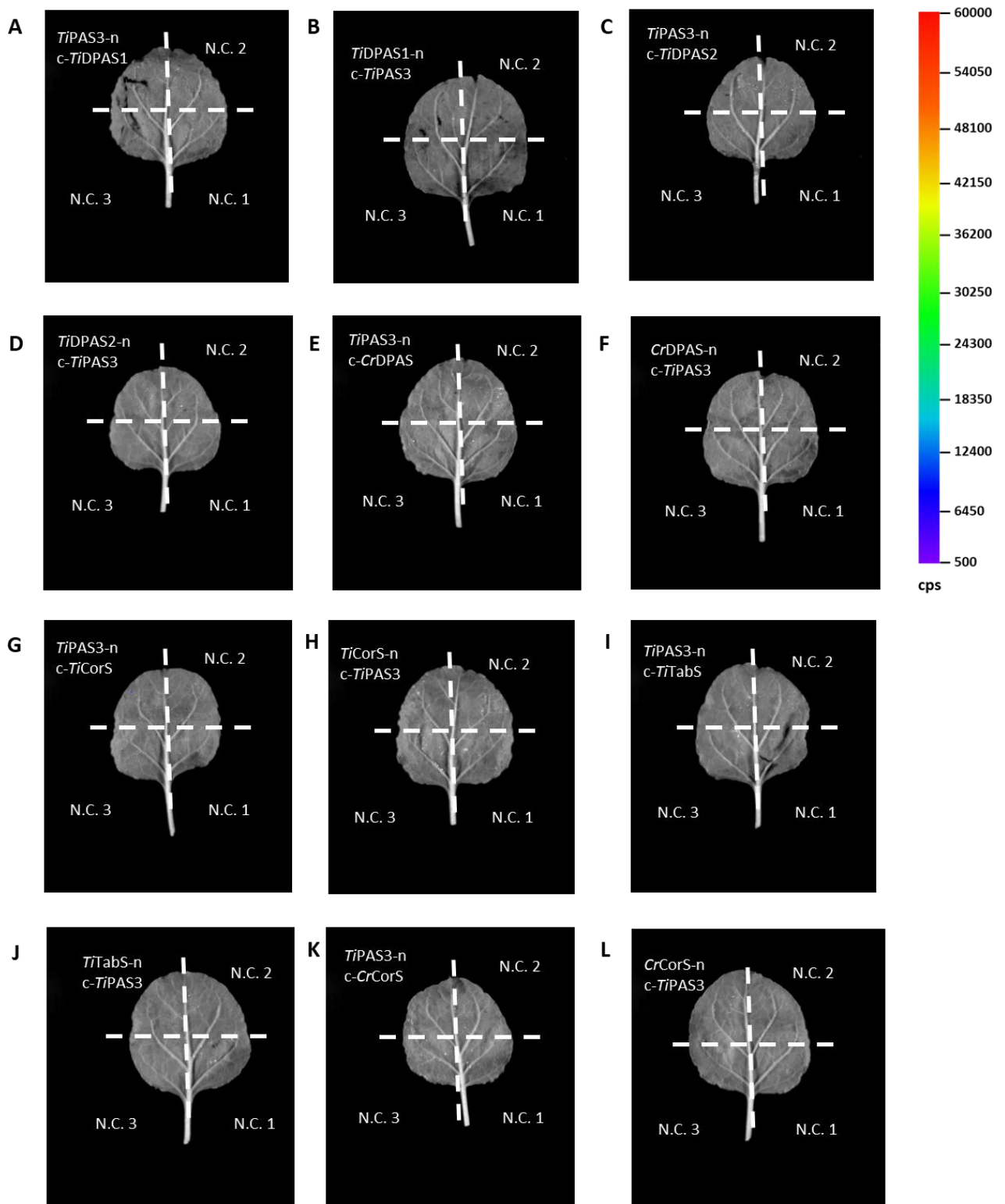


Figure 138. Representative images of pairwise interactions of *TiPAS2* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *TiPAS2-TiDPAS1*; **C-D.** *TiPAS2-TiDPAS2*; **E-F.** *TiPAS2-CrDPAS2*; **G-H.** *TiPAS2-TiCorS*; **I-J.** *TiPAS2-TiTabS*; **K-L.** *TiPAS2-CrCorS*; **M-N.** *TiPAS2-CrTS*; **O-P.** *TiPAS2-CrCS*.



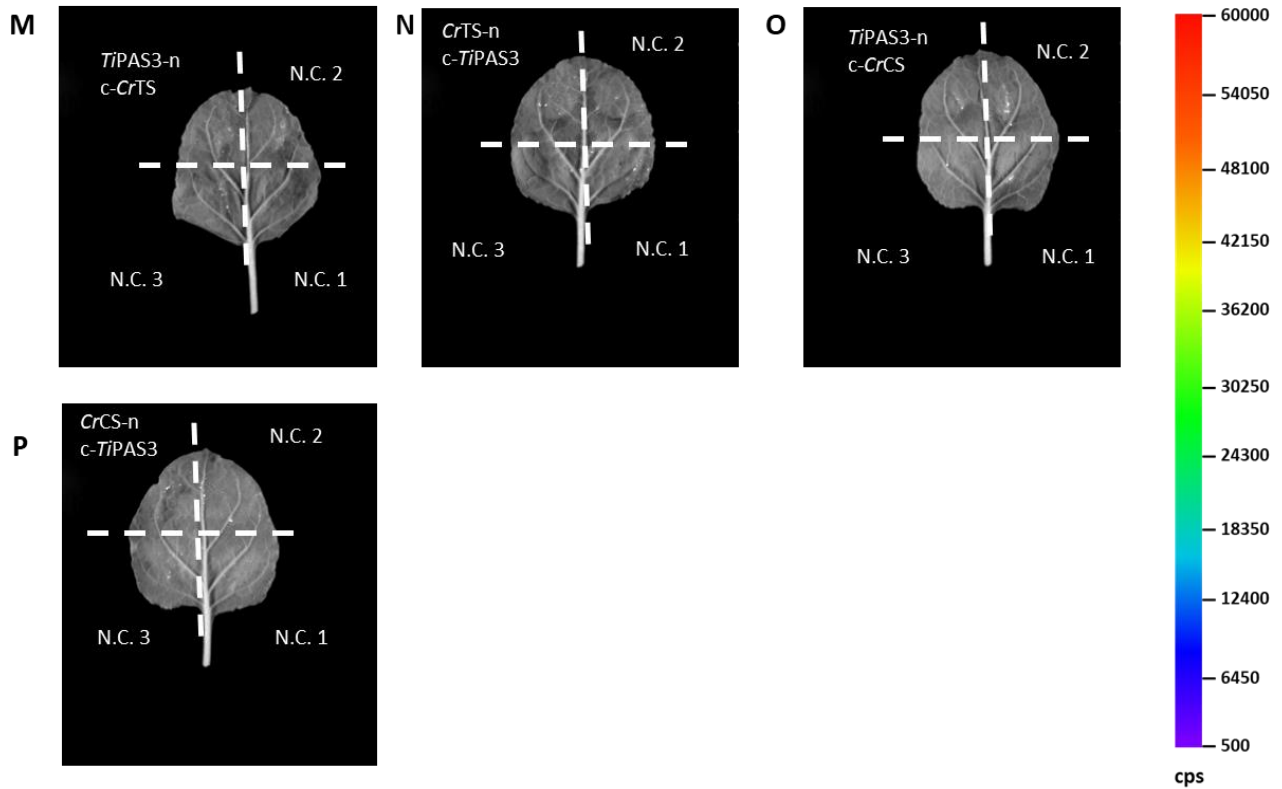


Figure 139. Representative images of pairwise interactions of *TiPAS3* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *TiPAS3-TiDPAS1*; **C-D.** *TiPAS3-TiDPAS2*; **E-F.** *TiPAS3-CrDPAS2*; **G-H.** *TiPAS3-TiCorS*; **I-J.** *TiPAS3-TiTabS*; **K-L.** *TiPAS3-CrCorS*; **M-N.** *TiPAS3-CrTS*; **O-P.** *TiPAS3-CrCS*.

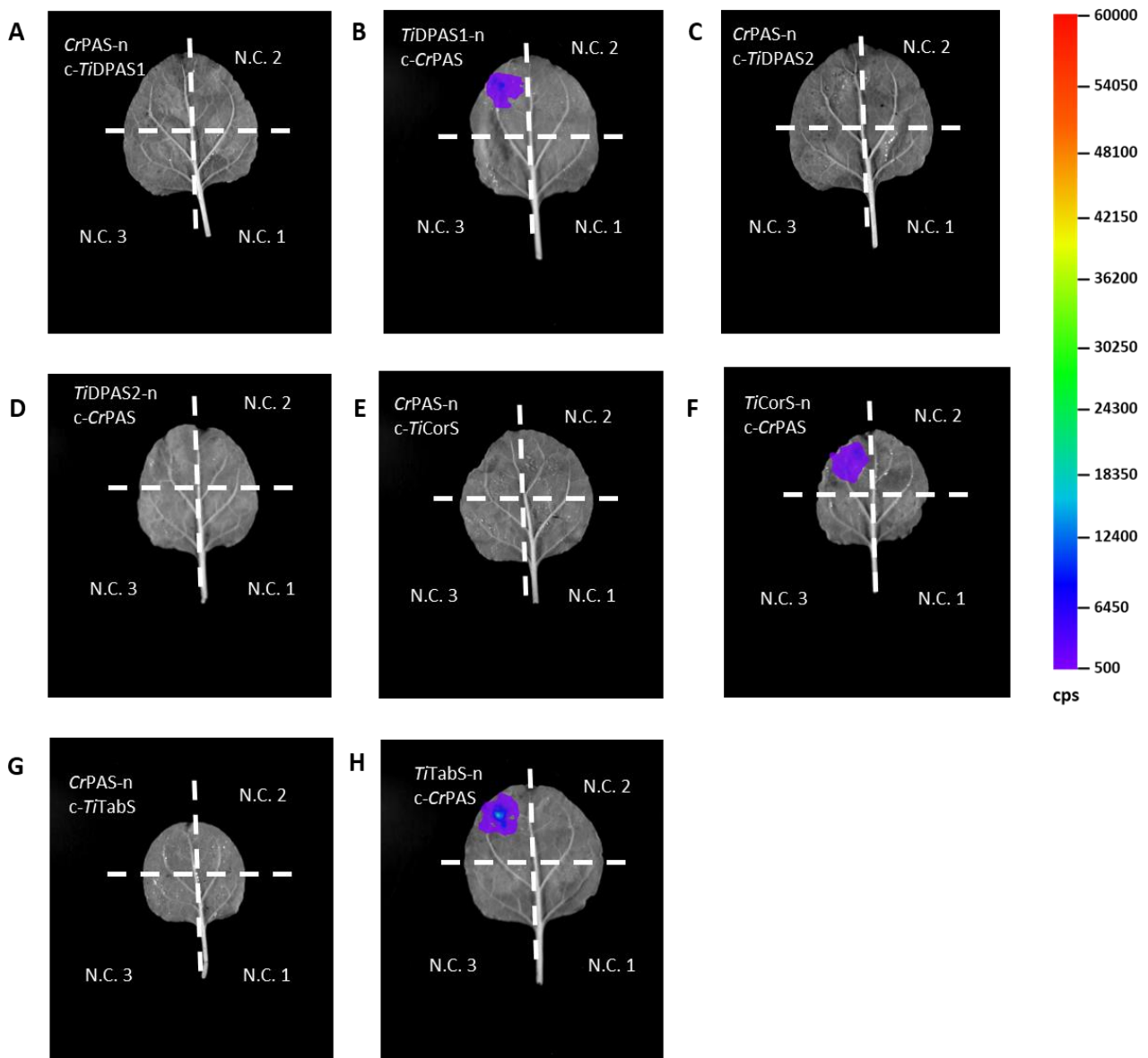


Figure 140. Representative images of pairwise interactions of *CrPAS* with *C. roseus* and *T. iboga* DPAS and cyclase enzymes tested by split-luciferase in *N. benthamiana*. –n represents constructs tagged with C-terminus nLuc luciferase fragment, c- represents constructs tagged with N-terminus cLuc luciferase fragment. N.C. 1 represents nLuc-tagged protein construct with empty cLuc fragment, N.C. 2 represents cLuc-tagged protein construct with empty nLuc fragment, N.C. 3 represents empty nLuc and empty cLuc fragment negative controls. Counts per second (cps) represented by false colour. **A-B.** *CrPAS-TiDPAS1*; **C-D.** *CrPAS-TiDPAS2*; **E-F.** *CrPAS-TiCorS*; **G-H.** *CrPAS-TiTabS*.