Norcoclaurine synthase:

The mechanism and biocatalytic potential of a Pictet-Spenglerase

Benjamin Robert Lichman

UCL

Thesis submitted to UCL for the degree of

Doctor of Philosophy

in

Biochemistry

September 2015

Declaration

The chemical syntheses of substrates and product standards were conducted by Dr Thomas Pesnot and Dr Eleanor Lamming. The cloning of Δ19*Tf*NCS and *Cj*NCS2 was carried out by Dr Thomas Pesnot and Dr Markus Gershater. The cloning of previously unreported transaminases was carried out by Dr Julio Rubéns Martínez-Torres and other members of the Ward group. The cloning of pQR1855 was carried out by Dr Altin Sula. The interpretation of X-ray electron density was conducted with the guidance of Dr Altin Sula and Prof Nicholas Keep. The crystallography project was carried out in co-operation with Dr Altin Sula and aided by Prof Nicholas Keep and Dr Ambrose Cole. All these contributions are indicated where appropriate.

A version of Chapter 4 has been published: Lichman, B. R., Gershater, M. C., Lamming, E. D., Pesnot, T., Sula, A., Keep, N. H., Hailes, H. C. & Ward, J. M. 'Dopamine-first' mechanism enables the rational engineering of the norcoclaurine synthase aldehyde activity profile. *FEBS J.* **282,** 1137–1151 (2015).

A version of Section 5.4 has been published: Lichman, B. R., Lamming, E. D., Pesnot, T., Smith, J. M., Hailes, H. C. & Ward, J. M. One-pot triangular chemoenzymatic cascades for the syntheses of chiral alkaloids from dopamine. *Green Chem.* **17,** 852–855 (2015).

I, Benjamin Robert Lichman confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

In plants, the enzyme norcoclaurine synthase (NCS) catalyses the formation of (*S*)-norcoclaurine *via* the Pictet–Spengler condensation of dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA). (*S*)-Norcoclaurine is the precursor to all benzylisoquinoline alkaloids (BIAs), a diverse group of over 2500 natural products.

The aim of this project was to elucidate the mechanism of NCS in order to enable the rational engineering of NCS activity. Variants of NCS were screened for activities with various substrates, forming novel tetrahydroisoquinolines (THIQs). NCS was combined with other enzymes in biocatalytic cascades to produce THIQs.

Initially, an N-terminally truncated NCS from *Thalictrum flavum* (Δ19*Tf*NCS) was expressed. Problems with the purification of Δ19*Tf*NCS led to the use of a different truncate, Δ29*Tf*NCS. This enzyme and variants were expressed and purified. The effect of mutations on the activity, kinetics and substrate specificity of Δ29*Tf*NCS led to the conclusion that NCS operates with a 'dopamine-first' mechanism. Computational analysis, including molecular dynamics and docking experiments, supported this conclusion. Furthermore, rational engineering of substrate specificity was demonstrated.

Next, the biocatalytic potential of NCS was investigated. Biotransformation conditions, such as enzyme or lysate loading, were optimised before demonstrative examples of milligram scale biotransformations were performed. Then, NCS and a transaminase were combined in a one-pot 'triangular' biocatalytic cascade to produce chiral BIAs. An additional chemical step led to the one-pot formation of chiral tetrahydroprotoberberines (berbines). The cascades were demonstrated on a milligram preparative scale.

Methods for screening NCS mutants were examined: various enzyme preparations, reaction conditions and reporter systems were tested and evaluated. Subsequently, NCS mutants were screened for activities with numerous amines and aldehydes. NCS activity was identified with some α-substituted aldehydes and ketones. Selected mutants demonstrated an increase in activity compared to wild-type for these unusual substrates. Notably, high conversions were revealed for cyclohexanone derivatives. A number of resulting cyclohexane-spiro-THIQs were characterised.

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my PhD supervisors Professors John Ward and Helen Hailes for their support, guidance and wisdom throughout the duration of this project. I was very lucky to be given the opportunity to take part in such a fascinating project, which has baffled and thrilled in equal measure. I was privileged to be supervised by two Professors with incredible knowledge in their respective fields.

I am very grateful to Dr Thomas Pesnot and Dr Markus Gershater for their excellent work on the project prior to my arrival, and for teaching and guiding me through the first year of the project. I am also thankful for the plentiful supply of aldehydes and amines, provided by Dr Eleanor Lamming, required to keep the enzyme sated.

It was just over two years ago that Dr Altin Sula and I began our collaboration. It has been a thoroughly enjoyable journey and I am completely indebted to Altin for his effort. I look forward to our enduring friendship and collaboration in the future. I am also grateful for the time, resources and support that Professor Nicholas Keep and Dr Ambrose Cole provided to me and Altin in the undertaking of this project.

In just four years I have seen the passing of generations in the Ward lab. Thank you all for the assistance, friendships and memories. Those at the beginning: Pedro, Vishal, John H, Alex and Markus. Some who have come and gone: Charlie, Laurent, John K and Lily. Many who remain: Aisha, Dave, Dragana, Fiona, Nadine, Maria, Mike, Phil and Yang. And, of course, Jack, who I think has just always been there. I am also grateful for the students and visiting PhDs I supervised, Charlotte and Josh, Vanessa and Jutta, who taught me it is possible to teach when you haven't yet learnt. I also extend my gratitude to members of the Hailes and Tabor groups, who always found me a fume hood in which I could break some glassware.

Thank you to Professor Paul Dalby for use of the Avacta Optim and CD machine. Thank you to Professor Francesco Gerevasio for sage advice on Molecular Dynamics. Thank you to Dr Rachel Morgan for teaching me how to purify a protein (the real problem was that I was purifying the wrong one). Thank you to Weiluo Lee for provocative discussions and processive polymerases.

I would like to give special thanks to Professor John Christodoulou for the unwavering encouragement, mentoring and career advice.

I am grateful to the Wellcome Trust, ISMB and Departments of Chemistry and Biochemical Engineering for facilitating this project. I am furthermore indebted to the international community of academics and doctoral students I have met at conferences who continue to inspire me.

I am extremely appreciative of the constant support from my family and friends.

Lastly, I have endless gratitude for Toto and her support, proofreading, understanding, food, company, proofreading and unwavering positivity—this thesis (and author) owes much to her inspiration.

Table of contents

List of figures

List of tables

List of abbreviations

1-MT	1-methyl-L-tryptophan
3,4-DHPAA	3,4-dihydrophenylacetealdehyde
3-HBA	3-hydroxybenzaldehyde
4-HBA	4-hydroxybenzaldehyde
4-HPA	4-hydroxyphenylacetone
4-HPAA	4-hydroxyphenylacetaldehyde
α-KG	α-ketoglutarate
ADH	Alcohol dehydrogenase
Am	*Argemone mexicana*
BA	Benzaldehyde
BBE	Berberine bridge enzyme
BIA	Benzylisoquinoline alkaloid
CD	Circular dichroism
CH	Cyclohexanone
CHCA	Cyclohexanecarboxaldehyde
CHMK	Cyclohexylmethylketone
Cj	*Coptis japonica*
COSY	Homonuclear correlation spectroscopy
CP	Cyclopropanone
CYP	Cytochrome P450
DCM	Dichloromethane
DMF	Dimethylformamide
DNA	Deoxyribonucleic acid
DMSO	Dimethylsulfoxide
E. coli	*Escherichia coli*
ee	Enantiomeric excess
FPS	Fungal Pictet-Spenglerase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Hexahistidine-tag
HPAA	4-hydroxyphenylacetaldehyde
HPLC	High-pressure liquid chromatrography
HRMS ES+	High resolution mass spectrometry, positive electrospray
HSQC	Heteronuclear single quantum coherence spectroscopy
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRED	Imine reductase
k_{cat}	Turnover number
K_m	Michaelis constant
MAO	Monoamine oxidase
MeCN	Acetonitrile
MD	Molecular dynamics
MS	Mass spectrometry
NCS	Norcoclaurine synthase
Ni-NTA	Nickel-nitrilotriacetic acid on agarose resin
NL	Norlaudanosoline
NMR	Nuclear magnetic resonance spectroscopy
NOE	Nuclear Overhauser effect

NOESY	Nuclear Overhauser effect spectroscopy
PA	Phenylacetone
PAA	Phenylacetaldehyde
Pb	*Papaver bracteatum*
PCR	Polymerase chain reaction
pK_a	Acid dissociation constant
PLP	Pyridoxal phosphate
PR	Pathogenesis-related
PS	Pictet-Spengler
Ps	*Papaver somniferum*
Rpm	Revolutions per minute
S. cerevisiae	*Saccharomyces cerevisiae*
STR	Strictosidine synthase
TAm	Transaminase/aminotransferase
TB	Terrific broth
Tf	*Thalictrum flavum*
THIQ	Tetrahydroisoquinoline
THPB	Tetrahydroprotoberberine/berbine
Tris	Tris(hydroxymethyl)aminomethane
WT	Wild-type

Standard chemical, nucleic acid and amino acid abbreviations are used.

1 Chapter 1: Introduction

1.1 Background

1.1.1 Secondary metabolism and evolution

Secondary, or specialised, metabolites are chemicals produced by organisms that are not required for their immediate survival. Sessile organisms produce large varieties and quantities of these chemicals in order to interact with their environment and increase organismal fitness.[1] In plants, these compounds have varied roles including pest and pathogen defense and UV protection.[2] These compounds have considerable practical applications ranging from medicine to fragrance, but the majority of secondary metabolites remain uncharacterised.

Recently, the notion of classical deterministic metabolic pathways in secondary metabolism has given way to an increasing appreciation of a metabolic networks, in which promiscuous enzymes catalyse transformations on multiple substrates in order to produce a cocktail of metabolites.[1] There may be an advantage for sessile organisms to producing mixtures of chemicals, especially in situations in which the environment is changing. The classical pathways may evolve when particular metabolites have a clear evolutionary advantage, and enzymes gain 'specialising' mutations, reducing promiscuity.

Enzyme promiscuity is considered crucial for the evolution of secondary metabolic pathways. Secondary metabolic pathways often have a 'signature' enzyme which has been recruited from primary metabolism *via* gene duplication.[2,3] Duplication allows the enzyme to drift and gain novel activity (neofunctionalisation) or may enable an existing promiscuous activity to increase (subfunctionalisation).[4] This signature enzyme defines the main structure of a particular secondary metabolite class. Subsequently, tailoring enzymes such as cytochrome P450s (CYPs), transferases and dehydrogenases are recruited *via* a similar duplication process; these enzymes form a metabolic network, capable of producing a variety of chemicals with potential fitness benefits. Individual metabolites providing distinctly beneficial effects will presumably lead to the reduction in the network-like character and an increase in pathway character as enzyme specificity increases. In a number of cases, these genes become clustered on the genome, potentially improving the ability of the organism to regulate the pathway.[2]

The chemical innovation possible with evolution is exemplified by plant secondary metabolism. Furthermore, secondary metabolism provides numerous examples of enzyme evolution and promiscuity. It is possible that the enzymes of secondary

metabolism have different properties to typical primary metabolism enzymes—they may have evolved to be more promiscuous.[3,5] Whilst the chemicals produced by secondary metabolism provide a rich resource for medicine and other industries, the enzymes found in secondary metabolism can teach us much about molecular evolution and chemical innovation.

1.1.2 Benzylisoquinoline alkaloids (BIAs)

The benzylisoquinoline alkaloids (BIAs) are a structurally diverse group of nitrogen containing secondary metabolites (alkaloids) found in plants.[6] These compounds, which all contain a benzylisoquinoline moiety, provide defence against pathogens and herbivores.[7] BIAs are mostly restricted to plants in the order Ranunculales (e.g. buttercup, poppy, barberry) and clade magnoliids (e.g. black pepper, magnolia, nutmeg), but their biosynthesis evolved monophyletically prior to the emergence of the eudicots (Fig. 1.1).[7]

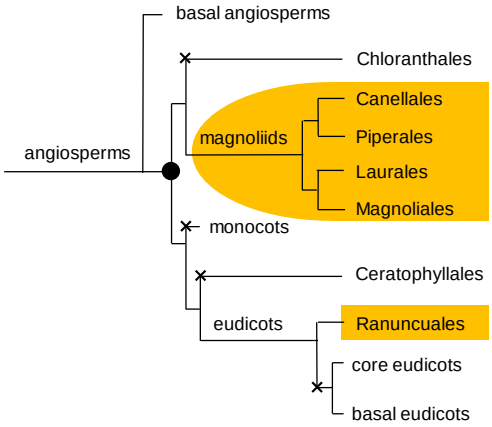


Fig. 1.1 The evolution of benzylisoquinoline alkaloid biosynthesis. The filled circle shows the hypothesised position of benzylisoquinoline alkaloid (BIA) evolution; the crosses show where BIA biosynthesis was lost. In orange are clades in which BIA biosynthesis is extant. All capitalised names are orders, all other clades are unranked. Basal eudicots are a polyphyletic group.

There are over 2500 BIAs,[8] including a number of molecules with significant and varied medicinal properties; examples of these bioactive BIAs include: morphine (analgesic),[9] noscapine (antitussive and anticancer)[10] and berberine (anti-microbial)[11] The BIAs can be separated into over 12 structural subgroups which are defined by C-C and C-O bonds added to the basic BIA structure at metabolic branch points.[6] One of these subgroups is the morphinan alkaloids, which includes morphine and thebaine, and is defined by a cytochrome P450 (CYP) catalysed C-C bond formation.[12] Other subgroups include the aporphine alkaloids, protoberberine alkaloids and the Bis-BIAs (Fig. 1.2). Different BIA producing species and strains produce these subgroups in varying proportions and purities.[13,14]

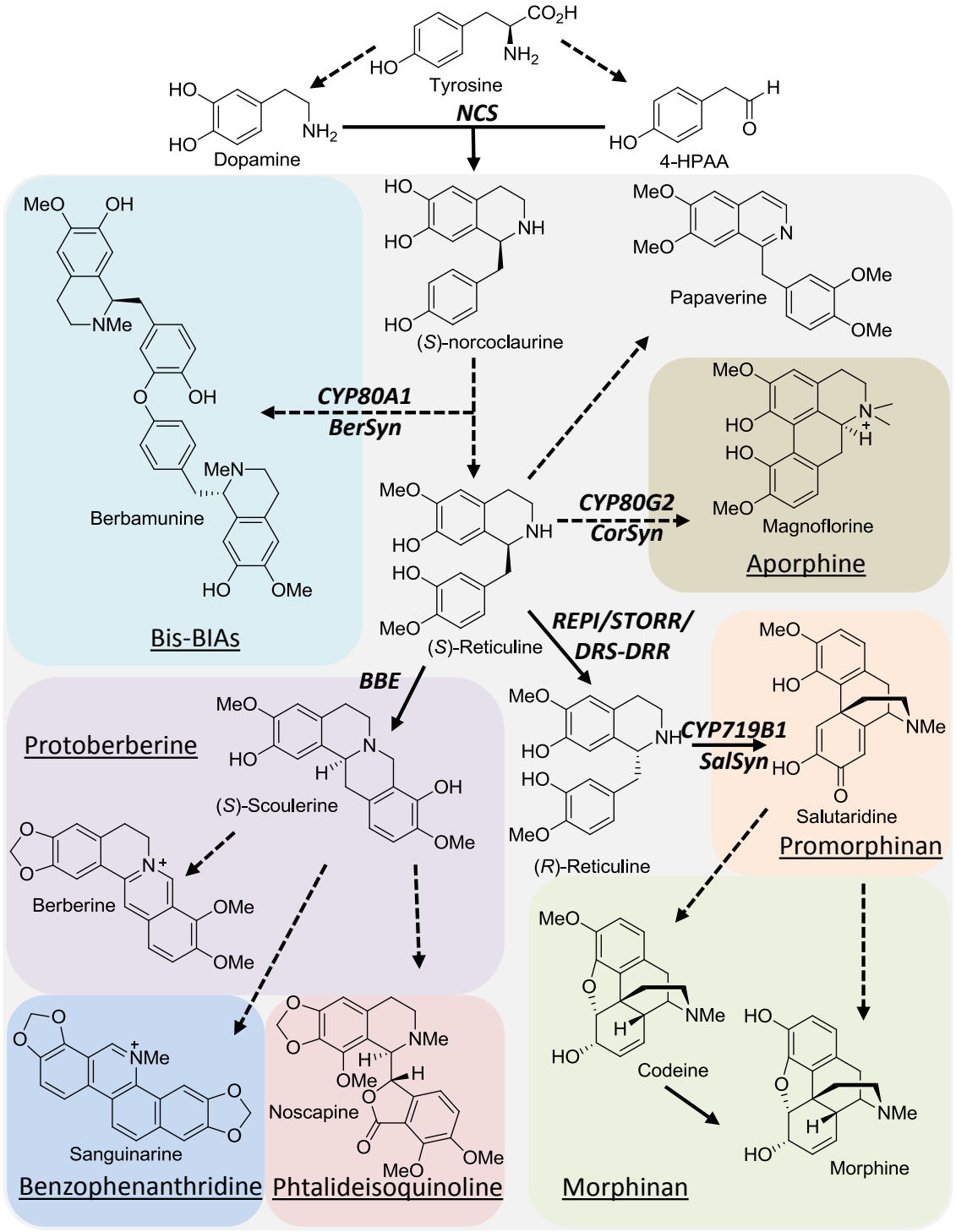


Fig. 1.2 Benzylisoquinoline alkaloid biosynthesis. The major BIA branch points norcoclaurine, (*S*)-reticuline, (*R*)-reticuline and (*S*)-scoulerine are depicted. Solid arrows represents single enzyme catalysed steps, dotted arrows represent multiple steps. Key enzymes are represented in italics: NCS (norcoclaurine synthase[15]), BerSyn (berbamunine synthase[16]), CorSyn (Corytuberine synthase[17]), SalSyn (Salutaridine synthase[12]), BBE (berberine bridge enzyme[18]) and REPI (reticuline epimerase[19], also named STORR (*S*-to-*R*-reticuline[20]) and DRS-DRR (1,2-dihydroreticuline synthase/reductase[21])). Coloured backgrounds and underlined titles represent BIA subgroups. In grey is the general BIA family. BIA subgroups not mentioned: protopine, rhoeadine, isopavine, pavine and cularine. The exact branch point(s) for berbamunine and other *bis*-BIAs is not known.

1.1.3 The Pictet-Spengler reaction

All plant BIAs have a common metabolic origin: (*S*)-norcoclaurine. This compound is the precursor to all BIAs; its formation is the first committed step into the BIA pathway. (*S*)-Norcoclaurine is formed *via* a Pictet-Spengler (PS) reaction between the tyrosine derivatives dopamine and 4-hydroxyphenylacetaldehyde (4-HPAA). The tetrahydroisoquinoline (THIQ) PS reaction was first observed over a century ago and was predicted to be involved in the biosynthesis of BIA natural products.[22,23] The PS reaction can be seen as a combination of a Mannich reaction and a Friedel-Crafts electrophilic aromatic substitution. In an analogous fashion to the Mannich reaction, an amine and an aldehyde (or ketone) combine to form an electrophilic iminium cation (Schiff base), which is attacked by a nucleophile. As in the Friedel-Crafts reaction, the nucleophile is an aromatic group. The iminium electrophile undergoes an intramolecular attack from the electron rich aromatic ring system; final rearomatisation yields the product (Fig. 1.3).[24] The PS reaction results in the formation of a C-N bond, a C-C bond and a chiral centre in a single reaction—this increase in molecular complexity in one synthetic transformation makes the reaction very useful for constructing complex molecules. This utility is shown by the fact that nature appears to have independently evolved use of the PS reaction for the formation of secondary metabolites in numerous cases (see section 1.7).

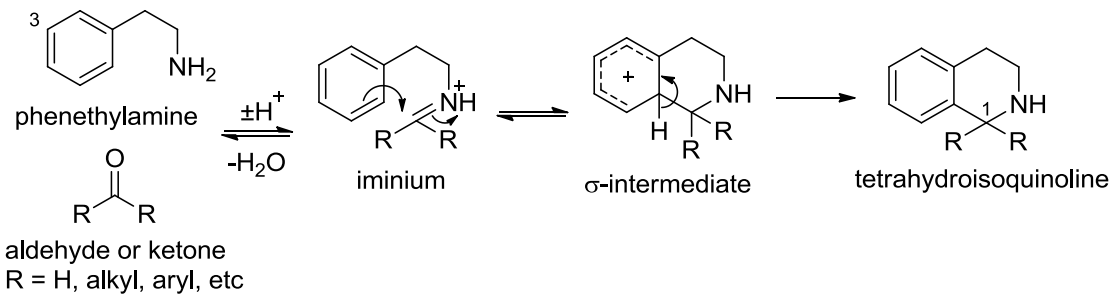


Fig. 1.3 Tetrahydroisoquinoline Pictet-Spengler reaction. Electron donating substituents on the phenethylamine are possible, including groups to stabilise positive charge on the aromatic ring (e.g. the 3-OH of the phenethylamine). If an aldehyde (or asymmetric ketone) is used then the THIQ is chiral at position 1. Ketones are possible reactants but are typically less reactive than aldehydes. Mechanistically, a spirocyclic intermediate is also possible, but experimentally has been shown to not be involved in most cases.

1.2 The benzylisoquinoline alkaloid Pictet-Spenglerase

1.2.1 Norlaudanosoline synthase

It was originally postulated that (*S*)-norlaudanosoline, not (*S*)-norcoclaurine, was the product of the initial BIA PS reaction.[25,26] Feeding studies of radioactive norlaudanosoline provided evidence for its role.[27] Also, racemic norlaudanosoline formation was observed in the spontaneous PS condensation between dopamine and

its corresponding aldehyde (3,4-dihydroxyphenylacetaldehyde, 3,4-DHPAA, DOPAL).[28,29]

The BIA PS enzyme was first isolated from *Eschscholzia tenuifolia* in the early 1980s, and was originally annotated as a norlaudanosoline synthase.[25,30] The enzyme, both in crude lysate and in its purified form, was shown to catalyse the condensation of dopamine with either 4-HPAA, forming norcoclaurine, or 3,4-DHPAA, forming norlaudanosoline. These reactions were catalysed with similar efficiency (Fig. 1.4).[25,30] The enzyme did not accept 3,4-dihydroxyphenylpyruvate, a postulated substrate.[30]

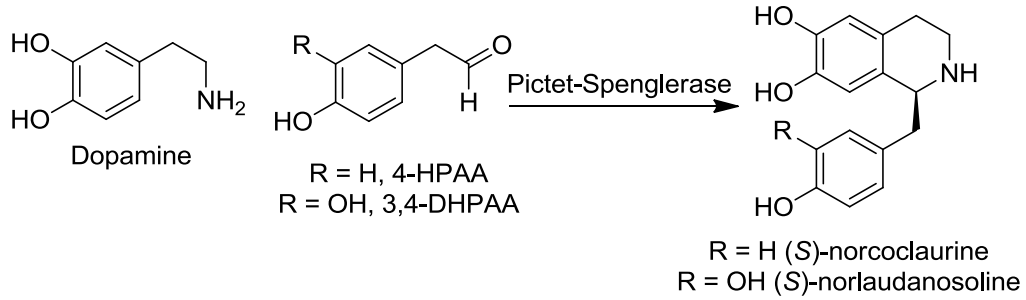


Fig. 1.4 Activities of Pictet-Spenglerase extracted from *E. tenuifolia*[30]

Later, radioactive feeding experiments showed that dopamine was not the precursor to the benzylic component of BIAs; this discovery led to a revision of the BIA pathway, and norcoclaurine was identified as the key intermediate. The hydroxylation of the aldehyde at the 3-position of the aromatic ring, was found to occur after the PS reaction, once the BIA skeleton had been formed. The natural activity and name of the BIA PS enzyme was correctly revised and the enzyme was re-annotated to (*S*)-norcoclaurine synthase (NCS).[26,31,32]

Despite predating subsequent NCS papers by almost twenty years, this early work on NCS (or norlaudanosoline synthase) conducted by Zenk's group contains valuable information. Crucially, they identified that a background chemical reaction produces significant quantities of racemic product, and took this into careful consideration when purifying and analysing the protein.[30] The enzyme was purified from *E. tenuifolia* plant cultures through a series of columns, and emerged as a 15 kDa monomer, with a pI of 7.6. There was evidence for three other isoenzymes, with pIs of 4.6, 5.3 and 6.7. Based on recent work we can speculate that these may have been different N- and C-terminal truncates. The enzyme was found to be most active at pH 7.4 for 4-HPAA and at pH 7.8 for 3,4-DHPAA. Activity dropped off gradually at more acidic pHs, but more sharply in basic conditions. Kinetic experiments showed dopamine K_m = 1.5 mM, 3,4-

DHPAA K_m = 0.7 mM and 4-HPAA K_m = 0.9 mM.[30] No evidence of dimerisation, co-operation, product inhibition or substrate inhibition was observed.

1.2.2 Norcoclaurine synthase

The subsequent studies on NCS were carried out by Facchini and co-workers. In 2001 Samanani and Facchini presented data from three partially characterised NCSs from *Papaver somniferum*, *E. tenuifolia* and *Thalictrum flavum*.[33] Enzymes were not purified, but extracted from whole plant tissue. The NCSs showed different behaviour to the enzyme purified in prior studies. All three enzymes had the highest activity at around pH 6.5-7 and 50 °C. Dopamine kinetics were observed as being sigmoidal, suggesting cooperativity, whilst 4-HPAA showed typical Michaelis-Menten kinetics with K_ms between 0.4-1 mM.[33]

Next, *T. flavum* NCS (*Tf*NCS) was purified and characterised further. Similar to Zenk's work, 4 isoenzymes of NCS were identified; although in this instance NCS was characterised by SDS-PAGE as a 28 kDa dimer, with monomers of 15 kDa.[15] Kinetic experiments of this pure protein supported previous data collected with the crude extracts;[33] both experiments identified cooperativity with respect to dopamine. This work also indicated that NCS was inhibited by norlaudanosoline. The mechanistic conclusion from these studies was that NCS follows an iso-ordered bi-uni mechanism with 4-HPAA binding before dopamine. The difference in the kinetic parameters determined by Samamani and Facchini compared to those determined by Zenk and co-workers may be partially accounted for by the use of different substrate saturating concentrations. The former used 280 µM of dopamine and 2 mM 4-HPAA; the latter used higher concentrations of approximately 5 mM for both substrates.[15,30]

In 2004 the Facchini group isolated the gene encoding NCS from *T. flavum*. This was achieved by constructing primers based on the peptide sequence obtained by analysis of the purified plant enzyme, and using a polymerase chain reaction (PCR) to retrieve the complementary deoxyribonucleic acid (cDNA).[34] The recombinant protein did not express in *Escherichia coli* in its full length form; N-terminal truncations (Δ10 or Δ19) were required for successful expression. The enzyme showed a mass of approximately 32 kDa on SDS-PAGE analysis, possibly corresponding to a dimer. Kinetic parameters were performed on crude cell extracts and provided similar values to those obtained from the native purified enzyme.[15] Computational analysis led to the identification of a 19 residue N-terminal sequence as a signal peptide for subcellular localisation.[34] Although only a single NCS gene was isolated from *T. flavum*, Southern blot analysis showed that there are three to five NCS genes in the *T. flavum* genome.

1.3 PR-10 family

1.3.1 PR-10/Bet v 1

The gene encoding for *Tf*NCS was found to have substantial homology to the Bet v 1 allergen and pathogenesis related-10 (PR-10) protein families, found in a variety of plant species.[34] It has particularly close identity (35%) with the postulated hypericin synthase HYP1 from St John's wort (*Hypericum perforatum*).[35] However, HYP1 activity has not been consistently observed and this protein's function remains unclear.[36–38] NCS shows no homology to the other well characterised Pictet-Spenglerase from plants, strictosidine synthase (STR).

Plant genes that are upregulated in times of stress are referred to as pathogenesis related (PR) genes. The PR-10 family share an evolutionarily origin and a structure: the Bet v 1 fold (Fig. 1.5).[39] The Bet v 1 fold is named after the major birch pollen allergen (Bet v 1), which was the first protein of its type to have an elucidated structure.[40] The Bet v 1 fold is distributed across all domains of life and is thus predicted to have an ancient origin in the last universal common ancestor.[41] Aside from the plant Bet v 1/PR10 family, the Bet v 1 superfamily includes families with diverse activities: lipid transfer (START family), hydroxylase activity, phosphatidylinositol transfer (PITP family), ATPase regulation (AHA1 family) and polyketide cyclase activity.[41] Bet v 1 superfamily members typically have a large, solvent-accessible hydrophobic cavity, which is able to bind bulky hydrophobic ligands. The diverse activities of Bet v 1 superfamily members suggest this scaffold is evolutionarily tractable for the innovation of novel functions, including enzyme activities.

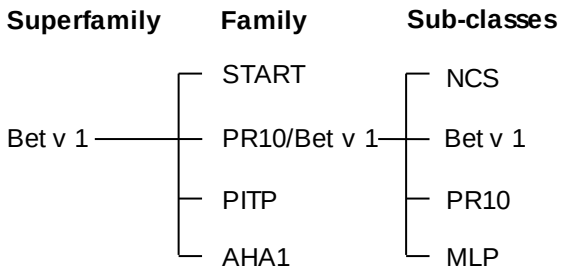


Fig. 1.5 Phylogeny of NCS proteins. Schematic describing the superfamily/fold and family relationships of the NCS clade. The diagram is illustrative and not comprehensive, as there are at least six other sub-classes in the PR-10/Bet v 1 family and other families in the Bet v 1 superfamily.

The PR-10/Bet v 1 family can be split up into eleven subclasses.[41] These include the classic PR-10s, the major pollen allergens (e.g. Bet v 1), the major latex proteins (MLP) and the NCS family.[39] Despite fairly low sequence similarities, members of the PR-10/Bet v 1 family are very similar structurally. They have a seven-stranded β-sheet

which curves around a long C-terminal α-helix. Two small α-helices connect the first N-terminal β-strand with the rest of the sheet (Fig. 1.6). Variations in the structure are typically localised to the α-helices. Compared to the rest of the sequence, the glycine rich loop L4 (between β-strands 2 and 3) is extremely well conserved and, despite the presence of glycines, the loop is surprisingly rigid in crystal structures.[39] The activities of these proteins are very diverse, and include RNase, anti-pathogenic, membrane binding and ligand binding activities. The Bet v 1 and MLP subclass of PR-10 proteins display strongly allergenic properties. This human immune response has led to investigations into the epitopes and mechanism of the immune response. Recently an epitope from Bet v 1 was grafted onto Δ29*Tf*NCS to enable elucidation of the amino acid constituents of the epitope.[42]

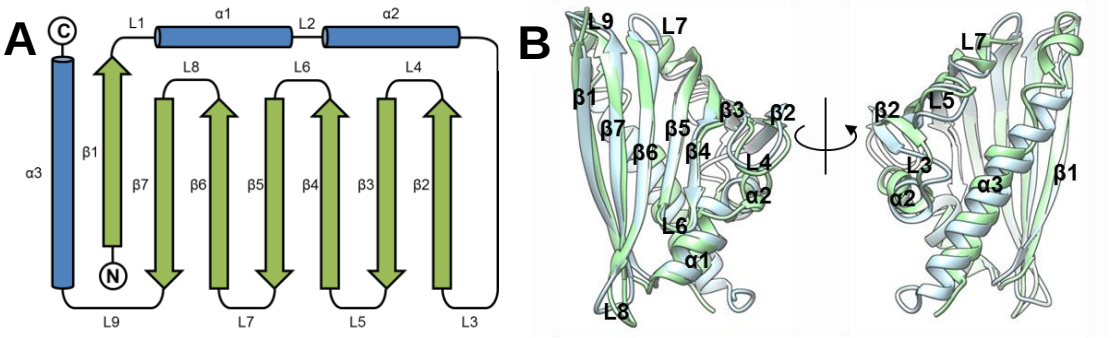


Fig. 1.6 PR-10/Bet v 1 family structure. A. Schematic of Bet v 1 structural topology.[39] **B.** Bet v 1 (light blue) and *Tf*NCS (light green) structural alignment, with secondary structure elements labelled. Structural alignment and figure constructed in Chimera (UCSF). PDB accession codes: Bet v 1[40], 1bv1; *Tf*NCS[43], 2VNE-A.

1.3.2 The NCS clade

Sequence homology (using *Tf*NCS) enabled the identification of two NCSs from opium poppy (*Papaver somniferum*), *Ps*NCS1 and *Ps*NCS2.[7] These two NCSs have 86% identity with each other, and both have approximately 40% identity with the *Tf*NCS homologue. These were shown to catalyse the NCS PS reaction. Other closely related PR-10 proteins, such as two others from *P. somniferum*, *Ps*PR10-1 and *Ps*PR10-2, did not catalyse the NCS reaction. From this study it was unclear whether active NCSs grouped together as a single clade phylogenetically. Further evidence for the NCS grouping was obtained through the testing of PR-10 proteins from Persian poppy (*P. bracteatum*) and Mexican prickly poppy (*Argemone mexicana*).[44] The *P. bracteatum* protein (*Pb*NCS) showed NCS activity (as does a closely related *Ps*NCS3[45]), but the *A. mexicana* PR-10 (*Am*PR10) did not. Phylogenetic analysis shows that the non-NCS proteins *Ps*PR10-1, *Ps*PR10-2 and *Am*PR10 are within the NCS clade but show no activity. The most likely explanation for this is that there was a monophyletic origin of NCS activity, but non-NCS PR-10 proteins within this clade subsequently lost activity.

This is considered more likely than convergent evolution of NCS activity (and BIA metabolism) in closely related PR-10 proteins.[7]

The Japanese gold thread plant (*Coptis japonica*) produces large quantities of berberine type BIAs. A library of expressed sequences tags from *C. japonica* was screened for NCS activity. Two proteins with NCS activity were identified: *Cj*PR10A (later renamed to *Cj*NCS2), and *Cj*NCS1. The former is closely related to *Tf*NCS, with 62% identity. Interestingly, it was shown to accept 4-hydroxyphenylpyruvate as a substrate.[46] The *Cj*NCS1 enzyme is not a PR-10 and has homology to 2-oxo-glutarate dependent, iron binding enzymes. It was subsequently determined that this enzyme does not catalyse the NCS reaction, but perhaps caused increased background reaction, providing an explanation for the erroneous annotation.[44] *Cj*NCS1 actually has close homology with demethylation enzymes involved in BIA biosynthesis, so it is possible that the activity observed was a result of an equilibrium shift in the cell extracts.

A further verification of the NCS monophyletic clade was the cloning and expression of putative NCS proteins from *Arabidopsis thaliana*.[47] One of these was a PR-10 protein, though lacked the crucial NCS catalytic residues. The other four NCSs were homologous to *Cj*NCS1. None of these proteins showed NCS activity. Although it is debated whether *A. thaliana* produces complex alkaloids, there remains no evidence for the presence of an NCS.[48]

The recent progress with plant transcriptomic libraries has led to a significant increase in the availability of plant gene sequences.[49,50] With these new resources and with the increasing fall in sequencing costs, the sequences of more NCSs will now become available.

1.4 NCS structure and mechanism

For an in depth discussion of structure and mechanism see section 4.1.

1.4.1 Overall mechanism

The mechanism of NCS was first investigated by Luk *et al* by examining the activity of the enzyme (Δ19*Tf*NCS) with various substrate analogues.[51] These analogues included methylated dopamine species: methylation of the 3-hydroxy abolished activity but methylation of the 4-hydroxyl did not (Fig. 1.7). This suggested cyclisation occurred via a σ-intermediate rather than a spirocyclic intermediate. Kinetic isotope effects investigated by the use of deuterated dopamine led to the identification of the reaction rate limiting steps. Overall the aromatic electrophilic addition was found to be rate

determining, whilst the subsequent rearomatisation step was shown to be partially rate determining and irreversible. Along with these mechanism determining experiments, a new method for determining kinetic rates was developed using circular-dichroism (CD) spectroscopy. The kinetic data obtained mostly supported previous work, though a much greater K_m for dopamine was observed: around 400 μM compared to previous values of 2.5 μM. This discrepancy was likely due to the low 'saturating' concentrations of substrates in previous work.[15,51]

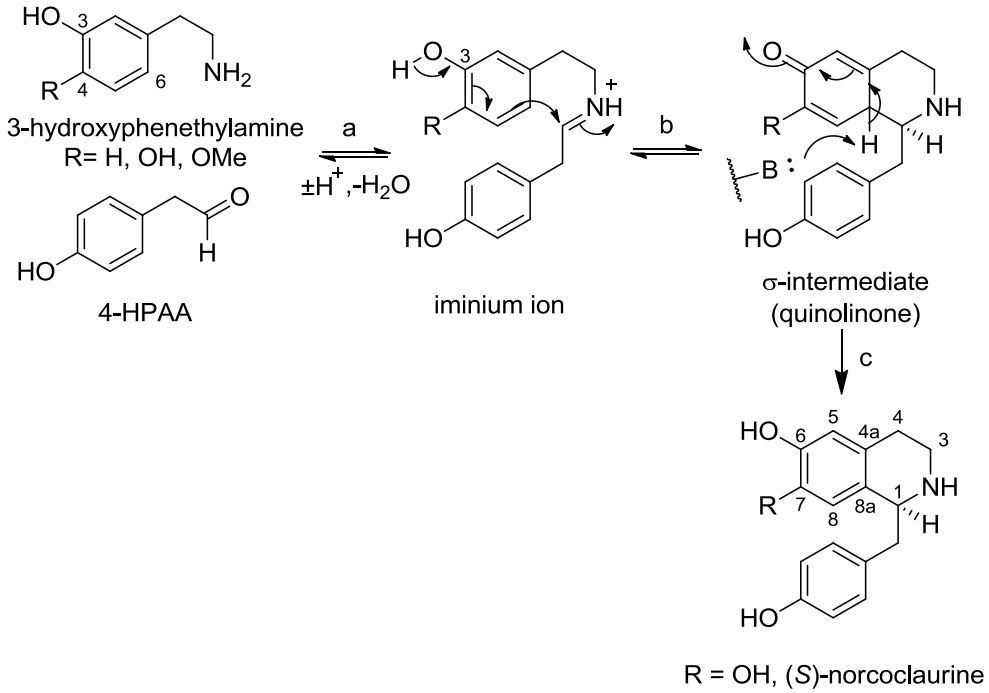


Fig. 1.7 Outline of mechanism proposed by Luk *et al*[51]. (a) iminium ion formation, (b) aromatic electrophilic addition (rate determining), (c) rearomatisation (partially rate determining). The 3-hydroxy group on the phenethylamine is vital for the reaction progression.

1.4.2 NMR investigation

In order to investigate the enzyme mechanism more thoroughly, it was necessary to conduct structural studies of NCS. As a small protein (~20 kDa), NCS was thought to be amenable to nuclear magnetic resonance spectroscopy (NMR) analysis. Thus the first of the structural studies was an NMR study, conducted by Rosch and co-workers. Their initial attempts at recombinant expression of Δ19*Tf*NCS led to the identification of 2 isoforms of NCS present during purification. One of these was the expected Δ19*Tf*NCS product, whilst the other was identified as a truncate: 10 N-terminal residues of Δ19*Tf*NCS were being removed during the expression/purification process, forming Δ29*Tf*NCS. A Δ29*Tf*NCS construct was directly expressed to avoid this problem.[52] This behaviour, upon purification, has not been reported by other groups who have expressed Δ19*Tf*NCS.[43,51,53] The expression of Δ29*Tf*NCS is biologically

justified as the N-terminal extension is not part of the globular structure of the enzyme and it is likely to be part of a signal peptide, which may be cleaved *in vivo*. For a more in depth discussion of the N-terminal extension of NCS, see section 3.6.

A homology model of Δ29*Tf*NCS was built based on the Bet v 1 structure and data from the NCS NMR measurements.[54] Aside from this model, the NMR study contained numerous valuable observations. Unlike observations from native NCS or Δ19*Tf*NCS, Δ29*Tf*NCS was shown to be monomeric in solution at concentrations of up to 1 mM. Monitoring chemical shifts during the titration of dopamine and a 4-HPAA substrate analogue (methyl 4-hydroxyphenylacetate) enabled the identification of substrate binding sites. Titration of the 4-HPAA substrate analogue led to significant shifts across the entire protein, pointing to a conformational change being triggered upon 4-HPAA binding.[54] Titration of dopamine caused a change in shifts in two regions: around residues 97-100 and residues 178-183; though these residues were also affected by non-specific binding of a phenylalanine control. It was observed that dopamine did not bind to the protein after the 4-HPAA analogue was bound. This was at odds with previous indications from kinetic investigations that 4-HPAA bound to the enzyme prior to dopamine.[34,51] The NMR data raised many questions about the enzyme mechanism, but the use of a 4-HPAA substrate analogue in place of the natural substrate undermined the significance of some of the results.

1.4.3 X-ray structure and mechanism

The X-ray structure of Δ19*Tf*NCS was solved in 2008, in both its *apo* form and a *holo* form with two ligands. As predicted through sequence analysis and the NMR homology model, the protein has a Bet v 1 type fold, characteristic of plant phytohormone carriers (see section 1.3).[43,55] Potentially important functional residues in the mostly hydrophobic active site were identified: Tyr-108, Glu-110, Lys-122 and Asp-141 (Fig. 1.8). The *holo* structure contained both dopamine and an aldehyde analogue, 4-hydroxybenzaldehyde (4-HBA), bound in the active site. 4-HBA was bound deep in the active site, the carbonyl group closely associated with Lys-122 (Fig. 1.8B). The dopamine was bound closer to the active site entrance. This supported a mechanism where the aldehyde binds to the enzyme prior to dopamine, a scheme originally proposed by the Facchini group based upon kinetic analyses.[15,51]

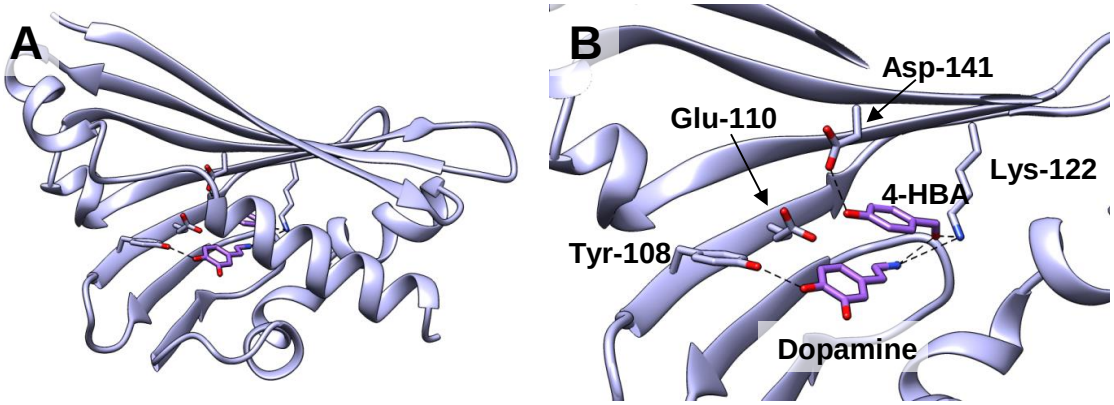


Fig. 1.8 Δ19*Tf*NCS *holo* structure.[43] **A.** Cartoon representation of NCS monomer. **B.** NCS active site (α3-helix removed). Sticks show active site residues with predicted catalytic role. 4-HBA (4-hydroxybenzaldehyde) and dopamine are ligands soaked into crystals. PDB accession number: 2VQ5 subunit B.

The observed binding modes of the ligands in the NCS active site led to the development of a mechanistic scheme (Fig. 1.9).[56] Although the mechanism could account for substrate binding, imine formation and the final rearomatisation step, it failed to provide a basic residue to abstract the 3-hydroxy proton from dopamine, a step that was previously shown to be partially rate determining.[51] The authors invoked the use of water as a base and Tyr-108 as a stabilising factor; but water is insufficiently basic to achieve this deprotonation, and no residue was proposed to interact with water to increase its basicity and enable a 'domino' deprotonation. In the next reaction step, the proposed Tyr-108 interaction does not seem to provide thermodynamic stabilisation to the intermediate quinolinone. The importance of Lys-122 was verified by the absence of activity in a K122A mutant. The mutations E110A and Y108F showed a fall in the k_{cat} (turnover number) of the enzyme, and an increase in K_ms (Michaelis constants), though the actual effect on the mechanism remained unclear. Both the kinetic experiments and the crystal structure of NCS provided no evidence of the cooperativity predicted in previous studies of NCS. Limitations of using these X-ray crystal structures as a basis of the enzyme mechanism derive from two features of the study. First, the protein was crystallised under acidic conditions (pH 4.5), conditions at which the enzyme is typically inactive. Second, 4-HBA has very different electronic properties to the natural substrate, 4-HPAA, and may interact with active site in a different fashion. For a more detailed discussion of the mechanism and crystal structure, see section 4.1.

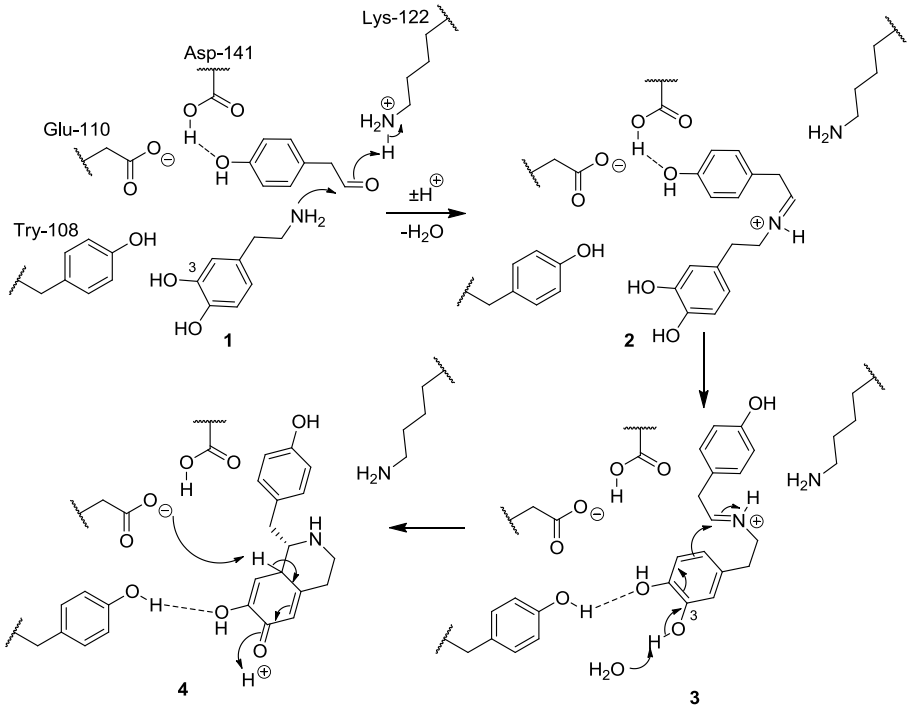


Fig. 1.9 NCS mechanism proposed based on X-ray studies.[43,56] (1) The aldehyde carbonyl is activated by Lys-122. Dopamine attacks the carbonyl and proton exchange with Lys-122 lead to the formation of the iminium ion (2). (3) Stabilised by Tyr-108, the electrophilic aromatic substitution occurs. (4) Glu-110 removes the proton causing rearomatisation and product formation.

1.4.4 Revision of mechanism

Whilst structurally diverse aldehydes are accepted by NCS, the majority of modifications of the amine substrate are not tolerated.[53,57] This behaviour seems at odds with the Ilari *et al* mechanism where the aldehyde is deeply buried in the active site, and the indispensable 3-hydroxy group of the dopamine is solvent exposed.[43,56] In order to understand the binding modes and mechanism further, Pesnot *et al* conducted computational docking experiments.[57] Reaction intermediates were docked into the active site of the Δ19*Tf*NCS crystal structure and these revealed a binding mode where the dopamine portion of the intermediate is buried in the active site, whilst the aldehyde R-group is partially solvent exposed. Such a binding mode could account for the different substrate promiscuities of the enzyme towards the two substrates. A novel mechanism was elucidated based on the binding mode observed from the docking calculations (Fig. 1.10).[57] For a full discussion and exploration of docking and mechanism see section 4.

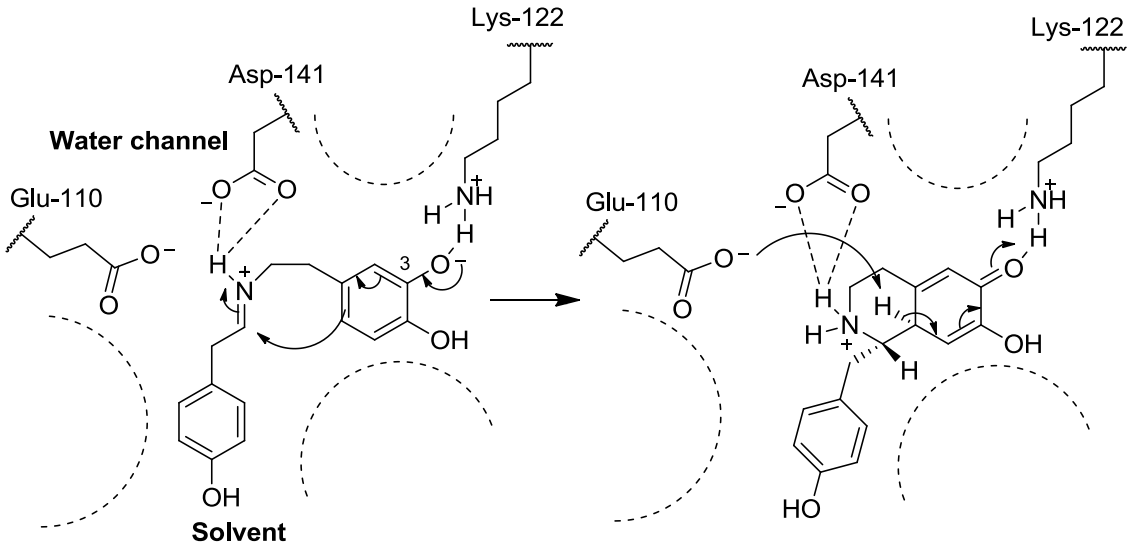


Fig. 1.10 'Dopamine first' mechanism.[57] The two key steps of the novel mechanism are shown. LHS: electrophilic alkylation, catalysed by abstraction of the 3-hydroxy by Lys-122. RHS: rearomatisation of the quinolone intermediate, catalysed by Glu-110. Dotted lines show hydrophobic interactions.

1.5 NCS biocatalysis and substrate scope

1.5.1 Hypochlorite biocatalysis

The tetrahydroisoquinoline (THIQ) moiety is feature of many bioactive natural products and synthetic drugs. A facile, stereoselective synthesis of THIQs would be of significant use to the fine chemical and pharmaceutical industries. The only gram scale synthesis using NCS was conducted with the natural substrates dopamine and 4-HPAA, using purified Δ19*Tf*NCS as the catalyst.[58] The method involved the oxidative decarboxylation of tyrosine by sodium hypochlorite to form 4-HPAA (Fig. 1.11). Upon completion of this reaction, dopamine, NCS and ascorbate were added to the reaction mixture. The *in situ* synthesis of 4-HPAA prevented the need to isolate and purify the aldehyde—it is a reactive compound, prone to oxidation and thus difficult to purify. The ascorbate prevented the autoxidation of dopamine, allowing concentrations to reach 10 mM. The use of activated carbon for extraction enabled the separation of unreacted dopamine and 4-HPAA from the norcoclaurine product, precluding extensive purification. The conversion was achieved with an enantiomeric excess (*ee*) of 93% and yield of 81%. Although the reaction was conducted in phosphate buffer, known to catalyse the racemic condensation of dopamine and 4-HPAA,[59] the enzyme acted at greater rate than the background reaction, allowing a route to (*S*)-norcoclaurine as the major enantiomer. A similar hypochlorite procedure was used to generate halogenated aldehydes *in situ* and these were successfully turned over (with dopamine) by Δ19*Tf*NCS.[60]

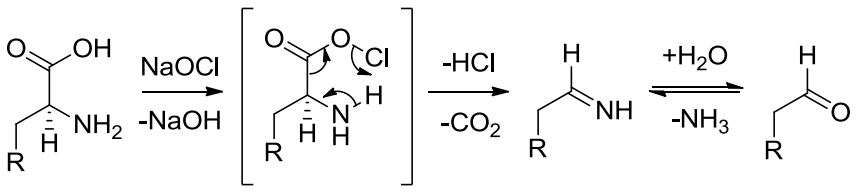


Fig. 1.11 Mechanism of aldehyde formation from amino acids by hypochlorite.

1.5.2 Substrate diversity

The catalytic versatility of NCS was first demonstrated in two papers published in 2012. One of these papers, by the group of O'Connor, was a screen of Δ19*Tf*NCS with a number of different aldehydes.[53] A high-pressure liquid chromatography (HPLC) assay detecting a reduction in dopamine concentration was used, and product formation was confirmed by mass spectrometry (MS) analysis and verified standards, however, *ee*s were not determined. Many different substituted phenylacetaldehyde substrates were turned over by the enzyme (Fig. 1.12). A number of aliphatic acetaldehydes were also accepted, though α-substituted aldehydes were not. Perhaps the most significant result of the study was the tolerance of the enzyme for 1-napthylacetaldehyde, a large substrate. This is notable because the mechanism proposed based on the X-ray structure would require both the 1-naphthyl acetaldehyde and dopamine bound in the active site simultaneously, which seems unlikely considering the comparative sizes of the active site and the substrates.[43,53]

The substrate promiscuity of NCS was explored by Pesnot *et al.* using *Cj*NCS2.[57] Reactions were screened using a novel fluorescamine based assay, which enabled the rapid monitoring of dopamine consumption. Results from amine screens confirmed the importance of the 3-hydroxy group, although some substitutions on the ethyl portion of the amine were tolerated.[51] For the aldehyde substrate it was shown that, in a similar fashion to Δ19*Tf*NCS, substitutions on the aromatic group were mostly tolerated by the enzyme; α-substituted aldehydes were not. The long chain aliphatic aldehyde heptanal was also shown to be turned over by the enzyme, a substrate very different to the natural substrate 4-HPAA (Fig. 1.12). Furthermore, the conversion of dopamine and hydrocinnamaldehyde by NCS to form 1-phenethyl-THIQ indicates that NCS or a homologue may be involved in the biosynthesis of phenethyl-THIQ alkaloids such as colchicine and dysoxyline.[61]

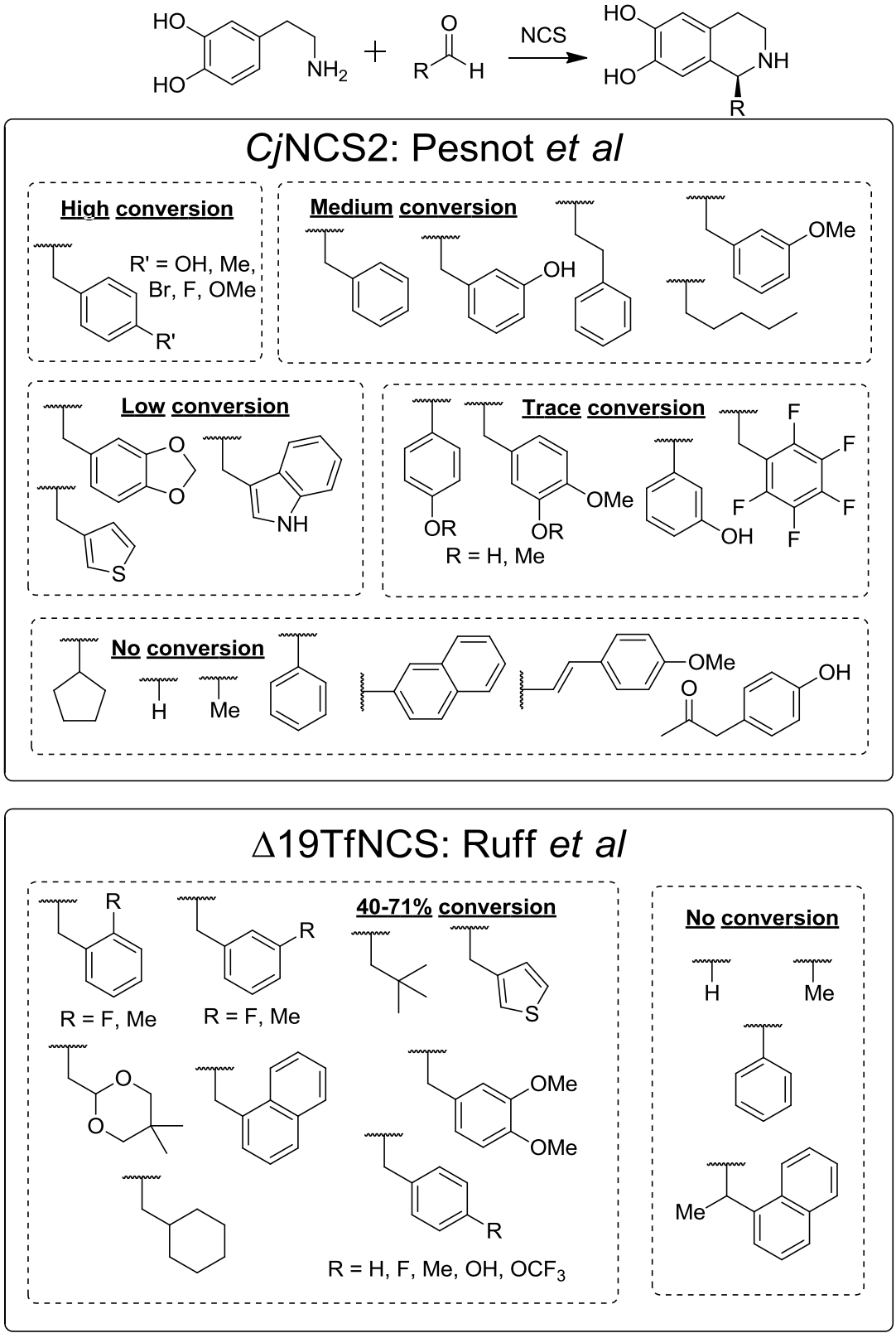


Fig. 1.12 Substrate tolerance of NCS. Various aldehydes accepted by *Cj*NCS2[57] and Δ19*Tf*NCS[53]. Aromatic acetaldehydes were mostly accepted. Notable aldehydes accepted here included the large indole and naphthyl ring systems, as well as a number of aliphatic groups. For *Cj*NCS2, conversions were measured by tracking consumption of dopamine with a fluorescamine assay. Conversions measured with loss of fluorescence: high = >80%, medium = 40-80%, trace = 10-25%, no conversions = <10%. For Δ19*Tf*NCS conversions were measured by tracking consumption of dopamine with HPLC

The in-depth study of the substrate tolerance of *Cj*NCS2 also included 4 representative milligram scale biotransformations.[57] The reactions involved the incubation of substrates with *Cj*NCS2 in HEPES buffer (4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid). Product isolation was achieved by preparative-HPLC. Three of the reactions involved the use of unnatural substrates. The products were fully characterised by NMR and Mosher's acid analysis, proving for the first time the absolute (*S*)-stereochemistry of the NCS product. Exceptional *ee*s were accounted for by the use of HEPES buffer rather than phosphate buffer.

1.6 Synthetic biology approaches towards BIAs

Whole cell conversions are a particularly attractive method for industrial biotransformations. Fermentative multistep chemical conversions can be achieved through the expression of multiple biocatalytic enzymes, often a biosynthetic pathway, in a single micro-organism. The development of these sophisticated biocatalysts is a major goal of synthetic biology; achievements thus far include the fermentative production of biofuels[62] and drug precursors[63]. The BIA pathway has attracted some interest within this field: the pathway is diverse and modular, many enzymes are characterised and the products have biomedical interest. Most notably the pathway includes morphine, which is an iconic target for synthetic efforts due to its historical significance, structural intricacy and continued medical importance.

1.6.1 Approaches without NCS

A number of synthetic biology approaches to BIAs use a BIA feedstock, rather than having the BIA structure made *in situ* by NCS.

The Smolke group used the yeast *Saccharomyces cerevisiae* as a host, and focussed their initial efforts on the pathways downstream of the formation of the THIQ skeleton by NCS. In one example *rac*-norlaudanosine was administered to a transgenic yeast and conversion to reticuline, sanguinarine and berberine precursors was demonstrated.[64] Co-expression of CYP2D6, a cytochrome P450 from human liver with broad substrate tolerance, was proposed to convert reticuline to salutaridine, the first morphinan alkaloid and a precursor to morphine. Although the enzyme seemed to accept reticuline as a substrate and catalysed the formation of a new C-C bond, the product was not fully characterised and was likely to be a mixture of numerous cyclisation products. Some subsequent work from the group focussed on developing yeast strains for the tailoring of thebaine, a morphinan alkaloid, into various natural and semi-synthetic morphinan alkaloids.[65]

The biosynthetic pathway from norcoclaurine to dihydrosanguinarine, a benzophenanthridine alkaloid, was successfully incorporated into *S. cerevisiae* first by Facchini, Martin and co-workers,[66] and later by Smolke and co-workers.[67] In both studies norlaudanosoline was used as a racemic feedstock. At the time of the former study, it was thought that the three enzymes converting norcoclaurine/ norlaudanosoline to reticuline were not enantioselective and therefore would produce racemic (*R*,*S*)-reticuline. With this expectation in mind, the formation of morphine in yeast was attempted using a metabolic short-cut in which the key intermediate (*R*)-reticuline would be produced directly from (*R*)-norlaudanosoline. This contrasts with the plant biosynthetic pathway where (*R*)-reticuline is formed from (*S*)-reticuline *via* an epimerisation. However, in the investigation, the formation of (*R*)-reticuline from *rac*-norlaudanosoline was not observed, and the three reticuline forming enzymes were revealed to be enantioselective, specific only to (*S*)-isomers.[68] This attempt to form morphine *via* a metabolic short-cut was not altogether unsuccessful, however, given that it resulted in the formation of codeine from (*R*)-reticuline.[68] Notably, this was the first verified demonstration of the CYP catalysed morphinan C-C bond forming step in yeast.

This metabolic short-cut was attempted as the plant enzyme(s) converting (*S*)-reticuline to (*R*)-reticuline had yet to be discovered. This missing enzyme, the final unknown step in the morphine biosynthesis pathway, was reported in 2015 in three separate publications by independent research groups.[19–21] Accordingly, the enzyme was given three different names: STORR (***S*-to-*R*-r**eticuline),[20] REPI (**r**eticuline **epi**merase)[19] and DRS-DRR (1,2-**d**ihydro**r**eticuline **s**ynthase/**r**eductase).[21] The enzyme is a fusion protein between a CYP, which oxidises (*S*)-reticuline to 1,2-dihydroreticuline, and a reductase, which stereoselectively reduces 1,2-dihydroreticuline to (*R*)-reticuline. It has already been incorporated successfully into transgenic yeast capable of the *de novo* production of morphinan alkaloids (see below, section 1.6.2).[21]

1.6.2 Approaches using NCS

The NCS catalysed PS step is the entry point to the BIA pathway. It is the key branch point between primary and BIA secondary metabolism. In plants, this step is controlled by cellular and subcellular localisation, and probably involves substrate channelling.[69] There is also a very high rate of spontaneous PS condensation between dopamine and aldehydes; this has been observed on many occasions[23,28,70,71] and is probably due primarily to the presence of phosphate in cells and tissues.[59] These factors have created some difficulties in generating microbial *in vivo* NCS activity.

The initial work in incorporating NCS into microbial systems was conducted by Sato and co-workers. They initially utilised *E. coli* as a host, and first attempted to incorporate a pathway beginning with dopamine. The biosynthesis of 4-HPAA from tyrosine was not fully characterised, so an alternative was constructed: 3,4-DHPAA was formed from dopamine by the action of a microbial monoamine oxidase (MAO), and this aldehyde was used in the Pictet-Spengler reaction instead of 4-HPAA, in a two enzyme one substrate 'triangular' scheme (Fig. 1.13, see also section 5.4). The formation of racemic norlaudanosoline from dopamine and MAO had been observed previously.[29,70] Furthermore, the use of norlaudanosoline rather than norcoclaurine is a feature of some synthetic biology approaches to BIA formation (probably due to its cost) and is reminiscent of the original annotation of norlaudanosoline as the product of NCS.[25,30] In the initial *E. coli* work, non-enzymatic condensation *in vivo* between dopamine and 3,4-DHPAA led to the production of racemic reticuline.[72] However, *in vitro* the same enzymes were shown to produce (*S*)-reticuline specifically. The rationale for this is that the environment of the cell is rich is in phosphate, which catalysed the non-enzymatic Pictet-Spengler reaction.[59] Upon lysis and dilution, the enzyme activity was observed above the background reaction.

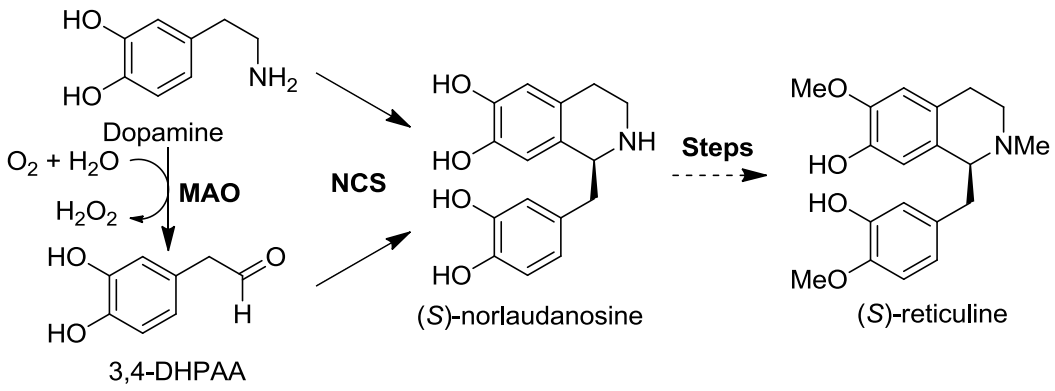


Fig. 1.13 Two enzyme, one substrate 'triangular' synthetic scheme. Monoamine oxidase converted a portion of dopamine to 3,4-DHPAA before cyclisation of dopamine and 3,4-DHPAA by NCS to form (*S*)-norlaudanosoline. Subsequent enzymatic steps led to the production of reticuline.

An extension to this work incorporated dopamine biosynthesis into the *E. coli* system, creating a system where (*S*)-reticuline was formed from glucose.[73] Interestingly, when the NCS enzyme was omitted from the system entirely, there was no effect on the *in vivo* production of (*S*)-reticuline. The explanation for this is that there is a high rate of chemical PS condensation, forming *rac*-norlaudanosoline in relatively large quantities. The enzymes downstream of NCS are enantioselective and thus form the single (*S*)-reticuline isomer. Although it is unclear why in the previous study this enzyme enantioselectivity was not observed,[72] the finding is in accordance with the later

observation that (*R*)-reticuline cannot be formed from *rac*-norlaudanosoline with the same enzymes.[68] Further optimisation of this system suggested that NCS expression levels were important for the (*S*)-reticuline yield, but the studies failed to fully verify that NCS was active *in vivo.*[74,75] Using a stepwise fermentation method, with one strain producing dopamine from glucose and a second strain expressing MAO, a system for producing *rac*-norlaudanosoline was developed; here the NCS was omitted completely and racemic products were formed.[71]

In 2015 the first fully verified case of *in vivo* NCS activity in microbes was demonstrated.[45] Dueber, Martin and co-workers constructed a *S. cerevisiae* strain able to produce (*S*)-reticuline from glucose, with NCS catalysing the key PS step. They used the plant biosynthetic pathway, with (*S*)-norcoclaurine as the intermediate. Although the production of (*S*)-reticuline from glucose was achieved previously in *E. coli*,[73] there was no verification that NCS was active. In this *S. cerevisiae* system, NCS stereoselective activity was verified through chiral HPLC and by the fact that without NCS (or other downstream enzymes) no norcoclaurine was detected (despite the presence of dopamine and 4-HPAA). The NCS used in this system was a newly detected *Ps*NCS3, which is very similar in sequence to *Pb*NCS (90% identity).[44] Interestingly the (*S*)-norcoclaurine concentration was 228-fold lower than the dopamine concentration, suggesting that the NCS catalysed step is very inefficient.

In a similar project conducted by Smolke and co-workers in which (*S*)-reticuline was produced from glucose in *S. cerevisiae*, the NCS enzymes *Cj*NCS, *Tf*NCS and *Ps*NCS were shown to have *in vivo* activity.[76] It was proposed that low concentrations of dopamine present reduced the background reaction. Activity was verified by comparing production with strains lacking the NCS enzyme. The variant *Tf*NCS had very low activity, only approximately 5 times greater than the background. The variants *Ps*NCS and *Cj*NCS had higher activity, with over 13-fold and 15-fold greater production of norcoclaurine than the background respectively.[76] This glucose to (*S*)-reticuline pathway was combined with the newly discovered DRS-DRR enzyme and downstream morphinan related enzymes to form a yeast strain capable of producing opioids in a fermentative fashion.[21] This was a milestone for synthetic biology and a significant step forward to the fermentative production of morphine. All the enzymes needed to achieve this are known and show activity in yeast. However, low titres of pathway intermediates and low enzyme conversions, especially in the final steps of the pathway, remain challenges that must be overcome before the microbial production of morphine can be finally achieved.[21,65]

1.7 Other Pictet-Spenglerases

There are over five known Pictet-Spenglerase enzymes, including NCS, and all contribute to secondary metabolism. The PS reaction is irreversible, causes a significant increase in molecular complexity and does not require any co-factors.. These factors have contributed towards its repeated evolution in secondary metabolism. There are also a number of secondary metabolite pathways with postulated PS reactions where an enzyme has not been identified.

1.7.1 Strictosidine synthase

The best characterised Pictet-Spenglerase, other than NCS, is strictosidine synthase (STR; EC 4.3.3.2), which catalyses the formation of 3-α-(*S*)-strictosidine from tryptamine and secologanin.[77] The activities of at least three different STRs have been verified.[78] The reaction it catalyses is the first committed step into the monoterpene indole alkaloids, a structurally diverse group of alkaloids that includes quinine, strychnine and vinblastine (Fig. 1.14). The reaction mechanism is different to NCS: no base is required to trigger the Friedel-Crafts step.[79] For more analysis of the STR reaction mechanism, see section 4.1.1. Modulation of STR activity has been achieved by mutations based upon the understanding of the binding modes of the substrates,[80] and by circular permutation.[81] The enzyme has also been found to turn over a variety of substrates, including an unusual substrate enabling the formation of a piperazino-indole structure (Fig. 1.14B).[82,83]

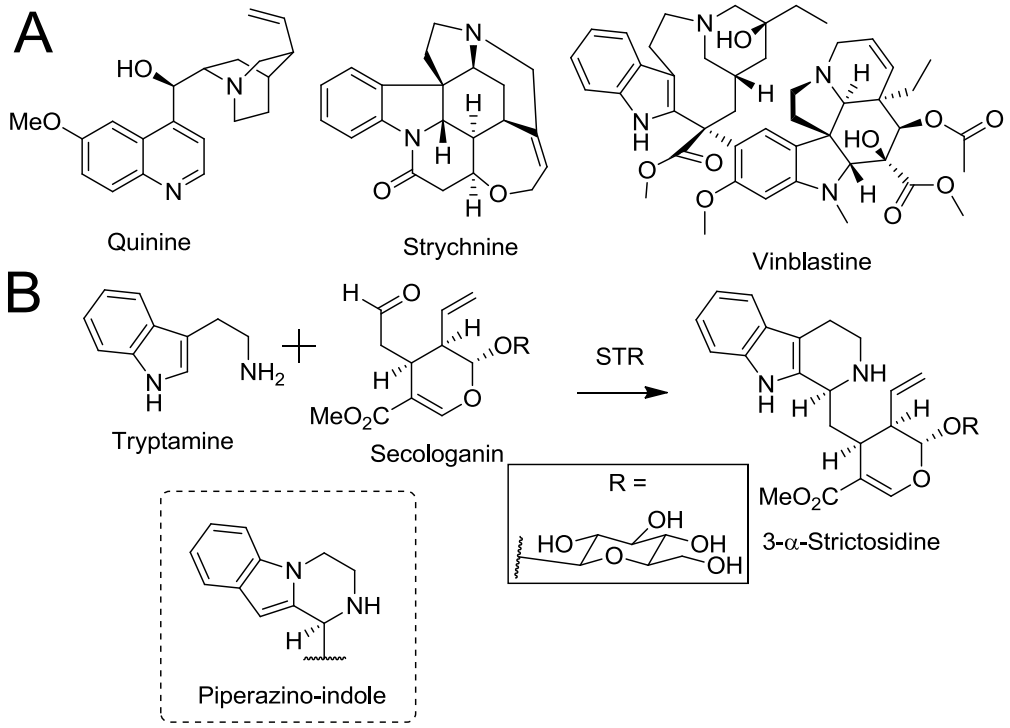


Fig. 1.14 A. Examples of monoterpene-indole alkaloids. B. Reaction catalysed by STR. STR catalyses the first committed step into the monoterpene indole alkaloids. In the dashed box is the unusual piperazino-indole structure also formed by this enzyme.

1.7.2 Fungal Pictet-Spenglerase

A remarkable study found that a fungi (*Chaetomium globsum*), which resides in the Calico grouper fish (*Epinephelus drummondhayi*) gut, generated a diverse range of indole alkaloids *via* a PS reaction.[84] The typically "silent" fungal Pictet-Spenglerase (FPS) gene was switched on with the addition of 1-methyl-*L*-trytophan (1-MT). The 1-MT then acted as a substrate for the FPS, combining *in situ* with the fungal dialdehyde flavipin. The enzyme products were then extensively modified by fungal enzymes, thereby triggering the formation of a gamut of "unnatural natural products" (Fig. 1.15). The Pictet-Spenglerase in question was annotated as STR-like; alignment between this FPS and *Rauvolfia serpentina* Strictosidine synthase showed 33% similarity. The automated annotation of the FPS places it within the SGL subgroup of the N6P superfamily. The Strictosidine Synthase-like (SSL) enzymes are found in a different subgroup. Phylogenetic evaluation of the N6P superfamily suggested only SSL enzymes have PS activity,[78] yet this study of FPS suggests that the PS STR-like activity may actually be more widely spread than previously considered.

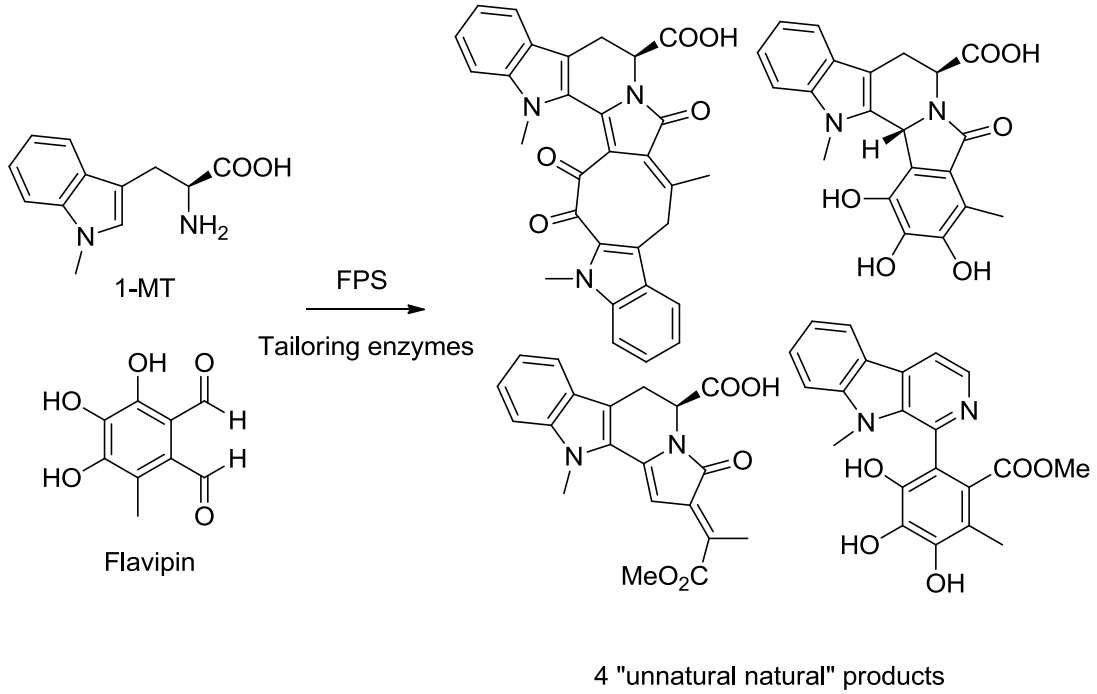


Fig. 1.15 Observed products of fungal Pictet-Spenglerase and tailoring enzymes.

1.7.3 Ipecacuanha alkaloids

The Ipecacuanha alkaloid (tetrahydroisoquinoline-monoterpene alkaloid) structure consists of the arylethylamine from the BIAs, dopamine, and the aldehyde from the indole alkaloids, secologanin. The reaction catalysed is more similar to the NCS reaction than the STR reaction, because it is the arylethylamine substrate that plays the larger role in the reaction mechanistically. Two Pictet-Spenglerase enzymes were identified from cell-free extracts of *Alangium lamarckii*: deacetylipecoside synthase (DIS, EC 4.3.3.4) and deacetylisoipecoside synthase (DIIS, EC 4.3.3.3). Both use dopamine and secologanin to make (1*R*)-deacetylipecoside and (1*S*)-deacetylisopecoside respectively, which spontaneously convert to demethyalangiside and demethyisoalaniside (Fig. 1.16).[85] DIS was purified from *A. lamarckii* and was shown to be a 30 kDa monomer with an optimum activity at pH 7.5 and 45 °C. It exhibited a high substrate specificity to dopamine; K_ms for dopamine and secologanin were 0.7 mM and 0.9 mM respectively.[86] The enzyme was inhibited by a number of Ipecacuanha alkaloids. Although only compared to STR at the time, DIS has properties more comparable to NCS. In order to evaluate the similarity of these enzymes the sequence of DIS must be elucidated. The fact that NCS can turn over a wide variety of aldehydes means that, until the DIS or DIIS enzyme sequences are determined, it remains possible that this PS reaction can be catalysed by an NCS type enzyme.

Incidentally, the ipecacuanha alkaloid pathway gives rise to the only known BIA that is not derived *via* the typical BIA pathway (Fig. 1.16). The compound bharatamine, a racemic protoberberine, is formed through the oxidation and reorganisation of the secologanin portion of deacetyl(iso)ipecoside.[87]

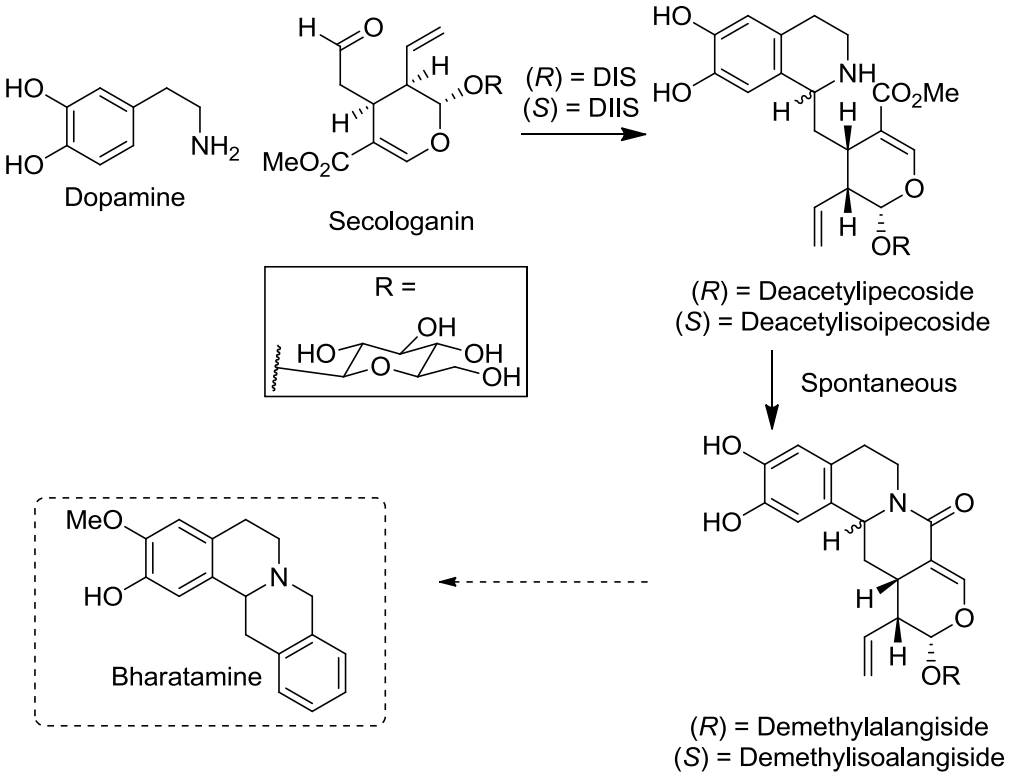


Fig. 1.16 Pictet-Spengler reaction catalysed by DIS and DIIS. The enzymes catalyse the first committed steps into the Ipecacuanha alkaloids. The compound bharatamine superficially appears to be from the BIA protoberberine biosynthesis pathway but is in fact an ipecacuanha alkaloid.

1.7.4 THIQ antitumour antibiotics

Tetrahydroisoquinoline (THIQ) antibiotics such as saframycin share a key structural motif with many plant alkaloids, but are derived from non-ribosomal peptides (NRPs). A key structural distinction between plant derived THIQs and the NRP derived THIQs, isolated from bacteria and invertebrates, is that the latter have an *ortho*-hydroxyl or ketone (position 8) and an absence of a *meta*-hydroxyl (position 6) (Fig. 1.17). The Pictet-Spenglerase domain responsible for this *ortho*-cyclisation has been partially characterised; it was isolated from *Streptomyces lavendulae*.[88] The Saframycin C non-ribosomal peptide synthetase (SfmC) condensation domain has been shown to have Pictet-Spenglerase activity, catalysing the condensation of an enzyme bound tyrosine-like arylethylamine with an aldehyde containing a long hydrophobic chain (Fig. 1.17). The regioselectivity of the cyclisation is an *ortho*-cyclisation, in contrast to the NCS *para*-cyclisation. This enzyme was also shown to be able to catalyse pyrrolidine formation *via* a Mannich-type reaction.[89]

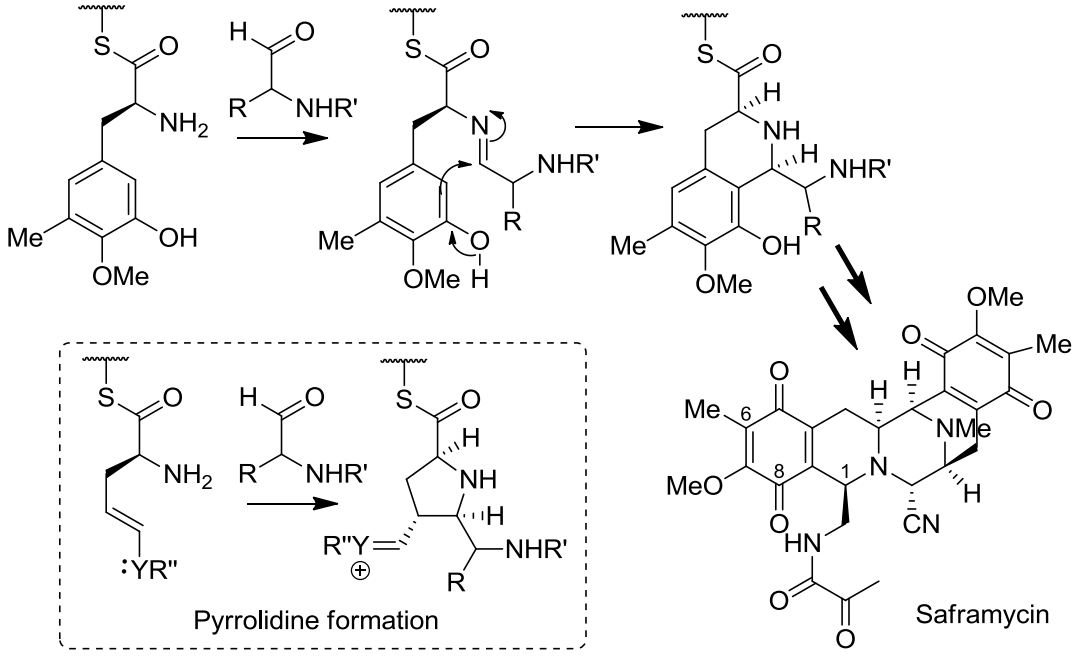


Fig. 1.17 Action of SfmC. SfmC catalyses an ortho directed Pictet-Spengler condensation in the biosynthesis of tetrahydroisoquinoline antibiotics in *Streptomyces*. Saframycin undergoes two iterative Pictet-Spengler condensations, the second being spontaneous. SfmC can also catalyse the formation of pyrrolidines.

1.7.5 β-Carboline

The β-carboline alkaloids are distributed throughout nature; they are found in higher plants, fungi, actinomycetes and marine sponges. The β-carboline Pictet-Spenglerase McbB was obtained from the actinomycete *Marinaactinospora thermotolerans*.[90] Carbon-13 labelling of substrates led to the identification of oxaloacetaldehyde as the aldehyde substrate, which is derived from oxaloacetic acid, found in the citric acid cycle. The arylethylamine was found to be the amino acid tryptophan (Fig. 1.18). Mutagenesis of Glu-97 in McbB knocked out activity completely, suggesting it has a key mechanistic role in the enzyme. It is likely that it acts a base, catalysing the rearomatisation and formation of the product. Although the enzyme has a similar mechanism to STR, it is structurally and evolutionarily distinct.[91] Originally, the variety of products observed indicated that McbB may have decarboxylase and oxidase activity, as well as the characterised Pictet-Spenglerase activity, but these steps have been found to be spontaneous and non-enzymatic. Furthermore, rational site directed mutagenesis of McbB has increased its capacity to accept bulky aldehydes.[91]

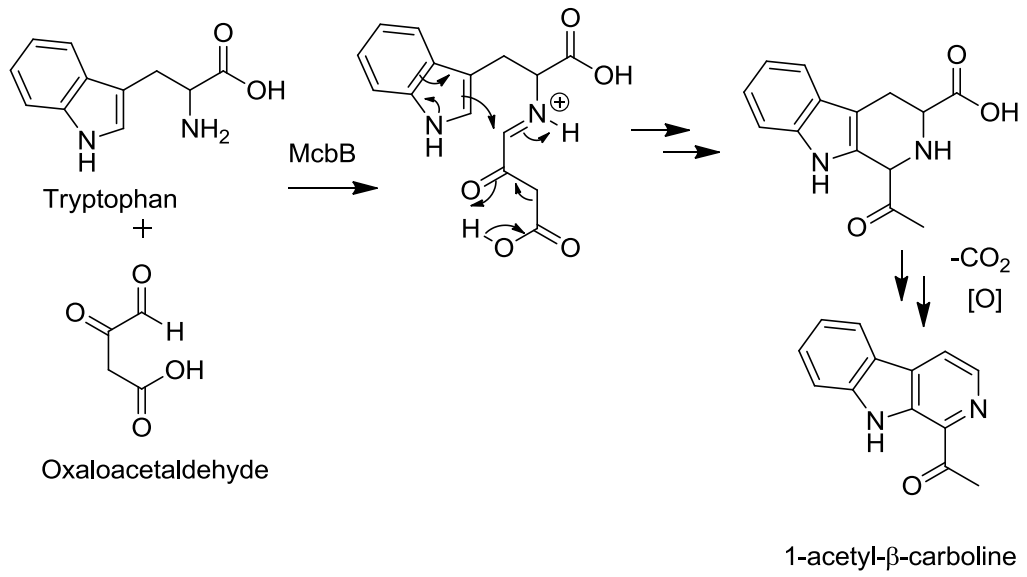


Fig. 1.18 Pictet-Spengler reaction for the formation of β-carboline alkaloids. The reactions here are catalysed by McbB. McbB has Pictet-Spengler activity; decarboxylation and oxidation occur spontaneously, which leads to the formation of a variety of products, including 1-acetyl-β-carboline, shown here.

1.8 Chemical and biocatalytic routes to THIQs

The THIQ moiety is a privileged scaffold in organic chemistry as it is found in numerous natural and artificial bioactive molecules.[92] Accordingly, there have been numerous chemical and biocatalytic methods developed for the synthesis of these compounds.

1.8.1 Chemical THIQ formation

There are many chemical methods for the synthesis of THIQs. The two key reactions for the formation of THIQs are the Pictet-Spengler reaction and the Bischler-Napieralski reaction. The Bischler-Napieralski reaction forms a cyclic imine (dihydroisoquinoline) from an amide intermediate and typically uses $POCl_3$ as a dehydrating agent. The imine can then be reduced to yield the THIQ (Fig. 1.19). Traditionally the Bischler-Napieralski reaction was considered to use milder reaction conditions than the PS reaction, in which high temperatures and acid conditions were required. This prevented the use of unstable amines and aldehydes in the PS reaction.

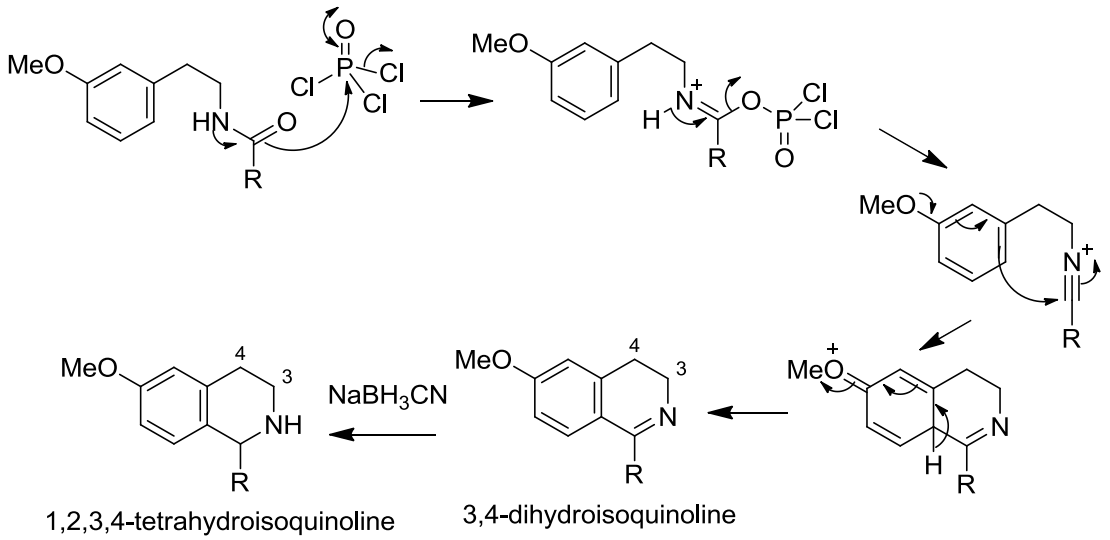


Fig. 1.19 Bischler-Napieralski reaction. Reaction mechanism with $POCl_3$. Optional reduction step yields THIQs. Other dehydrating agents can be used in place of $POCl_3$, such as PCl_5, $SOCl_2$ and $ZnCl_2$.

In 2010 Stambuli and co-workers established calcium hexafluoroisopropoxide as a PS catalyst for the formation of THIQs.[93] This catalyst was able to catalyse the conversion of unactivated ketones as well as aldehydes. Subsequently a biomimetic PS reaction using phosphate buffer was discovered by the groups of Hailes and Ward.[59] This mild, water based synthesis is applicable to a wide range of substrates. The major product of this phosphate reaction is the *para*-product, with the *ortho*-product present in small quantities.[94] The key requirement of both these PS THIQ reactions is that the arylethylamine has a 3-hydroxy group (or amine). This substrate requirement is therefore similar to the NCS reaction, and the reaction mechanisms may be comparable.

A further development of the phosphate catalysed PS reaction was the establishment of an organocatalytic reaction using BINOL derived chiral phosphate ligands.[95] The use of a hydrophobic chiral phosphate meant that the reactions had to be conducted in toluene, not water. The ligands only provided modest stereoselectivity compared to enzymes, the maximum *ee* being 86%. Crystallisation was used to further improve the product *ee*s.

1.8.2 Monoamine oxidase

There are a number of synthetic biocatalytic routes to chiral BIAs and THIQs that do not involve a Pictet-Spenglerase enzyme, however all require the formation of a racemic THIQ or a 1,2-dihydroisoquinoline prior to the enzyme catalysed step.[96] A deracemisation cascade developed by the Turner group used monoamine oxidase (MAO) and a mild reducing agent to form a chiral amine starting from either a racemate

or an imine. Enantioselective MAOs can be used catalyse the oxidation of a specific amine enantiomer to an imine using molecular oxygen. The imine produced can be reduced back to the amine (by a mild borane reducing agent) and over a number of redox cycles there will be a considerable enrichment of a single amine enantiomer. An evolved variant of MAO from *Aspergillus niger*, MAO-N-5, was used to deracemise the THIQ crispine A, providing an *ee* of 97%.[97] Another MAO variant, D11, was used for the deracemisation of a number of *N*-methylated BIAs (Fig. 1.20).[98]

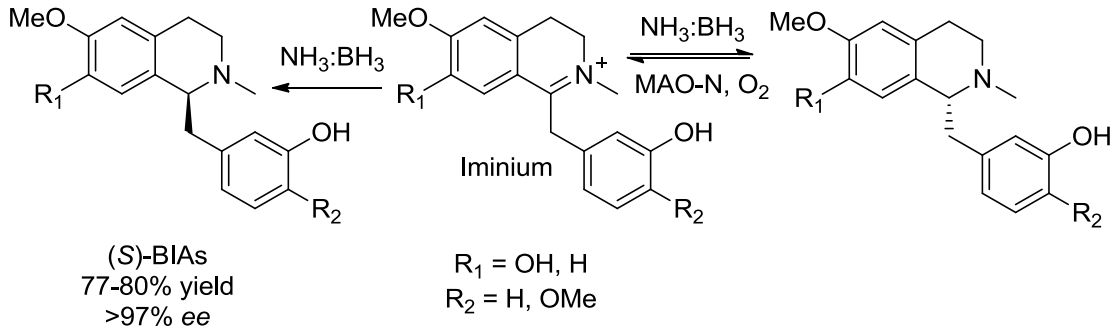


Fig. 1.20 MAO deracemisation method. Preparative scale deracemisation of *rac*-BIAs to (*S*)-BIAs. MAO-N can oxidise (*R*)-BIA but not (*S*)-BIA. With non-chiral reduction but enantiospecific oxidation, the population of (*S*)-BIA becomes enriched. The starting material is the racemic THIQ.

1.8.3 Berberine bridge enzyme

The berberine bridge enzyme (BBE) is a flavin dependent enzyme that catalyses the establishment of a C-C bond in the BIAs between an *N*-methyl group and the benzyl group.[18] It forms the protoberberine subgroup of the BIAs, which includes the compounds scoulerine, berberine and palmatine. The enzyme is enantioselective, so will only turn-over the (*S*)-isomer of a racemic *N*-methylated BIAs. This means that it can perform kinetic resolution if a racemate is used, and the maximum yield of the reaction can only be 50%. BBE has been used for the biocatalytic formation of protoberberine alkaloids (Fig. 1.21).[99–101] The use of some fluorinated substrates has been reported to result in a change in the enzyme regioselectivity.[102] Combination of the MAO deracemisation cascade (section 1.8.2) with the BBE enzyme enables conversions to reach 100%.[103]

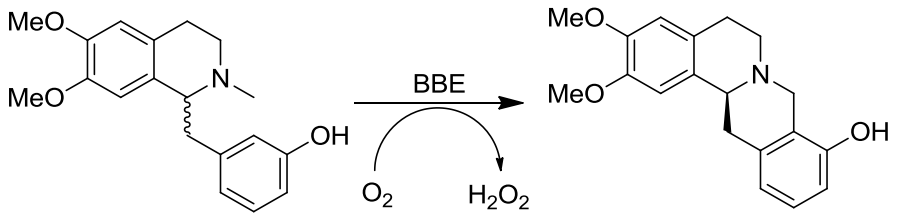


Fig. 1.21 Berberine bridge enzyme activity. A representative substrate of BBE. Catalase is typically used to break down hydrogen peroxide co-product. Only the (*S*)-enantiomer is turned over by the enzyme so maximum yield from a racemic BIA is 50%, unless a deracemisation step is incorporated (e.g. MAO).

1.8.4 Imine reductases

Recently, a new enzyme class has shown great potential in the biocatalytic formation of chiral alkaloids.[104] Imine reductases (IRED) catalyse the stereoselective reduction of imines to amines with NADPH. They have been found with complementary stereoselectivity,[105] thereby enabling the formation of both enantiomers from the same starting material. Some IREDs have been shown to be able to form THIQs,[106] although accurately predicting the stereoselectivity with novel substrates is currently not possible (Fig. 1.22). IREDs have advantages over the MAO system as a single enzyme is required rather than a coupled reducing agent and enzyme, though co-factor recycling is still required. However, for MAO, BBE and IRED methods the main structure of the molecule must be constructed prior to the biocatalytic step. Pictet-Spenglerases by comparison have the advantage that the key bond forming steps are enzymatic.

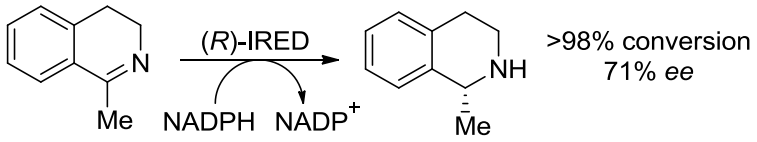


Fig. 1.22 Observed activity of an IRED. These enzymes show potential for the synthesis of chiral THIQs.

1.9 Aims

The overall aim of the project was to explore the biocatalytic utility of NCS.

1.9.1 Initial state of the project

At the beginning of this project, Dr Thomas Pesnot (Hailes group) and Dr Markus Gershater (Ward Group) were working on the chemical and enzymatic THIQ PS reaction. Dr Pesnot was primarily focussed on the chemical aspects of the reaction, while Dr Gershater was generating and testing different NCSs. The phosphate catalysed THIQ PS reaction had recently been discovered, and was being used to synthesise a variety of THIQs.[59] The biomimetic chemical syntheses of THIQs was later continued by Dr Eleanor Lamming. The THIQ products produced were investigated for antimicrobial activity.[94]

A number of synthetic genes for different NCSs had been obtained. Good activity was observed with *Cj*NCS2, and the enzyme's substrate scope was subsequently examined.[57] In order to rapidly screen NCS activities with numerous compounds, a novel fluorescamine-based assay was developed. This end-point assay measured consumption of dopamine by NCS and led to the observation that numerous aldehydes are accepted by *Cj*NCS2 (see section 1.5.2). This substrate scope could not be rationalised with the enzyme mechanism derived by the X-ray crystal structure,[58] and consequently substrate docking was used to propose a novel 'dopamine-first' mechanism (see section 1.4.4). Less progress had been made with the *Tf*NCS protein, but activity had been tentatively verified. Around the same time of the investigation into the *Cj*NCS2 substrate scope, the *Tf*NCS substrate scope was investigated by the O'Connor group and shown to be similar (see section 1.5.2).[53]

1.9.2 Enzyme mechanism

The first aim of the project was to use mutations to verify the enzyme mechanism. There are two proposed mechanisms for NCS, and it was hoped the activity of certain mutations would discriminate between the two mechanisms and reveal which was the most accurate. The enzyme *Tf*NCS would be used for this investigation, because a crystal structure was available and therefore mutations could be selected in a deterministic manner. Furthermore, the X-ray structure of *Tf*NCS formed the basis of both mechanisms (through the *holo* structure and computational docking) and therefore this was the ideal NCS to examine the enzyme mechanism.

Two types of mutations were proposed. Mechanistic mutations would probe residues involved in the enzyme mechanism: the residues Tyr-108, Glu-110, Lys-122 and Asp-

141 would be targeted. The second set of mutations would target residues more involved in aldehyde substrate binding. This latter set of mutations could potentially distinguish between the two mechanisms as the proposed aldehyde binding residues are very different in each mechanism. These mutants would be tested with a variety of aldehydes to identify how different mutations affect the activity with different aldehydes.

Selected mutations and substrates would then be investigated more thoroughly, using enzyme kinetics experiments. The change in k_{cat} or K_m revealed in kinetics could provide insight into the origin of activity changes. For mutations at aldehyde binding sites, it was expected that the K_m for the aldehyde would change, but the K_m for dopamine would remain mostly unchanged.

1.9.3 Enzyme engineering

In the process of understanding the enzyme mechanism *via* mutations, it was hoped that a number of variants with modified substrate specificities could be identified. Furthermore, verification of the enzyme mechanism would allow the undertaking of a considerable enzyme engineering effort. This enzyme engineering would begin with rational site-directed mutagenesis with the aim of modifying substrate specificity/tolerance.

The target for novel activities would initially be the modification and expansion of the aldehyde substrate scope. A range of aldehydes are already known to be accepted by NCS and expansion and improvement of these activities is achievable (see section 1.5.2).[53,57] The α-substituted, aliphatic and bulky aldehydes would be of particular interest. Subsequent effort would focus on the NCS amine tolerance, though this would likely to be more challenging as few amines are accepted. Key compounds to target would be substituted phenethylamines, including 3-aminophenethylamines and N/O methylated amines, and more diverse compounds such as phenylpropylamines (forming 7-membered ring products) and even phenethylalcohols (forming isochromans) (Fig. 1.23). More extreme modifications such as switching regioselectivity or stereoselectivity may also be considered.

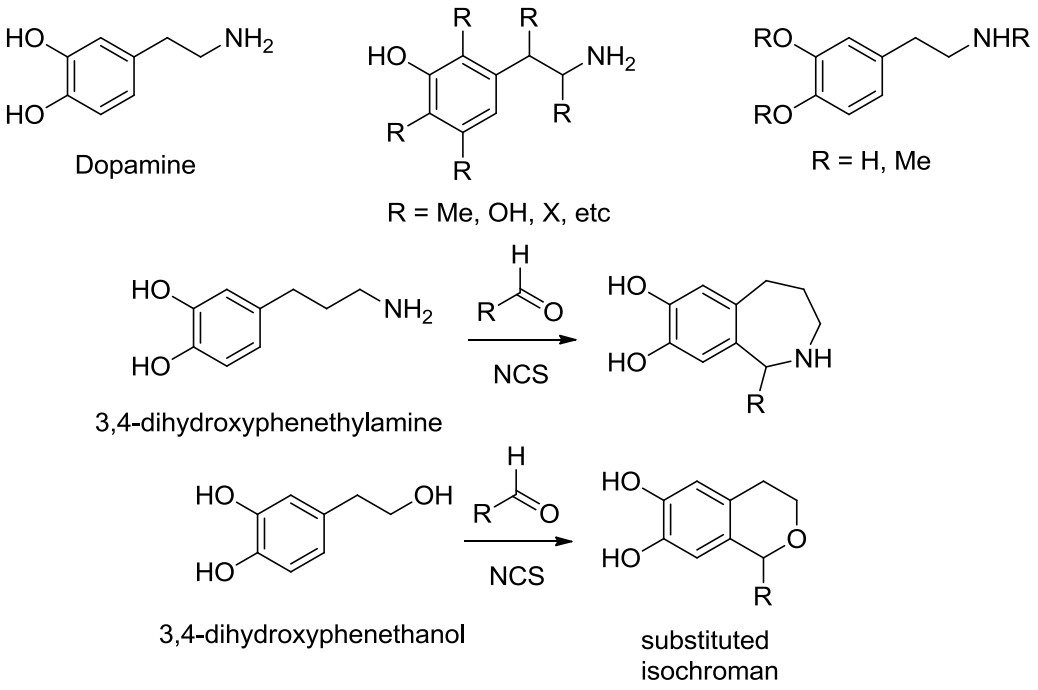


Fig. 1.23 Target amine substrates for enzyme engineering. Dopamine is the natural substrate. 3,4-dihyroxyphenylethanol is not an amine, but replaces an amine in the NCS reaction scheme.

Initially mutations would be rationally designed based on the mechanism and effected by site-directed mutagenesis. Depending on the project progression, saturation mutagenesis and more sophisticated direct evolution methods may be used.[107–111] The screening method would vary depending on the library size. Individual mutants can be purified and screened with fluorescamine detection. Hits can be verified with HPLC. For larger libraries purification may be inconvenient, and the fluorescamine assay may be inappropriate if protein purity and concentration cannot be controlled. In this instance, different screening methods may need to be developed and used.

Finally, once interesting mutants and substrates have been identified, preparative scale biotransformations will be performed. Novel products will be purified and characterised *via* typical chemistry analytic techniques such as NMR spectroscopy and mass spectrometry. The compounds may also be investigated for antibiotic activity, either in-house or by collaborators.

1.9.4 Multi-enzyme cascades

Depending on progress, it may be possible to combine NCS with other enzymes or chemicals to develop *in vitro* cascade reactions. These will generate chiral THIQs of interest. The nature of these cascades will depend greatly on the enzymes available. Enzymes will be combined with NCS in order to generate substrates (e.g. dopamine and aldehydes) or to modify THIQ reaction products. The Hailes and Ward labs have

considerable experience with transaminases and transketolases and these would especially be targeted for use in multi-enzyme cascades. The use of alcohol dehydrogenases and CYPs will also be considered (see section 5.4).

Finally, the use of NCS *in vivo* could be investigated. If successful, enzymes could be incorporated into *in vivo* pathways to mimic the natural plant biosynthesis pathway in *E. coli*. The use of different enzymes may enable novel biosynthesis pathways to be developed for the generation of novel chiral THIQs.

2 Chapter 2: Materials and methods

2.1 Genes and primers

2.1.1 Δ19*Tf*NCS

A plasmid containing the C-terminal hexahistidine tagged (His-tag) gene of Δ19*Tf*NCS was prepared by Dr Markus Gershater. Originally, synthesised genes were inserted into the pET29a(+) vector by restriction and subsequent ligation using the restriction enzymes NdeI and HindIII (Fig. 2.1).*

*CAT**ATG***CAAAAACTCATTCTCACAGGTCGGCCCTTTCTCCACCACCAAGGGATAAT
AAACCAAGTATCTACGGTCACGAAAGTAATTCATCATGAGCTGGAAGTTGCAGCAT
CAGCAGATGATATATGGACCGTTTATTCGTGGCCGGGGCTGGCTAAGCATCTGCC
GGACCTGCTCCCGGGGGCATTTGAAAAGTTGGAAATCATTGGTGATGGCGGTGTT
GGTACGATCCTTGACATGACGTTTGTACCGGGTGAATTTCCGCATGAATACAAGG
AGAAGTTTATATTAGTCGATAATGAGCATAGATTAAAGAAGGTACAGATGATTGAG
GGCGGTTATCTCGACCTGGGCGTAACGTACTACATGGACACGATCCATGTTGTTC
CGACCGGTAAAGATTCATGTGTTATTAAATCGTCAACCGAGTACCATGTAAAACCG
GAGTTTGTCAAAATCGTTGAACCGCTGATCACGACGGGTCCGTTAGCAGCTATGG
CAGACGCTATCTCAAAACTGGTTCTTGAACACAAATCGAAATCGAACTCAGATGAA
ATTGAGGCTGCAATAATAACGGTCCATCATCACCACCATCAC**TGATAG***GAAGCTT*

Fig. 2.1 Δ19*Tf*NCS (s6), pQR1024. DNA sense strand 5`-3`. Bold text shows start and stop codons. Italics shows restriction endonuclease recognition sites NdeI and HindIII used for cloning. Underlined is the 3'/C-terminal His-tag.*

All Δ19*Tf*NCS mutants were constructed using QuikChange® Lightening mutagenesis kit (Agilent, Santa Clara, CA, USA) and mutagenesis primers (purchased from Eurofins, Luxembourg) (Table 2.1). Standard QuikChange® Lightening recommended parameters and conditions were used for primer design and experimental conditions. All plasmid sequences were verified by Sanger sequencing provided at the UCL Wolfson Institute for Biomedical Research.

* The cloning of Δ19*Tf*NCS was conducted by Dr Markus Gershater

Table 2.1 Primer sequences for Δ19*Tf*NCS (s6) mutations. Nucleotide sequences numbered from start of NdeI site.

Amino acid variant	DNA mutation	Primer sequences (5` to 3`)
L76A	c175g, t176c	tctgccggacctggccccgggggcattt aaatgcccccggggccaggtccggcaga
A79F	g184t, c185t, a186t	gccggacctgctcccgggggtttttgaaaagttggaaatcat atgatttccaacttttcaaaaaaccccgggagcaggtccggc
A79I	g184a, c185t, a186t	gccggacctgctcccggggatttttgaaaagttggaaatcat atgatttccaacttttcaaaaatccccgggagcaggtccggc
F80L	t187c, t189g	gacctgctcccgggggcactggaaaagttggaaatcattg caatgatttccaacttttccagtgcccccgggagcaggtc
M97F	a238t, g240c	gttggtacgatccttgacttcacgtttgtaccgggtgaa ttcacccggtacaaacgtgaagtcaaggatcgtaccaac
M97L	a238c	tggtacgatccttgacctgacgtttgtaccggg cccggtacaaacgtcaggtcaaggatcgtacca
F99L	t244c, t246g	acgatccttgacatgacgctggtaccgggtgaatttccg cggaaattcacccggtaccagcgtcatgtcaaggatcgt
Y108F	a272t	ataaacttctccttgaattcatgcggaaattcacccgg ccgggtgaatttccgcatgaattcaaggagaagtttat
E110D	g279c	tcattatcgactaatataaacttgtccttgtattcatgcggaaattc gaatttccgcatgaatacaaggacaagtttatattagtcgataatga
E110Q	g277c	attatcgactaatataaacttctgcttgtattcatgcggaaattcac gtgaatttccgcatgaatacaagcagaagtttatattagtcgataat
D141E	c372g	acatggatcgtctccatgtagtacgttacgccc gggcgtaacgtactacatggagacgatccatgt
D141N	g370a	gaacaacatggatcgtgttcatgtagtacgttacgcc ggcgtaacgtactacatgaacacgatccatgttgttc
A182F	g493t, c494t	cgacgggtccgttagcatttatggcagacgctatct agatagcgtctgccataaatgctaacggacccgtcg
A182I	g493a, c494t	cgacgggtccgttagcaattatggcagacgctatct agatagcgtctgccataattgctaacggacccgtcg
M183F	a496t, g498t	tgagatagcgtctgcaaaagctgctaacggacccg cgggtccgttagcagcttttgcagacgctatctca
M183L	a496c	gcgtctgccagagctgctaacggacccgt acgggtccgttagcagctctggcagacgc
F80A[a]	t187g, t188c	tttccaacttttcagctgcccccgggagcaggtcc ggacctgctcccgggggcagctgaaaagttggaaa
F80H[a]	t187g	gatttccaacttttcaactgcccccgggagcagg cctgctcccgggggcagttgaaaagttggaaatc
F80V[a]	t187c, t188a	tttccaacttttcatgtgcccccgggagcaggtcc ggacctgctcccgggggcacatgaaaagttggaaa

[a] These mutants were produced during this project but were not examined.

2.1.2 Δ29*Tf*NCS

The plasmid containing the gene for C-terminal hexahistidine tagged Δ29*Tf*NCS was purchased from DNA2.0 (Menlo Park, CA, USA) (Fig. 2.2). The nucleotide sequences were designed using DNA2.0 codon optimisation algorithms for optimum expression in *E. coli*.[112] The synthesised genes were provided cloned into a pJ411 expression vector.

GTTTAACTTTTAGGAGGTAAAA*CAT**ATG***TTGCATCACCAGGGTATCATCAATCAAG
TTAGCACCGTCACGAAAGTAATTCATCACGAGCTGGAAGTTGCGGCATCCGCTGA
CGACATTTGGACCGTGTACAGCTGGCCGGGTCTGGCGAAGCACTTGCCGGATCT
GCTGCCTGGCGCGTTCGAAAAACTGGAGATTATCGGCGATGGCGGTGTTGGTAC
GATTCTGGACATGACCTTTGTCCCGGGTGAATTCCCGCACGAGTATAAAGAGAAA
TTCATCCTGGTTGATAACGAACATCGTCTGAAGAAGGTGCAGATGATCGAAGGCG
GCTATCTGGACCTGGGTGTGACGTATTACATGGACACGATTCACGTTGTGCCGAC
CGGTAAAGACAGCTGCGTCATCAAGAGCAGCACTGAGTACCACGTCAAGCCGGA
GTTTGTGAAGATTGTTGAGCCGCTGATCACCACCGGTCCACTGGCAGCCATGGCA
GATGCCATTAGCAAGTTGGTCCTGGAACATAAATCTAAAAGCAACTCCGATGAAAT
TGAGGCGGCGATCATCACCGTGCTGGAGCATCACCACCACCATCAC**TGATAA***AA
GCTT*CCCC

Fig. 2.2 Δ29TfNCS, pQR1046: DNA sense strand 5`-3`. Bold text shows start and stop codons. Italics shows restriction endonuclease recognition sites NdeI and HindIII . Underlined at the 5' is the Shine-Dalgarno sequence and at the 3'/C-terminus the His-tag. Sequence codon optimised for expression in *E. coli* by DNA2.0.

A number of Δ29*Tf*NCS mutants were purchased from DNA2.0 and provided in the same format. These were the variants: L76A, L76V, A79F, A79I, F80L, Y108F, E110D, E110Q, D141E and D141N. All other Δ29*Tf*NCS mutants were constructed using QuikChange® Lightening mutagenesis kit (Agilent, Santa Clara, CA, USA) and mutagenesis primers (purchased from Eurofins, Luxembourg) (Table 2.2). Standard QuikChange® Lightening recommended parameters and conditions were used for primer design and experimental conditions unless noted. All plasmid sequences were verified by Sanger sequencing provided by Source BioScience.

Table 2.2 Primer sequences for Δ29*Tf*NCS mutations. Nucleotide sequence numbered from AUG start codon.

Amino acid variant	DNA mutation	Primer sequences (5` to 3`)
M97FLIV[a,b]	a205n, g207c	attctggacntcacctttgtcccgggtgaattcccgcac ctaccgccacaaccatgctaagacctgnagtggaaacag
F112L[a]	t250c, c252g	aagagaaactgatcctggttgataacgaacatcgtctgaagaagg gggcccacttaagggcgtgctcatatttctctttgactaggacca
F112K	t250a, t251a, c252g	gaattcccgcacgagtataaagagaaaaagatcctggttgataacgaacatc gatgttcgttatcaaccaggatctttttctctttatactcgtgcgggaattc
K122F	a283t, a284t, g285t	gccttcgatcatctgcaccttaaacagacgatgttcgttatcaac gttgataacgaacatcgtctgtttaaggtgcagatgatcgaaggc
K122L	a283c, a283c, a284t	cttcgatcatctgcaccttcagcagacgatgttcgttatcaa ttgataacgaacatcgtctgctgaaggtgcagatgatcgaag
A69K	g121a, c122a	ccggcaagtgcttcttcagacccggccagc gctggccgggtctgaagaagcacttgccgg
L72K	t130a, t131a	gtctggcgaagcacaagccggatctgctgc gcagcagatccggcttgtgcttcgccagac
I85K	t170a, t171g	aacaccgccatcgccgatcttctccagtttttcgaacg cgttcgaaaaactggagaagatcggcgatggcggtgtt
L95K	c199a, t200a	aaggtcatgtccttaatcgtaccaacaccgccatcgc gcgatggcggtgttggtacgattaaggacatgacctt
M126K (K122L)[c]	t293a	agccgccttcgatcttctgcaccttcagc gctgaaggtgcagaagatcgaaggcggct
M183K	t464a	cggtccactggcagccaaggcagatgc gcatctgccttggctgccagtggaccg

[a] These primers were designed with 3` overhangs to minimise primer-primer annealing.
[b] The degenerate codon NTC was used, encoding for Phe, Leu, Ile and Val.
[c] This primer pair is specific for the K122L variant.

2.1.3 *Cj*NCS2

A plasmid containing the C-terminal hexahistidine tagged gene of *Cj*NCS2 was provided by Dr Thomas Pesnot. Originally, synthesised genes were inserted into the pET29a(+) vector by restriction and subsequent ligation using the restriction enzymes NdeI and HindIII (Fig. 2.3).*

*CAT**ATG***CGTATGGAAGTGGTGTTGGTGGTGTTCCTGATGTTCATCGGCACAATCA
ACTGTGAACGCCTGATCTTCAACGGCCGGCCCTTGTTGCATCGGGTAACCAAAGA
AGAAACCGTAATGCTGTATCATGAATTGGAAGTAGCGGCGAGCGCTGATGAAGTT
TGGAGCGTAGAAGGCTCACCGGAACTGGGGCTGCATCTGCCGGACCTGTTGCCG
GCGGGCATCTTTGCGAAATTTGAAATTACCGGCGATGGCGGCGAAGGCTCAATCT
TGGACATGACCTTCCCTCCGGGCCAATTTCCGCATCATTACCGCGAAAAATTCGTT
TTCTTCGATCATAAAAACCGGTACAAATTGGTAGAACAAATCGATGGCGATTTTTT
CGATTTGGGCGTGACCTACTATATGGATACCATCCGGGTGGTGGCTACCGGGCC
GGATAGCTGTGTAATCAAAAGCACCACCGAATACCATGTTAAACCGGAATTTGCTA
AAATCGTAAAACCGCTGATTGACACCGTGCCGTTGGCGATCATGAGCGAAGCTAT
TGCGAAAGTGGTGTTGGAAAATAAACATAAAAGCAGCGAACATCATCATCATCATC
AT**TAATAG***AAGCTT*

Fig. 2.3 *Cj*NCS2, pQR1025. DNA sense strand 5`-3`. Bold text shows start and stop codons. Italics shows restriction endonuclease recognition sites NdeI and HindIII used for cloning . Underlined at the 3'/C-terminus the His-tag.*****

The plasmid containing the gene for C-terminal hexahistidine tagged Δ21*Cj*NCS2 was purchased from DNA2.0 (Menlo Park, CA, USA) (Fig. 2.4). The nucleotide sequences were designed using DNA2.0 codon optimisation algorithms for optimum expression in *E. coli.*[112] The synthesised genes were provided cloned into a pJ411 expression vector.

ATGCTGATTTTCAATGGTCGTCCGCTGCTGCACCGTGTTACGAAAGAAGAAACCG
TTATGCTGTACCATGAGTTGGAAGTGGCGGCGAGCGCCGACGAAGTCTGGAGCG
TCGAGGGTAGCCCGGAGCTGGGCCTGCATCTGCCGGATCTGCTGCCGGCTGGTA
TTTTCGCAAAATTCGAGATCACGGGCGATGGCGGCGAAGGTAGCATTTTGGATAT
GACCTTTCCGCCAGGTCAATTTCCGCACCATTATCGTGAGAAGTTTGTGTTCTTTG
ATCACAAGAACCGCTACAAACTGGTGGAGCAGATCGATGGTGACTTCTTCGACCT
GGGTGTGACCTACTATATGGACACGATCCGCGTCGTGGCGACTGGCCCTGACAG
CTGCGTTATCAAGTCTACCACCGAATATCACGTTAAACCGGAGTTTGCGAAGATTG
TCAAGCCGCTGATTGACACCGTGCCGCTGGCCATCATGTCCGAGGCGATCGCAA
AAGTTGTTCTGGAAAACAAACACAAGAGCAGCGAGCATCACCATCACCACCAC**TA**
ACTCGAG

Fig. 2.4 Δ21*Cj*NCS2. DNA sense strand 5`-3`. Bold text shows start and stop codons. Underlined at the 3'/C-terminus the His-tag. Sequence codon optimised for expression in *E. coli* by DNA2.0.

* The cloning of *Cj*NCS2 was conducted by Dr Markus Gershater and Dr Thomas Pesnot

2.1.4 NCS plasmid list

Table 2.3 Complete list of NCS plasmids generated in this study. All these plasmids are stored both as the pure plasmid DNA and as BL21(DE3) glycerol stocks.

pQR	Vector	NCS gene	Mutation(s)	His Tag	Source	Method
1024	pET29a(+)[a]	Δ19*Tf*NCS	WT	C-terminal	M. Gershater	-
1721	pET29a(+)	Δ19*Tf*NCS	L76A	C-terminal	This study	SDM[f]
1722	pET29a(+)	Δ19*Tf*NCS	A79F	C-terminal	This study	SDM
1723	pET29a(+)	Δ19*Tf*NCS	A79I	C-terminal	This study	SDM
1724	pET29a(+)	Δ19*Tf*NCS	F80L	C-terminal	This study	SDM
1725	pET29a(+)	Δ19*Tf*NCS	M97F	C-terminal	This study	SDM
1726	pET29a(+)	Δ19*Tf*NCS	M97L	C-terminal	This study	SDM
1727	pET29a(+)	Δ19*Tf*NCS	F99L	C-terminal	This study	SDM
1728	pET29a(+)	Δ19*Tf*NCS	Y108F	C-terminal	This study	SDM
1729	pET29a(+)	Δ19*Tf*NCS	E110D	C-terminal	This study	SDM
1730	pET29a(+)	Δ19*Tf*NCS	E110Q	C-terminal	This study	SDM
1731	pET29a(+)	Δ19*Tf*NCS	D141E	C-terminal	This study	SDM
1732	pET29a(+)	Δ19*Tf*NCS	D141N	C-terminal	This study	SDM
1733	pET29a(+)	Δ19*Tf*NCS	A182F	C-terminal	This study	SDM
1734	pET29a(+)	Δ19*Tf*NCS	A182I	C-terminal	This study	SDM
1735	pET29a(+)	Δ19*Tf*NCS	M183F	C-terminal	This study	SDM
1736	pET29a(+)	Δ19*Tf*NCS	M183L	C-terminal	This study	SDM
1737	pET29a(+)	Δ19*Tf*NCS	F80A	C-terminal	This study	SDM
1738	pET29a(+)	Δ19*Tf*NCS	F80H	C-terminal	This study	SDM
1739	pET29a(+)	Δ19*Tf*NCS	F80V	C-terminal	This study	SDM
1046	pJ411[b]	Δ29*Tf*NCS	WT	C-terminal	This study	GS[g]
1831	pJ411	Δ29*Tf*NCS	L76A	C-terminal	This study	GS
1045	pJ411	Δ29*Tf*NCS	L76V	C-terminal	This study	GS
1832	pJ411	Δ29*Tf*NCS	A79F	C-terminal	This study	GS
1047	pJ411	Δ29*Tf*NCS	A79I	C-terminal	This study	GS
1833	pJ411	Δ29*Tf*NCS	F80L	C-terminal	This study	GS
1834	pJ411	Δ29*Tf*NCS	Y108F	C-terminal	This study	GS
1835	pJ411	Δ29*Tf*NCS	E110D	C-terminal	This study	GS
1836	pJ411	Δ29*Tf*NCS	E110Q	C-terminal	This study	GS
1837	pJ411	Δ29*Tf*NCS	D141E	C-terminal	This study	GS
1838	pJ411	Δ29*Tf*NCS	D141N	C-terminal	This study	GS
1839	pJ411	Δ29*Tf*NCS	M97F	C-terminal	This study	SDM
1840	pJ411	Δ29*Tf*NCS	M97L	C-terminal	This study	SDM
1841	pJ411	Δ29*Tf*NCS	M97V	C-terminal	This study	SDM
1842	pJ411	Δ29*Tf*NCS	F112L	C-terminal	This study	SDM
1843	pJ411	Δ29*Tf*NCS	F112K	C-terminal	This study	SDM
1844	pJ411	Δ29*Tf*NCS	K122L	C-terminal	This study	SDM
1845	pJ411	Δ29*Tf*NCS	K122F	C-terminal	This study	SDM
1846	pJ411	Δ29*Tf*NCS	K122L A69K	C-terminal	This study	SDM
1847	pJ411	Δ29*Tf*NCS	K122L L72K	C-terminal	This study	SDM
1848	pJ411	Δ29*Tf*NCS	K122L I85K	C-terminal	This study	SDM
1849	pJ411	Δ29*Tf*NCS	122L F112K	C-terminal	This study	SDM
1850	pJ411	Δ29*Tf*NCS	K122L L95K	C-terminal	This study	SDM
1851	pJ411	Δ29*Tf*NCS	122L M126K	C-terminal	This study	SDM
1852	pJ411	Δ29*Tf*NCS	122L M183K	C-terminal	This study	SDM
1853	pJ411	Δ29*Tf*NCS	122F F112K	C-terminal	This study	SDM
1025	pET29a(+)[a]	*Cj*NCS2	WT	C-terminal	T. Pesnot[e]	-
1854	pJ411	Δ21*Cj*NCS	WT	C-terminal	This study	GS
1855	pNic28-BSa4	Δ29*Tf*NCS	WT	N-term-TeV[d]	This study	LIC[h]
1856	pD451-SR[b]	ΔN33C196 *Tf*NCS[c]	WT	N-term-TeV[d]	This study	GS

[a] Gene inserted between NdeI and HindIII sites by Dr Markus Gershater. Genes are codon optimised for *E. coli*. [b] pJ411 and pD451-SR vectors are from DNA2.0. Both have: kanamycin resistance, high copy pUC origins, T7 IPTG inducible promoters and strong RBSs. [c] *Tf*NCS sequence from 1-33 and from 196-210 removed. [d] N-terminal HisTag with linker and TeV protease cleavage site: MHHHHHHSSGVDLGT**ENLYFQ^S.** [e] See Pesnot *et al*. [f] Site-directed mutagenesis (SDM). [g] Gene synthesis (GS) from DNA2.0, with *E. coli* codon optimisation. [h] Ligation independent cloning, performed by Dr Altin Sula

```
TfNCS              MMKME-VVFVFLMLLGTINCQKLILTGRPFLHHQGIINQVSTVTKVIHHELEVAASADDI   59
Δ19TfNCS           ------------------MQKLILTGRPFLHHQGIINQVSTVTKVIHHELEVAASADDI   41
Δ29TfNCS           ----------------------------MLHHQGIINQVSTVTKVIHHELEVAASADDI   31
ΔN33C196TfNCS      --------------------------------MGIINQVSTVTKVIHHELEVAASADDI   27
CjNCS2             -MRMEVVLVVFLMFIGTINCERLIFNGRPLLHRVTK-----EETVMLYHELEVAASADEV   54
Δ21CjNCS2          --------------------MLIFNGRPLLHRVTK-----EETVMLYHELEVAASADEV   34
                                                         * :::**********::

TfNCS              WTVYSWPGLAKHLPDLLP-GAFEKLEIIGDGGVGTILDMTFVPGEFPHEYKEKFILVDNE  118
Δ19TfNCS           WTVYSWPGLAKHLPDLLP-GAFEKLEIIGDGGVGTILDMTFVPGEFPHEYKEKFILVDNE  100
Δ29TfNCS           WTVYSWPGLAKHLPDLLP-GAFEKLEIIGDGGVGTILDMTFVPGEFPHEYKEKFILVDNE  90
ΔN33C196TfNCS      WTVYSWPGLAKHLPDLLP-GAFEKLEIIGDGGVGTILDMTFVPGEFPHEYKEKFILVDNE  86
CjNCS2             WSVEGSPELGLHLPDLLPAGIFAKFEITGDGGEGSILDMTFPPGQFPHHYREKFVFFDHK  114
Δ21CjNCS2          WSVEGSPELGLHLPDLLPAGIFAKFEITGDGGEGSILDMTFPPGQFPHHYREKFVFFDHK  94
                   *:* . * *. ******* * * *:** **** *:****** **:***.*:***::.*.:

TfNCS              HRLKKVQMIEGGYLDLGVTYYMDTIHVVPTGKDSCVIKSSTEYHVKPEFVKIVEPLITTG  178
Δ19TfNCS           HRLKKVQMIEGGYLDLGVTYYMDTIHVVPTGKDSCVIKSSTEYHVKPEFVKIVEPLITTG  160
Δ29TfNCS           HRLKKVQMIEGGYLDLGVTYYMDTIHVVPTGKDSCVIKSSTEYHVKPEFVKIVEPLITTG  150
ΔN33C196TfNCS      HRLKKVQMIEGGYLDLGVTYYMDTIHVVPTGKDSCVIKSSTEYHVKPEFVKIVEPLITTG  146
CjNCS2             NRYKLVEQIDGDFFDLGVTYYMDTIRVVATGPDSCVIKSTTEYHVKPEFAKIVKPLIDTV  174
Δ21CjNCS2          NRYKLVEQIDGDFFDLGVTYYMDTIRVVATGPDSCVIKSTTEYHVKPEFAKIVKPLIDTV  154
                   .* * *: *:* ::***********:** ** *******:*********.***:*** *

TfNCS              PLAAMADAISKLVLEHKSKSNSDEIEAAIITV------  210
Δ19TfNCS           PLAAMADAISKLVLEHKSKSNSDEIEAAIITVHHHHHH  198
Δ29TfNCS           PLAAMADAISKLVLEHKSKSNSDEIEAAIITVHHHHHH  188
ΔN33C196TfNCS      PLAAMADAISKLVLEHKS------------------  164
CjNCS2             PLAIMSEAIAKVVLENKHKSSE----------HHHHHH  202
Δ21CjNCS2          PLAIMSEAIAKVVLENKHKSSE----------HHHHHH  182
```

Fig. 2.5 Multiple sequence alignments of NCSs used in this study. Allignment constructed with Clustal omega.[113] Symbols: "." = conservation between groups of weakly similar properties, ":" = conservation between groups of strongly similar properties, and "*" = fully conserved residue. Important conserved residues emboldened: Tyr-108, Glu-110, Lys-122 and Asp-141 (numbered by *Tf*NCS from N-terminus). *Tf*NCS sequence from Saminami *et al*[34]. Δ19*Tf*NCS: pQR1024, Δ29*Tf*NCS: pQR1046, ΔN33C196*Tf*NCS: pQR1855, CjNCS2: pQR1025 and Δ21*Cj*NCS2: pQR1854 (see Table 2.3).

2.1.5 Transaminases

Transaminase (TAm) genes were amplified by PCR from genomic DNA and cloned into pET29a(+) plasmids by previous members of the Ward group.* Ten TAms from the Ward group TAm library and expressed from BL21 (DE3) cells.

Table 2.4 Transaminases used in this study. All enzymes obtained from the Ward group TAm library.

TAm	Organism	Uniprot entry name	Gene name	First reported
BSU_09260	*Bacillus subtilis*	YHXA_BACSU	yhxA	This thesis
CV_2025	*Chromobacterium violaceum*	Q7NWG4_CHRVO	CV_2025	Kaulmann *et al*[114]
Dgeo_1416	*Deinococcus geothermalis*	Q1IYH3_DEIGD	argD/lysJ	This thesis
KPN_00255	*Klebsiella pneumoniae*	A6T537_KLEP7	gabT	This thesis
PP_0596	*Pseudomonas putida*	Q88Q98_PSEPK	PP_0596	This thesis
PP_3718	*Pseudomonas putida*	Q88GK3_PSEPK	PP_3718	This thesis
SaV_2612	*Streptomyces avermitilis*	Q82JZ2_STRAW	SAV_2612	This thesis
SaV_4551	*Streptomyces avermitilis*	Q82ER2_STRAW	SaV_4551	This thesis
VF_JS17	*Vibrio fluvialis*	F2XBU9_VIBFL	JS17	Shin *et al*[115]

```
SAV_4551        MTPQPNPQVG-----AAVKAADRAHVFH---SWSAQELIDPLAVAGAEGSYFWDYDGRRY   52
PP0596          ---MNMPETGPA---GIASQLKLDAHWMPYTANRNF-QRDPRLIVAAEGNYLVDDHGRKI   53
CV_2025         MQKQRTT--------SQWRELDAAHHLHPFTDTASLNQAGARVMTRGEGVYLWDSEGNKI   52
VF_JS17         ---MNKP--------QSWEARAETYSLYGFTDMPSLHQRGTVVVTHGEGPYIVDVNGRRY   49
BSU09260        MEMMGMEN---IQQNQGLKQKDEQFVWH---AMKGAHQADSLIAQKAEGAWVTDTDGRRY   54
PP3718          ---MATPSKAFAIAHDPLVEADKAHYMHGYHVFDEHREQGALNIVAGEGAYIRDTHGNRF   57
SAV_2612        ---MGNP---IAVSKDL-SRTAYDHLWMHFTRMSSYENAPVPTIVRGEGTYIYDDKGKRY   53
Dgeo_1416       ---MTGTKTK----ASKWLDAE-------LRYDSGVYNKHQVVMVRGQGATVWDETGRAY   46
KPN_00255       ---MNSNKAM----MA---RRS-------DAVPRGVGQIHPIFAERAENCRVWDVEGREY   43
                                                          .:   . *  *.

SAV_4551        LDFTSGLVFTNIGYQHPKVVAAIQEQAASLTTFAPAF--AVEARSEAARLIAERTPGD-L   109
PP0596          FDALSGLWTCGAGHTRKEIADAVTRQLSTLDYSPAF-QFGHPLSFQLAEKIAELVPGN-L   111
CV_2025         IDGMAGLWCVNVGYGRKDFAEAARRQMEELPFYNTFFKTTHPAVVELSSLLAEVTPAG-F   111
VF_JS17         LDANSGLWNMVAGFDHKGLIDAAKAQYERFPGYHAFFGRMSDQTVMLSEKLVEVSPFD-S   108
BSU09260        LDAMSGLWCVNIGYGRKELAEAAYEQLKELPYYPLTQ--SHAPAIQLAEKLNEWLGGD-Y   111
PP3718          LDAVGGMWCTNIGLGREEMALAIVDQVRQLAYSNPFSDMANDVAIELCQKLAQLAPGD-L   116
SAV_2612        LDGLSGLFVVQAGHGRTELAETAFKQAQELAFFPVW-SYAHPKAVELAERLANYAPGD-L   111
Dgeo_1416       IDCVAGYGVANIGHCHPDVVKAIQEQAARLIVMPQT--LPNDKRAEFLTELVGVLPQG-L   103
KPN_00255       LDFAGGIAVLNTGHLHPQVVAAVEDQLKKLSHTCFQ-VLAYEPYLALCEKMNQKVPGDFA   102
                :*  .*      *  :  .  :   *   :                   :

SAV_4551        DKIFFTNGGADAIEHAVRMARIH------TGRPKVLSAYRSYHGGTQQAVNITGDPRRWA   163
PP0596          NHVFYTNSGSECADTALKMVRAYWRLKGQATKTKIIGRARGYHGVNIAGTSLGGVNGNRK   171
CV_2025         DRVFYTNSGSESVDTMIRMVRRYWDVQGKPEKKTLIGRWNGYHGSTIGGASLGGMKYMHE   171
VF_JS17         GRVFYTNSGSEANDTMVKMLWFLHAAEGKPQKRKILTRWNAYHGVTAVSASMTGKPYNSV   168
BSU09260        -VIFFSNSGSEANETAFKIARQYHLQNGDHSRYKFISRYRAYHGNTLGALSATGQAQRKY   170
PP3718          NHVFLTTGGSTAVDTAYRLIQYYQNCRGKPHKKHIIARYNAYHGSTTLTMSIGNKAADRV   176
SAV_2612        NKVFFTTGGGEAVETAWKLAKQYFKLQGKPTKYKVISRAVAYHGTPQGALSITGLPALKA   171
Dgeo_1416       ERVFLCNSGTEAMEAAKKFAITA------TGRSRFVSMKRGFSGRSLGALAFTWE-----   152
KPN_00255       KKTLLVTTGSEAVENAVKIARAA------TGRSGAIAFTGAYHGRTHYTLSLTGKVNPYS   156
                   :  . *  . :   ::           :   :    .: *
```

* Cloning of previously unreported transaminases conducted by previous Ward group members, including Julio Rubéns Martínez-Torres

```
SAV_4551        --SDSASAGVVHFWAPYLYRSRFYAETEQQECER-ALEHLET-TIAFEGPGTIAAIVLET       219
PP0596          -MFGQLLDV-DHLPHTVLPVNAFSKGLPEEGGIA-LADEMLK-LIELHDASNIAAVIVEP       227
CV_2025         -QGDLPIPGMAHIEQPWWYKHGKDM-TPDEFGVV-AARWLEE-KILEIGADKVAAFVGEP       227
VF_JS17         --FGLPLPGFVHLTCPHYWRYGEEGETEEQFVAR-LARELEE-TIQREGADTIAGFFAEP       224
BSU09260        -KYEPLSQGFLHAAPPDIYRNPDDADTLES------ANEIDR-IMTWELSETIAGVIMEP       222
PP3718          PEFDYHHDLIHHVSNPNPYRAPDDMDEA-EFLDF-LVAEFED-KILSLGADNVAAFFAEP       233
SAV_2612        P-FEPLVPGAHKVPNTNIYRAPLFGDDPEAFGRW-AADQIEQ-QILFEGPETVAAVFLEP       228
Dgeo_1416       ----------------PKYREPFGEAVDNKHVDFVTYGNIDE--LRAAVTDQTAAVILEP       194
KPN_00255       -------AGMG-LMPGHVYRALYPCALHGVSDD-EAIASIHRIFKNDAAPEDIAAIIIEP       207
                                                      :            *... *

SAV_4551        VPGTAGIMVPPPGYLAGVRELCDKYGIVFVLDEVMAGFGRTGEWFAAD-LFDVTPDLMTF       278
PP0596          LAGSAGVLPPPKGYLKRLREICTQHNILLIFDEVITGFGRMGAMTGSE-AFGVTPDLMCI       286
CV_2025         IQGAGGVIVPPATYWPEIERICRKYDVLLVADEVICGFGRTGEWFGHQ-HFGFQPDLFTA       286
VF_JS17         VMGAGGVIPPAKGYFQAILPILRKYDIPVISDEVICGFGRTGNTWGCV-TYDFTPDAIIS       283
BSU09260        IITGGGILMPPDGYMKKVEDICRRHGALLICDEVICGFGRTGEPFGFM-HYGVKPDIITM       281
PP3718          IMGSGGVIIPPEGYFQRMWQLCQTYDILFVADEVVTSFGRLGTFFASEELFGVTPDIITT       293
SAV_2612        VQNAGGCFPPPPGYFQRVREICDQYDVLLVSDEVICAFGRLGTMFACD-KFGYVPDMITC       287
Dgeo_1416       VQGEGGVRPVTPEFIRAAREITREKGALLILDEIQTGFCRTGKMFAAE-HFGVVPDGMTL       253
KPN_00255       VQGEGGFYAASPAFMQRLRALCDEHGIMLIADEVQSGAGRTGTLFAME-QMGVAADITTF       266
                :   .*       :      :       .: **:  .  * *   .         *

SAV_4551        AKGVNSGYVPLGGVAISGKIAETFG-------KRAYPGGLTYSGHPLACAAAVATINVMA       331
PP0596          AKQVTNGAIPMGAVIASSEIYQTFMNQPTPEYAVEFPHGYTYSAHPVACAAGLAALDLLQ       346
CV_2025         AKGLSSGYLPIGAVFVGKRVAE----GL--IAGGDFNHGFTYSGHPVCAAVAHANVAALR       340
VF_JS17         SKNLTAGFFPMGAVILGPELSKRLETAI--EAIEEFPHGFTASGHPVGCAIALKAIDVVM       341
BSU09260        AKGITSAYLPLSATAVKRDIFEAYQ-GE--APYDRFRHVNTFGGSPAACALALKNLQIME       338
PP3718          AKGLTSAYLPLGACIFSERIWQVIA-EP--GKGRCFTHGFTYSGHPVCCTAALKNIEIIE       350
SAV_2612        AKGMTSGYSPIGACIVSDRIAEPFY-K----GDNTFLHGYTFGGHPVSAAVGVANLDLFE       342
Dgeo_1416       AKAMAGG-VPVGAFAMTAEVADRMPAG---------GHGTTFGGNPLAMAAGIAAIRAMK       303
KPN_00255       AKSIAGG-FPLAGVTGRAEVMDAIAPG---------GLGGTYAGNPIACAAALAVLQIFE       316
                :* :  .  *:..      : .                 * .. *   : .   :  .

SAV_4551        EEGVVENAANLGARVIEPGLRELAERHPSVGEVRGVGMFWALELVKDRETREPLVPYNAA       391
PP0596          KENLVQSAAEL-APHFEKLLHG-VKGTKNIVDIRNYGLAGAIQIAARDGDAIVRPY----       400
CV_2025         DEGIVQRVKDDIGPYMQKRWRETFSRFEHVDDVRGVGMVQAFTLVKNKAKRELFPDFGEI       400
VF_JS17         NEGLAENVRRL-APRFEERLKH-IAERPNIGEYRGIGFMWALEAVKDKASKTPFDGNLSV       399
BSU09260        DEQLIQRSRDL-GAKLLGELQA-LREHPAVGDVRGKGLLIGIELVKDKLTKEPADAAKVN       396
PP3718          REQLLDHVNDV-GSYLEQRLQS-LRDLPLVGDVRCMKLMACVEFVANKASKALFADEVNI       408
SAV_2612        REGLNQHVLDN-ESAFLTTLQK-LHDLPIVGDVRGNGFFYGIELVKDKATKETFTDEESE       400
Dgeo_1416       NEKMAEQAREK-GAYFMERLRAI--RSPKIREVRGLGLMIGVELKEKSAPY---------       351
KPN_00255       QENLLEKANQL-GDTLRQGLLAIAEDHPEIGDVRGLGAMIAIELFEEGDHSRPNA-----       370
                 * : :        :            : : *      .

SAV_4551        GEA----NAPMAAFGAAAKANGLWPFINMNRTHVVPPCNVTEAEAKEGLAALDAALSVAD       447
PP0596          -----EAAMKLWKAGFYVRFGGDT-------LQFGPTFNTKPQELDRLFDAVGETLNLID       448
CV_2025         GTL---CRDIFFRNNLIMRACGDH-------IVSAPPLVMTRAEVDEMLAVAERCLEEFE       450
VF_JS17         SER---IANTCTDLGLICRPLGQS-------VVLCPPFILTEAQMDEMFDKLEKALDKVF       449
BSU09260        Q-----VVAACKEKGLIIGKNGDTVAGYNNVIQLAPPFCLTEEDLSFIVKTVKESFQTI-       450
PP3718          GER---IHSKAQEKGLLVRP----I---MHLNVMSPPLIITHAQVDEIVETLRQCIIETA       458
SAV_2612        RVLYGFVSKKLFEYGLYCRA----DDRGDPVIQLSPPLISNQSTFDEIESIIRQVLTEAW       456
Dgeo_1416       ------IAALEHEEGVLTLAATP------LVVRFLPPLTISREQIDQVVAAFERVLERVN       399
KPN_00255       -RLTADIVARARDKGLILLSCGPY----YNVLRILVPLTIEEAQIEQGLKIIADCFSEAK       425
                                                        .        :
```

```
SAV_4551        EYTV------------        451
PP0596          ----------------        448
CV_2025         QTLKARGLA-------        459
VF_JS17         AEVA------------        453
BSU09260        ----------------        450
PP3718          RELTALGLYQGR----        470
SAV_2612        TKL-------------        459
Dgeo_1416       PRAIPNQELREDKQTE        415
KPN_00255       QA--------------        427
```

Fig. 2.6 Multiple sequence alignment of transaminases used in this study. Allignment constructed with Clustal omega.[113] Symbols: "." = conservation between groups of weakly similar properties, ":" = conservation between groups of strongly similar properties, and "*" = fully conserved residue.

2.1.6 *E. coli* genotypes

Top-10® and XL10 Gold cells were used interchangeably in site directed mutagenesis procedures and for producing large quantities of plasmid DNA for further use.

Top-10® (Invitrogen): F^- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ^-

XL10 Gold (Stratagene): endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB $lacI^qZ\Delta M15$ Tn10(Tet^R Amy Cm^R)]

BL21 (DE3) cells (NEB) were used for protein expression.

BL21 (DE3): F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])

2.2 Computational tools

2.2.1 Statistics and plotting

Data were plotted in Microsoft Excel, R or QtGrace. Fitting conducted in R using the nonlinear least squares function using the default Gauss-Newton algorithm or the "port" algorithm. When standard deviations were available fitting was typically conducted with weights equal to inverse variance ($1/\sigma^2$). All fitted parameters have $p \leq 0.05$ unless mentioned. For Michalis-Menten or hyperbolic curves equation used was $y = \frac{Vx}{K+x}$, with gradient equal to $\frac{VK}{(K+x)^2}$. Exponential curves were: $y = C - Ce^{-Lx}$ and with the gradient LCe^{-Lx}. Substrate inhibition/Haldane curves: $y = \frac{Vx}{K_m+x+\frac{x^2}{K_i}}$.

2.2.2 Molecular dynamics (MD)

Simulations were conducted in GROMACS using the Charmm27 forcefield.[116,117] Each subunit of 2VQ5 was prepared by removing waters and ligands. The N- and C-termini were truncated, with the protein starting at Ser-40 and ending at Lys-197. The protein subunits were placed in a cube, then water (tip3p) and counter ions were added. The systems were energy minimised (F_{max} < 250 $kJ.mol^{-1}.nm^{-1}$). This was followed by NVT equilibration (100 ps) and NPT equilibration (100 ps). Three simulations for each subunit ran for 100 ns, at 300 K. The Phe-112 conformation was determined by calculating the angle between the Cα and Cζ of Phe-112 and the Cα of His-106. A value for this angle was calculated for every tenth frame. Data from MD simulations plotted with QtGrace. For more details, see Appendix A, section 8.

2.2.3 Molecular docking

Docking was conducted according to the protocol developed by Dr Thomas Pesnot.[57] Potential reaction intermediates were energy optimised using MM2 energy minimisation (ChemBio3D, CambridgeSoft) and docked in the active site of subunit A from the *Tf*NCS crystal structure 2VQ5 (after removal of ligands)[43]. Docking was performed using AutoDock Vina (exhaustiveness = 8).[118] The lowest free energy clusters were then visualised and their structure analysed. Unless noted otherwise, only the top ranked binding modes were used in this investigation.

2.2.4 pK_a predictions

PROPKA 3.1 (http://propka.ki.ku.dk/) was used to predict pK_as of residues in the enzyme active site.[119] The *Tf*NCS enzyme structure model was obtained from crystal structure 2VQ5 subunit A.[43] Protein-ligand structures derived from docking calculations were also analysed in this manner.

2.2.5 Structural depiction and analysis

The Δ19*Tf*NCS X-ray crystal structures 2VNE and 2VQ5 were used.[43] Depiction was achieved with Chimera (UCSF). Depiction of small molecules was achieved with ChemBioDraw (CambridgeSoft). The Δ19*Tf*NCS sequence was submitted into Betascan (http://groups.csail.mit.edu/cb/betascan/) and analysed.[120] The online tool scored sequences for pairwise β-strand formation and single strand formation. Residue conformations and Ramachandran angles of X-ray crystal structures were analysed using MOLPROBITY.[121] Alternative conformations of side chains and ligands were explored using Coot and resulting density maps were refined with PHENIX.[122]

2.3 Protein preparation

Analysis of proteins was achieved by using SDS-PAGE (BioRad), typically with 12% polyacrylamide gels unless noted otherwise. Gels were stained using Imperial Protein Stain (ThermoScientific) or Instant blue (Expedeon). Protein purity and running mass on SDS-PAGE were measured with AlphaView software (ProteinSimple).

2.3.1 NCS expression

Plasmids containing NCS genes were transformed into *E. coli* BL21(DE3) cells by a standard heat-shock protocol. The transformed cells were stored as glycerol stocks at -80 °C. An aliquot from a frozen glycerol stock of plasmid containing *E. coli* BL21(DE3) was inoculated into 20 mL of terrific broth (TB) medium (containing 50 $\mu g.mL^{-1}$ kanamycin) (Miller, Merck Millipore). Starter cultures were incubated 16 h at 37 °C, shaking at 250 revolutions per minute (rpm). This overnight culture was transferred into fresh media (4% $v.v^{-1}$, TB with 50 $\mu g.mL^{-1}$ kanamycin). These expression cultures were incubated for 2 hours at 37 °C, followed by 1 hour at 25 °C, whilst shaking at 250 rpm. Expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG, final concentration 500 μM). Cultures were incubated for a further 3 hours at 25 °C prior to harvesting, whilst shaking at 250 rpm. Cells were harvested by centrifugation at 10,000 × g for 10 mins at 4 °C. Supernatant was removed and the cell pellets were stored at -20 °C until purification.

2.3.2 NCS purification

Optimum purification conditions for NCS proteins were explored using an automated ÄKTA purifier system and 5 mL HisTrap columns (GE Healthcare, Buckinghamshire, UK). Once the correct purification conditions were found, purification was performed on the bench-top in the process described below.

Cell pellets were thawed and suspended in BugBuster (10% culture volume) (Merck Millipore, Darmstadt, Germany). If necessary, the cells were homogenised by sonication. The insoluble portion of the lysate was pelleted by centrifugation at 10,000 × g for 30 minutes at 4 °C. The supernatant was removed and filtered through a glass fibre prefilter and 0.2 µm cellulose acetate syringe filter (Sartorius Stedim Biotech, Göttingen, Germany). An empty and clean PD-10 column (GE) was charged with Ni-Sepharose HP resin (2 mL, GE). The column was washed with distilled water (10 mL), followed by binding buffer (0.1 M HEPES, 20 mM imidazole, 100 mM NaCl, pH 7.5; 10 mL). The filtered supernatant was passed through the Ni-Sepharose column, and the column was then washed with binding buffer (10 mL) followed by wash buffer (0.1 M HEPES, 40 mM imidazole, 100 mM NaCl, pH 7.5; 20 mL) or until no protein could be detected coming off the column by a standard Bradford's assay. The bound protein was then eluted with elution buffer (0.1 M HEPES, 500 mM imidazole, 100 mM NaCl, pH 7.5; 5 mL). The eluant containing pure enzyme was buffer exchanged into assay buffer (50 or 100 mM HEPES pH 7.5), using a PD-10 column.

Protein purity was established by SDS-PAGE (12% $w.v^{-1}$ polyacrylamide). If sufficiently pure, protein concentration was determined using absorbance at 280 nm on a NanoDrop® 2000C (with molar absorbance calculated from Expasy ProtParam[123]). For storage, glycerol was added (10% $v.v^{-1}$), the protein was divided into small samples (50-500 µL) and then flash-frozen in liquid nitrogen. The purified protein was stored at -80 °C. For Δ29*Tf*NCS, the yield of purified protein from culture was approximately 100 $mg.L^{-1}$.

2.3.3 Desalted lysate preparation

Desalted lysate was prepared for biotransformation reactions by the following protocol: Cell pellets were thawed and suspended in BugBuster (10% culture volume). If necessary, lysate was sonicated until mixture became homogenous. The insoluble portion of the lysate was pelleted by centrifugation at 10,000 × g for 45 mins at 4 °C. Lysate was buffer exchanged into assay buffer 0.1 M HEPES pH 7.5 using PD-10 columns (GE). The HEPES-lysate elution was used immediately in biotransformations.

2.3.4 Transaminases

Transaminase expression: The Ward group have a library of TAm enzymes all expressed from the pET29a(+) vector in BL21(DE3) and stored as glycerol stocks. Selected enzymes from this library were plated out on LB-agar (Miller, Merck Millipore) plates with kanamycin (100 $µg.mL^{-1}$). Single colonies from these plates were used to inoculate 5 mL 2xYT media (Sigma) and these were grown for 16 hours at 37 °C with

250 rpm orbital shaking. 300 µL of these overnight cultures were used to inoculate 75 mL of 2xYT media. These were incubated (37 °C, 250 rpm) until OD_{600} = 0.8. Cultures were then induced with IPTG (500 µM) and incubated at 30 °C for 8 hours. The cells were then harvested by centrifugation (10,000 × g, 10 mins) and the cell pellets stored at -80 °C.

Transaminase lysate preparation: TAm cell pellets were resuspended in lysis buffer (10% culture volume, 0.1 M potassium phosphate, 1 mM PLP, pH 7.5) and homogenised by sonication (10 s intervals, 5 cycles). The insoluble portion of the lysate was pelleted by centrifugation (10,000 × g, 30 mins). The supernatant was flash frozen in liquid nitrogen and stored at -80 °C.

CV2025 preparation for one-pot two-enzyme cascade: The transaminase from *Chromobacterium violaceum*, CV2025, was expressed in 2xYT using the general TAm method described above. The CV2025 pellets were resuspended in lysis buffer (50 mM HEPES, 10 mM PLP, pH 7.5) and homogenised by sonication (10 s intervals, 5 cycles). The insoluble portion of the lysate was pelleted by centrifugation (10,000 × g, 30 mins). The supernatant was flash frozen in liquid nitrogen and stored at -80 °C.

2.4 Analytical methods

2.4.1 HPLC

Analytical HPLC: Methods were performed with a HPLC system consisting of a LC Packing FAMOS Autosampler, a Dionex P680 HPLC Pump, a Dionex TCC-100 Column oven and a Dionex UVD170U Ultraviolet detector.

Method 1, achiral: Analytic reverse phase analysis method was used for achiral quantitative analyses. Separation was achieved with a HiChrom ACE C18-5 (150 × 4.6 mm) column and a 1 $mL.min^{-1}$ gradient of H_2O (0.1% TFA) and acetonitrile 90:10 to 30:70 over 5 minutes at 30 °C (Fig. 2.7). Injection volumes were 20 µL. Product was detected via UV absorbance at 280 nm. Product retention times and concentrations were determined by chemically verified standards (see section 2.8). All non-chiral analytical HPLC analysis used this method.

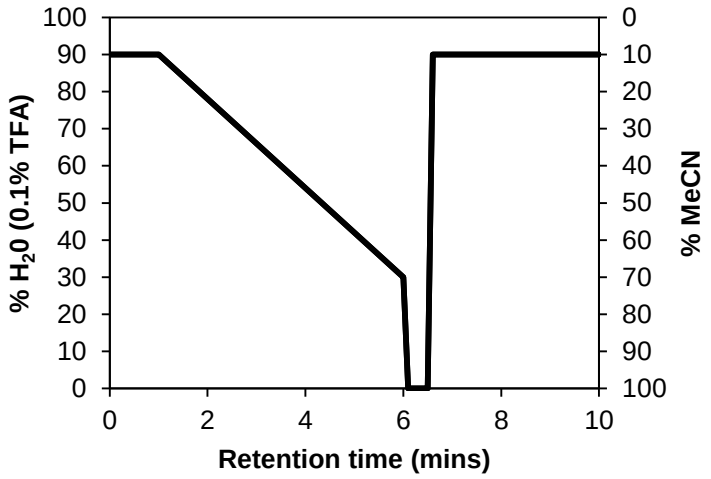


Fig. 2.7 Analytical HPLC method.

Method 2, chiral: Chiral separation was achieved with a Supelco Astec Chirobiotic T column and an isocratic mobile phase 20 mM NH_4OAc pH 4:MeOH (70:30) mobile phase at 0.2 $mL.min^{-1}$ and 30 °C. Injection volumes were 5 µL. Compounds were detected by UV absorbance at 230 nm. Product retention times and concentrations were determined by chemically verified standards.

Method 3, chiral: Chiral separation was achieved with a Supelco Astec Chirobiotic T2 column and an isocratic MeOH (0.1% TFA, 0.2% TEA) mobile phase at 1 $mL.min^{-1}$ and 30 °C. Injection volumes were 5 µL. Compounds were detected by UV absorbance at 230 nm. Product retention times and concentrations were determined by chemically verified standards.

Prep HPLC: Preparative HPLC was performed on a Varian Prostar instrument equipped with a UV-visible detector. Elutions were monitored at 280 nm and collected manually.

Method 4: DiscoveryBIO wide Pore C18-10 column (25 × 2.12 cm), gradient: 5% to 40% of acetonitrile/water (0.1% TFA).

Method 5: Ascentis C18 150 x 21.2 mm, 5 µm. Gradient 5-20% of acetonitrile/water (0.1% TFA).

2.4.2 Chemical characterisation

^{1}H and ^{13}C NMR spectra were recorded at 298 K at the field indicated using Bruker Avance 500 and Avance 600 machines. Coupling constants were measured in Hertz (Hz) and referenced to the deuterated solvent used. High resolution mass spectra (electrospray ionisation) performed with Waters Autosampler Manager 2777C connected to Waters LCT Premier XE, with positive ion detection, or a Q-TOF Aqilent 6510.

2.4.3 Plate readers

Purpald® assays were read on a Tecan Safire^{2} plate reader. Absorbance spectra were recorded between 350 and 600 nm, with a step of 1 nm. Three reads were taken per well, each 1 ms apart. Typically absorbance at 522 nm was used for the aldehyde-Purpald® complex.

Fluorescamine assays were recorded on a BMG LABTECH CLARIOstar platereader. The gain was adjusted to highest well. Excitation wavelength: 390 nm with 15 nm bandwidth. Emission wavelenfth: 480 nm with 20 nm bandwidth.

2.4.4 Circular dichroism (CD)

CD spectroscopy performed on an AVIV 400, using Hellma QS quartz cuvettes. CD signal of standards were recorded by measuring 1 minute time-courses and averaging signal. All samples and reactions maintained at 37 °C. Dynode voltage was maintained below 500 V.

2.5 Enzyme assays

This refers to the methods used in section 4.

2.5.1 General

Reactions were typically conducted using 20 or 80 $\mu g.mL^{-1}$ of purified enzyme in 50 mM HEPES, with varying concentrations of dopamine and aldehydes, in a total volume of 100 µL. Reaction components were equilibrated at 37 °C and the reaction was started by the addition of enzyme to a substrate mixture. Typically the aldehyde stock solutions were made using acetonitrile (MeCN), as the hydrophobic aldehydes used were insoluble in water. Stock solutions were generally made at 10x concentration, thus all reactions contained 10% $v.v^{-1}$ MeCN, unless noted. The reaction was incubated at 37 °C and terminated by the addition of 20 µL of 1 M HCl. Reactions were performed in triplicate; all errors are standard deviations. The background reactions were recorded

by performing the reaction in HEPES buffer without enzyme and this was subtracted from the enzymatic reactions.

2.5.2 Conversions

Reactions were conducted with 80 µg.mL^{-1} enzyme, 10 mM dopamine and 10 mM aldehyde. Reactions were conducted with 50 mM HEPES pH 7.5 and 10% v.v^{-1} MeCN unless noted otherwise. Reactions were performed for 1 hour, and were analysed by analytical HPLC method 1 and chiral HPLC method 3.

2.5.3 Initial rates

The method described here is used in the main section on the enzyme mechanism (section 4.3). Reactions were performed with an enzyme concentration of 20 µg.mL^{-1}.Reactions were conducted with 50 mM HEPES pH 7.5 and 10% v.v^{-1} MeCN unless noted otherwise. Initial rates were determined by measuring the conversion after 30 seconds, within the linear phase of the reaction. Initial rates were determined with 10 mM of one substrate and 2.5 mM of the other. Reactions were analysed by analytical HPLC method 1.

2.5.4 Kinetics

To determine kinetic parameters, initial rates (see section 2.5.3) were measured with varying substrate concentrations. Dopamine concentrations were varied between 500 µM and 20 mM in the presence of 2.5 mM aldehyde. Aldehyde concentrations were varied between 250 µM and 15 mM in the presence of 2.5 mM dopamine. The concentrations of substrates that were not varied during the assays were kept below saturating levels (typically 2.5 mM) to limit the non-enzymatic reaction. Similarly, higher concentrations of the varying substrates were not used as the increased background reactions led to increased errors. The kinetic data were calculated by fitting initial rates to a Michaelis-Menten or Substrate Inhibition equation using the non-linear least-squares analysis in the R software environment (see section 2.2.1).

2.5.5 Melting temperatures

Enzyme melting and aggregation temperatures were determined using an Avacta Optim 1000 (Avacta Analytical Ltd., York, UK). Runs were performed in triplicate. T_m is calculated by observing changes in the intrinsic protein fluorescence spectrum; T_{agg} is calculated by changes in light scattering intensity.

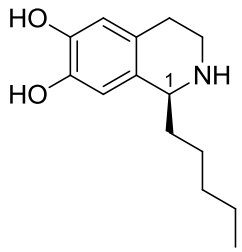
2.6 Biotransformations

2.6.1 Demonstrative biotransformations

Biotransformations (see section 5) could be performed with a variety of substrates and enzyme preparations. Two demonstrative examples appear below. HPLC retention times and results are shown in a table below (Table 2.5). Compounds were identified by NMR and mass spectrometry.

Hexanal and purified L76V- Δ29*Tf*NCS: Substrate solutions of dopamine-HCl (38 mg, 0.20 mmol) in water (2 mL) and hexanal (30 mg, 0.30 mmol) in MeCN (2 mL) were prepared. Purified L76V- Δ29*Tf*NCS was prepared as described above (section 2.3.2). The reaction mixtures were prepared by the addition of the dopamine solution (2 mL, 100 mM, 0.20 mmol), the hexanal solution (2 mL, 150 mM, 0.30 mmol) and purified enzyme (1.0 $mg.mL^{-1}$, 2 mL) to water (14 mL). The final reaction mixtures (20 mL) contained 10 mM dopamine, 15 mM hexanal and approximately 0.1 $mg.mL^{-1}$ purified enzyme. The reaction was incubated at 37 °C for 3 hours before being quenched by the addition of 1M HCl (2 mL). The mixture was centrifuged (10,000 × g for 30 mins) to remove protein precipitate and then filtered before being concentrated under reduced pressure. Samples were taken at this point and analysed by HPLC methods 1 and 3 to determine *ee* and conversion yields. The product was purified via preparative-HPLC using method 5. Solvent was removed under reduced pressure and the product washed with MeOH to remove excess TFA, yielding the previously unreported **(*S*)-1-pentyl-1,2,3,4-tetrahydro-isoquinoline-6,7-diol** as the TFA salt (40 mg, 57%, >98% *ee*)[6].

(*S*)-1-Pentyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol (TFA salt)[6]



^{1}H NMR (500 MHz; CD_3OD) δ = 0.93 (3H, t, *J* = 7.1 Hz, $(CH_2)_3CH_3$), 1.38–1.50 (6H, m, $CH_2)_3CH_3$), 1.87 (1H, m, CHC*H*H), 2.01 (1H, m, CHCH*H*), 2.86–3.00 (2H, m, 4-H_2), 3.29–3.34 (1H, m, 3-*H*H), 3.50 (1H, m, 3-H*H*), 4.34 (1H, dd, *J* = 8.3 and 5.0 Hz, 1-H), 6.60 (1H, s, 5-H), 6.64 (1H, s, 8-H); ^{13}C NMR (125 MHz; CD_3OD) δ 14.3, 23.5, 25.7, 26.1, 32.7, 35.1, 41.0, 56.7, 113.9, 116.2, 118.3 (q, $^1J_{CF}$ 295 Hz, CF_3), 123.7, 124.3,

145.9, 146.7, 163.2 (br, CF_3CO_2); *m/z* [HRMS ES+] found $[M+H]^+$ 236.1639. $C_{14}H_{22}NO_2$ requires 236.1651.

B: Phenylacetaldehyde and *Cj*NCS2 lysate: Substrate solutions of dopamine-HCl (38 mg, 0.20 mmol) in water (2 mL) and phenylacetaldehyde (36 mg, 0.30 mmol) in MeCN (2 mL), and were prepared. HEPES-lysate of *Cj*NCS2 was prepared as described as in section 2.3.3. The reaction mixtures were prepared by the addition of the dopamine solution (2 mL, 0.20 mmol), the phenylacetaldehyde solution (2 mL, 0.30 mmol) and HEPES-lysate (7 mL) to water (9 mL). The final reaction mixtures (20 mL) contained 10 mM dopamine and 15 mM phenylacetaldehyde. The reaction was incubated at 37 °C for 1 hour before being quenched by the addition of 1M HCl (2 mL). The mixture was centrifuged (10,000 × g for 30 mins) to remove protein precipitate and then filtered. Samples were taken at this point and analysed by HPLC methods 1 and 3 to determine *ee* and conversion yields. The product was purified via preparative-HPLC using method 5. Solvent was removed under reduced pressure and the product washed with MeOH to remove excess TFA, yielding **(*S*)-1-benzyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol** as the TFA salt (33 mg, 45%, >97% *ee*) with spectroscopic data consistent that previously reported.[6]

Table 2.5 HPLC retention times for biotransformations to give (1*S*)-1-pentyl-1,2,3,4-tetrahydro-isoquinoline-6,7-diol and (1*S*)-1-benzyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol.

Product		Pentyl		Benzyl	
Analysis	**HPLC method**	**RT (min)**	**Yield**	**RT (min)**	**Yield**
Conversion	1	5.9	70%	5.5	81%
ee	3	12.7 (*S*) 15.9 (*R*)	>98%	12.0 (*S*) 16.8 (*R*)	>97%
Purified	5	32.0	57%	18.4	45%

2.6.2 Triangular cascades

These following methods relate to section 5.4.

Transaminase screen

Transaminases were screened for activity with dopamine and pyruvate through the formation of *rac*-norlaudanosoline. The reaction conditions: 50 mM dopamine, 25 mM pyruvate, 1 mM PLP, 10% $v.v^{-1}$ TAm lysate. The reactions (50 µL) were incubated at 37 °C with 500 rpm orbital rotation. The reactions were quenched by addition of 10 µL 1M HCl. Reactions were quenched after 30 mins, 1, 2 and 4 hours. The insoluble portions of the reactions were removed by centrifugation and the soluble fractions were analysed by analytic HPLC (method 1) and compared to chemically verified standards.

One-pot two-enzyme cascade for the synthesis of (*S*)-norlaudanosoline

Reactions were conducted to screen for optimum conditions. The reactions (100 µL) contained dopamine, pyruvate, purified Δ29*Tf*NCS and CV2025 lysate in varying concentrations. Conditions were: either 20 or 50 mM dopamine (and 10 mM or 25 mM sodium pyruvate respectively), 0.1 or 0.5 $mg.mL^{-1}$ NCS, 10, 20 or 30% CV2025 lysate. Reactions were quenched with 20 µL 1 M HCl. Reactions were incubated at 37 °C with 500 rpm orbital rotation and were quenched after 3 hours. Each condition was performed in triplicate. The insoluble material was removed by centrifugation and the reaction were analysed by analytical and chiral HPLC (methods 1 and 2).

One-pot two-enzyme cascade for the synthesis of (*S*)-1-(3-Hydroxybenzyl)-1,2,3,4-tetrahydro-isoquinolin-6-ol

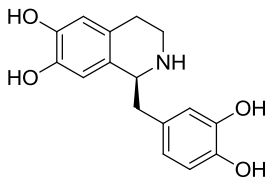
Reaction components 3-hydroxyphenethylamine (20 mM), pyruvate (10 mM), CV2025 lysate (20% $v.v^{-1}$) and Δ29*Tf*NCS (0.5 $mg.mL^{-1}$) were combined (total volume 100 µL) and incubated at 37 °C with 500 rpm shaking for 3 hours. Reactions were quenched with 1M HCl (20 µL), centrifuged and then analysed by analytical and chiral HPLC (methods 1 and 3).

One-pot two-enzyme chemoenzymatic cascade for the synthesis of tetrahydroprotoberberines

First, synthesis of (*S*)-norlaudanosoline was achieved with conditions determined previously: reactions (100 µL) consisting of dopamine (20 mM), pyruvate (10 mM), Δ29*Tf*NCS (0.5 $mg.mL^{-1}$) and CV2025 lysate (20% $v.v^{-1}$) were incubated at 37 °C with 500 rpm shaking for 3 hours. To these reactions was added 100 µL of formaldehyde/phosphate solution (40 mM formaldehyde, 1 M sodium phosphate, pH 6) and the reactions were incubated for 30 minutes at 37 °C and with 500 rpm shaking. Reactions were centrifuged to remove insoluble matter and then analysed by analytical and chiral HPLC (methods 1 and 2).

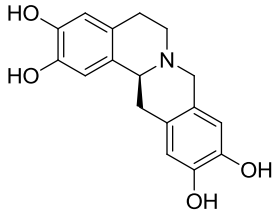
Preparatory scale enzyme cascade reactions

(*S*)-1-(3,4-Dihydroxybenzyl)-1,2,3,4-tetrahydroisoquinoline-6,7-diol



Dopamine (0.5 mmole, 94.5 mg), pyruvate (0.25 mmole, 27.5 mg), CV2025 lysate (5 mL) and Δ29*Tf*NCS (12.5 mg) were combined with water to a total volume of 25 mL. The reaction was stirred at 37 °C, checking reaction progress by HPLC analysis (method 1). Once the reaction had completed (86% conversion, 2 hours), the reaction was quenched by the addition of HCl (1M, 5 mL). Insoluble material was removed through centrifugation and ultrafiltration, and the mixture was concentrated *in vacuo*. The product was purified by preparative-HPLC (method 4) for characterisation purposes (TFA salt, 63 mg, 62% isolated yield). ^{1}H NMR and chiral HPLC (method 4) confirmed the product was the expected **(*S*)-1-(3,4-dihydroxybenzyl)-1,2,3,4-tetrahydroisoquinoline-6,7-diol** (>95% *ee*). See section 2.8.5 for synthesis and characterisation of chemical standards. See Appendix B (section 9) for NMR spectra.

(*S*)-6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,10,11-tetraol



Dopamine (0.5 mmole, 94.5 mg), pyruvate (0.25 mmole, 27.5 mg), CV2025 lysate (5 mL) and Δ29*Tf*NCS (12.5 mg) were combined with water to a total volume of 25 mL. The reaction was stirred at 37 °C for 2 hours. Formaldehyde (40 mM in 1M phosphate pH 6, 25 mL) was added to the reaction and it was stirred at 37 °C for 30 minutes (47% conversion). Insoluble material was removed through centrifugation and ultrafiltration, and the mixture was concentrated *in vacuo*. The product was purified by preparative-HPLC (method 4) for characterisation (TFA salt, 43 mg, 42% isolated yield). ^{1}H NMR and chiral HPLC (method 2) confirmed the product was the expected **(*S*)-6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,10,11-tetraol** (>95% *ee*). See section 2.8.5 for synthesis and characterisation of chemical standards. See Appendix B (section 9) for NMR spectra.

2.6.3 Spiro-THIQs

Refers to Section 6.6.

Enzyme preparation: A Δ29*Tf*NCS variant was expressed as described in section 2.3.1. The frozen cell pellet was thawed and resuspended in 50 mM HEPES buffer pH 7.5 (10% volume of original culture). The suspension was sonicated (10 s ON, 10 s OFF) until homogenous. The insoluble portion of the lysate was pelleted by centrifugation at 10,000 × g for 30 minutes at 4 °C. The supernatant was removed and filtered through a glass fibre prefilter and 0.2 µm cellulose acetate syringe filter.

Reaction conditions: Reactions were conducted in 5 mL volumes and contained: 20% $v.v^{-1}$ lysate, 5 mM ascorbic acid, 10 mM substituted cyclohexanone (CH), 15 mM dopamine, 10% $v.v^{-1}$ DMSO and 50 mM HEPES pH 7.5. The components were mixed, with the lysate added last. The reaction was incubates for 6 hours at 37 ˚C, shaking at 250 rpm. The reactions were stored at -20 ˚C until further processing.

Extraction procedure: The reaction was diluted to 10 mL with water. The solution was extracted with ethyl acetate (3 x 20 mL). The fractions were pooled and washed with brine (3 x 20 mL). The organic layer was dried with magnesium sulfate, filtered and the solvent removed *in vacuo*. The material was resuspended in 20 mL of a 1:1 mixture of dichloromethane (DCM) and 0.1 M HCl. The DCM layer was extracted with 0.1 M HCl (2 x 10 mL). The aqueous layers were combined and water was removed *in vacuo* by a rotary evaporator until 5 mL remained. Then the final compound was freeze dried to yield the product as a solid hydrochloride salt (Table 2.6).

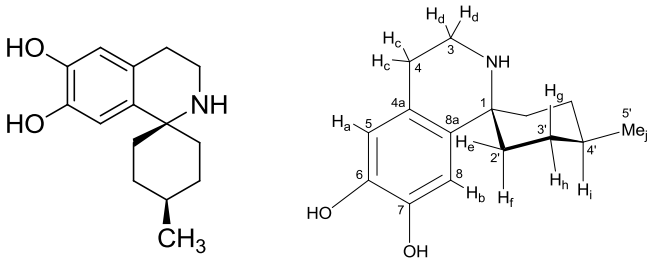
Table 2.6 Purification of cyclohexanone derived spiro-THIQ products

Cyclohexanone Substrate	Purified mass (mg)	Yield %	Appearance
4-methyl	7.3	51	Grey solid
4-phenyl	10.0	58	White solid
4-*tert*-butyl	5.4	33	White solid
3-methyl (racemate)	8.5	60	Yellow solid

Characterisation data:

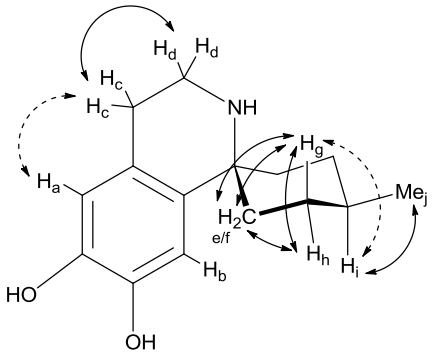
4-Methylcyclohexane-spiro-THIQ

(1*s*,4*s*)-4-methyl-3',4'-dihydro-2'H-spiro[cyclohexane-1,1'-isoquinoline]-6',7'-diol

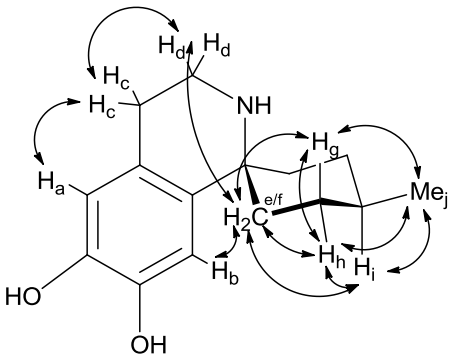


^{1}H NMR (600 MHz; CD_3OD): δ = 1.04 (3H, d, *J* 6.5 Hz, Me_j), 1.31 (2H, m, H_g), 1.65 (1H, m, H_i), 1.78 (1H, app. d, *J* 14.3 Hz, H_h), 2.05-2.10 (4H, m, H_e and H_f), 2.96 (2H, t, *J* 6.3 Hz, H_c), 3.41 (2H, t, *J* 6.3 Hz, H_d), 6.56 (1H, s, H_a), 6.73 (1H, s, H_b); ^{13}C NMR (150 MHz; CD_3OD): δ = 22.2 (5'), 26.2 (4), 30.1 (3'), 32.4 (4'), 36.9 (2'), 38.7 (3), 60.6 (1), 113.2 (8), 116.1 (5), 123.1 (8a), 129.5 (4a), 146.1(6/7), 146.5 (6/7); *m/z* [HRMS ES+] found $[M+H]^+$ 248.1650. $C_{15}H_{22}NO_2$ requires 248.1651.

For NMR spectra, see Appendix B (section 9).



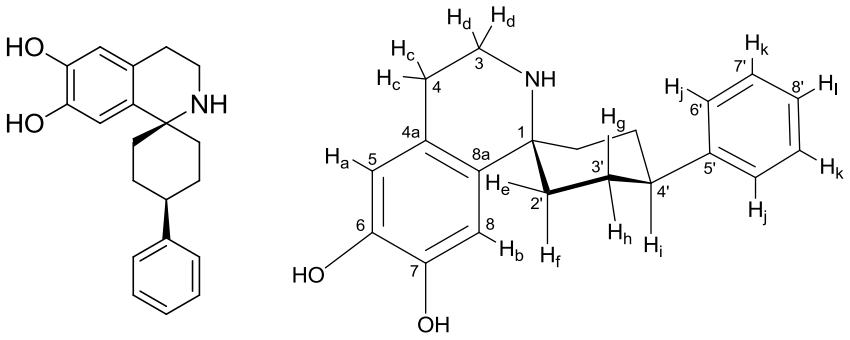
COSY correlations
Weaker signals are dashed



NOESY correlations

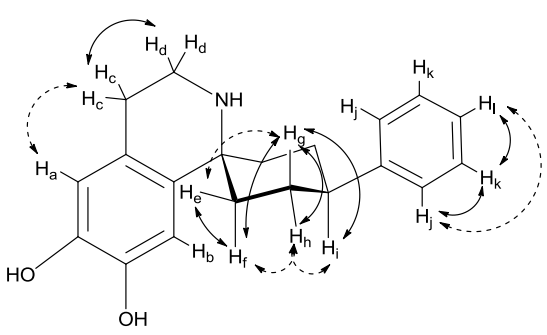
4-Phenylcyclohexane-spiro-THIQ

(1*s*,4*s*)-4-phenyl-3',4'-dihydro-2'H-spiro[cyclohexane-1,1'-isoquinoline]-6',7'-diol

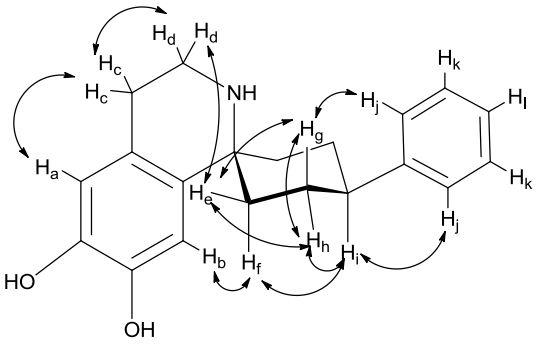


^{1}H NMR (600 MHz; CD_3OD): δ = 1.84 (2H, m, H_g), 1.97 (2H, d, *J* 14.6 Hz, H_h), 2.21 (2H, d, *J* 14.9 Hz, H_e), 2.28 (2H, m, H_f), 2.83 (1H, tt, *J* 12.3 and 3.4 Hz, H_i), 3.00 (2H, t, *J* 6.3 Hz, H_c), 3.49 (2H, t, *J* 6.3 Hz, H_d), 6.59 (1H, s, H_a), 6.83 (1H, s, H_b), 7.20 (1H, tt, *J* 7.3 and 1.2 Hz, H_l), 7.31 (2H, m, H_k), 7.36 (2H, app. d, *J* 7.9 Hz, H_j); ^{13}C NMR (150 MHz; CD_3OD): δ = 26.3 (4), 29.5 (3'), 37.3 (2'), 38.8 (3), 44.0 (4'), 60.6 (1), 113.3 (8), 116.1 (5), 123.2 (8a), 127.5 (8'), 127.9 (7'), 129.3 (4a), 129.6 (6'), 146.2 (6/7), 146.6 (6/7), 146.9 (5'); *m/z* [HRMS ES+] found $[M+H]^+$ 310.1808. $C_{20}H_{24}NO_2$ requires 310.1807.

For NMR spectra, see Appendix B (section 9).



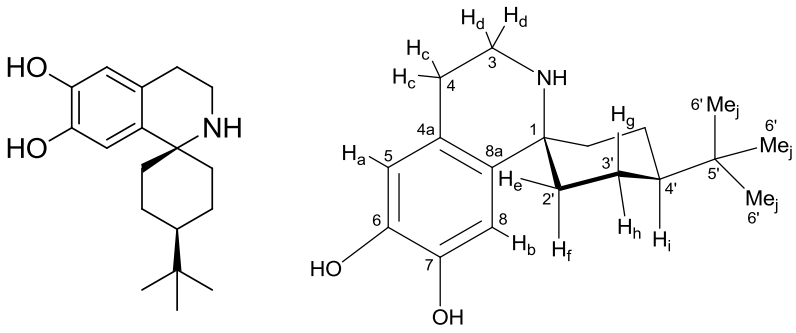
COSY correlations
Weaker signals are dashed



NOESY correlations

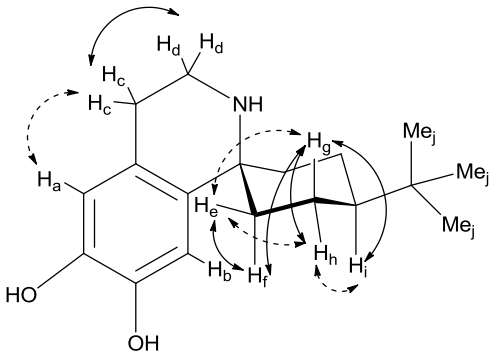
4-*Tert*-butylcyclohexane-spiro-THIQ

(1*s*,4*s*)-4-(*tert*-butyl)-3',4'-dihydro-2'H-spiro[cyclohexane-1,1'-isoquinoline]-6',7'-diol

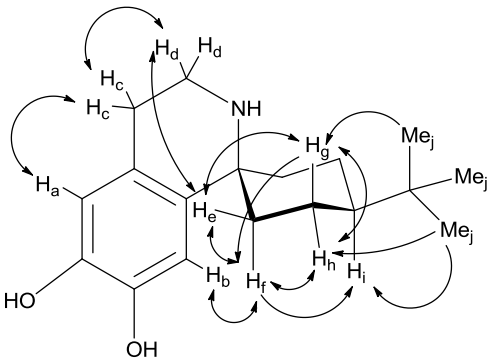


^{1}H NMR (600 MHz; CD_3OD): δ = 0.96 (9H, s, H_j), 1.28-1.40 (3H, m, H_i and H_g), 1.87 (2H, app. d, *J* 13.0 Hz, H_h), 2.05 (2H, m, *J* 13.4 and 15.0 Hz, H_f) 2.13 (2H, d, *J* 15.0 Hz, H_e), 2.96 (2H, t, *J* 6.3 Hz, H_c), 3.42 (2H, t, 6.3 Hz, H_d), 6.57 (1H, s, H_a), 6.73 (1H, s, H_b); ^{13}C NMR (150 MHz; CD_3OD): δ = 23.2 (3'), 26.3 (4), 27.9 (6'), 33.3 (5'), 37.7 (2'), 38.9 (3), 48.4 (4'), 60.8 (1), 113.3 (8), 116.0 (5), 123.3 (8a), 129.4 (4a), 146.0 (6/7), 146.5 (6/7); *m/z* [HRMS ES+] found $[M+H]^+$ 290.2120. $C_{18}H_{28}NO_2$ requires 290.2120.

For NMR spectra, see Appendix B (section 9).



COSY correlations
Weaker signals are dashed

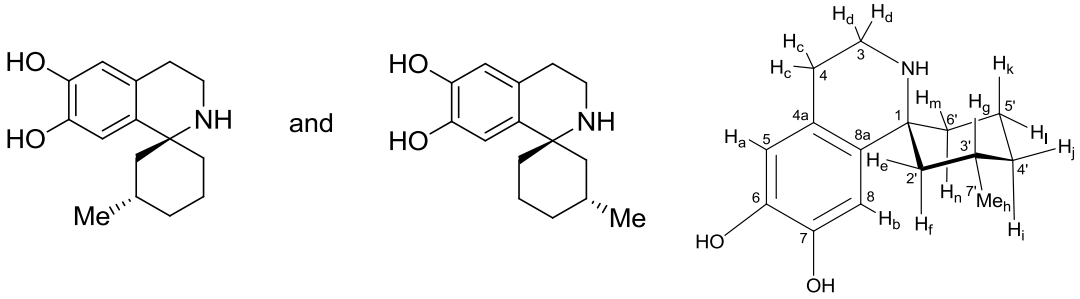


NOESY correlations

***Rac*-3-methylcyclohexane-spiro-THIQ**

(1*S*,3*S*)-3-methyl-3',4'-dihydro-2'H-spiro[cyclohexane-1,1'-isoquinoline]-6',7'-diol

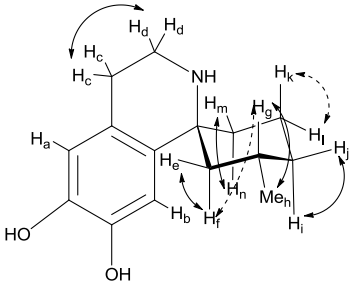
(1*R*,3*R*)-3-methyl-3',4'-dihydro-2'H-spiro[cyclohexane-1,1'-isoquinoline]-6',7'-diol



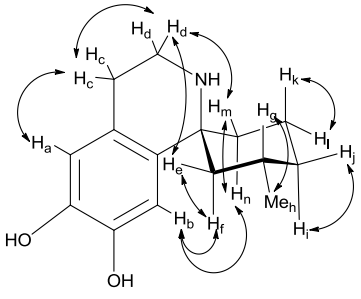
^{1}H NMR (600 MHz; CD_3OD): δ = 1.00 (3H, d, *J* 6.32 Hz, Me_h), 1.13 (1H, dq, *J* 3.7 and 12.6 Hz, H_i),1.61-1.69 (2H, m, H_f and H_k), 1.74 (1H, br m, H_g), 1.85 (2H, m, H_j and H_l), 1.93 (1H, dt, *J* 4.3 and 14.4 Hz, H_n), 2.04 (2H, m, H_e and H_m), 2.96 (2H, dt, *J* 2.3 and 6.2 Hz, H_c), 3.41 (2H, m, *J* 4.3 and 6.2 Hz, H_d), 6.57 (1H, s, H_a), 6.73 (1H, s, H_b); ^{13}C NMR (150 MHz; CD_3OD): δ = 21.6 (5'), 22.5 (7'), 26.2 (4), 28.3 (3'), 34.1 (4'), 36.3 (6'), 38.7 (3), 45.3 (2'), 61.7 (1), 113.3 (8), 116.1 (5), 123.0 (4a), 129.6 (5a), 146.1 (6/7), 145.5 (6/7); *m/z* [HRMS ES+] found $[M+H]^+$ 248.1649. $C_{15}H_{22}NO_2$ requires 248.1651.

HSQC ^{1}H (600 MHz; CD_3OD): δ = 1.60-1.68 (t, H_f), 1.60-1.69 (q, H_k), 1.74 (br m, H_g), 1.81-1.87 (d, H_l), 1.83-1.88 (d, H_j), 1.93 (dt, H_n), 2.00-2.05 (d, H_e), 2.03-2.07 (d, H_m).

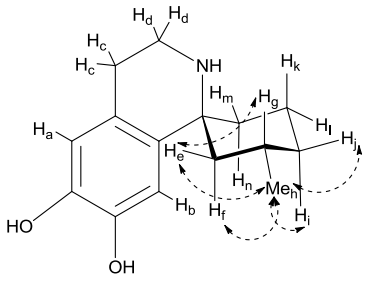
For NMR spectra and further analysis, see Appendix B (section 9).



COSY correlations
Dashed lines show weak interactions



Strong NOESY correlations



Weak NOESY correlations

2.7 Screening protocols

Refers to section 6.

2.7.1 Fluorescamine methods

Dopamine (100 µM), aldehyde (1 mM) and purified enzyme (25 µg.mL^{-1}) were incubated for 1-3 h at 37 ˚C. Fluorescamine (10 mM in MeCN, 25% volume of the reaction) was added, the mixture was incubated for 5 minutes, and then the fluorescence was read on a plate reader (see section 2.4.3).

2.7.2 Purpald® progress curve

Mixtures containing 10 µL dopamine (100 mM) and 10 µL aldehyde (100 mM) and 80 µL desalted lysate (0.1 mg.mL^{-1}) in 0.1 M HEPES buffer (pH 7.5) were prepared. After 10 minutes 40 µL of the reaction mixture was added to 40 µL Purpald® (100 mg.mL^{-1} in 1N NaOH) and 40 µL was added to 10 µL 1M HCl. 40 µL of the acid quenched mixture was diluted with 100 µL H_2O and analysed by HPLC by method 1. After 30 minutes the absorbance of the Purpald® mixtures were read according to the plate reading method described above (section 2.4.3). Concentration of aldehyde was determined by absorbance at 522 nm and quantified by comparison with standards.

2.7.3 Full micro-scale screening protocol

An aliquot from a frozen glycerol stock of plasmid containing *E. coli* BL21(DE3) was inoculated into 20 mL of TB medium (containing 50 µg.mL^{-1} kanamycin). Starter cultures were incubated overnight at 37 °C, shaking at 250 rpm. 24 µL of starter culture was used to inoculate 600 µL TB medium (containing 50 µg.mL^{-1} kanamycin) in deep-well microtitre plates. The expression cultures were incubated for 2 hours at 37 °C followed by 1 hour at 25 °C, whilst shaking at 1,100 rpm. Expression was induced by the addition of 6 µL of 50 mM IPTG. Cultures were incubated for a further 3 hours at 25 °C prior to harvesting, whilst shaking at 250 rpm. Cells were harvested by centrifugation at 4,000 × g for 10 mins at 4 °C. Supernatant was removed and the cells stored at -20 °C until purification. Cell pellets were thawed and suspended in BugBuster (60 µL). Lysate was transferred to U-bottomed microtitre plates. The insoluble portion of the lysate was pelleted by centrifugation at 4,000 × g for 10 mins at 4 °C. 50 µL of the supernatant was desalted using a Zeba desalting plate (>7 kDa MW, prepared according to manufacturer's instructions (ThermoFisher, Massachusetts, USA). Purpald® (10 mg.mL^{-1} in 1N NaOH), dopamine (100 mM in water) and aldehyde (100 mM in MeCN) stocks were prepared. 5 µL of dopamine stock and 5 µL of aldehyde stock were added to 40 µL of lysate (substrate final concentration of 10 mM). Reaction was mixed and incubated at 37 °C. After 10 minutes, 50 µL Purpald® solution

was added to quench the reaction. After 30 minutes incubation at room temperature, the absorbance of each mixture was recorded using the Purpald® plate reading method (see section 2.4.3).

2.8 Chemicals and syntheses

All reagents were obtained from commercial sources (Sigma-Aldrich, St. Louis, MO, USA) and used as received unless otherwise stated. Non-commercially available compounds were provided by Dr Thomas Pesnot and Dr Eleanor Lamming.

2.8.1 Aldehydes

4-hydroxyphenylacetaldehyde (4-HPAA), 3,4-dihydroxyphenylacetaldehyde (3,4-DHPAA) and 3,4-methylenedioxyphenylacetaldehyde were synthesised by Dr Eleanor Lamming and Dr Thomas Pesnot as previously described.[57,59,94]

2.8.2 Amines

Syntheses of 3-hydroxyphenethylamine and 2-bromo,5-hydroxyphenylethylamine have been previously described.[57,94] 3,4-dihydroxyphenylpropylamine, *N*-methyl-3-hydroxy-phenethylamine, 4,*N*-dimethyl-dopamine, and 3,5-dihydroxy-phenethylamine were synthesised by Dr Eleanor Lamming.

2.8.3 Benzylisoquinoline alkaloids (BIAs)

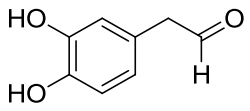
These compounds were used as analytical HPLC standards. Norcoclaurine, 1-(3-hydroxyphenyl)-1,2,3,4-tetrahydroisoquinoline-6,7-diol (3-hydroxbenzaldehyde product), 1-(2,6-dimethylhept-5-en-1-yl)-1,2,3,4-tetrahydro-isoquinoline-6,7-diol (citronellal product), 1-(3,4-methylenedioxybenzyl)-1,2,3,4-tetrahydroisoquinoline-6,7-diol (methylene-dioxy-phenylacetaldehyde product) and 1-benzyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol (phenylacetaldehyde product) syntheses have been described previously.[59,94] The characterisation of the novel hexanal product (Pentyl-1,2,3,4-tetrahydroisoquinoline-6,7-diol) is described above (section 2.6.1). Compounds norlaudanosoline and 1-(3-Hydroxybenzyl)-1,2,3,4-tetrahydroisoquinolin-6-ol (from triangular cascade, section 5.4.8) were synthesised by Dr Eleanor Lamming according to the standard phosphate method and are described below.[59] Racemic tetrahydroprotoberberines (THPBs) were synthesised and characterised by Dr Tom Pesnot and are described below, section 2.8.5.

2.8.4 Chemical syntheses of BIAs

These syntheses were performed by Dr Eleanor Lamming, but have been included for completeness.

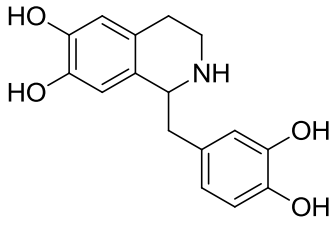
Synthesis of norlaudanosoline

(3,4-Dihydroxyphenyl)acetaldehyde*



(3,4-Dihydroxyphenyl)acetaldehyde (3,4-DHPAA) was prepared by the Parikh-Doering oxidation of 4-(2-hydroxyethyl)benzene-1,2-diol[124] (500 mg, 3.24 mmol). To a solution of 4-(2-hydroxyethyl)benzene-1,2-diol (1 eq.) in DMSO/CH_2Cl_2 (15 mL) was added *N*,*N*-diisopropylethylamine (2.5 eq.) followed by a solution of SO_3.pyridine (2.5 eq.) in a 1:1 mixture of DMSO/CH_2Cl_2 (10 mL), added over 30 min at -15 ºC. The mixture was stirred for 1 h at -15 ºC and quenched by the addition of ice-cold water (50 mL). The aqueous layer was extracted with CH_2Cl_2 (3 × 50 mL) and the organic layers were combined and concentrated under reduced pressure. The crude material was purified using flash silica chromatography (60% EtOAc in hexane) to give (3,4-dihydroxyphenyl)acetaldehyde as a pale yellow oil (212 mg, 43%).[57,125] R_f 0.15 (EtOAc:hexane, 1:1); ^{1}H NMR (600 MHz; $CDCl_3$) δ = 3.59 (2H, d, *J* = 2.4 Hz, CH_2), 5.51 (1H, br s, OH), 5.76 (1H, br s, OH), 6.64 (1H, dd, *J* = 8.0 and 2.0 Hz, 6-H), 6.70 (1H, d, *J* = 2.0 Hz, 2-H), 6.85 (1H, d, *J* = 8.0 Hz, 5-H), 9.69 (1H, t, *J* 2.4 Hz, CHO); ^{13}C NMR (150 MHz; $CDCl_3$) δ = 49.9, 115.9, 116.6, 122.3, 124.2, 143.1, 144.0, 200.6; *m/z* [HRMS ES+] found $[M]^+$ 152.04679. $C_8H_8O_3$ requires 152.04680.

Norlaudanosoline*



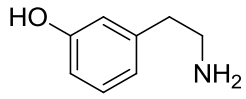
Norlaudanosoline/tetrahydropapaveroline was prepared by a biomimetic Pictet-Spengler reaction. Dopamine.HCl (75 mg, 0.39 mmol) and (3,4-dihydroxyphenyl)acetaldehyde (40 mg, 0.26 mmol) were added to 6 mL of a 1:1

*This synthesis was conducted by Dr Eleanor Lamming
*This synthesis was conducted by Dr Eleanor Lamming

mixture of acetonitrile/potassium phosphate buffer (0.1 M solution at pH 6). The resulting solution was stirred at 50 ºC for 12 h. The crude product was concentrated *in vacuo*, then purified by preparative-HPLC (method 4, r_t (retention time) 11.5 min). Fractions containing the desired product were combined, concentrated and co-evaporated with methanol (3 × 20 mL) to give **1-(3,4-dihydroxybenzyl)-1,2,3,4-tetrahydroisoquinoline-6,7-diol** as a colourless oil (46 mg, 61%).[126] ^{1}H NMR (600 MHz; CD_3OD) δ = 2.86–3.02 (3H, m, 4-H_2 and NCHC*H*H), 3.23 (1H, app. quintet, *J* = 6.0 Hz, 3-*H*H), 3.30–3.36 (1H, m, NCHCH*H*), 3.44 (1H, app. quintet, *J* = 5.9 Hz, 3-H*H*), 4.54 (1H, dd, *J* = 8.8 and 5.6 Hz, 1-H), 6.61–6.65 (3H, m, 5-H, 8-H and 6'-H), 6.74 (1H, d, *J* = 1.7 Hz, 2'-H), 6.78 (1H, d, *J* = 8.0 Hz, 5'-H); ^{13}C NMR (150 MHz; CD_3OD) δ = 25.7, 40.7, 40.9, 57.9, 114.0, 116.1, 116.9, 117.4, 121.8, 123.6, 123.8, 127.7, 145.8, 146.1, 146.6, 146.8; *m/z* [HRMS ES+] found $[M+H]^+$ 288.1236. $C_{16}H_{18}NO_4$ requires 288.1233.

Synthesis of 1-(3-hydroxybenzyl)-1,2,3,4-tetrahydro-isoquinolin-6-ol

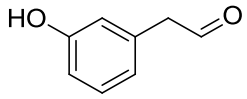
2-(3-Hydroxyphenyl)ethylamine hydrobromide*



The reaction was performed under anhydrous conditions. A solution of 1 M boron tribromide in CH_2Cl_2 (10.2 mL, 10.2 mmol) was added to a stirred solution of 2-(3-methoxyphenyl)ethan-1-amine (0.700 g, 4.63 mmol) in CH_2Cl_2 (20 mL) at -78 ^{o}C under Ar. The reaction was warmed to room temperature and stirred for 24 h. The reaction was then cooled to 0 ^{o}C and quenched by addition of methanol (40 mL). The solution was stirred at room temperature for 3 h. The solution was concentrated *in vacuo* to give a brown oil. Further methanol (20 mL) was added, and solvent evaporated. This was repeated until no white fumes were observed upon addition of methanol, to give 2-(3-hydroxyphenyl)ethylamine as a pale brown solid as the hydrobromide salt which was used without further purification (0.899 g, 89%).[127] m.p. 102-104 ^{o}C; ^{1}H NMR (600 MHz; CD_3OD) δ = 2.88 (2H, t, *J* = 8.0 Hz, CH_2CH_2N), 3.15 (2H, t, *J* = 8.0 Hz, CH_2N), 6.68-6.72 (2H, m, 2-H and 4-H), 6.73 (1H, d, *J* = 7.5 Hz, 6-H), 7.16 (1H, t, *J* = 7.5 Hz, 5-H); ^{13}C NMR (150 MHz; CD_3OD) δ = 34.5, 41.9, 115.2, 116.6, 120.7, 131.0, 139.3, 159.1; *m/z* [HRMS ES+] found $[M+H]^+$ found 138.091213. $C_8H_{11}ON$ requires 138.09189.

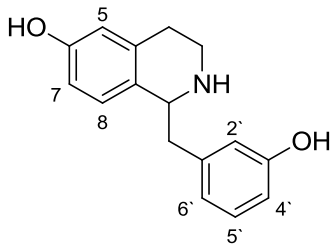
* This synthesis was conducted by Dr Eleanor Lamming

2-(3-Hydroxyphenyl)acetaldehyde*



2-(3-Hydroxyphenyl)acetaldehyde was prepared by Parikh-Doering oxidation of 2-(3-hydroxylphenyl)ethanol (500 mg, 3.62 mmol). To a solution of 2-(3-hydroxylphenyl)ethanol (1 eq.) in DMSO/CH_2Cl_2 (15 mL) was added *N*,*N*-diisopropylethylamine (2.5 eq.) followed by a solution of SO_3.pyridine (2.5 eq.) in a 1:1 mixture of DMSO/CH_2Cl_2 (10 mL) added over 30 min at -15 ºC. The mixture was stirred for 1 h at -15 ºC and quenched by the addition of ice-cold water (50 mL). The aqueous layer was extracted with CH_2Cl_2 (3 × 50 mL) and the organic layers were combined and concentrated under reduced pressure. The crude material was purified using silica flash chromatography (eluent 20% EtOAc in hexane). Fractions containing the desired product were combined and concentrated under reduced pressure. Purification by flash silica chromatography (15% EtOAc in hexane) yielded 2-(3-hydroxyphenyl)acetaldehyde as a pale yellow oil (100 mg, 20%).[128] R_f 0.47 (EtOAc:hexane, 1:1); ^{1}H NMR (500 MHz; $CDCl_3$) δ = 3.64 (2H, d, J = 2.4 Hz, CH_2), 5.46 (1H, br s, OH), 6.69 (1H, t, J = 2.0 Hz, 2-H), 6.75–6.79 (2H, m, 4-H and 6-H), 7.23 (1H, t, J = 6.4 Hz, 5-H), 9.73 (1H, t, J = 2.4 Hz, CHO); ^{13}C NMR (125 MHz; $CDCl_3$) δ = 50.4, 114.6, 116.6, 122.0, 130.3, 133.4, 156.3, 199.8; *m/z* [HRMS ES+] found $[M]^+$ 136.05206. $C_8H_8O_2$ requires 136.05188.

1-(3-Hydroxybenzyl)-1,2,3,4-tetrahydroisoquinolin-6-ol*



A solution of 2-(3-hydroxyphenyl)acetaldehyde (27.2 mg, 0.200 mmol) in acetonitrile (2 mL) was added to a solution of 2-(3-hydroxyphenyl)ethylamine (35.0 mg, 0.160 mmol) in potassium phosphate buffer (2 mL, 0.1 M, pH 6). The solution was stirred at 50 °C for 17 h. The reaction was concentrated and purified by preparative-HPLC (method 5, r_t = 30 min). Fractions containing the desired product were combined, concentrated and co-evaporated with methanol (3 x 10 mL), to give 1-(3-Hydroxybenzyl)-1,2,3,4-

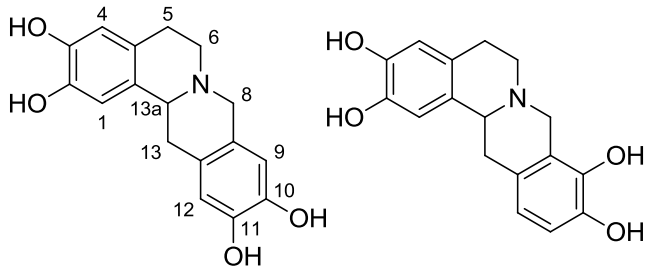
* This synthesis was conducted by Dr Eleanor Lamming

tetrahydro-isoquinolin-6-ol as a pale yellow oil (15.0 mg, 37%). ^{1}H NMR (CD_3OD; 600 MHz) δ = 2.96-3.03 (2H, m, 4-*H*H and NCHC*H*H) 3.05-3.12 (1H, m, 4-H*H*), 3.26-3.31 (1H, m, 3-*H*H), 3.40-3.45 (1H, m, NCHCH*H*), 3.46-3.52 (1H, m, 3-H*H*), 4.66-4.70 (1H, m, 1-H), 6.65 (1H, s, 5-H), 6.68 (1H, d, *J* = 8.5 Hz, 7-H), 6.73-6.80 (3H, m, 2`-H, 6`-H, 4`-H), 7.03 (1H, d, *J* = 8.5 Hz, 8-H), 7.12 (1H, t, *J* = 7.5 Hz, 5`-H); ^{13}C NMR (150 MHz; CD_3OD) δ = 24.9, 39.1, 39.6, 56.2, 114.1 (2 x signals), 114.5, 115.8, 119.9, 122.0, 127.5, 129.7, 132.4, 136.4, 157.0, 157.7); *m/z* [HRMS ES+] found $[M+H]^+$ 256.133461. $C_{16}H_{18}O_2N$ requires 256.13375.

2.8.5 Chemical syntheses of THPBs

These syntheses were performed by Dr Thomas Pesnot, but have been included for completeness.

Synthesis of 6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,10,11-tetraol, and **6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,9,10-tetraol***



Norlaudanosoline (30 mg, 100 µmol) and formaldehyde (15 µL, 0.20 mmol) were added to a 1:1 mixture of acetonitrile/potassium phosphate buffer (0.5 M solution at pH 6, 3 mL). The solution was stirred at 40 ºC for 0.5 h. The crude product was purified by preparative-HPLC (method 4) and fractions containing the desired product were combined, concentrated and co-evaporated with methanol (3 × 20 mL). Two products were isolated 6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,10,11-tetraol[129] (22.6 mg, 72%, RT (retention time) 11.6 min) and 6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,9,10-tetraol[129] (5.3 mg, 17% RT 12.6 min).

Major regioisomer **6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,10,11-tetraol**: ^{1}H NMR (600 MHz; CD_3OD) δ = 2.88–3.03 (2H, m, 5-*H*H and 13-*H*H), 3.12–3.23 (1H, m, 5-H*H*), 3.48 (1H, td, *J* = 12.3 and 4.5 Hz, 6-*H*H), 3.58–3.66 (1H, m, 13-H*H*), 3.75–3.82 (1H, m, 6-H*H*), 4.38-4.50 (2H, m, 8-H_2), 4.67 (1H, dd, *J* = 11.9 and 4.6 Hz, 13a-H), 6.62 (1H, s, 9-H), 6.64 (1H, s, 4-H), 6.71 (1H, br.s, 12-H),

* This synthesis was conducted by Dr Thomas Pesnot

6.79 (1H, br.s, 1-H); ^{13}C NMR (150 MHz; CD_3OD) δ = 24.8, 32.7, 51.0, 55.1, 60.5, 111.3, 111.7, 114.4, 114.5, 117.6, 117.9, 121.6, 122.3, 144.7, 144.8, 145.4, 145.6; *m/z* [HRMS ES+] found $[M+H]^+$ 300.1228. $C_{17}H_{18}NO_4$ requires 300.1236.

Minor regioisomer **6,8,13,13a-tetrahydro-5H-isoquinolino[3,2-a]isoquinoline-2,3,9,10-tetraol**: ^{1}H NMR (600 MHz; CD_3OD) δ = 2.93 (1H, br.d, *J* = 16.8 Hz, 5-*H*H), 3.02 (1H, dd, *J* = 16.8 and 12.3 Hz, 13-*H*H), 3.16-3.25 (1H, m, 5-H*H*), 3.53 (1H, td, *J* = 12.2 and 4.7 Hz, 6-*H*H), 3.67 (1H, dd, *J* = 16.8 and 4.0 Hz, 13-H*H*), 3.88 (1H, dd, *J* = 12.2 and 4.6 Hz, 6-H*H*), 4.31 (1H, d, *J* = 15.7 Hz, 8-*H*H), 4.66 (1H, dd, *J* = 12.3 and 4.0 Hz, 13a-H), 4.73 (1H, d, *J* = 15.7 Hz, 8-H*H*), 6.65 (1H, s, 4-H), 6.67 (1H, d, *J* = 8.2 Hz, 12-H), 6.77-6.82 (2H, m, 1-H and 11-H); ^{13}C NMR (150 MHz; CD_3OD) δ = 26.5, 34.3, 52.7, 53.4, 61.7, 112.9, 115.9, 166.4, 120.3, 123.3, 123.5, 123.6, 123.8, 143.1, 144.6, 146.5, 146.9; *m/z* [HRMS ES+] found $[M+H]^+$ 300.1219. $C_{17}H_{18}NO_4$ requires 300.1236.

3 Chapter 3: Expression and purification of *Tf*NCS

3.1 Introduction

The enzyme Δ19*Tf*NCS was selected to investigate the mechanism of NCS because it is the most studied enzyme in the literature. It was the first NCS to be isolated[15] and subsequently the first to be cloned and recombinantly expressed.[34] It has been examined with regards to its mechanism,[51] its substrate scope[53] and its structure.[43,55] The availability of Δ19*Tf*NCS X-ray structures was the crucial factor in selecting the enzyme, because mutations and their effect on mechanism can only be fully understood in a structural context.

Samanami *et al* originally obtained the cDNA for the *Tf*NCS gene from *Thalictrum flavum* cell cultures, using PCR with degenerate primers based on the peptide sequence of the isolated enzyme.[34] However, the full length gene did not produce recombinant protein. Genes encoding for N-terminal truncates of 10 and 19 residues (and a C-terminal His-Tag) both produced detectable and active recombinant NCS. Comparison with related PR-10 proteins revealed unique N- and C-terminal extensions on *Tf*NCS. A 19 residue N-terminal signal peptide was predicted from the sequence which led to the original decision to use Δ19*Tf*NCS for mechanistic and structural studies.[34]

3.2 Expression of Δ19*Tf*NCS

3.2.1 'Dimer' observation

A gene encoding a C-terminal His-tagged Δ19*Tf*NCS was obtained and cloned into a pET29a vector (see section 2.1.1).* Recombinant expression and purification of Δ19*Tf*NCS appeared to provide a 44 kDa protein, with approximately 80% purity (Fig. 3.1A). This protein was approximately double the expected size of Δ19*Tf*NCS (22.1 kDa), and was initially proposed to be a dimer form of the enzyme. However, incubation of the dimer with increasing concentrations of DTT failed to result in the observation of any monomer species by SDS-PAGE, perhaps suggesting a particularly tight dimer association (Fig. 3.1B).

* The cloning of Δ19*Tf*NCS was performed by Dr Markus Gershater

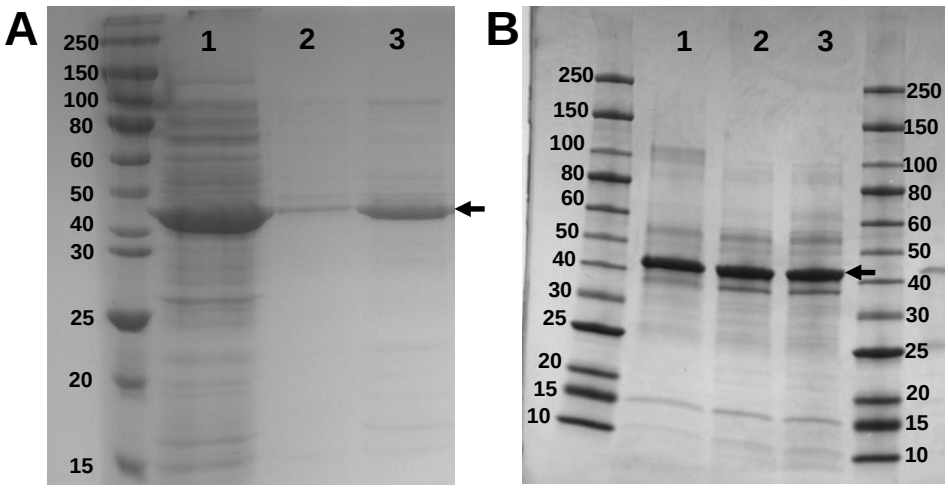


Fig. 3.1 Purification of 44 kDa protein. A: Purification of 44 kDa protein from Δ19*Tf*NCS expressing cells. Lane 1: lysate; 2: wash 20mM imidazole; 3: elution 250 mM imidazole. The 44 kDa protein (marked with arrow) was obtained with approximately 80% purity. These fractions were used for initial Δ19*Tf*NCS activity assays (see **Fig. 3.2**). **B**: Treatment of 44 kDa protein fraction with DTT. Lane 1: 0 M DTT; 2: 0.2 M DTT; 3: 0.5 M DTT. 44 kDa protein of interest marked with arrow. Gradient gel (4-20%) used.

The elution fraction containing this protein was shown to exhibit Pictet-Spengler activity (Fig. 3.2). For example, approximately 10% conversion was observed with wild-type (WT) enzyme over 3 hours (0.45 $mg.mL^{-1}$, 1 mM dopamine, 1.5 mM 4-HPAA). Although it was not possible to directly compare this with previously measured specific activities or kinetic parameters, the low conversions suggested orders-of-magnitude lower activity than expected. Variants of Δ19*Tf*NCS were produced by site-directed-mutagenesis with the intention of examining the substrate binding and mechanism of the enzyme. These demonstrated a diverse range of activities. The mutations Y108F, E110D/Q, D141N and M183F/L removed nearly all enzyme activity. L76A, F80L, F99L and A182F seemed to cause a relative shift in substrate specificity. The mutation A79I appeared to cause an increase in activity. However, despite these promising preliminary results, attempts to repeat the purification and assays of Δ19*Tf*NCS and its variants proved unsuccessful.

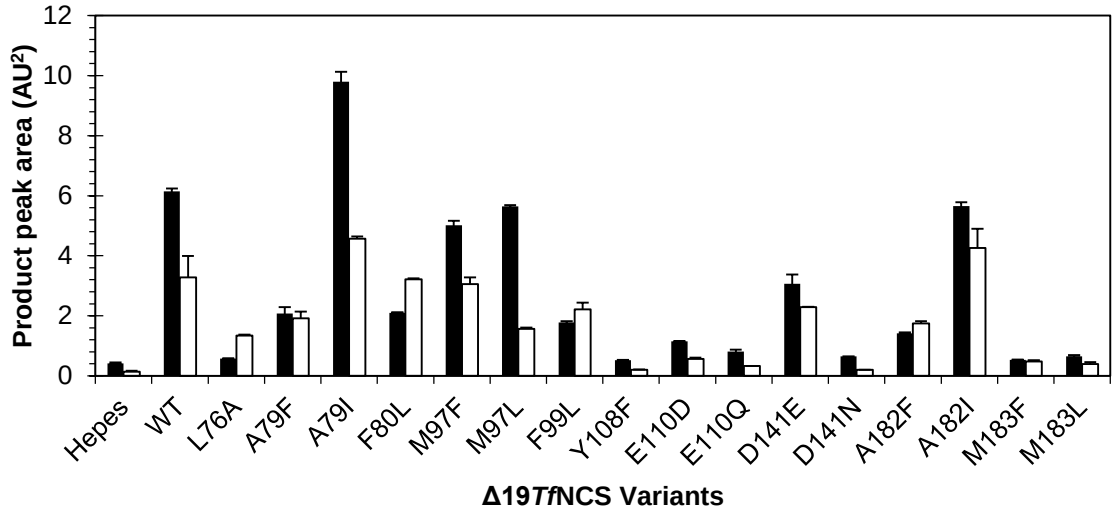


Fig. 3.2 Δ19*Tf*NCS activities (44 kDa fraction). Reaction conditions: 1 mM dopamine, 1.5 mM aldehyde. Final protein concentration varied between 0.05 – 0.75 mg.mL^{-1} (approximately 80% of this is the 44 kDa protein). Solid bars: 4-HPAA, white bars: heptanal. Product peak area is raw HPLC data and does not account for differences in absorbance in different products. Values are the mean of two measurements, error bars are standard deviations.

Expression of an empty vector control revealed a 44 kDa protein that bound to the nickel column at concentrations greater than 40 mM imidazole (Fig. 3.3A). This suggested that the 44 kDa NCS 'dimer' was merely a native *E. coli* His-rich protein. The most likely candidate for this protein was ODO2, component E2 of the dihydrolipoamide succinyl transferase, which has previously been shown to bind to His-trap columns and only elute at imidazole concentrations of approximately 50 mM.[130] The addition of a wash step with higher imidazole concentrations (up to 50 mM) removed this contaminant from the Δ19*Tf*NCS enzyme preparation (Fig. 3.3B, Lane 7). Furthermore, it became apparent that the Δ19*Tf*NCS bound to the column with high affinity; the use of 250 mM imidazole was insufficient to remove the majority of the protein from the column and 500 mM was required (Fig. 3.3B, Lane 8). The origin of NCS activity observed in the fractions containing mostly the 44 kDa protein was likely to be small contaminating quantities of monomeric Δ19*Tf*NCS, not detectable by SDS-PAGE.

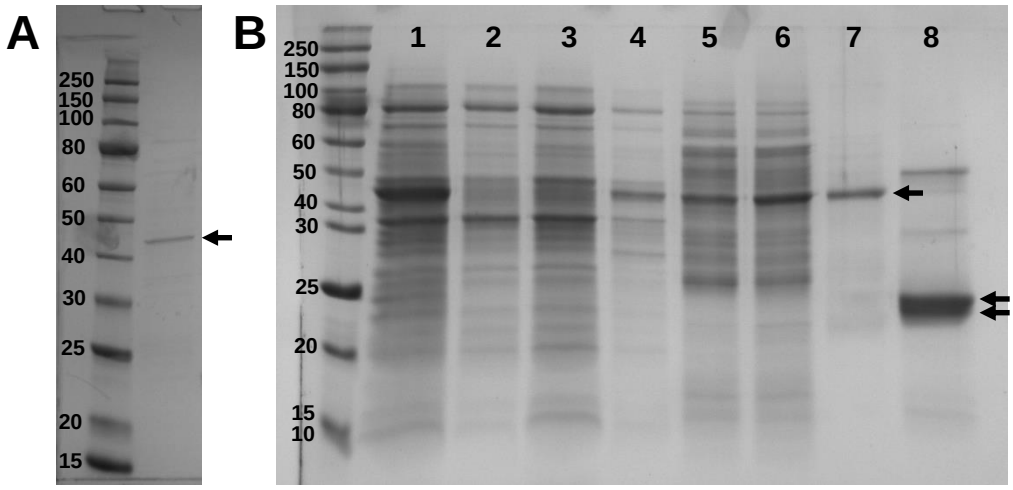


Fig. 3.3 Purification of monomeric Δ19*Tf*NCS. A: Purification of 44 kDa His-rich protein from empty pET vector control. His-trap media was washed with 40 mM imidazole wash buffer (not shown). The major component of elution with 50 mM EDTA was a 44 kDa protein (marked by arrow). **B**: Purification of Δ19*Tf*NCS (L76V). Lane 1: Supernatant; 2-4: column flow through; 5-6: Wash (20-30 mM imidazole); 7: Wash (40-50 mM imidazole); 8: Elution (500mM imidazole). The His-rich 44 kDa protein (Lane 7, marked with an arrow) elutes from the column at 40-50 mM. Monomeric Δ19*Tf*NCS elutes at high imidazole concentrations (lane 8); there appears to be two closely running bands (marked with arrows).

3.2.2 Evidence of cleavage *in vivo*

A more thorough exploration of the purification conditions of Δ19*Tf*NCS led to the identification of two isoforms of NCS; these bound to the nickel column more tightly than the 44 kDa protein and eluted from the column at a higher imidazole concentration (>250 mM imidazole) (Fig. 3.3B, Lane 8; Fig. 3.4). These forms seemed to correspond to the Δ19*Tf*NCS monomer and the N-terminal truncate Δ29*Tf*NCS, the latter previously identified through N-terminal sequencing by Berkner *et al*.[52] Cleavage of recombinant Δ19*Tf*NCS in *E. coli* was not affected by the presence of protease inhibitors. However, it was not clear what conditions cause the cleavage event to occur: a number of research groups have expressed recombinant Δ19*Tf*NCS in BL21(DE3) without reporting purification problems or this cleavage event.[43,51,53]

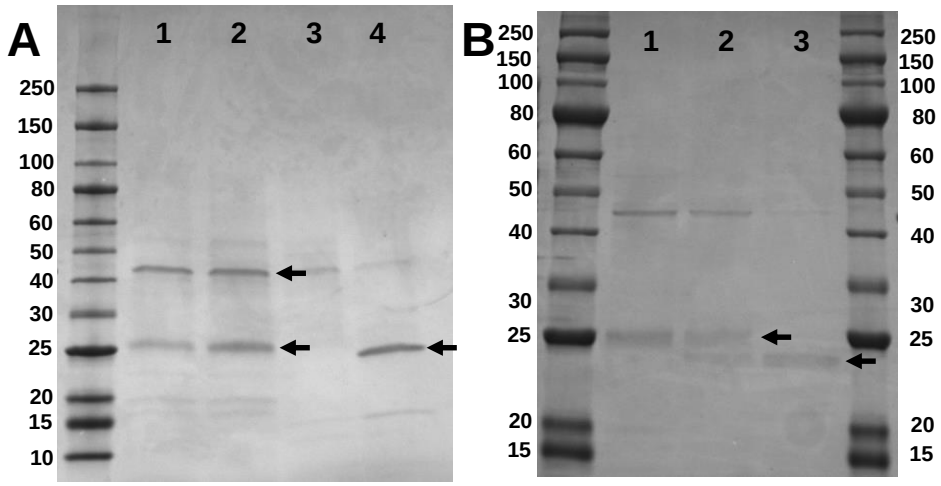


Fig. 3.4 Δ19*Tf*NCS *in situ* cleavage. A: Δ19*Tf*NCS purification. His-Trap column washed with 50 mM imidazole (not shown). Lanes 1-4: Elutions with increasing imidazole concentrations (100-500 mM). His-rich 44 kDa protein observed (lane 1 and 2, marked with an arrow). Possible Δ19*Tf*NCS observed in lane 2, marked with arrow. Possible Δ29*Tf*NCS observed in lane 4, marked with arrow. Gradient gel (4-20%) used. **B**: Verification of two eluents with different sizes. Lane 1 is identical to fraction 2 from (A); lane 3 is fraction 4 from (A); lane 2 is mixture of fractions. Lane 1 is Δ19*Tf*NCS (22.1 kDa but runs as 25.0 kDa); lane 3 is Δ29*Tf*NCS (21.2 kDa but runs as 23.5 kDa).

3.2.3 Δ19*Tf*NCS aggregation

In Δ19*Tf*NCS expressing cells, significant Pictet-Spengler activity was observed in the insoluble fraction after cell lysis (Fig. 3.5A). SDS-PAGE analysis of the insoluble fraction showed the presence of monomeric Δ19*Tf*NCS (Fig. 3.5B). It was notable that the insoluble protein did not form an inactive precipitate, but an active oligomer.

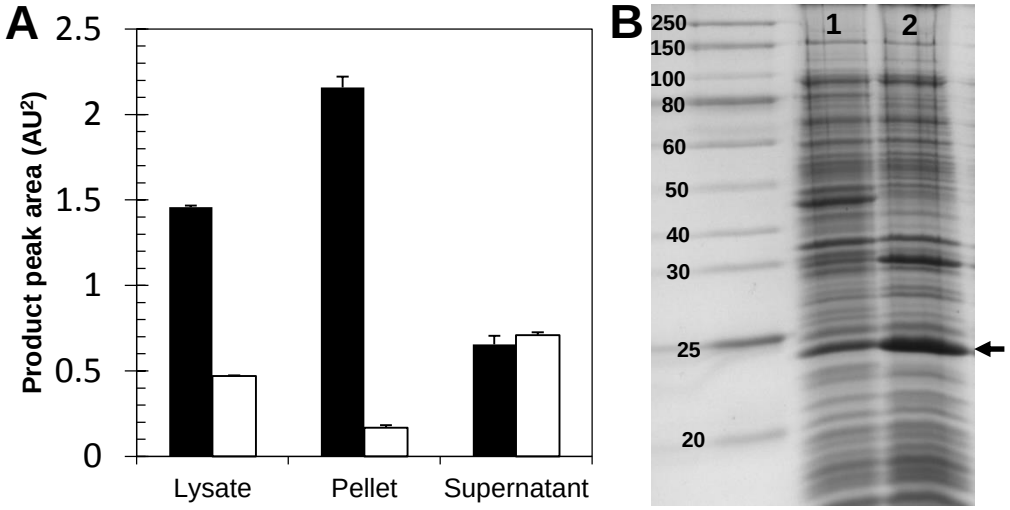


Fig. 3.5 Activity in the Δ19*Tf*NCS insoluble fraction. A: Pictet-Spengler activity in Δ19*Tf*NCS lysate before and after centrifugation. Activity is present in the insoluble fraction. Solid bars: non-boiled samples; clear bars: boiled controls. After centrifugation, the insoluble fraction was resuspended in HEPES buffer for use in assay. Activity in the supernatant control is a result of the non-enzymatic chemical rate. Values are the mean of two measurements, error bars are standard deviations; data is from HPLC peak area. **B**: SDS-PAGE analysis of Δ19*Tf*NCS insoluble fraction: Lane 1: Lysate; 2: Pellet. Δ19*Tf*NCS monomer is enriched in the pellet (arrow).

A resuspension of the insoluble Δ19*Tf*NCS oligomer with lysozyme and BugBuster produced a solution that demonstrated norcoclaurine synthase activity. This activity was removed through centrifugation, showing that the original precipitation was not caused by inefficient lysis of the *E. coli* but was indeed insoluble NCS. If insufficient lysis was occurring, then the second lysozyme/BugBuster step would leave fully soluble NCS that would not be removed by centrifugation. Further resuspension and filtration of the precipitate led to complete loss of Δ19*Tf*NCS, indicating that the protein remained insoluble in particles greater than 0.2 μm (Fig. 3.6). It is worth noting that conditions required for SDS-PAGE preparation (50 mM DTT, 95 °C 5 mins) was sufficient to break any oligomers into Δ19*Tf*NCS monomers.

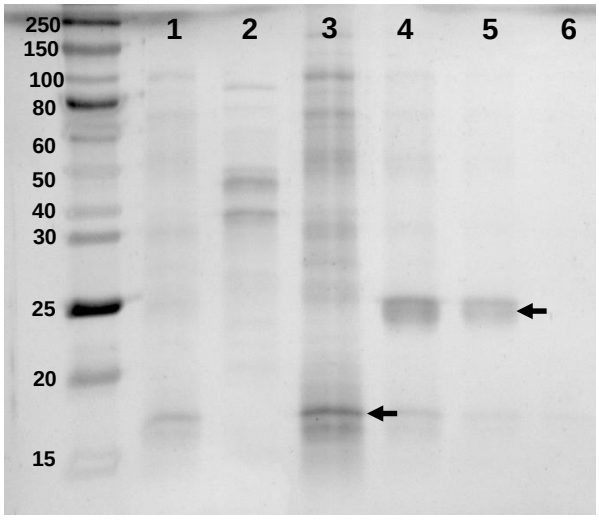


Fig. 3.6 Purification of Δ19*Tf*NCS insoluble oligomers. SDS-PAGE analysis of purification of insoluble oligomers. Lane 1: Pellet; 2: supernatant; 3: post-lysozyme supernatant; 4: post-lysozyme pellet; 5: post-lysozyme pellet resuspension; 6: post-lysozyme pellet resuspension filtered. Lysozyme highlighted in lane 3; Δ19*Tf*NCS insoluble monomer highlighted in lane 4 and 5.

3.3 Expression and purification of Δ29*Tf*NCS

To eliminate problems associated with expressing Δ19*Tf*NCS, a gene encoding Δ29*Tf*NCS was obtained.[52] The Δ29*Tf*NCS protein was highly over-expressed and soluble. The enzyme was purified to 99% purity with bench-top affinity chromatography alone (Fig. 3.7). A wash step with 40 mM imidazole was sufficient to remove all native *E. coli* His-rich proteins and 500 mM imidazole was used to elute the protein. This enzyme was active and was used for subsequent investigations. In a similar manner to the Δ29*Tf*NCS obtained via *in situ* cleavage (Fig. 3.4), the 21.2 kDa protein was observed at around 23.5 kDa on the SDS-PAGE gel.

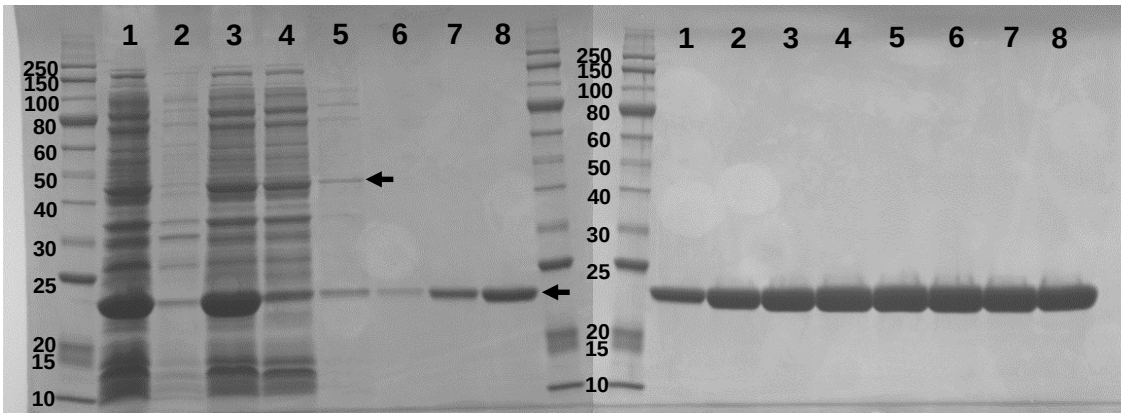


Fig. 3.7 Δ29*Tf*NCS purification. LHS Lane 1: lysate; 2: pellet; 3: flow-through; 4-5: 20 mM imidazole wash; 6-8: 40 mM imidazole wash; RHS Lane 1-8: 500 mM imidazole elution. Native 45 kDa protein highlighted in LHS lane 5; overexpressed Δ29*Tf*NCS monomer highlighted in LHS lane 8.

3.4 *Cj*NCS2 expression

Following on from the work conducted by Pesnot *et al*, *Cj*NCS2 was expressed and purified for use in biotransformations.[57] The gene for *Cj*NCS2 used has no N-terminal truncation and therefore contains the entire N-terminal signal peptide (Table 3.1). Unlike *Tf*NCS, however, this peptide does not preclude purification of the active monomer (Fig. 3.8).[34]

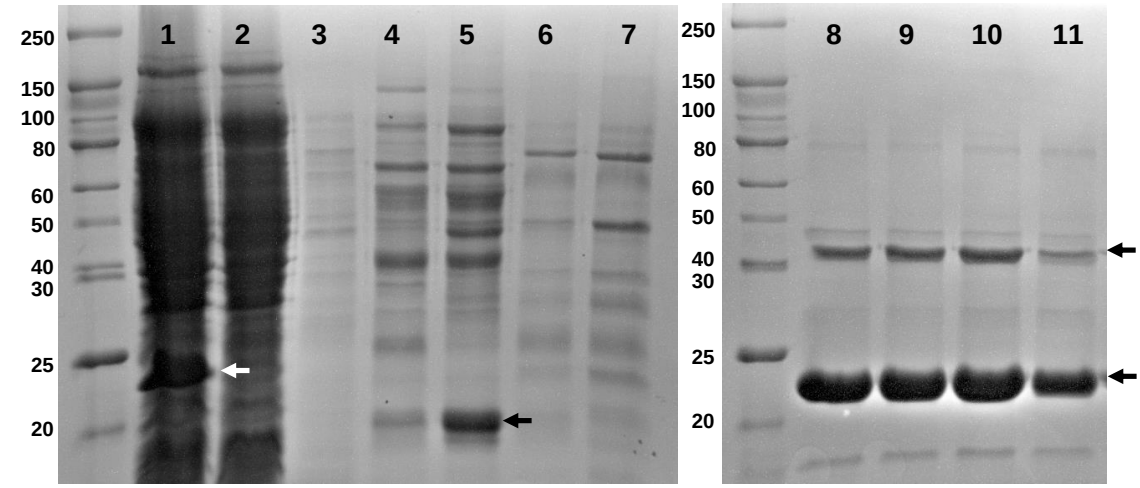


Fig. 3.8 *Cj*NCS2 expression. SDS-PAGE analysis of *Cj*NCS2 purification. Lane 1: Cell lysate; 2-3:Flow-through; 4-5: 30 mM imidazole; 6-7: 100 mM imidazole; 8-11: 200 mM imidazole. *Cj*NCS2 highlighted in Lane 1. Potential N-terminal cleaved *Cj*NCS2 highlighted in lane 5. *Cj*NCS2 and 44 kDa His-rich protein highlighted in lanes 8-11.

*Cj*NCS2 precipitated at high concentrations, behaviour seen with Δ19*Tf*NCS but not Δ29*Tf*NCS. Also, the pelleted insoluble fraction after cell lysis contained considerable quantities of the protein monomer (Fig. 3.9; Lane 1). It is possible that it is the N-terminal extension of *Cj*NCS2 that causes these purification characteristics. Despite this, *Cj*NCS2 monomer purified sufficiently well for application in biotransformations.

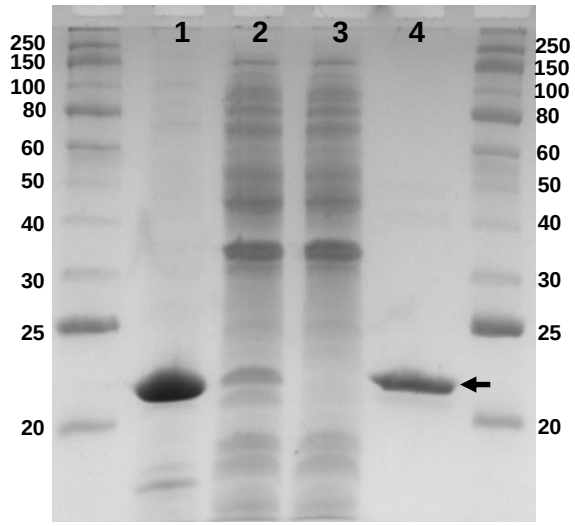


Fig. 3.9 CjNCS2 pellet. SDS-PAGE analysis of *Cj*NCS2 purification. Lane 1: pellet (insoluble fraction); 2: Lysate; 3: Flow-through; 4: Purified *Cj*NCS2. Monomeric *Cj*NCS2 is highlighted (seen in Lanes 1 and 4).

3.5 N-terminal truncates of *Cj*NCS2 have been recently examined by the Sato group.[61] The Δ10*Cj*NCS2, Δ19*Cj*NCS2 and Δ29*Cj*NCS2 proteins appeared to express better than *Cj*NCS2, but Δ19*Cj*NCS2 and Δ29*Cj*NCS2 showed increased insolubility. The protein Δ42*Cj*NCS2 did not appear to express. The Δ29*Cj*NCS2 enzyme was shown to have the highest activity compared to the other proteins and was then tested with a number of different substrates. This study further highlights the importance of using truncations when expressing eukaryotic proteins in *E. coli*. Δ19*Tf*NCS structural analysis

The effect of residues 19 to 28 on the behaviour of recombinant NCS was demonstrated by the differences between Δ19*Tf*NCS and Δ29*Tf*NCS. The former was largely insoluble (though remains active) and was cleaved within *E. coli*. The latter was soluble and highly expressed.

The X-ray crystal structures of Δ19*Tf*NCS provide interesting insight into how the residues 19-28 may affect the behaviour of the protein. In the crystal structures from Ilari *et al* these residues appear to mediate the formation of the crystallographic tetramer.[43] The tetramers feature 2 asymmetric units, each containing a single dimer. The dimer interaction is stabilised by the presence of a β-strand between the two monomers; this causes the formation of a continuous β-sheet through the length of the dimer (Fig. 3.10). The β-strand in question is the N-terminal sequence from a monomer

in the other symmetry related dimer (Fig. 3.10A). This interaction leads to stable, tight tetrameric packing, and ultimately to a dense network of interconnected β-sheets stretching through 2 dimensions of the crystal (Fig. 3.10B).

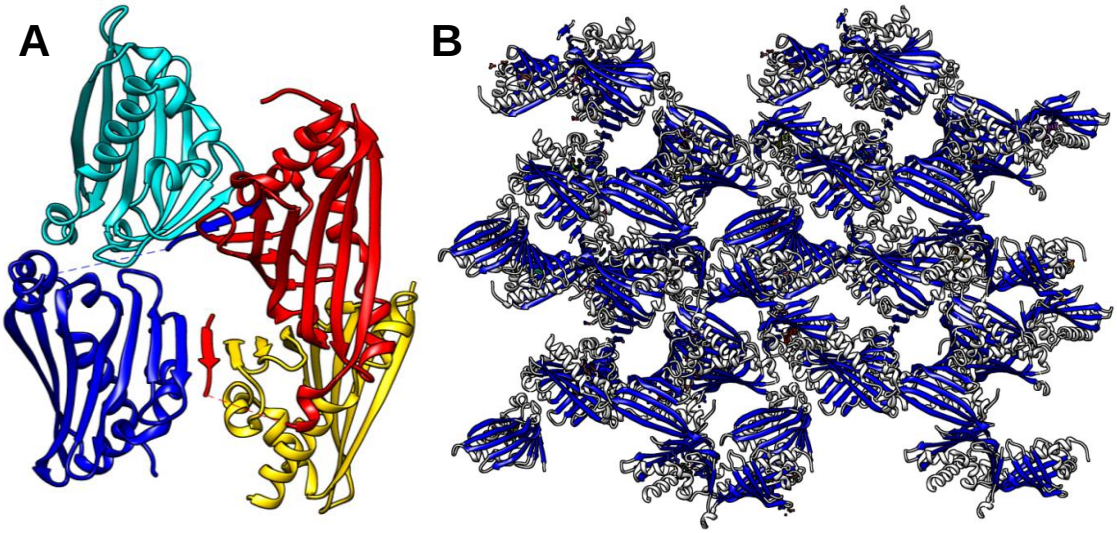


Fig. 3.10 Δ19*Tf*NCS crystal structure packing. A: Cartoon representation of asymmetric units in 2VQ5[43]. Red and blue cartoons show subunits A. Cyan and yellow show subunits B. Note the subunit A N-terminal β-strand that mediates the interaction between subunits A and B in the adjacent dimer. **B:** Cartoon representation of crystal packing interactions across four crystallographic unit cells (expanded in *y* and *z* dimensions). β-sheets are coloured blue, loops and helices are white with outlines. Notice the continuous β-sheets formed across unit cells in the crystal due to the N-terminal β-strand.

Computational analysis of the Δ19*Tf*NCS sequence demonstrates that the N-terminal residues have a strong tendency toward forming β-strands. Betascan analysis shows a strong probability of β-strand formation in this region: it strongly predicts a β-strand involving residues 3 to 8 (Fig. 3.11).[120] As this N-terminal β-strand is not part of the general globular fold, it is highly probably that it will form β-strand interactions in a promiscuous manner, such as with other monomers to form oligomers.

It is possible that the N-terminal β-strands mediate the polymerisation of recombinant Δ19*Tf*NCS to form an insoluble, active precipitates. The oligomer may exist as a 'beads-on-a-string' type structure: a long continuous amyloid-like β-sheet running through its length, monomers connected *via* the N-terminal β-strand, NCS active sites on each surface. This would be a similar structure to that observed in the crystal structure packing (Fig. 3.10B). For biotechnological applications of NCS, these active oligomers may be of some use as they are essentially immobilised enzymes. It may even be possible to graft different enzymes onto the structure of such a structure, making it a more general tool for immobilising biocatalysts. Further characterisation of this insoluble Δ19*Tf*NCS is required to understand its structure and potential application.

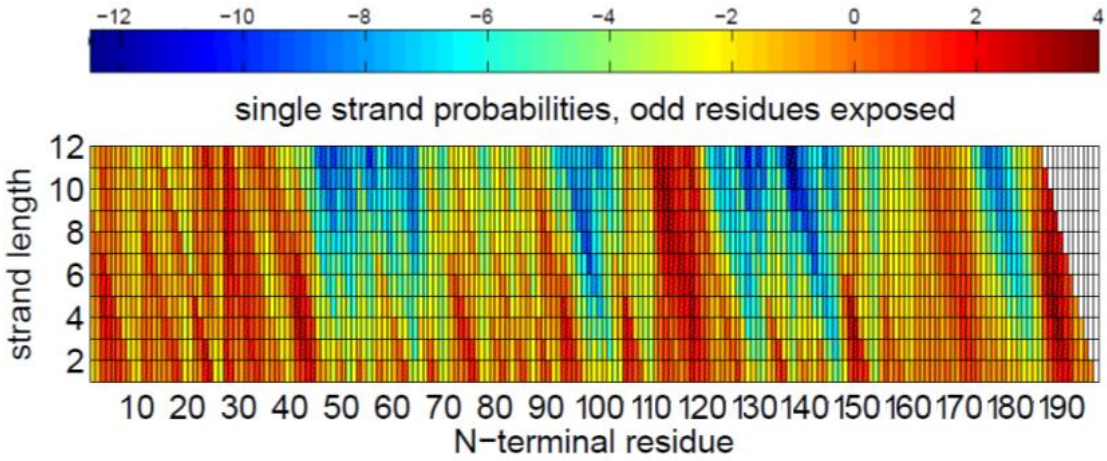


Fig. 3.11 Betascan analysis of Δ19*Tf*NCS. Heat-map of β-strand probabilities (odd residues exposed). Red shows high probability of β-strand formation, blue a low probability. Residues 2 to 8 of the expressed protein have a high probability of forming a β-strand (7 residues long). However, these residues do not form part of the globular structure of the protein and so may form promiscuous intermolecular connections.

3.6 Discussion

It is unclear what role the N-terminal sequence plays in the native enzyme and whether it remains attached to active NCS *in planta*. It is known from the original cDNA derived sequence that the native protein includes a longer N-terminal extension that acts as a signal peptide, causing relocation to an organelle within the plant. The first 19 residues at the N-terminus of *Tf*NCS (and *Cj*NCS2) are predicted to be a signal peptide by SignalP, an algorithm capable of predicting signal peptides (Table 3.1).[131] The enzymes *Ps*NCS1 and *Ps*NCS2, active in *Papaver somniferum*, also have a similar N-terminal extension, though, *Pb*NCS from *P. bracteatum* does not.[44] Signal peptides cause proteins to localise to the endoplasmic reticulum, often before transportation into a vesicle for secretion or for localisation into other membrane bound organelles.

The crystal structure of Δ19*Tf*NCS shows that a number of residues on the N- and C-termini of the enzyme—that are not accounted for by signal peptide prediction—are not part of the main globular fold of the enzyme. At the N-terminal, the main globular fold of the protein only begins with the β-strand beginning at residue 42. In subunit A residues 19-27 form the crystal packing mediating β-strand (see section 3.5), 28-33 are not present (disordered) and 36-41 form a small α-helix with limited contacts with the main fold. Residues 19-41 of subunit B are disordered and not visible in the crystal structure (Table 3.1).

The situation at the C-terminal is similar; the globular fold ends around residue 193, but there are around 20 subsequent residues not present. In subunits A and B residues 197-217 and 194-217 are disordered, respectively. Although it is common for the ends

of proteins to be disordered in crystal structures, in this instance significant lengths of the protein are unstructured and extend beyond the main fold. It is unlikely that these portions of the enzyme are required for activity, as the enzyme reaction requires no co-factor recycling and there is little suggestion of any significant conformational changes occurring in catalysis.[54] Instead these sequences may play a role in the localisation of NCS and its association with other proteins and membranes.

By using an NCS-GFP fusion, with GFP fused to the C-terminus of *Ps*NCS2, it was observed in *P. somniferum* that after expression NCS is transported to the central vacuole, via the endoplasmic reticulum.[44] In general, vacuolar localisation is directed by a propeptide, which can be found either at the N- or C-terminus of the protein. When at the N-terminus, the vacuolar directing propeptide sequence typically sits between an N-terminal signal peptide and the C-terminal protein sequence.[132] There are approximately ten residues between the signal peptide and the globular NCS protein—these residues are possibly the vacuolar-targeting propeptide. This would be an analogous case to the Berberine Bridge Enzyme (BBE), which has both a signal peptide and vacuolar sorting element on the N-terminus. BBE is localised to the same tissues and cellular compartments as NCS.[18,44] It is notable that a different signal peptide prediction algorithm, SOSUIsignal, predicts the N-terminal peptide of *Tf*NCS to extend for 27 residues: this prediction incorporates the proposed vacuolar localisation sequence (Table 3.1).[133] The C-terminus of NCS may also have a sorting role: *Ps*NCS1, *Ps*NCS2 and *Pb*NCS have significant C-terminal extensions of around 30 residues beyond the predicted domain; *Tf*NCS and *Cj*NCS2 have shorter extensions of 16 and 6 residues respectively.[44]

This close association of NCS and other BIA enzymes with the membrane and membrane bound organelles suggests that there is a role for the transmembrane trafficking of pathway intermediates. This prediction has implications for NCS, its background reaction and attempts to use it in recombinant *in vivo* scenarios: in plants the substrates may be synthesised in close proximity to an enzyme that catalyses the next step, increasing the rate (through the increase in local concentration) and reducing the chance of background catalysis.

Table 3.1 Sequence and structural analysis of *Tf*NCS N-terminus. Signal peptide prediction using SignalP[131] and SOSUISIGNAL.[133] X-ray structure analysis using subunits A and B of 2VQ5.[43] Key for structure: n/p, electron density not present; α, α-helix; β, β-strand. For *Tf*NCS SignalP predicts a 19 residue signal peptide (for ER targeting). SOSUISIGNAL predicts a 27 residue signal peptide. The 8 residue difference can be accounted for by vacuolar propeptide signal. **X-ray structure analysis** shows main globular fold only begins at residue 42. Prior to residue 42 much of the polypeptide is not present due to disorder. Subunit A forms an N-terminal β-strand interaction and a short α-helix prior to the main fold. Lines in table show N-terminal starting points of the truncates Δ19*Tf*NCS and Δ29*Tf*NCS. SignalP and SOSUISIGNAL are both signal peptide predictors but use different algorithms.

*Tf*NCS		Peptide prediction		X-ray 2VQ5	
Residue		SignalP	SOSUI	A	B
1	M				
2	M				
3	K				
4	M				
5	E				
6	V				
7	V				
8	F				
9	V				
10	F				
11	L				
12	M				
13	L				
14	L				
15	G				
16	T				
17	I				
18	N				
19	C				
20	Q			β	n/p
21	K			β	n/p
22	L			β	n/p
23	I			β	n/p
24	L			β	n/p
25	T			β	n/p
26	G			β	n/p
27	R			β	n/p
28	P			n/p	n/p
29	F			n/p	n/p
30	L			n/p	n/p
31	H			n/p	n/p
32	H			n/p	n/p
33	Q			n/p	n/p
34	G			loop	n/p
35	I			loop	n/p
36	I			α	n/p
37	N			α	n/p
38	Q			α	n/p
39	V			α	n/p
40	S			α	loop
41	T			loop	loop
42	V			**Main fold**	
43	T			**Main fold**	
44	K			**Main fold**	

There is also a fascinating role for protein aggregation in the intracellular trafficking of enzymes. Aggregation is accepted to have a role in sorting proteins into vacuoles.[134,135] For example, secreted proteins may be in dense vesicles as aggregates prior to secretion.[134] BIA enzymes are expected to be synthesised in companion cells before being secreted into sieve elements, so aggregation-sorting is one possible mechanism for this.[136] The identification of the insoluble Δ19*Tf*NCS certainly shows that the N-terminal extension increases the aggregation propensity of the enzyme, and the observation of activity even raises the possibility that the enzyme could be an active oligomer in plants. The ideal experiments to elucidate the roles of the protein termini would be to introduce genes coding for fluorescently labelled fusion proteins with different peptide extensions on the C- and N-termini. The localisation of these proteins could then be followed visually.

3.7 Summary

The original enzyme intended for use, Δ19*Tf*NCS was chosen as it is well studied and structural information is available. However, it was prone to aggregation and unsuitable for use. Removing 10 residues from the N-terminus was sufficient to correct expression and purification problems. The resulting enzyme Δ29*Tf*NCS expressed with high yields, was soluble and was easy to purify. It was therefore an ideal enzyme to use in further mechanistic and biocatalytic studies. The N-terminus of NCS that effects enzyme solubility is likely to be involved *in planta* with cellular localisation and membrane association.

4 Chapter 4: NCS Mechanism

4.1 Introduction

4.1.1 The Pictet-Spengler reaction

The Pictet-Spengler reaction involves the formation of an imine (Schiff base) followed by an electrophilic aromatic substitution. The two classic Pictet-Spengler reactions are tetrahydroisoquinoline (THIQ) synthesis and tetrahydro-β-carboline synthesis (Fig. 4.1).[24] In plants, these reactions are catalysed by NCS and strictosidine synthase (STR) respectively.

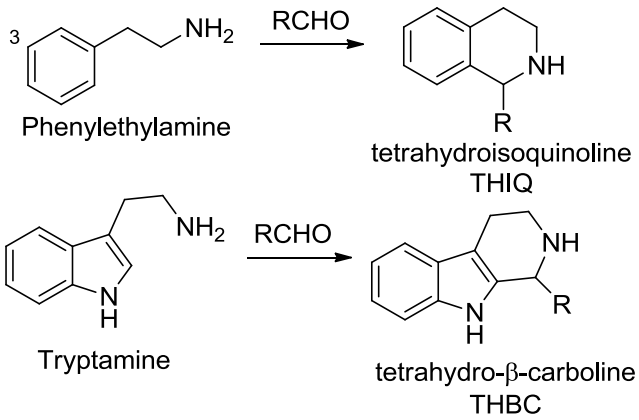


Fig. 4.1 Two classic Pictet-Spengler reactions. Pictet-Spengler reaction between an aldehyde and a phenethylamine provides a tetrahydroisoquinoline. Chemically this can be catalysed by acids. A 3-hydroxy group on the amine allows catalysis by phosphate or NCS. The Pictet-Spengler condensation between an aldehyde and tryptamine leads to the formation of tetrahydro-β-carboline. This can also be catalysed by acid. In the plant indole alkaloid pathway this step is catalysed by strictosidine synthase (STR) using secologanin as the aldehyde.

STR catalyses the Pictet-Spengler reaction between tryptamine and the terpene derived aldehyde secologanin. The product, strictosidine, is the precursor to all indole alkaloids. Maresh *et al* were able to define the mechanism of STR using methods including pH profiling, mutations, isotope effects, structural studies and theoretical calculations.[79] The key enzymatic residue of the enzyme was determined to be Glu-309, which catalyses both the formation of the Schiff base and the electrophilic aromatic substitution (Fig. 4.2). Through kinetic isotope effects, the final deprotonation step was identified as the rate limiting; this was the same in the chemical reaction, suggesting the enzyme does not significantly alter the reaction mechanism. Free energy calculations also ruled out the presence of a spiroindolenine intermediate.[79]

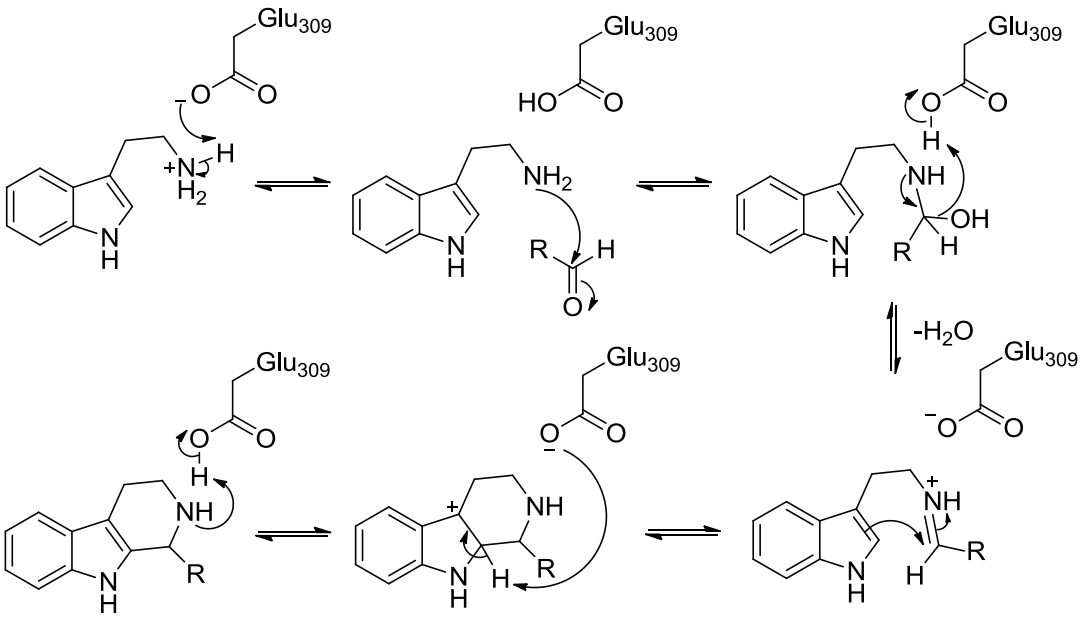


Fig. 4.2 Catalytic mechanism of strictosidine synthase. As proposed by Maresh *et al.*[79]

In plants NCS catalyses the condensation of dopamine (3,4-dihydroxyphenethylamine) and 4-HPAA (4-hydroxyphenylacetaldehyde). Investigation by Luk *et al* into the mechanism of NCS—through the exploration of its substrate scope, kinetics and isotope effects—determined an overall mechanistic scheme for the enzyme (Fig. 4.3).[51] Turnover of 3-hydroxyphenylethylamine suggested the electronic effect of the 4-hydroxyl group found on dopamine was not important, thus ruling out a spirocyclic intermediate. On the other hand the 3-hydroxy group was vital for catalysis. Isotope effects were not definitive: unlike in STR, the irreversible rearomatisation step was found to be only partially rate determining. The prior substrate binding, imine formation or the electrophilic addition played an unquantified rate limiting role. Furthermore, without structural information, no protein residues could be assigned any roles in the mechanism.

Structural information on NCS was initially provided by an NMR study.[54] A homology model of Δ29*Tf*NCS was built based on NMR nuclear Overhauser effects (NOEs), chemical shifts and the Bet-v-1 structure. Observation of chemical shift perturbations from the titration of NCS with substrates led to the identification of a binding site centred on the residue Glu-103. However, relatively small chemical shift changes in titrations prevented assignment of catalytic roles to amino acid residues.

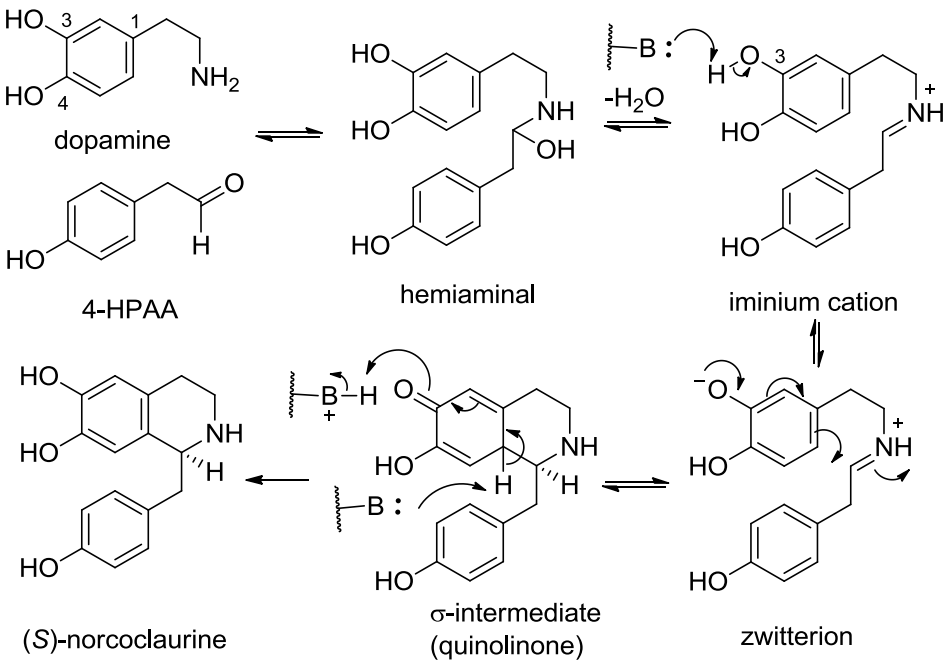


Fig. 4.3 Overall mechanism of NCS catalysis. As proposed by Luk *et al*.[51]

4.1.2 HPAA-first mechanism

Solution of the Δ19*Tf*NCS crystal structure enabled the identification of the enzyme structure and active site residues (Fig. 4.4.[43,55] The NCS active site is positioned in a catalytic tunnel and contains three charged and one polar reside: Tyr-108, Glu-110, Lys-122 and Asp-141 (Fig. 4.4B). The *holo* crystal structure contained dopamine and a non-productive aldehyde. An enzyme mechanism was proposed using evidence from the *holo* structure ligand binding modes, with support from a number of active site mutations (Fig. 4.5).[43,56] This mechanism involves the binding of 4-hydroxyphenylacetaldehyde (HPAA) to the enzyme prior to dopamine (the HPAA-first mechanism); substrate binding order was previously proposed on the basis of kinetic analysis.[15]

In the HPAA-first mechanism (Fig. 4.5), the aldehyde is activated by Lys-122, enabling nucleophilic attack from dopamine (step a). Loss of water, forming the iminium, is catalysed by the Lys-122 (b). Next, electrophilic addition and stabilisation of the resulting quinolinone intermediate is aided by Tyr-108 (d). Finally, rearomatisation is catalysed by Glu-110, forming the product (e). The behaviour of NCS mutants supported elements of this mechanism. The variant K122A eliminated activity, and the mutations Y108F and E110A reduced activity.[43]

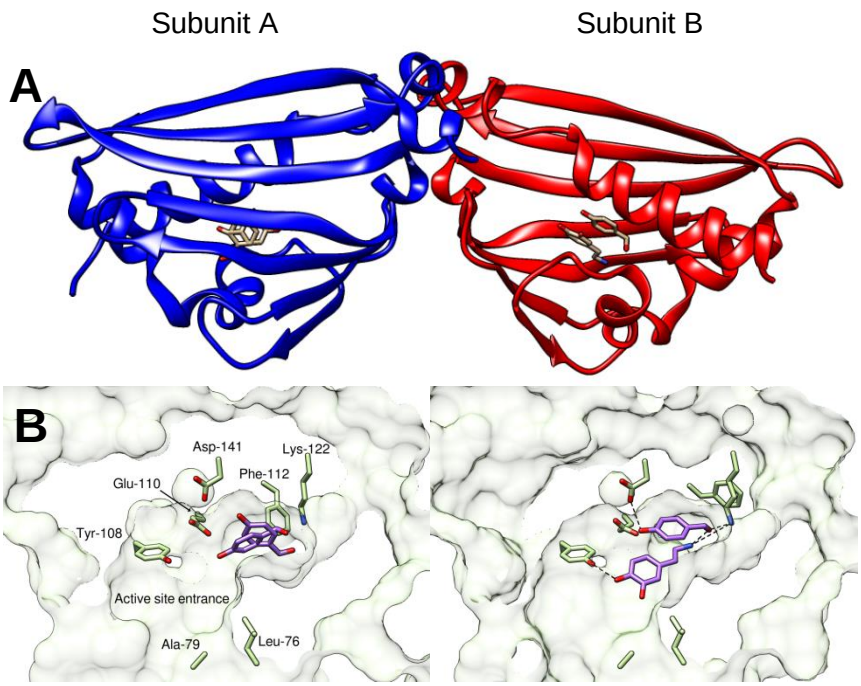


Fig. 4.4 NCS X-ray structure. A. NCS globular structure. Δ19*Tf*NCS *holo* crystal structure (2VQ5). Subunit A (blue) and subunit B (red) backbones depicted as cartoons. Bound ligands depicted as sticks. First ten N-terminal residues of subunit A removed. **B. Active site cross section.** Sections of the protein solvent accessible surface are shown and selective active site residues are represented by yellow sticks. Ligands are represented as purple sticks. Orientation of protein relative to overall structure seen in **Fig. 4.4**A is similar to that observed for subunit B. Figure adapted from Lichman *et al*.[137]

There are a number of problems with the HPAA-first mechanism. Firstly, the intermediate iminium is proposed to undergo a rearrangement, which includes a *cis-trans* isomerisation and the rotation of the entire aldehyde derived portion of the intermediate (Fig. 4.5, step c). This step is vital for the subsequent rate-determining cyclisation and deprotonation, which ultimately establish the chirality of the product. However, there is no evidence for it; the *holo* crystal structure only provides evidence for the initial arrangement of the substrates. Such a large rearrangement seems unlikely to occur in the confines of the active site, especially with bulky aldehydes which have been shown to be accepted by the enzyme (see section 4.1.4).

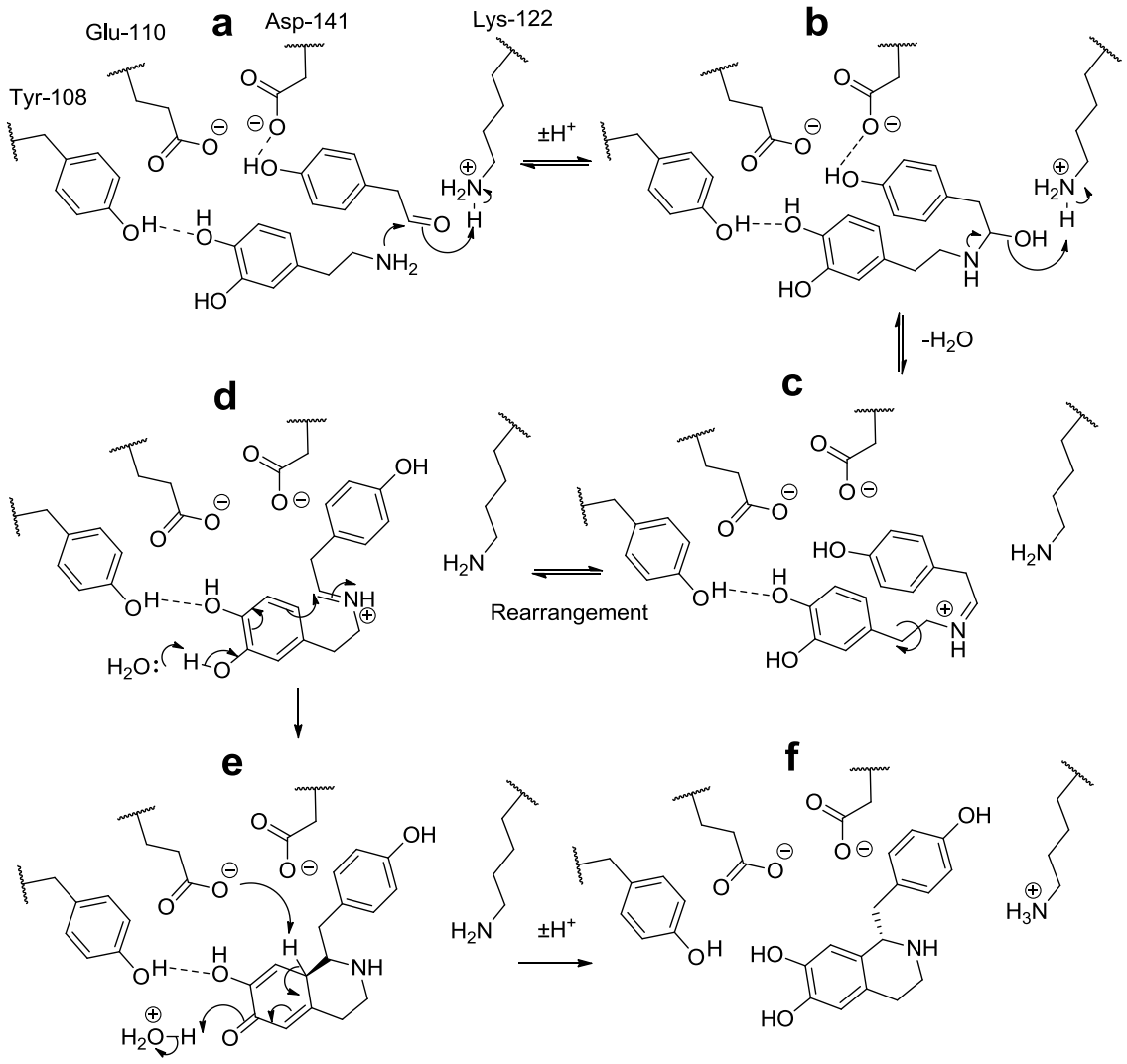


Fig. 4.5 HPAA-first mechanism. As proposed by Ilari *et al*[43] and Bonamore *et al*[56] on the basis of crystal structure information.

The chemical basis of the mechanism also has limitations. No residue was proposed for interaction with the 3-hydroxy group in the electrophilic aromatic substitution steps. A partial positive charge is expected on the oxygen as the hydroxyl group stabilises the positive charge on the aromatic ring.[51] To achieve this, a base is required to deprotonate or stabilise the 3-hydroxy group. In this mechanism, water was proposed to perform this role, with enzymatic stabilisation provided by Tyr-108 interaction with the 4-hydroxyl group (Fig. 4.5, steps d and e). However, water is not sufficiently basic to deprotonate this position (pKb = -1.7); a basic active site residue is necessary.[138] It is possible that a residue may interact with the water to increase its basicity, in a 'domino' type mechanism. However, the only charged residue that is a reasonable position to interact with the water is Glu-103, which is not conserved in all NCSs. Finally, the H-bond between the quinolinone and Tyr-108 would not provide the electrostatic

stabilisation to the intermediate as proposed, and cannot account for the successful reaction when the 4-hydroxy moiety is not present.

4.1.3 X-ray structure

The HPAA-first mechanism was based on and supported by the observed ligand binding modes in the *holo* NCS crystal structure (2VQ5), which had a resolution of 2.1 Å.[43,56] This structure was obtained by soaking NCS crystals in dopamine and 4-hydroxybenzaldehyde (4-HBA), an electron deficient non-productive 4-HPAA analogue. Subunit A of the *holo* crystal structure is shown to bind 4-HBA in two different conformations, whilst subunit B binds both dopamine and 4-HBA simultaneously (Fig. 4.4B, Fig. 4.6).

The ligands modelled into the active site of the *holo* crystal structure do not fit the electron density convincingly (Fig. 4.6). The density observed in subunit A does not support the presence of two conformations of 4-HBA, as modelled. In subunit B, the methylene from dopamine does not fit the density, and there is no electron density in the area where the oxygen atoms must locate. Generally, the electron density appears to have insufficient definition to describe the identity of ligands present. This observed ambiguity may be a result of low occupancy or a heterogeneous sample.

The difference in the binding behaviour of the two subunits was possibly a consequence of the different rotameric conformations of Phe-112 (Fig. 4.7A). Analysis of the NCS structure shows that whilst the Phe-112 in conformation A has favourable Ramachandran and rotamer angles, conformation B is less favoured: the residue is not in the favourable region of the Ramachandran plot, and the rotamer is unfavourable (analysis by MOLPROBITY).[121] Interestingly, the unrefined electron density of subunit B may show that conformations A and B are both partially occupied in subunit B (Fig. 4.7B). In the final published solution the density showing evidence of conformation A in subunit B was modelled as a water molecule.*

* X-ray structure electron density analysis performed in collaboration with Dr Altin Sula and Prof Nicholas Keep

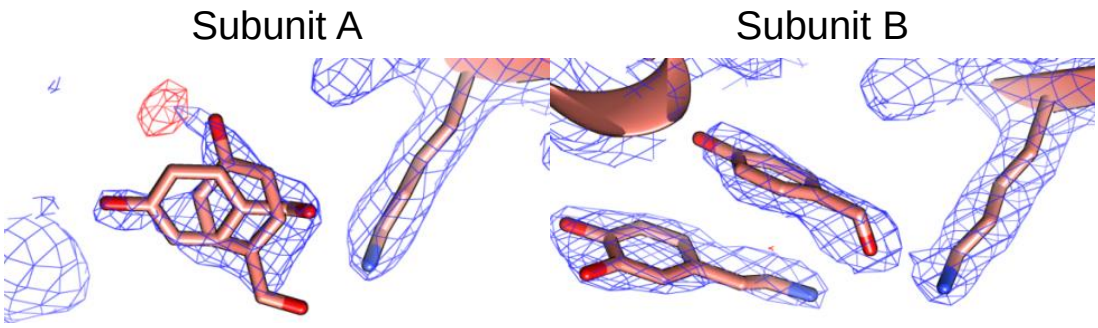


Fig. 4.6 Ligands in the NCS *holo* crystal structure. Electron density maps (blue, $2F_o$ - F_c, contoured at 1σ) and difference map (red, $2mF_o$ - DF_c, contoured at 3 σ) of 2VQ5, with Lys-122 and ligands represented by sticks. Subunit A contains a 4-HBA molecule in two conformations. Subunit B contains a 4-HBA and dopamine ligand. Note that in both subunits the density does not fit around certain ligand atoms. Figure adapted from Lichman *et al*.[137]

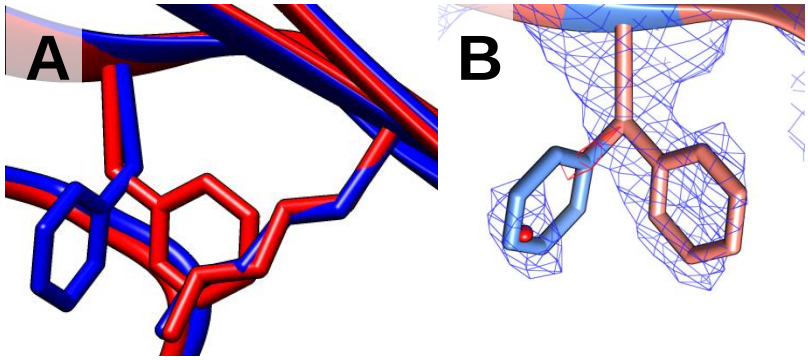


Fig. 4.7 Rotamers of Phe-112. A. Structural alignment of subunits A (blue) and B (red). Side chains of Phe-112 and Lys-122 represented as sticks. Backbone represented as cartoon. **B.** Subunit B original Phe-112 and water depicted in red. Remodelled Phe-112 side chain to fit water molecule density in blue (modelled with Coot and PHENIX). 2VQ5 Electron density maps (blue, $2F_o$ - F_c, contoured at 1σ) and difference map (red, $2mF_o$ - DF_c, contoured at 3 σ).*

The differences in the subunits' active sites may have been a knock-on effect from the tight crystal packing. The asymmetric dimer was held together by a non-native β-strand formed from the first nine N-terminal residues of subunit A in the adjacent dimer (see section 3.5). This N-terminal sequence is not necessary for enzyme activity, and is possibly part of a cleaved signal peptide.[52,54]

Overall, the binding modes derived from the *holo* structure appear to be speculative on the basis of the observed electron density. This, in turn, adversely affects the reliability of the HPAA-first mechanism.

4.1.4 Kinetic assays

The Pictet-Spengler tetrahydroisoquinoline synthesis reaction can be catalysed by phosphate.[59] Efficient catalysts for this reaction include inorganic phosphate (i.e. potassium or sodium salts) as well as organic phosphates such as uridine

* X-ray structure electron density analysis performed in collaboration with Dr Altin Sula and Prof Nicholas Keep

monophosphate and glucose-1-phosphate. The phosphate catalyses reactions between phenylethylamines and aldehydes, but the amines must have a 3-hydroxy group for the reaction to proceed; this is similar to NCS substrate requirements.. The catalytic mechanism is currently unknown, but the phosphate is hypothesised to interact with the 3-hydroxy group at the aromatic cyclisation step, stabilising a build up of positive charge on the oxygen probably through deprotonation. The phosphate may also directly attack the imine, forming a covalent intermediate.

Prior to the discovery of the phosphate mediated reaction, NCS assays were typically conducted in phosphate buffer[43,51] or Tris (tris(hydroxymethyl)aminomethane) buffer.[15,33,34] Although in all cases care was taken to account for the background reaction versus the enzymatic activity, the presence of phosphate and thus a high background reaction invalidates the observed kinetic parameters. The Pictet-Spengler reaction has multiple steps: hemiaminal formation, iminium formation, aromatic electrophilic addition and rearomatisation. Phosphate must be able to catalyse any of these steps, so although in phosphate NCS can raise conversions above background levels, its mechanistic contribution to this increase is unknown. Furthermore, the increased depletion of substrates by the phosphate reaction adversely affects enzyme time course assays. The reaction is also incompatible with Tris buffer as aldehydes can react with Tris to form oxazolidines, affecting the results (Fig. 4.8).

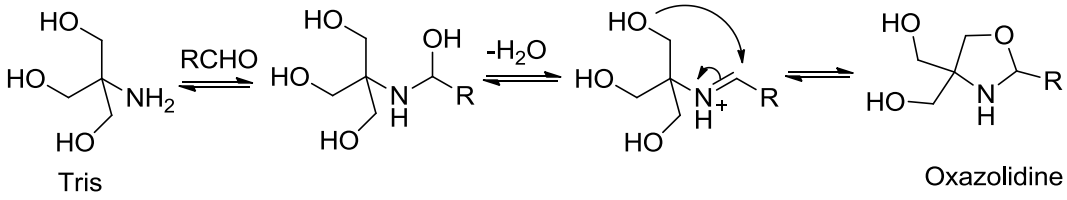


Fig. 4.8 Reaction between Tris and aldehydes.

The key difference between the enzyme reaction and the phosphate catalysed reaction is the stereoselectivity. In an effort to account for the background reaction and just observe enzymatic products, a circular dichroism (CD) spectroscopy method for measuring kinetic activity was developed.[43,51] As only enantiomeric excesses are optically active, chiral enzyme products can be observed by CD spectroscopy, but racemic products cannot. Though it is true that in the NCS CD kinetics method any products detected by CD would be enzyme products, only the reaction step that establishes the chiral centre would actually need to be enzyme catalysed.

In the NCS reaction, there are two factors which could influence chirality. First, the face of the pro-chiral iminium on which the aromatic electrophilic addition step occurs (which

itself is influenced by the iminium geometric isomer present). Second, a spatial preference for deprotonation in the final rearomatisation step. The electrophilic aromatic addition causes the formation of two chiral centres: the 1 position, which is maintained in the product, and the 8a position, which is deprotonated in the final reaction step. If this final irreversible deprotonation/rearomatisation step is enantioselective, then an enantiomeric excess may be formed in the product, regardless of the stereoselectivity of previous reversible steps (Fig. 4.9). The result of all this is that phosphate may catalyse three out of four steps in the reaction, and the enzyme only one step, and a CD signal would still be observed.

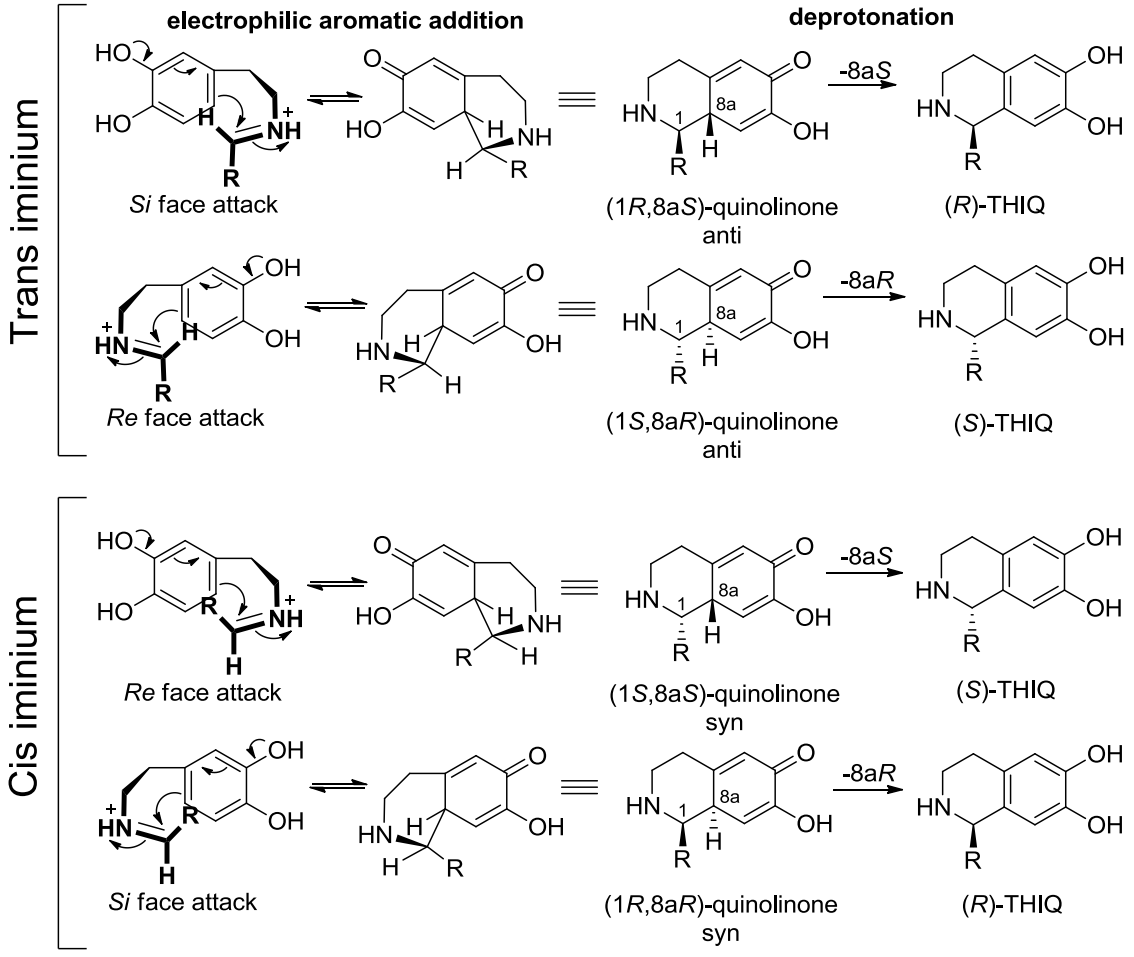


Fig. 4.9 Origin of chirality in NCS reaction. Four quinolinone isomers can be formed by electrophilic aromatic addition, depending on the isomer of the iminium and the face of the iminium that is attacked. This step is reversible—so the final irreversible enantioselective deprotonation may also influence chirality of the final product.

A further problem with the CD method was the establishment of the molar ellipticity constant used for calibration of the CD signal. A reaction with dopamine, 4-HPAA and NCS in phosphate buffer was observed to go to completion (as measured by proton NMR), the product was assumed to be entirely enzymatic (i.e. a single enantiomer) and so the CD signal of this product was used as a calibration. However, as this reaction

was performed in phosphate buffer, a significant proportion of this product was racemic and the result of non-enzymatic catalysis. This was demonstrated by the fact that the product was found to be racemic when purified. Any optical activity observed by CD was likely to be the result of a small enantiomeric enrichment caused by the presence of the enzyme. Overall, problems with the calibration constant and the presence of phosphate buffer in the reaction invalidate kinetic parameters determined with this sophisticated method.

Kinetic assays conducted with these problematic methods led to the observation of positive cooperativity with respect to dopamine.[15,33,34,51] This, coupled with chromatography data on the isolated plant enzyme,[15] led to the prediction that NCS was dimeric in solution. However, investigations Δ29*Tf*NCS by NMR and size-exclusion chromatography demonstrated that the enzyme is predominantly monomeric at concentrations around 10 μM.[54] Kinetic assays (with Tris buffer) in the form of Hanes plots, with norlaudanosoline as a product inhibitor, led to the prediction that 4-HPAA binds to NCS prior to dopamine.[15] This substrate binding order was supported by the X-ray *holo* structure[43] but is not yet satisfactorily accounted for mechanistically.

4.1.5 Substrate scope

NCSs from *Thalictrum flavum* (*Tf*NCS) and *Coptis japonica* (*Cj*NCS2) have been shown to have a broad aldehyde substrate scope, enabling the enzymatic formation of diverse (*S*)-tetrahydroisoquinolines (THIQs).[53,57,61] Both enzymes accepted most phenylacetaldehydes but did turn over α-substituted aldehydes well. Particularly notable was the turnover of a linear aliphatic aldehyde by *Cj*NCS2[57] and the large 1-napthylacetaldehyde by *Tf*NCS,[53] as these substrates have very different properties to the natural substrate 4-HPAA. In contrast to the aldehyde, little variation in dopamine structure was tolerated—the 3-hydroxy moiety is crucial for successful turnover.

As previously described, the HPAA-first mechanism was proposed based on the arrangement of the substrates in subunit B (Fig. 4.5). A major drawback of the HPAA-first mechanism (made evident by these substrate scope results) is that it cannot account for the fact that NCS reacts with a wide variety of aldehydes. First, the proposed interaction between Asp-141 and the phenol of 4-HPAA would not be possible with many aldehydes. Second, it is unlikely that the active site can incorporate large aldehydes (such as 1-napthylacetaldehyde[53]) and dopamine simultaneously and then rotate the intermediate in an important mechanistic step (Fig 5, step c).

4.1.6 Dopamine-first mechanism

The inability of the HPAA-first mechanism to fully account for NCS activity was noted during the investigation into the *Cj*NCS2 substrate scope.[57] This led to a re-examination of the enzyme mechanism through docking studies of reaction intermediates into the *Tf*NCS crystal structure. Docking studies revealed novel binding modes of reaction intermediates. The lowest energy binding modes showed the dopamine portion of the reaction intermediates buried in the enzyme active site and the aldehyde R-group exposed to the solvent. This arrangement can account for the large variety of aldehydes accepted by the enzyme. It also accounts for low activity with substituted amines such as metaraminol, as the amine binds in a sterically hindered pocket.[57] The docking investigation led to the proposal of a novel mechanism for the enzyme. In this mechanism the Lys-122 plays a key role as the base that deprotonates the 3-hydroxy group to trigger electrophilic aromatic addition. Glu-110 catalyses the rearomatisation step, and Asp-141 is proposed to interact with the charged nitrogen on the substrate and intermediates, holding them in place and providing electrostatic stabilisation (Fig. 4.10).

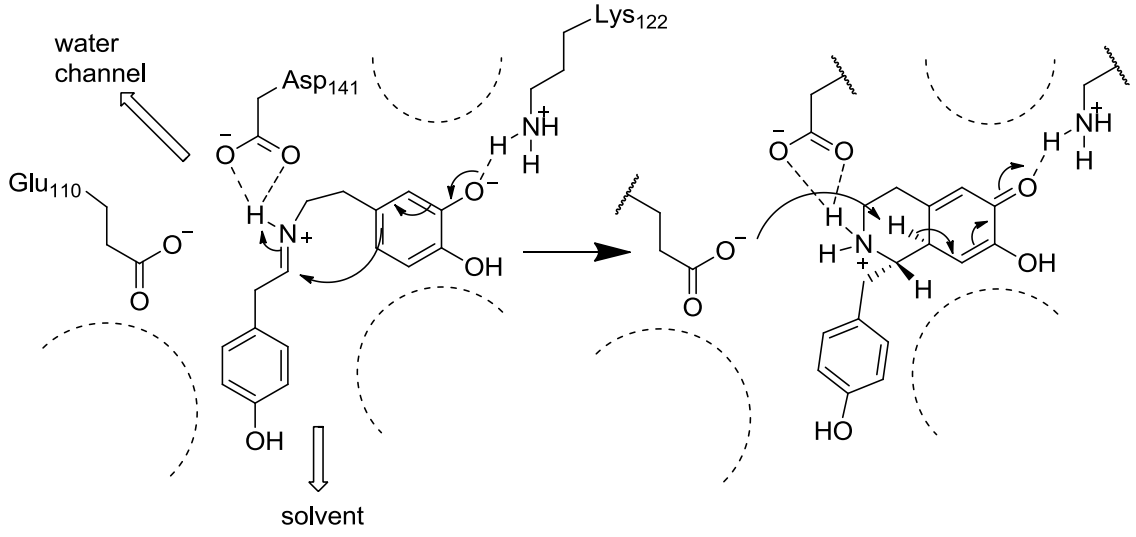


Fig. 4.10 Dopamine-first mechanism. As proposed by Pesnot *et al*.[57] The key feature of the mechanism is the arrangement of the intermediate so the aldehyde R-group can be partially solvent exposed. The residues Lys-122 and Glu-110 are proposed to be important in the mechanism. Dotted semi-circles show regions with hydrophobic residues.

This 'dopamine-first' mechanism appears to account for the wide aldehyde substrate scope of NCS reaction more effectively than the HPAA-first mechanism. However, in order to more fully compare the two mechanisms, the enzyme must be examined more closely. Therefore, an investigation was conducted into the enzyme using computational and experimental methods, focussing on the effect of mutations on the mechanism and substrate specificity.

4.2 Computational analysis

4.2.1 Molecular dynamics and Phe-112

As described in section 4.1.3, the crystal structure 2VQ5 features two subunits, A and B that primarily differ in the nature of N-terminal interactions. The only significant difference in the active site of the two subunits is the conformation of the residue Phe-112. In subunit A, the phenyl ring of the residue projects into the active site, whilst in subunit B it is situated behind Lys-122 (Fig. 4.7A). Structurally, this conformational difference has a number of implications. Firstly, the active site is larger in subunit B, as the Phe-112 phenyl is absent. Secondly, the position of Phe-112 in subunit B partially blocks a water channel that is present connecting Lys-122 to the solvent. The conformational organisation of the Phe-112 seems to affect substrate binding. In subunit A, with the smaller active site, it appears that a single molecule binds whereas in subunit B two molecules are bound: both dopamine and 4-HBA occupy the active site, in a stacked conformation (Fig. 4.4B). Analysis of the electron density of the ligands and the Phe-112 is described above (section 4.1.3). In order to investigate the influence of the residue using a different method, molecular dynamics simulations of NCS were performed. The crystal structure 2VQ5 was used as the starting point, and subunits A and B were investigated separately (see section 2.2.2 and Appendix A, section 8).

For each subunit three separate molecular dynamics simulations of 100 ns were calculated. The behaviour of the Phe-112 residue was examined by tracking two separate angles. The rotation of the phenyl ring was tracked through the changing dihedral angle between the Cα, Cβ, Cγ and $C\delta^1$ atoms (Fig. 4.11). Although there are two chemically identical Cδ atoms (so a 0° angle is identical to a 180° angle), by tracking the angle it was possible to follow the rotation of the side chain. The phenyl ring typically occupied a position of approximately 0° or 180°, which is expected for phenylalanine (Fig. 4.12). Notably the phenyl ring rotated at least once in all simulations (and at most eight times), suggesting that it was not packed tightly with other residues in the enzyme active site.

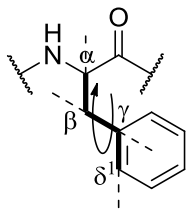


Fig. 4.11 Method for measuring Phe-112 rotation. The dihedral angle depicted was tracked to measure Phe-112 side chain rotation.

The position of Phe-112 relative to the entire protein was monitored by the angle formed between the Phe-112 Cα, Cζ (at the tip of the phenyl ring) and the Cα of His-106, which is on the same β-strand as Phe-112 but at the other end of the protein (Fig. 4.13). For simulations of subunit A, this angle fluctuated around 65°. The Phe-112 did not change its position relative to the protein at any point in the 300 ns of subunit A simulations (Fig. 4.14). This must suggest, to some degree, that this Phe-112 conformation is stable. Subunit B simulations showed very different characteristics. Firstly, the starting angle of Phe-112 was 110-130°. In all simulations the Phe-112 conformation switched to the 65° conformation at some point, suggesting the 110-130° conformation is not stable (Fig. 4.14). For simulation 1 (yellow), this happened immediately (within 5 ns) and the residue remained in this position for the rest of the simulation (though rotated a number of times). In simulation 2 (blue) the Phe-112 conformation decayed to 65° only after 70 ns, but remained there for the rest of the simulation. Simulation 3 (red) started similar to simulation 1 in that the residue immediately switched to the 65° conformation. However, in this simulation the side chain appeared slightly more unstable and there was brief reversion to the 110-130° conformation at 60-75 ns before switching again to 65°.

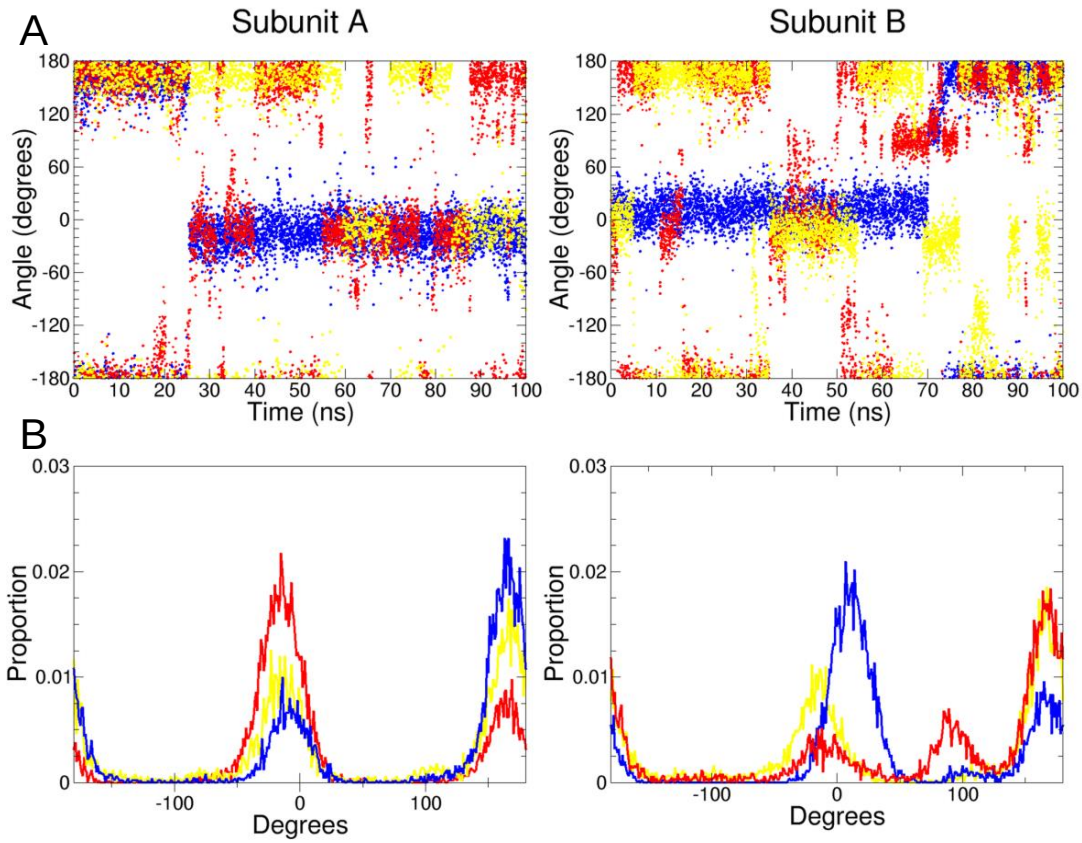


Fig. 4.12 Phe-112 rotation in simulations. A. Angle of dihedral over course of simulations (Fig. 4.11). **B. Distribution of angles in simulation.** Different colours depict different simulations, subunits were examined independently. Note how for all simulations the Phe-112 rotates, and typically occupies positions around 0˚ and 180˚. The exception is the red simulation of subunit B which also occupies a position around 100 ˚.

Overall the data show that Phe-112 was dynamic in the active site during the molecular dynamics simulations. The phenyl ring rotated and the rotamer conformation changed. Generally, it seems that the rotamer at 65° (observed in subunit A) is more stable than the rotamer at 110-130° (observed in subunit B). Although no switching from 65° to 110-130° was observed in subunit A in these simulations, the observation of this switch in subunit B simulation 3 suggests that it is a possible transition—it is not a completely kinetically trapped conformation. This molecular dynamics simulation data adds to the evidence from the electron density analysis suggesting that the Phe-112 conformation and active site arrangement observed in the crystal structure subunit B is atypical. Overall, the simulations suggested that the Phe-112 conformation most prevalent in solution is that observed in subunit A.

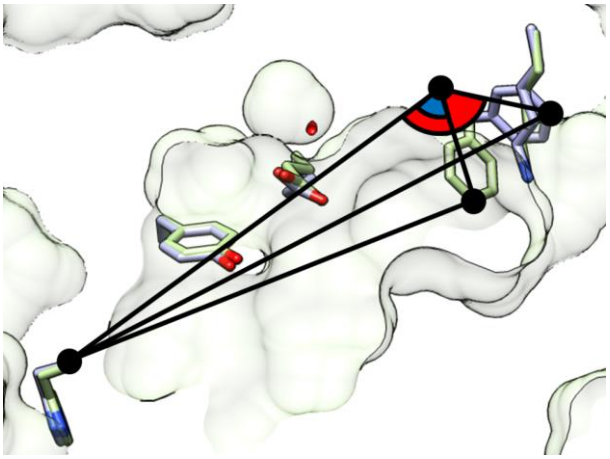


Fig. 4.13 Method for measuring Phe-112 position. The Cα atoms on His-106 and the Cα and Cζ Phe-112 are marked by black dots. The angle formed between them was recorded. Depicted are the subunits A (blue) and B (red) from the crystal structure. The angle for subunit A is approximately 65˚, the angle for subunit B around 120˚.

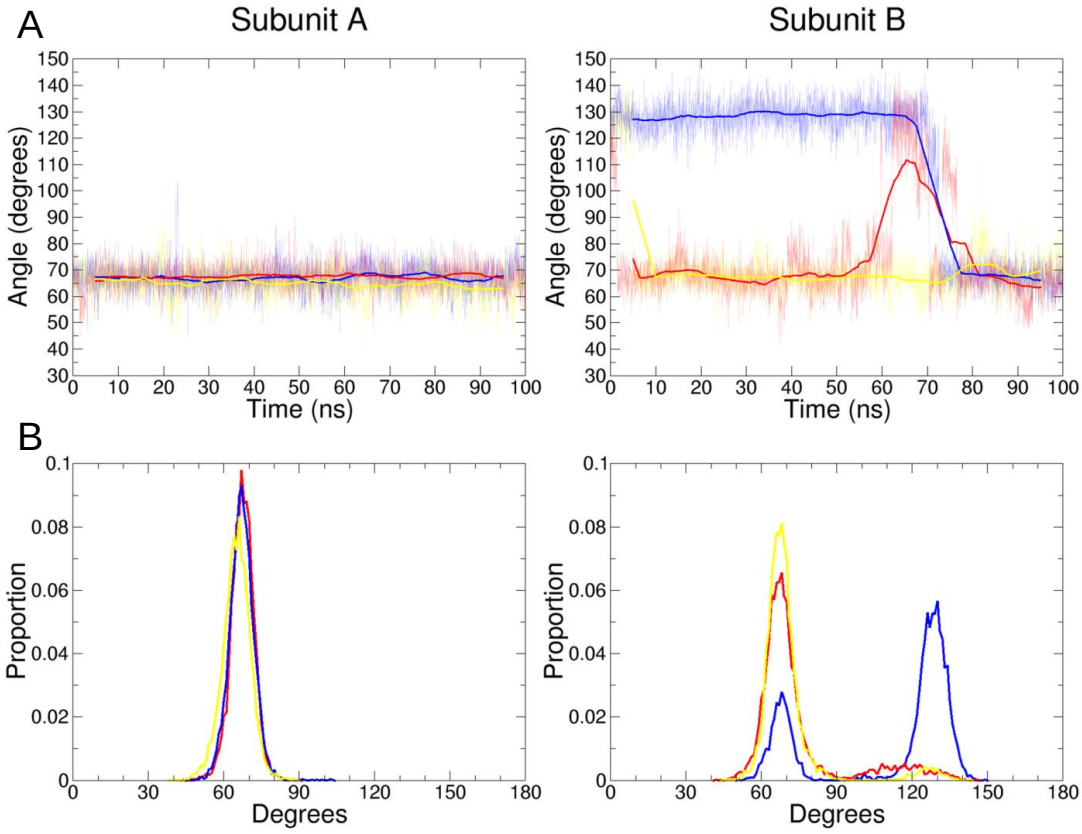


Fig. 4.14 Position of Phe-112 residue in simulations. A. Value of angle (Fig. 4.13) over course of simulation. The thin lines are the values for this angle; the thick lines are a moving average (0.5 ns window). Different colours represent independent simulations. **B. Distribution of angles in simulation.** Note how for all simulations of subunit A the Phe-112 angle does not vary whilst in subunit B all simulations decay to 65 ˚. Figure adapted from Lichman *et al*.[137]

4.2.2 Molecular docking

In order to explore the NCS enzyme mechanism, reaction intermediates were docked into the active site of *Tf*NCS (2VQ5, subunit A) using AutoDock Vina (Fig. 4.15). This

work was a replication and extension of the docking experiments conducted by Pesnot *et al*.[57] Subunit A was used for docking studies as it is a better candidate than subunit B on the basis of molecular dynamic simulations (see section 4.2.1). The *holo* crystal structure 2VQ5 (with ligands removed) was preferred to *apo* structure 2VNE on the basis of its superior resolution.

The lowest energy docked structure of dopamine features the catechol portion of the substrate in close proximity with Lys-122, with the 3-hydroxy closest to the side chain amine (Fig. 4.16A, Table 4.1). The positively charged nitrogen on dopamine is near to the negatively charged carboxylic acid residues Glu-110 and Asp-141. The aldehyde 4-HPAA binds in a similar manner, with the phenol group in close proximity to Lys-122, and the carbonyl near to the carboxylic acid residues (Fig. 4.16B, Table 4.1). This binding mode is different to that predicted in the HPAA-first mechanism; in this mechanism the aldehyde carbonyl binds to the Lys-122 (Fig. 4.5).

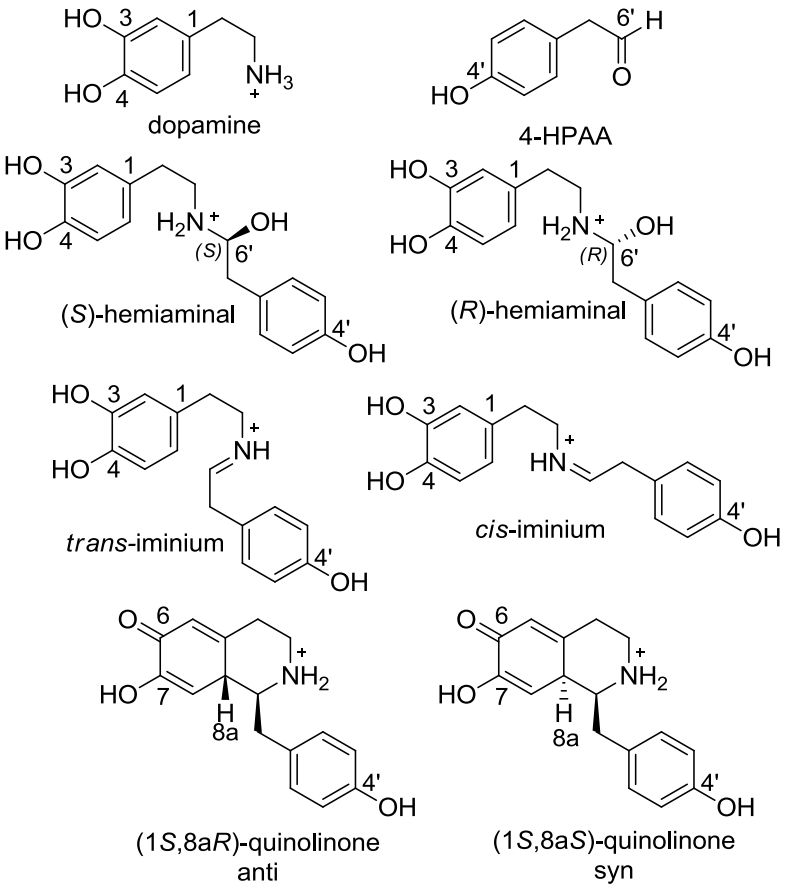


Fig. 4.15 Structures of intermediates used in docking calculations. Figure adapted from Lichman *et al*.[137]

Two hemiaminal intermediates can be formed by the reaction between dopamine and 4-HPAA. These are stereoisomers, with chirality at the hydroxyl group (Fig. 4.15).

Docking the (*S*)-hemiaminal provides a 'dopamine first' arrangement of the intermediate, with the dopamine catechol buried in the active site, close to Lys-122, in a similar manner to that observed with docking dopamine alone (Fig. 4.17A, Table 4.1). The charged nitrogen of the hemiaminal interacts with Glu-110 and Asp-141, whilst the hydroxyl group occupies a space between Glu-110 and Tyr-108, in the entrance to a water channel. The aldehyde R-group extends past Tyr-108 at the active site tunnel entrance, in close proximity to Leu-76. In contrast, the (*R*)-hemiaminal docked structure binds in the opposite arrangement, with the 4-HPAA R-group buried in the active site, in a similar manner observed with the 4-HPAA docking (Fig. 4.16B, Table 4.1). The hemiaminal moiety is close to the carboxylic acid residues, but the hydroxyl is deeper into the active site than the nitrogen.

Table 4.1 Binding modes calculated from *in silico* docking. Predicted binding affinity and protein-ligand interaction distances of docked reaction intermediates in the NCS active site (2VQ5, subunit A). Conformations predicted by AUTODOCK VINA. For full structures of intermediates and atom numbering information, see Fig. 4.15. Fig. 4.16, Fig. 4.17, Fig. 4.18 and Fig. 4.19 show the structures of selected binding modes. All binding modes described were the highest ranked by the software. Table adapted from Lichman *et al*.[137]

Ligand	**Binding affinity**	**Heteroatom distances Residues**										
		Tyr-108		Glu-110(O1)[a]		Glu-110(O2)[a]			Lys-122		Asp-141	
		Atom on ligand[b]										
		N	6′O	N	6′O	N	6′O	8aH	3O	4′O	N	6′O
	kcal.mol^{-1}	Å										
Dopamine	-5.3	4.5		2.9					3.1		3.0	
4-HPAA	-5.2		4.4		3.0		3.4			3.1		3.4
(*S*)-hemiaminal	-7.3	5.0	2.8	3.1	2.9				3.2		3.2	4.2
(*R*)-hemiaminal	-7.1	4.6	4.4	3.8	3.0	3.1	3.7		>10	3.1	5.1	3.0
Cis-iminium	-7.5	4.0		3.0		3.2			3.1		3.9	
Trans-iminium	-7.2	4.3		3.8		3.6			3.2		4.6	
anti-quinolinone	-7	5.3		4.3				3.3	2.8		4.0	
syn-quinolinone	-6.8	8.3		>10				9.6	>10		>10	

[a] Both oxygen atoms on the Glu-110 side chain are involved in docking interactions.

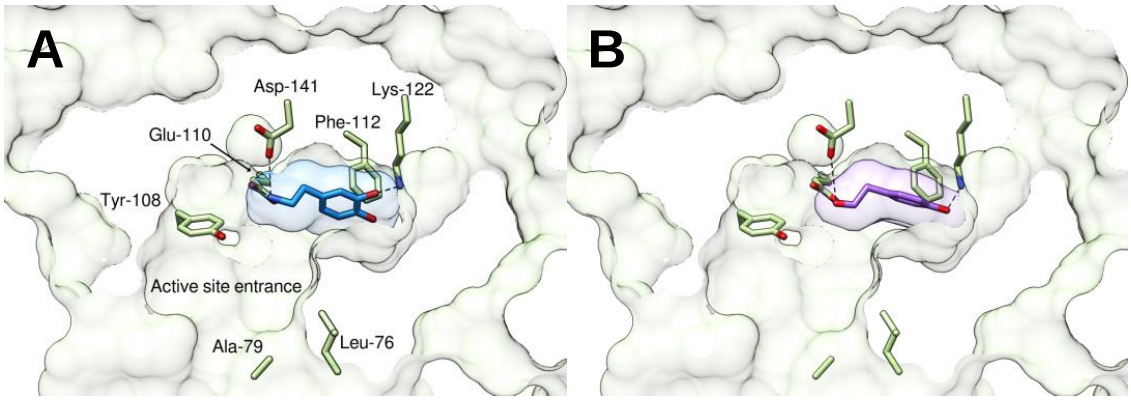


Fig. 4.16 Predicted conformations of substrates in NCS active site. A. Dopamine. B. 4-HPAA. Structures of ligands can be found on Fig. 4.15. More detailed information on protein-ligand interactions can be found in Table 4.1. Black dotted lines show predicted electrostatic and catalytic interactions. Active site residues carbons are coloured yellow; docked ligands carbons are coloured purple. Ligand accessible surfaces are represented by transparent purple. A section of the protein solvent accessible surface is represented in green. Conformations depicted are lowest energy predictions. Figure adapted from Lichman *et al*.[137]

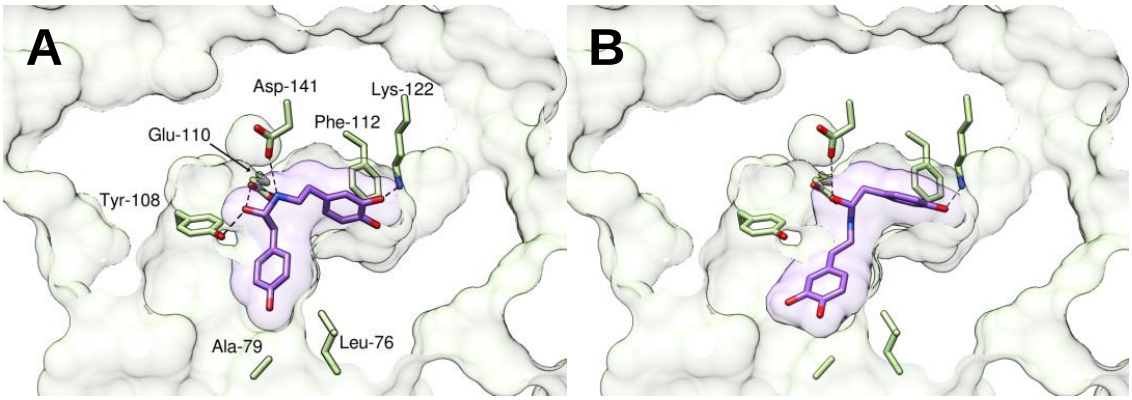


Fig. 4.17 Predicted conformations of hemiaminal intermediates substrates in NCS active site. A. (*S*)-hemiaminal. B. (*R*)-hemiaminal. Figure details described above in Fig. 4.16. Figure adapted from Lichman *et al*.[137]

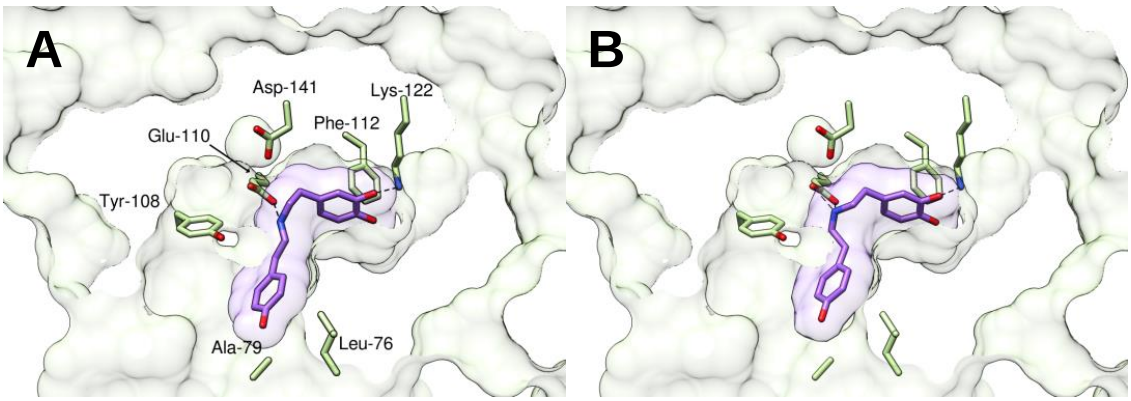


Fig. 4.18 Predicted conformations of iminium intermediates in NCS active site. A. *Trans*-iminium. B. *Cis*-iminium. Figure details described above in Fig. 4.16. Figure adapted from Lichman *et al*.[137]

There are two possible iminium intermediates depending on the *cis* or *trans* geometry of the double bond (Fig. 4.15). Typically, *trans* double bonds are more stable due to steric effects; however, in the confines of an enzyme active site this preference may be reversed. Both isomers dock into the enzyme active site in a similar dopamine-first manner: the dopamine catechol binds to Lys-122; the nitrogen is in close proximity to Tyr-108, Glu-110 and Asp-141; and the aldehyde R-group is at the active site entrance, pointing out towards the solvent (Fig. 4.18, Table 4.1). The key difference between the *cis*- and *trans*-iminium intermediates is the chiral centre that would form in the next electrophilic addition step. The electrophilic addition into *trans*-iminium would be at the *Re*-face of the iminium, forming the correct (1*S*)-product (Fig. 4.18A). The *cis*-iminium conformation would result in a *Si*-face attack, causing formation of the incorrect (1*R*)-isomer (Fig. 4.18B, also see Fig. 4.9).

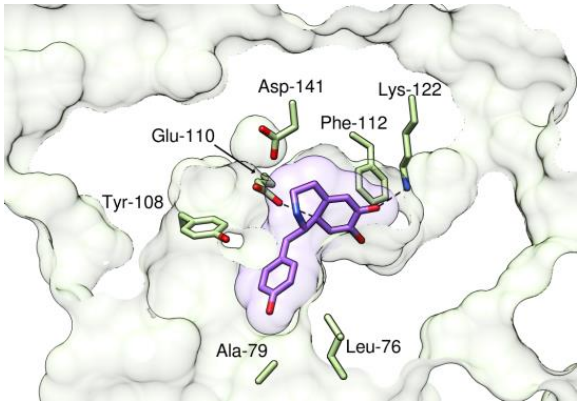


Fig. 4.19 Predicted conformation of *anti*-quinolinone intermediate in NCS active site. The *syn*-quinolinone was not predicted to bind in the active site. Figure details described above in Fig. 4.16. Figure adapted from Lichman *et al*.[137]

The final reaction intermediate is the quinolinone. Again, there are two possible isomers of this intermediate (assuming an (*S*)-isomer product) (Fig. 4.15, also Fig. 4.9). For these intermediates the difference is the 8a-stereocentre. This second chiral centre is removed in the rearomatisation step, but its configuration is important for the involvement of the enzyme in this process. The two quinolinone intermediates are described either as *syn,* where both protons are on the same face of the molecule, or *anti*, where the protons are on opposite faces. Typically *anti* products are more stable due to steric and conformational effects. The docking of these two intermediates provides very distinct results. The *syn*-(1*S*,8a*S*)-quinolinone does not bind in the active site at all, but on the edge of the active site tunnel entrance. The *anti* (1*S*,8a*R*)-quinolinone binds in a dopamine-first manner: the catechol is close to Lys-122 (< 3 Å) and the aldehyde R-group is at the active site entrance (Fig. 4.19). The positive nitrogen was further away from the carboxylic acid residues than with the docking of

other reaction intermediates, but there is a close interaction with Glu-110 and the 8a-proton, which is removed in the final reaction step.

A key factor in the biocatalytic utility of NCS is its ability to turn over a wide variety of structurally diverse aldehydes. These docking calculations provide a structural explanation for this observation: in the most favourable binding modes for reaction intermediates only the aldehyde R-group is partially exposed to the solvent. The poor tolerance of α-substituted aldehydes by NCS may be rationalised on steric grounds, as the α-carbon is buried in the active site. Large aldehydes (but with no α-substitutions) may be accepted as the R-group bulk can be in the solvent, away from the enzyme. The docking calculations also aid in understanding the mechanism itself by showing potential amino acid interactions with reaction intermediates and by demonstrating the different binding modes of reaction intermediate isomers.

4.2.3 pK_a predictions

Predicted pK_a values of titratable amino acids in the NCS active site indicate their protonation states at different points in the reaction mechanism (Table 4.2). These residues are Tyr-108, Glu-110, Lys-122 and Asp-141. All of these residues were identified as interacting with reaction intermediates on the basis of the docking experiments.

Table 4.2 Active site residue pK_a predictions. Acid dissociation constants (pK_a) of TfNCS active site and reaction intermediate ionisable groups predicted by PROPKA, version 3.1[139]. The enzyme model was the X-ray structure 2VQ5, subunit A. Ligand conformations were predicted by computational docking of reaction intermediates into the enzyme active site with AUTODOCK VINA (**Table 4.1**). Table adapted from Lichman *et al*.[137]

	Residue				
	Tyr-108	**Glu-110**	**Lys-122**	**Asp-141**	**Ligand N**
			pK_a		
Typical[a]	10.0	4.5	12.5	3.8	
Ligand[b]					
None	15.5	8.1	7.9	6.7	
Dopamine	13.3	5.7	7.2	4.1	15.8
(*S*)-hemiaminal	15.2	5.6	7.2	4.2	13.6
Cis-iminium	16.4	5.1	7.1	6.8	11.6
Trans-iminium	16.2	5.4	7.1	7.0	11.4
Anti-quinolinone	16.3	6.9	7.4	5.6	10.5

[a] Model pK_a values used by PROPKA 3.1 at the start of calculations

The phenol group of tyrosine has a typical pK_a of 10 but is known to deprotonate in some reaction mechanisms. In the NCS active site it is predicted to have a higher pK_a, ranging from 13.3 to 16.4 depending on the ligand present; this is too high for any significant deprotonation during the reaction mechanism. These predictions therefore

suggest that Tyr-108 does not play any active acid/base role in the mechanism, but may contribute electrostatic stabilisation or H-bond interactions.

The carboxylic acid residues glutamic acid and aspartic acid are typically deprotonated and negatively charged (model pK_a of 4.5 and 3.9 respectively). However, in the *apo* NCS active site both Glu-110 and Asp-141 are predicted to have a significantly increased pK_as (8.1 and 6.7 respectively), suggesting a proportion of the residues are protonated and not charged. There would be repulsion between two negatively charged residues in close proximity in a hydrophobic active site, and the pK_a is affected accordingly. It is possible that at pH 7.5 there is one negative charge (and one proton) shared between the two residues. The binding of a ligand with a charged nitrogen changes the predicted pK_a noticeably—both residues are expected to be deprotonated and negatively charged to stabilise the positively charged nitrogen. There is a predicted increase in the basicity of Asp-141 when the iminium was bound, probably caused by the greater distance between the positive charge and the residue. There is also a predicted increase in basicity of Glu-110 when the quinolinone intermediate is bound. At this point, Glu-110 appears in close proximity to the 8aH proton, which is removed in the next reaction step (Fig. 4.19). This prediction points towards Glu-110 as the mediator of this deprotonation/rearomatisation.

Lysine residues are typically protonated and positively charged (model pK_a 12.5). In the hydrophobic confines of the NCS active site, this tendency for positive charge is removed, and the predicted pK_a of Lys-122 accordingly is reduced to 7.9. This would result in a proportion of Lys-122 residues being neutral and non-protonated at pH 7.5. Upon substrate binding this effect appears greater and the pK_a falls to 7.2-7.4 depending on the substrate bound. This increased neutrality is in accordance with Lys-122 catalysing the cyclisation step in the dopamine-first mechanism. In order for the lysine nitrogen to interact/bind to the 3-hydroxy group of the substrate, it must be neutral. The pK_{a2} of phosphate, which also catalyses this reaction, is 7.2, suggesting that this level of acidity/basicity is ideal for catalysing this mechanism. Therefore, the predicted pK_a of Lys-122 with substrates bound is ideal for acid-base catalysis in enzyme mechanisms at pH 7.5.

4.3 Experimental analysis

4.3.1 Selection of mutants and experiments

In order to investigate the mechanism of NCS, the behaviours of ten enzyme variants, including wild-type (WT), were examined. Residues predicted to play a role in the two proposed enzyme mechanisms were mutated: Tyr-108 to phenylalanine, Glu-110 to aspartic acid and glutamine, Lys-122 to leucine, and Asp-141 to glutamic acid and asparagine. Residues were also mutated at the aldehyde binding site predicted by the dopamine-first mechanism. Leu-76 was mutated to alanine and valine, and Ala-79 was mutated to isoleucine. These aldehyde-binding mutations in particular were chosen to distinguish between the HPAA-first and dopamine-first mechanisms; it was proposed that, based on the dopamine-first mechanism, these mutations would have aldehyde specific effects. If the enzyme operated by the HPAA-first mechanism, mutations of these residues would affect activity in a manner general to all aldehydes, as they are not near to the proposed aldehyde binding site adjacent to Lys-122.

All NCS variants expressed and purified successfully (Fig. 4.20). Melting and aggregation temperatures were measured (for all variants except A79I) to examine the effect of amino acid substitutions on thermodynamic stability. Biotransformation conversion yields were measured, along with enantiomeric excesses for enzymes demonstrating sufficient conversions. Enzyme activities were determined by measuring conversions after thirty seconds—this was within the linear phase of the reaction time course and thus a reasonable proxy for initial rate (Fig. 4.21). For selected variants, kinetic parameters were determined using the same method. In order to examine any aldehyde-specific effects from amino acid substitutions both 4-HPAA and hexanal were examined in enzyme assays (dopamine was always the amine used). 4-HPAA (the natural substrate for the enzyme) and hexanal have different aldehyde R-groups and so were expected to interact with the enzyme differently to each other.

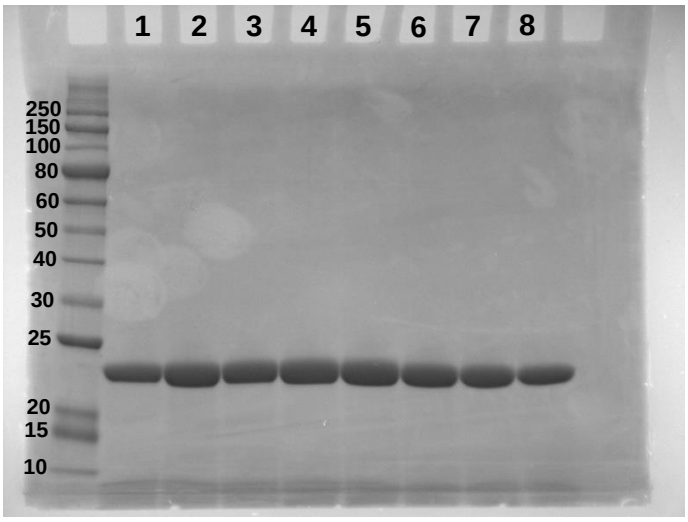


Fig. 4.20 SDS-PAGE showing purified Δ29*Tf*NCS mutants. From lane 1 to 8: WT, L76A, L76V, Y108F, E110D, E110Q, D141E, D141N. Mutants A79I and K122L were also purified in a similar manner.

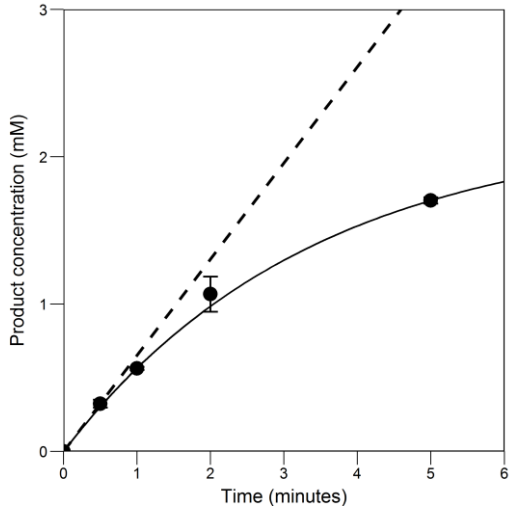


Fig. 4.21 Time course of reaction with 10 mM dopamine, 2.5 mM 4-HPAA and 20 μg.mL^{-1} WT-Δ29*Tf*NCS. Time points taken at 30 sec, 1 min, 2 min and 5 min. Solid line describes fitted exponential decay, dotted line shows initial rate. This experiment demonstrates that a reaction time of 30 seconds is within the linear regime of the enzyme and thus is a reasonable proxy for the initial rate.

4.3.2 General observations

The, By using a method that analyses fluorescence spectra and static light scattering to define protein unfolding transitions, the melting temperature (T_m) of WT Δ29*Tf*NCS was determined to be 60.7 °C (Table 4.3). A previous study, which used CD spectroscopy to determine melting temperatures, detected a T_m of 66 °C.[54] The difference between these two values is likely to be due to differences in method and sample preparation.

Table 4.3 Melting and aggregation temperatures (°C) of TfNCS variants. Data generated using an Avacta Optim instrument. Over a temperature ramp, T_m was calculated by observing changes in the intrinsic protein fluorescence spectrum; T_{agg} was calculated by changes in light scattering intensity. Table adapted from Lichman *et al*.[137]

Protein	T_m			T_{agg}		
	°C					
WT	60.7	±	0.1	60.8	±	0.1
L76A	54.9	±	0.1	54.4	±	0.6
L76V	58.7	±	0.1	59.1	±	0.5
Y108F	61.5	±	0.3	62.4	±	0.4
E110D	57.2	±	0.2	57.4	±	0.5
E110Q	66.0	±	0.4	65.8	±	0.1
K122L	69.8	±	0.3	67.9	±	1.8
D141E	62.0	±	0.3	61.8	±	0.2
D141N	63.9	±	0.1	62.8	±	0.9

Previous investigations into NCS had been adversely affected by the use of phosphate or Tris buffers (see section 4.1.4).[57] To avoid these problem, HEPES buffer was used in all enzyme assays; this buffer exhibited significantly lower background reactions than phosphate and does not react with aldehydes.[57] However, in order to solubilise the aldehydes, the reactions contained 10% $v.v^{-1}$ MeCN, which may have been slightly detrimental to enzyme activities (see section 6.5.2 for more exploration of co-solvent). Under the conditions used for the biotransformation between dopamine and 4-HPAA, phosphate buffer exhibited 69% conversion, without enzyme being present. HEPES, on the other hand, showed just 5% conversion (Fig. 4.22). High substrate concentrations increase the non-enzymatic conversion; this ultimately limits the concentrations that can be reached in biotransformations, enzyme activity assays and kinetic experiments. Enzyme activities with both substrates at 10 mM were difficult to measure due to high background conversions, so activities were measured with one substrate at 10 mM and the other 2.5 mM. In kinetic experiments, one substrate was maintained at 2.5 mM, the varied substrate was tested varied up to 20 mM for dopamine and 15 mM for aldehydes. High concentrations could not be reached due to high background conversions and concentration dependent substrate degradation.

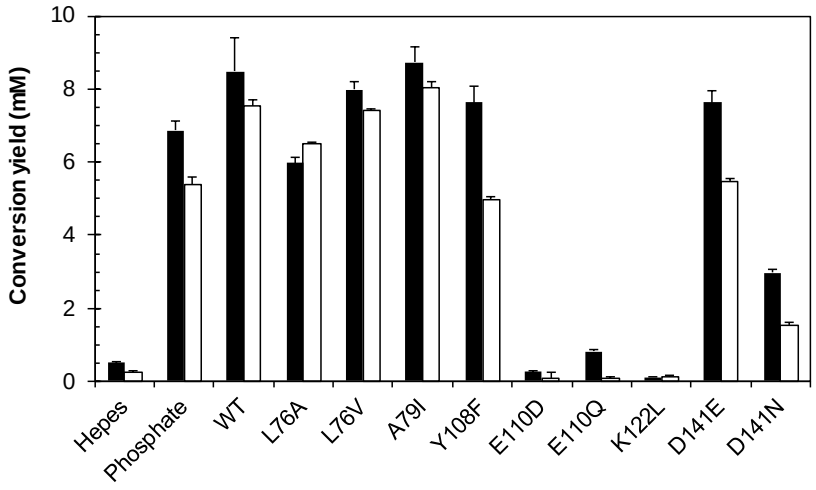


Fig. 4.22 Conversion yields of biotransformations. Enzymes and buffers were incubated with dopamine and aldehydes for 1 h at 37 °C. These conversion yields are the product concentrations at the end of the hour reactions. Both starting substrate concentrations were 10 mM, so that the maximum product concentration possible is 10 mM. Aldehydes: 4-HPAA (black bars) and hexanal (white bars). Conversion yields were obtained by HPLC analysis and quantified by comparison with verified standards. Values are the means of three separate reactions, error bars indicate SDs. Hepes background conversions have been minused from all enzyme samples. Figure adapted from Lichman *et al*.[137]

Generally, conversions with 4-HPAA were higher than with hexanal (Fig. 4.22). This difference was observed also with the chemical phosphate reaction, so is not an enzyme dependent effect. Instead it is possibly due to intrinsic differences between 4-HPAA and hexanal with respect to solubility, reactivity and propensity to exist in solution as hydrates. This difference in reactivity was also reflected in the activity data, and it was more apparent in activities than conversions.For WT, activity with 4-HPAA is three times that with hexanal, 426 and 142 nkatal.mg^{-1} respectively. This is observed because activity reflects initial rates rather conversions over time; with biotransformations, over time the hexanal conversion converges with the 4-HPAA conversion.

Chiral HPLC analysis of the biotransformations showed that THIQ products produced by NCS were single enantiomers (Fig. 4.23). The enzyme products were compared to those produced by the racemic phosphate reaction: phosphate catalysed reactions gave two enantiomer peaks and a dopamine starting material peak. For all conversions above 10% the product peaks could be detected by chiral HPLC. For these conversions only a single product peak could be identified (>95% *ee*), demonstrating the chiral nature of the enzyme reaction, and also showing the successful repression of the racemic background reaction. For variants E110D, E110Q and K122L, which showed conversions below 10%, peaks on the chiral HPLC could not be identified and integrated with sufficient confidence to define the reaction stereoselectivity.

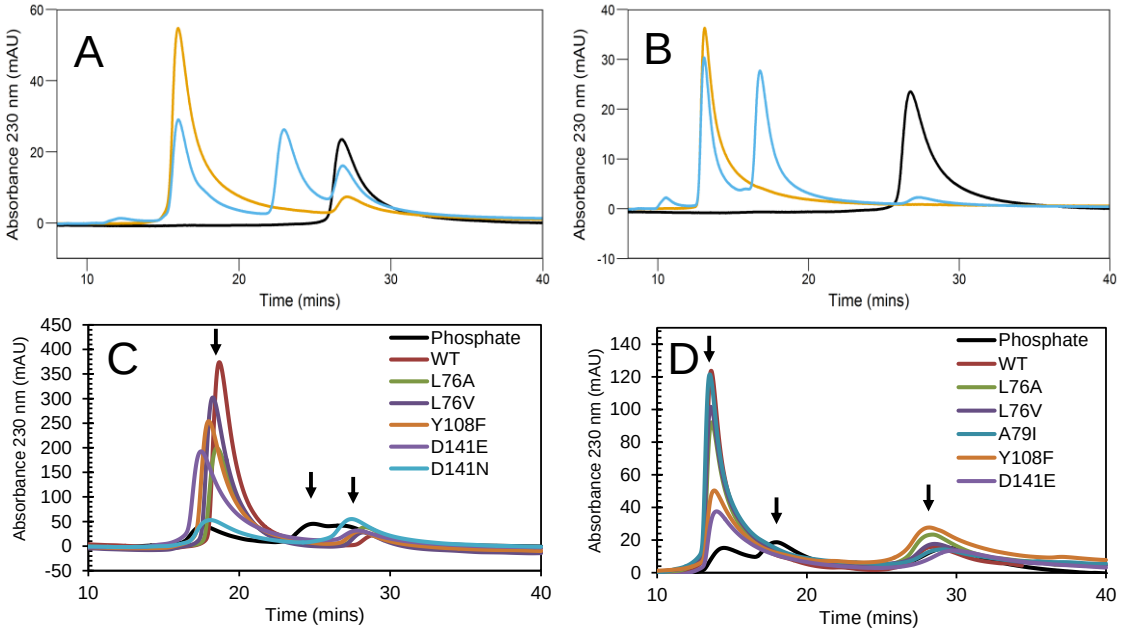


Fig. 4.23 Chiral HPLC analysis of enzymatic reactions. (A) Dopamine and 4-HPAA with WT. (B) Dopamine and hexanal with WT. Dopamine standard (black trace) appeared at 27 min. Phosphate buffer catalyzed reactions (blue trace) showed two equal sized peaks for the racemic products. Wild-type enzyme catalyzed reactions (orange trace) showed a single isomer peak (> 95% ee). **(C) Dopamine and 4-HPAA with variants.** Conversion with A79I not shown (but variant was shown to be stereoselective in plot D). **(D) Dopamine and hexanal with variants.** Conversion with D141N was too low to observe on chiral HPLC. For both C and D, conversions with E110D, E110Q and K122L were too low to observe. Arrows, from left, show (*S*)-product, (*R*)-product and dopamine. HPLC method 3 (see section 2.4.1). Figure adapted from Lichman *et al*.[137]

Due to the high background reaction at high substrate concentrations, enzyme activities were measured with one substrate at 10 mM and the other at 2.5 mM (rather than both at 10 mM as in the biotransformations). This resulted in two enzyme activity measurements for each substrate combination. In all cases, higher concentrations of dopamine resulted in higher rates (Fig. 4.24). Possible explanations for this (based on this data alone) is that 2.5 mM dopamine was not at a saturating concentration or that high concentrations of aldehyde could have been inhibitory. Subsequent kinetic analysis of WT with 4-HPAA or hexanal failed to show evidence of substrate inhibition and also showed that neither dopamine nor 4-HPAA or hexanal were saturating even at 10 mM (Fig. 4.25,Table 4.4).

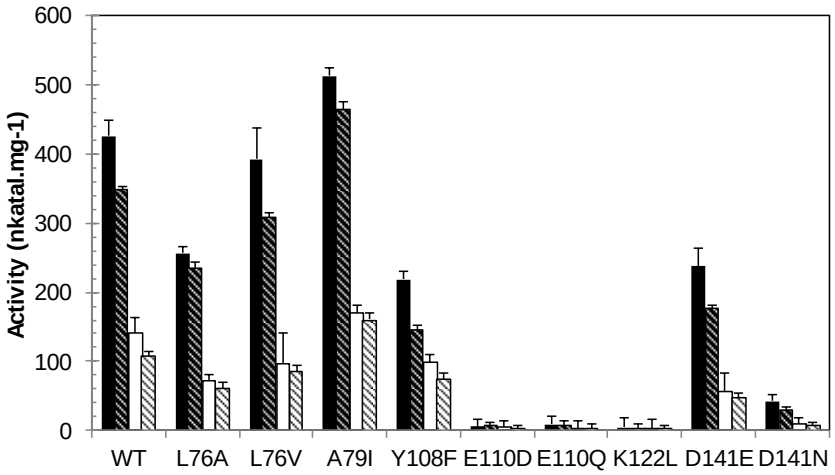


Fig. 4.24 Enzyme activities. Initial rates of reactions between dopamine and aldehydes catalysed by NCS variants. 4-HPAA (black bars) and hexanal (white bars); Solid bars show 10 mM dopamine and 2.5 mM aldehyde, lined bars show 2.5 mM dopamine and 10 mM aldehyde. Values are the mean of three separate measurements, error bars indicate SDs. The Hepes background activity has been minused from all measurements.

Kinetic experiments were conducted by keeping one substrate concentration at 2.5 mM and varying the other up to 15 mM (for aldehydes) or 20 mM (for dopamine). With WT: dopamine, 4-HPAA and hexanal did not reach saturating concentrations in these experiments.For two-substrate kinetics, the non-varied substrate concentration should ideally be at a saturating concentration. The fact that this was not achieved in these experiments prevents straightforward comparisons between kinetic parameters for dopamine and for aldehydes; this in turn prohibits satisfactory explanations for the effect of substrate concentrations on activity. It is possible, though, to compare the effect of different aldehydes and amino acid substitutions on kinetic parameters. For example, it is clear that reduction in activity caused by the use of hexanal instead of 4-HPAA was manifested in k_{cat}^{app} rather than K_m^{app}. Dopamine K_m^{app}, as expected, did not change significantly, but the k_{cat}^{app} fell by over three-fold. Interestingly, although hexanal had a lower k_{cat}^{app} than 4-HPAA, it also had a lower K_m^{app}. This was perhaps reflective of hexanal's increased hydrophobicity and reduced solubility, which could increase affinity towards a hydrophobic enzyme active site. Taking into account the reduced chemical conversions observed with hexanal instead of 4-HPAA, it is possible that the reduced k_{cat}^{app} was a chemical reactivity effect rather than an enzymatic effect. The aldehydes exist in water in equilibrium with the hydrate, and it is possible that for hexanal the equilibrium is further towards the hydrate, reducing reactivity.

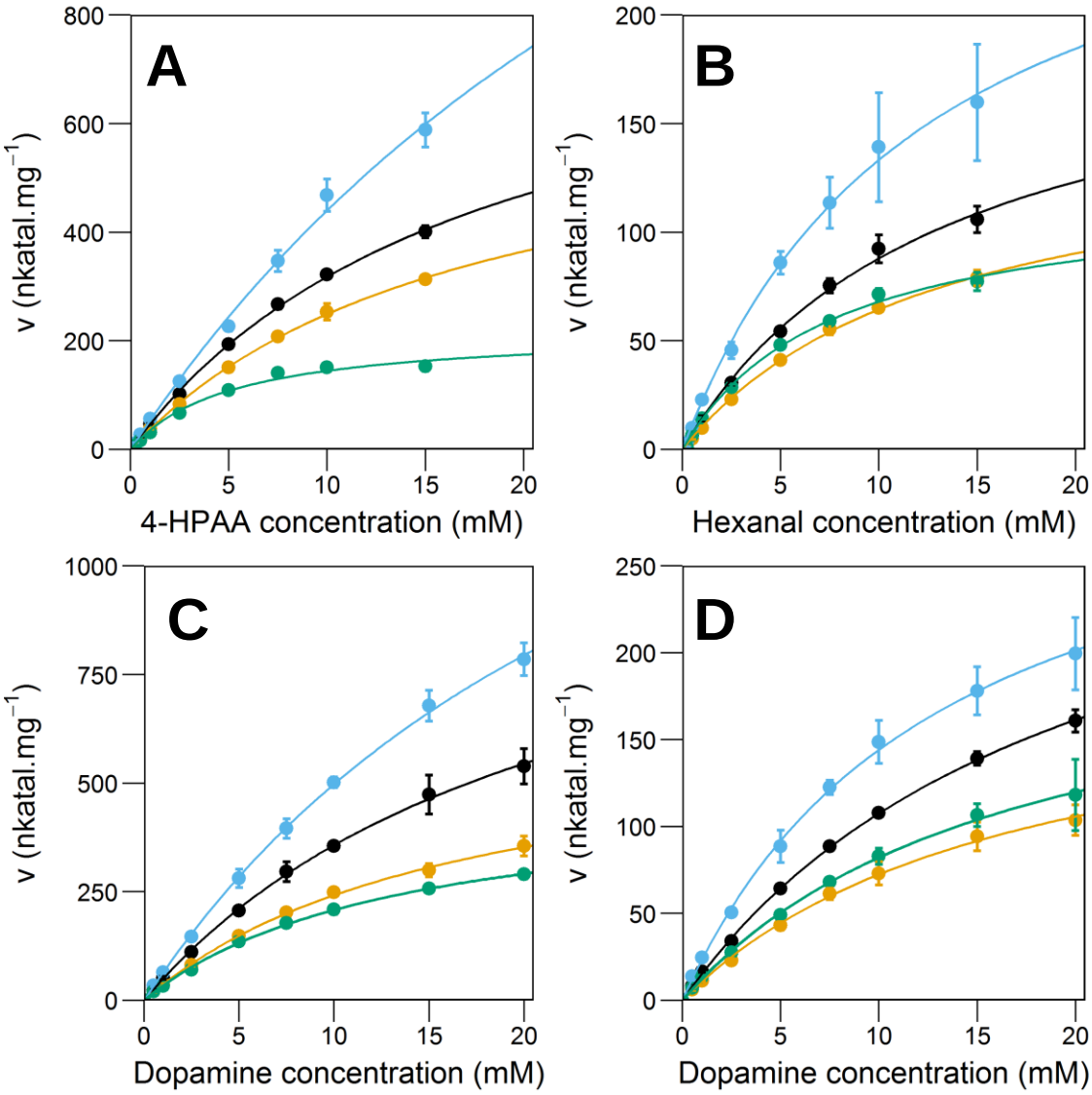


Fig. 4.25 Velocity/substrate concentration curves for wild-type (WT, black), L76A (orange), A79I (blue) and Y108F (red). (A) Varying 4-HPAA concentrations with 2.5 mM dopamine. (B) Varying hexanal concentrations with 2.5 mM dopamine. (C) Varying dopamine concentrations with 2.5 mM 4-HPAA. (D) Varying dopamine concentrations with 2.5 mM hexanal. All fits are Michaelis-Menten kinetics. Data points are the mean from three measurements; error bars indicate SDs of three reactions. For parameters, see **Table 4.4**. Figure adapted from Lichman *et al*.[137]

Table 4.4 Apparent kinetic parameters for TfNCS with 4-HPAA and hexanal. Corresponding velocity/substrate concentration curves can be found in **Fig. 4.25**. Table adapted from Lichman *et al*.[137]

Varied Substrate	Constant Substrate	Enzyme	Michaelis-Menten			
			V_{max}^{app}	k_{cat}^{app}	K_m^{app}	k_{cat}^{app}/K_m^{app}
			nkatal.mg^{-1}	s^{-1}	mM	s^{-1}.mM^{-1}
Dopamine[a]	4-HPAA	WT	1150 ± 48	24 ± 1	22.2 ± 1.5	1.09 ± 0.08
		L76A	649 ± 22	13.6 ± 0.5	16.8 ± 1.0	0.81 ± 0.06
		A79I	1972 ± 95	41 ± 2	29.6 ± 2.1	1.39 ± 0.12
		Y108F	485 ± 13	10.2 ± 0.3	13.3 ± 0.7	0.76 ± 0.05
Dopamine[a]	Hexanal	WT	323 ± 7	6.8 ± 0.1	20.1 ± 0.7	0.34 ± 0.01
		L76A	196 ± 10	4.1 ± 0.2	17.0 ± 1.5	0.24 ± 0.02
		A79I	334 ± 11	7.0 ± 0.2	13.2 ± 0.8	0.53 ± 0.04
		Y108F	221 ± 8	4.6 ± 0.2	16.9 ± 1.1	0.27 ± 0.02
4-HPAA[b]	Dopamine	WT	893 ± 43	18.7 ± 0.9	18.1 ± 1.4	1.03 ± 0.09
		L76A	682 ± 24	14.3 ± 0.5	17.4 ± 1.0	0.82 ± 0.05
		A79I	2174 ± 429	46 ± 9	39.4 ± 9.9	1.16 ± 0.37
		Y108F	219 ± 16	4.6 ± 0.3	5.2 ± 0.9	0.89 ± 0.17
Hexanal[b]	Dopamine	WT	205 ± 16	4.3 ± 0.3	13.4 ± 1.8	0.32 ± 0.05
		L76A	150 ± 5	3.2 ± 0.1	13.4 ± 0.8	0.24 ± 0.02
		A79I	300 ± 21	6.3 ± 0.4	12.5 ± 1.5	0.50 ± 0.07
		Y108F	119 ± 7	2.5 ± 0.1	7.6 ± 0.9	0.33 ± 0.04

[a] All kinetic parameters for varied dopamine were recorded with 2.5 mM aldehyde, and dopamine varied from 500 µM to 20mM

[b] All kinetic parameters for varied aldehydes were recorded with 2.5 mM dopamine, and aldehyde varied from 250 µM to 15 mM

4.3.3 Lys-122

The exchange of lysine to leucine at position 122 (K122L) was designed to probe the role of the Lys-122 nitrogen. Leucine was selected as it lacks the amine moiety, does not have β-branching, and has no additional functionality or significant bulk that could potentially give rise to misleading results. Methionine may also be a reasonable amino acid to substitute at this position. The variant K122L was expressed and purified in a similar manner to WT. The K122L variant demonstrated essentially no activity or conversions above background with either 4-HPAA or hexanal (Fig. 4.24). This was in agreement with results from Ilari *et al* where no activity was identified with a K122A mutant (using Δ19*Tf*NCS and the CD spectroscopy method, see section 4.1.4).[43] This data confirmed that Lys-122 is necessary for NCS activity; it plays a key role in the enzyme mechanism.

Interestingly, the K122L amino acid substitution caused an increase in the enzyme stability with regards to both the melting and aggregation temperatures (Table 4.3). The increase in T_m was fairly large: 9 °C. The destabilising effect of the Lys-122 is caused by the energetic cost of burying a positively charged residue in a hydrophobic region. Although Lys-122 is accessible to water from two directions, which can stabilise a charged residue, the adjacent amino acid residues are hydrophobic, and tend not to

form stabilising interactions with charged side chains. The exception to this is Phe-112, adjacent to Lys-122, which could form π-cation interactions with a positively charged lysine.[140]

The destabilising effect of burying a charged residue was also manifested in the pK_a predictions, in which Lys-122 is predicted to have a shifted pK_a, increasing its propensity to be neutral (Table 4.2). The large change in T_m in K122L, however, does suggest that a large proportion of the residue is charged in the *apo* enzyme. There is presumably a balance between the energetic cost of burying a charge and the intrinsic basicity of the primary amine; at pH 7.5 this equilibrium allows the lysine to be primed for acid-base catalysis. There is also likely to be a mechanism for Lys-122 to gain or lose protons aside from interactions with substrates—access to a proton sink is required. This could be mediated by water *via* the tunnel next to Lys-122. Alternatively main-chain amide groups could provide access to the pool of protons on the protein surface.

Both the dopamine-first and HPAA-first mechanisms propose acid-base catalysis for Lys-122. In the HPAA-first mechanism, the positively charged lysine initially activates the aldehyde and then catalyses the formation of the hemiaminal and then the loss of water to form the imine (Fig. 4.5).[56] In the dopamine-first mechanism the Lys-122 interacts with the 3-hydroxy group of dopamine to catalyse the electrophilic aromatic substitution step (Fig. 4.10).[57] The data concerning K122L were insufficient to distinguish between the HPAA-first and dopamine-first mechanisms.

4.3.4 Glu-110

The residue Glu-110 is proposed to play the same role in both the HPAA-first and dopamine-first mechanisms: it catalyses the proton abstraction from the quinolinone intermediate to trigger rearomatisation and product formation. This step has been proposed to be partially rate determining.[51] In order to separate the spatial and charge effects of the glutamic acid carboxyl group, two variants were investigated. The E110Q variant, with an amide in place of the carboxyl group, was intended to preserve as much structural and positional features of Glu-110 as possible, but with the charge removed. The E110D variant, on the other hand, maintained the original charge but the position of this charge was changed. Both variants expressed and purified in a similar manner to WT (Fig. 4.20). Both variants also showed no significant activity or conversions above background, except E110Q which demonstrated conversion with 4-HPAA slightly above the HEPES background (Fig. 4.22, Fig. 4.24). There was insufficient conversion to measure stereoselectivity accurately with the chiral HPLC

method, so it was unclear whether this was an enzyme catalysed conversion. A tentative explanation for this increased conversion can be found later in this section.

The two amino acid substitutions showed contrasting impacts on enzyme stability. E110Q had a melting temperature 5 °C greater than WT, whilst E110D destabilised the enzyme by just over 2 °C (Table 4.3). The explanation for this is similar to that for K122L stability changes: placing a charged group in a buried hydrophobic region of a protein exhibits an energetic cost. The variant E110Q loses the charge so is stabilised, whilst in E110D the charge is buried deeper in the protein, so increases the destabilising effect. It is also possible some interactions between Glu-110 and adjacent residues are maintained in E110Q but lost in E110D due to the change in functional group position—this could also explain the reduction in melting temperature seen for E110D. Again, as for Lys-122, this energetic cost of burying a charge was reflected in pK_a predictions: in the *apo* active site the residue was predicted to be more neutral than charged at pH 7.5 (Table 4.2). When a positively charged nitrogen is placed near the residue, as predicted by the dopamine-first mechanism, the pK_a of the residue is predicted to fall, and the residue would be predominantly negatively charged.

The severe loss of activity with both E110D and E110Q shows that both the position of the residue and the charge of the residue are of utmost importance to the NCS mechanism. This suggests that the residue does not merely provide electrostatic stabilisation to the substrates and intermediates, but plays an active role in the enzyme mechanism. In both proposed mechanisms the residue acts as base, extracting the 8a proton from the quinolinone intermediate—the exact position of the residue determines the success of this step. The docking of the *anti*-(1*S*,8a*R*)-quinolinone in the active site shows the proximity of the 8a proton with the Glu-110 (3.3 Å to the 8a carbon) (Fig. 4.19). This provides a structural insight into how this step may occur in the dopamine-first mechanism. There was also, accordingly, a slight increase in the predicted pK_a of Glu-110 at this point (Table 4.2).

Data on Glu-110 variants are insufficient to distinguish between the HPAA-first and dopamine-first mechanisms, especially as both mechanisms have them performing the same main function. However, previous work on a Glu-110 mutant (E110A) by Ilari *et al* provides a very interesting comparison to the work here. Whereas no activity is observed with Glu-110 variants in this work, Ilari *et al* observed reduced, but significant, activity in the E110A variant.[43] The kinetic assays were conducted using a CD spectroscopy method that detects specific formation of (*S*)-products. Importantly, the assays were conducted in phosphate buffer, which catalysed the racemic reaction. The

presence of activity with phosphate and not with HEPES (the buffer used here) signifies that phosphate must be catalysing the reaction step that Glu-110 typically catalyses. The fact that a chiral product was observed with E110A in CD experiments shows that chirality is not established by the Glu-110 catalysed step. The mutant K122A, on the other hand, showed no formation of chiral products, even though phosphate was present and could catalyse any Lys-122 catalysed step. This shows that chirality is established by the action of Lys-122. This may also account for the very slightly increased conversions observed from E110Q compared to K122L (Fig. 4.22).

This understanding also provides insight into substrate binding in the E110A NCS active site, as in order for phosphate to catalyse the step typically catalysed by Glu-110, either phosphate must enter and bind in the active site (in the Glu-110 position), or reaction intermediates must diffuse out of the enzyme into the buffer. In either scenario this suggests that the active site binds substrates fairly loosely. The high K_m^{app} values determined through kinetic assays suggest this weak binding is a general feature of NCS (Table 4.4).

The *anti*-(1*S*,8a*R*)-quinolinone can only be derived from the *trans*-iminium; it is inaccessible from the *cis*-iminium (see Fig. 4.9). Observations of Glu-110 mutant behaviours coupled with results from docking studies provide strong indication that the *anti*-(1*S*,8a*R*)-quinolinone is the reaction intermediate found in NCS. This, therefore, must also mean that the iminium is found in a *trans* configuration in the reaction mechanism.

Docking investigations of dopamine, hemiaminal and iminium reaction intermediates suggest that Glu-110 is in close proximity to the intermediate nitrogen throughout the reaction mechanism (Table 4.1). This suggests that in the dopamine-first mechanism Glu-110 may play a role in the early steps in the reaction mechanism. It may help catalyse the initial nucleophilic attack to form the hemiaminal, and the subsequent dehydration to form the iminium intermediate.

4.3.5 Asp-141

In both proposed enzyme mechanisms, Asp-141 plays a minor role compared to Glu-110. In the HPAA-first mechanism, it binds to the 4-HPAA phenol (Fig. 4.5).[56] Substrate tolerance studies have shown this cannot be an important interaction for the enzyme mechanism as aldehydes without this phenol group are readily accepted.[53,57,61] In the dopamine-first mechanism the residue is proposed to play a role in binding and stabilising the positively charged nitrogen of the substrate and intermediates. In a

similar manner to investigation into Glu-110, two variants of Asp-141 were studied: D141E probed the spatial effect of the charge and D141N probed the presence of the charge itself. Both variants expressed and purified in a similar manner to WT (Fig. 4.20). The amino acid substitutions also caused a slight increase in melting temperatures (Table 4.3). D141N raised the melting temperature by 3 °C, as removing a charge from a hydrophobic active site increases thermodynamic stability. D141E raised the melting temperature by 1 °C: charges are more detrimental to stability when more buried. Again, pK_a predictions reflected melting point observations: the pK_a of Asp-141 was shifted slightly towards neutrality (at pH 7.5) (Table 4.2).

Activities and conversions with all substrates were reduced as a result of the Asp-141 amino acid substitutions. The removal of the charge (D141N) caused a greater reduction in activity than moving the charge (D141E). The variant D141N caused reduction in all activities by 90% or more; in contrast D141E retained 55% activity for 4-HPAA and 40% activity for hexanal (dopamine 10 mM, aldehyde 2.5 mM) (Fig. 4.24). This effect was reflected in conversions, though it was not so pronounced (Fig. 4.22). Both variants retained stereoselectivity (Fig. 4.23), though D141N conversions with hexanal were too low to detect *via* chiral HPLC. Interestingly, the substitution D141E affected activities with hexanal more than activities with 4-HPAA (Fig. 4.26). The HPAA-first mechanism predicts Asp-141 binds to the 4-HPAA phenol; if so then it would be expected that aldehydes with this phenol interaction would be more affected by Asp-141 amino acid substitutions than those without. The opposite effect was observed—hexanal was more effected by the substitutions than 4-HPAA. The dopamine-first mechanism proposes a more general role for Asp-141 which is not aldehyde specific, so superficially this also cannot account for the slight shift in substrate specificity of D141E. However, the intrinsic reactivity of different aldehydes could account for the aldehyde specific effect: as hexanal is less reactive than 4-HPAA (as shown by reactions with phosphate) it benefits proportionally more from enzyme catalysis than 4-HPAA. When an element in the enzyme catalysis is lost (i.e. the Asp-141 stabilising/charge interactions), the activity towards the less reactive aldehyde is reduced proportionally more. This explanation is speculative; kinetic experiments may be able to account for the aldehyde specific behaviour of Asp-141 variants.

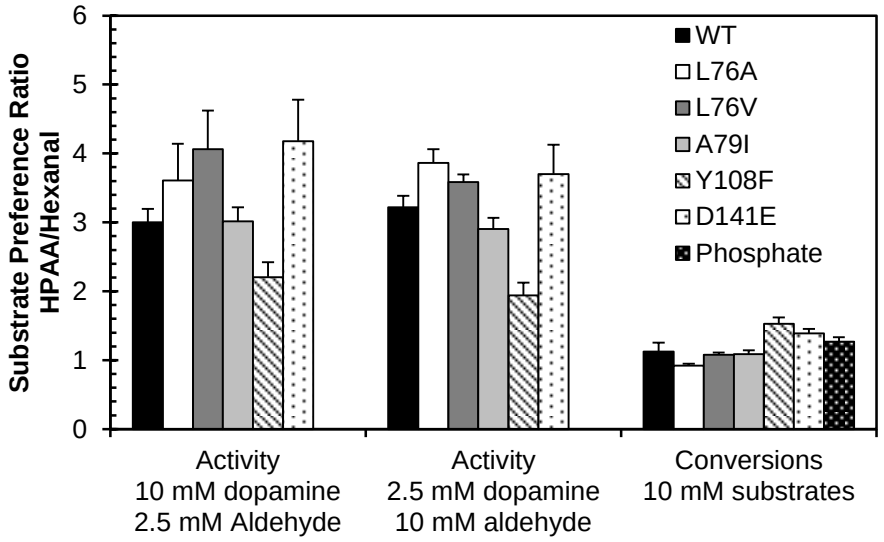


Fig. 4.26 Relative substrate preference of variants. The value for substrate preference is obtained by dividing the activity or conversion with HPAA by the value for hexanal. The resulting value describes the magnitude of HPAA activity/conversion relative to hexanal. Therefore, higher values show increased preference for HPAA compared to hexanal, and lower values relative improvement in activity/conversion for hexanal rather than HPAA.

The behaviour of Asp-141 variants cannot be fully accounted for by either the HPAA-first mechanism or the dopamine-first mechanism. However, the severe reduction in activity (93%) observed with the substitution D141N and the substrate hexanal is completely incongruent with the proposed Asp-141 role in the HPAA-first mechanism (Fig. 4.24). It can, however, be rationalised with the dopamine-first mechanism, in which it does not interact with the aldehyde.

The difference in behaviour between D141E and D141N is indicative of a residue that provides electrostatic stabilisation rather than base catalysis: for Asp-141 the charge was much more important than the exact position of the charge. This contrasts with Glu-110 variant behaviour, where the exact position and charge were both vital. Computational docking (supporting the dopamine-first mechanism) showed Asp-141 stabilising the dopamine and hemiaminal positive charges at the first stage of the reaction, perhaps catalysing the initial nucleophilic attack and subsequent dehydration (Table 4.1). This behaviour would correspond to the experimental data collected.

4.3.6 Tyr-108

In NCS, Tyr-108 forms part of a narrow access tunnel to the large internal active site, restricting the entrance of substrates. It is also adjacent to the important Glu-110 residue. In the HPAA-first mechanism it is proposed to H-bond with the dopamine 4-hydroxy group, stabilising the quinolinone intermediate (Fig. 4.5). This proposed interaction, however, cannot account for successful activity observed when the 4-

hydroxy group was not present.[51,57] In the dopamine-first mechanism Tyr-108 may interact with the iminium cation, but otherwise no clear role is assigned.[57] Computational docking of reaction intermediates in the active site suggested that Tyr-108 interacts with the hemiaminal hydroxy, and perhaps aids initial aldehyde substrate binding and subsequent formation of the hemiaminal and imine (Fig. 4.17A).

To investigate the role of Tyr-108, the variant Y108F was expressed and purified (Fig. 4.20). This amino acid substitution allowed specific interrogation into the role of the phenol, in terms of its H-bond and electrostatic interactions. The variant had a very slightly increased stability compared to wild-type (1 °C), probably because hydrophobic interactions were improved (Table 4.3). The variant showed a 50% loss of activity for 4-HPAA and 30% loss for hexanal (dopamine 10 mM, aldehyde 2.5 mM) (Fig. 4.24). The amino acid substitution had a greater detrimental impact on 4-HPAA activity than on hexanal activity (Fig. 4.26). In order to investigate this aldehyde specific effect more, kinetic parameters for both substrate combinations were determined (Fig. 4.25, Table 4.4).

The variant Y108F showed a reduction in both k_{cat}^{app} and K_m^{app} for every substrate combination (Fig. 4.27, Table 4.4). The most significant reductions in k_{cat}^{app} and K_m^{app} were observed with varying 4-HPAA concentrations. The decreases in k_{cat}^{app} and K_m^{app} were proportionally larger when 4-HPAA was used than when hexanal was used; this was apparent both when dopamine and when the aldehyde was varied. Reductions in K_m^{app} for all substrates with Y108F can be rationalised simply by the increased size of the active entrance allowing substrates easier access. This, however, seems to be the opposite result to what would be expected if the phenol binds to the aldehyde. The decrease in K_m^{app} was more prominent for 4-HPAA than hexanal as it is the bulkier aldehyde and its binding is impacted the greatest due to this change. The fact that the K_m^{app} reductions were proportionally larger for aldehydes than dopamine perhaps can be best accounted for by the dopamine-first mechanism where Tyr-108 is part of the aldehyde binding site. In contrast, the HPAA-first mechanism predicts Tyr-108 to bind to dopamine, not the aldehyde, therefore it would be expected that Y108F would weaken binding (increasing K_m^{app}) to dopamine—this is not what was observed experimentally.

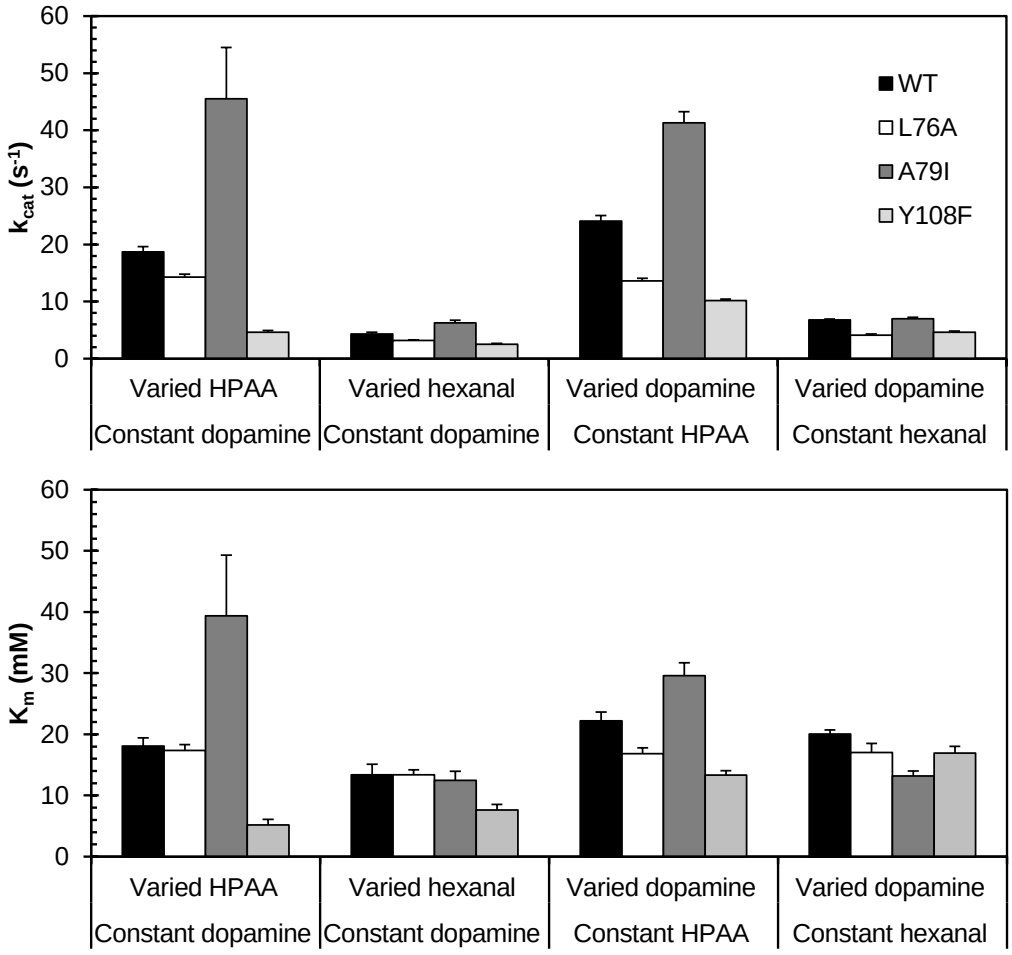


Fig. 4.27 Apparent kinetic parameters for TfNCS with 4-HPAA and hexanal. Parameters found in Table 4.4.

The reduction in k_{cat}^{app} with Y108F suggested that Tyr-108 plays a role in the mechanism of the enzyme. In the HPAA-first mechanism this role would be the binding to the quinolinone intermediate via the 4-hydroxyl group. In the dopamine-first mechanism this role is not as well defined, but based on its interactions with (*S*)-hemiaminal in docking calculations (Fig. 4.17A) it may aid with the initial aldehyde binding, formation of the hemiaminal and dehydration to form the iminium. The reduction of k_{cat}^{app} was greater for 4-HPAA than hexanal (Fig. 4.27). It was unclear, though, why the reduction in k_{cat}^{app} was not consistent with different aldehydes. An origin of this difference (according to the dopamine-first mechanism) may be that the Tyr-108 phenol group sterically constrains the intermediates during the reaction, encouraging them to be in conformations conducive to reaction progression. Hexanal is less affected by this loss of the tyrosine phenol as it is less bulky, and so is less sensitive to its presence. This explanation is, of course, speculative; presently there is insufficient information to explain the aldehyde-specific effects of Y108F. Ultimately for 4-HPAA there was a

greater reduction in k_{cat}^{app} than K_m^{app}; this resulted in an overall reduction in k_{cat}^{app}/K_m^{app}. Hexanal, on the other hand, had an equal reduction in these kinetic parameters, so overall k_{cat}^{app}/K_m^{app} was the same as WT (Table 4.4).

Interestingly, although Y108F demonstrated poorer activities with 4-HPAA than with hexanal (compared to WT), the conversions showed the opposite result (Fig. 4.26). The comparative quantity of 4-HPAA product compared to hexanal product was slightly increased. A possible explanation could be if there is a correlation of activity with product inhibition. This, however, has not been observed.

4.3.7 Leu-76

The residue Leu-76 was examined in order to identify aldehyde specific effects. Although the residue is not proposed to play an active role in either enzyme mechanism, it was selected for investigation as it is proposed to interact with substrates in a very different manner in each mechanism. In the HPAA-first mechanism it is not near either the dopamine or the aldehyde substrate so would not be expected to have any influence on the enzyme activity (Fig. 4.6, subunit B). Conversely, in the dopamine-first mechanism it contacts the aldehyde R-group (Fig. 4.18). If the dopamine-first mechanism were correct, it should be possible to mutate this residue and observe aldehyde specific effects on the enzyme activity. Two Leu-76 variants were selected, to examine a progressive change in the residue sidechain: L76V and L76A. These variant were expressed and purified (Fig. 4.20), and were shown to demonstrate reduced stabilities with respect to WT (Table 4.3). The reduction in stability correlates with loss of hydrophobic interactions: the progressive loss of each carbon resulted in a 2 °C reduction in melting temperature.

The removal of methyl groups on the side chain of residue 76 (from leucine to valine to alanine) caused a reduction of activities and conversions (with 4-HPAA and hexanal) (Fig. 4.22, Fig. 4.24). For activities, the relative decrease in 4-HPAA activities was slightly less pronounced than the decrease for hexanal (Fig. 4.26). Strangely, conversions showed the opposite trend, with 4-HPAA conversions falling more sharply than those with hexanal. This was similar but opposite to the trends seen for Y108F, where activities showed a relative switch towards hexanal but conversions moved in the opposite direction. A tentative explanation for this observation is product inhibition. Increased activities may correlate with increased product inhibition which ultimately results in reduced conversions. More experiments are required to unravel these observations.

In order to examine the origin of the fall in activity with Leu-76 variants, kinetic parameters were measured for L76A (Fig. 4.25). If Leu-76 were to form part of the aldehyde binding site it may be expected that the loss of activity would be caused by increases to aldehyde K_m^{app}. This was not observed: aldehyde K_m^{app} values were identical to WT. Dopamine K_m^{app} values were slightly reduced (Fig. 4.27, Table 4.4). The origin of the loss of activity was the reduction in k_{cat}^{app}, which was observed for all substrate combinations. As observed with K_m^{app} values, changes in k_{cat}^{app} were larger when dopamine is the varied substrate. The impact of L76A on k_{cat}^{app} was unexpected, and suggests the residue is not involved in the initial binding of substrates but perhaps in the position of the intermediates as the reaction progresses. Overall, the L76A kinetic parameters demonstrate that there no clear difference in behaviour between 4-HPAA or hexanal.

4.3.8 Ala-79

Sequence alignments between *Tf*NCS and *Cj*NCS2 show differences in the constitution of loop 3 at the entrance of the active site. In *Tf*NCS the sequence from 76 to 80 is LPGAF; *Cj*NCS2 has an alanine inserted between 77 and 78, and an isoleucine in position 79: LPAGIF. The *Tf*NCS variant A79I was examined to investigate the effect of having an isoleucine at position 79. The mutant was expressed and purified in a similar manner to WT. No melting point was recorded so the effect of the mutation on enzyme stability is unknown.

The A79I showed increased activities (between 20-50%) with both 4-HPAA and hexanal (Fig. 4.24). The increased activities were proportional with all substrates: there was no aldehyde specific effect (Fig. 4.26). Kinetic experiments performed to examine the origin of the activity increase were unable to approach substrate saturation, so some parameters have large errors associated (Fig. 4.25), though all fits were statistically significant ($p<0.05$). The origins of the increased activities seem to depend on the aldehyde tested. For HPAA, both K_m^{app} and k_{cat}^{app} values became larger with the amino acid substitution (Fig. 4.27, Table 4.1). Hexanal activity was affected through slight reduction in dopamine K_m^{app} and an increase in hexanal k_{cat}^{app}. The changed dopamine K_m^{app} when different aldehydes were examined is difficult to rationalise with the dopamine-first mechanism as dopamine binding would not be expected to change. This could support the HPAA-first mechanism, or could merely demonstrate the complexity of the kinetics in this reaction—K_m^{app} does not simply correspond to substrate binding.

It is interesting that A79I affected activities with the two aldehydes in different ways. The difference between the aldehydes can be partially rationalised on the basis of the aldehyde bulk. The A79I variant may reduce affinity towards 4-HPAA as an increased side chain bulk disrupts binding. This additional protein bulk does not disrupt the hexanal binding has less bulk. Increases in k_{cat}^{app} may have been the result of increased restriction in reaction intermediates conformations. The isoleucine could also prevent intermediates from exiting the active site prematurely, or encourage the adoption of conformations that progress the reaction.

4.3.9 3,4-Methylenedioxyphenylacetaldehyde

Studying the activities of variants with dopamine and the aldehydes 4-HPAA and hexanal led to the observation of some aldehyde-specific effects (Fig. 4.26). D141E and L76V showed proportionally greater losses of activity for hexanal than for 4-HPAA; Y108F demonstrated the opposite effect, with proportionally greater losses in activity for 4-HPAA than hexanal. The amino acid substitution A79I had similar overall impact on both 4-HPAA and hexanal activity, but the kinetic origins of these changes were different (Fig. 4.27).

In an attempt to draw out more aldehyde-specific effects, especially for variants of the dopamine-first aldehyde binding site (Leu-76, Ala-79, Tyr-108), the activity of variants with a large substrate was examined. The substrate selected was 3,4-methylenedioxy-phenylacetaldehyde (Fig. 4.28A), which had not been previously tested with NCS, but was expected to be turned over based on substrate studies.[53,57] The methylenedioxy group attached to the aromatic ring results in a bicyclic ring, similar to the previously tested 1-naphthylacetaldehyde substrate.

Activities with 3,4-methylenedioxyphenylacetaldehyde were between 10-30% lower than with 4-HPAA (Fig. 4.28B). L76A and Y108F showed the smallest reduction in activities relative to 4-HPAA and A79I the largest (Fig. 4.28D). According to the dopamine-first mechanism, these relative changes could be related to varying steric clashes due to the increased bulk of the aldehyde. However these differences are small: overall, the pattern of activities for the different variants compared to WT was the same as for 4-HPAA. No clear aldehyde-specific effects were observed and so no further assays were conducted with this aldehyde.

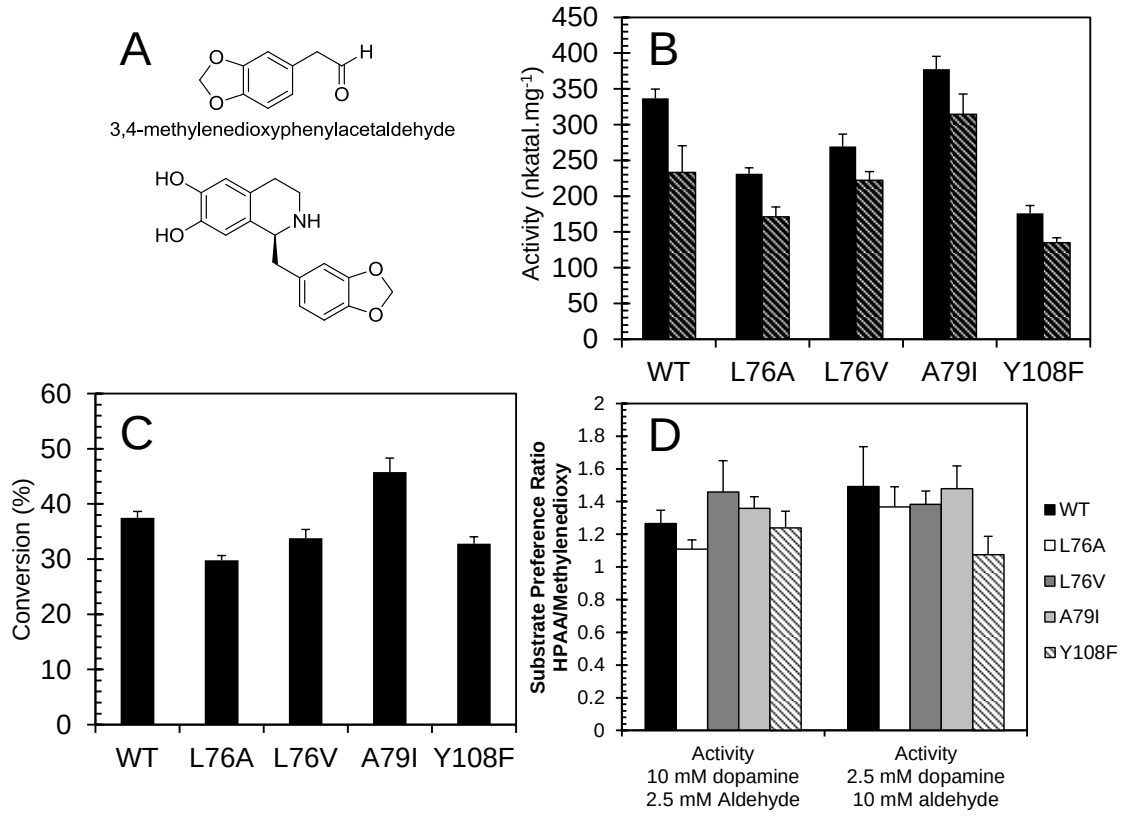


Fig. 4.28 Activities and conversions with 3,4-methylenedioxyphenylacetaldehyde. A. Structure of 3,4-methylenedioxyphenylacetaldehyde and (*S*)-THIQ product with dopamine. **B.** Enzyme activities. Solid bars show dopamine 10 mM, aldehyde 2.5 mM. Lined bars show dopamine 2.5 mM, aldehyde 10 mM. **C.** Conversions, 10 mM dopamine, 10 mM 3,4-methylenedioxyphenylacetaldehyde. **D.** Substrate preference for activities (activity with HPAA divided by activities with 3,4-methylenedioxyphenylacetaldehyde).

4.3.10 Citronellal

Clear aldehyde-specific effects were observed with the variant L76A and the aliphatic aldehydes (*S*)- and (*R*)-citronellal (Fig. 4.29A). L76A demonstrated increased activities compared to WT for both citronellals (with dopamine) (Fig. 4.29 B and C). With 10 mM dopamine and 2.5 mM citronellal, activities with (*R*)-citronellal increased by 30% and activities with (*S*)-citronellal more than doubled. In contrast, activities with hexanal, also an aliphatic aldehyde, fell by 40%. The fact that the two citronellal enantiomers demonstrated different activities with WT and L76A is a strong indication that the effects observed are enzymatic/structural effects rather than just due to general chemical reactivity.

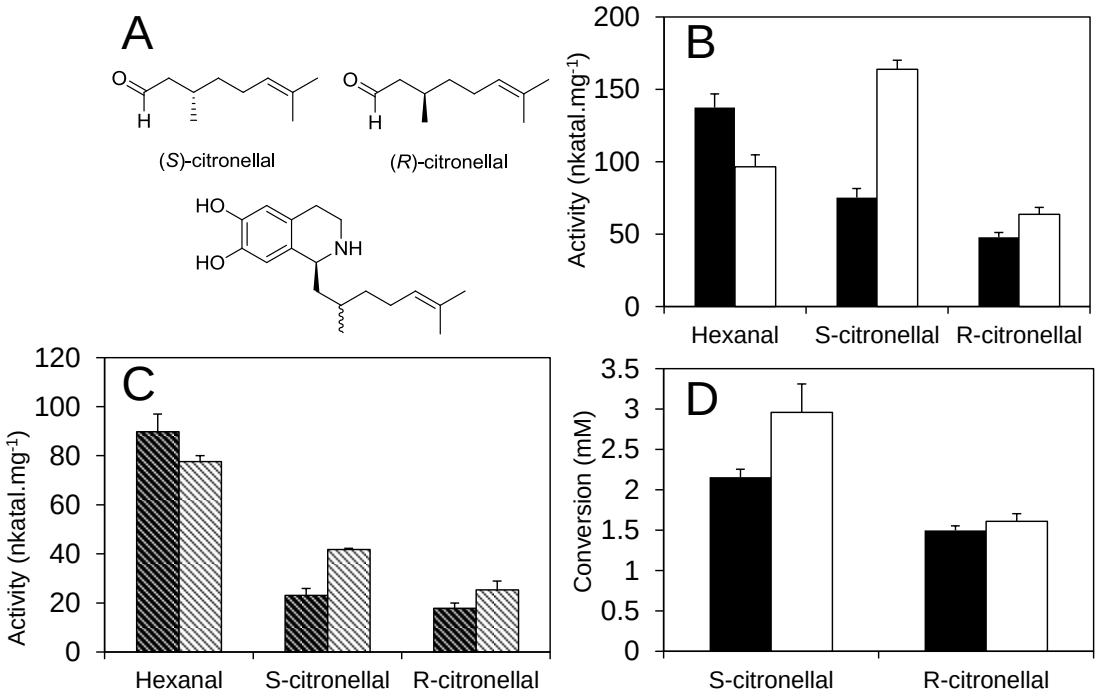


Fig. 4.29 Activities and conversions with (*S*) and (*R*)-citronellal. A. Structure of (*S*) and (*R*)-citronellal. **B** and **C.** Enzyme activities. Black bars are WT, white bars are L76A. Solid bars (B) show dopamine 10 mM, aldehyde 2.5 mM, lined bars (**C**) show dopamine 2.5 mM, aldehyde 10 mM. D**.** Conversions, 10 mM dopamine, 10 mM citronellal. Black bars are WT, white bars are L76A.

The HPAA-first mechanism cannot account for these results, as Leu-76 is not especially near the proposed aldehyde binding site. Docking calculations of imine-citronellal intermediates in the WT active site (Fig. 4.30) predicted a dopamine-first arrangement of the intermediate, with the dopamine portion in close proximity to Lys-122. Notably, the aldehyde R-group is predicted to be in contact with Leu-76. The 3`-methyl groups of the (*S*)- and (*R*)-citronellal occupy positions in the active site tunnel around 3.65 and 3.75 Å away from Leu-76 respectively. Whilst the 3`-methyl from the (*S*)-citronellal points away from the enzyme bulk, the (*R*)-citronellal 3`-methyl points into the enzyme, towards the residues Leu-72, Phe-80 and Met-97. Although the substitution L76A frees space near this methyl group for both citronellal isomers, there remain more steric restrictions on the (*R*)-citronellal. This could account for the different effects between the aldehydes.

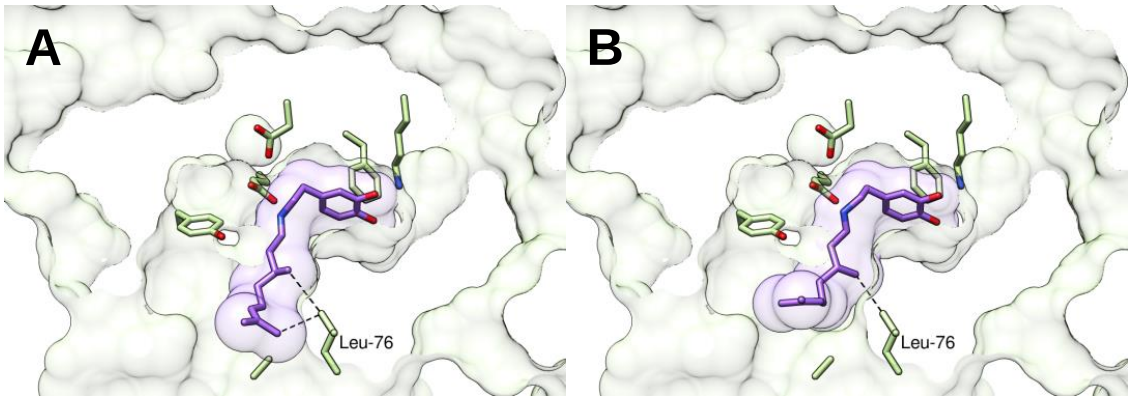


Fig. 4.30 Predicted conformations of citronellal iminium intermediates in the NCS active site. A. Docking of (*S*)-citronellal-*trans*-iminium. B. Docking of (*R*)-citronellal-*trans*-iminium. Black dotted lines show potential steric or hydrophobic interactions between the intermediate methyl groups and L76A. Active site residues carbons are coloured yellow; docked ligands carbons are coloured purple. Ligand accessible surfaces are represented by transparent purple. A section of the protein solvent accessible surface is represented in green. Conformations depicted are lowest energy predictions. Figure adapted from Lichman *et al*.[137]

Kinetic assays were conducted to explore the origin of the aldehyde-specific activity increase with L76A (Fig. 4.31, Table 4.5). For both WT and L76A, dopamine K_m^{app} values with citronellal were lower than K_m^{app} values in assays with 4-HPAA or hexanal. WT had a slightly lower K_m^{app} than L76A in these reactions. For these assays, where dopamine concentration was varied, the origin of the activity increase in L76A was manifested in k_{cat}^{app}, rather than K_m^{app}. The kinetic plots for varying citronellal concentrations (Fig. 4.31) are very different to plots with 4-HPAA or hexanal (Fig. 4.25). When interpreted with Michaelis-Menten kinetics the K_m^{app} values are an order of magnitude lower than any other recorded NCS K_m^{app}: 400-600 μM, compared to the next lowest, 5.2 mM for HPAA and Y108F (Table 4.4). The k_{cat}^{app} values for citronellal isomers are also an order of magnitude lower than those previously recorded. The origin of the L76A increased activity with citronellal appears to be entirely due to improvements in k_{cat}^{app}, rather than reductions in K_m^{app}. The L76A substitution may not affect initial binding of the aldehyde (which is perhaps directed more by Tyr-108) but may affect the behaviour of reaction intermediates. Removing the steric bulk close to the citronellal methyl moiety may enable the intermediate to more easily occupy favourable conformations that are conducive to reaction progression.

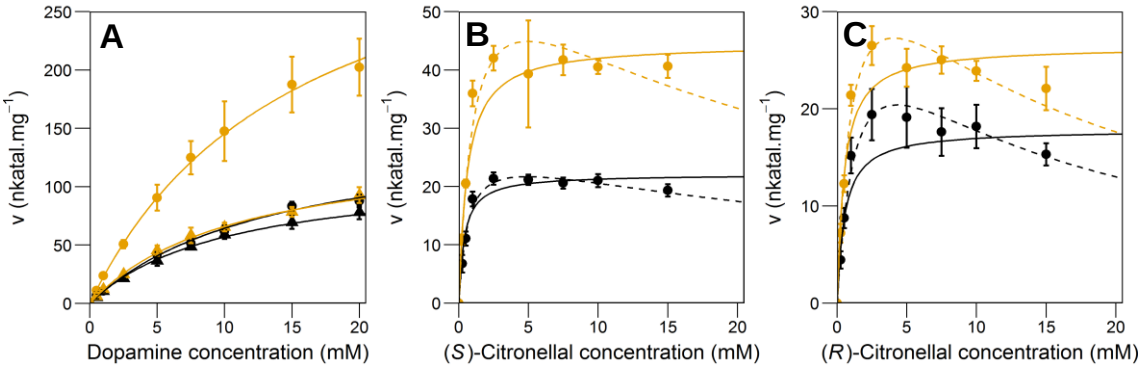


Fig. 4.31 Velocity/substrate concentration curves with citronellal for wild-type (WT, black) and L76A (orange). (A) Varying dopamine concentrations with 2.5 mM (*S*)-citronellal (circles) or (*R*)-citronellal (triangles). (B) Varying (*S*)-citronellal concentrations with 2.5 mM dopamine. (C) Varying (*R*)-citronellal concentrations with 2.5 mM dopamine. Nonlinear least-squares fits: solid lines Michaelis-Menten kinetics, dashed lines substrate inhibition kinetics. Substrate inhibition equation: rate = $V_{max}.[S]/(K_m + [S] + ([S]^2/K_i))$. Data points are the mean from three measurements; error bars indicate SDs of three reactions. For parameters, see Table 4.5. Figure adapted from Lichman *et al*.[137]

Table 4.5 Apparent kinetic parameters for *Tf*NCS with (*R*)- and (*S*)-citronellal (citron.). Corresponding velocity/substrate concentration curves can be found in Fig. 4.31. Table adapted from Lichman *et al*.[137]

Varied Substrate	Constant Substrate	Enzyme	Michaelis-Menten			
			V_{max}^{app}	k_{cat}^{app}	K_m^{app}	k_{cat}^{app}/K_m^{app}
			nkatal.mg^{-1}	s^{-1}	mM	s^{-1}.mM^{-1}
Dopamine[a]	(*S*)-citronellal	WT	153 ± 8	3.2 ± 0.2	13.8 ± 1.0	0.23 ± 0.02
		L76A	368 ± 19	7.7 ± 0.4	15.3 ± 1.0	0.50 ± 0.04
Dopamine[a]	(*R*)-citronellal	WT	115 ± 3	2.4 ± 0.1	10.3 ± 0.4	0.24 ± 0.01
		L76A	138 ± 5	2.9 ± 0.1	11.0 ± 0.6	0.26 ± 0.02
(*S*)-citronellal[b]	Dopamine	WT	22 ± 1	0.46 ± 0.02	0.38 ± 0.10	1.20 ± 0.31
		L76A	45 ± 2	0.93 ± 0.05	0.59 ± 0.09	1.57 ± 0.25
(*R*)-citronellal[b]	Dopamine	WT	17.9 ± 1.4	0.37 ± 0.03	0.53 ± 0.14	0.71 ± 0.20
		L76A	26.4 ± 1.6	0.55 ± 0.03	0.49 ± 0.11	1.13 ± 0.27
			Substrate Inhibition[c]			
			$V_{max,SI}^{app}$	$k_{cat,SI}^{app}$	$K_{m,SI}^{app}$	$K_{i,SI}^{app}$
			nkatal.mg^{-1}	s^{-1}	mM	mM
(*S*)-citronellal[b]	Dopamine	WT	27.4 ± 1.8	0.57 ± 0.04	0.64 ± 0.12	37 ± 11
		L76A	64.1 ± 9.6	1.34 ± 0.20	1.06 ± 0.25	23 ± 11[d]
(*R*)-citronellal[b]	Dopamine	WT	33.9 ± 4.8	0.71 ± 0.10	1.42 ± 0.31	13 ± 4
		L76A	41.9 ± 6.6	0.88 ± 0.14	1.09 ± 0.29	15 ± 6[e]

[a] All kinetic parameters for varied dopamine were recorded with 2.5 mM aldehyde, and dopamine varied from 500 µM to 20mM

[b] All kinetic parameters for varied aldehydes were recorded with 2.5 mM dopamine, and aldehyde varied from 250 µM to 15 mM

[c] Substrate inhibition equation: $Rate = V_{max}.[S]/(K_m + [S] + ([S]^2/K_i))$

[d] Calculated *p*-value > 0.05 for this coefficient (0.081)

[e] Calculated *p*-value > 0.05 for this coefficient (0.054)

The plots for varying citronellal concentrations also seem to show evidence of substrate inhibition (Fig. 4.31, dotted lines). Although substrate inhibition equations appear by eye to fit the data better than the Michaelis-Menten equations, the Michaelis-Menten fits are slightly more statistically significant. For both L76A plots the substrate inhibition $K_{i,SI}^{app}$ term has a *p*-value>0.05. Furthermore, typically the simplest possible model (with fewest parameters) should ideally be used to interpret data (Occam's razor)—in this context this is the Michaelis-Menten fit. More data points would be required at higher substrate concentrations to fully distinguish between these two fits. If substrate inhibition with respect to the aldehyde were present, it would fit with the NMR observation that dopamine did not bind to the enzyme subsequent to aldehyde binding.[54] With the dopamine-first mechanism aldehyde substrate inhibition is straightforward to imagine: the aldehyde binds deep in the active site, preventing dopamine binding and the reaction progressing. There is no evidence, however, for substrate inhibition with 4-HPAA or hexanal.

(*R*)-Citronellal and (*S*)-citronellal demonstrated strikingly different kinetic parameters than 4-HPAA and hexanal: an order of magnitude difference for both K_m^{app} and k_{cat}^{app}. This is slightly surprising, given that citronellal and hexanal appear structurally more similar than hexanal and 4-HPAA. It is also notable that the citronellal concentration with maximum rate was 2.5 mM, which was the concentration of dopamine in the experiments. It appears that citronellal binds to the enzyme more tightly than other aldehydes tested. This may be a result of its increased hydrophobicity. It would be interesting to determine if this reduction in K_m^{app} is a gradual effect caused by increasing hydrophobicity or there is a cut-off where K_m^{app} reduces more suddenly. A more sudden change would be indicative of a shift in kinetic mechanism or of the rate determining step. This could be examined by testing various related substrates with increasingly long alkyl chains.

Overall, the increased activity with citronellal caused by the amino acid substitution L76A is evidence in favour of the dopamine-first mechanism. Although the actual origin of this activity increase cannot be defined based on activity and kinetics data alone, the computational docking data shows the close proximity of Leu-76 and the citronellal R-group in the dopamine-first mechanism. Furthermore, the fact that the (*R*)- and (*S*)-citronellal behaved differently despite having the same chemical reactivity proves this was an effect caused by the enzyme.

4.4 Discussion of mechanism

Overall, computational and experimental data provide more support for the dopamine-first mechanism than the HPAA-first mechanism. The key evidence supporting the dopamine-first mechanism is the increased activity of L76A with citronellal. The effect of this amino acid substitution on activity with citronellal can be partially rationalised on the basis of the computational docking.

4.4.1 Electrophilic aromatic substitution

Due to the corroboration between predicted binding modes and experimental data, the mechanism of the final electrophilic aromatic substitution step in the reaction seems clear (Fig. 4.32). Electrophilic addition occurs on the *Re*-face of the *trans*-iminium cation, establishing the 1*S* chiral centre. This step is catalysed by the interaction of Lys-122 with the 3-hydroxy group of the catechol. In order to achieve the correct (neutral) protonation state to catalyse this step, Lys-122 can exchange protons with the solvent via an adjacent water channel. The electrophilic aromatic addition forms the (1*S*, 8a*R*)-*anti*-quinolinone, defining the product chirality, and the 8a*R* proton is then removed by Glu-110, causing rearomatisation and formation of the (*S*)-THIQ.

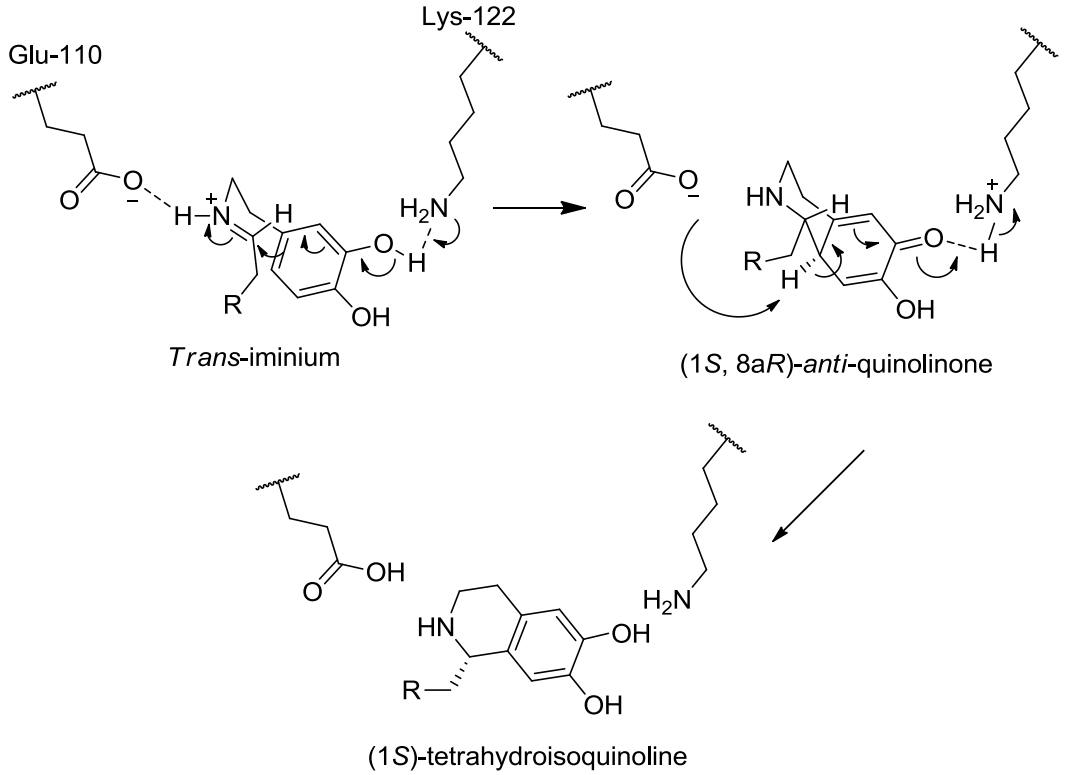


Fig. 4.32 Electrophilic aromatic substitution. Final steps of NCS mechanism.

4.4.2 Iminium formation

The first step in the reaction mechanism is the binding of dopamine in the active site with the catechol adjacent to the Lys-122 and the nitrogen interacting with Glu-110 and Asp-141. The mechanism of the next steps, the formation of the iminium intermediate, is unclear, both with respect to the specific identity of the reaction intermediates and the roles and protonation states of the enzyme residues. The aldehyde must bind: candidates residues for binding to the aldehyde moiety are Tyr-108 and Glu-110. The aldehyde is then attacked by the dopamine; the face of the aldehyde attacked determines whether the intermediate formed is the (*S*)- or (*R*)-hemiaminal. Docking calculations show a strong preference for the (*S*)-hemiaminal intermediate compared to the (*R*)-isomer but there is no experimental evidence to validate this. Overall, one side of the active site entrance (around Tyr-108) is more hydrophilic than the opposite—this difference may indicate a tendency for forming the (*S*)-hemiaminal.

If dopamine is positively charged, it must be deprotonated to enable nucleophilic attack—this can be catalysed by a negatively charged Glu-110 or Asp-141. The electrophilic aldehyde then requires a proton to neutralise the resulting oxyanion. In solution mechanisms this proton is obtained *via* exchange with the charged nitrogen (Fig. 4.33). This is possible in the enzyme mechanism, but the exchange may instead be mediated by a residue, probably Glu-110, based on the docking conformations. It is also possible the aldehyde may enter the enzyme active site as a hydrate. In this scenario, dehydration of the hydrate to form an activated (protonated) aldehyde would require a proton. This could also be taken from Glu-110. Thus regardless of whether the aldehyde enters as a hydrate or an non-hydrated aldehyde, one proton is required from the enzyme (probably Glu-110) to form the hydroxy.

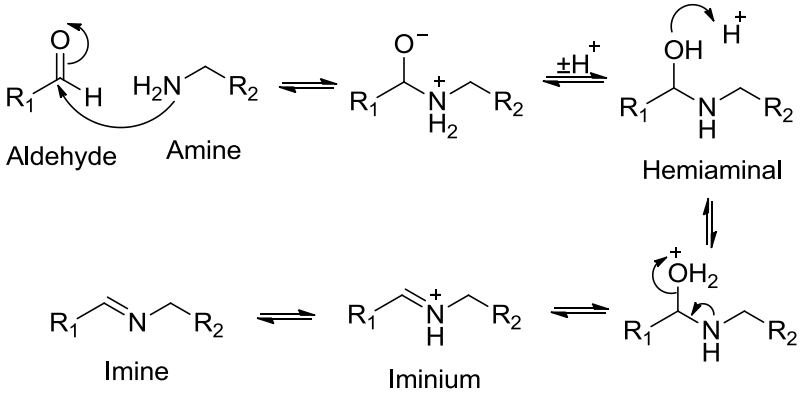


Fig. 4.33 Imine formation. Typical acid-catalysed mechanism for imine formation in solution. An important step is the protonation of the hemiaminal to form a good leaving group (water).

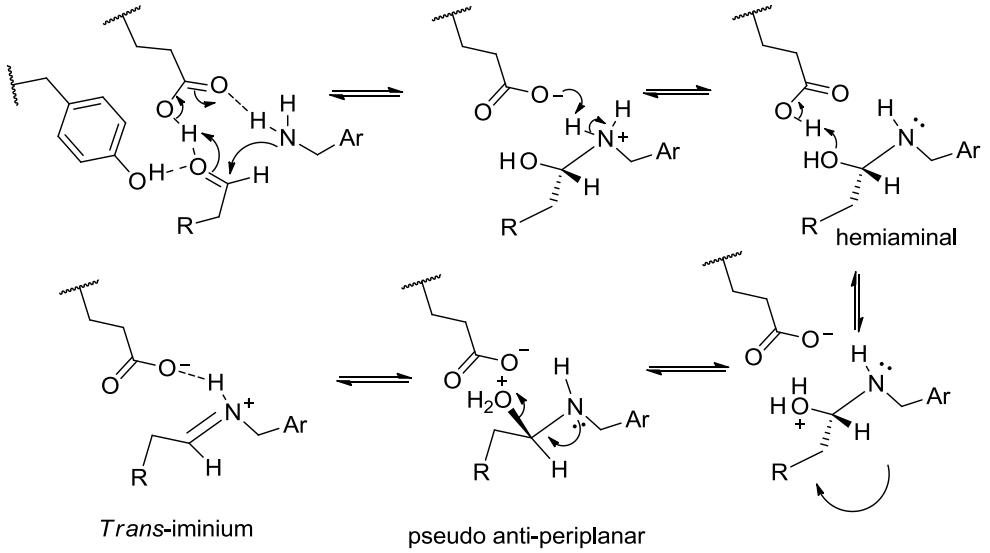


Fig. 4.34 Enzymatic *trans*-iminium formation. Proposed mechanism for *trans*-iminium formation. Prior to the steps drawn here, the dopamine must be deprotonated. The aldehyde may enter as a hydrate – the outcome would be identical. The small bond rotation depicted prior to formation of the iminium double bond is required to establish the *trans*-isomer of the iminium. It is possible that this is not part of the mechanism but rather a matter of depicting the intermediate in a clearer manner.

In the solution mechanism for formation of a Schiff base, elimination of water from the hemiaminal to form an iminium first requires protonation of the hydroxy group to form water, a good leaving group (Fig. 4.33). The docking of the (*S*)-hemiaminal shows the Glu-110 oxygen less than 3 Å away from the hydroxy group, and this residue is therefore the best candidate for protonating the hydroxy group (Fig. 4.34). But, as described above, the Glu-110 is likely to be deprotonated, having donated a proton to the hydroxy group (or by dehydrating the aldehyde hydrate). A proton can be provided to the hydroxy group by the adjacent protonated secondary amine. This proton exchange may be modulated by Glu-110 or direct exchange with the amine may occur. In all reasonable mechanistic scenarios, the amine is deprotonated prior to formation of the iminium. The presence of a lone pair on nitrogen at this point in the mechanism indicates that no base is required at the elimination step, as electron donation will come directly from the nitrogen lone pair. The neutral nitrogen is also able to invert its configuration, and is able to have optimal overlap with the C-O σ*-orbital at elimination (Fig. 4.34).

The prediction that Glu-110 or the oxyanion deprotonates the secondary amine means Asp-141 does not catalyse the elimination step, despite its close association with the (*S*)-hemiaminal nitrogen. Mutational investigation into Asp-141 showed that the chemistry/charge of the residue's side chain is influential in the reaction mechanism, but its position and H-bonding is less important. Its charge may stabilise reaction

intermediates and modulate Glu-110 pK_a. The only possible catalytic role Asp-141 may play is catalysing the initial nucleophilic attack of dopamine to form the hemiaminal.

There is an entrance to the active site on the other face of Tyr-108 to the main entrance. This tunnel meets with the main active site between Tyr-108 and Glu-110. It is lined with hydrophilic residues: Glu-103, Tyr-131, Tyr-139, Thr-159 and Asp-141 (Fig. 4.35). This is a water channel, and waters are present in this channel in the crystal structures. In the docking studies, the (*S*)-hemiaminal hydroxyl group is perfectly positioned to leave through this water channel when formation of the iminium occurs.

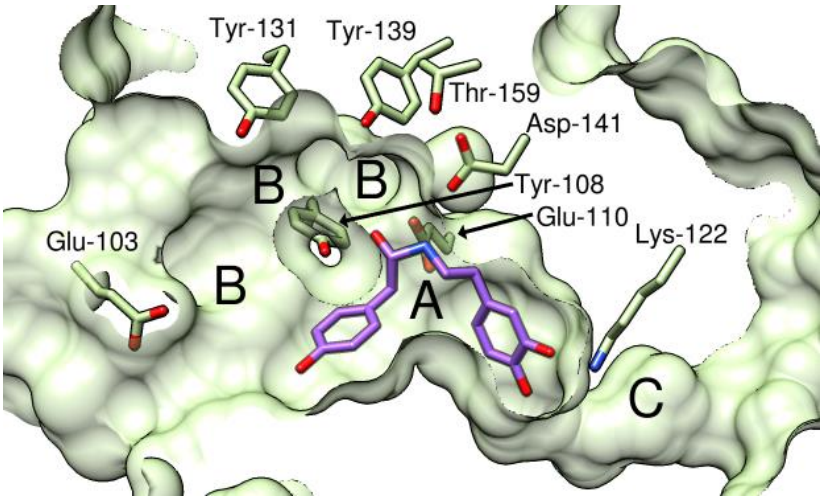


Fig. 4.35 Water channels in NCS active site. A. Main substrate binding site (showing (*S*)-hemiaminal intermediate docked, purple sticks). **B.** Route of water channel, around Tyr-108, lined with hydrophilic residues. **C.** Water channel near Lys-122, allowing exchange of protons. Side chains depicted as yellow sticks. A section of the protein solvent accessible surface is represented in green. Taken from PDB file 2VQ5, subunit A.

The exact protonation states of Glu-110, Asp-141 and the substrate nitrogen at the start of the mechanism are unknown: there are eight potential arrangements (2^3). It is, however, only possible to draw reasonable mechanisms from five of these starting points based on key mechanistic requirements: Glu-110 protonation of the hemiaminal hydroxyl leaving group and Glu-110 deprotonation of quinolinone intermediate (Fig. 4.32). It is not possible to fully differentiate between these mechanisms based on the present experimental evidence. The only differences between the mechanisms are the overall protonation state of Asp-141 and the initial exchange of protons to ensure a neutral nitrogen nucleophile and a protonated Glu-110 (Fig. 4.36).

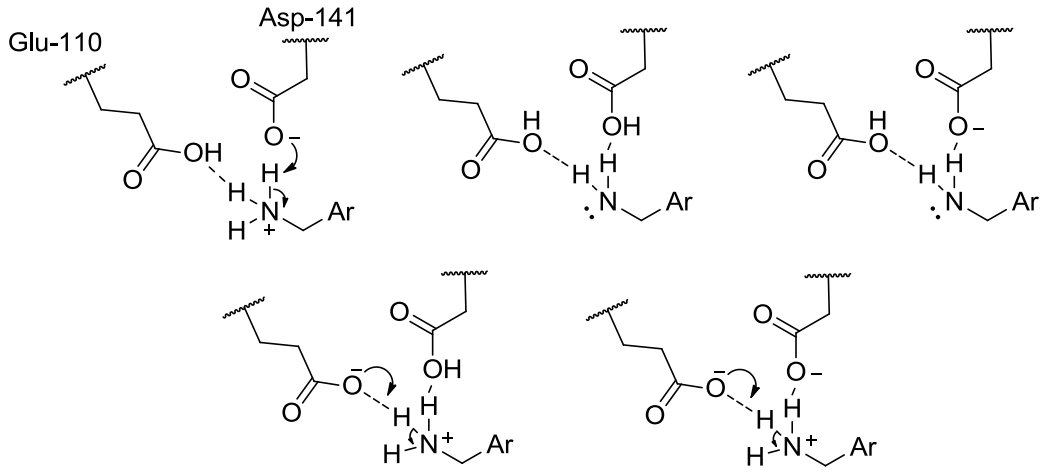


Fig. 4.36 Productive protonation states of active site. Different possible stating states of active site generated by varying Glu-110, Asp-141 and dopamine protonation states. Arrows show exchange of protons required to reach active site state conducive to reaction progression. Three other possible states exist but reasonable mechanisms for them cannot be drawn. Leftmost state is that predicted by computational predictions: Glu-110 neutral, Asp-141 charged and dopamine protonated.

4.4.3 Conformational change

The predicted conformations of the (*S*)-hemiaminal and the *trans*-iminium in the active site showed a crucial difference. Whilst in the (*S*)-hemiaminal binding mode the nitrogen occupied a position near Glu-110 and Asp-141, in the iminium conformation the nitrogen moved away from Asp-141 and towards the catechol ring (Fig. 4.17A, Fig. 4.18A). This change was enabled by a bond rotation in the ethylamine part of dopamine. The ethylamine in the (*S*)-hemiaminal was in a staggered conformation, but in the iminium had a pseudo-chair conformation (Fig. 4.37). Mechanistically, this conformational change is crucial for enabling the electrophilic aromatic substitution step, and it is remarkable that the docking investigation revealed this in the binding modes. It is worth bearing in mind that this proposed bond rotation is not as extreme as the rearrangement required in the HPAA-first mechanism (section 4.1.2), and, furthermore, the aldehyde R-group that moves does so in the relatively wide active site entrance. Nevertheless, the apparent movement of the aldehyde R-group in this step is extremely interesting—the group appears to move further towards the residues Leu-76 and Ala-79, suggesting closer interactions in the iminium than the hemiaminal.

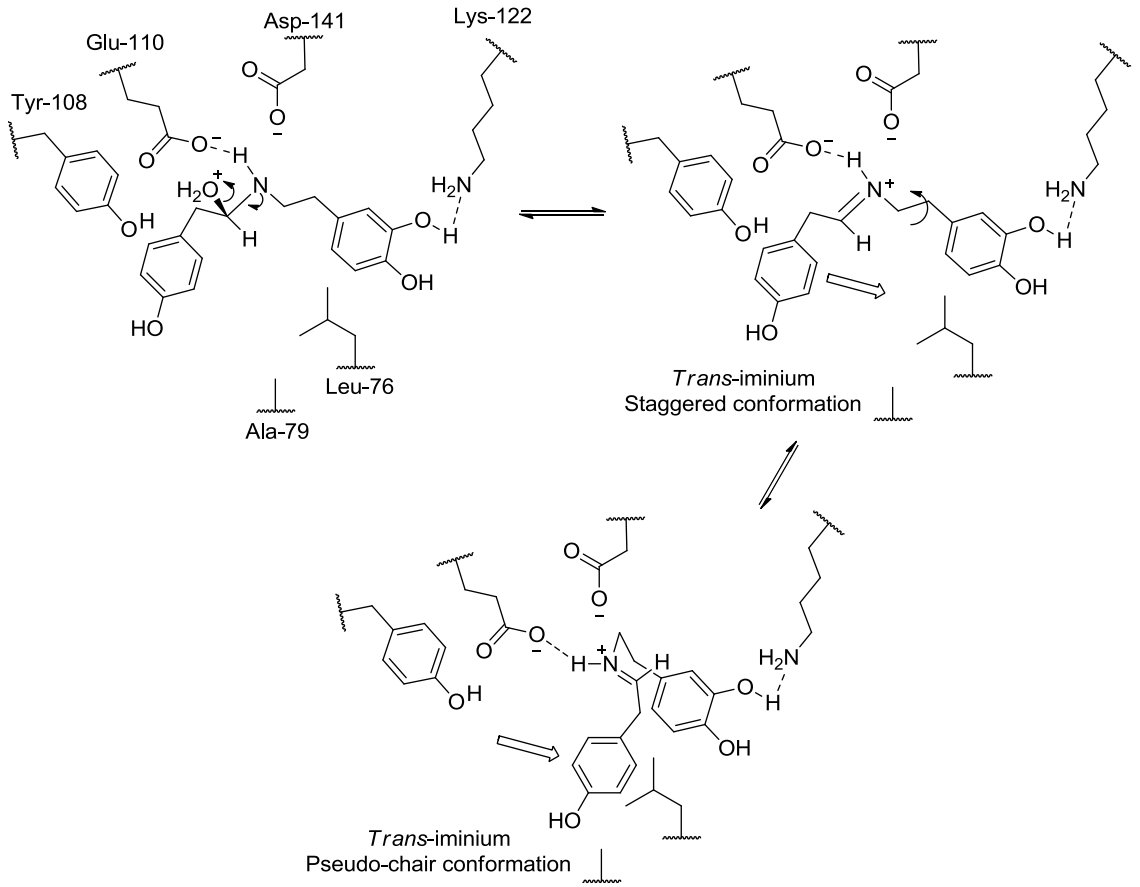


Fig. 4.37 Proposed conformational change in *Trans*-iminium. The thin black arrows show either movement of electrons or bond rotation. The clear block arrows show the movement in the aldehyde R-group during the reaction mechanism. This uses the aldehyde 4-HPAA as an example.

It is likely that this rearrangement plays a role in the enzyme mechanism as a pseudo-chair conformation of the intermediate would be the lowest energy conformation of the iminium at the intramolecular cyclisation step. Inclusion of this step in the reaction mechanism may aid rationalisation of some kinetic parameters, especially those for L76A. The activity changes caused by the L76A substitution are primarily due to k_{cat}^{app} and not aldehyde K_m^{app} values, despite Leu-76 being in the proposed aldehyde binding site. The mutation, though, may have greatest influence not on the initial substrate binding, but on this conformational change. The increased ease of this conformational change would result in an increased k_{cat}^{app}. The substitution L76A may therefore stabilise this conformational change for the bulky citronellal iminium intermediates.

4.4.4 Mechanistic unknowns

Based on computation and experimental data, it has been possible to propose a full mechanism for NCS (Fig. 4.38). Despite this, there remain many uncertainties about the NCS mechanism. Structural information with ligands (e.g. dopamine, reduced iminium) would test the veracity of the dopamine-first mechanism, which currently relies heavily on computational docking. Testing enzyme activities over different pH values may reveal the real pK_a values of active site residues; this can be compared to pK_a predictions (Table 4.2).

The specific roles of numerous residues are unknown. For example, based on structural, computational and experimental data it is expected that Tyr-108 and Phe-112 are expected to be somehow involved in the enzyme mechanism. Preliminary screening results with the Phe-112 variant F112L shows activity was present, but reduced compared to WT (see section 6.6). More in depth investigation into this interesting residue is required to understand its role in substrate binding, its interaction with Lys-122 and its apparent dynamic nature. This screening also showed that Met-97 has an influence on the enzyme activity; some Met-97 variants had improved activities with certain substrates (section 6.6). The residue Tyr-108 occupies a crucial position in the enzyme active site—it helps define the shape of the active site entrance and the water channel. It was also predicted to bind to the aldehyde and hemiaminal. The kinetics of the variant Y108F showed a reduction in K_m^{app} for the aldehydes HPAA and hexanal. This could not be rationalised simply, and the precise role of the residue remains unclear.

A key prediction of the dopamine-first mechanism is the binding order of substrates: dopamine binds in the active site before the aldehyde. A verification of this would be more evidence for this enzyme mechanism. Attempts to determine the binding order of substrates using product inhibition and Hanes plots were unsuccessful, primarily due to the background reaction and high substrate concentrations. Alternatively, the observation of substrate inhibition with the aldehyde may also add to evidence for the dopamine-first mechanism. Reactions with citronellal possibly show substrate inhibition, but clearer data with a variety of aldehydes is required to prove it is a feature of the enzyme kinetics.

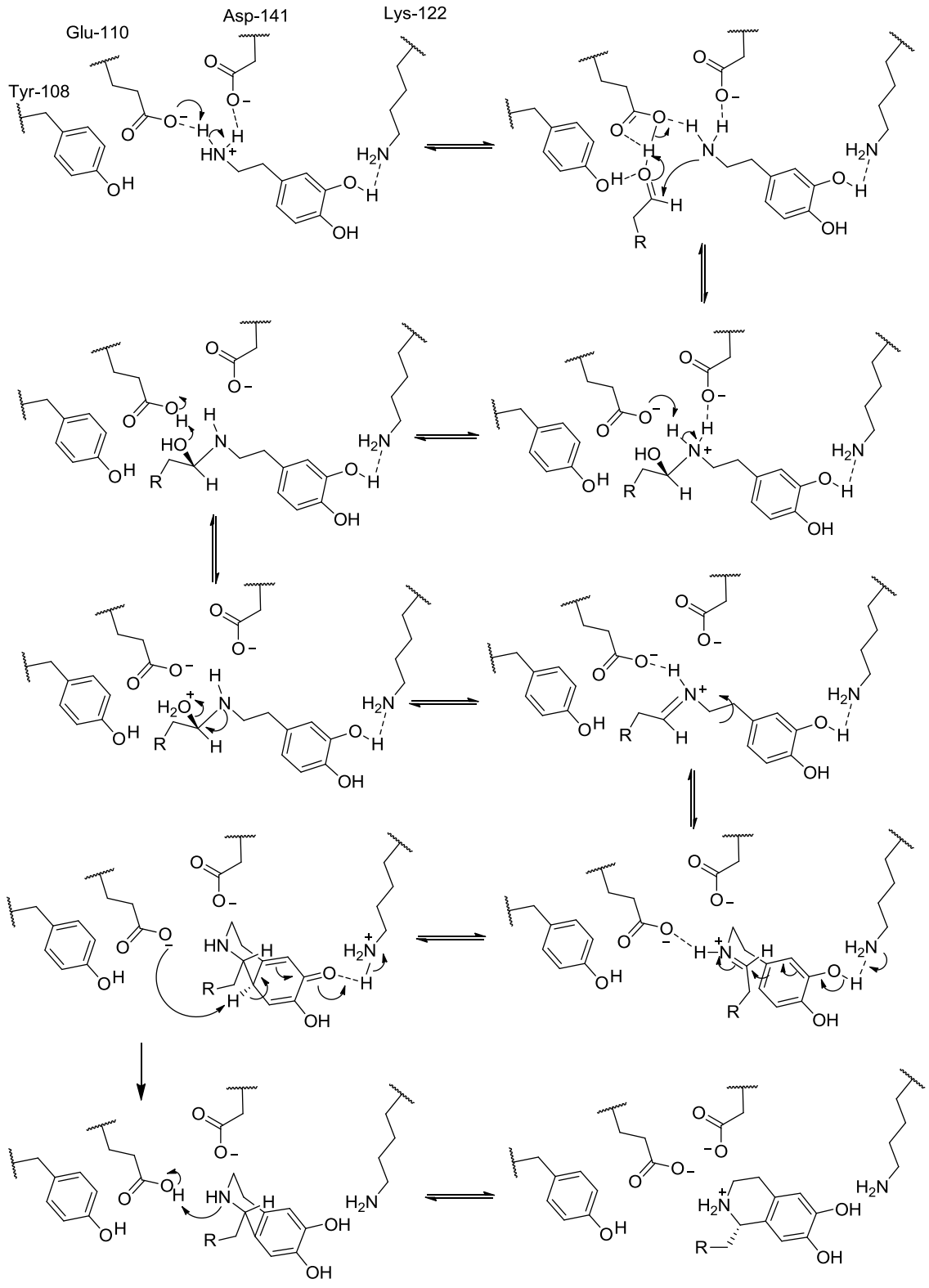


Fig. 4.38 Proposed dopamine-first mechanism for NCS. The roles of Tyr-108 and Asp-141 are unclear but suggested interactions are included as an hypothesis.

4.4.5 Kinetic challenges

Overall, the kinetic data has proven to be difficult to rationalise in terms of the dopamine-first mechanism. This is influenced by a range of factors including the high background reaction, the complexity of the mechanism and the remarkably high K_m^{app} values. It is important to note that when dealing with a bimolecular reaction mechanism, the apparent Michaelis-Menten parameters k_{cat}^{app} and K_m^{app} are collections of numerous rate constants and as such the parameters cannot be interpreted as they can for simpler reactions. The fact that the substrates kept constant were not at saturating concentrations is very detrimental to analysis of the kinetic parameters—using a saturating concentration of substrate is vital so the effects of varying one substrate can be observed clearly.

The kinetics of the Pictet-Spengler reaction catalysed by NCS is exceptionally complex as the reaction has multiple reversible steps, numerous possible intermediates and reaction steps that can occur outside the enzyme. There is also some tentative evidence of substrate and product inhibition, and that reaction intermediates are able to diffuse in and out of the active site. For example, the first step in the NCS mechanism, iminium formation, occurs readily in solution around neutral pH, so it is possible that these intermediates could diffuse into the active site from the solvent. The iminium may be more favourable in solution for some aldehydes than others so different substrates may have different kinetic mechanisms (this is a possible origin of the kinetic differences of HPAA/hexanal and citronellal).

The effect of the co-solvent on NCS activity is currently unknown. In order to solubilise the aldehydes, all enzyme activity assays contributing towards the kinetic parameters contained 10% v.v^{-1} MeCN, The presence and concentration of this organic solvent may have affected the kinetic parameters, perhaps contributing to the high K_m^{app} values. Investigation of different aldehydes and ketones as substrates for NCS demonstrated that co-solvent can be very influential for activity, but its affect differs with the substrate investigated (see sections 6.5.2 and 6.6.4). The degree to which the co-solvent influences the enzyme stability, substrate solubility and imine formation is unknown, but it probably plays a role in all these factors. Further exploration of the NCS activity and kinetics must include a thorough investigation on the effect of co-solvent.The combination of very high K_m^{app} values with large background reaction rates at high substrate concentrations is a major problem that prevents kinetics experiments from reaching saturating concentrations. At high substrate concentrations there would be increased formation of imines outside the enzymes and possibly increased catalysis

from dopamine itself. Improved experimental methods such as quenched flow or CD stopped flow may enable the collection of improved kinetic parameters.

4.4.6 Catalytic inefficiency

Despite the challenges experienced in collecting kinetic data, the apparent kinetic parameters collected may still be indicative of the enzyme behaviour. The parameters describe an enzyme that is catalytically inefficient. For its natural reaction NCS exhibited an apparent k_{cat}/K_m of 10^3 $s^{-1}.M^{-1}$ (varied HPAA). In a global analysis of kinetic parameters, where thousands of kinetic enzyme reactions were compared, the median value for k_{cat}/K_m was determined to be 10^5 $s^{-1}.M^{-1}$. NCS appears to be amongst the least efficient enzymes, operating with two orders of magnitude lower catalytic efficiency than the median values.[5] The origin of this comparative inefficiency is the K_m^{app} value, which was in the region of 20 mM (the median is 0.13 mM). The k_{cat}^{app} (around 20 s^{-1}) was, on the other hand, above the median value of 13.7 s^{-1}.

The high K_m^{app} means that high substrate concentrations are required for the enzyme to demonstrate turnover that approaches its fastest rate. In *Thalictrum flavum*, NCS was present in the roots and rhizomes,[141] and in *Papaver somniferum,* it was found in the sieve tube elements of the phloem.[44,69] The concentration of NCS substrates in *Thalictrum flavum* are unknown, but metabolomics studies in *Papaver somniferum* and *Papaver bracteatum* offer some insight into the absolute metabolite concentrations in alkaloid producing plants.[142,143] In a low-alkaloid *P. somniferum* strain tyramine (the precursor to dopamine) accumulated to 5 mM in the latex (NMR analysis). Other metabolites which do not accumulate (e.g. amino acids) were in the approximate range 0.5-10 mM (NMR analysis). In *P. somniferum* and *P. bracteatum* dopamine was found at concentrations of 1 $mg.mL^{-1}$ (5 mM) in the latex.[144] This surprising abundance of the NCS substrate dopamine in alkaloid producing tissues will enable NCS to have a reasonable turnover despite a high K_m^{app}. Subcellular trafficking of NCS and its substrates will also play a role in regulating turnover.

NCS catalyses the first committed step into the benzylisoquinoline alkaloids, and, as such, it is vital for alkaloid production.[44] It is abundant in alkaloid producing tissues: in *P. somniferum* stems NCS is the most abundant protein of all in the morphine pathway, and it is in the top thirty of all proteins present.[69] The position of NCS at the entry to alkaloid pathways allows it to control the metabolic flux through these pathways. Assuming *in vitro* kinetics are similar to those *in vivo*, the high K_m^{app} of the enzyme will ultimately result in entry to alkaloid pathways being proportional to the concentrations of its precursors. The substrates dopamine and 4-HPAA are tyrosine derivatives and

therefore are closely linked to primary metabolism. In times of stress, when primary metabolites are relatively scarce, it would be undesirable for the plant to deplete the primary metabolites in order to produced specialised chemicals that are not vital for its survival. The high K_m^{app} values of NCS prevent significant flux through the alkaloid pathways when dopamine or HPAA concentrations are low. The NCS kinetics may result in the enzyme acting as the gatekeeper to the BIA pathway—any change in substrate concentration will elicit a corresponding change of flux into the BIA pathway. The high K_m^{app} values therefore prevent complete depletion of substrates that may be required in other pathways.

4.4.7 Enzyme engineering

Plant secondary metabolism enzymes often display low catalytic efficiency, perhaps due to low selection pressure, or, as is possible in the case of NCS, requirement for metabolic control.[5,145] Although the enzyme may be optimised for its particular task in the plant, it may not be fully optimised for its actual catalytic activity. It may be desirable to improve the enzyme catalysis for use in *in vitro* biocatalysis, *in vivo* metabolic engineering or even *in planta* genetic modification. Recombinant NCS has been used in alkaloid producing microbial systems. However, in a number of these systems, NCS does not appear to be active.[72–74] In a yeast system that produces (*S*)-reticuline from glucose, *Ps*NCS3 was shown to be active, but the inefficiency of the step caused a 228-fold drop in yield.[45] In another yeast system, two copies of the NCS gene were required to increase yield.[21]

The variant A79I is evidence that it is possible to increase the catalytic efficiency of *Tf*NCS towards its natural substrate through mutation. This variant had an overall increased apparent k_{cat}/K_m, but also increased K_m^{app}s. Variants with reduced K_m^{app} values also had reduced k_{cat}^{app}s (and overall reduced apparent k_{cat}/K_m). In order to generate significant improvements in the NCS catalytic efficiency then a separation of the correlation between K_m^{app} and k_{cat}^{app} is desired. Ideal variants for metabolic engineering applications are those with reduced K_m^{app} values and increased k_{cat}^{app} values. It is possible, though, that the mechanism of the reaction places a speed-limit on the enzyme kinetics. The high K_m^{app} may be a necessary result of the deep substrate binding site, and the fact that substrates may enter in the wrong order or orientations and block activity.[146]

The activity of the *Tf*NCS variant L76A with (*S*)- and (*R*)-citronellal is the first demonstration that the NCS substrate preference can be modified through mutation. The rationalisation of this change with the dopamine-first mechanism signifies that it is

possible to design mutations based on this mechanism to engineer changes into the enzyme. This can lead to the future generation of NCS variants capable of catalysing the stereoselective formation of diverse tetrahydroisoquinolines.

5 Chapter 5: Biotransformations and cascades

5.1 Introduction

The THIQ moiety is a “privileged scaffold” for drug design.[92] Privileged scaffolds are small organic structures that have shown significant precedent for modulating biological systems, through their presence in numerous bioactive natural products and/or their prevalence in synthetic drugs (Fig. 5.1). The THIQ moiety is found in numerous plant BIAs with biological activities, such as glaucine (antitussive),[147] norcoclaurine (anti-HIV),[148] and tubocurarine (anesthetic/poison) (see section 1.1.2).[149] Non-ribosomal peptide derived THIQs are produced by micro-organisms and include saframycin[88,89] and trabectedin[150] (both have anti-cancer properties) (see section 1.7.4). The THIQ is also present in synthetic drugs such as almorexant (anti-insomnia),[151] and solifenacin (antimuscarinic).[152] The (*S*)-THIQ CKD712, known to be a product of *Tf*NCS,[53] has recently been shown to have a number of bioactivities, including inhibition of cell signalling pathways[153] and wound-closing effects.[154]

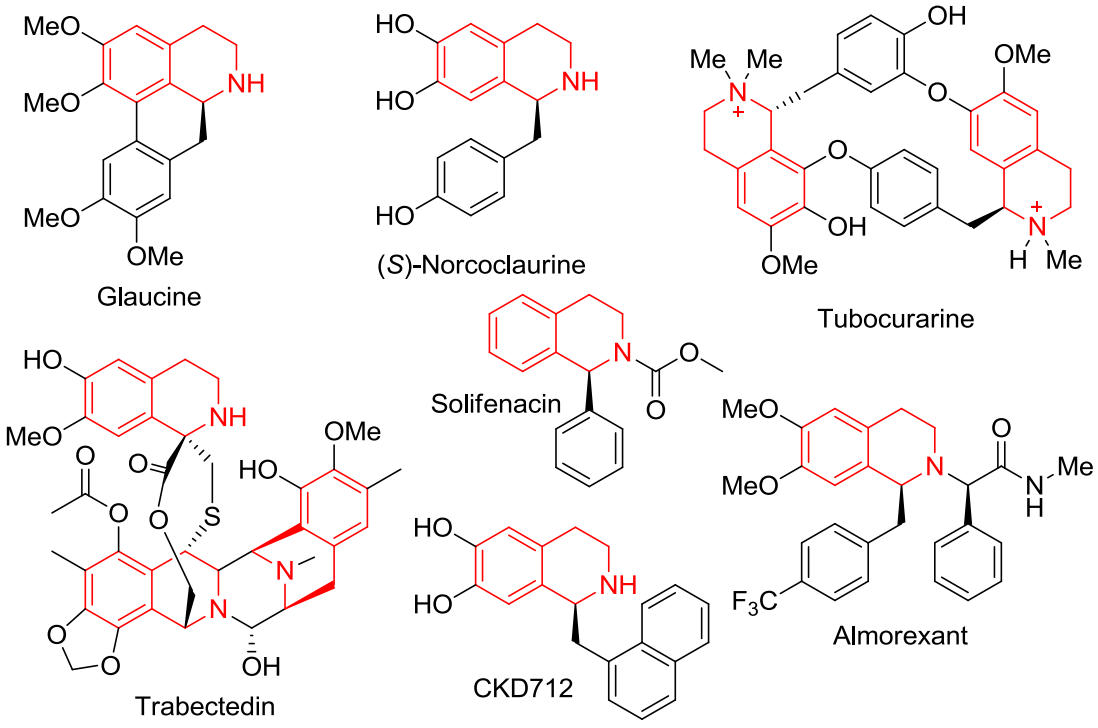


Fig. 5.1 The THIQ moiety in natural products and drugs. THIQ moiety highlighted in red.

Compounds with privileged scaffolds are considered good targets for drug discovery, and therefore they are common targets for synthetic organic chemistry. Discussion of synthetic approaches to THIQ can be found in section 1.8. To summarise, recently, two

catalytic methods for the facile PS formation of THIQs have been developed. The first uses a calcium hexafluoroisopropoxide catalyst in organic conditions (and can also accept ketones, see section 6.6).[93] The second, developed by Hailes and co-workers, uses phosphate buffer in mild aqueous conditions.[59] A chiral method based on the phosphate catalysed reaction has also been developed, by a different group.[95] In the last decade there also been numerous advances in methods for forming chiral THIQs *via* hydrogenation[155] and non-classical asymmetric approaches.[156]

Enzymes are exceptional catalysts for use in sustainable organic chemistry.[157] They exhibit high chemo-, regio-, stereo- and enantioselectivities and operate in mild, aqueous conditions. Developments in directed evolution means that enzyme properties can now be modified relatively easily—changes to substrate scope, thermostability, solvent tolerance and even reaction selectivities are regularly reported.[107–111] It is also becoming possible to construct whole syntheses using enzymes,[158] and use them in a single pot (see section 5.4.1).[159] The chiral THIQ is an ideal target for biocatalysis.

As described in section 1.8, a number of enzymes have been used for the biocatalytic formation of chiral THIQs: BBE, IRED and MAO.[104] Crucially, these methods all require prior synthesis of THIQ or the 1,2-dihydroisoquinoline moiety before the biocatalytic step. This means that using these methods still requires a considerable synthetic effort. NCS, on the other hand, performs the whole PS reaction, converting an arylethylamine and an aldehyde to a (*S*)-THIQ. The synthesis of the THIQ and formation of the chiral centre is concurrent. Furthermore, the enzyme does not require cofactors or regeneration.

There are four publications describing preparative scale production of chiral THIQs using NCS. The first, in 2010, was a gram scale synthesis of (*S*)-norcoclaurine, using hypochlorite to generate 4-HPAA from tyrosine, and *Tf*NCS as the catalyst.[58] A recent study has expanded this method to halogenated compounds.[60] The other studies have used *Cj*NCS2 to produce a variety of (*S*)-THIQs including 1-alkyl substituted compounds.[57,61]

This section describes efforts to use *Tf*NCS and *Cj*NCS2 in biotransformations. This involved the measuring of specific activities, the investigation into different conditions for biotransformations and, finally, the establishment of a multi-enzyme cascade method.

5.2 Specific activities

Specific activities of purified Δ29*Tf*NCS and *Cj*NCS2 were measured by incubation of the enzymes with 10 mM dopamine and 10 mM 4-HPAA at 37 °C (Fig. 5.2). Enzyme concentrations in these experiments were 0.11 mg.mL^{-1} for Δ29*Tf*NCS and 0.08 mg.mL^{-1} for *Cj*NCS2. Enzyme activities plateau well below full conversion (10 mM): the fitted curves suggest maximum conversions are 5.9 mM and 6.6 mM for Δ29*Tf*NCS and *Cj*NCS2 respectively. This plateauing is perhaps due to product inhibition, enzyme instability or degradation of substrates/products. For both enzymes data points at 30 minutes are lower than the expected value; this, again, was possibly due to product degradation.

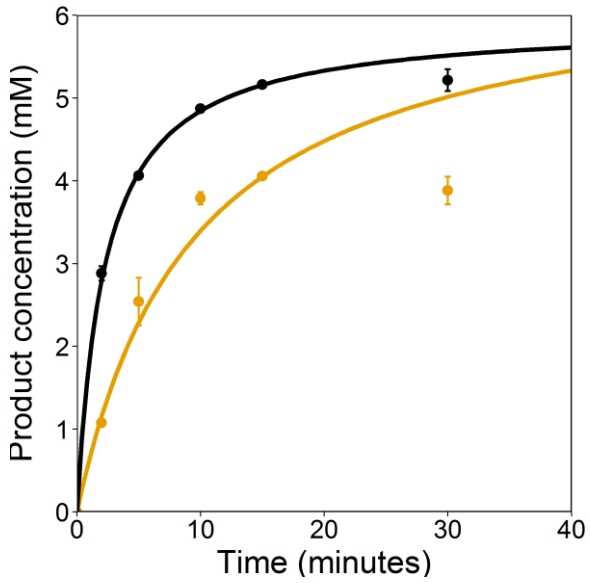


Fig. 5.2 Specific activity of NCS. Δ29*Tf*NCS (black) and *Cj*NCS2 (orange). Points are the mean of three measurements values, error bars are one standard deviation from mean. Lines are equations: $= \frac{Vt}{t+K}$, the fits have been weighted (to 1/SD^2). Enzyme concentrations are 0.106 mg.mL^{-1} for Δ29*Tf*NCS and 0.08 mg.mL^{-1} for *Cj*NCS2.

Conversions were measured after 2, 5, 10, 15 and 30 minutes. The specific activity of *Cj*NCS2 can be estimated by assuming the first three points of the time course (including zero) are linear and calculating the gradient of the straight line. This was the method used previously by Pesnot *et al* to calculate the specific activity of *Cj*NCS2.[57] Using this method here to calculate the specific activity of *Cj*NCS2 provided a value of 6.32 U.mg^{-1}. This was in the same order of magnitude as the value of 4.4 U.mg^{-1} quoted in Pesnot *et al*.[57] The difference between these specific activity measurements may be in the conversions from the raw data (HPLC peak area) to product concentration, and the presence of TFA in the chemical standards (see section 5.3.5). If a TFA counterion was accounted for in the Pesnot *et al* calculations then the *Cj*NCS2 activity rose to 6.25 U.mg^{-1}, very similar to that calculated here.

Assuming linearity in the first few points in a time course is not appropriate in many cases. When rates are particularly fast, as with Δ29*Tf*NCS, there is no section of the curve which is approximately straight: the rate is constantly changing as the substrates are consumed (Fig. 5.2). One solution to this is to use lower enzyme concentrations to obtain time courses that are approximately linear. However, it is also possible to use a non-linear regression fit to calculate the initial rate numerically. The fitted equation can be differentiated and the gradient calculated at any given time. Unsurprisingly, with this method the gradient varies significantly depending on which time point is selected. Strictly, specific activity should be calculated from the initial gradient, at t=0, which for *Cj*NCS2 gave a specific activity of 8.75 U.mg^{-1} (Table 5.1), around 40% larger than that calculated with the linear rate method.

Table 5.1 Slope of time course (rate) at different time points calculated by differentiation of fitted equation: $\boldsymbol{c = \frac{Vt}{t+K}}$. Derivative function: $\boldsymbol{Rate = \frac{K+V}{(t+K)^2}}$. Raw data in Fig. 5.2.

NCS	Time (min)	0	0.1	0.5	1	2	5
Tf	Rate (mM.min^{-1})	2.66	2.43	1.77	1.27	0.74	0.25
Tf	S.A. (U.mg^{-1})	25.1	22.9	16.7	12	6.98	2.36
Tf	S.A. (nkatal.mg^{-1})	418	382	278	200	116	39.3
Cj	Rate (mM.min^{-1})	0.7	0.68	0.63	0.57	0.48	0.3
Cj	S.A. (U.mg^{-1})	8.75	8.5	7.88	7.13	6	3.75
Cj	S.A. (nkatal.mg^{-1})	146	142	131	119	100	63

The specific activity of Δ29*Tf*NCS (using numerical method at t=0) was 25.1 U.mg^{-1}. This value was in accordance with activity measurements for Δ29*Tf*NCS in the mechanism studies, though those experiments used lower concentrations of substrates and a different experimental method (Chapter 4). In the mechanism investigation, the specific activity of WT Δ29*Tf*NCS with 10 mM dopamine and 2.5 mM HPAA was determined to be 426 ± 22 nkatal.mg^{-1} and with 2.5 mM dopamine and 10 mM HPAA, 348 ± 12 nkatal.mg^{-1} (see section 4.3). These values were in reasonable agreement with the value determined here, with both substrates at 10 mM: 418 nkatal.mg^{-1}.

Δ29*Tf*NCS had a larger specific activity than *Cj*NCS2 but also experienced a sharper plateauing of activity through the time course. This was reflective of the high K_m values calculated for Δ29*Tf*NCS (section 4.3); the poor binding of substrates means that high substrate concentrations are required to maintain good activities. There was also a possible effect of product inhibition, though this has not been verified.

5.3 Biotransformations

5.3.1 *Cj*NCS2 and *ortho*-product

A biotransformation was performed with purified *Cj*NCS2, dopamine and hexanal. The reaction was conducted in 85 mL HEPES buffer (50 mM, pH 7.5), with 10 mM dopamine, 15 mM hexanal and 0.014 mg.mL^{-1} protein. After 3 hours (at 37 ˚C) the reaction was quenched with acid, filtered, and purified by preparative-HPLC. Two two fractions were obtained. NMR analysis confirmed the first fraction was the expected *para* substituted product and the second fraction a mixture between the *ortho* product (66%) and *para* product (33%). The major *para* product was obtained with overall 57% purified yield (170 mg). Comparison of the purified *para* compound with a racemic standard by chiral HPLC showed the major biotransformation product had an *ee* of 78% (Fig. 5.3). This was inadequate for a chiral biotransformation and showed the presence of the competing racemic chemical reaction in these conditions—a simple way to improve the *ee* would be to have increased the enzyme concentration.

The *ortho* product was also analysed by chiral HPLC, and although a verified racemic *ortho*-standard was not available, comparison of the data with the racemic *para* product suggested that the *ortho* product was racemic. The production of the *ortho*-THIQ as a minor product in a biotransformation was reported by Pesnot *et al* (with heptanal rather than hexanal) but chiral analysis of this product was not performed.[57] The analysis presented here suggests that the product was not chiral and therefore not likely to be derived from enzyme activity (Fig. 5.3). The chemical Pictet-Spengler THIQ synthesis reaction typically produces a mixture of *ortho* and *para* products but the enzyme reaction seems to exclusively produce the *para*.[94] The presence of *ortho* in this biotransformation was indicative of a high background reaction, which was also represented in the relatively low *ee*s.

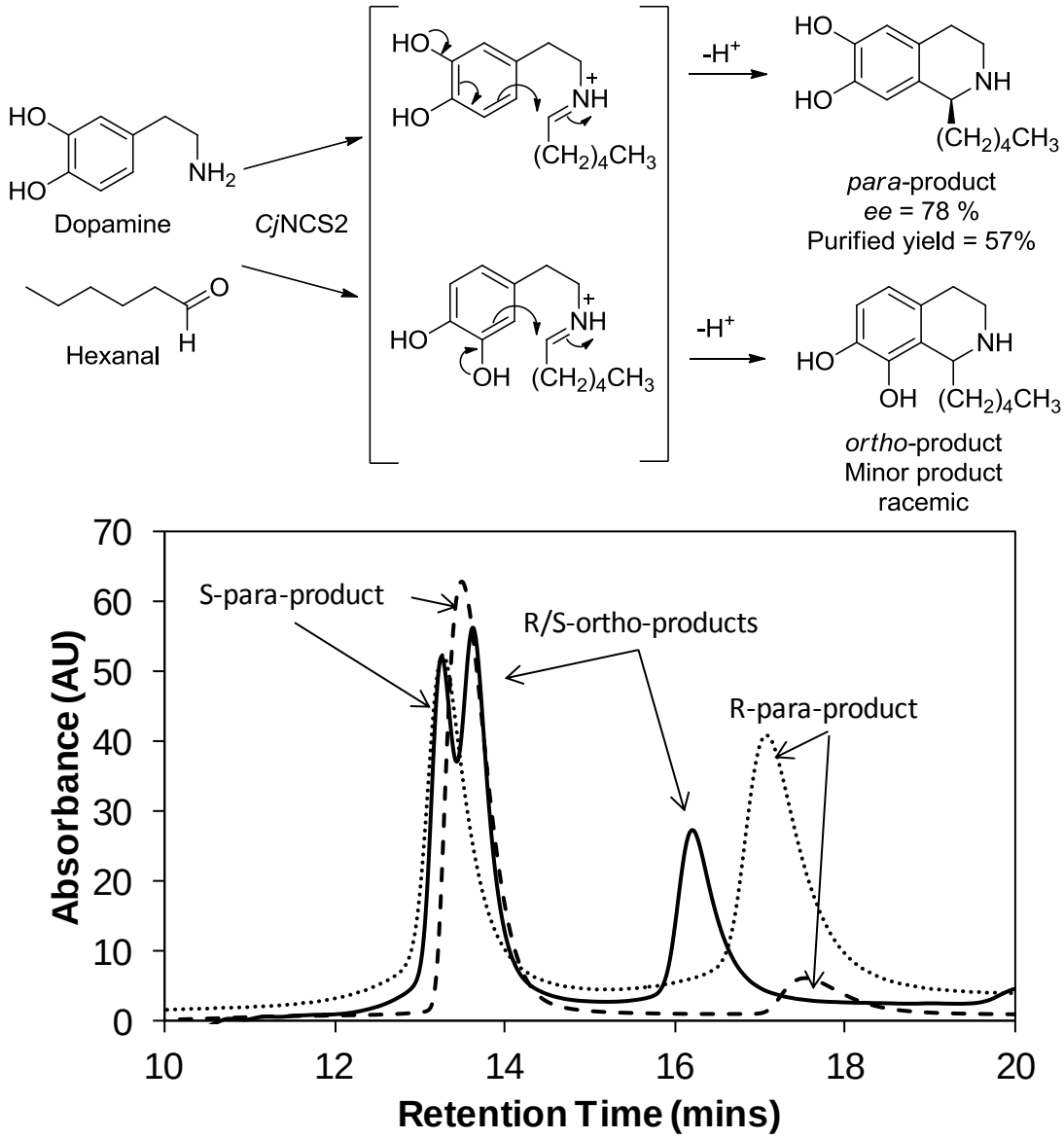


Fig. 5.3 Biotransformation with dopamine and hexanal. Top: formation of *ortho* and *para* product. Bottom: chiral HPLC analysis of products (method 3). Dotted line: chemically verified racemic *para*-product. (*S*)-isomer RT: 13.2 (*R*)-isomer RT: 17.0-17.5. Dashed line: purified *para*-product (*ee* = 78%). Solid line: *ortho*-product (containing small quantities of (*S*)-*para*-product) double peak at 13-14 is (*S or R*)-*ortho* and (*S*)-*para* and peak at 16.2 is the other *ortho* enantiomer. Therefore *ortho*-product appears mostly racemic. Magnitudes of absorbance adjusted to enable comparison. HPLC method 3 used for *ee*s (see method 2.4.1).

5.3.2 Enzyme loading

The quantity of purified enzyme required to achieve high enantiomeric excesses (*ee*s) in biotransformations was investigated (Fig. 5.4). Purified *Cj*NCS2 at 0.014 mg.mL^{-1} was shown to give inadequate *ee*s of 78% (Fig. 5.3). Different concentrations of *Cj*NCS2 (diluted in HEPES buffer) were used in biotransformations (with dopamine and hexanal) and the *ee*s of the conversions were measured. Even at the lowest

concentration tested (0.16 mg.mL^{-1}), *ee*s were high (97%) and satisfactory for biotransformations.

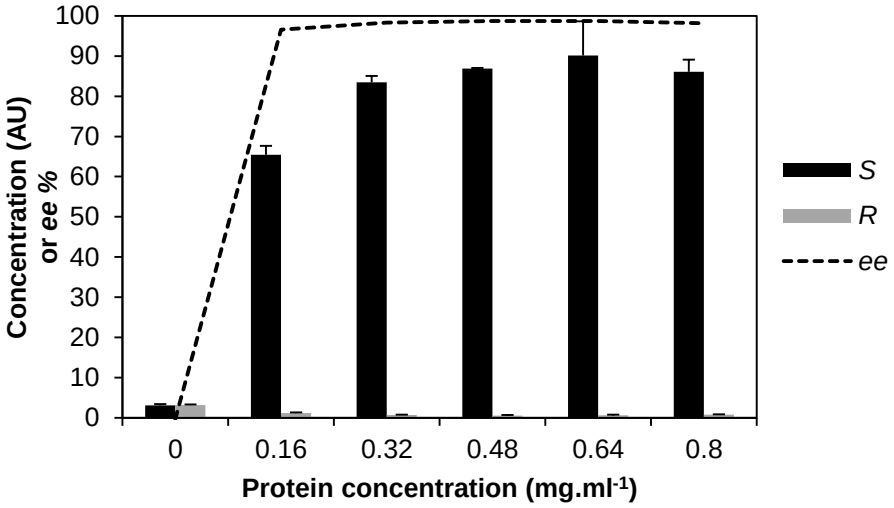


Fig. 5.4 The effect of enzyme concentration on the *ee* of a reaction between dopamine and hexanal. Bars show the quantities of the *S* and *R* enantiomers produced, in HPLC peak area. Measurements are means of triplicate experiments, error bars show one standard deviation from the mean. The dashed line shows the *ee* of the reaction, as a percentage. Reaction conditions are 10 mM dopamine and 15 mM hexanal. HPLC method 3 used for *ee*s.

Additionally, two types of biotransformation were investigated: the 'batch process' and the 'step process'. In the 'batch process', all substrates were added simultaneously (10 mM dopamine and 15 mM hexanal), whilst in the 'step process' they were added gradually so the concentrations of substrates did not reach the same maxima. In the step reactions approximately 20% of the final substrate concentration was added every 15 mins. There was no observable difference between the two reactions, both had similar conversions and excellent *ee*s of >99%. Both biotransformations showed superior *ee*s than reactions attempted previously (e.g. Fig. 5.4, maximum *ee* >98%), despite lower enzyme concentrations (0.1 mg.mL^{-1}). This is likely to be because water was used to dilute the reactions, whilst previously HEPES was used. Although HEPES buffer catalyses the chemical PS reaction to a much lower degree than phosphate buffer, it still has a minor catalytic effect, greater than that of pure water.[59] Thus HEPES very slightly reduces *ee*s compared to water. There is likely to be a compromise to be made between the buffering strength required in the reaction and the background reaction caused by the buffer.

5.3.3 Buffer choice

For maximum *ee*s the ratio of enzymatic conversion over the background must be optimised. The type and concentration of buffer used in the reactions was investigated. The concentration of buffers was varied between 10 and 50 mM. Due to the previous

success with using HEPES at pH 7.5, only similar Good's buffers (MOPS and MES) at pH 7.5 were investigated.

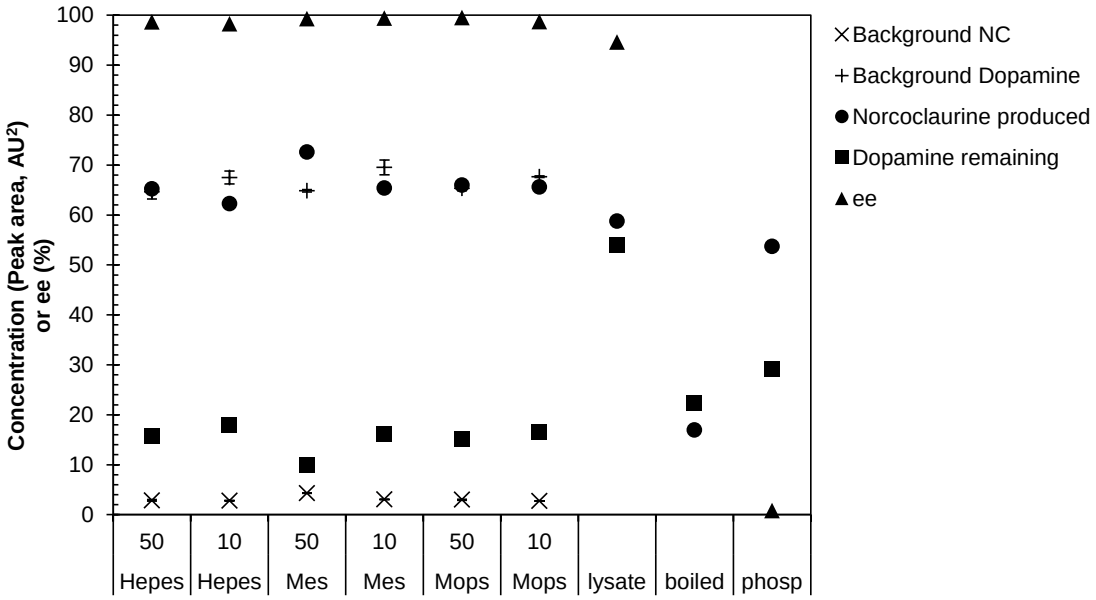


Fig. 5.5 Biotransformations in different buffers. Data points are averages of triplicate raw peak areas. Error bars are one standard deviation from mean. Numbers above buffer type refer to concentration (mM). 'Background' measurements are those recorded without enzyme present. 'Lysate' is activity in soluble fraction after lysis with BugBuster. 'Boiled' is activity in lysate after 30 minutes incubation at 100 ˚C. 'Phosp' is activity in phosphate buffer. NC = norcoclaurine. HPLC method 3 used for *ee*s. The product concentration of the boiled control was too low for *ee* detection.

The product formation and substrate consumption, both with and without purified *Cj*NCS2, was measured for 1 hour reactions (Fig. 5.5). Not surprisingly, non-desired background reactions were greater with higher concentrations of buffer. However, conversions with enzymes present were also slightly raised, and as the *ee*s did not appeared to change between different concentrations of buffers, it appeared that there was a slight increase in enzyme activity with higher buffer concentrations. The buffers HEPES and MOPS, which have similar operational pH ranges, provided similar results. MES had marginally greater conversions than other buffers, both with and without enzymes. However, it was subsequently realised that pH 7.5 lies outside the buffering range of MES (pH 5.2-7.2), so this was likely to be a pH effect rather than an effect on MES. It was decided that the reasonably low background conversions and high *ee*s in 50 mM HEPES pH 7.5 was sufficient for biotransformations.

5.3.4 Lysate reactions

Enzyme purification is costly in terms of both time and resources. It is desirable to develop reaction conditions where the enzyme does not require purification, so the reaction can be performed as either a whole cell conversion or in crude lysate. The ability to use crude lysate also means that screening for enzyme activity with diverse

substrates and mutants can be achieved much more cheaply and quickly (see section 6). The use of crude lysate is typically a problem with NCS as phosphate (present in the cell lysate), catalyses the background reaction.[59] The correct controls must be put into place to ensure that any observed activity or conversion is enzymatic. The most sensitive method to use would be chiral-HPLC, but very long run times make this method impractical.

It was possible to remove much of the background reaction in the cell lysate by desalting. Desalting removes small molecular weight molecules from the lysate. These include phosphate itself, as well as other phosphate containing molecules such as ATP and sugar-phosphates. It did not reduce the background reaction to the extent that a full purification does, but it reduced it sufficiently to enable the effective use of lysate in some screens and biotransformations (Fig. 5.6). Although purification of a His-tagged enzyme is generally fairly straightforward, and would reduce the background reaction effectively, desalting avoided the use of the expensive nickel resins. Furthermore, elution of proteins from nickel columns typically used high concentrations of imidazole, which inhibited the enzyme.

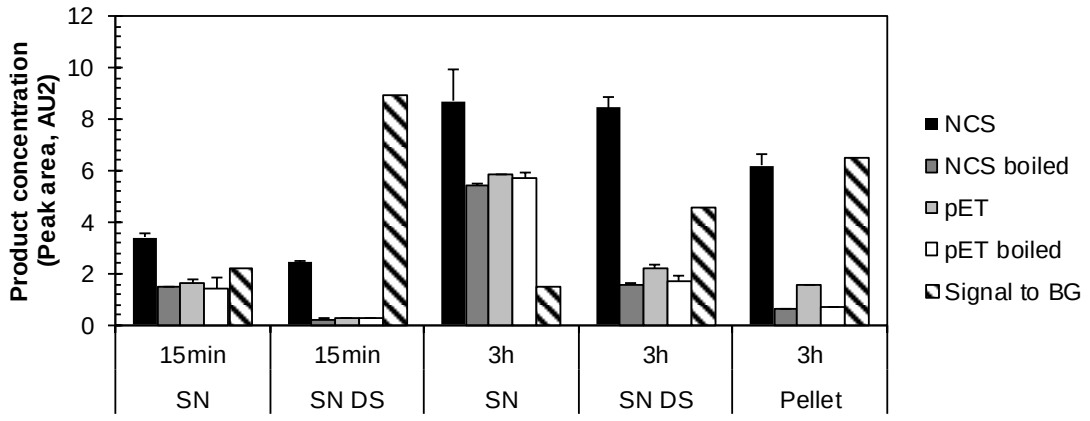


Fig. 5.6 Lysate Reactions. SN: supernatant, DS: Desalted, pET = empty vector. Signal to BG is the enzyme catalysed reaction divided by the background reaction observed. Δ19*Tf*NCS used. Conditions: 60% lysate, 0.5 mM dopa, 2.5 mM PAA.

A number of different biotransformations were attempted with *Cj*NCS2 desalted lysate using different substrates and different loading quantities (Table 5.2). A 35% loading was sufficient to provide excellent *ee*s for phenylacetaldehydes. Hexanal, a substrate less well accepted than the phenylacetaldehydes, has a slightly lower but nonetheless acceptable *ee*. Decreasing the loading to 9% reduced the *ee* of the hexanal reaction to below 90%, which was inadequate.

Table 5.2 Lysate reactions. *Cj*NCS2 desalted lysate reactions with different substrates and with different loading concentrations. HPLC method 3 used for *ee*s. Reactions: 10 mM dopamine, 15mM hexanal. PAA = phenylacetaldehyde. HPAA = 4-hydroxyphenylacetaldehyde. Dopamine and product peaks described as raw peak areas from chiral HPLC analysis (AU^2).

Substrate	Loading	Dopamine	(*S*)-isomer	(*R*)-isomer	*ee*
	%	AU^2	AU^2	AU^2	%
HPAA	35	30.7	250.0	3.0	>97
PAA	35	21.0	135.0	0.5	>99
Hexanal	35	37.3	84.7	3.5	92
Hexanal	9	62.2	41.4	3.1	86

5.3.5 Identification of TFA salt

In measuring some biotransformation conversion yields by HPLC, some yields were calculated to be over 100%. The origin of this error was in the conversion of HPLC peak areas to concentrations, which was governed by calibration curves. These curves were obtained running HPLC measurements of verified chemical standards with different known concentrations. These standards were obtained by performing a chemical or enzymatic reaction, quenching the reaction with acid, and then purifying the products by reverse phase preparative-HPLC. The HPLC solvents used are acidic and contain trifluoroacetic acid (TFA). Under acidic conditions the secondary amine product will be protonated, and trifluoroacetate was the most likely counter ion. After isolation of products by preparative-HPLC, the solvent was removed *in vacuo.* Two methods were used to remove excess TFA and isolate the pure product: either repeated addition and evaporation of methanol (as performed by Pesnot *et al*[57]) or by freeze drying in small quantities of water. For oxidation sensitive compounds the latter method was preferred as it avoided heating the products.

The dissolving and evaporation steps with methanol were proposed to remove TFA from the sample, leaving neutral products.[57] Although this process removed excess TFA *via* an azeotrope, it does not necessarily leave the product as a free base, but may leave it as a trifluoroacetate salt. Similarly, repeated freeze drying removes excess TFA, but still leaves the product as a cation with a trifluoroacetate counter ion. Basification of the protonated product, which has a pK_a of around 10-11, would typically require the addition of a base with at least a similar pK_a (e.g. bicarbonate).[160] After neutralisation, the product can be extracted into an organic solvent. The pK_a of trifluoroacetate ~ 0 and so is insufficiently basic to extract the product proton. Although the protonation state of a compound is at an equilibrium, here the energy barrier between the protonated product and TFA is probably too high to allow exchange.

The presence of TF-acetate in these compounds was confirmed by carbon-13 NMR. The TF-acetate ion gives carbon-13 peaks with low signal intensity, due to the

muliticipity (splitting by F^{19}) and the fact they have no protons attached. Repeated NMR scans were required to increase signal-to-noise and revealed TF-acetate peaks in purified biotransformation products. By taking into account the additional TFA molecular weight in the purified compounds, the concentrations of the standards were revealed to be lower than originally determined. Thus, when the calibration curves were recalculated, the product concentrations for enzyme conversions were reduced to reasonable values, below 100%.

5.3.6 Demonstrative biotransformations

To demonstrate the scope of reactions possible with NCS, two fully characterised biotransformations were performed. Both used dopamine as the amine, but differed in enzyme, enzyme preparation method and aldehyde. The first reaction used a purified Δ29*Tf*NCS mutant (L76V) turning over hexanal and dopamine, whilst the second used *Cj*NCS2 desalted lysate and phenylacetaldehyde (Fig. 5.7). Both these biotransformations had novel elements: the former was the first NCS reaction to report the mutant and the latter the first to use lysate. The mutant L76V was chosen for use in the biotransformation due to preliminary data which suggested that was superior to WT at turning over aliphatic aldehydes. Unfortunately, this initial observation was not subsequently verified, and the mutant was determined to perform the reaction equally as well as WT, but no better.

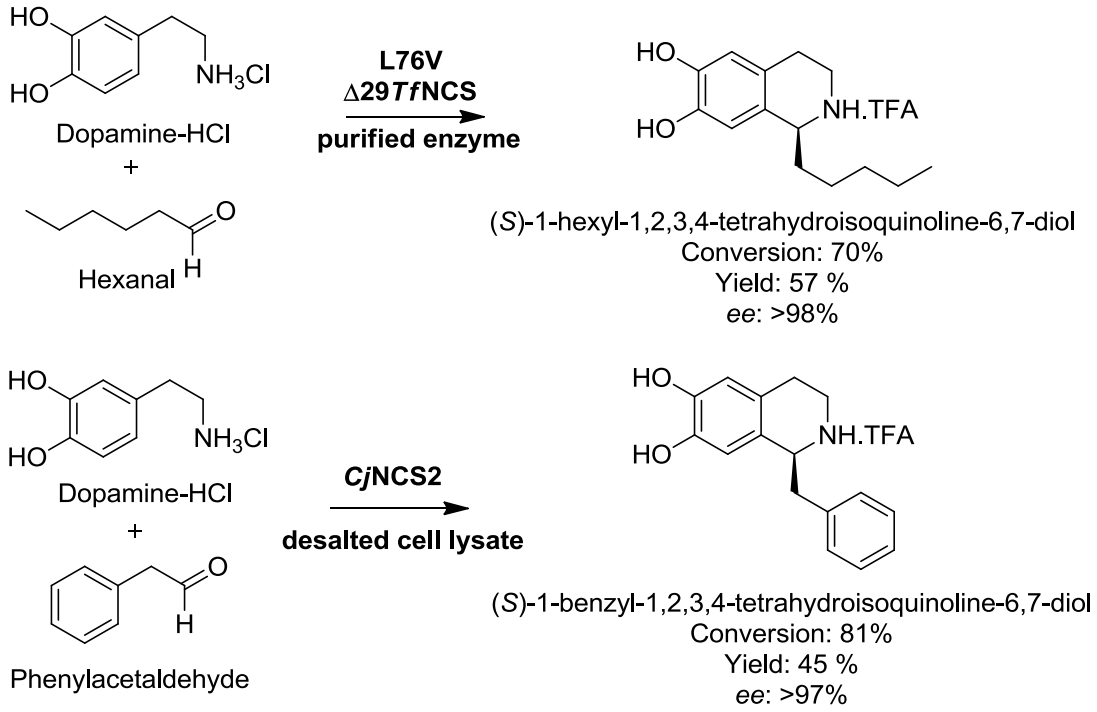


Fig. 5.7 Biotransformations. Two demonstrative biotransformations using different enzymes, preparations and substrates. Conversion describes the quantity of product at the end of the reaction in the

crude mixtures. Yield refers to the quantity of purified product isolated. Both products were isolated as TFA salts. HPLC method 3 used for *ee*s.

The reactions were performed at milligram scale, using 40 mg of dopamine starting material (with 1.5 molar equivalents of aldehyde) and obtaining 40 mg and 33 mg of product for the hexanal and phenylacetaldehyde reaction respectively. Both the reactions provided excellent *ee*s and good conversions, but relatively modest purified yields.

Steps can be taken to improve the purification process. Although the *ee*s cannot be improved, both the conversions and yields could ideally be higher. It was apparent that much of the product is lost during the preparative-HPLC purification. Furthermore, with increased reaction scales (>100 mg) this purification method is not practical. In a reported biocatalytic synthesis of (*S*)-norcoclaurine with NCS, chromatographic purification was avoided by using activated carbon from which the substrates and products were eluted.[58]

Recent work by Maresh *et al* has demonstrated that it is possible to drive the reactions to completion and simultaneously avoid chromatographic separation by the use of excess dopamine.[60] The tetrahydroisoquinolines can then be extracted into an organic ethyl acetate layer, separating them from dopamine. Removal of solvent should result in the pure product. The purity of the product will depend on whether the reactions go to completion and whether there are numerous side products. Addition of the antioxidant sodium ascorbate into the reaction mixture may help prevent the oxidation and degradation of substrates and products.[58,60] This reaction and organic extraction method was adapted and used for biotransformations described in section 6.6.9 (methods section 2.6.3).

Although the methods developed here resulted in the production of single enantiomer products with reasonable yields, it may be worth investigating other strategies and purification methods for biotransformations in the future.

5.4 Multi-enzyme cascades

The majority of the following work can be found in Lichman *et al*.[161]

5.4.1 Introduction

Enzymatic cascade reactions enable the combination of numerous biocatalytic steps in a single pot, bypassing isolation and purification steps, and providing complex chemical transformations in an efficient manner.[162] The formation of these cascades is greatly aided by the fact that enzymes generally operate in very similar conditions (pH 6-8, 20-

50 ˚C, aqueous).[163] Furthermore, the increasing emphasis on environmentally benign, non-toxic synthesis methods has led to a greater interest in the potential for enzyme cascades in organic synthesis.

The variety of enzymes now available enables various combinations of enzymatic transformations and substrates. It is now possible to consider a 'biocatalytic retrosynthesis' approach and construct syntheses in a purely biocatalytic manner.[158] Similarly, the concept of 'systems biocatalysis' is to use multiple enzymes *in vitro* creating an 'artificial metabolism' that can be carefully and deterministically controlled.[159] Systems biocatalysis has a similar aim to synthetic biology/metabolic engineering (the 'biological' production of valuable chemicals) but has a significant difference in methodology: the former operates *in vitro* and the latter *in vivo*. Using enzymes *in vitro* makes reaction control relatively simple—for example, to slow down a reaction step rate, just add less enzyme. In contrast, *in vivo* approaches would require genetic manipulation of ribosome binding sites or promoters to achieve the same goal. The *in vitro* systems biocatalysis approach enables the establishment of specific cascade designs that enable recycling of co-factors and shifting of equilibria.[162–164]

There has been considerable focus on the use of NCS in *in vivo* synthetic biology systems, but few attempts at *in vitro* cascades.[21,45,72,73,76] The exception is the formation of aldehydes *in situ* from amino acids using hypochlorite prior to addition of NCS; this is an example a sequential chemoenzymatic cascade (see section 1.5.1).[58,60] When using NCS, the high non-enzymatic background PS reaction must be controlled to obtain high *ee*s. This is more manageable *in vitro* where enzyme concentrations and buffer components can be modified. However, despite the challenges of the background reaction, recent work has demonstrated NCS activity is possible *in vivo*.[21,45,76]

There are numerous biocatalytic steps possible both before and after an NCS catalysed step that could be combined in an *in vitro* cascade. The available options are, of course, exemplified by natural BIA biosynthesis. Dopamine could potentially be formed enymatically from tyrosine *via* a 3-hydroxylation and decarboxylation. Enzymatic modifications of the BIA moiety may be achieved for example through methylation, methylenedioxy bridge formation and aromatic oxidative coupling (Fig. 5.8).

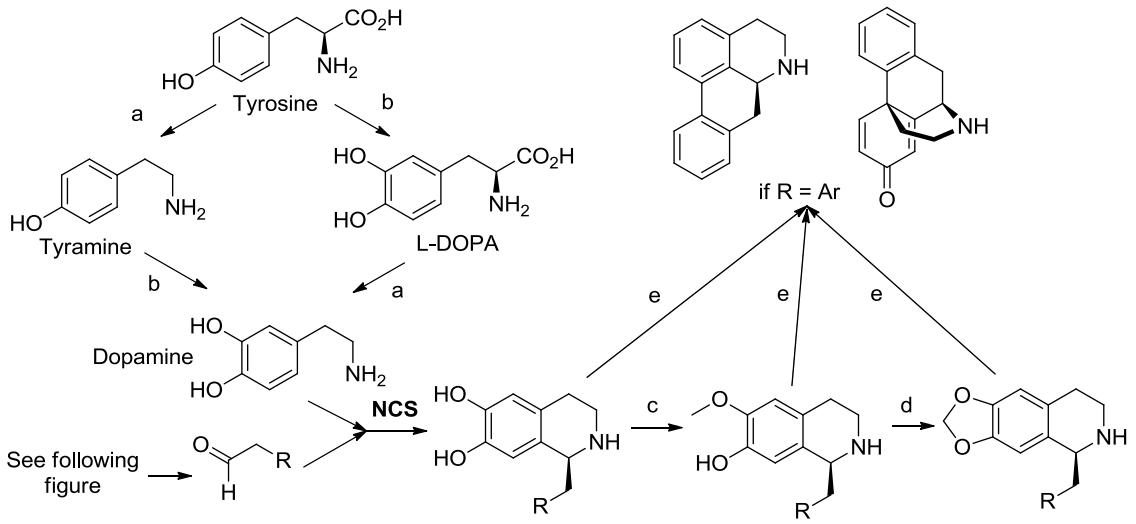


Fig. 5.8 Possible biocatalytic steps preceding or following the NCS reaction. Dopamine can be formed from tyrosine *via* (a) decarboxylation catalysed by tyrosine/DOPA decarboxylase, and (b) hydroxylation catalysed by tyrosine hydroxylase or tyrosinase. Modification of hydroxyl groups is possible through (c) methylation catalysed by methyltransferases and (d) methylenedioxy bridge formation, catalysed by CYPs. Modification of the carbon skeleton may be possible through aromatic oxidative coupling (e), also catalysed by CYPs. Routes to aldehydes are discussed below in **Fig. 5.9**.

The large aldehyde substrate scope of NCS led to the decision to focus on cascade designs enabling the *in situ* formation of aldehydes.[53,57,61] Aldehydes are very reactive and relatively difficult to synthesise, so the *in situ* formation of these intermediates would be very valuable for the NCS reaction and may ultimately allow for the one-pot synthesis of diverse THIQs. An aldehyde synthesis method could potentially be useful for other biocatalytic reactions requiring aldehydes. There are a number of enzymes that can form aldehydes. These enzymes include: alcohol dehydrogenases,[165] carboxylic acid reductases,[166] laccases[167] and transaminases (Fig. 5.9).[114]

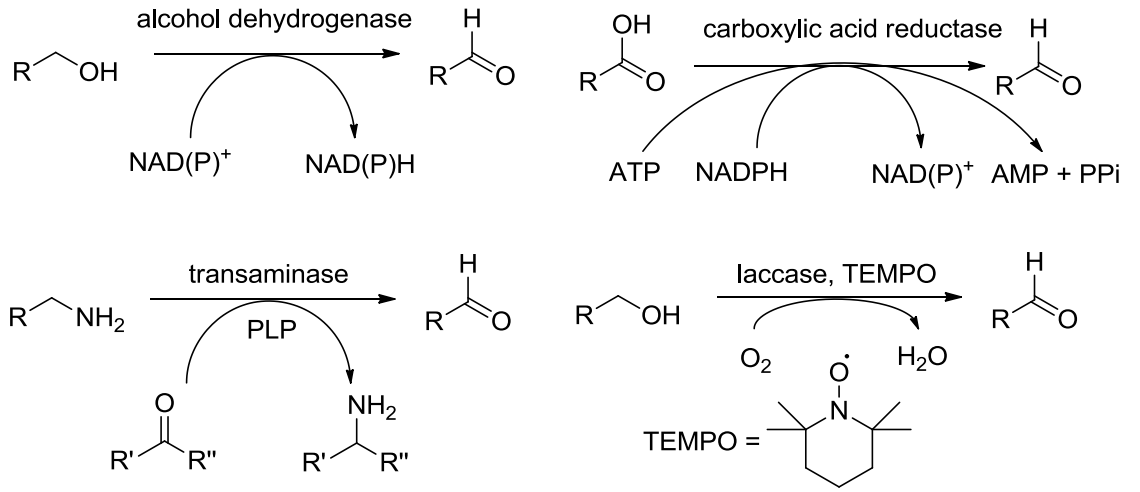


Fig. 5.9 Possible biocatalytic routes to aldehydes.

5.4.2 Aim

The aim was to construct a biocatalytic cascade using a transaminase (TAm) and NCS. The TAm can catalyse the formation of an aldehyde from an amine. The aldehyde would then be a substrate for NCS in the formation of a THIQ. NCS also requires an amine substrate, and although it may be possible to use two different amines as starting substrates (and select enzymes and substrates with no cross-reactivity), it was decided to use the same amine as substrates for both enzymes.

The amine compound chosen was dopamine. Dopamine was expected to be converted into the aldehyde 3,4-dihydroxyphenylacetaldehyde (3,4-DHPAA) by a TAm. This aldehyde can be combined with dopamine in a PS reaction catalysed by NCS, forming the chiral THIQ (*S*)-norlaudanosoline ((*S*)-NL, also known as tetrahydropapaveroline, THP) (Fig. 5.10).[25] It was also hypothesised that the addition of formaldehyde to norlaudanosoline could trigger a second PS reaction, forming a tetracyclic tetrahydroprotoberberine (THPB, also known as a berbine) compound.[168]

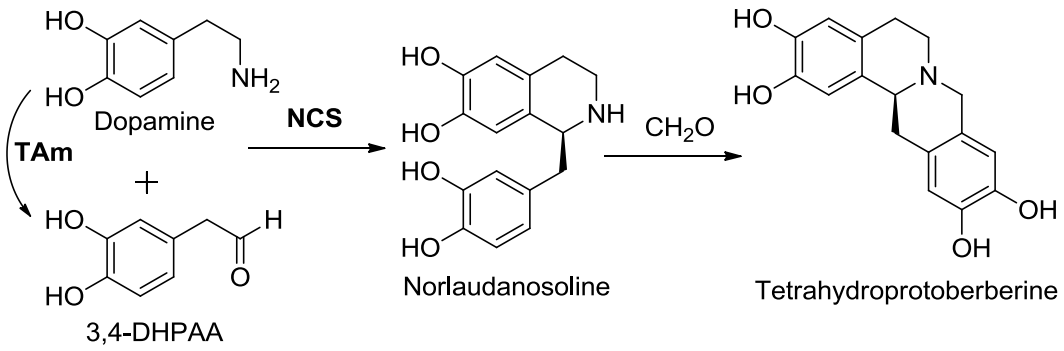


Fig. 5.10 Proposed biocatalytic scheme.

5.4.3 Target compounds

The key target compound of this cascade is norlaudanosoline (NL). It is a BIA found in mammals, where neurologically it is acts as a depressor and has been shown to have roles in Parkinson's disease[169] and drug addiction.[170] Perhaps unsurprisingly, considering its formation from dopamine, it has been shown to regulate dopamine levels in the brain.[171] Intriguingly, it has been shown to be the precursor to 'endogenous' morphine, which is produced in mammalian brains.[172–174] The proposed mammalian morphine biosynthesis route, *via* NL, provides a fascinating counterpart to the plant biosynthesis route, which is shown to proceed *via* (*S*)-norcoclaurine (though was originally thought to proceed *via* NL).[25,30–32] The presence of a morphine biosynthesis route in mammals is a remarkable example of the convergent evolution of the formation of very complex molecule. There is also evidence that only the (*S*)-isomer

of NL is present in mammalian brains; this points towards the presence of a mammalian BIA Pictet-Spenglerase.[173,175]

Synthetically, NL is a versatile compound and can serve as a precursor to numerous BIAs with biological activities. NL has also been employed in synthetic biology/metabolic engineering routes to BIAs and in a number of cases it serves as a vital feedstock.[64–66] Typically these systems are fed with *rac*-NL, which results in an immediate 50% loss of material due to the (*S*)-selectivity of downstream enzymes.[68] For synthetic biology, medicinal chemistry and natural product investigations, a stereoselective synthesis of (*S*)-NL would be valuable.

The tetrahydroprotoberberine alkaloids (THPBs, also known as berbines) are a subgroup of BIAs closely related to berberine, a compound with anticancer and antimicrobial activities.[11,176,177] They are found in both plants and animals[174] and include compounds such as canadine[178] and spinosine.[179] Although in plants they are typically formed enzymatically *via* the berberine bridge enzyme,[18] *in vitro* they can be accessed chemically from norlaudanosoline *via* a PS reaction with formaldehyde (Fig. 5.11).[168,179] The products of this reaction have potential in cancer chemotherapeutics.[129] The expected regioselectivity of this formaldehyde cyclisation is the least sterically hindered product (major product: 10,11-dihydroxy). This contrasts with the regioisomerism of the 9,10-dihydroxy-THPBs formed by BBE.[101,103]

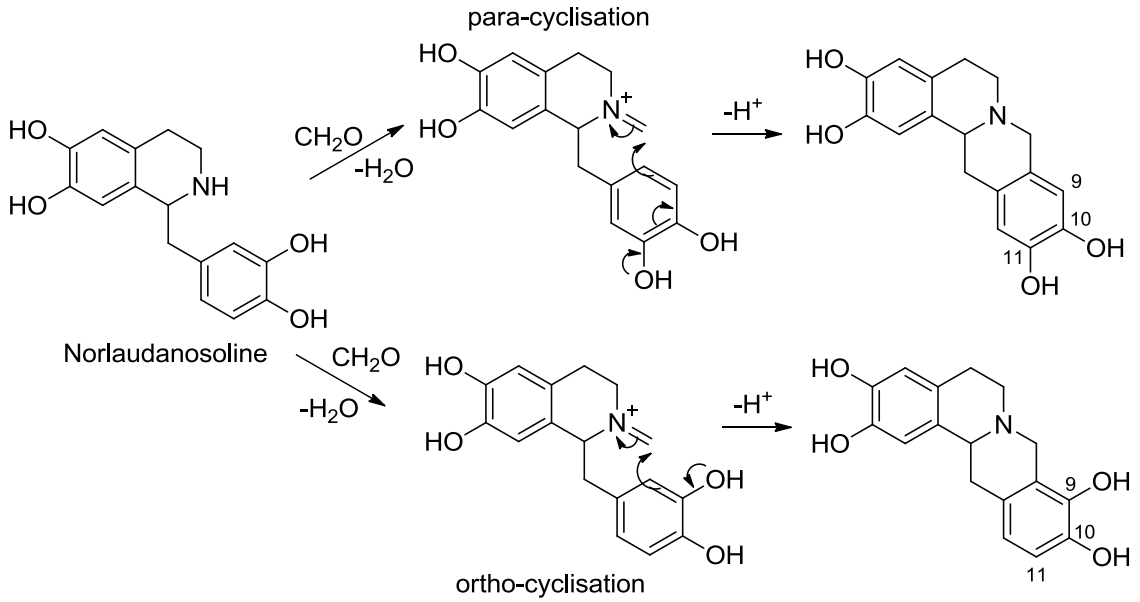


Fig. 5.11 Tetrahydroprotoberberine production from norlaudanosoline with formaldehyde.

5.4.4 Transaminases

Transaminases were chosen for this cascade as they can convert amines to aldehydes, and could potentially accept dopamine as a substrate. The Ward and Hailes groups have considerable expertise in using TAms, and the Ward group has a large library of TAms available to screen. TAms have been extensively studied in biocatalysis, both in one-pot reactions and multi-enzyme cascades.[163] In a case often considered a landmark in biocatalysis, a TAm was engineered to accept a very bulky substrate and contribute to the industrial synthesis of the commercial pharmaceutical sitagliptin.[180]

Transaminases catalyse the conversion of a carbonyl moiety to a primary amine, and *vice-versa*. Typically the carbonyl in question is a pro-chiral ketone and the desired product is a chiral amine. The reversed reaction, from an amine to a carbonyl, is possible but less typically desired as it may involve the destruction of chirality. Superficially, the reaction appears to be a reductive amination, but instead of a reducing agent the reaction utilises a co-substrate which exchanges its amine/carbonyl function group with the substrate. The co-factor pyridoxal phosphate (PLP, vitamin B_6) is vital for the reaction mechanism, essentially acting as an electron sink (Fig. 5.12).

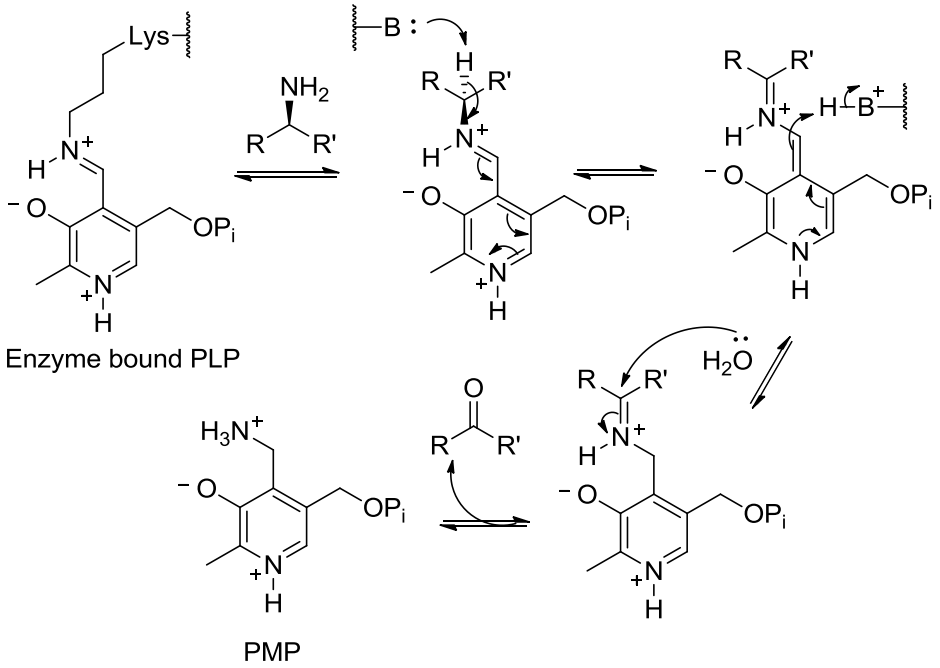


Fig. 5.12 Transaminase mechanism. Half the reaction is depicted—in order for the PLP to regenerate the reverse reaction must occur with the co-substrate. PLP: pyridoxal phosphate; PMP: pyridoxamine phosphate; P_i: inorganic phosphate.

Due to the reversible nature of the reaction, TAms often suffer from problems relating to the equilibrium of the reaction—the isolated reaction will not typically go to completion. These unfavourable equilibria can be helped by numerous cascade designs.[163] Methods for shifting equilibria include: excess co-substrates,[180] removal of the co-products,[180] enzymatic[181] or chemical[182] consumption of the (co)product, and cascades to recycle the co-substrate.[183,184] Cascades with TAms must be designed with shifting the equilibrium in mind.

5.4.5 Cascade design

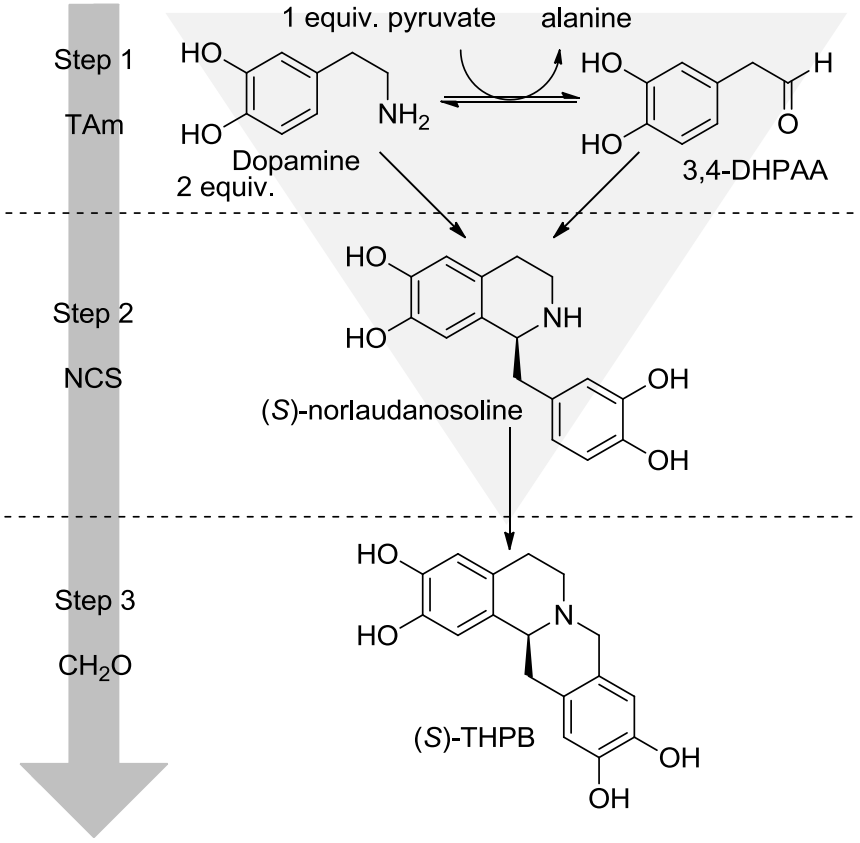


Fig. 5.13 'Triangular' cascade design. Figure adapted from Lichman *et al*.[161]

The overall reaction design described here is a one-substrate, two-enzyme, one-pot cascade. This three sided design can be described as a 'triangular' cascade (Fig. 5.13).

Similar cascade designs had previously been achieved using monoamine oxidase (MAO) to convert dopamine to 3,4-DHPAA in the first step. MAOs convert amines to imines using molecular oxygen. The imines are then hydrolysed by water to form aldehydes. Originally the dopamine-MAO cascade forming racemic NL (THP) was observed using mammalian tissue extracts containing MAO.[70] Subsequently *Aspergillus niger* was shown to have sufficient MAO activity to enact the same

transformation.[29] More recently, a stepwise fermentation method was developed to produce racemic NL *de novo* using two *E. coli* strains, one containing dopamine biosynthesis genes and the other overexpressing a MAO.[71] In these three preceding examples, the PS reaction between dopamine and 3,4-DHPAA occurs spontaneously, and thus a racemic product is formed.

There is precedence for NCS accepting 3,4-DHPAA as a substrate in place of the natural substrate 4-HPAA.[25] Some *in vivo* synthetic biology approaches to BIAs have attempted to add NCS to the dopamine-MAO cascade, in order to form (*S*)-NL at the key branch BIA point.[72–75] However, it is not clear that NCS is fully active in these systems; the reported formation of (*S*)-BIAs such as (*S*)-reticuline may be a result of (*S*)-selective enzymes downstream of NCS. More recently, *in vivo* approaches to BIAs have eschewed this MAO route and focussed on the natural biosynthetic route *via* 4-HPAA.[21,45,76]

There are significant advantages to using a TAm rather than an MAO in this triangular cascade. The requirement of a co-substrate for TAms allows for stoichiometric control of the reaction. As the PS reaction requires equimolar quantities dopamine and 3,4-DHPAA, the TAm co-substrate can be limited to ensure exactly half of the original dopamine is converted to 3,4-DHPAA. This control is not possible with MAO as oxygen is the 'co-substrate' and controlling oxygen concentrations is more challenging. This exacting stoichiometric control is a feature of the *in vitro* systems biocatalysis methodology—such control would not be possible in an *in vivo* system. The triangular cascade design also solves problems with the TAm equilibrium. The PS reaction consumes both the starting material and product from the TAm reaction. Assuming the PS catalyst is stable and sufficiently active, the actual TAm equilibrium position is unimportant and the reaction progression should be determined by the limiting quantity of co-substrate.

5.4.6 Transaminase screen

The initial stage towards the formation of the triangular cascade was the screening of various TAms to identify an enzyme that could efficiently convert dopamine into 3,4-DHPAA. Ten TAms were chosen for screening from the in-house TAm library in the Ward group (see methods sections 2.1.5 and 2.3.4). These were selected partly as activities had previously been detected with these enzymes. Analysis of lysed cells post-induction showed all constructs demonstrated overexpression of TAm (Fig. 5.14). Only two homologues appeared to be insoluble, and the remainder demonstrated soluble expression of recombinant protein with the expected size.

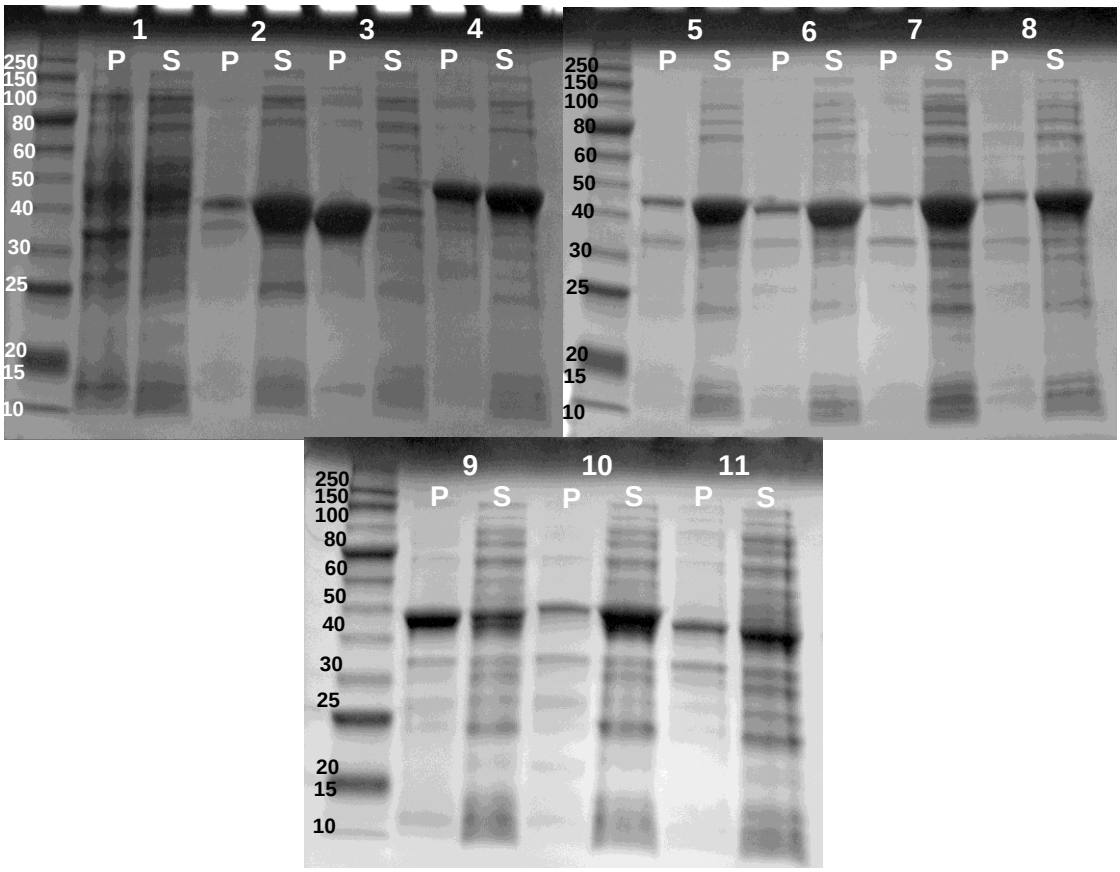
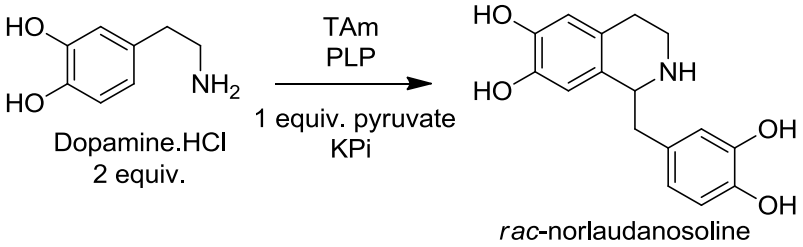


Fig. 5.14 Expression of TAms. P = pellet, S = supernatant. Enzymes: 1 = empty vector, 2 = BSU_09260 (1), 3 = BSU_09260 (2), 4 = CV_2025, 5 = Dgeo_1416, 6 = KPN_00255, 7 = PP_0596, 8 = PP_3718, 9 = SaV_2612, 10 = SaV_4551, 11 = VF_JS17. Of the two BSU_09260 constructs, only the one producing soluble protein (2) was taken on for further investigation. Figure taken from Lichman *et al*.[161]

The TAm screening reactions were conducted in phosphate buffer, which catalysed the *in situ* formation of racemic norlaudanosoline from dopamine and 3,4-DHPAA. Therefore, the reaction progress was tracked on HPLC by formation of norlaudanosoline. Following the relatively stable norlaudanosoline was simpler than tracking the formation of the reactive aldehyde, which, unless shepherded into the PS reaction with dopamine, would probably be degraded by native *E. coli* enzymes present. The TAm co-substrate pyruvate was present in the reaction at half the molar concentration of dopamine, thus overall allowing the TAm to convert only half of the dopamine present.

The enzymes CV2025 (*Chromobacterium violaceum*)[114] and PP_3718 (*Pseudomonas putida*) demonstrated significant activity in these screening conditions, with 21% and 15% conversion respectively (Table 5.3, Fig. 5.15). The enzymes PP_0596 (*P. putida*), SaV_2612 (*Streptomyces avermitilis*) and VF_JS17 (*Vibrio fluvialis*)[115] displayed trace activities. The well-studied enzyme CV2025 exhibited the greatest conversions and was used in the completed cascades.

Table 5.3 TAm mediated synthesis of *rac*-norlaudanosoline from dopamine. Table adapted from Lichman *et al*.[161]



Entry	TAm	Organism	Uniprot entry name	Gene name	Conv. %[b]
1	Empty vector				n.d. [c]
2	BSU_09260	*Bacillus subtilis*	YHXA_BACSU	yhxA	n.d.
3	CV_2025[114]	*Chromobacterium violaceum*	Q7NWG4_CHRVO	CV_2025	21
4	Dgeo_1416	*Deinococcus geothermalis*	Q1IYH3_DEIGD	argD/lysJ	n.d.
5	KPN_00255	*Klebsiella pneumoniae*	A6T537_KLEP7	gabT	n.d.
6	PP_0596	*Pseudomonas putida*	Q88Q98_PSEPK	PP_0596	Trace
7	PP_3718	*Pseudomonas putida*	Q88GK3_PSEPK	PP_3718	14
8	SaV_2612	*Streptomyces avermitilis*	Q82JZ2_STRAW	SAV_2612	Trace
9	SaV_4551	*Streptomyces avermitilis*	Q82ER2_STRAW	SaV_4551	n.d.
10	VF_JS17[115]	*Vibrio fluvialis*	F2XBU9_VIBFL	JS17	Trace

50 mM dopamine, 25 mM pyruvate, 1 mM PLP, 10% v.v^{-1} TAm, 37 °C, 4 h. [b] Concentration of *rac*-norlaudanosoline, determined by analytical HPLC. [c] n.d. = not detected. See Fig. 5.15 for reaction time course.

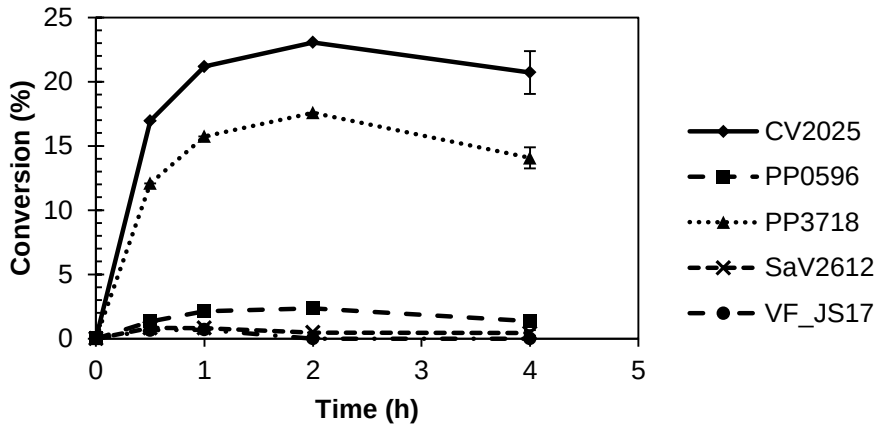


Fig. 5.15 Progress curves charting the accumulation of *rac*-norlaudanosoline over time with different TAms. The final measurement at 4 hours was performed in triplicate (error bars are one standard deviation). Loss of product between 2 and 4 hours is likely to be due to the instability and oxidation of the product. Only TAms showing activity towards dopamine are represented here. Conditions: 50 mM dopamine, 25 mM pyruvate, 1 mM PLP, 10% v.v^{-1} TAm, 37 °C, 4 h. Figure adapted from Lichman *et al*.[161]

5.4.7 Optimisation

To enable the formation of (*S*)-norlaudanosoline from dopamine, the racemic PS reaction was minimised by changing the buffer from phosphate to HEPES and purified Δ29*Tf*NCS was added to catalyse the (*S*)-selective PS reaction. The optimum reaction conditions for the cascade were investigated with the aim of maximising the conversion and *ee*. The factors that were varied were the starting concentrations of dopamine, NCS and TAm. The molar concentration of the TAm co-substrate pyruvate always remained half of dopamine. It was hoped that this optimisation would enable the rates of the two enzymatic steps to be approximately matched. If the TAm rate were considerably faster than the NCS step then a building up of 3,4-DHPAA would occur and result in an increase of non-productive side-reactions (including the non-enzymatic PS), ultimately reducing the yield and *ee*.

Time courses of product formation were recorded under different reaction conditions (Fig. 5.16). The factors varied were dopamine concentration (20 and 50 mM), lysate concentration (10 or 30% $v.v^{-1}$) and NCS concentration (100 or 500 $\mu g.mL^{-1}$). Unsurprisingly, higher concentrations of enzyme led to greatest and fastest conversions. With dopamine at 50 mM it was clear that the limiting step was the transamination;an increase in TAm concentration greatly increased initial rates. Overall it seems that the initial rate was determined by TAm concentration, but the point at which the conversion plateaus was partially determined by NCS concentration.

Conversions and stereoselectivities (after three hours) were measured in a similar experiment (Table 5.4). The component concentrations were varied: dopamine concentration (20 or 50 mM), lysate concentration (10, 20 or 30% $v.v^{-1}$) and NCS concentration (100 or 500 $\mu g.mL^{-1}$).

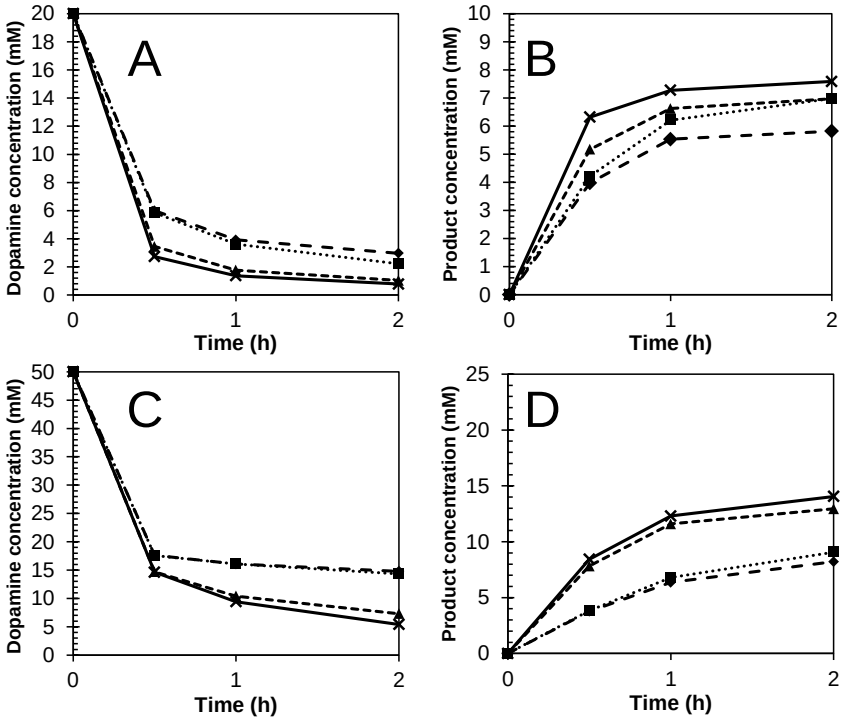
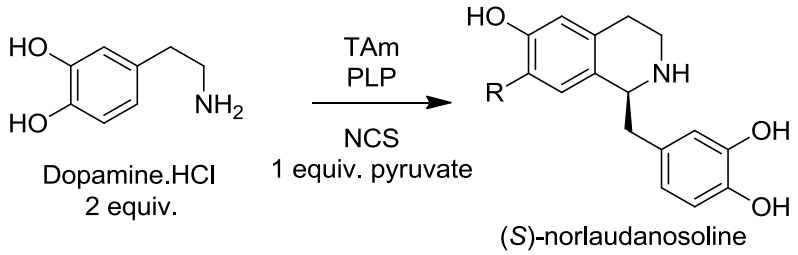


Fig. 5.16 Reaction time courses. A: Dopamine depletion from 20 mM. B: Product formation from 20 mM dopamine. C: Dopamine depletion from 50 mM. D: Product formation from 50 mM. Diamonds, large dashed line: 10% $v.v^{-1}$ TAm lysate, 0.1 $mg.mL^{-1}$ NCS. Squares, dotted line: 10% $v.v^{-1}$ TAm lysate, 0.5 $mg.mL^{-1}$ NCS. Triangles, short dashed line: 30% $v.v^{-1}$ lysate, 0.1 $mg.mL^{-1}$ NCS. Crosses, solid line: 30% $v.v^{-1}$ lysate 0.5 $mg.mL^{-1}$. Data points are means of triplicates, error bars of one standard deviation are smaller than the markers so were not depicted.

Table 5.4 One pot synthesis of (*S*)-BIAs. Table adapted from Lichman *et al*.[161]



Dopamine	CV2025	NCS	Conv.	*ee*
mM	% $v.v^{-1}$	µg.mL^{-1}	%[b]	%[c]
20	10	100	74 (74)	99
20	20	100	86 (89)	99
20	30	100	82 (93)	99
20	10	500	77 (82)	99
20	20	500	87 (92)	99
20	30	500	84 (95)	98
50	10	100	36 (42)	98
50	20	100	55 (66)	97
50	30	100	70 (78)	99
50	10	500	42 (48)	98
50	20	500	65 (73)	98
50	30	500	72 (85)	96

General reaction conditions: 2 equivalents dopamine, 1 equivalent pyruvate, purified Δ29*Tf*NCS and CV2025 lysate, 50 mM HEPES pH 7.5, 37 °C, 3 h. [b] Determined by analytical HPLC. Conversions in brackets refer to depletion of dopamine. [c] Determined by chiral HPLC of the crude product, method 2 for dopamine.

Stereoselectivities were determined by chiral HPLC of the crude reaction. Product peaks were identified by comparison with verified chemical standards. Chromatograms of reactions contained some small peaks that could not be identified. The major (*S*)-norlaudanosoline product was easily identified due to its size and retention times. However, based on the chemical standard alone, it was not immediately clear which peak in the crude reactions represented the minor (*R*)-norlaudanosoline**:** there were two small peaks around the correct retention time that were possible candidates. Enantiomeric excesses were calculated for both peak options, and were all found to be >95%, regardless of the (*R*)-norlaudanosoline peak selected. By plotting calculated *ee*s against NCS concentration it could be seen that with the (*R*)-norlaudanosoline as peak 1, *ee*s fall with increasing NCS concentration. On the other hand, if *ee*s are calculated with peak 2 they rise with increasing NCS concentration (Fig. 5.17). As NCS is only known to make the (*S*)-isomer, increasing NCS concentration will increase *ee*s. Therefore, peak 2 and not peak 1 must be the (*R*)-isomer. By establishing this, it was then possible to calculate the *ee* with greater accuracy. This peak identification problem would have also been solved by doping a run of the chiral HPLC analysis with racemic NL.

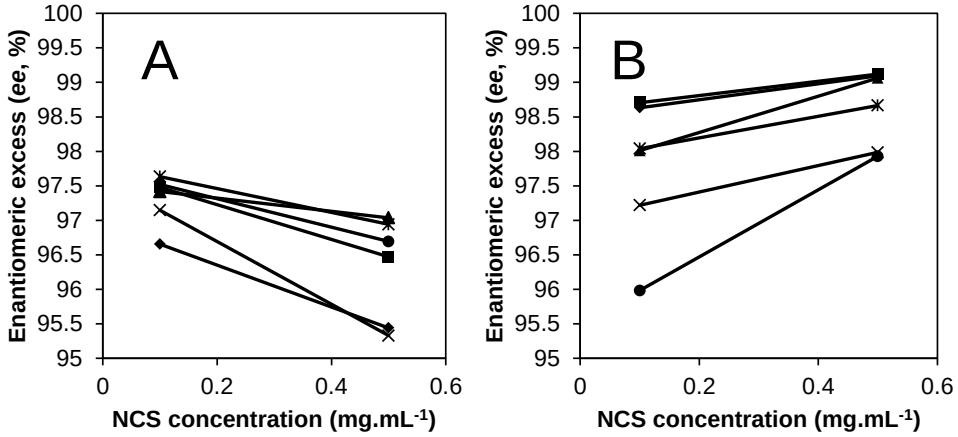


Fig. 5.17 Enantiomeric excesses when calculated with peak 1 as (*R*)-isomer (A) or peak 2 as (*R*)-isomer (B). Each different shape marker represents different reactions condition: 20 or 50 mM dopamine, 10, 20 or 30% v.v^{-1} TAm lysate. Solid lines connect identical reaction conditions (except NCS concentration, which is represented on the x-axis). Notice the gradient of all lines in A is negative and in B is positive. HPLC method 2.

The optimum conditions resulted in 87% conversion of dopamine to (*S*)-norlaudanosoline with an *ee* of 99% (Table 5.4). The reaction contained 20 mM dopamine, 500 µg.mL^{-1} NCS and 20% v.v^{-1} CV2025 lysate. The highest concentration of NCS, together with the lowest concentration of substrates, resulted in the maximum

conversions and yields. Similarly, increased TAm concentrations always resulted in increased dopamine consumption, though this did not always translate into increased (*S*)-norlaudanosoline concentrations. At 50 mM dopamine the maximum TAm concentrations (30% $v.v^{-1}$) resulted in greater conversions, but at the lower dopamine concentration (20 mM), the lower TAm concentration of 20% $v.v^{-1}$ was optimal. This difference was probably due to aldehyde side reactions, as the aldehyde was being produced more rapidly than NCS can turn it over. There was also marginally poorer stereoselectivity with 30% $v.v^{-1}$ lysate—the chemical composition of lysate and the increased concentration of reactive aldehyde resulted in increased racemic PS chemical conversion. In terms of conversions, NCS concentration had more of an impact when higher concentrations of dopamine were present.

5.4.8 Modifying the amine

The scope of the triangular cascade is limited because the arylamine must be a substrate for two enzymes, and the TAm product must also be a substrate for NCS. Due to the large library of TAms available, and the broad substrate scope of these enzymes, the limiting enzyme is unlikely to be the TAm. The NCS, on the other hand, has very specific substrate requirements, especially regarding the amine substrate. Though the 4-hydroxy group of dopamine is superfluous, the 3-hydroxy is vital for the reaction progression.[51,53,57] Bulkier chiral amines such as metaraminol are substrates for NCS, but the TAm products would be ketones, and as such unlikely to be suitable carbonyl substrates for the enzyme (see sections 6.3.2, 6.5 and 6.6).

As described, the simplest modification of the reaction cascade that could be applied was to use 2-(3-hydroxyphenyl)ethylamine in place of dopamine. This compound is essentially dopamine with the 4-hydroxy group removed. Using the optimum conditions determined for dopamine, the compound 2-(3-hydroxyphenyl)ethylamine was successfully turned over by the cascade, producing a novel (*S*)-BIA (Fig. 5.18). The conversion of 56% and *ee* of 90% were considerably lower than for dopamine, and this is perhaps reflective of either the poorer affinity of NCS to 2-(3-hydroxyphenyl)ethylamine or the corresponding 3-hydroxyphenylacetaldehyde compared to dopamine and 3,4-DHPAA.

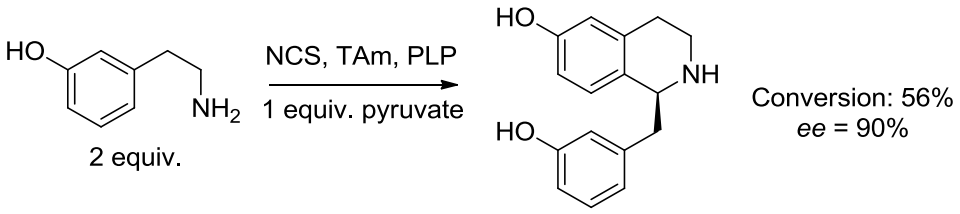


Fig. 5.18 Cascade performed with 2-(3-hydroxyphenyl)ethylamine. Conditions: 20 mM amine, 10 mM pyruvate, 20% v.v^{-1} CV2025, 500 µg.mL^{-1} Δ29*Tf*NCS, 50 mM HEPES pH 7.5, 37 °C, 3 h. Conversion (formation of product) determined by analytical HPLC, *ee* measured by chiral HPLC method 3.

5.4.9 Tetrahydroprotoberberines

Norlaudanosoline can be converted into a tetrahydroprotoberberine (THPB) through a second PS reaction with formaldehyde. Formaldehyde can be added to the cascade reaction after the reaction has gone to completion. The formaldehyde destroys the enzymes present and reacts rapidly with the norlaudanosoline. In an analogous manner to the preceding PS reaction, the 3-hydroxy group on the 3,4-DHPAA is crucial for this cyclisation.

Under optimised conditions, the cascade for the formation of (*S*)-norlaudanosoline from dopamine reached completion after approximately 3 hours. Formaldehyde (in phosphate buffer) was added to the reaction at this point and allowed to react for 30 minutes. Analysis of this reaction revealed the formation of two THPB products, the *para*-cyclisation product (10,11-dihydroxy) and the *ortho*-cyclisation product (9,10-dihydroxy) in a ratio of approximately 7:1 (Fig. 5.19). Based on previous studies and steric effects, this is the expected distribution of product.[168] The second (formaldehyde) PS step occurred with 74% conversion (64% and 9% for major *para* and minor *ortho* product respectively); overall from dopamine this is 64% conversion (56% and 8%). The chirality of the product is established enzymatically by NCS, and the addition of formaldehyde does not affect the *ee*. Only the (*S*)-enantiomer of the major product was observed by chiral HPLC, though the HPLC method used was of insufficient sensitivity to determine the *ee* of the minor product.

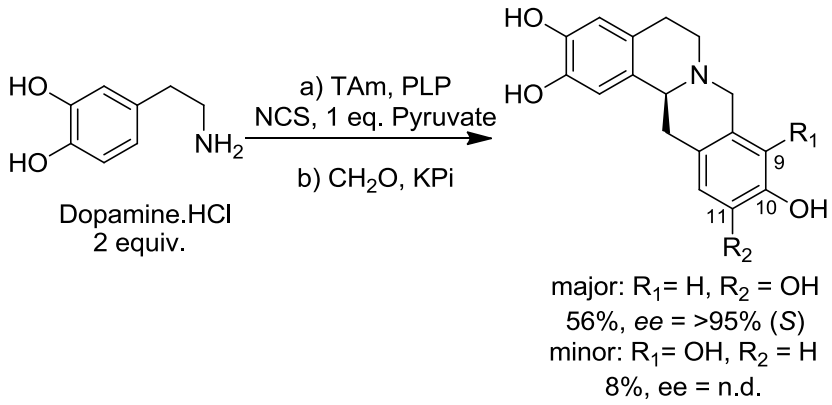


Fig. 5.19 One-pot chemoenzymatic synthesis of tetrahydroprotoberberines. Reaction conditions: (a) 20 mM dopamine, 10 mM sodium pyruvate, 500 μg.mL^{-1} NCS and 20% v.v^{-1} CV2025 lysate, 50 mM HEPES pH 7.5, 37 °C, 3 h. (b) 40 mM formaldehyde, 1 M sodium phosphate, pH 6, 30 min, 37 °C. HPLC method 2 used for *ee*s. Figure adapted from Lichman *et al*.[161]

THPBs are formed biosynthetically by the berberine bridge enzyme (BBE), a flavin dependent oxidase. This enzyme catalyses the formation of exclusively (*S*)-THPBs *en route* to berberine, and typically forms products with 9,10-dihydroxy regioselectivity.[101] The enzyme has been employed in kinetic resolutions and deracemisation cascades for the biocatalytic production of 9,10-dihydroxy-(*S*)-THPBs.[99–101,103] The formaldehyde method developed here produces the alternative 10,11-dihydroxy isomer, and can therefore be seen as complementary to the BBE approach—there are now different chemoenzymatic routes to both 10,11-dihydroxy-(*S*)-THPBs and 9,10-dihydroxy-(*S*)-THPBs. Interestingly, using a fluorinated substrate led to a switch of the typical BBE regioselectivity.[102] Similarly, it has been shown that modification of pH[168] or buffer concentration[60] may shift the regioselectivity of the non-enzymatic PS reaction.

5.4.10 Scale up

Two demonstrative preparative biotransformations at milligram scale (0.2 mmole) were performed using the cascades, demonstrating the scalability of the system. The syntheses were performed on a 0.5 mmole scale, using identical reaction conditions to those used at the micro-scale. The products were isolated and purified by preparative-HPLC. The cascade for the synthesis (*S*)-norlaudanosoline was completed in 2 hours, with 86% conversion; the pure compound was obtained with 62% isolated yield and >95% *ee*. The 10,11-dihydroxy-(*S*)-THPB was formed from dopamine within 2.5 hours with 47% conversion; the pure compound was obtained with 42% isolated yield and >95% *ee*. The product identities were confirmed by NMR spectroscopy (see methods section 2.6.2 and for NMR spectra see Appendix B section 9.1).

5.4.11 Summary

In summary, a novel 'triangular' cascade for the synthesis of BIAs has been established, using the enzymes TAm and NCS. Although similar cascades have been designed with MAOs, the cascade developed here is unique due to the use of a TAm, and the stoichiometric control of conversions afforded by the triangular cascade arrangement. The stereoselectivity of the process is also unique. The syntheses demonstrate that with chemoenzymatic cascades it is possible to obtain remarkably rapid increases in chemical complexity with high atom economy. For example, the major THPB product is formed from a low-cost starting material, in under 3 hours and the reaction involves the formation of 2 C-N bonds, 2 C-C bonds and a chiral centre.

5.4.12 Future work

In order to make these systems truly scalable, methods for facile purification must be developed as preparative-HPLC is not practical for scales of over 1 mmole. Furthermore, the loss of material at the purification step would ideally be reduced. Possible purification options include adsorption to resin[58] or extraction into organic solvent.[60] In order for purifications to be most effective, very high conversions are desired, to minimise contaminating substrates and side products.

Subsequent to the work on the triangular cascades, NCS activity was examined with different substrates (Chapter 6). In this study, an organic extraction method was developed capable of purifying secondary amine NCS products away from starting materials without the requirement of chromatography (see section 6.6.9 and methods section 2.6.3). This method could be easily applied to the multi-enzyme cascade synthesis of THIQs and THPBs described here. This would enable further reaction scale-up.

The scope of the triangular cascade reaction with respect to the starting substrate has been demonstrated through use of 2-(3-hydroxyphenyl)ethylamine. Other phenethylamines with a 3-hydroxy group are likely to be potential substrates also. For example, norepinephrine, L-DOPA and metaraminol may be suitable substrates, as may some halogenated dopamine analogues. If CV2025 does not turn over the amine, it will be possible to screen the numerous reporting TAms for activity. Therefore, the limiting factor in this system is likely to be acceptance of substrates into NCS.

It is also possible to use different aldehydes instead of formaldehyde for the second cyclisation step. As this is not an enzyme catalysed step, many different aldehydes can be used in this step, potentially forming numerous 8-substituted THPBs. Modifying the aldehyde vastly increases the possible compounds that can be produced by this

cascade. Similar 8-substituted THPBs have previously been chemically synthesised from the oxidised berberine by attacking the electrophilic 8-position with Gringard reagents.[185–187] The resulting compounds have been examined for antimicrobial properties,[185] and show promise especially against *Mycobacterium tuberculosis*.[186,187] In contrast to the fully oxidised berberine derivatives, the addition of an R-group onto the THPBs produced here will result in the formation of a new chiral centre at the 8-position. It is possible that the first chiral centre, established using NCS, will determine the isomer formed in the subsequent chemical cyclisation (Fig. 5.20). Whether or not this is the case, the characterisation or separation of products from such a reaction will be challenging.

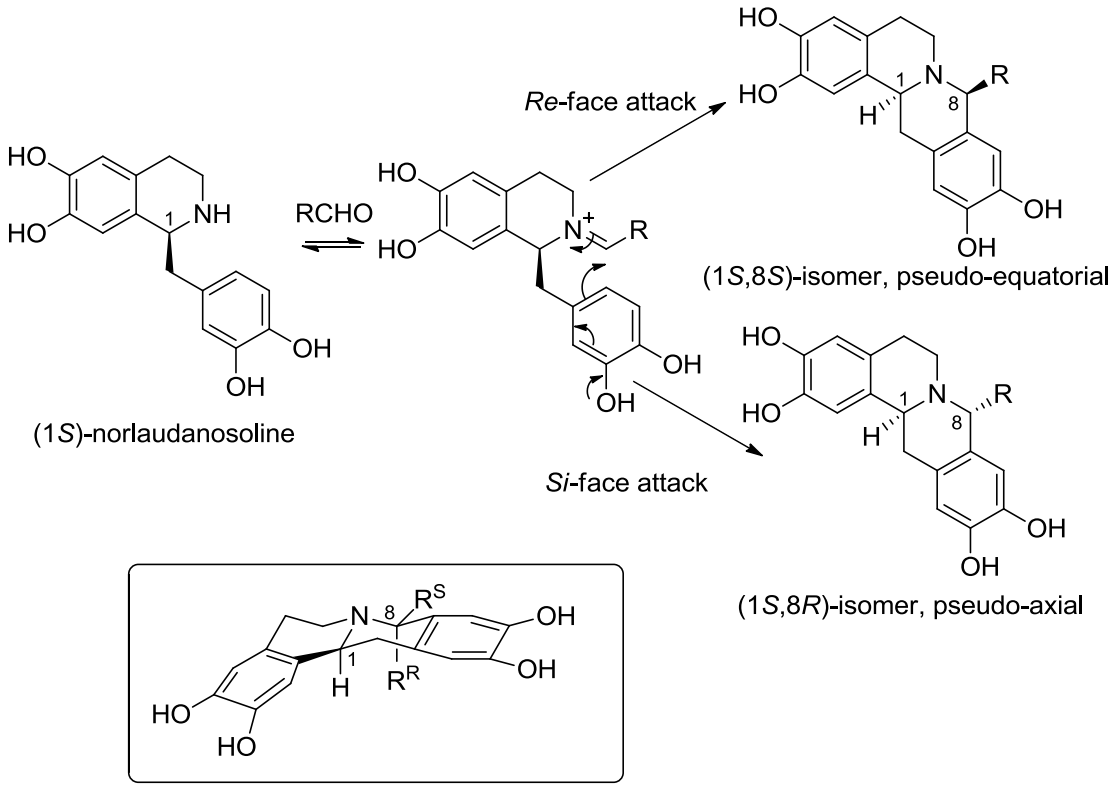


Fig. 5.20 Proposed stereoselective formation of 8-substituted-THPBs. The (*S*)-stereocenter set up by NCS could influence the stereoselective of a subsequent PS reaction, forming a THPB with two stereocenters. For simplicity, only the more favourable *para*-cyclised regioisomer is considered here. For the THPB products, it is assumed that the *trans*-bicyclic conformation will be more favourable than the *cis* conformation, so only the *trans* is considered here. The *Re*-face attack of the imine by the catechol results in the R-group in a pseudo-equatorial position, whereas a *Si*-face attack results in the R-group being in a less favourable pseudo-axial position. A 3D depiction of the THPB is shown in the box. The expected major products are the pseudo-equatorial (1*S*,8*S*)-THPB. The ratio of the diastereomers formed may depend on the size of the aldehyde R-group.

Although the triangular cascade has elegance in terms of the single starting material and the TAm equilibrium position, it is ultimately a ‘dimerisation’ cascade and so is very limited by the NCS mechanism requirements. A simple modification to the cascade would be the use of different amines for NCS and TAm. The crucial step here would be

the acquisition of a TAm that did not accept dopamine as a substrate. It would then be possible to introduce different amines that could be converted *in situ* to aldehydes, which, in turn, would combine with dopamine in a PS reaction catalysed by NCS. This system would have challenges with respect to the cross-reactivity of substrates, but considering the large array of TAms available to screen this would be a sensible next step.

In order to diversify NCS cascade products further, other multi-enzyme systems must developed. As described in section 5.4.1, formation of aldehydes can be catalysed by alcohol dehydrogenases, laccases or carboxylic acid reductases (Fig. 5.9). Dopamine itself can be formed from tyrosine with a combination of enzymes (Fig. 5.8). THIQ reaction products can be modified with methyltransferases and various different CYPs. In all, there are an incredible number of possible cascade designs centred on the NCS Pictet-Spengler reaction. The formation of these will allow the biocatalytic, green production of diverse molecules.

6 Chapter 6: Screens and substrates

6.1 Introduction

6.1.1 NCS engineering

It has been demonstrated that it is possible to engineer NCS towards increased activities for both natural and unnatural substrates (section 4.4.7). The poor kinetic parameters for NCS suggest it may be possible to improve the catalytic efficiency of the enzyme, especially with respect to the K_m^{app}—the mutations A79I and Y108F showed that modification of kinetic parameters is possible. The substrate promiscuity, especially with respect to the aldehyde, also confirmed that the enzyme has the potential to catalyse the formation of diverse chiral tetrahydroisoquinolines (THIQs). Notably, the mutation L76A demonstrated that activities towards specific substrates can be significantly improved.

Using directed evolution methods, it may be possible to develop a 'toolkit' of NCSs. This would consist of numerous enzymes, each evolved to specialise in the synthesis of a particular chiral THIQ. Directed evolution would consist of accumulating multiple mutations onto NCS; these mutations could result in new desirable activities.

There are numerous methods for mutating enzymes in directed evolution experiments. These include error-prone PCR, site-saturation mutagenesis, recombination, phylogenetic and combinatorial methods.[107–111] The choice of mutation method depends on the desired outcome of the evolution, but also on the information available on the system being investigated. The data presented in the project has indicated that NCS operates with a dopamine-first mechanism (section 4) and therefore it is now possible to mutate NCS in a rational manner. For example, the mutation L76A altered aldehyde specificity, and Leu-76 was selected for mutation as it is in the aldehyde-binding site.

There are a number of exciting potential avenues for evolving NCS activities. This includes the aforementioned modification of kinetic parameters and aldehyde substrate specificity. There are also certain aldehyde substrates that are currently poorly accepted by NCS: in particular α-substituted bulky aldehydes. Furthermore, it may also be possible to turn over ketones. There could be potential for the modification of the amine tolerance, but this is perhaps more challenging than modifying aldehyde specificity because the amine is bound deep into the active site. It may also be conceivable to modify the stereo- or regio-selectivity of the NCS reaction, providing access to valuable (*R*)- or *ortho*-THIQs respectively.

6.1.2 Screening requirement

In directed evolution investigations a crucial limitation on the experimental outcome is the quantity of mutants that can be screened. Screening activities with purified enzyme and HPLC provides excellent sensitivity, but is often too low-throughput for effective directed evolution. More rapid screening methods are desired for the effective directed evolution of NCS. This section describes attempts to develop experimental methods suitable for the future directed evolution of NCS.

The best methods for directed evolution are selection methods, where genotype and phenotype are linked and only active variants survive to the next evolution round.[188] Unfortunately, developing a selection strategy for NCS activity would be a considerable challenge and this was not pursued. A similar excellent option is a colony based assay, where screening reactions can occur on the agar plates and a colorimetric change is observed. Again, no plate based assay for NCS was considered, mostly as in these methods the high background reaction would be difficult to control.

The background chemical Pictet-Spengler reaction is promoted by cell lysate, even without the presence of recombinant NCS. The use of purified enzymes enables the greatest control of the background reaction. However, purification of enzymes is costly in both time and resources. Furthermore, NCS activity is inhibited by imidazole, so if NCS is purified with nickel and eluted with imidazole, an extra desalting step is required for optimal expression. It was discovered that it is possible to significantly reduce the background reaction in cell lysates that had been treated with a desalting step, without the requirement for purification (see section 5.3.4). This desalting method for enzyme preparation is reasonable to pursue in screening methods as a compromise between enzyme purification and use of cell lysate.

A screening method was developed by Pesnot *et al* using fluorescamine to detect primary amines.[57] Fluorescamine reacts with primary amines to form a fluorescent product. This method detected the loss in dopamine concentration associated with a successful reaction. The secondary amine THIQ product does not form fluorescent products with fluorescamine. However, fluorescamine also reacts with proteins to form a fluorescent product. Therefore, this method is effective for screening a single purified enzyme with multiple aldehydes, as with a purified enzyme the protein concentration can be controlled and minimised. It would be less effective for screening cell lysates of multiple mutants, where protein concentrations are unpredictable.

Alternative screening methods for NCS activity were sought which would be more compatible with unpurified cell lysate. A method which could specifically detect

secondary amines would be ideal as this would enable product formation to be followed rather than substrate depletion. With this in mind, the chloranil test was investigated. This test is often used for detecting amines on peptide synthesis resins. Chloranil appeared to be insoluble upon mixture with aqueous solution, and no readable output was possible. This mixing with water is necessary as the enzyme reactions occur in aqueous solution—therefore the chloranil method was not explored further.

6.2 Purpald®

6.2.1 Proof of concept

The compound Purpald® reacts with aldehydes to form a purple coloured product (Fig. 6.1).[189] This method is often used to detect formaldehyde production. In a recent directed evolution study, Purpald® was used in an enzyme coupled assay for the detection of methanol.[190] Purpald® was investigated as a screening method for NCS. Although this method detects loss of substrate rather than formation of product, it was thought that it would be more appropriate for screening mutant lysate reactions than fluorescamine as it would not be sensitive to proteins present.

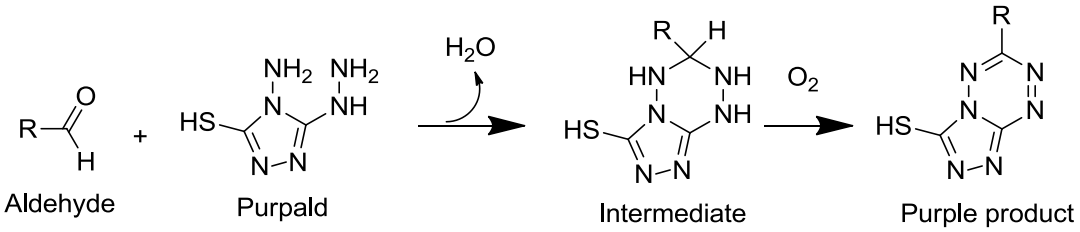


Fig. 6.1 Mechanism of aldehyde detection with Purpald®. Purpald® reacts specifically with aldehydes to form an intermediate complex. Oxidation of the complex in air over a few minutes will give rise to a purple complex. It is possible to use an oxidant such as $NaIO_4$ to accelerate oxidation.

Test reactions were prepared to probe the behaviour of Purpald® with different aldehydes (Fig. 6.2). Hexanal, heptanal and phenylacetaldehyde (PAA) were investigated: all showed reasonable sensitivity between 0.25 and 10 mM aldehyde, which is suitable for NCS reaction screening where aldehyde concentrations are around 1-10 mM. Aldehyde concentrations above 10 mM caused saturation of absorbance detection. Absorbance at 550 nm and 595 nm was monitored, with 550 nm showing greater values. The sensitivity of the assay depended on the aldehyde investigated, with PAA giving double the absorbance of hexanal which, in turn gave twice the signal intensity of heptanal.

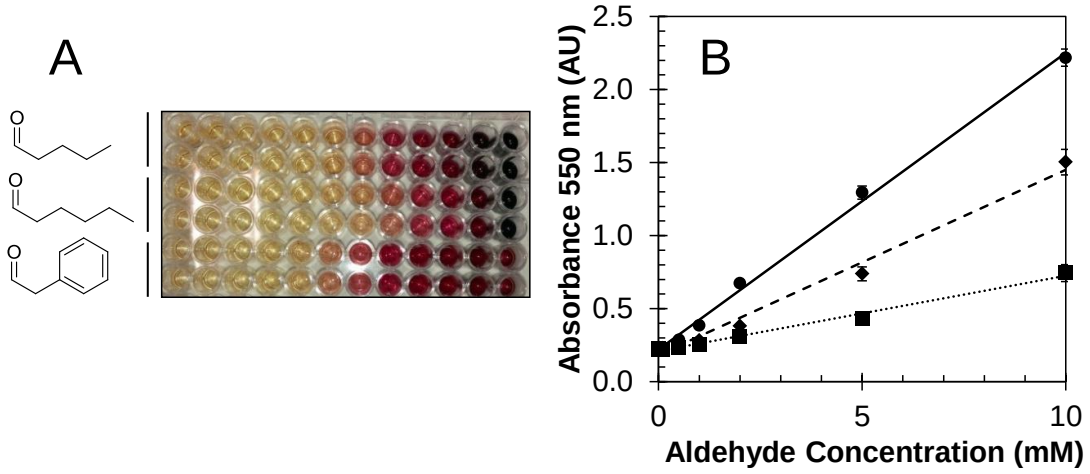


Fig. 6.2 Purpald® reaction with different aldehydes. A: Picture of increasing aldehyde concentrations from L to R: 0, 0.01, 0.1, 0.5, 1, 2, 5, 10, 15, 20, 50, 250 mM. **B: Calibration curve.** Absorbance measured at 550nm. Aldehydes: Phenylacetaldehyde (circles, solid line), Hexanal (diamond, dashed) and Heptanal (square, dotted line). Each measurement is mean of duplicates, error bars show one standard deviation. Reaction conditions: 30% $v.v^{-1}$desalted boiled lysate, 10 mM dopamine, 0-10 mM aldehyde, then 1:1 combination with Purpald® (10 $mg.mL^{-1}$ in 1M NaOH).

The absorbance spectrum of the Purpald® complex with PAA was measured, showing a peak at 527 nm (Fig. 6.3).

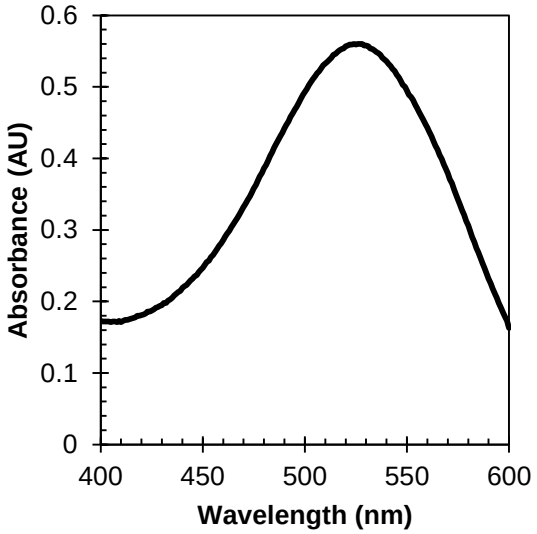


Fig. 6.3 Purpald®-PAA complex. Sample conditions: 20% $v.v^{-1}$ desalted lysate, 50mM PAA, 50% $v.v^{-1}$ Purpald® (10 $mg.mL^{-1}$ in 1M NaOH). All diluted 1:100 in MES pH 7.5.

6.2.2 Time courses and amine reaction

A number of experiments were conducted in order to investigate the behaviour of Purpald® and to optimise conditions for its use in screening assays. It became evident that it was possible to track the consumption of aldehyde in a reaction using a standard curve to convert absorbance values into concentrations (Fig. 6.4). The pET control, without NCS, however, was observed to show consumption of 50% aldehyde perhaps

due to background Pictet-Spengler reaction or consumption of aldehyde by other processes.

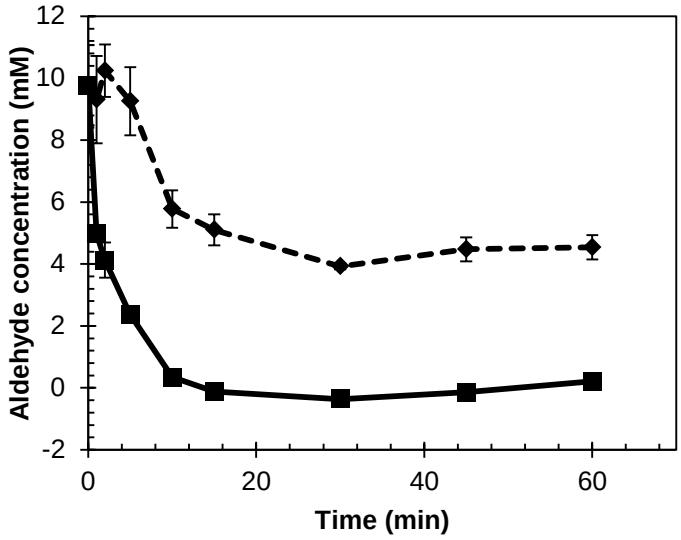


Fig. 6.4 Time course reaction using Purpald®. Sample conditions: 20% v.v^{-1} desalted lysate, 10mM dopamine and 10 mM PAA. Reactions quenched at time points with 50% v.v^{-1} Purpald® (10 mg.mL^{-1} in 1M NaOH). WT is squares, solid line. pET control is diamonds with dashed line. Aldehyde concentrations calculated using standard curve. Measurements are triplicates, error bars show one standard deviation.

It was also noticed that Purpald® formed a yellow colour when incubated with dopamine, even without the presence of aldehyde. This was confirmed by measuring the absorbance of different samples containing control pET lysate and combinations of substrates at 10 mM (Fig. 6.5). PAA-Purpald® produces an absorbance peak at around 530 nm, whilst dopamine-Purpald® gave a smaller peak at around 410 nm. When both dopamine and PAA were present in the sample, both peaks were observed. The aldehyde peak was reduced in size, but the dopamine peak seemed to become much more intense. The reduction in the size of the aldehyde peak may be due to the consumption of Purpald® by dopamine or due to formation of non-reactive imines between aldehyde and dopamine. Purpald® also reacts with secondary amines in a similar manner to primary amines. The identity of the amine-Purpald® reaction product is unknown, but may involve nucleophilic attack from the amine onto Purpald®. Overall, this previously unobserved reaction between amines and Purpald® reduced the sensitivity of this method for the detection of aldehydes.

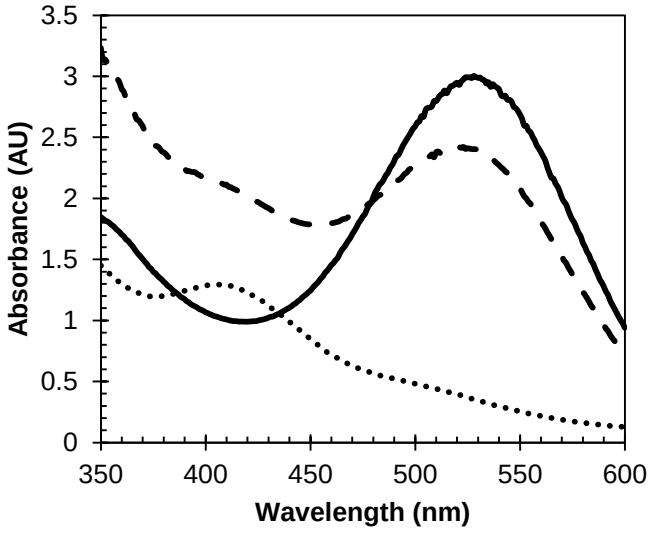


Fig. 6.5 Spectra of Purpald® products. Dopamine (dotted line), aldehyde (solid line), both (dashed). All substrates at 10 mM.

Despite the reaction between amines and Purpald®, it was still possible to effectively track the aldehyde consumption with Purpald®. Comparison of Purpald® readings with HPLC product detection showed that (with dopamine, PAA and WT desalted lysate) aldehyde consumption mirrored product formation (Fig. 6.6). However, there remained a large decrease in aldehyde concentration with pET controls, due to the consumption of aldehydes or background Pictet-Spengler activity. This suggested that although Purpald® detection can follow successful reactions well, it may not be ideal for the detection of unknown activities.

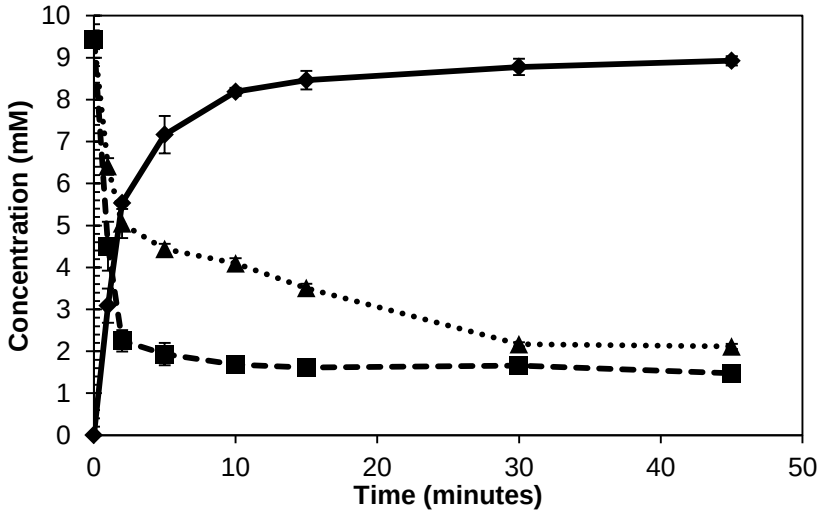


Fig. 6.6 Time course of reaction. Squares, dashed line: PAA with WT (Purpald® detection). Diamonds, solid line: THIQ product with WT (HPLC detection). Triangles, dotted lines: PAA with pET (Purpald® detection). PAA detection with Purpald® at 522 nm and converted to concentration with standard curve. Measurements are means of triplicates, error bars are one standard deviation. Reaction conditions: 80% $v.v^{-1}$ desalted lysate, 10 mM dopamine, 10 mM PAA. Then dilution 1:1 with Purpald® (10 $mg.mL^{-1}$ in 1M NaOH).

6.2.3 Optimisation

A screen of *Tf*NCS mutants with various aldehydes was conducted using Purpald® as the detection method (Fig. 6.7). The selections of mutants are discussed below (section 6.3.1). Although reasonable activities were identified with HPAA and PAA, high background readings or low activities for all other aldehydes prevented useful data being collected. It was known that WT has significant activity with hexanal compared to background, but in this screen this activity was not evident, therefore it was clear that there were problems with the assay. The origin of this could have been the very high loading concentration of lysate (80% $v.v^{-1}$).

In order to optimise the screening conditions, the optimum lysate loading concentration was investigated. Hexanal was used as the aldehyde as this showed poor sensitivity in the previous assay. Reactions with different desalted lysate concentrations were measured, comparing WT to a pET control (Fig. 6.8). Surprisingly, losses in aldehyde in the controls were shown to be consistent regardless of lysate concentration. This suggested that the aldehyde was intrinsically unstable or perhaps was forming imines with dopamine. The best condition to detect activity over background was after 5 minutes with 50% $v.v^{-1}$ lysate loading. In all lysate loadings, after 30 minutes aldehyde consumption was greater in controls than with enzymes. Although it is possible to detect activity with the Purpald® assay, the small dynamic range, and high background conversion suggests it was not be suitable as a general method.

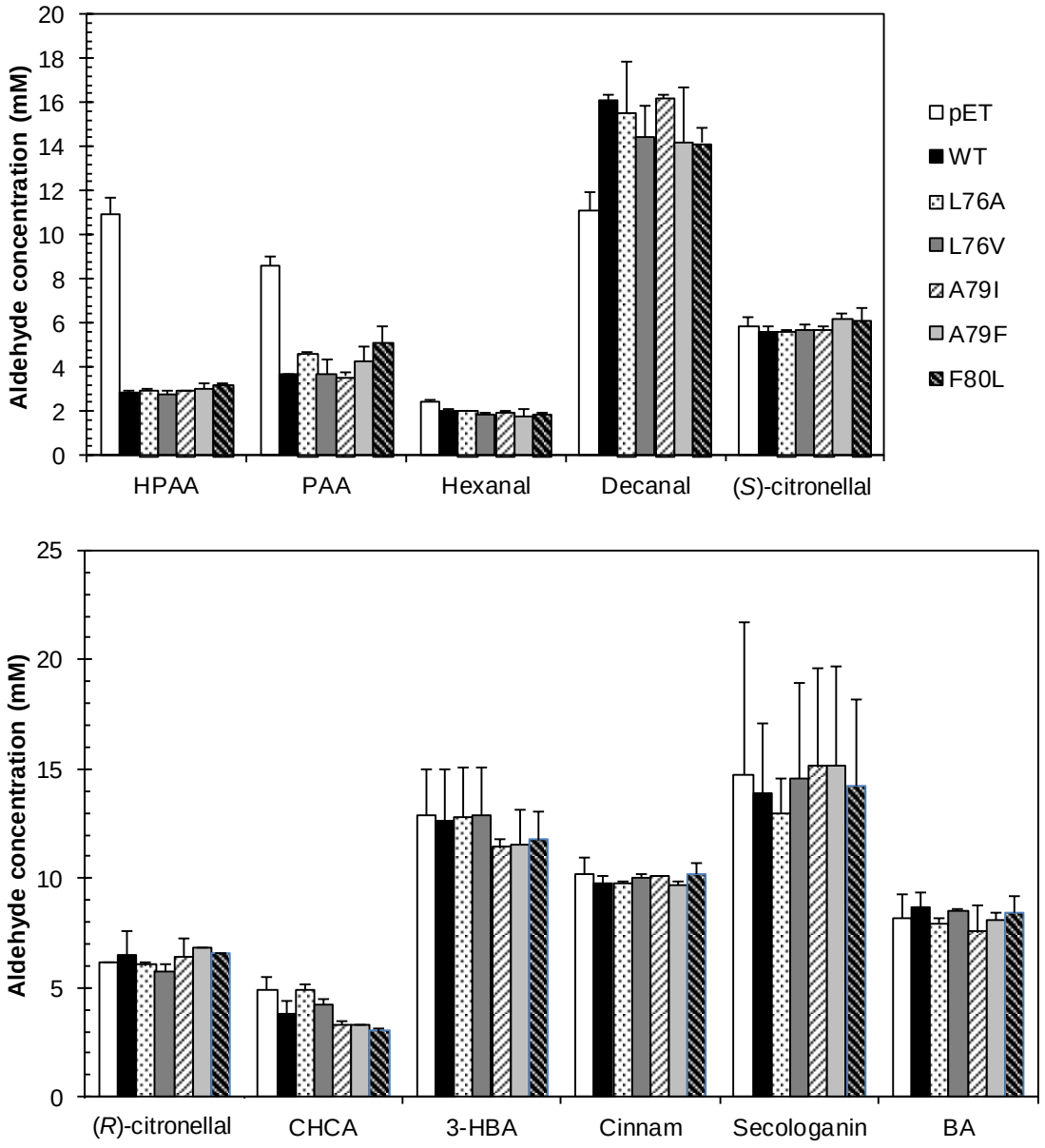


Fig. 6.7 Aldehyde Purpald® screen 1. HPAA: 4-hydroxyphenylacetaldehyde, PAA: phenylacetaldehyde, CHCA: cyclohexanecarboxaldehyde, 3-HBA: 3-hydroxybenzaldehyde, cinnam: cinnamaldehyde, BA: benzaldehyde. Conditions: desalted lysate 80% $v.v^{-1}$, 10 mM dopamine, 10 mM aldehyde. Then dilution 1:1 with Purpald® (10 $mg.mL^{-1}$ in 1M NaOH). Aldehyde concentrations determined by standard curves.

A final attempt at using Purpald® to screen mutants with various aldehydes was conducted, using purified enzymes rather than lysates, and doubling the concentration of Purpald® used for detection. In constructing standard curves, there appeared to be very little Purpald® response to benzaldehyde and 3-hydroxybenzaldehyde. This suggests that Purpald® is not generally applicable in screens with all aldehydes. Activity was observed with HPAA, PAA, Hexanal and (*S*)-citronellal. Overall, despite

efforts to optimise reaction conditions, it was clear that Purpald® was not a good detection method for NCS reactions.

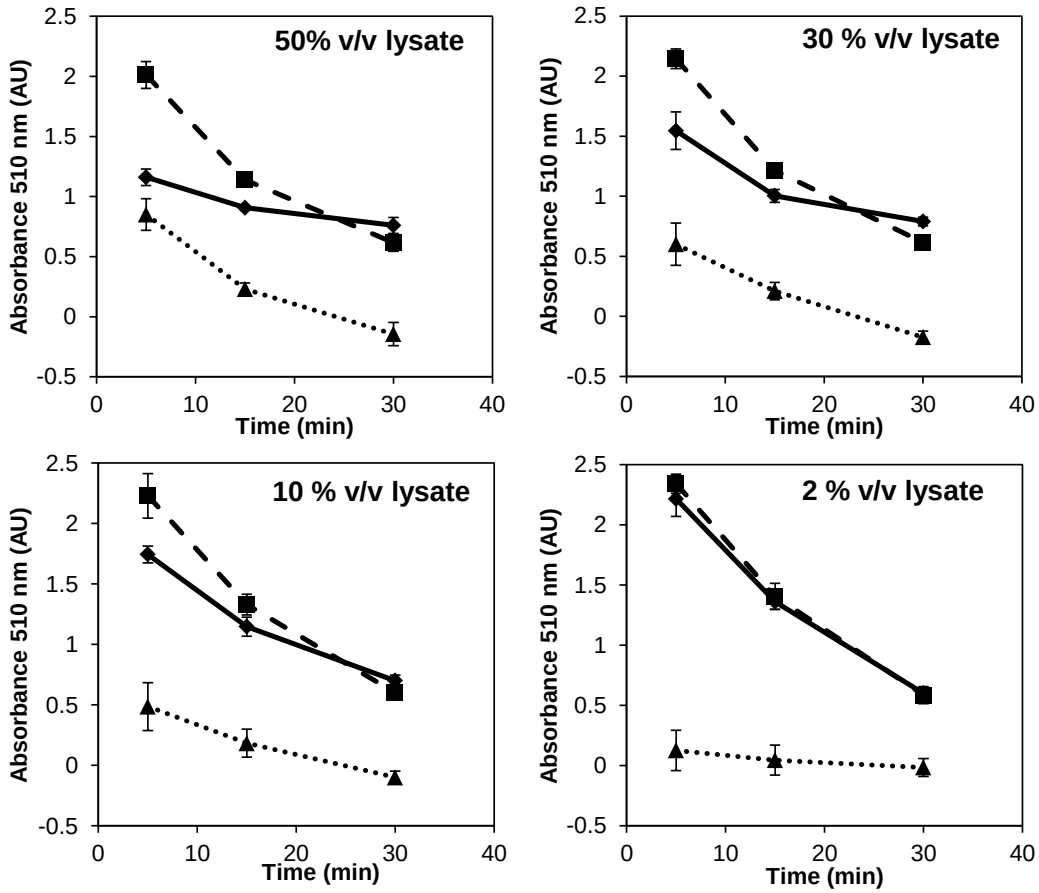


Fig. 6.8 Lysate loading examination. WT (diamonds, solid line), pET (square, dashed line), difference (triangle, dotted line). Conditions: 10 mM dopamine, 10 mM hexanal, varying lysate concentration. Then dilution 1:1 with Purpald® (10 mg.mL^{-1} in 1M NaOH).

6.3 Substrate screens

6.3.1 Mutants

Mutants of *Tf*NCS were screened with a variety of substrates. The mutants can be split into two main categories. The residues Leu-76, Ala-79 and Phe-80 are all in the region of active site where aldehydes R-groups are expected to bind based on the dopamine-first mechanism (Fig. 6.9). The other residues investigated in this section were Tyr-108, Glu-110 and Asp-141. All of these are 'mechanistic' residues, that are expected to play a role in the mechanism of the enzyme, rather than just substrate binding. They were targeted when amines were varied, especially when the substrates require more extreme modifications of activity (e.g. *N*-methylated amines). Formation of mutants is described in method section 2.1.

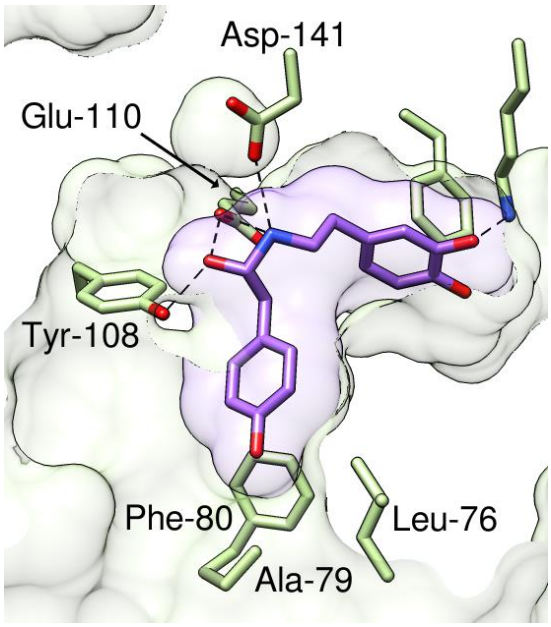


Fig. 6.9 Residues targeted in screening assays. Green sticks: residue side chains. Green surface: solvent accessible protein surface. Purple surface: solvent accessible ligand surface. Purple sticks: docked structure of (*S*)-aminol to demonstrate dopamine-first mechanism interactions. See section 4 for more details on mechanism.

6.3.2 Fluorescamine and aldehydes

Fluorescamine was used to screen mutants with aldehydes in a method similar to that reported by Pesnot *et al*.[57] This method requires the protein to be purified. The first attempt using a fluorescamine screen, with a 1 hour reaction time, provided low signals—WT *Tf*NCS with HPAA showed only about 40% conversion. To improve signal-to-noise, reaction times were increased to three hours.

The enzymes Δ29*Tf*NCS WT, L76A, L76V, A79F, A79I and F80L were examined, along with *Cj*NCS2 (see section 2.1.2 for creation of mutants). In the reactions, the

amine used was dopamine. No activities were observed with acetaldehyde, cyclohexanecarboxaldehyde (CHCA), cinnamaldehyde, 3-hydroxybenzaldehyde (3-HBA), benzaldehyde (BA) and the ketone 4-hydroxyphenylacetone (4-HPA). This is in line with previous observations that α-substituted aldehydes and ketones are poor substrates for NCS.[57] Activities were also not observed with 3-methoxylbutanal, pentanal and cyclopropanecarboxaldehyde, though these three compounds were not freshly distilled prior to use and may have oxidised.

Activity was observed with 4-hydroxyphenylacetaldehyde (HPAA), phenylacetaldehyde (PAA), hexanal, heptanal, decanal, (*S*)-citronellal and (*R*)-citronellal (Fig. 6.10). When this reaction was initially conducted, activity with hexanal, decanal and citronellal had not previously been identified.

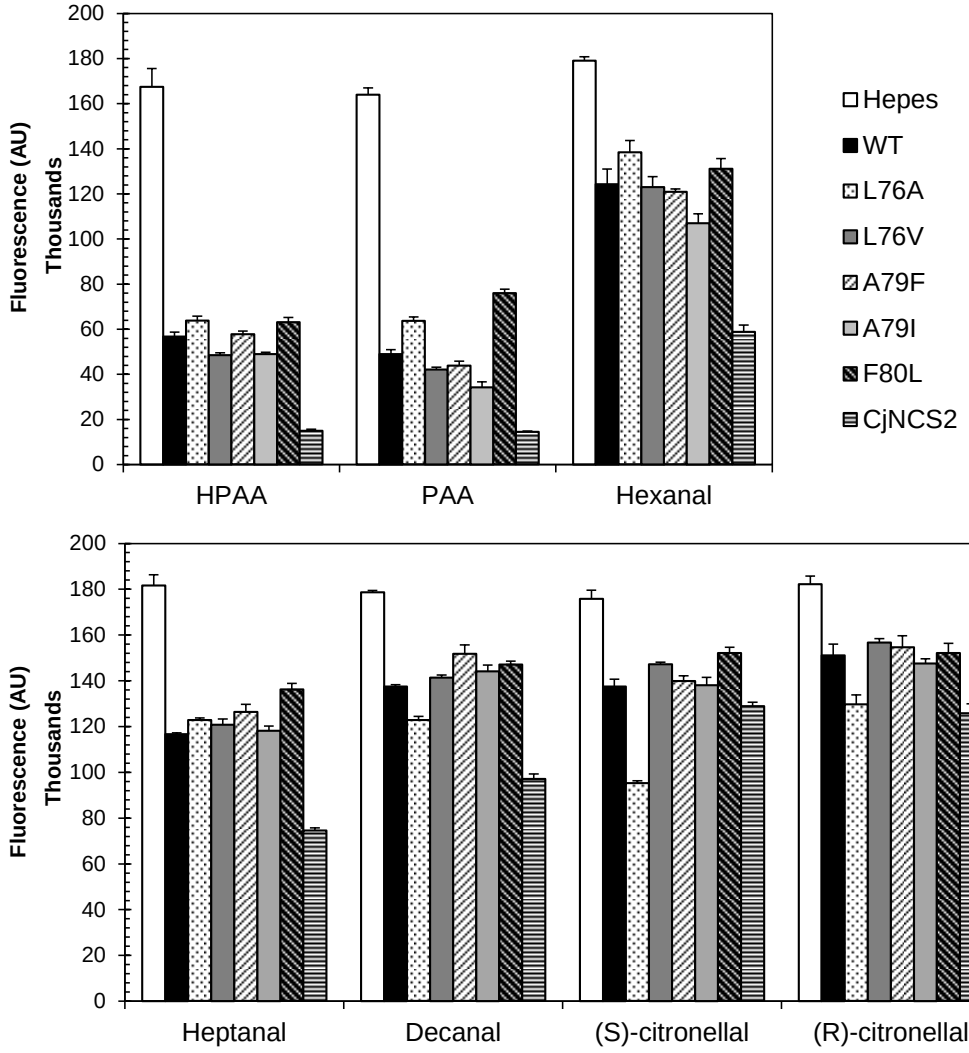


Fig. 6.10 Aldehyde mutant fluorescamine screen. Conditions: 100 μM dopamine, 1 mM aldehyde, 25 μg/mL purified enzyme, 3 h incubation then addition of fluorescamine (10 mM in MeCN 0.25x volume of reaction).

For all substrates, except for (*S*)-citronellal, *Cj*NCS2 showed less fluorescence than any *Tf*NCS mutant. Although it is clear that *Cj*NCS2 had activity for all these aldehydes, direct comparison with *Tf*NCS mutants was difficult as the background fluorescence of the proteins were different. The purified protein *Cj*NCS2 had less than half the fluorescence than purified WT-Δ29*Tf*NCS at the same concentrations.

It is possible to compare *Tf*NCS mutants with each other, as they demonstrate similar fluorescent responses. The mutant A79I showed increased activity compared to WT for HPAA, PAA and hexanal. This was observed in kinetic assays and discussed in section 4.4.3. The mutant L76A showed increased activity for both citronellal aldehydes and decanal. The response of L76A to citronellal was discussed in section 4.3.7. The increased activity with decanal had not previously been observed but, as for citronellal, can be accounted for by steric effects. The mutants L76V, A79F and F80L showed similar activities to WT, and did not show any significant improvement in activity with any particular aldehydes.

6.3.3 Fluorescamine and amines

Primary amines were screened with a different selection of mutants (Fig. 6.11). These mutants included aldehyde-binding mutants known to cause changes in activity (L76A and A79I), and mutants further inside the active site near to where the amine nitrogen is expected to bind: Y108F, E110D and D141E (see section 4). The amines screened (with PAA) included dopamine and metaraminol, which was previously shown to be turned over by *Cj*NCS2.[57] More unusual amines tested were 2-bromo,5-hydroxyphenylethylamine and 3,4-dihydroxyphenylpropylamine.* In the bromo amine, the bromide blocks the typical cyclisation position in the Pictet-Spengler reaction, so the intention here was to force formation of an *ortho*-product. The latter, a propylamine, would form a seven-membered ring in a Pictet-Spengler reaction.

Activities with dopamine were similar to those expected from studies of the mechanism: activity was knocked out with E110D and reduced with L76A, Y108F and D141E. Metaraminol activity was observed but only for WT and A79I. Notably, it is known that *Cj*NCS has some activity with metaraminol, but this was not observed in this assay. This suggests the sensitivity of this assay was poor. No activities were shown with the brominated amine. Strangely, the propyl amine did not give a fluorescent response at all, even in the HEPES background.

* These compounds were synthesised by Dr Eleanor Lamming

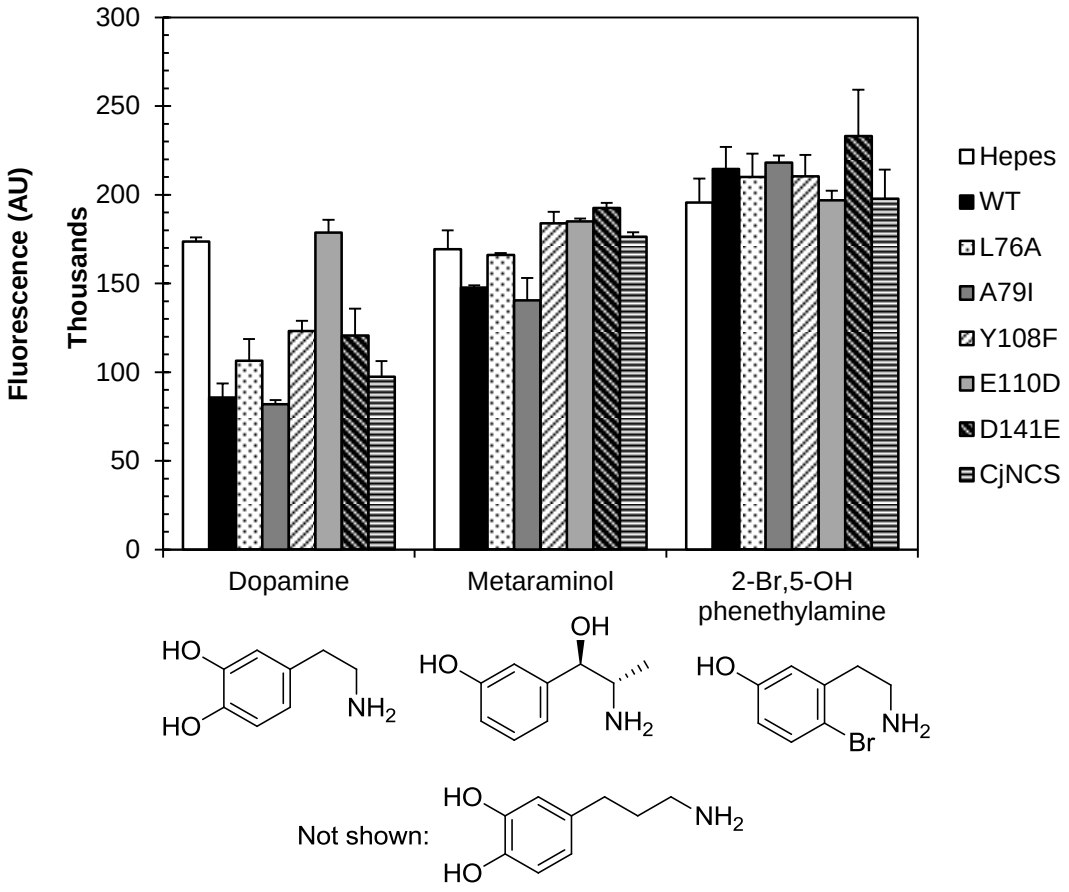


Fig. 6.11 Amines with fluorescamine. Conditions: 500 µM amines, 1 mM HPAA, 25 µg/mL purified enzyme, 3 h reaction time then addition of fluorescamine (10 mM in MeCN 0.25x volume of reaction).

6.3.4 Purpald® and secondary amines

Secondary amines were tested for activity with the same set of mutants as section 6.3.3. As the compounds tested were secondary amines, fluorescamine could not be used, so a Purpald® assay was conducted instead. The aldehyde used was PAA, which in the Purpald® assay performed relatively well (see section 6.2). A standard curve was used to convert absorbance values into concentrations; however it was clear that there was a large background reaction (Fig. 6.12). Regardless, it was possible to see that the results for dopamine matched very well with the fluorescamine results (Fig. 6.11), so the assay could certainly identify some activities. Despite this, no activity was observed for the secondary amines: *N*-methyl-3-hydroxyphenethylamine*, 4,*N*-dimethyldopamine* and isoprenaline (Fig. 6.12). This was not especially surprising considering the nitrogen-enzyme interaction is vital; disrupting this interaction by adding alkyl groups is detrimental for NCS activity.

* These compounds were synthesised by Dr Eleanor Lamming

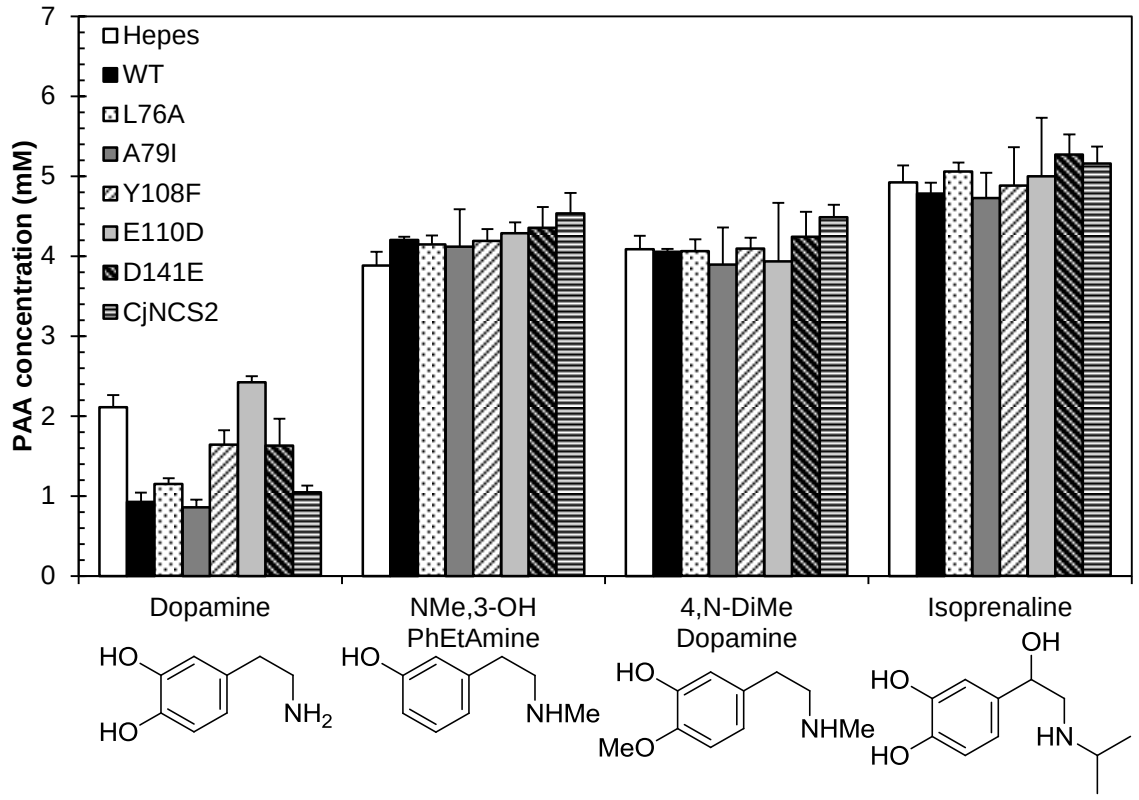


Fig. 6.12 Purpald® amines. Conditions. 45 µg/mL purified enzyme, 18 mM amines, 9 mM PAA. 1 hour incubation then 1:1 addition of Purpald® (20 mg.mL^{-1} in 2N NaOH).

6.4 Regioselectivity screens

6.4.1 Introduction

Two different regioisomers are possible in the Pictet-Spengler reaction with dopamine, depending on the orientation of the catechol ring at the electrophilic aromatic addition step. In the chemical phosphate catalysed reaction the major product is the *para*-regioisomer (tetrahydroisoquinoline-6,7-diol) and the minor product is the *ortho*-regioisomer (tetrahydroisoquinoline-7,8-diol). Recently, a number of *ortho*-substituted THIQs have been shown to have superior antibiotic activities than similar *para*-THIQs for inhibiting *Mycobacterium tuberculosis* growth.[94] To date NCS produces exclusively *para*-THIQs (see section 5.3.1). Enzymatic formation of o*rtho*-THIQs is known: the biosynthesis of the antibiotic saframycin for example involves an *ortho*-Pictet-Spengler cyclisation step, though the mechanism of this non-ribosomal peptide associated enzyme is not known (see section 1.7.4).[89]

An enzymatic route to *ortho*-THIQs would enable the facile production of interesting bioactive compounds and would thus be very valuable. It would also be beneficial if the products could be single enantiomers. NCS was rationally engineered with the intention of producing *ortho*-THIQs. The engineering was based on the dopamine-first

mechanism. The key electrophilic addition cyclisation step that establishes the regioisomer of the product is catalysed by Lys-122. By repositioning the catalytic lysine in the active site, it was hoped that the cyclisation directionality could be adjusted, and form *ortho*-THIQs preferentially to the typical *para*-product (Fig. 6.13). The substrate nitrogen position is dictated by the carboxylic acid residues Glu-110 and Asp-141, so it was hoped that moving the lysine residue would only affect the catechol position and not the entire substrate orientation.

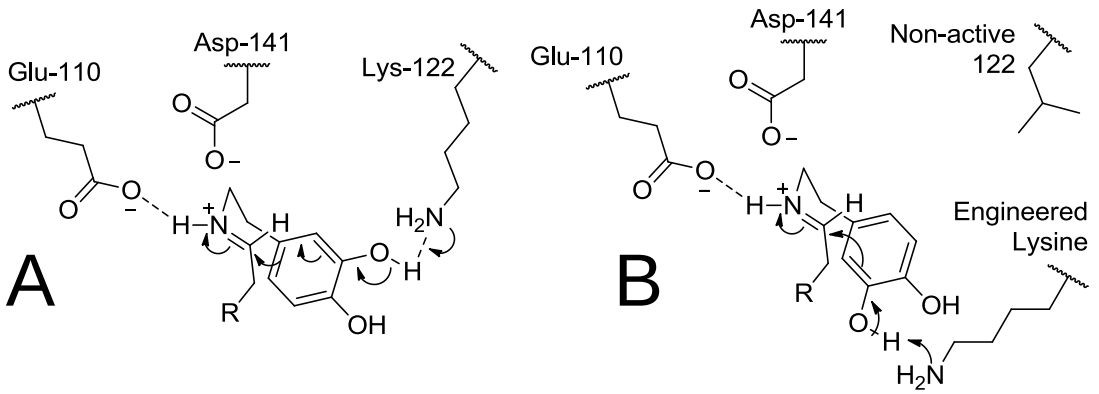


Fig. 6.13 Cyclisation step of NCS reaction. A. The *Para*-cyclisation step in dopamine-first mechanism. **B.** Proposed *ortho*-cyclisation step in engineered NCS.

An alternative strategy for engineering the enzyme towards the formation of *ortho*-THIQs may be to attempt to modify the chemical mechanism of the cyclisation, and encourage the formation of a spiro intermediate (Fig. 6.14). This intermediate could then undergo a C-C bond migration to form a bicyclic compound that shares a C-C bond, before losing a proton to form the product. The spiro intermediate could be formed if, rather than the 3-hydroxy group of the dopamine providing the key electron density, the 4-hydroxyl group could be more dominant. This would be encouraged by removal of the 3-hydroxy group (i.e. tyramine) or by methylation. The subsequent C-C migration would be controlled by the electron donating nitrogen, but two products remain possible. A similar spiro-mechanism has previously been observed for the related Bischler-Napieralski reaction.[191]

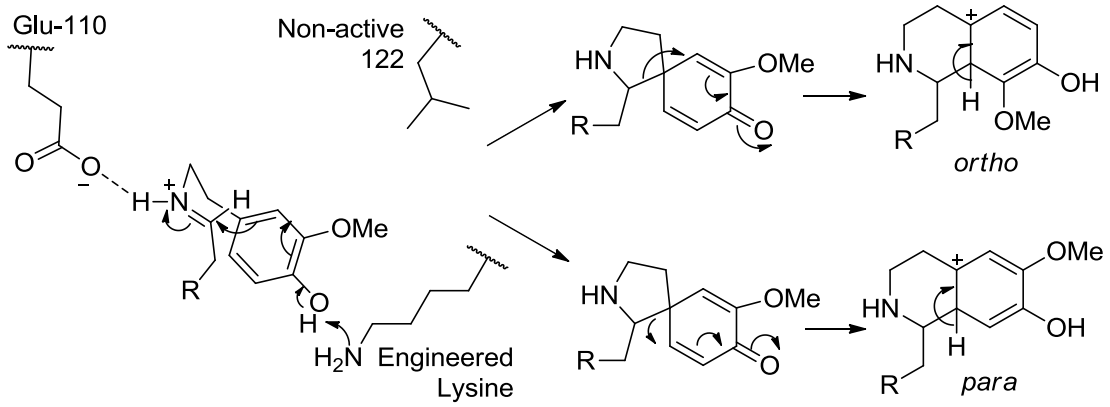


Fig. 6.14 Theoretical spiro-mechanism. Nitrogen is an electron donating atom so would dictate the bond that migrates.

Both strategies for altering the regioselectivity of the enzyme catalysed reaction are extremely speculative, but the former *ortho*-cyclisation approach is probably the best as there is clear precedent in the phosphate catalysed reaction. The dopamine-first mechanism was used as a guide to engineer the enzyme.

The strategy for this engineering was first mutation of the Lys-122 to a non-active hydrophobic residue (leucine or phenylalanine). Then, by mutating a different residue to a lysine, the catalytic nitrogen would be moved around the active site. The best candidate residues for this are those near an '*ortho*' position: Ala-69, Leu-72, Ile-85 and Leu-95 (Fig. 6.15). The residues near the faces of the catechol—Phe-112 and Met-183—were also tested. The residue Met-126 was also examined despite its considerable distance from the catechol. Of course, the catalytic Lys-122 in the typical mechanism has evolved to be in a particular environment to optimise catalysis. This includes modification of the pK_a and position of a water channel (see section 4.3.3). Any activity observed in these double mutants would be extremely small—subsequent directed evolution would be required to make the enzyme a useful biocatalyst. Due to the very small activities expected the screening was conducted with HPLC detection, as this has best sensitivity and reliability.

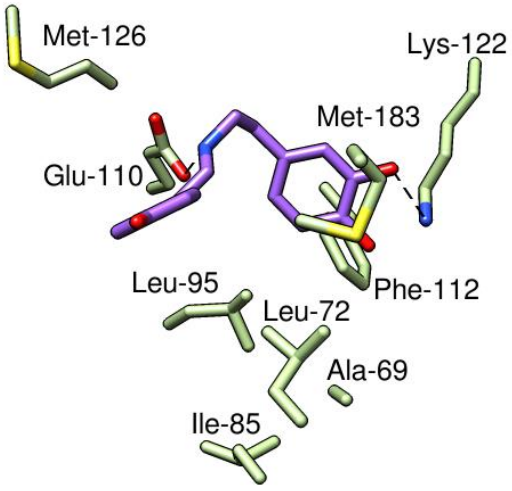


Fig. 6.15 Residues modified to lysine in regioselectivity screen. Lys-122 was mutated to leucine and phenylalanine. Glu-110 was not changed but is depicted to demonstrate active site orientation.

The single mutants tested in this study were K122L, K122F and F112K. To the K122L mutant background the mutants A69K, L72K, I85K, L95K, F112K, M126K and M183K were added and examined. The double mutant K122F F112K was also investigated.

6.4.2 Phenethylamines

In order to speed up the enzyme preparation, a rapid, small scale method was developed. Pellets from small scale cultures (2 mL) were lysed (200 µL BugBuster) and centrifuged. Nickel agarose suspension (50 µL) was added to the supernatant and the mixture incubated. The Ni-agarose was isolated, washed (40 mM imidazole) and the protein eluted (50 mM EDTA) (Fig. 6.16). EDTA was used for elution rather than imidazole as imidazole inhibits the NCS reaction.

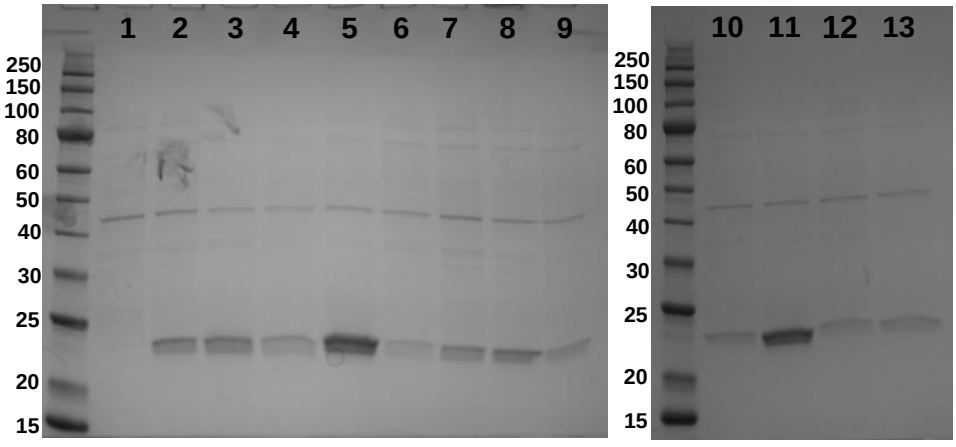


Fig. 6.16 SDS-PAGE of Lysine mutants. Lanes: 1, pET control; 2, WT; 3, K122L; 4, K122L-69K; 5, K122L-72K; 6, K122L-85K; 7, K122L-95K; 8, K122L-112K; 9, K122L-126K; 10, K122L-183K; 11, K122F; 12, F122L; 13, K122F-F112K. Purified from small scale cultures.

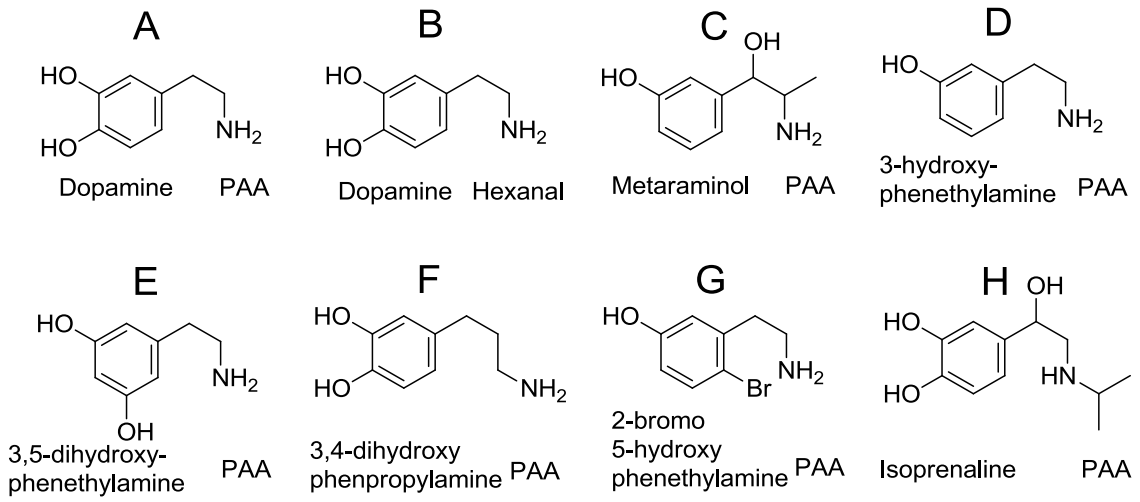


Fig. 6.17 Compounds tested with Lys mutants. Assay conditions: 12.5 mM amines (except D, 5mM), Aldehydes 10 mM. 20% $v.v^{-1}$ enzyme, 1 h, 37 °C. Reactions analysed by HPLC. PAA: phenylacetaldehyde. Compounds 3-hydroxyphenethylamine, 3,5-dihydroxyphenethylamine, 3,4-dihydroxyphenylpropylamine and 2-bromo,5-hydroxyphenylethylamine were synthesised by Dr Eleanor Lamming.

The purified enzymes were tested for activity with a variety of amines, with the intention of observing either novel or unusual activities (Fig. 6.17).* Of those tested, WT NCS was known to have activity with dopamine, metaraminol and 3-hydroxyphenethylamine.

HPLC analysis of the reactions showed WT activity with the expected compounds: dopamine (A, B), metaraminol (C) and 3-hydroxyphenethylamine (D). The 3,5-dihydroxy-phenethylamine (E) has very high reactivity so conversion was observed in all samples including the pET vector control. WT did not appear to have activity with compounds E, F and G. All mutants (including F112K) had no apparent activities with any amines. There were three exceptions: new peaks were observed for WT with isoprenaline (H), F112K-K122F with B (dopamine and hexanal), and F112K-K122L with E.

These hits were re-tested to eliminate false positive data—reactions were replicated and analysed by HPLC. All three of the hits observed in the screen were not observed in the replicate and thus the original screening results appeared to be false positives. Control reactions with dopamine and hexanal were performed for mutants F112K-K122F and F112K-K122L. Although the originally observed peaks from screening were not observed, these two mutants appeared to catalyse the typical *para* Pictet-Spengler reaction.

* Compounds 3-hydroxyphenethylamine, 3,5-dihydroxyphenethylamine, 3,4-dihydroxy-phenylpropylamine and 2-bromo,5-hydroxyphenylethylamine were synthesised by Dr Eleanor Lamming.

Although no novel activities were noted, the fact that some activity is maintained upon moving a key catalytic residue is surprising and interesting. The Phe-112 side chain does point towards the catechol, so a lysine here can perhaps interact with the 3-hydroxy group. The position of Phe-112, next to Lys-122, means that both are in a similar local environment. Therefore a lysine in position 112 may have similar properties to one at 122.

6.4.3 Other Pictet-Spengler reactions

The set of lysine mutants were tested for activity with different Pictet-Spengler reactions (Fig. 6.18). The reaction mechanisms for tryptamine and histamine Pictet-Spengler reactions are similar to that with dopamine; all require the formation of an iminium intermediate before electrophilic aromatic cyclisation. The key differences are the aromatic rings and the nature of the electrophilic aromatic substitution step. The residues that are thought to be involved in the formation of the iminium and the final deprotonation step (which is conserved in all reactions) are Glu-110 and Asp-141, and remain unchanged in these mutants. The Lys-122 interacts with the catechol ring on dopamine, so it was thought that different aromatic rings may require different positions of Lysine, or perhaps an absence of this residue.

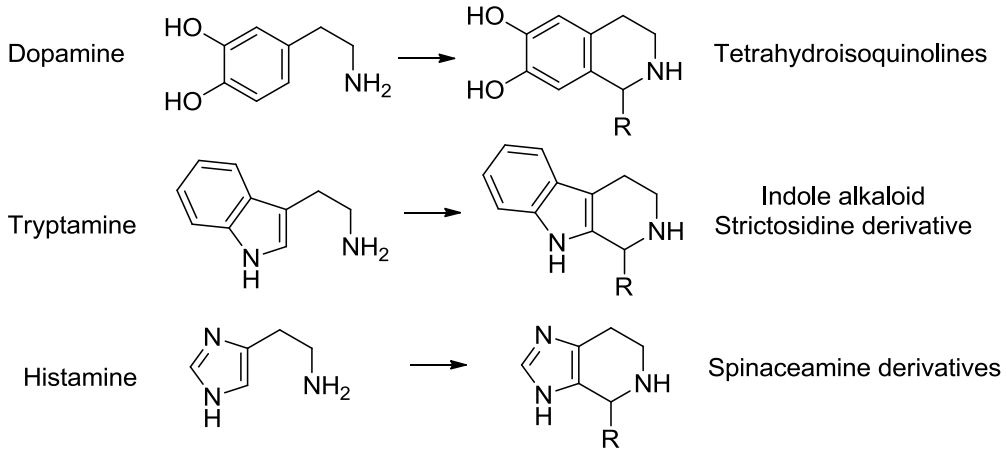


Fig. 6.18 Pictet-Spengler reactions tested and their corresponding products

To screen mutants in a rapid manner, reactions were conducted on 96-well microtiter plates coated with nickel sepharose. These types of plates are typically designed for ELISA assays, which do not require large protein loadings. Due to this, many manufacturers' plates did not show sufficient protein binding for observable enzyme catalysis. However, HIS-Select® high capacity plates from Sigma-Aldrich demonstrated sufficient loading for NCS reactions to occur (using the fluorescamine assay detection, see section 6.3.2) (Fig. 6.19). Unfortunately the reaction damaged the plates, rendering them single use only.

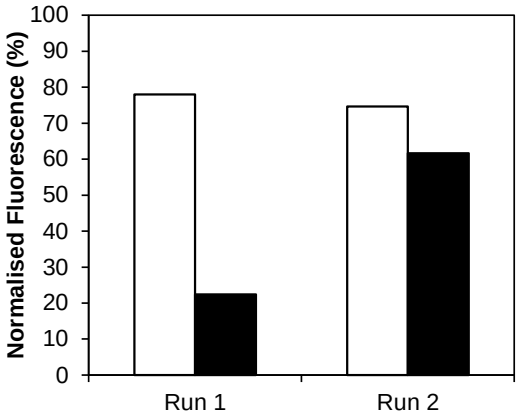


Fig. 6.19 Assay using nickel coated plates. Clear bars – pET control, Solid bars – WT. Fluorescence shows presence of amine normalised to maximum amine concentration (1 mM). Reaction: dopamine and PAA.

It was thought that the presence of 10% $v.v^{-1}$ organic solvent in the reaction mixture may damage the nickel coating of the plate. The solvent is required to solubilise the aldehyde. Different solvents were tested with the plates to determine which was most compatible (Fig. 6.20). Upon addition of the reaction mixture to the nickel coated wells, all solutions became cloudy, except for those with MeCN. This cloudiness was possibly detachment of the coating from the well wall caused by the solvents. In terms of the fluorescence response, the best solvent was MeCN. This solvent was used in future work with the nickel coated plates. Elution of protein from the nickel coated plates with varying concentrations of imidazole or EDTA was examined, as this may enable multiple reactions with a single purification. However, purification from the plates was not possible, and the best activities were observed with the enzyme immobilised on the well surface.

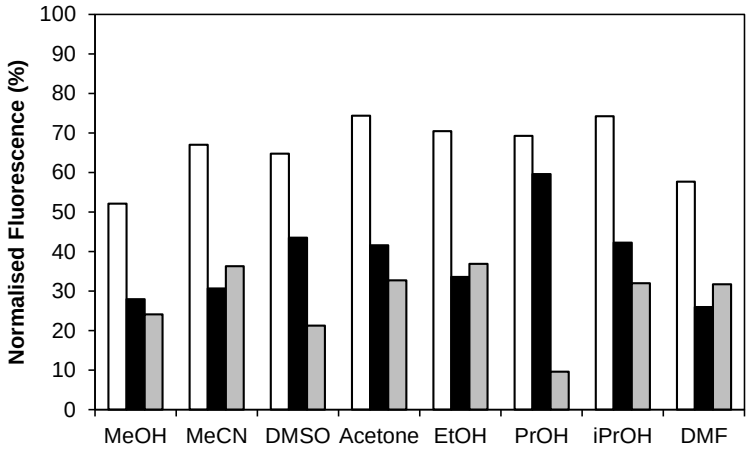


Fig. 6.20 Solvent compatibility with nickel plates. Clear bars – pET control, solid bars – WT, grey bars –difference (pET-WT). Fluorescence shows presence of amine normalised to maximum amine concentration (1 mM). All solvents except MeCN caused cloudiness/precipitation. Propanol only caused small amount of cloudiness, whilst for DMSO the precipitate settled at the bottom of the well.

The nickel plate method was used to detect activity with the lysine library and amines (Fig. 6.18). Fluorescamine was used to screen for a reduction in amine present. As expected WT and phosphate activity was detected with dopamine but no mutant demonstrated a clear loss in fluorescence due to activity. Tryptamine alone showed a very low fluorescent response, suggesting a problem with the assay for this compound. The histamine fluorescent response was similar to dopamine, but activities were not observed for any enzymes, except a slight reduction of fluorescence was observed with phosphate. It is not clear whether this background reading was due to lack of a protein signal in the phosphate control (which will cause reduction in signal regardless of activity) or genuine activity. Phosphate catalysed Pictet-Spengler reactions with histamine have not previously been explored, so it may be worth verifying this signal.

The fluorescamine assay is not especially sensitive, and does not seem appropriate for the observation of very low level promiscuous activities in enzymes. Therefore the same samples as analysed by the fluorescamine assay above were also analysed by HPLC. The chromatograms were examined for the presence of new peaks. Unfortunately the histamine samples did not seem to run well using the typical HPLC method. This is perhaps as they are too polar for the method which is optimised for THIQ analysis. The HPLC traces showed a few unknown peaks but these peaks were considered too small to warrant further examination.

Overall, the lack of positive results for the lysine mutants was not unexpected. Larger mutant libraries and more sensitive assay methods (i.e. MS) would increase the chances of finding a hit. Regardless, such extreme rational engineering is ambitious; it would require large efforts in directed evolution and a considerable dose of luck for a successful novel enzyme to be generated in this manner.

6.5 α-Substituted aldehydes

6.5.1 Identification of activity

The WT Δ29*Tf*NCS was screened with 3-hydroxybenzaldehyde (3-HBA) and cyclohexanecarboxaldehyde (CHCA), using a nickel coated plate and a fluorescamine assay. Activity with these substrates was detected. Only trace activities with WT NCS had ever been previously observed for these substrates, and therefore this result was explored further by HPLC analysis.[57]

Reactions with purified WT *Tf*NCS using dopamine with 3-HBA and CHCA were analysed by HPLC. Both dopamine consumption and new peaks in the expected region of THIQs were observed (Fig. 6.21). Phosphate catalysed reactions also formed the

same potential product peaks. The phosphate reaction was greater than the enzyme reaction for 3-HBA. The 3-HBA product peak was verified with a chemically characterised standard. For CHCA the phosphate reaction seemed to form very little product compared to CHCA, so the product peak at the time was unverified.

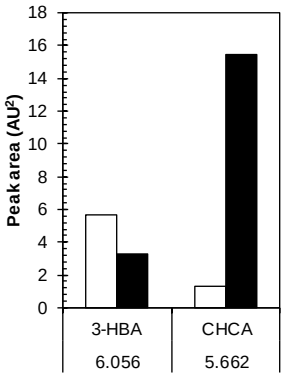


Fig. 6.21 α-Substituted aldehyde product formation. Open bars: Phosphate, solid bars: WT NCS. 3-HBA: 3-hydroxybenzaldehyde. CHCA: cyclohexanecarboxaldehyde. Conditions: 1 mM dopamine, 10 mM aldehyde. 1 mg.mL^{-1} purified WT *Tf*NCS. 3-HBA product peak verified with characterised standard. CHCA product peak unverified. Peak areas are not standardised so comparing different products directly is not possible.

6.5.2 The effect of solvent

A profound impact of solvents on the NCS reaction became apparent during lysate experiments. The aldehydes used in the Pictet-Spengler reactions are typically not very soluble in water, and so are solubilised in an organic solvent at 10% v.v^{-1} of the final reaction mixture. Reactions with dopamine, PAA and crude lysate showed that the use of DMSO as a co-solvent greatly enhances enzyme activity (Fig. 6.22). Heptane, even though it is not miscible in water, was slightly superior to MeCN as a co-solvent. Both were inferior to DMSO. Adding neat aldehyde with no co-solvent showed poorest conversions

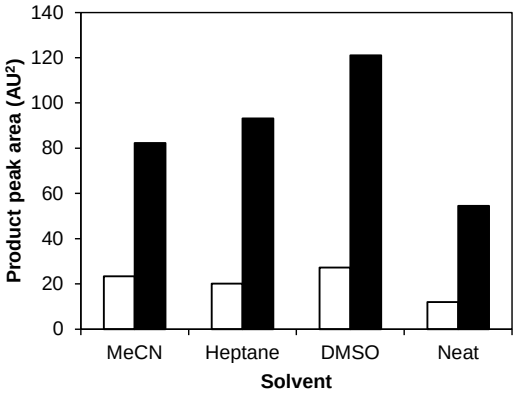


Fig. 6.22 Solvent effects with crude lysate. Open bars: pET control, solid bars: WT *Tf*NCS. Conditions: Dopamine 10 mM, PAA 15 mM, 5% v.v^{-1} cells, 15% v.v^{-1} solvent, 2.5 h incubation. Peak area of dopamine shows a similar trend.

6.5.3 Verification of activity

The aldehydes phenylacetaldehyde (PAA), benzaldehyde (BA), 3-hydroxy-benzaldehyde (3-HBA) and cyclohexanecarboxaldehyde (CHCA) were tested with purified WT *Tf*NCS, dopamine and with MeCN or DMSO as a co-solvent. Activity above background was observed with PAA, BA and CHCA (Fig. 6.23). There seemed to be slight activity above background for 3-HBA. CHCA and BA strongly favoured MeCN as a co-solvent whilst PAA showed a slight preference for DMSO. The product peaks for PAA and 3-HBA peaks were validated with verified standards. The product peak for CHCA was isolated by collecting the analytic HPLC eluent and its identity was verified by high resolution mass spectrometry ([HRMS ES+] found $[M+H]^+$ 248.1653, $C_{15}H_{22}NO_2$ requires 248.1651). The product peak for BA was unverified, but in the expected region of THIQs. It was notable that the enzyme has superior activity for CHCA than the phosphate reaction. This suggests that the enzyme provides catalytic rate enhancement in a different manner to phosphate—some substrates that are poorly turned over by phosphate may be good substrates for the enzyme.

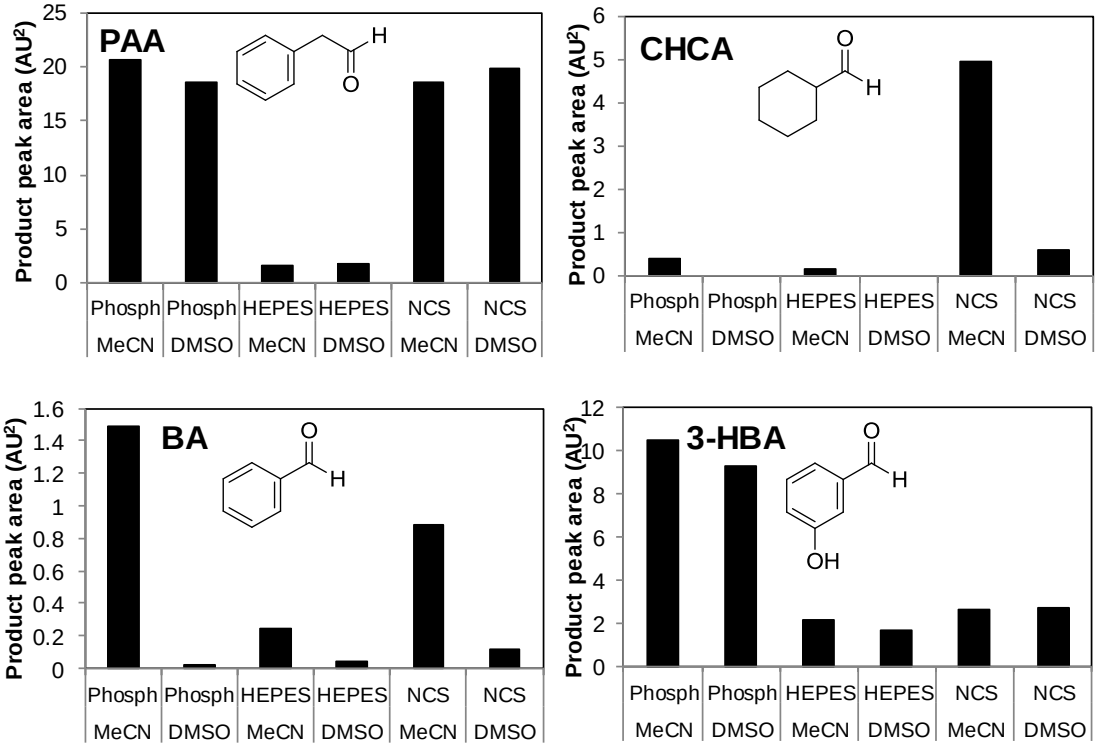


Fig. 6.23 α-Substituted aldehyde product formation. Conditions: Dopamine 1 mM, aldehydes 10 mM, 50 µg/mL purified WT *Tf*NCS, 10% $v.v^{-1}$ solvent, 3 h, 37 °C. PAA: phenylacetaldehyde, CHCA: cyclohexanecarboxaldehyde, BA: benzaldehyde, 3-HBA: 3-hydroxybenzaldehyde. MeCN/DMSO refers to solvent used. PAA and 3-HBA peaks validated with verified standards. CHCA peak was subsequently purified and analysed by mass-spectrometry. The BA product peak was unverified.

In the previous experiment (Fig. 6.23), the activity of the enzyme with 3-HBA was just above the control HEPES reaction. To verify that the small increase in peak area was

an enzyme mediated effect, the reaction was replicated with higher concentrations of NCS. The data showed the reaction was indeed enzyme catalysed (Fig. 6.24). Surprisingly, the enzyme reaction clearly favoured DMSO as a co-solvent, which was the opposite of the results observed for the very similar substrate BA.

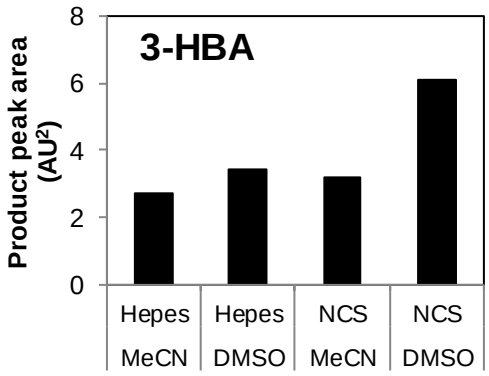


Fig. 6.24 3-Hydroxybenzaldehyde activity. Conditions: Dopamine 1 mM, ketone 10 mM, 93 µg/mL purified WT *Tf*NCS, 10% v.v^{-1} solvent, 3 h, 37 °C. MeCN/DMSO refers to solvent for aldehyde.

The BA product identity remains unverified. In the future, it will be necessary to purify larger quantities of these novel products in order to analyse them by NMR and chiral HPLC.

6.6 Ketones

6.6.1 Chemical synthesis of 1,1-disubstituted-THIQs

The 1,1-disubstituted-THIQ moiety is found in numerous natural products, and it is an attractive synthetic target for active pharmaceutical intermediates (APIs). Analogously to 1-monosubstituted-THIQs, these disubstituted moieties can be formed *via* a PS reaction between a phenethylamine and a ketone. Ketones are generally less reactive than aldehydes and consequently whilst there are many methods for the synthesis of 1-monosubstituted-THIQs, general methods for the disubstituted compounds are more unusual. For example, the facile phosphate catalysed formation of a THIQ from dopamine and an aldehydes does not extend to ketone starting materials.[59]

There are a number of methods for the synthesis of 1,1-disubstituted THIQs (Fig. 6.25). Some methods rely on using specific ketones, which limit the reaction scope.[192,193] Other more general methods require unique catalysts: including a zeolite catalyst,[194] titanium catalysts[195,196] and a calcium catalyst.[93] There are also few methods for formation of 1,1-substituted tetrahydro-β-carbolines *via* PS reactions with ketones. These include the use of microwaves[197] or unusual solvents such as 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP)[198] to increase the slow rate of the reaction. There are

currently no general methods for the construction of chiral 1,1-disubstituted THIQs.[156] Some chiral applications are possible with a lithiation-substitution method that turns the THIQ 1-position into a nucleophile (Fig. 6.25G).[199] Other methods involve chiral phase transfer catalysis, in which chiral ligands stereoselectivity catalyse alkylation or hydrolysis in biphasic systems (Fig. 6.25H).[200,201]

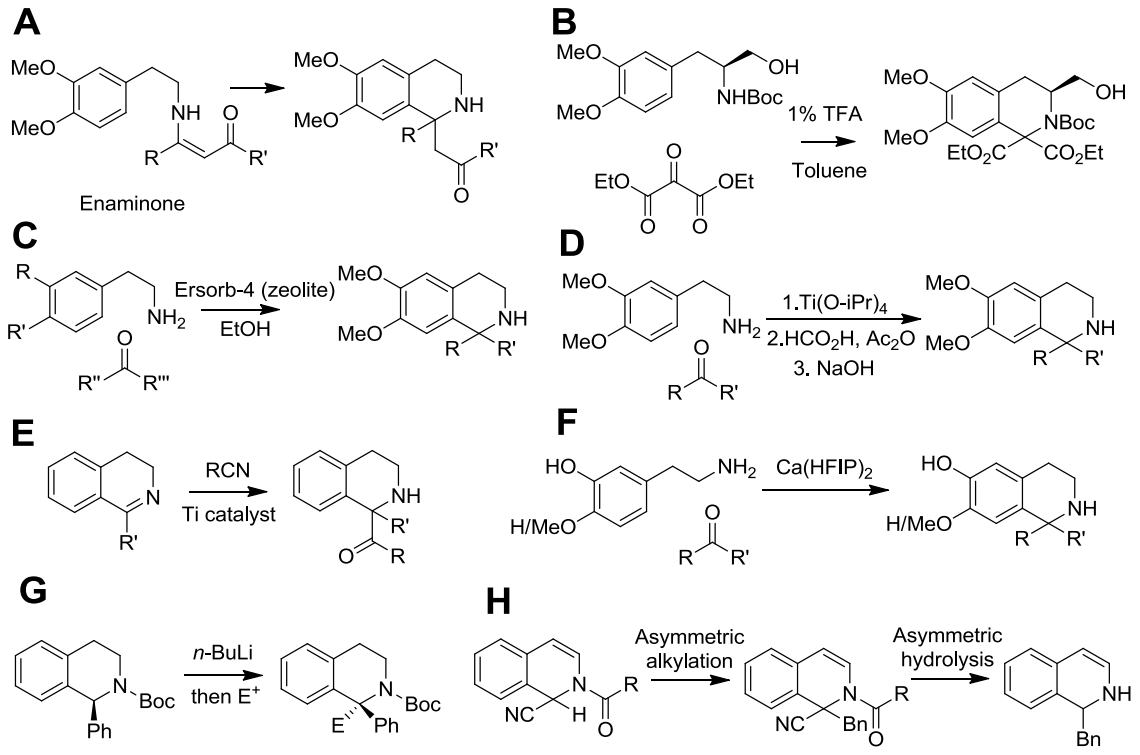


Fig. 6.25 Chemical syntheses of 1,1-disubstituted THIQs. A. Synthesis via enaminones[192]. **B.** Synthesis with diethyl oxomalonate[193]. **C.** Zeolite catalysis[194]. **D.** Titanium (IV) catalysis[195]. **E.** 'Umpolung' synthesis with titanium (III) catalysis[196]. **F.** Calcium catalysis[93]. **G.** Lithiation-substitution reaction, using chiral starting materials[199]. **H.** Asymmetric alkylation/hydrolysis of Reissert compounds forming chiral products[201].

Through reactions with ketones, NCS may be able to produce 1,1-disubstituted THIQs. Based on its catalysis with aldehydes, it is possible that these reactions could occur in a chiral manner. Verified activity with ketones would make NCS a very valuable and unique catalyst, able to produce chiral 1,1-disubstituted THIQs.

6.6.2 Biosynthesis of 1,1-disubstituted-THIQs

There are BIA subgroups which contain the 1,1-disubstituted-THIQ structure. The spiro-BIAs[202] are formed *in planta* from the protoberberines,[203] and in turn are converted into the azepine BIA subgroup (Fig. 6.26A). The tetracyclic erythrina alkaloids[204] are biosynthesised from the BIA (*S*)-norreticuline, through a rearrangement of the carbon skeleton *via* a 9-membered ring intermediate (Fig. 6.26C).[205]

The complex biosynthesis of these compounds derives ultimately from the 1-monosubstituted-THIQ (*S*)-norcoclaurine, with the PS step between dopamine and 4-HPAA catalysed by NCS and extensive elaboration required to reach the 1,1-disubstituted molecular skeletons. It may, however, be possible to access similar 1,1-disubstituted compounds directly *via* a PS reaction with a ketone. Direct access to the spiro-BIAs from a PS reaction (perhaps catalysed by NCS) would require conversion of ketones similar to 2-indanone (Fig. 6.26B). Tetracyclic erythrina alkaloids may be accessible through reaction with a cyclohexanone derivative (Fig. 6.26D). Formation of the final 5-membered ring can be achieved by cyclisation onto the nitrogen, which will require a 2-substitutent on the cyclohexanone.

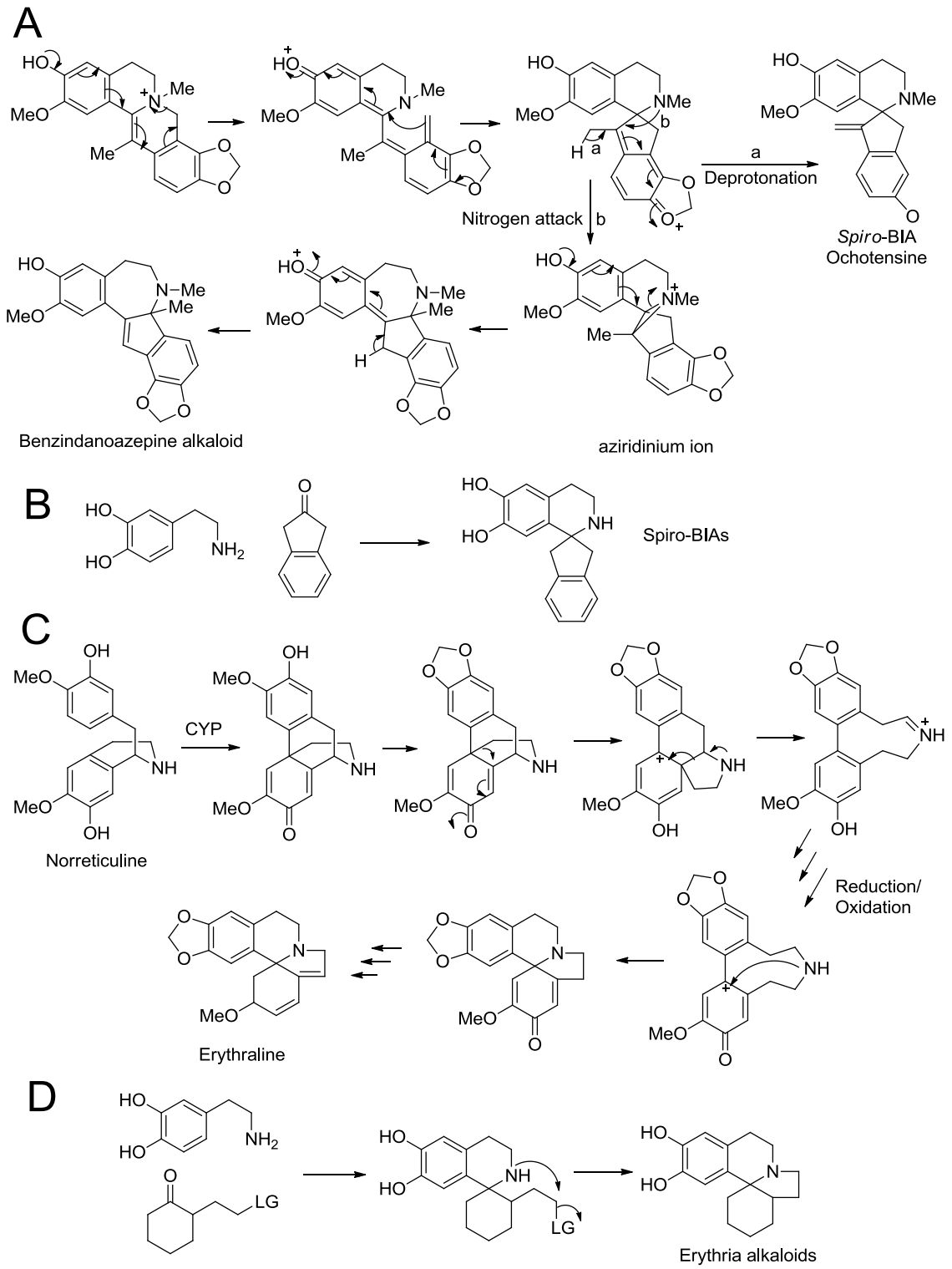


Fig. 6.26 Biosynthetic pathways to BIA subgroups and metabolic short-cuts *via* NCS. A: proposed biosynthetic route to spiro-BIAs and benzindanoazepine alkaloids from protoberberines.[203] **B**: Pictet-Spengler shortcut to spiro-BIAs. **C**: Biosynthetic route to erythrina alkaloids *via* norreticuline and a morphinan-like CYP step and 9 membered ring intermediate.[205] **D:** Pictet-Spengler shortcut to erythrina alkaloids. LG = leaving group.

6.6.3 Identification of activity

Wild-type Δ29*Tf*NCS was tested with 4-hydroxyphenylacetone (4-HPA), the ketone most similar to the natural NCS substrate 4-hydroxyphenyl-acetaldehyde. A fluorescamine assay seemed to show some activity with 4-HPA. Analysis of the reactions by HPLC showed the presence of a new peak with a retention time similar to other THIQs.

Purified WT *Tf*NCS was tested for activity with dopamine and 4-HPA and 2-hexanone. The effect of co-solvent on these reactions was also examined. The predicted 4-HPA product peak observed in the initial screen was observed again in this experiment, with DMSO greatly enhancing the amount produced (Fig. 6.27). This product was not observed in the phosphate reaction. For 2-hexanone, a very small peak was observed in the THIQ region. This peak was larger when MeCN was used as a co-solvent, and there was some evidence of this peak in the phosphate reaction. The reactions were repeated with double the concentration of enzyme, and an identical trend was observed. This suggested these new products formed were not merely side-products or contamination but indicative of a real enzymatic effect.

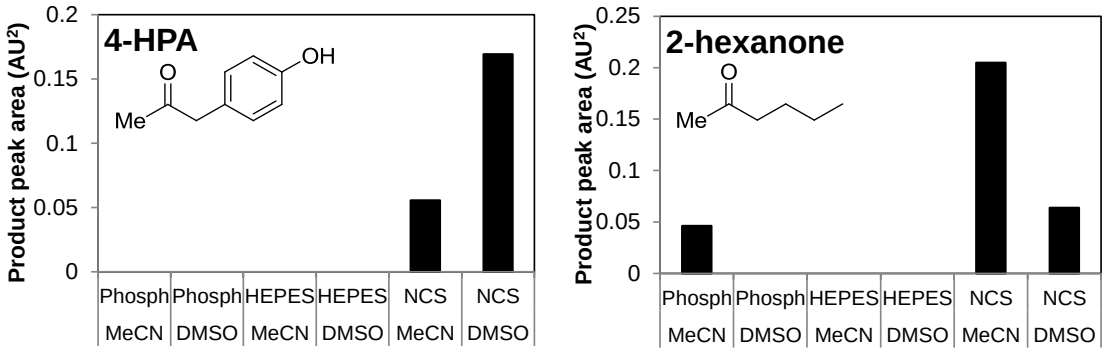


Fig. 6.27 Ketone product formation. Conditions: Dopamine 1 mM, ketones 10 mM, 50 µg/mL purified WT *Tf*NCS, 10% v.v^{-1} solvent, 3 h, 37 °C. 4-HPA: 4-hydroxyphenylacetone. MeCN/DMSO refers to solvent for ketone solubilisation.

6.6.4 Verification of activity

The very low or zero conversions observed for the phosphate reaction with ketone substrates suggested that clarified cell lysate could be used in place of purified enzymes without concern of a high background reaction. Accordingly, ketone reactions (4-HPA and 2-hexanone) were conducted with 50% v.v^{-1} clarified cell lysate loading, simplifying the reaction preparation. The choices of co-solvent, enzyme and substrate proportion were also investigated. For both 4-HPA and 2-hexanone, conversions appeared greatest with DMSO rather than MeCN, with *Tf*NCS rather than *Cj*NCS2, and with dopamine in excess rather than the ketone (Fig. 6.28). The enzyme *Cj*NCS2

showed no activity for ketones. The use of DMSO rather than MeCN increases conversion of 4-HPA by over 4-fold, and for 2-hexanone around 3-fold. In contrast, for both ketones, the background reaction with phosphate or pET seemed to be larger with MeCN than DMSO. The preference of the 2-hexanone conversion for DMSO was not expected based on previous reactions with pure enzyme, where MeCN seemed preferable (Fig. 6.27). The improved conversion observed with excess dopamine has been shown previously,[60] though the 5% v.v^{-1} less solvent present in these reactions with dopamine in excess may also have an effect on conversions.

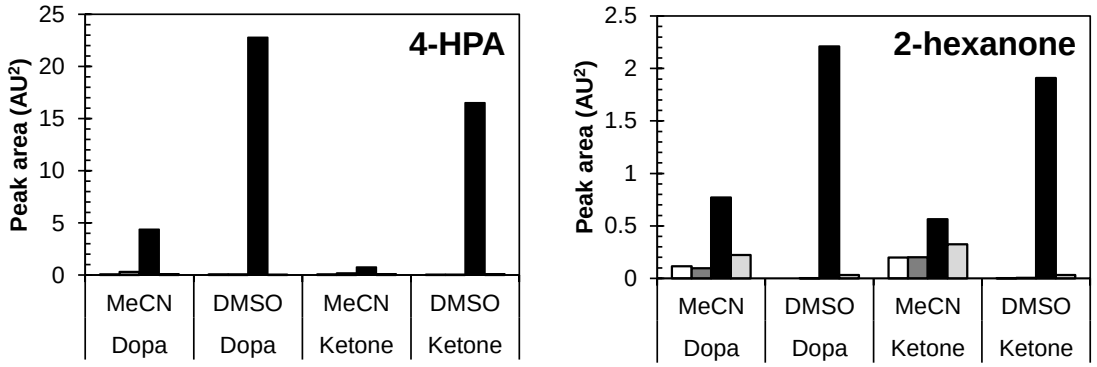


Fig. 6.28 Ketones and lysate. Conditions: Dopamine 10 or 15 mM, ketones 10 or 15 mM, 50% v.v^{-1} lysate, 10-15% v.v^{-1} solvent, 3 h, 37 °C. 4-HPA: 4-hydroxyphenylacetone. MeCN/DMSO refers to solvent for ketone. Dopa/Ketone refers to which substrate in excess. Clear bars, phosphate; dark grey bars, pET; black bars, *Tf*NCS; light grey bars, *Cj*NCS2.

The optimum reaction conditions were explored further. In a single experiment, the effects of DMSO, enzyme preparation and ascorbate were investigated (Fig. 6.29). Ascorbate is a common additive to NCS reactions—here the addition of ascorbate prevented the reactions turning a brown/black colour, and increased the product yield.[58,60] Dopamine is prone to auto-oxidation and polymerisation; ascorbate is an anti-oxidant and prevents this side reaction from occurring.[206] Without ascorbate no conversion at all was observed with pure enzyme. Generally, pure enzyme was observed to be an inferior catalyst compared to cell lysate. Although the conditions were not controlled with respect to absolute enzyme quantity in lysate, the quantity of pure enzyme used here was typical for a biotransformation (80 μg.mL^{-1}). The addition of DMSO did not serve to improve yields when using pure enzyme, but did when lysate was used. For lysate, the optimum DMSO concentration tested was 10% v.v^{-1}, with yields falling to zero as concentration was increased to 50% v.v^{-1}.

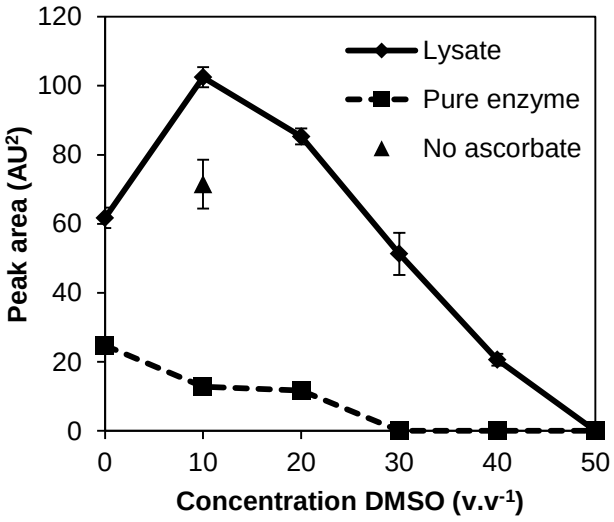


Fig. 6.29 Reaction conditions for conversions of ketones. Conditions: 20% v.v^{-1} Δ29*Tf*NCS lysate or 80 μg.mL^{-1} pure Δ29*Tf*NCS, 5 mM ascorbate (except in 'no ascorbate' reaction), 10 mM 2-hexanone, 15 mM dopamine. All data points are averages of triplicates; error bars show one standard deviation from the mean. No activity was observed for controls lacking substrates, an empty vector lysate control and for pure enzyme reaction without ascorbate present. The peak area used was the predicted product peak area.

To verify the reaction products and HPLC signals, high resolution mass spectrometry (HRMS) analysis was conducted. The predicted THIQ product peaks for dopamine with 4-HPA and dopamine with 2-hexanone were isolated by collecting the analytical HPLC eluent. These purified peaks were then analysed by high-resolution MS. The 4-HPA product was verified as the expected product ([HRMS ES+] found $[M+H]^+$ 286.1457, $C_{17}H_{20}NO_3$ requires 286.1443). The 2-hexanone product HRMS data was not satisfactory because only very low concentrations of the product were available. However, the m/z value strongly indicated the product was the expected product ([HRMS ES+] found $[M+H]^+$ 236.1245, $C_{14}H_{22}NO_2$ requires 236.1651). Overall, this MS data verified that *Tf*NCS is capable of forming 1,1-disubstituted-THIQs *via* a Pictet-Spengler reaction with dopamine and a ketone.

6.6.5 Mutant screen

The range of ketone substrates that could be accepted by NCS was probed (Fig. 6.30). All available NCS mutants were screened for activity with dopamine and four different ketones. 4-HPA and 2-hexanone, with which activity had already been verified, were screened. The α-ketoacid α-ketoglutarate (α-KG) was examined—NCS activity had been observed for the α-ketoacid 3-hydroxyphenylpyruvic acid[25,46] and 1-carboxylic acid THIQs were historically proposed as key intermediates in BIA synthesis.[207,208] Cyclohexanone (CH) was investigated as representative cyclic ketone. The screen was conducted using clarified cell lysate, 10% v.v^{-1} DMSO, excess dopamine and analysis by HPLC.

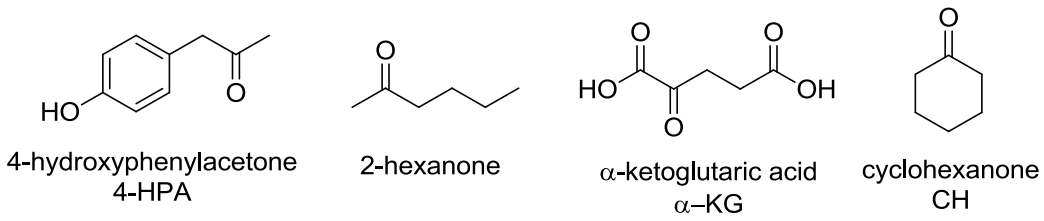


Fig. 6.30 Ketones screened with all NCS mutants.

Conversions (with verified HPLC peaks) were observed with both 4-HPA and 2-hexanone, with the greatest turnover demonstrated by mutants A79I and M97F respectively (Fig. 6.31). Activity was observed with α-KG, again with A79I and M97F showing the highest conversions. Conversions with CH appeared to be very high—analysis of dopamine concentrations suggested conversions were greater than 80% (for WT *Tf*NCS). It is important to note that the product peak for α-KG has to date not been verified by MS or NMR. The identity of the CH product has since been verified by MS (see section 6.6.9 below) but its retention time has not yet been confirmed. These are priorities for future work.

The behaviour of different mutants in these ketone substrate screens provides some interesting information regarding NCS. It is clear than *Cj*NCS2 cannot turnover ketones; this is consistent with previous substrate screens.[57] The mutant F112L (not previously examined) had 10-20% activity of WT for all substrates. This showed Phe-112 is important for activity—probably for controlling the position and pK_a of Lys-122—but it is not necessary for the enzyme reaction to occur. Improvement of activities with the mutation A79I has been observed previously and seemed to impact a variety of substrates (section 4.3.8). The reduction in activities with the L76A mutation is perhaps greater than would be expected based on previous work (section 4.3.7). The Met-97 mutants have not previously been tested; here for 4-HPA, 2-hexanone and α-KG they show the same trend, with activities increasing from M97V to L to F, with M97F having greater activities than WT. For CH this trend was not observed and M97L activity was equal to WT, with M97F and M97V showing lower conversions. The CH conversions have relatively high background reactions, so activities observed for mutants that are typically inactive (i.e. the K122 mutants), may not be a real enzymatic effect.

Mutants that appeared to demonstrate good conversions with ketones were then examined with a greater range of substrates.

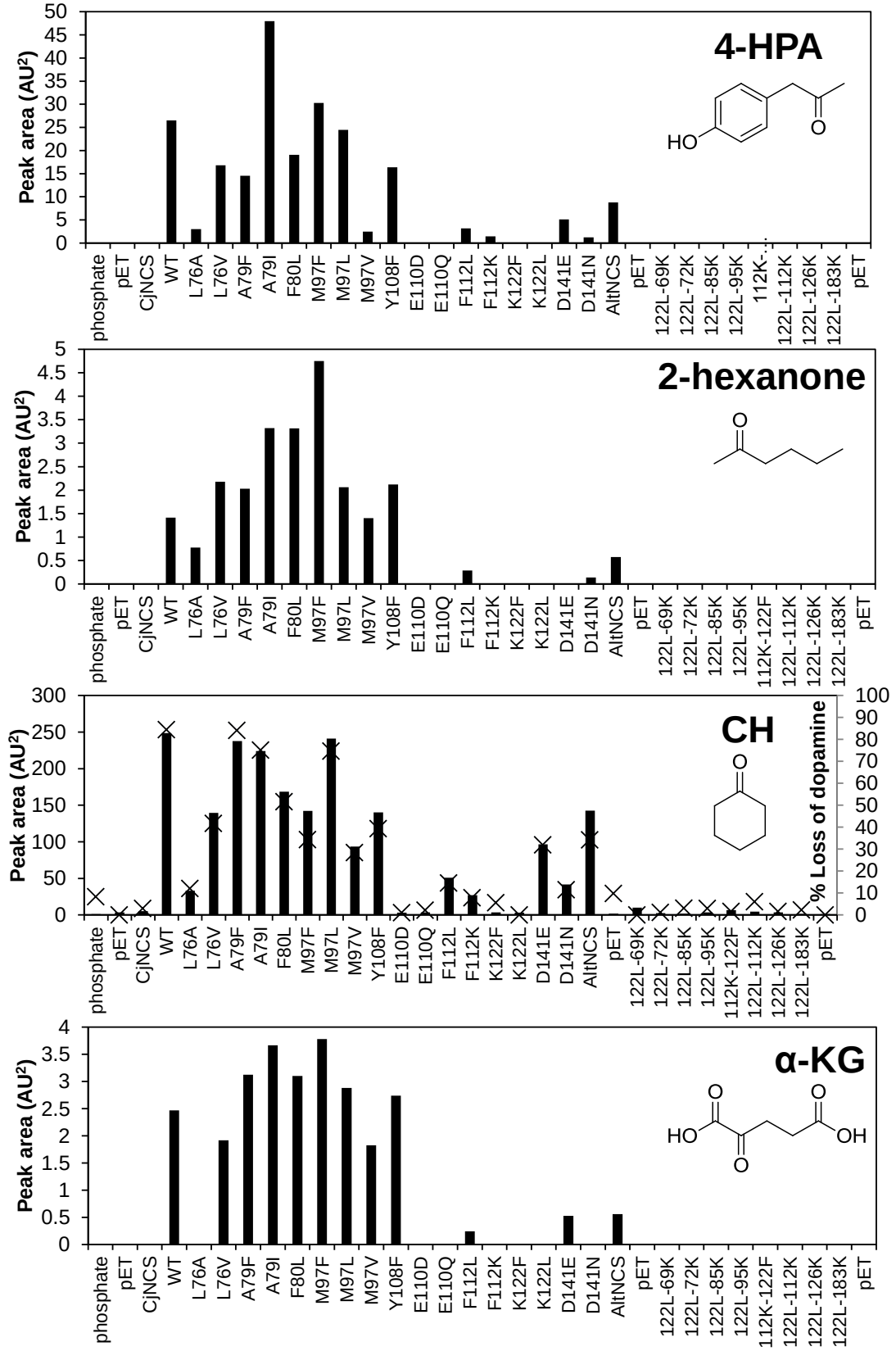


Fig. 6.31 Mutant ketone screen. Conditions: Dopamine 15 mM, ketones 10 mM, 45% v.v^{-1} lysate, 10% v.v^{-1} DMSO, 3 h, 37 °C. 4-HPA: 4-hydroxyphenylacetone, CH: cyclohexanone, α-KG: α-ketoglutarate. Solid bars are product peak areas. For 4-HPA and 2-hexanone peaks were verified by mass spec. For CH and α-KG peaks are not verified. For CH plot, crosses show consumption of dopamine. *Alt*NCS is WT Δ29*Tf*NCS without C-terminus His-tag used in crystallisation trials (pQR1855).

6.6.6 Alkyl ketones

NCS activity with the alkyl ketone 2-hexanone had been partially verified by MS analysis. The screen showed that a number of mutants had greater activity for 2-hexanone than WT, in particular A79I, F80L and M97F. A variety of alkyl ketones with different chain lengths were then investigated (Fig. 6.32).

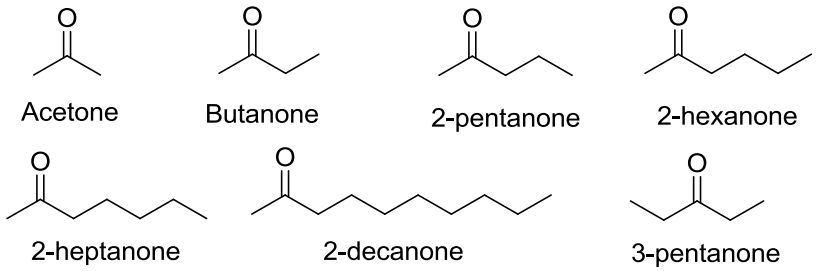


Fig. 6.32 Alkyl ketones examined.

Alkyl methyl ketones of C_3, C_4, C_5, C_6, C_7 and C_{10} length were examined. Activity was observed for all of these compounds except for acetone, the most volatile (Fig. 6.33). The mutant A79F showed highest conversions at short chain lengths, and this seemed to decrease as chain length increased. The mutants A79I and F80L showed the opposite trend: increased conversions as chain length increased (up to C_7, 2-heptanone). These trends seem to correspond with increasing or decreasing bulk at the active site entrance to compensate for the substrate chain length. When the ketone becomes 10 carbons long activity is reduced for all mutants, probably as the solubility of the ketone becomes very poor. It is unclear with the standard HPLC methods whether activity is present with acetone. No activity was observed with the symmetric 3-pentanone, probably because the ketone functional group becomes more difficult to access due to increased bulk of the ethyl group rather than methyl substituent.

In the previous screen the M97F mutant appeared to turnover 2-hexanone with the greatest efficiency. This result was not replicated in this screen; here M97F generally appeared to have poorer conversions than WT. The differences between these two screens may be due to the slightly different conditions: the concentration of lysate is halved and reaction times are extended. There may also be differences in expression levels of mutants between the two experiments.

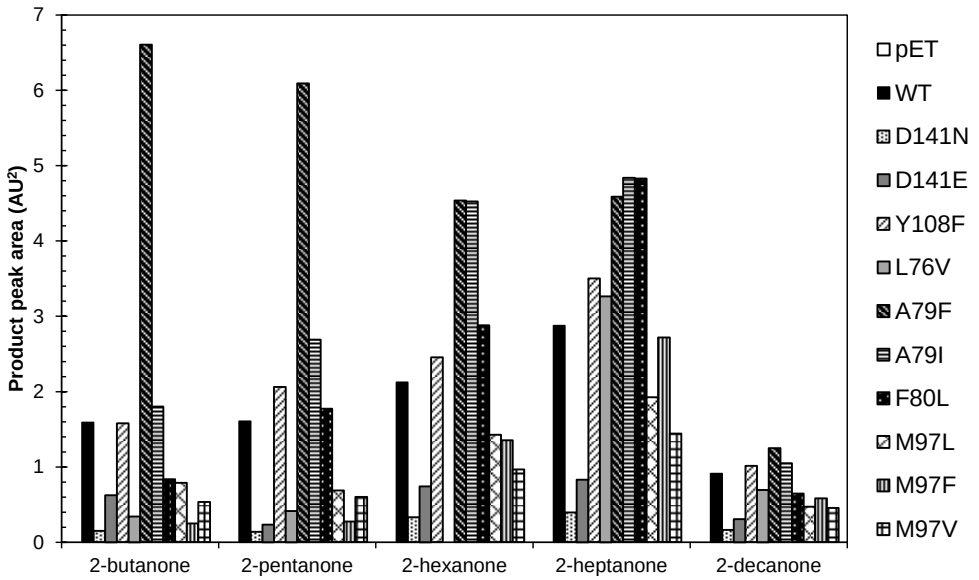


Fig. 6.33 Alkyl ketones. Conditions: 20% $v.v^{-1}$ lysate, 37 ˚C, 6 hours, 15mM dopa,10 mM ketone, 10% $v.v^{-1}$ DMSO. All product peaks are unverified, except for 2-hexanone. No data present for L76V with 2-heptanone. These products are expected to have similar absorbance coefficients so peak areas are comparable.

6.6.7 Bulky ketones

A variety of ketones with bulky cyclic R-groups were examined (Fig. 6.34). These included the phenyl methyl ketones phenylacetone (PA) and 4-hydroxyphenylacetone (4-HPA), the cyclo methyl ketones cyclohexylmethylketone (CHMK) and adamantylmethylketone, and the aromatic ketones acetophenone and propiophenone. No activities were observed with the conjugated ketones acetophone or propiophenone. For either substrates this is probably because of the increased stability of the ketone through conjugation, and for propiophenone the increased bulk of the ethyl adds another steric challenge for the enzyme. The adamantylmethylketone was not turned over by any enzyme, probably due to its extreme bulk, resultant steric clashes and inaccessibility of the ketone.

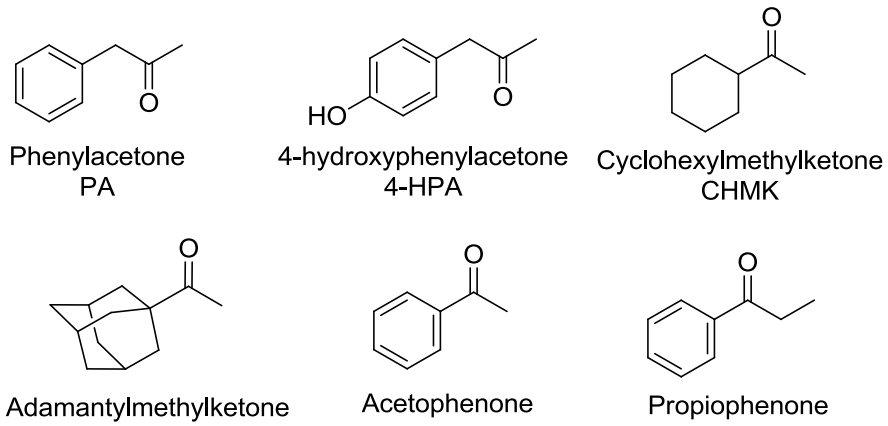


Fig. 6.34 Bulky ketones examined.

Activity was previously observed and verified with 4-HPA and so it was not surprising that phenylacetone (PA), a very similar compound, is also turned over (Fig. 6.35A). Conversions were observed but were low with cyclohexylmethylketone (CHMK). The seemingly lower conversions were probably due to accessibility difficulties with the α-substitution present (Fig. 6.35B).

Compounds 4-HPA and PA showed highest conversions with A79I, around double the WT values. Variants A79I and A79F had slightly increased conversions for CHMK. Potential explantions for the improved activity with Ala-79 mutants can be found in section 4.4.3. Briefly, increased bulk in this position may prevent the intermediates leaving the active site and may also encourage the reaction intermediate to adopt conformations condusive to reaction progression.

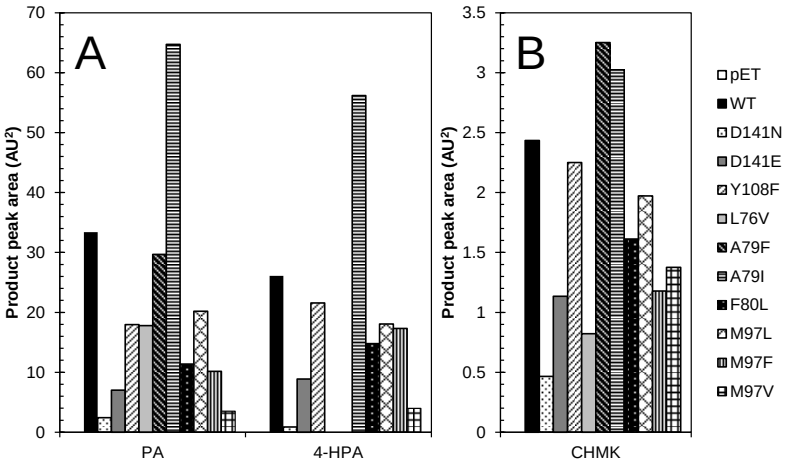


Fig. 6.35 A. Phenyl ketones. B. Cyclohexylmethylketone. Conditions: 20% $v.v^{-1}$ lysate, 37 ˚C, 6 hours, 15mM dopa,10 mM ketone, 10% $v.v^{-1}$ DMSO. PA: phenylacetone, 4-HPA: 4-hydroxyphenylacetone, CHMK: cyclohexylmethylketone. All product peaks are unverified, except for 4-HPA. No data present for L76V and A79F with 4-HPA. Products in A are expected to have fairly similar absorbance coefficients so peak areas may be comparable.

6.6.8 Cyclohexanones

The high conversions observed for NCS with cyclohexanone (CH) was unexpected and rather remarkable considering that the cyclic ketone is structurally very different to the natural NCS 4-HPAA substrate. A variety of different substituted cyclohexanones and related compounds were investigated with a selection of mutants (Fig. 6.36).

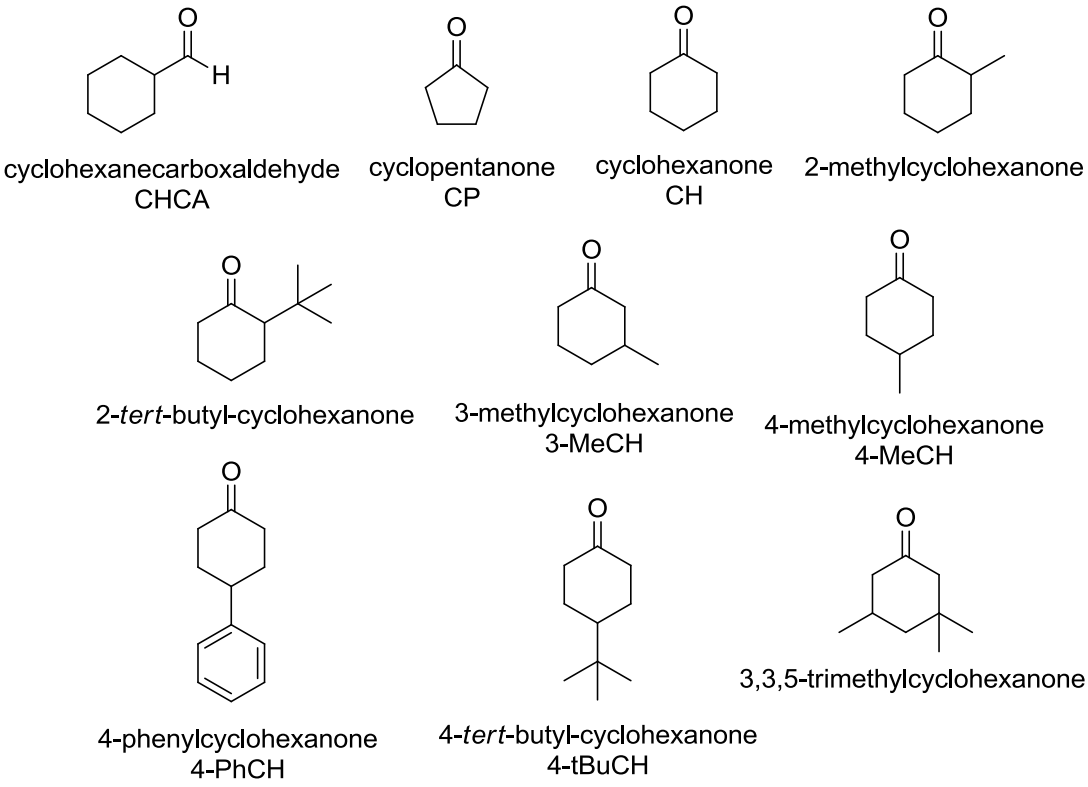


Fig. 6.36 Cyclic ketones examined.

The conversions with different cyclohexanone derivatives seemed to be influenced by the steric bulk of the substrate. Substituents in the 2-position were not accepted (e.g. *tert*-butyl or methyl), probably due to steric clashes. There was possibly a very small product peak with 2-methylcyclohexanone and A79I, but a larger peak would need to have been observed to give this observation any confidence. Substituents in the 3- and 4- positions of cyclohexanone were generally accepted (Fig. 6.37), except when numerous substitutions were present (e.g. the trisubsituted 3,3,5-trimethylcyclohexanone). The size of substituents in the 4 position seemed to slightly affect conversions, with the smaller methyl group substituent showing larger product peaks than the phenyl and *tert*-butyl groups. The 4-PhCH and 4-tBuCH substrates showed very similar conversions, but the 4-phenyl showed generally slightly larger peaks (though this could have been due to differences in absorbance).

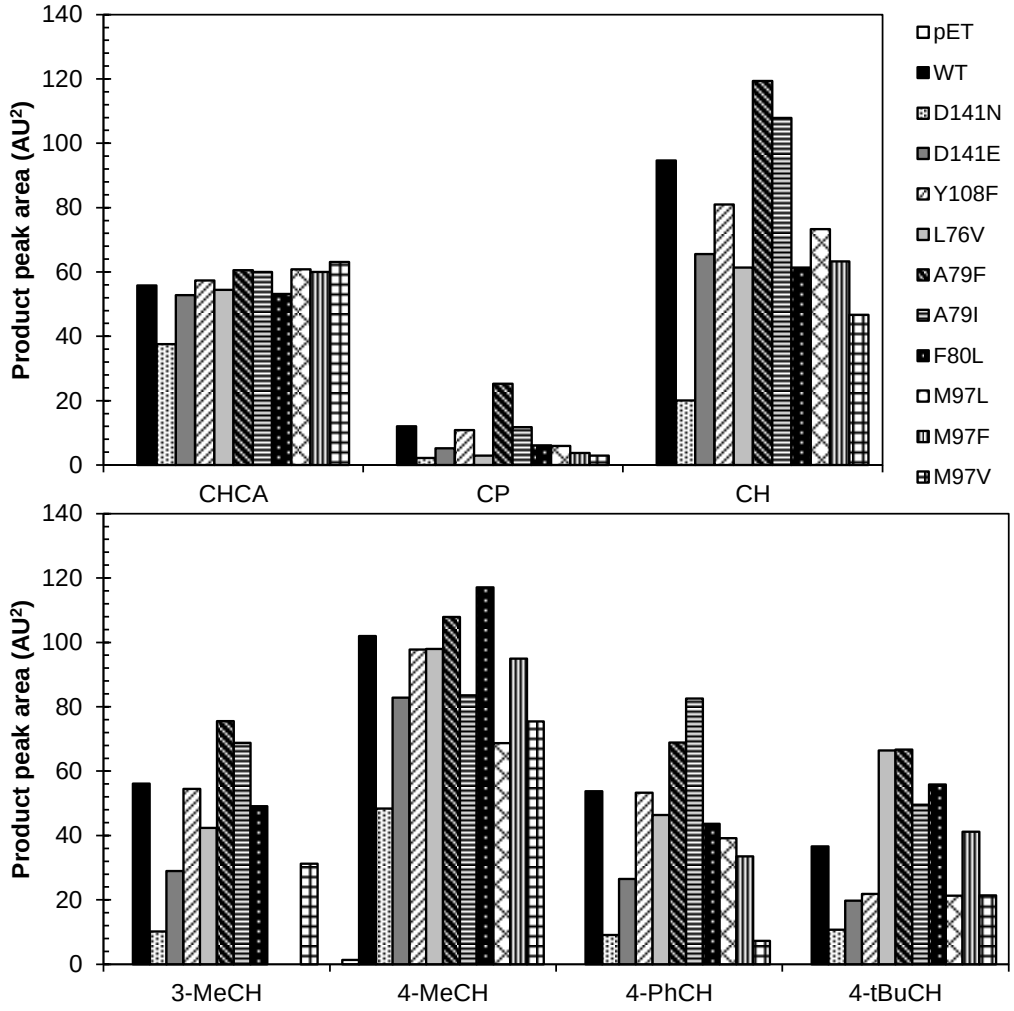


Fig. 6.37 Cyclohexanones. Conditions: 20% $v.v^{-1}$ lysate, 37 ˚C, 6 hours, 15mM dopa, 10 mM ketone, 10% $v.v^{-1}$ DMSO. CHCA: cyclohexanecarboxaldehyde, CP: cyclopropanone, CH: cyclohexanone, 3-MeCH: 3-methylcyclohexanone, 4-MeCH: 4-methylcyclohexanone, 4-PhCH: 4-phenylcyclohexanone, 4-tBuCH: 4-*tert*-butylcyclohexananonehydroxyphenylacetone. All product peaks are unverified, except for CHCA. No data present for M97F and M97L for 3-MeCH. Products are expected to have fairly similar absorbance coefficients so peak areas may be comparable.

Cyclopropanone seemed to show considerably lower conversions than cyclohexanone. This was probably a general non-enzymatic feature of cyclic ketone reactivity. Attack of the ketone by a nucleophile and change of hybridisation from sp^2 to sp^3 will result in a change of the torsional strain in the ring. For the six membered ring, this strain change is favourable: in cyclohexanone the carbonyl is partially eclipsed with neighbouring hydrogens; after attack, all bonds can be in staggered arrangements. This is not the case for cyclopentanone, where the change in strain is unfavourable as bonds become more eclipsed upon attack (Fig. 6.38).

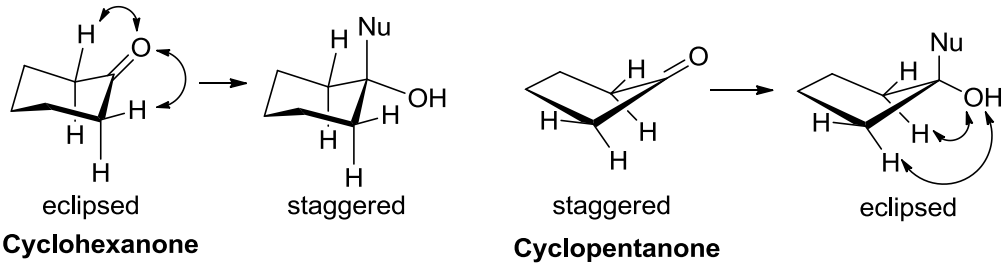


Fig. 6.38 Reactivity of cyclic ketones.

The effects of mutations on the conversion of cyclohexanone derivatives were fairly general: A79F and A79I tended to show increased conversion compared to WT. Deviations from this were observed for 4-tBuCH, where L76V showed equally high conversions as A79F. For 4-MeCH, F80L seemed to be the best performing mutant. The origin of these improvements was probably shape complementarity between the substrate and active site binding region.

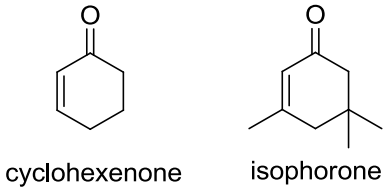


Fig. 6.39 Enones examined

The cyclic enones cyclohexenone and isophorone were investigated and neither appeared to be turned over by NCS. There were some small extra peaks present with 2-cyclohexen-1-one with the mutants A79F and A79I but these were too small to indicate conversion with any confidence. The lack of activity with these compounds was likely to be due to their reduced reactivity caused by the conjugated system present.

6.6.9 Biotransformations

Larger scale biotransformations were conducted for selected substituted cyclohexanones and Δ29*Tf*NCS mutants. The aim was to purify and fully characterise the spiro-products. Based on the screening data, substituted cyclohexanones and mutants selected were: CH (A79F), 4-tBuCH (A79F), 4-PhCH (A79I), 4-MeCH (A79F), CP (A79F), 4-HPA (A79I) and 3-MeCH (A79F). The biotransformations reactions were performed in a volume of 5 mL, with 50 mM HEPES pH 7.5, 20% $v.v^{-1}$ clarified, filtered lysate, 5 mM ascorbate, 10 mM ketone, 15 mM dopamine and 10% $v.v^{-1}$ DMSO. The reactions were incubated for 6 hours at 37 ˚C. At this point the samples were frozen until further processing.

Previously, preparative-HPLC was used to purify biotransformation products. However, preparative-HPLC is not especially convenient for large scale reactions, and alternative purification methods were explored. Recent work described that at around pH 7 it may be possible to extract THIQ products into ethyl acetate whilst leaving excess dopamine in the aqueous phase.[60] This extraction method was attempted with the cyclohexanone reaction. The THIQ product successfully extracted into the ethyl acetate, and all dopamine remained in the aqueous layer. A number of brine washes were required to remove protein and other cell debris from the organic layer. However, upon removal of the solvent, the product was not sufficiently pure—the yield was calculated to be over 100%, and proton NMR verified impurities were present. High resolution mass spec analysis verified that the expected product was present ([HRMS ES+] found $[M+H]^+$ 234.1497. $C_{14}H_{20}NO_2$ requires 234.1494). In subsequent attempts to purify THIQs in this manner, significant amounts of the ketone starting material were found to co-purify with the THIQs product. When extracting into ethyl acetate, any remaining ketone (or aldehyde) will move into the organic solvent along with the product. Thus, this method is only appropriate when 100% conversion is known to have occurred.

A second extraction step was added to the procedure to remove ketones. After removal of the ethyl acetate, the material was resuspended in a 1:1 mixture of dichloromethane and 0.1 M HCl. The acid conditions protonated the THIQ. The charged THIQ partitioned with the aqueous phase whilst the neutral ketone remained in the DCM. After washing the DCM with more dilute HCl, the aqueous layers were combined and water was removed, initially by rotary evaporation and the last few millilitres by freeze drying (this prevented overheating of the product). The compounds purified by this method were obtained as hydrochloride salts and were pure (as analysed by NMR). The purified yields were reasonable, ranging from 33-60% for four different compounds. The range was perhaps related to slight variations in the extraction procedures, or the pK_as of the products. The conversion yields for these reactions are presently unknown.

The establishment of this purification method is significant as it should be general to most NCS reactions: typically a primary amine combines with a neutral ketone/aldehyde to form a secondary amine. It will allow the scale up of reactions and remove the requirement for preparative-HPLC purification. There is also scope for improving the method further, for example more concentrated acid can be used to ensure full protonation of the THIQs. Also, it is probably possible to acidify the ethyl acetate directly, protonating the THIQ at this earlier stage. This would forgo an evaporation step. It may be apposite to add a third final extraction step: basifying the

aqueous layer and neutralising the THIQ, moving it back into a low-boiling point organic solvent. Removing organic solvent by evaporation to yield the product would be quicker and more energy efficient than removing water. It is worth noting that these acid-base extraction procedures are classic methods for extracting and purifying alkaloids from natural sources.

6.6.10 Spiro-THIQs

Four novel cyclohexane-spiro-THIQ products have been purified using the extraction method. They have been characterised by ^{1}H and ^{13}C NMR, and high-resolution mass spectrometry has also confirmed their elemental identity (see methods section 2.6.3 and appendix B, section 9.2). Typically, five characterisation methods are required for novel compounds, therefore melting points, infra-red absorbance spectra and/or retention times/R_f-values will be measured for these compounds to complete their characterisation.

Three of the four compounds purified by this method are symmetric and non-chiral. They each have a substituent on the 4-cyclohexane position (Fig. 6.40).

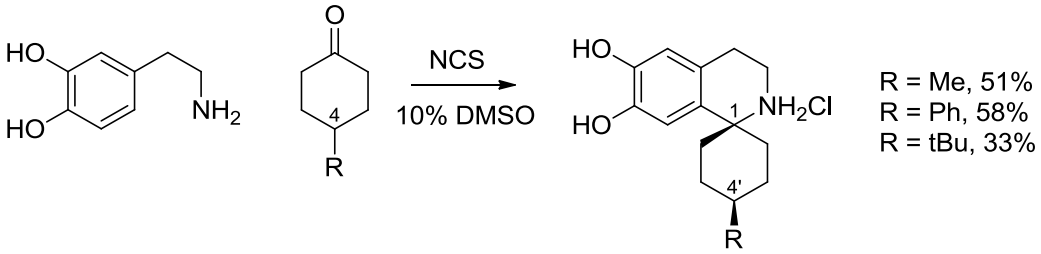


Fig. 6.40 Synthesis of 4'-substituted-spiro-cyclohexane-1,1-THIQs. Reaction conditions: 50 mM HEPES pH 7.5, 20% $v.v^{-1}$ filtered lysate, 5 mM ascorbate, 10 mM ketone, 15 mM dopamine and 10% $v.v^{-1}$ DMSO. The reactions were incubated for 6 hours at 37 ˚C. Purification was achieved by organic then acid extraction. See methods section 2.6.3 for more details.

Though these compounds are not chiral, the 4' position on these compounds is a stereocentre. This stereoisomerism arises due to the configuration of the 4' substituent relative to the THIQ moiety. Although the reaction could theoretically provide either stereoisomer, the outcome has a significant steric consequence. When depicted in three dimensions it is possible to see that, assuming the cyclohexane is in a chair conformation, the 4' substituent can either be in an axial or equatorial position. Theoretically the spiro-fused ring system can ring-flip, changing the axial/equatorial distribution, but one conformation seems much more favourable (Fig. 6.41). The 4' substituent is expected to favour an equatorial position due to steric effects.

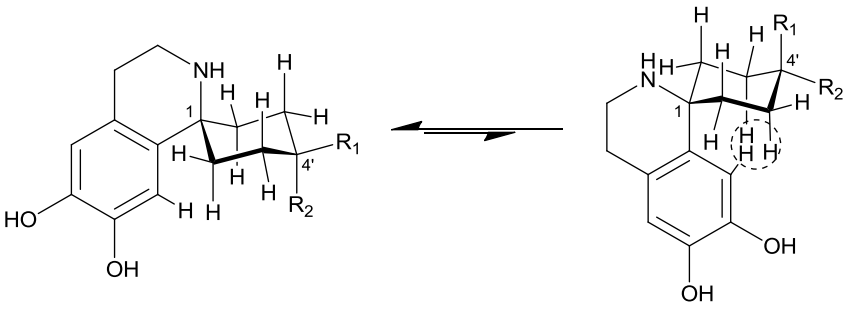


Fig. 6.41 Conformations of cyclohexane-spiro-THIQ. The two ring-flip conformations that can be occupied by spiro-THIQs. The axial/equatorial positions of 4'-substitutents are depicted (R_1/R_2). The LHS conformation is expected to be favoured as it has fewer steric clashes. The RHS conformation has considerable steric clashes, shown circled. For 4'-substituents the LHS R_1 equatorial position/conformation is most favoured.

The predicted stereochemical and conformational outcome of these reactions was confirmed by NMR spectroscopy—all three products had purely equatorial 4'-substituents. The crucial NMR experiment for assigning the configuration was nuclear Overhauser effect spectroscopy (NOESY), which identified the relative position of protons through space (Fig. 6.42). NOESY correlations between the isoquinoline aromatic proton (H_b) and cyclohexane protons (H_f) verified the dominant conformation of the molecules. The correlations from the 4' R-group to adjacent protons confirmed the equatorial position.

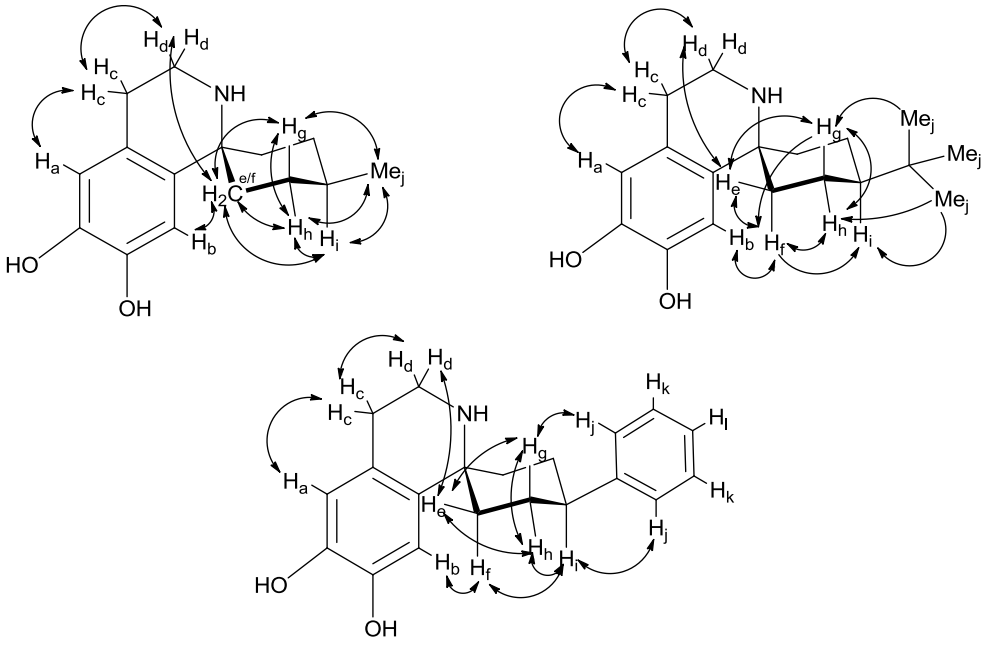


Fig. 6.42 NOESY correlations. These interactions verified the equatorial conformation of the 4'-subsituent. Due to the plane of symmetry through the cyclohexane ring, not all protons are depicted. If the 4' substituents were axial, then correlations between the 4' R-groups and H_f would be expected: these are not observed. Furthermore, correlations between the 4' R-groups and H_g would not be present if the substituent was axial. Arrowheads show the directionality of the correlation; correlations with the tBu group could only be confirmed in one direction. For complete ^{1}H and ^{13}C NMR assignment see methods section 2.6.3 and for NMR spectra see appendix B section 9.2.

6.6.11 3-Methylcyclohexanone

A reaction was conducted with A79F-Δ29*Tf*NCS lysate, dopamine and racemic 3-methylcyclohexanone (Fig. 6.44). In a similar manner to the 4-subsituted cyclohexanones, the most sterically favoured product would have an equatorial configuration. However, because the starting material is chiral, but used as a racemate, there are four possible products. If the substituent is in an equatorial position, then two possible enantiomers can be formed (Fig. 6.43). These enantiomers have two chiral centres, the 3'-centre defined by the starting material and a new centre at the 1-position. If the yield is over 50% then the enzyme must accept both enantiomers of 3-methylcyclohexanone, though it is possible one may be more favoured than the other. The nature of the products formed is determined by the whether the 1-position or the 3'-position dictates the reaction progression. If the 1-position dominates (i.e. through enzyme control—only the (1*S*)-THIQ is formed) then the two products will not be enantiomers, but diastereomers. One of these will have an axial methyl group. If the 3'-position defines the reaction progress (i.e. through substrate/steric control) then the products will be enantiomers and the enzyme will produce both an (*S*)-isoquinoline and a (*R*)-isoquinoline in one reaction.

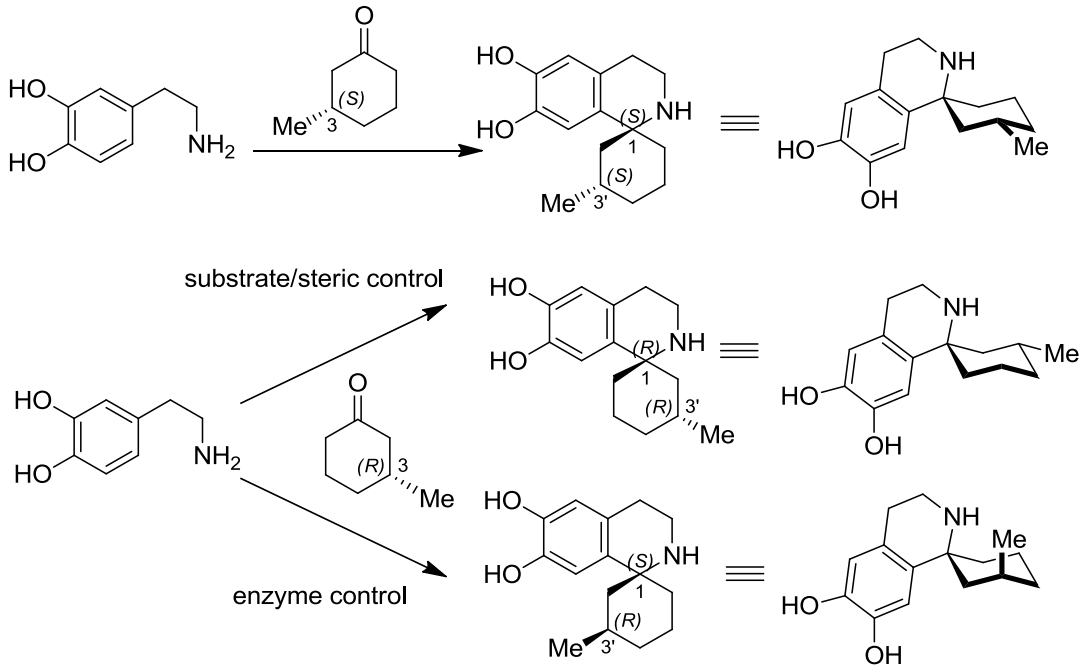


Fig. 6.43 *Rac*-3-Methylcyclohexananone products. If the more sterically favoured equatorial products are formed (substrate/steric control), two enantiomeric products may be produced. If 'enzyme control' dominates, and the 1-position controls the reaction, then (*R*)-3-methyl-cyclohexananone would form a product with an axial substituent.

The product was obtained with a purified yield of 60%, showing that both 3-methylcyclohexanone enantiomers are accepted (Fig. 6.44). NMR analysis showed only a single compound was formed, suggesting two enantiomers are formed.

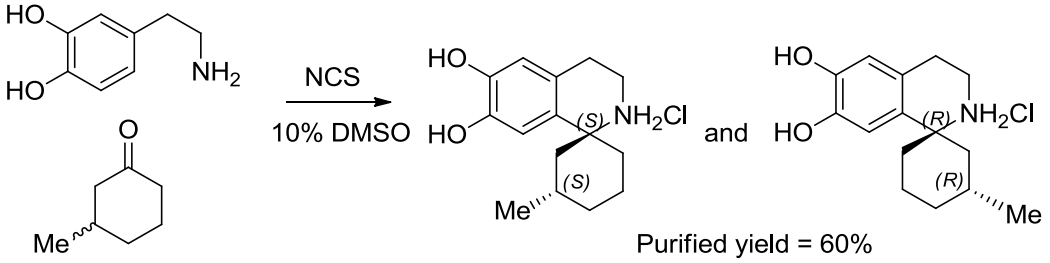


Fig. 6.44 3'-Methylcyclohexane-spiro-THIQ formation. Reaction conditions: 50 mM HEPES pH 7.5, 20% v.v^{-1} A79F-Δ29*Tf*NCS lysate, 5 mM ascorbate, 10 mM ketone, 15 mM dopamine and 10% v.v^{-1} DMSO. The reactions were incubated for 6 hours at 37 ˚C. Purification was achieved by organic then acid extraction. Product mass was confirmed by HRMS. For more details see methods section 2.6.3 and for NMR spectra see Appendix B section 9.2.

The NMR spectra were fully assigned, and this confirmed the identity of the enantiomeric 3'-methyl equatorial product(s). Again, the NOESY was vital in confirming the identity of the compound (Fig. 6.45).

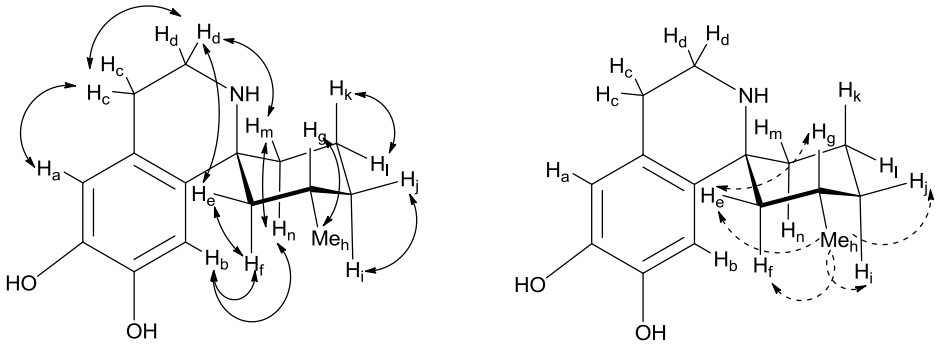


Fig. 6.45 NOESY correlations. Solid lines (LHS) show strong correlations, dotted lines (RHS) show weaker correlations. There is probably correlation between H_g and H_k but signals are very close so appear positive on spectra (correlations are typically negative). Due to similar shifts, it is also difficult to distinguish between H_f and H_k, so some correlations could not be assigned. Arrowheads show the directionality of the correlation; correlations with the Me group are only described in one direction. Crucially, no evidence of correlation is observed between the 3'-methyl and H_n, H_m, H_l or H_k, which would be present if the 3'-methyl group was positioned axially. For more details and NMR assignments see Methods section 2.6.3 and for NMR spectra and more detailed assignments see Appendix B section 9.2.

The relative proportions of the (1*S*,3'*S*) and (1*R*,3'*R*) products are unknown, and it is possible the enzyme prefers one 3-methylcyclohexanone enantiomer over another. This distribution could be investigated with chiral HPLC or derivitisation of the products. Comparing this racemic reaction with an enantiomerically pure 3-methylcyclohexanone reaction would also be informative.

This reaction is notable for being the first demonstration of the formation of a (1*R*)-isoquinoline by NCS, though admittedly this is more a quirk of nomenclature than a novel reaction. Nevertheless, the fact that these equatorial products are formed shows that, with respect to stereochemistry, these substituted cyclohexanone reactions seem to be mostly controlled by substrates/sterics than the enzyme.

6.6.12 Docking studies

Preliminary computational docking studies into the cyclohexanone reactions have been conducted in order to understand how the substrates may fit into the NCS active site. So far, only the 4-*tert*-butylcyclohexanane-iminium intermediate has been investigated. The most favourable binding modes for this compound fit precisely with the expected conformation according to previous docking studies and the dopamine-first mechanism (Fig. 6.46) (see section 4.2.2). The intermediate binds in a dopamine-first conformation, with the catechol bound to the Lys-122. The iminium is in a *trans* configuration, with the nitrogen near Tyr-108 and Glu-110. The cyclohexane is in a chair conformation, with the *tert*-butyl group equatorial, and points directly out the active site.

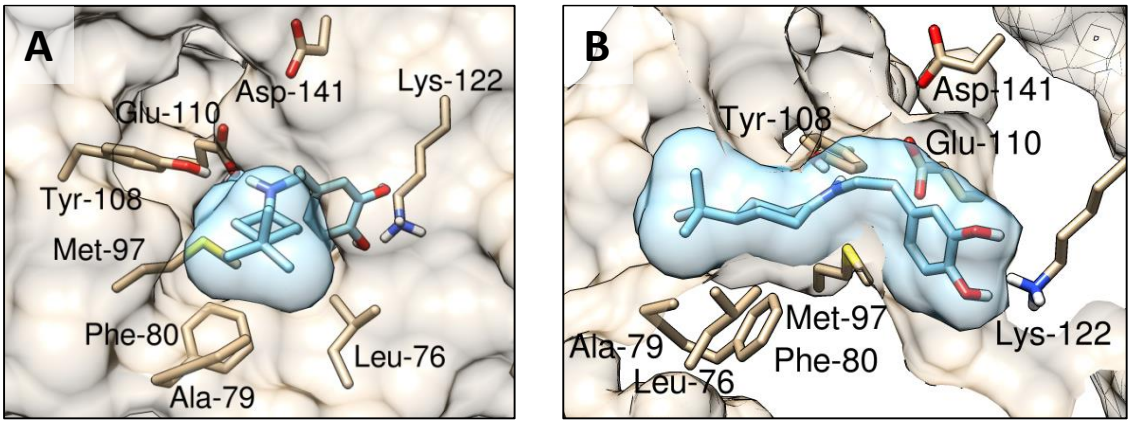


Fig. 6.46 4-*Tert*-butylcyclohexane-dopamine iminium docking. The imine intermediate of dopamine/4-tBucyclohexanone docked in *Tf*NCS active site (see methods section 2.2.3). This is the most energetically favourable binding mode. The nitrogen is in close proximity to Tyr-108 and Glu-110, as predicted by the dopamine-first mechanism (see section 4.2.2).

Here it appears that the alternative axial conformation of the *tert*-butyl group would not be tolerated—it seems, therefore, that the general chemical sterics of the reaction match the enzyme steric requirements. This model also shows how substitutions on the cyclohexane 2-position would be sterically unflavoured in the context of the enzyme. This matches experimental observation (section 6.6.8). Similarly, the origin of the failure to convert 3,3,5,trimethyl-cyclohexanone seems clear—an axial methyl group on the 3-position would be sterically unflavoured.

There appear to be relatively close interactions between residues at the active site entrance and cyclohexanone: Leu-76, Ala-79 and Phe-80. These residues were targeted in the past for modification of aldehyde substrate specificity. The mutations A79I, A79F and F80L had a positive impact on conversions with cyclohexanone (Fig. 6.37); though data was not normalised to protein concentration. It is possible that these mutations modify the shape of the cyclohexanone binding site, increasing shape complementarity. Nevertheless, it does seem counter-intuitive that increasing bulk at the active site entrance (A79F/I) can have a positive impact for bulky substrates such as 4-phenyl- or 4-*tert*-butylcyclohexanone as docking seems to show there is no space for the larger amino acid side chains. It is possible to account for increased activities by proposing that the bulky residues do not contribute to the initial ketone binding but to the subsequent reaction progression. The increased bulk could stabilise conformations of the imine intermediate that enable the crucial cyclisation step. This hypothesis is discussed in section 4.4.3.

The A79I mutation was originally designed as *Cj*NCS2 has an isoleucine in position 79. Yet, in *Tf*NCS A79I appears to improve conversions with cyclohexanones whilst *Cj*NCS2 has no activity with ketones at all. The difference in activity is possibly the arrangements of loops around the active site entrance. Although there is generally strong sequence identity between *Cj*NCS2 and *Tf*NCS, there is some variation on the make-up of loops, especially the positions of proline residues. These could severely affect the active site entrance interactions, removing promiscuous activities to ketones. An alternate explanation would be that the catalytic machinery in *Cj*NCS2 is somehow 'weaker' and it cannot react with the more stable ketones, regardless of the initial binding. Modelling the *Cj*NCS2 active site would greatly aid understanding of why the similar enzymes behave so differently.

6.6.13 The origin of activity with ketones

Aside from the differences between the two NCSs tested, the other key unknowns are the effect that the organic co-solvent has on the reaction, and the preference for lysate rather than pure enzyme. The reaction appears to proceed efficiently with 10% $v.v^{-1}$ DMSO, but does not proceed with 10% $v.v^{-1}$ MeCN. There are two possible origins to this observation. The co-solvent may affect the formation of the imine intermediate outside the active site, or it may alter the activity of the enzyme itself. Recent work on Δ29*Cj*NCS2 also described the positive effect of DMSO on reaction conversions, though this was for aldehydes and pure enzyme, rather than ketones and lysate.[61]

The PS reaction can be split into two parts: imine formation and cyclisation. For ketones, imine formation may be possible outside the enzyme active site, but the lack of reaction with the phosphate buffer catalyst suggests that cyclisation is definitely not possible outside the enzyme active site. It is known that the co-solvent can significantly affect imine formation in aqueous media. In a study of the related Mannich reaction between an amine, formaldehyde and butyraldehyde, it was found that DMF and DMSO were the best co-solvents due to their 'homogenising' effect.[209] Higher concentrations of these solvents provided improved rates. In a similar study of aqueous imine (and enamine) formation *via* a fluorescent reaction reporter, DMSO and DMF were again found to be most effective co-solvents. MeCN demonstrated far inferior rates.[210] A related hypothesis for NCS is that the presence of DMSO (rather than MeCN) dramatically shifts changes the equilibrium between amine/ketone, imine and enamine enabling enzymatic reaction progression.

The effect of organic solvents on enzyme activity is widely acknowledged: stability, rate, stereo-selectivity and regio-selectivity can be modified positively with the addition of organic solvents.[211] There is a particular precedent for the action of DMSO on enzymes. For lysozyme, the addition of small quantities of DMSO (<5% $v.v^{-1}$) increases the rigidity of the enzyme, reducing non-productive conformations and increasing the rate.[212] DMSO has also been shown to improve the enantioselectivity of some lipases.[213] The co-solvent may be affecting the NCS structure enabling it to catalyse the formation of more challenging (both sterically and energetically) products.

The results have some implication regarding NCS in plant metabolism. The enzyme appears to demonstrate significant promiscuity with respect to its substrate. The observed activity with ketones does not quite qualify as catalytic promiscuity as the reaction has not changed. However, turning over ketones is certainly more energetically demanding than aldehydes; this is demonstrated by the fact that the phosphate cannot catalyse this reaction but turns over aldehydes. The observed activity with ketones is not merely substrate promiscuity, which is generally based on steric and not energetic requirements.[214] It is interesting that the enzyme seems to have 'extra' catalytic ability compared to its natural reaction, as it may be expected that evolution would maintain only the necessary catalytic power in the enzyme. The ketone reactivity does not necessarily co-evolve with aldehyde reactivity, as demonstrated by *Cj*NCS2. It is not known whether the conditions in plant vacuoles, where NCS is expected to reside, are condusive to the ketone activity, or whether this activity is only revealed *in vitro*. However, the promiscuity of the enzyme may explain partially why conversions of 4-hydroxyphenylpyruvate were previously observed with NCS

enzymes,[25,46] and why 1-carboxylic 1,1-disubstituted THIQs were observed in plant extracts and believed to be BIA precursors.[207,208]

6.6.14 Future work

The ability of NCS to turn over ketones and form 1,1-disubstituted-THIQs is remarkable, but more data is required to complete the investigation. The products of the reactions with the ketones cyclohexanone, cyclopentanone, phenylacetone, 4-hydroxyphenylacetone and 4-methyoxyphenylacetone will be purified and characterised, using the extraction method described above. The products of the α-keto-acids 4-hydroxyphenylpyruvate, phenylpyruvate and 4-methyoxyphenylpyruvate will also be purified and characterised, though these compounds will exist as zwitterions so may not purify using organic extraction methods; preparative-HPLC may be used instead. Purification of the (*R*)-3-methylcyclohexanone product will provide an interesting comparison with the racemate. Enzymatic conversion data for these reactions will be measured by calculating dopamine consumption *via* HPLC: samples have been saved for this analysis. This will also give a good indication of the efficiency of the purification procedure.

The stereochemical characteristics of the methyl ketone reactions are unknown. Once obtained and purified, the products will be analysed by chiral-HPLC to determine their enantiomeric excess. The reactions with methyl ketones could also be performed with a chiral arylethylamines, such as metaraminol or L-DOPA, causing formation of a diastereomeric product which can be analysed *via* NMR. However, NCS activities are low for these substituted amines. So far the purified compounds are all solids: it may be possible to obtain a crystal structure of a compound, which will enable the assignment of an enantiomeric product without the requirement to form a diastereomer. Conversions with aliphatic methyl ketones are currently insufficient for product purification and characterisation. Conditions producing higher quantities of product will be determined prior to preparative scale reactions.

As described above (section 6.6.2), it may be possible to access natural BIA subgroups with a ketone PS reaction, catalysed by NCS. The spiro-BIAs and erythrina alkaloids may be accessed with 2-indanones or 2-substituted cyclohexanones respectively. Presently there is no evidence for successful activity with 2-indanone or 2-substituted cyclohexanones, but similar ketones such as cyclopentanone and 3-substituted cyclohexanones show turnover. Targeted mutations of NCS may enable modification of substrate tolerance and allow access to the more challenging cyclic ketones, eventually providing a ‘metabolic short-cut’ to the spiro-BIAs and erythrina alkaloids.

6.7 Conclusion

NCS is an excellent candidate for directed evolution but unfortunately no reliable high-throughput method has been established for rapid screening. The high background reaction, instability of the aldehyde substrates and variety of possible products formed by the enzyme has proved problematic in the formation of such a method. The fluorescamine screen remains the most reliable but is not ideal for lysate screens or variable protein concentrations. The method that has proved most fruitful for screening NCS activities has been HPLC analysis. This is the most sensitive method and will produce fewer false positives or negatives than more rapid methods. It enabled the identification of activities with α-substituted aldehydes and ketones, activities which were largely undetected by other methods.

Directed evolution benefits not just from rapid screening methods but also from the use of small, smart libraries.[109] The knowledge of the enzyme mechanism enables design of combinatorial rational mutation libraries specifically targeting particular enzyme activities.[215,216] These smaller libraries require less screening effort and so lower-throughput, more sensitive methods like HPLC can be used.

The identification of activities with ketones demonstrates the potential of the NCS for the synthesis of more diverse isoquinolines. Mutations with improved activities towards unusual substrates have already been identified. Directed evolution could improve these activities further. Another excellent target for directed evolution is the kinetic parameters of NCS for its natural substrates. NCS appears to have fairly low catalytic efficiency (see section 4.4.6); an improved NCS would be useful for biocatalytic and synthetic biology routes to THIQs. Ultimately a toolkit of NCSs, each a specialist at a particular PS THIQ reaction, could be produced through multiple directed evolution experiments.

7 Chapter 7: Conclusion and future work

7.1 Mechanism and kinetics

7.1.1 Dopamine-first mechanism

The overall aim of the project was to explore the use of NCS for biocatalysis. The first step was to confirm the mechanism of the enzyme; this would provide a starting point for future enzyme engineering. Computational and experimental data collected during this study have provided considerable evidence in favour of the dopamine-first mechanism for NCS (Chapter 4). Roles have been proposed for catalytic residues in the enzyme active site, and the positions of the substrate binding sites have been established. The successful rational engineering of the aldehyde substrate profile demonstrated the veracity of the dopamine-first mechanism. Based on all current understanding, a detailed mechanism for NCS has been proposed (Fig. 4.38).

7.1.2 Kinetics

There have been considerable challenges in achieving a full understanding of NCS reaction kinetics, but some progress has been made (section 4.4.5). A new method for measuring kinetics was developed which avoided problems associated with incompatible buffers. Based on this method, it has been possible to compare the kinetics of different mutants and substrate combinations, though it has not always been possible to fully rationalise the observed effects. The presence of an organic co-solvent in the enzyme assays may also have affected the kinetic parameters calculated.

The apparent kinetic parameters derived from this work describe an enzyme which requires high substrate concentrations (very high apparent K_m values) to achieve high turnovers. It is possible that *in vivo* this sensitivity acts as a gatekeeper into BIA biosynthesis: only when there is high substrate concentrations are present will metabolic flux be diverted into this secondary metabolism pathway. This could prevent energy/material from being misdirected in the plant metabolism. The high apparent K_m values may also have impacted efforts to use NCS in microbial systems, though recent work has shown this *in vivo* activity is possible.[21,45,76] The kinetic assays were found to be extremely sensitive to substrate concentrations, mostly because of the non-enzymatic background reaction. Despite the effort made to establish a good method, ultimately significant compromises were made. Overall it has not been possible to obtain a satisfactory understanding of the enzyme kinetics.

7.1.3 Future work

Most work on the NCS mechanism, kinetics and structure has been conducted with *Tf*NCS. It would be interesting to examine other NCSs and see how they compare. In the final section of this study it was shown that *Tf*NCS and *Cj*NCS2 have considerably different substrate tolerances (section 6.6.5); the origin of these differences could be examined. There has also been recent work showing significant differences in the *in vivo* activities of different NCSs.[76] This may be due to their kinetics—perhaps only *Tf*NCS is 'inefficient', the other NCSs may have different kinetic properties.

The complete exploration of NCS kinetics is ultimately very important to fully understand the enzyme mechanism. However, the system is not easy to control, and there are many factors to deconvolute. Perhaps first, a kinetic understanding of the chemical phosphate catalysed PS reaction is necessary before tackling the enzyme kinetics. Also, before the complex kinetics is examined, there are many basic questions to be answered about the enzyme mechanism, especially regarding the steps lead to the formation of the imine.

A structural examination of NCS and its mutants would significantly aid any further investigation into the enzyme mechanism and would also be invaluable for enzyme engineering efforts. Currently, the effort to establish NCS crystallisation conditions for X-ray structural analysis has not yielded fruit, but for now this work is continuing. It is hoped that a new structure would correct some ambiguities in the previous structure, such as the orientation of Phe-112. An NMR structural investigation would also provide interesting data to this regard, but NMR studies require significant time and resources to establish.

It may even be possible to examine NCS using a biophysical method, such as intrinsic tryptophan fluorescence. There is tryptophan in within 7 Å to the interesting Phe-112 and Lys-122 residues. Once other tryptophans (and perhaps tyrosines) are removed by mutation, it may be possible to track the movement of Phe-112 and substrate binding through changes to tryptophan fluorescence in single molecule experiments. This would provide a dynamic picture of NCS.

Finally, there is still much to understand about *in planta* NCS behaviour. The full functions of the N- and C-termini are unknown (section 3.6). Furthermore, there is the intruiging possibility of functional aggregated NCS inside plant cells. There are also experiments that can be done to examine the hypothesis that NCS is inefficient due to metabolic flux control.

7.2 Biotransformations and cascades

7.2.1 Biotransformation methods

Methods for biotranformation procedures using NCS have been improved over the course of the project. Significantly, (desalted) lysate was shown to be a suitable method of enzyme preparation, despite the background reaction. The first use of mutants for biotransformations was also demonstrated. Despite this, most NCS biotransformations demonstrated in this project were not superior to previous examples. The use of preparative-HPLC was not ideal for the purification of products from larger scale conversion (>100 mg), and concentrations of substrates used are far too low for any industrially relevant transformations.

However, the biotransformations conducted with substituted cyclohexanones have led to significant improvements in biotransformation methodology (section 6.6.9). The adoption of the anti-oxidant ascorbate, use of excess dopamine and development of an extraction method has led to a cleaner, quicker and more scalable biotransformation protocol. The use of ketone substrates, which do not suffer from the background reaction, will enable much higher substrate concentrations in the future.

7.2.2 Triangular cascade

A multi-enzyme cascade was developed for the synthesis of (*S*)-norlaudanosoline and an (*S*)-tetrahydroprotoberberine (section 5.4). During the optimisation of the cascade, seven previously unreported transaminases were examined. The one-substrate, two-enzyme, one-pot 'triangular' arrangement was a novel method to pull the TAm equilibrium reactions to completion. Overall, with the additional formaldehyde chemical step, these cascades demonstrate a remarkable increase in molecular complexity over a short time: one chiral centre, two C-N bonds and two C-C bonds can be formed within about 3 hours.

7.2.3 Future work

Although the triangular cascade is limited with regards to its substrate/product scope, there is potential for scaling the reaction. Conditions can be modified to improve the regioselectivity of the final step. The newly developed extraction protocol can be used for product purification, rather than preparative-HPLC.

The extension of the cascade reaction that can be worked on as soon as possible is the use of different aldehydes in the final chemical step. This will increases the possible range of products significantly. The stereocentre set by NCS could influence the subsequent PS reaction, resulting in products with two chiral centres.

There are many more enzymes that can be use with NCS to build cascades both *in vitro* and *in vivo.* For example, a different TAm can be used, which does not turn over dopamine—this will allow the use of two different amines as starting materials for THIQs. Other enzymes that form aldehydes can be explored: alcohol dehydrogenase, carboxylic acid reductase, etc. Enzyme can be used to form dopamine (e.g. tyrosinase and decarboxylase) and can be used to modify the THIQ product (e.g. CYPs). The newly discovered NCS ketone/cyclohexanone activity could be combined with enone reductases.

The incorporation of the complete BIA pathway into microbes is currently a particularly exciting area, with significant progress being reported on an almost monthly basis.[19–21,45,65,76] However, biosynthetic pathways do not have to just be copied from a natural product producing organism into microbes. Instead, novel biosynthetic pathways can be constructed, using promiscuous enzymes from various sources. This will enable the production of novel and non-natural chemicals *in vivo*.

7.3 Screens and substrates

7.3.1 Screening methodology

Screening methods were developed with the intention of using them for screening libraries of NCS mutants in an enzyme engineering/directed evolution effort. Although the fluorescamine screen for NCS activity had already been developed, it was not suitable for lysate or non-purified enzymes, so different methods were investigated. The investigation into Purpald® did not result in a good screening method. Aldehydes are readily consumed by native *E. coli* proteins so there is little reliability in detecting aldehyde concentration. Also, Purpald® was found to react with amines. Some improvements were made with regards to increasing the throughput, by using small cultures and nickel coated plates. Ultimately, however, no significantly improved method was developed. Ultimately, the low-throughput HPLC proved most useful screening activities.

7.3.2 Screening results

NCS activity was observed with a range of aliphatic aldehydes, which had not been investigated beforehand. Especially interesting was the activity with the citronellals. A number of mutants were found to have improved activities, including L76A for (*S*)-citronellal and A79I for numerous aldehydes including the natural substrate. No activities were identified with *N*-alkyated amines. Furthermore, there was no success in the effort to switch enzyme regioselectivity. However, the double mutants F122K-

K122F/L showed native activity, suggesting a certain robustness to the enzyme mechanism. Activities were observed with the α-substituted aldehydes 3-hydroxybenzaldeyde, benzaldehyde and cyclohexanecarboxaldehyde.

7.3.3 Ketones

The identification of NCS activity with ketones is probably the most significant results of this entire project (section 6.6). It is especially remarkable how high the activities are with derivatives of cyclohexanone. Furthermore, there is considerable intrigue with regards to the solvent dependence of the reaction, and the lack of activity with *Cj*NCS2. The novel spiro compounds generated are also of interest—there are very few methods to synthesise them. This discovery is of wide interest, with potential impact within enzymology, synthetic biology and chemistry.

7.3.4 Future work

Future screening and engineering is likely to use HPLC for maximum sensitivity. Although a rapid screening method is useful, if a relatively small number of mutations need to be screened, then sensitive methods are preferred. Mutations will be site directed, and informed by the dopamine-first mechanism and docking investigations. If directed evolution is to be attempted, libraries will be rationally designed, small and smart to maximise hits and minimise screening effort.[109,111]

A toolkit of many different evolved NCSs can be generated, each specialising in a particular substrate. It may be worth evolving the seemingly catalytically inefficient enzyme towards improved kinetic parameters. Activities with aliphatic compounds and α-substituted compounds can be significantly improved.

There is scope for considerable improvement in the use of ketone substrates. The aliphatic methyl ketones seem to have very poor yields—these can be improved through evolution. The products of these reactions are likely to be chiral: (1*S*)-1,1-disubstituted-THIQs. This will be one of the only methods, chemical or enzymatic, for the formation of such compounds.[156] Finally, other ketones currently not accepted by NCS may be accepted by evolved enzymes—finding an NCS that turns over the 2-substituted cyclohexanones and/or 2-indanone will provide a metabolic short cut into the erythrina and spiro isoquinoline families (section 6.6.2).

7.4 Perspective

Ultimately the project fulfilled its key objective: to improve the understanding of the NCS mechanism *via* mutation and then to use this knowledge to engineer the enzyme activity. This was most clearly achieved with the improved activity of L76A for (*S*)-citronellal.[137] The aim to explore the biocatalytic potential of the enzyme was also achieved, especially with the development of a novel triangular cascade method.[161]

The most significant result of study was the observation and verification of NCS activity with ketones, in particular substituted cyclohexanones. Interestingly, in contrast to the project's objective, the discovery of this activity was not a result of the mechanistic understanding of the enzyme. Instead, it was achieved through a mixture of chemical intuition (the ketone-aldehyde relationship), serendipity (the use of DMSO) and, perhaps most importantly, a willingness to 'follow one's nose' (i.e. recognise promising preliminary results and follow them up). Many aspects of the ketone activities are completely unpredictable: the effect of the solvent, the preference for lysate, and the differences in activity between closely related enzymes.

Why does the reaction occur? There is no clear evidence for these types of reactions occurring in plants. It appears, through the behaviour of NCS homologs, that the NCS ketone activity does not necessary co-evolve with the aldehyde activity. One would expect, through genetic drift, promiscuous reactions such as the ketone reaction to not occur. Though it is possible that the formation of 1,1-disubsituted-THIQs by NCS is a purely *in vitro* phenomenon, there are historic reports of 1-carboxylic acid THIQ derivatives in *Papaver* species.[207,208] It is possible there are undiscovered 1,1-disubsituted-THIQ natural products in *Thalictrum flavum* and other BIA producing species formed by NCS.

NCS may also play a role in other biosynthetic pathways. *Cj*NCS2 has been shown to turn over dihydrocinnamaldehyde, forming a THIQ similar to those on the biosynthetic pathway to colchicine.[57,61] The Pictet-Spenglerases forming the ipecacuanha alkaloids remain unknown; an enzyme closely related to NCS could catalyse these reactions. The Pictet-Spengler reactions with dopamine form at least three branch points of plant secondary metabolism; all these pathways may have diverged from a Pictet-Spengler reaction producing various THIQ products, catalysed by a promiscuous NCS ancestor.

Intriguingly, certain enzymes, especially those involved in secondary metabolism, may have evolved to be promiscuous. When a change in the (chemical) environment of an organism occurs, these promiscuous enzymes would be able to respond by producing

mixtures of novel compounds. If one of these compounds has a positive impact on survival, its formation would be selected for: enzyme expression would be upregulated and the enzymes would acquire ‘specialising’ mutations, increasing their catalytic efficiency. This is when the classical biosynthetic pathways would arise. Yet underlying these upregulated pathways, there may still remain low-level, plastic, promiscuous metabolic networks able to produce novel compounds in response to new environmental conditions.[3] Regardless of specific evolutionary mechanism, there are certainly countless enzymes in secondary metabolism with promiscuous activities that have not yet been explored.

As this work demonstrates, there are aspects of enzyme behaviour that are not easy to rationalise and certainly very difficult to predict. Despite this, there are bold efforts to design novel enzymes from scratch using computational methods.[217,218] Typically, these designed proteins are inferior to extant enzymes and are considerably improved by directed evolution, thereby demonstrating both imperfect design methods and the blind power of evolution.[219,220] An alternative approach to developing ‘new’ enzymes is to use a combination of chemical intuition and directed evolution.[108] By exploiting the existing catalytic machinery in extant enzymes and harnessing the power of (directed) evolution, it is possible to develop enzymes with novel activities.[221–223] This is similar to how nearly all enzymes have come into existence. This approach shows an appreciation that there remain significant mysteries in nature—unknowns in enzyme mechanism, promiscuity and evolution. By attempting to unravel these mysteries novel methods and materials for industrial biocatalysis and synthetic biology will be revealed.

8 Appendix A: Molecular Dynamics

8.1 Introduction

8.1.1 Protein dynamics

In solution, proteins are dynamic structures: bonds stretch, surfaces are buffeted by solvent molecules, loops flex, side chains rotate and, in some cases, whole portions of a structure move relative to each other. Overall, these motions contribute enormously to the function of a protein, though there remains much debate about the exact roles motions play.[233] On a very simplistic level, the overall conformation of an enzyme structure—its tertiary fold, secondary structures, side chains and loops—will determine its ability to both bind a substrate and catalyse a reaction. In some cases these two prerequisite events of catalysis may require different protein conformations.[234]

By considering the vast possible structures an enzyme amino acid sequence can occupy it becomes immediately clear that only very few conformations would be conducive to substrate binding or catalysis. Therefore limiting the accessibility of non-productive conformations (reducing conformational entropy) is vital for efficient catalysis.[235] This is generally achieved by increasing the rigidity/stability of a productive conformation, or at least severely limiting the number of easily accessable protein conformations.[146] By considering that impact of flexibility, stability and conformation in catalysis, the role of residues beyond the enzyme active site becomes clearer: generally, the active site conformation determines the catalysis, the overall fold determines the relative positions of the active site residues and amino acid networks control the exact conformations of the active site residues.[236,237] Altogether the combination of these interactions result in an active enzyme.

8.1.2 Molecular Dynamic Simulations

The motions within proteins determine the major structural conformations and therefore activities; in turn, these motions will impact experimentally observable phenomena. Despite this, protein motions are very difficult to detect directly. This is partly due to the fact that most experimental techniques deal with ensemble measurements and not single molecules, but the small size- and time-scale of the changes also limits their detectability. Experimental methods such as NMR and single molecule FRET can detect exchange/motion between different protein conformations.

Using computational simulations it is possible to ‘observe’ protein motions. Presently, computational simulations can offer hypotheses which can then be verified experimentally. Simulations can also be used to provide explanations for experimental

observations. The former hypothesis-generating approach is more scientifically rigorous and therefore preferable.

All-atom molecular dynamics (MD) simulations are typically used to compute protein behaviour over a few hundred nanoseconds. This time period can detect fast motions such as bond fluctuations and sidechain conformations. It is also possible to observe coupled motions between residue networks and loops. Longer scale motions, such as conversions between secondary structures and motions between domains are typically of larger timescales so must be examined with coarse-grained models or a 'metadynamics' type approach.

8.1.3 NCS

Norcoclaurine synthase (NCS) was considered a good candidate for MD simulations as it is a relatively small (21 kDa) single domain protein, with a high resolution crystal structure available.[43] Despite the high-resolution of the crystal structure, there were concerns about the interdomain N-terminal interaction and the impact this may have on protein conformation. Furthermore, there appeared to be a difference in the active site conformation in the two protein subunits—the rotamer of the Phe-112 was different (see section 4.1.3). This rotamer was expected to have a significant knock-on effect on catalysis as it is adjacent to the vital Lys-122 residue. Docking studies, which proved vital for establishment of the dopamine-first mechanism for NCS, were severely affected by the Phe-112 conformation (section 4.2.2). In subunit A, where the Phe-112 points into the active site, docking produced excellent data which enabled the establishment of the new mechanism. In subunit B the Phe-112 was modelled pointing away from the active site, in a seemingly non-favoured position (as determined by MOLPROBITY).[121] The docking for subunit B, on the other hand, provided very different data—substrates, when bound, were very deep in the active site and in arrangements that did not appear conducive to catalysis. The active site cavity revealed by the Phe-112 rotameric configuration significantly modified the docking results. Finally, there was also the suggestion in the X-ray electron density that in subunit B the Phe-112 could be observed in two conformations, one similar to that in subunit A.

These interesting observations, coupled to a general interest in protein motions and dynamics, led to the decision to conduct MD simulations for NCS. Of primary interest was the motion of Phe-112 in the active site.

8.2 Method

8.2.1 Preparation of NCS

The crystal structure of *holo* Δ19*Tf*NCS (2VQ5) was selected in favour of the *apo* (2VNE) courtesy of its superior resolution (2.1 Å compared to 2.7 Å).[43] The A and B subunits were separated into separate files. Water molecules, ions and ligands were removed from the structure. The N- and C-termini were truncated, with the protein starting at Ser-40 and ending at Lys-197. This ensured that only the main globular domain was observed in the simulations.

8.2.2 MD Commands

The commands for the MD simulations are described, numbered, below. The outputs for a single simulation of subunits A and B are described. The two other simulations for each subunit were conducted in an identical manner and provided similar results.

1. pdb2gmx -f 2VQ5A.pdb -o 2VQ5Ap.pdb -water tip3p -forcefield charmm27

Simulations were conducted in GROMACS using the CHARMM27 forcefield.[116,117] Enzyme protonation states were left unchanged: Glu-110 and Asp-141 were negatively charged and Lys-122 was positively charged. Although this may not be the catalytically active form of the enzyme, this is likely to a major form of the enzyme in solution and it is likely that this form of the enzyme binds the substrates.

2. editconf -f 2VQ5Bp.pdb -o 2VQ5Bbox.pdb -c -d 1.0 -bt cubic

The protein was centred in a cube and positioned at least 1.0 nm from the edge.

3. genbox -cp 2VQ5Abox.pdb -cs spc216.gro -o 2VQ5Asolv.pdb -p topol.top

The box was solvated with water. The Tip3p explicit water model was used for this as this is appropriate for use with CHARMM forcefields.

4. grompp -f ions.mdp -c 2VQ5Asolv.pdb -p topol.top -o ions.tpr
5. genion -s ions.tpr -o 2VQ5ions.pdb -p topol.top -pname NA -nname CL -np 8

Counterions to the protein charge were added. In this instance sodium positive ions were added to counter the overall protein negative charge. Water molecules were replaced with the ions.

6. grompp -f minim.mdp -c 2VQ5Aions.pdb -p topol.top -o em.tpr
7. mdrun -v -deffnm em

Structures with solvent and ions were energy minimised. Minimisation was stopped once F_{max} reduced below 250 (Table 8.1).

Table 8.1 Energy minimisation output.

Subunit	Potential Energy	Max force	Norm of force	Steps
A	-6.49E+05	2.4399084e+02 (atom 310)	8.02E+00	2980
B	-6.50E+05	2.4718457e+02 (atom 310)	7.556655	3239

8. grompp -f nvt.mdp -c em2.gro -p topol.top -o nvt.tpr
9. mdrun -v -deffnm nvt

NVT equilibration for 100 ps. This phase involves a constant **N**umber of particles, **V**olume and **T**emperature. It is known as "isothermal-isochoric" or "canonical". Temperature is expected to stabilise once equilibration is complete, which it was observed to do in all simulations (Table 8.2).

Table 8.2 NVT equilibration output.

Subunit		Potential Energy (kJ/mol)	Kinetic Energy (kJ/mol)	Total Energy (kJ/mol)	Temp (K)	Pressure (bar)
A	Average	-528020	103228	-424792	299.62	-791
	Err.Est.	370	76	450	0.22	15
	RMSD	5795.74	1200.11	6225.53	3.48335	345.121
	Tot-Drift	2331.66	506.323	2837.98	1.46961	110.578
B	Average	-528932	103144	-425788	299.643	-768
	Err.Est.	390	68	460	0.2	13
	RMSD	5830.16	1218.17	6289.27	3.53893	330.21
	Tot-Drift	2512.69	429.182	2941.88	1.24682	76.8577

10. grompp -f npt.mdp -c nvt.gro -p topol.top -o npt.tpr
11. mdrun -v -deffnm npt

NPT equilibration stabilises the pressure and the density of the system. **N**umber of particles, **P**ressure and **T**emperature are kept constant in this step. It is known as the "isothermal-isobaric" ensemble. Similar to the NVT step, this was conducted with a 100 ps simulation. The density was expected to converge on the value 1000 kg m^{-3}, the experimental value for water. In all simulations this convergence was observed (Table 8.3).

Table 8.3 NPT equilibration output.

Subunit	Energy	Potential (kJ/mol)	Kinetic (kJ/mol)	Total (kJ/mol)	Temp (K)	Pressure (bar)	Density (kg/m^3)	Volume (nm^3)
A	Average	-533970	103354	-430616	299.986	-0.79143	1005.13	415.737
	Err.Est.	64	24	74	0.071	3.1	0.38	0.16
	RMSD	797.369	460.299	904.811	1.33603	135.949	4.01689	1.7122
	Tot-Drift	-261.715	4.77326	-256.943	0.013855	1.27769	2.3275	-0.99008
B	Average	-534642	103220	-431422	299.865	4.39353	1005.03	415.435
	Err.Est.	110	16	110	0.047	5.8	0.39	0.17
	RMSD	835.758	469.415	950.211	1.3637	133.957	4.03061	1.71255
	Tot-Drift	-253.287	-37.1174	-290.403	-0.10783	1.16311	1.33054	-0.5769

12. gerun mdrun_mpi -deffnm md_100 -cpi -cpo -cpt 1

Three 100 ns simulations were run for each subunit (Table 8.4). The overall computational efficiency for each simulation was 22 ns/day.

Table 8.4 MD simulation output (100 ns).

	Subunit					
	A			B		
Energy	Average	Err.Est.	Tot-Drift	Average	Err.Est.	Tot-Drift
Pressure (bar)	0.921108	0.13	0.407032	1.08316	0.09	-0.41284
Potential (kJ/mol)	-534699	14	26.7634	-535491	4.5	0.667088
Total Energy (kJ/mol)	-431341	15	31.0167	-432226	5.1	4.93421
Temperature (K)	300	0.0035	0.012346	299.995	0.0026	0.012396
Density (kg/m^3)	1004.76	0.1	-0.39464	1004.66	0.014	0.065392
Kinetic En. (kJ/mol)	103358	1.2	4.25354	103265	0.9	4.26691
Constr. Rmsd	6.40E-11	6.40E-11	-3.84E-10	6.11E-11	6.10E-11	-3.67E-10
Volume (nm^3)	415.886	0.041	0.163388	415.584	0.006	-0.02689

Post processing was performed to ensure protein was whole throughout the simulation, and in the centre of the simulation box.

13. trjconv -s md_100.tpr -f md_100.xtc -o mdA_100A.xtc -pbc mol -center -ur compact -skip 10
14. trjconv –s md_100.tpr –f mdA_100A.xtc –o mdA_100Bfit.xtc –fit rot+trans

The “skip 10” command resulted in only one frame being obtained out of every ten. This made the output file a more manageable size, speeding up subsequent calculations and processes.

8.2.3 Calculating bond angles

The behaviour of the Phe-112 residue was examined by tracking two separate angles (using the 'angle' function). See section 4.2.1 for more details. Briefly: the rotation of the phenyl ring was tracked through the changing dihedral angle between the Cα, Cβ, Cγ and $C\delta^1$ atoms (Fig. 4.11). The position of Phe-112 relative to the entire protein was tracked by the angle formed between the Phe-112 Cα, Cζ (at the tip of the phenyl ring) and the Cα of His-106, which was on the same β-strand as Phe-112 but at the other end of the protein (Fig. 4.13). The distributions of these angles were measured both over time and as frequencies (Fig. 4.12, Fig. 4.14).

8.3 Results

8.3.1 General observations

The pressures, temperatures, densities, constrained RMSDs, volumes and energies (potential, total, kinetic) of all six simulations (three for each subunit) were stable over the 100 ns duration. No large motion or unfolding was observed for the simulations, except for the C-terminus, which was very mobile (Fig. 8.1, backbone RMS fluctuation). Subunit B appeared to be slightly more stable over the 100 ns, with fewer sharp peaks in RMSD (Fig. 8.1, backbone RMSD).

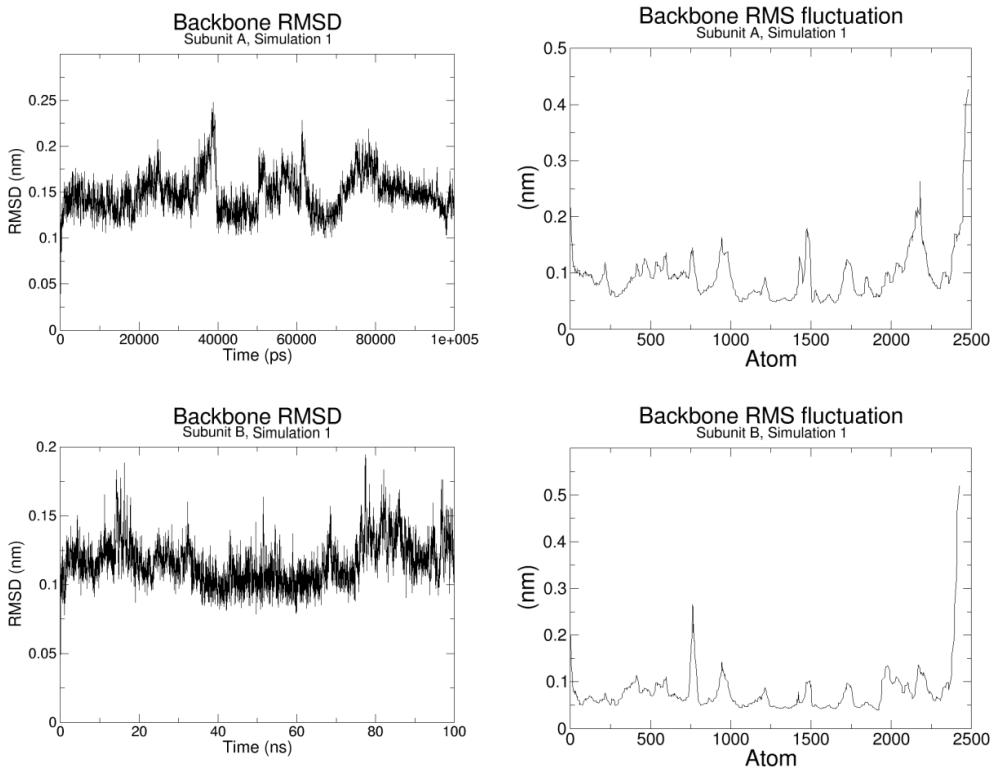


Fig. 8.1 Backbone root mean square deviations and fluctuations. Simulation 1 shown. Simulations 2 and 3 provided similar data.

8.3.2 Phe-112 dynamics

As described in section 4.2.1, the Phe-112 side chain appears to be flexible in the enzyme site. In all simulations, the residue aromatic ring flips, suggesting packing is not especially tight. More importantly, in all three subunit B simulations the Phe-112 residue rotamer changes conformation, adopting the conformation observed in subunit A. Subunit A simulations show no changes in Phe-112 rotamer conformation. Overall, this data suggests that the conformation seen in subunit A (at around 65°, Fig. 4.14) is the most favourable in solution. It is unclear whether the starting subunit B conformation (110-130°) is a crystallographic artefact or a conformation actually accessible in solution. If it is the latter, then it would be expected that the exchange from conformation A (65°) to conformation B (110-130°) can occur, but this was not observed in subunit A simulations. Longer simulations may reveal this exchange.

8.3.3 Phe-112 mutations

The mutant F112L was designed to investigate the Phe-112 residue. Full examination of this mutant has not yet been conducted, but it was included in substrate screens (section 6.6.5). This screen showed that the mutation F112L does not completely remove NCS activity, but significantly reduces it. The effect of the mutation on enzyme stability is unknown. Interestingly the mutants F112K-K122F and F112K-K122L show some NCS activity, but at very low levels (section 6.4.2).

The position of Phe-112, adjacent to the catalytically important Lys-122, means that although it has no direct role in the enzyme mechanism, it influences the positioning and environment of Lys-122. It is conserved in all known NCSs. It is likely to help maintain the pK_a of the lysine site chain, which is vital for the reaction progression. Based on preliminary data from mutations, however, its role does not appear to be so vital to make it a necessary residue for the NCS reaction.

8.4 Discussion

8.4.1 Phe-112 conformations

If the two different Phe-112 conformations are real features of the enzyme in solution, then they would have significant impacts on the enzyme activity. Docking investigations on the two different subunits demonstrated that the different Phe-112 conformations cause very different binding modes. Preliminary experimental mutation data suggests that Phe-112 is a very important residue for the enzyme mechanism, if not vital.

It has been demonstrated that NCS has relatively high apparent K_m values for its natural substrates (section 4.4.6). Non-productive/non-binding active site

conformations would result in high K_ms—there is a possibility Phe-112 conformations could impact on this value.[146] If this were to be the case, then many questions would be raised regarding the role of residue and its conservation. Perhaps the high apparent K_m is favoured due to *in planta* metabolic control—if so, the unfavourable conformations of Phe-112 could be evolutionarily favoured. Alternatively, perhaps no other amino acid is tolerated in its place (adjacent to the important Lys-122), but the unfavourable conformation is a necessary but unflavoured consequence of this. It is also may have a role in controlling the water channel adjacent to the residue.

8.5 Conclusion and further work

The MD simulations successfully demonstrated that the Phe-112 observed in subunit B of the Δ19*Tf*NCS crystal structures (2VNE and 2VQ5) are not stable in solution and decay to the more favourable conformation seen in subunit A. It remains unclear, however, whether this conformation B is crystallographic artefact or a real feature of NCS in solution.

If the alternative Phe-112 conformations are real, they would impact substrate binding, as suggested by docking studies. If this were the case, in order to improve the enzymes' apparent K_ms, the Phe-112 residue could be targeted for mutation. First, the alternative aromatic residues tyrosine and tryptophan can be investigated in the 112 position. Secondly, improving the packing of the Phe-112 side chain by mutating adjacent residues can be attempted—hopefully fixing the side chain into the favourable conformation.

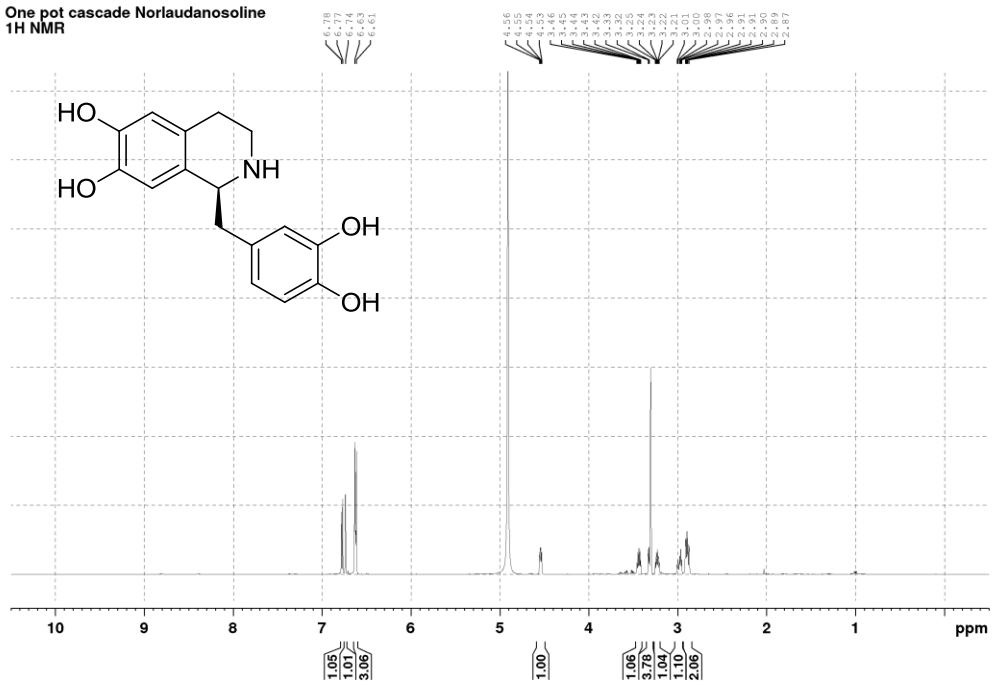
The motion of the Phe-112 remains a hypothesis based on MD simulations; the only tentative experimental data for this is the X-ray crystal structure electron density. Experimental data for the residue motion could be obtained through NMR spectroscopy, perhaps through specific labelling of phenylalanine residues. These NMR experiments are not trivial, and would require significant time and resources. Slightly more straightforward experiments could use fluorescence methods: either intrinsic fluorescence (perhaps from Trp-61, which is 6.5-7 Å from Lys-122 and Phe-112) or from FRET probes. Ideally these experiments would be done in a single molecule manner, to identify discrete conformational states.

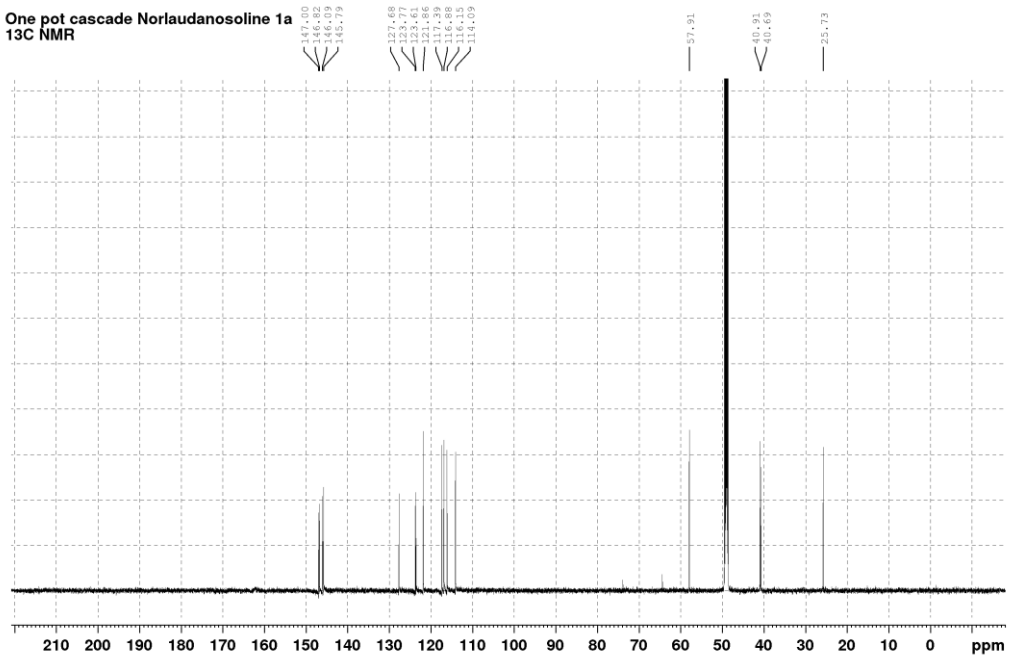
9 Appendix B: NMR spectra

9.1 Triangular cascade products

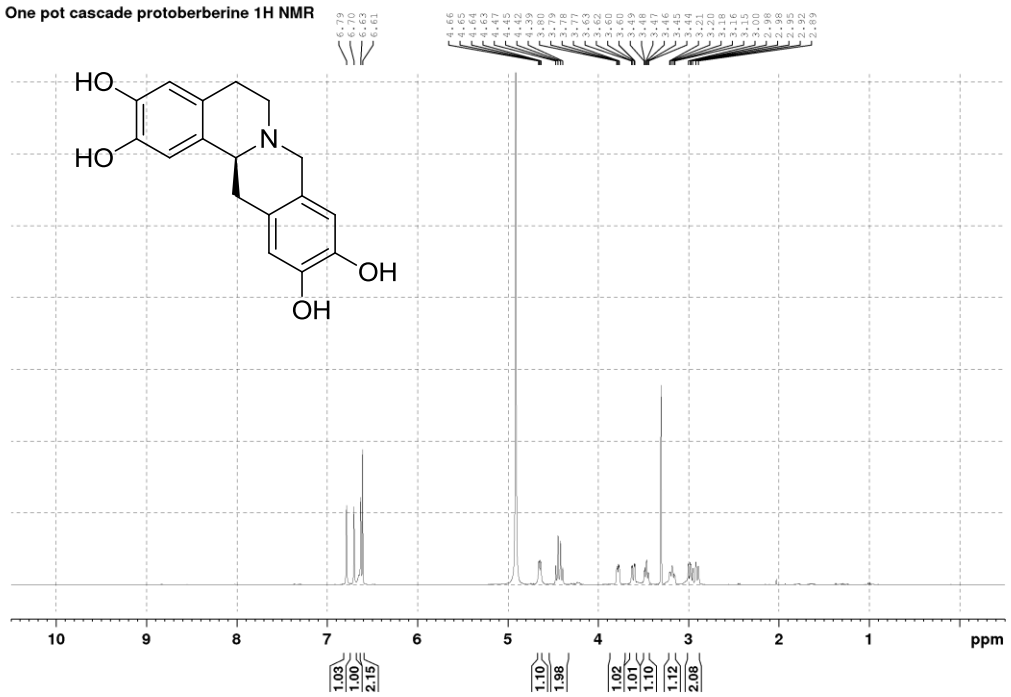
From methods section 2.6.2 and section 5.4.10.

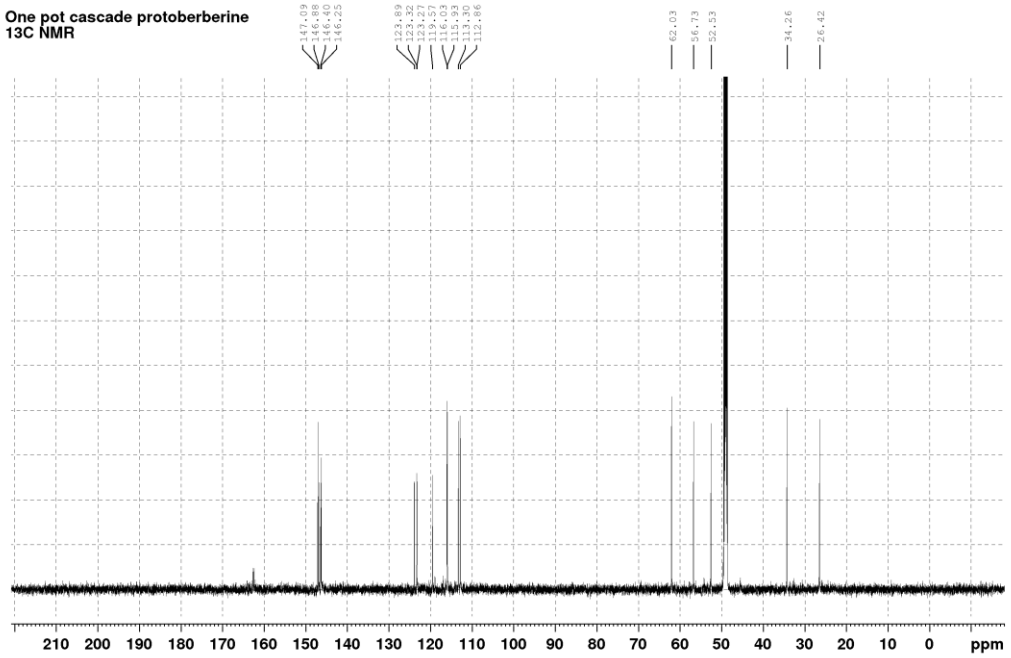
9.1.1 Norlaudanosoline





9.1.2 THPB

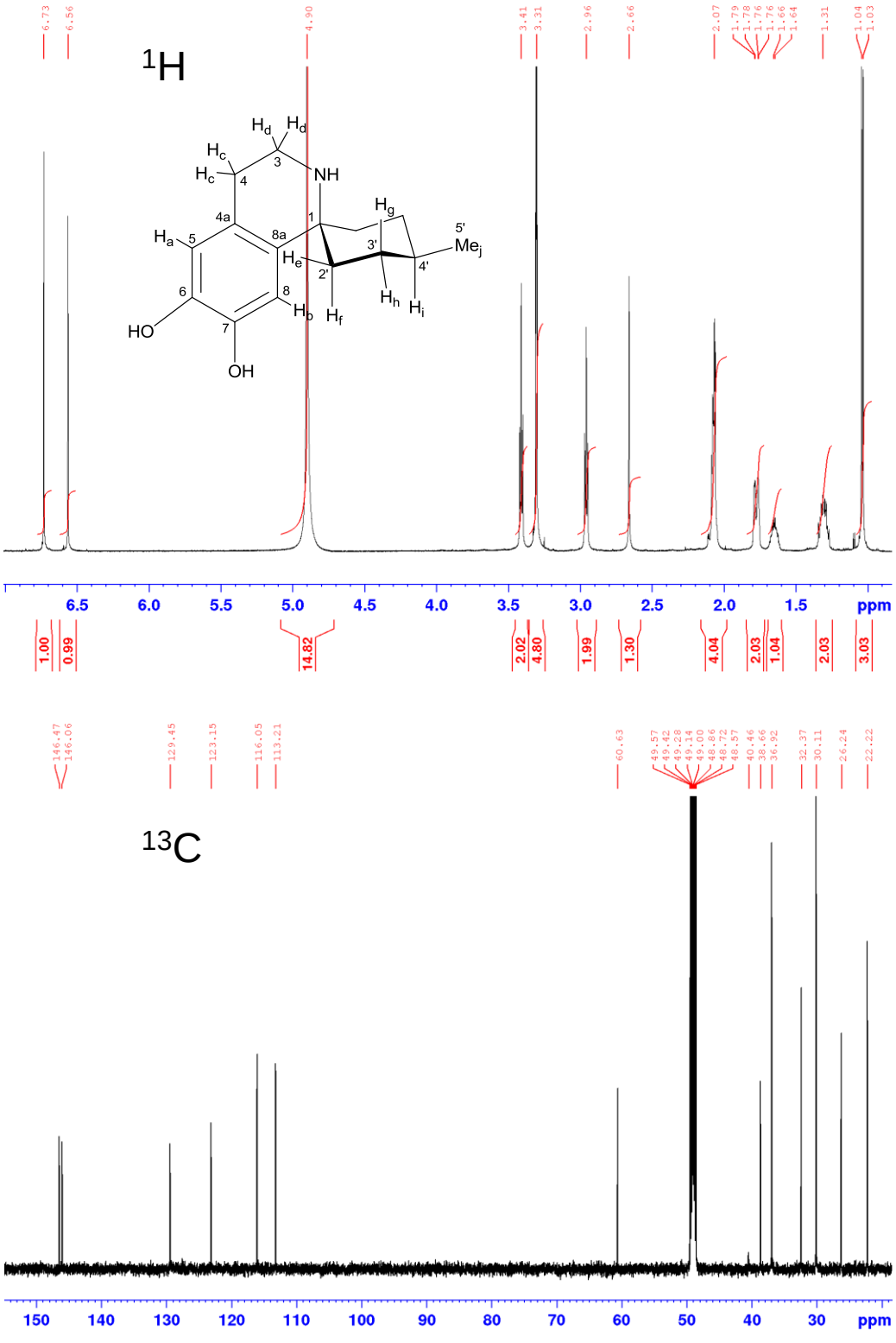

One pot cascade protoberberine 1H NMR
HO
HO
N
OH
OH
ppm


One pot cascade protoberberine
13C NMR
ppm

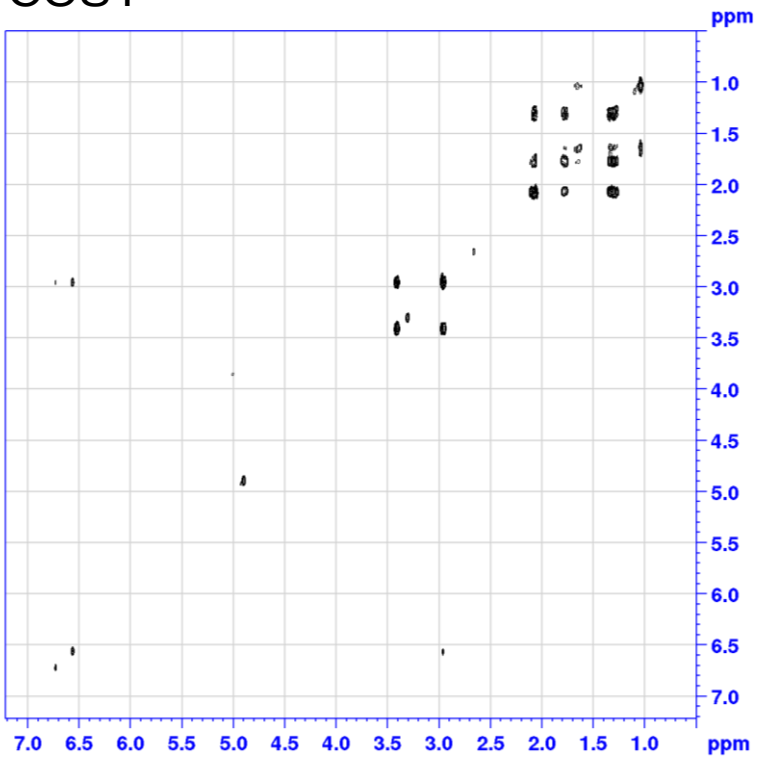
9.2 Spiro-1,1-THIQs

From methods section 2.6.3 and section 6.6.9.

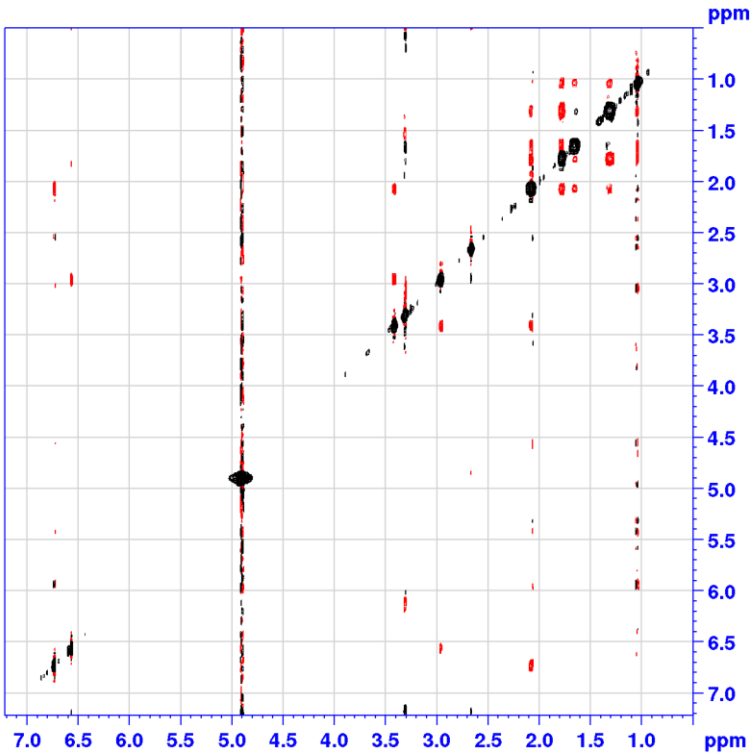
9.2.1 4-Methylcyclohexane



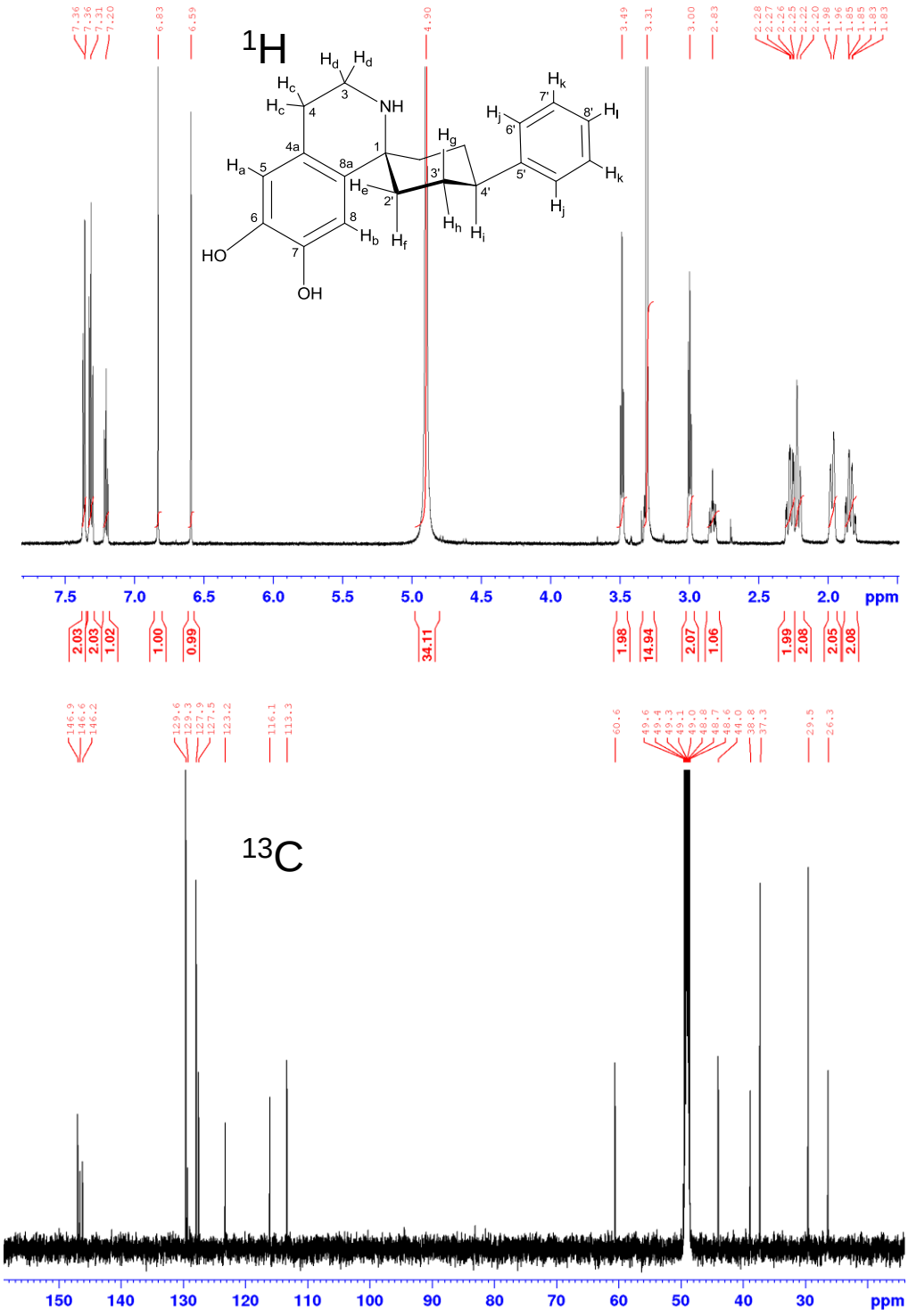
COSY


ppm
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0
ppm

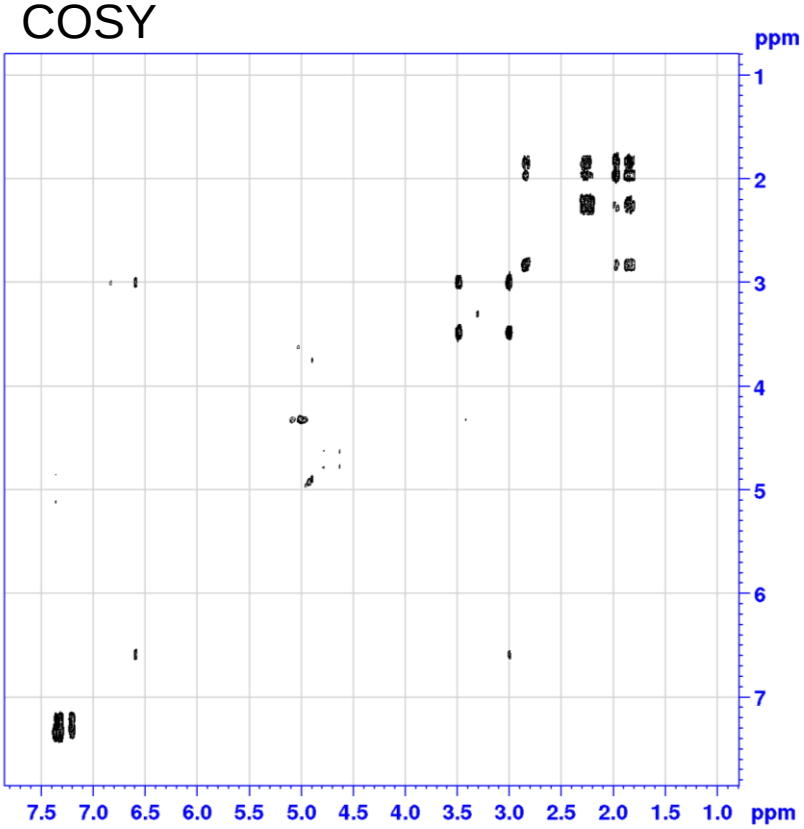
NOESY


ppm
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0
ppm

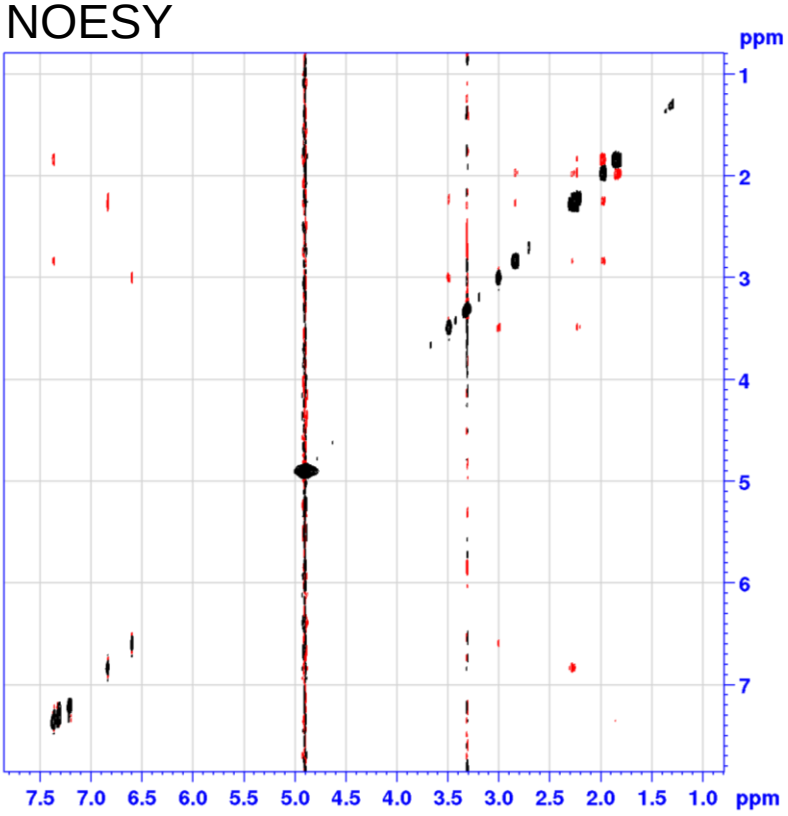
9.2.2 4-Phenylcyclohexane



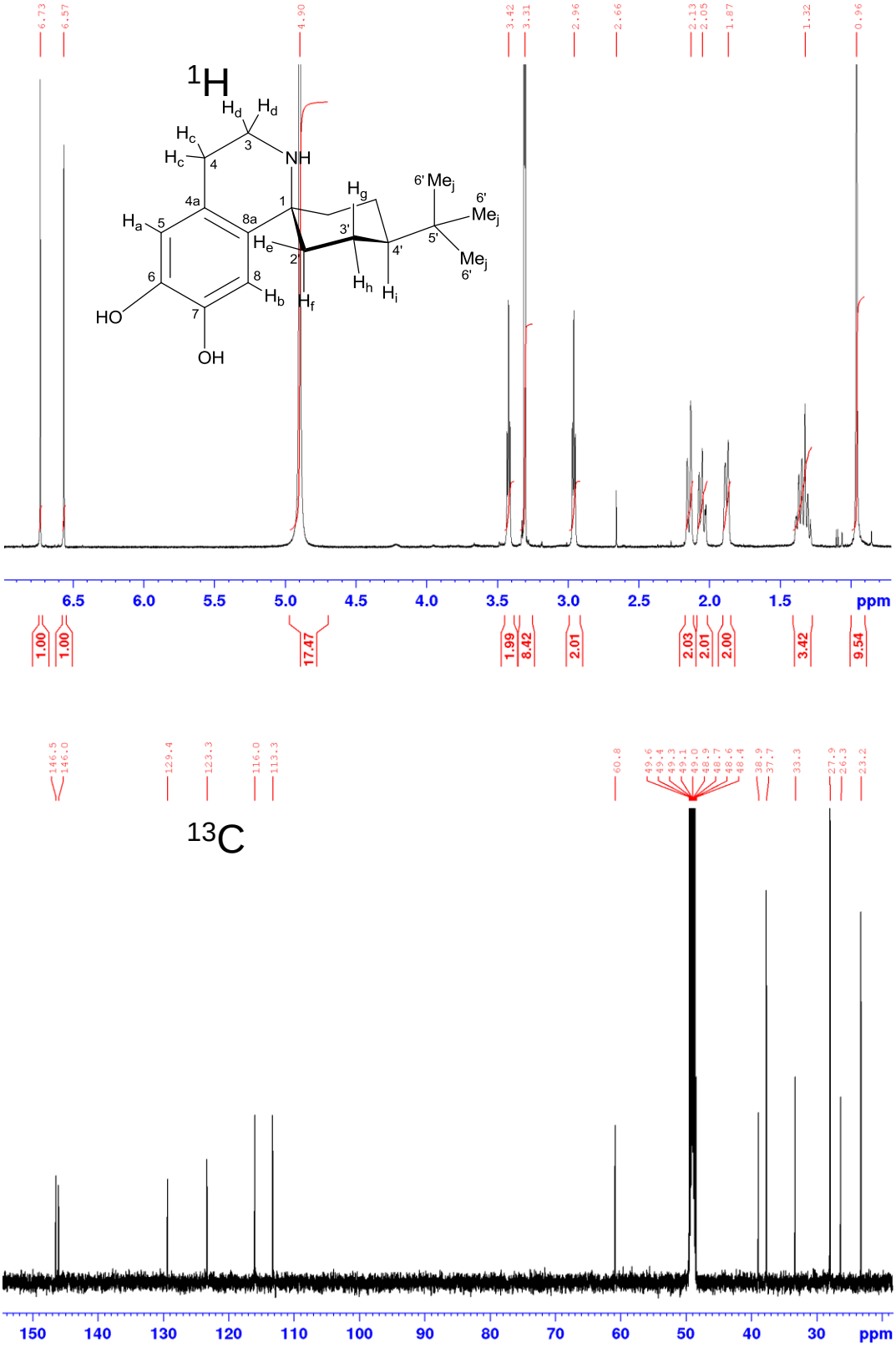
COSY


ppm
1
2
3
4
5
6
7
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm

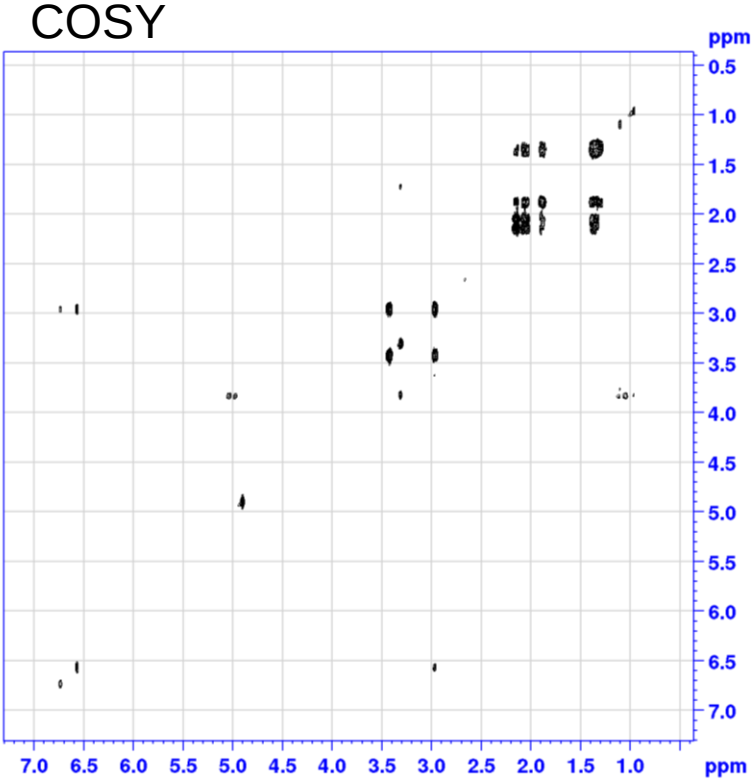
NOESY


ppm
1
2
3
4
5
6
7
7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm

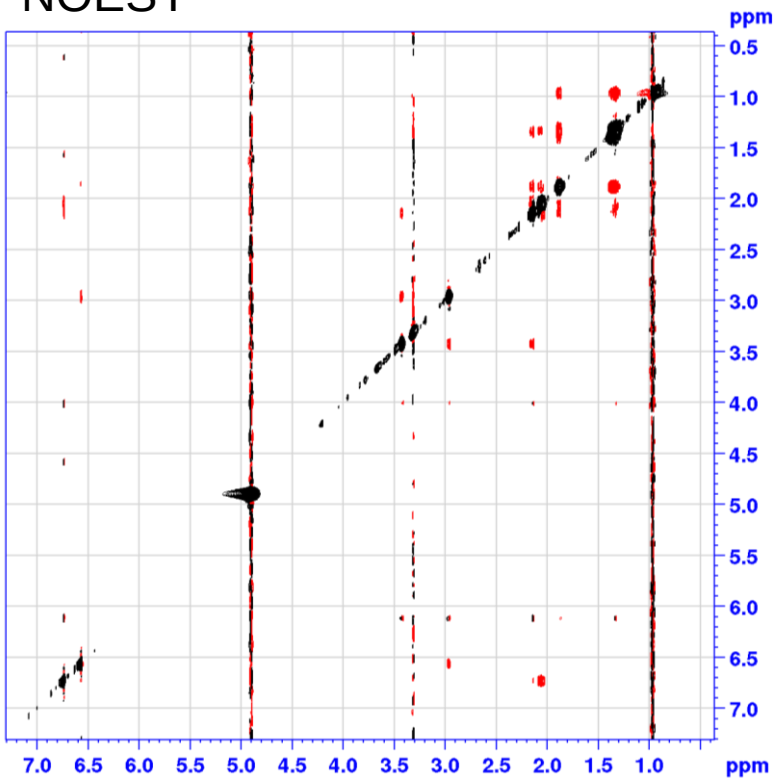
9.2.3 4-*Tert*-butylcyclohexane



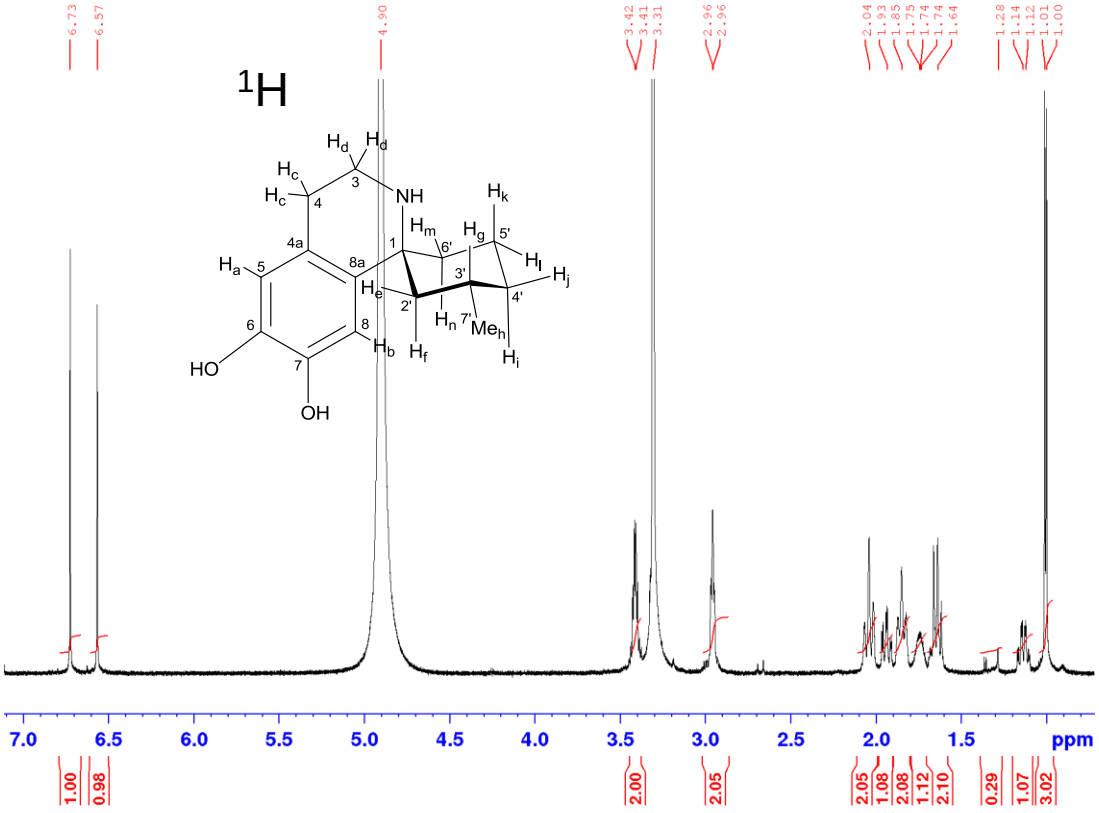
COSY

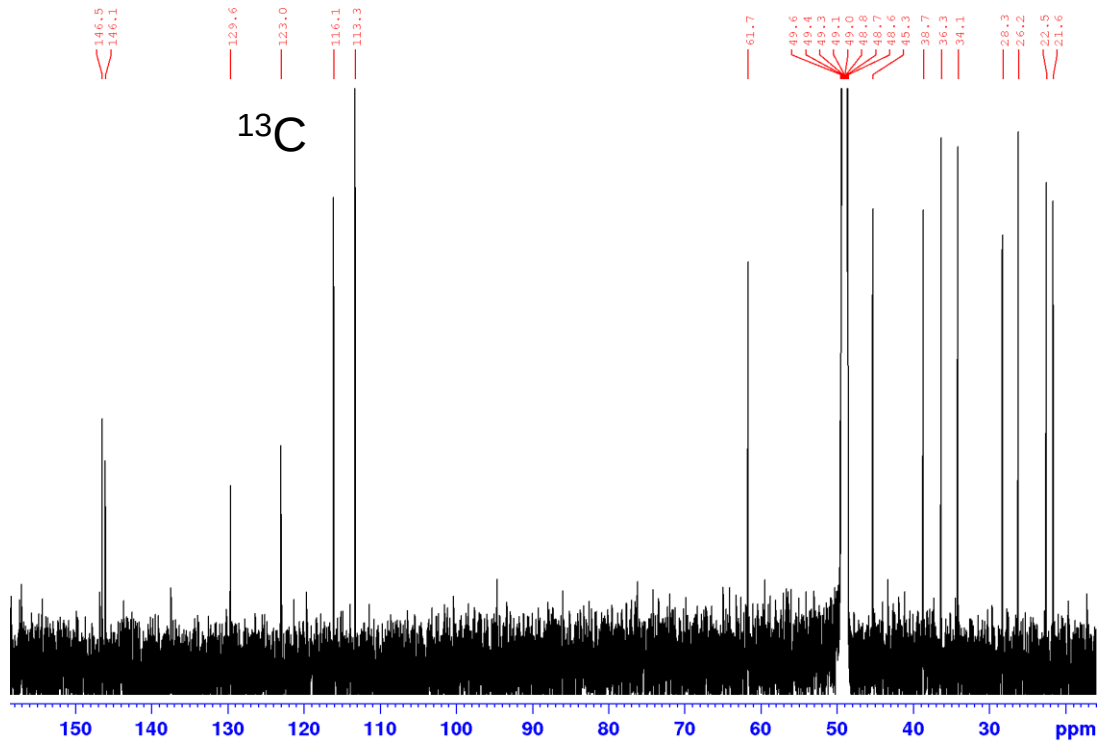

ppm
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0
ppm

NOESY

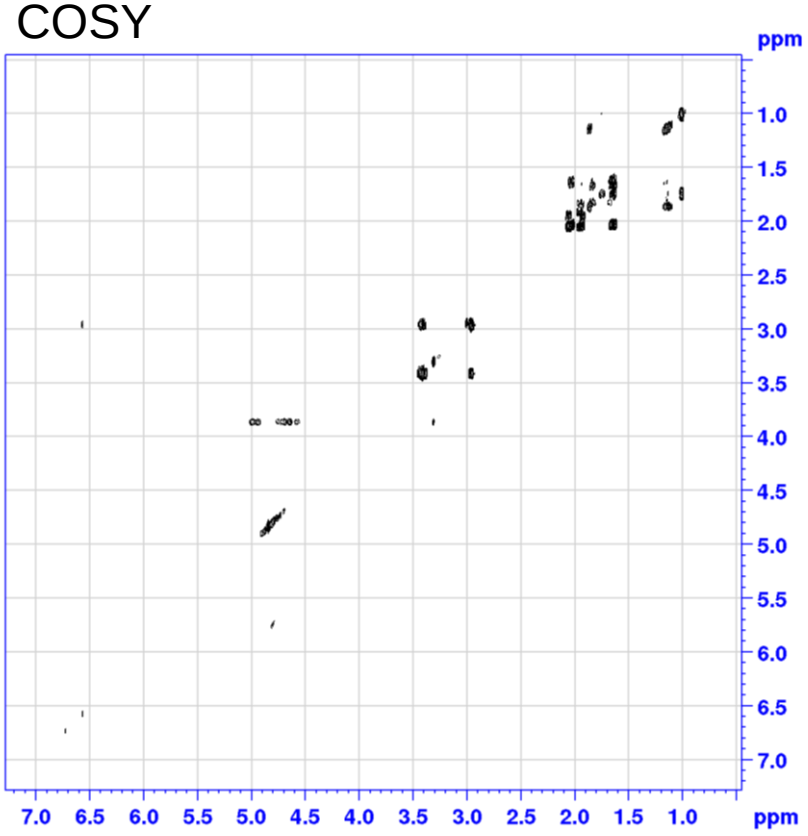

ppm
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0
ppm

9.2.4 *Rac*-3-methylcyclohexane

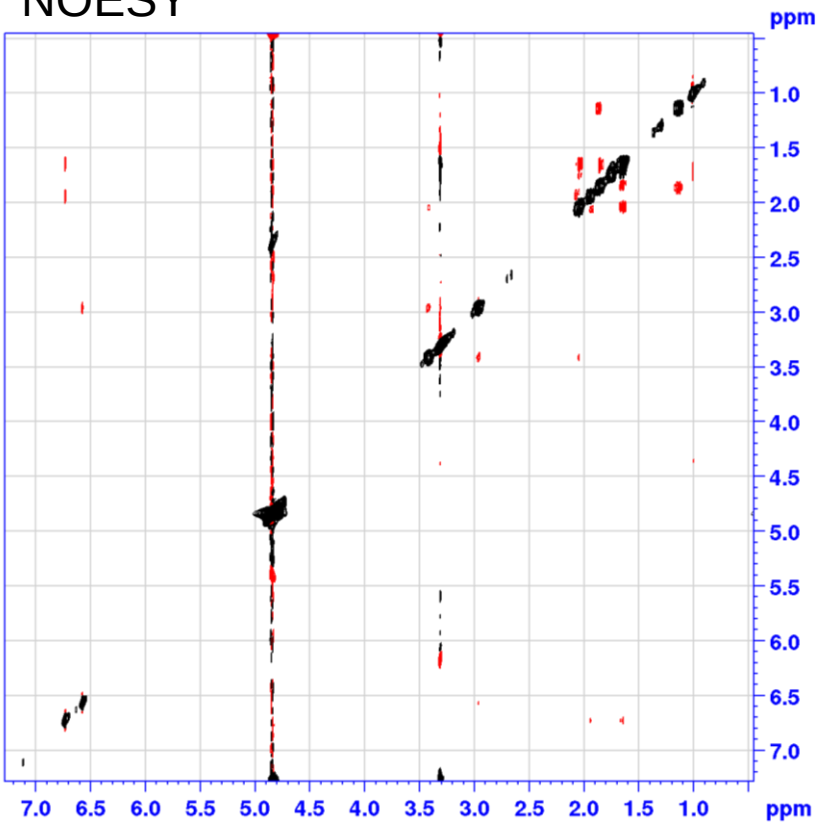




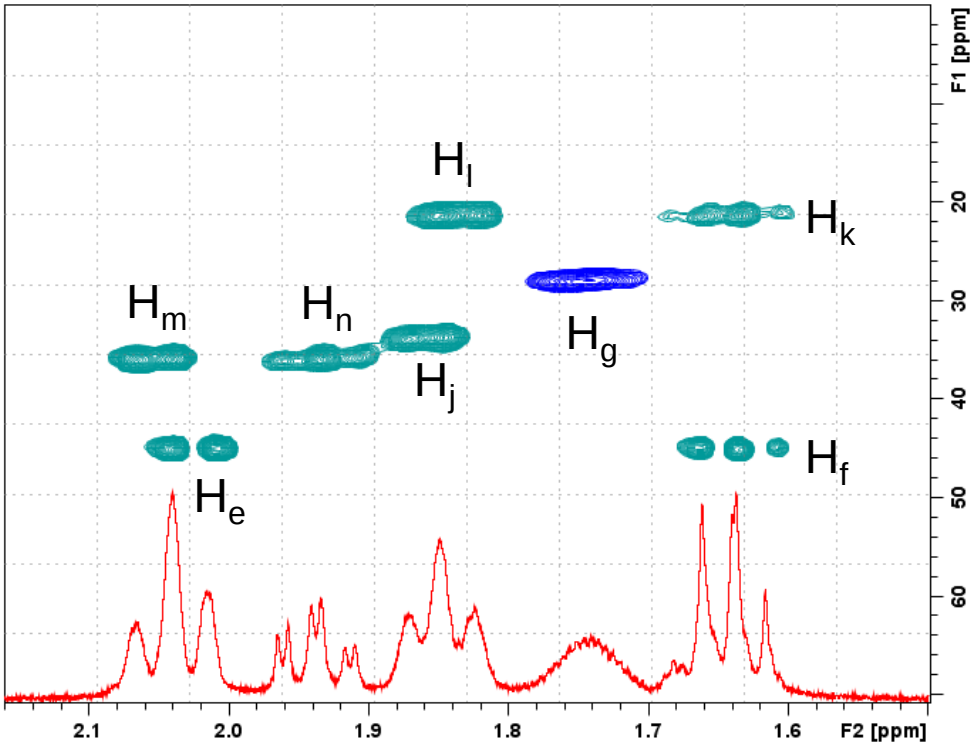
COSY


ppm
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0
ppm

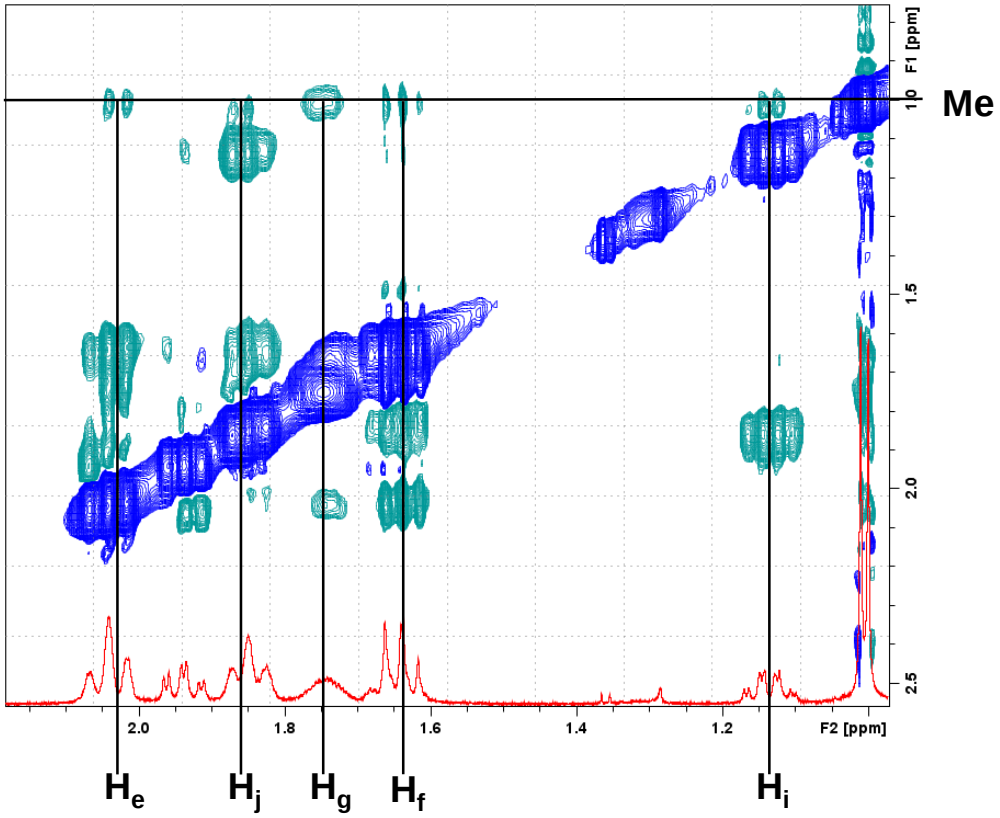
NOESY


ppm
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0
ppm

HSQC signals used to assign protons



NOESY signals for determining methyl position



10 **Bibliography**

1. Weng, J. K. & Noel, J. P. The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* **77,** 309–320 (2012).

2. Nützmann, H. W. & Osbourn, A. Gene clustering in plant specialized metabolism. *Curr. Opin. Biotechnol.* **26,** 91–99 (2014).

3. Weng, J.-K., Philippe, R. N. & Noel, J. P. The Rise of Chemodiversity in Plants. *Science* **336,** 1667–1670 (2012).

4. Nasvall, J., Sun, L., Roth, J. R. & Andersson, D. I. Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. *Science (80-.).* **338,** 384–387 (2012).

5. Bar-even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S. & Milo, R. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50,** 4402–4410 (2011).

6. Hagel, J. M. & Facchini, P. J. Benzylisoquinoline Alkaloid Metabolism: A Century of Discovery and a Brave New World. *Plant Cell Physiol.* **54,** 647–672 (2013).

7. Liscombe, D. K., Macleod, B. P., Loukanina, N., Nandi, O. I. & Facchini, P. J. Evidence for the monophyletic evolution of benzylisoquinoline alkaloid biosynthesis in angiosperms. *Phytochemistry* **66,** 1374–1393 (2005).

8. Singla, D., Sharma, A., Kaur, J., Panwar, B. & Raghava, G. P. BIAdb: A curated database of benzylisoquinoline alkaloids. *BMC Pharmacol.* **10,** 4 (2010).

9. Ziegler, J., Facchini, P. J., Geissler, R., Schmidt, J., Ammer, C., Kramell, R., Voigtländer, S., Gesell, A., Pienkny, S. & Brandt, W. Evolution of morphine biosynthesis in opium poppy. *Phytochemistry* **70,** 1696–1707 (2009).

10. Winzer, T., Gazda, V., He, Z., Kaminski, F., Kern, M., Larson, T. R., Li, Y., Meade, F., Teodor, R., Vaistij, F. E., Walker, C., Bowser, T. A. & Graham, I. A. A *Papaver somniferum* 10-gene cluster for synthesis of the anticancer alkaloid noscapine. *Science* **336,** 1704–1708 (2012).

11. Jabbarzadeh Kaboli, P., Rahmat, A., Ismail, P. & Ling, K.-H. Targets and mechanisms of berberine, a natural drug with potential to treat cancer with special focus on breast cancer. *Eur. J. Pharmacol.* **740,** 584–595 (2014).

12. Gesell, A., Rolf, M., Ziegler, J., Díaz Chávez, M. L., Huang, F.-C. & Kutchan, T. M. CYP719B1 is salutaridine synthase, the C-C phenol-coupling enzyme of morphine biosynthesis in opium poppy. *J. Biol. Chem.* **284,** 24432–24442 (2009).

13. Desgagné-Penix, I., Farrow, S. C., Cram, D., Nowak, J. & Facchini, P. J. Integration of deep transcript and targeted metabolite profiles for eight cultivars of opium poppy. *Plant Mol. Biol.* **79,** 295–313 (2012).

14. Farrow, S. C., Hagel, J. M. & Facchini, P. J. Transcript and metabolite profiling in cell cultures of 18 plant species that produce benzylisoquinoline alkaloids. *Phytochemistry* **77,** 79–88 (2012).

15. Samanani, N. & Facchini, P. J. Purification and characterization of norcoclaurine synthase. The first committed enzyme in benzylisoquinoline alkaloid biosynthesis in plants. *J. Biol. Chem.* **277,** 33878–33883 (2002).

16. Kraus, P. F. & Kutchan, T. M. Molecular cloning and heterologous expression of a cDNA encoding berbamunine synthase, a C-O phenol-coupling cytochrome

P450 from the higher plant *Berberis stolonifera*. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 2071–2075 (1995).

17. Ikezawa, N., Iwasa, K. & Sato, F. Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C-C phenol coupling of (*S*)-reticuline in magnoflorine biosynthesis, from cultured Coptis japonica cells. *J. Biol. Chem.* **283,** 8810–8821 (2008).

18. Bird, D. A. & Facchini, P. J. Berberine bridge enzyme, a key branch-point enzyme in benzylisoquinoline alkaloid biosynthesis, contains a vacuolar sorting determinant. *Planta* **213,** 888–897 (2001).

19. Farrow, S. C., Hagel, J. M., Beaudoin, G. A. W., Burns, D. C. & Facchini, P. J. Stereochemical inversion of (*S*)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat. Chem. Biol.* **11,** 728–732 (2015).

20. Winzer, T., Kern, M., King, A. J., Larson, T. R., Teodor, R. I., Donninger, S. L., Li, Y., Dowle, A. A., Cartwright, J., Bates, R., Ashford, D., Thomas, J., Walker, C., Bowser, T. A. & Graham, I. A. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* **349,** 309–312 (2015).

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

22. Pictet, A. & Spengler, T. On the formation of isoquinoline derivatives by the action of methylal on Phenylethylamine, Phenylalanine and tyrosine. *Ber. Dtsch. Chem. Ges.* **44,** 2030–2036 (1911).

23. Schopf, C. & Salzer, W. On the question of biogenesis of 1-benzyl-1,2,3,4-tetrahydroisoquinoline alkaloids. *Liebigs Ann.* **544,** 1–30 (1940).

24. Stöckigt, J., Antonchick, A. P., Wu, F. & Waldmann, H. The Pictet-Spengler reaction in nature and in Organic Chemistry. *Angew. Chem. Int. Ed.* **50,** 8538–8564 (2011).

25. Rueffer, M., El-shagi, H., Nagakura, N. & Zenk, M. H. (*S*)-Norlaudanosoline synthase: The first enzyme in the benzylisoquinoline biosynthetic pathway. *FEBS Lett.* **129,** 5–9 (1981).

26. Stadler, R., Kutchan, T. M., Loeffler, S., Nagakura, N., Cassels, B. & Zenk, M. H. Revision of the early steps of Reticuline Biosynthesis. *Tetrahedron Lett.* **28,** 1251–1254 (1987).

27. Battersby, A. R., Binks, R., Francis, R. J., McCaldin, D. J. & Ramuz, H. Alkaloid Biosynthesis. Part IV. 1-Benzylisoquinolines as precursors of Thebaine, Codeine, and Morphine. *J. Chem. Soc* 3600–3610 (1963).

28. Lorenz, T., Legge, R. L. & Moo-Young, M. Production of morphine alkaloids: (S)-norlaudanosoline, a key intermediate. *Enzyme Microb. Technol.* **10,** 219–226 (1988).

29. Hoover, L. K., Moo-Young, M. & Legge, R. L. Biotransformation of Dopamine to Norlaudanosoline by *Aspergillus niger*. *Biotechnol. Bioeng.* **38,** 1029–1033 (1991).

30. Schumacher, H. M., Ruffer, M., Nagakura, N. & Zenk, M. H. Partial Purification and Properties of (*S*)-Norlaudanosoline Synthase from *Eschscholtzia tenuifolia* Cell Cultures. *Planta Med.* **48,** 212–220 (1983).

31. Stadler, R., Kutchan, T. M. & Zenk, H. (*S*)-Norcoclaurine is the central intermediate in benzylisoquinoline alkaloid biosynthesis. *Phytochemistry* **28,**

1083–1086 (1989).

32. Stadler, R. & Zenk, M. H. A revision of the generally accepted pathway for the biosynthesis of the benzyltetrahydroisoquinoline alkaloid reticuline. *Liebigs Ann. der Chemie* **1990,** 555–562 (1990).

33. Samanani, N. & Facchini, P. J. Isolation and partial characterization of norcoclaurine synthase, the first committed step in benzylisoquinoline alkaloid biosynthesis, from opium poppy. *Planta* **213,** 898–906 (2001).

34. Samanani, N., Liscombe, D. K. & Facchini, P. J. Molecular cloning and characterization of norcoclaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. *Plant J.* **40,** 302–313 (2004).

35. Bais, H. P., Vepachedu, R., Lawrence, C. B., Stermitz, F. R. & Vivanco, J. M. Molecular and biochemical characterization of an enzyme responsible for the formation of hypericin in St. John's Wort (*Hypericum perforatum* L.). *J. Biol. Chem.* **278,** 32413–32422 (2003).

36. Košuth, J., Katkovčinová, Z., Olexová, P. & Čellárová, E. Expression of the *hyp-1* gene in early stages of development of *Hypericum perforatum* L. *Plant Cell Rep.* **26,** 211–217 (2007).

37. Kusari, S., Zu, S., Spiteller, M., Africa, N. & America, N. Light-Independent Metabolomics of Endophytic *Thielavia subthermophila* Provides Insight into Microbial Hypericin Biosynthesis. *J. Nat. Prod.* **72,** 1825–1835 (2009).

38. Michalska, K., Fernandes, H., Sikorski, M. & Jaskolski, M. Crystal structure of Hyp-1, a St. John's Wort protein implicated in the biosynthesis of hypericin. *J. Struct. Biol.* **169,** 161–171 (2010).

39. Fernandes, H., Michalska, K., Sikorski, M. & Jaskolski, M. Structural and functional aspects of PR-10 proteins. *FEBS J.* **280,** 1169–1199 (2013).

40. Marković-Housley, Z., Degano, M., Lamba, D., Von Roepenack-Lahaye, E., Clemens, S., Susani, M., Ferreira, F., Scheiner, O. & Breiteneder, H. Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. *J. Mol. Biol.* **325,** 123–133 (2003).

41. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol.* **8,** 286 (2008).

42. Berkner, H., Loetzen, C. S. Von, Hartl, M., Randow, S., Gubesch, M., Vogel, L., Husslik, F., Reuter, A., Lidholm, J., Schiller, D., Ballmer-Weber, B., Vieths, S. & Ro, P. Enlarging the Toolbox for Allergen Epitope Definition with an Allergen-Type Model Protein. *PLoS One* **9,** e111691 (2014).

43. Ilari, A., Franceschini, S., Bonamore, A., Arenghi, F., Botta, B., Macone, A., Pasquo, A., Bellucci, L. & Boffi, A. Structural basis of enzymatic (*S*)-norcoclaurine biosynthesis. *J. Biol. Chem.* **284,** 897–904 (2009).

44. Lee, E.-J. & Facchini, P. Norcoclaurine synthase is a member of the pathogenesis-related 10/Bet v1 protein family. *Plant Cell* **22,** 3489–3503 (2010).

45. DeLoache, W. C., Russ, Z. N., Narcross, L., Gonzales, A. M., Martin, V. J. J. & Dueber, J. E. An enzyme-coupled biosensor enables (*S*)-reticuline production in yeast from glucose. *Nat. Chem. Biol.* **11,** 465–471 (2015).

46. Minami, H., Dubouzet, E., Iwasa, K. & Sato, F. Functional analysis of

norcoclaurine synthase in *Coptis japonica*. *J. Biol. Chem.* **282,** 6274–6282 (2007).

47. Zhang, X., Ma, L., Tian, Y., Zhang, G. & Luo, Y. Molecular cloning and heterologous expression of putative (*S*)-norcoclaurine synthases from *Arabidopsis thaliana*. *Chinese J. Appl. Environ. Biol.* **19,** 61–68 (2013).

48. Facchini, P. J., Bird, D. A. & St-Pierre, B. Can Arabidopsis make complex alkaloids? *Trends Plant Sci.* **9,** 116–122 (2004).

49. Xiao, M., Zhang, Y., Chen, X., Lee, E.-J., Barber, C. J. S., Chakrabarty, R., Desgagné-Penix, I., Haslam, T. M., Kim, Y.-B., Liu, E., MacNevin, G., Masada-Atsumi, S., Reed, D. W., Stout, J. M., Zerbe, P., Zhang, Y., Bohlmann, J., Covello, P. S., De Luca, V., Page, J. E., Ro, D.-K., Martin, V. J. J., Facchini, P. J. & Sensen, C. W. Transcriptome analysis based on next-generation sequencing of non-model plants producing specialized metabolites of biotechnological interest. *J. Biotechnol.* **166,** 122–134 (2013).

50. Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E. J., Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Barker, M., Burleigh, J. G., Gitzendanner, M. a, Wafula, E., Der, J. P., dePamphilis, C. W., Roure, B., Philippe, H., Ruhfel, B. R., Miles, N. W., Graham, S. W., Mathews, S., Surek, B., Melkonian, M., Soltis, D. E., Soltis, P. S., Rothfels, C., Pokorny, L., Shaw, J. a, DeGironimo, L., Stevenson, D. W., Villarreal, J. C., Chen, T., Kutchan, T. M., Rolf, M., Baucom, R. S., Deyholos, M. K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J. & Wong, G. K.-S. Data access for the 1,000 Plants (1KP) project. *Gigascience* **3,** 17 (2014).

51. Luk, L. Y. P., Bunn, S., Liscombe, D. K., Facchini, P. J. & Tanner, M. E. Mechanistic studies on norcoclaurine synthase of benzylisoquinoline alkaloid biosynthesis: an enzymatic Pictet-Spengler reaction. *Biochemistry* **46,** 10153–10161 (2007).

52. Berkner, H., Engelhorn, J., Liscombe, D. K., Schweimer, K., Wöhrl, B. M., Facchini, P. J., Rösch, P. & Matecko, I. High-yield expression and purification of isotopically labeled norcoclaurine synthase, a Bet v 1-homologous enzyme, from *Thalictrum flavum* for NMR studies. *Protein Expres. Purif.* **56,** 197–204 (2007).

53. Ruff, B. M., Bräse, S. & O'Connor, S. E. Biocatalytic production of tetrahydroisoquinolines. *Tetrahedron Lett.* **53,** 1071–1074 (2012).

54. Berkner, H., Schweimer, K., Matecko, I. & Rösch, P. Conformation, catalytic site, and enzymatic mechanism of the PR10 allergen-related enzyme norcoclaurine synthase. *Biochem. J.* **413,** 281–290 (2008).

55. Pasquo, A., Bonamore, A., Franceschini, S., Macone, A., Boffi, A. & Ilari, A. Cloning, expression, crystallization and preliminary X-ray data analysis of norcoclaurine synthase from *Thalictrum flavum*. *Acta Crystallogr.* **F64,** 281–283 (2008).

56. Bonamore, A., Barba, M., Botta, B., Boffi, A. & Macone, A. Norcoclaurine synthase: mechanism of an enantioselective pictet-spengler catalyzing enzyme. *Molecules* **15,** 2070–2078 (2010).

57. Pesnot, T., Gershater, M. C., Ward, J. M. & Hailes, H. C. The catalytic potential of *Coptis japonica* NCS2 revealed - Development and utilisation of a fluorescamine-based assay. *Adv. Synth. Catal.* **354,** 2997–3008 (2012).

58. Bonamore, A., Rovardi, I., Gasparrini, F., Baiocco, P., Barba, M., Molinaro, C., Botta, B., Boffi, A. & Macone, A. An enzymatic, stereoselective synthesis of (*S*)-

norcoclaurine. *Green Chem.* **12,** 1623–1627 (2010).

59. Pesnot, T., Gershater, M. C., Ward, J. M. & Hailes, H. C. Phosphate mediated biomimetic synthesis of tetrahydroisoquinoline alkaloids. *Chem. Commun.* **47,** 3242–3244 (2011).

60. Maresh, J. J., Crowe, S. O., Ralko, A. A., Aparece, M. D., Murphy, C. M., Krzeszowiec, M. & Mullowney, M. W. Facile one-pot synthesis of tetrahydroisoquinolines from amino acids via hypochlorite-mediated decarboxylation and Pictet–Spengler condensation. *Tetrahedron Lett.* **55,** 5047–5051 (2014).

61. Nishihachijo, M., Hirai, Y., Kawano, S., Nishiyama, A., Minami, H., Katayama, T., Yasohara, Y., Sato, F. & Kumagai, H. Asymmetric synthesis of tetrahydroisoquinolines by enzymatic Pictet-Spengler reaction. *Biosci. Biotechnol. Biochem.* **78,** 701–707 (2014).

62. Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B. & Keasling, J. D. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463,** 559–562 (2010).

63. Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. a, Eachus, R. a, Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R. & Keasling, J. D. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440,** 940–943 (2006).

64. Hawkins, K. M. & Smolke, C. D. Production of benzylisoquinoline alkaloids in *Saccharomyces cerevisiae*. *Nat. Chem. Biol.* **4,** 564–573 (2008).

65. Thodey, K., Galanie, S. & Smolke, C. D. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat. Chem. Biol.* **10,** 837–844 (2014).

66. Fossati, E., Ekins, A., Narcross, L., Zhu, Y., Falgueyret, J.-P., Beaudoin, G. A. W., Facchini, P. J. & Martin, V. J. J. Reconstitution of a 10-gene pathway for synthesis of the plant alkaloid dihydrosanguinarine in *Saccharomyces cerevisiae*. *Nat. Commun.* **5,** 3283 (2014).

67. Trenchard, I. J. & Smolke, C. D. Engineering strategies for the fermentative production of plant alkaloids in yeast. *Metab. Eng.* **30,** 96–104 (2015).

68. Fossati, E., Narcross, L., Ekins, A., Falgueyret, J.-P. & Martin, V. J. J. Synthesis of Morphinan Alkaloids in *Saccharomyces cerevisiae*. *PLoS One* **10,** e0124459 (2015).

69. Onoyovwe, A., Hagel, J. M., Chen, X., Khan, M. F., Schriemer, D. C. & Facchini, P. J. Morphine biosynthesis in opium poppy involves two cell types: sieve elements and laticifers. *Plant Cell* **25,** 4110–4122 (2013).

70. Holtz, P., Stock, K. & Westermann, E. Formation of Tetrahydropapaveroline from Dopamine in vitro. *Nature* **203,** 656–658 (1964).

71. Nakagawa, A., Matsuzaki, C., Matsumura, E., Koyanagi, T., Katayama, T., Yamamoto, K., Sato, F., Kumagai, H. & Minami, H. (*R*,*S*)-Tetrahydropapaveroline production by stepwise fermentation using engineered *Escherichia coli*. *Sci. Rep.* **4,** 6695 (2014).

72. Minami, H., Kim, J.-S., Ikezawa, N., Takemura, T., Katayama, T., Kumagai, H. & Sato, F. Microbial production of plant benzylisoquinoline alkaloids. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7393–7398 (2008).

73. Nakagawa, A., Minami, H., Kim, J.-S., Koyanagi, T., Katayama, T., Sato, F. &

Kumagai, H. A bacterial platform for fermentative production of plant alkaloids. *Nat. Commun.* **2,** 326 (2011).

74. Nakagawa, A., Minami, H., Kim, J.-S., Koyanagi, T., Katayama, T., Sato, F. & Kumagai, H. Bench-top fermentative production of plant benzylisoquinoline alkaloids using a bacterial platform. *Bioeng. Bugs* **3,** 49–53 (2012).

75. Kim, J.-S., Nakagawa, A., Yamazaki, Y., Matsumura, E., Koyanagi, T., Minami, H., Katayama, T., Sato, F. & Kumagai, H. Improvement of reticuline productivity from dopamine by using engineered *Escherichia coli*. *Biosci. Biotechnol. Biochem.* **77,** 2166–2168 (2013).

76. Trenchard, I. J., Siddiqui, M. S., Thodey, K. & Smolke, C. D. De novo production of the key branch point benzylisoquinoline alkaloid reticuline in yeast. *Metab. Eng.* **31,** 74–83 (2015).

77. Stöckigt, J., Barleben, L., Panjikar, S. & Loris, E. A. 3D-Structure and function of strictosidine synthase--the key enzyme of monoterpenoid indole alkaloid biosynthesis. *Plant Physiol. Biochem.* **46,** 340–355 (2008).

78. Hicks, M. A., Barber, A. E., Giddings, L. A., Caldwell, J., Connor, S. E. & Babbitt, P. C. The evolution of function in strictosidine synthase-like proteins. *Proteins Struct. Funct. Bioinforma.* **79,** 3082–3098 (2011).

79. Maresh, J. J., Giddings, L. A., Friedrich, A., Loris, E. A., Panjikar, S., Trout, B. L., Stöckigt, J., Peters, B. & O'Connor, S. E. Strictosidine synthase: Mechanism of a Pictet-Spengler catalyzing enzyme. *J. Am. Chem. Soc.* **130,** 710–723 (2008).

80. Chen, S., Galan, M. C., Coltharp, C. & O'Connor, S. E. Redesign of a central enzyme in alkaloid biosynthesis. *Chem. Biol.* **13,** 1137–1141 (2006).

81. Fischereder, E., Pressnitz, D., Kroutil, W. & Lutz, S. Engineering strictosidine synthase: Rational design of a small, focused circular permutation library of the β-propeller fold enzyme. *Bioorganic Med. Chem.* **22,** 5633–5637 (2014).

82. McCoy, E., Galan, M. C. & O'Connor, S. E. Substrate specificity of strictosidine synthase. *Bioorganic Med. Chem. Lett.* **16,** 2475–2478 (2006).

83. Wu, F., Zhu, H., Sun, L., Rajendran, C., Wang, M., Ren, X., Panjikar, S., Cherkasov, A., Zou, H. & Stöckigt, J. Scaffold tailoring by a newly detected Pictet-Spenglerase activity of strictosidine synthase: from the common tryptoline skeleton to the rare piperazino-indole framework. *J. Am. Chem. Soc.* **134,** 1498–1500 (2012).

84. Yan, W., Ge, H. M., Wang, G., Jiang, N., Mei, Y. N., Jiang, R., Li, S. J., Chen, C. J., Jiao, R. H., Xu, Q., Ng, S. W. & Tan, R. X. Pictet–Spengler reaction-based biosynthetic machinery in fungi. *Proc. Natl. Acad. Sci.* **111,** 18138–18143 (2014).

85. De-Eknamkul, W., Ounaroon, A., Tanahashi, T., Kutchan, T. M. & Zenk, M. H. Enzymatic condensation of dopamine and secologanin by cell-free extracts of *Alangium lamarckii*. *Phytochemistry* **45,** 477–484 (1997).

86. De-Eknamkul, W., Suttipanta, N. & Kutchan, T. M. Purification and characterization of deacetylipecoside synthase from Alangium lamarckii Thw. *Phytochemistry* **55,** 177–181 (2000).

87. Pakrashi, S. C., Mukhopadhyay, R., Dastidar, P. P. G., Bhattacharjya, A. & Ali, E. Bharatamine - a unique protoberberine alkaloid from *Alangium lamarckii* Thw., biogenetically derived from monoterpenoid precursor. *Tetrahedron Lett.* **24,** 291–294 (1983).

88. Koketsu, K., Watanabe, K., Suda, H., Oguri, H. & Oikawa, H. Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms. *Nat. Chem. Biol.* **6,** 408–410 (2010).

89. Koketsu, K., Minami, A., Watanabe, K., Oguri, H. & Oikawa, H. Pictet-Spenglerase involved in tetrahydroisoquinoline antibiotic biosynthesis. *Curr. Opin. Chem. Biol.* **16,** 142–149 (2012).

90. Chen, Q., Ji, C., Song, Y., Huang, H., Ma, J., Tian, X. & Ju, J. Discovery of McbB, an Enzyme Catalyzing the β-Carboline Skeleton Construction in the Marinacarboline Biosynthetic Pathway. *Angew. Chem. Int. Ed.* **52,** 9980–9984 (2013).

91. Mori, T., Hoshino, S., Sahashi, S., Wakimoto, T., Matsui, T., Morita, H. & Abe, I. Structural Basis for β-Carboline Alkaloid Production by the Microbial Homodimeric Enzyme McbB. *Chem. Biol.* **22,** 898–906 (2015).

92. Welsch, M. E., Snyder, S. a. & Stockwell, B. R. Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.* **14,** 347–361 (2010).

93. Vanden Eynden, M. J., Kunchithapatham, K. & Stambuli, J. P. Calcium-promoted Pictet-Spengler reactions of ketones and aldehydes. *J. Org. Chem.* **75,** 8542–8549 (2010).

94. Guzman, J. D., Pesnot, T., Barrera, D. A., Davies, H. M., McMahon, E., Evangelopoulos, D., Mortazavi, P. N., Munshi, T., Maitra, A., Lamming, E. D., Angell, R., Gershater, M. C., Redmond, J. M., Needham, D., Ward, J. M., Cuca, L. E., Hailes, H. C. & Bhakta, S. Tetrahydroisoquinolines affect the whole-cell phenotype of *Mycobacterium tuberculosis* by inhibiting the ATP-dependent MurE ligase. *J. Antimicrob. Chemother.* **70,** 1691–1703 (2015).

95. Mons, E., Wanner, M. J., Ingemann, S., Van Maarseveen, J. H. & Hiemstra, H. Organocatalytic enantioselective pictet-spengler reactions for the syntheses of 1-substituted 1,2,3,4-tetrahydroisoquinolines. *J. Org. Chem.* **79,** 7380–7390 (2014).

96. Schrittwieser, J. H. & Resch, V. The role of biocatalysis in the asymmetric synthesis of alkaloids. *RSC Adv.* **3,** 17602 (2013).

97. Bailey, K. R., Ellis, A. J., Reiss, R., Snape, T. J. & Turner, N. J. A template-based mnemonic for monoamine oxidase (MAO-N) catalyzed reactions and its application to the chemo-enzymatic deracemisation of the alkaloid (+/-)-crispine A. *Chem. Commun. (Camb).* **44,** 3640–3642 (2007).

98. Schrittwieser, J. H., Groenendaal, B., Willies, S. C., Ghislieri, D., Rowles, I., Resch, V., Sattler, J. H., Fischereder, E., Grischek, B., Lienhart, W., Turner, N. J. & Kroutil, W. Deracemisation of benzylisoquinoline alkaloids employing monoamine oxidase variants. *Catal. Sci. Technol.* **4,** 3657–3664 (2014).

99. Resch, V., Schrittwieser, J. H., Wallner, S., Macheroux, P. & Kroutil, W. Biocatalytic Oxidative C-C Bond Formation Catalysed by the Berberine Bridge Enzyme: Optimal Reaction Conditions. *Adv. Synth. Catal.* **353,** 2377–2383 (2011).

100. Schrittwieser, J. H., Resch, V., Sattler, J. H., Lienhart, W.-D., Durchschein, K., Winkler, A., Gruber, K., Macheroux, P. & Kroutil, W. Biocatalytic enantioselective oxidative C-C coupling by aerobic C-H activation. *Angew. Chem. Int. Ed.* **50,** 1068–1071 (2011).

101. Schrittwieser, J. H., Resch, V., Wallner, S., Lienhart, W.-D., Sattler, J. H., Resch,

J., Macheroux, P. & Kroutil, W. Biocatalytic organic synthesis of optically pure (S)-scoulerine and berbine and benzylisoquinoline alkaloids. *J. Org. Chem.* **76,** 6703–6714 (2011).

102. Resch, V., Lechner, H., Schrittwieser, J. H., Wallner, S., Gruber, K., Macheroux, P. & Kroutil, W. Inverting the regioselectivity of the berberine bridge enzyme by employing customized fluorine-containing substrates. *Chem. Eur. J.* **18,** 13173–13179 (2012).

103. Schrittwieser, J. H., Groenendaal, B., Resch, V., Ghislieri, D., Wallner, S., Fischereder, E.-M., Fuchs, E., Grischek, B., Sattler, J. H., Macheroux, P., Turner, N. J. & Kroutil, W. Deracemization by simultaneous bio-oxidative kinetic resolution and stereoinversion. *Angew. Chem. Int. Ed.* **53,** 3731–3734 (2014).

104. Kohls, H., Steffen-Munsberg, F. & Höhne, M. Recent achievements in developing the biocatalytic toolbox for chiral amine synthesis. *Curr. Opin. Chem. Biol.* **19,** 180–192 (2014).

105. Scheller, P. N., Fademrecht, S., Hofelzer, S., Pleiss, J., Leipold, F., Turner, N. J., Nestl, B. M. & Hauer, B. Enzyme Toolbox: Novel Enantiocomplementary Imine Reductases. *ChemBioChem* **15,** 2201–2204 (2014).

106. Hussain, S., Leipold, F., Man, H., Wells, E., France, S. P., Mulholland, K. R., Grogan, G. & Turner, N. J. An (*R*)-Imine Reductase Biocatalyst for the Asymmetric Reduction of Cyclic Imines. *ChemCatChem* **7,** 579–583 (2015).

107. Dalby, P. A. Strategy and success for the directed evolution of enzymes. *Curr. Opin. Stuc. Biol.* **21,** 473–480 (2011).

108. Brustad, E. M. & Arnold, F. H. Optimizing non-natural protein function with directed evolution. *Curr. Opin. Chem. Biol.* **15,** 201–210 (2011).

109. Goldsmith, M. & Tawfik, D. S. Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* **22,** 406–412 (2012).

110. Lane, M. D. & Seelig, B. Advances in the directed evolution of proteins. *Curr. Opin. Chem. Biol.* **22,** 129–136 (2014).

111. Currin, A., Swainston, N., Day, P. J. & Kell, D. B. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **44,** 1172–1239 (2014).

112. Gustafsson, C., Minshull, J., Govindarajan, S., Ness, J., Villalobos, A. & Welch, M. Engineering genes for predictable protein expression. *Protein Expr. Purif.* **83,** 37–46 (2012).

113. Sievers, F. & Higgins, D. G. Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol. Biol.* **1079,** 105–116 (2014).

114. Kaulmann, U., Smithies, K., Smith, M. E. B., Hailes, H. C. & Ward, J. M. Substrate spectrum of ω-transaminase from *Chromobacterium violaceum* DSM30191 and its potential for biocatalysis. *Enzyme Microb. Technol.* **41,** 628–637 (2007).

115. Shin, J.-S., Yun, H., Jang, J.-W., Park, I. & Kim, B.-G. Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from *Vibrio fluvialis* JS17. *Appl. Microbiol. Biotechnol.* **61,** 463–471 (2003).

116. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B. & Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular

simulation toolkit. *Bioinformatics* **29,** 845–854 (2013).

117. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25,** 1400–1415 (2004).

118. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31,** 455–461 (2010).

119. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p*K*a Values. *J. Chem. Theory Comput.* **7,** 2284–2295 (2011).

120. Bryan, A. W., Menke, M., Cowen, L. J., Lindquist, S. L. & Berger, B. BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis. *PLoS Comput. Biol.* **5,** e1000333 (2009).

121. Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

122. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 213–221 (2010).

123. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607 (2005). doi:10.1385/1-59259-890-0:571

124. Lucas, R., Alcantara, D. & Morales, J. C. A concise synthesis of glucuronide metabolites of urolithin-B, resveratrol, and hydroxytyrosol. *Carbohydr. Res.* **344,** 1340–1346 (2009).

125. Li, S. W., Spaziano, V. T. & Burke, W. J. Synthesis of a Biochemically Important Aldehyde, 3,4-Dihydroxyphenylacetaldehyde. *Bioorg. Chem.* **26,** 45–50 (1998).

126. Teitel, S., O'Brien, J. & Brossi, A. Alkaloids in Mammalian Tissues. 2.' Synthesis of (+)- and (-)-I Substituted-6,7-dihydroxyl-1,2,3,4-tetrahydroisoquinolines. *J. Med. Chem.* **15,** 845–846 (1972).

127. Guengerich, F. P., Miller, G. P., Hanna, I. H., Sato, H. & Martin, M. V. Oxidation of methoxyphenethylamines by cytochrome P450 2D6. Analysis of rate-limiting steps. *J. Biol. Chem.* **277,** 33711–33719 (2002).

128. Tahirovic, Y. A., Geballe, M., Gruszecka-Kowalik, E., Myers, S. J., Lyuboslavsky, P., Le, P., French, A., Irier, H., Choi, W. B., Easterling, K., Yuan, H., Wilson, L. J., Kotloski, R., McNamara, J. O., Dingledine, R., Liotta, D. C., Traynelis, S. F. & Snyder, J. P. Enantiomeric propanolamines as selective N-methyl-D-aspartate 2B receptor antagonists. *J. Med. Chem.* **51,** 5506–5521 (2008).

129. Cui, W., Iwasa, K., Tokuda, H., Kashihara, A., Mitani, Y., Hasegawa, T., Nishiyama, Y., Moriyasu, M., Nishino, H., Hanaoka, M., Mukai, C. & Takeda, K. Potential cancer chemopreventive activity of simple isoquinolines, 1-benzylisoquinolines, and protoberberines. *Phytochemistry* **67,** 70–79 (2006).

130. Bolanos-Garcia, V. M. & Davies, O. R. Structural analysis and classification of native proteins from E. coli commonly co-purified by immobilised metal affinity chromatography. *Biochim. Biophys. Acta* **1760,** 1304–1313 (2006).

131. Petersen, T. N., Brunak, S., von Heijne, G. & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8,** 785–786 (2011).

132. Nakamura, K. & Matsuoka, K. Protein targeting to the vacuole in plant cells. *Plant Physiol.* **101,** 1–5 (1993).

133. Gomi, M., Sonoyama, M. & Mitaku, S. High performance system for signal peptide prediction: SOSUIsignal. *Chem-Bio Informatics J.* **4,** 142–147 (2005).

134. Robinson, D. G., Oliviusson, P. & Hinz, G. Protein sorting to the storage vacuoles of plants: A critical appraisal. *Traffic* **6,** 615–625 (2005).

135. Xiang, L., Etxeberria, E. & Van Den Ende, W. Vacuolar protein sorting mechanisms in plants. *FEBS J.* **280,** 979–993 (2013).

136. Beaudoin, G. a W. & Facchini, P. J. Benzylisoquinoline alkaloid biosynthesis in opium poppy. *Planta* **240,** 19–32 (2014).

137. Lichman, B. R., Gershater, M. C., Lamming, E. D., Pesnot, T., Sula, A., Keep, N. H., Hailes, H. C. & Ward, J. M. 'Dopamine-first' mechanism enables the rational engineering of the norcoclaurine synthase aldehyde activity profile. *FEBS J.* **282,** 1137–1151 (2015).

138. Taft, R. W., Wolf, J. F., Beauchamp, J. L., Scorrano, G. & Arnett, E. M. Solvent Effects of Water and Fluorosulfuric Acid on Proton Transfer Equilibria and the Energies of Solvation of Gaseous Onium Ions. *J. Am. Chem. Soc.* **100,** 1240–1249 (1978).

139. Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. & Baker, N. A. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35,** W522–W525 (2007).

140. Burley, S. K. & Petsko, G. A. Amino-aromatic interactions in proteins. *FEBS Lett.* **203,** 139–143 (1986).

141. Samanani, N., Park, S.-U. & Facchini, P. J. Cell type-specific localization of transcripts encoding nine consecutive enzymes involved in protoberberine alkaloid biosynthesis. *Plant Cell* **17,** 915–926 (2005).

142. Hagel, J. M., Weljie, A. M., Vogel, H. J. & Facchini, P. J. Quantitative 1H nuclear magnetic resonance metabolite profiling as a functional genomics platform to investigate alkaloid biosynthesis in opium poppy. *Plant Physiol.* **147,** 1805–1821 (2008).

143. Zulak, K. G., Weljie, A. M., Vogel, H. J. & Facchini, P. J. Quantitative 1H NMR metabolomics reveals extensive metabolic reprogramming of primary and secondary metabolism in elicitor-treated opium poppy cell cultures. *BMC Plant Biol.* **8,** 5 (2008).

144. Roberts, M. F., McCarthy, D., Kutchan, T. M. & Coscia, C. J. Localization of enzymes and alkaloidal metabolites in Papaver latex. *Arch. Biochem. Biophys.* **222,** 599–609 (1983).

145. Bar-Even, A. & Tawfik, D. S. Engineering specialized metabolic pathways-is there a room for enzyme improvements? *Curr. Opin. Biotechnol.* **24,** 310–319

(2012).

146. Bar-Even, A., Milo, R., Noor, E. & Tawfik, D. S. The moderately efficient enzyme: futile encounters and enzyme floppiness. *Biochemistry* **54,** 4969–4977 (2015).

147. Dicpinigaitis, P. V, Morice, A. H., Birring, S. S., Mcgarvey, L., Smith, J. A., Canning, B. J. & Page, C. P. Antitussive Drugs — Past, Present, and Future. *Pharmacol. Rev.* **66,** 468–512 (2014).

148. Kashiwada, Y., Aoshima, A., Ikeshiro, Y., Chen, Y. P., Furukawa, H., Itoigawa, M., Fujioka, T., Mihashi, K., Cosentino, L. M., Morris-Natschke, S. L. & Lee, K. H. Anti-HIV benzylisoquinoline alkaloids and flavonoids from the leaves of Nelumbo nucifera, and structure-activity correlations with related alkaloids. *Bioorganic Med. Chem.* **13,** 443–448 (2005).

149. Brown, T. C. K. From arrow poison to neuromuscular blockers. *Paediatr. Anaesth.* **23,** 865–867 (2013).

150. D'Incalci, M., Badri, N., Galmarini, C. M. & Allavena, P. Trabectedin, a drug acting on both cancer cells and the tumour microenvironment. *Br. J. Cancer* **111,** 646–650 (2014).

151. Neubauer, D. N. Almorexant, a dual orexin receptor antagonist for the treatment of insomnia. *Curr. Opin. Investig. Drugs* **11,** 101–110 (2010).

152. Chapple, C. R., Cardozo, L., Steers, W. D. & Govier, F. E. Solifenacin significantly improves all symptoms of overactive bladder syndrome. *Int. J. Clin. Pract.* **60,** 959–966 (2006).

153. Tsoyi, K., Kim, W. S., Kim, Y. M., Kim, H. J., Seo, H. G., Lee, J. H., Yun-Choi, H. S. & Chang, K. C. Upregulation of PTEN by CKD712, a synthetic tetrahydroisoquinoline alkaloid, selectively inhibits lipopolysaccharide-induced VCAM-1 but not ICAM-1 expression in human endothelial cells. *Atherosclerosis* **207,** 412–419 (2009).

154. Jang, H. J., Tsoyi, K., Kim, Y. M., Park, E. J., Park, S. W., Kim, H. J., Lee, J. H. & Chang, K. C. (*S*)-1-α-naphthylmethyl-6,7-dihydroxy-1,2,3,4-tetrahydroisoquinoline (CKD712), promotes wound closure by producing VEGF through HO-1 induction in human dermal fibroblasts and mouse skin. *Br. J. Pharmacol.* **168,** 1485–1496 (2013).

155. Zhao, D. & Glorius, F. Enantioselective hydrogenation of isoquinolines. *Angew. Chem. Int. Ed.* **52,** 9616–9618 (2013).

156. Liu, W., Liu, S., Jin, R., Guo, H. & Zhao, J. Novel strategies for catalytic asymmetric synthesis of C1-chiral 1,2,3,4-tetrahydroisoquinolines and 3,4-dihydrotetrahydroisoquinolines. *Org. Chem. Front.* **2,** 288–299 (2015).

157. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C. & Robins, K. Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).

158. Turner, N. J. & O'Reilly, E. Biocatalytic retrosynthesis. *Nat. Chem. Biol.* **9,** 285–288 (2013).

159. Fessner, W.-D. Systems Biocatalysis: Development and engineering of cell-free 'artificial metabolisms' for preparative multi-enzymatic synthesis. *N. Biotechnol.* **32,** 658–664 (2014).

160. Roux, S., Zékri, E., Rousseau, B., Paternostre, M., Cintrat, J.-C. & Fay, N. Elimination and exchange of trifluoroacetate counter-ion from cationic peptides:

a critical evaluation of different approaches. *J. Pept. Sci.* **14,** 354–359 (2008).

161. Lichman, B. R., Lamming, E. D., Pesnot, T., Smith, J. M., Hailes, H. C. & Ward, J. M. One-pot triangular chemoenzymatic cascades for the syntheses of chiral alkaloids from dopamine. *Green Chem.* **17,** 852–855 (2015).

162. Ricca, E., Brucher, B. & Schrittwieser, J. H. Multi-Enzymatic Cascade Reactions: Overview and Perspectives. *Adv. Synth. Catal.* **353,** 2239–2262 (2011).

163. Simon, R. C., Richter, N., Busto, E. & Kroutil, W. Recent Developments of Cascade Reactions Involving ω-Transaminases. *ACS Catal.* **4,** 129–143 (2014).

164. Sehl, T., Hailes, H. C., Ward, J. M., Wardenga, R., von Lieres, E., Offermann, H., Westphal, R., Pohl, M. & Rother, D. Two steps in one pot: enzyme cascade for the synthesis of nor(pseudo)ephedrine from inexpensive starting materials. *Angew. Chem. Int. Ed.* **52,** 6772–5 (2013).

165. Könst, P., Merkens, H., Kara, S., Kochius, S., Vogel, A., Zuhse, R., Holtmann, D., Arends, I. W. C. E. & Hollmann, F. Enantioselective oxidation of aldehydes catalyzed by alcohol dehydrogenase. *Angew. Chem. Int. Ed.* **51,** 9914–9917 (2012).

166. Akhtar, M. K., Turner, N. J. & Jones, P. R. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 87–92 (2013).

167. Fabbrini, M., Galli, C., Gentili, P. & Macchitella, D. An oxidation of alcohols by oxygen with the enzyme laccase and mediation by TEMPO. *Tetrahedron Lett.* **42,** 7551–7553 (2001).

168. McMurtrey, K. D., Meyerson, L. R., Cashaw, J. L. & Davis, V. E. Kinetic and Product Distribution in Pictet-Spengler Cyclization of Tetrahydropapaveroline to Tetrahydroprotoberberine Alkaloids. *J. Org. Chem.* **49,** 947–948 (1984).

169. Collins, M. A. Tetrahydropapaveroline in Parkinson's disease and alcoholism: a look back in honor of Merton Sandler. *Neurotoxicology* **25,** 117–120 (2004).

170. Yao, L., Fan, P., Arolfo, M., Jiang, Z., Olive, M. F., Zablocki, J., Sun, H.-L., Chu, N., Lee, J., Kim, H.-Y., Leung, K., Shryock, J., Blackburn, B. & Diamond, I. Inhibition of aldehyde dehydrogenase-2 suppresses cocaine seeking by generating THP, a cocaine use-dependent inhibitor of dopamine synthesis. *Nat. Med.* **16,** 1024–1028 (2010).

171. Okada, T., Shimada, S., Sato, K., Kotake, Y., Kawai, H., Ohta, S., Tohyama, M. & Nishimura, T. Tetrahydropapaveroline and its derivatives inhibit dopamine uptake through dopamine transporter expressed in HEK293 cells. *Neurosci. Res.* **30,** 87–90 (1998).

172. Poeaknapo, C., Schmidt, J., Brandsch, M., Dräger, B. & Zenk, M. H. Endogenous formation of morphine in human cells. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 14091–14096 (2004).

173. Boettcher, C., Fellermeier, M., Boettcher, C., Dräger, B. & Zenk, M. H. How human neuroblastoma cells make morphine. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 8495–8500 (2005).

174. Grobe, N., Lamshöft, M., Orth, R. G., Dräger, B., Kutchan, T. M., Zenk, M. H. & Spiteller, M. Urinary excretion of morphine and biosynthetic precursors in mice. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 8147–8152 (2010).

175. Sango, K., Maruyama, W., Matsubara, K., Dostert, P., Minami, C., Kawai, M. &

Naoi, M. Enantio-selective occurrence of (S)-tetrahydropapaveroline in human brain. *Neurosci. Lett.* **283,** 224–226 (2000).

176. Stermitz, F. R., Lorenz, P., Tawara, J. N., Zenewicz, L. A. & Lewis, K. Synergy in a medicinal plant: antimicrobial action of berberine potentiated by 5'-methoxyhydnocarpin, a multidrug pump inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 1433–1437 (2000).

177. Yu, H.-H., Kim, K.-J., Cha, J.-D., Kim, H.-K., Lee, Y.-E., Choi, N.-Y. & You, Y.-O. Antimicrobial activity of berberine alone and in combination with ampicillin or oxacillin against methicillin-resistant Staphylococcus aureus. *J. Med. Food* **8,** 454–461 (2005).

178. Queiroz, E. F., Roblot, F., Cavé, A., Paulo, M. de Q. & Fournet, A. Pessoine and Spinosine, Two Catecholic Berbines from *Annona spinescens*. *J. Nat. Prod.* **7,** 438–440 (1996).

179. Iwasa, K., Cui, W., Takahashi, T., Nishiyama, Y., Kamigauchi, M., Koyama, J., Takeuchi, A., Moriyasu, M. & Takeda, K. Biotransformation of phenolic tetrahydroprotoberberines in plant cell cultures followed by LC-NMR, LC-MS, and LC-CD. *J. Nat. Prod.* **73,** 115–122 (2010).

180. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W. & Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329,** 305–309 (2010).

181. Shin, J. S. & Kim, B. G. Asymmetric synthesis of chiral amines with omega-transaminase. *Biotechnol. Bioeng.* **65,** 206–211 (1999).

182. Park, E. S., Kim, M. & Shin, J. S. Molecular determinants for substrate selectivity of omega-transaminases. *Appl. Microbiol. Biotechnol.* **93,** 2425–2435 (2012).

183. Sattler, J. H., Fuchs, M., Tauber, K., Mutti, F. G., Faber, K., Pfeffer, J., Haas, T. & Kroutil, W. Redox Self-Sufficient Biocatalyst Network for the Amination of Primary Alcohols. *Angew. Chem. Int. Ed.* **51,** 9156–9159 (2012).

184. Fuchs, M., Tauber, K., Sattler, J., Lechner, H., Pfeffer, J., Kroutil, W. & Faber, K. Amination of benzylic and cinnamic alcohols via a biocatalytic, aerobic, oxidation–transamination cascade. *RSC Adv.* **2,** 6262–6265 (2012).

185. Iwasa, K., Lee, D. U., Kang, S. I. & Wiegrebe, W. Antimicrobial activity of 8-alkyl- and 8-phenyl-substituted berberines and their 12-bromo derivatives. *J. Nat. Prod.* **61,** 1150–1153 (1998).

186. Wang, Y., Fu, H., Li, Y., Jiang, J. & Song, D. Synthesis and biological evaluation of 8-substituted berberine derivatives as novel anti-mycobacterial agents. *Acta Pharm. Sin. B* **2,** 581–587 (2012).

187. Li, Y.-H., Fu, H.-G., Su, F., Gao, L.-M., Tang, S., Bi, C.-W., Li, Y.-H., Wang, Y.-X. & Song, D.-Q. Synthesis and structure-activity relationship of 8-substituted protoberberine derivatives as a novel class of antitubercular agents. *Chem. Cent. J.* **7,** 117 (2013).

188. Boersma, Y. L., Dröge, M. J. & Quax, W. J. Selection strategies for improved biocatalysts. *FEBS J.* **274,** 2181–2195 (2007).

189. Hopps, H. B. Purpald: A reagent that turns aldehydes purple! *Aldrichimica Acta* **33,** 28–30 (2000).

190. Lauchli, R., Rabe, K. S., Kalbarczyk, K. Z., Tata, A., Heel, T., Kitto, R. Z. &

Arnold, F. H. High-throughput screening for terpene-synthase-cyclization activity and directed evolution of a terpene synthase. *Angew. Chem. Int. Ed.* **52,** 5571–5574 (2013).

191. Doi, S., Shirai, N. & Sato, Y. Abnormal products in the Bischler – Napieralski isoquinoline synthesis. *J. Chem. Soc., Perkin Trans. 1* 2217–2221 (1997).

192. Venkov, A. P. & Angelov, P. a. 1,1-Disubstituted Tetrahydroisoquinolines from Enaminones via Pictet–Spengler Reaction. *Synth. Commun.* **33,** 3025–3033 (2003).

193. Bois-Choussy, M., Cadet, S., De Paolis, M. & Zhu, J. Diethyl oxomalonate as a three carbon synthon for synthesis of functionalized 1,1′-disubstituted tetrahydroisoquinoline. *Tetrahedron Lett.* **42,** 4503–4506 (2001).

194. Hegedüs, A. & Hell, Z. One-step preparation of 1-substituted tetrahydroisoquinolines via the Pictet-Spengler reaction using zeolite catalysts. *Tetrahedron Lett.* **45,** 8553–8555 (2004).

195. Horiguchi, Y., Kodama, H., Nakamura, M., Yoshimura, T., Hanezi, K., Hamada, H., Saitoh, T. & Sano, T. A convenient synthesis of 1,1-disubstituted 1,2,3,4-tetrahydroisoquinolines via Pictet-Spengler reaction using titanium(IV) isopropoxide and acetic-formic anhydride. *Chem. Pharm. Bull. (Tokyo).* **50,** 253–257 (2002).

196. Luu, H.-T., Wiesler, S., Frey, G. & Streuff, J. A Titanium(III)-Catalyzed Reductive Umpolung Reaction for the Synthesis of 1,1-Disubstituted Tetrahydroisoquinolines. *Org. Lett.* **17,** 2478–2481 (2015).

197. Kuo, F. M., Tseng, M. C., Yen, Y. H. & Chu, Y. H. Microwave accelerated Pictet-Spengler reactions of tryptophan with ketones directed toward the preparation of 1,1-disubstituted indole alkaloids. *Tetrahedron* **60,** 12075–12084 (2004).

198. Wang, L.-N., Shen, S.-L. & Qu, J. Simple and efficient synthesis of tetrahydro-β-carbolines via the Pictet–Spengler reaction in 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP). *RSC Adv.* **4,** 30733–30741 (2014).

199. Li, X. & Coldham, I. Synthesis of 1,1-disubstituted tetrahydroisoquinolines by lithiation and substitution, with in situ IR spectroscopy and configurational stability studies. *J. Am. Chem. Soc.* **136,** 5551–5554 (2014).

200. Brozda, D., Hoffman, K. & Rozwadowska, M. D. Enantioselective Alkylation of Reissert Compounds in Phase Transfer Catalysed Reactions. *Heterocycles* **67,** 119–122 (2005).

201. Frisch, K. & Jørgensen, K. A. Organocatalytic asymmetric destruction of 1-benzylated Reissert compounds catalysed by quaternary cinchona alkaloids. *Org. Biomol. Chem.* **5,** 2966–2974 (2007).

202. Preisner, R. & Shamma, M. The spirobenzylisoquinoline alkaloids. *J. Nat. Prod.* **43,** 305–318 (1980).

203. Shamma, M. & Nugent, J. F. The Protoberberine to Spirobenzylisoquinoline to Dibenzocyclopent[b]azepine rearrangement. *Tetrahedron* **29,** 1265–1272 (1973).

204. Amer, M., Shamma, M. & Freyer, A. J. The tetracyclic Erythrina alkaloids. *J. Nat. Prod.* **54,** 329–363 (1991).

205. Maier, U. H., Rödl, W., Deus-Neumann, B. & Zenk, M. H. Biosynthesis of Erythrina alkaloids in Erythrina crista-galli. *Phytochemistry* **52,** 373–382 (1999).

206. Dreyer, D. R., Miller, D. J., Freeman, B. D., Paul, D. R. & Bielawski, C. W. Perspectives on poly(dopamine). *Chem. Sci.* **4,** 3796–3802 (2013).

207. Wilson, M. L. & Coscia, C. J. Studies on the Early Stages of *Papaver* Alkaloid Biogenesis. *J. Am. Chem. Soc.* **97,** 431–432 (1974).

208. Battersby, A. R., Jones, R. C. F. & Kazlauskas, R. Experiments on the Early Steps of Morphine Biosynthesis. *Tetrahedron Lett.* **22,** 1873–1876 (1975).

209. Marshalok, G. A., Karpyak, N. M., Makitra, R. G. & Polyuzhin, I. P. Solvent Effect on the Rate of the Mannich Reaction of Butyraldehyde with Formaldehyde. *Russ. J. Org. Chemsitry* **42,** 1264–1268 (2006).

210. Tanaka, F., Thayumanavan, R., Mase, N. & Barbas, C. F. Rapid analysis of solvent effects on enamine formation by fluorescence: How might enzymes facilitate enamine chemistry with primary amines? *Tetrahedron Lett.* **45,** 325–328 (2004).

211. Klibanov, A. M. Improving enzymes by using them in organic solvents. *Nature* **409,** 241–246 (2001).

212. Roy, S., Jana, B. & Bagchi, B. Dimethyl sulfoxide induced structural transformations and non-monotonic concentration dependence of conformational fluctuation around active site of lysozyme. *J. Chem. Phys.* **136,** 115103 (2012).

213. Watanabe, K. & Ueji, S. Dimethyl sulfoxide as a co-solvent dramatically enhances the enantioselectivity in lipase-catalysed resolutions of 2-phenoxypropionic acyl derivatives. *J. Chem. Soc. Perkin Trans. 1* 1386–1390 (2001).

214. Khersonsky, O., Malitsky, S., Rogachev, I. & Tawfik, D. S. Role of Chemistry versus Substrate Binding in Recruiting Promiscuous Enzyme Functions. *Biochemistry* **50,** 2683–2690 (2011).

215. Currin, A., Swainston, N., Day, P. J. & Kell, D. B. SpeedyGenes: an improved gene synthesis method for the efficient production of error-corrected, synthetic protein libraries for directed evolution. *Protein Eng. Des. Sel.* **9,** 273–280 (2014).

216. Herman, A. & Tawfik, D. S. Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): A versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20,** 219–226 (2007).

217. Kries, H., Blomberg, R. & Hilvert, D. De novo enzymes by computational design. *Curr. Opin. Chem. Biol.* **17,** 221–228 (2013).

218. Mak, W. S. & Siegel, J. B. Computational enzyme design: Transitioning from catalytic proteins to enzymes. *Curr. Opin. Struct. Biol.* **27,** 87–94 (2014).

219. Khersonsky, O., Röthlisberger, D., Dym, O., Albeck, S., Jackson, C. J., Baker, D. & Tawfik, D. S. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *J. Mol. Biol.* **396,** 1025–1042 (2010).

220. Khersonsky, O., Röthlisberger, D., Wollacott, A. M., Murphy, P., Dym, O., Albeck, S., Kiss, G., Houk, K. N., Baker, D. & Tawfik, D. S. Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *J. Mol. Biol.* **407,** 391–412 (2011).

221. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339,** 307–10 (2013).

222. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the Enzyme Universe:

Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angew. Chemie Int. Ed.* **54,** 3351–3367 (2015).

223. Arnold, F. H. The nature of chemical innovation: new enzymes by evolution. *Q. Rev. Biophys.* 1–7 (2015). doi:10.1017/S003358351500013X

224. King, E. L. & Altman, C. A Schematic Method of Deriving the Rate Laws for Enzyme-Catalyzed Reaction. *J. Phys. Chem.* **60,** 1375–1378 (1956).

225. Lam, C. F. & Priest, D. G. Enzyme kinetics. Systematic generation of valid King-Altman patterns. *Biophys. J.* **12,** 248–256 (1972).

226. Goličnik, M. Exact and approximate solutions for the decades-old Michaelis-Menten equation: Progress-curve analysis through integrated rate equations. *Biochem. Mol. Biol. Educ.* **39,** 117–125 (2011).

227. Garman, E. F. Developments in X-ray Crystallographic Biological Macromolecules. *Science* **343,** 1102–1108 (2014).

228. Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18,** 6069–6074 (1990).

229. Cowtan, K., Emsley, P. & Wilson, K. S. From crystal to structure with CCP4. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67,** 233–234 (2011).

230. Keegan, R. M. & Winn, M. D. MrBUMP: An automated pipeline for molecular replacement. in *Acta Crystallogr. Sect. D Biol. Crystallogr.* **64,** 119–124 (2007).

231. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **53,** 240–255 (1997).

232. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 486–501 (2010).

233. Kamerlin, S. C. L. & Warshel, A. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* **78,** 1339–1375 (2010).

234. Masterson, L. R., Cheng, C., Yu, T., Tonelli, M., Kornev, A., Taylor, S. S. & Veglia, G. Dynamics connect substrate recognition to catalysis in protein kinase A. *Nat. Chem. Biol.* **6,** 821–828 (2010).

235. Ma, B. & Nussinov, R. Enzyme dynamics point to stepwise conformational selection in catalysis. *Curr. Opin. Chem. Biol.* **14,** 652–659 (2010).

236. Strafford, J., Payongsri, P., Hibbert, E. G., Morris, P., Batth, S. S., Steadman, D., Smith, M. E. B., Ward, J. M., Hailes, H. C. & Dalby, P. A. Directed evolution to re-adapt a co-evolved network within an enzyme. *J. Biotechnol.* **157,** 237–245 (2011).

237. Sunden, F., Peck, A., Salzman, J., Ressl, S. & Herschlag, D. Extensive site-directed mutagenesis reveals interconnected functional units in the alkaline phosphatase active site. *Elife* **4,** 1–31 (2015).