Bionanotechnology Platforms for Biocatalysis

HELENA KATE PHILPOTT

THIS DISSERTATION IS SUBMITTED FOR THE DEGREE OF

Doctor of Philosophy

Supervisors:

Professor Helen Hailes

Professor Stefan Howorka

Submitted: 05/12/2021

Declaration

I, Helena Kate Philpott, 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.

Signed

05/12/2021 Date

Abstract

Biocatalysis, the use of enzymes to perform chemical reactions, is widely accepted as a greener, more sustainable alternative to traditional chemical synthesis. Enzymes provide exquisite regio- and stereo-selectivity and can open synthetic routes to compounds which are inaccessible by other chemical means. Enzymes have been extensively investigated and exploited for these properties, and due to the recent advances in bioinformatics, high-throughput screening and gene synthesis, we now have an almost infinite library of potential enzymes. Despite this library of enzymes being available there are still only a few commercially large scale processes which use enzymes. The use of enzymes in industrial processes pose a number of problems including low yields due to low substrate loading, slow reaction rates, cross reactivities and issues with downstream processing. In order to tackle these problems several methods, including various immobilisation and encapsulation techniques, have been developed.

This thesis investigated two such methods for immobilisation of enzymes. These methods, based upon cholesterol membrane insertion and complementary DNA binding, were designed and synthesised to tether enzymes to a lipid bilayer membrane. The tethering of individual enzymes to a surface can improve enzyme stability, activity, selectivity, specificity and reduce inhibition. These methods built upon this, extending to an enzymatic cascade with a design to eliminate three-dimensional diffusion, increase reaction rates and yields, and improve flux. Testing of these methods was divided into a proof of concept, utilising fluorescent proteins to examine the efficacy of immobilisation through confocal microscopy; and testing with individual and cascaded enzymes in order to determine the benefits of the immobilisation techniques. A novel cascade was selected which would utilise redox and functional group conversion reactions and involve three enzyme classes: alcohol dehydrogenases, ene-reductases and transaminases. It is expected that these techniques will help to overcome the problems faced by biocatalysis and provide new bionanotechnology reaction platforms.

Impact Statement

In the fight against global warming, green chemistry has the potential to greatly reduce the impact of chemical processes by reducing the amount of harmful chemicals involved, the amount of energy used, and by providing alternative routes which do not rely on petrochemicals. The work presented in this thesis has the potential to have a an impact as part of a global drive towards sustainable chemical practices.

The use of biocatalysis, particularly in industry, is expanding rapidly and as a result the research into this area is also increasing. There are many ways to approach integrating biocatalysts into the synthesis of various chemicals. Biocatalysis has huge potential but its uptake and wide spread use has been slow. If the costs of biocatalysis can, either directly, or through reduction in emissions and feedstock costs, be reduced to the point where it is worth investing in a new process, not just for sustainability, then the uptake will be drastically increased. The research presented in this thesis aims to address some of the current problems faced with biocatalysis through developing novel immobilisation platforms. Enzyme immobilisation is hoped to make using enzymes an easier process and could be the way to enable biocatalysis to really challenge and replace traditional chemical synthesis. Developing new enzymatic cascades, whether immobilised or not, can also help to reduce the use of harmful chemicals and solvents, reduce pollution, and potentially provide alternative green route to useful chemicals. The combination of new enzyme cascades, novel immobilisation techniques, and all other developments in biocatalysis from academia to industry can all help to reduce the impact of the chemical industry on our planet.

The work presented here was disseminated through a number of poster and oral presentations at various conferences. Although the work was not published, it provides a starting point for investigating novel immobilisation platforms using lipid bilayers and chemical tethers.

Acknowledgements

The research presented in this thesis would not have been possible without the support of a number of people who helped me both academically and personally throughout my time as a PhD student.

First, I would like to thank my two supervisors, Professor Helen Hailes and Professor Stefan Howorka, for welcoming me into their research groups and providing invaluable guidance throughout my time at UCL. I would also like to thank my secondary supervisor Professor John Ward who allowed me to work in his laboratory and also provided valuable discussions on my research.

There are a huge number of scientists who helped train me and provide me with the skills I needed to complete this very broad range of experiments. Three amazing postdocs from the Hailes group, Dr. Fabiana Subrizi, Dr. Laure Benhamou, and Dr. Daniel Mendez-Sanchez, guided me through my time as a fledgling scientist, without whom I would have greatly struggled. Dr. Rachael Dickman and Dr. Stephen McCarthy, from the Tabor group, also provided invaluable support throughout my research and helped me with my chemistry and peptide synthesis. Dr. Daniel Offenbartl-Stiegert and Dr. Conor Lanphere, from the Howorka group, who guided me through the world of DNA and confocal microscopy. Thanks also to Dr. Rachael Dickman, Dr. Stephen McCarthy, Dr. Daniel Offenbartl-Stiegert, Dr. Rebecca Roddan, and Eve Carter for reading various chapters of this thesis and providing useful feedback.

I would also like to thank a number of collaborators and scientists from other research groups who worked with me, trained me, or assisted me in my research. Dr. Nadine Tappertzhofen, Dr. Dragana Dobrijevic, Dr. Maria Bawn, and Dr. Daniel Mendez-Sanchez all provided me with plasmids or enzymes I required for developing enzymatic cascades and Dr. James Allen provided plasmids for the fluorescent proteins. Dr. Abil Aliev ran and provided helpful discussions on NMR, whilst Dr. Kersti Karu and Dr. Maggie Puchnarewicz ran and provided helpful discussions on mass spectrometry. Dr. Christoph Zimmer provided training for ITC and helped me set up my first experiments. Dr. Alan Greig provided training for confocal microscopy and helped set up initial experiments.

On a personal note, I would like to thank everyone from Lab 435 and the Howorka group, past

and present, who have been amazing co-workers and friends throughout my time at UCL. Rebecca Roddan, Eve Carter, Rachael Dickman, Billy Darling, Ben Thair, Esther Ambrose-Dempster, Simon Guillaume, Liam Martin, and Stephen McCarthy from 435, Daniel Offenbartl-Stiegert, Conor Lanphere, Katya Ahmad, Elena Georgiou, Alexia Rottensteiner, Jonah Ciccone, Catherine Webley, and Adam Dorey from the Howorka group. I don't think I would have survived all the ups and downs without all of you being there for me, thank you all.

I also need to thank my family and all of my friends who have been there for me throughout all of this supporting me in whatever way they could. Thank you to my Mum and Dad for their support, encouragement, pride, and belief in me. Finally, I need to thank my partner Luke for his unending love and support, for always being there for me, and for always knowing how to cheer me up no matter what.

Table of Contents

List of Symbols and Abbreviations

A	Adenine
A_{260}	Absorbance (at a given wavelength (nm))
ACN	Acetonitrile
ADH	Alcohol Dehydrogenase
ADP	Adenosine Diphosphate
AldDH	Aldehyde Dehydrogenase
Am'	Ampicillin
AMP	Adenosine Monophosphate
aq	Aqueous
ATP	Adenosine Triphosphate
a.u.	Absorbance Units
Boc	*tert*-butoxycarbonyl
BMC	Bacterial Microcompartment
c	Concentration
C	Cytosine
°C	Degrees Celsius
CCMV	Cowpea Chlorotic Mottle Virus
CFP	Cyan Fluorescent Protein
CH_2Cl_2	Dichloromethane
$CHCl_3$	Chloroform
Chol	Cholesterol
CLEA	Cross-Linked Enzyme Aggregate
CLEC	Cross-Linked Enzyme Crystal
cm	Centimetre
Cm'	Chloramphenicol
CMC	Critical Micelle Concentration
δ	Chemical Shift
DIPEA	*N*,*N*-Diisopropylethylamine
DMAP	4-(Dimethylamino)pyridine
DMF	Dimethylformamide
DMSO	Dimethylsulphoxide

DNA	Deoxyribonucleic Acid
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
dsDNA	Double-Stranded DNA
ϵ	Extinction Coefficient
EDTA	Ethylenediaminetetraacetic Acid
EP	Encapsulation Peptide
eq	Equivalents
ER	Ene-Reductase
ESI-MS	Electrospray Ionisation Mass Spectrometry
Et_2O	Diethyl Ether
EtOAc	Ethyl Acetate
EtOH	Ethanol
E4P	Erythrose-4-Phosphate
FDH	Formate Dehydrogenase
FMN	Flavin Mononucleotide
$FMNH_2$	Reduced Flavin Mononucleotide
Fmoc	9-Fluorenylmethyloxycarbonyl
FRET	Fluorescence Resonance Energy Transfer
F6P	Frutose-6-Phosphate
g	Graviational Force
g	Gram
G	Guanine
GC	Gas Chromatography
GDH	Glucose Dehydrogenase
GFP	Green Fluorescent Protein
GPI	Glycosylphosphatidylinositol
GUV	Giant Unilamellar Vesicle
G3P	Glyceraldehyde-3-Phosphate
G6P	Glucose-6-Phosphate
G6PDH	Glucose-6-Phosphate Dehydrogenase
h	Hours
Hz	Hertz

HCl	Hydrochloric Acid
His-Tag	Hexahistidine Tag
HPA	Hydroxypyruvic Acid
HRMS	High Resolution Mass-Spectrometry
H_2O	Water
H_2SO_4	Sulfuric Acid
IBX	2-Iodoxybenzoic Acid
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
J	Coupling Constant
Km'	Kanamycin
K_2CO_3	Potassium Carbonate
l	Path Length
L	Litre
LB	Luria-Bertani
LC-MS	Liquid Chromatography-Mass Spectrometry
LUV	Large Unilamellar Vesicle
μg	Microgram
μL	Microlitre
μM	Micromolar
μmol	Micromole
M	Molar
mA	Milliamp
mg	Milligram
mL	Millilitre
mM	Millimolar
$mmol$	Millimole
mol	Mole
MDR	Medium-Chain Dehydrogenase/Reductase
MeOH	Methanol
$MgSO_4$	Magnesium Sulphate
min	minute
MOPS	3-(N-Morpholino)propanesulfonic Acid

MW	Molecular Weight
MWCO	Molecular Weight Cut Off
nm	Nanometres
nM	Nanomolar
$nmol$	Nanomole
NAD(P)$^+$	Nicotinamide Adenine Dinucleotide (Phosphate)
NAD(P)H	Reduced Nicotinamide Adenine Dinucleotide (Phosphate)
NaOH	Sodium Hydroxide
Na_2SO_4	Sodium Sulphate
NHS	*N*-hydroxysuccinimide
NMR	Nuclear Magnetic Resonance
NTA	Nitrilotriacetic Acid
OD_{600}	Optical Density (at a given wavelength (nm))
OPOE	*n*-Octylpolyoxyethylene
OYE	Old Yellow Enzyme Family
ppm	Parts per Million
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PC	Phosphatidylcholine
PCC	Pyridinium Chlorochromate
PCR	Polymerase Chain Reaction
PE	Phosphatidylethanolamine
PEG	Polyethylene Glycol
pI	Isoelectric Point
PLP	Pyridoxal-5'-phosphate
PMP	Pyridoxamine-5'-phosphate
POPC	1-Palmitoyl-2-Oleoyl-Glycero-3-Phosphocholine
PPP	Pentose Phosphate Pathway
PTAC	Phosphotransacylase
rpm	Revolutions per Minute
RFP	Red Fluorescent Protein
r.t.	Room Temperature
RT	Retention Time

RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
SDR	Short-Chain Dehydrogenase/Reductase
SDS	Sodium Dodecyl Sulfate
SOC	Super Optimal Broth with Catabolite Repression
SPPS	Solid-Phase Peptide Synthesis
ssDNA	Single-Stranded DNA
SUV	Small Unilamellar Vesicle
S7P	Sedoheptulose-7-Phosphate
T	Thymine
TAE	Tris-Acetate-EDTA
TAm	Transaminase
TB	Terrific Broth
tBBA	*tert*-Butylbromoacetate
^{t}Bu	*tertiary*-Butyl
TCEP	Tris (2-Carboxyethyl) phosphine
TEA	Triethylamine
TEG	Triethylene Glycol
TFA	Trifluoroacetic Acid
TGX	Tris-Glycine
ThDP	Thiamine Diphosphate
THF	Tetrahydrofuran
TIPS	Triisopropylsilyl
TK	Transketolase
TLC	Thin Layer Chromatography
Tris	2-Amino-2-Hydroxymethyl-Propane-1,3-Diol
UV/Vis	Ultraviolet/Visible Light
w/v	Weight per Volume
WT	Wild-Type
V	Volts
v/v	Volume per Volume
X5P	Xylulose-5-Phosphate
YFP	Yellow Fluorescent Protein

List of Figures

List of Tables

List of Schemes

4 - DNA Linker Synthesis and Testing

1 Introduction

1.1 Green Chemistry

Global warming has been identified as one of the greatest anthropogenic threats to global ecosystems.[1] Reductions in pollution, raw material usage, and energy usage are key in combating global warming. Towards this aim, many public and private sector institutions have established their own environmental agencies such as the US Environmental Protection Agency,[2] established in 1970, and the European Environmental Protection Agency,[3] established in 1990. These agencies provide information to global governments, companies and individuals and work with them in order to reduce their impacts on the environment. Chemicals manufacturing is one of the largest global industries and has a significant environmental impact. It accounts for 10% of the global energy demand, 5.5% of the global CO_2 emissions, and is heavily reliant on raw materials from fossil fuels.[4]

The chemical industry is relied upon heavily for a wide range of products from manufactured goods, to pharmaceuticals, to energy storage, to transportation, and provide these goods and services across the globe. This reliance means that the chemical industry cannot simply cut back on production but must adapt and innovate in order to supply these services, whilst minimising the impact on the environment. One of the most significant changes the chemical industry has implemented over the years involves the way chemists and chemical companies assess the impact of their reactions and processes on the environment. In 1998, Anastas and Warner introduced the idea of green chemistry and gave scientists a set of twelve principles to follow in order to reduce their impact on the environment.[5] The principles of green chemistry (Table 1.1) are designed to not only reduce the hazards and waste associated with chemical reactions, but to improve the whole life-cycle of products, from feedstocks to degredation and recycling, whilst considering the energy and pollution costs across the whole process.

Table 1.1: The 12 Principles of Green Chemistry, taken from Anastas and Warner[5]

Principle	Description
1 - Prevention	It is better to prevent waste than to treat or clean up waste after it has been created.
2 - Atom Economy	Synthetic methods should be designed to maximise incorporation of all materials used in the process into the final product.
3 - Less Hazardous Synthesis	Wherever practicable, synthetic methods should be designed to use and generate substances that possess little or no toxicity to human health and the environment.
4 - Designing Safer Chemicals	Chemical products should be designed to preserve efficacy of function while reducing toxicity.
5 - Safer Solvents and Auxiliaries	The use of auxiliary substances (e.g., solvents, separation agents, etc.) should be made unnecessary wherever possible, and innocuous when used.
6 - Design for Energy Efficiency	Energy requirements should be recognised for their environmental and economic impacts and should be minimised. Synthetic methods should be conducted at ambient temperature and pressure.
7 - Use of Renewable Feedstocks	A raw material or feedstock should be renewable rather than depleting whenever technically and economically practicable.
8 - Reduce Derivatives	Unnecessary derivatisation (use of blocking groups, protection/deprotection, temporary modification of physical/chemical processes) should be minimised or avoided if possible, because such steps require additional reagents and can generate waste.
9 - Catalysis	Catalytic reagents (as selective as possible) are superior to stoichiometric reagents.
10 - Design for Degradation	Chemical products should be designed so that at the end of their function they break down into innocuous degradation products and do not persist in the environment.
11 - Real-Time Analysis for Pollution Prevention	Analytical methodologies need to be further developed to allow for real-time, in-process monitoring and control prior to the formation of hazardous substances.
12 - Inherently Safer Chemistry for Accident Prevention	Substances and the form of a substance used in a chemical process should be chosen to minimise the potential for chemical accidents, including releases, explosions, and fires.

One of the key synthetic methods within green chemistry is catalysis. Catalysis not only increases the atom and step economy of reactions, to make them more sustainable, but also reduces costs by reducing the number of stoichiometric reagents required.[6] Many of the traditional stoichiometric reagents, including metals (Na, Zn), metal hydrides ($LiAlH_4$, $NaBH_4$), mineral acids (H_2SO_4, HF) and Lewis acids ($AlCl_3$, $ZnCl_2$), can easily be replaced by using a catalyst to perform the same reactions.[7] Specialised catalysts are often required in asymmetric synthesis to selectively synthesise single enantiomers, particularly in the pharmaceutical industry where manufacture of a single enantiomer is often a requirement for regulators, as many biologically active molecules have very different activities depending on their absolute stereochemistry.[8] This can be achieved using chiral metal complexes of rhodium, ruthenium, iridium and other transition metals[9] with chiral ligands such as DuPhos, TADDOL or BINAP.[10] These complexes can produce highly enantiopure compounds, however, the metals are often very expensive and the chiral ligands even more so. Another problem is faced when disposing of these complexes, as they can be highly toxic to the environment, so a greener alternative to achieve the same enantiopurity is preferential[6].

Many synthetic processes have had to be redesigned in order to accommodate the principles of the green chemistry. One of the most radical of these redesigns has been moving away from traditional organic chemistry altogether and looking to synthetic biology. The developments within synthetic biology, particularly in bioinformatics, high-throughput DNA sequencing, gene synthesis and enzyme optimisation methods, have resulted in commercial biocatalysis becoming a viable alternative to traditional chemical synthesis.[11,12] Biocatalysis, the use of enzymes to perform chemical reactions, is an inherently green and more sustainable alternative to traditional organic chemistry synthesis, being compatible with all twelve of the green chemistry principles.

1.2 Biocatalysis

1.2.1 Enzymes

Enzymes are protein or RNA based biological catalysts which catalyse specific biochemical reactions by providing an alternative, more energetically favourable, reaction pathway through an enzyme-substrate complex. The decrease in the Gibbs free energy of activation from this more energetically favourable pathway results in a more stable transition state and greatly increases the rate of the reaction. The reaction rates are often 10^6-10^{12} times greater than the uncatalysed reaction and often several orders of magnitude above the equivalent chemical catalysis.[13]

Enzymes are incredibly valuable tools, particularly in the synthesis of fine chemicals, due to their high regio- and stereo-selectivity. This exquisite regio- and stereo-control often eliminates the need for protecting group chemistry, specific chiral reagents and expensive, environmentally detrimental, metal catalysts.[13] Additionally, enzymes often function optimally under benign conditions avoiding high temperatures, solvents, harsh reaction conditions, and extensive work-up procedures. Enzymes can reduce hazardous waste and come from renewable feedstocks, making them excellent green catalysts. Enzymes can also perform highly complex and intricate chemical reactions which, despite centuries of developments, still cannot be performed chemically.[14] However, despite this myriad of advantages there are still a few disadvantages to using enzymes, such as narrow substrate scopes, dependence on cofactors, sensitivity to solvents and temperatures, although enzyme engineering has helped to overcome some of these issues, and the difficulties in expressing enzymes from unculturable organisms, although this has been greatly improved with the development of *in silico* and other sequencing methods.[15]

1.2.2 Enzyme Toolbox

There are seven main classes of enzymes which can be utilised to perform chemical reactions.[16] This broad range of enzymes can be exploited, and engineered, to perform a vast array of reactions to match, and in some cases exceed, the capabilities of traditional organic synthesis. The seven enzyme classes and the types of reactions they perform are outlined in Table 1.2.

Table 1.2: The seven main classes of enzymes and the types of reactions they catalyse.[16]

EC Class	Description
1 - Oxidoreductases	Catalyse oxido-reductions, the transfer of hydrogen, oxygen or electrons from one substance to another.
2 - Transferases	Transfer a functional group from one compound, the donor, to another compound, the acceptor.
3 - Hydrolases	Catalyse the hydrolytic cleavage of various bonds, mainly C-O, C-N, C-C bonds.
4 - Lyases	Catalyse cleavage of C-O, C-N, C-C, and other bonds by elimination (not hydrolysis or oxidation), leaving double bonds or rings, or conversely adding groups to double bonds.
5 - Isomerases	Catalyse geometric or structural changes within one molecule.
6 - Ligases	Catalyse the joining together of two molecules coupled with the hydrolysis of a diphosphate bond in ATP or a similar triphosphate.
7 - Translocases	Catalyse the movement of ions or molecules across membranes or their separation within membranes

1.2.3 Industrial Biocatalysis

The use of enzymes in industrial processes has increased dramatically in the last two decades from often niche uses and only four subclasses of enzymes: proteases and lipases, for kinetic resolution of alcohols, amines and carboxylic acids; ketoreductases, for formation of chiral secondary alcohols from prochiral ketones; and nitrile hydratase to make acrylamide from acrylonitrile.[17] There have been several major developments in biocatalysis which have facilitated this rapid increase in its use in industrial processes. Enzyme discovery through *in silico* methods, namely through identification of putative enzymes by metagenomics, has drastically increased the number of enzymes available through sequence and structural alignments to identify putative enzymes without needing to test them experimentally.[18] A novel β-glucosidase was discovered in this manner by sequencing a microbial metagenome from Lake Poraquê in the Amazon jungle.[19] The genes for these putative enzymes can be synthesised and expressed rapidly due to the decrease in costs and production times for synthetic genes. Improvements in high-throughput screening allow vast enzyme panels to be analysed against any desired substrate. If any of these enzymes show activity they can then be engineered through various methods, such as directed evolution for which Frances Arnold won the Nobel Prize in chemistry for in 2018.[20] Directed evolution has developed from a method designed to predominately increase enzyme stability to a method which can generate enzyme capable of performing reactions which had previously not been possible with any single enzyme. An example of this is the evolution of a cytochrome P450, by changing only 0.2% of its sequence, to generate a cytochrome P411 which is capable of catalysing the non-natural C-C bond forming cyclopropanation of styrene by ethyl-diazoacetate.[21] These methods can improve the activity and stability of these novel enzymes and make them suitable for large scale industrial processes. Another development has been a change in attitude towards biocatalysis from being a niche method, used predominately in research laboratories, to a viable alternative to industrial chemical processes.[22] Enzymes were thought to be expensive, unstable, dependent upon expensive cofactors, difficult to develop and required extensive development the corresponding industrial process. However, recombinant expression, enzyme immobilisation, rational design, use of recycling systems, bioreaction engineering, and the development of interdisciplinary research teams now exist to overcome these problems making biocatalysis more accessible for industrial processes.

Integration of biotransformations into the traditional retrosynthetic repertoire is also an important development towards incorporation of enzymes into synthetic routes, where enzymes can be used

in cascades along or in combination with traditional chemical reactions for formation of fine and bulk chemicals.[23] Generating a catalogue of enzymes which can be searched to identify, and potentially purchase, an appropriate enzyme for a given transformation would provide a modular approach to synthetic route development enabling easier incorporation of enzymes. Towards this, numerous reviews[23–25] and a book[26] have now been published which introduce synthetic chemists to the benefits of biocatalysis and the broad range of enzymes available for synthesis (Figure 1.1). Having enzymes sit alongside chemical methods in a synthetic chemist's toolbox will help to shift the collective mindset away from automatically selecting chemical methods and considering alternatives. The shift of focus from individual labs onto interdisciplinary teams will also help to facilitate the generation of combined synthetic routes through the collective expertise of researchers from across chemical and biological fields. Industrial biocatalysis will continue to develop rapidly as new technologies become available and a need for greener processes become ever more important.

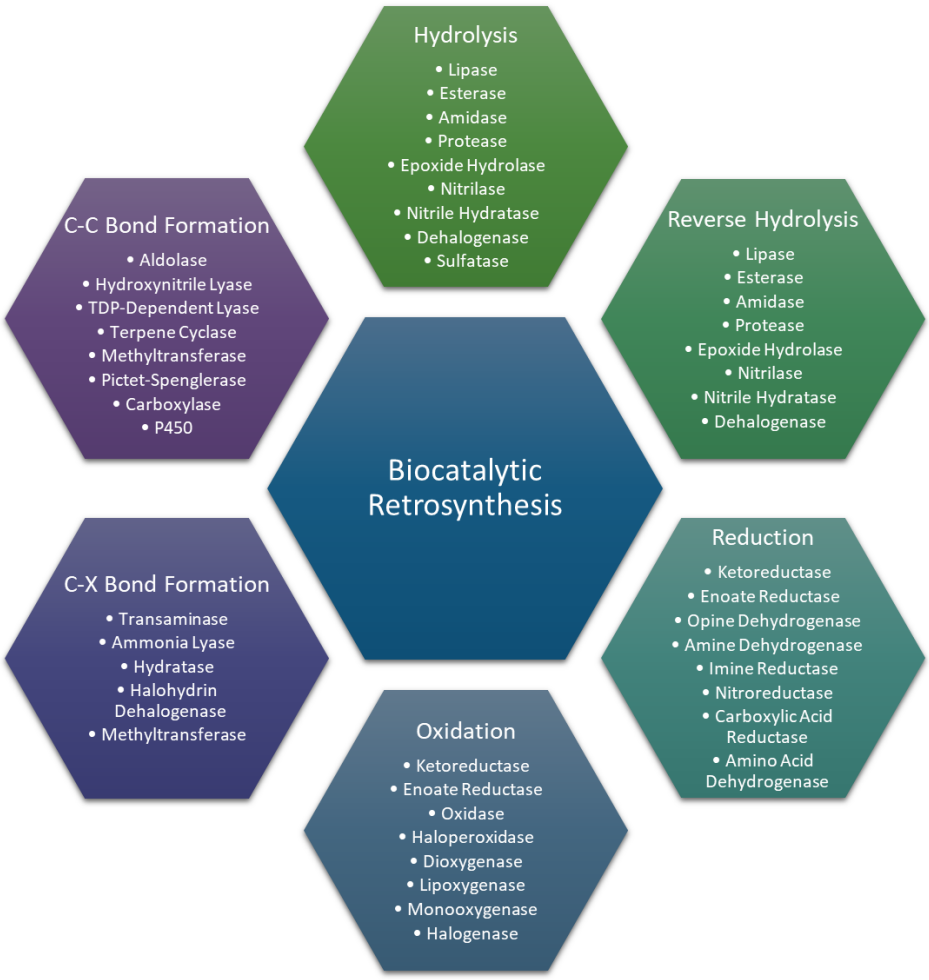


Figure 1.1: Overview of biocatalytic retrosynthesis toolbox showing the range of enzymes available for each reaction class.[26]

1.2.4 *In Vitro* vs. *In Vivo*

There are two main methods for performing enzymatic reactions: either *in vitro*, using purified enzymes, as well as cofactors and corresponding recycling systems if required; or *in vivo*, expressing the enzyme within a host organism and utilising its internal machinery.[27,28] Both of these methods have their advantages and disadvantages (Table 1.3). Whole cell biocatalysis does not require isolation or purification of the enzymes, greatly reducing the number of upstream steps required, which is particularly useful for larger industrial processes. Enzymes often have a higher stability in their native cellular environment and can be regenerated by the host cell if they are exhausted or degraded. The cell's internal metabolism can also be utilised as a recycling system for any cofactors required for the enzyme to function as well as providing other additives such as molecular oxygen from peroxidases. The microbial host cells can be engineered to do this without sequestering the substrate or desired product into their normal metabolism. However, this engineering can be very time consuming and labour intensive, and is required to control the desired reaction when any changes are made to it. The reduction in upstream steps can be offset by the increase in complexity and number of downstream steps. The product must be isolated from all the cellular components and any unwanted side-products, both of which can reduce the yield and purity of the product.[29] The yield and reaction rates can also be affected by membrane diffusion rates and competition from similar endogenous enzymes,[30] as well as being affected by toxicity from high substrate or organic co-solvent concentrations.[31]

The alternative to whole-cell biocatalysis is using isolated enzymes.[27,28] This is often used to simplify a reaction, making it easier to study, by removing all components except for those directly involved in the reaction. The reaction can be more easily controlled and optimised by simply changing the reaction conditions such as concentrations, pH or temperature. Without cellular components, reaction rates are not limited by membrane diffusion, side reactions are reduced, and organic co-solvents or higher substrate concentrations are better tolerated. The downstream processing is also significantly easier as the product has to be separated from far fewer components. As with whole-cell biocatalysis, the easier process, in this case downstream, is offset by making the other end of the process, upstream, more complex. Isolation, purification, and characterisation of enzymes is required before any reactions can be performed. If any cofactors or other additives are required, they either need to be added stoichiometrically, or recycling systems have to be isolated to regenerate the cofactors or provide the additives. Any enzymes involved are likely to be less stable in a non-cellular environment and cannot be regenerated by cellular

protein expression when they are depleted.

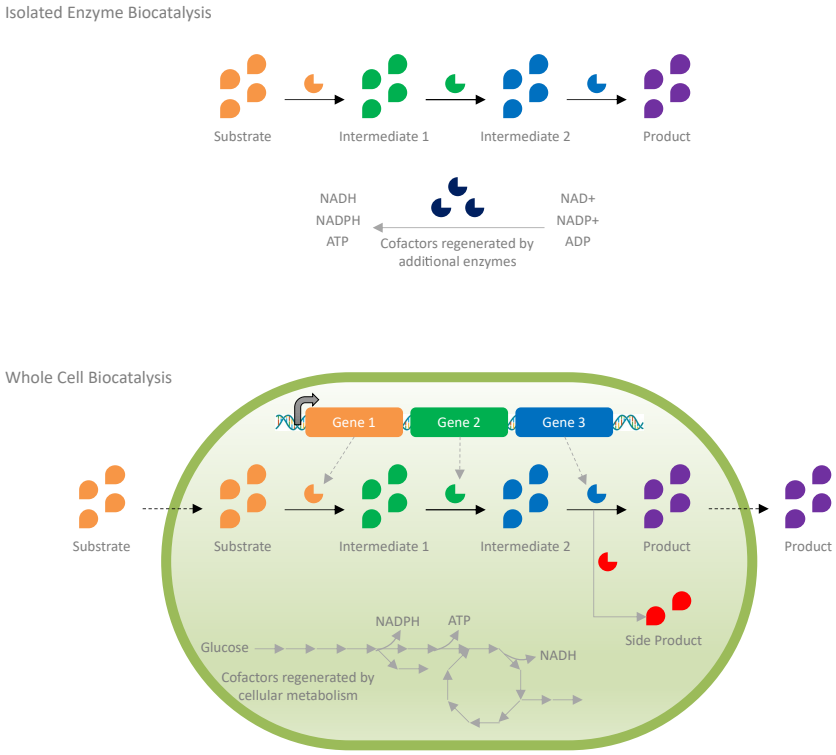


Figure 1.2: Isolated enzyme and whole cell biocatalysis - Schematic showing the differences between isolated enzyme and whole cell biocatalysis. The isolated enzyme biocatalysis shows how three purified enzymes can convert a substrate to a product and that the cofactors require additional enzymes to recycle them. The whole cell biocatalysis also shows three enzymes, which are translated from a plasmid, converting a substrate to a product. It also shows how the cell's metabolism regenerates cofactors and how other enzymes can interfere to form side products. Figure adapted from Wu and Li [32].

Table 1.3: Overview of the advantages and disadvantages of *in vitro* and *in vivo* enzymatic reactions.

	In Vivo	***In Vitro***
Advantages	• No enzyme isolation required • Higher enzyme stability in cellular environment • Cofactors can be regenerated by the host cell • Depleted enzymes can be regenerated by cellular protein expression	• Side reactions reduced • Reaction rates not limited by membrane diffusion • Reaction more easily controlled and optimised • Solvents and higher substrate concentrations better tolerated • Simpler isolation of products
Disadvantages	• Control of the reaction complex and requires host cell engineering • Isolation of products more complicated • Reaction rates limited by membranes and intracellular competition • Cells more likely to be affected by solvents and high substrate toxicity • Side reactions reduce yield and purity	• Isolation, purification and chacterisation of enzymes required • Cofactor recycling may be required • Enzymes cannot be regenerated once depleted • Enzymes likely to be less stable in non-cellular environments

There is also a third alternative method, generally preferred in industry, which provides a middle ground between whole cells and purified enzymes, crude cell lysates.[33] Using lysates overcomes the issues with using whole cells without needing to have enzymes engineered with affinity tags which can be fully purified. In fact, using lysates overcomes a number of issues with *in vitro* and *in vivo* catalysis discussed above, but does still face issues with enzyme stability, enzyme regeneration, side reactions, and product isolation. The native cofactor systems are also still available, although these often have to be supplemented by co-expressing recycling enzymes.[33] When working at industrial scales, purification and isolation of enzymes is highly impractical and expensive and would render many biocatalysts unusable. Once a biocatlaytic process has been studied and understood *in vitro*, on a small scale, using purified enzymes a switch can be made to using lysates, which only requires cell disruption. This process is hugely important in industry, as well as in large scale academic processes, and in 2020 the cell lysis and disruption market size was valued at $4.09 million USD.[34] The choice between the three ways of performing biocatalysis will vary depending on the stage of research, the scale, the funding available, the type of enzyme, and a host of other factors.

1.2.5 Enzymatic Cascades

Engineering an enzyme to perform a single, highly specific reaction is an exceedingly useful tool, however, the synthesis of complex chemicals often requires several steps. In a traditional synthesis, isolation and purification of intermediates may be required at each step as the different reaction conditions, solvents, and side products can be incompatible. This process can be highly time consuming as well as being very solvent intensive. Here, enzymes are once again able to provide a greener, more sustainable alternative to purely chemical methods. In nature, enzymes are infrequently designed as standalone reactors, they are much more likely to be found in a cascade or cycle, of which there are countless examples.[27,35] This feature means that enzymes are the ideal candidate for designing several step syntheses which do not require protecting group chemistry or isolation of the reaction intermediates.

The expansive number of enzymes available, both commercially (Section 1.2.3) and in individual laboratories, means that designing fully, or partly, biocatalytic syntheses is now possible. The ideas of retrosynthesis, which have been used in chemical synthesis since the 1960s, can now be expanded to include biocatalysts in the toolbox of available transformations.[23] Enzymes can be used in place of various chemical reactions and catalysts to provide greener, more sustainable,

alternatives and even be used in novel routes to known compounds. Biocatalysts are often used commercially to replace a single step or in the synthesis of enantiopure starting materials, however, advances in synthetic biology mean that multi-step biocatalytic syntheses are frequently found in research laboratories.[36–38] These developments, combined with the ability to express several genes in a single organism, could lead to the creation of designer organisms capable of transforming a starting material into a single desired product over several enzymatic steps.[39]

One of the current goals of biocatalytic research is to replicate the conditions found within cells, where reaction rates and yields are significantly higher and cross reactivities are controlled.[40] Anabolic and catabolic reactions occur within the same cell and must be separated in order for the opposing pathways to be controlled, such as with glycolysis and gluconeogenesis.[41] Within cells, particularly eukaryotes, enzymes are often segregated into biological compartments with enzymes of related function. For example, in eukaryotes, the eight key enzymes involved in the Krebs Cycle are co-localised to the mitochondria, maximising flux through the pathway by limiting diffusion distances and preventing reactions with other compounds in the cytosol.[42] Applying these ideas to synthetic systems could provide the solution to many of the current problems faced by biocatalysis.

1.3 Compartmentalisation

Compartmentalisation by biomolecular membranes plays a key role in cells and in controlling cellular metabolism. These biomembranes not only maximise the internal surface area of a cell, to allow for all the key membrane bound processes to occur, they provide enclosed compartments with tunable properties to generate functionally specialised aqueous environments.[43] As mentioned previously, these compartments separate opposing enzymatic pathways, however, they can also separate incompatible reactions which require highly different reaction conditions. Within many eukaryotic organelles the pH is significantly lower than that of the cytosol, allowing the optimum environment for enzymes within the organelle without affecting those in the cytosol.[44] Compartmentalisation also overcomes many of the intrinsic problems associated with metabolic reactions: enzyme flux is optimised by controlling local substrate concentrations to increase kinetics; undesired side product production is limited by separating competing pathways; toxicity to the whole cell is reduced by controlling harmful product localisation; and reaction intermediates can be sequestered by having an internal membrane with a specific composition, which prevents loss through the cell membrane (Figure 1.3).[45]

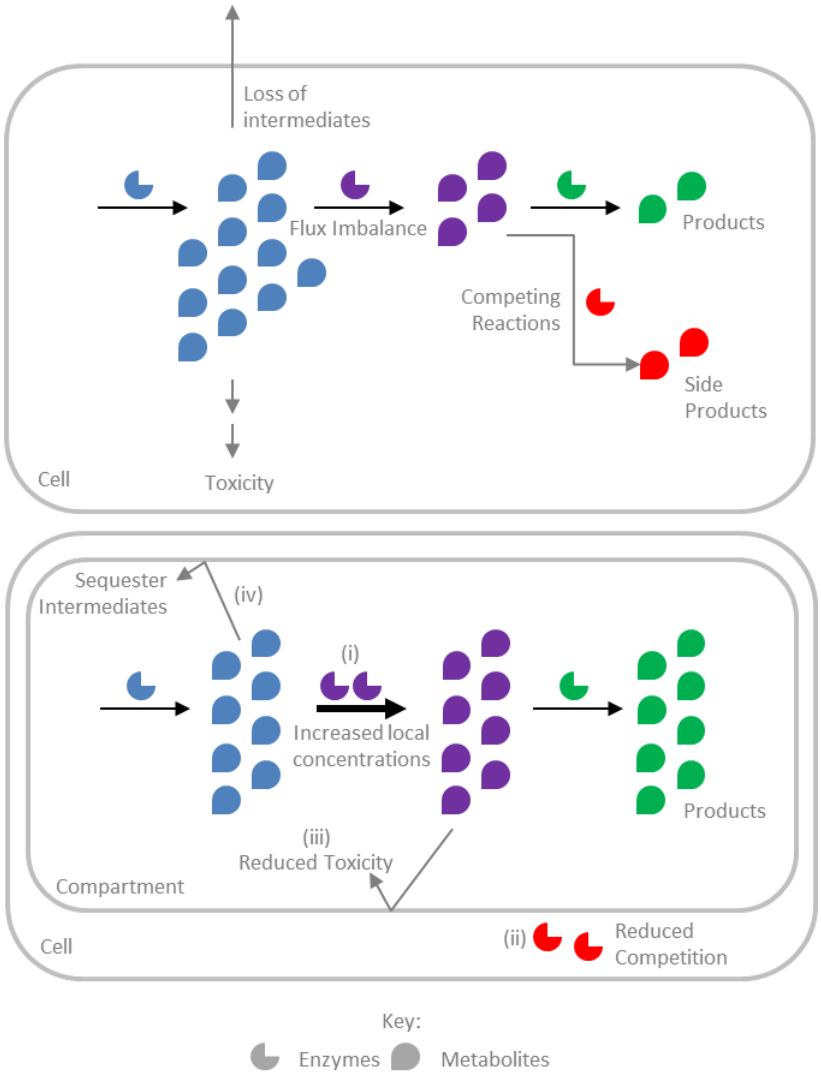


Figure 1.3: Benefits of cellular compartmentalisation: i) enzyme flux is optimised by controlling local substrate concentrations to increase kinetics; ii) undesired side product production is limited by separating competing pathways; iii) toxicity is reduced by controlling harmful product localisation; and iv) reaction intermediates can be sequestered to be used in further reactions. Figure adapted from Lee *et al.* [45].

Eukaryotic cells use phospholipid bilayer membranes to divide the cell into discrete organelles within the cytosol, as well as protecting the cells from the surrounding environment. Prokaryotic cells also use internal compartmentalisation through protein based microcompartments, such as carboxysomes and encapsulin shells,[46] as well as lipid membranes. Another method of compartmentalisation, not involving membranes, is through multi-enzyme complexes. These complexes contain enzymes from the same pathway to allow substrate channelling to occur.[47] The reaction intermediates are passed directly from one enzyme to the next without being released into the environment. This prevents the intermediates being lost, minimises cross reactions and increases reaction yields. These complexes also improve the flux through the pathway and prevent volatile or toxic intermediates from being lost or released into the cell. Several examples of these exist including polyketide synthase,[48] where the acyl carrier protein shuttles the intermediates

between the enzymes, and tryptophan synthetase[49] where the intermediates are directly channelled between the enzyme active sites.

1.3.1 Lipid Compartmentalisation

1.3.1.1 Natural Lipid Compartmentalisation

Cells from all domains of life utilise a variety of lipids to form their external plasma membranes and internal organelle membranes. There are three major classes of lipids found in biological membranes: glyercophospholipids, sphingolipids and sterols (Figure 1.4).[50] These lipids can be very structurally diverse, however, they are all amphiphilic, meaning that they can spontaneously form bilayers in aqueous environments. Glycerophospholipids are the major lipid component of biological membranes and are composed of two fatty acid tails attached to glycerol-3-phosphate, via an ester linkage. Additional polar groups, such as choline or ethanolamine, can then be attached to the phosphate OH to give structural diversity. Sphingolipids contain a molecule of sphingosine, a C18 amino alcohol, attached to a fatty acid chain, to give a ceramide. This ceramide can then be phosphorylated and attached to a polar group, to give sphingophospholipids, or can be attached to a variety of sugar groups to give a range of glycolipids. Sterols, namely cholesterol, are found in eukaryotic membranes at 10-20% of the total lipid composition. The rigid fused ring structure of sterols give them a high packing density and provide mechanical stability to the membrane.[50]

In eukaryotic cells, lipids are utilised throughout the cell from the plasma membrane, through the entire endosomal system, to the nucleus itself. Nearly 50% of the internal volume of eukaryotic cells is separated from the cytosol into discrete organelles in order to allow a huge number of metabolic reactions to occur concurrently (Figure 1.5).[52] Eukaryotic membrane-bound organelles can be broadly divided into two categories: those with single membranes, and those with double membranes. The entire endosomal system is made up of single-membrane organelles which are constantly exchanging cargo and membrane lipids through transport vesicles. The different organelles are distinguished by the reactions they can perform and the lipid composition of their membranes. Cholesterol and sphingolipids (lipid rafts) make membranes more rigid so are present at a higher concentration in the plasma membrane and vesicle-like organelles with a lower membrane curvature. The concentration of these is greatly decreased in the Golgi and endoplasmic reticulum as their tubular structures require a highly flexible membrane and a high percentage of glycerophospholipids.[53]

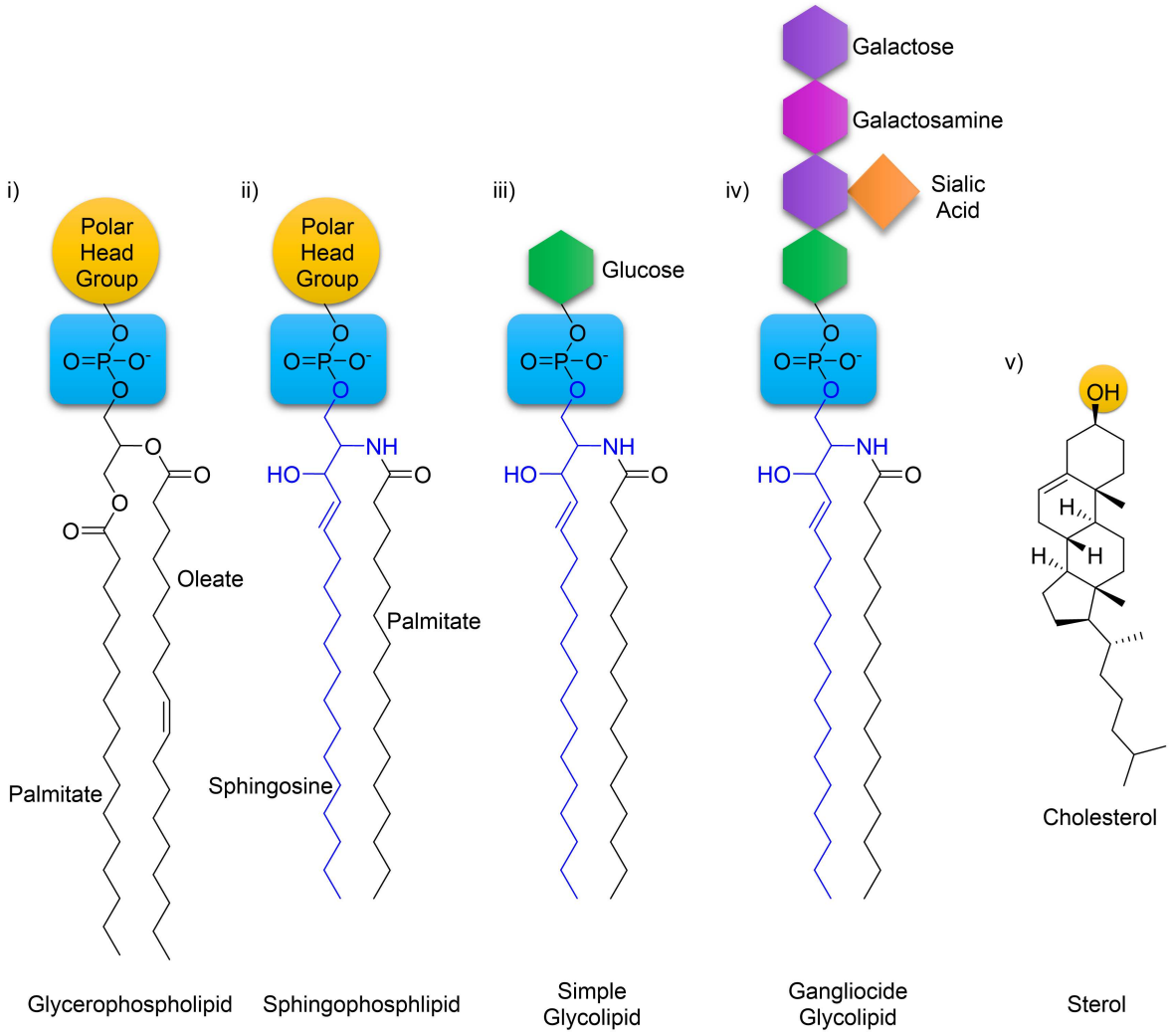


Figure 1.4: Structure of common membrane lipids from the three major classes: glyercophospholipids (i), sphingolipids (ii-iv) and sterols (v). Figure adapted from Van Meer and De Kroon [51].

The different organelles for endocytosis, from early endosomes through to lysosomes, all require a different internal pH to provide optimal conditions for their enzymes, and an increasingly harsh environment for internalised cargo. These compartments prevent damage to the cell from a harmful cargo before and after it has been broken down. These organelles also allow for any cargo to be trafficked to the correct location after endocytosis, such as to the Golgi, endoplasmic reticulum or recylcing endosome. The Golgi and endoplasmic reticulum networks are key in metabolic processes (Figure 1.5) as well as processing and transport of proteins from the nucleus. One of the most important organelles, the nucleus, is also enclosed by a single membrane which extends to form the endoplasmic reticulum. The nuclear envelope is a single contiguous membrane that has an outer face and an inner face; these meet at the nuclear pore complexes, where the inner and outer surfaces connect. The nuclear envelope protects the chromatin and provides the optimal reaction conditions for transcription and ribosome synthesis to occur.

Double-membrane organelles are those associated with endosymbiotic theory, chloroplasts and mitochondria. Endosymbiotic theory is that these organelles were bacteria endocytosed by a pre-karyote which were not digested and provided an evolutionary advantage to the host organism. The double membrane is made up of the pre-karyotic plasma membrane which engulfed the bacteria during endocytosis and the bacterial membrane. This double membrane is important for both mitochondira and chloroplasts, not just for separating their reactions from the cytosol, but providing an inter-membrane space which can be used to generate the proton gradient required for ATP synthesis. The outer membrane protects the rest of the cell from the low pH present in the inter-membrane space and keeps the protons from being lost into the cytosol.

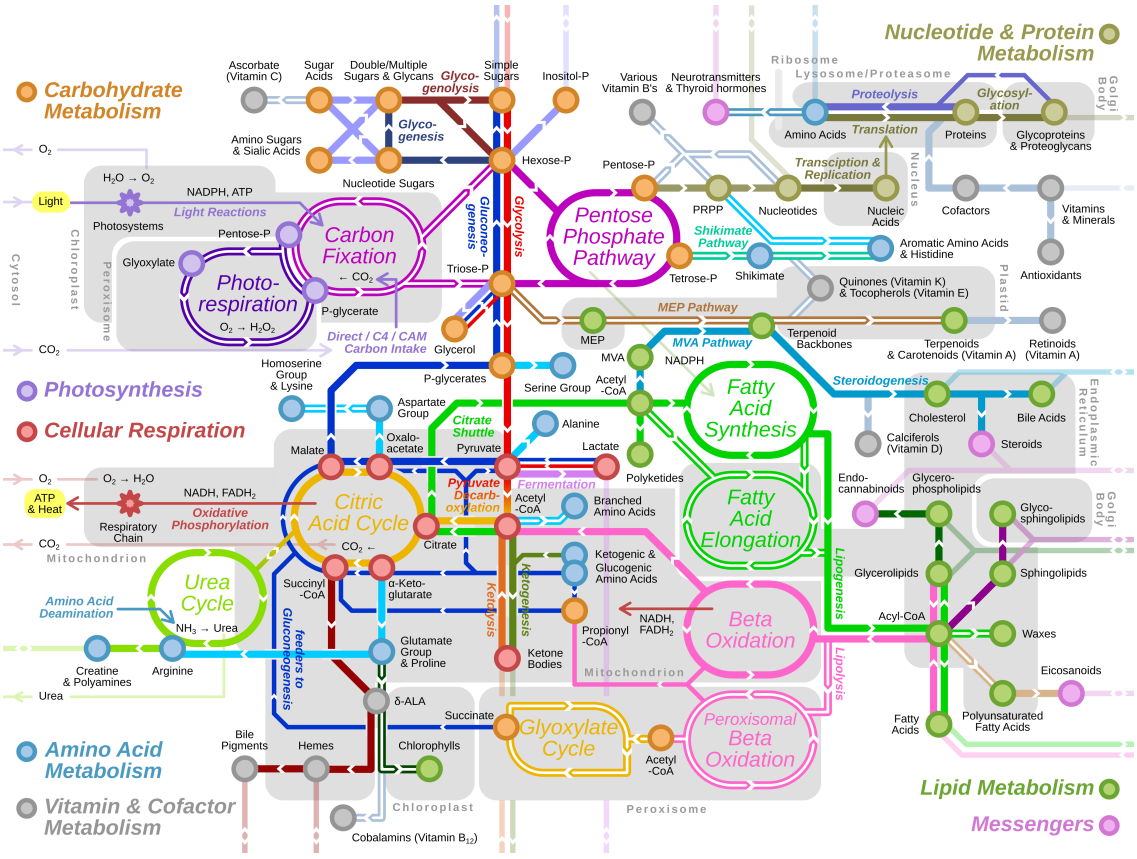


Figure 1.5: Metabolism Map showing the many metabolic reactions carried out within eukaryotic cells and those which are compartmentalised into various lipid membrane bound organelles. Figure taken from Metro Map of Metabolism - The Overview [54].

Prokaryotic cells were thought for many years to have little or no internal organisation, however, this has been shown to be untrue. As in eukaryotic cells, a lipid bilayer is used to form the plasma membrane and separate the cytosol from the external environment. Prokaryotic cells can also contain both monolayer and bilayer internal lipid structures in the form of specialised organelles. Monolayer structures are used generally for storage such as with lipid bodies and vacuoles. Bilayer

structures are more diverse and are more similar to 'traditional' eukaryotic organelles. These organelles give the bacteria specialised functions and are usually specific to a particular taxonomic group. Magnetotatic bacteria utilise geomagnetic field lines to guide them to their preferred redox conditions using specialised organelles called magnetosomes.[55] Magnetosomes consist of a lipid bilayer surrounding a magnetic crystal of magnetite (Fe_3O_4) or greigite (Fe_3S_4). These can be arranged into chains allowing the bacteria to passively line up along the magnetic field lines. Several unique proteins are associated specifically with the magnetosome membrane to provide a more acidic environment and a point of nucleation for the iron crystals to form.[56]

Photosynthetic bacteria also contain a number of different types of lipid bilayer organelles. There are three main groups of bacteria which utilise photosynthesis: purple photosynthetic bacteria, containing chromatophores; green photosynthetic bacteria, containing chlorosome compartments; and cyanobacteria, the prerequisite to chloroplasts, containing various thylakoid membranes. In photosynthesis the lipid membrane not only provides a location for the integral photosynthetic proteins, but also provides a physical barrier to allow the required proton gradient to be generated. This proton gradient impacts the pH of both sides of the membrane, so isolating this process from the rest of the cell is important to ensure no other metabolic processes are impacted.

Chromatophores appear as membrane associated organelles or invaginations of the inner membrane. They house the photosynthetic machinery required by numerous photosynthetic bacteria. These membranes are highly folded to provide the maximum number of sites for ATP synthesis. Chlorosomes are larger, simpler organelles also associated with, and formed from, the cytoplasmic membrane and are filled with bacteriochlorophylls. These pigments accumulate in the inter-membrane space of the inner membrane and cause it to swell to form flattened, ellipsoidal structures surrounded by a monolayer of lipids. This accumulation of bacteriochlorophylls, 150,000–300,000 per organelle, allows for the green photosynthetic bacteria to harvest light even at the lowest light intensities, including 2390 metres below the surface of the Pacific Ocean using only infrared light from hydrothermal vents.[57] The final type of bacterial photosynthetic organelles are the thylakoid membranes. These membranes, like the chromatophores, provide the sites for photosynthesis but are not associated with, or made from, the cytoplasmic membrane. They form stacks of lipid bilayers around the whole interior of the cell, which are connected through lipid bilayer bridges and not only facilitate photosynthesis but also long range transportation and communication across the organelle.

The most similar prokaryotic organelles, and probable bacterial ancestors, to eukaryotic organelles are found in the phylum *Planctomycetes*. The pirellulosome, found in *Pirellula*, is a large lipid membrane bound organelle which acts as a pseudo-nucleus separating the chromosome and ribosomes from the cytosol. In *Gemmata obscuriglobus* this organelle appears even more like a eukaryotic nucleus, with the chromosome compacted and housed in a double-membrane to form a nuclear body with lipid membrane rearrangements to form pores. The ribosomes are both in and out of the nuclear body physically, separating some transcription and translation reactions as in eukaryotic cells.

1.3.1.2 Engineered Lipid Compartmentalisation

Lipids were the first biomolecules used to form synthetic compartments in the form of simple vesicles. The formation of vesicles, both in nature and synthetically, relies upon the amphiphilic nature of the molecules used.[58] In order to minimise the unfavourable interactions of the hydrophobic regions with an aqueous environment, the amphiphilic molecules self-assemble into fully enclosed shells that have an aqueous interior and exterior (Figure 1.7). These amphiphilic molecules are traditionally phospholipids, such as phosphatidylcholines (PCs), phosphatidyl ethanolamines (PEs) and sphingomyelins which are all found in different cellular membranes and usually form a bilayer structure.[59] More recently polymersomes have also been investigated as a more stable, synthetic alternative to liposomes, particularly in relation to drug delivery. Polymersomes are composed of block copolymers, which consist of two or more 'blocks' of different polymer monomers chemically attached to each other. Depending upon the properties of these monomers, the block copolymers form distinct mono- or bilayer structures.[60]

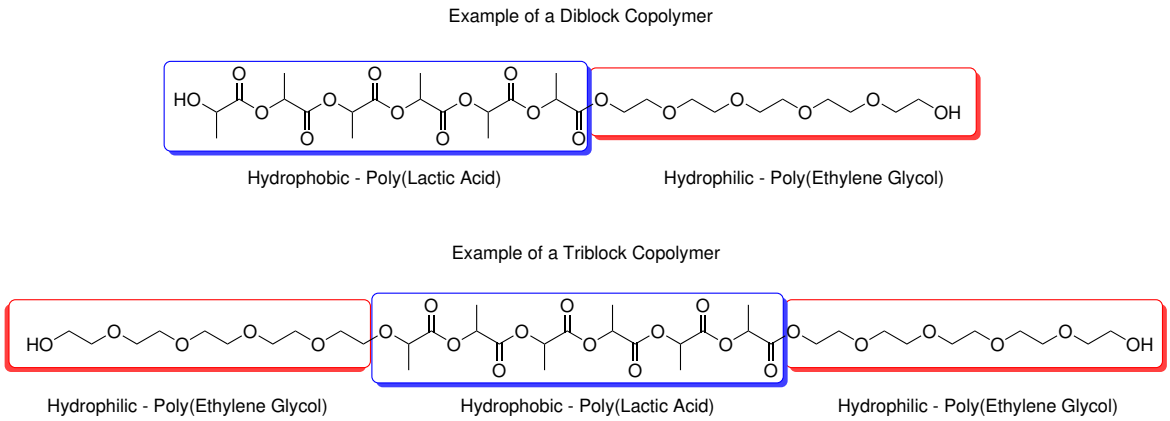


Figure 1.6: Examples of Block Copolymers - The block copolymers which make up polymersomes are amphiphilic molecules most often being di- or tri-blocks. Di-blocks are similar to lipids having a hydrophilic and a hydrophobic portion and can form mono- or bilayers. Tri-blocks have a hydrophobic centre surrounded by two hydrophilic blocks and will form a monolayer which behaves like a liposome as it can have an aqueous centre and exterior.

In comparison to liposomes, polymersomes are more versatile as the physical properties of the monomers can be finely tuned to precisely alter the vesicle properties, such as membrane thickness, stability and permeability.[61] Polymersomes are more stable than liposomes, being able to withstand greater physical stresses, as well as being more resistant to chemical and biochemical degradation. They are also less permeable than liposomes making them an appealing target for controlled and long-term drug delivery and release.[62]

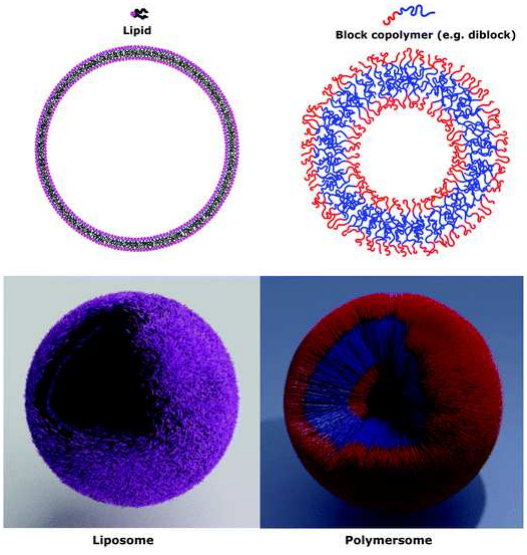


Figure 1.7: Structure of liposomes and polymersomes. Liposomes and polymersomes are both composed of amphiphilic molecules with liposomes using phospholipds and polymersomes using block copolymers. Figure adapted from Rideau *et al.* [63]

When investigating cell membranes mimics, however, liposomes are still the predominant choice as they are often composed of the lipids found in native cell membranes. Being composed of lipids means that liposomes have many of the properties of cell membranes and are also

much more compatible with other biological components, such as membrane proteins and DNA. Polymersomes can be punctured by protein membrane pores, however, there are only a few native pores which have the stability to withstand being combined with a synthetic membrane, limiting the usefulness of this feature.[64] Although being more permeable, the greater membrane fluidity in liposomes provides a more dynamic and responsive membrane, again more closely replicating the features of a true cell. Liposomes can also be tuned depending upon the properties of the monomers used, although, this process is more challenging than for the polymer monomers.[61] Different lipid monomers can be used individually or in combination with other lipids to alter the membrane properties such as permeability and size.

Liposomes are classed depending upon their size and number of bilayers with each class being used for different applications. Giant unilamellar vesicles (GUVs) have a diameter greater than one micrometer, similar to a cell, and are most often used as reduced cell mimics. Small unilamellar vesicles (SUVs, <100 nm) are used to study membrane protein activity and membrane interactions, while large unilamellar vesicles (LUVs, 100-1000 nm) are used as biological drug carriers. Various vesicles can also be used in combination with each other or other vesicles, like the polymersomes, to even more closely mimic a cell and its internal compartmentalisation.[65] The ability to closely mimic cellular compartmentalisation means that lipid based vesicles provide a useful platform from which to investigate native cellular processes as well as utilising them for synthetic purposes.

1.3.2 Protein Compartmentalisation

1.3.2.1 Natural Protein Compartmentalisation

Bacterial Microcompartments

The most common protein based compartments are the bacterial microcompartments (BMCs), spherical organelles formed entirely of shell proteins with a diameter of approximately 100 nm.[66] The proteinaceous membranes of BMCs have different properties to phospholipid membranes to allow polar, rather than non-polar, molecules to freely diffuse across the membrane. The first BMCs to be discovered were from the chemoautotrophic bacterium *Halothiobacillus neapolitanus*, which contained carbonic anhydrase and an integral enzyme of the Calvin cycle, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO).[67] These BMCs were named carboxysomes due to their ability to increase the local concentration of carbon dioxide to maximise the very slow reaction rate of RuBisCO. This is thought to be due to the very low permeability of the membrane

to gases, which limits the loss of carbon dioxide, generated by carbonic anhydrase, and limits oxygen entry, to minimise the production of the toxic by-product 2-phosphoglycolate from the oxygenase reaction.[68]

Carboxysomes are the only known anabolic type of BMC, they build molecules from smaller subunits, all others which have been discovered are catabolic, they break large molecules down; these are the metabolosomes and they are found across bacterial phyla. Metabolosomes are further subdivided into groups depending upon the substrates they use, and are usually only expressed when this substrate is present. Despite the wide range of substrates used by metabolosomes, they share a common core of enzymes: a specific 'signature enzyme', which converts the substrate to an aldehyde; an alcohol dehydrogenase, to convert the aldehyde to the alcohol product; an aldehyde dehydrogenase (AldDH); and an phosphotransacylase (PTAC) (Figure 1.8).[69] The AldDH and PTAC are involved in a separate reaction pathway to recycle NADH and coenzyme A generating a phosphorylated product. There are six main classes of metabolosome which have been characterised: PDU BMCs which utilise 1,2-propanediol; EUT BMCs which utilise ethanolamine; PDU/EUT which can utilise either substrate; GRMs (Glycyl Radical Enzyme-Containing Microcompartment) contain a glycyl radical enzyme and an activating enzyme which act as the 'signature enzyme' and some utilise deoxy sugars such as L-fucose and L-rhamnose; PVMs (Planctomycetes and Verrucomicrobia-Type) are found exclusively in the phyla Planctomycetes and Verrucomicrobia and utilise L-fucose and L-rhamnose; and ETU BMCs which utilise ethanol. There are also a number of metabolsome loci which have been identified, but their substrates and enzymes have yet to be characterised.[69]

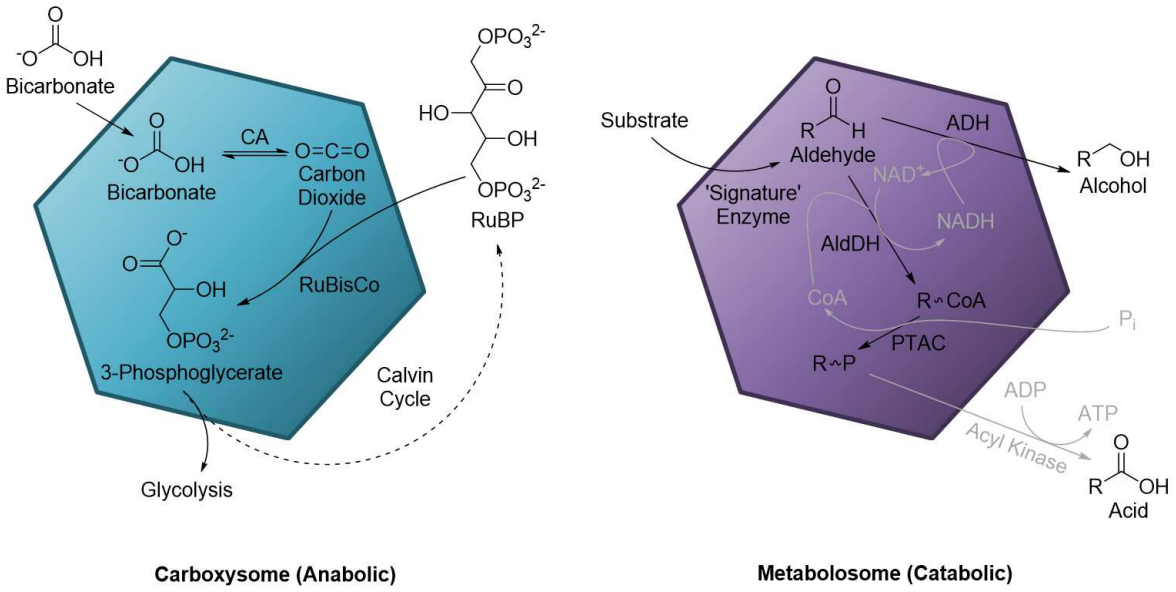


Figure 1.8: Generalised schemes of characterised bacterial microcompartments: a carboxysome (left), the only know example of an anabolic BMC, performing the key carbon fixation reaction of the Calvin Cycle; and a catabolic metabolosome (right), showing the BMC core enzymes found in all subtypes required to generate the main alcohol product. Peripheral reactions to regenerate the required cofactors are shown in grey. Figure adapted from Axen *et al.* [69].

BMC shells are all composed of the same shell protein building blocks: BMC-H, contain a single copy of the Pfam 00936 domain and form hexamers; BMC-T contain a fusion of two Pfam 00936 domains and form trimers, which appear as pseudohexamers; BMC-P contain a single Pfam 03319 domain and assemble into cyclic pentamers.[69] The hexamers and pseudohexamers of BMC-H and BMC-T form the faces of the icosahedra while the pentamers of BMC-P form the twelve vertices to completely seal the shell. The shell proteins are packed tightly together to provide a diffusional barrier, however, the (pseudo)hexameric structures contain central pores.[70,71] These pores allow molecules to enter and leave the shell and, opposite to phosphoslipid membranes, have a preference for polar or charged molecules.[66,72]

Virus Capsids

Structurally similar, but evolutionarily unrelated, to BMCs are virus capsids. These nanometre-sized protein shells are designed to encapsulate viral genomes, transport them, and release them into a new host.[73] These capsids are symmetrical shells composed of protein subunits, capsomeres, which are in turn formed from aggregates of one or a few viral proteins. These viral proteins, known as major capsid proteins, come in a range of structures and can range from 42 to over 1400 amino acids (Uniprot). This subunit assembly of a capsid means that it can: self-assemble, requiring no energy input; minimise the genomic requirements for structural proteins; and generate a more stable structure with numerous repeating, tightly packed interactions.[74] Viral

capsids exist in three main forms (Figure 1.9): helical, with the capsomeres assembling into circles and in turn into a tube structure to surround a single nucleic acid molecule; icosahedral, which are similar to BMCs with hexameric and/or pentameric subunits making up a three dimensional cage to house the viral genome; and complex, these capsids do not fall under the first two categories but are generally found in the more complex bacteriophages, such as the T4 bacteriophage. The icosahedral structures, similar to BMCs, consist of a minimum of 20 faces and required 60 subunits, 3 non-triangular subunits per triangular facet. The majority of viral capsids are composed of a multiple of 60 subunits with increasing numbers of hexameric units between the pentameric verticies (Figure 1.9).[75]

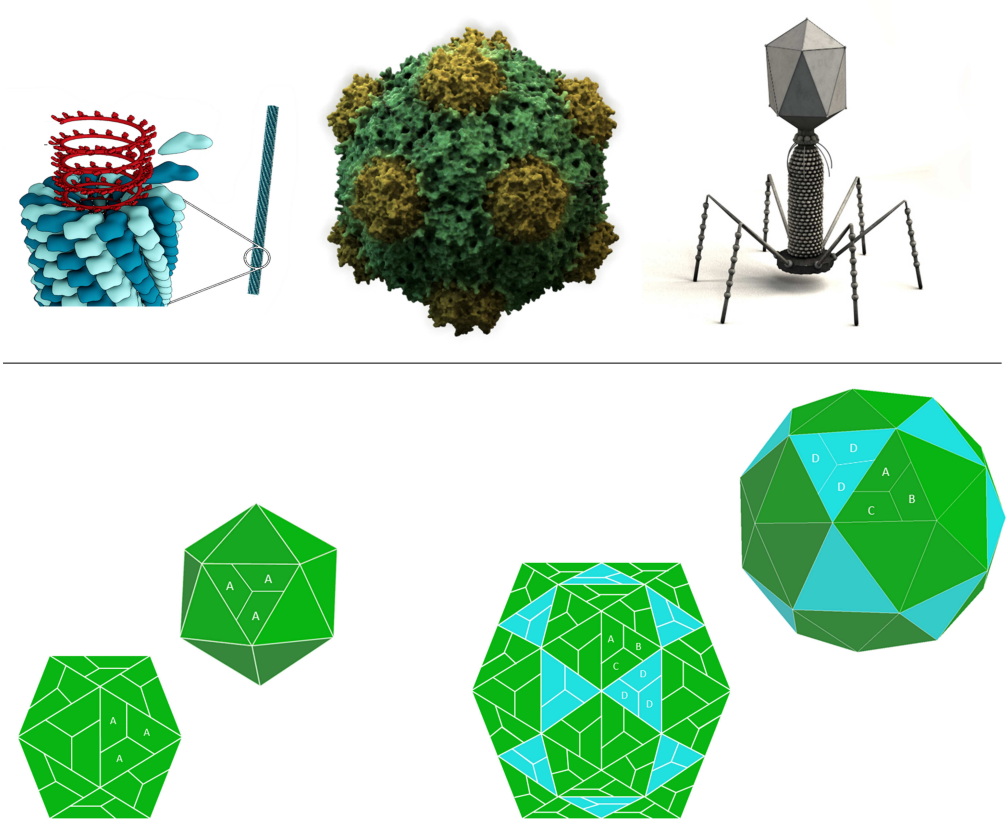


Figure 1.9: Viral capsids - i) Examples of the three main forms of viral capsids: helical - tobacco mosaic virus; icosahedral - cowpea mosaic virus; and complex - T4 bacteriophage. ii) Icosahedral structures showing how size increase can be achieved by introducing hexamers between the pentamers.

Encapsulins

Encapsulins are the smallest proteinacious nanocompartments at 24-42 nm and are structurally similar to both virus capsids and BMCs, forming icosahedral structures. Encapsulins are more structurally similar, and evolutionarily related, to virus capsids, with their shell proteins being homologous to the gp5-HK97 phage main capsid protein, and their icosahedral shells being comprised of a multiple of 60 subunits.[76] Encapsulins are far simpler than virus capsids as their shells are comprised of a single protomer, which self-assembles into the encapsulin structure. Encapsulin

shells form a number of small pores, 5-6 Å, at the points of particle symmetry and between the subunits which can be negatively, neutrally, or positively charged to control the passage of molecules between the cytosol and encapsulin lumen.[77] Internally, encapsulins are more similar to BMCs as they have evolved to house a range of cargo proteins which are encapsulated via a cargo loading peptide, a short conserved sequence which targets the cargo to the encapsulin, or are fused directly to the shell proteins.[78] Encapsulins are involved in the metabolism of bacteria and archaea, however, there is limited knowledge as to the specific role encapsulins play. Encapsulins have been shown to house, experimentally, ferritin-like proteins and DyP-type peroxidases, and putatively, haemerythrins and di-iron proteins.[79] The majority of these proteins are involved in iron-storage and/or mitigation of oxidative stress which are key for cellular survival.

1.3.2.2 Engineered Protein Compartmentalisation

The various proteinacious compartments discussed above can be readily engineered for synthetic applications. Protein compartments, comparatively to lipid vesicles, have been investigated to a much lesser degree as a possible engineering target for compartmentalisation, due to their more recent discovery and characterisation.

In order for BMCs to be useful for forming synthetic vesicles, the genes for forming and assembling the shells needed to be identified and be able to be heterologously expressed. The genes for Pdu, Eut, and carboxysome shell proteins were identified in *Citrobacter freundii*, *Salmonella enterica*, and *Halothiobacillus neopolitanus* respectively.[80] All of these BMC shell proteins, either alone or part of the full operon, were expressed heterologously in *E. coli* and were able to successfully produce empty, or fully functional, BMCs. The BMCs self-assemble in one of two ways: inside out assembly, where an enzyme core forms before the shell assembles around it; and simultaneous assembly, where enzymes and shell proteins interact to form the core and shell simultaneously.[81,82] However, the process by which this occurs to give stable structures with the correct stoichiometry is unclear. It is thought that the electrostaticity of the shell proteins combined with some steric effects may lead to the self-assembly.

Proteins are targeted to the interior of BMCs through a short ∼20 amino acid sequence termed the encapsulation peptide (EP) which binds to the shell proteins during BMC assembly.[83] This EP sequence can be genetically fused to a non-native protein, such as GFP, to target it to the interior of the corresponding BMC. The EPs can be the native sequence for the BMC or can be synthetic to improve features such as cargo capture efficiency.[84] This process can also be applied

to other enzymes to generate novel nanoreactors which separate the enzyme from the cytosol without impacting its reactivity even when they are multimeric or are cofactor dependent.[85] The modular design of BMCs make them an appealing target for engineering, however, the complexity of BMCs present a number of challenges including inefficient assembly in heterologous hosts. Developments in rational design of novel BMCs and *in vitro* production of the various components are required for BMCs to be a readily available method for enzyme encapsulation.[86]

Virus capsids have evolved to house and protect RNA and DNA from degradation within a confined volume, with a well defined structure making them ideal candidates for nano-reactor formation. The capsid shell proteins, like the BMC shell proteins, can be heterologously expressed and functionalised by directed packaging of a desired cargo.[87] The first capsid nanoreactor was generated when horseradish peroxidase was loaded into a Cowpea Chlorotic Mottle Virus (CCMV) capsid utilising the pH dependent disassembly/assembly process of this virus.[88] Encapsulation of specific enzymes was achieved through use of heterodimeric coiled-coil peptide oligomers, with one tethered to a cargo protein and a corresponding oligomer attached to a capsid protein allowing for assembly of a capsid around a cargo protein.[89] This encapsulation was also possible by functionalising a cargo protein with DNA, which enable it to be readily incorporated into a capsid through non-covalent electrostatic interactions.[90] Alternative methods for encapsulation include using protein–RNA interactions which use RNA aptamers to bind to protein tags such as Rev Tags; and SpyTag/SpyCatcher systems which form a covalent isopeptide bond between an already immobilised SpyCatcher protein and a short SpyTag attached to the protein of interest.[91,92]

While CCMV, and similar viruses, form small capsids and are limited to the number of enzymes they can house, the bacteriophage P22 can form a far larger capsid and has been used as an alternative nanoreactor. Cargo proteins can be fused to a scaffolding protein tag, similar to the encapsulation peptide for BMCs, which is required for capsid assembly and allows for co-encapsulation of multiple proteins.[93] The large size of this capsid allowed for generation of an encapsulated three enzyme cascade for the metabolism of lactose (Figure 1.10).[94] The three enzymes were co-expressed as a multi enzyme fusion with a scaffolding protein tag which could then be encapsulated into the capsid providing enhanced enzyme activity and stability. This principle could be applied to a broad variety of enzymatic cascades to generate countless nanoreactors.

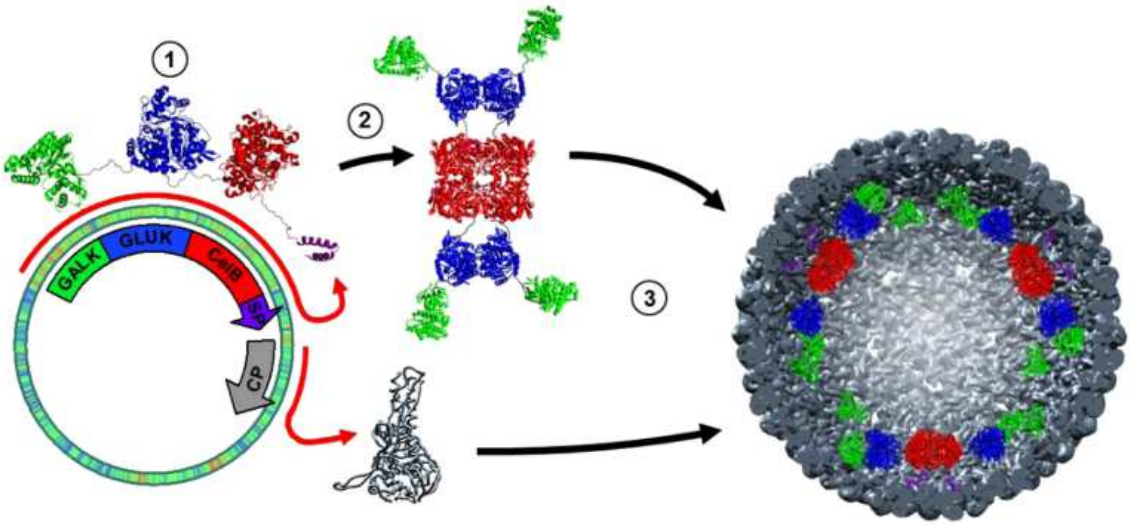


Figure 1.10: Encapsulation of Enzymes for Lactose Metabolism - Three enzyme for lactose metabolism, CelB, GALK, GLUK, were encapsulated within a P22 virus-like particle. 1 - Expression of the coat protein (CP) and the three enzyme fusion with the scaffolding protein domain (SP); 2 - association of the multi-enzyme fusion to form the CelB tetramer and GLUK dimer; 3 - capsid formation facilitated by the interactions of the SP domains and CP subunits leading to encapsulation of the multi-enzyme fusion. Figure adapted from Patterson *et al.* [94]

Encapsulins, like BMCs, have evolved as natural protein nanoreactors making them easier to engineer as synthetic nanoreactors. As with both BMCs and virus capsids, the targeting peptide for encapsulation can be commandeered for synthetic purposes. In one key example of encapsulin nanoreactors, the EncA gene from taken from *Myxococcus xanthus* and expressed in *Saccharomyces cerevisiae*, alongside a pyruvate decarboxylase enzyme (Aro10p), with a targeting peptide. This generated a self-assembled artificial organelle capable of converting 4-hydroxyphenylpyruvate to 4-hydroxyphenylacetaldehyde, a key step in the synthesis of benzylisoquinoline alkaloids (Figure 1.11).[95] This encapsulin from *M. xanthus* was also expressed in mammalian cells with the three native cargo proteins (*EncB,C,D*), as well as split luciferase fragments fused to the native proteins to generate enhanced bioluminescence.[96] They also incorporated a tyrosinase fused to a native protein to generate melanin from tyrosine, a toxic process which was successfully contained within the encapsulin as a non-native organelle.

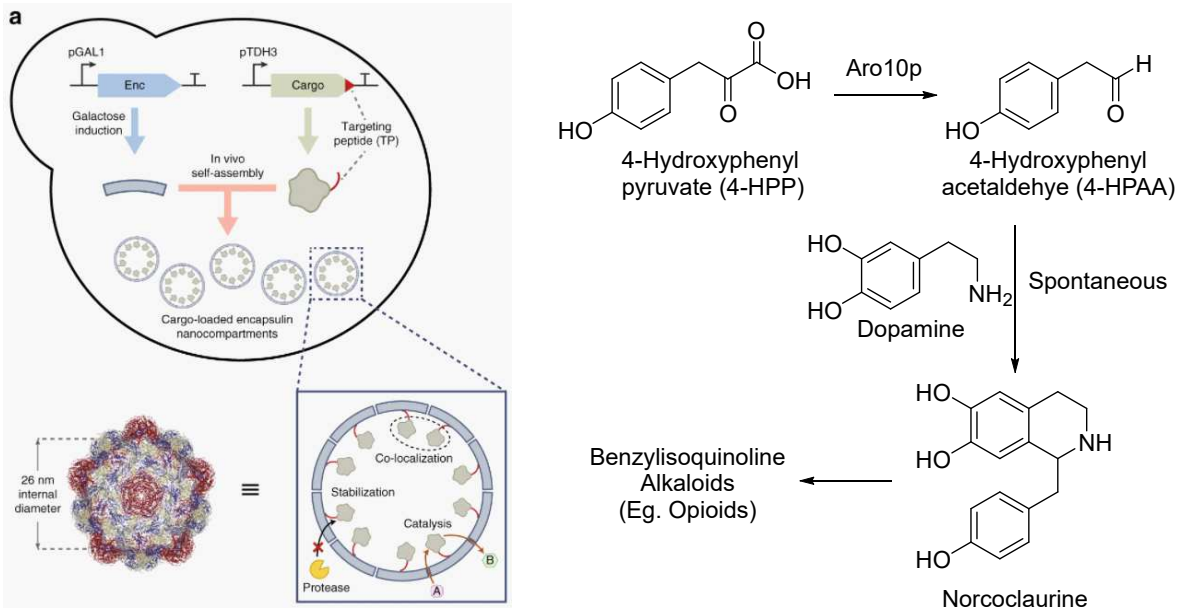


Figure 1.11: Encapsulation of Aro10p for Synthesis of Benzylisoquinoline Alkaloids - The EncA gene, which forms an encapsulin shell, was co-expressed with Aro10p fused to a targeting peptide to form an encapsulin nanoreactor in yeast. This nanoreactor was capable of converting 4-hydroxyphenylpyruvate (4-HPP) to 4-hydroxyphenylacetaldehyde (4-HPAA), a key step in the synthesis of benzylisoquinoline alkaloids. Figure adapted from Lau *et al.* [95]

De novo protein design has also been utilised in the pursuit of novel self-assembling nanoreactors to generate protein structures such as a 60-subunit icosahedral structure using engineered keto-hydroxyglutarate aldolase from *Thermotoga maritima* to form the shell.[97] This novel engineered structure has also been developed as a scaffold for a nanoreactor by fusing (+)-γ-lactamase to the modified aldolase, which self-assembled to encapsulate the lactamase.[98] The lactamase enzymatically resolves Vince lactam, a key intermediate in the synthesis of carbocyclic nucleoside medicines, and this activity was retained upon encapsulation, whilst the enzyme stability to temperature, pH, solvent and proteases was greatly increased. This technology, along with the other engineered compartments, could be broadly applied to sustainable synthesis and green chemistry to generate a range of nanoreactors capable of performing product synthesis on an industrial scale.

1.3.3 DNA and Other Synthetic Compartmentalisation

A truly synthetic form of compartmentalisation not observed anywhere in nature involves the use of DNA, not as genetic material, but as a building material.[99] In nature, DNA is a biopolymer composed of two polynucleotide chains which form a double helix which encodes the genetic information for organisms across all domains of life as well as for some viruses. The subunits or monomers of DNA are the nucleotides which are made up of a pentose sugar, deoxyribose for DNA or ribose for RNA, a phosphate group, and a nucleobase[100] (Figure 1.12). The nucleotides

are connected together through a phosphodiester bond between the 3' carbon of the ribose sugar on one nucleotide and the phosphate group of the next to form the sugar-phosphate backbone. There are five standard nucleobases: the bicyclic purines, guanine (G) and adenine (A); and the monocyclic pyrimidines, cytosine (C), thymine (T, DNA only), and uracil (U, RNA only). Hydrogen bonds form specifically between pairs of these bases, one on each polynucleotide chain when the chains are anti-parallel (running in the opposite direction) to one another, two bonds between A and T/U and three between G and C known as Watson–Crick–Franklin base pairing.[101]

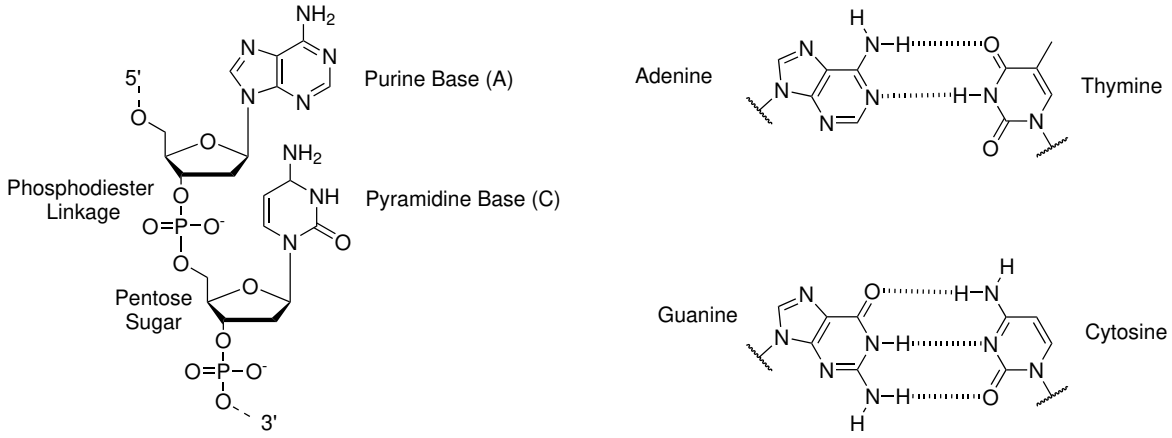


Figure 1.12: DNA structure and base pairing - Left - Two nucleotides connected with a phosphodiester bond showing the pentose sugar, phosphate group, and nucleobase. Right - Adenine and thymine/uracil form two hydrogen bonds, whilst guanine and cytosine form three. Figure adapted from Voet *et al.* [100].

DNA can adopt a number of different confirmations, with the three biologically active double helical structures being B-, A-, and Z-DNA[102] (Figure 1.13). B-DNA is the most biologically common form of DNA having the structural features first observed by Watson, Crick, and Franklin: a right-handed double helix with 10 bases per turn, a diameter of 20 Å, with the nucleobases perpendicular to the helix. A-DNA is observed under high salt, dehydrating conditions, which simply compresses the B-DNA structure to give a shorter, more compact helix with a wider diameter of 26 Å, 11.6 bases per turn, and narrow and deep major grooves.[103] The nucleobases in A-DNA are also puckered away from the perpendicular and the sugars adopt a different conformation. Z-DNA is the odd one out of these three forming a left-handed helix, under high salt conditions, with a sequence of strictly alternating purines and pyrimidines. Z-DNA has a diameter of 18 Åwith 12 bases per turn, and a flat major groove.

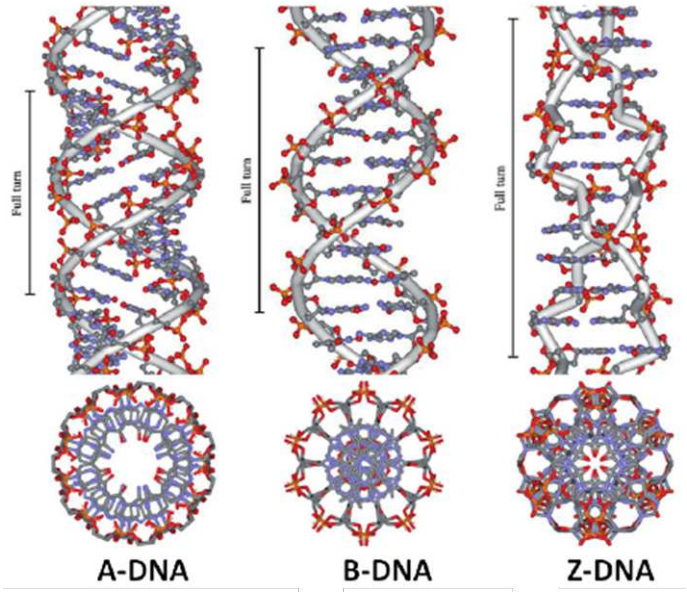


Figure 1.13: DNA Conformations - The three biologically active conformations of DNA, B-DNA, A-DNA, and Z-DNA. Figure adapted from Carlos García-Ramos *et al.* [104].

The specific base pairing of DNA can be exploited to generate structures able to act as micro-compartments using a technique known as DNA origami.[105] This technique uses the strong and specific base pairing observed between the nucleotide bases to form two-dimensional structures. DNA origami uses a single long scaffold strand, which runs throughout the structure and several shorter oligonucleotides called staple strands. The shorter staple strands specifically bind to sections of the scaffold strand in order to manipulate it into a certain orientation. The combination of all of the manipulations of the structure by the staple strands work together to form the final structure. This method has been extended to form three-dimensional structures, allowing the formation of nanocompartments. The most successful method for achieving this involves forming polyhedral meshes from the scaffold strand, which can self-assemble into a final three-dimensional structure[106] (Figure 1.14). These structures have ranged from simple cubes, to the Stanford Bunny, to dynamic structures[107] such as a waving man and an enclosed box able to open and shut with "DNA keys".[108]

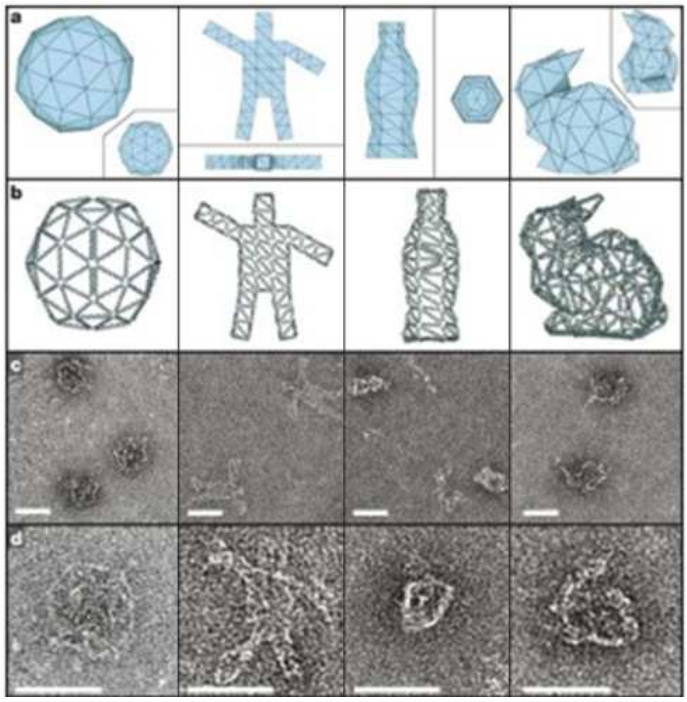


Figure 1.14: DNA origami structures showing the design, TEM, and cryo-EM images of various strucutres. The structures include a ball, a stickman, a soda bottle and a reduced polygon version of the Stanford bunny. Figure taken from Benson *et al.* [106].

The specificity of DNA base pairing allows the formation of these structures to be highly self-controlled and, due to the developments in design and synthesis of DNA, can be designed with ease and specificity for individual applications, making them attractive potential nanoreactors. The ability of DNA to form a variety of coiled structures, and to break and reform its base pairings, makes it suitable for creating dynamic structures capable of changing shape in response to oligonucleotide binding or environmental change. These structures, therefore, have potential as nanoreactors with applications in medicine.

Another synthetic method of enzyme encapsulation that has been widely used to generate heterogeneous biocatalysts are sol-gels. Enzymes are trapped within a gel matrix, most commonly consisting of silicon alkoxide monomers, by chemical condensation of the silica gel network around the enzyme[109] (Figure 1.15). This process does not require harsh conditions so the enzymes can survive entrapment, and as they are trapped physically, they also do not have to be chemically modified to be encapsulated. The properties of these silica gels, such as pH, gelation time, shaping, or hydrophobicity, can be varied to provide the optimal conditions for the enzyme being studied.

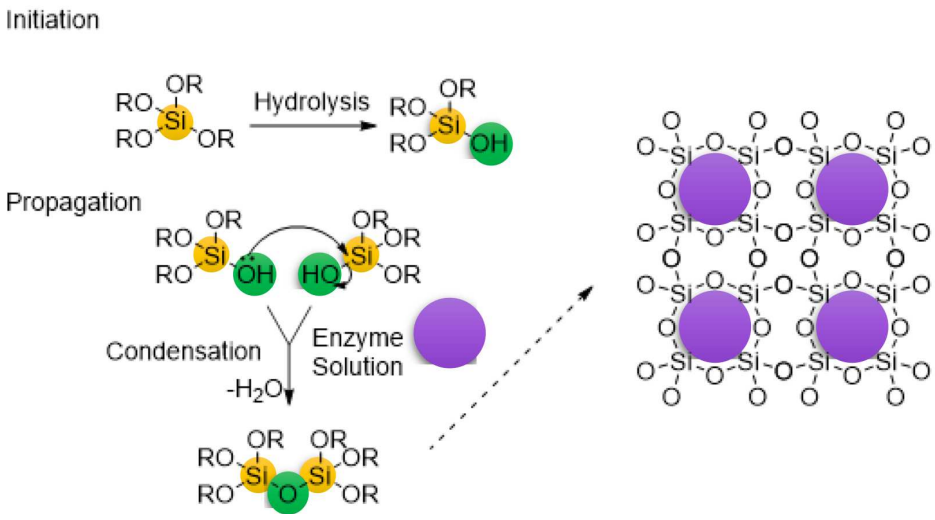


Figure 1.15: Sol-Gel formation: The silica alkoxide ($Si(OR)_4$) is hydrolysed to give $Si(OR)_3(OH)$ which undergoes a condensation reaction, encapsulating an enzyme (purple) in the gel network. Figure adapted from Campás and Marty [110].

One of the earliest uses of these sol-gels was for entrapment of alkaline phosphatase, which only retained 30% enzyme activity but showed a greatly enhanced stability, retaining full activity in water for two months. [111] The applications of this technique were shown when glucose oxidase and peroxidase were co-entrapped and were able to act as an optical glucose sensor. [112] Since the first studies, many biosensors have been investigated alongside various enzymes and matrices including lipases, horseradish peroxidase, and trypsin. [109,113,114] Sol-gels have also been shown to be capable of encapsulating whole cells and have interesting potential as biomedical devices as well as providing a robust system for stabilising and recycling enzymes for synthetic and sensor uses. [115,116]

Another, more recent, compartmentalisation method uses peptide and lipid nanotubes to encapsulate enzymes in the internal environment of the nanotube. The peptide or lipid monomers self-assemble into nanotubes through non-covalent interactions such as van der Waals forces, hydrophobic and ionic interactions, and hydrogen bonding. [117] A variety of peptides and lipid types can be used to achieve this, enabling tuning of the surface for different applications. [118] Enzymes are bound to the internal surface through a simple incubation process where hydrogen bonding occurs between amide groups of the nanotube and the complementary functional groups of the enzyme. [119] The enzymes are initially adsorbed to the interior of the nanotube due to the capillary effect so, if enzyme concentrations are controlled, adsorption will only occur on the internal surface. [120] A model enzyme, *Candida rugosa* lipase, was encapsulated within a protein nanotube and its hydrolysis rate of *p*-nitrophenyl butyrate was increased by 33% (Figure 1.16). [121]

As with other encapsulation methods, the nanotube also greatly increased the enzyme stability resulting in 70% higher activity, than free enzyme, at 65 °C. Horseradish peroxidase was also encapsulated within a protein nanotube and showed greatly enhanced stability against temperature, denaturants, and retained 90% activity after 18 days storage.[122] Protein and lipid nanotubes have an interesting application as potential biosensors due to their biocompatability, tunability, and stability. They also have applications as potential nanoreactors and could be used for enzyme recycling as they can be decorated with magnetic particles making for easy removal in a process system.[121,123]

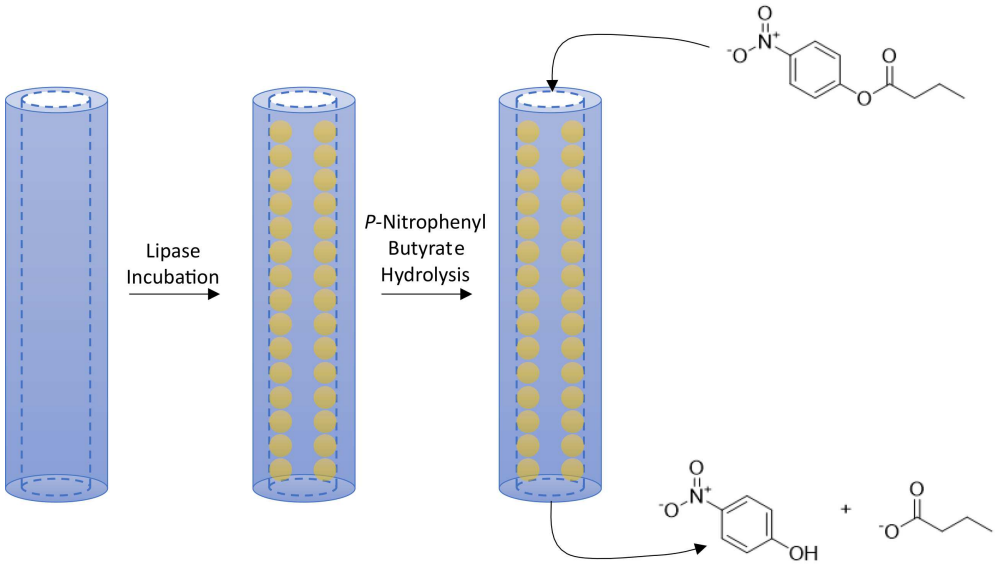


Figure 1.16: Protein nanotube encapsulation: *Candida rugosa* lipase (yellow) was encapsulated within a protein nanotube (blue) through adsorption which enhanced its hydrolysis reaction rate by 33% and its stability at 65 °C by 70%. Figure adapted from Yu *et al.* [121].

1.3.4 Encapsulation

The structures discussed previously have mostly focused upon encapsulation of enzymes within a confined space as this is commonly found in nature. These structures help to increase local substrate concentrations, allow enzymes to be encapsulated, with little or no modifications, and help to improve flux through various pathways.[124,125] Enzymes have been encapsulated within lipid vesicles,[126] within polymersomes punctured with DNA nanopores,[64] targeted for encapsulation within encapsulins and BMCs,[95,127] and within DNA nanocages[128] and more dynamic structures such as DNA nanovaults.[129] However, these structures, although they more closely mimicking natural systems, require a large amount of biological engineering to get the desired encapsulation system established. Many of the benefits of encapsulation, increased concentrations and improved flux, enzyme stability and reduction of toxic intermediates can all be achieved through surface immobilisation systems which are far simpler to engineer, whilst achieving the same effects.[130]

1.4 Immobilisation

Immobilising enzymes on a solid support can improve the properties of the enzymes to make them more amenable to industrial biocatalysis. Immobilisation can also help to improve enzyme stability, activity, selectivity, specificity, and reduce inhibition.[130] This can be achieved by tethering enzymes to a solid support, or to the surface of synthetic compartments, and can be used to form defined enzymatic cascades.

1.4.1 Natural Protein Immobilisation

In nature, many proteins, including enzymes, are found within, or associated to, a phospholipid membrane using a variety of different tethering mechanisms. Proteins can be either permanently or temporarily associated with a membrane.[131] Integral membrane proteins are permanently attached to a membrane either to one side (monotopic) or, more commonly, across the membrane (polytopic) to form a transmembrane protein. Integral membrane proteins have key roles as: structural proteins; channels and shuttles to transport material across and inside the membrane; energy transducers; receptors; enzymes; and cell adhesion mediators.[132] Many integral proteins are embedded into membrane due to the hydrophobicity of the peptide side chains generating either: a transmembrane protein with hydrophilic sections on either side of the membrane; or a monotopic protein with the hydrophilic section of the protein on a single side. These integral proteins are often complex and challenging to express recombinantly as they require large genes and additional proteins to target the protein to the membrane and ensure it folds correctly.

Some integral proteins do not directly interact with the membrane and are instead anchored through covalent interactions with a lipid or glycolipid chain embedded into the membrane.[133] The type of anchor used determines which side of the membrane it will be embedded in and allows the protein to be targeted to the exoplasmic or cytosolic face of the membrane. Glycosylphosphatidylinositol (GPI) is a glycolipid anchor attached post-translationally to the *C*-terminus of many eukaryotic proteins, and inserts on the exoplasmic face of a phospholipid membrane.[134] The GPI anchor consists of two fatty acyl groups, *N*-acetylglucosamine, mannose, and inositol (Figure 1.17). Over 250 different proteins contain a GPI anchor covering a wide range of biological functions including catalysis, cell-cell interactions, complement regulation and as antigens. Some key examples include: the hydrolytic enzymes alkaline phosphatase and acetylcholinesterase; the T-cell marker Thy-1; and CD55 (decay-accelerating factor) which regulates the complement cascade in human red-blood cells.[135]

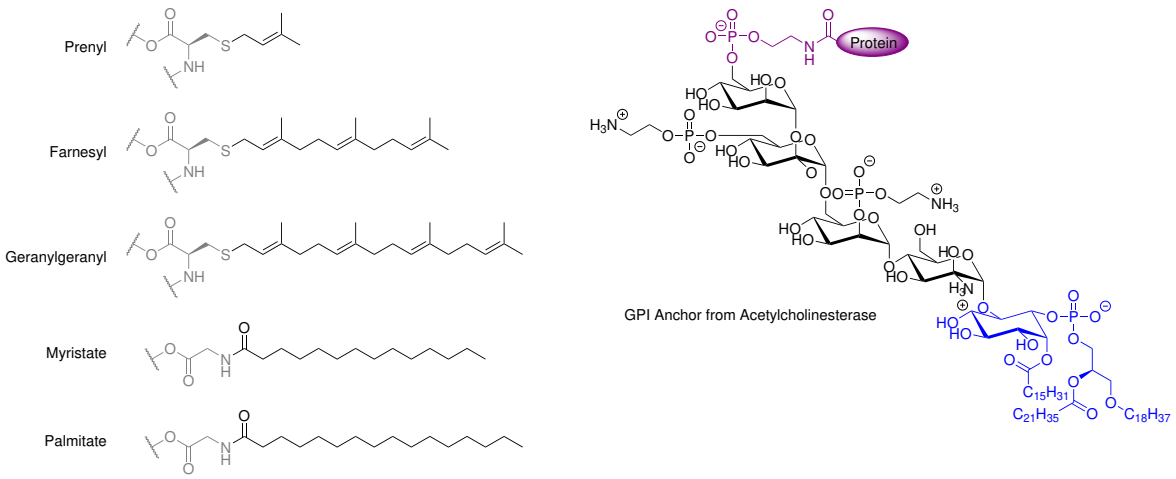


Figure 1.17: Examples of biological membrane anchors: Hydrocarbon (prenyl, farnesyl and geranylgeranyl) and fatty acid (myristate and palmitate) moieties are shown on the left; and a glycosylphosphatidylinositol (GPI) anchor for human acetylcholinesterase. Protein glycine and cysteine residues are shown in grey.

Cytosolic proteins are tethered to the cytosolic face of the membrane through hydrocarbon or fatty acid moeities. Prenyl, farnesyl, and geranylgeranyl groups are the most common anchors and require chemical modification of the peptide chain. These groups are covalently linked to a cysteine four residues from the *C*-terminus of the protein, through a thioether bond.[133] Another example of lipid-anchored tethering involves the formation of an amide bond between the *N*-terminal glycine of a protein and a fatty acyl group, such as myristate or palmitate, which then embeds in the cytosolic side of the membrane.[136]

Peripheral membrane proteins are associated with the membrane though interactions with the lipid head-groups, or integral membrane proteins without entering the hydrophobic core of the membrane. Due to this feature many peripheral proteins are water soluble as these are hydrophilic molecules, which cannot cross or enter the membrane. Most peripheral proteins interact with the cytosolic face of the membrane and often shuttle between the cytosol and its target membrane or receptor. Many of these proteins form regulatory units of integral membrane proteins and are involved in signal transduction in a huge number of pathways. Cytochrome c is a peripheral membrane protein involved in the electron transport chain and is associated to the inner mitochondrial membrane in the inter-membrane space. Cytochrome c is an important, highly soluble, hemeprotein which transfers electrons between two integral membrane proteins: cytochrome bc_1 complex and cytochrome c oxidase.[137] The membrane skeleton of human erythrocytes is also made up of peripheral proteins, namely actin and spectrin, which give rise to the distinctive erythrocyte shape.[138]

1.4.2 Synthetic Immobilisation of Enzymes

Synthetic immobilisation of enzymes can be achieved through several methods which vary depending upon the type of bonds use to tether the enzyme[139] (Figure 1.18). The type of bond used determines the strength of the enzyme binding and all have their advantages and disadvantages.

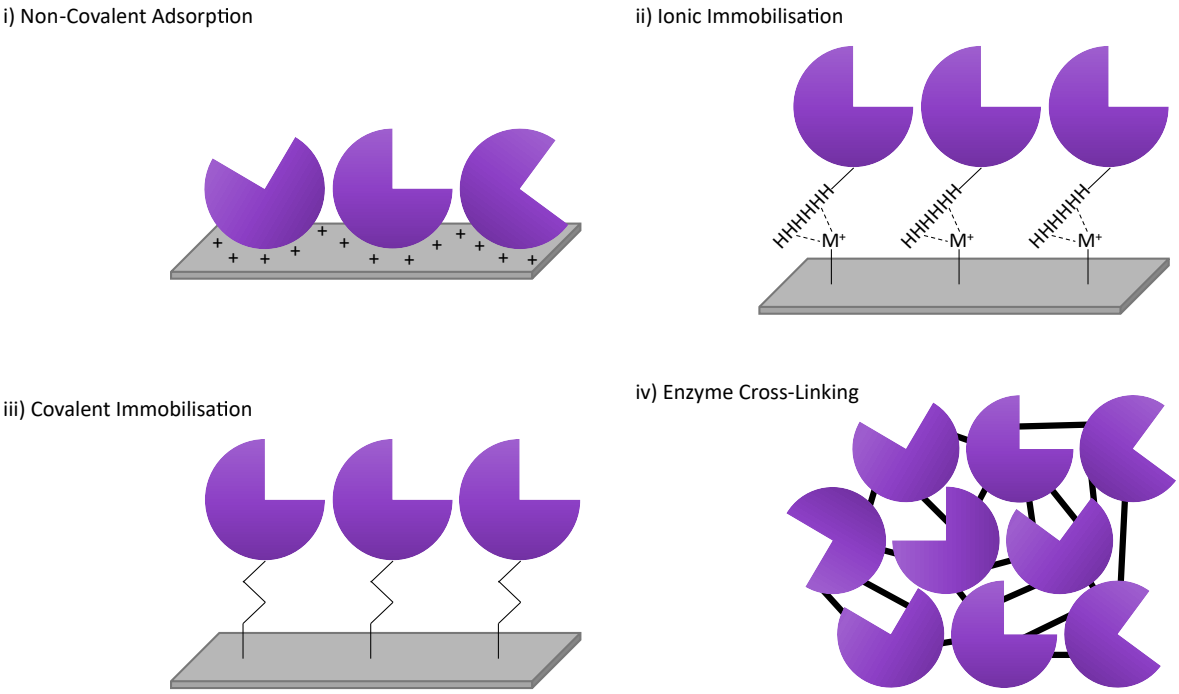


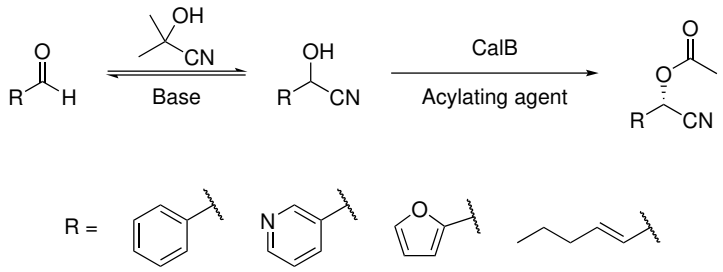
Figure 1.18: Enzyme Immobilisation Techniques - i) adsorption of an enzyme onto a surface through non-covalent hydrogen bonds, van der Waals forces, entropy changes or hydrophobic/hydrophilic interactions; ii) Ionic immobilisation through chelation from an intrinsic or added residue to a support; iii) Covalent immobilisation through covalent bond formation between a support and the enzyme; iv) Cross-linking between enzymes to make a cross-linked enzyme crystal or aggregate.

1.4.2.1 Non-Covalent Immoblisation

Non-covalent interactions rely on hydrogen bonds, van der Waals forces, entropy changes, or hydrophobic/hydrophilic interactions.[140] The principal advantage of non-covalent adsorption techniques is that the enzymes do not need to be chemically or genetically modified prior to immobilisation, greatly reducing the number of steps required. Instead these techniques rely on the intrinsic properties of the surface residues of the enzyme, such as hydrophobicity and ability to form hydrogen bonds. The main problem with relying on these weaker interactions, however, is that the immobilised enzymes often leach into the bulk solution. The immobilisation conditions have to be tuned to maximise the strength of the interaction between the enzyme and the surface to minimise leaching and maximise enzyme activity.

Enzymes bury their hydrophobic residues within their structure to limit their exposure to the aqueous environment, an important driving force behind all protein folding, so there are only a few enzymes amenable to immobilisation solely through van der Waals forces. Lipases are an exception to this, as they often have a lid protecting the active site which is opened when the lipase is at the interface between water and oil, exposing the hydrophobic core around the active site.[141] The interaction between these hydrophobic residues and a hydrophobic support stabilises the lipase in this open, active, form increasing its activity. This immobilisation method relies not only on the van der Waals forces between the two surfaces, but on the entropic changes associated with a single enzyme displacing multiple water molecules on the support.

Since few enzymes have solvent-exposed hydrophobic residues, methods that exploit interactions with hydrophilic surface residues, capable of forming hydrogen bonds, are more common. These residues can also be glycosylated, which increases the hydrophilicity and ability to form hydrogen bonds with a hydrophilic support. Celite is a common hydrophilic support and has been used in the immobilisation of the lipase CALB (*Candida antarctica* lipase B).[142] The hydrophilic surface of Celite helped to absorb water and greatly improved the conversion and enatioselectivity of the synthesis of cyanohydrin esters via dynamic kinetic resolution (Scheme 1.1). Various enzymes have also been immobilised on other hydrophilic supports such as cellulose, lignine, Avicel, porous glass, clay, and silica gel.[139]

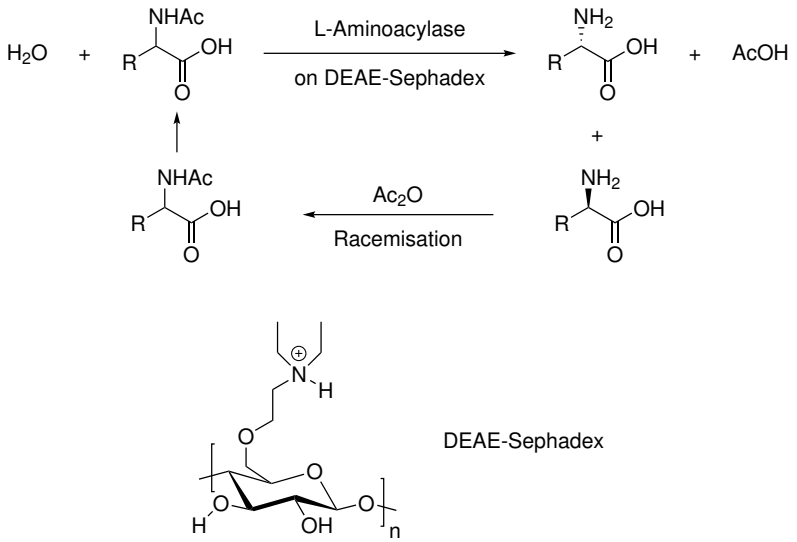


Scheme 1.1: Enantioselective synthesis of cyanohydrin esters via a dynamic kinetic resolution using immobilised CalB. Scheme adapted from Veum *et al.* [142].

1.4.2.2 Ionic Immobilisation

The next step up from adsorption immobilisation in binding strength is ionic immobilisation. Stronger ionic bonds can help overcome the problems of enzyme leaching associated with adsorption, however, these bonds can be disrupted by pH, temperature and salt concentrations.[143] The trade off for this stronger bonding is that the enzyme often has to undergo a chemical or genetic modification, if it does not have a strong enough surface charge, to be able to form an ionic bond with the support.

Depending on the isoelectric point of an enzyme, at a given pH it may carry an overall surface charge. This surface charge can be used to bind to any oppositely charged ion-exchange resin without requiring modification of the enzyme.[144] The first industrial use of an immobilised enzyme was using this method involving the binding of a L-aminoacylase to a DEAE-Sephadex resin to produce an L-amino acid in 1969 (Scheme 1.2).[145] Another example of this method is the multipoint immobilisation of CALB (*Candida antarctica* lipase B) on polyethylenimine, an anion-exchanger.[146] Depending upon the immobilisation conditions, its activity and enantioselectivity for the hydrolysis of *R,S*-mandelic acid methyl ester, were greatly enhanced even in unfavourable reaction conditions.



Scheme 1.2: L-Amino acid catalysed synthesis of L-amino acids using an L-aminoacylase immobilised on DEAE-Sephadex. Figure adapted from Sheldon *et al.* [145].

One of the most common enzyme modifications is the genetic addition of a six histidine tag (His-Tag) at the *C*- or *N*-terminus of a recombinant enzyme.[145] This enables it to bind to an ion-exchange column, most commonly agarose beads bearing nickel-nitrilotriacetic acid (NTA) groups, through multidentate interactions between the His nitrogen atoms and the positively charged nickel ions. This form of immobilisation is used frequently to aid enzyme purification and as such many enzymes are expressed with a His-tag, or similar, making ionic linkages the most common form of immobilisation.[147] An interesting example of this immobilisation was shown when two proteins, a phosphodiesterase and an alkaline phosphatase, were purified and immobilised within a continuous flow, vortex fluidic device using a Ni-NTA resin.[148]

1.4.2.3 Covalent Immobilisation

The strongest form of immobilisation involves using covalent linkages between the enzyme and the support.[145] Covalent linkages prevent enzyme leaching and limit problems with downstream processing and enzyme contamination. The formation of multiple covalent bonds can also stabilise the enzyme and protect it from denaturing. However, greater chemical modifications of the support are often required to generate the reactive groups to form these bonds. Another issue with covalent linkages is that if an enzyme becomes irreversibly inhibited, both the enzyme and the support become unusable.

Many amino acid side chains contain functional groups suitable for covalent bond formation: the amino groups of the *N*-terminus, lysine and arginine; the carboxyl groups of the *C*-terminus, aspartic and glutamic acids; the hydroxyl groups of serine, threonine and tyrosine; the thiol group of cysteine; the imidazole group of histidine; and the indole group of tryptophan.[149]

Lysine is the most commonly used residue for bioconjugation as there are often several of these residues on the external surface of the protein and they can be readily reacted with activated esters (such as NHS esters), isocyanates, isothiocyanates, benzoyl fluorides, and epoxide- and aldehyde-functionalised supports.[150] Lysine is useful when the enzyme is robust and is not disrupted by the bioconjugation, however, due to the number of lysines naturally present in proteins, this immobilisation is not site-specific as a number of lysines can be modified in the bioconjugation reaction. This can affect the binding and orientation of the protein and can greatly impact enzyme activity, so lysine bioconjugation has to be carefully considered or involve site-selection for the most reactive lysines.[151] Cysteine has a relatively low abundance in natural proteins so is used when a more site-directed immobilisation is required, either through native residues or those introduced by site-directed mutagenesis.[152] Cysteine is mostly commonly coupled to solid supports by reacting with iodoacetamides, maleimides, vinyl sulfones, acrylamides, or forming disulfides with other thiols. The carboxyl groups on aspartic and glutamic acid can be used for conjugation, but these have to be converted *in situ* into their corresponding activated esters using a carbodiimide coupling agent, such as EDC or DDC, and an auxiliary nucleophile, such as NHS, so are used less often.[153] These are the most common forms of covalent immobilisation, but there are many other options available using the full range of reactivities from the residues, peptide chain, and other post-translational modifications including adding additional protein domains for SNAP and HALO tagging or adding peptide-tagging sequences for biotin ligase or bacterial

sortase.[152–155]

1.4.2.4 Cross-linking

Immobilisation of enzymes on solid supports is very beneficial to help improve enzyme stability and reduce problems with downstream processing, however, it can also significantly reduce space-time yields due to the space taken up by the non-catalytic support itself.[156] Supports are also often costly so an alternative method, which does not rely on a support and used the enzymes themselves as supports, would avoid these problems. Cross-linking of enzymes provides this alternative as an irreversible, covalent, support-free form of immobilisation. Enzymes can be attached to one another through their surface lysine residues using the same reactions used for standard covalent immobilisation, namely through the use of glutaraldehyde as a bifuncitonal coupling reagent.

Cross-linked enzymes retain catalytic activity but are not very stable and are difficult to work with, so cross-linked enzyme crystals (CLECs) were developed.[157] CLECs are formed from enzymes that have been crystallised from an aqueous buffer and then cross-linked, requiring only that the enzyme be able to be crystallised. CLECs are more stable, more robust and easier to recycle than their unlinked counterparts, and provide high catalytic efficiency. However, the actual process and purity requirements of crystallisation have made this process prohibitively expensive and due to this CLECs have been made obsolete.

Enzyme precipitation is a far simpler process than crystallisation and generates stable enzyme aggregates which can be easily cross-linked to generate cross-linked enzyme aggregates (CLEAs).[158] This process can be combined with enzyme purification as addition of ammonium sulfate or polyethylene glycol is often used to precipitate an enzyme in the purification process and the aggregate can then be directly cross-linked. CLEAs provide the same improvements in stability and productivity that CLECs provide without the need for the time-consuming and costly crystallisation process. CLEAs have been applied across the enzyme classes with wide selection of hydrolases, oxidoreductases, lyases, transferases, and isomerases being used successfully in organic synthesis. CLEAs can also be combined with magnetic nanoparticles, for easier separation and use in magnetically stabilised fluidised bed reactors, or combined with a monomer, such as siloxane, to undergo co-polymerisation to give CLEA–polymer composites with tunable physical properties.[158]

1.4.2.5 Multi-Enzyme Complexes and Immobilised Cascades

Multi-enzyme clusters provide an ideal way of performing enzymatic cascades, and their advantages can be recreated by immobilising several sequential enzymes to a solid support provided they are within 10 nm of each other.[159] This inter-enzyme distance allows for the substrates to be easily transferred between the enzymes, which increases individual reaction rates as well as the overall flux through the cascade. All the previously discussed immobilisation and compartmentalisation techniques, as well as an array of other methods, can be used to form immobilised cascades providing varying levels of control over the stoichiometry, orientation, and spacing of the enzymes in a cascade. One example of this is a DNA-based "dimer nanoreactor" where glucose oxidase and horseradish peroxidase were co-localised on DNA origami building blocks using neutravidin tethers which greatly increased the reaction rates (Figure 1.19 i).[160] A more dynamic example, which achieves the same effect, has been explored which incorporates a swinging arm to transfer substrates between enzymes with an inter-enzyme distance of 14 nm, similar to the acyl-carrier protein in polyketide synthase (Figure 1.19 ii).[161] This has also been applied to support-free immobilisation through combi-CLEAs, such as the immobilisation of an enoate reductase and a glucose dehydrogenase to successfully reduce 4-(4-Methoxyphenyl)-3-buten-2-one and recycle the required NAD cofactor.[162]

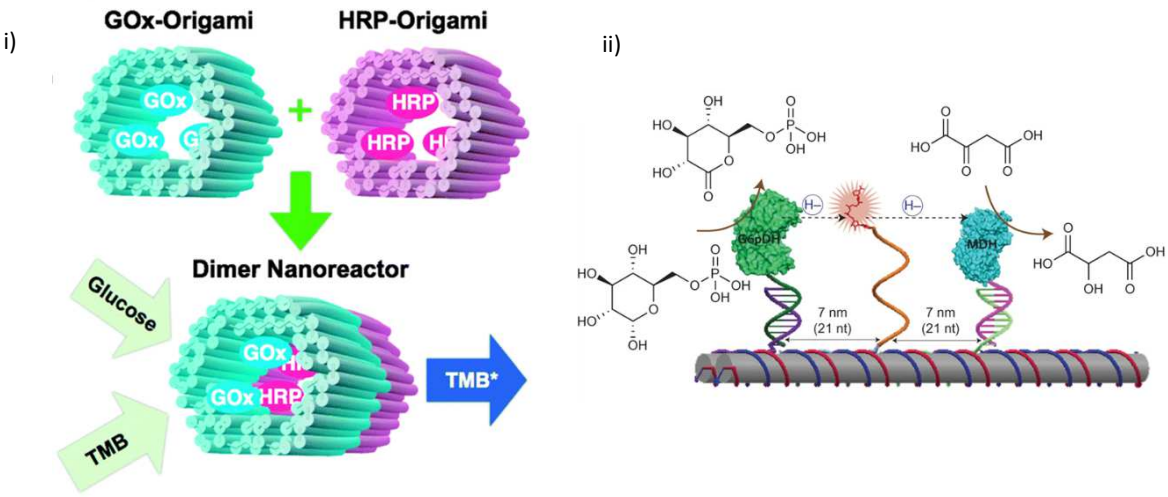


Figure 1.19: Examples of DNA-based nanoreactors - i) Two separately fabricated origami units are equipped with glucose oxidase (GOx) or horseradish peroxidase (HRP) and are combined to form a DNA-based "dimer nanoreactor" capable of producing hydrogen peroxide and converting 3,3',5,5'-tetramethylbenzidine to 3,3',5,5'-tetramethylbenzidine diimine. ii) A nanostructured complex consisting of G6PDH and MDH organized on a DNA tile. An NAD^+-modified DNA strand is positioned halfway between the two enzymes, facilitating the transfer of hydrides and recycling of the cofactor. Figures adapted from Linko *et al.* [160], Fu *et al.* [161].

Immobilisation of multi-enzyme cascades combines the benefits of cascades and immobilisation in improving enzyme stability and increasing flux through a cascade. However, there are drawbacks

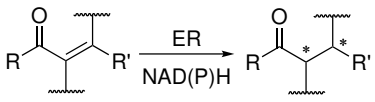
to using these methods, particularly in relation to reductions in catalytic activity, so new methods which improve the efficiency of cascades are important in advancing biocatalysis.

1.5 Enzymes

Developing an immobilised enzyme cascade requires first a robust development and understanding of the cascade as a free *in vitro* system. The following enzyme classes were of particular interest as potential targets for forming a multi-enzyme cascade to generate a useful product, so a brief introduction to these specific enzymes is included below.

1.5.1 Ene-Reductases

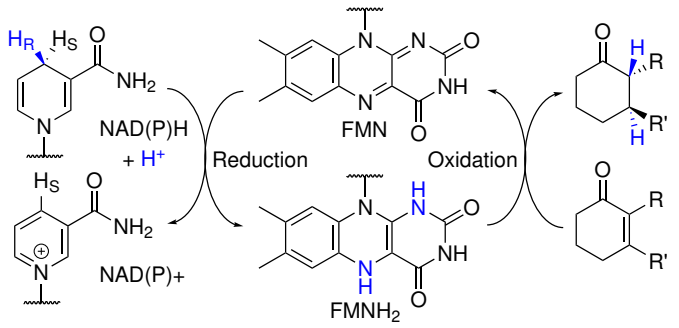
Ene-reductases (ERs) catalyse the reduction of α,β-unsaturated carbonyls activated by electron withdrawing substituents in the γ-position[163] (Scheme 1.3). This reduction introduces, depending on the molecule, up to two chiral centres in a trans-specific manner. ERs are both flavin and iron-sulfur cluster dependent and require a nicotinamide cofactor to provide the hydride.



Scheme 1.3: ER Reaction - An α,β-unsaturated carbonyl is stereospecifically reduced to the corresponding saturated ketone or aldehyde at the expense of NAD(P)H requiring a flavin cofactor.

There are four families of ERs: those from the Old Yellow Enzyme (OYE) family; enoate reductases; medium-chain dehydrogenase/reductase (MDRs); and short-chain dehydrogenase/ reductase (SDRs).[164] Enoate reductases are highly stereo- and regio-specific but are extremely oxygen sensitive so few have been investigated for biocatalysis. MDRs and SDRs have recently been investigated with some zinc and flavin independent enzymes being characterised. OYEs have been studied in much greater depth with structural and kinetic data, reaction mechanisms and substrate scopes all being readily available. OYEs are flavin mononucleotide (FMN) dependent and follow a bi-bi ping-pong reaction mechanism[165] (Scheme 1.4). In the first half reaction hydride is transferred from the reduced nicotinamide cofactor to FMN which reduces it to $FMNH_2$. In the second half reaction, this hydride is transferred on the unsaturated substrate which reoxidises $FMNH_2$ and a second hydride is transferred from a tyrosine or cysteine residue to produce the reduced substrate. In an *in vivo* system the NAD(P)H is regenerated by the host cell's metabolism but in an *in vitro* system an additional recycling system needs to be added to prevent the need for costly stoichiometric amounts of NAD(P)H.[166] These systems can also help to increase reaction yields. The main recycling systems used include: glucose dehydrogenase(GDH)/glucose, glucose-

6-phosphate dehydrogenase (G6PDH)/glucose-6-phosphate (G6P) and formate dehydrogenase (FDH)/formate.

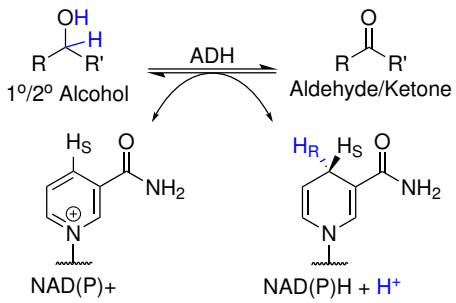


Scheme 1.4: ER Catalytic Cycle - The reductive half reaction involves the reduction of FMN to $FMNH_2$ and the oxidation of NAD(P)H to $NAD(P)^+$. The oxidative half reaction transfers the hydride from $FMNH_2$ to the α,β-unsaturated carbonyl to give the reduced product and reoxidises $FMNH_2$ to FMN.

ERs have been shown to accept a wide range of α,β-unsaturated substrates with most of the families accepting the classical substrates including aldehydes, cyclic ketones and methyl ketones.[167] OYE1-3 have also been shown to accept carboxylic acids, terpenoids and OYE1 accepts nitrate esters and nitroaromatics. The OYE homologs, NCR, XenA and YqjM have also been shown to accept similar substrates with YqjM additionally accepting explosives such as nitroglycerin and TNT.[168] These wide substrate scopes make ERs particularly useful biocatalysts as many can be used without requiring mutagenesis.

1.5.2 Alcohol Dehydrogenases

Alcohol dehydrogenases (ADHs) catalyse the reversible oxidation of primary or secondary alcohols to aldehydes or ketones requiring a nicotinamide cofactor to act as a hydride acceptor (Scheme 1.5).[169] ADHs can be used in either the oxidative or the reductive direction and have been shown to oxidise aldehydes to carboxylic acids under basic conditions. ADHs have been isolated from organisms across the spectrum of life and fall into three non-homologous families.[170] Medium-chain (Type I) ADHs exist as dimers or tetramers and have subunits with approximately 370 residues and are generally zinc dependent (structural and catalytic). Short-chain (Type II) ADHs are found in insects, do not contain metals and are approximately 250 residues. Long-chain (Type III) ADHs are iron dependent and are between 380 and 900 residues in length. Type I ADHs have been most studied and two classical examples, horse liver ADH and yeast ADH1, fall under this family.



Scheme 1.5: ADH Reaction - A primary or secondary alcohol is oxidised to an aldehyde or ketone, respectively, whilst the nicotinamide cofactor is reduced to provide hydride.

Type I ADHs have a conserved tertiary structure consisting of a cofactor binding domain and a substrate binding domain.[171] The active site contains two cysteine residues and one histidine residue which coordinate to the zinc ion. These ADHs follow an ordered bi-bi reaction mechanism whereby the oxidised nicotinamide cofactor binds into the active site followed by the alcohol substrate. The alcohol substrate is then converted to a zinc-alkoxide which releases a proton into the solvent via a hydrogen bonded network in the active site. Hydride is then transferred from the alkoxide anion directly to NAD^+ which generates the zinc-bound product and NADH. The aldehyde or ketone is then released followed by NADH.

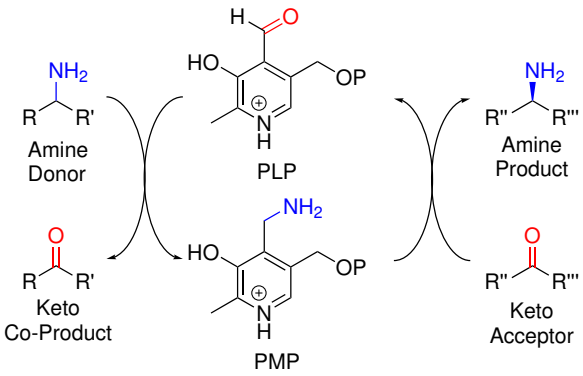
ADHs exhibit broad substrate specificity as they are found in a diverse range of native enzymatic pathways and are required in the detoxification of a range of alcohols. This makes ADHs a particularly attractive target for industrial biocatalysis as little genetic engineering is required to switch between various substrates. Similarly to ERs the recycling of the nicotinamide cofactor has to be considered, but this can be achieved in the same manner by using cofactor recycling systems with cheap sacrificial substrates.

1.5.3 Transaminases

Transaminases (TAms) catalyse the stereospecific interconversion between ketones and primary amines through the transfer of an amine group or keto group from a donor substrate. TAms are widely used in nature as the acceptor substrate, the carbonyl, is often an α-keto acid and the reaction is used to stereospecifically generate many natural amino acids. TAms require pyridoxal-5'-phosphate (PLP) as a cofactor which binds to an active site lysine as a Schiff base.[172]

TAms exhibit a bi-bi ping-pong reaction mechanism with two half reactions (Scheme 1.6). In the first half reaction, the amine donor forms a Schiff base with the PLP and transfers its amine group onto the PLP to form pyridoxamine-5'-phosphate (PMP) and the ketone co-product. In the

second half reaction, the amine group is transferred from the PMP onto the acceptor substrate to regenerate the PLP and release the amine product.



Scheme 1.6: TAm Catalytic Cycle: An amine group is transferred stereospecifically from a donor amine substrate to a keto acceptor substrate requiring a PLP cofactor.

TAms exist in two forms: α-TAms which only accept α-amino and α-keto acids; and ω-TAms which accept substrates with a distal carboxyl group. There is also a subgroup of ω-TAms, amine TAms, which can accept substrates completely lacking in a carbonyl group which would allow for generation of a chiral amine from a prochiral ketone. The wild-type enzymes have been shown to be relatively promiscuous towards the amine donor and have been shown to accept a range of aledhydes and ketones, however, the enzyme active site contains two binding pockets, one small and one large, which limits the range of ketones it can accept to those with at least one smaller group (C1 or C2) attached. The substrate scope of TAms has been expanded by various techniques including gene mining and protein engineering which has enabled them to be used successfully in commercial biocatlysis, the most notable being in the synthesis of Sitagliptin, an anti-diabetic drug, using an engineered (R)-selective ω-TA.[173]

The biggest problem when using TAms is the highly reversible nature of the reaction, especially when using an amino acid as the amine donor. In order to achieve high conversion rates the carbonyl co-product needs to be removed in order to prevent the reverse reaction. The most common way of achieving this is through using either alanine or isopropyl amine which generate pyruvate and acetone respectively which can be removed *in situ*. However, this removal, although established, is often complex and requires specialised enzymes. The best way to overcome this problem is to use the enzymes as they are in nature, in a cascade. This would ensure that the products generated are transferred to the next enzyme pulling the reaction equilibrium towards the products to help ensure high reaction rates, yields and flux through a cascade.

1.6 Aims

The main aim of this project was to develop novel immobilisation platforms for biocatalysis using cholesterol-based linkers and ionic tethering methods to immobilise an enzymatic cascade. This platform aims to overcome some of the problems currently faced by biocatalysis such as low yields, slow reaction rates, cross reactivities and issues with downstream processing. Two types of immobilisation platform were selected for this project, one which relies upon a cholesterol-linker which is chemically modified to attach to a protein, and one which relies upon a cholesterol-DNA-linker. A secondary aim within this development, was to chemically synthesise the two linkers in order to construct the immobilisation platforms and test its efficacy.

Alongside developing novel immobilisation platforms, another aim of this project was to investigate and develop a novel enzymatic cascade, which could be used as a model system to test the immobilisation platforms. This novel enzymatic cascade involved combining alcohol dehydrogenases with ene-reductases and transaminases to generate a three-step cascade with the potential to produce useful products, nepetalactone derivatives. A secondary aim for this part of the project was also synthetic and involved producing the starting material and key intermediates for the three-enzyme cascade and to test these against a panel of each of the enzymes of interest.

2 | Substrate Synthesis and Enzyme Assays

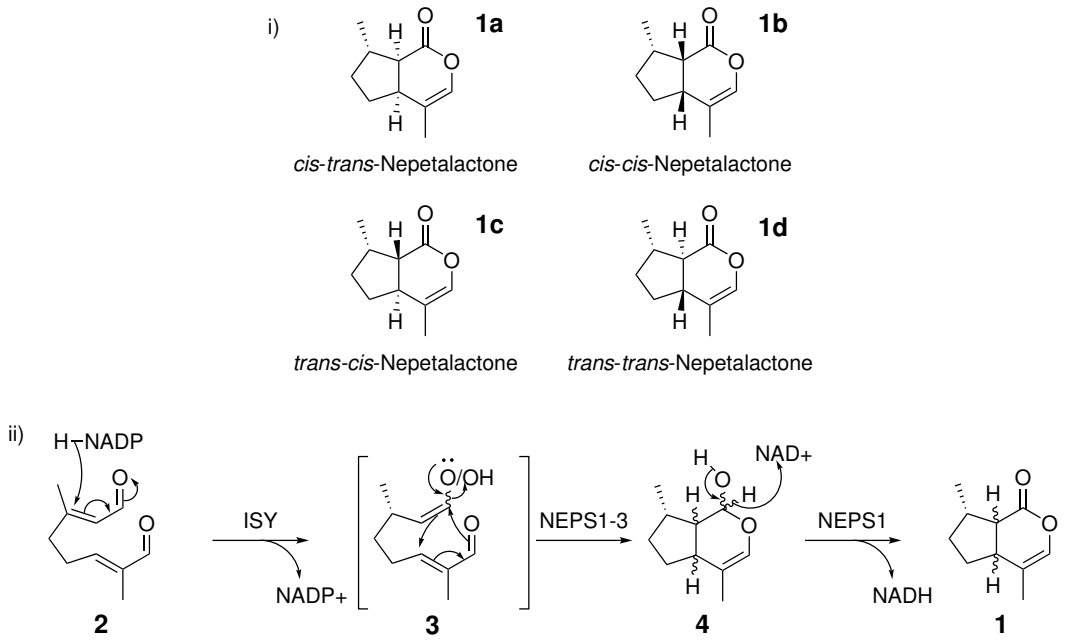
This chapter describes and discusses approaches towards the development of an enzymatic cascade incorporating alcohol dehydrogenases, ene-reductases and transaminases, designed to use as a model system for enzyme immobilisation.

2.1 Introduction

Enzymatic cascades, as discussed in section 1.2.5, provide a useful platform for green substrate synthesis. Enzymatic cascades are able to perform multi-step reactions, generate stereo- and enantio-pure products and produce compounds not accessible by chemical means, without the usual isolation, toxic, and solvent intensive requirements found in traditional chemical synthesis. However, there are downsides to using enzymatic cascades, enzymes often require different optimum conditions, such as pH or temperature, so the ideal conditions for one enzyme may be very different to another, and, as with other multi-step synthesises, the reactions may not be compatible with one another and a cascade which has been developed for one substrate may not be able to transform a similar substrate. A key factor in cascade development involves careful testing and optimisation to maximise individual enzyme activity and flux through the whole cascade.

The three enzyme classes, discussed in section 1.5, that were selected for the cascade cover redox reactions, alcohol dehydrogenases (ADHs) and ene-reductases (ERs), and functional group interconversions, transaminases (TAms). Cascades were considered to showcase the use of all of these enzymes in a one-pot system, with the aim of generating a useful product. Nepetalactone (**1**), a name which refers to multiple iridoid analog stereoisomers (Scheme 2.1 (i)), is a volatile compound found in *Nepeta cataria*, more commonly known as catnip, which is thought to mimic cat pheromones.[174] Alongside producing a range of effects in numerous feline species, nepetalactone is a potent insect repellent, with efficiencies comparable to the synthetic repellent *N,N*-diethyl-*meta*-toluamide (DEET), designed to protect the plant from herbivorous insects.[175] The biosynthesis of nepetalactone has recently been resolved and has been shown to start from 8-oxogeranial (**2**) and involves three enzymes which can vary depending on the stereoisomer being produced (Scheme 2.1 (ii)).[176] The 8-oxogeranial (**2**) is converted to the reactive intermediate 8-oxocitronllyl enol/enolate (**3**) by an NADPH-dependent iridoid synthase (ISY). This intermediate is then cyclised to give a nepetalactol (**4**) either spontaneously or by a nepetalactol-related

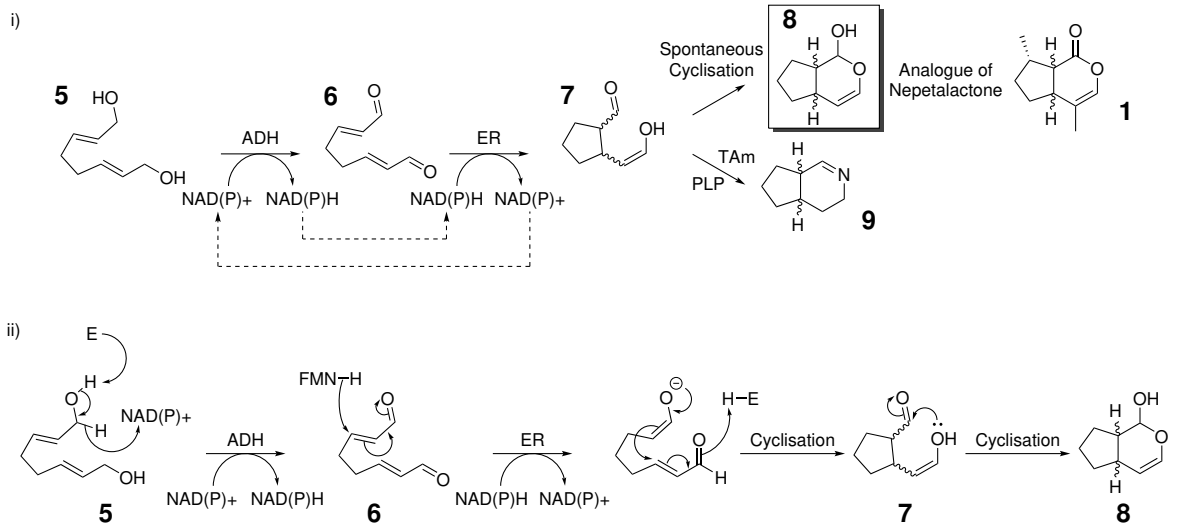
short-chain dehydrogenase (NEPS), NEPS1-3, depending on the stereoisomer being made. The nepetalactol, existing as any of the four stereoisomers, is then oxidised by NAD-dependent NEPS1 to give the final nepetalactone product (**1**). Nepetalactone can then be extracted from the plant by steam distillation and has been shown to be an effective natural, if somewhat potent, insect repellent, particularly in malaria ridden areas.[177] Generation of analogues of this compound would be useful to provide a synthetic route to a wider range of iridoids and provide alternatives that may have improved properties, as the scent of these compounds is a main issue against their widespread usage.



Scheme 2.1: Nepetalactone biosynthesis - i) The different stereoisomers of nepetalactone found in the *Nepeta* species. ii) The biosynthesis of nepetalactone starting from 8-oxogeranial (**2**) using an ISY and two NEPS enzymes to convert it into the final nepetalactone product (**1**).

Building upon the reported biosynthesis, an initial cascade (Scheme 2.2) was designed to convert an unsaturated diol (**5**) through two steps, via an unsaturated dialdehyde (**6**), into a cyclic alcohol/aldehyde (**7**) which can spontaneously cyclise to give a nepetalactone analogue (**8**). The first step of this cascade involves the oxidation of the unsaturated diol (**5**) by an ADH to the unsaturated dialdehyde (**6**) at the expense of NAD(P)$^+$. This unsaturated dialdehyde (**6**) has an analogous structure to 8-oxogeranial (**2**), the starting point for nepetalactone biosynthesis. One of the alkene groups on dialdehyde intermediate is then reduced by an ER and the intermediate can undergo a cyclisation step to give a cyclic alcohol/aldehyde (**7**), this can then undergo another spontaneous cyclisation to give the final product (**8**), a nepetalactone analogue. Alternatively, the cyclic alcohol/aldehyde intermediate (**7**) can be transformed with a TAm to give a nitrogen heterocycle product (**9**). Using the ADH and ER in tandem, one requiring NAD(P)$^+$ and the other

NAD(P)H, is a useful feature of this cascade as it provides its own internal cofactor recycling system and only needs to be supplied with a catalytic amount of the expensive cofactor, provided they both accept either NAD or NADP.

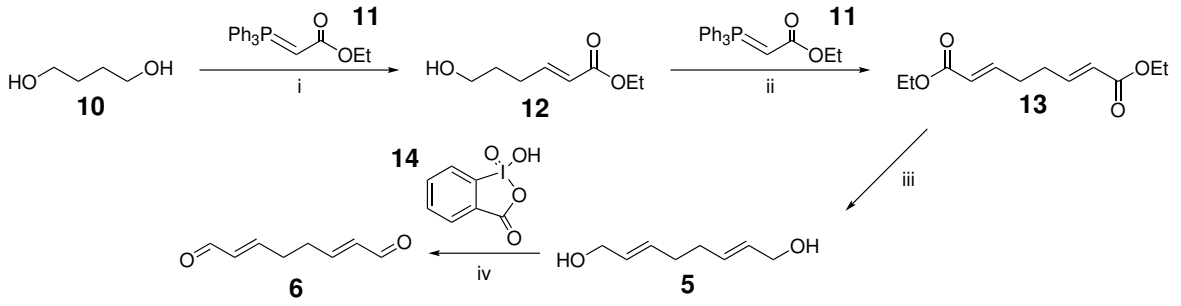


Scheme 2.2: Enzymatic cascade design - i) The production of a nepetalactone analogue (**8**) through a two step reaction using an ADH and an ER. An alternative route using a transaminase to give a cyclopentapyridine derivative is also shown. ii) Mechanism of the reaction in the cascade to produce the nepetalactone analogue (**8**).

2.2 Substrate Synthesis

2.2.1 (2*E*,6*E*)-Octa-2,6-dienedial Synthesis

In order to investigate the proposed enzyme cascade (Scheme 2.2), chemical synthesis of the α/β-unsaturated dialcohol, (2*E*,6*E*)-Octa-2,6-diene-1,8-diol (**5**), and the dialdehyde, (2*E*,6*E*)-Octa-2,6-dienedial (**6**), was required. These were selected as they could potentially give intermediates with the capability to form the nepetalactone analogues and would both be used to test the ADHs and ERs as novel substrates and provide chemical standards. A four-step synthesis (Scheme 2.3) to generate both of these compounds was designed using a commercially available starting material, 1,4-butanediol (**10**).



Scheme 2.3: Synthetic route for (2*E*,6*E*)-octa-2,6-dienedial (**6**) from 1,4-butanediol (**10**). i) MnO_2, CH_2Cl_2, 5 d, rt, 42%; ii) PCC, imidazole, CH_2Cl_2, 24 h, rt, 48%; iii) DIBAL-H, THF, 3 h, -78 °C, 94%; iv) DMSO, 1 h, rt.

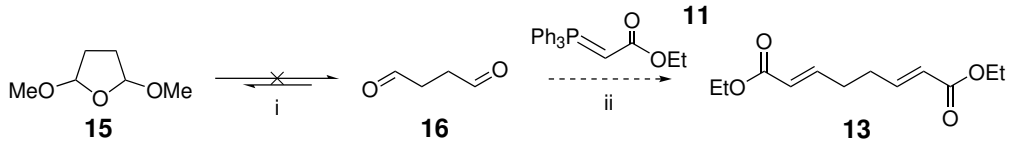
The first step of the synthesis (i), a Wittig reaction, required the synthesis of a ylide (**11**). This was successfully synthesised from triphenylphosphine and ethyl bromoacetate on a 5 g scale following a literature method, by first forming the phosphonium salt in toluene and then adding the salt to aqueous NaOH to generate the ylide (**11**) in 63% yield.[178] The ylide was then reacted with 1,4-butanediol (**10**) in a reported oxidative Wittig coupling reaction, using manganese dioxide (MnO_2) as the oxidant, which successfully produced ethyl (*E*)-6-hydroxyhex-2-enoate (**12**) in 42% yield after purification by column chromatography.[179] Wittig reactions are traditionally performed with aldehydes directly rather than *in situ* generation from their alcohol counterparts, so the second Wittig reaction required the stronger oxidising conditions provided by pyridinium chlorochromate (PCC). This second Wittig reaction lead to the formation of diethyl (2*E*,6*E*)-octa-2,6-dienedioate (**13**) which was isolated in a 48% yield after purification by column chromatography.[179] These two reactions successfully generated the diester product but were time consuming, taking roughly a week to complete, and required the use of toxic oxidising reagents.

The next step was to reduce the diester to a dialcohol (**5**), without reducing the alkenes, which was achieved using DIBAL-H with a crude yield of 94%.[180] ^{1}H NMR spectroscopic analysis of **5** showed peaks corresponding the the presence of ethyl acetate. However, the dialcohol was taken with these apparent solvent impurities through to the oxidation step as the residual ethyl acetate should not interfere with subsequent reactions and could be removed at a later stage. Oxidation to the dialdehyde (**6**) under mild conditions was achieved using 2-iodoxybenzoic acid (IBX, **14**) as the oxidant. IBX was selected for this oxidation step as it has been reported to oxidise this specific dialcohol to the dialdehyde in a 97% yield.[180] ^{1}H NMR spectroscopic analysis indicated that the product was produced, however, some IBX still remained and there was insufficient material to perform further purifications and complete the analysis. Several attempts were made to generate more material, however, these were unable to generate enough material so a different synthetic route was explored.

2.2.2 (2*E*,7*E*)-Nona-2,7-dienedial Synthesis

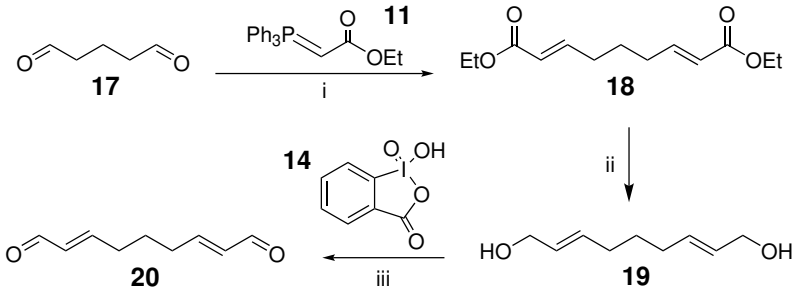
The original synthetic route (Scheme 2.3) was time intensive, so an alternative, shorter, route was explored. The Wittig reaction also requires a carbonyl group so the methods used above required the *in situ* oxidation of the hydroxyl to a carbonyl using either manganese dioxide or PCC. Both of these reagents are toxic oxidants so a reaction to replace these two reagents, and combine the two Wittig reactions, was preferable.

An attempt was made to use 2,5-dimethoxytetrahydrofuran (**15**) as a starting material to synthesise succinaldehyde (**16**).[180] This could then be used *in situ* to form (2*E*,6*E*)-octa-2,6-dienedioate (**13**) using a Wittig reaction (Scheme 2.4). However, after several attempts this method was abandoned in favour of synthesising a longer chain equivalent from commercially available glutaraldehyde (**17**) (Scheme 2.5).[180]



Scheme 2.4: Synthetic route to diethyl (2*E*,6*E*)-octa-2,6-dienedioate (**13**) from 2,5-dimethoxytetrahydrofuran (**15**). i) HCl (0.6 M), 1 h, 80 °C; ii) H_2O, 17 h, rt.

The longer chain diester, (2*E*,7*E*)-nona-2,7-dienedioate (**18**), was successfully synthesised and isolated in a 69% yield from (**17**). This diester was then fully reduced to **19**, the dialcohol, using the same method as previously used for **13**, the shorter chain diester. This reaction was repeated several times with initially very low isolated yields of 6% or no conversion at all, possibly due to issues with the DIBAL-H reducing agent being inactive. As with **5**, the shorter chain dialcohol, an impure product (**19**) was obtained in a very high yield (> 100%), containing the solvents THF and ethyl acetate from the reaction and column chromatography steps. However, this impure product was taken through to the next step as again these should not interfere with the oxidation reaction and could be removed at a later stage.



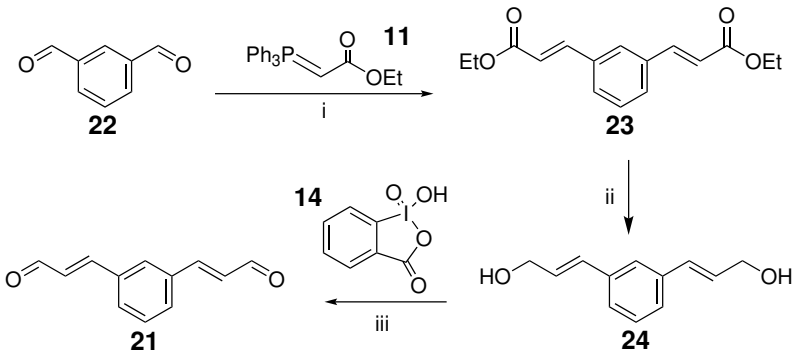
Scheme 2.5: Synthetic route to (2*E*,7*E*)-nona-2,7-dienedial (**20**) from glutaraldehyde (**17**). i) CH_2Cl_2, 17 h, rt, 69%; ii) DIBAL-H, THF, 3 h, -78 °C, 23%; iii) DMSO, 1 h, rt, 59%.

The dialdehyde, (2*E*,7*E*)-nona-2,7-dienedial (**20**), was successfully synthesised, again using an analogous method to that followed for the shorter chain dialdehyde (**6**), however, the work up procedure had to be altered to minimise IBX contamination. Extraction of the product into ether, a brine wash, drying, filtration, and a silica column were unable to remove the by-product of the reaction, reduced IBX, so an alternative method was required. Centrifugation of a small sample of

20 helped to remove some of the by-product, however, a high level of reduced IBX was still visible by NMR spectroscopicanalysis. This sample was then washed with potassium hydrogen carbonate, to basify the reduced IBX and move it into the aqueous layer, before drying, filtration and solvent removal. This small sample showed minimal IBX contamination by NMR spectroscopic analysis so this method was applied to the remainder of the crude product. This gave the dialdehyde, **20**, in a 59% yield with minimal IBX contamination (approx. 4% by NMR) and could be taken through to substrate screening with a panel of ERs.

2.2.3 (2*E*,2'*E*)-3,3'-(1,3-Phenylene)diacrylaldehyde Synthesis

Once a robust synthetic route to forming α/β-unsaturated aldehydes was established this was used to generate a more conformationally rigid dialdehyde containing a phenyl ring (**21**). This substrate would likely behave differently to the straight chain dialdehyde (**20**) and would provide an alternative cascade which could be monitored by HPLC rather than GC due to the presence of a chromaphore. A commercially available dialdehyde, isophthalaldehyde (**22**), was selected for this adapted synthetic route (Scheme 2.6), similarly to glutaraldehyde (**17**), to avoid having to generate the dialdehyde *in situ* from the equivalent dialcohol.



Scheme 2.6: Synthetic route to (2*E*,2'*E*)-3,3'-(1,3-phenylene)diacrylaldehyde (**21**) from isophthalaldehyde (**22**). i) CH_2Cl_2, 17 h, rt, 71%; ii) DIBAL-H, THF, 3 h, -78 °C, 66%; iii) DMSO, 1 h, rt, 18%.

The diester, diethyl 3,3'-(1,3-phenylene)(2*E*,2'*E*)-diacrylate (**23**), was successfully synthesised from isophthalaldehyde (**22**) by a Wittig reaction in 71% yield, after purification by column chromatography. A new DIBAL-H reagent was purchased after the issues with the previous syntheses and the benzyl dialcohol (**24**) was successfully synthesised and isolated in 66% yield, after purification by column chromatography. The final oxidation step and work up procedure, to generate the benzyl dialdehyde (**21**), combined the step introduced for successful IBX removal in the purification of **20**. The reaction was performed as previously followed by extraction of the aqueous phase using diethyl ether, a brine wash and a base wash, using potassium hydrogen

carbonate, before drying and removal of solvents *in vacuo*. The reaction was successful, however, even after the washing steps and purification by column chromatography the IBX contamination was significant (approx. 50%), giving **21** in 18% yield, considering the contaminant present.

2.3 Protein Expression and Purification

Once the substrates had been generated, the various enzymes (ERs, ADHs and TAms) could be expressed and purified ready for testing against the synthesised substrates and other commercially available model substrates.

2.3.1 Ene-Reductases

Seven ERs from the old yellow enzyme (OYE) family were selected for use within this project as they have previously been shown to be highly active and have a diverse substrate scope which covers a range of chemical space (section 1.5.1). OYE1[181] from *Saccharomyces pastorianus*, OYE2 and OYE3[182] from *Saccharomyces cerevisiae* and a number of OYE homologs, NCR,[183] XenA,[184] MorA,[185] and YqjM[186] were selected. All seven of these enzymes were available at UCL and were transformed into BL21 (DE3)*E. coli* expression strains and made into glycerol stocks[†]. The strains were inoculated into 500 mL cultures with protein expression induced by the addition of IPTG, followed by cell harvesting, lysis, and clarification to produce crude cell lysates (Figure 2.1). OYE1-3 were previously cloned without His-tags so could not be purified further and were not used beyond this initial point. The four homologs, NCR, XenA, MorA and YqjM, were expressed with His-tags so could be purified using nickel-nitrilotriacetic acid (NTA) affinity columns. The enzymes were then used as crude cell lysates or purified proteins to investigate their activity and substrate specificity.

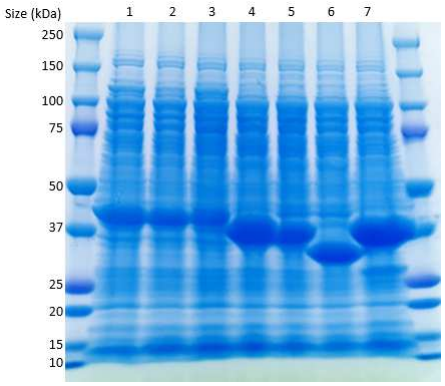


Figure 2.1: ER protein expression gels - SDS-PAGE analysis of soluble lysate fractions staining with coomassie blue. 1 - OYE1 (45.0 kDa), 2 - OYE2 (45.0 kDa), 3 - OYE3 (44.9 kDa), 4 - NCR (40.6 kDa), 5 - XenA (40.9 kDa), 6 - MorA (33.2 kDa) and 7 - YqjM (38.7 kDa).

[†]Procedure carried out by Nadine Tappertzhofen at UCL.

2.3.2 Alcohol Dehydrogenases

Sixteen genes from the UCL Drain Metagenome, pQR1492-1503 and pQR1918-1921, were previously identified as putative alcohol or aldehyde dehydrogenases and were available in pET29a plasmids as glycerol stocks in BL21 (DE3) *E. coli*.[‡] All of these genes were previously expressed and analysed [‡] and were expressed again on a small scale (20 mL) using IPTG for induction of expression. An additional five ADHs, ADH-6 (*Thermobifida fusca YX*), ADH-9 (*Thermobifida fusca YX*), BtADH (*Bacillus stearothermophilus*[187]), HLADH (*Equus caballus*[188]), and RaADH (*Rhodococcus aetherivorans*[189]) were also available having been transformed and expressed previously.[§] These enzymes were all taken forward to be tested for activity in the initial reaction in the enzymatic cascade for the nepetalactone analogue biosynthesis, the oxidation of a dialcohol (**5**) to a dialdehyde (**6**) (Scheme 2.2).

2.3.3 Transaminases

Initially, only one TAm enzyme from *Chromobacterium violaceum* (*Cv*)-TAm[190] was selected as the cascades and substrate scopes needed to be investigated with the ERs and ADHs prior to incorporating a TAm. (*Cv*)-TAm was also available at UCL and was expressed, harvested as with the ADHs and ERs, and was used when investigating the overall cascades (Scheme 2.2).

2.4 Enzymatic Assays

2.4.1 Ene-Reductase Spectrophotometric Assay

Investigating the substrate specificity of the ERs required an appropriate substrate panel to cover the chemical space related to (2*E*,7*E*)-nona-2,7-dienedial (**20**). Eight aliphatic and one aromatic α,β-unsaturated substrates were selected to fulfil this role (Figure 2.2). 2-Cyclohexen-1-one (**25**), cinnamaldehyde (**26**), 2-methylpent-2-enal (**27**), and citral (**28**) were selected as they have been used previously as substrates for investigating the activity of ERs, citral was of particular interest due to its structural similarity to the synthesised substrate **20**.[191–194] The other substrates (**29**-**33**) covered chemical space which had not previously been tested against these ERs and several of these would react to produce chiral products which would allow for investigation of the stereoselectivity of the ERs.

[‡]Procedure carried out by Dragana Dobrijevic at UCL.
[§]Procedure carried out by Maria Bawn and Daniel Mendez-Sanchez at UCL.

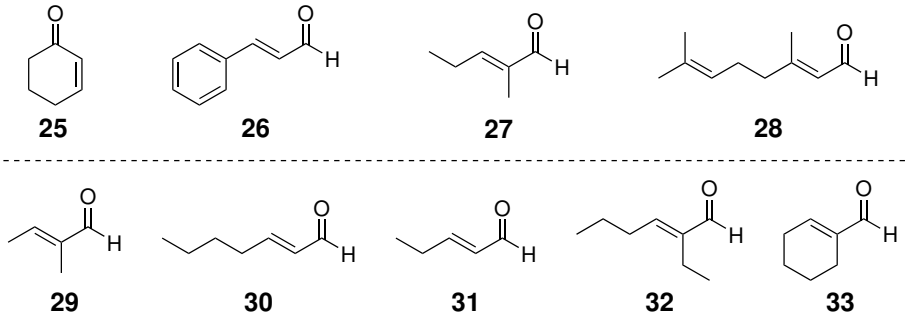


Figure 2.2: ER substrate panel - The nine substrates selected for screening against the ERs: **25**-**28** have previously been shown to be accepted by several ERs; **29**-**33** cover untested chemical space. **25**-**27** and **29** were used for initial spectrophotometric testing.

An initial investigation of the substrate scopes was performed using crude cell lysates of the four homologs (NCR, YqjM, XenA and MorA) as these have been shown to have broader substrate scopes than OYE1-3.[168] These were tested against a selection of four substrates from the panel (**25**, **26**, **27** & **29**). An established spectrophotometric assay was selected to monitor the depletion of the reduced nicotinamide cofactor (NADH) which is taken to correspond to enzyme activity.[195] These reactions were performed in triplicate on a 200 μL scale using clarified lysates (50% v/v), NADH (1 mM), and substrates (1 mM) in Tris buffer (50 mM) (method 6.1.5.1). The reactions were monitored over 1 h at 340 nm to detect the oxidation of NADH to NAD^+ (Figure 2.3).

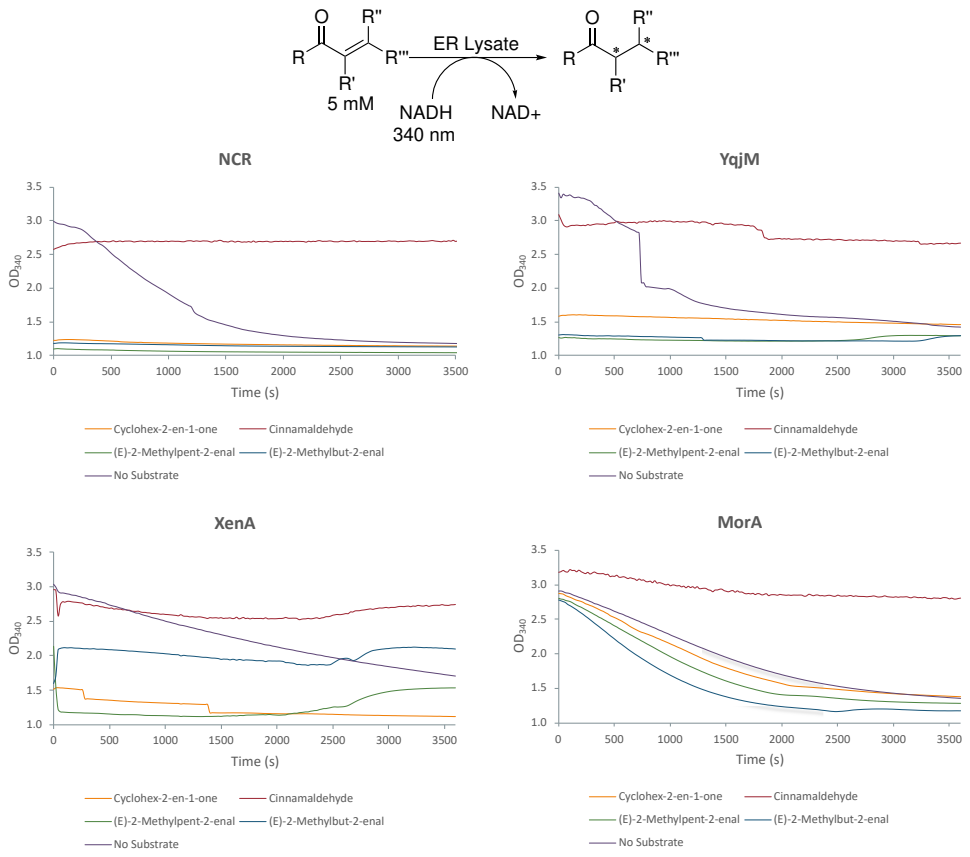


Figure 2.3: Overview of the reactions performed with NCR, YqjM, XenA and MorA lysates. Key: orange - 2-cyclohexen-1-one (**25**); red - cinnamaldehyde (**26**); blue - 2-methylpent-2-enal (**27**); green - 2-methylbut-2-enal (**29**); purple - no substrate control. 200 μL reactions were performed in triplicate with ER lysates (50% v/v), NADH (1 mM), substrate (1 mM), and Tris buffer (50 mM, pH 7.4). Reactions were monitored at 340 nm at 20 s intervals for 1 h.

These initial spectrophotometric screens showed that NCR and YqjM had good activity with three of the substrates, **25**, **27**, and **29**, depleting the NADH before the first reading. This matched with the literature where NCR and YqjM have been shown to accept **25** and **27**.[196,197] It was also unsurprising that **29** was accepted as this has a very similar structure to **27**. XenA also appeared to have similar activity, accepting the substrate it has been shown to accept **25**[198] and also **27** and **29**, however, the results were unclear so an alternative quantitative assay was required. MorA has previously been shown to accept **25**[185] and it appeared to have very slight activity towards **25** as well as **27** and **29**, however, they were all close to the background reaction so this activity would also need to be verified. All of the lysates showed a strong background reaction which was unsurprising with lysates as NADH is used in a number of native *E. coli* metabolic pathways. None of the enzymes appeared to have activity with cinnamaldehyde (**26**), however, the absorbance at 340 nm did not deplete at the background reaction rate, which would be expected if there was no activity. This was surprising as cinnamaldehyde has been shown to be accepted by XenA[199] and NCR,[200] and YqjM has shown activity with α-methyl-cinnamaldehyde.[201] It is possible that the cinnamaldehyde was diverted into another pathway or was accepted by another enzyme present

in the lysate, which recycled the NADH causing the absorbance to remain high.

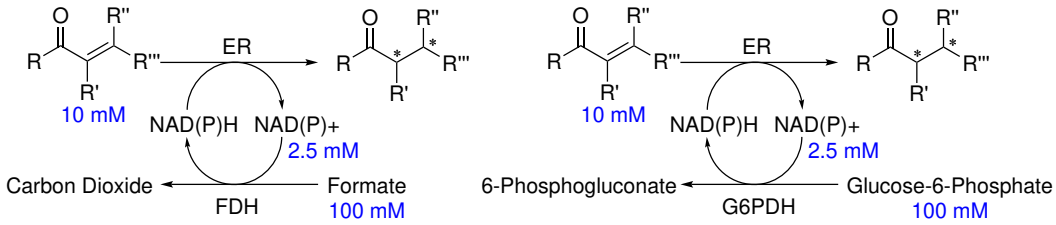
Spectrophotometric assays are a useful starting point when investigating enzymatic activity, however, it relies upon detecting cofactor depletion rather than substrate formation, it is a coupled assay. Further investigation of the substrate scopes of these enzymes required a different method to directly monitor substrate formation, gas-chromatography (GC).

2.4.2 Ene-Reductase Gas Chromatography Assays

The enzyme NCR provided the most promising initial results so was selected to perform small scale biotransformations for analysis by GC, which allows for quantitative determination of product formation or substrate depletion rather than monitoring cofactor depletion. Calibration curves of the reduced products were prepared from commercial standards, from 0-10 mM, to enable this quantitative determination of product formation. Method development was performed using the substrates and their reduced products. Two GC methods were required to ensure separation of the starting materials from the products, and any enantiomers using a chiral column (Table 6.6). The calibration curves were then prepared using these methods (Figure A.1).

The stoichiometric use of nicotinamide cofactors is not feasible at a large scale, particularly in industry, due to their high expense.[166] Cofactor recycling systems can be used, which reduce or oxidise the cofactor back to required oxidation state, to prevent the need for stoichiometric amounts of the cofactor. In order to perform the assay with all nine substrates an FDH/formate system was selected which oxidises NAD^+ back to NADH using formate as a cheap sacrificial substrate. Glycerol stocks of NCR co-expressed with FDH from *S. cerevisae* in *E. coli*[†] were grown and lysed following the same procedure as for the individual ERs. Small scale, 250 μL, biotransformations were performed in triplicate with clarified NCR/FDH lysate (10% v/v), one of the nine substrates, NAD^+, formate and DMSO (10% v/v) in Tris buffer (50 mM, pH 7.4) (Scheme 2.7).

[†]Perfomed by Nadine Tappertzhofen at UCL.



Scheme 2.7: ER biotransformations with FDH and G6PDH recycling systems - formate and glucose-6-phosphate are used as cheap sacrificial substrates to reduce the catalytic amounts of $NAD(P)^+$ cofactor back to NAD(P)H, to the correct oxidation state, to be used by the ER again. (Conditions: NCR/FDH or NCR/G6PDH lysates, 25 °C, 450 rpm, 18 h, Tris buffer (50 mM, pH 7.4))

Despite showing promising activity in the spectrophotometric assay, NCR did not show activity with any of the substrates (**25**-**33**) in the GC assay. This was highly unexpected especially as some of the substrates have previously been reported to have good activity and high conversions with NCR, **25** 83% conversion, **26** 98% conversion, **27** 49% conversion, and **28** 37% conversion.[191,200] The FDH/formate recycling system was used as this requires NAD^+ which is less expensive than $NADP^+$, however, NCR has been shown to have higher activity with NADPH over NADH. Previous work using NCR with citral (**28**) as the substrate showed conversions of 93% and 94% with NADH and NADPH respectively and when using a recycling system the conversions were increased to 98% and >99% respectively.[202] NADPH is also the physiological substrate for the reduction of FMN in the reductive half reaction for the OYEs and other ERs.[168] The combination of these factors could rationalise the lack of activity, therefore, a new assay using a glucose-6-phosphate dehydrogenase (G6PDH)/glucose-6-phosphate (G6P) recycling system and NADPH was designed.

The new assay utilising G6PDH was performed on a 200 μL scale with clarified NCR and G6PDH lysates (10% v/v each, expressed separately), substrate, $NADP^+$, G6P and DMSO (10% v/v) in Tris buffer (50 mM, pH 7.4) (Scheme 2.7, method 6.1.5.2). A smaller panel of substrates was selected for these reactions, **25**-**31**, and the recycling system was expressed separately as only NCR was available co-expressed with G6PDH and this could have caused some of the previous issues with activity due to differing expression levels. This assay was also extended to include all four of the ERs which were used in the initial spectrophotometric assay, NCR, XenA, MorA, and YqjM.

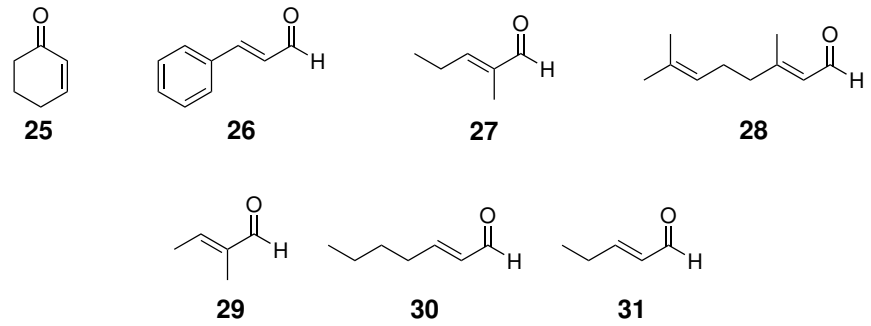


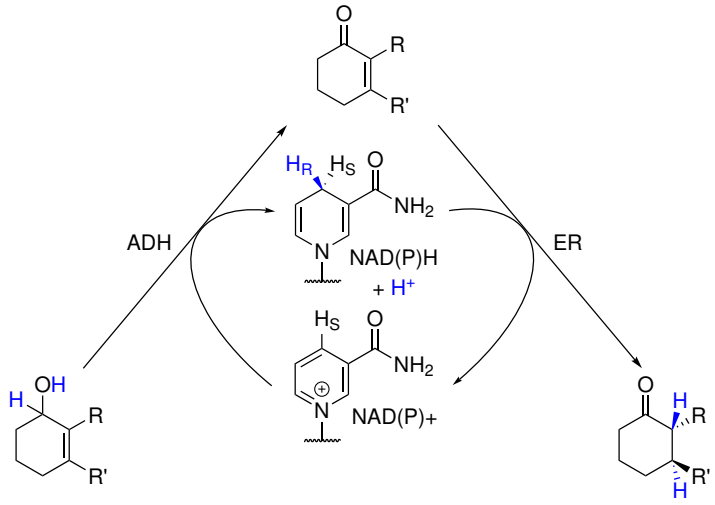
Figure 2.4: ER and G6PDH GC assay substrate panel - The reduced substrate panel from Figure 2.2 used in the ER GC assay using NADP and a G6PDH/G6P recycling system.

This assay was more successful with NCR displaying activity towards **25**, **26**, and **27**, which it has been shown to be active with previously,[196,197] **29**, and very slight activity towards **31**. The conversions for these reactions, however, were all <10%, which does not match with the literature values reported in the previous paragraph.[191,200] YqjM had a similar activity to NCR with the exception of not accepting **26**, which matches with the previous data and literature, however, **26** was expected to be accepted as YqjM accepts α-methylcinnamaldehyde.[201] XenA had a slightly lower substrate promiscuity displaying activity towards **25**, **29**, and **31**, but not towards **26**, which it has been shown to have activity with.[199] MorA showed the lowest activity only towards **25** and slight activity towards **27**. Compound **30** has not been tested previously, to the best of our knowledge, however, it is a similar structure to **28** which has been shown to be accepted by NCR,[191] YqjM,[194] and XenA.[199] However, none of the enzymes showed activity towards either **28** or **30**, these compounds should have been accepted so there may have been another issue in the extraction or detection of the compounds. The conversions for all of the successful reactions were below 10%, considerably lower than observed previously, however, this could be more to do with the assay and extraction protocol rather than the actual enzymatic activity. These conversions could be improved by further work developing a better extraction protocol or by using purified enzymes to determine the conversion and specific activity of the enzymes. Despite the very low conversions, the assay provided an indication of which substrates were accepted by each enzyme and this information could be taken forward to inform the next assays.

2.4.3 Alcohol Dehydrogenase Gas Chromatography Assays

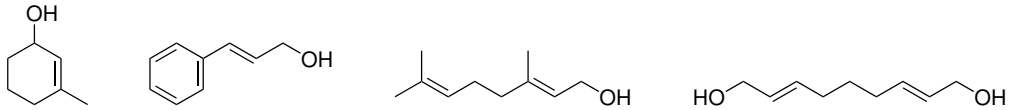
Once the preliminary activity of the ERs had been assessed, it was then possible to test the ADHs using the ERs to recycle the cofactor to eliminate the need for another recycling enzyme or sacrificial substrate. In the cascade, the ADHs were to be used in the less favoured direction,

oxidising an alcohol to an aldehyde or ketone, meaning they require an oxidised cofactor, in the form of NAD(P)$^+$, to act as a hydride acceptor. The ERs can then use the reduced NAD(P)H and transfer the hydride, and a proton, back to the substrate to reduce the double bond, completing the redox cycle and recycling the cofactor to NAD(P)$^+$ to be used again by the ADH (Scheme 2.8).



Scheme 2.8: ADH and ER Recycling System - The ADH converts NAD(P)$^+$ to NAD(P)H to oxidise the alcohol of the substrate. The ER recycles the NAD(P)H back to NAD(P)$^+$ to reduce the double bond of the substrate.

NCR was selected as the ER to act as the co-factor recycling system enzyme as it had shown activity with the widest range of substrates in the individual ER assays. Eleven ADHs were selected for initial screening: six from the metagenome, which expressed under the initial conditions, these were expressed from putative gene fragments with pQR numbers 1492, 1494, 1495, 1503, 1917 and 1920; and five available pre-expressed in house, two putative ADHs (ADH-6 and ADH-9) and three known ADHs (BtADH, HLADH and RaADH) (section 2.3.2). Four substrates were selected for this assay, the corresponding alcohols of the aldehydes and ketones used in the ER assays: 3-methylcyclohex-2-enol **34**; cinnamyl alcohol **35**; geraniol **36**, which, despite its aldehyde equivalent being unsuccessful in the previous assay (Scheme 2.7), has been shown to be accepted previously by NCR[200] and is structurally similar to the synthesised substrate; and the synthesised substrate **19** (Figure 2.5). NCR has also been used previously as part of a ADH-ER-ADH cascade using cinnamyl alcohol (**35**) and geraniol (**36**) as starting substrates.[200]



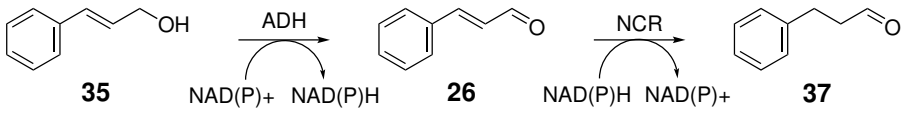
34 **35** **36** **19**

Figure 2.5: ADH and ER Assay Substrates

This ADH/ER assay was performed on a 200 μL scale with clarified ADH and ER lysates (10%

v/v each, expressed separately), substrate (10 mM), $NADP^+$ (2.5 mM), and DMSO (10% v/v) in Tris buffer (50 mM, pH 7.4). BtADH, HLADH and RaADH all showed trace activities (<5%) towards cinnamyl alcohol (**35**) but did not show full conversion or any activity with the other three substrates, including the synthesised substrate (**19**). None of the metagenomic ADHs, ADH-6, or ADH-9, showed any activity with these four substrates (**34**, **35**, **36** and **19**), however, these were only putative aldehyde/alcohol dehydrogenases so may have a very limited substrate scope or may not have dehydrogenase activity at all.

NCR has been shown to have a reduction in its activity when coupled with an ADH in this manner. Reich *et al.* [200] showed that conversions of cinnamaldehyde and citral when using NCR as a standalone enzyme were 68% and 30% respectively, but, when NCR was part of a cascade with equine ADH the overall conversions were just 4% and 3%, respectively. In order to try and compensate for this reduction in conversions, another assay was performed using purified NCR, which would enable the concentration of NCR to be increased, reduce the amount of lysate, and reduce the diversion of the NADP cofactor into other intrinsic pathways. The ADHs and ERs may also have differing cofactor preferences between NAD and NADP which could affect the activity of the overall cascade. This next assay investigated these possible affects by running concurrent reactions, one with NAD^+ and the other with $NADP^+$. The assay was expanded to include all 15 of the metagenomic ADHs alongside BtADH, HLADH and RaADH but was only run using a single substrate, cinnamyl alcohol (**35**) (Scheme 2.9). A method was developed on the GC which could separate all three cinnamaldehyde derivatives in the pathway (Methods Table 6.6). This was used to generate calibration curves of cinnamyl alcohol (**35**), cinnamaldehyde (**26**), and hydrocinnamaldehyde, to quantify the reaction (Figure A.2).



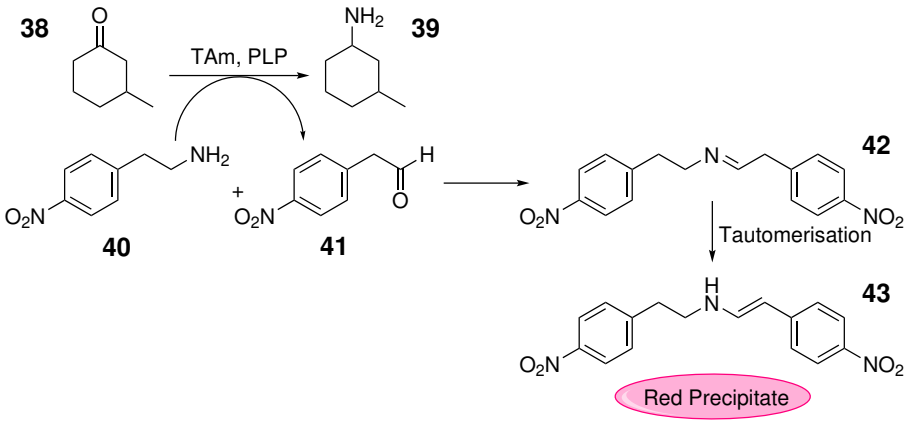
Scheme 2.9: ADH and ER GC assay with cinnamyl alcohol - Two step redox reaction of cinnamyl alcohol (**35**) using an ADH lysate and purified NCR to produce hydrocinnamaldehyde (**37**) via cinnamaldehyde (**26**).

Only one reaction, of the 36 reactions run, seemed to show conversion through the cascade with depletion of the alcohol peak for **35** and small peaks corresponding to cinnamaldehyde (**26**) and hydrocinnamaldehyde (**37**). This reaction was performed using a metagenomic ADH expressed from pQR1921 and $NADP^+$. The peak for the final product only corresponded to a concentration of 0.2 mM, a 2% conversion, however, in the reactions that did not work the peaks for the

starting material only correspond to 1.4 mM, despite 10 mM being added initially, a loss of 86%, indicating that there was an issue with the reaction work-up and extraction into ethyl acetate. This was possibly due to the ethyl acetate containing a contaminant which reacted with and caused polymerisation of the polypropylene 96-well plates. Using 1.4 mM as the total amount that could be extracted, the conversion with the metagenomic ADH expressed from pQR1921 increases to 14% which is still low but much more significant. The results for the known ADHs, BtADH, HLADH, and RaADH, seemed to show conversion to cinnamaldehyde (**26**), but not through to hydrocinnamaldehyde (**37**) with both NAD^+ and $NADP^+$. This was also seen with enzymes expressed from pQR1492, pQR1494, and pQR1495. This pattern was unexpected as NCR should have accepted and converted the aldehyde, as shown previously in section 2.4.2. This indicated that there may have been an issue with using the cascade to recycle the cofactor, such as the cofactor being diverted into other pathways present in the lysate, which would also explain the low conversions. The individual reaction components, such as the cofactor, NCR, and substrates, would need to be investigated further to determine where the problem with the cascade lies and the reactions would have to be performed in a glass plate to prevent any potential problems with extraction.

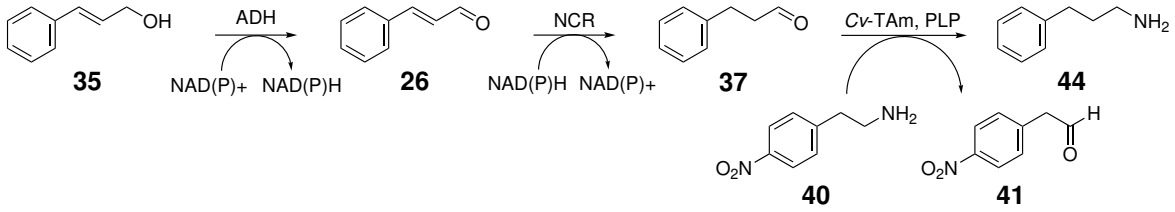
2.4.4 Transaminase Colorimetric Assay

Investigation of the full cascade, ADH-ER-TAm, was performed in parallel with the ADH-ER assays but could be investigated in a different manner. A colorimetric assay was previously developed which could be used to determine TAm activity, conversion of a carbonyl (**38**) to an amine (**39**), by using 4-nitrophenylethylamine (**40**) as the amine donor. If **38** and **40** are accepted by the TAm, **40** is converted to 4-nitrophenylacetaldehyde (**41**) which can then react with another molecule of **40** to give an imine (**42**) which tautomerises to produce a red precipitate (**43**) (Scheme 2.11).[203] The more red precipitate formed equates to a higher TAm activity with greater product formation.



Scheme 2.10: TAm Colorimetric Assay - When 4-nitrophenylethylamine (**40**) is used as the sacrificial substrate for the transaminase reaction (**38**→**39**) it produces 4-nitrophenylacetaldehyde (**41**). These two can react with one another to give an imine (**42**) which tautomerises to give a red precipitate (**43**) which can be used as a colorimetric assay for TAm activity. Scheme adapted from Baud *et al.* [203].

This colorimetric assay was used to determine if *Cv*-TAm could be coupled on to the end of the cascade after the ADH and ER reactions. The four substrates used previously, **34**, **35**, **36**, and **19**, were selected for use in this assay, along with NCR and eight ADHs, ADH-6, ADH-9, HLADH, RaADH and four expressed from pQR1492, pQR1494, pQR1917, and pQR1920. These reactions were performed on a 200 μL scale following the established literature protocol: ADH and *Cv*-TAm lysate (10% v/v each), purified NCR (0.5 mg/mL), $NADP^+$, substrate (10 mM), PLP, 4-nitrophenylethylamine (**40**), and phosphate buffer (100 mM, pH 7.5) (Figure 2.6) (section 6.1.5.3).[203]



Scheme 2.11: Full Cascade with ADH, NCR, and *Cv*-TAm - Example of the full cascade using cinnamyl alcohol (**35**) as the starting substrate. The alcohol is oxidised to cinnamaldehyde (**26**) by an ADH, which is then reduced by NCR to hydrocinnamaldehyde (**37**). **40** is then used as an amine donor by *Cv*-TAm to give 3-phenylpropylamine (**44**).

The first time this assay was performed (Figure 2.6 (i)), the results were consistent with the control across the plate with one exception, ADH-9 seemed to show activity with 3-methyl-2-cyclohexenol (**34**) above the no enzyme control, however, the no substrate control had shown slight colour change. This assay was repeated to determine if this result was valid (Figure 2.6 (ii)). ADH-9 and now ADH-6 showed activity above the no enzyme control, but again, the no substrate control was also darker for these two than the other controls. This showed possible activity, but this would need to be confirmed through another analytical method, such as GC or

HPLC, to confirm these results.

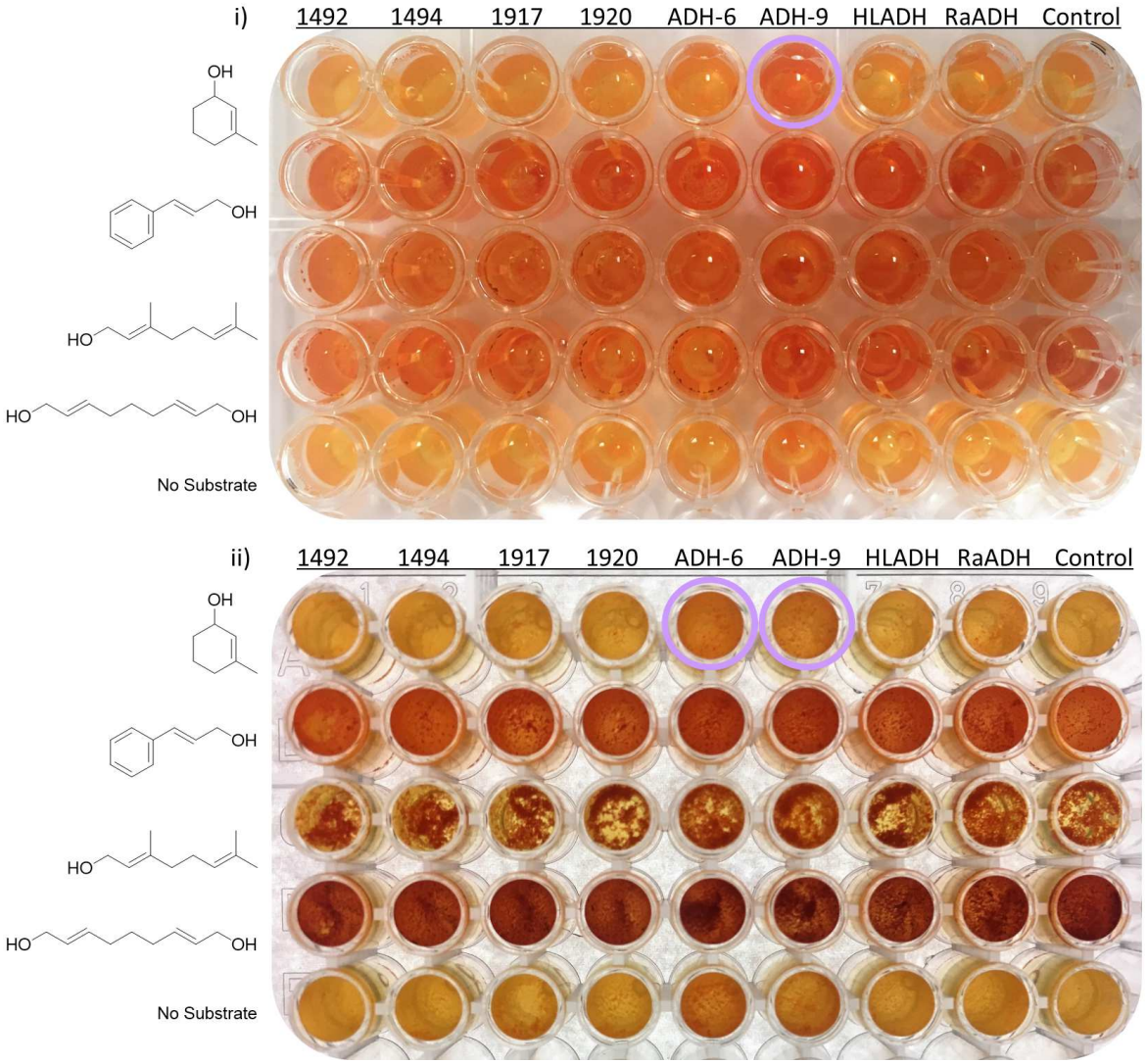


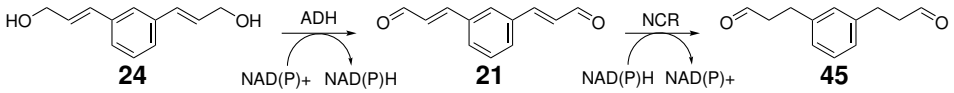
Figure 2.6: TAm colorimetric assay - The substrate for each reaction is shown on the left (top to bottom: **34**, **35**, **36**, and **19**), whilst the ADH used is shown along the top of each panel. i) ADH-9 with **34**, circled, appears to show a colour change which indicated full activity through the cascade using the ADH-9, NCR and *Cv*-TAm; none of the other reactions produce a significant colour change. ii) ADH-6 and ADH-9 show a colour change in the repeat reactions indicating activity as with (i). Note - the pictures look very different due to taking pictures from above and below the plate showing the precipitation more clearly in the second picture.

2.4.5 Alcohol Dehydrogenase and Ene-Reductase Assay with the Synthesised Substrates

Once the benzyl dialcohol (**24**) and benzyl dialdehyde (**21**) had been synthesised they were screened against the ERs and ADHs. The presence of a phenyl ring in these structures meant that they could be detected by a UV detector so HPLC analysis could be used. GC analysis is particularly useful for analysing low molecular weight aliphatic compounds with high volatility, however, as the flame ionisation detector detects all molecules that are injected, it can make analysis, particularly of complicated mixtures like enzyme assays, more challenging. Using HPLC with a UV detector will only detect UV active molecules so a number of peaks which would show

up by GC do not show up by HPLC. GC usually requires extraction of aqueous solutions, as it is run with low boiling point, volatile solvents, whereas reverse phase HPLC uses acetonitrile and water so aqueous solutions do not need to be extracted. This simplifies the analysis further and eliminates the extraction step, which had been problematic, affecting yields, and preventing accurate determination of product formation.

The benzyl dialcohol (**24**) and benzyl dialdehyde (**21**) could both be detected using a 20 minute HPLC method from 5-95% acetonitrile in water with 0.1% TFA. The retention times for **24** and **21** were 8.9 and 10.1 minutes respectively (Appendix Figure A.3). The three known ADHs, BtADH, HLADH and RaADH, were screened against the benzyl dialcohol (**24**), using NCR to recycle either NADPH or NADH, and NCR was screened as a standalone enzyme against the benzyl dialdehyde (**21**), using a stoichiometric amount of NAD(P)H (Scheme 2.12).



Scheme 2.12: Benzyl dialcohol reaction - Two step redox reaction of **24** using the known ADHs, BtADH, HLADH and RaADH, and NCR to produce **45** via **21**.

Unfortunately, the tandem assay with NCR and the three ADHs did not show any conversion of **24** to **21**. However, NCR appeared to show possible successful reduction of the alkenes of **21** to give **45** by a shift of the HPLC peak from 10.1 to 10.5 minutes (Appendix Figure A.3). A standard of **45** would have to be synthesised chemically in order to confirm this result and to be able to accurately calculate the conversions of the substrate to the product. This reaction could also be scaled up to produce enough product for NMR or LC-MS analysis as an alternative method to validate this result. If this result was confirmed this assay could be repeated expanding the ADH panel to find two enzymes which would work together and then extending further to include a TAm to complete the full cascade.

2.5 Conclusions

This chapter described the successful synthesis of two α/β-unsaturated dialcohols and dialdehydes and the development of a robust synthetic route to generate additional α/β-unsaturated substrates. The four substrates generated from this synthetic effort, **19**, **20**, **21**, and **24**, were then tested against a variety of ERs and ADHs.

In total 23 enzymes, seven ERs, fifteen ADHs, and one TAm, were successfully recombinantly expressed and, along with five additional ADHs available at UCL, were used to test the cascade using the synthesised substrates, to attempt to generate a three enzyme cascade. A panel of additional substrates was also used to further probe the substrate scope of these enzymes. Two of the ERs, NCR and YqjM, showed good initial activity towards five and four of the initial substrates, respectively. However, none of the enzymes showed activity with hex-2-enal or citral, which have the most structural similarity to the synthesised substrates, indicating that these enzyme were possibly not suitable for the designed cascade.

When the investigation of the cascade was extended to include ERs and ADHs, the recycling system, which had to be used for the ERs, was in principle no longer required as the two enzymes could recycle the NAD(P) cofactor during the redox cycle of the substrate. Using NCR as the ER, the known ADHs, BtADH, HLADH and RaADH, all showed some activity with cinnamyl alcohol, converting it to cinnamaldehyde for NCR to convert to hydrocinnamaldehyde. However, when using purified NCR and a wider panel of ADHs, this conversion through the pathway was not seen and only a low level of conversion was seen with one of the metagenomic ADHs expressed from pQR1921. Several of the reactions seemed to show conversion to cinnamaldehyde but not to hydrocinnamaldehyde showing that there was a problem with the NCR catalysed reaction. This could have been due to the NCR being affected by the ratio of NAD(P)H/NAD(P)$^+$ when used as a purified enzyme. This has been shown to drastically reduce NCR activity as the NAD$^+$ concentration increases in relation to NADH.[200] Without the fast recycling system of G6PDH, the concentration of NAD$^+$ may simply be too high and negatively impact the activity of NCR. Supplementing the cascade with an external recycling system may help to overcome these problems for future assays. This assay could also have been greatly improved through optimisation studies, such as DoE (design of experiments), which would allow for the different reaction conditions, such as temperature, pH, substrate concentration, and cofactors, to be properly analysed to determine the optimal conditions for this coupled assay.

The full cascade, ADH-ER-TAm, was investigated using a TAm-based colorimetric assay and showed potential activity with ADH-6 or ADH-9, NCR and *Cv*-TAm using 3-methyl-2-cyclohexenol. These reactions would have to be scaled up and analysed using another analytical method, such as GC or HPLC, to confirm if full conversion through the cascade was achieved and at which percentage of conversion. Unfortunately, the full cascade, as designed, could not be realised as the substrate

required for producing nepetalactone analogues was not accepted by any of the enzymes tested. NCR has been shown to accept citral well, which is structurally similar to the substrates in the designed cascade, so as with the ADH-ER assay, this designed cascade may also have required optimisation. The phenyl derivative of this substrate showed possible activity with NCR, but again this could not be confirmed on the scale it was performed at, nor using the analytical methods available. This could be investigated further by using larger scale biotransformations to produce standards for analysis, or through synthetic chemical routes. The panels of ADHs and ERs could also be expanded to try and find two compatible enzymes to work as a co-factor recycling system for the cascade, or supplementing the cascade with a more efficient co-factor recycling system. The panel of TAms could also be expanded once the ADH-ER recycling system is established beyond *Cv*-TAm to find three compatible enzymes to complete the full cascade.

This cascade was designed as a model system for immobilisation. Once two or three compatible enzymes (ADH/ER or ADH/ER/TAm) have been identified, this cascade could be used to test immobilisation platforms, such as those discussed and developed in the following chapters.

3 | Cholesterol Linker Synthesis and Testing

This chapter discusses approaches towards the development of a novel enzyme immobilisation platform using a cholesterol-NTA linker to tether His-tagged proteins to a lipid membrane surface.

3.1 Introduction

Immobilisation of enzymes, as discussed in section 1.4, is a widely used and useful technique, particularly in relation to industrial biocatalysis, as it can help to improve a number of enzyme features including stability, selectivity, and recyclability. Immobilisation of an enzyme cascade can provide further benefits including increasing flux, preventing toxic intermediates from being released, reducing inhibition, and possible substrate channelling. The work described in this chapter aimed to develop a biocatalytic platform which could provide all these benefits whilst trying to overcome some of the drawbacks, such as reduction in catalytic activity upon immobilisation, and issues associated with deactivation of immobilised enzymes.

The four main types of immobilisation, non-covalent, ionic, covalent, and cross-linking, all have their advantages and disadvantages (see section 1.4.2) and require different levels of modification of the enzyme and support. Ionic immobilisation was selected for this platform as it is relies on stronger bonds than non-covalent immobilisation, but also requires limited modification of the enzyme. His-tags are routinely incorporated into enzymes to aid purification, so this method of enzyme modification was selected to maximise the scope of the immobilisation platform as, in theory, any His-tagged enzyme could be bound to the linker. Ionic immobilisation is also reversible, making it more applicable for industrial applications as inactive enzymes could be removed and replaced.

The immobilisation platform was designed to incorporate the ionic binding of, in this case, up to three enzymes through a metal complex to a cholesterol based linker and then rely upon the intrinsic membrane insertion properties of cholesterol to tether the enzymes to a lipid membrane (Figure 3.1). Lipid membranes were selected as the support for the immobilisation as they are relatively inexpensive and straightforward to form. More importantly, they provide a fluid surface, which would enable anything tethered to the surface some degree of mobility to ensure that the enzymes can access the substrates. This would likely limit three-dimensional diffusion of substrates away from the surface and increase the flux through a cascade. A cholesterol-

based linker was selected to complement the membrane flexibility, as it readily inserts into lipid membranes and provides some rigidity to the membrane to ensure that, whilst still fluid, the membrane remains stable with the tethered enzymes attached.

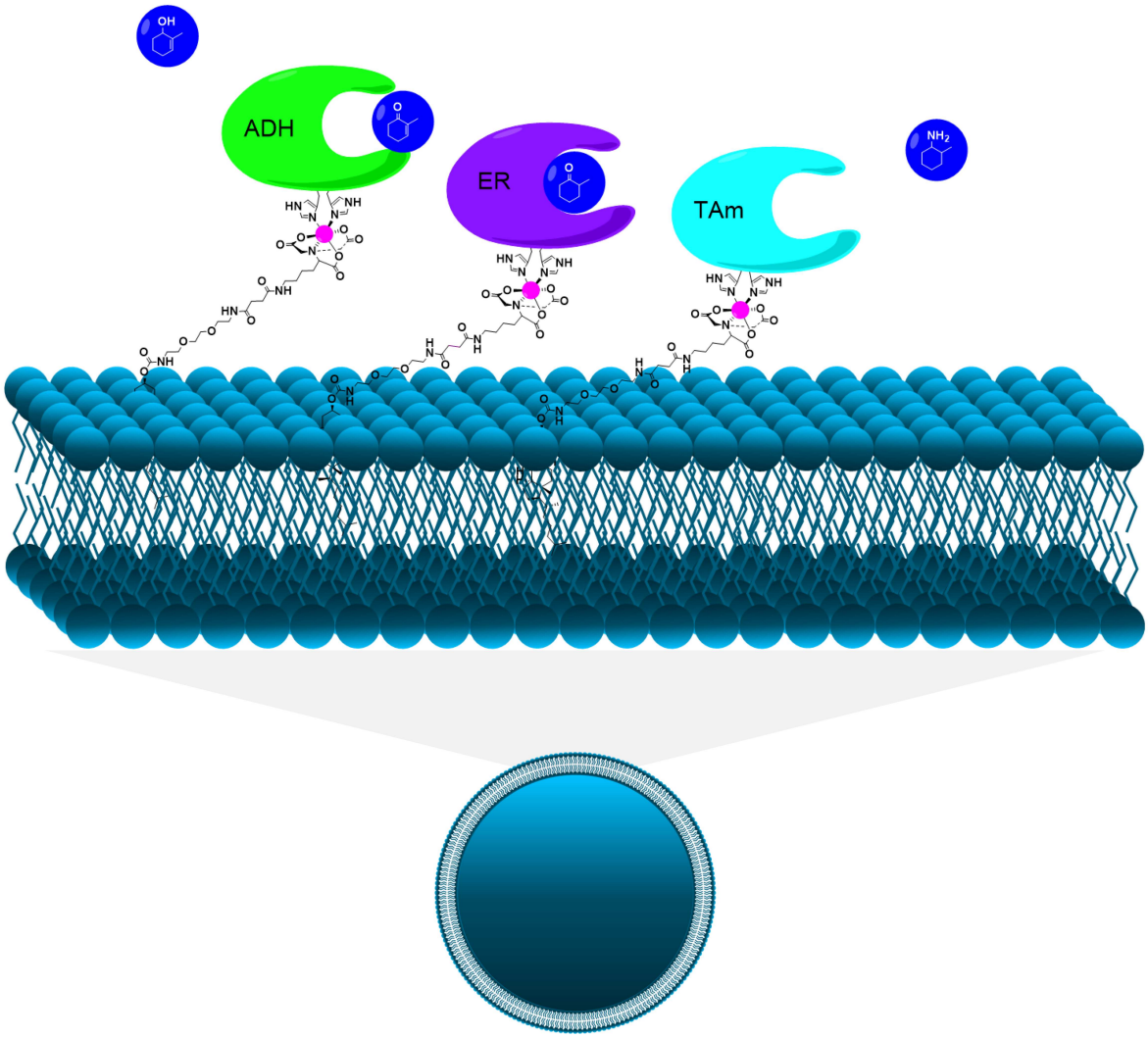


Figure 3.1: Design of the cholesterol linker immobilisation platform - The three enzymes are tethered to a lipid bilayer, which is part of a GUV, through a cholesterol-NTA linker. This platform relies upon the ionic bond formed between the His-tag on the protein and the cobalt ion on the linker, and upon the membrane inserting characteristic of cholesterol.

3.2 Cholesterol Linker Design

The first stage in developing this novel immobilisation technique required the design of a cholesterol based linker which would be be able to tether a protein to a lipid bilayer surface. The design of this linker went through several iterations until a final design was selected. The linker had several requirements, it needed: a hydrophobic portion to embed the linker into a lipid membrane; a spacer to extend the linker away from the membrane surface; and a means of attaching the linker to an enzyme.

The most challenging part of the design was establishing a method to attach the enzyme to the linker, as this would require a linkage which would be strong enough to tether the enzyme without affecting its activity. One approach to achieve the desired linkage was by taking inspiration from enzyme purification techniques which use nickel-NTA resins and His6-tags.[147] Many enzymes, including those used in the previous chapter, have His6-tags introduced genetically in order to aid purification from untagged cellular proteins. Therefore, the use of a nickel-NTA based linkage is beneficial in making this design widely applicable for a range of enzymes. Ni-NTA resins utilise these NTA groups, attached to agarose beads through hydrocarbon chains, which consist of three carboxylic acid groups attached to a nitrogen which chelate to nickel(II) through four ionic bonds (Figure 3.2). NTA groups are most commonly formed by attaching two additional ethanoic acid groups to an amino acid. Lysine is usually used as this amino acid so it was included as part of the linker design.

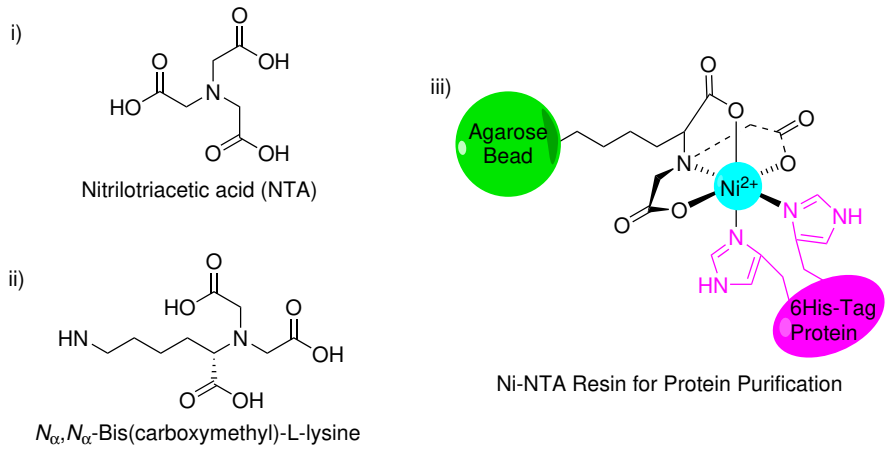


Figure 3.2: Various NTA groups - i) Structure of NTA. ii) Structure of lysine-NTA, one of the most common ways of introducing an NTA group. iii) A Ni-NTA protein purification system, the NTA group is attached to an agarose bead and chelates a Ni^{2+} ion which can then bind to a His6-tagged protein.

The bonds formed between a nickel-NTA group and a His6-tag are ionic, meaning that they are transient and affected by salt concentrations, which is a useful feature during protein purification in order to remove the purified protein. However, this could cause problems when trying to more permanently bind a protein to the linker for use in flow systems, for example. In order to overcome this, a different metal ion, cobalt, was selected to provide a more stable bond. Cobalt(II), like other divalent ions, can be used to directly replace the nickel(II) with similar affinities, however, when it is oxidised to cobalt(III) a highly stable complex with the protein is observed.[204] Cobalt(III) is a d^6-ion which forms low-spin octahedral paramagnetic complexes. These complexes have significantly higher formation constants and undergo significantly slower ligand exchange than their divalent counterparts. For example, the water complex of Co(II) has an

exchange rate of 3x10^6 s^{-1} while the Co(III) complex has an exchange rate of less than 10^{-6} s^{-1}. However, [Co(III)NTA(His6)] complexes form very slowly so an indirect method of formation is required, first forming the Co(II) complex and then oxidising it, with an oxidant such as hydrogen peroxide. Hydrogen peroxide has been shown to not denature some proteins, for example His6-GFP immobilised on Co(II)-NTA, up to a concentration of 100 mM so this indirect oxidation method was included in the design and synthetic route towards the cholesterol linker.[204]

The two remaining portions of the linker were based upon cholesterol and triethylene glycol (TEG), which are regularly used in tethering DNA strands to vesicles.[205] Cholesterol is a highly hydrophobic moiety which means it easily embeds into lipid membranes and it is used universally in membranes, both natural and synthetic, to give the membrane stability.[206] Cholesterol is also mostly unreactive under physiological conditions making it compatible with the enzymes in the designed system, and a good choice for the hydrophobic portion of the linker. This was derived from cholesteryl chloroformate to provide, not only a commercially available starting material, but also a reactive centre from which to start the system.

A hydrocarbon chain, derived from 12-aminododecanoic acid, was initially considered to form the spacer section (Figure 3.3), however, this was not selected as it was thought to be too hydrophobic and may have prevented the linker from protruding from the membrane and attaching to an enzyme. The next approach selected utilised a single TEG spacer, which overcame the problem with hydrophobicity, however, this would have required a weaker carbonate bond between the spacer and the cholesterol which could be cleaved under relatively mild reaction conditions during assembly or use. The final spacer design, selected for the cholesterol linker, was derived from 1,8-diamino-3,6-dioxaoctane and combined the hydrophilic TEG linker with the stronger carbamate bond. Finally, this spacer section needed to be able to react with the lysine to form the chelating group so an additional section, derived from succinic anhydride, was added to extend the spacer further and also to provide a carboxylic acid group for the amide bond formation.

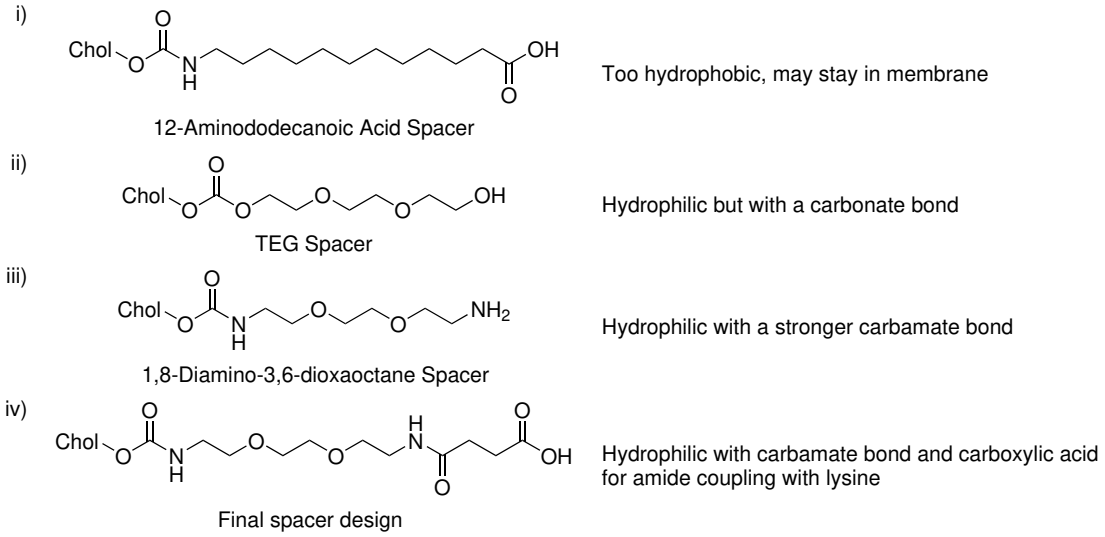


Figure 3.3: Development of the linker spacer design. i) 12-Aminododecanoic acid spacer - too hydrophobic; ii) TEG spacer - overcame the hydrophobicity, but the carbonate bond is weak. iii) 1,8-Diamino-3,6-dioxaoctane spacer - overcame the hydrophobicity issues and uses the stronger carbamate bond. iv) Spacer iii was extended with succinic anhydride to provide a carboxylic acid end to react with lysine.

The final design combined all of the desired elements (Figure 3.4): a hydrophobic cholesterol moiety (blue) to embed the linker into a lipid bilayer; a hydrophilic spacer (green and purple) to extend the other half of the linker out of the membrane and provide the desired reactivity; and the protein attachment point (red) derived from L-lysine and two pendant carboxylic acid groups, which can chelate a cobalt(II) ion and by extension a 6His-tagged protein. Once this coordination has occurred the cobalt can be oxidised to cobalt(III) using hydrogen peroxide to make the coordination non-labile.

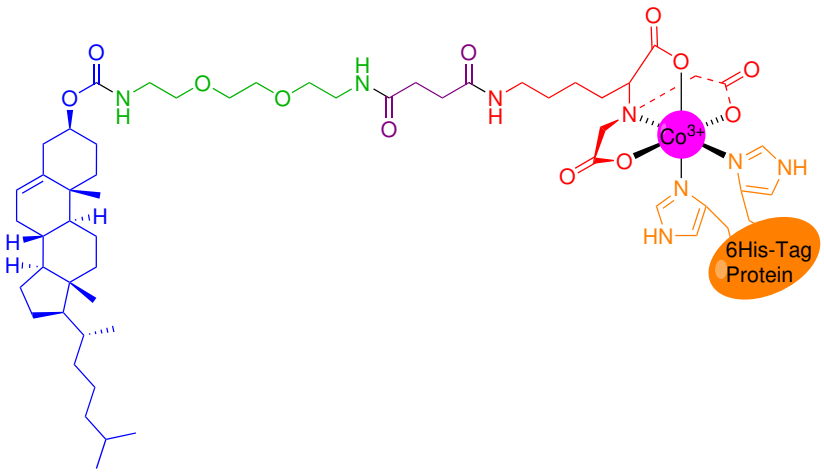


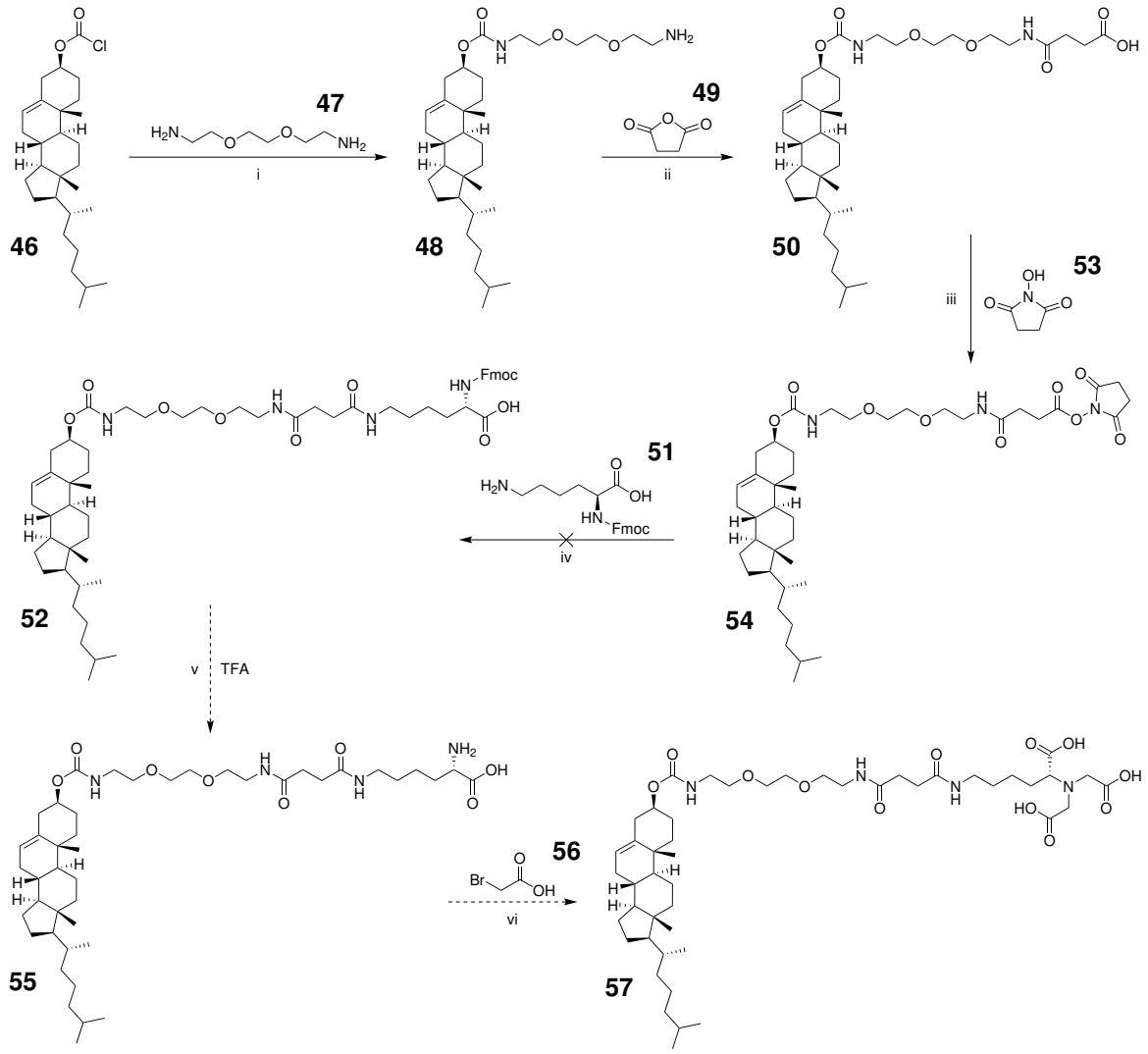
Figure 3.4: Final design of the cholesterol based linker - The linker is divided into four sections: a hydrophobic cholesterol section (blue); a hydrophilic spacer (green) derived from 1,8-diamino-3,6-dioxaoctane and extended (purple) with succinic anhydride to provide the desired carboxylic acid group for amide bond formation; and the NTA ligand (red) derived from L-lysine and two acetic acid groups which chelates cobalt (II). The cobalt(II) can then coordinate to an enzyme with a His-tag, before oxidation to cobalt(III), strongly tethering the protein to a membrane.

3.3 Cholesterol Linker Synthesis

The synthesis of the cholesterol linker initially involved an eight-step synthesis starting from cholesteryl chloroformate (**46**) (Scheme 3.1). The initial step (i) involved forming a carbamate bond between **46** and 1,8-diamino-3,6-dioxaoctane (**47**), using triethylamine (TEA) as the base, which was achieved in 90% yield (5 g scale) to give **48**.[207] This reaction was successfully performed on small scales from 280 mg-1 g as well as on larger scales up to 10 g, with the larger scale reactions giving higher yields due to easier product separation during silica column chromatography. The carbamate, **48**, was then reacted with succinic anhydride (**49**) to extend the linker and provide the carboxylic acid group to react with lysine in the next step (ii).[208] This reaction was successfully performed on small scales from 200 mg-1 g to give **50**. When performing the reaction on a small scale, the reaction product could be extracted into ethyl acetate prior to column chromatography, however, when the reaction was scaled up, to 5 or 10 g, a precipitate was formed, which caused problems with the extraction procedure. This step was removed in favour of directly purifying the reaction mixture by silica column chromatography, which lead to the highest yield of 76% being achieved at the largest scale of 10 g.

Fmoc protected lysine (**51**) was selected for the amide coupling step to generate **52**. The addition of Fmoc-lysine-OH (**51**) required *in situ* activation of the carboxylic acid group of **50** using *N*-hydroxysuccinimide (NHS, **53**) and *N*-(3-Dimethylaminopropyl)-*N*'-ethylcarbodiimide hydrochloride (EDC.HCl) (iii) to generate **54**, prior to solution phase amide bond formation (iv) to generate **52**. However, these steps (iii/iv) proved problematic and the mass spectrometry analysis showed only starting material (**50**) and a side product with a m/z of 973 which did not correspond to any of the expected side products. After several failed attempts to generate **52**, due to unknown products being formed, by mass spectrometry analysis of the reaction mixture, a different method using Fmoc solid phase peptide synthesis (SPPS) methodology was selected (Scheme 3.2). This would enable the carboxylic acid group of Boc-lysine (**58**) to be protected, by attachment to a resin, to ensure the side chain amine group was the only group available for coupling to **50**. Performing the reaction on a resin would also simplify the purification of the final product as any unreacted starting material or side products could simply be washed away before the product is cleaved from the resin. A high-loading 2-chlorotrityl chloride resin was selected for this as it only requires a low concentration of acid, only 2% TFA, in order to cleave the final product, which ensures that other hydrolysable groups in the molecule remain intact.[209] The resin also prevents racemisation of the attached amino acid, which could affect its orientation

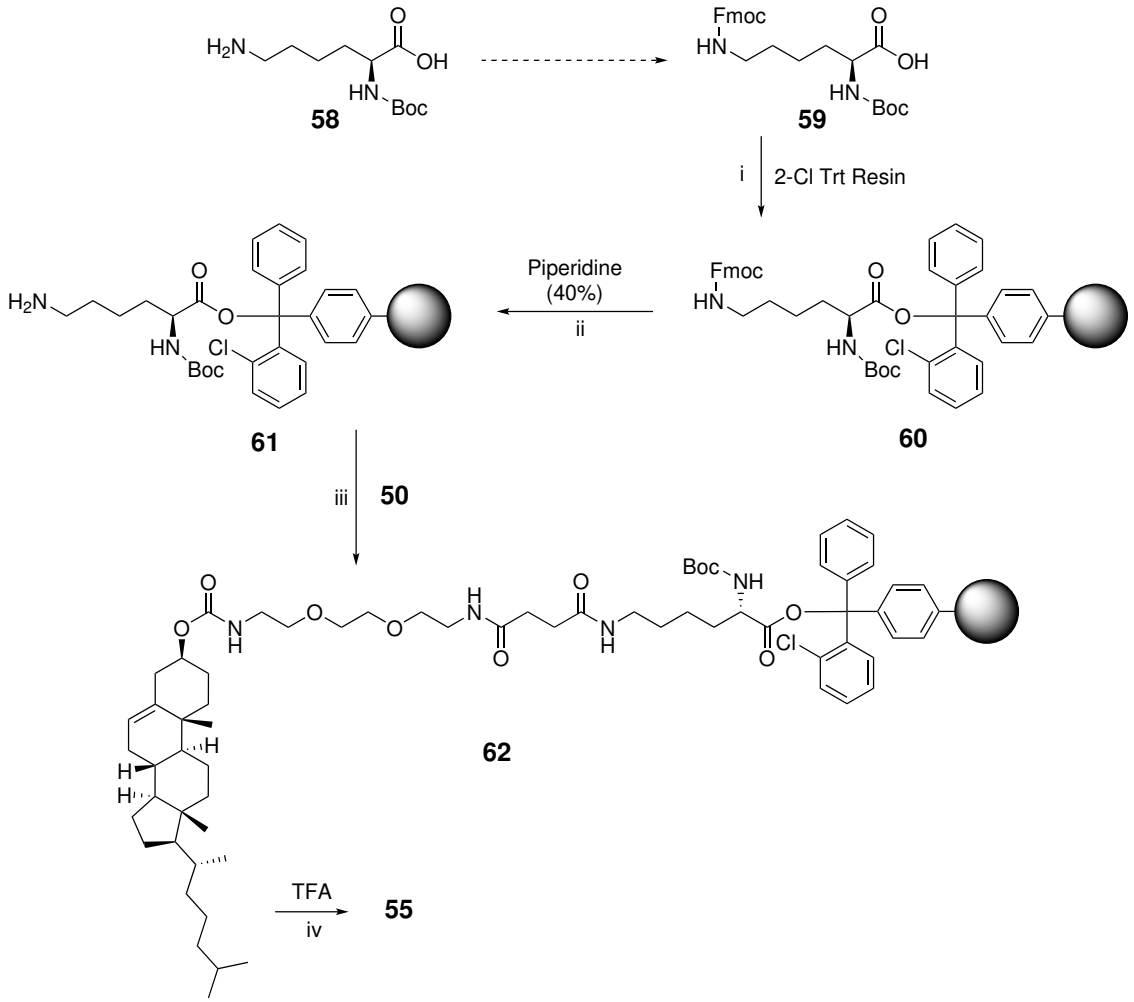
for chelation later on, and cleavage of the product with trifluoroacetic acid (TFA) would also simultaneously cleave the Boc protecting group.



Scheme 3.1: Synthetic route for the cholesterol linker (**57**) from cholesteryl chloroformate (**46**) - i) TEA, CH_2Cl_2, 4 h, rt, 90%; ii) THF, 3 h, rt, 76%; iii) EDC.HCl, DMSO-water, 15 min, rt; iv) DMSO, 2 h, rt; v) CH_2Cl_2, 2 h, rt; vi) TEA, $CHCl_3$, 0 °C (1 h), rt (12 h), 50 °C (2 h).

In order to attach Boc-lysine (**58**) to the resin, the di-protected equivalent Boc-Lys(Fmoc)-OH (**59**) was required to ensure only the carboxylic acid could attach to the resin. The Boc-Lys(Fmoc)-OH (**59**) was available commercially and was loaded onto the resin, using *N,N*-diisopropylethylamine (DIPEA) as a base, to give a loading of 0.51 mmol/g for **60**, as determined by a standard Fmoc loading test (see section 6.1.6.1 for details).[210,211] The side chain amine group was then de-protected using a piperidine solution to remove the Fmoc group and give **61**, with the reaction being stopped when the flow through was no longer active under a UV lamp. The final step on the resin was reacting the side chain amine of **61** with **50** to form an amide using standard coupling conditions of 1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) and DIPEA to give **62**. This reaction was monitored using a

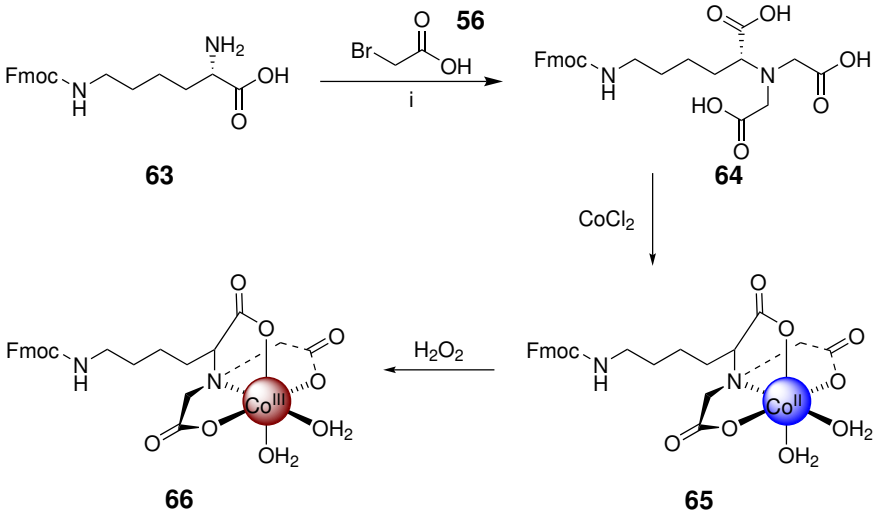
ninhydrin test to determine if the free amine had been reacted successfully (6.1.6.2).[211] The final resin product (**62**) was then acid cleaved with TFA to release the product and simultaneously cleave the Boc group to give **55** in a 98% yield, over the two reactions with **50** on the resin (iii and iv).[208]



Scheme 3.2: Solid phase peptide synthesis of **55** combining Boc-Lysine-OH (**58**) and **50** - i) DIPEA, CH_2Cl_2, 2 h, rt, 70% loading; ii) DMF, 15 min x3; iii) HATU, DIPEA, DMF, 2 h, rt; iv) TFA (25%), TIPS (2.5%), H_2O (1%), CH_2Cl_2, 45 min x2, rt, 98% (two steps).

The final step, before introducing cobalt(II), required forming the NTA ligand, which would chelate the metal ion. This involved introducing two additional carboxylic acids to the α-amino group of the lysine moiety.[212] Before performing this reaction on the cholesterol linker, a reaction using lysine(Fmoc)-OH (**63**) was carried out to test the procedure (Scheme 3.3). In order to produce the Lys(Fmoc)-OH (**63**), Boc-Lys(Fmoc)-OH (**59**) was Boc deprotected, using a solution of TFA (25%) in dichloromethane, to give **63** which was reacted *in situ* with two equivalents of bromoacetic acid to add the two acetic acid groups to form the NTA tridentate ligand. Initially this reaction was worked up using a basic wash (1 M sodium hydroxide) which generated a large amount of material, however, even after multiple washes and extractions the characterisation data

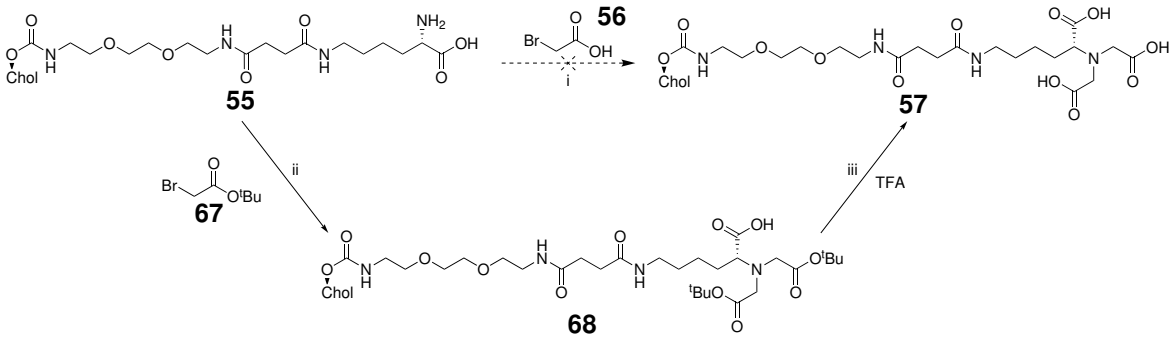
for **64** remained inconclusive. It was likely that the reaction did not occur, the TFA could have affected the addition step, or a sodium salt was formed during the basic wash that could not be identified. This reaction was repeated as before, however, as the next step required the chelation of cobalt(II) this was incorporated into the work up in place of the sodium hydroxide used previously. Cobalt(II) hydroxide was initially selected, however, due to the very low solubility of this compound in water it was replaced with cobalt(II) chloride. This appeared to be successful, on a small sample, generating a blue solid (**65**), which oxidised to a brown solid on the addition of hydrogen peroxide, suggesting the presence of cobalt in the compound (**66**). Due to the paramagnetic nature of cobalt(II) the analysis of this compound could not be performed using NMR spectroscopy and due to its insolubility in water, methanol, and dichloromethane it could not be analysed using mass spectrometry. Despite this, the procedure, and the extraction method, were taken forward and applied to the cholesterol linker (**55**).



Scheme 3.3: Synthesis of the NTA ligand using Lys(Fmoc)-OH (**63**) as a mimic for the cholesterol linker - i) TEA, $CHCl_3$, 0 °C (1 h), rt (18 h), 50 °C (2 h). **65** was obtained through extraction of **64** from $CHCl_3$ with anhydrous cobalt(II) chloride (excess). Oxidation of cobalt(II) to cobalt(III) with hydrogen peroxide (10 mM) induced a colour change from blue to brown.

The attachment of the two acetic acid groups to **55** using 2-bromoacetic acid, unlike with the test compound **63**, proved problematic. The mass spectrometry analysis of this reaction showed that there was only starting material present and no coupling had occurred. This reaction was repeated varying the base from TEA to DIPEA and K_2CO_3 as well as changing the concentration of the base and switching the solvent from CH_2Cl_2 to DMF to no avail. The addition of carboxylic acids can be problematic due to issues with solubility, anion formation, complexation with other species, and zwitterion formation of intermediates. Therefore, protecting the acid groups for the coupling reaction could help overcome some of these issues and provide an alternative route to the product. A *tert*-butyl protecting group was selected as this can be removed easily under acidic

conditions with TFA. Commercially available *tert*-butyl-2-bromoacetate (tBBA, **67**) was reacted on a small scale with **55** using either triethylamine (TEA) or DIPEA as the base and following the same reaction conditions as for the 2-bromoacetic acid. The reaction using DIPEA proved successful, confirmed by an *m/z* of 1018 by mass spectrometry analysis, and gave **68** in 54% yield after purification by column chromatography. This protected product, **68**, was deprotected using TFA (25%) to give the final product, **57**, with an *m/z* of 904 (negative ion mode) in 76% yield.



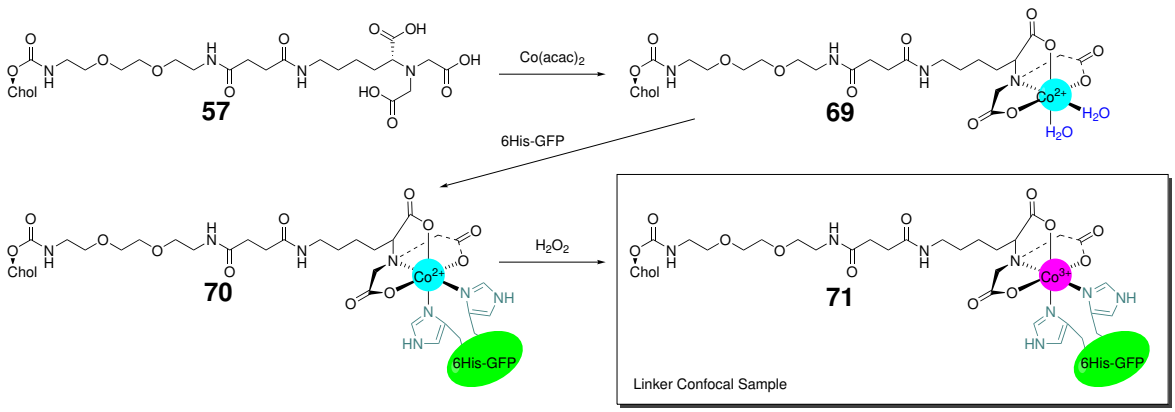
Scheme 3.4: Synthesis of the NTA ligand on the cholesterol linker (**55**) - i) TEA/DIPEA, $CHCl_3$, 0 °C (1 h), rt (18 h), 50 °C (2 h); ii) as i, 54%; iii) TFA (25%), CH_2Cl_2, 2 h, rt, 76%.

3.4 Initial Confocal Imagery

In order to analyse whether the cholesterol linker (**57**) would perform as a protein tether, a method was required that would allow visual confirmation of any binding. Confocal microscopy allows for this type of visual analysis and can confirm whether a fluorescent protein is bound to a surface through laser excitation. Therefore, a His-tagged fluorescent protein was required to form the first component of this analysis. His-tagged eGFP[213] (enhanced green fluorescent protein) was available in house as a glycerol stock (from Dr. James Allen (UCL)) so this was selected and was expressed and purified by affinity chromatography (Methods 6.1.4). The cholesterol linker (**57**) forms the second component, once cobalt(II) is added to the NTA group. The final component is the lipid bilayer for the cholesterol to embed into and tether the GFP to the surface. The easiest method for forming lipid bilayers is through the generation of vesicles as these structures form spontaneously in aqueous solutions. Giant unilamellar vesicles (GUVs) are large enough to be visualised under a confocal microscope so these were selected to act as the lipid bilayer.

The cholesterol linker (**57**) needed to undergo cobalt addition, chelation to eGFP, and oxidation before it could undergo analysis using confocal microscopy (Scheme 3.5). When the mimic (**64**) was coordinated to cobalt(II) it generated a compound that was only soluble in DMSO or DMF so

there was a concern that the coordinated cholesterol linker (**69**) would have a similar solubility. Like many proteins, eGFP can tolerate low percentages of organic co-solvents, so the linker (**57**) could be coordinated to cobalt(II) in DMSO before being added to an aqueous buffer solution. $CoCl_2$ is insoluble in DMSO so $Co(acac)_2$ was selected as an alternative source of cobalt(II). The final sample for confocal analysis (**71**) was prepared by adding $Co(acac)_2$ to 1 mg of linker (**57**) in DMSO to give the coordinated linker (Linker[Co(II)], **69**). The coordinated linker was then added to a Tris buffer solution containing eGFP, to a concentration of 5% DMSO, to allow the eGFP to coordinate to the Co(II) (Linker[Co(II)][eGFP], **70**). **70** was then oxidised with hydrogen peroxide to give the final conjugate **71** (Linker[Co(III)][eGFP]).



Scheme 3.5: Preparation of the cholesterol linker for confocal microscopy - i) The linker was first added to $Co(acac)_2$ to give **69**; ii) The GFP was then coordinated to this to give **70**; iii) Finally the whole sample was oxidised with H_2O_2 to give the final sample (**71**) ready to be added to GUVs.

Once the linker conjugate, **71**, had been prepared, it needed to be analysed under confocal microscopy to determine whether it could bind to vesicles. GUVs were prepared from the neutral lipid POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) using the inverted emulsion method.[214] POPC is a neutral, uncharged, lipid and was used to prevent any possible issues of repulsion between a charged lipid and the cholesterol moiety, which could affect membrane insertion. The inverted emulsion method involves forming a water-in-oil droplet, with a lipid monolayer, in an oil phase and transferring it into an aqueous phase, through a lipid monolayer at the phase interface, by centrifugation to give a vesicle with a lipid bilayer (Methods 6.1.8). These GUVs were then analysed by dynamic light scattering (DLS) to determine the size distribution of the population of vesicles. The majority of these vesicles were over 1000 nm meaning they can be classed as GUVs.

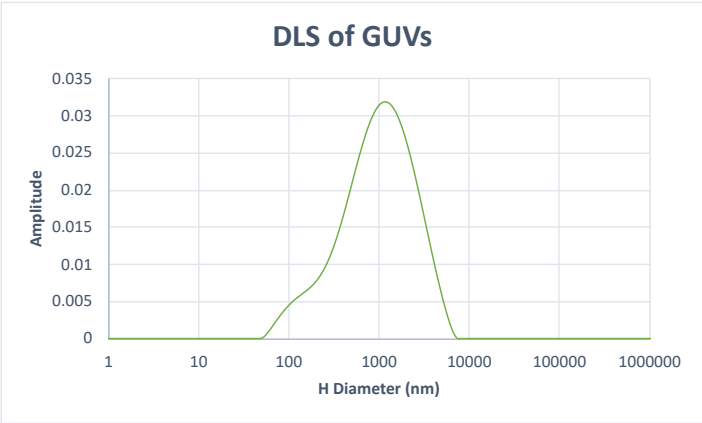


Figure 3.5: Dynamic Light Scattering (DLS) Analysis of GUVs - The GUVs formed by the inverted emulsion method were analysed by DLS which determines the size distribution of the population of vesicles. The majority of these vesicles were over 1000 nm meaning they can be classed as GUVs.

The initial experiments showed no binding between the linker conjugate (**71**) and the vesicles (Figure 3.6 (iv)), matching the negative controls (Figure 3.6 (i/ii)) . These experiments did show, however, aggregation of the linker conjugate as bright eGFP signal spots (Figure 3.6 (iii/iv)), most likely due to the amphiphilic nature of the linker itself. The critical micelle concentration (CMC) of cholesterol is 25-40 nM so a more amphiphilic molecule is likely to have an even lower CMC.[215] In order to overcome this interaction a small amount of a mild detergent, n-octylpolyoxyethylene (OPOE), was added to allow the linker conjugate (**71**) to be dispersed and bind to the vesicles. At 1% OPOE, binding to vesicles as well as some aggregation was observed (Figure 3.6 (v)), but, at 5% there was very little aggregation and clear tethering of the eGFP to the vesicles (Figure 3.6 (vi)). It is possible that OPOE could affect other enzymes in the future, however, these results provided a positive proof of principle and an excellent starting point from which to develop a more complex system.

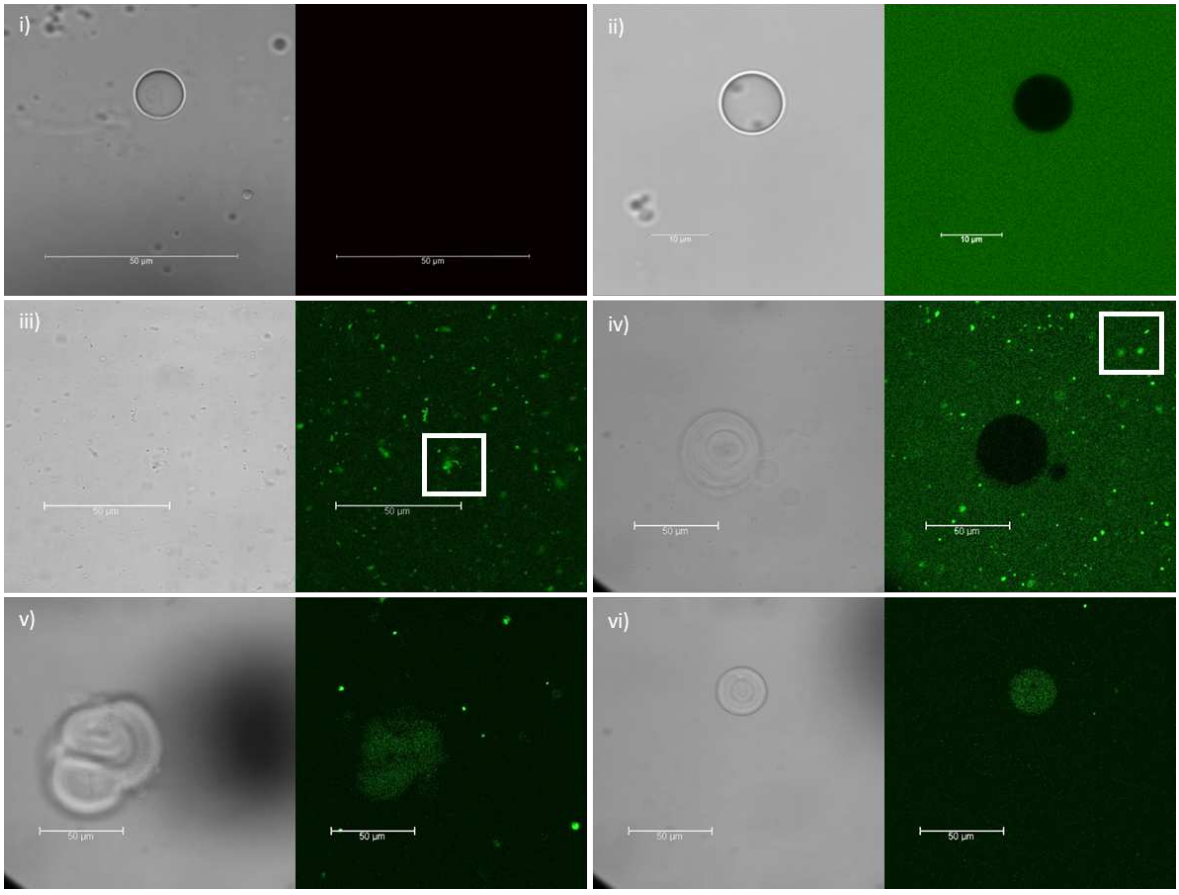


Figure 3.6: Initial confocal experiments - i) GUVs only, only visible under transmitted light; ii) GUVs and GFP negative control, GFP does not bind or enter the GUV; iii) Linker-GFP only, aggregation shown by bright GFP spots and highlighted in the white box; iv) GUVs and Linker-GFP, no binding and aggregation shown in the white box; v) 1% OPOE, binding to GUVs as well as some aggregation; and vi) 5% OPOE, clear tethering of GFP to vesicles with minimal aggregation.

3.5 Three Protein FRET System

Once initial testing of the system showed promising data with eGFP, the next step was to determine whether the system could be extended to two or three proteins. This would also be analogous to the three enzyme system envisaged in chapter 2. A good way to determine the proximity of fluorescent proteins to each other, and whether the positioning might limit diffusion of substrates away from the surface when they are replaced with enzymes, is through fluorescence resonance energy transfer (FRET). FRET is the transfer of energy from one excited chromophore, the donor, to another, the acceptor, through non-radiative dipole–dipole coupling.[216] This energy transfer is highly distance dependent, being inversely proportional to the sixth power of the distance between donor and acceptor,[217] meaning that FRET can be used to accurately measure the distance between the chromophores in the angstrom range. FRET is most efficient when the chromophores are within the Förster radius, the distance at which half of the excitation energy of the donor is transferred to the acceptor, typically 3-6 nm. Therefore, if two, or three, fluorescent

proteins within the system can undergo FRET, it would indicate that the equivalent enzymes would be sufficiently close to limit three-dimensional diffusion of the cascade substrates.

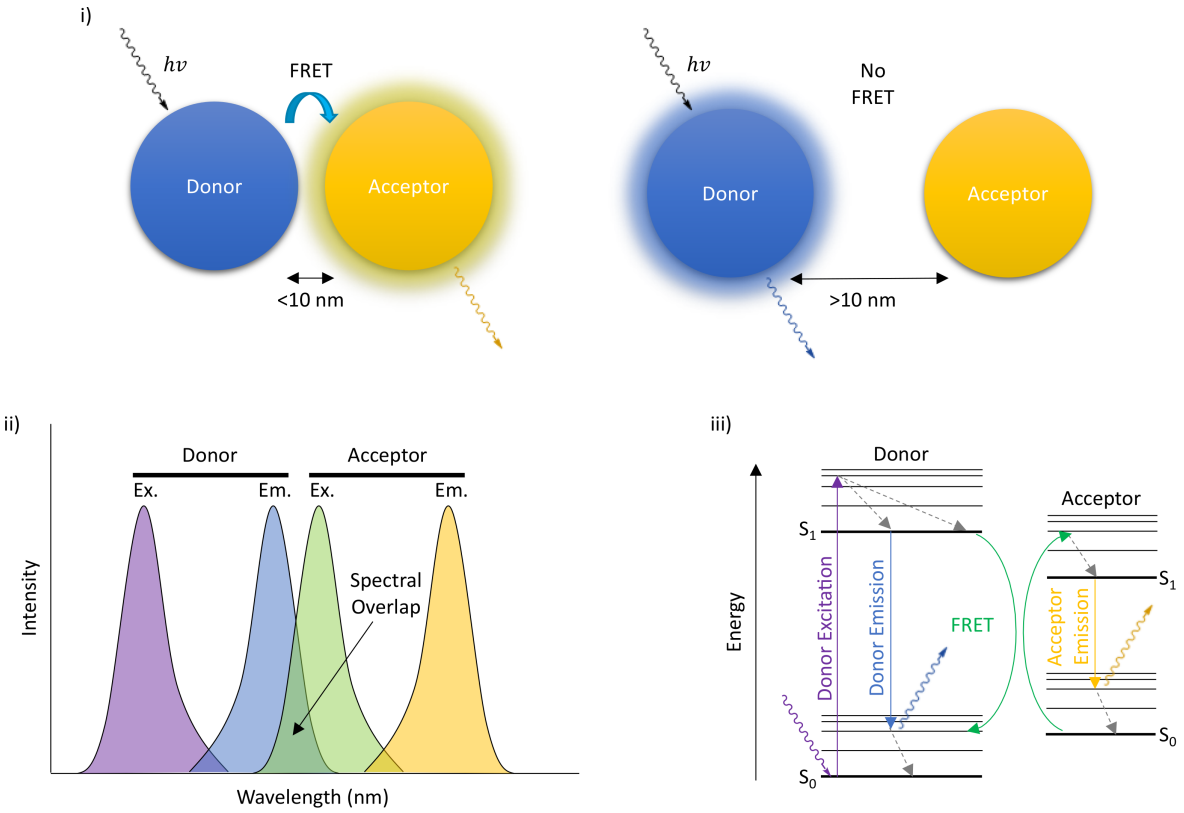


Figure 3.7: Fluorescence resonance energy transfer (FRET) - i) FRET is highly distance dependent and will only occur when the donor and acceptor are within 10 nm of each other. When FRET occurs the acceptor emission signal is observed, otherwise the donor emission signal is observed. ii) FRET only occurs when there is a spectral overlap between the emission of the donor and the excitation of the acceptor. iii) Jablonski diagram of FRET between a donor and an acceptor. The donor is excited from its ground state (purple) before undergoing non-radiative relaxation and fluorescence emission (blue) or FRET which excites the acceptor from its ground state (green) before final relaxation and fluorescence emission from the acceptor (yellow). Figure adapted from Skruzny *et al.* [218] and Chen *et al.* [219].

Three fluorescent proteins were required that could undergo a two-step FRET transfer, in order to mimic a three enzyme system. Three proteins, available from ATUM ProteinPaintbox,[220] which could undergo this transfer, were selected: FrostyCFP (Ex/Em 400/495); YetiYFP (Ex/Em 518/546); and FresnoRFP (Ex/Em 553/592). The cloning plasmids for these proteins had been amplified by Dr. James Allen (UCL), but the genes needed to be transferred into an expression plasmid with a 6His-tag sequence to allow for purification and provide the binding site for the cholesterol linker.

3.5.1 Cloning and Expression of Three Fluorescent Proteins

In order to clone the fluorescent proteins into pET28a, an expression vector, they needed to be PCR amplified from the cloning vector and have appropriate restriction sites introduced for subcloning. Primers were designed to amplify the DNA and introduce NdeI and XhoI restriction

sites (Table 6.1). Each of the fragments were amplified using a high-fidelity Q5 polymerase according to the manufacturer's protocol (NEB) before being purified via gel extraction (Methods 6.1.3.2 and 6.1.3.4). The three fragments and the expression vector, pET28a, were digested with NdeI and XhoI restriction enzymes before being purified via either gel extraction (fragments) or a PCR clean up kit (vector) all according to the manufacturer's protocols (NEB) (Methods 6.1.3.5). The inserts and linearised plasmid were ligated using T4 ligase before being transformed into a cloning *E. coli* strain, NEB10β (Methods 6.1.3.6 and 6.1.3.7). Colony PCR was performed to determine which colonies contained the correct inserts, and those which did were transformed into an expression *E. coli* strain, BL21(DE3) (Methods 6.1.3.2).

The three proteins were successfully expressed and purified using the newly incorporated His6-tags (Methods 6.1.4). The additional amino acids from adding the His6-tag and restriction sites could have altered the protein structure, so fluorescence scans were carried out to determine if these additions had made any changes to the reported excitation or emission maxima. The scans showed that these values had all shifted between 2 and 36 nm from the reported maxima, but the new values were still compatible with the three-protein FRET system (Table 3.1).

Table 3.1: Excitation and emission values of the cloned fluorescent proteins showing the reported values[220] and the experimental values

Protein	***Reported*** λ_{Ex}	***Reported*** λ_{Em}	***Experimental*** λ_{Ex}	***Experimental*** λ_{Em}
FrostyCFP	400	495	420	497
YetiYFP	518	546	482	537
FresnoRFP	553	592	526	585

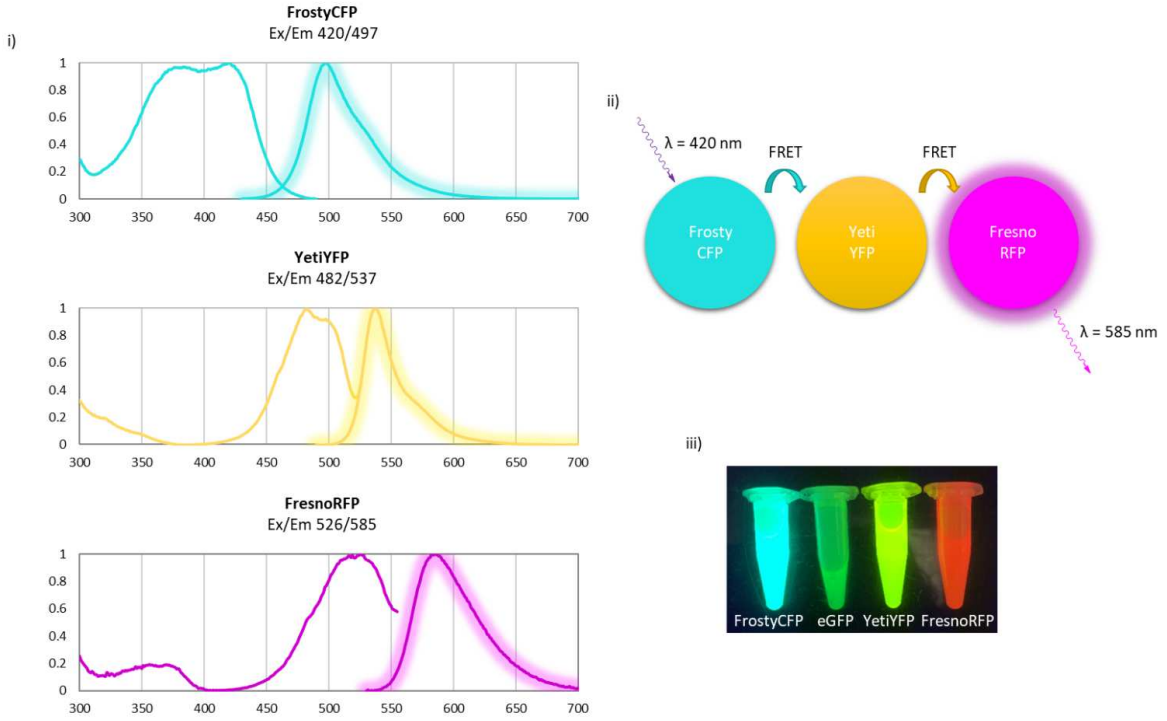


Figure 3.8: Three fluorescent protein FRET system - i) Experimental excitation and emission profiles of FrostyCFP, YetiYFP and FresnoRFP, λ_{max} values are displayed. ii) FRET system utilising the three proteins. iii) Image of the three proteins and eGFP excited under a UV lamp.

3.5.2 Investigation of Newly Expressed Fluorescent Proteins

The three proteins, FrostyCFP, YetiYFP and FresnoRFP, needed to be tested individually to confirm that each of them were able to be tethered to a vesicle surface. Unfortunately, under the same conditions as used for the eGFP experiments, a loss of fluorescence was observed when making up the samples, which had not been observed previously. In order to determine which reaction component was causing the problem, each component was added individually to the four fluorescent proteins (Figure 3.9). This assay showed two problems: one, that the cobalt(acac)$_2$ used is not very soluble in water or DMSO, so it needed to be used at a low concentration; and two, that the linker itself (**57**) was causing quenching of the fluorescence. The pH of the solutions containing **57** were measured and were shown to be acidic (pH 4). This low pH is likely due to the final step in the synthesis of the linker where TFA is used to remove the two *tert*-butyl groups and a **57**.TFA salt is formed. The theoretical pI of the fluorescent proteins ranges from 6.15-8.20 meaning that at pH 4 they would be highly positively charged which could affect their structure and solubility resulting in the loss of fluorescence. This quenching problem was overcome by adjusting the pH of the linker solution to pH 7 using sodium hydroxide prior to addition of the fluorescent proteins.

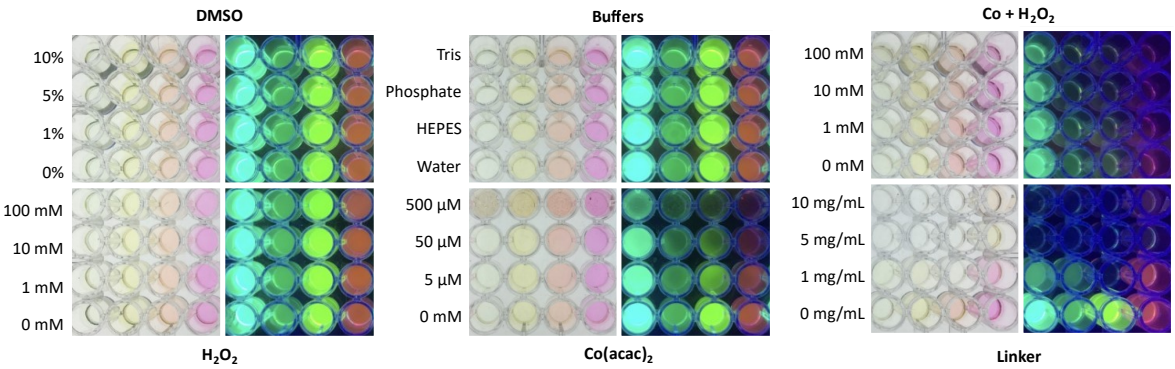


Figure 3.9: Cholesterol linker confocal sample reaction components - DMSO, H_2O_2, and the different buffers showed no affect on the fluorescence; $Co(acac)_2$ by itself and with H_2O_2 appears to show some reduction in fluorescence but this is only due to the low solubility and precipitation of the cobalt compound and the angle of the picture; the linker showed clear reduction in fluorescence under UV and visible light showing this was the component causing the problem.

Once the sample preparation had been modified, the fluorescent proteins could be analysed using confocal microscopy. However, the results from the initial experiments with vesicles could not be replicated under similar conditions even with eGFP. It was unclear why these results could not be replicated, but, it was possible that over time the linker (**57**) had been affected by the acidity of the TFA salt so was not behaving in the same manner as it had done previously. In order to try and resolve these issues the interactions between the components needed to be analysed individually to determine where the system was failing. A set of initial conditions were identified and 13 further experiments to test the variables were selected. The initial conditions were as follows: Linker (**57**, 0.1 mM), PBS buffer, DMSO (5%), $CoCl_2$ (0.5 mM), FrostyCFP (7 μM), OPOE (0.1%), H_2O_2 (10 mM). A 30 minute incubation at 37 °C was carried out after FrostyCFP addition; 1 hour incubation at rt was carried out after H_2O_2 addition; and a 1 hour incubation at 55 °C was carried out after GUV addition. The 14 experiments performed are summarised in Table 3.2 below:

Table 3.2: Confocal Conditions

Experiment Number	*Condition Changed*	*New Value*
1	-	-
2	Buffer	PBS + 0.5 M NaCl
3	[DMSO]	10%
4	[FrostyCFP]	1.75 μM
5	[FrostyCFP]	3.5 μM
6	[FrostyCFP]	14 μM
7	$[H_2O_2]$	0 mM
8	$[H_2O_2]$	1 mM
9	$[H_2O_2]$	5 mM
10	$[H_2O_2]$	20 mM
11	[OPOE]	0.5%
12	[OPOE]	1%
13	[OPOE]	5%
14	GUV Incubation	1 hr, 37 °C

Unfortunately none of these experiments proved successful and it was still not possible to determine which components were causing problems it was not possible to analyse the steps prior to using confocal microscopy.

3.5.3 Isothermal Titration Calorimetry

It was not possible to determine whether the proteins were binding to the cholesterol linker through traditional methods: cobalt is paramagnetic so interferes with NMR spectroscopic characterisations; and the ionic bond between the linker and the protein is too weak to withstand ionising mass spectrometry. In order to determine whether the fluorescent proteins were binding to the linker a different method was required. Isothermal titration calorimetry (ITC) measures the interaction between macromolecules and small molecules by measuring millionth of a degree temperature changes that result from hydrophobic, ionic, or covalent interactions. A small molecule is titrated against a macromolecule to determine whether the molecules bind, and the data from the experiment can be used to determine the stoichiometry, K_D and ΔH of the interaction (Figure 3.10).

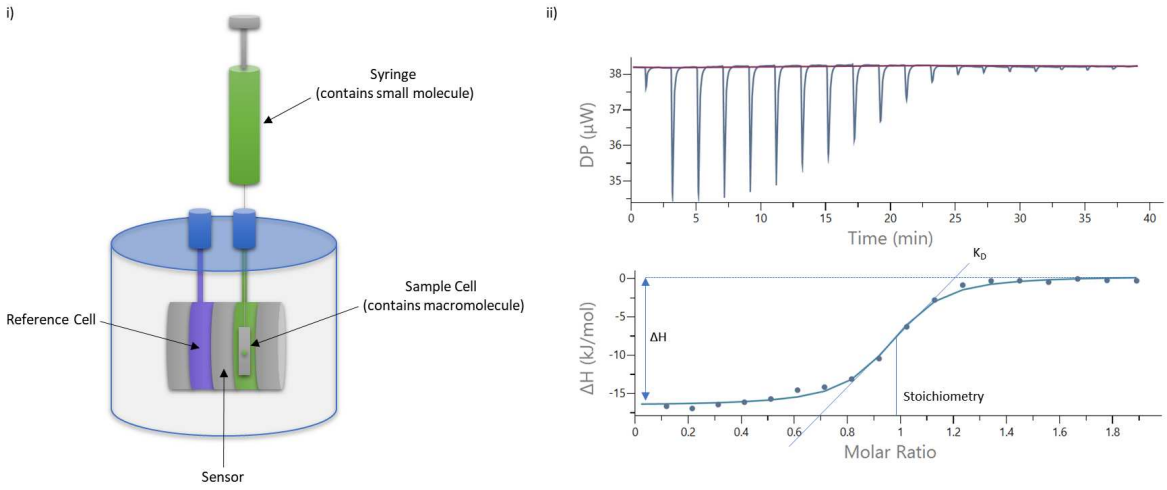


Figure 3.10: Isothermal titration calorimetry (ITC) - i) a typical isothermal titration calorimeter, the macromolecule is loaded into the sample cell and the small molecule is titrated against it through a series of small injections. The power feedback/temperature difference between the sample and reference cell is used to calculate the molecular binding affinities; ii) a typical ITC trace showing titration of calcium chloride (100 μM) against EDTA (10 μM) showing how the stoichiometry, K_D and ΔH can be calculated.

The linker (**57**), as the small molecule, and a fluorescent protein, as the macromolecule, should bind in a 1:1 molar ratio (one NTA group to one 6His-tag) as with the example given in Figure 3.10, however, this technique relies more upon the K_D of the binding interaction. The K_D of Ni(II) and a 6His-tag has been reported as 14 ± 4 μM so this was used to calculate the initial concentrations for the linker and protein.[221] The suggested concentrations from the ITC machine manufacturer (Malvern) for this K_D are 50 μM for the macromolecule and 500 μM for the small molecule. A linker conjugate was prepared with Co(II) following the modified procedure from section 3.5.2 and the fluorescent proteins, eGFP or FresnoRFP, were prepared in the same buffer. The samples were prepared in the same buffers to ensure that the temperature changes associated with dilution do not affect the binding results.

The first experiment (Figure 3.11 (i)) seemed to show a jump after 35 minutes indicating that full binding had occurred at a 1.5 molar ratio, however, the data could not be fitted to a curve for a K_D value so this was repeated. The second experiment (Figure 3.11 (ii)) did not match the first data set, showing that there may have been a loading or injection error that resulted in a null-response. However, despite not showing full binding in this experiment, the data indicated that binding had occurred with a K_D of 17.2 ± 6.5 μM, comparable to the literature values for Ni(II) and a 6His-tag.[221] Subsequent experiments were performed to try and replicate this data and adjust the concentrations to show full binding between the linker and fluorescent protein. Unfortunately, these experiments were not able to confirm binding, repetition using the same concentrations of linker (**57**, 500 μM) and RFP (50 μM) gave a K_D of 427 ± 7340 μM (Figure 3.11 (iii)), and a

lower RFP concentration of 30 μM gave a K_D of 366 ± 7130 μM (Figure 3.11 (iv)) with the huge errors also confirming no reliable binding had occurred.

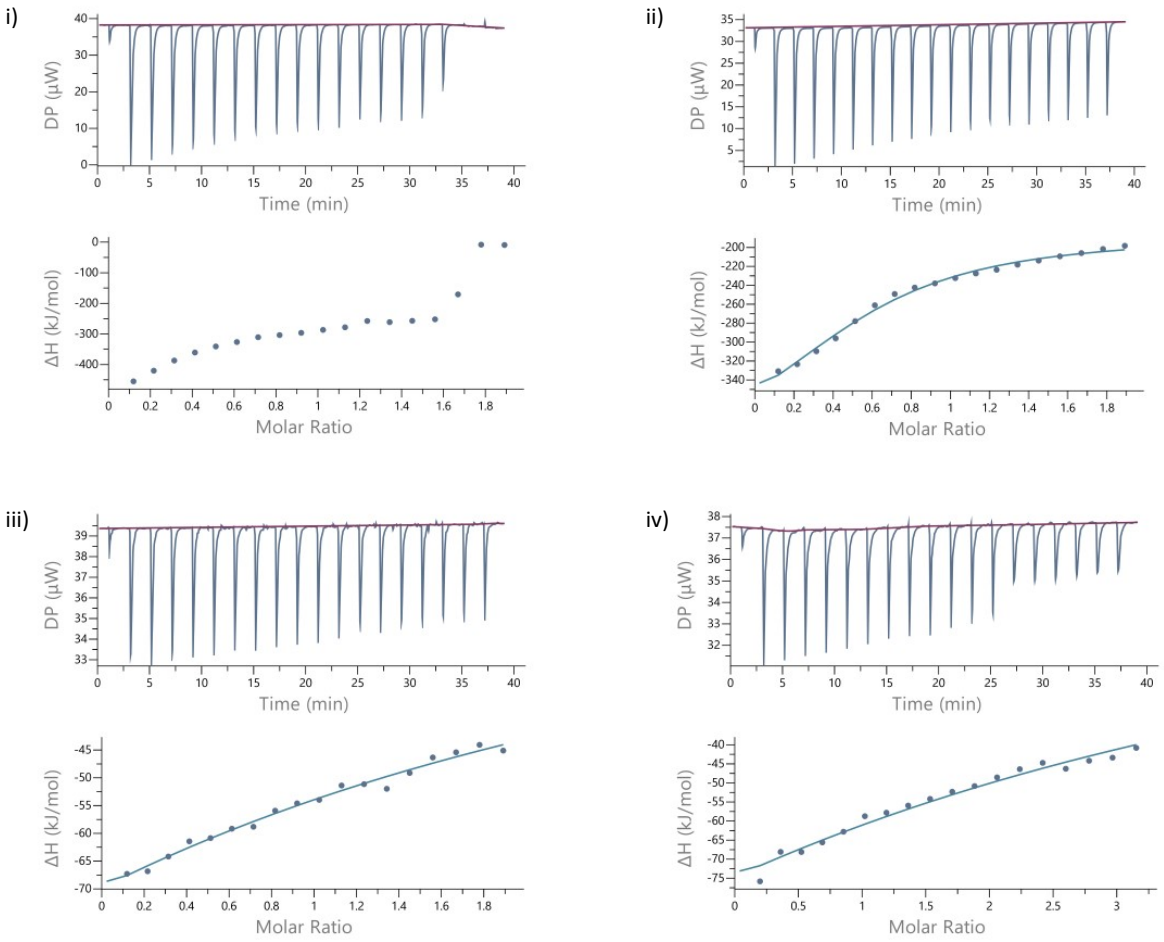


Figure 3.11: ITC Binding Experiments - i) **57** 500 μM and eGFP 50 μM; ii) **57** 500 μM and eGFP 50 μM; iii) **57** 500 μM and FresnoRFP 50 μM; iv) **57** 500 μM and FresnoRFP 30 μM

This final experiment (iv) clearly did not show binding between the linker and protein, however, it had a jump corresponding to the injection where the linker concentration reached 50 nM. The linker, **57**, was designed to be amphiphilic in order to sit within a lipid membrane and protrude into the aqueous solution to bind to a protein, however, this could be causing the linker to aggregate into pseudo-micelles. The critical micelle concentration (CMC) for cholesterol is 25 to 40 nM,[215] therefore, it is possible that this jump corresponds to the linker reaching its CMC with a value of approximately 50 nM (Figure 3.12). This micelle formation would result in crowding of the NTA groups and could prevent either: all of the protein from binding, as the groups are not accessible; or would only allow a small number of proteins to bind, as steric clashes between them would prevent a 1:1 binding ratio. This matches with the initial confocal experiments which showed aggregation of the linker conjugate, however, the small injection volumes, and low initial concentrations used in the ITC experiments should have prevented the linker, **57**, reaching its CMC prior to binding to the protein. Due to the ionic and labile nature of the cobalt(II) complex

it is also likely that the dissociation rate (K_{off}) would be greater than the association rate (K_{on}). This would have resulted in a large concentration of dissociated linker which would be able to form micelles, even if it had been conjugated to a protein previously.

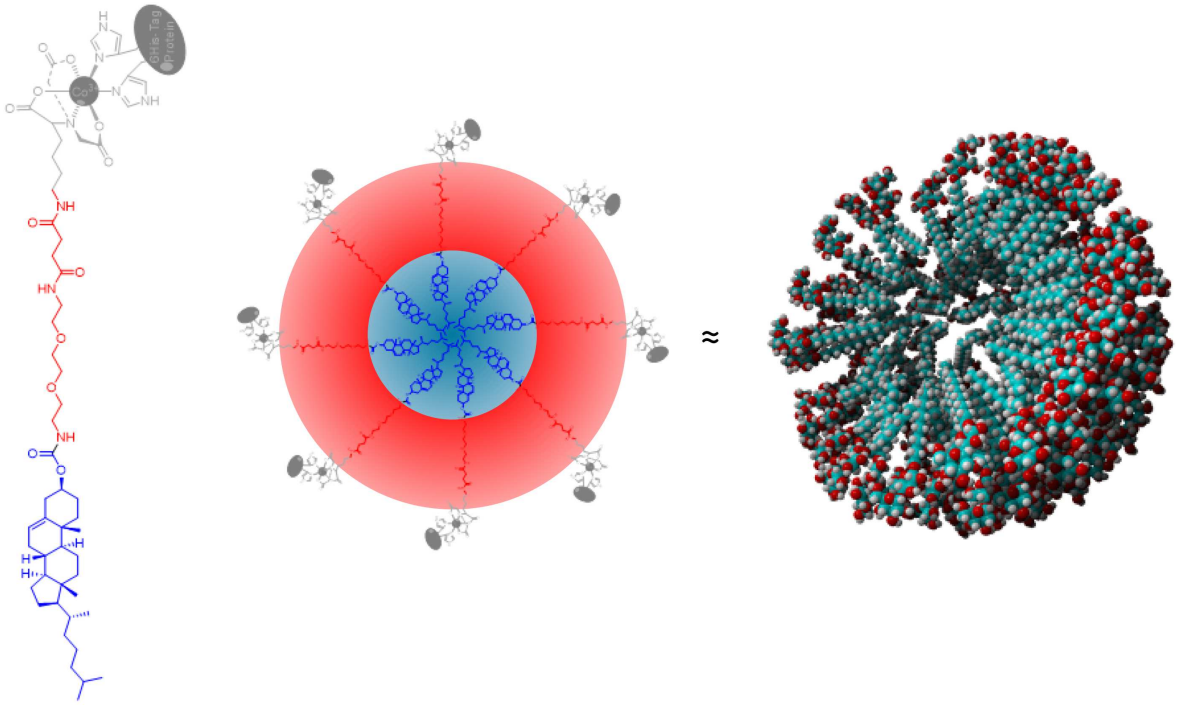


Figure 3.12: Linker Micelle Formation - The hydrophobic cholesterol (blue) would likely be protected from the aqueous solution by sitting in the centre of the micelle surrounded by the hydrophilic section of the linker (red). This structure mimics a true lipid micelle (right).[222]

3.6 Conclusions

This work was able to show a novel design and successful synthesis of a cholesterol-based linker, **57**, for enzyme immobilisation. The synthesis of the linker was refined and streamlined to provide the best route to the final compound, combining traditional chemical synthesis and SPPS techniques. Initial testing using confocal imagery looked promising, indicating successful binding and immobilisation of eGFP to POPC GUVs, after addition of a mild detergent. The requirement for OPOE was an indication that the linker may have been forming pseudo-micelles, which appeared as aggregates. The linker was designed to be amphiphilic with the hydrophobic cholesterol in the membrane and the hydrophilic spacer and chelating group to enable protein attachment, however, this likely resulted in a number of the problems discussed above.

Three fluorescent proteins were successfully cloned and purified to act as a FRET system to investigate the behaviour of the linker. The first problem faced with these proteins was the quenching of fluorescence when making up the linker conjugates. This problem was successfully

identified as the acidic pH carried over from the TFA deprotection step in the linker synthesis, and was overcome by a neutralisation step being added to the sample preparation. However, despite trying a multitude of different conditions for the sample preparation the results from the initial confocal images could not be replicated with the newly expressed fluorescent proteins.

The biggest problem faced with the cholesterol-based linker, **57**, was the inability to determine where the issues with the system were due to the number of components required which could not be independently tested. It was hoped that the ITC experiments would shine a light on any problems by determining if the protein was actually binding to the linker, as this could not be performed by traditional chemical methods. Unfortunately this analysis was inconclusive, showing potential binding, but also possibly indicating that the linker was aggregating in the reaction chamber.

Once all these analytical experiments had been performed it was still not possible to determine the key failure points, only that the linker was not performing as designed. It is possible that continuing the strategy, but using another approach, such as a supported bilayer, or using a different vesicle lipid composition, may have resulted in successful outcomes due to the different conditions that would have been used.

4 | DNA-Based Linker Synthesis and Testing

This chapter describes approaches towards the development of a second novel enzyme immobilisation platform using a DNA-based NTA linker to immobilise His-tagged proteins on a lipid membrane surface.

4.1 Introduction

DNA, as discussed in section 1.3.3, has been shown to be a robust and reliable building material for everything from plates and cages to nanoreactors and nanovaults.[106,108] DNA origami, the technique these structures rely upon, uses the specificity of DNA base pairing to provide precise spatial control over the DNA building blocks, which allows complex structures to be developed through rational design.[105] This approach can also be used to attach biomolecules, such as enzymes, to specific locations on the DNA surface. Biomolecules can be attached to designed DNA structures through protruding 'capture strands' which can be positioned to control the number, location, orientation, and position of proteins in a network.[223] Enzymes can be immobilised onto these strands through a number of immobilisation techniques such as biotin–streptavidin, Ni-NTA–6His-tag, aptamer–protein, and covalent cross-linking, each requiring different levels of modification of the enzyme and the capture strand.[224] This technique of immobilising an enzyme, or a whole cascade, onto a DNA structure has been achieved by multiple groups ranging from: static immobilisation of two enzymes on a DNA support,[159,225,226] to dynamic three-dimensional systems.[128,129,161,227] These examples have demonstrated the benefits of using DNA for immobilisation and shown how the spatial control it offers can maximise pathway efficiency.

On a smaller scale, the reliable and controllable DNA hybridisation can be used to design simple systems which rely on the formation of a single duplex. One of the problems faced with the cholesterol linker in the previous chapter, was not being able to determine where the system was falling down. Replacing part of the linker with DNA would allow the linker to be more easily visualised, through gel electrophoresis and ethidium bromide staining, enabling analysis of the individual linker components, which was not possible previously. Using DNA hybridisation also allows for two halves of the linker to be prepared separately and then combined once they have undergone any modifications. Another feature of DNA duplexes is that they can be displaced by another strand, if they have an overhang, resulting in a system which can be modified during use, making it more amenable for use in an industrial setting.[228] The idea behind this immobilisation

platform built upon the ideas from chapter 3, using a linker to tether an enzymatic cascade to a lipid surface, to develop a more dynamic system by using DNA hybridisation.

4.2 DNA Linker Design

Developing a second novel immobilisation technique to adapt and improve upon the cholesterol linker required a modified design to overcome some of the problems faced previously. The design of the DNA-based linker followed the same general structure as the cholesterol linker (Section 3.2) with a hydrophobic section, a hydrophilic spacer/linking section, and a NTA group to coordinate to a His-tagged protein. A cholesterol moiety was again selected as the hydrophobic portion as this has been robustly shown to embed into membranes.[229] An NTA group derived from lysine was again selected as this can coordinate to a His-tagged protein through an oxidised cobalt ion, which is an established method for non-labile ionic protein binding.[147] The connection between these two ends of the linker, however, is very different from the original design. One of the problems with the cholesterol linker was the difficultly in analysing the binding of various components to one another, mostly due to the limitations of the analytical methods available. DNA, on the other hand, is highly stable, and easy to manipulate and visualise through gel electrophoresis and was selected as the link between the cholesterol and the NTA. This also allowed the two halves to be prepared separately, prior to hybridisation of two single strands, making the synthesis and the analysis of this linker more straightforward than previously with the cholesterol linker.

The DNA-based linker (Figure 4.1) was divided into two halves, a cholesterol-modified DNA strand (**72**), and an NTA-modified DNA strand (**73**). Cholesterol modified DNA strands are commercially available (IDT), with a TEG spacer between the DNA and the cholesterol, so this half of the linker could be easily obtained without requiring further modifications. In order to attach a DNA strand to a protein, a modification is needed to provide a group to attach to a protein, such as an NTA group. DNA strands can be easily modified with a thiol group, also available commercially (IDT), to provide a reactive centre from which a small molecule can be attached. One molecule which has been used to tether proteins to DNA origami structures is maleimido-C3-NTA.[230] This molecule consists of a maleimide group, which is specifically and highly reactive with thiols, a small hydrocarbon spacer, and a lysine-NTA moiety. This molecule can also be synthesised in a small number of steps before being attached to a commercial thio-modified DNA strand. These halves could then be combined through DNA hybridisation between the two strands to give the final linker (**74**).

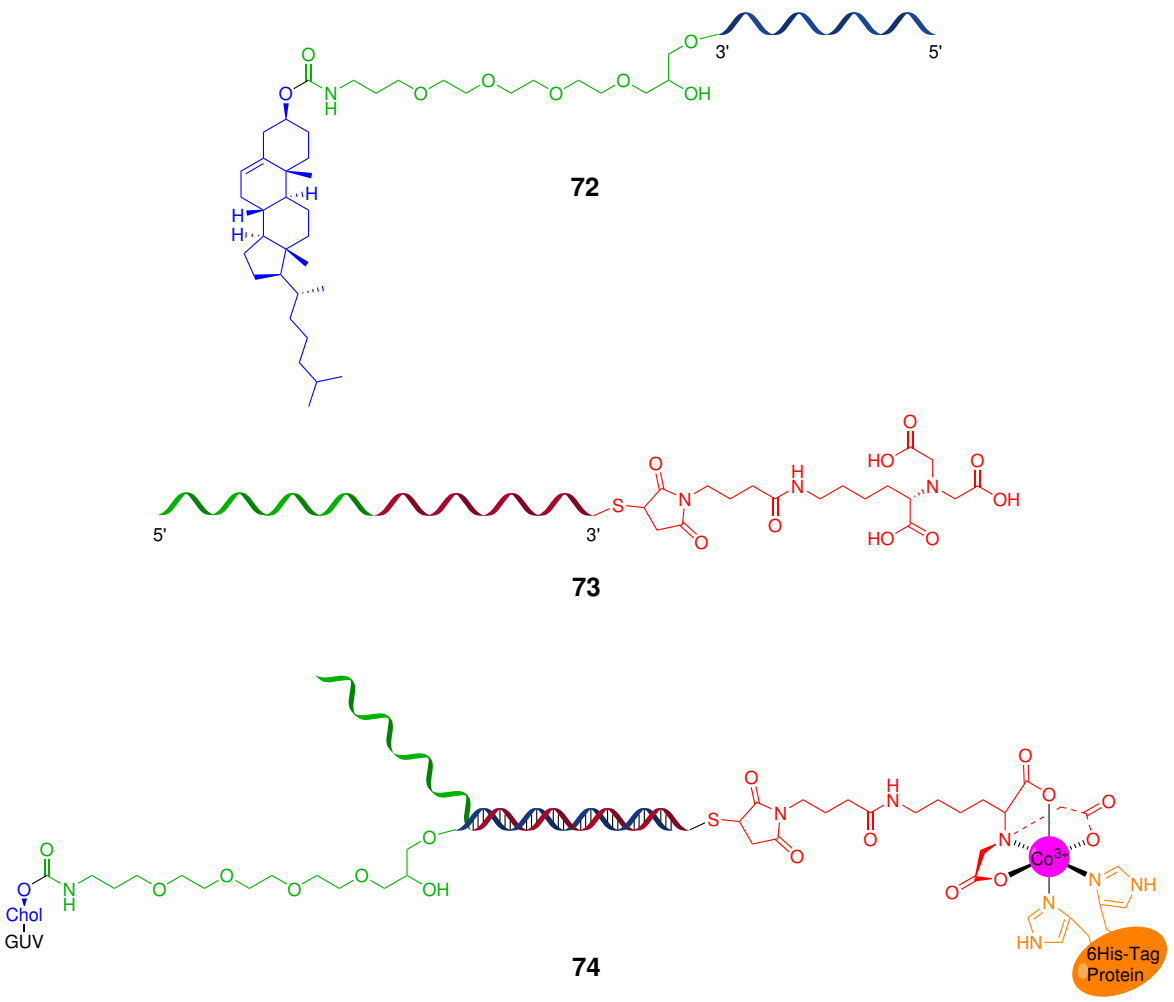


Figure 4.1: DNA linker design - The DNA linker (**74**) is formed of a cholesterol-modified strand (**72**) and a thio-NTA-modified strand (**73**) and is divided into four sections: a hydrophobic cholesterol section (blue); a hydrophilic TEG spacer (green); two complementary DNA strands to attach the cholesterol-TEG modified strand to the thio-NTA modified strand with an overhang to accommodate toehold mediated strand displacement; and maleimido-C3-NTA (red) which can chelate a cobalt ion and coordinate to a His-tagged protein.

This design fulfilled the same criteria as the cholesterol linker, however, due to the highly specific binding of DNA, it was adapted further to include a method for detaching the two halves (Figure 4.2). Enzymes can be degraded over time so having a way of removing and replacing them would be a useful feature of a tethering system. Hybridised DNA structures can be separated from one another by heating the strands to break the hydrogen bonding, however, this would denature any proteins and disrupt any lipid structures. Toe-hold mediated strand displacement can be used to remove the hybridisation between two strands by introducing a third strand to displace one of the hybridised ones (Figure 4.2).[228] In this case, the protein-bound strand would hybridise, with an overhang, to the cholesterol strand, which would be partly embedded in a vesicle membrane. The overhang, un-hybridised part, would sit freely until the protein or enzyme became degraded and then would be available to hybridise with a displacement strand. The displacement strand would gradually displace the cholesterol strand from the protein-bound strand. The displacement

strand and protein-bound strand would fully hybridise and then could be washed away leaving the embedded cholesterol strand free to bind to a fresh protein-bound strand.

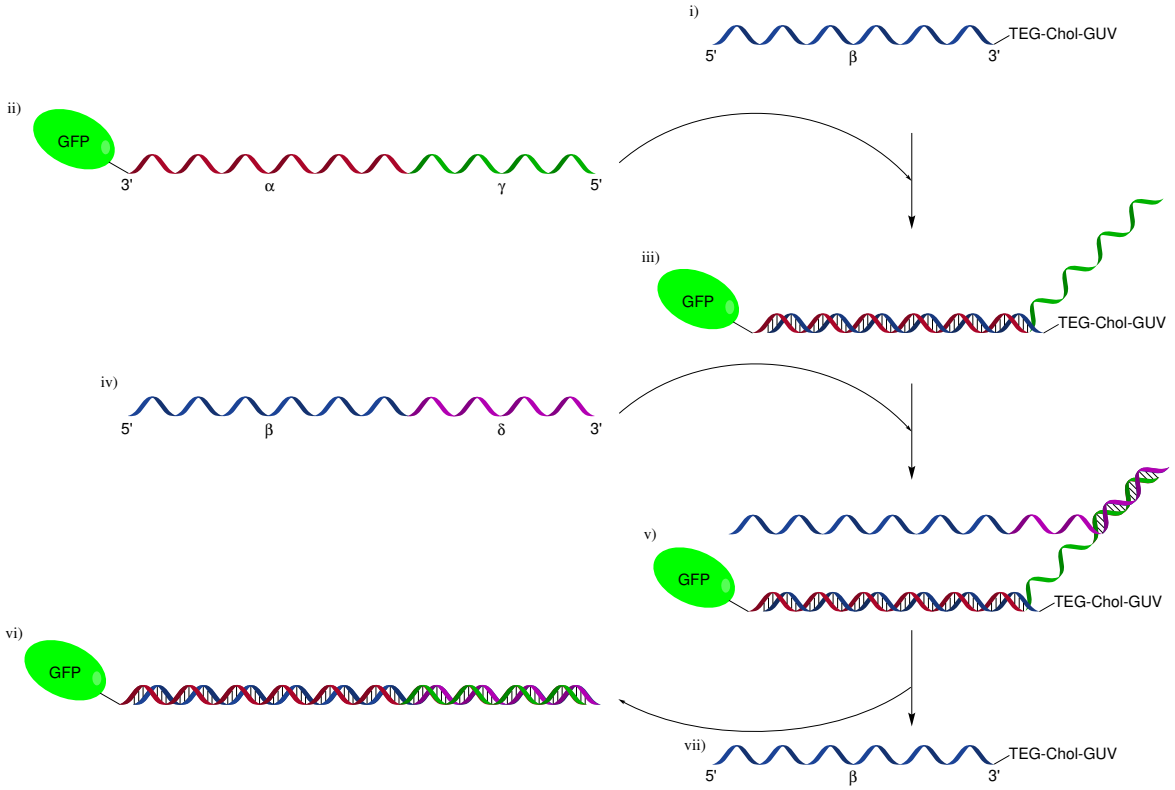


Figure 4.2: Toehold mediated strand displacement - i) The cholesterol modified strand (blue) is tethered to a vesicle (GUV); ii) The thio-NTA modified strand is coordinated to His-tagged GFP and has a sequence complementary to the cholesterol strand (red) and an overhang of several nucleotides (green); iii) The thio-NTA strand and cholesterol strand hybridise (red-blue) to tether the GFP to the vesicle surface; iv) The displacement strand (blue/pink) has a complementary sequence to the full length of the thio-NTA strand; v) The displacement strand hybridises to the overhang of the thio-NTA strand (green-pink), to gain a toehold; vi) The displacement strand continues to hybridise to the thio-NTA strand (pink-green/blue-red) gradually displacing the cholesterol strand and releasing the GFP from the vesicle surface; vii) The cholesterol strand is then ready to hybridise with another thio-NTA strand.

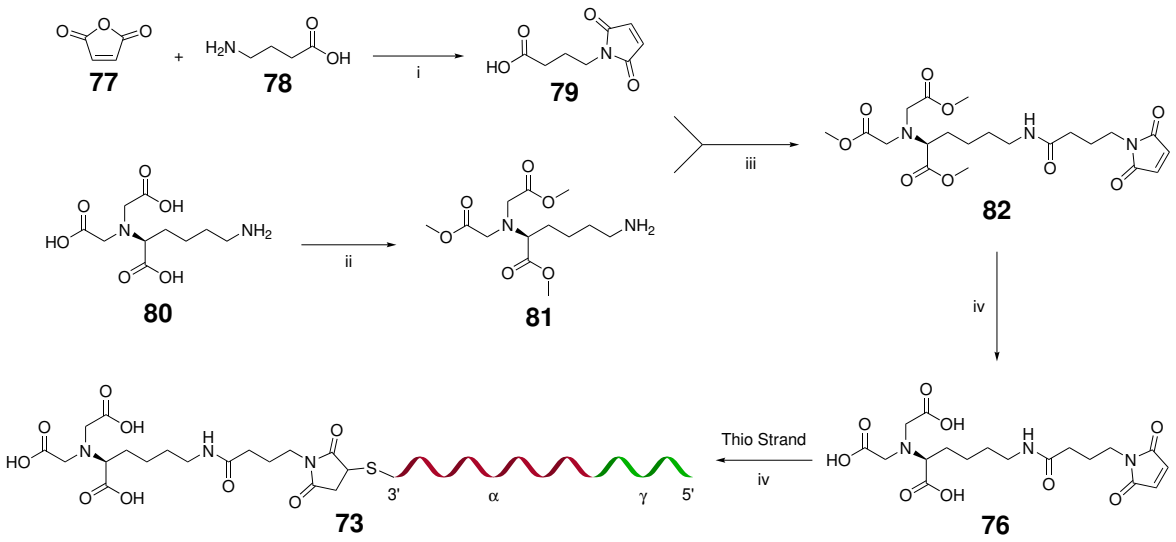
The strands in this system did not require a particular nucleotide sequence, only the correct complementarity to form the desired hybridisation structures. The thio-NTA strand (**73**) needed to be partly complementary (α) to the cholesterol strand (**72**, β) and fully complementary (α-γ) to the displacement strand (**75**, β-δ). For this work, sequences published by Zhang and Winfree [228] were selected as they have been shown to undergo toehold displacement and their kinetics have been analysed (Table 4.1).

Table 4.1: DNA linker sequences - The three strands required for the DNA linker and toehold strand displacement shown in Figure 4.2. The colours and labels also correspond to Figure 4.2.

Strand	***Sequence (5'→3')***
Cholesterol Strand (**72**, β)	CGTAGGGTATTGAATGAGGG-*TEG-Chol*
Thio-NTA Strand (**73**, α/γ)	TCTCCATGTCACTTCCCCTCATTCAATACCCTACG-*Thio-NTA*
Displacement Strand (**75**, β/δ)	CGTAGGGTATTGAATGAGGGGAAGTGACATGGAGA

4.3 DNA Linker Synthesis

The synthesis of the DNA linker first involved the synthesis of maleimido-C3-NTA (**76**), which would act as a bridge between a DNA strand and a His-tagged protein (Scheme 4.2). This involved a four-step converging synthesis starting from maleic anhydride (**77**), 4-aminobutyric acid (**78**) and lysine-NTA (**80**). A direct coupling method between 4-maleimidobutyric (**79**) and lysine-NTA (**80**) was considered but the number of carboxylic acid groups on the lysine-NTA could have led to a lack of selectivity in the reaction, so the synthetic route shown in Scheme 4.2, with the NTA acid groups protected, was selected. The first step was to generate 4-maleimidobutyric acid (**79**) in a nucleophilic acyl substitution reaction between 4-aminobutyric acid (**78**) and maleic anhydride (**77**).[230] This reaction was performed in acetic acid to facilitate the maleimide ring opening and water loss, and gave the product in a 67% yield after 6 hours without requiring any purification steps. The other arm of the converging synthesis involved the protection of the acid groups on lysine-NTA (**80**) by first converting the carboxylic acids to acyl chlorides using thionyl chloride ($SOCl_2$).[231] This gave a better leaving group for the esterification reaction with methanol which gave tri-methyl protected lysine-NTA (**81**) in 100% yield after 48 hours and again no purification was required.[231] Compound **81** was then reacted with 4-maleimidobutyric acid (**79**) in a standard amide coupling reaction with HATU and DIPEA in DMF to give **82** as a hydrate, observed as an $[M+H_3O]^+$ ion by mass spectrometry, in a 15% yield after purification by column chromatography.[230] A fully assigned carbon spectrum could not be attained for this compound as it was too small a sample, however, due to the mass spectrometry and proton NMR being the expected values this compound was carried into the next step of the synthesis. The carboxylic acids of **82** were then deprotected with lithium hydroxide and the lithium salt was protonated using a proton exchange resin (Dowex 50Wx8) to give maleimido-C3-NTA (**76**) as a hydrate, observed as an $[M+H_3O]^+$ ion by mass spectrometry, after 48 hours in an 82% yield after lyophilisation.[230]



Scheme 4.1: Synthesis of the thio-NTA-modified DNA strand (**73**) - i) AcOH, H_2O, 6 h, rt, 67%; ii) $SOCl_2$, MeOH, 48 h, 55 °C, 100%; iii) HATU, DIPEA, DMF/CH_2Cl_2, 2 h, rt, 15%; iv) LiOH, H_2O:MeOH (1:1), 48 h, rt, 82%; v) HEPES buffer (10 mM, 50 mM NaCl, pH 7.3), 18 h, rt.

The final stage in forming the thio-NTA strand (**73**) was to attach the maleimido-C3-NTA (**76**) to the thio-modified strand. The strand was produced commercially as a protected disulphide (DNA-S-SR) so it needed to be reduced before it could be used. The strand was reduced with a 100x excess of TCEP for 2 hours at room temperature and was then used directly. The free sulfur (DNA-SH) was reacted with the maleimide of **76**, according to the manufacturer's protocol (IDT) reacting in HEPES buffer for 18 hours at room temperature, and purified using a 3 kDa molecular weight cut off (MWCO) spin filter to give the thio-NTA modified DNA strand, **73**, ready to be attached to a protein or combined with the cholesterol strand (**72**) to form the full linker (**74**). The success of this reaction was determined through agarose gel electrophoresis when the strand was used to form a duplex and is discussed below.

4.4 DNA Linker Testing

4.4.1 DNA Duplex Formation

The first step in confirming formation of the complete DNA linker (**74**) was to check that the cholesterol and thio-NTA strands could still form a duplex after the modifications. This analysis was performed by gel electrophoresis as the strands can be easily visualised, with ethidium bromide staining, under ultra-violet light. Due to the strands being both under 40 nucleotides, the first agarose gel used 3% agarose as this slows the movement of all DNA and makes smaller strands easier to analyse. This first gel (Figure 4.3 (i)) showed that ethidium bromide, the standard DNA stain, was not ideal as it did not stain the single stranded DNA (ssDNA) of the thio-NTA strand. It

did stain the cholesterol strand, however, and showed an upward gel shift when the two strands were combined which indicated duplex formation.

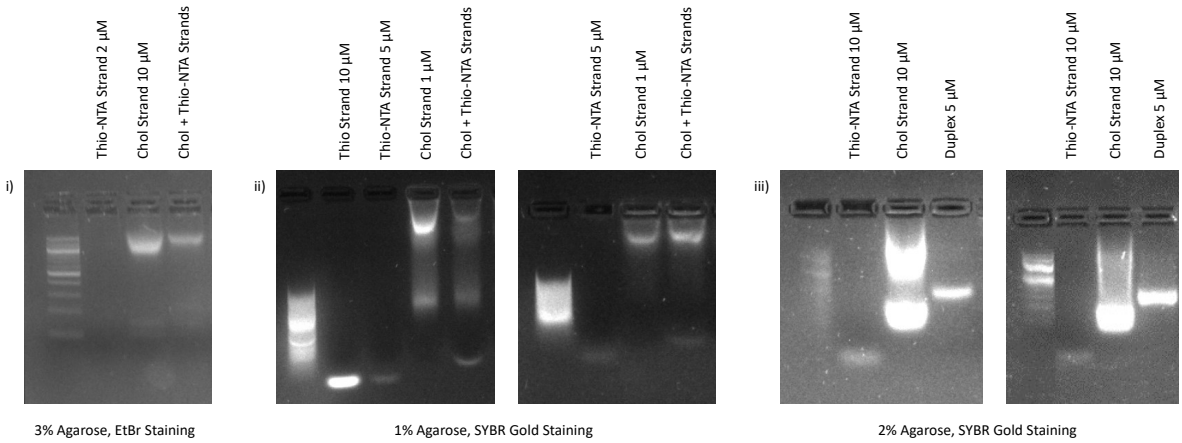


Figure 4.3: DNA duplex formation - i) 3% agarose, EtBr Staining, duplex formation shown but shift unclear; ii) 1% agarose, SYBR Gold staining, ssDNA visible but shifts very slight; iii) 2% agarose, SYBR Gold staining, heating to 95 °C and cooling at 6 °C/min, ssDNA visible and shifts clearly show duplex formation. (Note - the two bands observed for the cholesterol strand in the left gel are due to cholesterol aggregation in an older sample generating a second band with a higher molecular weight)

The second set of gels (Figure 4.3 (ii)) were 1% agarose gels to allow the DNA to migrate further down the gel as the cholesterol strand stayed at the top for the first gel. SYBR Gold was selected for staining as it is designed to stain both ssDNA and dsDNA. The modification of the thio strand was seen here as a small upshift from the first to the second lane indicating the addition of the maleimido-C3-NTA. Both of these gels showed the upshift of the thio-NTA and cholesterol strands when they were added together which indicated binding. The two gels also showed how differently the cholesterol strand behaved depending on the freshness of the solutions, the right hand gel showed a cleaner band with a freshly made solution, over the cholesterol aggregation and double band seen in the left hand gel. These three gels all seemed to indicate binding but a clear band of a duplex was not appearing. This may be due to the fact that the duplex samples were prepared at room temperature, and not using an exact 1:1 ratio.

The duplex samples for the final gels were prepared according to Arnott *et al.*[232], heating to 95 °C and cooling slowly to room temperature which would break any non-specific interactions and ensure only a fully hybridised duplex remained. These gels (Figure 4.3 (iii)) both showed this clear duplex band with no excess of either single strand remaining, even with the older cholesterol sample in the left gel, clearly confirming that the duplex between the two strands could form after their respective modifications. The hydrophobic nature of cholesterol resulted in the streaking seen with this strand across all the gels, however, when the duplex formed it resolved to a single sharp band. This is due to the overhang for toehold mediated strand displacement shielding the

cholesterol upon binding, meaning that the cholesterol does not interact with the agarose matrix and prevents streaking.[233]

4.4.2 Fluorescent Protein Binding

Once the duplex formation was observed, the next step was to confirm whether the fully formed DNA linker (**74**) could coordinate to a His-tagged protein. The protein should bind in a 1:1 ratio with the linker, with two histidines bound to each NTA-cobalt group, however, binding was likely to start occurring at lower concentrations and would be affected by the strength of the binding, especially under electrophoretic conditions. The best way to determine a binding point is through a titration, altering the concentration of the protein against a constant concentration of the DNA linker. His-tagged FresnoRFP was arbitrarily selected for these experiments. A concentration of 1.25 μM was selected for the duplex-Co(II) and concentrations from 0-1.75 μM were selected for the RFP. The first gel was run using 2% agarose, to match the conditions for the duplex formation, using hydrogen peroxide to oxidise the cobalt(II) to cobalt(III) (Figure 4.4 (i)). This first gel showed possible binding at the two highest concentrations, but this was unclear as the RFP appeared to remain in the well. However, this could have be due to an excess of RFP to linker as the positively charged RFP would have behaved very differently when attached to negatively charged DNA. This also showed that agarose gels may not have been the best way to visualise this binding, so a shift was made to PAGE gels in the hope that any gel shifts would be easier to identify and analyse.

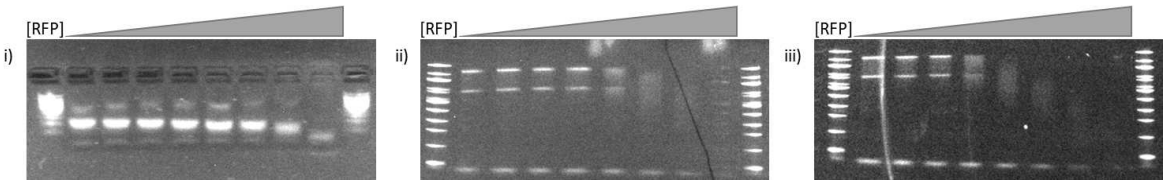


Figure 4.4: DNA linker RFP titrations - All gels show [RFP] = 0, 0.035, 0.086, 0.175, 0.35, 0.525, 0.875, 1.75 μM and [Linker-Co(II)] = 1.25 μM. i) 2% agarose, [H_2O_2] = 20 mM; ii) 10% Native PAGE, no H_2O_2, binding from 0.35 μM RFP; iii) 10% Native PAGE, [H_2O_2] = 20 mM, binding from 0.175 μM RFP.

In order to determine whether oxidation of the cobalt(II) was affecting the binding, two titrations were performed, one with, and one without H_2O_2 oxidation. The unoxidised titration (Figure 4.4 (ii)) showed a gel shift from 0.35 μM RFP with only a very faint band present at 1.75 μM RFP. The oxidised titration (Figure 4.4 (iii)) also showed a gel shift and only a very faint band present at 1.75 μM RFP but started showing a gel shift from 0.175 μM indicating that the oxidation had improved binding between the linker and RFP.

4.4.3 SUV Binding

The final step in analysing the DNA linker (**74**), by gel electrophoresis, was to determine if the linker could bind to a vesicle. GUVs are too large for gel electrophoresis so SUVs, also formed with POPC, were selected as an alternative. SUVs are still very large so will sit within the wells during gel electrophoresis so a shift of the DNA band towards the well will be indicative of the linker binding to the SUVs. The maximum number of linkers that can bind to a vesicle can be calculated using the size and concentration of SUVs, their surface area, and the maximum packing density for DNA-cholesterol conjugates.[234] The diameter of the SUVs made for this experiment was determined using dynamic light scattering (DLS) to be 209.2 nm (Figure 4.5).

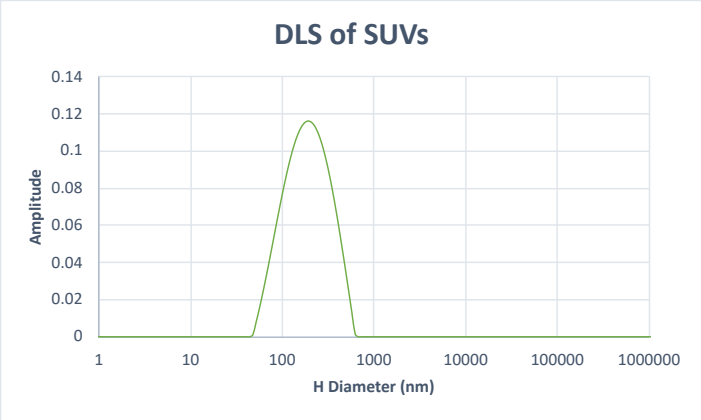


Figure 4.5: DLS Analysis of SUVs - The SUVs formed were analysed by DLS which determines the size distribution of the population of vesicles. The average size of these vesicles was 209.2 nm.

Assuming a perfect sphere, the external surface area of the SUVs (SA_{SUV}) was calculated using the following equation:

$$SA_{SUV} = 4\pi r^2,\ r = 104.6\,\text{nm}$$

The concentration of SUVs was then determined using the following equation:

$$[SUV] = \frac{A_l[lipid]}{2 \times SA_{SUV}}$$

Where A_l = lipid head group area, for POPC $A_l = 6.7 \times 10^{-19}$ m^2; the factor 1/2 accounts for the two leaflets in the lipid bilayer.[234] With a POPC concentration of 1.97 mM this equation gave an SUV concentration of 4.89 nM. The maximum loading was then calculated using the surface area and the surface density (SD):

$$Maximum\,loading = SA_{SUV} \times SD_{SUV}$$

For POPC/POPE vesicles in PBS with 20 nucleotide double stranded DNA cholesterol modified strands, the surface density is 0.047 ± 0.005 molecules/nm^2.[232] This value would likely be lower than this for pure POPC vesicles as they have a slightly smaller head group, however, it provided a good estimate. This calculation gave a maximum loading of 6462 molecules/vesicle, which was used to calculate the starting conditions for the SUV titrations. Given a duplex concentration of 2.5 μM in 20 μL being 0.05 nmol, this equated to 7.74 fmol SUVs. Using the final volume of 20 μL, this equated to an SUV concentration of 0.387 nM which was used as the midpoint for the titrations.

The first titration (Figure 4.6 (i)) was performed with a 30 minute room temperature incubation, prior to running the gel, using the calculated concentrations of SUVs, 0-5 nM. This titration showed the duplex migrating to the middle of the gel with very faint binding at the highest two SUV concentrations. This was promising and prompted an increase in SUV concentration, as the calculated values represented the maximum optimum loading value, so more vesicles may be required. The incubation temperature was also increased to 37 °C as this was likely to improve the probability of binding occurring. These results (Figure 4.6 (ii)) improved upon the first titration showing an increase in binding from 0.2 nM SUVs, however, full binding was not observed and a band for the unbound duplex was visible across the gel. The concentrations were again increased to maximise the potential for full binding and the incubation temperature was increased to 55 °C. This temperature has been shown to be required to break cholesterol-cholesterol interactions which could be preventing binding from occurring.[233] This final titration (Figure 4.6 (iii)) proved successful showing full binding at 20 nM and confirmed that the duplex could bind to a vesicle once formed.

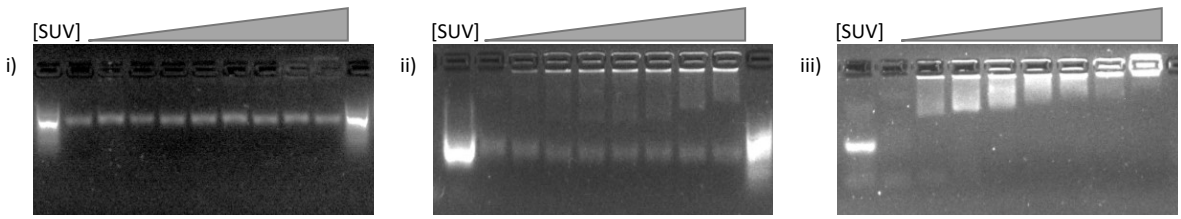
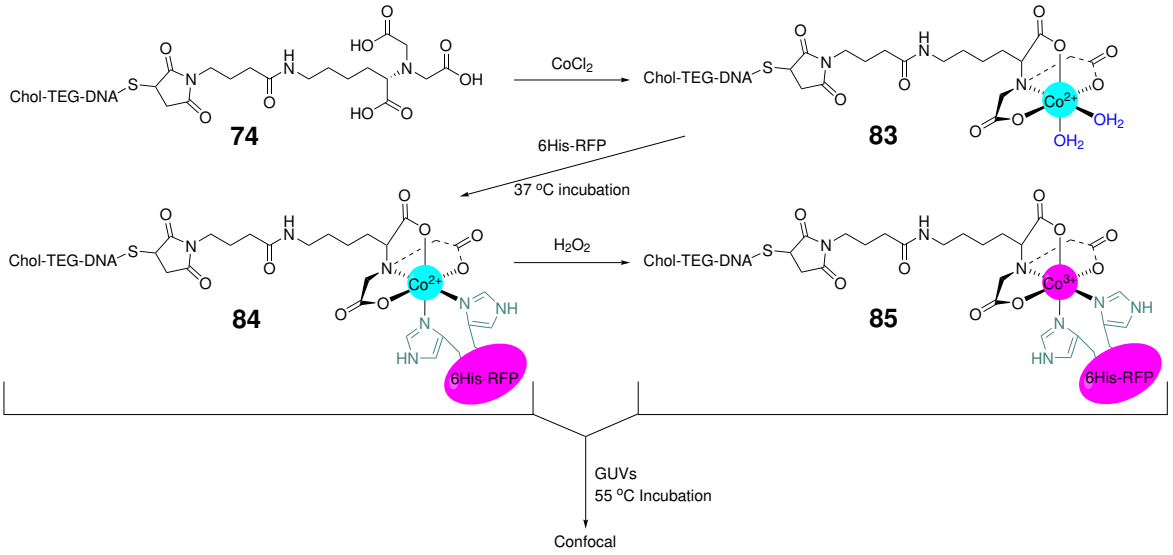


Figure 4.6: DNA linker SUV titrations - All gels show [Linker] = 2.5 μM i) Room temperature incubation, [SUV] = 0, 0.05, 0.1, 0.25, 0.5, 0.75, 1, 2.5, 5 nM; ii) 37 °C incubation, [SUV] = 0, 0.2, 0.5, 1, 1.5, 2, 5, 10 nM; iii) 55 °C incubation, [SUV] = 0, 0.4, 1, 2, 3, 4, 10, 20 nM.

4.4.4 Confocal Microscopy

Once the individual components had been tested, the full linker could then be tested using confocal microscopy. The linker samples were prepared as follows: $CoCl_2$ (150 μM) was added

to the duplex (5 μM) to give **83**; RFP (35 μM) was added to the duplex-Co(II) complex (4.3 μM) and incubated for 30 minutes at 37 °C to give **84**; half the sample was oxidised using H_2O_2 (100 μM) to give **85**; the two samples were then added to GUVs and incubated for one hour at 55 °C before analysis by confocal microscopy.



Scheme 4.2: Preparation of the DNA Linker **74** for confocal microscopy - i) The linker **74** was first added to $CoCl_2$ to give **83**; ii) FresnoRFP was then added and the mixture was then incubated at 37 °C for 30 minutes to give **84**; iii) Half of the sample was oxidised with H_2O_2 to give **85**; iv) Both samples, **84** and **85** were then added to GUVs and incubated at 55 °C for 1 hour to form the confocal samples.

These first samples, **84** and **85**, did not show any binding whilst the positive control of a cholesterol-DNA-Cy3 modified strand showed clear binding (Figure 4.7). The concentrations of the duplex were then increased to give a final duplex-Co(III)-RFP concentration of 2.5 μM, which was then added to GUVs and incubated or added directly under the microscope. These experiments also proved unsuccessful showing no binding, giving the same results as with the negative control of just adding RFP, so a different approach was required to get the linker to bind.

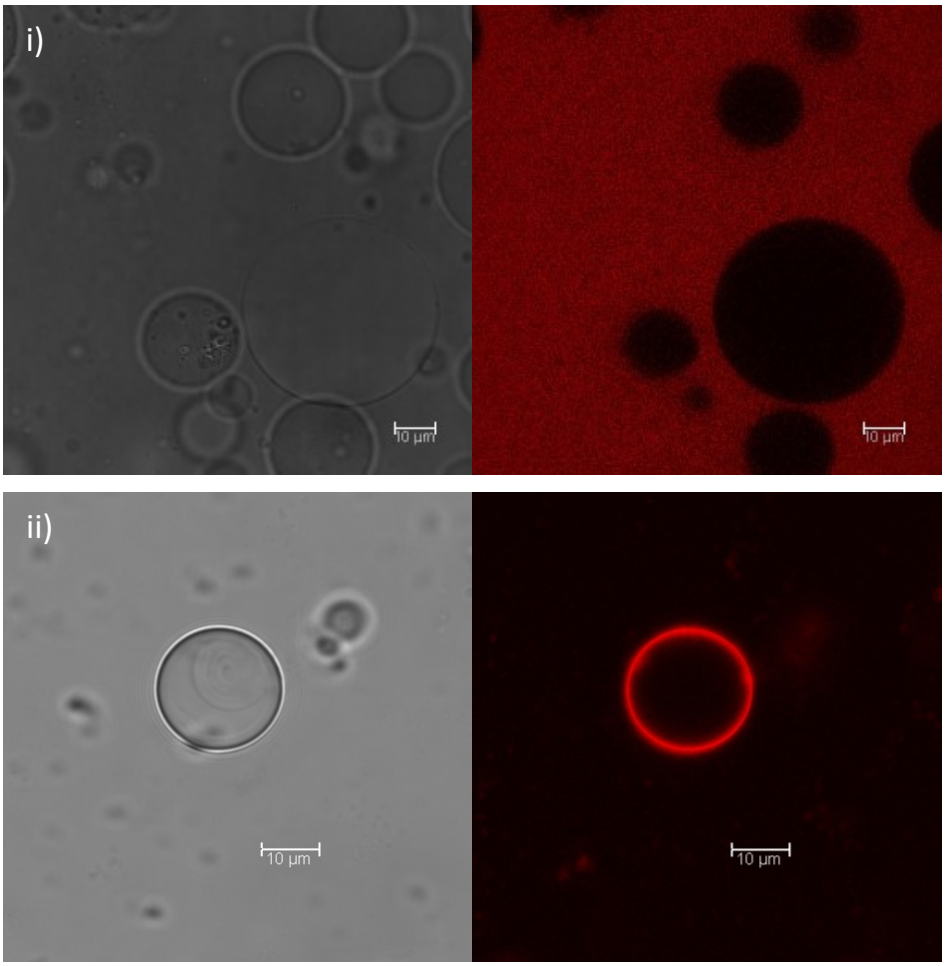


Figure 4.7: DNA linker confocal images - i) Linker sample **84** showing no binding to the GUVs; ii) Cholesterol-DNA-Cy3 control strand showing clear binding to the GUVs as a ring of fluorescence.

4.4.5 New Cholesterol Strands

The original design for the DNA linker included a large overhang, 15 nucleotides, to enable toe-hold mediated strand displacement. As seen with the duplex formation gels (Figure 4.3), this overhang wrapped around the cholesterol and reduced its interaction with the gel, shown through the loss of streaking once the duplex was formed (Figure 4.8). This is a useful feature for gel electrophoresis, however, this may have caused the linker to not bind to the GUVs, even with the higher incubation temperatures to break the cholesterol-cholesterol interactions. Two new cholesterol strands were ordered with no overhang and a smaller 5 nucleotide overhang, which has been shown to not affect cholesterol binding (Table 4.2).[233]

Table 4.2: New cholesterol strand sequences - The two new strands with 0 and 5 nucleotide (nt) overhangs. The colours and labels correspond to Figure 4.2.

Strand	***Sequence (5'→3')***
Cholesterol Strand No Overhang	CGTAGGGTATTGAATGAGGGGAAGTGACATGGAGA-*TEG-Chol*
Cholesterol Strand 5 nt Overhang	CGTAGGGTATTGAATGAGGGGAAGTGACAT-*TEG-Chol*

The duplex formation for each strand with the original thio-NTA strand (**73**) was carried out using the conditions developed in section 4.4.1, heating to 95 °C for 2 minutes followed by cooling to 20 °C at 6 °C/minute. The effect of the different lengths of the overhang was very apparent when all three overhang lengths were run on the same gel (Figure 4.8). The linker with no overhang (**86**) was streaky and sits nearer the top of the gel showing how the cholesterol slowed the DNA migration. The two duplexes with overhangs sat in clear single bands towards the bottom of the gel. The 5 nucleotide overhang linker (**87**) exhibited slightly more streaking than the original 15 nucleotide overhang linker (**74**) which showed how the length of the overhang affected the level of interaction the cholesterol had with the gel. These differences in the interactions matched the streaking patterns observed by Ohmann *et al.* [233].

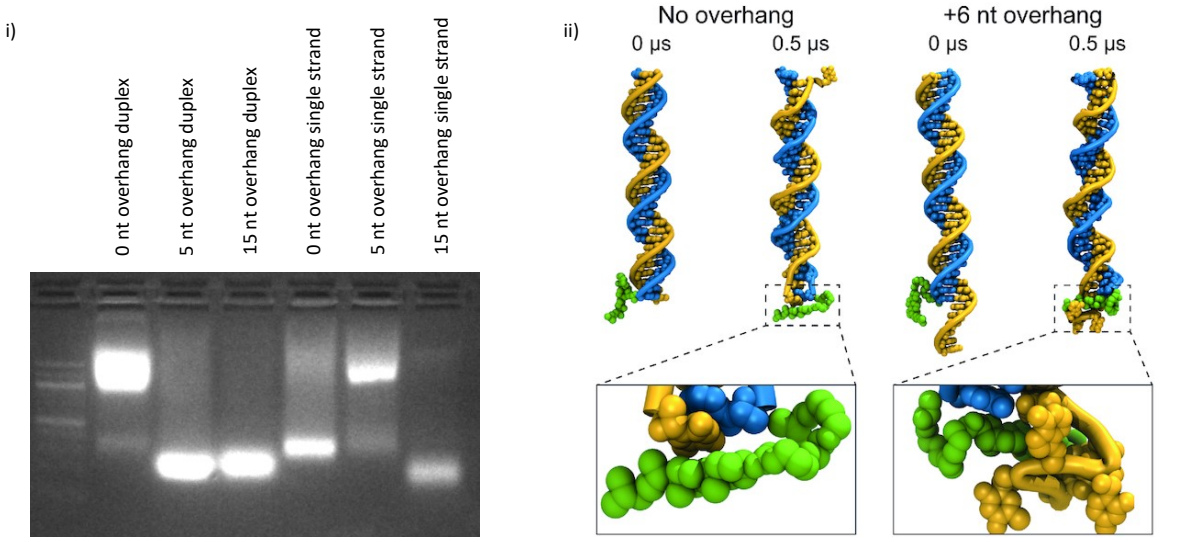


Figure 4.8: New cholesterol strands duplex formation - i) agarose gel showing the three different cholesterol strand overhang lengths as duplexes and single strands, showing how the longer overhangs shield the cholesterol from interacting with the gel; ii) Figure adapted from Ohmann *et al.* [233] showing all-atom MD simulations to show how the cholesterol (green) is free when there is no overhang and how it is shielded by the single stranded overhang (yellow).

Once the duplex formation had been confirmed the new linkers, **86** and **87**, could be tested against SUVs to determine the effects of the different overhangs on membrane tethering. These SUV titrations were performed using the conditions identified in section 4.4.3, incubating the duplexes with POPC SUVs at 55 °C for 1 hour prior to gel electrophoresis. As with the duplex formation the different overhang lengths greatly affected the migration patterns of the linkers through the gel (Figure 4.9). Both of the new linkers (**86** and **87**) showed clear binding at the higher SUV concentrations, not quite full binding, as was seen with the original linker (**74**), but, this can be accounted for as the additional nucleotides add steric bulk, which reduced the packing density each linker can achieve.

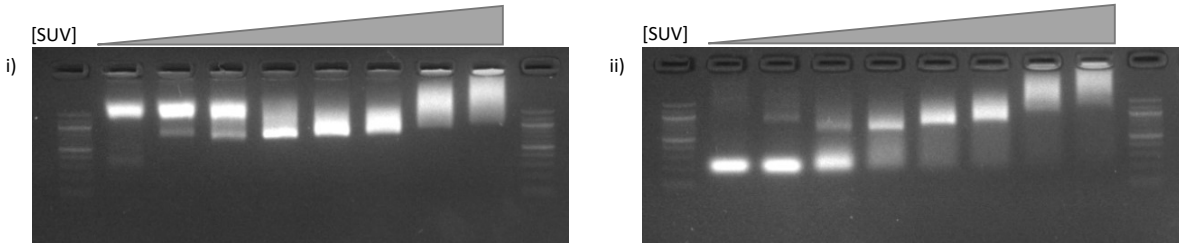


Figure 4.9: New cholesterol strands SUV titrations - i) No overhang linker (**86**) 2.5 μM, [SUV] = 0, 0.4, 1, 2, 3, 4, 10, 20 nM ; ii) 5 nt overhang linker (**87**) 2.5 μM, [SUV] = 0, 0.4, 1, 2, 3, 4, 10, 20 nM.

The next step for testing the new linkers was to perform the fluorescent protein titrations to confirm binding to the duplex. As with the SUV titrations, these were performed following the conditions developed previously in section 4.4.2, binding each linker to cobalt(II), followed by incubation with CFP for 30 minutes at 37 °C, and oxidation of cobalt(II) using H_2O_2. These were then analysed as before using PAGE and higher protein concentrations in the hope that binding would be clearly visible (Figure 4.10). Unfortunately, the new cholesterol linkers (**86** and **87**) clearly showed no binding to the protein as the bands for the linkers were consistent across the gels and no shifts were observed. The gels were also very different to the previous titrations (Figure 4.4), showing less streaking, most likely due to the different cholesterol interactions with the gel. Using CFP, which is excited at a similar wavelength to ethidium bromide, also showed how the large fluorescent protein remained in the well. The lack of binding may have been due the ionic bonds between the linkers and protein not being strong enough to withstand electrophoretic conditions, or, there was a problem with the linkers which prevented protein binding from occurring.

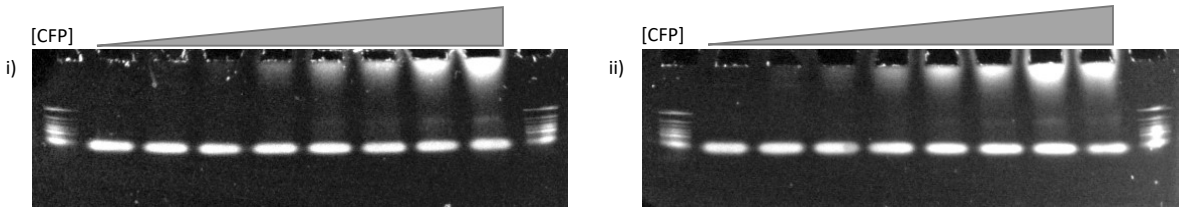


Figure 4.10: New cholesterol strands CFP titrations - i) No overhang linker (**86**) 2.5 μM, [CFP] = 0, 0.35, 0.7, 1.75, 3.5, 5.25, 8.75, 17.5 μM; ii) 5 nt overhang linker (**87**) 2.5 μM, [CFP] = 0, 0.35, 0.7, 1.75, 3.5, 5.25, 8.75, 17.5 μM.

Another method, which would not break ionic bonds, to confirm protein binding to a linker, size exclusion chromatography (SEC), was used to attempt to isolate a fraction containing just the linker (**86**) bound to CFP. This was performed using an AKTA Purifier and a sephadex 200 SEC column. The linker and protein could be separated from one another using this method, however, after attempts using both cobalt(II) and cobalt(III), no additional peaks were observed. This confirmed that the protein was not coordinating to the linker under these conditions, despite

the NTA group being expected to easily coordinate to the CFP.

Electrophoretic and possibly SEC conditions may have caused the bond between the linker and protein to break, so a switch was made back to confocal microscopy to provide an alternative method to analyse the linker. Unfortunately, under the conditions used to prepare the linker (**86**) for the CFP titrations, using H_2O_2 for the oxidation step, no binding to GUVs was observed. Another component, which could have prevented binding, was the high concentration of cobalt(II) which was in a 2 fold excess of the linker (**86**). This could have affected the binding as the cobalt(II) could coordinate to the linker (**86**) and the protein separately and leave no free histidines available to bind, preventing the linker coordinating to the protein. In case this was preventing binding, after the cobalt(II) had been added to the linker (**86**), any un-coordinated cobalt was removed by spin filtration (3,000 MWCO) so no excess would interfere with the binding. A spin filtration step was also incorporated, with a higher MWCO of 30,000, after the protein had been added in the hope that any unbound linker or protein would pass through the filter and only the linker-CFP conjugate would be retained. Fluorescence was observed above and below the filter which indicated that there may have been some conjugate present. However, when this sample was analysed by confocal microscopy, including addition of OPOE up to 5%, no binding was observed, so the fluorescence observed above the filter may have been due to the presence of unbound linker or protein aggregates.

If, however, the linker (**86**) had actually bound to the protein but was not inserting in to the membrane, this result would have been the same. The SUV titrations clearly showed binding between all of the linkers and the lipid membrane, however, the difference between SUVs and GUVs, particularly membrane curvature, could have prevented any of the linkers from binding during the confocal microscopy experiments. In order to try and investigate this, a method for covalently labelling the no overhang linker (**86**), through the thio strand, was attempted. The free sulfur group on the reduced thio strand was reacted with fluorescein maleimide. This labelled strand was then used to form a duplex with the 5 nucleotide overhang cholesterol strand, which could be tested by confocal microscopy. However, this labelled duplex did not show any binding to the GUVs. This indicated that, despite the previous titrations clearly showing binding to SUVs, these results were not transferable to the confocal conditions and binding to GUVs was not possible. This may be due to the different membrane curvatures, the type of lipid, or the different conditions, such as glucose content, required for confocal analysis.

4.5 Conclusions

This work aimed to develop another enzyme immobilisation platform utilising a DNA-based linker, incorporating toe-hold mediated strand displacement to provide recyclability by enabling specific removal of degraded enzymes. This linker consisted of two halves, one DNA strand modified with addition of a cholesterol-TEG group (**72**) and a complementary DNA strand modified with the addition of an NTA group (**73**). The synthesis of the NTA-modified strand was achieved through five synthetic steps and demonstrated a robust route to the high value intermediate maleimido-C3-NTA (**76**), which was then coupled on to the DNA strand. Once the DNA strand had been modified with the NTA group (**73**), it showed successful hybridisation, with the cholesterol-modified strand (**72**) to form the linker duplex (**74**), after a suitable method was developed. The DNA-based linker (**74**) was then ready to be tested to determine if it could coordinate to a His-tagged protein and embed into a lipid membrane.

The first titration of this linker (**74**), against different concentrations of RFP, appeared to show binding of the protein as a gel shift from 0.175 μM and 0.350 μM RFP using Co(II) and Co(III) as the coordinating ion, respectively. This seemed to confirm the increase in binding strength between the His-tagged protein and the NTA group when Co(II) was oxidised to Co(III) which was also observed by Wegner and Spatz[204]. Oxidation of the cobalt(II) was, therefore, incorporated into further experiments to maximise binding, along with the sample preparation used for these titrations. The titration against SUVs required more method development as the initial experiment did not show any binding of the linker to the SUVs. An increase in temperature showed that the room temperature incubation did not provide sufficient energy to overcome the binding barriers. The pre-incubation temperature had to be increased to 55 °C before full binding could occur, as this temperature breaks any cholesterol-cholesterol interactions and enable the linker to bind to the SUVs.[233] These two titrations demonstrated that both desired functions of the linker, vesicle and protein binding, appeared to be working.

When it came to confocal microscopy analysis, however, these desired functions appeared to be absent as no binding to GUVs was observed. Despite using the conditions which were successful for the gel titrations, only the positive control, a cholesterol-DNA-Cy3 strand, showed any binding to the GUVs. This was possibly due to the large ssDNA overhang of 15 nucleotides wrapping around the cholesterol preventing it from binding to the GUVs. Evidence of this was observed during gel electrophoresis by comparing the sharp band for the linker duplex (**74**), where the

cholesterol was protected by the overhang, against the streaky band for the ssDNA cholesterol strand (**72**), where the cholesterol interacts strongly with the gel matrix.[233] New cholesterol strands with 0 and 5 nucleotide overhangs aimed to overcome this problem by allowing the cholesterol to be more available for binding.

The new cholesterol strands showed successful duplex formation with the thio-NTA strands to form two new linkers, **86** and **87**, and showed successful SUV binding under the conditions developed for the original DNA linker (**74**), although not as complete binding due to the extra steric bulk from the additional nucleotides. When it came to the fluorescent protein binding, however, the titrations with CFP clearly showed no binding for either **86** or **87**, even with the higher CFP concentrations. CFP was used as it excites at a similar wavelength to ethidium bromide so was visible under UV light, which showed that the CFP did not migrate through the gel and did not interact with either linker. When comparing this data to the earlier RFP titrations, the new linkers (**86** and **87**) behaved very differently, likely due to the cholesterol interactions, and contradicted the previous results. It is possible that the slight negative charge on the surface of the CFP repelled the negatively charged DNA further whilst it remained in the well. The additional testing of the fluorescent protein binding through size exclusion chromatography and confocal microscopy confirmed that neither linker, **86** or **87**, were binding to CFP, despite the NTA group being commonly used for protein coordination. The binding of the linkers to the SUVs was also clear by titration, however, when fluorescin was used to label one of the linkers (**86**) covalently, this seemed to show no binding of the linker to GUVs, further contradicting the gel electrophoresis data.

Despite all the method development and previous successful electrophoretic data, the two functions of the linker could not be confirmed visually, by confocal microscopy. It is possible that the cholesterol binding may not have been transferable from SUVs to GUVs due the different membrane curvatures, or, more likely, the different conditions required for confocal microscopy, such as the requirement for glucose and sucrose to form GUVs, could have had an impact on the cholesterol binding. Using or incorporating another lipid, such as POPE, could increase the possibility of binding and the stability of the DNA.[232] More problematic than GUV binding, however, was getting any of the linkers to coordinate to a protein. This was initially unclear in the electrophoresis experiments and then concluded as unsuccessful by subsequent experiments. This could be due to a number reasons: the bond formed between the linker and the protein

may have been too weak or too labile, despite the oxidation of Co(II) to Co(III) increasing the strength of the interactions; the cobalt(II) may not be coordinating to the synthesised NTA group or His-tag, in the experimental time frame; or a much higher or lower concentration of either the linker or the protein may be required which was not established during this work. When NTA groups are used for protein purification they are attached to agarose beads and are in a much higher stoichiometry and in a confined environment. It is possible that having any of the linkers binding in solution, the chance of a linker encountering a protein is much lower than during protein purification procedure, so future work could take this approach and work with linkers attached to a membrane before adding a protein. Similar work investigating NTA modified lipids showed that when only one NTA group, from a supported bilayer, was coordinated to a His10-GFP the binding was labile until an additional NTA group was coordinated to another histidine on the protein.[235] The additional histidine groups may have given this method an advantage, and like with the Ni-NTA purification, having a confined location for the NTA groups on the supported bilayer may have also increased binding. The DNA linker could be investigated with supported bilayers, His10-tags, flow systems, other lipids, and other systems in the future.

5 | Conclusions and Future Work

The aim of this thesis was to investigate enzymatic cascades and develop novel platforms for immobilisation of these cascades.

Towards this aim a cascade involving ene-reductases, alcohol dehydrogenases and transaminases was investigated as a potential pathway to produce nepetalactone analogues from α/β-unsaturated diols. A number of substrates were successfully synthesised through chemical methods as starting materials for these cascades. A range of 23 enzymes were also successfully produced through recombinant expression and some of these were further purified through nickel affinity chromatography for use as purified proteins in the subsequent assays. Unfortunately, despite some initial promising results, the enzymes which had been selected for this cascade were only active on a small number of substrates and had too narrow a substrate scope to accept the substrates required for the designed cascade. There was also not a set of three enzymes identified that had compatible reactivity to produce a three enzyme cascade with any of the substrates, which could have been used to test the immobilisation platforms irrespective of the product they produced. This work was also disrupted by lack of access to equipment, namely GC, due to technical problems as well as by a lack of access to the laboratory during the lockdown period.

The enzyme panel used in this thesis was not particularly expansive so future work could focus on the same cascade with a wider, more promiscuous, enzyme panel. This panel could also be expanded to include enzymes with a higher efficiency so the cofactors required are more effectively recycled and do not require supplementation with an additional recycling enzyme. The enzymes used in this project were also all wild-type enzymes so these enzyme could have been engineered to enable a successful cascade to be developed. Although this was beyond the scope of this thesis, enzyme engineering efforts, such as mutagenesis and directed evolution, could be used in future work to produce a robust cascade for producing these nepetalactone analogues.

The first enzyme immobilisation platform developed relied upon a synthetic cholesterol based linker to tether enzymes through a His-tag to a lipid surface through a chelating cobalt-NTA group. The development of this platform required the design and synthesis of the cholesterol linker. The linker design went through several iterations due to synthetic constraints and ensuring that the linker had the required functions, to have a hydrophobic cholesterol to embed into a lipid

membrane, a hydrophilic spacer, and a chelating group. A robust synthetic route to the linker was developed and the linker was eventually synthesised successfully through a five-step synthesis combining solution and solid-phase synthetic techniques. Initial results using GFP as a model protein were promising and showed that the linker could tether GFP to a membrane surface after aggregation barriers had been overcome. In order to test this linker further, three fluorescent proteins were successfully cloned, expressed, and purified to act as a three-protein FRET system. Unfortunately, despite overcoming problems with fluorescence quenching, experiments with these new proteins could not replicate the results seen with GFP, throwing into question these previous results. The linker had been left for several months between these experiments and the high acidity observed may have caused the linker to decompose. Due to the initial results not being replicated and the linker not show binding to the protein during ITC experiments, the linker was resynthesised but time restraints prevented proper analysis of this newly synthesised linker.

One of the main problems with this linker was the inability to test the individual components to determine where the system was falling down. This resulted in the development of the DNA linker as an alternative enzyme immobilisation platform. This second immobilisation platform had a similar design to the first, a cholesterol based linker to tether enzymes to a lipid surface through chelation of a His-tag. Using DNA made this design easier to analyse each step of the process as DNA can be easily visualised. This design also incorporated a system for regenerating the catalyst if it became inactivated which would not have been possible with the original chemically synthesised cholesterol linker. Again, a robust synthetic route to generating this linker was established and initial results were promising, showing SUV and protein binding after some method development. However, the confocal imagery results did not confirm these initial results so additional linkers, without overhangs, were designed. These linkers showed binding to SUVs but did not show binding to the fluorescent proteins. Unfortunately, the full function of the linker could never be realised as the linker could not tether the protein to the lipid surface.

The problems seen with both of the immobilisation platforms came down to the linker not binding to the protein through the NTA-cobalt group, a feature which was consistent across both platforms. A number of possible reasons for this included: the bond formed being too weak; lack of coordination between the NTA and His-tag; and unknown stoichiometries. The most likely explanation, however, is the system being solution based rather than using a solid support which is the basis of Ni-NTA protein purification. NTA groups are used widely in enzyme immobilisation due to the

easy of introducing a His-tag onto a protein of interest, however, they are almost exclusively used when attached to solid supports.

The idea behind this project was to allow some level of movement of the linker across the surface to maximise the flux through the cascade, this was a similar idea to the commercially available 18:1 DGS-NTA (Avanti Polar Lipids)[212,236] a lipid with an NTA group covalently attached to its head group. There are a wide range of applications using NTA groups to attach enzymes to solid surfaces, such as: carbon nanotubes,[237] supported bilayers,[235] magnetic nanoparticles,[238] and poly(methyl methacrylate) (PMMA) microfluidic devices.[239] Microfluidic and other flow systems are becoming the leading technologies for enzyme immobilisation by combing high selectivity and benign conditions of biocatalysis with enhanced mass transfer and resource efficiency of flow chemistry.[240] The immobilisation platforms developed in this thesis may have been more likely to work as flow systems, having the linker immobilised prior to addition of the protein would make the system more like the established immobilisation platforms. Having the enzymes separated, the ADH and ER in one section to recycle the cofactor and the TAm in another where the sacrificial substrate could be added separately. If the immobilisation had been successful transferring the system to flow would have been the next step to making the system a more commercially viable product and applicable to industrial systems.

Developing greener, more efficient ways to produce chemicals is paramount to the evolution of the chemical industry to a fully sustainable industry in this time of global warming and global responsibility to preserve the planet. Combining biological techniques,[241,242] large scale process chemistry, smart technology,[243] green energy, sustainable feedstocks[244,245] and other new technologies will be required for this green evolution.

6 Materials and Methods

6.1 Experimental

6.1.1 General Experimental

All chemicals were purchased from Sigma-Aldrich and used as provided, unless otherwise stated, and all solvents were purchased from Fisher Scientific, unless otherwise stated. Oven (Memmet) dried glassware was used where appropriate. Flash chromatography was carried out using Geduran Silica gel 60 (40- 60 μm). Thin layer chromatography (TLC) was carried out using Silica gel 60 F_{254} plates (Millipore) and were visualised using ultra-violet (UV) light, potassium permangamate, phosphomolybdic acid, ninhydrin or bromocresol green stains. All anhydrous reactions were performed under argon. Sonication of cell lysates was performed using a probe Sonifier 150 (Branson) sonicator. The following centrifuges were used: Allegras x-15R centrifuge (Beckman Coulter), Centrifuge 5415R (Eppendorf), Centrifuge 5810R (Eppendorf), Centrifuge 5430R (Eppendorf). The following equipment was used where applicable: SI600C shaking incubator (Stuart Equipment), class I biological safety cabinet (Microflow), ultrasonic cleaner (VWR), rotary evaporators (Büchi) and freeze dryer (SP Scientific). Growth media and apparatus were autoclaved (Priorclave) at 121 °C for 15 minutes where required.

6.1.2 Spectroscopy

6.1.2.1 Nuclear Magnetic Resonance Spectroscopy (NMR)

^{1}H and ^{13}C NMR spectroscopy was carried out using the following instruments (Bruker): Avance Neo 700, Avance III 600, Avance Neo 500, Avance III 400. All chemical shifts (δ) are reported in ppm, relative to the residual protonated solvent signal. The multiplicity is given by the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, p = pentet, m = multiplet, and br. = broad signal. The coupling constants (J) were recorded in hertz (Hz). Data was viewed and processed in MestReNova (Mestrelab Research).

6.1.2.2 Low-Resolution Liquid Chromatography-Mass Spectrometry (LC-MS)

Low-resolution LC-MS data was acquired using an Acquity UPLC (Waters) equipped with a UV detector and coupled to a Single Quad mass spectrometer (Waters). Analytes were separated using a BEH C18 (50 x 2.1 mm, 1.7 μm beads) column (Thermo Scientific) and a 5-95% gradient

of acetonitrile (+0.1% formic acid) in water (+0.1% formic acid) over 5 minutes. Data was analysed using MassLynx software (Waters).

6.1.2.3 High-Resolution Mass Spectrometry (HRMS)

High-Resolution MS data was acquired by the UCL Mass Spectrometry Department using a Waters LCT Premier XE and ESI. Data was analysed using MassHunter software (Agilent).

6.1.3 Molecular Cloning Methods

6.1.3.1 Plasmid Preparation and Isolation

100 ng DNA was used to transform NEB10-β competent E.coli cells (NEB), which were plated onto LB agar containing kanamycin (50 μg/mL) and incubated at 37 °C overnight. Single colonies were used to inoculate LB media (10 mL), containing kanamycin (50 μg/mL), and were incubated overnight at 37 °C, 200 rpm. DNA stocks were prepared using a Monarch Plasmid Miniprep Kit (NEB) following the manufacturer's instructions. DNA was eluted in 30 μL elution buffer. DNA concentrations were calculated using a NanoDrop ND-1000 Spectrophotometer (Labtech International) using Monarch Elute buffer as a blank.

6.1.3.2 High Fidelity PCR

Primers for PCR were ordered from Integrated DNA Technologies (IDT) to amplify the DNA and introduce the appropriate restriction sites given in Table 6.1.

Table 6.1: PCR primers for fluorescent protein gene amplification. Sequences recognised by the corresponding restriction enzyme are underlined.

Primer	*Sequence (5'→3')*	*Tm Q5*	*Cloning Site*
YetiYFP_NdeI_Fwd	GAGGTAAACATATGACGGCATTGACGGAAGG	72	NdeI
YetiYFP_XhoI_Rev	CACCACCTCGAGTTAGCGATACGTCTCCAG	75	XhoI
FrostyCFP_NdeI_Fwd	GAGGTAAACATATGACGGCATTGACGGAAGG	72	NdeI
FrostyCFP_XhoI_Rev	CACCACCTCGAGTTACTGATACGTGTCCA	73	XhoI
FresnoRFP_NdeI_Fwd	CGGCAGCCATATGAATAGCCTGATTAA	69	NdeI
FresnoRFP_XhoI_Rev	CACCACCTCGAGTTATTTGTACAGTTCGTCCA	73	XhoI

High fidelity PCR amplification was performed, when the PCR product was to be introduced into a vector, using Q5 polymerase (NEB). A generalised PCR reaction is shown in Table 6.2. When PCR was required to confirm insertion of a gene, the template DNA was replaced with a colony resuspended in nuclease-free water (10 μL) and the amount of water reduced to 10 μL. PCR reactions were performed in a T100 thermocycler (Bio-Rad) using the protocol given in Table 6.3.

Table 6.2: PCR Reaction Components

Reaction Component	*Volume (μL)*
Q5 High-Fidelity MasterMix (2x)	25
Forward Primer (10 μM)	2.5
Reverse Primer (10 μM)	2.5
Template DNA	1
Nuclease-Free Water	19
Total	50

Table 6.3: PCR Reaction Conditions

Step	*Temperature (°C)*	*Time (s)*
Initial Denaturation	98	30
Chain Extension - 25 Cycles	98	5
	70-72	10
	72	20
Final Extension	72	120
Hold	4	∞

6.1.3.3 Agarose Gel Electrophoresis

1% agarose gels were prepared using UltraPure Agarose (Invitrogen) and SYBR Safe DNA Gel Stain (0.004% v/v, ThermoFisher) in TAE buffer (23.5 mM Tris base, 20.0 mM acetic acid and 1.3 mM EDTA). PCR reactions were loaded with 6x purple loading dye (NEB) alongside Quick-Load Purple 100 bp or 1 kb DNA Ladder (NEB). Gels were run in a Sub-Cell GT Cell (Bio-Rad) in TAE buffer at 100 V until separation was achieved. Gels were visualised with an Azure 200 gel imager (Cambridge Bioscience).

6.1.3.4 Gel Extraction

DNA fragments from PCR or restriction digests were extracted from agarose gels and purified using a Monarch DNA Gel Extraction Kit (NEB) following the manufacturer's protocol. Elution was carried out in 20 μL elution buffer.

6.1.3.5 Restriction Digests

Restriction digests were carried out using up to 1000 ng of purified DNA using NdeI and XhoI Time-Saver restriction enzymes (NEB) and NEB reagents. A generalised reaction is shown in Table 6.4. Reactions were carried out at 37 °C for 30 minutes, followed by 10 minutes at 80 °C, to heat inactivate the enzymes. Digests were analysed on agarose gels followed by gel extraction of the desired fragment.

Table 6.4: Restriction Digest Reaction Components

Reaction Component	*Volume (μL)*
NdeI	1
XhoI	1
CutSmart Buffer	5
Insert or Vector DNA	20
Nuclease-Free Water	23
Total	50

6.1.3.6 Ligations

Restriction digested vector and insert fragments were ligated at a molar ratio of 1:3, 1:5 or 1:7 vector to insert, with 50 ng of vector DNA. Ligation reactions were carried out using T4 DNA Ligase (Thermo Scientific) and a generalised reaction is shown in Table 6.5.

Table 6.5: Ligation Reaction Components

Reaction Component	*Volume (μL)*
T4 Ligation Buffer (10x)	2
Vector (50 ng)	2.0-6.7
Insert	0.2-1.6
T4 DNA Ligase	1
Nuclease-Free Water	8.7-14.8
Total	20

Ligations were carried out for 10 minutes at room temperature or overnight at 16 °C before being heat inactivated at 65 °C for 10 minutes. Ligations were then placed on ice ready for transformation into chemically competent NEB10β *E. coli* cells.

6.1.3.7 E. coli Transformation

All transformations were performed with storage NEB10β or expression BL21(DE3) strains (NEB). 2 μL of DNA was added to 50 μL of competent cells and incubated on ice for 30 minutes. Cells were heat-shocked in a water bath at 42 °C for 45 s before cooling on ice for 5 minutes. 900 μL of pre-incubated SOC or 10β outgrowth media (NEB) was added and cells were incubated at 37 °C with shaking (900 rpm, 1 h). Cells were pelleted (8000 rpm, 30 s) before being plated out onto LB agar plates containing kanamycin (50 μg/mL) before being incubated at 37 °C overnight.

6.1.4 Protein Expression, Purification and Analysis

6.1.4.1 General Protein Expression Protocol

500 or 20 mL LB cultures, containing kanamycin (50 μg/mL), were inoculated with 1% (v/v) inoculum from a 5 mL starter culture (BL21 (DE3)). Cultures were grown at 37 °C, 200 rpm, until they reached an OD_{600} of 0.6-0.8. Expression was induced by the addition of IPTG (1 mM final) and cultures were incubated overnight at 25 °C, 200 rpm. The cells were harvested by centrifugation (30 min, 6000 *g*) and cell pellets were either used directly or stored at -80 °C until required.

6.1.4.2 Cell Lysis

Small scale lysis (to confirm protein expression) was performed by re-suspending a cell pellet (from 1 mL of culture) in 1 mL lysis buffer (0.1% w/v lysozyme, 1% v/v Triton X-100, 20 mM Tris, 0.5 M NaCl, pH 8.0) or BugBuster (Sigma-Aldrich), left to lyse (30 min, rt) and cleared by centrifugation (30 s, 14,200 *g*). Large scale lysis (for assays or purification) was performed by re-suspending a cell pellet (from 500 mL of culture) in Tris buffer (10, mL, 50 mM, pH 7.4), sonicated (10 s on, 10 s off, 10-15 cycles) and cleared by centrifugation (30 min, 6,000 *g*).

6.1.4.3 Protein Purification

Clarified lysate (10 mL) was purified using an ÄKTA purifier 900 (GE) equipped with a 5 mL HisTrap HP (GE). Buffer A (50 mM Tris, 150 mM NaCl, 10 mM imidazole, pH 7.4) and buffer B (50 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 7.4) were prepared in house and were filtered and degassed prior to use. A step gradient of two column volumes of 0, 20, 40 and 80% buffer B and a flow rate 1.5 mL/min were used to purify the protein. Protein was detected using a UV detector monitoring at 260 and 280 nm. Fractions containing the desired protein were then concentrated using a centrifugal concentrator (Vivaspin 6) to 0.5 mL. Buffer exchange, and additional purification, was performed using a Superdex 75 10/300 GL size exclusion column (GE) into a storage buffer (50 mM Tris, 300 mM NaCl, 10 mM $MgCl_2$, 10% glycerol, pH 7.4). Alternatively a PD10 column (GE) was utilised to perform manual buffer exchange. Purified proteins were either used directly or stored at -80 °C until required.

6.1.4.4 SDS-PAGE

A sample from the soluble and total cell lysate fractions was mixed with 4x Laemmli Sample Buffer (Bio-Rad) and heated to 95 °C for 10 minutes. 20 μL from each sample was loaded into an SDS-

PAGE 4-20% Novex WedgeWell Tris-Glycine Mini Protein Gel (Invitrogen) alongside Precision Plus Protein Dual Color Standards (Bio-Rad). Gels were run in an XCell SureLock Mini-Cell (Invitrogen) in 1x Tris-Glycine Buffer (Sigma-Aldrich) at 225 V until separation was achieved. Protein bands were visualised with InstantBlue Protein Stain (Expedeon) and imaged with an Azure 200 gel imager (Cambridge Bioscience).

6.1.4.5 Protein Concentration Determination

Protein concentrations were determined by Bradford assay. A protein standard solution (Bovine Serum Albuin) at 2 mg/mL concentration (Sigma Aldrich) was used to prepare standard solutions at 0.2, 0.4, 0.6, 0.8, and 1 mg/mL in water. Serial dilutions were used to prepare 10x and 20x dilutions of the sample of interest. 5 μL of each solution was pipetted into a 96 well plate (standards in duplicate, samples in triplicate). 250 μL of Quick Start Bradford 1x Dye Reagent (BioRad Laboratories) was then added to each well and left to react for at least 5 minutes. The absorbance of each well at 595 nm was recorded using a SpectraMax plate reader (Thermo Scientific). A calibration curve was prepared using the standards and the concentration of the samples was calculated from the absorbance values against a buffer reference. Only absorbance values in the linear range were used for quantification of protein concentration.

6.1.5 Enzymatic Assays

6.1.5.1 Spectrophotometric Assays

200 μL reactions were prepared as follows: ER lysate (50% v/v); substrate (1 mM, from 100 mM stock in DMSO); NADH (1 mM); DMSO (5%) and Tris buffer (50 mM). Reactions were performed in triplicate and were monitored at 340 nm every 20 s over 1 h using a SpectraMax plate reader.

6.1.5.2 Gas Chromatography (GC) Assays

500 μL biotransformations were prepared as follows: purified enzyme (0.1-1 mg/mL); lysate (10-20% v/v); substrate (5-10 mM, from 100 mM stock in DMSO); $NAD(P)^+$/NAD(P)H (0.1-10 mM); DMSO (0-10%); and Tris Buffer (50 mM, 150 mM NaCl, pH 7.4). Reactions were incubated for 18 h at 30 °C with shaking, 450 rpm. Reaction products were extracted into EtOAc (500 μL) and dried using Na_2SO_4 prior to analysis. Samples were analysed using an Agilent 7820A Gas Chromatograph equipped with a chiral column (Beta DEX 225, fused silica capillary column 30 m x 0.25 mm x 0.25 μm) with 5 μL of sample being injected (Splitless). The methods used

for analysis of chemical standards and biotransformation reactions are shown in the table below (Table 6.6).

Table 6.6: GC Methods for ER and ER/ADH Assays

Rate (°C/min)	*Temperature (°C)*	*Hold Time (min)*
ER Assay Method 1		
Initial	40	1
40	100	8
20	200	1
ER Assay Method 2		
Initial	80	1
40	100	8
20	200	1
ER/ADH Assay Method		
Initial	60	1
20	150	2
5	850	0
40	220	1

6.1.5.3 Transaminase Colorimetric Assays

200 μL reactions were prepared as follows: ADH and *Cv*-TAm lysates (10% v/v each), purified NCR (0.5 mg/mL), $NADP^+$ (0.5 mM), substrate (10 mM, from 100 mM stock in DMSO), PLP (0.5 mM), 4-nitrophenylethylamine (25 mM), and phosphate buffer (100 mM, pH 7.5). The reactions were incubated at 30 °C at 400 rpm for 18 hours before being imaged.

6.1.6 Cholesterol Linker Methods

6.1.6.1 Fmoc Resin Loading Test[211]

5 mg of the resin to be tested was weighed into a 15 mL Falcon tube. 2 mL of DBU (1,8-Diazabicyclo[5.4.0]undec-7-ene) (2% v/v) in DMF was added to the resin, and the reaction mixture was shaken at room temperature for 30 minutes. A reference containing only 2 mL of DBU (2% v/v) in DMF was also prepared. After this time, 8 mL acetonitrile was added to each tube. 2 mL of each diluted solution was removed into 50 mL Falcon tubes, and further diluted with

23 mL acetonitrile. 2 mL of each solution was transferred to 1.5 mL disposable cuvettes (GmbH), and the absorbance recorded at 304 nm using a Lambda 365 UV/Vis spectrometer (Perkin Elmer). The reference solution was used as a blank. Loading was calculated using:

$$Loading\,(mmol/g) = \frac{Abs_{304} \times 16.4}{mass\,of\ resin\,(mg)}$$

6.1.6.2 Ninhydrin Test

A small amount of the resin to be tested (typically enough to coat the tip of a spatula) was placed in a 1 mL glass vial. A second identical vial was prepared using an equivalent amount of unreacted resin. Approximately 200 μL of ninhydrin solution (1.5% (w/v) ninhydrin in *n*-butanol with 3% acetic acid) was added to each vial, and the vials were placed in a 100 °C oven for 5 minutes. A positive ninhydrin test is recorded when the pale yellow resin beads turn a dark blue/purple colour; no change in the colour is observed for a negative result.

6.1.6.3 Isothermal Titration Calorimetry

ITC was performed using a MicroCal PEAQ-ITC (Malvern) and the associated analysis software. The reference cell was filled with ddH_2O (280 μL). The sample cell was filled with the given concentration of fluorescent protein (280 μL). Titrations were performed using 19 injections of 2-4 μL of ligand, at the given concentrations, at 150 s intervals with a 1 minute initial delay, 750 rpm, 25 °C.

6.1.7 DNA Linker Methods

6.1.7.1 DNA Oligonucleotides

DNA oligonucleotides were ordered from IDT DNA at 100 nmol scale. Oligos were purified by IDT DNA using HPLC and delivered freeze-dried. They were made up to 100 μM solutions using nuclease-free water. If required, DNA concentrations were measured using a NanoDrop ND-1000 Spectrophotometer (Labtech International) and calculated using the Beer-Lambert Law:

$$A_{260} = \epsilon_{260} cl$$

Table 6.7: DNA Oligo Sequences

Strand	***Sequence (5'→3')***	***MW***	ϵ_{260}
72	CGTAGGGTATTGAATGAGGG-*TEG-Chol*	7033	206900
73	TCTCCATGTCACTTCCCCTCATTCAATACCCTACG-*Thio*	10717	309900
75	CGTAGGGTATTGAATGAGGGGAAGTGACATGGAGA	11029	369200
86	CGTAGGGTATTGAATGAGGGGAAGTGACATGGAGA-*TEG-Chol*	11785	369200
87	CGTAGGGTATTGAATGAGGGGAAGTGACAT-*TEG-Chol*	10171	311800

6.1.7.2 Duplex Formation

Duplexes were formed at a 1:1 ratio of the corresponding strands with a final concentration of 1-10 μM and final volume of 20-100 μL. Strands were heated to 95 °C for 2 minutes before being cooled at 6 °C/min to 20 °C in a T100 Thermal Cycler (BioRad) to form the duplex.

6.1.7.3 Agarose Gel Electrophoresis

Agarose gels (1% for dsDNA, 2% for ssDNA and dsDNA, and 3% for ssDNA) were prepared using UltraPure Agarose (Invitrogen) and ethidium bromide (dsDNA) or SYBR Gold DNA Gel Stain (0.001% v/v, ThermoFisher, ssDNA) in 1x UltraPure TAE buffer (Invitrogen). ssDNA and dsDNA were loaded with 6x purple loading dye (NEB) alongside Quick-Load low MW, 100 bp or 1 kb DNA Ladder (NEB). Gels were run in a Sub-Cell GT Cell (Bio-Rad) in TAE buffer at 60-70 V for 45 min - 1 h. Gels were visualised with an Azure 200 gel imager (Cambridge Bioscience).

6.1.7.4 Native-PAGE

10% Mini-PROTEAN TGX Native PAGE gels (BioRad) were equilibrated in 1x UltraPure TAE buffer (Invitrogen) for 30 minutes prior to loading with sample. Samples were loaded with 6x purple loading dye (NEB) alongside Quick-Load low MW, 100 bp or 1 kb DNA Ladder (NEB). Gels were run in a Mini-PROTEAN Tetra Cell (BioRad) at 60-70 V for 90 minutes before being stained in ethiduium bromide and visualised with an Azure 200 gel imager (Cambridge Bioscience).

6.1.7.5 Small Unilamellar Vesicle (SUV) Preparation

POPC (150 μL, 10 mg/mL) in chloroform was added to a 1 mL glass vial, the solvent was removed *in vacuo*. The lipids were then resuspended in 1x PBS (1 mL, pH 8, diluted from pre-made 10x PBS (Sigma)) and sonicated in a sonicator bath for 30 minutes. The DNA linker was incubated with the SUVs prior running in the agarose gel to allow for binding to the membranes. Vesicles were used directly or stored at 4 °C for up to a week.

6.1.8 Confocal Microscopy

6.1.8.1 Giant Unilamellar Vesicle (GUV) Preparation[214]

POPC or DOPE (150 µL, 10 mg/mL) in chloroform were added to a 1 mL glass vial, the solvent was removed *in vacuo*. The thin film generated was resuspended in mineral oil (150 μL) by alternating between vortexing and sonicating in a sonicator bath (10 s each for 5 cycles). The inner solution (25 μL, 0.4 M sucrose, 0.3 M potassium glutamate) was added, followed by additional mineral oil (150 μL). The suspension was alternately vortexed and sonicated at room temperature (10 s each for 5 cycles) before being carefully added to the top of the outer solution (1 mL, 0.5 M glucose, 0.3 M potassium glutamate) in an eppendorf. Vesicles were formed by centrifugation (13,000 *g*, 10 min). The mineral oil top layer and the majority of the glucose layer ($\sim$800 μL) were carefully removed by pipetting. The pelleted vesicles were gently resuspended and transferred to a clean plastic vial. Vesicles were used directly or stored at 4 °C for up to a week.

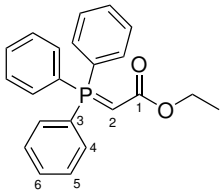
6.1.8.2 Sample Preparation and Microscopy Method

Confocal samples were analysed using a Leica SPE Inverted microscope and LAS-X software. Samples were prepared in 35 mm FluoroDish cell culture dishes (Fischer) and were viewed under 63,000 x magnification, exciting with 405, 488, 532 or 635 nm lasers.

6.2 Chemical Synthesis

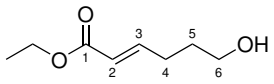
6.2.1 Enzyme Substrate Synthesis

6.2.1.1 (Ethoxycarbonylmethylene)triphenylphosphine (11)[178]



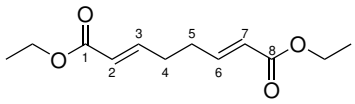
To a solution of PPh_3 (5.25 g, 20.0 mmol) in toluene (40 mL) was added ethyl bromoacetate (3.34 g, 20.0 mmol) over 15 minutes. The soultion was stirred for 2–3 h and the precipitate filtered, washed with cold toluene (3×5 mL) and dried. The collected phosphonium salt was dissolved in H_2O (60 mL) and the mixture was cooled in an ice bath and saturated aqueous NaOH was added dropwise until pH 10. The solid was filtered and washed with cold H_2O, collected and dried *in vacuo* to afford **11** (4.39 g, 12.6 mmol, 63%). ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.65 (6H, m, 6 x 5-H), 7.54 (3H, m, 3x 6-H), 7.45 (6H, m, 6 x 4-H), 3.96 (2H, q, $J = 7.0$ Hz, CH_2OCO), 2.88 (1H, s, 2-H), 1.04 (3H, m, CH_3CH_2O); ^{13}C NMR (151 MHz, $CDCl_3$) δ ppm 172.1 (C-1), 133.1 (C-4), 132.0 (C-6), 128.9 (C-5), 128.8 (C-3), 58.0 (CH_2OCO), 30.6 (C-2), 14.9 (CH_3CH_2O). LC-MS (ESI): *m/z* calc. 348 ($C_{22}H_{21}O_2P$), found $[M+H]^+$ 349.

6.2.1.2 Ethyl (E)-6-hydroxyhex-2-enoate (12)[179]



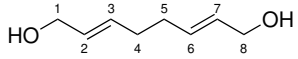
To a solution of 1,4-butanediol (**10**, 0.57 g, 6.32 mmol) in CH_2Cl_2 (100 mL) was added **11** (2.63 g, 7.56 mmol) and MnO_2 (1.80 g, 21.2 mmol) and the mixture was stirred at rt. After 24 h and 48 h, respectively, two additional portions of MnO_2 (2 x 1.80 g) were added. The mixture was stirred for a further 72 h, filtered through Celite^{TM} and washed with CH_2Cl_2 (50 mL). The solvent was evaporated under reduced pressure and the residue was treated with Et_2O to allow precipitation of Ph_3PO. After filtration and solvent evaporation *in vacuo*, the crude product was purified by flash chromatography (hexanes/EtOAc 1:1), affording **12** (420 mg, 2.65 mmol, 42%). ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 6.98 (1H, dt, $J = 15.6$, 6.8 Hz, 3-H), 5.85 (1H, dt, $J = 15.6$, 1.6 Hz, 2-H), 4.19 (2H, q, $J = 7.1$ Hz, CH_2OCO), 3.68 (2H, m, 6-H), 2.31 (2H, m, 4-H), 1.73 (2H, m, 5-H), 1.29 (3H, t, $J = 7.1$ Hz, CH_3CH_2O); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 166.8 (C-1), 148.5 (C-3), 121.9 (C-2), 62.2 (C-6), 60.5 (CH_2OCO), 31.0 (C-5), 28.6 (C-4), 14.4 (CH_3CH_2O); LC-MS (ESI): *m/z* calc. 158 ($C_8H_{14}O_3$), found $[M+H]^+$ 159.

6.2.1.3 Diethyl (2E,6E)-octa-2,6-dienedioate (13)[179]



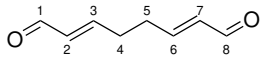
A mixture of **12** (1.00 g, 6.32 mmol) and PCC (1.36 g, 12.6 mmol, ground with 2.72 g of silica) in CH_2Cl_2 (100 mL) was stirred at rt for 5 h. Imidazole was added (0.43 g, 12.6 mmol), and the reaction mixture was stirred for an additional 1 h. After addition of [(ethoxycarbonyl)methylene]-triphenylphosphorane (**11**, 2.75 g, 15.8 mmol), stirring was continued for 24 h. The mixture was filtered through Celite^{TM} and washed with CH_2Cl_2 (50 mL). The solvent was removed under reduced pressure, and the crude material was treated with Et_2O to promote precipitation of Ph_3PO. After filtration, the solvent was evaporated, and the residue was purified by flash chromatography (Hexanes/EtOAc 8:2) to afford **13** (0.69 g, 3.03 mmol, 48%) (a low percentage of cis isomers is also observed in the spectra, <10%). ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 6.92 (2H, m, 3-H 6-H), 5.85 (2H, d, $J = 15.6$ Hz, 2-H, 7-H), 4.18 (4H, q, $J = 7.2$ Hz, CH_2OCO), 2.37 (4H, m, 4-H, 5-H), 1.28 (6H, t, $J = 7.2$ Hz, 2x CH_3CH_2O); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 166.7 (C-1, C-8), 147.2 (C-3, C-6), 122.7 (C-2, C-7), 60.6 (CH_2OCO), 30.8 (C-4, C-5), 14.6 (CH_3CH_2O); LC-MS (ESI): m/z calc. 226 ($C_{12}H_{18}O_4$), found $[M+H]^+$ 227.

6.2.1.4 (2E,6E)-Octa-2,6-diene-1,8-diol (5)[180]



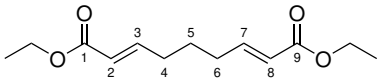
To a stirring solution of **13** (0.34 g, 1.50 mmol) in dry THF (20 mL), DIBAL-H (1 M in THF, 9.0 mL, 9.0 mmol), was added under argon atmosphere at -78 °C, and stirred for 1.5 h. The mixture was allowed to warm to 0 °C and stirred for 1 h. After dropwise addition of MeOH (2 mL) and a saturated solution of potassium sodium tartrate (20 mL) a white emulsion was formed. The phases were separated after 2 h of stirring. The aqueous phase was extracted with CH_2Cl_2 (3 x 80 mL). The combined organic phases were dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure after filtration. The crude product was purified by column chromatography (Hexanes/EtOAc 7:3 2:3) to afford **5** (0.20 g, 1.41 mmol, 94%) as colorless oil. ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 5.65 (4H, m, 2-H, 3-H, 6-H, 7-H), 4.08 (4H, m, 1-H, 8-H), 2.14 (4H, m, 4-H, 5-H), 1.76 (2H, br. s., 2x OH); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 132.3 (C-2, C-7), 129.8 (C-3, C-6), 63.8 (C-1, C-8), 31.8 (C-4, C-5); LC-MS (ESI): m/z calc. 142 ($C_8H_{14}O_2$), found $[M+H]^+$ 143.

6.2.1.5 (2E,6E)-Octa-2,6-dienedial (6)[180]



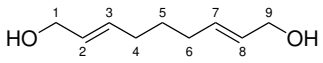
To a solution of **5** (0.20 g, 1.41 mmol) in DMSO (40 mL) a solution of 2-iodoxybenzoic acid (IBX) (1.97 g, 7.05 mmol) in DMSO (40 mL) was added and stirred for 1 h at rt. H_2O (80 mL) was added and the resulting cloudy suspension was filtered through a course fritted funnel and washed with Et_2O (50 mL). The aqueous phase was extracted with Et_2O (3 x 40 mL) and the combined organic phases were washed with brine (60 mL), dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure after filtration. The reaction crude was purified by column chromatography (Hexanes/Et_2O 2:1 1:2), however, IBX remained and could not be removed by further purification. ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 9.54 (2H, d, $J = 7.7$ Hz, 1-H, 8-H), 6.83 (2H, m, 3-H, 6-H), 6.18 (2H, dd, $J = 15.7, 7.7$ Hz, 2-H, 7-H), 2.59 (4H, m, 4-H, 5-H); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 193.6 (C-1, C-8), 155.3 (C-3, C-6), 133.9 (C-2, C-7), 30.7 (C-4, C-5).

6.2.1.6 Diethyl (2E,7E)-nona-2,7-dienedioate (18)[180]



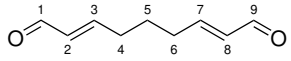
To a solution of glutaraldehyde (**17**) (1.00 g, 10.0 mmol) in CH_2Cl_2 was added **11** (8.71 g, 25.0 mmol) and stirred for 17 h at rt. The mixture was filtered through Celite and washed with CH_2Cl_2. The solvent was evaporated under reduced pressure and the residue was treated with Et_2O to allow precipitation of Ph_3PO. After filtration and solvent evaporation, the crude product was purified by flash chromatography (Hexanes/EtOAc 9:1 4:1), affording **18** (1.65 g, 6.88 mmol, 69%) (a low percentage of cis isomers is also observed in the spectra, <10%). ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 6.92 (2H, dt, $J = 15.6, 6.9$ Hz, 3-H, 7-H), 5.82 (2H, m, 2-H, 8-H), 4.18 (4H, q, $J = 7.1$ Hz, CH_2OCO), 2.23 (4H, m, 4-H, 6-H), 1.63 (2H, m, 5-H), 1.28 (6H, t, $J = 7.1$ Hz, 2x CH_3CH_2O); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 166.7 (C-1, C-9), 148.2 (C-3, C-7), 122.1 (C-2, C-8), 60.4 (CH_2OCO), 31.5 (C-4, C-6), 26.5 (C-5), 14.4 (CH_3CH_2O); LC-MS (ESI): m/z calc. 240 ($C_{13}H_{20}O_4$), found $[M+H]^+$ 241.

6.2.1.7 (2E,7E)-nona-2,7-diene-1,9-diol (19)[180]



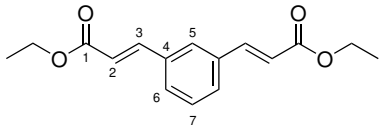
To a stirring solution of **18** (2.0 g, 8.33 mmol) in dry THF (20 mL), DIBAL-H (1 M in THF), (9.0 mL, 9.0 mmol), was added under an argon atmosphere at -78 °C, and the reaction was stirred for 1.5 h. The mixture was allowed to warm to 0 °C and stirred for 1 h. After dropwise addition of MeOH (2 mL) and a saturated solution of potassium sodium tartrate (20 mL) a white emulsion was formed. The phases were separated after 2 h of stirring. The aqueous phase was extracted with Et_2O (3 x 10 mL). The combined organic phases were dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure after filtration. The crude product was purified by column chromatography (Hexanes/EtOAc 1:1 3:7) to afford **19** (301 mg, 1.92 mmol, 23%) as a colorless oil, contaminated with partly-reduced starting material, possible aldehyde peaks present. ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 5.63 (4H, m, 2-H, 3-H, 7-H, 8-H), 4.09 (4H, m, 1-H, 9-H), 2.51 (2H, br. s., 2x OH), 2.08 (4H, m, 4-H, 6-H), 1.49 (2H, m, 5-H); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 133.1 (C-2, C-8), 129.3 (C-3, C-7), 63.9 (C-1, C-9), 31.8 (C-4, C-6), 28.3 (C-5); LC-MS (ESI): *m/z* calc. 156 ($C_9H_{16}O_2$), found $[M+H]^+$ 157.

6.2.1.8 (2E,7E)-Nona-2,7-dienedial (20)[180]



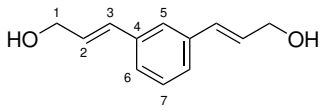
To a solution of the **19** (100 mg, 0.64 mmol) in DMSO (40 mL) a solution of 2-iodoxybenzoic acid (IBX) (896 mg, 3.20 mmol) in DMSO (40 mL) was added and the reaction was stirred for 1 h at rt. H_2O (80 mL) was added and the resulting cloudy suspension was filtered through a coarse fritted funnel and washed with Et_2O (50 mL). The aqueous phase was extracted with Et_2O (3 x 40 mL) and the combined organic phases were washed with brine (60 mL), dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure after filtration. The reaction crude was purified by column chromatography (Hexanes/Et_2O 2:1 1:2). To remove the remaining IBX, the crude product was dissolved in diethyl ether, washed with $KHCO_3$, dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure after filtration to afford **20** (57 mg, 0.38 mmol, 59%). ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 9.53 (2H, d, $J = 7.8$ Hz, 1-H, 9-H), 6.83 (2H, dt, $J = 15.7$, 6.7 Hz, 3-H, 7-H), 6.15 (2H, ddt, $J = 15.7$, 7.8, 1.5 Hz 4-H, 6-H), 2.41 (4H, m, 2-H, 8-H), 1.76 (2H, p, $J = 7.5$ Hz, 5-H); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 193.9 (C-1, C-9), 156.9 (C-3, C-7), 133.7 (C-2, C-8), 32.1 (C-4, C-6), 26.1 (C-5); LC-MS (ESI): *m/z* calc. 152 ($C_9H_{12}O_2$), found $[M+H]^+$ 153.

6.2.1.9 Diethyl 3,3'-(1,3-phenylene)(2E,2'E)-diacrylate (23)[180]



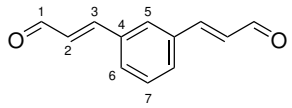
To a solution of isophthalaldehyde (**22**) (1.34 g, 10.0 mmol) in CH_2Cl_2 (100 mL) was added **11** (8.71 g, 25.0 mmol) and the reaction was stirred for 17 h at rt. The mixture was filtered through Celite^{TM} and washed with CH_2Cl_2 (50 mL). The solvent was evaporated under reduced pressure and the residue was treated with Et_2O to allow precipitation of Ph_3PO. After filtration and solvent evaporation, the crude product was purified by flash chromatography (Et_2O/EtOAc 1:4 1:1), affording **23** (1.95 g, 7.1 mmol, 71%). ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 7.68 (1H, m, 7-H), 7.66 (2H, m, 3-H), 7.53 (2H, m, 6-H), 7.41 (1H, m, 5-H), 6.46 (2H, m, 2-H), 4.27 (4H, q, $J = 7.1$ Hz, CH_2OCO), 1.34 (6H, t, $J = 7.1$ Hz, CH_3CH_2O); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 166.8 (C-1), 143.8 (C-3), 135.3 (C-4), 129.6 (C-7), 129.5 (C-6), 127.7 (C-5), 119.4 (C-2), 60.8 (CH_2OCO), 14.4 (CH_3CH_2O); LC-MS (ESI): *m/z* calc. 274 ($C_{16}H_{18}O_4$), found $[M+H]^+$ 275.

6.2.1.10 (2E,2'E)-3,3'-(1,3-phenylene)bis(prop-2-en-1-ol) (24)[180]



To a stirring solution of **23** (1.00 g, 3.65 mmol) in dry THF (100 mL), DIBAL-H (1 M in THF), (22.0 mL, 22.0 mmol), was added under argon atmosphere at -78 °C, and the reaction was stirred for 1.5 h. The mixture was allowed to warm to 0 °C and stirred for 1 h. After dropwise addition of MeOH (5 mL) and a saturated solution of potassium sodium tartrate (50 mL) a white emulsion was formed. The phases were separated after 2 h of stirring. The aqueous phase was extracted with Et_2O (3 x 10 mL). The combined organic phases were dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure after filtration. The crude product was purified by column chromatography (Hexanes/EtOAc 3:2 1:1) to afford **24** (457 mg, 2.41 mmol, 66%). ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 7.39 (1H, m, 7-H), 7.28 (3H, m, 5-H, 6-H), 6.62 (2H, dt, $J =$ 15.9, 1.6 Hz, 3-H), 6.38 (2H, dt, $J = 15.9$, 5.7 Hz, 2-H), 4.34 (4H, dd, $J = 5.7$, 1.6 Hz, 1-H); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 137.1 (C-4), 130.9 (C-3), 129.0 (C-2), 129.0 (C-7), 125.9 (C-6), 124.9 (C-5), 63.8 (C-1); LC-MS (ESI): *m/z* calc. 190 ($C_{12}H_{14}O_2$), found $[M-H]^-$ 189.

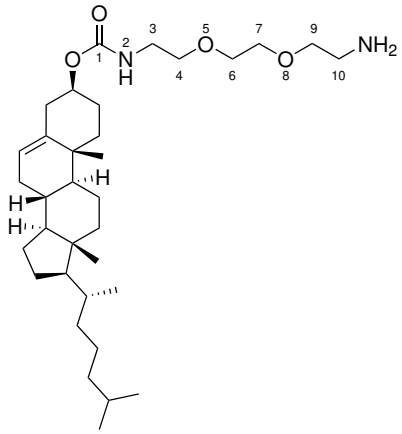
6.2.1.11 (2E,2'E)-3,3'-(1,3-phenylene)diacrylaldehyde (21)[180]



To a solution of **24** (400 mg, 2.10 mmol) in DMSO (40 mL) a solution of 2-iodoxybenzoic acid (IBX) (6.53 g, 10.5 mmol) in DMSO (40 mL) was added and the reaction was stirred for 1 h at rt. H_2O (80 mL) was added and the resulting cloudy suspension was filtered through a coarse fritted funnel and washed with Et_2O (50 mL). The aqueous phase was extracted with Et_2O (3 x 40 mL) and the combined organic phases were washed with brine (60 mL) and $KHCO_3$ (60 mL), dried over anhydrous $MgSO_4$ and the solvent was removed under reduced pressure after filtration. The crude reaction mixture was purified by column chromatography (Hexanes/Et_2O 3:2 l:1) to afford **21** (143 mg, 0.77 mmol, assuming 50% is IBX, 18% yield). ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 9.75 (2H, d, J = 7.6 Hz, 1-H), 7.72 (1H, m, 5-H), 7.65 (3H, m, 3-H, 7-H), 7.53 (2H, m, 6-H), 6.77 (2H, dd, J = 16.0, 7.6 Hz, 2-H); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 193.9 (C-1), 151.7 (C-3), 135.0 (C-4), 129.6 (C-2), 129.5 (C-7), 128.8 (C-6), 128.6 (C-5); LC-MS (ESI): m/z calc. 186 ($C_{12}H_{10}O_2$), found $[M+H]^+$ 187.

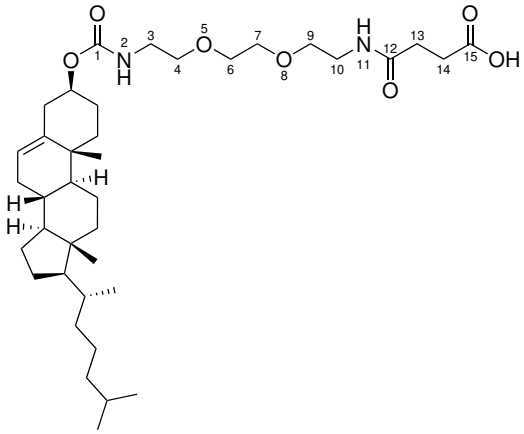
6.2.2 Cholesterol Linker Synthesis

6.2.2.1 Cholesteryl (2-(2-(2-aminoethoxy)ethoxy)ethyl)carbamate (48)[207]



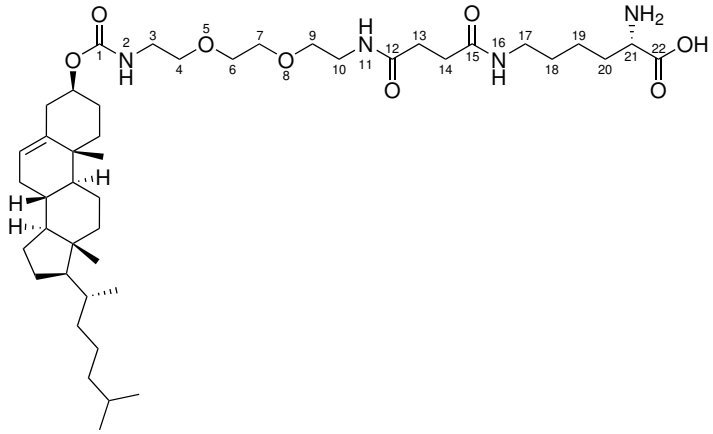
To a solution of 1,8-diamino-3,6-dioxaoctane (2.67 g, 18.0 mmol) and TEA (2.5 mL, 18.0 mmol) in dry CH_2Cl_2 (25 mL) a solution of cholesteryl chloroformate (4.04 g, 9.0 mmol) in dry CH_2Cl_2 (50 mL) was added and the reaction was stirred for 4 h at rt. The voltailes were removed under reduced pressure and the crude product was purified by column chromatography ($CHCl_3$/MeOH 19:1 9:1) to afford **48** (4.48 g, 8.1 mmol, 90%) as an off-white gum. ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 5.37 (1H, m, olefinic H), 5.32 (1H, br. s, NH), 4.49 (1H, m, OC*H* of Cholesteryl), 3.63 (4H, s, 6-H, 7-H), 3.56 (4H, m, 4-H, 9-H), 3.49 (2H, s, NH_2), 3.37 (2H, m, 3-H), 2.93 (2H, m, 10-H), 0.92-2.37 (28H 6 x CH, 11 x CH_2 cholesteryl), 1.01 (3H, s, CH_3 cholesteryl), 0.91 (3H, d, $J = 6.6$ Hz, CH_3CH cholesteryl), 0.86 (6H, dd, $J = 6.6$, 3.2 Hz, 2 x CH_3), 0.67 (3H, s, CH_3 cholesteryl); ^{13}C NMR (176 MHz, $CDCl_3$) 156.4 (C-1), 140.0 (Olefinic C), 122.6 (Olefinic C), 74.5 (O*C*H of cholesteryl), 72.7 (C-9), 70.4 (C-6), 70.3 (C-7), 56.8 (C-4), 56.3, 51.0, 50.2, 42.5, 41.6, 40.8 (C-10), 39.9 (C-3), 39.7, 38.7, 37.1, 36.7, 36.3, 35.9, 32.0, 32.0, 28.4, 28.3, 28.2, 24.4, 24.0, 23.0, 22.7, 21.2, 19.5, 18.9, 12.0; MS-ESI: m/z calc. 560 ($C_{34}H_{60}N_2O_4$), found $[M+H]^+$ 561.

6.2.2.2 1-Cholesteryloxy-1,12-dioxo-5,8-dioxa-2,11-diazapentadecan-15-oic acid (50)[208]



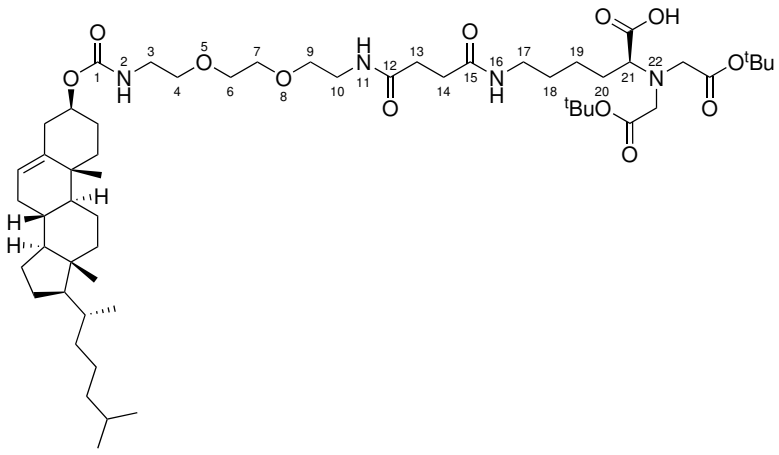
To a stirring solution of compound **48** (7.00 g, 12.5 mmol) in dry THF (5 mL), a solution of succinic anhydride (2.5 g, 25.0 mmol) in dry THF (5 mL) was added dropwise and the reaction was stirred for 3 h at rt under an argon atmosphere. The solvent was removed under vacuum and the residue was redissolved in ethyl acetate (50 mL). The organic layer was then washed successively with water and brine (50 mL each), dried over anhydrous Na_2SO_4 and the solvent removed *in vacuo*. The crude product was then purified by column chromatography ($CHCl_3$/MeOH 19:1 9:1) to afford **50** (6.25 g, 9.5 mmol, 76%) as an off-white gum. ^{1}H NMR (700 MHz, $CDCl_3$) δ ppm 7.40 (1H, s, OH), 7.12 (1H, s, 11-H), 6.57 (1H, s, 2-H), 5.37 (1H, s, olefinic H), 4.54 (1H, m, OC*H* cholesteryl), 3.62 (4H, m, 6-H, 7-H), 3.56 (4H, m, 4-H, 9-H), 3.46 (2H, m, 10-H), 3.38 (2H, m, 3-H), 2.67 (4H, m, 13-H, 14-H), 2.51 (2H, m, 2x CH_AH$_B$ cholesteryl), 2.33 (2H, m, 2x CH_AH$_B$ cholesteryl), 1.98 (2H, m, 2x CH_AH$_B$ cholesteryl), 1.84 (4H, m, 4x CH_AH$_B$ cholesteryl), 1.57 (5H, m, 4x CH_AH$_B$, CH cholesteryl), 1.35 (3H, m, 2x CH_AH$_B$, CH cholesteryl), 1.26 (1H, m, CH cholesteryl), 1.11 (7H, m, 2x CH_AH$_B$, 3x CH cholesteryl), 1.01 (3H, m, CH_3 cholesteryl), 0.96 (2H, m, CH_2 cholesteryl), 0.91 (3H, d, $J = 6.6$ Hz, CH_3CH), 0.86 (6H, dd, $J = 6.6$, 3.2 Hz, 2 x CH_3), 0.68 (3H, s, CH_3 cholesteryl); ^{13}C NMR (176 MHz, $CDCl_3$) δ ppm 175.3 (C-15), 172.8 (C-12), 158.2 (C-1), 139.6 (Olefinic C), 123.0 (Olefinic C), 75.7 (O*C*H of cholesteryl), 70.5 (C-9), 70.3 (2C, C-6, C-7), 69.7 (C-4), 56.8, 56.3, 52.1, 42.4, 41.4 (C-10), 39.9 (C-3), 39.7, 39.4, 38.6, 37.1, 36.7, 36.3, 35.9, 32.0, 32.0, 30.5 (C-14), 28.9 (C-13), 28.7, 28.4, 28.2, 24.4, 24.0, 23.0 (CH_3), 22.7 (CH_3), 21.2, 19.5, 18.9, 12.0; MS-ESI: *m/z* calc. 660 ($C_{38}H_{64}N_2O_7$), found $[M+H]^+$ 661.

6.2.2.3 (S)-21-Amino-1-cholesteryloxy-1,12,15-trioxo-5,8-dioxa-2,11,16-triazadocosan-22-oic acid (55)[210]



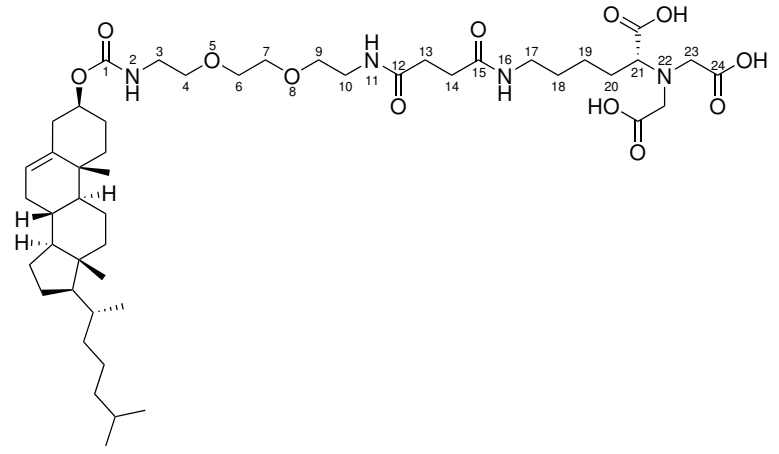
(i) A solution of Boc-lysine(Fmoc)-OH (**59**) (1.65 g, 3.53 mmol) and DIPEA (2.5 mL, 14.1 mmol) in dry CH_2Cl_2 (20 mL) was added to 2-Cl Trt resin (2.00 g, 1.47 mmol/g loading) and stirred for 2 h. The unreacted resin was then capped for 5 minutes with CH_2Cl_2/MeOH/DIPEA (17:2:1) before being washed with 3 x CH_2Cl_2, 2 x DMF, 2 x CH_2Cl_2, 2 x MeOH, 1 x MeCN (1 x = 10 mL) and dried *in vacuo* overnight. The resin was swelled for 30 minutes using CH_2Cl_2 before being drained for the next reaction. (ii) 5 mL of 40% piperidine in DMF was applied to the resin, stirred for 15 minutes and washed through. This was repeated until the flow-through was no longer UV active. (iii) A solution of **50** (1.08 g, 1.64 mmol), HATU (624 mg, 1.64 mmol) and DIPEA (286 μL, 1.64 mmol) in dry DMF (10 mL) were added to the resin and stirred for 2 h. The resin was then washed with 3 x CH_2Cl_2, 3 x MeOH, 3 x ether and dried *in vacuo*. (iv) A 25 mL solution of TFA (25%), H_2O (2.5%) and TIPS (1%) in CH_2Cl_2 was applied to the resin and stirred for 45 minutes. The resin was then drained and the process repeated. The two flow throughs were then combined and added to cold diethyl ether. The product was then isolated using centrifugation to yield **55** (776 mg, 3.46 mmol, 98%). MS-ESI: *m/z* calc. 789 ($C_{44}H_{76}N_4O_8$), found $[M+H]^+$ 790.

6.2.2.4 (S)-21-(bis(2-(tert-butoxy)-2-oxoethyl)amino)-1-cholesteryloxy-1,12,15-trioxo-5,8-dioxa-2,11,16-triazadocosan-22-oic acid (68)[212]



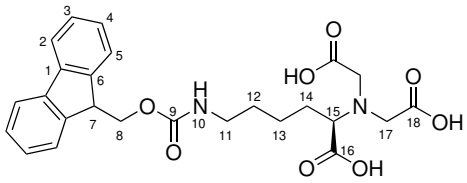
To an ice-cooled solution of tert-butyl bromoacetate (124 µL, 0.84 mmol) and DIPEA (183 µL, 1.05 mmol) in $CHCl_3$ (10 mL) was added dropwise **55** (163 mg, 0.21 mmol) and DIPEA (183 µL, 1.05 mmol) in $CHCl_3$ (10 mL). After cooling for 1 h, the solution was stirred at 50 °C overnight. The volatiles were removed *in vacuo* and the crude product was purified by column chromatography ($CHCl_3$/MeOH 9:1) to afford **57** (55 mg, 0.45 mmol, 54%). MS-ESI: *m/z* calc. 1017 ($C_{56}H_{96}N_4O_{12}$), found $[M+H]^+$ 1018.

6.2.2.5 (S)-21-carboxy-22-(carboxymethyl)-1-((Cholesteryl)oxy)-1,12,15-trioxo-5,8-dioxa-2,11,16,22-tetraazatetracosan-24-oic acid (57)[212]



68 (50 mg, 0.05 mmol) was treated with TFA in CH_2Cl_2 (1:3, 4 mL) for 2 h before the volatiles were removed *in vacuo* to afford **57** (34 mg, 0.04 mmol, 76%) as a brown oil. MS-ESI: *m/z* calc. 905 ($C_{48}H_{80}N_4O_{12}$), found $[M-H]^-$ 904.

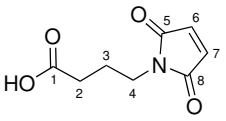
6.2.2.6 (R)-2,2'-((4-((Fmoc)amino)-1-carboxybutyl)azanediyl)diacetic acid (64)[212]



i) Boc-Lysine(Fmoc)-OH (**59**) (500 mg, 1.0 mmol) was treated with TFA in $CHCl_3$ (1:3, 4 mL) for 2 h at room temperature. The volatiles were removed *in vacuo* and the intermediate was dissolved in CH_2Cl_2 and washed with saturated aqueous Na_2CO_3 and brine. The organic layer was dried with anhydrous Mg_2SO_4 and concentrated *in vacuo*. ii) To an ice-cooled solution of bromoacetic acid (277 mg, 2.0 mmol) and Et_3N (303 μL, 2.2 mmol) in $CHCl_3$ (10 mL) was added dropwise a solution of the intermediate and Et_3N (394 μL, 2.8 mmol) in $CHCl_3$ (20 mL). After cooling for 1 h the solution was stirred at room temperature over-night and then heated to 50 °C for 2 h. The solution containing **64** was extracted into $CoCl_2$ (1 M, 30 mL) to give **65** (amount of purified product could not be obtained as the reaction was extracted directly with aqueous $CoCl_2$).

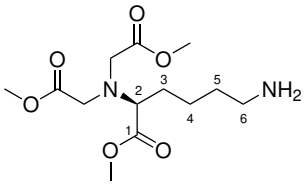
6.2.3 DNA Linker Synthesis

6.2.3.1 4-Maleimidobutyric acid[230] (79)



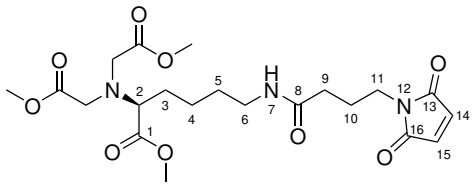
To a solution of maleic anhydride (0.98 g, 10.0 mmol) in glacial acetic acid (10 mL) was added 4-aminobutyric acid (1.03 g, 10.0 mmol) under argon with vigorous stirring. The resulting slurry was stirred for 8 hours at room temperature. The white precipitate was collected by filtration, and washed with H_2O (10 mL) twice. The white solid was dispersed in H_2O (20 mL) and heated to reflux for 30 minutes. H_2O was removed *in vacuo* to afford **79** (1.23 g, 6.7 mmol, 67%) as a white powder. ^{1}H NMR (700 MHz, DMSO-d_6) δ ppm 7.70 (1H, br. s., COOH), 6.03 (2H, s, 6-H, 7-H), 2.81 (2H, t, $J = 7.4$ Hz, 4-H), 2.31 (2H, t, $J = 7.4$ Hz, 2-H), 1.74 (2H, p, $J = 15.0, 7.4$ Hz, 3-H); ^{13}C NMR (176 MHz, DMSO-d_6) δ ppm 173.8 (C-1), 167.4 (C-5, C-8), 136.1 (C-6, C-7), 38.3 (C-4), 30.4 (C-2), 22.5 (C-3); MS-ESI: m/z calc. 183 ($C_8H_9NO_4$), found $[M+H_3O]^+$ 202.

6.2.3.2 Dimethyl 2,2'-((6-amino-1-methoxy-1-oxohexan-2-yl)azanediyl)(S)-diacetate[231] (81)



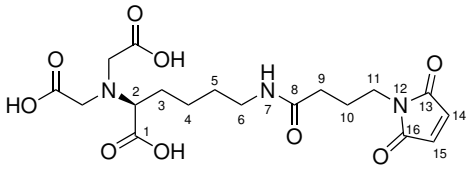
N_α, N_α-Bis(carboxymethyl)-L-lysine (100 mg, 0.43 mmol) was dissolved in methanol (30 mL) and the solution cooled to 0 °C. Thionyl chloride (623 µL, 8.58 mmol) was added in a dropwise manner and the mixture was stirred for 48 hours under reflux at 55 °C. The solvent was removed *in vacuo* to afford **81** with an assumed yield of 100% and was used in the next step without further purification. ^{1}H NMR (700 MHz, D_2O) δ ppm 3.75 (4H, m, CH_2N), 3.74 (3H, s, CH_3O), 3.72 (6H, s, CH_3OCO), 3.58 (1H, t, $J = 7.6$ Hz, 2-H), 2.99 (2H, t, $J = 7.6$ Hz, 6-H),1.74 (4H, m, 3-H, 5-H), 1.43 (2H, m, 4-H); ^{13}C NMR (176 MHz, D_2O) δ ppm 175.3 (C-1), 174.4 (COO), 65.2 (C-2), 53.3 (CH_2N), 52.9 (CH_3OCO), 52.7 (CH_3O), 39.8 (C-6), 29.3 (C-5), 26.9 (C-3), 22.8 (C-4); MS-ESI: m/z calc. 304 ($C_{13}H_{24}N_2O_6$), found $[M+H]^+$ 305.

6.2.3.3 (S)-2,2'-((1-carboxy-5-(4-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)butanamido)pentyl)azanediyl)diacetic acid[230] (82)



To a solution of 4-maleimidobutyric acid (**79**)(163 mg, 0.89 mmol) in DMF (2 mL) was added HATU (677 mg, 1.78 mmol). After 5 minutes, a solution of DIPEA (620 μL, 3.56 mmol) and **81** (353 mg, 1.16 mmol) in DMF (1 mL) and CH_2Cl_2 (4 mL) was added, and the reaction mixture was stirred for 2 hours before dilution with CH_2Cl_2 (50 mL). The organic phase was washed with 5% acetic acid (20 mL), water (20 mL) and brine (20 mL), dried over anhydrous $MgSO_4$ and the solvent removed *in vacuo*. The reaction crude was purified by column chromatography (CH_2Cl_2/MeOH 7:3 0:1) to afford **82** (65 mg, 0.13 mmol, 15%). ^{1}H NMR (700 MHz, D_2O) δ ppm 6.28 (2H, s, 14-H, 15-H), 3.73 (3H, s, CH_3O), 3.72 (6H, s, CH_3OCO), 3.70 (4H, m, CH_2N), 3.52 (1H, m, 2-H), 3.25 (2H, m, 11-H), 3.17 (2H, m, 6-H), 2.28 (2H, m, 9-H), 1.81 (2H, m, 10-H), 1.70 (2H, m, 3-H), 1.52 (2H, m, 5-H), 1.33 (2H, m, 4-H); ^{13}C NMR (176 MHz, D_2O) δ ppm 174.7, 135.9 (C-13, C-16), 53.3 (CH_2N), 52.8 (CH_3O), 39.6 (C-11), 39.0 (C-6), 33.7 (C-9), 29.6 (C-5), 28.5 (C-3), 23.2 (C-10), 23.2 (C-4) (a more concentrated sample would be required to fully assign the carbonyl region of this compound); MS-ESI: *m/z* calc. 469 ($C_{21}H_{31}N_3O_9$), found $[M+H_3O]^+$ 488.

6.2.3.4 (S)-2,2'-((1-carboxy-5-(4-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)butanamido)pentyl)azanediyl)diacetic acid[230] (76)



To a solution of $LiOH{\cdot}H_2O$ (58.7 mg, 1.40 mmol) in MeOH (5 mL) and H_2O (5 mL) was added **82** (64.5 mg, 0.14 mmol) and the reaction mixture was stirred for 48 hours. Resin (Dowex® 50Wx8 hydrogen form) was used to remove Li^+ ions by adjusting pH to 6, then the solvent was removed *in vacuo* and water was removed by lyophilization to afford **76** (49.2 mg, 0.11 mmol, 82% yield). ^{1}H NMR (700 MHz, D_2O) δ ppm 6.36 (3H, s, 14-H, 15-H), 3.90 (4H, m, CH_2N), 3.24 (4H, m, 11-H, 6-H), 3.04 (1H, m, 2-H) 2.34 (2H, m, 9-H), 1.97 (2H, m, 10-H), 1.86 (2H, m, 3-H), 1.58 (2H, m, 5-H), 1.31 (2H, m, 4-H); ^{13}C NMR (176 MHz, D_2O) δ ppm 173.4 (C-1), 171.2 (*C*OOH), 170.0 (C-13, C-16), 167.3 (C-8), 132.1 (C-14, C-15), 69.0 (C-2), 61.5 (CH_2N), 39.6 (C-11), 38.6 (C-6), 35.1 (C-9), 31.6 (C-5), 29.5 (C-3), 27.3 (C-10), 23.7 (C-4); MS-ESI: *m/z* calc. 427 ($C_{18}H_{25}N_3O_9$), found $[M+H_3O]^+$ 446.

7 References

1 - P. Kareiva and V. Carranza, *Futures*, 2018, **102**, 39–50.

2 - The Origins of EPA, *https://www.epa.gov/history/origins-epa, (Accessed April 2020)*.

3 - Who we are — European Environment Agency, *https://www.eea.europa.eu/about-us/who, (Accessed April 2020)*.

4 - M. van der Van, Y. Kobayashi and R. Diercks, *Technology Roadmap: Energy and GHG Reductions in the Chemical Industry via Catalytic Processes*, 2013.

5 - P. T. Anastas and J. C. Warner, *Green Chemistry: Theory and Practice*, Oxford University Press, Inc., Oxford, 1998, pp. 1–135.

6 - P. T. Anastas, M. M. Kirchhoff and T. C. Williamson, *Applied Catalysis A: General*, 2001, **221**, 3–13.

7 - R. A. Sheldon, I. W. C. E. Arends and U. Hanefeld, in *Green Chemistry and Catalysis*, Wiley-VCH Verlag GmbH & Co. KGaA, 2007, ch. 1, pp. 1–47.

8 - J. J. Verendel, *Ph.D. thesis*, Uppsala universitet, Uppsala, 2012.

9 - E. B. Bauer, *Chemical Society Reviews*, 2012, **41**, 3153–3167.

10 - A. Pfaltz and W. J. Drury, *Proceedings of the National Academy of Sciences*, 2004, **101**, 5723–5726.

11 - T. Masuch, A. Kusnezowa, S. Nilewski, J. T. Bautista, R. Kourist and L. I. Leichert, *Frontiers in Microbiology*, 2015, **6**, 1110.

12 - R. Woodyer, W. Chen and H. Zhao, *Journal of Chemical Education*, 2004, **81**, 126.

13 - D. Voet, J. G. Voet and C. W. Pratt, in *Principles of Biochemistry*, John Wiley & Sons, Inc, Singapore, 4th edn., 2013, ch. 11, pp. 315–370.

14 - C. M. Clouthier and J. N. Pelletier, *Chemical Society Reviews*, 2012, **41**, 1585–1605.

15 - R. A. Sheldon and D. Brady, *Chemical Communications*, 2018, **54**, 6088–6104.

16 - L. Jeske, S. Placzek, I. Schomburg, A. Chang and D. Schomburg, *Nucleic Acids Research*, 2019, **47**, D542–D549.

17 - A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts and B. Witholt, *Nature*, 2001, **409**, 258–268.

18 - A. Kamble, S. Srinivasan and H. Singh, *Molecular Biotechnology*, 2019, **61**, 53–59.

19 - D. Toyama, M. A. B. de Morais, F. C. Ramos, L. M. Zanphorlin, C. C. C. Tonoli, A. F. Balula, F. P. de Miranda, V. M. Almeida, S. R. Marana, R. Ruller, M. T. Murakami and F. Henrique-Silva, *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 2018, **1866**, 569–579.

20 - F. H. Arnold, *Angewandte Chemie - International Edition*, 2019, **58**, 14420–14426.

21 - P. S. Coelho, Z. J. Wang, M. E. Ener, S. A. Baril, A. Kannan, F. H. Arnold and E. M. Brustad, *Nature Chemical Biology*, 2013, **9**, 485–487.

22 - S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius and U. T. Bornscheuer, *Angewandte Chemie International Edition*, 2021, **60**, 88–119.

23 - N. J. Turner and E. O'Reilly, *Nature Chemical Biology*, 2013, **9**, 285–288.

24 - A. P. Green and N. J. Turner, *Perspectives in Science*, 2016, **9**, 42–48.

25 - R. O. M. A. de Souza, L. S. M. Miranda and U. T. Bornscheuer, *Chemistry - A European Journal*, 2017, **23**, 12040–12063.

26 - N. J. Turner and L. Humphreys, *Biocatalysis in Organic Synthesis*, The Royal Society of Chemistry, 2018.

27 - W. D. Fessner, *New Biotechnology*, 2015, **32**, 658–664.

28 - I. Oroz-Guinea and E. García-Junceda, *Current Opinion in Chemical Biology*, 2013, **17**, 236–249.

29 - W. A. Duetz, J. B. Van Beilen and B. Witholt, *Current Opinion in Biotechnology*, 2001, **12**, 419–425.

30 - H. Jiang, K. V. Wood and J. A. Morgan, *Applied and Environmental Microbiology*, 2005, **71**, 2962–2969.

31 - S. Atsumi, T. Hanai and J. C. Liao, *Nature*, 2008, **451**, 86–89.

32 - S. Wu and Z. Li, *ChemCatChem*, 2018, **10**, 2164–2178.

33 - A. Alissandratos, *Biophysical Reviews*, 2020, **12**, 175.

34 - Cell Lysis & Disruption Market Analysis Report, *https://www.grandviewresearch.com/industry-analysis/cell-lysis-dissociation-market (Accessed November 2021)*.

35 - B. R. Lichman, E. D. Lamming, T. Pesnot, J. M. Smith, H. C. Hailes and J. M. Ward, *Green Chem.*, 2015, **17**, 852–855.

36 - W. Finnigan, R. Cutlan, R. Snajdrova, J. P. Adams, J. A. Littlechild and N. J. Harmer, *ChemCatChem*, 2019, **11**, 3474–3489.

37 - M. Bawn, F. Subrizi, G. J. Lye, T. D. Sheppard, H. C. Hailes and J. M. Ward, *Enzyme and Microbial Technology*, 2018, **116**, 16–22.

38 - H. K. Philpott, P. J. Thomas, D. Tew, D. E. Fuerst and S. L. Lovelock, *Green Chemistry*, 2018, **20**, 3426–3431.

39 - A. S. Khalil and J. J. Collins, *Nature Reviews Genetics*, 2010, **11**, 367–379.

40 - C. M. Agapakis, P. M. Boyle and P. A. Silver, *Nature Chemical Biology*, 2012, **8**, 527–535.

41 - D. Voet, J. G. Voet and C. W. Pratt, in *Principles of Biochemistry*, John Wiley & Sons, Inc, Singapore, 4th edn., 2013, ch. 16.4, pp. 538–545.

42 - F. Wu, L. N. Pelster and S. D. Minteer, *Chemical Communications*, 2015, **51**, 1244–1247.

43 - B. C. Buddingh' and J. C. Van Hest, *Accounts of Chemical Research*, 2017, **50**, 769–777.

44 - J. Shen, Y. Zeng, X. Zhuang, L. Sun, X. Yao, P. Pimpl and L. Jiang, *Molecular Plant*, 2013, **6**, 1419–1437.

45 - H. Lee, W. C. DeLoache and J. E. Dueber, *Metabolic Engineering*, 2012, **14**, 242–251.

46 - C. A. Kerfeld, S. Heinhorst and G. C. Cannon, *Annual review of microbiology*, 2010, **64**, 391–408.

47 - R. J. Conrado, J. D. Varner and M. P. DeLisa, *Current Opinion in Biotechnology*, 2008, **19**, 492–499.

48 - A. M. Hill and J. Staunton, in *Comprehensive Natural Products II*, ed. L. Mander and H.-W. B. Liu, Elsevier Ltd., 2010, ch. 1.10, pp. 385–452.

49 - S. Raboni, S. Bettati and A. Mozzarelli, *Cellular and Molecular Life Sciences*, 2009, **66**, 2391–2403.

50 - D. Voet, J. G. Voet and C. W. Pratt, in *Principles of Biochemistry*, John Wiley & Sons, Inc, Singapore, 4th edn., 2013, ch. 9.1, pp. 242–254.

51 - G. Van Meer and A. I. De Kroon, *Journal of Cell Science*, 2011, **124**, 5–8.

52 - B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, in *Molecular Biology of the Cell, 4th Edition*, Garland Science, New York, 2002, ch. 12, pp. 641–694.

53 - F. Hullin-Matsuda, T. Taguchi, P. Greimel and T. Kobayashi, *Seminars in Cell and Developmental Biology*, 2014, **31**, 48–56.

54 - Metro Map of Metabolism - The Overview, *https://www.behance.net/gallery/38270165/Metro-Map-of-Metabolism-The-Overview, (Accessed April 2020)*.

55 - A. Komeili, H. Vali, T. J. Beveridge and D. K. Newman, *Proceedings of the National Academy of Sciences of the United States of America*, 2004, **101**, 3839–3844.

56 - H. Nudelman, Y. Z. Lee, Y. L. Hung, S. Kolusheva, A. Upcher, Y. C. Chen, J. Y. Chen, S. C. Sue and R. Zarivach, *Frontiers in Microbiology*, 2018, **9:2480**, 1–14.

57 - J. T. Beatty, J. Overmann, M. T. Lince, A. K. Manske, A. S. Lang, R. E. Blankenship, C. L. Van Dover, T. A. Martinson and F. G. Plumley, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 9306–9310.

58 - G. Bozzuto and A. Molinari, *International Journal of Nanomedicine*, 2015, **10**, 975–999.

59 - J. M. Berg, J. L. Tymoczko and L. Stryer, in *Biochemistry. 5th edition.*, W. H. Freeman, New York, 5th edn., 2002, ch. 12.3.

60 - T. Trantidou, M. Friddin, Y. Elani, N. J. Brooks, R. V. Law, J. M. Seddon and O. Ces, *ACS Nano*, 2017, **11**, 6549–6565.

61 - C. Schmitt, A. H. Lippert, N. Bonakdar, V. Sandoghdar and L. M. Voll, *Frontiers in Bioengineering and Biotechnology*, 2016, **4**, 19.

62 - T. Anajafi and S. Mallik, *Therapeutic Delivery*, 2015, **6**, 521–534.

63 - E. Rideau, R. Dimova, P. Schwille, F. R. Wurm and K. Landfester, *Chemical Society Reviews*, 2018, **47**, 8572–8610.

64 - L. Messager, J. R. Burns, J. Kim, D. Cecchin, J. Hindley, A. L. Pyne, J. Gaitzsch, G. Battaglia and S. Howorka, *Angewandte Chemie - International Edition*, 2016, **55**, 11106–11109.

65 - S. A. Walker, M. T. Kennedy and J. A. Zasadzinski, *Nature*, 1997, **387**, 61–64.

66 - C. A. Kerfeld, M. R. Sawaya, S. Tanaka, C. V. Nguyen, M. Philips, M. Beeby and T. O. Yeates, *Science*, 2005, **309**, 936–938.

67 - J. M. Shively, F. Ball, D. H. Brown and R. E. Saunders, *Science*, 1973, **182**, 584–586.

68 - Z. Dou, S. Heinhorst, E. B. Williams, C. D. Murin, J. M. Shively and G. C. Cannon, *Journal of Biological Chemistry*, 2008, **283**, 10377–10384.

69 - S. D. Axen, O. Erbilgin and C. A. Kerfeld, *PLoS Computational Biology*, 2014, **10**, e1003898.

70 - C. Chowdhury, S. Chun, A. Pang, M. R. Sawaya, S. Sinha, T. O. Yeates and T. A. Bobik, *Proceedings of the National Academy of Sciences*, 2015, **112**, 2990–2995.

71 - M. G. Klein, P. Zwart, S. C. Bagby, F. Cai, S. W. Chisholm, S. Heinhorst, G. C. Cannon and C. A. Kerfeld, *Journal of Molecular Biology*, 2009, **392**, 319–333.

72 - I. Stojiljkovic, A. J. Baumler and F. Heffron, *Journal of Bacteriology*, 1995, **177**, 1357–1366.

73 - W. H. Roos, I. L. Ivanovska, A. Evilevitch and G. J. L. Wuite, *Cellular and Molecular Life Sciences*, 2007, **64**, 1484–1497.

74 - W.-S. Ryu, in *Molecular Virology of Human Pathogenic Viruses*, Academic Press, London, 2017, pp. 21–29.

75 - M. G. Mateu, *Archives of Biochemistry and Biophysics*, 2013, **531**, 65–79.

76 - W. R. Wikoff, L. Liljas, R. L. Duda, H. Tsuruta, R. W. Hendrix and J. E. Johnson, *Science*, 2000, **289**, 2129–2133.

77 - M. Sutter, D. Boehringer, S. Gutmann, S. Günther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban and N. Ban, *Nature Structural and Molecular Biology*, 2008, **15**, 939–947.

78 - A. N. Gabashvili, N. S. Chmelyuk, M. V. Efremova, J. A. Malinovskaya, A. S. Semkina and M. A. Abakumov, *Biomolecules*, 2020, **10**, 1–12.

79 - T. W. Giessen and P. A. Silver, *Nature Microbiology*, 2017, **2**, 17029.

80 - J. B. Parsons, S. Frank, D. Bhella, M. Liang, M. B. Prentice, D. P. Mulvihill and M. J. Warren, *Molecular Cell*, 2010, **38**, 305–315.

81 - C. A. Kerfeld and M. R. Melnicki, *Current Opinion in Plant Biology*, 2016, **31**, 66–75.

82 - H. Wang, X. Yan, H. Aigner, A. Bracher, N. D. Nguyen, W. Y. Hee, B. M. Long, G. D. Price, F. U. Hartl and M. Hayer-Hartl, *Nature*, 2019, **566**, 131–135.

83 - C. Fan, S. Cheng, Y. Liu, C. M. Escobar, C. S. Crowley, R. E. Jefferson, T. O. Yeates and T. A. Bobik, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 7509–7514.

84 - C. M. Jakobson, M. F. Slininger Lee and D. Tullman-Ercek, *Protein Science*, 2017, **26**, 1086–1092.

85 - H. J. Wagner, C. C. Capitain, K. Richter, M. Nessling and J. Mampel, *Engineering in Life Sciences*, 2017, **17**, 36–46.

86 - S. Planamente and S. Frank, *Biochemical Society Transactions*, 2019, **47**, 765–777.

87 - D. P. Patterson, P. E. Prevelige and T. Douglas, *ACS Nano*, 2012, **6**, 5000–5009.

88 - M. Comellas-Aragonès, H. Engelkamp, V. I. Claessen, N. A. Sommerdijk, A. E. Rowan, P. C. Christianen, J. C. Maan, B. J. Verduin, J. J. Cornelissen and R. J. Nolte, *Nature Nanotechnology*, 2007, **2**, 635–639.

89 - I. J. Minten, L. J. Hendriks, R. J. Nolte and J. J. Cornelissen, *Journal of the American Chemical Society*, 2009, **131**, 17771–17773.

90 - M. Brasch, R. M. Putri, M. V. De Ruiter, D. Luque, M. S. Koay, J. R. Castón and J. J. Cornelissen, *Journal of the American Chemical Society*, 2017, **139**, 1512–1519.

91 - J. D. Fiedler, S. D. Brown, J. L. Lau and M. G. Finn, *Angewandte Chemie - International Edition*, 2010, **49**, 9648–9651.

92 - T. W. Giessen and P. A. Silver, *ChemBioChem*, 2016, **17**, 1931–1935.

93 - H. Ren, S. Zhu and G. Zheng, *International Journal of Molecular Sciences*, 2019, **20**, 592.

94 - D. P. Patterson, B. Schwarz, R. S. Waters, T. Gedeon and T. Douglas, *ACS Chemical Biology*, 2014, **9**, 359–365.

95 - Y. H. Lau, T. W. Giessen, W. J. Altenburg and P. A. Silver, *Nature Communications*, 2018, **9**, 1311.

96 - F. Sigmund, C. Massner, P. Erdmann, A. Stelzl, H. Rolbieski, M. Desai, S. Bricault, T. P. Wörner, J. Snijder, A. Geerlof, H. Fuchs, M. H. De Angelis, A. J. Heck, A. Jasanoff, V. Ntziachristos, J. Plitzko and G. G. Westmeyer, *Nature Communications*, 2018, **9**, 1–14.

97 - Y. Hsia, J. B. Bale, S. Gonen, D. Shi, W. Sheffler, K. K. Fong, U. Nattermann, C. Xu, P. S. Huang, R. Ravichandran, S. Yi, T. N. Davis, T. Gonen, N. P. King and D. Baker, *Nature*, 2016, **535**, 136–139.

98 - J. Li, A. Johnson-Buck, Y. R. Yang, W. M. Shih, H. Yan and N. G. Walter, *Nature Nanotechnology*, 2018, **13**, 723–729.

99 - N. C. Seeman, *Journal of Theoretical Biology*, 1982, **99**, 237–247.

100 - D. Voet, J. G. Voet and C. W. Pratt, in *Principles of Biochemistry*, 4th edn., 2013, ch. 3, pp. 40–48.

101 - M. Spencer, *Acta Crystallographica*, 1959, **12**, 66–71.

102 - R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini and M. L. Kopka, *Science*, 1982, **216**, 475–485.

103 - D. Voet, J. G. Voet and C. W. Pratt, in *Principles of Biochemistry*, 4th edn., 2013, ch. 24, pp. 821–837.

104 - J. Carlos García-Ramos, R. Galindo-Murillo, F. Cortés-Guzmán and L. Ruiz-Azuara, *Chem. Soc*, 2013, **57**, 245–259.

105 - P. W. Rothemund, *Nature*, 2006, **440**, 297–302.

106 - E. Benson, A. Mohammed, J. Gardell, S. Masich, E. Czeizler, P. Orponen and B. Högberg, *Nature*, 2015, **523**, 441–444.

107 - F. Wang, B. Willner and I. Willner, *Topics in Current Chemistry*, 2014, **354**, 279–338.

108 - R. M. Zadegan, M. D. E. Jepsen, K. E. Thomsen, A. H. Okholm, D. H. Schaffert, E. S. Andersen, V. Birkedal and J. Kjems, *ACS Nano*, 2012, **6**, 10050–10053.

109 - A. C. Pierre, *Biocatalysis and Biotransformation*, 2004, **22**, 145–170.

110 - M. Campás and J.-L. Marty, in *Immobilization of Enzymes and Cells*, ed. J. M. Guisan, Humana Press, Totowa, NJ, 2006, pp. 77–85.

111 - S. Braun, S. Rappoport, R. Zusman, D. Avnir and M. Ottolenghi, *Materials Letters*, 1990, **10**, 1–5.

112 - S. Braun, S. Shtelzer, S. Rappoport, D. Avnir and M. Ottolenghi, *Journal of Non-Crystalline Solids*, 1992, **147-148**, 739–743.

113 - D. Avnir, S. Braun, O. Lev and M. Ottolenghi, *Chemistry of Materials*, 1994, **6**, 1605–1614.

114 - I. Gill, *Chemistry of Materials*, 2001, **13**, 3404–3421.

115 - D. Avnir, O. Lev and J. Livage, *Journal of Materials Chemistry*, 2006, **16**, 1013–1030.

116 - G. J. Owens, R. K. Singh, F. Foroutan, M. Alqaysi, C. M. Han, C. Mahapatra, H. W. Kim and J. C. Knowles, *Progress in Materials Science*, 2016, **77**, 1–79.

117 - E. Gazit, *Chemical Society Reviews*, 2007, **36**, 1263–1269.

118 - S. Scanlon and A. Aggeli, *Nano Today*, 2008, **3**, 22–30.

119 - G. E. Douberly, S. Pan, D. Walters and H. Matsui, *Journal of Physical Chemistry B*, 2001, **105**, 7612–7618.

120 - R. Djalali, Y. F. Chen and H. Matsui, *Journal of the American Chemical Society*, 2003, **125**, 5873–5879.

121 - L. Yu, I. A. Banerjee, X. Gao, N. Nuraje and H. Matsui, *Bioconjugate Chemistry*, 2005, **16**, 1484–1487.

122 - B. W. Park, K. A. Ko, D. Y. Yoon and D. S. Kim, *Enzyme and Microbial Technology*, 2012, **51**, 81–85.

123 - B. W. Park, D. Y. Yoon and D. S. Kim, *Recent progress in bio-sensing techniques with encapsulated enzymes*, 2010.

124 - M. Nasseau, Y. Boublik, W. Meier, M. Winterhalter and D. Fournier, *Biotechnology and Bioengineering*, 2001, **75**, 615–618.

125 - B. Chaize, J. P. Colletier, M. Winterhalter and D. Fournier, *Artificial Cells, Blood Substitutes, and Immobilization Biotechnology*, 2004, **32**, 67–75.

126 - G. Sessa and G. Weissmann, *Journal of Biological Chemistry*, 1970, **245**, 3295–3301.

127 - E. Y. Kim and D. Tullman-Ercek, *Current Opinion in Biotechnology*, 2013, **24**, 627–632.

128 - Z. Zhao, J. Fu, S. Dhakal, A. Johnson-Buck, M. Liu, T. Zhang, N. W. Woodbury, Y. Liu, N. G. Walter and H. Yan, *Nature Communications*, 2016, **7**, 10619.

129 - G. Grossi, M. D. E. Jepsen, J. Kjems and E. S. Andersen, *Nature Communications*, 2017, **8**, 992.

130 - C. Mateo, J. M. Palomo, G. Fernandez-Lorente, J. M. Guisan and R. Fernandez-Lafuente, *Enzyme and Microbial Technology*, 2007, **40**, 1451–1463.

131 - B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, in *Molecular Biology of the Cell*, Garland Science, New York, 4th edn., 2002, ch. 10, pp. 593–612.

132 - Z. Cournia, T. W. Allen, I. Andricioaei, B. Antonny, D. Baum, G. Brannigan, N. V. Buchete, J. T. Deckman, L. Delemotte, C. del Val, R. Friedman, P. Gkeka, H. C. Hege, J. Hénin, M. A. Kasimova, A. Kolocouris, M. L. Klein, S. Khalid, M. J. Lemieux, N. Lindow, M. Roy, J. Selent, M. Tarek, F. Tofoleanu, S. Vanni, S. Urban, D. J. Wales, J. C. Smith and A. N. Bondar, *Journal of Membrane Biology*, 2015, **248**, 611–640.

133 - L. Harvey, B. Arnold, Z. S Lawrence, l. M. Pau, B. David and D. James, in *Molecular Cell Biology. 4th edition*, W. H. Freeman, New York, 2000, ch. 3.4.

134 - M. A. Deeg, D. R. Humphrey, Shun Hua Yang, T. R. Ferguson, V. N. Reinhold and T. L. Rosenberry, *Journal of Biological Chemistry*, 1992, **267**, 18573–18580.

135 - M. G. Paulick and C. R. Bertozzi, *Biochemistry*, 2008, **47**, 6991–7000.

136 - D. I. Udenwobele, R. C. Su, S. V. Good, T. B. Ball, S. V. Shrivastav and A. Shrivastav, *Frontiers in Immunology*, 2017, **8**, 1.

137 - M. Hüttemann, P. Pecina, M. Rainbolt, T. H. Sanderson, V. E. Kagan, L. Samavati, J. W. Doan and I. Lee, *Mitochondrion*, 2011, **11**, 369–381.

138 - P. L. Yeagle, in *The Membranes of Cells*, Academic Press, 3rd edn., 2016, ch. 10, pp. 219–268.

139 - L. Cao, *Carrier-bound Immobilized Enzymes: Principles, Application and Design*, Wiley-VCH Verlag GmbH & Co., 2006, pp. 1–563.

140 - Y. Z. Chen, C. T. Yang, C. B. Ching and R. Xu, *Langmuir*, 2008, **24**, 8877–8884.

141 - R. Fernandez-Lafuente, P. Armisén, P. Sabuquillo, G. Fernández-Lorente and J. M. Guisán, *Chemistry and Physics of Lipids*, 1998, **93**, 185–197.

142 - L. Veum, L. T. Kanerva, P. J. Halling, T. Maschmeyer and U. Hanefeld, *Advanced Synthesis and Catalysis*, 2005, **347**, 1015–1021.

143 - U. Hanefeld, L. Gardossi and E. Magner, *Chem. Soc. Rev.*, 2009, **38**, 453–468.

144 - R. P. Vaz and E. X. F. Filho, in *Applications of Ion Exchange Materials in Biomedical Industries*, Springer International Publishing, 2019, pp. 13–27.

145 - R. A. Sheldon, A. Basso and D. Brady, *Chemical Society Reviews*, 2021, 1.

146 - R. Torres, C. Ortiz, B. C. Pessela, J. M. Palomo, C. Mateo, J. M. Guisán and R. Fernández-Lafuente, *Enzyme and Microbial Technology*, 2006, **39**, 167–171.

147 - S. A. Lauer and J. P. Nolan, *Cytometry*, 2002, **48**, 136–145.

148 - J. Britton, R. P. Dyer, S. Majumdar, C. L. Raston and G. A. Weiss, *Angewandte Chemie International Edition*, 2017, **56**, 2296–2301.

149 - J. Křenková and F. Foret, *Electrophoresis*, 2004, **25**, 3550–3563.

150 - I. Dovgan, S. Ursuegui, S. Erb, C. Michel, S. Kolodych, S. Cianférani and A. Wagner, *Bioconjugate Chemistry*, 2017, **28**, 1452–1457.

151 - M. Haque, N. Forte and J. R. Baker, *Chemical Communications (Cambridge, England)*, 2021, **57**, 10689.

152 - O. Koniev and A. Wagner, *Chemical Society Reviews*, 2015, **44**, 5495–5551.

153 - A. A. Homaei, R. Sariri, F. Vianello and R. Stevanato, *Journal of Chemical Biology*, 2013, **6**, 185–205.

154 - G. T. Hermanson, in *Bioconjugate Techniques*, 2013, pp. 259–273.

155 - N. Stephanopoulos and M. B. Francis, *Nature Chemical Biology*, 2011, **7**, 876–884.

156 - R. A. Sheldon, *Applied Microbiology and Biotechnology*, 2011, **92**, 467–477.

157 - R. A. Sheldon and S. Van Pelt, *Chem. Soc. Rev. Chem. Soc. Rev*, 2013, **42**, 6223–6235.

158 - R. A. Sheldon, *Catalysts*, 2019, **9**, 261.

159 - J. Fu, M. Liu, Y. Liu, N. W. Woodbury and H. Yan, *Journal of the American Chemical Society*, 2012, **134**, 5516–5519.

160 - V. Linko, M. Eerikäinen and M. A. Kostiainen, *Chem. Commun.*, 2015, **51**, 5351–5354.

161 - J. Fu, Y. R. Yang, A. Johnson-Buck, M. Liu, Y. Liu, N. G. Walter, N. W. Woodbury and H. Yan, *Nature Nanotechnology*, 2014, **9**, 531–536.

162 - H. Li, W. Xiao, P. Xie and L. Zheng, *Enzyme and Microbial Technology*, 2018, **109**, 66–73.

163 - R. Stuermer, B. Hauer, M. Hall and K. Faber, *Current Opinion in Chemical Biology*, 2007, **11**, 203–213.

164 - H. Toogood, D. Mansell, J. Gardiner and N. Scrutton, in *Comprehensive Chirality, Elsevier*, ed. N. J. Turner, Elsiver, Amsterdam, Vol 1 edn., 2012, vol. ch. 7, pp. 216–255.

165 - R. M. Kohli and V. Massey, *Journal of Biological Chemistry*, 1998, **273**, 32763–32770.

166 - U. Kragl, W. Kruse, W. Hummel and C. Wandrey, *Biotechnology and Bioengineering*, 1996, **52**, 309–319.

167 - M. Goretti, C. Ponzoni, E. Caselli, E. Marchigiani, M. R. Cramarossa, B. Turchetti, P. Buzzini and L. Forti, *Enzyme and Microbial Technology*, 2009, **45**, 463–468.

168 - H. S. Toogood, J. M. Gardiner and N. S. Scrutton, *ChemCatChem*, 2010, **2**, 892–914.

169 - T. McKee and J. R. Mckee, *Biochemistry: The Molecular Basis of Life*, Oxford University Press, Inc., New York, Internatio edn., 2009, pp. 197–199.

170 - K. L. Kavanagh, H. Jörnvall, B. Persson and U. Oppermann, *Cellular and Molecular Life Sciences*, 2008, **65**, 3895–3906.

171 - H. Jornvall and J. I. Harris, *European Journal of Biochemistry*, 1970, **13**, 565–576.

172 - C. Sayer, R. J. Martinez-Torres, N. Richter, M. N. Isupov, H. C. Hailes, J. A. Littlechild and J. M. Ward, *FEBS Journal*, 2014, **281**, 2240–2253.

173 - C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, *Science*, 2010, **329**, 305–309.

174 - B. R. Lichman, G. T. Godden, J. P. Hamilton, L. Palmer, M. O. Kamileen, D. Zhao, B. Vaillancourt, J. C. Wood, M. Sun, T. J. Kinser, L. K. Henry, C. Rodriguez-Lopez, N. Dudareva, D. E. Soltis, P. S. Soltis, C. Robin Buell and S. E. O'Connor, *Science Advances*, 2020, **6**, eaba0721.

175 - M. A. Birkett, A. Hassanali, S. Hoglund, J. Pettersson and J. A. Pickett, *Phytochemistry*, 2011, **72**, 109–114.

176 - B. R. Lichman, M. O. Kamileen, G. R. Titchiner, G. Saalbach, C. E. Stevenson, D. M. Lawson and S. E. O'Connor, *Nature Chemical Biology*, 2019, **15**, 71–79.

177 - G. S. Patience, G. Karirekinyana, F. Galli, N. A. Patience, C. Kubwabo, G. Collin, J. C. Bizimana and D. C. Boffito, *Scientific Reports*, 2018, **8**, 2235.

178 - L. Rout and A. M. Harned, *Chemistry - A European Journal*, 2009, **15**, 12926–12928.

179 - V. M. Carneiro, C. M. Avila, M. J. Balunas, W. H. Gerwick and R. A. Pilli, *Journal of Organic Chemistry*, 2014, **79**, 630–642.

180 - E. Marqués-López, R. P. Herrera, T. Marks, W. C. Jacobs and M. Christmann, *Synthesis (Germany)*, 2013, **45**, 1016–1028.

181 - K. Saito, D. J. Thiele, M. Davio, O. Lockridge and V. Massey, *Journal of Biological Chemistry*, 1991, **266**, 20720–20724.

182 - P. A. Karplus, K. M. Fox and V. Massey, *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 1995, **9**, 1518–26.

183 - A. Müller, B. Hauer and B. Rosche, *Biotechnology and Bioengineering*, 2007, **98**, 22–29.

184 - D. S. Blehert, B. G. Fox and G. H. Chambliss, *Journal of Bacteriology*, 1999, **181**, 6254–6263.

185 - C. E. French and N. C. Bruce, *The Biochemical journal*, 1994, **301**, 97–103.

186 - T. B. Fitzpatrick, N. Amrhein and P. Macheroux, *Journal of Biological Chemistry*, 2003, **278**, 19891–19897.

187 - C. Ceccarelli, Z. X. Liang, M. Strickler, G. Prehna, B. M. Goldstein, J. P. Klinman and B. J. Bahnson, *Biochemistry*, 2004, **43**, 5266–5277.

188 - D. Quaglia, J. A. Irwin and F. Paradisi, *Molecular Biotechnology*, 2012, **52**, 244–250.

189 - B. Kosjek, W. Stampfer, M. Pogorevc, W. Goessler, K. Faber and W. Kroutil, *Biotechnology and Bioengineering*, 2004, **86**, 55–62.

190 - M. S. Humble, K. E. Cassimjee, M. Häkansson, Y. R. Kimbung, B. Walse, V. Abedi, H. J. Federsel, P. Berglund and D. T. Logan, *FEBS Journal*, 2012, **279**, 779–792.

191 - S. Reich, H. W. Hoeffken, B. Rosche, B. M. Nestl and B. Hauer, *ChemBioChem*, 2012, **13**, 2400–2407.

192 - D. Dobrijevic, L. Benhamou, A. E. Aliev, D. Méndez-Sánchez, N. Dawson, D. Baud, N. Tappertzhofen, T. S. Moody, C. A. Orengo, H. C. Hailes and J. M. Ward, *RSC Adv.*, 2019, **9**, 36608–36614.

193 - N. Kress, J. Rapp and B. Hauer, *ChemBioChem*, 2017, **18**, 717–720.

194 - T. Knaus, H. S. Toogood and N. S. Scrutton, in *Green Biocatalysis*, John Wiley & Sons, Inc, Hoboken, NJ, 2016, ch. 18, pp. 473–488.

195 - A. Dasgupta and A. Wahed, in *Clinical Chemistry, Immunology and Laboratory Quality Control*, ed. A. Dasgupta and A. Wahed, Elsevier, San Diego, 2014, pp. 1–18.

196 - M. Hall, C. Stueckler, H. Ehammer, E. Pointner, G. Oberdorfer, K. Gruber, B. Hauer, R. Stuermer, W. Kroutil, P. Macheroux and K. Faber, *Advanced Synthesis and Catalysis*, 2008, **350**, 411–418.

197 - C. Stueckler, M. Hall, H. Ehammer, E. Pointner, W. Kroutil, P. Macheroux and K. Faber, *Organic Letters*, 2007, **9**, 5409–5411.

198 - Y. Yanto, H.-H. Yu, M. Hall and A. S. Bommarius, *Chem. Commun.*, 2010, **46**, 8809–8811.

199 - J. F. Chaparro-Riggers, T. A. Rogers, E. Vazquez-Figueroa, K. M. Polizzi and A. S. Bommarius, *Advanced Synthesis & Catalysis*, 2007, **349**, 1521–1531.

200 - S. Reich, B. M. Nestl and B. Hauer, *ChemBioChem*, 2016, **17**, 561–565.

201 - C. Stueckler, N. J. Mueller, C. K. Winkler, S. M. Glueck, K. Gruber, G. Steinkellner and K. Faber, *Dalton Trans.*, 2010, **39**, 8472–8476.

202 - M. Hall, C. Stueckler, B. Hauer, R. Stuermer, T. Friedrich, M. Breuer, W. Kroutil and K. Faber, *European Journal of Organic Chemistry*, 2008, **2008**, 1511–1516.

203 - D. Baud, N. Ladkau, T. S. Moody, J. M. Ward and H. C. Hailes, *Chemical Communications*, 2015, **51**, 17225–17228.

204 - S. V. Wegner and J. P. Spatz, *Angewandte Chemie - International Edition*, 2013, **52**, 7593–7596.

205 - Y. Dayani and N. Malmstadt, *Biomacromolecules*, 2013, **14**, 3380–3385.

206 - M. R. Krause and S. L. Regen, *Accounts of Chemical Research*, 2014, **47**, 3512–3521.

207 - X. Pan, G. Wu, W. Yang, R. F. Barth, W. Tjarks and R. J. Lee, *Bioconjugate Chemistry*, 2007, **18**, 101–108.

208 - X. Han, A. S. Achalkumar, M. R. Cheetham, S. D. A. Connell, B. R. G. Johnson, R. J. Bushby and S. D. Evans, *ChemPhysChem*, 2010, **11**, 569–574.

209 - M. Alhassan, O. Al Musaimi, J. M. Collins, F. Albericio and B. G. De La Torre, *Green Chemistry*, 2020, **22**, 2840–2845.

210 - K. Barlos, D. Gatos, G. Papaphotiou and W. Schäfer, *Liebigs Annalen der Chemie*, 1993, **1993**, 215–220.

211 - M. Gude, J. Ryf and P. D. White, *International Journal of Peptide Research and Therapeutics*, 2002, **9**, 203–206.

212 - L. Schmitt, C. Dietrich and R. Tampé, *Journal of the American Chemical Society*, 1994, **116**, 8485–8491.

213 - B. P. Cormack, R. H. Valdivia and S. Falkow, *Gene*, 1996, **173**, 33–38.

214 - G. Tsuji, T. Sunami and N. Ichihashi, *Journal of Bioscience and Bioengineering*, 2018, **126**, 540–545.

215 - M. E. Haberland and J. A. Reynolds, *Proc. Nat. Acad. Sci.*, 1973, **70**, 2313–2316.

216 - V. Helms, in *Principles of Computational Cell Biology: From Protein Complexes to Cellular Networks*, Wiley-VCH Verlag GmbH & Co. KGaA, 2008, ch. 8, p. 202.

217 - R. B. Sekar and A. Periasamy, *Journal of Cell Biology*, 2003, **160**, 629–633.

218 - M. Skruzny, E. Pohl and M. Abella, *Biosensors*, 2019, **9**, 122.

219 - T. Chen, B. He, J. Tao, Y. He, H. Deng, X. Wang and Y. Zheng, *Advanced Drug Delivery Reviews*, 2019, **143**, 177–205.

220 - Protein Paintbox of Synthetic Fluorescent & Chromogenic Proteins, *https://www.atum.bio/catalog/reagents/protein-paint-box#__fluorescent-overview (Accessed November 2021)*.

221 - S. Lata, A. Reichel, R. Brock, R. Tampé and J. Piehler, *Journal of the American Chemical Society*, 2005, **127**, 10205–10215.

222 - What are Micelles?, *http://whataremicelles.blogspot.com/2017/05/what-are-micelles.html*.

223 - G. Kong, M. Xiong, L. Liu, L. Hu, H. M. Meng, G. Ke, X. B. Zhang and W. Tan, *Chemical Society Reviews*, 2021, **50**, 1846–1873.

224 - T. A. Ngo, H. Dinh, T. M. Nguyen, F. F. Liew, E. Nakata and T. Morii, *Chemical Communications*, 2019, **55**, 12428–12446.

225 - O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman and I. Willner, *Nature Nanotechnology*, 2009, **4**, 249–254.

226 - T. A. Ngo, E. Nakata, M. Saimura and T. Morii, *Journal of the American Chemical Society*, 2016, **138**, 3012–3021.

227 - S. M. Douglas, I. Bachelet and G. M. Church, *Science*, 2012, **335**, 831–834.

228 - D. Y. Zhang and E. Winfree, *Journal of the American Chemical Society*, 2009, **131**, 17303–17314.

229 - G. Cooper, in *The Cell: A Molecular Approach*, Sinauer Associates, Sunderland (MA), 2nd edn., 2000.

230 - H. Y. Song, M. H. Ngai, Z. Y. Song, P. A. MacAry, J. Hobley and M. J. Lear, *Organic and Biomolecular Chemistry*, 2009, **7**, 3400–3406.

231 - Y. T. Lai, Y. Y. Chang, L. Hu, Y. Yang, A. Chao, Z. Y. Du, J. A. Tanner, M. L. Chye, C. Qian, K. M. Ng, H. Li and H. Sun, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 2948–2953.

232 - P. M. Arnott, H. Joshi, A. Aksimentiev and S. Howorka, *Langmuir*, 2018, **34**, 15084–15092.

233 - A. Ohmann, K. Göpfrich, H. Joshi, R. F. Thompson, D. Sobota, N. A. Ranson, A. Aksimentiev and U. F. Keyser, *Nucleic acids research*, 2019, **47**, 11441–11451.

234 - C. Lanphere, D. Offenbartl-Stiegert, A. Dorey, G. Pugh, E. Georgiou, Y. Xing, J. R. Burns and S. Howorka, *Nature Protocols*, 2021, **16**, 86–130.

235 - J. A. Nye and J. T. Groves, *Langmuir*, 2008, **24**, 4145–4149.

236 - G. Raghunath and R. B. Dyer, *Langmuir*, 2019, **35**, 12550–12561.

237 - R. A. Graff, T. M. Swanson and M. S. Strano, *Chemistry of Materials*, 2008, **20**, 1824–1829.

238 - C. Xu, K. Xu, H. Gu, X. Zhong, Z. Guo, R. Zheng, X. Zhang and B. Xu, *Journal of the American Chemical Society*, 2004, **126**, 3392–3393.

239 - G. Kulsharova, N. Dimov, M. P. Marques, N. Szita and F. Baganz, *New Biotechnology*, 2018, **47**, 31–38.

240 - P. Fernandes and C. C. de Carvalho, *Processes*, 2021, **9**, 1–23.

241 - C. Schmidt-Dannert, *Microbial Biotechnology*, 2017, **10**, 1164–1166.

242 - R. A. Sheldon and J. M. Woodley, *Chemical Reviews*, 2018, **118**, 801–838.

243 - C. L. Gargalo, I. Udugama, K. Pontius, P. C. Lopez, R. F. Nielsen, A. Hasanzadeh, S. S. Mansouri, C. Bayer, H. Junicke and K. V. Gernaey, *Journal of Industrial Microbiology and Biotechnology*, 2020, **47**, 947–964.

244 - P. G. Levi and J. M. Cullen, *Environmental Science and Technology*, 2018, **52**, 1725–1734.

245 - C. J. Clarke, W. C. Tu, O. Levers, A. Bröhl and J. P. Hallett, *Chemical Reviews*, 2018, **118**, 747–800.

A | Appendix

A.1 GC Calibration Curves

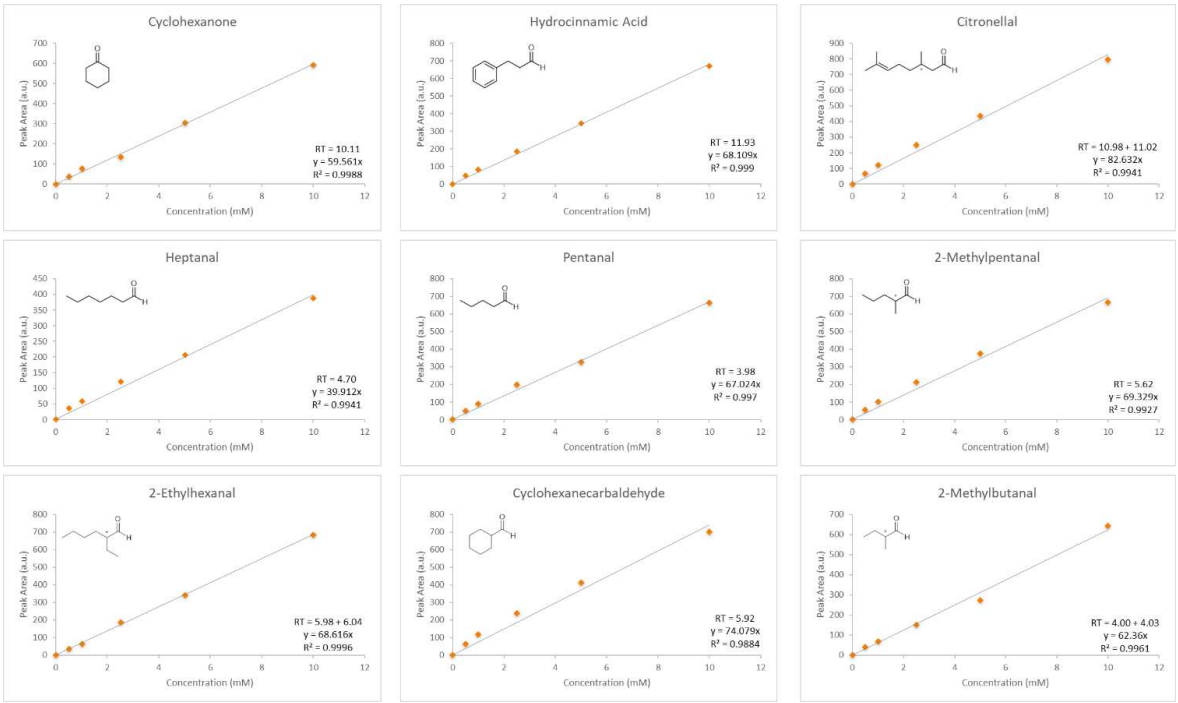


Figure A.1: GC Calibration Curves for ER Assay

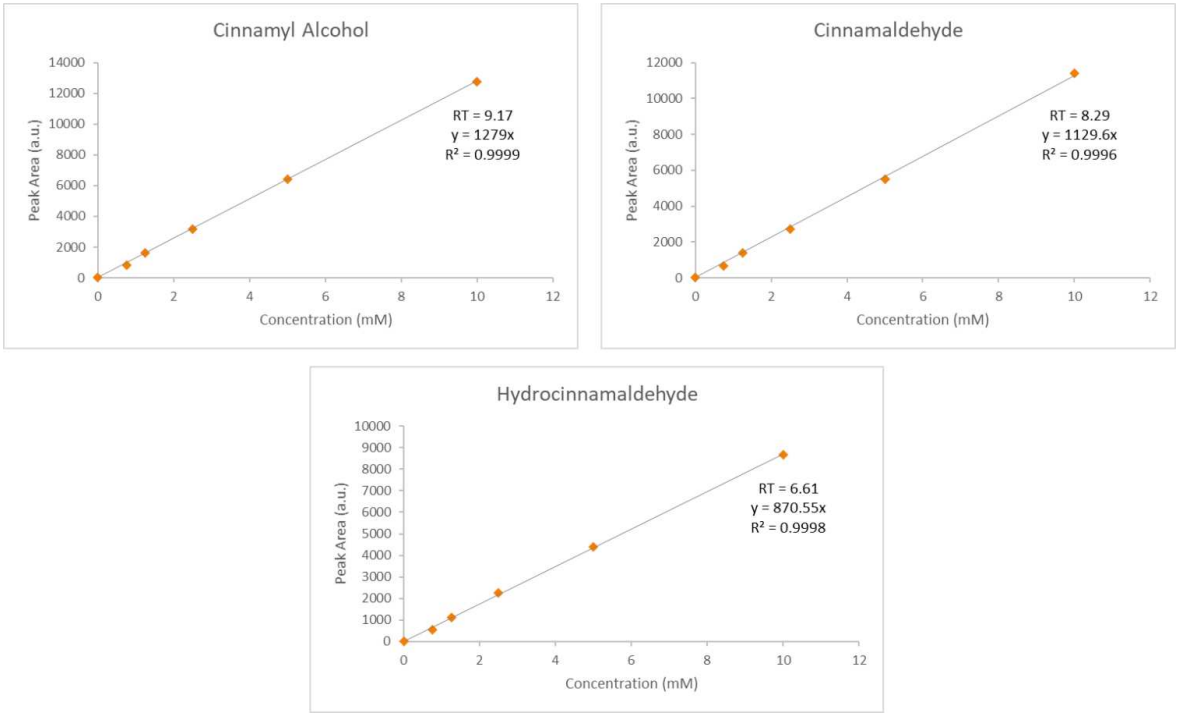


Figure A.2: GC Calibration Curves for ER/ADH Assay

A.2 HPLC Traces

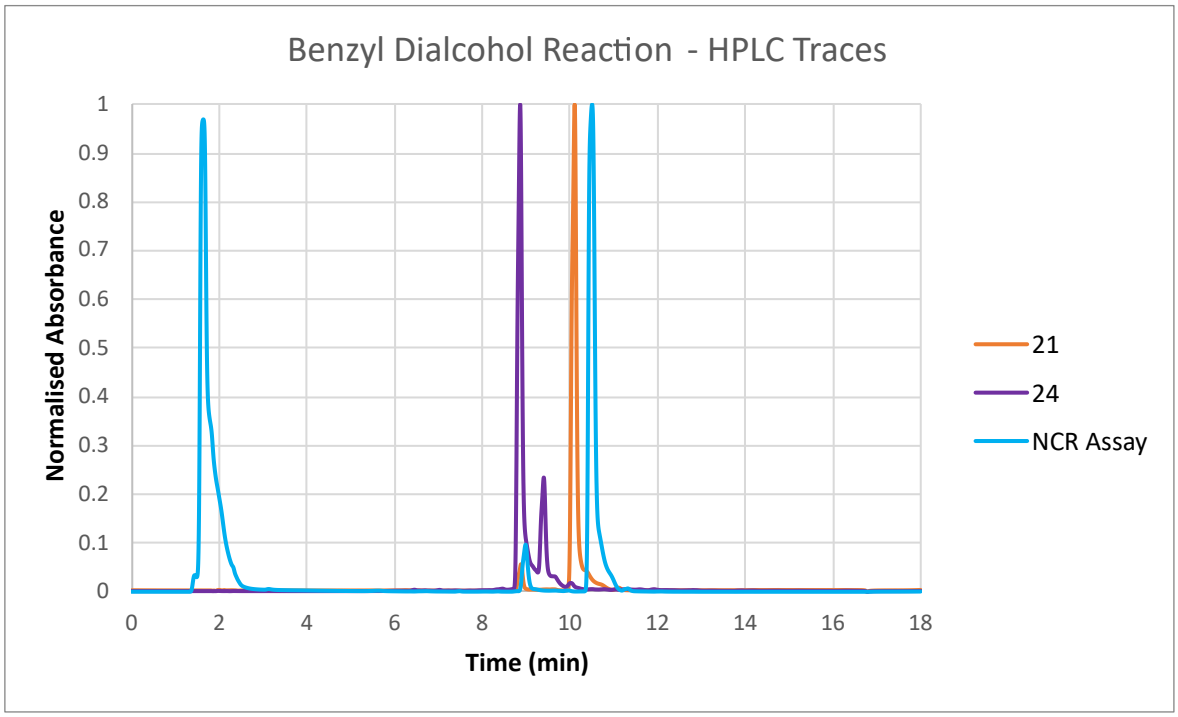


Figure A.3: HPLC Traces for the Benzyl Dialcohol Reaction - 24 = Benzyl Dialcohol, RT = 8.9; 21 = Benzyl Dialdehyde, RT = 10.1.

A.3 Protein Sequences

A.3.1 Alcohol Dehydrogenases

ADH6 (40,616 Da)
MKAVVYKEPFTVAVEEVADPRIEDPNDVIIRVTTSAICGSDLHMYEGRTGATPGMVFGHENMGIVEEV
GSGVASIRKGDRVVLPFNVACGFCKNCLAGKTGFCLTVNPGFAGGAYGYVAMGPYQGGQSEYLRVPF
ADFNCLKLPPGTEFEDDFALLADIFPTGYHATELAQVSPGETVSVFGAGPVGLMAAYSAILRGASRVFV
VDRVPSRLSLAERIGAIPIDYSKGDPVAQIKEHTNGEGTDKGIDAVGYQATAPEGEEQPALVLNTLVDT
VRPTGALGVVGLYVTEDPGAPSESAKQGELLFAMGKLFEKGQRMGTGQANVKAYNRQLRDLIIAGRA
KPSFVVSQRIPLQRAPEAYEKFDKREEGWTKVLLKPQQES

ADH9 (27,379 Da)
MAISGRLTDKIALITGAGSGFGRASALRFAAEGASVVCVDRNADAAHSAADDIAAAGGTALALTADVS
SAADAERMTTTTLEHFGRIDIVFANAGIPGSGDAHTTTEEEWDRVIAINLKGVWLTSKYALPHMVERR
SGVITNQASVGGLIGIPGIFPYAAAKGGVISMTRQMAAAYAPHNIRVNAICPGGVYTPLVELSRQKRGL
TASSVEEANAIAARNYPLGRLGTVEEIASLALFLASDEAAWITGGIYPVDGGRGAVGTIPTD

HLADH (42,099 Da)
MGSSHHHHHHSSGLVPRGSHMSTAGKVIKCKAAVLWEEKKPFSIEEVEVAPPKAHEVRIKMVATGIC
RSDDHVVSGTLVTPLPVIAGHEAAGIVESIGEGVTTVRPGDKVIPLFTPQCGKCRVCKHPEGNFCLKND
LSMPRGTMQDGTSRFTCRGKPIHHFLGTSTFSQYTVVDEISVAKIDAASPLEKVCLIGCGFSTGYGSAV
KVAKVTQGSTCAVFGLGGVGLSVIMGCKAAGAARIIGVDINKDKFAKAKEVGATECVNPQDYKKPIQE
VLTEMSNGGVDFSFEVIGRLDTMVTALSCCQEAYGVSVIVGVPPDSQNLSMNPMLLLSGRTWKGAIF
GGFKSKDSVPKLVADFMAKKFALDPLITHVLPFEKINEGFDLLRSGESIRTILTF

BtADH (36,324 Da)
MKAAVVEQFKEPLKIKEVEKPTISYGEVLVRIKACGVCHTDLHAAHGDWPVKPKLPLIPGHEGVGIVEE
VGPGVTHLKVGDRVGIPWLYSACGHCDYCLSGQETLCEHQKNAGYSVDGGYAEYCRAAADYVVKIP
DNLSFEEAAPIFCAGVTTYKALKVTGAKPGEWVAIYGIGGLGHVAVQYAKAMGLNVVAVDIGDEKLEL
AKELGADLVVNPLKEDAAKFMKEKVGGVHAAVVTAVSKPAFQSAYNSIRRGGACVLVGLPPEEMPVPI
FDTVLNGIKIIGSIVGTRKDLQEALQFAAEGKVKTIIEVQPLEKINEVFDRMLKGQINGRVVLTLEDK

RaADH (35,192 Da)
MKAVQYTEIGSEPVVVDIPTPTPGPGEILLKVTAAGLCHSDIFVMDMPAAQYAYGLPLTLGHEGVGTV
AELGEGVTGFGVGDAVAVYGPWGCGACHACARGRENYCTRAADLGITPPGLGSPGSMAEYMIVDSA
RHLVPIGDLDPVAAAPLTDAGLTPYHAISRVLPLLGPGSTAVVIGVGGLGHVGIQILRAVSAARVIAVDL
DDDRLALAREVGADAAVMSGAGAADAIRELTGGQGATAVFDFVGAQSTIDTAQQVVAVDGHISVVGI
HAGAHAKVGFFMIPFGASVVTPYWGTRSELMEVVALARAGRLDIHTETFTLDEGPAAYRRLREGSIRG
RGVVVP

pQR1492 (36,078 Da)
MQGLMKAAIVHEFGKPLIIENVPIPVPGPGEVLVKVAACGVCHTDLHAASGDWPVKPSVPFIPGHEVA
GTVAALGPGVTDFAIGDAVGIAWLHDACMRCEHCETGWETLCEHQHNTGYSCDGGFAEYAIANAAF
AARLPDDVDFTAVAPILCAGVTTYKGLKETEVRPGQWVAISGIGGLGQVAIQYAKAMGLKVIALDVTP
EKLALARRTGADVALDARSPDAIAEALEATGGGAHGVLVTAVSPPAFSQALQLVRRKGTVALVGLPPG
EFATPIFDVVLKRITVRGSIVGTRRDLDEAIAFFAEGKVKAQVTPAPLEDINRIFEELKAGRVEGRVVLD
MSSKAA

pQR1493 (35,653 Da)
MRAWQLGAFDIASLALADRLAPVPQAGEVLVDITAVSLNYRDLAIAVGTYAPTQRLPLIPASDAVGTIS
AIGAGVTKWKVGDRVIGCYMQTWERGPSKPSDRQNTLGSPLDGVLCEQRVFPQGSIVAAPARLTDH
ECAALPIAAVTAWCALFEHGEARPGETMLVQGSGGVSTFAIQLAAAAGLRVAAVSRSAQKLEAARRL
GASLLIESTATPDWSRAVLDLTQQVGVDVILDVGGQGTLRESLRAAAQNGRIISIGFLGGTSSEINLGQ
VITKNLTVRGVTVGSRTSFENLLRFLDQTGIKPVIDSVFSFEEAPLAFARLASGEQQGKIVIRMEGHT

pQR1494 (39,368/ Da)
MKTRAAIAWAAGQPLSIHEVDLAPPRAGEVLVEIKATGICHTDYYTLSGADPEGRFPAILGHEGAGVV
LEVGEGVKSLKAGDHVIPLYTPECRQCKFCLSQKTNLCQAIRSTQGQGLMPDGTSRFSLDGQPILHYM
GTSTFANHIVVPEIALAKIREDAPFDKVCYIGCGVTTGVGAVLFTAKVEAGANVVVFGLGGIGLNVIQG
AKMVGAGKIIGVDINPARQEMARKFGMTHFINPNEVENVVDAIIQLTDGGADYSFECIGNTQVMRQA
LECCHKGWGQSIIIGVAEAGKEIATRPFQLVTGRVWKGSAFGGARGRTDVPKIVDWYMDGKLNIDDLI
THHLTLDEINKGFDLMKSGESIRSVVVY

pQR1495 (52,253 Da)
MEQCTIAQMQACLRQQQQAYQQDPYPSLAVRMDRLDRLLKMSEVATEALIEAISADFGNRSRHETL
LAELSMSQGAIRNAKRHLKGWMKPRRVATPLAYMPSSNQMIRQPLGVVGVVAPWNYPYQLSIGPTIC
AIAAGNRVMIKPSELTPRFSALLAQIIAEHFAPDEISVITGDASVAQDFVSLPFDHLVFTGSTAVGRKVA
EAAAKNLTPVTLELGGKSPAVFDSSVDLGLIAPRISFAKFVNAGQTCIAPDYLMVPTGQTETVIKALQQ
ATQTLYPTTQTNGDYTSIINERHYQRLMHLLQDAEQKGAQIQRLTGEAPDAQSRKIPPTVVVNVRPD
MQIMQEEIFGPIFPIMEYSSVDEAIDYINQRDRPLALYWFGNNPRHQQRFLTETISGGVTINDCVWHF
AQETQPFGGVGSSGQGAYHGEWGFRTFSKEKPVFIQSRLSAIKLLYPPYGAMFSRMMALLKIIT

pQR1496 (51,470 Da)
MMHTTTQELDRLFAAQKAAFGRNMMPTLDERLDRLRSLEAAMVRNRKGFQETLAADFGSHSPMVT
DLFETGGVLGRNRYIQTQLAEWMQPSERPLNAYVHGQSSARVIRQPKGVIGNLAPWNFPIECALVMV
NDMLAAGNRVIVKASEQTEATSNLLRQAISEAFDESELAVVCGGDIAFAEYFCTLPWDHITYTGGGKI
GRNVLRLAAERLTPVTLELGGKNPTVFSESGVSADLVERFLYNRVFKGGQVCTSPDYAFVPAHRLDEW
IALAKSTWSAMYPSYVGHPDATGTINRRHYDRILGYLSEAEAAGCTVISLNGDTPDPERLQIPMYLVID
PGEALGVMTDEIFGPVTAVKPYTSLDEAVDYINGHDRPLAAYFVGRERDLADVFQTSVIAGGIGINVFG
LQGAEPVLPFGGIGASGMGCHSGYEGFIGFTHSKSVFECADDSPLMAALKGPYGAMTQAFADAVFPP
A

pQR1497 (52,460 Da)
MATAIKQDLAGETARMHEVLAAQKASFTAAMPESLSVRTDRIDRAVALLVDNAEDFAKAVSEDFGHR
SREQTLMTDIMPSVSALKHAKKHMASWAKGEKRKPTFPLGLLGAKAEVVFQPKGVVGIVAPWNFPVG
MVFVPMAGVLAAGNRAMIKPSEFTENVSALMARLVPDYFDESEMAVFTGDSDVGVAFSKLAFDHMIF
TGATSVGRHIMRAAADNLVPVTLELGGKSPTFIGRSANKDLVGQRVALGKMMNAGQICLAPDYLLVA
EDQEGEVIDSVTKGATALYPTLLANDDYTSVVNTRNYDRLQSYLADAREKGAEVIEVNPGGEDFASSN
GHKMPLHIVRNPTDDMRVMQEEIFGPILPVKTYKHIDDAIDYVNANDRPLGLYYFGQDKSEEDRVLTR
TISGGVTVNDVLFHNAMEDLPFGGVGPSGMGNYHGVDGFRTFSHARAVYRQPKLDVAGLAGFKPPY
GKATAKTLAKELKK

pQR1498 (36,864 Da)
MNLAGIYRPPTERQEAVKVLRRAYERGVTFFDTAQLYGPFLSEEQVGDALAPVRDKVVIATKFGYEIDP
ETRQFRGLNSRPEYIKQATEASLRRLRTDYIDLYYQHRIDPNVPIEDVAGAVQDLIRQGKVRHFGLSEA
GAATIRRAHAVQALTSVTNEYSVWTRDPEIEVIAACEELGIGLSPWSPLGPGFLTGAITSATVLDPTDA
RMSYRFPRFTPEAIRANTPIVELLRRVGQCHDATPGQVALAWLLARKPFIVPIPGTTRIEHLEENLGAL
NVRLSAEDMAEVESVFARIGVEGARFPPEVLALSDIGAVLGTSSLGGHGKSPLPVRAVR

pQR1499 (34,803 Da)
MQAALNRLAAAPRNDIPYRKLGRTGVEVSIVGVGGFHIGEKPTEDEAIRIVRTALDHGVNFLDNCWDY
NKGASEIRMGKALRDGYRDRAFLMSKIDSHSHAGAARQIDESLQRLQTDHVDLMQFHEIIRPDDPEKI
FAKGGPMEAMLAAQKAGKVRFIGFTGHKHPDIHAHMLEVARERKFHFDAVQLPLNVMDAHYESFED
KVLPVLVKEEIGVLGMKPLGSGKILESKTVEAIECLHYAMSLPTSVVITGCDSLAVLKQALQAARTFKPL
SEKQRETLLTRSSPASPGGKFEGYKTTNEFDGTIQHPEWLL

pQR1500 (36,720 Da)
MEYRQLGSSGLRVPALSLGTGTFGGQGPLFSAWGTSDAAEARRLIDISLDAGVTLFDTADVYSNGASE
AILGEAIKGRRDAVLISTKTGLPMGDGPHDWGASRARLIGAVEASLRRLGTDHIDLLQLHAFDASTPV
DELMDTLATLIAAGKLRYAGVSNYPGWQLMKAQAAADRLGVPRFVAHQVYYSLIGRAYEADLMPLAV
DQGIGALVWSPLGWGRLTGKIGRGRPVPAGSRLHETEQFAPPVAEELLYRVIDALEAAAAETGKTVPQ
VAINWLLRRPTVSSVIIGARNEEQLRQNLGAVGWELTAEQVAVLDGASDVLPPYPHTPYRQQAGFARL
NPPLV

pQR1501 (38,325 Da)
MENQTFAVAADGVDPSRVPYKTLYTGARIPSIGIGTFGSDRFSGEEVADAVLGAISVGYRHVDCASVYS
NEHLIGESLQKAVAGGIKREDLFITSKVWNNMHGDGDILLSVAKTLKDLRLDYLDLFLVHWPFPNHHA
TGVSVDSRDPHAVPYIHENFMKTWRQMERLVDMGLVKAIGTSNVTVPKLKLILRDARIKPAVNEMEL
HPHFQQPALFQFCLDNNIVPIGYSPIGSPNRPDRDKTEADTVDVEDPVVVAIGNRLGVHPALVCLKWA
VQRGQIPIPFSIHRKNYLANIRSITENLLTDEEMKAMAGIDKNCRLIKGQVFLWPSAKDWPDLWDEDG
TIAGA

pQR1502 (37,889 Da)
MEYRYLGASGFRVPVLSFGTGTFGGQGALFSAWGSTDVAEARRLVDICLDAGLTMFDSADVYSKGAA
EEILGEAIKDRPRDALIISTKATFRFGDGENQVGSSRHHLINAVDAALKRLGTDYIDLFQLHGFDARPPV
EEVLSTLDTLVKAGKIRYLGVSNFSGWHLMKSLAAADRHGWTRYVAHQAYYSLVGRDYEWELMPLA
ADQGVGAVVWSPLGWGRLTGKIRRGQPLPENSRLQSQTAHDAGPKVDLEYLYDVVDALDEVAQETG
KTIPQVALNWLLQRPTVSTVVIGARNEAQLRQNLGAVGWNLTPEQVAKLDKASQRPKPYPYWHQAG
FAYRNPTPV

pQR1503 (42,704 Da)
MTQRVFDEQRRKFIGTSMGVASGAAVAVASSGLLGCAQAQTQSGAANSGRSAAAQISTRRKLGRLE
VSRVGLGVQNMHRTYQTTVPARPEMIDIIRTAYERGVTFFDTAEAYGPHECERILGEAIAPFRDQVVIT
SKFGWNIDIETGARGPGLISRPDHIKLAVEGSLKRLRTDRIDLLYQHRVDPQVPIEDVAGAVKDLMTQ
GKVLHWGLSEMGLKTLRRAHAALPVTAVQSEYSMLWRGPEKEVLPSCEELGIGFVPWSPLGVGFLTG
AIDEKTNFADGDFRKTETRFSSENRAHNLALVKLVKTWAERKNTTPARIALAWLLAQKPWIVPIPGTT
QMAHMFDNIGAAEVTFSADELQALNAALANIVIQGERLPPAVQAFSDVEAPLK

pQR1917 (34,964 Da)
MKAVAYQQSLPIDAANALVDVELPAPVAGEHDLLVAVRAISVNPVDTKVRRGVTPAAGEWKVLGWD
VAGEVVGVGAAVRGFAPGDRVWYAGSIARPGGNSELHAVDARIAAKMPSTVDFAQAAALPLTAITAW
EILFDRLGLHAASTGTLLVIGAAGGVGSILVQLARQLTGLTVVGTASRAETRDWLLDLGAHAVVDHGA
PLAAGLRAAGYAEADYVVSLNQSEQHFAEVVELIAPQGKFALIDDPASLDIVPLKRKSVSVHWEFMFTR
PLFATADLARQGELLAELARLVDDGRVRTTLAANLGRIDAANLRRAHALIESGKARGKLVLAGF

pQR1918 (35,903 Da)
MKAVGYKVPGPIAEDASLVDIDLPRPVAEGGDILVEVKAVSVNPVDYKIRSSTPPADGDWKVLGWDAA
GIVQEVGPEVTQFAVGDEVYYAGSLIRPGTNAEFHLVDARIVGRKPASLDWAEAAALPLTTLTAWEAM
FDRLDVTKPVPGAAPAILIIGGAGGVGSIAIQIARQRTDLIVISTASRPETQEWVKGLGAHHVIDHSRPL
APQIAELNIGAPAFVFSTTHTEQHASDIAELIAPQGRFGFIDDPKALAVMLFKRKAVSIHHELMFTRSLY
GTPDMDEQGKILNSLAVLVDDGKIRTTLTEKLSPINAANLKTVHALIESGAARGKIVLEGF

pQR1919 (36,603 Da)
MAKALVLDKTLDLKLRDIDVPLVVGPRDVKIEIRNVGVCGSDLHYFLHGRIGPYVVEQPMILGHEASGV
VAEVGSEVTHLKVGDRVCMEPGIPDSSSRATMEGLYNLDPAVSFWATPPIHGCLTPHVVHPAAFTFK
MPDNVSFAEGAFVEPLAIGLQAAKKAAIKPGDVAVVIGAGTIGAMTALAALAGGAARVILADLVPEKLA
LFKDDPAVTTVNVKEQSLAEVVKAATDGWGANIVFEASGGAKAFEHIFELLCPGGCVVLVGIPVDKVP
FDIVAIQAKEARIESVFRYANIFPRALALISSGKIDVKRFISRVFPFEQGVEAFEEAAKARPGDVKIQIAVT
E

pQR1920 (50,674 Da)
MTIDRDFRMLIGGQLVPGAAQLAVENPATGDIFAAAPDASPTDLDAAVAAARAAFPGWRDTPVEARR
EAIIALADRLIAEVEPMKRLLTREQGKPHKDAAEELLGAASWLKAAASLDIPVTVNEDSAERYSETRHV
PIGVVAAIAPWNYPLILASFKLGPALLAGNCVILKPSPFTPLTALKLGELAADIFPAGVLNVISGGDALGP
WLTSHAGIDKVSFTGSTATGKKVMAAASDTLKRVTLELGGNDPAIVMHDVDAEAIAEQIFWAAFGNG
GQICIASKRIYVHEQVYDALRDALVAYARTVKVGDGAELGTQIGPVNNSLQYKRVLELIRDARAQGYEF
LTGGEEAPGTGHFIPVTIIDNPPEDSRIVQEEQFGPVLPLLKFGDYDEAIARANACEYGLGATVWGADE
DSAWALAERIESGNVWVNETRHLSPFVAFAGHKQSGFGVENGQEGLLEYLLPKTITRKRAPGSQ

pQR1921 (38,613 Da)
MAWQPADTRYDAMLYNRCGNSGLKLPAISLGLWHNFGDDTPHQTKREICQKAFDLGITHFDLANNY
GPPPGSAELAFGEILRKDFAGYRDELIISSKAGYNMWPGPYGEWGSRKYVLASLDQSLKRMGLDYVDI
FYSHRYDPETPLEETMGALDHAVRSGKALYVGISSYNSQRTREAAAILRSLGTPFIIHQPSYSMINRWV
EDDGLLDTLDGLGVGSIVFSPLAQGMLTGKYLGGIPQDSRAAQGKSLRQSFLNDETLEKVRALNGIAE
KRGQTLAQLALAWVLRKGRVTSALIGASRPSQVEDCVGAIKNLDFSDAELAEIDRYAKDANINLWAKS
AERDGPERKK

A.3.2 Ene-Reductases

NCR (40,570 Da)
MHHHHHHGGGSPSLFDPIRFGAFTAKNRIWMAPLTRGRATRDHVPTEIMAEYYAQRASAGLIISEAT
GISQEGLGWPYAPGIWSDAQVEAWLPITQAVHDAGGLIFAQLWHMGRMVPSNVSGMQPVAPSASQ
APGLGHTYDGKKPYDVARALRLDEIPRLLDDYEKAARHALKAGFDGVQIHAANGYLIDEFIRDSTNHR
HDEYGGAVENRIRLLKDVTERVIATIGKERTAVRLSPNGEIQGTVDSHPEQVFIPAAKMLSDLDIAFLG
MREGAVDGTFGKTDQPKLSPEIRKVFKPPLVLNQDYTFETAQAALDSGVADAISFGRPFIGNPDLPRRF
FEKAPLTKDVIETWYTQTPKGYTDYPLLGD

YqjM (38,665 Da)
MHHHHHHGGGSARKLFTPITIKDMTLKNRIVMSPMCMYSSHEKDGKLTPFHMAHYISRAIGQVGLII
VEASAVNPQGRITDQDLGIWSDEHIEGFAKLTEQVKEQGSKIGIQLAHAGRKAELEGDIFAPSAIAFDE
QSATPVEMSAEKVKETVQEFKQAAARAKEAGFDVIEIHAAHGYLIHEFLSPLSNHRTDEYGGSPENRY
RFLREIIDEVKQVWDGPLFVRVSASDYTDKGLDIADHIGFAKWMKEQGVDLIDCSSGALVHADINVFP
GYQVSFAEKIREQADMATGAVGMITDGSMAEEILQNGRADLIFIGRELLRDPFFARTAAKQLNTEIPAP
VQYERGW

XenA (40,819 Da)
MHHHHHHGGGSSALFEPYTLKDVTLRNRIAIPPMCQYMAEDGMINDWHHVHLAGLARGGAGLLVV
EATAVAPEGRITGCAGIWSDAHAQAFVPVVQAIKAAGSVPGIQIAHAGRKASANRPWEGDDHIAADD
ARGWETIAPSAIAFGAHLPKVPREMTLDDIARVKQDFVDAARRARDAGFEWIELHFAHGYLGQSFFSE
HSNKRTDAYGGSFDNRSRFLLETLAAVREVWPENLPLTARFGVLEYDGRDEQTLEESIELARRFKAGG
LDLLSVSVGFTIPDTNIPWGPAFMGPIAERVRREAKLPVTSAWGFGTPQLAEAALQANQLDLVSVGRA
HLADPHWAYFAAKELGVEKASWTLPAPYAHWLERYR

MorA (33,205 Da)
MHHHHHHGGGSAGKSPLINLNNGVKMPALGLGVFAASAEETASAIASAISSGYRLIDTARSYNNEAQ
VGEGIRNSGVDRAEMFVTTKLFNCDYGYERALRAFDESLGRLGLDYVDLYLLHWPTKDWNATIQSW
KAAEKILGDGRARAIGVCNFLEDQLDELIAASDVVPAVNQIELHPYFAQKPLLAKNRALGIVTEAWSPI
GGAINDGDGDNHGGRKHPLTDPVITTIAEAHGRSAAQVILRWHFQNDVVAIPKSVNPERIAKNIDVFD
FALSDAEMAQLDELDTGVRIGPDPRDVDTSSFAEFV

A.3.3 Transaminases

CvTAm (52,036 Da)
MQKQRTTSQWRELDAAHHLHPFTDTASLNQAGARVMTRGEGVYLWDSEGNKIIDGMAGLWCVNV
GYGRKDFAEAARRQMEELPFYNTFFKTTHPAVVELSSLLAEVTPAGFDRVFYTNSGSESVDTMIRMVR
RYWDVQGKPEKKTLIGRWNGYHGSTIGGASLGGMKYMHEQGDLPIPGMAHIEQPWWYKHGKDMT
PDEFGVVAARWLEEKILEIGADKVAAFVGEPIQGAGGVIVPPATYWPEIERICRKYDVLLVADEVICGFG
RTGEWFGHQHFGFQPDLFTAAKGLSSGYLPIGAVFVGKRVAEGLIAGGDFNHGFTYSGHPVCAAVAH
ANVAALRDEGIVQRVKDDIGPYMQKRWRETFSRFEHVDDVRGVGMVQAFTLVKNKAKRELFPDFGEI
GTLCRDIFFRNNLIMRACGDHIVSAPPLVMTRAEVDEMLAVAERCLEEFEQTLKARGLAHHHHHH

A.3.4 Fluorescent Proteins

FresnoRFP (28,153 Da)
MGSSHHHHHHSSGLVPRGSHMNSLIKENMHMKLYMEGTVNNHHFKCTSEGEGKPYEGTQTMRIK
VVEGGPLPFAFDILATSFMYGSRTFIKYPKGIPDFFKQSFPEGFTWERVTRYEDGGVVTATQDTSLEDG
CLVYHVQVRGVNFPSNGPVMQKKTLGWEPNTEMLYPADGGLEGRSDMALKLVGGGHLSCSFVTTYR
SKKTVGNIKMPGIHAVDHRLVRIKEADKETYVEQHEVAVAKFAGLGGGMDELYK

FrostyCFP (28,640 Da)
MGSSHHHHHHSSGLVPRGSHMTALTEGTKLFEKEIPYITELEGDVNGMKFTIHGKGTGDATTGTIKA
KYICTTGDLPVPWATLVSTLSYGVQCFAKYPSHIKDFFKSAMPEGYVQERTITFEGDGVFKTRAEVTFE
NGSVYNRVKLNGQGFKKDGHILRKNVAFQCPPDMVYILPDTVNNGIRVEFNQAYDIEGVTEKLVTKCS
QMNRPLAGSAAVHIPRYHHLSKHTKLSKDRDERRDHMCLVEVVKAVDLDTYQ

YetiYFP (28,568 Da)
MGSSHHHHHHSSGLVPRGSHMTALTEGAKLFEKEIPYITELEGDVEGMKFIIKGEGTGDASVGKIDAQ
FICTTGDVPVPWSTLVTTLTYGAQCFAKYPRHIADFFKSCMPEGYVQERTITFEGDGVFKTRAEVTFEN
GSVYNRVKLNGQGFKKDGHVLGKNLEFNFTPHCLYIWGDQANHGLKSAFKIMHEITGSKEDFIVADH
TQMNTPIGGGPVHVPEYHHITYHVTLSKDVTDHRDHLNIVEVIKAVDLETYR

eGFP (28,211 Da)
MKHHHHHHHMRSGKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVP
WPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRI
ELKGIDFKEDGNILGHKLEYNYNGHNVYITADKQKNGIKANFKIRHNIEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

A.4 NMR Spectra

A.4.1 Enzyme Substrate Synthesis

A.4.1.1 [(Ethoxycarbonyl)methylene]triphenylphosphine (11)

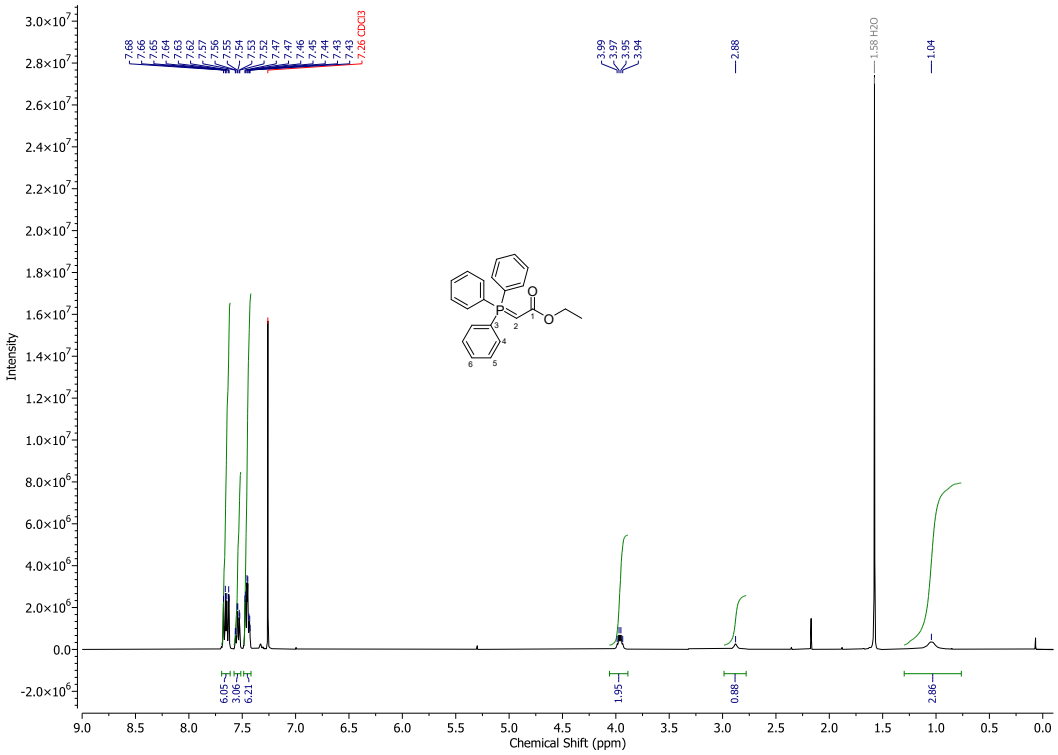


Figure A.4: 11 ^{1}H NMR

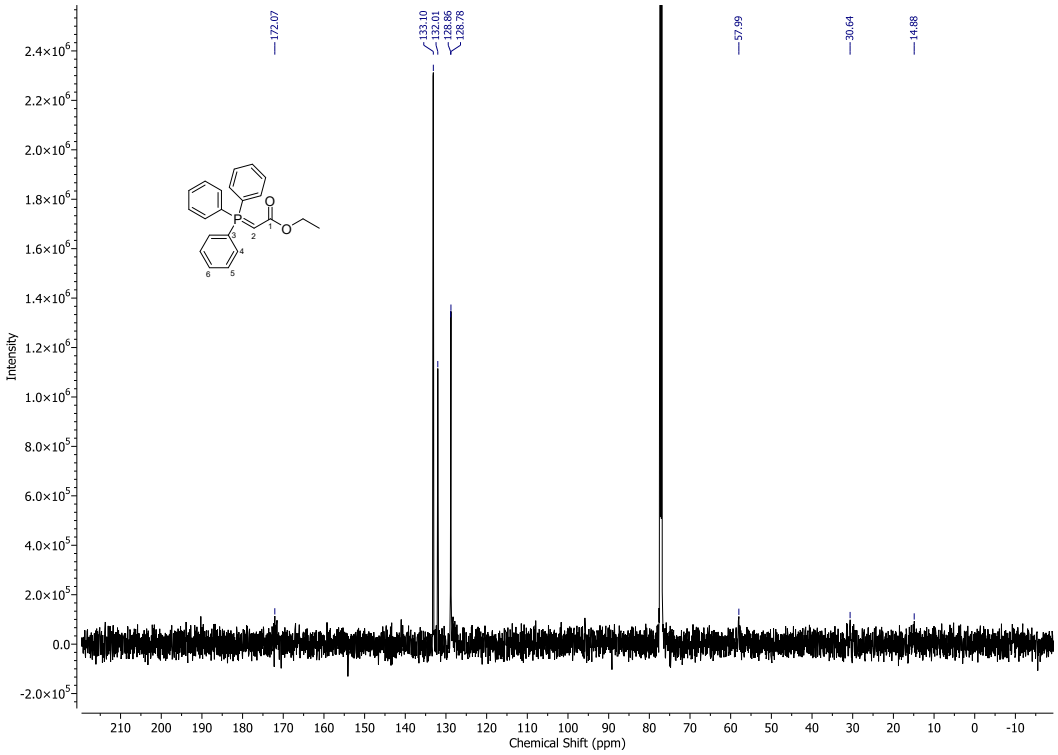


Figure A.5: 11 ^{13}C NMR

A.4.1.2 Ethyl (E)-6-hydroxyhex-2-enoate (12)

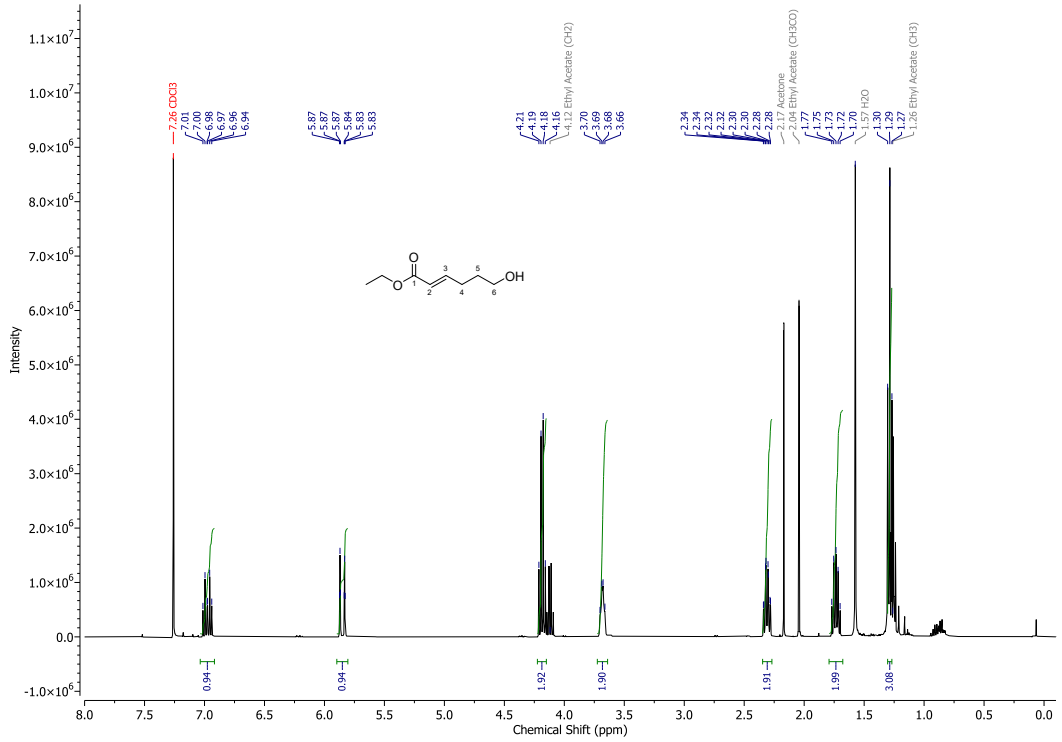


Figure A.6: 12 ^{1}H NMR

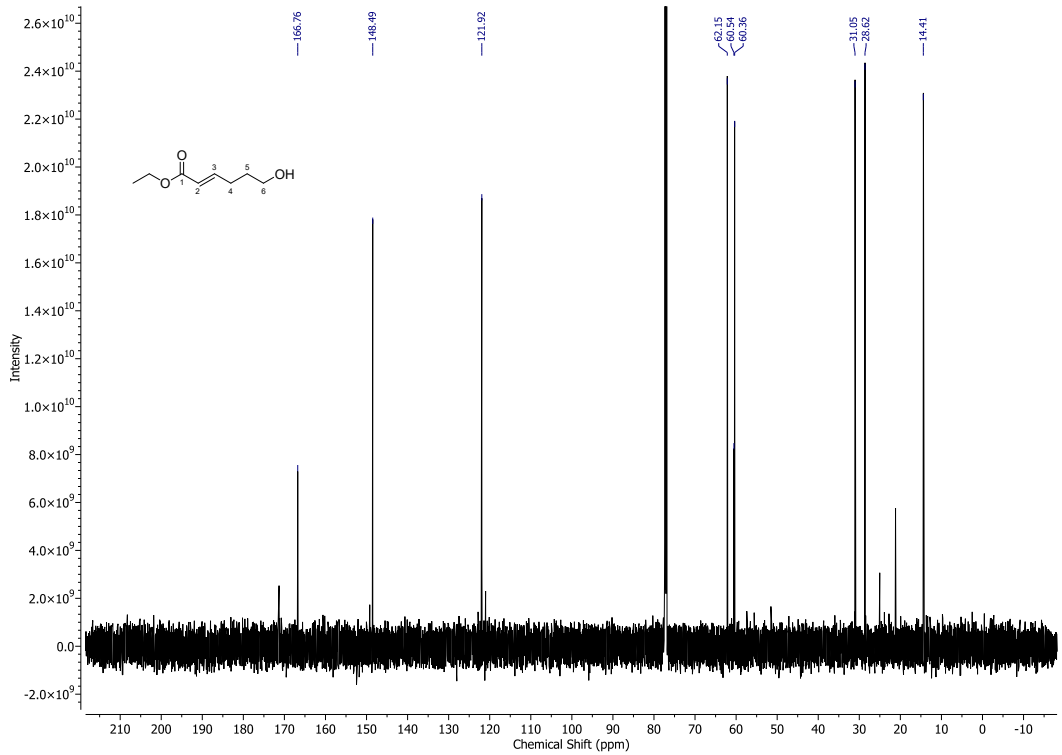


Figure A.7: 12 ^{13}C NMR

A.4.1.3 Diethyl (2E,6E)-octa-2,6-dienedioate (13)

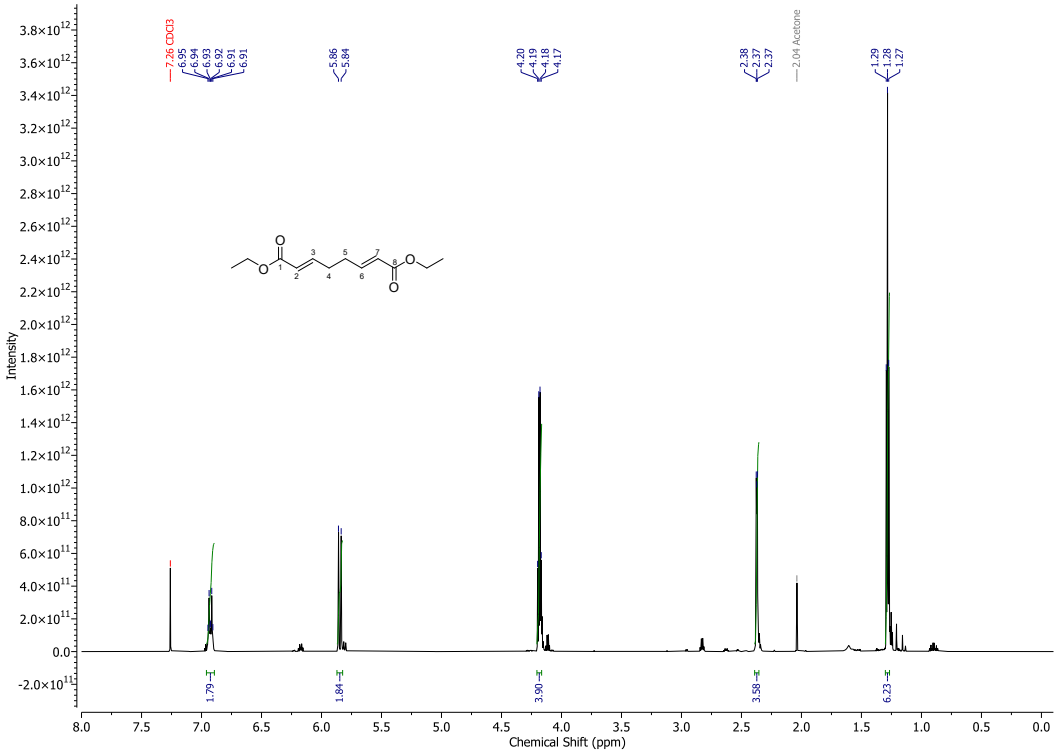


Figure A.8: 13 ^{1}H NMR

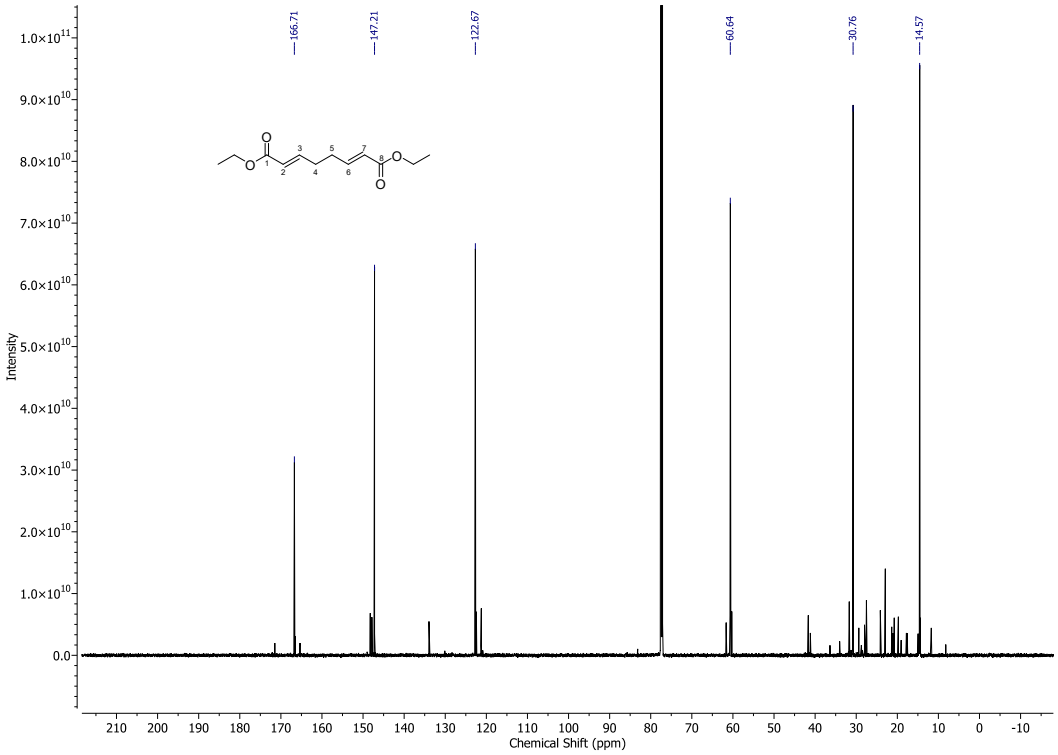


Figure A.9: 13 ^{13}C NMR

A.4.1.4 (2E,6E)-Octa-2,6-diene-1,8-diol (5)

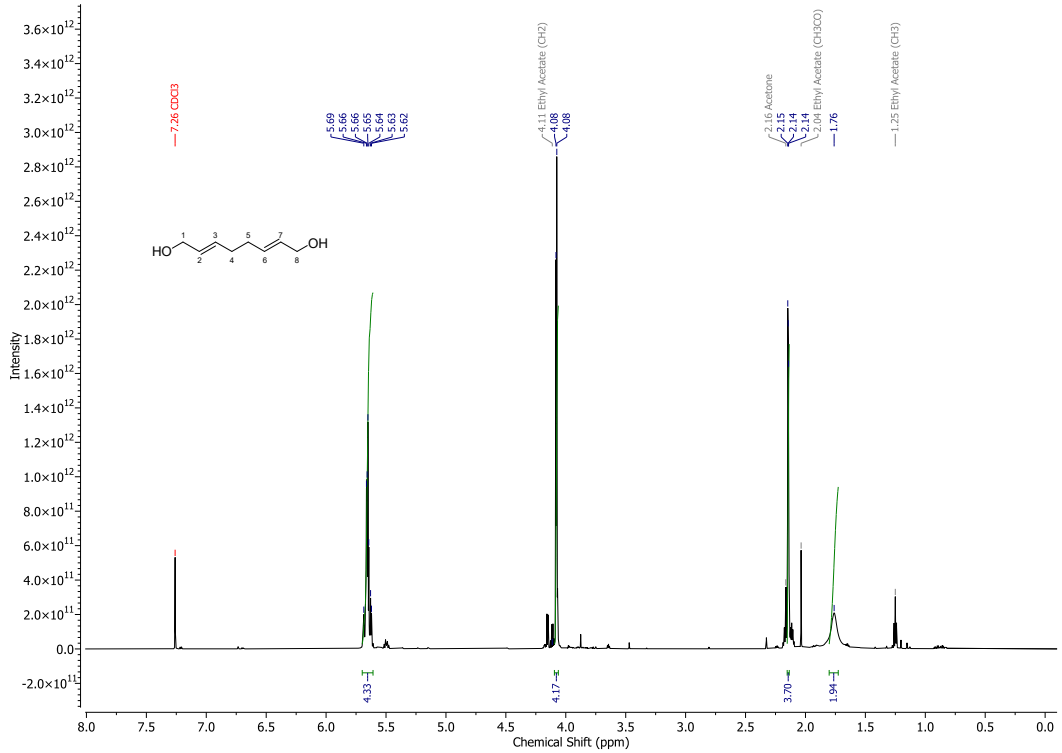


Figure A.10: 5 ^{1}H NMR

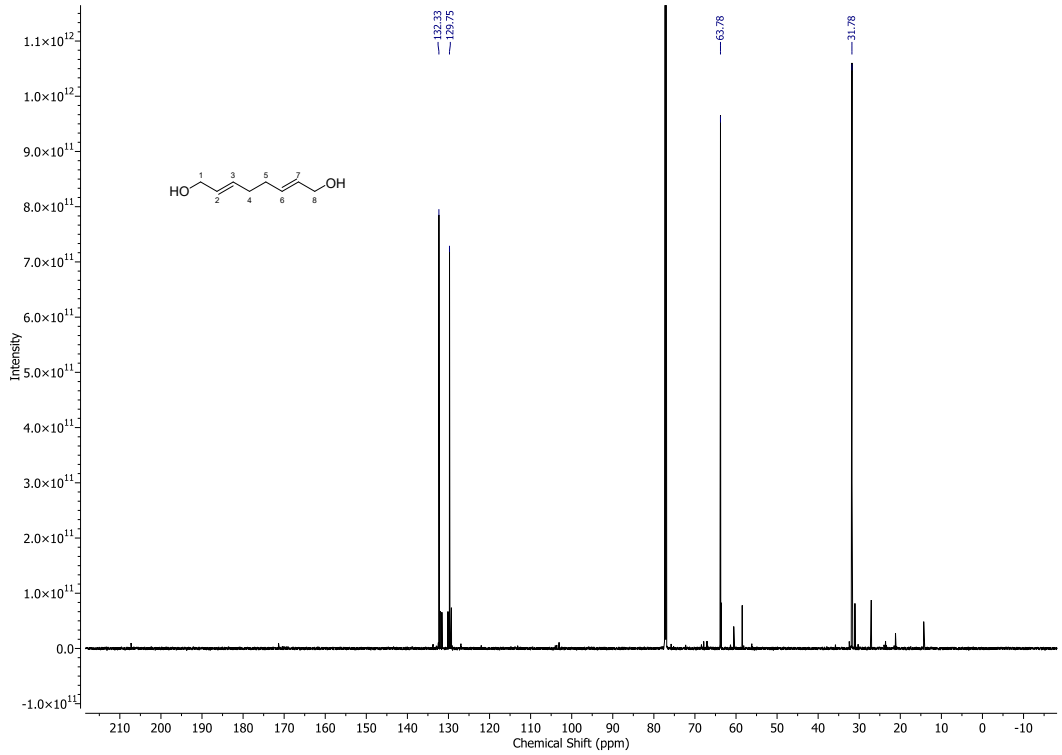


Figure A.11: 5 ^{13}C NMR

A.4.1.5 (2E,6E)-Octa-2,6-dienedial (6)

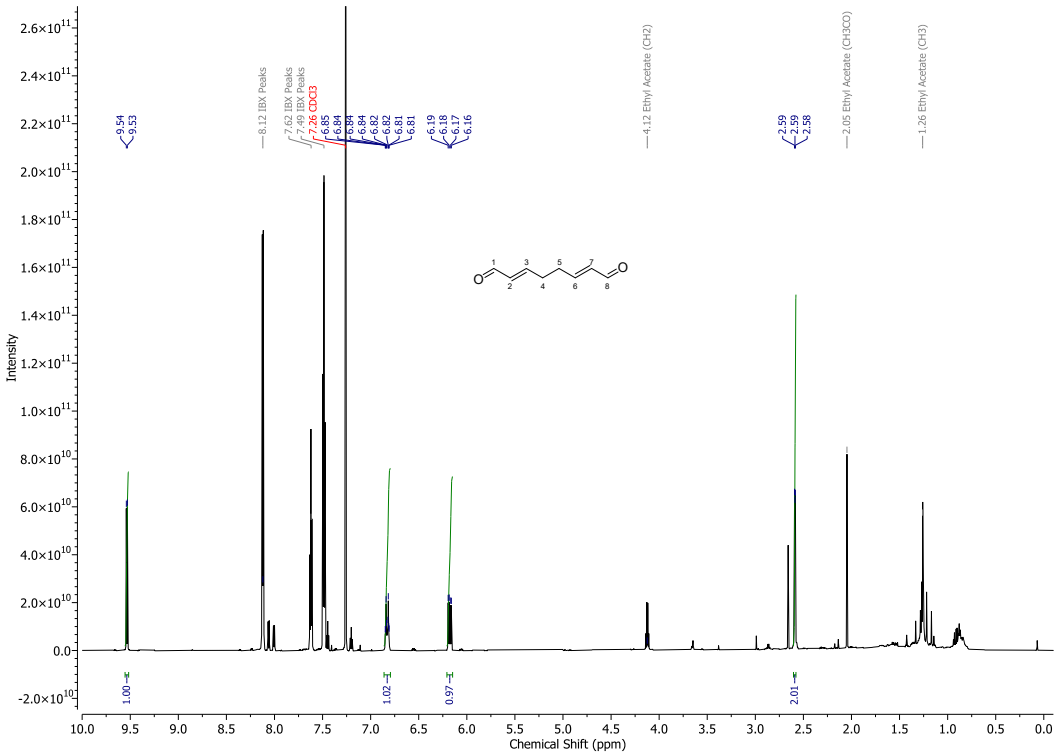


Figure A.12: 6 ^{1}H NMR

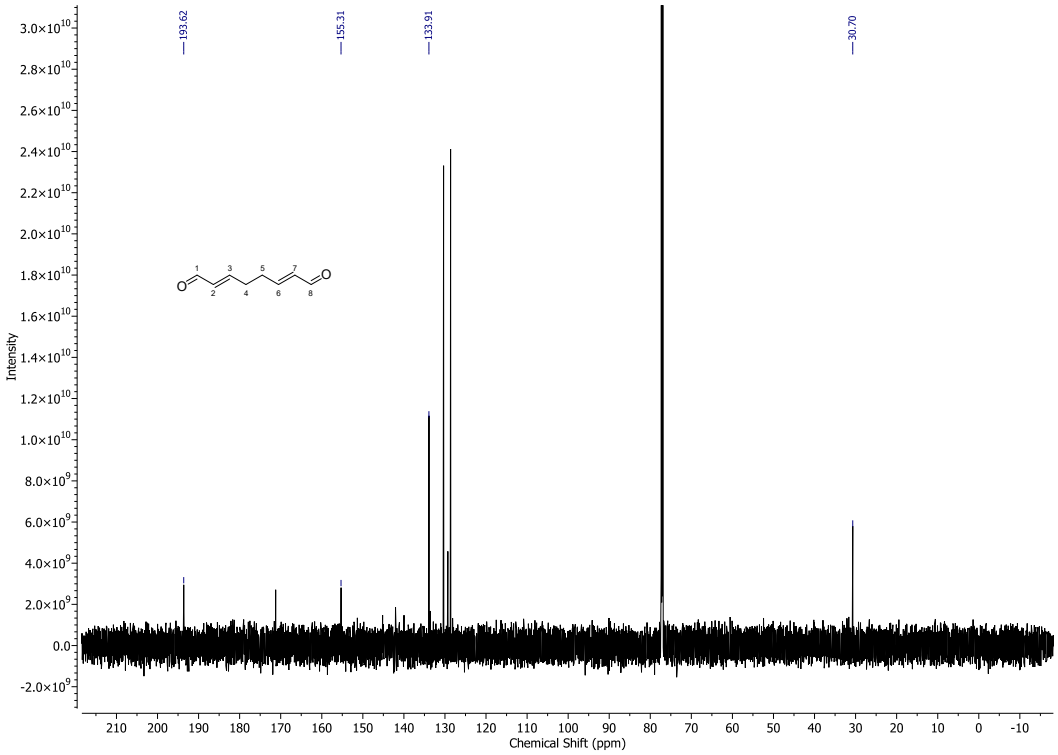


Figure A.13: 6 ^{13}C NMR

A.4.1.6 (2E,7E)-nona-2,7-dienedioate (18)

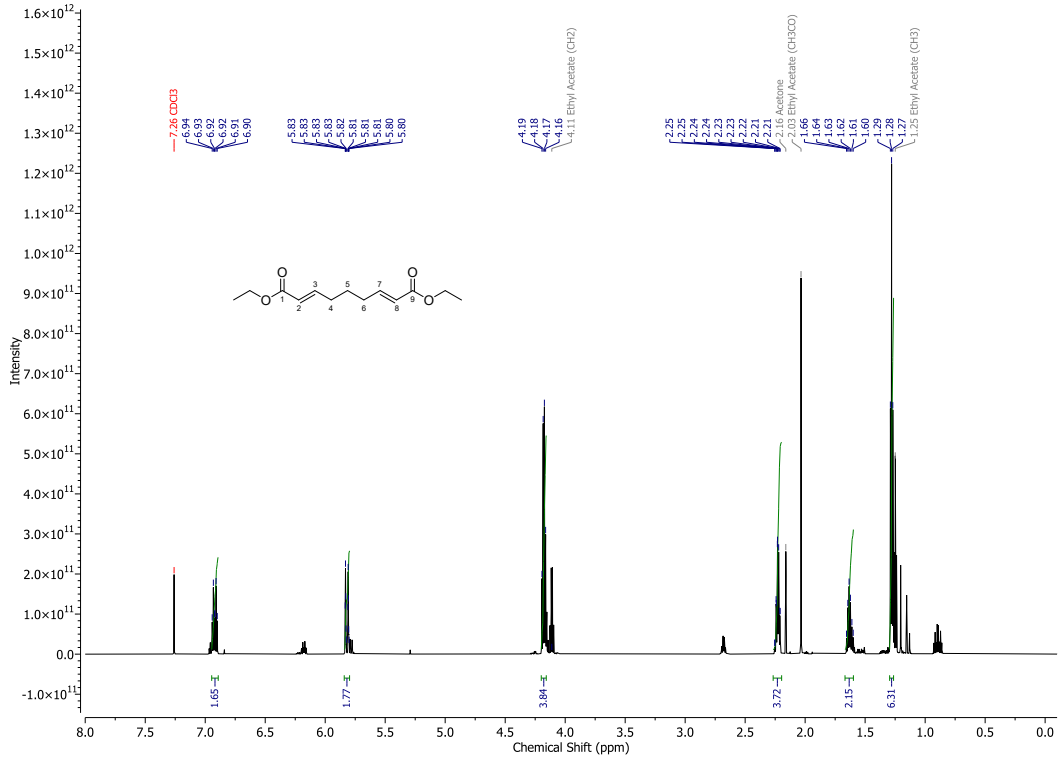


Figure A.14: 18 ^{1}H NMR

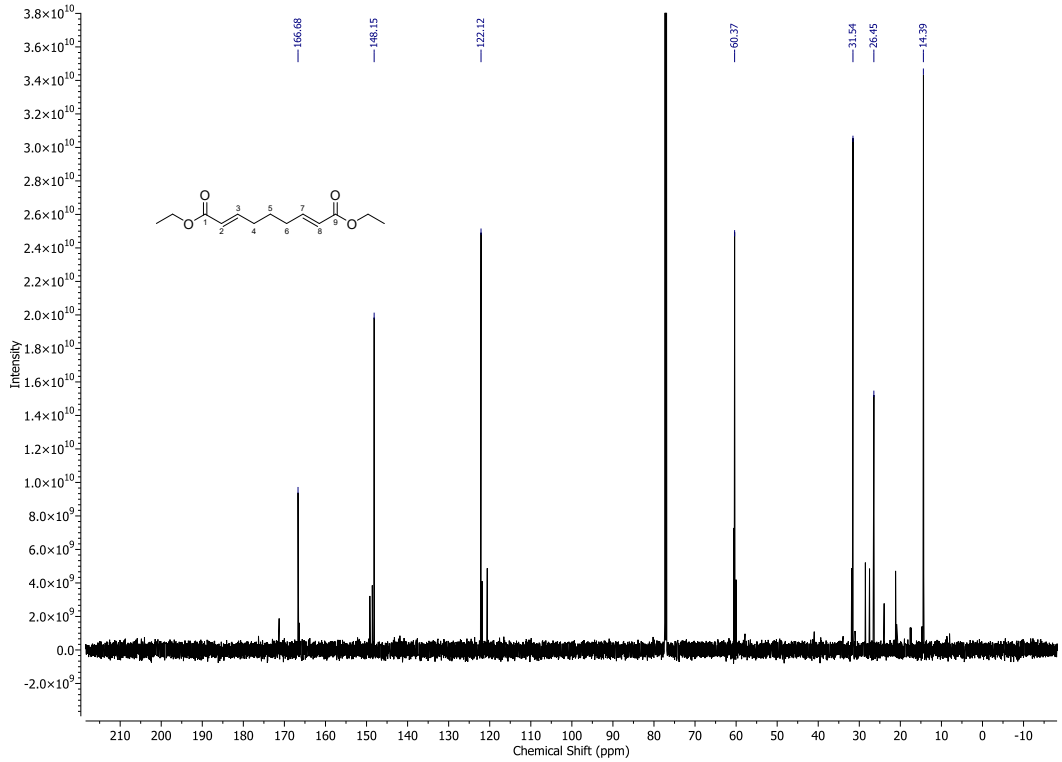


Figure A.15: 18 ^{13}C NMR

A.4.1.7 (2E,7E)-nona-2,7-diene-1,9-diol (19)

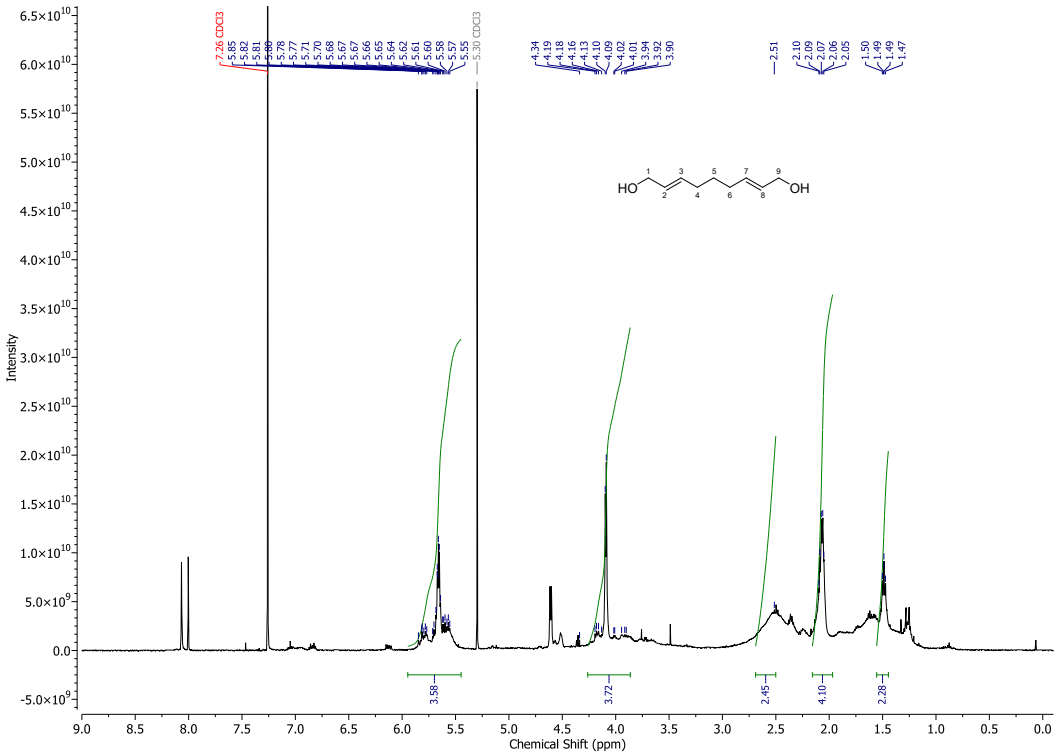


Figure A.16: 19 ^{1}H NMR

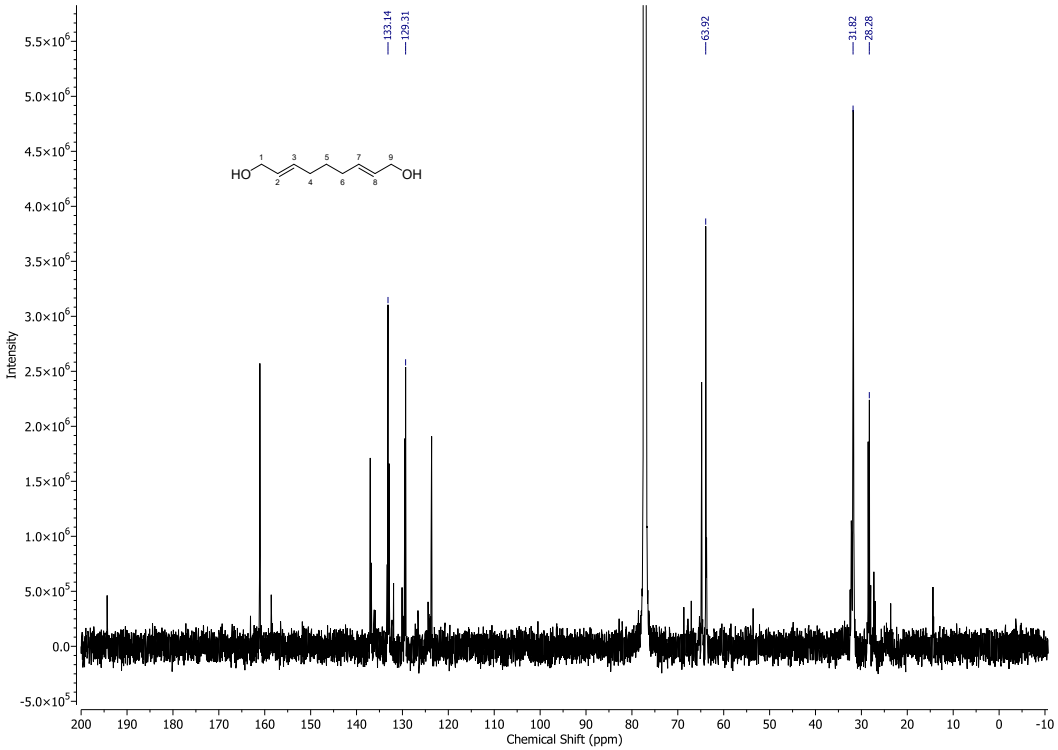


Figure A.17: 19 ^{13}C NMR

A.4.1.8 (2E,7E)-nona-2,7-dienedial (20)

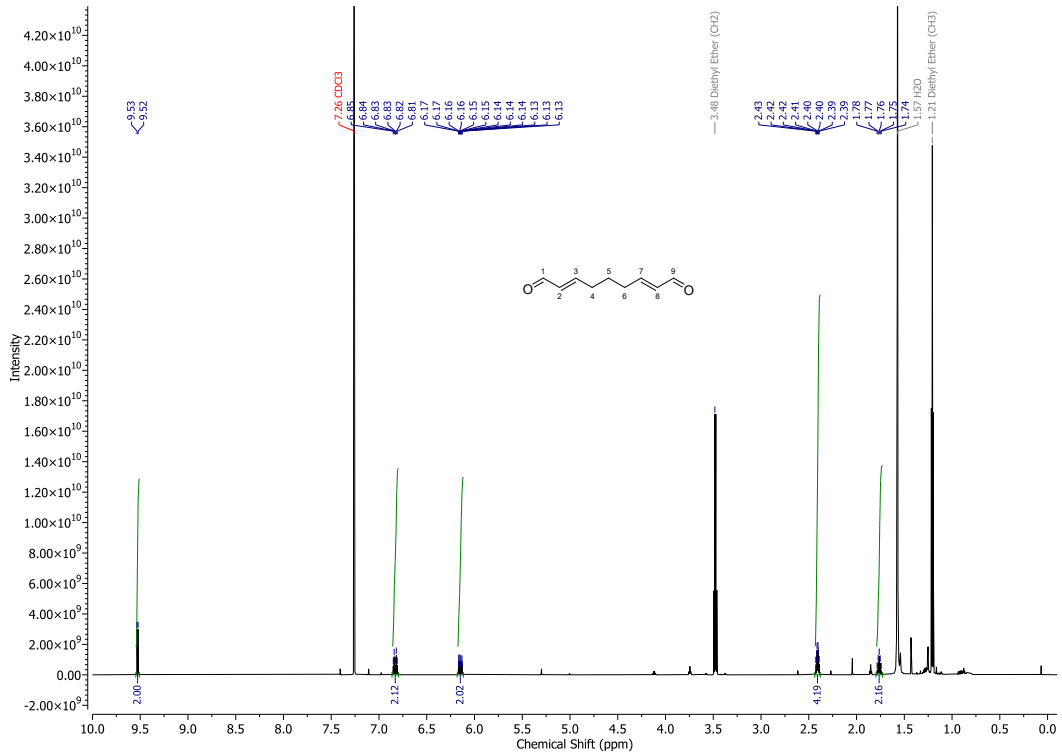


Figure A.18: 20 ^{1}H NMR

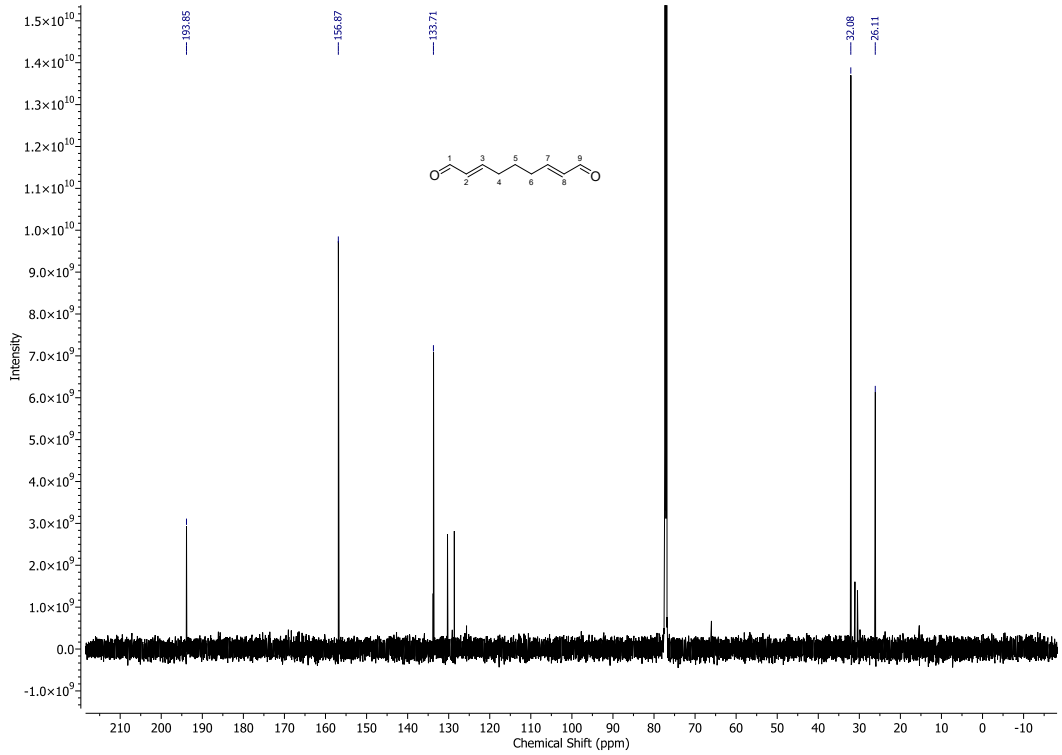


Figure A.19: 20 ^{13}C NMR

A.4.1.9 Diethyl 3,3'-(1,3-phenylene)(2E,2'E)-diacrylate (23)

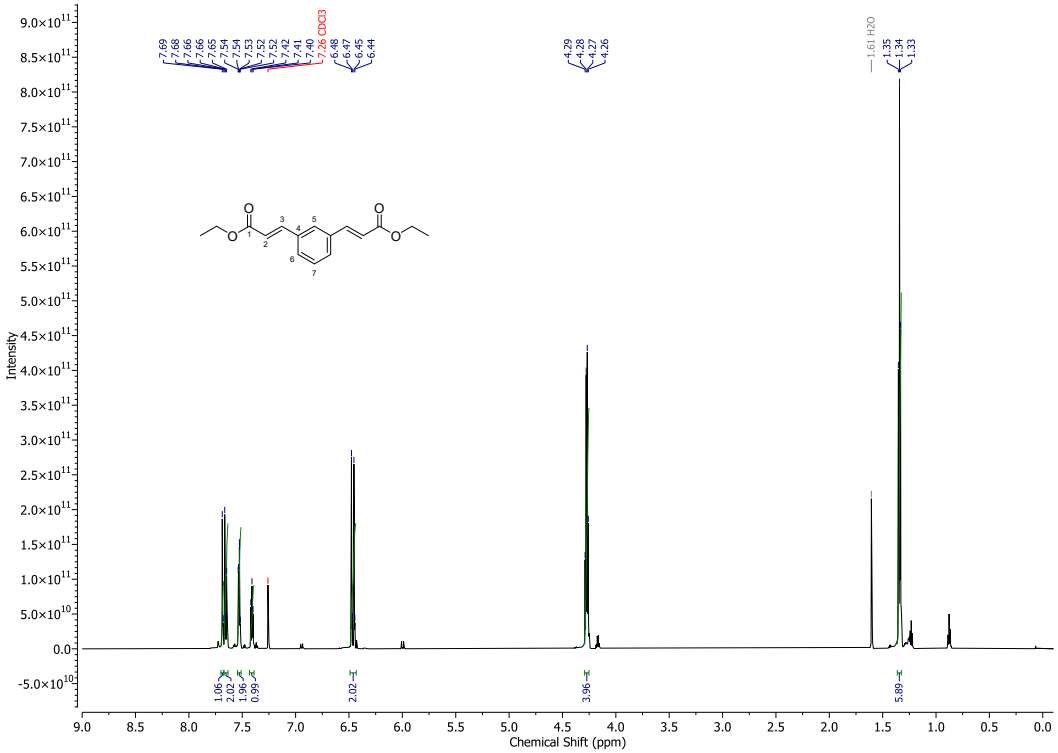


Figure A.20: 23 ^{1}H NMR

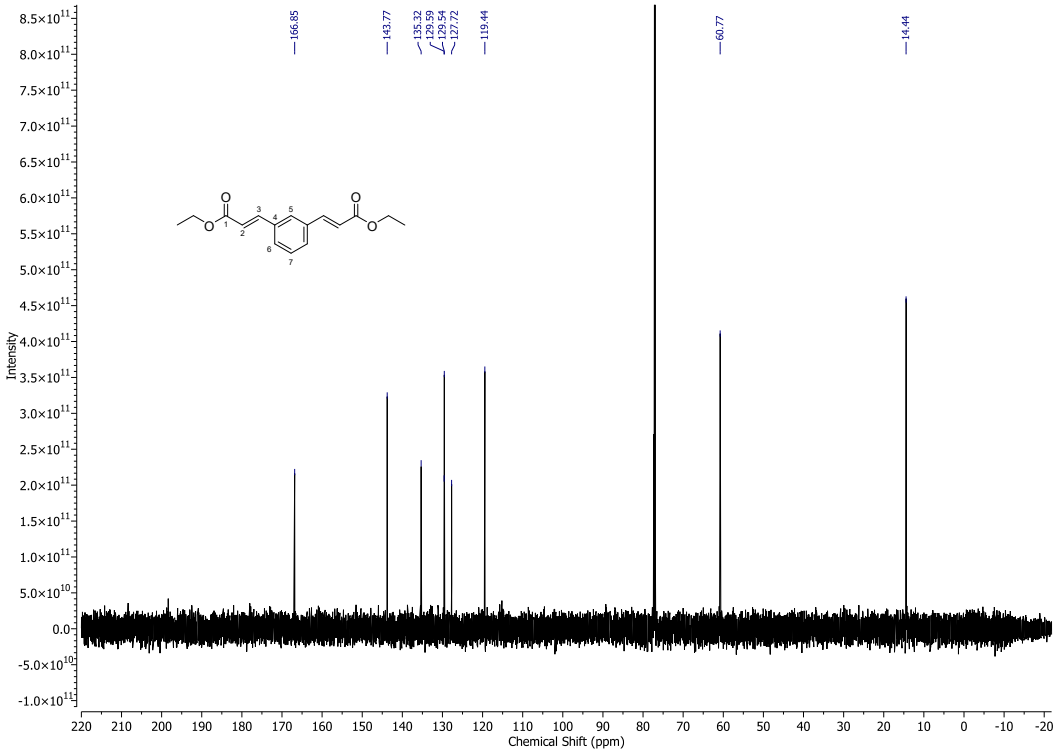


Figure A.21: 23 ^{13}C NMR

A.4.1.10 (2E,2'E)-3,3'-(1,3-phenylene)bis(prop-2-en-1-ol) (24)

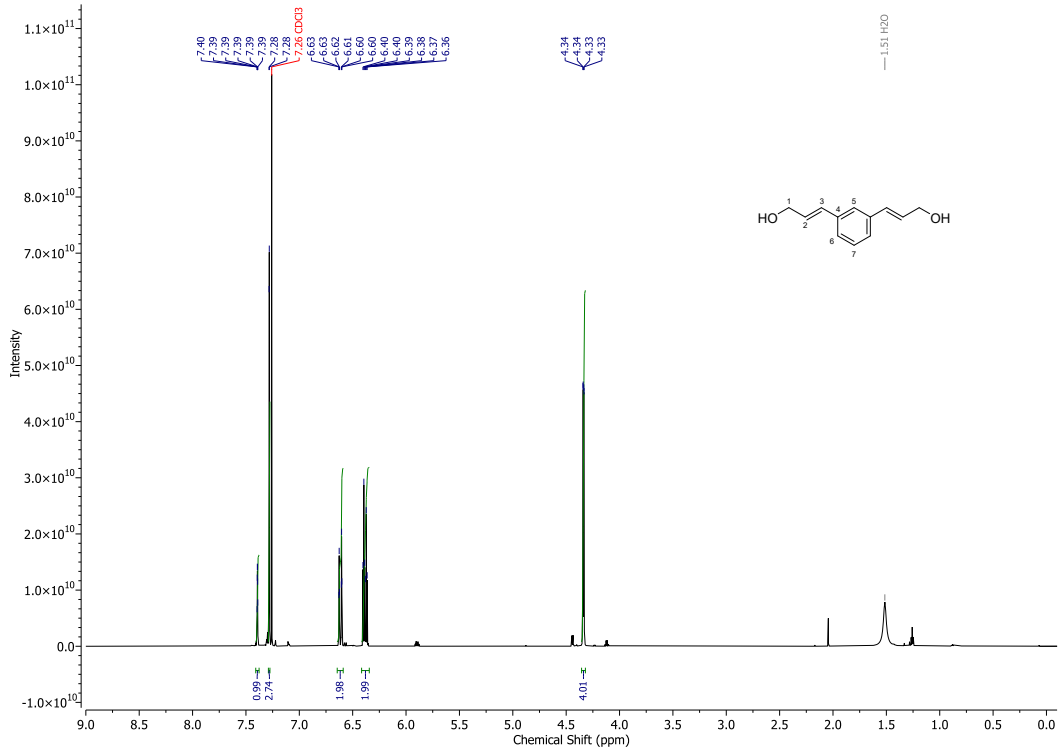


Figure A.22: 24 ^{1}H NMR

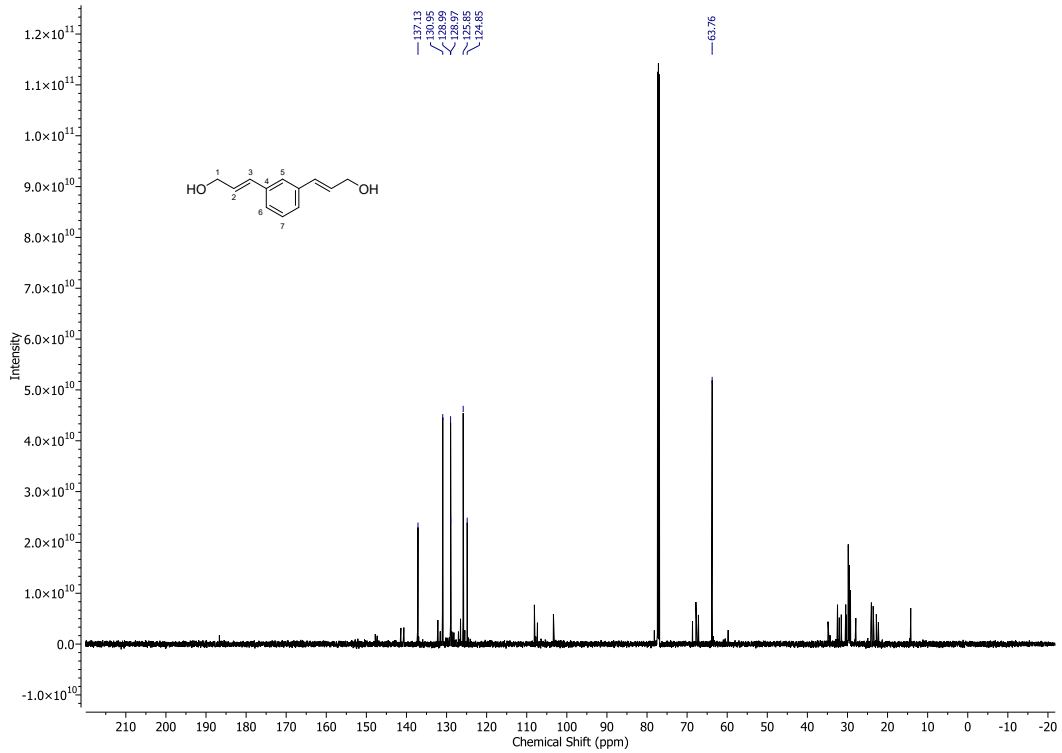


Figure A.23: 24 ^{13}C NMR

A.4.1.11 (2E,2'E)-3,3'-(1,3-phenylene)diacrylaldehyde (21)

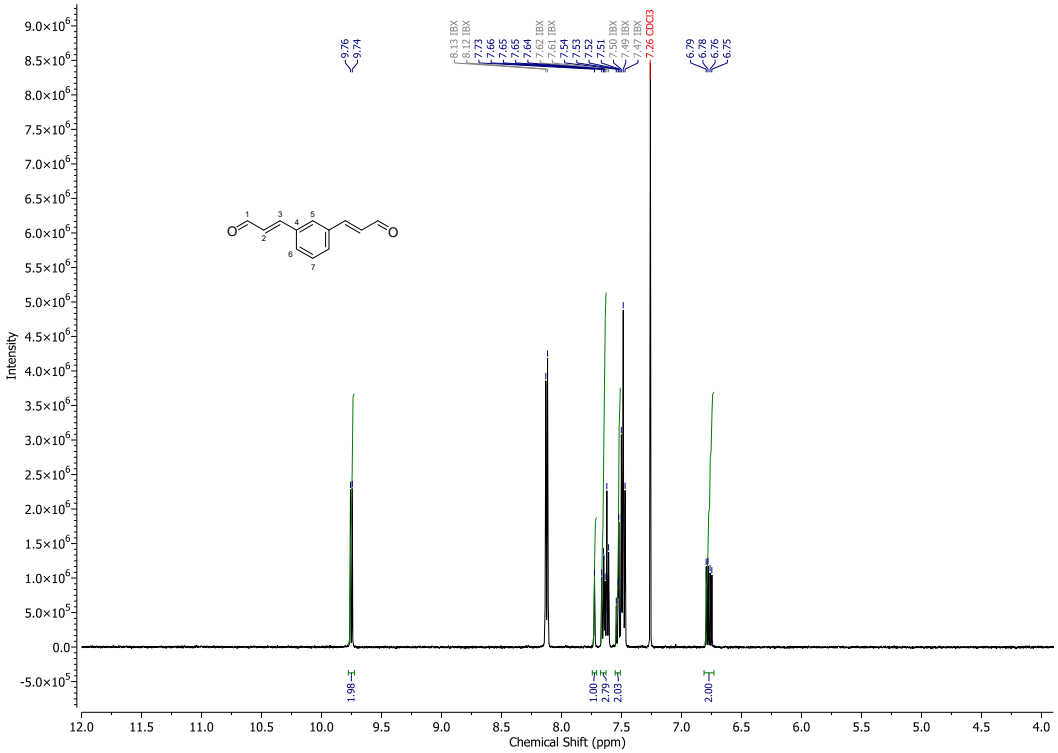


Figure A.24: 21 ^{1}H NMR

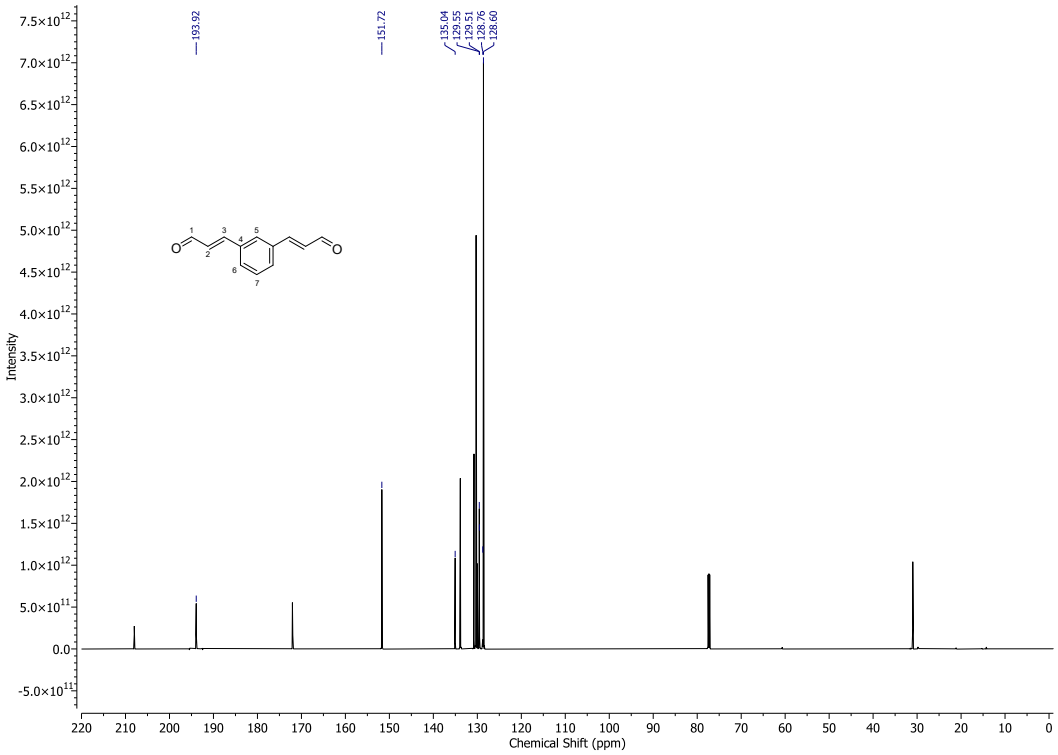


Figure A.25: 21 ^{13}C NMR

A.4.2 Cholesterol Linker Synthesis

A.4.2.1 (2-(2-(2-aminoethoxy)ethoxy)ethyl)carbamate (48)

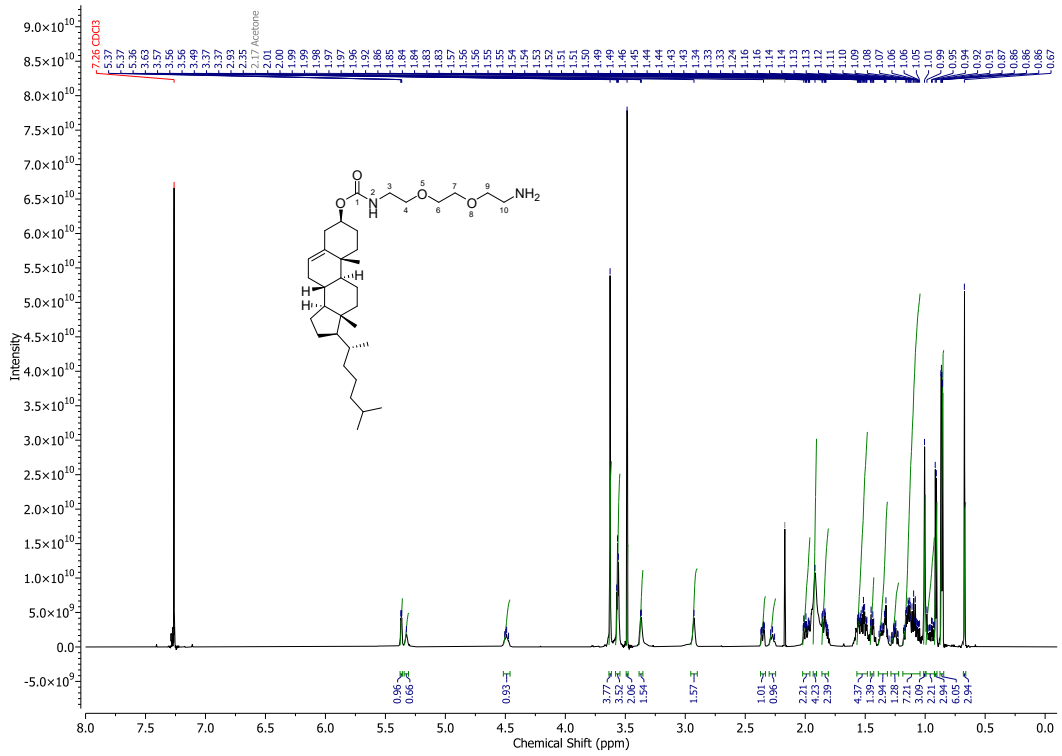


Figure A.26: 48 ^{1}H NMR

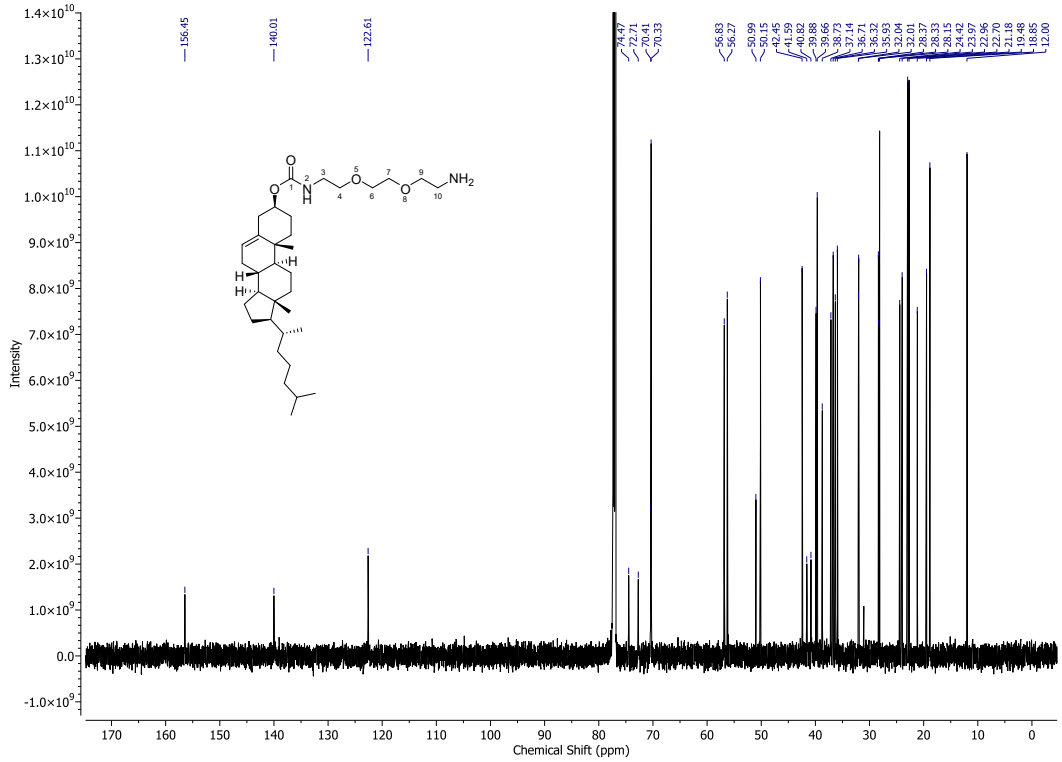


Figure A.27: 48 ^{13}C NMR

A.4.2.2 1-((Cholesteryl)oxy)-1,12-dioxo-5,8-dioxa-2,11-diazapentadecan-15-oic acid (50)

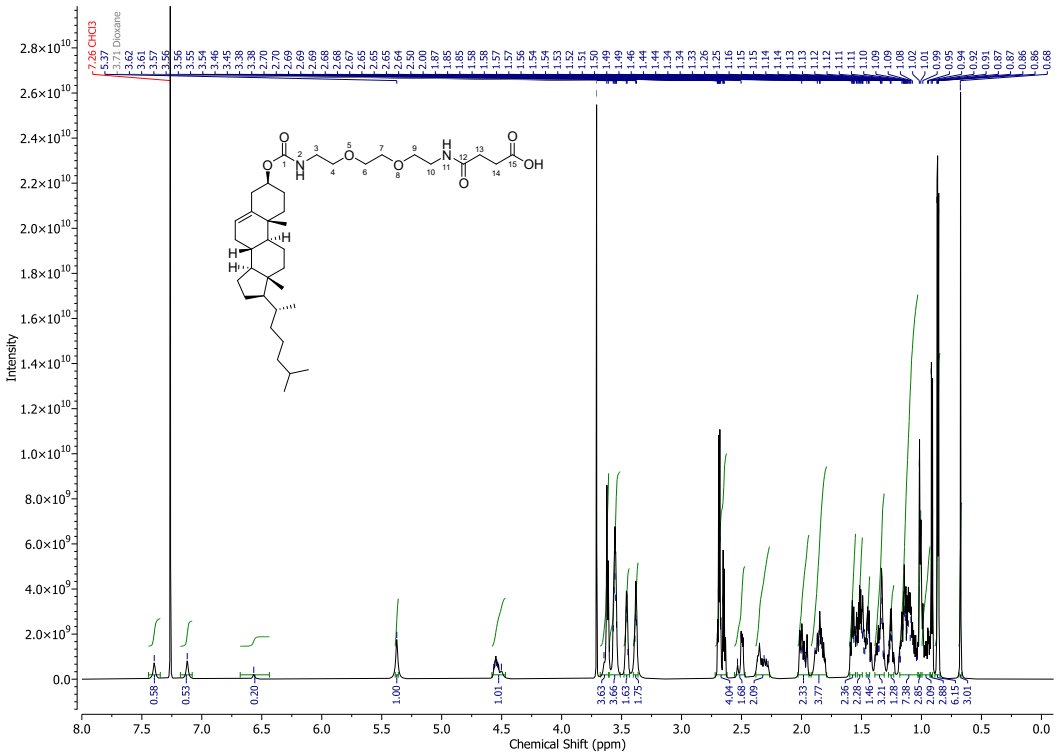


Figure A.28: 50 ^{1}H NMR

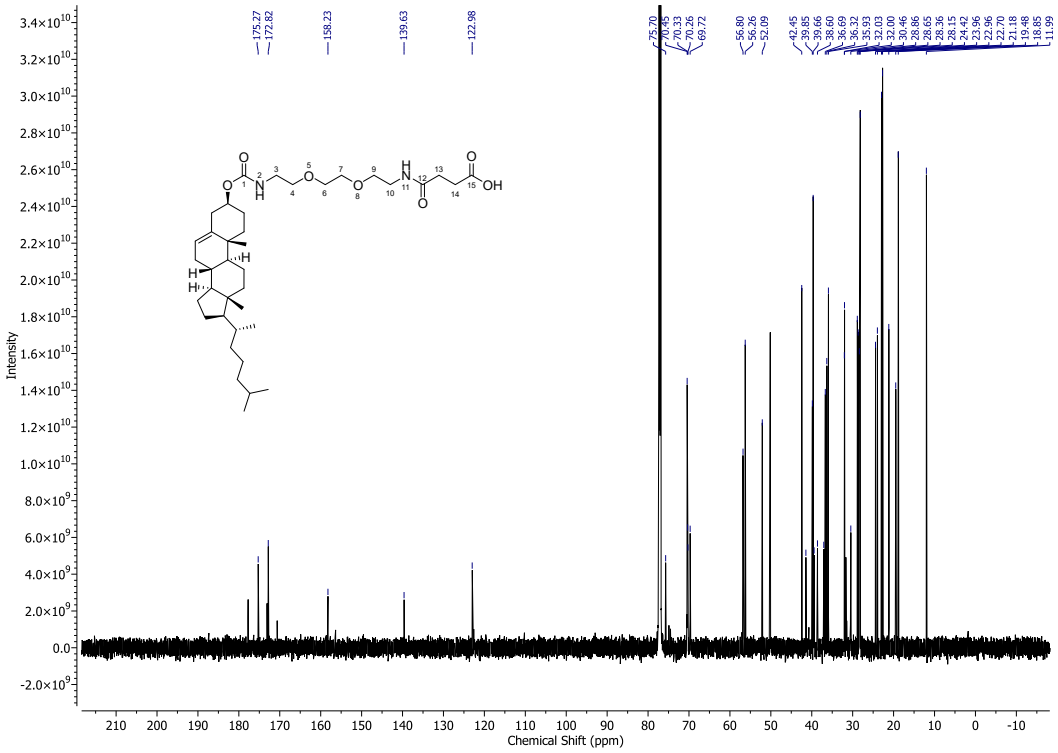


Figure A.29: 50 ^{13}C NMR

A.4.3 DNA Linker Synthesis

A.4.3.1 4-Maleimidobutyric acid (79)

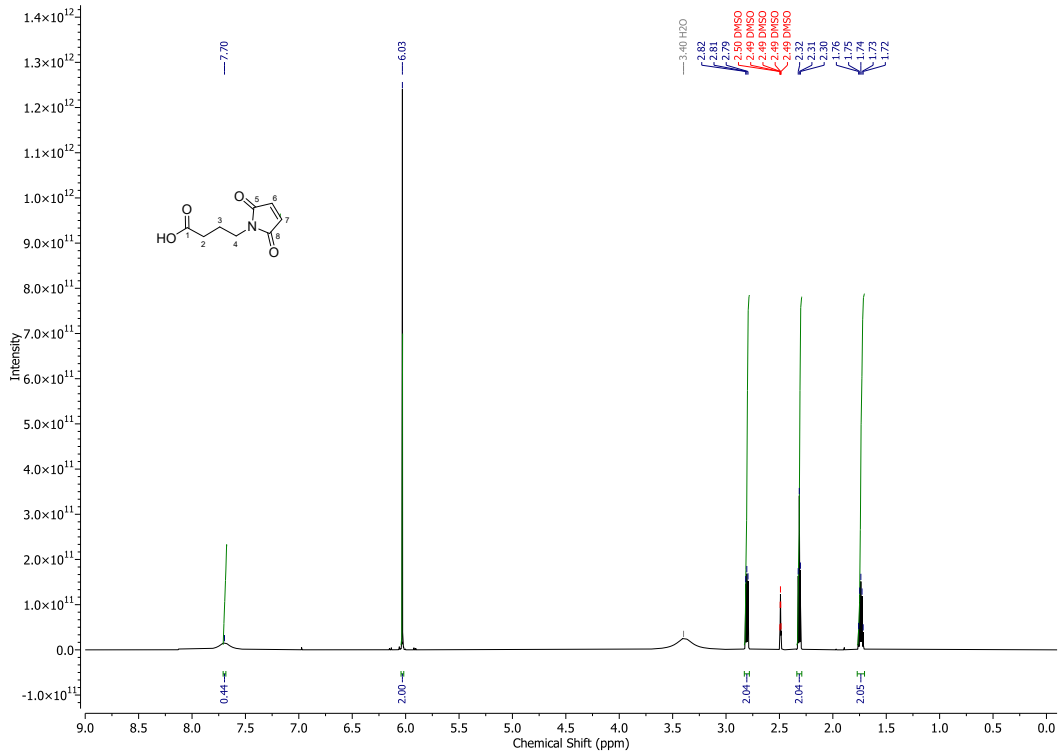


Figure A.30: 79 ^{1}H NMR

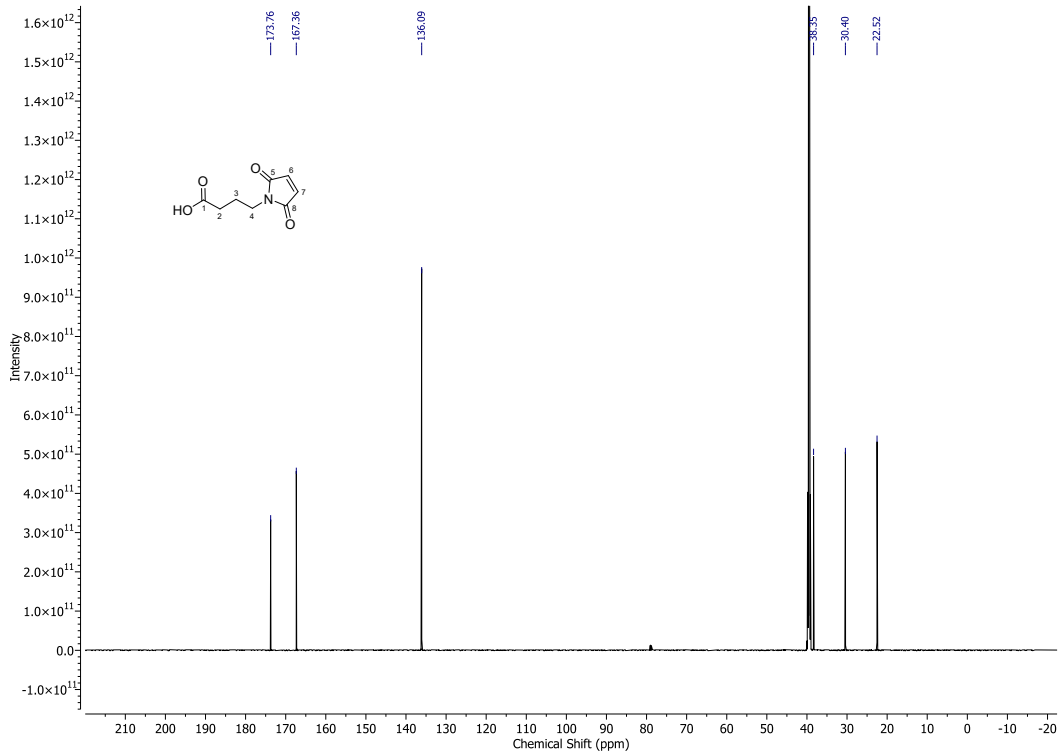


Figure A.31: 79 ^{13}C NMR

A.4.3.2 Dimethyl 2,2'-((6-amino-1-methoxy-1-oxohexan-2-yl)azanediyl)(S)-diacetate (81)

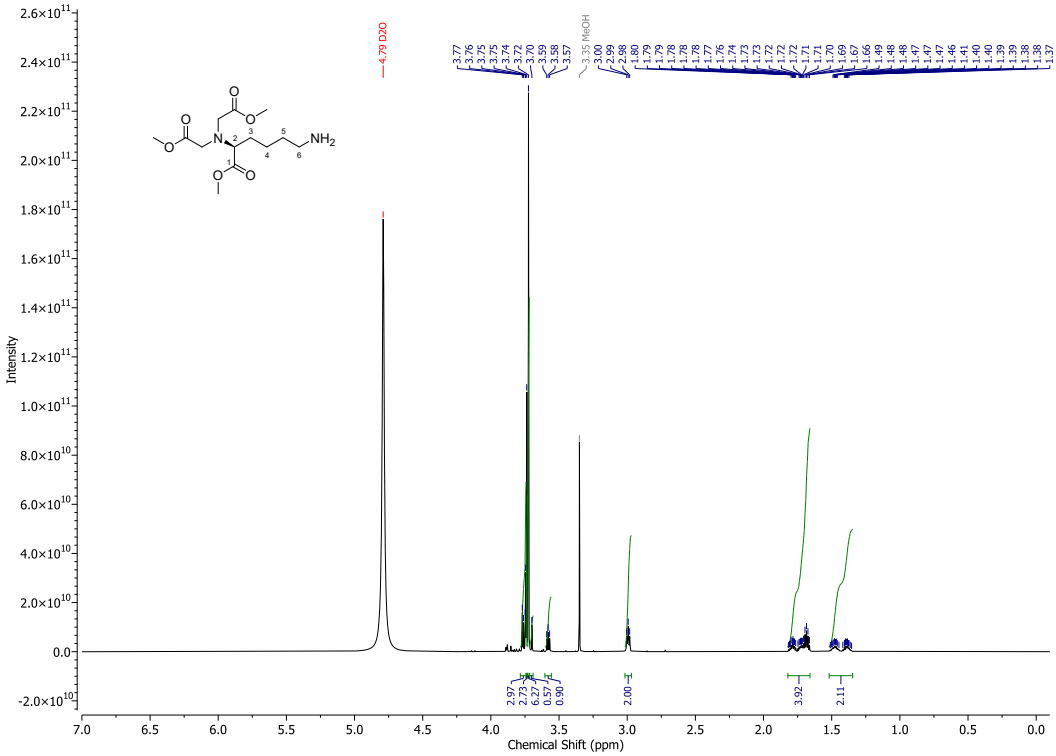


Figure A.32: 81 ^{1}H NMR

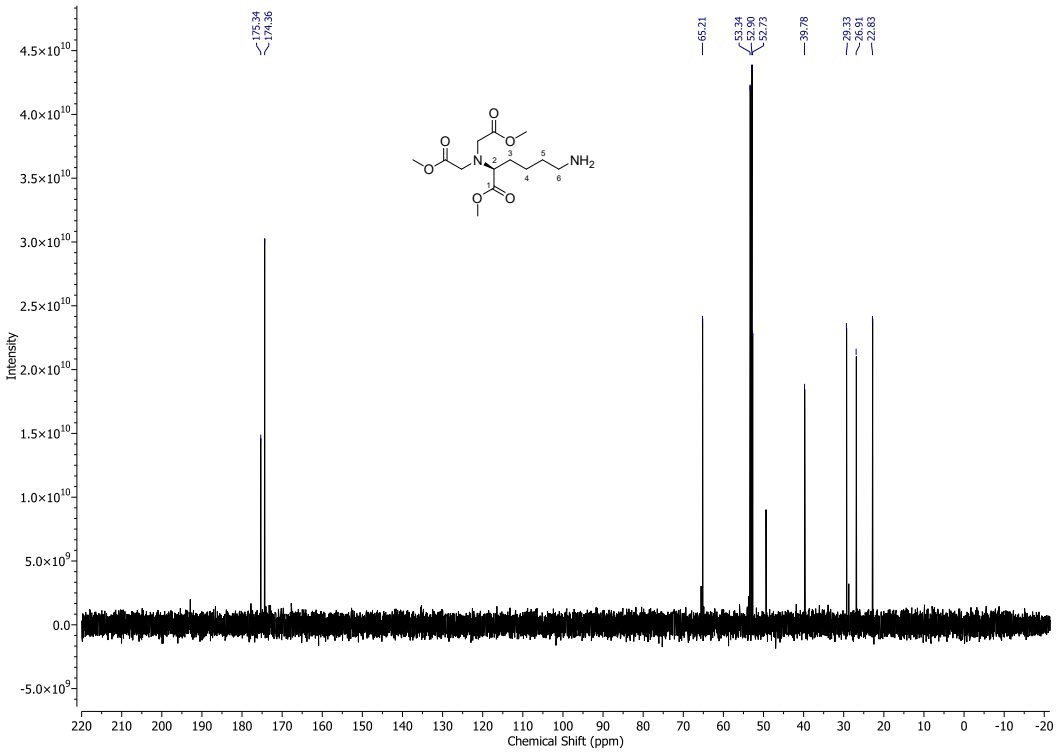


Figure A.33: 81 ^{13}C NMR

A.4.3.3 (S)-2,2'-((1-carboxy-5-(4-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)butanamido)pentyl)azanediyl)diacetic acid (82)

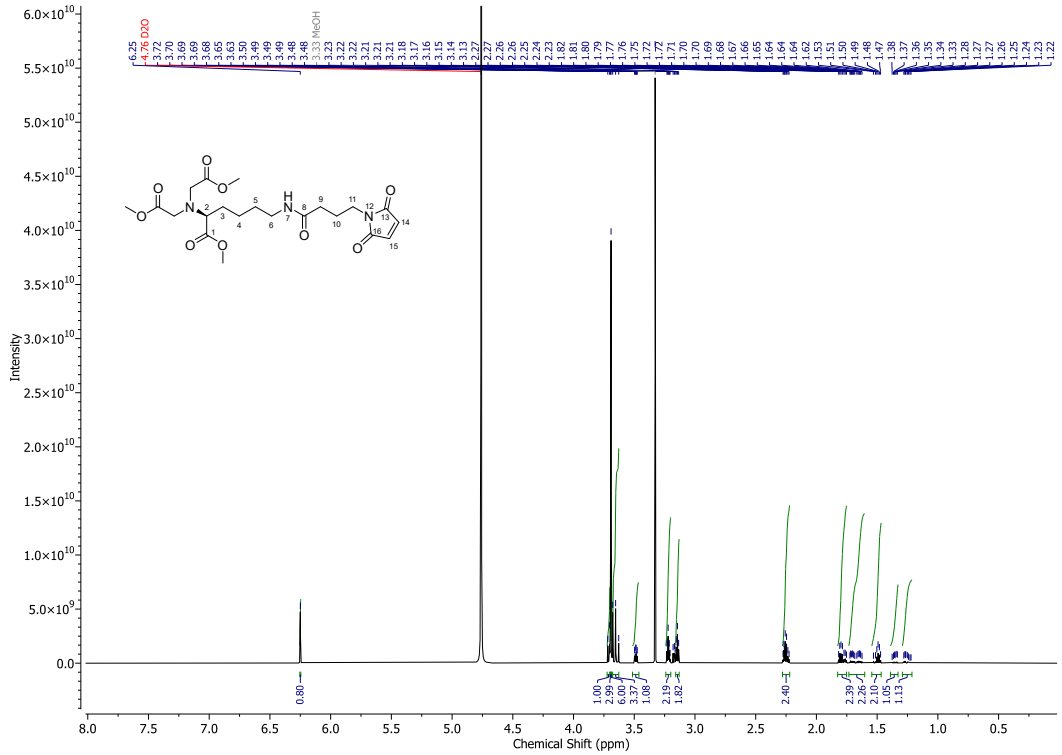


Figure A.34: 82 ^{1}H NMR

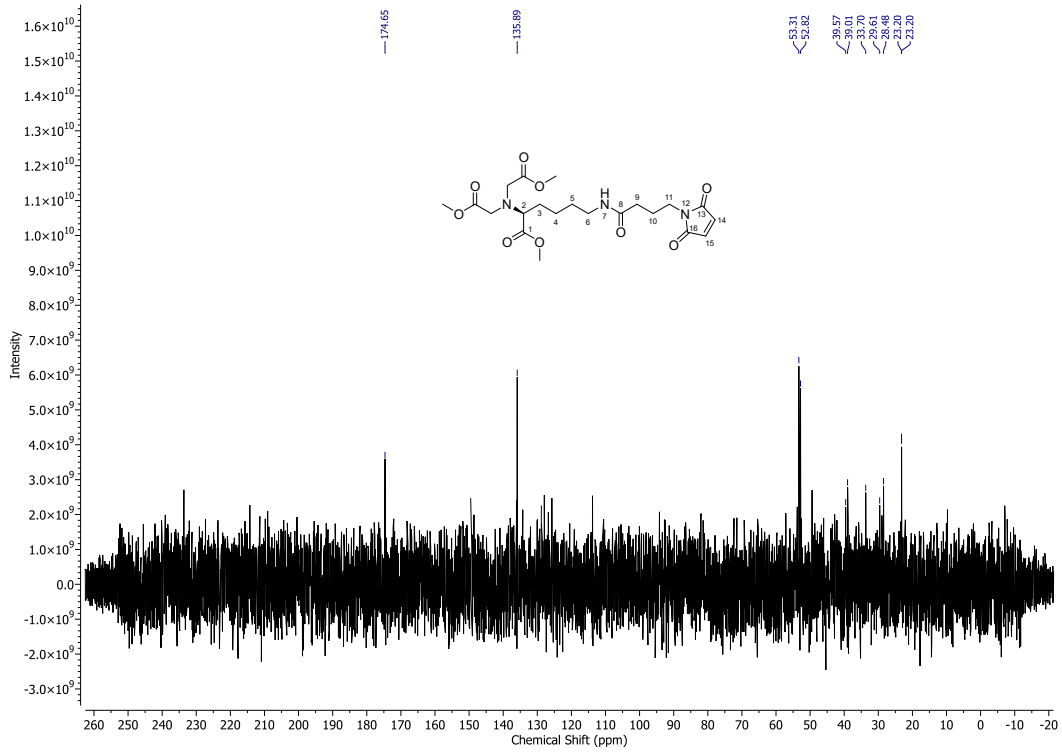


Figure A.35: 82 ^{1}H NMR

A.4.3.4 (S)-2,2'-((1-carboxy-5-(4-(2,5-dioxo-2,5-dihydro-1H-pyrrol-1-yl)butanamido)pentyl)azanediyl)diacetic acid (76)

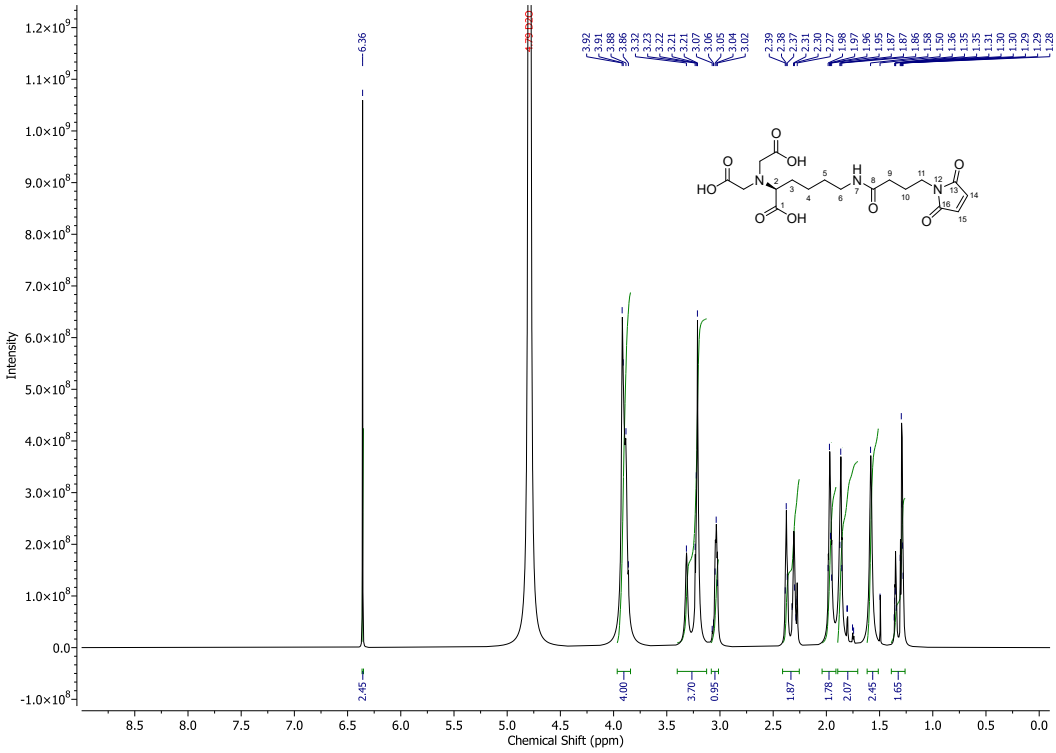


Figure A.36: 76 ^{1}H NMR

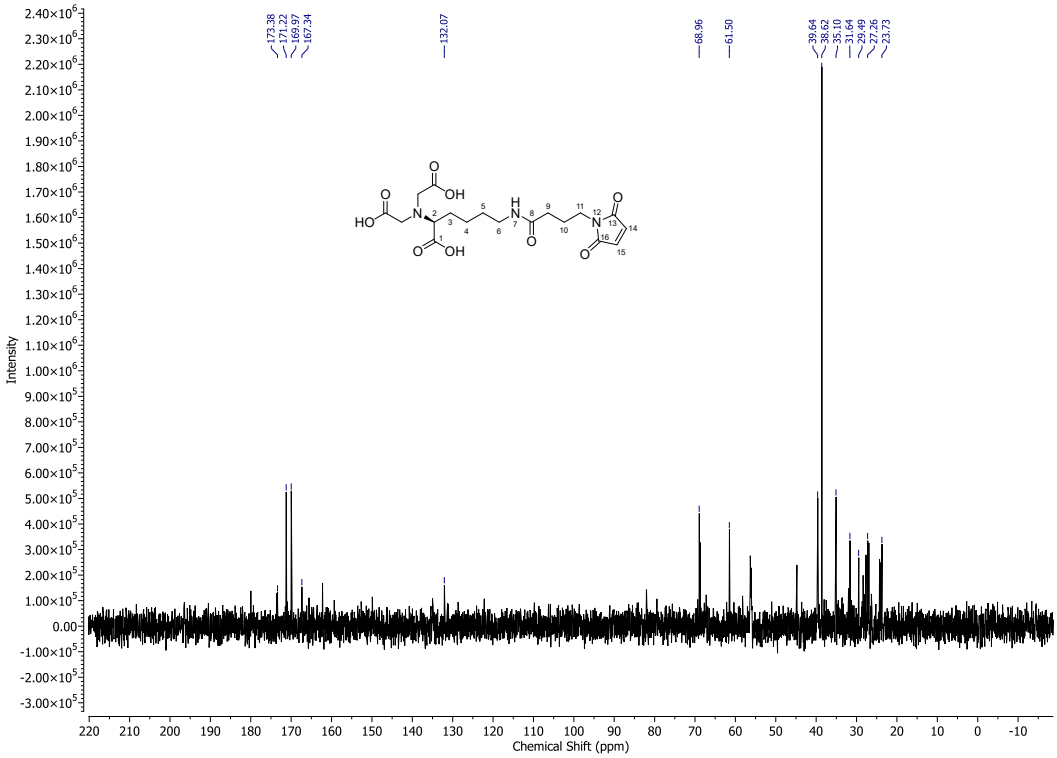


Figure A.37: 76 ^{1}H NMR