Expanding the Scope of DNA Compatible Chemistry for the Application within DNA Encoded Libraries

PhD Thesis

James H. Hunter

Supervisor:

Prof. Michael J. Waring

Declaration

The work described in this thesis was carried out between September 2016 and November 2020 in the Medicinal Chemistry Laboratories (Bedson building, Newcastle centre of Cancer Research, Newcastle University, Newcastle-Upon-Tyne, UK, NE1 7RU), the Cancer Structural Biology Laboratories (Paul O'Gorman Building, Newcastle centre of Cancer Research, Newcastle University, Newcastle-Upon-Tyne, UK, NE2 4HH) and AstraZeneca Mölndal Laboratories (Pepparedsleden 1, 431 50 Mölndal, Sweden). The research was conducted in collaboration with and funded by AstraZeneca.

All the research described in this thesis is original in context and does not incorporate material or ideas previously published or presented by other authors except where due reference is given in the text.

No part of this thesis has been previously submitted for a degree, diploma, or any other qualification at any other University.

Acknowledgements

Firstly, I would like to thank my PhD supervisor, Prof. Mike Waring for all his support and encouragement over the past 4 years. His advice and guidance have been invaluable over the course of my PhD and it has been a great privilege to have his supervision. My appreciation also extends to my Industrial supervisors Dr Garry Pairaudeau and Dr Andy Madin for all their knowledge and input on the project and the incredible opportunities they have given me. I would also like to thank Prof. Bernard Golding, Dr Céline Cano and Dr Ian Hardcastle for their support, encouragement, and expertise throughout the project.

I would like to thank all members, past and present, in the NICR medicinal chemistry team. It has been an absolute pleasure to work with you all and grow together; Dr Duncan Miller, Dr Suzannah Harnor, Dr Hannah Stewart, Dr Stephen Hobson, Dr James Sanderson, Dr Kirsty Wilson, Dr Jennyfer Goujon-Ricci, Dr Kate McAdam, Dr Christine Basmadjian, Dr Greg Aldred, Dr Harry Shrives, Dr Cinzia Bordoni, Dr Marie Lawson, Dr Mohammed Al Yasiri, Dr Jane Totobenazara, Dr Edwige Picazo, Dr Alfie Brennan, Dr Amy Heptinstall, Dr Alexia Papaionannou, Dr Dani Lopez, Dr Shaun Stevens, Conghao Gai, Gemma Davison, Mélanie Uguen, Shaimaa Khalifa, Ayaz Ahmad, Islam Al-Khawaldeh, Selma Dormen, Luisa Camerin, Jess Graham, Catherine Salvini, Isaline Castan and Lucas Sprenger. I would also like to thank the Newcastle University chemistry NMR and crystallography teams for their help in some of the difficult analyses; Prof. William McFarlane, Dr Corinne Wills and Dr Paul Waddell.

I would like to thank the biological NICR team for their help with some of the DNA related aspects of the project; Prof Martin Noble, Dr Mathew Martin, Dr Ian Hickson, Dr Lisa Prendergast, Dr Richard Heath, Dr Hilal Sarac, and Dr Tom McAllister. Their help with DNA design, ligations and sequencing was invaluable. I would additionally like to thank Dr Michael Kossenjans and the iLab in AstraZeneca Gothenburg for their hospitality and help during my brief time in their lab. My thanks also extend to Lou Valente for his expert knowledge on JMP and factorial experiment design.

I would like to acknowledge AstraZeneca for providing the funding to carry out and perform this research, and to Cancer Research UK for funding our laboratories.

A huge thank you goes to Lish Lin, Dan Brough and Pasquale Morese for making my time in Newcastle amazing, even though you all support ridiculous football teams, and to my “mint

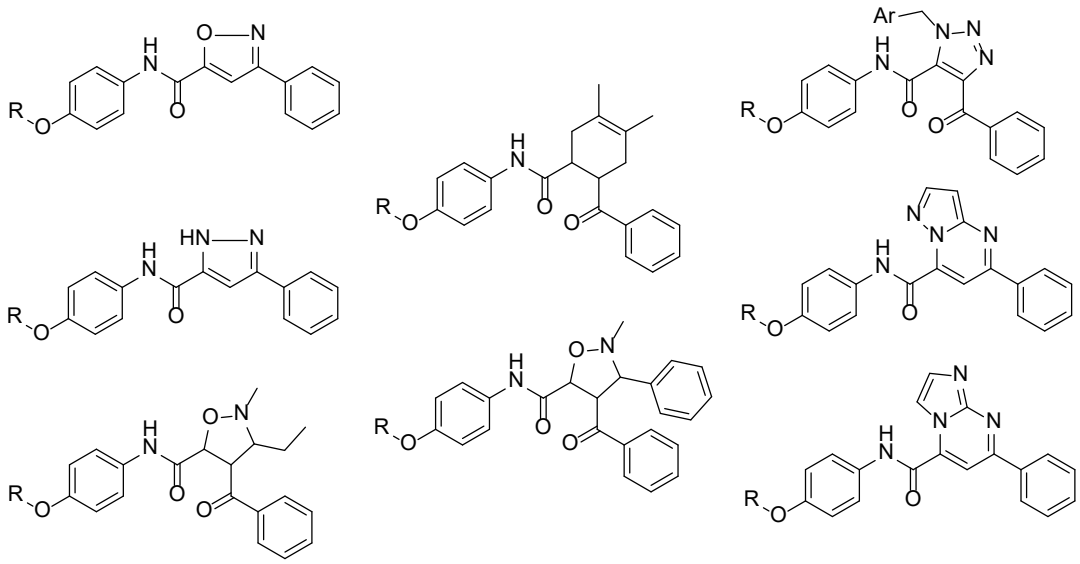
colleague" Harriet Stanway for keeping me sane throughout the last 2 years. You will all be missed immensely!

I would like to recognize my amazing friends. Even though this long trek up North has meant I only see you a couple of times a year, it always seems as though we have never been apart. You are all amazing and I couldn't have asked for a better group of people to call my friends. My amazing girlfriend Katy, who I wouldn't have even met without coming to do a PhD in these labs, I look forward to our next chapter together! And my brother Si who has always been there for me even though we always seem to be in different places.

Finally, I would like to dedicate this work to my Mum and Dad, I couldn't have asked for more supportive and loving parents and I wouldn't be the person I am today without them.

Abstract

The discovery of potential lead like molecules is a crucial phase for any drug discovery program. Current methods to identify lead molecules are often resource intensive, often requiring millions of compounds to be screened in biological assays. Reducing the time and cost taken to generate, store and screen large compound libraries would have a positive impact on academia and small pharmaceutical companies. DNA encoded libraries (DELs) aim to improve upon this by screening an entire library in a single vessel against a target and utilising the DNA tag to identify potential inhibitors. Current methods of preparing DELs are limited to chemistry that is compatible with DNA. These chemical methods are often limited to simple chemical reactions, such as cross-couplings and amide bond formation, which are used combinatoriality to generate vast libraries. As analogous chemical reactions are used, current libraries are often populated with compounds with similar physical properties and have limited structural diversity. The reactions used to generate DELs are often low yielding, or are limited in substrate scope, further reducing the diversity of potential libraries. Development of new approaches with DEL synthesis will increase the ability to synthesise libraries with greater chemical diversity and improved physical property profiles.

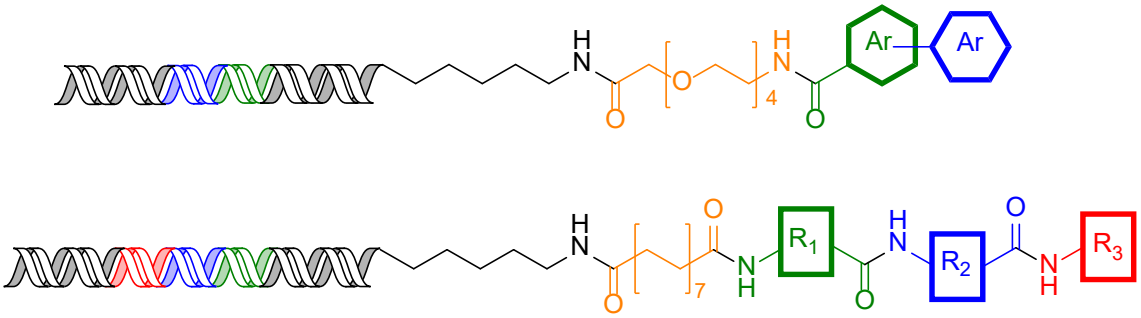


Products produced from the encoded transformation paradigm.

A new paradigm termed “encoded transformations” was introduced. This technique involves encoding a specific chemical transformation of a reactive core molecule instead of coding the addition of a building block. This would lead to a potential library of compounds with reduced overall molecular weight and more lead-like physical properties. A common reactive 2-

oxobetenamide core was used and several chemical transformations have been successfully employed. These chemical transformations were developed off-DNA and provide a means to prepare libraries with significant scaffold diversity using DNA compatible chemistry.

Micellar solvents have been shown to improve normal phase organic chemical reactions. Applying this technique to on-DNA synthesis vastly improved the scope and diversity of the reactants used. Both amide couplings and Suzuki-Miyaura reactions were successfully optimised utilising micellar media as a solvent. Optimisation was carried out in both examples by factorial experimental design (FED), which revealed second order relationships between variables tested that would not likely be discovered using conventional techniques.



Products produced from novel micellar promoted Suzuki-Miyaura and amide reactions.

These novel reactions were used to synthesise a prototype 6x6 library, which was PCR amplified and sequenced proving that reactions in micellar media do not cause DNA damage. The reverse amide coupling was then used to create a 99,405-member library, including compounds designed to specifically target the SARs-CoV-2 main protease (M^{pro}) binding site, using an array of both covalent and non-covalent inhibitors.

Abbreviations

4CzIPN	2,4,5,6-Tetra(9H-carbazol-9-yl)isophthalonitrile
14mer	14 base single stranded DNA
2,4-DNP	2,4-dinitrophenylhydrazone
AcOH	Acetic acid
ATP	Adenosine triphosphate
AUC	Area under the curve
BB1	Building block 1
BB2	Building block 2
BB3	Building block 3
BOC	*tert*-butyloxycarbonyl
BOP	benzotriazol-1-yloxytris(dimethylamino)phosphonium hexafluorophosphate
BP	Base pairs
Bu_4NOAc	Tetrabutylammonium acetate
CD_3OD	Deuterated methanol
$CDCl_3$	Deuterated Chloroform
cLogP	Logarithm of a compounds partition coefficient between n-octanol and water
CMC	Critical micelle concentration
COMU	(1-Cyano-2-ethoxy-2-oxoethylidenaminooxy)dimethylamino morpholino-carbenium hexafluorophosphate
COSMO-RS	Conductor-like Screening Model for Realistic Solvents
COSY	^1H-^1H Correlation Spectroscopy
CROs	Contract research organization
Cs_2CO_3	Caesium carbonate
CuAAC	Copper promoted azide-alkyne Huisgen cycloaddition
Da	Daltons
DABCO	1,4-diazabicyclo[2.2.2]octane
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene

DCC	N,N’-Dicyclohexylcarbodiimide
DCE	1,2-Dichloroethane
DCM	Dichloromethane
DDR1	Discoidin domain receptor 1
DDR2	Discoidin domain receptor 2
DEAE	Diethylaminoethyl
DEL	DNA encoded libraries
DH	Dehalogenation
DHP	3,4-Dihydropyran
DIAD	Diisopropyl azodicarboxylate
DIC	N,N’-Diisopropylcarbodiimide
DIPEA	N,N-Diisopropylethylamine
DMA	Dimethylacetamide
DMAP	4-Dimethylaminopyridine
DMF	Dimethylformamide
DMSO	Dimethyl sulphoxide
DMSO-d6	Deuterated DMSO
DMT-MM	4-(4,6-Dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride
DNA	Deoxyribonucleic acid
DOE	Design of experiment
DP	Desired product
DS	Double stranded
EDC	1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide
ELSD	Evaporative light scattering detector
EtOH	Ethanol
FACS	Fluorescence-activated cell sorting
FED	Factorial experimental design
GC content	Guanine-Cytosine content
HATU	1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate

HFIP	Hexafluoroisopropanol
HI	Hydrogen iodide
HOAT	1-Hydroxy-7-azabenzotriazole
HP	Head piece
HPLC	High-performance liquid chromatography
HTS	High-throughput screening
IC_{50}	Half maximal inhibitory concentration
IEDDA	Inverse-electron-demand-Diels-Alder
IPA	Isopropylalcohol / 2-propanol
K_2CO_3	Potassium carbonate
K_3PO_4	Potassium phosphate
K_d	Dissociation constant
KOAc	Potassium acetate
LCMS	Liquid chromatography mass spectrometry
LED	Light-emitting diode
MeCN	Acetonitrile
MEK2	Dual-specificity protein kinase
MERS	Middle East Respiratory Syndrome
MHz	Megahertz
MIDA	Methyliminodiacetic acid
MMT	Monomethoxytrityl
MOPS	3-(N-morpholino)propanesulfonic acid
MPEG	Methyl-polyethylene glycol
M^{pro}	SARS-CoV-2 Main Protease
MS	Mass spectrometry
Na_2CO_3	Sodium carbonate
$NaClO_4$	Sodium perchlorate
NGS	Next generation sequencing
NICR	Northern Institute of Cancer Research

NMI	1-Methylimidazole
NMP	N-Methyl-2-pyrrolidone
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
Nok	β-Sitosterol methoxyethyleneglycol succinate
o/n	Overnight
OH	Overhang
OTf	Triflate
PBOP	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
Pd/C	Palladium on activated carbon
PEG	Polyethylene glycol
PEG-4	2,2'-[Oxybis(2,1-ethanediyloxy)]diethanol
PL^{pro}	Papain-like protease
PNA	Peptide Nucleic Acid
PPh_3	Triphenylphosphine
RCM	Ring closing metathesis
R_f	Retention factor
RIP1	Receptor interacting protein 1
RNA	Ribonucleic acid
ROESY	Rotating-frame nuclear Overhauser effect
RPKA	Reaction progress kinetic analysis
RT	Room temperature
SAR	Structure activity relationship
SARs	Severe acute respiratory syndrome
SCX	SiliaBond Tosic Acid
sEH	Soluble epoxide hydrolase
SM	Starting material

S_N2	Nucleophilic Substitution
S_NAr	Nucleophilic aromatic substitution
SPR	Surface plasmon resonance
SS	Single stranded
t-BuOH	*tert*-Butanol
t-BuOK	Potassium-*tert*-butoxide
TCA	Trichloroacetic acid
TEA	Triethylamine
TEMPO	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TNKS1	Tankyrase 1
Tosyl	Paratoluenesulphonyl
TPGS-750-M	DL-α-Tocopherol methoxypolyethylene glycol succinate
TPPTS	3,3',3''-Phosphanetriyltris(benzenesulfonic acid) trisodium salt
UV	Ultra-violet

Contents

Chapter 1 – Introduction to DNA Encoded Libraries

1.1 – Inspiration for DNA Encoded Libraries

Discovery of small molecules of biological importance remains a major challenge for the pharmaceutical industry. Developing a drug from the initial hit to the completion of clinical trials is a complex process which can take up to 15 years and nearly 1 billion dollars.[1] Any method capable of reducing the cost or time frame of developing a drug candidate would be a huge benefit to both industrial and academic research labs.

Most novel drug candidates are produced via an iterative process categorised into hit generation, hit-to-lead and lead optimisation phases. This process begins with the initial discovery of a hit compound, a biologically active compound that can be used as a start point for optimisation. Due to Lipinski's rule of 5,[2] and requiring compounds with a favourable pharmacokinetic profile, the hit compound should be of low molecular weight and clogP. The majority of marketed drugs have a molecular weight of less than 400 and a clogP of less than 4.[1] Currently high-throughput screening (HTS) is the most commonly used method for the discovery of biologically active compounds. This is a resource intensive procedure of which millions of compounds are screened in a biological assay and is very demanding in terms of cost, logistics and synthesis of the libraries of this size. The process is often deemed too impractical for many small pharmaceutical and academic laboratories to employ, and any methods that reduce the cost and time taken to generate hits could positively impact on the number of biologically active small molecules that are discovered.

Phage display has been used extensively to discover protein interactions, and these selection-based approaches have been employed in high throughput screening of protein interactions, allowing for discovery of specific antibodies against a huge variety of target proteins. This process is rapid and more cost effective in comparison to the HTS process in drug discovery. A phage display library is constructed by inserting a range of randomly synthesised oligonucleotides into phage vectors, which then produce the corresponding translated peptides [Figure 1].[3] Once synthesised the library is then incubated with the biological target. The target is then washed, removing the peptides that have not bound to the target, and the peptides that had bound to the target protein are then cloned inside the phage and amplified. This process is then repeated multiple times to improve the data quality and selectivity of the

peptide. The DNA is then decoded to identify the peptides that have bound, allowing for them to be synthesised as potential new biologically active compounds against the target protein.

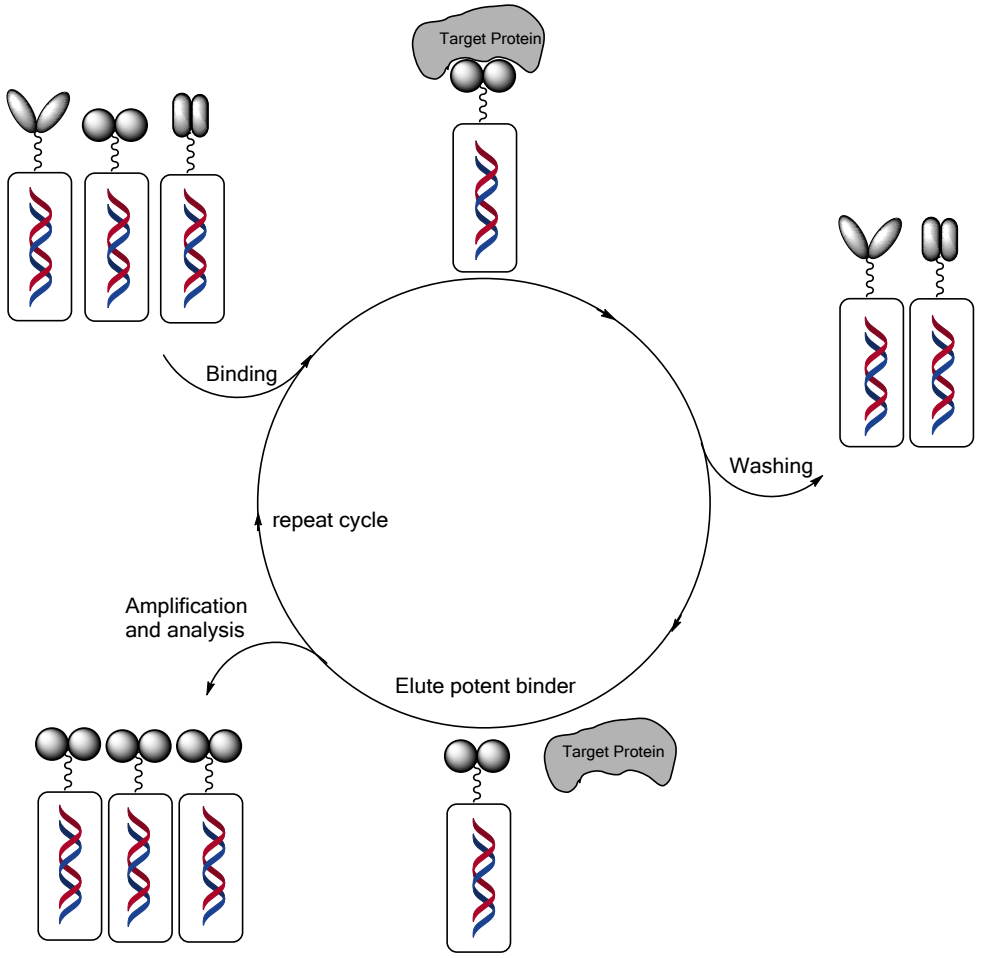


Figure 1 - Technique used by phage display to find potent binders. Once a library of phage vectors has been produced, they are incubated with the target protein and washed to remove the non-binding substituents. The binders are then eluted from the protein and the cycle repeated to increase selectivity. The binders are then amplified by PCR and analysed.

1.2 – Foundations of DNA Encoded Libraries

An early stage method from which DNA encoded libraries evolved was first suggested by Brenner and Lerner in 1992 when they proposed applying the techniques employed in phage display to chemical library synthesis. Phage display technology is limited to protein, peptide and oligonucleotide discovery due to the specificity of the enzymes used to sequence the library. However, phage display shows that using DNA to encode libraries of molecules is possible.

Brenner and Lerner suggested the first chemically synthesised libraries utilising DNA-encoding in a paper in 1992.[3] Using a split and pool approach, they described a library synthesis using

solid phase peptide and oligonucleotide synthesis. Solid supported peptides would be constructed linked to an oligonucleotide tag, and the synthesis would be terminated by liberating the compound from the pore glass matrix [Figure 2]. This would allow for the creation of a library of peptides attached to an encoding oligonucleotide tag. To identify the attached peptide the oligonucleotides could be cloned and sequenced. The liberation of the peptide-oligonucleotide from the solid bead sets this approach apart from other solid-phase split and pool libraries.

In 1993,[4] a year after this first conceptual paper, Brenner and Lerner published a second paper with experimental proof of the success of the initial concept. Utilising known peptide ligands for specific antibodies it was shown that binding activity was retained even when attached to an oligonucleotide tag. It was also shown that such compounds could be amplified using PCR. No novel ligands were found using this initial library, and due to the early stage proof of concept, there were a few flaws with their approach. Firstly, using a solid supported approach negatively impacts the overall number of compounds that can be synthesised in this way, due to the number of beads that are available. Also, the chemistry that is used in this approach must be compatible with oligonucleotide synthesis, which really limits the diversity of the library. However, even with the flaws to the design, Brenner and Lerner initiated a vast number of advancements to improve upon this initial method.

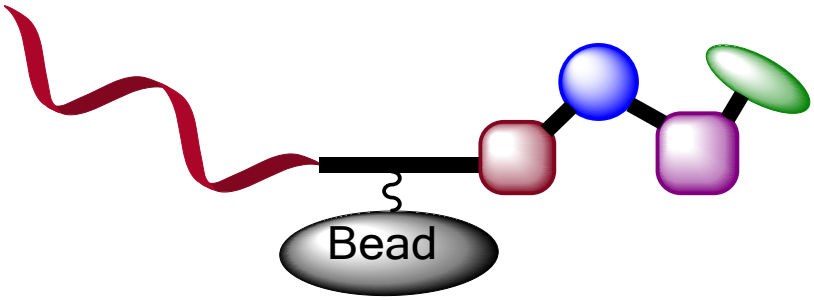


Figure 2 - The initial concept by Brenner and Lerner was the synthesis of peptide sequences in a combinatorial fashion together with the corresponding amplifiable DNA strands. Note that in their initial concept, the synthesis was carried out on a solid support.

This initial concept was further improved by a group at Saitama University.[5] Using combinatorial methodology, an oligonucleotide tagging system was developed utilising a split and pool synthetic route, resulting in an encoded library of compounds. This method improved one of the flaws of the initial concept as it utilised a solution phase synthesis, and therefore was not limited to the number of beads. However, using this approach, library

synthesis would need to be compatible in the presence of an oligonucleotide tag rather than solid supported oligonucleotide synthesis. A drawback of this method, and one that still exists to this day, is that if the tag is modified by any synthesis step the integrity of the sequencing data deteriorates rapidly. This process conflicts with the benefit of rapid cloning and sequencing using PCR to determine the compound that the tag is attached to.

1.3 – Templated DNA Synthesis

The subsequent examples of successful chemical transformations in the presence of DNA came in 2001 from Gartner and Liu at Harvard University.[6] It was proven that generic chemistry was achievable in the presence of DNA by using DNA to rapidly increase the reaction rate of a Michael addition of a thiol reagent onto a maleimide to form the thioether. The rate of this simple solution phase reaction was increased dramatically by using a sequence directed approach. This is where reacting molecule pairs are linked to complementary oligonucleotides to bring them into proximity. The effect of the proximity of the two reactants increased the reaction rate by a striking amount. Two different templates were used to test the reaction, both a hairpin (H) and end-of-helix (E) [Figure 3]. Both templates showed a comparable reaction rate, and when a non-complementary mismatched sequence was used, no reaction occurred. The reaction was still as successful when the reactants were employed up to 30 base pairs away. A library of 1025 maleimide-linked E templates, each with an individual encoded DNA sequence was produced and the group showed that amplification using PCR was possible.

Expansion of the reaction scope was conducted using DNA-templated reactions.[7] Reactions such as nitro-aldol addition, nitro-Michael addition, Wittig olefination, dipolar cycloaddition, and Heck couplings were demonstrated to be compatible with a DNA-templated synthesis. This was the first breakthrough showing that a variety of solution phase chemistry could be achieved outside of the usual amide coupling and reductive amination chemistry. The same group then synthesised a DNA-templated library to create a 65-member peptide macrocycle library. Three amide couplings using different amino acids were used, then a final cyclisation step utilizing a templated Wittig reaction yielded the desired macrocycle. Each of the different library members were attached to an individual amplifiable DNA tag. They found that a single member of their library had activity against carbonic anhydrase, the first ligand found to be active utilizing a DNA tagged library.

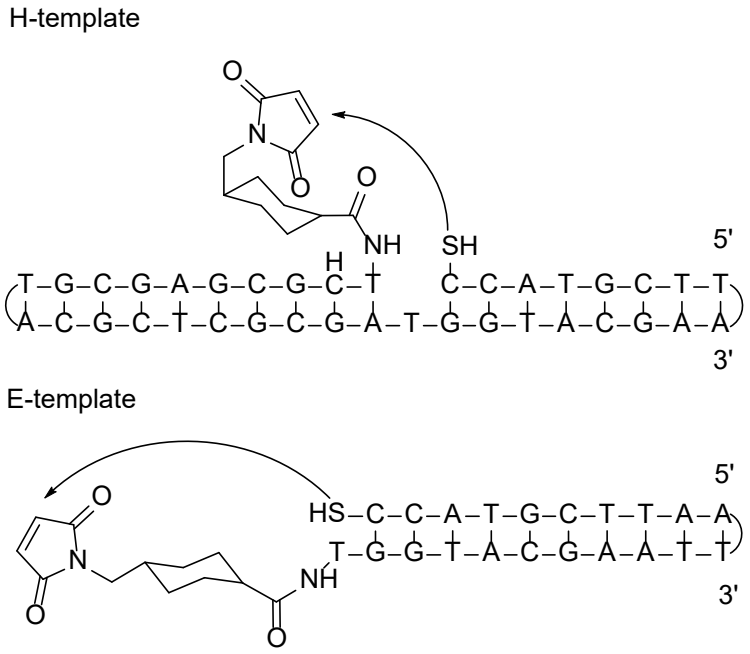


Figure 3 - The two examples of templates used to support hybridisation-dependent chemistry. Interestingly, the 2 reactants can be up to 30 base pairs apart.

The Neri group in Zurich took a slightly different approach.[8] Utilising the ability for complimentary oligonucleotide strands to anneal to each other, they attached a target molecule to the 5' and 3' ends of the complimentary strands. The hybridisation of the 2 oligonucleotides holds the two molecules together in space allowing them to be easily screened against a target protein. This technique resembles a fragment-based approach and differs from before as there are no chemical reactions that take place between the encoded fragments, which in turn decreases the possibility of sequence deterioration. The Harbury group took a similar approach[9-11] using an initial 340-base oligonucleotide which is capped at the 5' end with an amine group, and could undergo a combinatorial split and pool synthesis to form a molecule with 6 points of variation. The DNA strand contains lengths of 20 base noncoding and coding regions. All library members contain the same seven noncoding regions; however, the six coding regions differ in each example. A unique reactive site is attached to the end of a 20 base oligonucleotide, and when this hybridizes with the complimentary bases on the original 340 base strand, a reaction will take place between the 5' amine and the unique reactive site. This technique allows for the straightforward synthesis of a library with 6 points of variation, all controlled by the code of the attached DNA strand. A library was constructed of one million different peptides using this method.

However, there are a few challenges when using a sequence directed approach. Firstly, large chemical libraries are difficult to synthesise due to the limited number of oligonucleotide tagged molecules available. The molecules need to have complimentary strands to each other, which is a significant problem for Neri's approach. Also, mistranslation can occur quite readily in these examples which introduces miscoding in the DNA sequence. This increases linearly with the size of the library, which also limits the overall size of a library that can by synthesised by this approach. These initial approaches were also expanded upon in more recent years to iron out some of the problems; nevertheless, the size of the libraries and the chemical space probed by this technology is insufficient related to comparable methods.

1.4 - Sequence Recorded Encoding

This technique differs greatly from templated synthesis. Instead of using the base pair coding to dictate chemical reactions, the code is used only as a tag for the building block added and is not involved in the actual synthesis. Utilising a step wise combinatorial approach allows a large diverse library to be synthesised with relative ease. This is still a solution phase approach, so includes many of the limitations that the DNA-templated reactions had in terms of chemical compatibility.

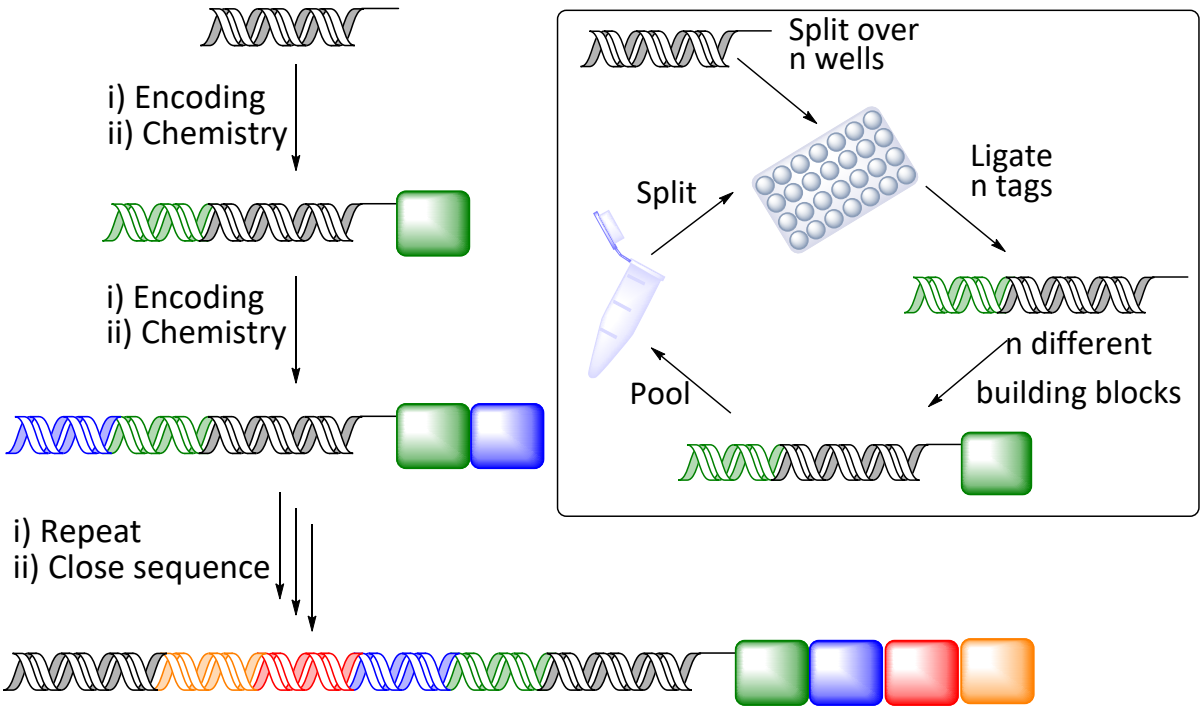


Figure 4 - Split and pool DEL synthesis used to create DELs. First a DNA primer attached to a liner molecule terminated with a reactive group is split over several wells in a plate. The first building block addition is performed followed by the ligation.

The first step when applying a split and pool approach is to attach a functional group (often a PEG linked amine) to a primer DNA strand. This modified base acts as an anchor for any future chemical reactions, and a PEG spacer is often used to increase the distance between the DNA strand and the chemical probe to prevent the DNA tag interfering with binding. The library is then split into wells in a plate where the addition of the first building block to the functional group occurs along with the ligation of a unique DNA strand for each building block. The individual DNA encoded molecules are then pooled together, purified as a bulk and the mixture split again into individual wells. The process is then repeated for as many times as required to achieve the desired library size and diversity. This methodology requires chemistry that is fully compatible with DNA in every step of the library formation, which can be difficult.

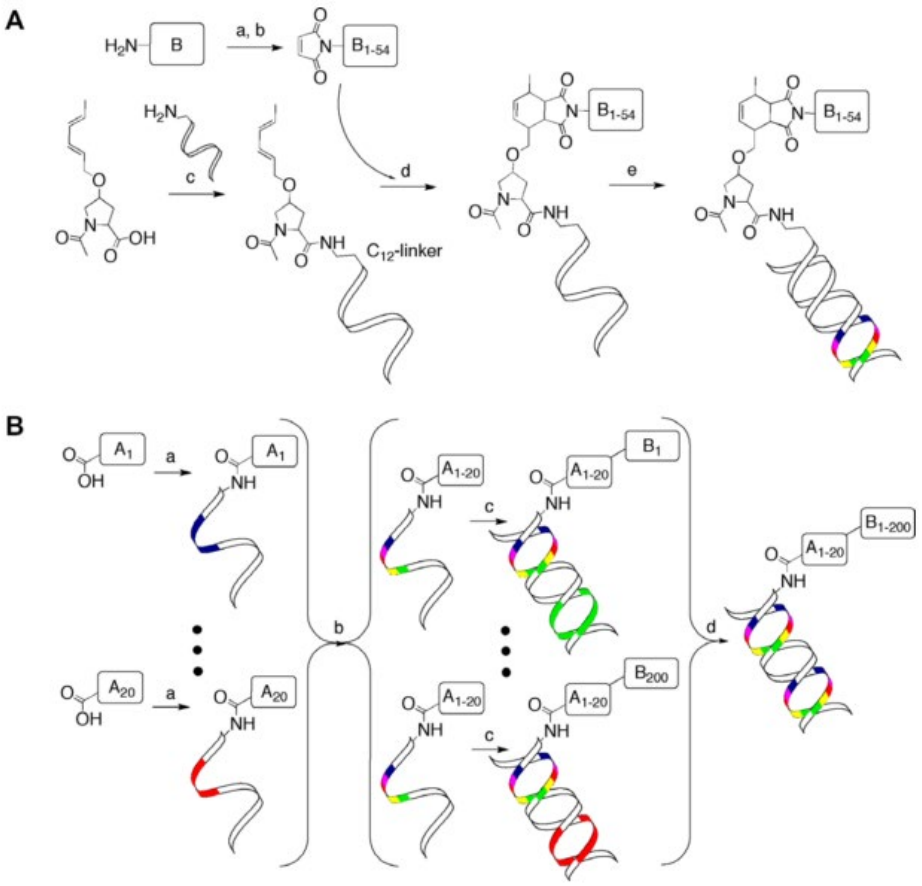


Figure 5 - A) Reaction scheme showing the synthesis of the basic 54 member Diels-Alder library. First the addition of the dienyl moiety to the DNA strand using an EDC coupling. Then the Diels-Alder reaction was carried out with 54 different maleimides. Finally a Klenow-Polymerase mediated encoding step was employed; B) Representation of the 4000 member library where 20 different dienyls and 200 different maleimides were used.[12]

The first libraries that were produced using this technology were from Praecis Pharmaceuticals[13] and Nuevolution A/S.[14] Praecis described an enzymatic ligation step that enabled encoding of a DNA strand attached to a linker comprising of an active site at which construction of a small molecule is possible. They described the synthesis of several libraries using this technique. The Nuevolution patent describes something similar, however they also describe the formation of a library using split and pool methodology utilising a range of chemical reactions including reductive amination, acylation, isocyanate addition, nucleophilic substitution and sulphonylation. Both patents also show that the libraries can be used to find ligands for exemplar proteins.

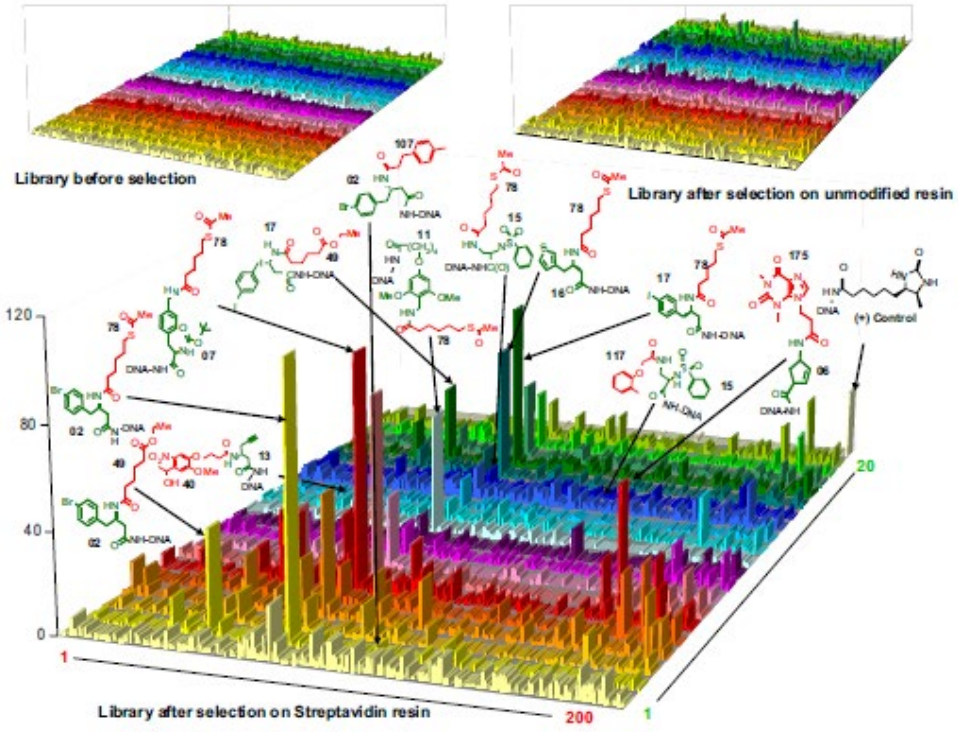


Figure 6 - A typical plot showing how building blocks are represented after high-throughput sequencing. 2 controls were run showing the 4000-member library before selection and after selection on empty resin. The main plot is after selection on streptavidin resin and reveals which building blocks have bound to streptavidin.[15]

The Neri group produced many libraries using these techniques and possibly show the first novel ligands to bind to non-pre-selected target proteins. A library was constructed using a Diels-Alder cycloaddition between a maleimide and a dienophile [Figure 5]. The library they developed utilised two diversity steps.[12] Firstly they coupled 20 different carboxylic acids containing the reactive dienophile moiety to 20 unique 5'-amino modified oligonucleotides. They then used split and pool methodology, where 200 unique maleimides were added and

uniquely encoded. The 200 wells were then pooled resulting in a 4000-member library. A second library was synthesised concentrating on amide couplings and an addition of 200 amides to an original 20-member sub-library to produce a second 4000-member library.[15] They showed that even with low concentrations of <1 nM, binding molecules could still be effectively recovered using a biotinylated target protein followed by a streptavidin capture. The library was incubated with streptavidin-Sepharose resin, which was then washed to remove any molecules that did not bind, and the remaining oligonucleotides were amplified using PCR and sequenced [Figure 6]. Decoding of the streptavidin selection shows a preferential enrichment of the genes that code for certain classes of related compounds. Novel ligands against protein polyclonal human IgG were consequently found using this technique, with K_D values as low as 350 nM. The same library was also screened against IgG showing potential binders.

The developments made by these groups indicates that DEL synthesis can be performed practically to acquire new binders of proteins. Many groups in both industry and academia have successfully synthesised this type of split and pool library to find novel ligands,[15-18] and some of these groups have found successful lead compounds. Currently there are no drugs on the market or in current clinical trials that have arisen from any DEL, but this is a very early stage technology that is currently being expanded. However there have been clinical candidates in the past utilising this technique. This may also be due to the negative parameters that these libraries still possess, such as high average molecular weight and compounds with poor pharmacokinetics.

1.5 – The Synthesis of a DNA Encoded Library

Chemical synthesis has historically been carried out in organic solvents, even with the recent upsurge of “green” chemistry. DNA is insoluble in organic solvents, and thus reactions must be carried out in aqueous media. Small amounts of miscible organic solvents are compatible when used to solvate building blocks. DNA is degraded by strongly acidic conditions, due to it being liable to depurination of the guanine or adenine bases [Figure 7]. This in turn denatures the DNA making it impossible to sequence and therefore reducing the fidelity of the overall library. For chemistry to be compatible in DEL synthesis, it must be compatible with aqueous media and work at a pH >4. Standard chemical transformations have proven successful when applied to DNA. However, many chemical reactions are inherently incompatible with DNA due

to the ability to denature and modify the coding sequence, deeming research in this area vital for increasing the diversity of a DEL.

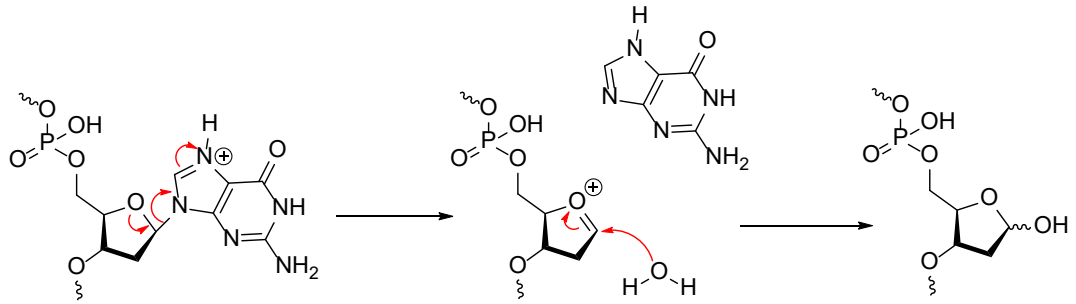


Figure 7 - DNA damage by acid promoted depurination.

Due to the importance of discovering chemistry that is compatible with DNA, research has been conducted to work out which reactions are truly compatible.[19] A method that can be used to analyse the reaction yield as well as the impact of the reaction on the integrity of DNA has been used to study chemical reactions. This was achieved by synthesising a DNA strand on a magnetic bead and carrying out reactions in the presence of this bead. The magnetic beads were easily removed from the mixture after the reaction had taken place, amplified by PCR and sequenced. This showed not only what the overall yield of the reaction was, but the percentage of amplifiable DNA remaining. This method can then be used to test a multitude of possible DNA compatible reactions. It was found using this method that a majority of reactions thought to be fully compatible damaged the DNA to some degree.

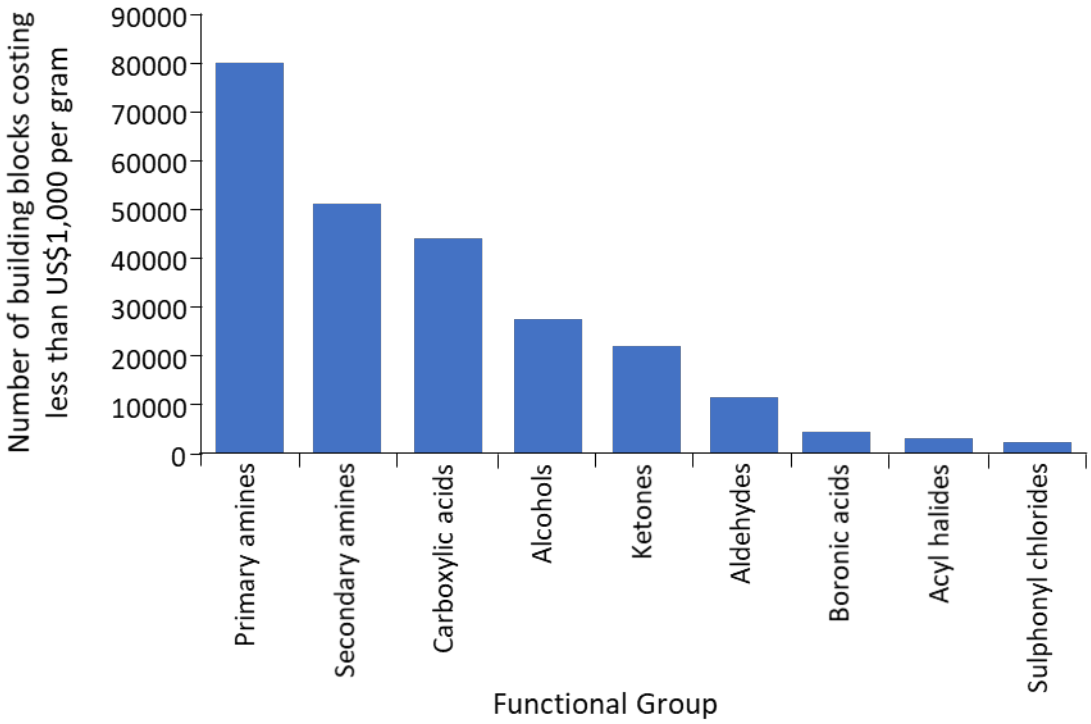


Figure 8 - The availability of building blocks by functional group from suppliers. A limit was used of 250 Da for molecular weight, as well as a price of less than $1000 per gram. This data has been extracted from the eMolecules database by Kalliokoski.[20] The data was then plotted on a bar graph.[21]

As well as the number of different chemical reactions, to build a diverse large library of compounds requires a diverse selection of building blocks. A large number of the building blocks available commercially are primary and secondary amines, as well as carboxylic acids and alcohols [Figure 8].[22] Difunctional and trifunctional building blocks, used as cores or templates, are even fewer in number, with a majority of these being made up from amino acids.[23] The liability of the building blocks that are commercially available is also of concern to any library design. Not all building blocks available are appropriate for a drug-like library, and some contain unwanted functional groups or are closely related to others. Therefore, a great deal of discretion must be employed in order to pick the overall building blocks from which the final library will be synthesised.

1.6 – DNA Compatible Chemistry

The most common method for DEL synthesis relies upon using a split and pool combinatorial approach, especially to build very large (eg. ~billion member) libraries. This includes several repetitive steps of ligating a DNA tag, followed by a chemical reaction, deprotections of functional groups if needed, then often purification of the entire library subset. Due to the

nature of combinatorial chemistry, high conversion rates and high-fidelity reactions are required. If successful, this approach can generate libraries of billions of unique compounds, which can be screened against targets of interest. For a reaction to be deemed compatible it must be run on a nmol scale, with limited miscible organic cosolvents, a reaction pH of 4 and above, lower than 100°C and use no harsh oxidising or reducing agents. If these conditions are not met, the reaction is likely to be unsuccessful as increasing amounts of DNA damage will occur, resulting in a low performance library.

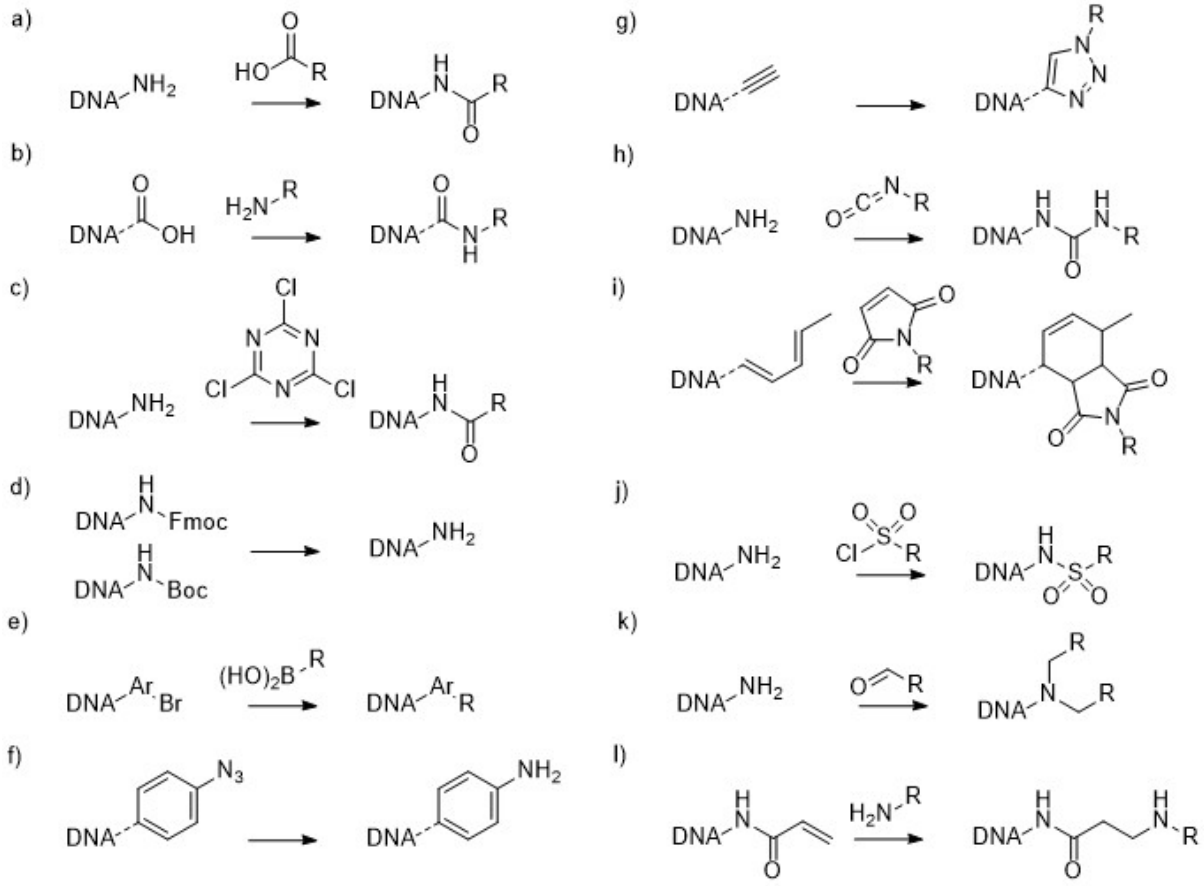


Figure 9 – a,b) amide couplings,[17, 24] c) S_NAR,[14, 17] d) Fmoc and Boc deprotections,[14, 17] e) Suzuki-Miyaura coupling,[25] f) Azide reduction,[26] g) "click" triazol synthesis,[26] h) Urea formation,[14] i) Diels-Alder reactions,[12, 27] j) Acylation with sulfonyl chloride,[28] k) Reductive aminations,[14] l) Michael additions.[6]

The rate at which novel DNA compatible reactions are being produced has increased significantly over the past 5 years. In 2014[23] an extremely limited number of DNA compatible reactions were shown to be compatible [Figure 9]. Many of these reactions were simple couplings of building blocks such as amide coupling, S_NAr, Suzuki cross-coupling and reductive amination, as well as deprotections for straightforward sequential addition of building blocks in the synthesis of the library. There is a great need for new compatible reactions and

improvements of existing ones, as overall library design is limited by the building blocks and reactions. A significant quantity of research has been carried out in this area as shown by successful publications of novel reactions, as well as increasing the success of those already known. Many new reactions have been shown to perform well with DELs over the past few years. The steady expansion of novel reactions will have a significant impact on the fidelity and chemical composition of new libraries.

1.6.1 – Recent Advances of Simple Organic Reactions

Simple organic reactions are vital for the construction of DELs. As shown previously [Figure 8] there is a vast assortment of available organic substrates containing amines, carboxylic acids, and other simple functionality. The basis of designing any practical DEL relies upon high fidelity, high yielding, simple organic reactions, such as amide coupling, reductive aminations, S_NAr, and deprotections. These reactions also form the basis of many medicinal chemistry manuscripts [Figure 10], in which amide couplings, S_NAr, S_N2 and cross-coupling reactions, as well as amine and ester deprotections are extremely common. The fact that many marketed drugs also contain the linkages formed from these reactions shows how vital simple, high-fidelity reactions are in library design and synthesis.

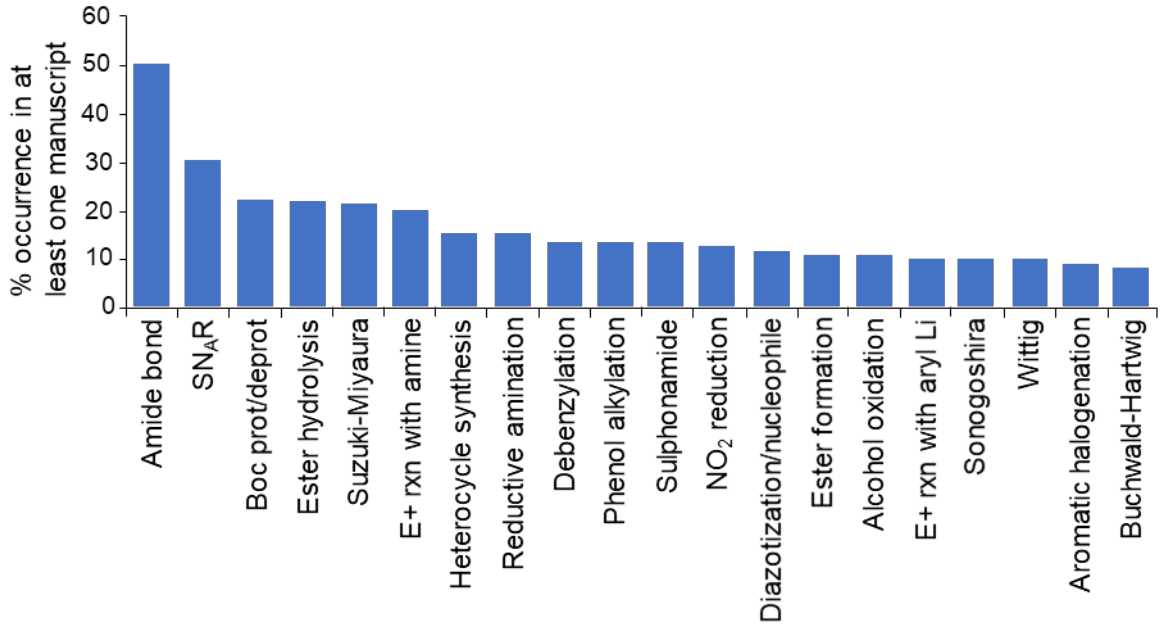


Figure 10 – Occurrence of each reaction plotted as a percentage of use in at least one manuscript involving medicinal chemical reactions. Blue shows reactions in 2014, red is reactions taken in 1984. The arrows with years show the first citation of these particular reactions.[29]

Amide couplings are an extremely important class of transformations and a majority of DNA encoded libraries will utilise at least one amide coupling. One of the major concerns with amide couplings is that the activated acid species can also react with the water present in the reaction, therefore slowing the coupling, particularly when the amine in question is of low reactivity. This implies that there is always a limit to which building blocks can be used in library design, as most of the time non-activated amines and acids must be discarded in the validation phase. This can greatly limit the chemical diversity of the library. Although many libraries have been synthesised utilizing the previously described DMT-MM couplings,[16, 18, 30, 31] it has been demonstrated that using this procedure drastically limits the number of acids that are compatible [Figure 11]. However, using EDC as a coupling agent with HOAt and DIPEA in DMSO and water it was found that 78% of the acids tested yielded product at over 75% conversion rates. This increase in yield would drastically improve the number of building blocks available for the library build, significantly diversifying the library. An advantage of increasing the number of building blocks available would also affect peptide and macrocycle synthesis. Macrocyclic peptides are used often in libraries, and can be synthesised using multiple amide coupling steps, followed by an on-DNA, copper catalysed click reaction.[30] The macrocycle closing click reaction was achieved in good yields, allowing for the synthesis of a 2.4 billion member library.

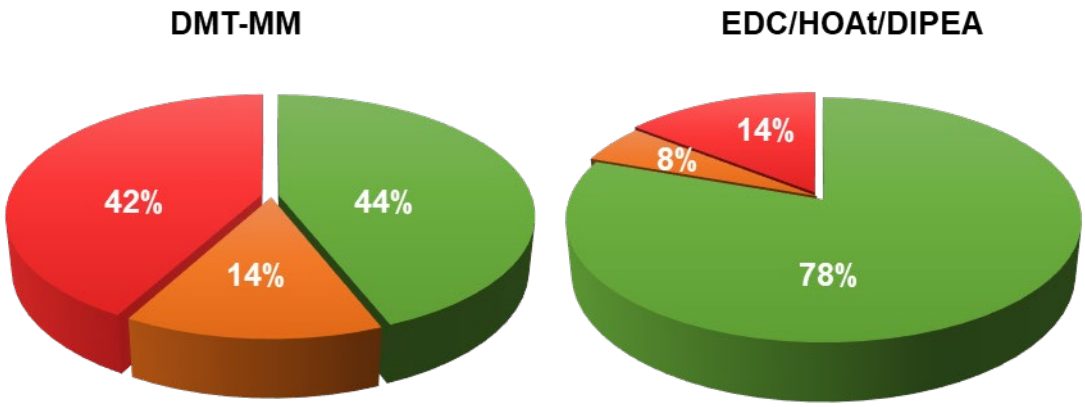


Figure 11 – Yields of couplings of 543 carboxylic acids and a DNA linked amine, using DMT-MM or EDC/HOAt/DIPEA coupling conditions (green >75%, orange 50 - 75%, red <50%).[32]

Nitro groups are an important tool in synthesis by way of a masked amine, allowing access to some scaffolds previously inaccessible with protected amines. This is especially true when requiring access to bi, or trifunctional building blocks and can be used to synthesise libraries

containing the privileged benzimidazole scaffold.[33, 34] Originally Raney nickel and hydrazine were used for the reduction of nitro groups,[35] producing a phenyl functionalised methyl benzimidazole product, but this was not feasible with multiple functional groups [Figure 12]. The use of $Na_2S_2O_4$ increased the scope of the reaction[36] considerably, with many functional groups, such as pyridine and sulphones tolerated. It was also found that using iron sulphate, under basic conditions was a tolerated method for reduction.[34] However, problems arose using electron deficient species. Conditions that are tolerated on other aromatic nitro compounds have also been studied using tetrahydroxydiboron under basic conditions and ethanol as a co-solvent.[37] Using these conditions, 38 examples containing a variety of nitro groups, including ortho-aminonitroarenes were successfully reduced. The conditions are also milder than those previously reported, showing moderate yields with greater than 40% conversion over all examples listed. These examples increase the diversity of functionality that would not be accessible under normal amine protection/deprotection chemistry and allows for a vital new pathway in the synthesis of DELs.

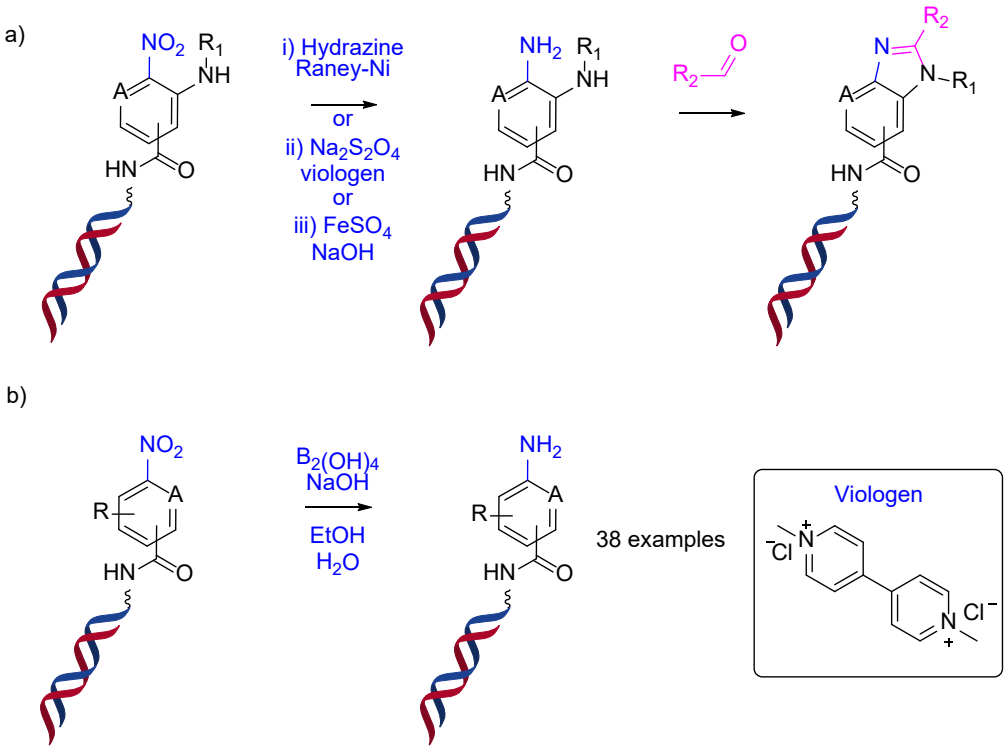


Figure 12 - Benzimidazole formations utilising nitro reduction reactions under different conditions.[33-37]

Cycloaddition reactions, such as the Diels-Alder reaction, can play an important role in generating compounds with a variety of stereochemistry and chemical diversity. Previously,

these reactions have been utilised, however some significant advances have been made, such as the synthesis of pyridazines, a previously difficult substrate to employ in a library.[38] Pyridazine containing structures are thought to be privileged scaffolds in medicinal chemistry and are often difficult to synthesize from cross-coupling or other general methods. However, using an inverse-electron-demand-Diels-Alder (IEDDA) reaction, these scaffolds can be synthesised from 1,2,4,5-tetrazines [Figure 13]. It was discovered that tetrazines were able to react with DNA conjugated alkenes in DMSO and water. When using a DNA conjugated tetrazine, the reaction required a second reaction to oxidise the product using copper perchlorate and TEMPO. Good yields were found across most substrates, where 16 out of the 20 tested yielded over 50% conversion to product. A mixture of products was produced using some of the substrates in this reaction, thereby providing the incentive to use conditions that do not require an oxidation step. Generating an enamine in situ from the ketone or aldehyde together with proline, allowed access to the oxidised pyridazine structure, with no need for a separate oxidation step. Ten out of the twelve substrates yielded a greater than 70% conversion rate. The reaction was also applied to substituents with further functionality attached to the pyridazine, such as bromo, amino and dichloropyridazine, providing handles for further diversification steps.

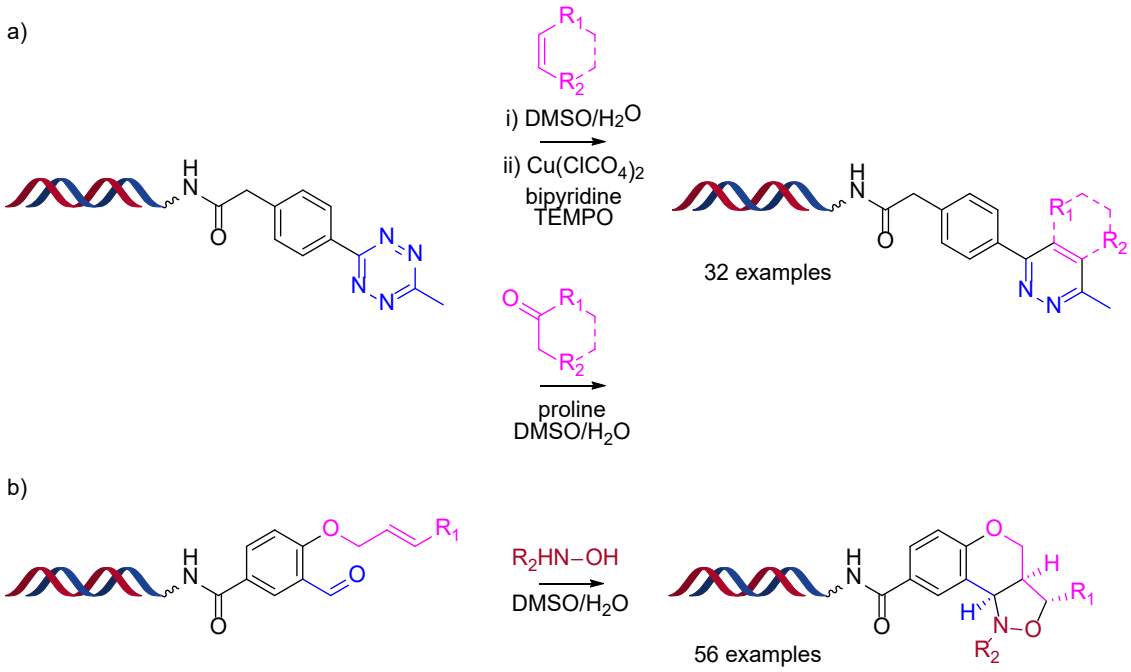


Figure 13 – Advances to cycloaddition reaction compatible with DNA.[38, 39]

Nitrone-olefin [3+2] cycloaddition can also be achieved with high compatibility with DNA. Initially the reaction was performed in a templated manner reacting simple nitrones with activated alkenes.[40] This reaction has been adapted to DEL synthesis to form sp^3 rich isoxazolidines, with at least 2 stereogenic elements from sp^2 rich starting materials.[39] This motif is found in alkaloid natural products and other bioactive compounds and can be used to introduce diversity through the hydroxylamine residues. It was discovered that using sodium acetate buffer, good conversion to product was achieved with a variety of hydroxylamine residues as well as enriched diversity of the attached alkyne R group. A three-compound library was then synthesised, and a successful PCR amplification with no detected loss of DNA integrity showed that this reaction was DEL compatible. Improved diversity can be achieved by utilising simple organic reaction such as cycloadditions to maximise the chemical diversity of the library.

1.6.2 – Transition Metal Catalysed Reactions

DEL synthesis often requires excess reagents to increase the reactivity in the aqueous environment in which reactions are performed. Excess reagents can have a detrimental effect on DNA, increasing the risk of damage. The use of transition metal catalysed reactions in medicinal chemistry has risen significantly, with many modern medicinal chemical syntheses using these reactions at some stage. Suzuki-Miyaura couplings were the second most used organic reaction in medicinal chemistry in 2014, with over 20% of manuscripts using the reaction [Figure 10]. Interestingly, Sonogoshira and Buchwald-Hartwig reactions also make the list at over 10%. There is also a large variety of iodo/bromo building blocks, as well as boronic acids and amines [Figure 6] available to design and synthesise a DNA encoded library. Advances in this area are vital to unlock a vast area of chemical space and library design. To date, many successful reactions have been produced using copper and palladium catalysed reactions with some utilised in library synthesis.[16, 18, 41, 42] However, advances are still required to further increase the number of metal catalysed reactions and also substrate scope.

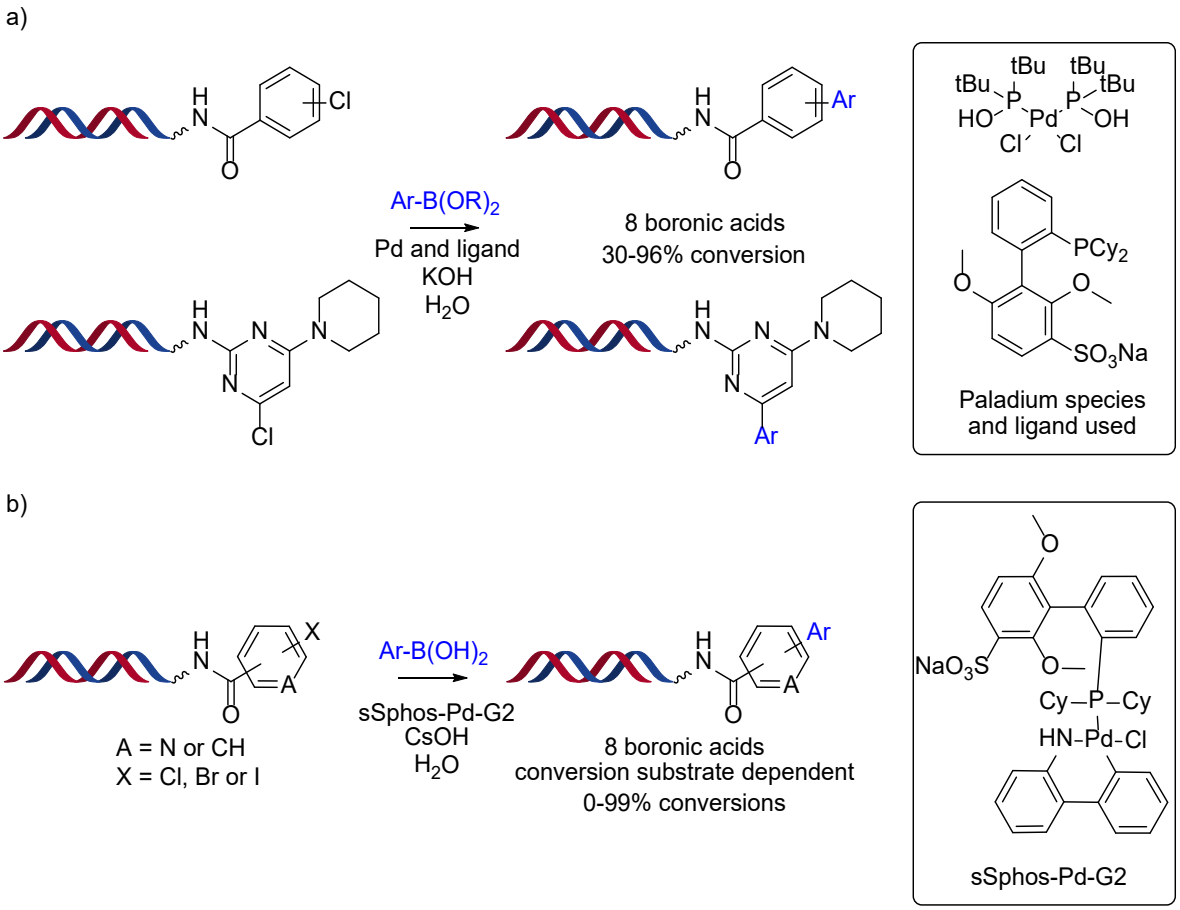


Figure 14 – Suzuki-Miyaura reactions in current literature.

Suzuki-Myaura reactions have been vital component of a medicinal chemist's toolbox since their invention in 1981. Original Suzuki-Miyaura reactions were achieved using tetrakis(triphenylphosphine)palladium(0) in water. However, yields varied greatly depending on both the halogen and boronate species used.[43] Even though these conditions were far from ideal, they were employed in the synthesis of a large DNA encoded library.[35] One impact of using ligands such as tetrakis is that it is highly insoluble in aqueous solvents, requiring increasing amounts of co-solvent for reaction to take place. To improve upon this methodology a novel ligand system was used[44] making use of a sodium sulphate salt to help solubilize the active catalyse species [Figure 14 a)]. These conditions were applied to an activated pyrimidine system where modest yields of 70% plus were achieved across 8 simple boronic acids and esters, including electron poor boronates and heterocycles. Applying these same conditions to a phenylchloride substituent was not as encouraging as many conversion rates dropped considerably. Electron poor para-trifluoromethylphenyl boronic acid was not tolerated when a chloro was in the meta position as well as poor conversion for 3-

pyridylboronic acids across the board. This reaction was shown to be successful in validation for a DNA encoded library synthesis where 107 out of 153 boronic acids/esters passed the validation of greater than 70% yield.[31] Advances in Suzuki-Miyaura couplings followed with the invention of palladium species sSphos-Pd-G2 [Figure 14 b)]. This species further utilises the solubilising properties of the ligand but incorporates it into a stable precatalyst.[45] In this example the reaction proceeded with good yields on the test substrate that was used, showing conversion rates of 70% plus on all boronic acids used. However, changing the halogenated headpiece resulted in a mixed effect on the coupling of each boronic acid. It could be therefore suggested that these conditions are substrate dependent and rely upon a specific halogenated building block for successful reactions. The DNA was also shown to be intact after this reaction as a model library member was PCR amplified and sequenced using Sanger sequencing.

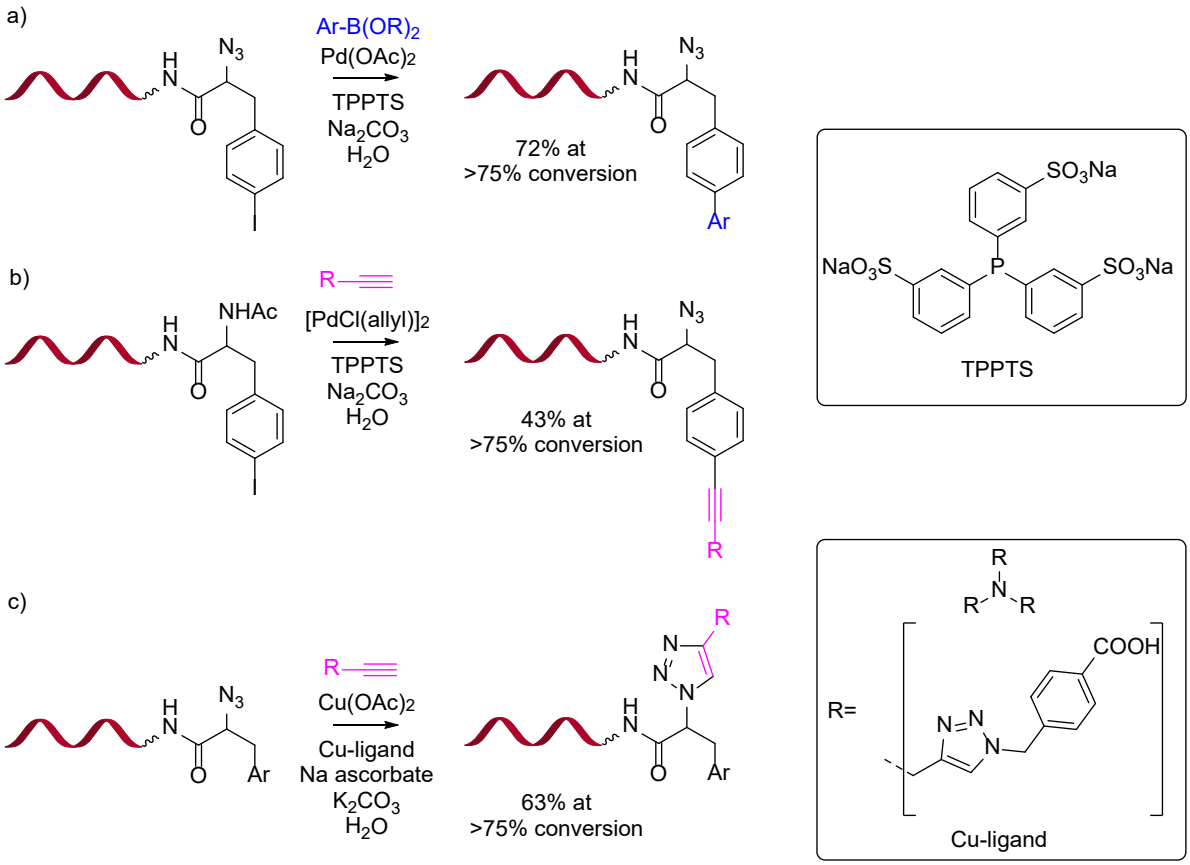


Figure 15 – Palladium catalysed Suzuki-Miyaura and Sonogoshira and copper catalysed CuAAC reactions using novel ligands.[46]

With an aim of improving Suzuki-Miyaura reactions, Sonogoshira couplings and copper catalysed alkyne-azide cycloadditions (CuAAC) were implemented on single stranded DNA for

encoded self-assembling chemistry.[46] Using palladium acetate as the catalyst and TPPTS [Figure 15] as a water soluble ligand, 72% of the boronic acids tested achieved conversion rates of greater than 75%. The problematic boronic acids were typically aliphatic boronate, acrylboronate, *p*-pyridinylboronic acid derivatives, and insoluble substrates. These results were achieved on one test iodo substrate. The test conditions were slightly modified for Sonogoshira couplings to use $[PdCl(allyl)]_2$ due to improved conversion rates. 43% of alkynes proceeded with a conversion rate of >75%. However, the more difficult alkynes were found to decompose in the reaction mixture. Finally, the use of a novel ligand for CuAAC reactions was observed to be a success. For this reaction to be successfully carried out in aqueous conditions, a novel copper ligand was synthesised. The complex was synthesised in situ by sodium ascorbate, obtaining the Cu(I) species required to promote the reaction successfully. These three examples show how many extra variables need to be considered when developing new reaction conditions. However, reactant or catalyst insolubility are recurring themes in this reaction class.

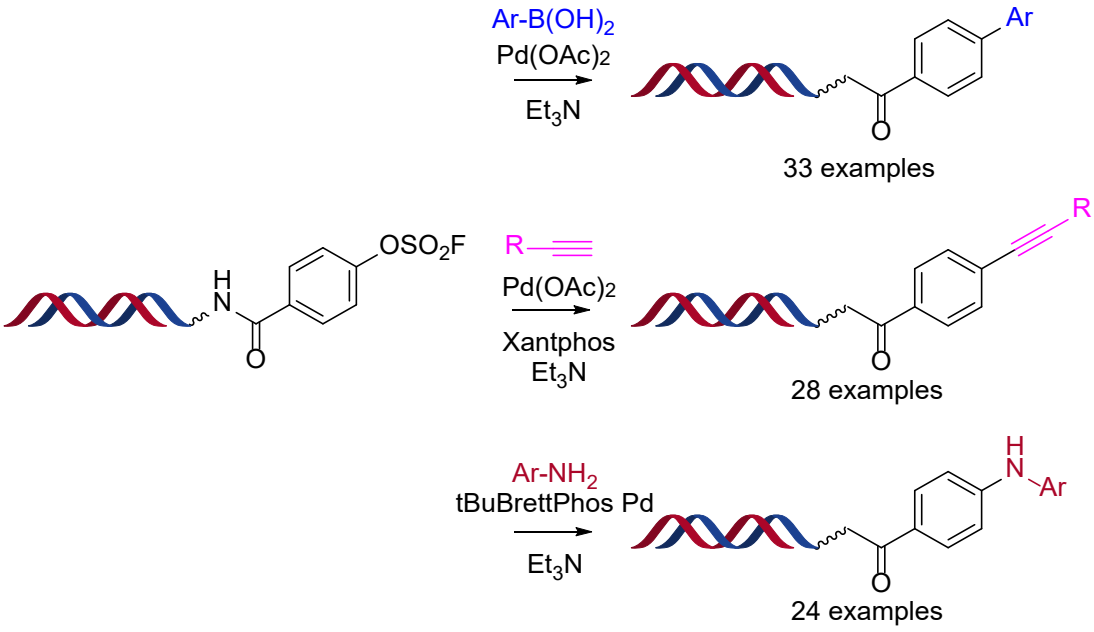


Figure 16 - Palladium catalysed Suzuki-Miyaura, Sonogoshira and Buchwald-Hartwig cross-couplings.[47]

Improved palladium coupling reactions can also be achieved by using aryl fluorosulfonates.[47] Fluorosulfonates are thought to be an atom efficient and economical alternative to triflates, as well as being successfully synthesised from phenol using sulfuryl fluoride.[48] They have also been successfully applied to traditional transition metal catalysed reactions using the

fluorosulfonate as the electrophile. When applied to DNA conjugated phenylfluorosulfonate [Figure 16], high conversions were seen for 33 examples of Suzuki-Miyaura coupling partners, at >80% conversions for a majority of boronic acids, including heterocycles, at room temperature. This was also true when applying the same system to alkynes in the Sonogoshira reaction but required higher temperatures to proceed. Finally, using t-butylBrettPhos palladium precatalyst 21 out of 24 aromatic primary amines reacted at >80% conversion rates. Using aromatic fluorosulfonates, this technology opens a new window of opportunity for metal catalysed reactions. However, although fluorosulfonates are simple to synthesise, they are rarely available commercially, which either adds extra work in the lab or limits the number of building blocks that can be used. Also, it was only proven to work on a single simple aromatic fluorosulfonate but was not shown to work on the more privileged heteroatom containing structures, such as pyrimidine or pyridine, which would be desired in pharmaceutical libraries.

On DNA palladium catalysed Heck reactions have also been studied. Although not a widely used chemical reaction especially in library chemistry, it has been used in the synthesis of available drugs and offers the ability to add different functionality to a chemical library, such as the addition of a alkyl species. A broad variety of DNA conjugated styrene/acrylamide or aryl iodides were subjected to an on-DNA Heck reaction with reasonable substrate scope and yields.[49] This reaction was carried out under relatively simple reaction conditions using $PdCl_2$(COD) and base in a DMA/water solvent mixture.

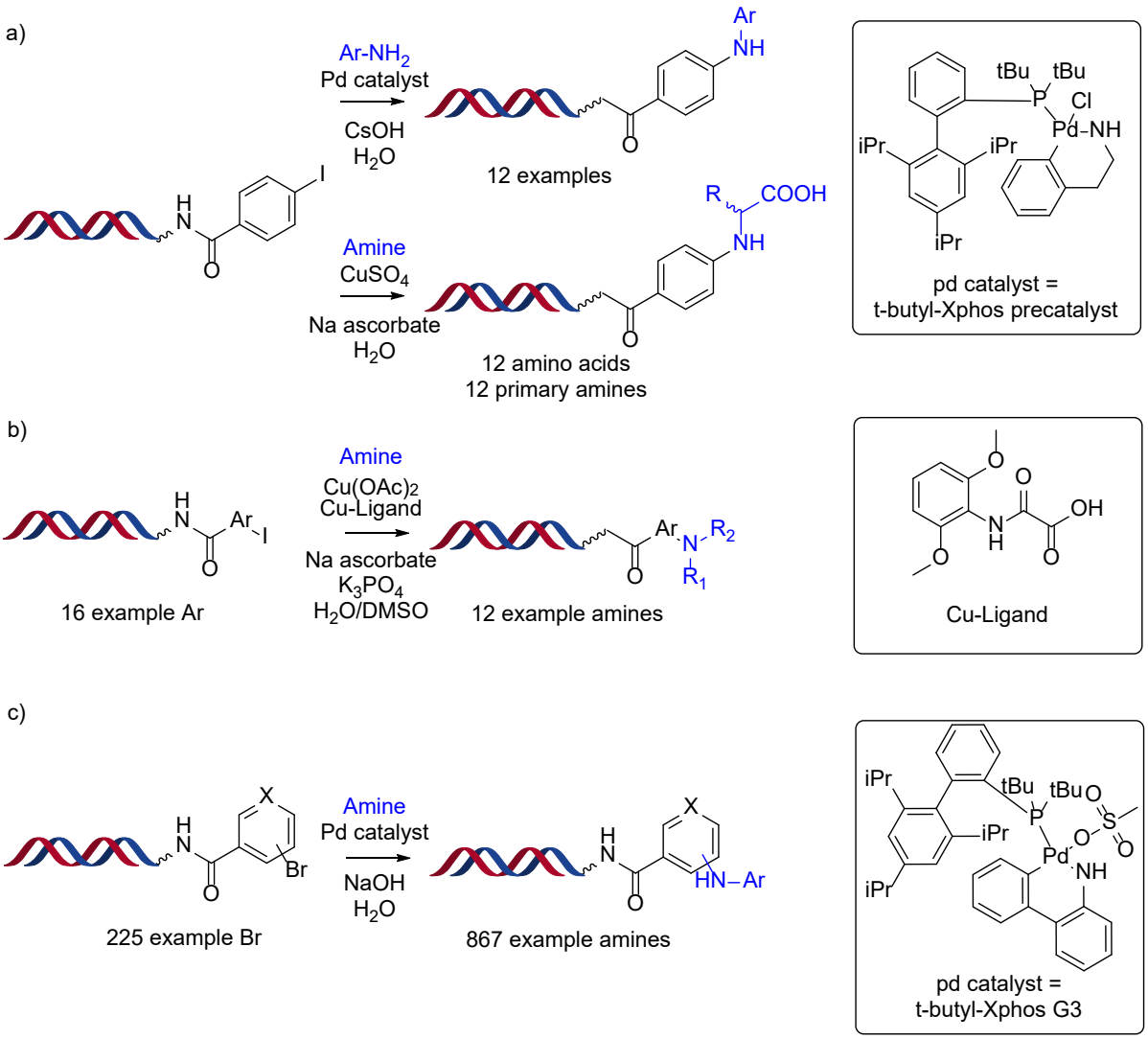


Figure 17 – Palladium and copper catalysed C-N coupling reactions.[50-52]

One of the most challenging coupling reactions to achieve on DNA is the aromatic C-N cross-coupling. Due to the huge variety of amines available, a sizable amount of chemical space could be accessed by a well employed C-N coupling. Using a Buchwald *t*-butyl-Xphos precatalyst, a reaction between a DNA conjugated aromatic iodo species and a range of amines was achieved in good yields [Figure 17 a)].[50] However, limitations did arise in this procedure where DNA damage was observed in a substrate dependent manner, and no reactions occurred for aliphatic amines. Ullmann type copper couplings were also achieved by using in situ generated Cu(I) using sodium ascorbate and copper sulphate pentahydrate. This allowed for much improved coupling conversions for both amino acids and aliphatic primary amines. These copper promoted reaction conditions were further optimised by using a specific copper ligand [Figure 17 b)] which has improved solubility and increases the reactivity at the Cu(I) active site.[51] Efficient coupling of both primary and secondary aliphatic

amines was achieved using these conditions and a variety of aromatic iodides were also tolerated. Using palladium and a G3 *t*-butylxphos palladium precatalyst, it was discovered that a variety of aromatic primary amines were able to be successfully coupled.[52] Under these conditions 36% of the 867 primary amines were successfully coupled with conversion rates of >70%. These were further coupled to 225 different aromatic bromines with 105 substrates reacting successfully. This shows that even with some great advances in C-N bond formation, there is still a huge demand for conditions that are not only applicable to both aromatic and aliphatic amines, but also to improve the scope of the reaction. In the best examples, less than 50% of the amines are successfully converted under these reaction conditions, truly limiting the diversity of the resulting chemical library.

Ring closing metathesis (RCM) reactions are widely used for the formation of unsaturated rings in the chemical and pharmaceutical industry. Macrocycle libraries are widely employed in the drug discovery process, with many synthesised using RCM reactions. Ruthenium reagents are widely used for this process and catalyse the alkene bond formation between 2 terminal alkenes. One of the issues for intramolecular RCM reactions is that they often require high dilutions to forego the intermolecular side reaction that could occur. High dilution factors are not an issue with DNA chemistry as they are already carried out in at high dilutions. However, ruthenium "Grubbs" catalysts are also insoluble in water and require a large amount of co-solvent to solubilise, as well as inducing DNA decomposition under high ruthenium loadings. It was found that the addition of large amounts of Mg^{2+} could be used to protect peptides and proteins from ruthenium induced degradation.[53] This reasoning was also applied to the phosphate backbone of DNA preventing interaction with ruthenium [Figure 18 a)]. It was observed that the use of 150 equivalents of ruthenium Grubbs Catalyst M37a, with 8000 equivalents of magnesium dichloride, in water and *tert*-butanol, high conversion rates of 85% were present in simple diene RCM reactions.[54] Other more challenging substrates were found to yield macrocycles with conversion rates of >50%, proving that these catalytic systems can be compatible with DNA encoded library synthesis which allow a routine method for the synthesis of macrocycle libraries. Similar reaction conditions were also applied to an example cross-metathesis reaction with yields of 50%.

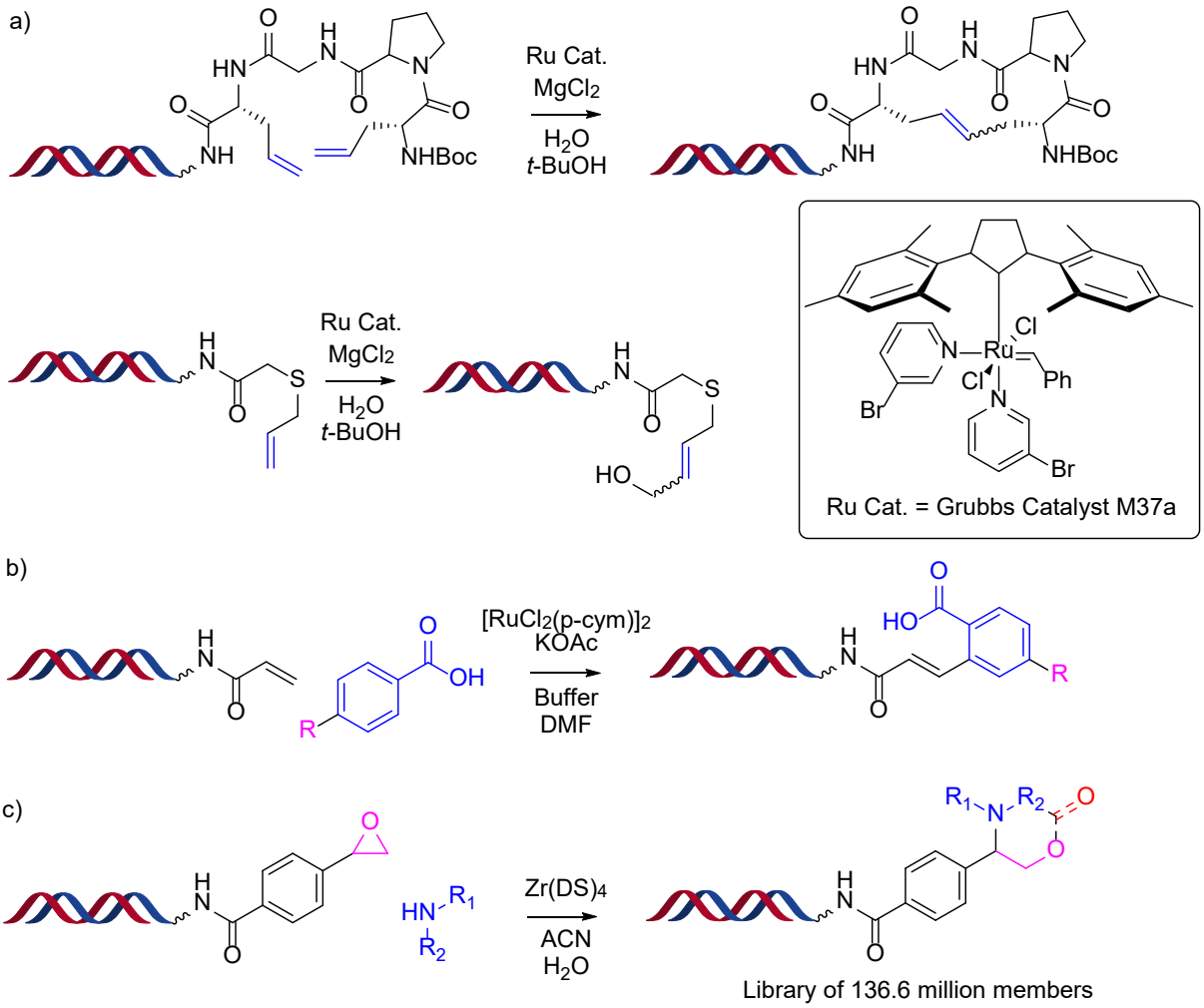


Figure 18 – Ruthenium catalysed RCM and C-H coupling reactions, and zirconium catalysed ring opening of epoxides.[54-56]

Ruthenium catalysed reactions have also been developed in the C-H activation of acrylamide and aromatic acids. Using simple aromatic acids a C-H coupling can be achieved, utilising the versatility of the aromatic acid building block, libraries in alternate chemical space are predicted to be accessible.[55] Utilising commercially available dichloro(p-cymene)ruthenium(II) dimer with potassium acetate, borate buffer and DMF, variable conversion rates were found on a variety of different acids [Figure 18 b)]. The scope of this reaction is limited by the acids used but does add promise for integrating novel ruthenium C-H activation chemistry with DNA conjugated compounds.

Zirconium has also been used as a catalyst for aminolysis of epoxides and employed in DNA synthesis.[56] This reaction is useful for creating β-amino alcohols and has been shown to work in aqueous conditions required for DEL synthesis. A Lewis acid catalysed zirconium tetrakis(dodecyl sulfate) ring opening of on-DNA epoxides has been successfully developed

and applied to a large DEL synthesis. It was found on initial testing that 8 of the 12 amines would yield over 60% conversion to product using zirconium in water and acetonitrile [Figure 18 C)]. A library of 136.6 million members was then synthesised using these conditions as a second building block addition after a diversifying amide coupling. Further functionalisation was added to the library by forming oxazolidin-2-one and morpholin-3-one heterocycle motifs from different amines used in the epoxide opening step.

Transition metal catalysed chemistry has improved significantly over the past few years in respect to DNA compatibility and substrate scope. There are however still significant gaps in the outcome of certain chemical transformations and a vast amount of research is still needed to improve these reactions to a degree more like solution phase chemistry. The major issues with this chemistry seem to be the solubility of reagents and catalysts, high catalyst loadings required and catalyst degradation. In turn this means harsh conditions are required which can damage DNA or poison the catalysts in question. If these issues are addressed then this chemistry could become a vital tool to create vast diverse encoded libraries.

1.6.3 – Photoredox and Radical Reactions

Access to C-C bond formation could enhance the diversity accessible in chemical library design and synthesis. C-C bond formation is often a difficult reaction to access as many reactions require a metallated organic fragment to progress which are often not tolerated under the aqueous conditions required for on-DNA synthesis. However, using redox promoted radical reactions it is possible to access sp^3 C-C bond formation chemistry. The first example of this chemistry was the decarboxylative Giese reaction, using a 1,4 addition of radicals to DNA conjugated acrylates.[57] A 6-stage outline of how to translate organic reactions to a DEL platform has been used to achieve a novel DNA compatible redox radical reaction. The first step was to evaluate compatible reactions that are not air or water sensitive, then to determine if the concentration has a positive driving force using a reaction progress kinetic analysis (RPKA). In the third step, the evaluation of solvents, additives and temperatures is used to determine further compatibility. The starting conditions are then applied to a DNA conjugated system in the fourth step, where a further evaluation of scope would need to be made in the fifth step. Finally, the optimised conditions are then moved to a DEL platform for synthesis of a library. Using these steps, the chosen Giese reaction was optimised for use on DNA using a redox active ester to access diversity. It was found that using zinc nanopowder

as an electron source the redox active esters were able to react, with no additives, with reasonable conversion rates of 40-90% depending on the substrate used [Figure 19 a)].

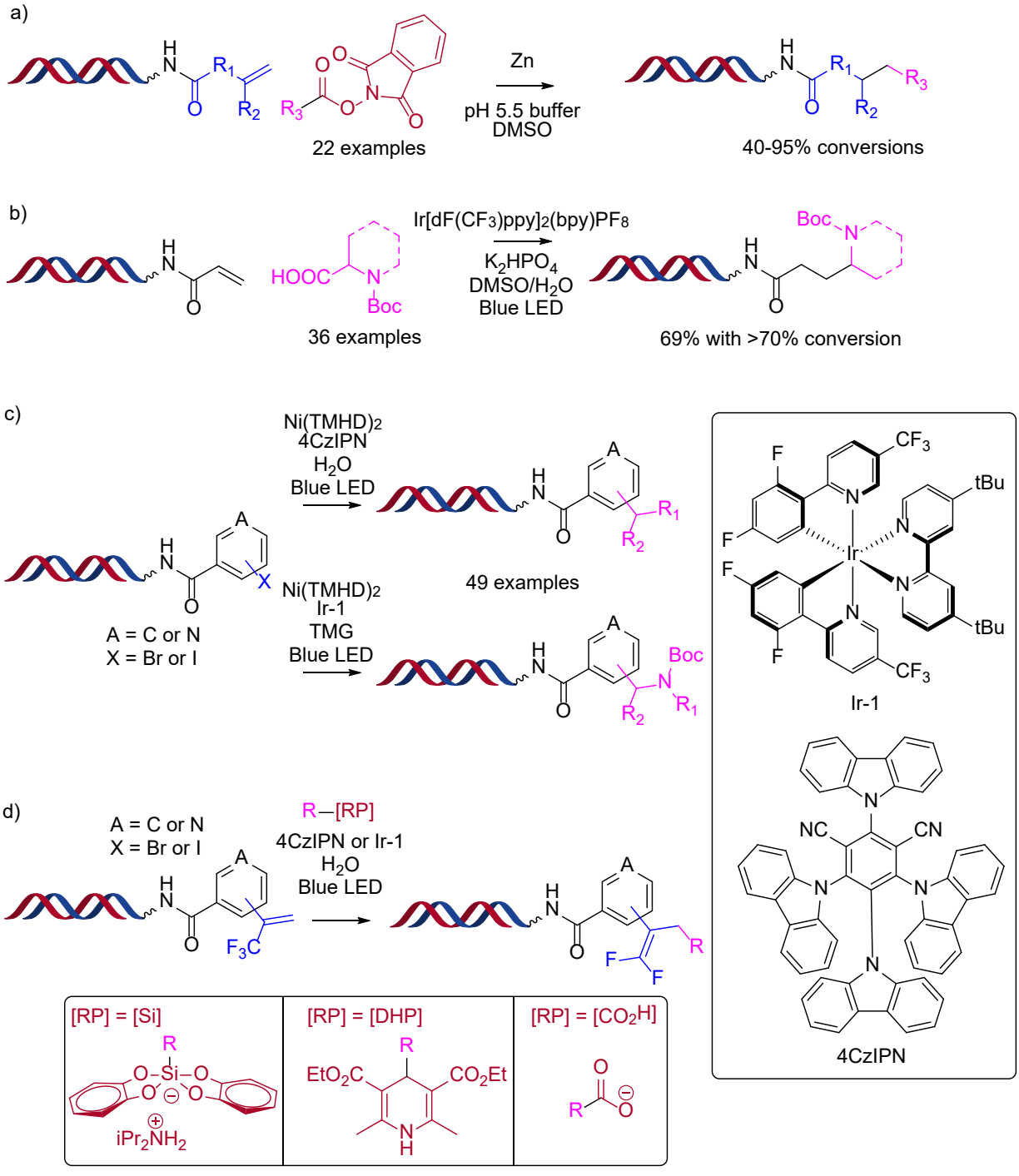


Figure 19 - Redox promoted radical reactions compatible with DNA.[57-59]

Photoredox promoted chemistry can also be applied to DELs. An sp^3 C-C bond formation was achieved using a radical created from an α-amino acid addition to a Michael acceptor and styrene conjugated to DNA.[58] In this example, an iridium complex is excited by a blue LED light source, activating a radical cascade by a decarboxylation event on an α-amino acid. This

radical species is then quenched by a Michael acceptor in a Giese type coupling [Figure 19 b)]. In this example, 36 α-amino acids were subjected to the conditions, and 69% of these progressed in the reaction at over 70% conversion rates. This reaction was then subjected to a potential 3-cycle DEL design library. The first step is to couple 100 Michael acceptors, then 500 protected amino acids are subjected to the radical Giese type reaction, finishing with a capping reaction with 1500 compounds to make a 75 million library.

Dual catalytic photoredox reactions have also been shown to occur on DNA.[59] A 4-alkyl-1,4-dihydropyridine $C(sp^2)$-$C(sp^3)$ cross-coupling, using nickel and 4CzIPN as the photocatalyst [Figure 19 c)], successfully converted DHP functionalised molecules with a variety of aromatic halogenated DNA conjugates. An iridium complex (Ir-1) was used as a photocatalyst alongside nickel, to cross-couple Boc protected amino acids with the same aromatic halogenated DNA conjugates. These reactions were both carried out at ambient temperature, over very short time frames and can be easily scaled to working in 96 well plates.

A defluorinative alkylation of trifluoromethylalkenes was also achieved under photochemical conditions [Figure 19 d)]. Here a radical precursor, such as silicate, DHP or a carboxylic acid can be used to generate the radical in the presence of 4CzIPN and blue light, which is then quenched by the trifluoromethylalkene. The gem-difluoroalkene motif has been proposed to act as a more metabolically stable carbonyl isostere.[60] A further 76 examples of compatible building blocks have been produced to greatly increase the scope of the reaction.[61] Given that radical induced reactions have the ability to damage DNA, the products from the defluorinative alkylation were subject to PCR amplification and sequencing, showing no detrimental DNA damage or lower amplification.

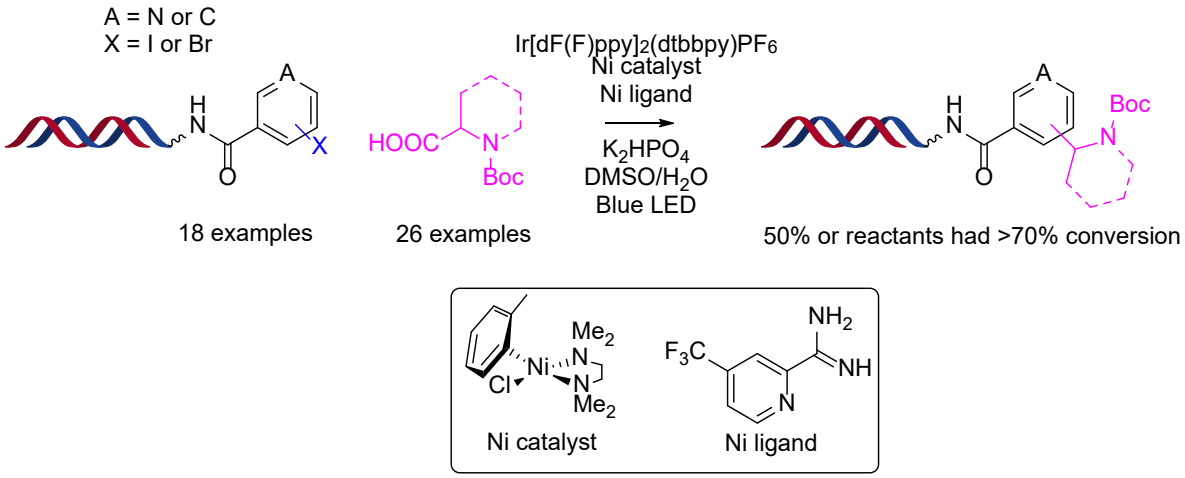


Figure 20 - C(sp^2)-C(sp^3) cross-coupling of Boc protected amino acids using nickel and iridium photocatalysis.[62]

A similar reaction was also developed to make use of C(sp^2)-C(sp^3) cross-coupling of Boc protected amino acids using nickel and iridium photocatalysis [Figure 20].[62] In this example, a diverse set of both amino acids and DNA conjugated heterocycles were subjected to the reaction, with 50% of the reactions proceeding with greater than 70% conversion. Interestingly, the reaction was shown to be scalable with 96-well chemistry, using a blue LED powered array. Four products were successfully reacted under these conditions, importantly proving that photoredox chemistry can be easily scalable for a split and pool combinatorial type library. The reaction was then used in a potential library design, utilising 3-cycle diversity, a potential 60-million-member library, with good chemical coverage could be synthesised. This approach seems very feasible to build large libraries and will surely be used in future library builds.

Utilising photoredox and radical reactions could allow access to a vast amount of diverse chemical scaffolds previously not thought possible. This chemistry also has the added benefit of often being compatible with aqueous conditions using low temperatures and being easily scaled to operate in a library format. It also seems to be compatible with DNA with little to no damage seen in the examples that have been reported. The scope of these reactions is still limited since there are often limited building blocks or specific substrates required. However, it is extremely promising and the use of this sort of chemistry could access a variety of novel chemical scaffolds.

1.6.4 – Solid Supported Reactions

Reactions must be performed in aqueous media for DNA to be in solution. However, a variety of novel techniques have been produced to carry out reactions on a solid support. One of the major benefits of using solid supported DNA is that reactions can be conducted in organic solvents using reactants that would usually be incompatible with aqueous reaction conditions. This advance could allow for libraries to be designed with significantly higher diversity as chemistry that is previously inaccessible can be utilised. In turn, it could increase the number and variety of hits developed when screening the library. This approach is distinct from the favoured solution phase approach and rediscovers the initial premise that Brenner and Lerner implied using a solid supported method to build the DNA tag and the peptide library.

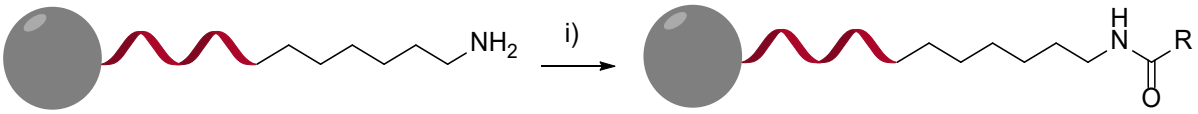


Figure 21 – Amide coupling of acids to DNA attached to commercially available solid support;[41] i) HATU, DIPEA, DMF.

Simple reactions such as amide couplings can be achieved on large scale using solid supported material. DNA is synthesised on a bead as mentioned previously and is usually cleaved from the solid support prior to synthetic manipulation. However, if the DNA is retained on the solid support, then chemistry can be applied to the modified DNA. This technique allows for coupling of long chain "linkers" or even first building block addition utilizing the solid support on which the DNA is synthesised.[41] The major concern with this technique is that only a single building block addition can be carried out on solid support as there is no way of coding additional building blocks after the first step. The DNA is cleaved from the bead and coded after the first building block addition. Using solid supported processes does not necessarily protect the DNA, and depurination events can occur in acidic or other denaturing environments. To improve the reaction diversity using solid support, purines can be removed from the initial DNA sequence.[63-66]

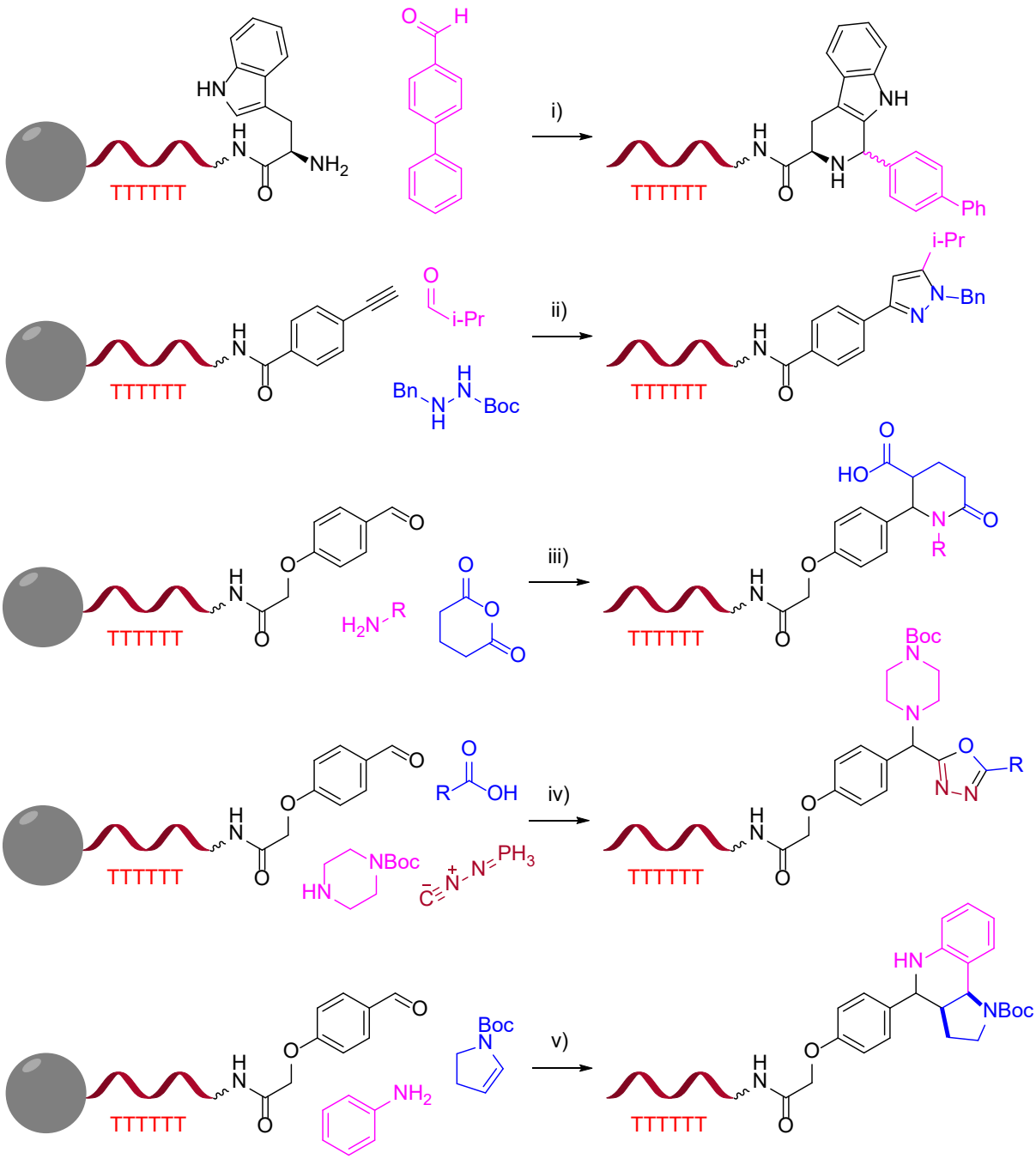


Figure 22 - Scope of "HexT" solid supported reactions, i) β-carboline synthesis using 1% TFA in DCM;[63] ii) Pyrazol synthesis using AcOH;[63]; iii) Castagnoli–Cushman reaction using $Yb(OTf)_3$ in DCM;[64] iv) 1,3,4-oxadiazol synthesis using DCE;[66] v) Povarov reaction using EtOH and TFA.[65]

Using an initial DNA sequence consisting of 6 thymidines, called HexT, reactions were performed on solid support under conditions that would usually lead to depurination of the DNA sequence. The first examples were shown using 1% TFA in DCM [Figure 22 i)] to carry out a solid supported β-carboline synthesis. In this example, 100% conversion rates were produced, applying conditions that would be completely incompatible with solution phase DNA. After this initial example, multiple component reactions were all successfully deployed

on this substrate. Heterocycle synthesis is one of the most sought-after chemical transformations for inclusion in libraries as many heterocycles are privileged structures in medicinal chemistry. Oxadiazole and pyrazole syntheses have both been achieved using multicomponent reaction protocols [Figure 22 ii) & iv)] in excellent yields and with a variety of substrate scope. Interestingly, the aza-Wittig reaction was accomplished using these conditions which would not be otherwise compatible with aqueous solution phase DNA encoded synthesis. Also, multicomponent named reactions such as the Povarov and Castagnoli-Cushman reactions have been performed with high conversion rates [Figure 22 iii) & v)]. These solid supported reactions also benefit from the fact that they can be easily purified by filtration and washing with organic solvents to remove many impurities. The DNA is then easily cleaved from the bead, and a sequence that codes for each transformation or reactant can then be ligated. However, one of the issues with this technique is that the transformation must be carried out as the initial reaction in the library synthesis. This limits the capacity of the library due to the design having to incorporate the transformation as the initial coding step. Also, this technique has not been proven to work with large library sizes, maybe due to the limited number of reactants that are tolerated in the library build or to handling issues of hundreds of wells of solid supported material. This technique does, however, have scope for further development and could become a vital tool for the design of novel and diverse chemical libraries.

Solid phase chemistry is not limited to using a “HexT” strategy and performing only the initial transformation on solid support. Often when solid supported chemistry is performed on a small molecule, it is attached to a resin by an irreversibly cleavable covalent linker. This is a similar method seen in the above “HexT” examples. However, if reactions are to be carried out in the second, third, or even fourth cycle, the DNA would need to be able to be reversibly bound to a solid support. This would allow the DNA to be bound to the solid support for the reaction and then eluted for encoding purposes. The library could then be pooled together and repeated allowing a combinatorial synthesis to be used. This approach offers the stability of reactions in organic media with the ability to then encode each reaction step. For this to occur an efficient media is required for binding and releasing the DNA, which has a robust solvent integrity and is resistant to the desired chemical transformations. Critically, there is a large difference in binding kinetics of biomacromolecules and small molecules, which allows

for adsorbed bio molecules to react with the small molecule reagents.[67] The very first solid supported resin that was found to have excellent bind and release properties was DEAE Sepharose.[11, 68, 69] Using this resin it was shown that single stranded DNA of up to 340 bases length was able to be bound and eluted with high efficiency and displayed no resin compression in any organic solvent. It was shown that peptide coupling could be achieved at 90% conversion rates in all natural, D and protected amino acids using this solid support. Also, after the reaction had taken place, elution of the DNA was near to 100%, and the reaction was purified with simple washes and filtrations.

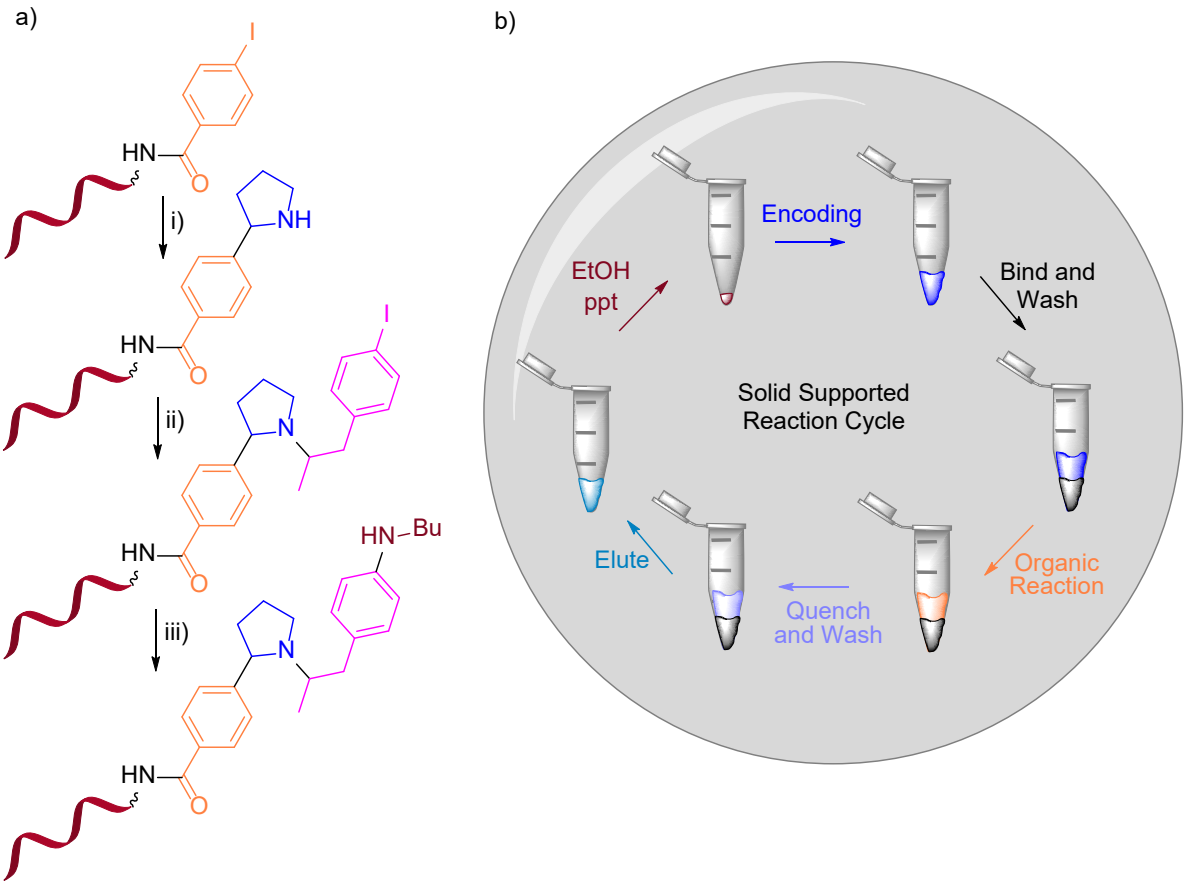


Figure 23 - A) DEL rehearsal reaction scheme using 3 solid supported reactions, i) Decarbonylative cross-coupling using Ni:^{t}bubpy, RAE, K_2CO_3, Rubin's Silane, ii) Reductive amination using $B(OH)_3$, $NaBH_3CN$, H_2O, NMP, iii) e-Amination using $Ni(bpy)_3Br_2$, graphite electrodes; B) The solid supported reaction cycle used in each organic synthesis step.[70]

After the success of using solid supported peptide and protein immobilisation for synthetic chemistry, a paradigm was suggested to use solid support for more challenging chemical transformations on DNA.[70] Hydrophobic resins were screened, but weak binding of DNA led to premature elution. Weak anion exchange resins were ruled out as they hold basic, nucleophilic, or reactive moieties. DEAE-Sepharose was also screened but ruled out due to

the presence of abundant hydroxylated functionality which would limit the scope of employable reactions. The solid support that was therefore used was a polystyrene bound anion exchange resin which overcomes the issue of Sepharose. It also incorporates hydrophobic interactions as well as an electrostatic interaction which efficiently anchored the DNA in place. Bind and release was achieved in a simple manner using PBS to load the DNA and eluting with $NaClO_4$ buffer. The DNA could then be pelleted afterwards for use in subsequent reactions. The reaction profile of using this solid support was then tested employing an on-DNA decarboxylative sp^2-sp^3 cross-coupling. Under aqueous conditions this reaction did not proceed and completely degraded the DNA tag. However, applying solid support and a nickel catalyst, the reaction produced the desired product with yields of up to 80%. The substrate scope was also analyzed and shown to successfully cross couple a variety of acids with varying yields of 50% plus, also using iodo heterocycles with 40% plus conversion rates. The reaction was demonstrated to work considerably better than when using aqueous conditions. However the substrate scope is still limited, and high yields were only achieved with simple iodo aromatics and low functionality sp^3 centres.

Another interesting use of this technology is the first application of electrochemistry to DNA encoded libraries,[70] which was applied to an amination reaction. These reactions are used consistently to provide great diversity in organic chemistry and the technology was applied to DELs using a nickel catalysed electrochemical process. Again, under normal aqueous conditions it led to complete DNA degradation, however, when applied to a system of solid supported DNA, yields of between 30 and 75% were produced. This advance shows that using solid support in situ with electrochemistry, could allow access to a diverse range of chemical transformations that would be applicable to DNA encoded library synthesis.

This technology was then applied to a test system [Figure 23] in which a three-step synthesis using all three solid supported reactions was used as a 1x1x1 proof of concept for a library build. Using a cut-off point of at least 30% recovery for each reaction, 3 reactants were picked for this synthesis. The headpiece was ligated to a 20mer primer, and then subjected to a three round synthesis starting with a sp^2-sp^3 coupling of Fmoc protected proline. The resulting product was deprotected, a reductive amination performed on the free amine and then a final electrochemical amination to yield an example library product. This reaction process was

achieved with an overall yield of 9%, but no attempt was made to PCR and sequence the product.

There have been some great advancements in the field of solid supported chemistry and its application to DELs. The simplicity of using a covalently linked "HexT" solid support means that initial building blocks can be designed using complex reactions that would typically be incompatible with DNA. However, as further solid supported chemistry cannot take place, these conjugates must be cleaved from the support prior to further chemical building block additions. Using an ion exchange resin allows for subsequent reactions to take place, as well as allowing access to chemistry that had previously been out of reach of DNA compatibility. However, the negative of this technology is that the reactions that have been employed on DNA either have poor substrate scope or poor conversion rates. Often a cut-off point for reagents is taken at 50% conversion, which is very low and would create large amounts of false negatives in the library. However, using the solid supported media, even on test systems the conversion rates are around 50%. This would massively limit the number of compatible reagents and therefore limit the size and variety of the libraries created using this technology. Also, the electrochemical reaction, in respect to the chemistry, is a novel method for carrying out aminations but has not been shown to scale to a 96 or 384 well reaction plate. Although huge steps have been taken to create novel and compatible reactions on solid support, there would need to be considerable improvements to scope, conversion rates and also scale up of these reactions for them to be truly compatible to a DNA encoded library platform.

1.7 – DNA Encoded Library Screening

Screening of an entire library can be achieved with relative ease compared to other high-throughput methods of drug discovery. Therefore, a well-designed DEL can be an effective tool for initial ligand discovery. The entire library is incubated with the target protein, either attached to an affinity tag[17] or utilising covalent capture techniques.[71] The protein is then washed eluting any non-binding library members and retaining the binding ones. This method is then repeated a further 1-3 times depending on the library size, to allow detection of only the bound compounds and reducing the likelihood of non-binding ligands being present, resulting in fewer false positives in the screen. Once repeated, the binding compounds are eluted, amplified by PCR and sequenced. This process is relatively straightforward and swift

compared to many other screening methods, allowing it to be carried out in standard laboratory conditions and producing results within 10 days.[72]

The above method can only be utilised to carry out identification of non-covalently bound ligands. However, in recent years there have been reports of a library containing Michael acceptors to identify cysteine reactive ligands. Using this method it was demonstrated that it was possible to discriminate between high affinity noncovalent binders and the desired covalent bound inhibitors. The library was produced by a split and pool approach, using PNA recording, to produce a 10,000 member library that was used to identify novel covalent binders of MEK2.[73]

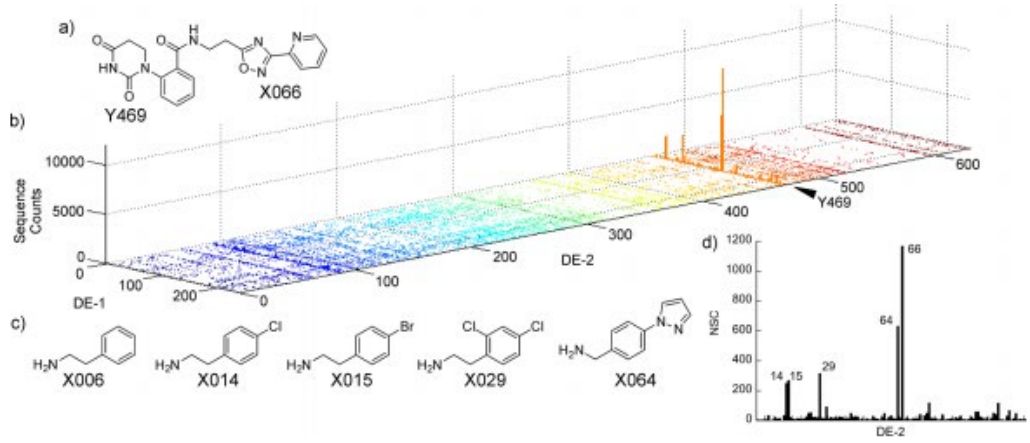


Figure 24 - A 3D plot of results gained after sequencing results for TNKS1 binding; a) the chemical structure of the most enriched hit compound; b) Affinity table against the target protein. Each peak corresponds to an encoded sequence, and the height of the peak corresponds to the number of sequences obtained; c) shows the structure of the most enriched amines.[73]

One of the concerns with screening a DEL is that off-DNA follow up compounds need to be synthesised in order to confirm that the hit fragments are genuinely active. Once the binding compounds have been amplified and sequenced, the data is then analysed using graphical representations [Figure 24]. This shows which fragments have bound to the target by matching the enriched DNA sequences to the building blocks. The fragments can then be combined to replicate the original library hits and synthesised off-DNA. One issue with this method is where to "cut" the molecule from the DNA tag when resynthesizing the compound from DNA. Due to the close proximity of the linker moiety to the organic molecule, the linker could be involved in the binding of the compound,[74] or it could be detrimental. This often leads to several variations of the same hit compounds being synthesised, either leaving a small

element of the linker[75] or complete removal of the actual linker as well as the attachment from the building block.[76] Due to the nature of the DNA tag and linker, the actual compounds that are screened could have high hydrophobicity and keeping parts of the solubilising tag may be beneficial. Utilising a position that can tolerate a linker can be very beneficial when it comes to target validation of a hit molecule. Fluorophores are likely to be tolerated at the same position as the linker molecule allowing for easy access to hit characterisation. Wu et. al. used this tool to add a DNA-fluor in the position of the original DNA tag and analysed by FACS to find a 12-fold increase in fluorescence for the positive cells compared to the control.[77]

1.8 – DNA Encoded Library Case Studies

DNA encoded library technology has been present for over 15 years as a tool for drug discovery. The ability to create huge libraries relatively simply and inexpensively, as well as screen these libraries rapidly, has meant that many libraries utilising a variety of reactions have been synthesised. Although there are many limits to the synthesis and building block selections, it has been used as an effective tool for generating drug candidates in drug discovery. Historically, HTS and fragment screens have been used to screen bioactive targets. However, the ability to synthesize huge libraries has meant that DNA encoded technology has been used as a further tool to aid hit discovery.

1.8.1 - Receptor Interacting Protein 1 (RIP1) Kinase Inhibitors

RIP1 has emerged as an important driver of tumour necrosis factor mediated inflammation and pathology. It has been shown that activated RIP1 can directly regulate proinflammatory cytokine production and apoptosis, as well as being a critical downstream driver of various other pathways.[78, 79] This allows for potential RIP1 activation for multiple inflammatory disease.

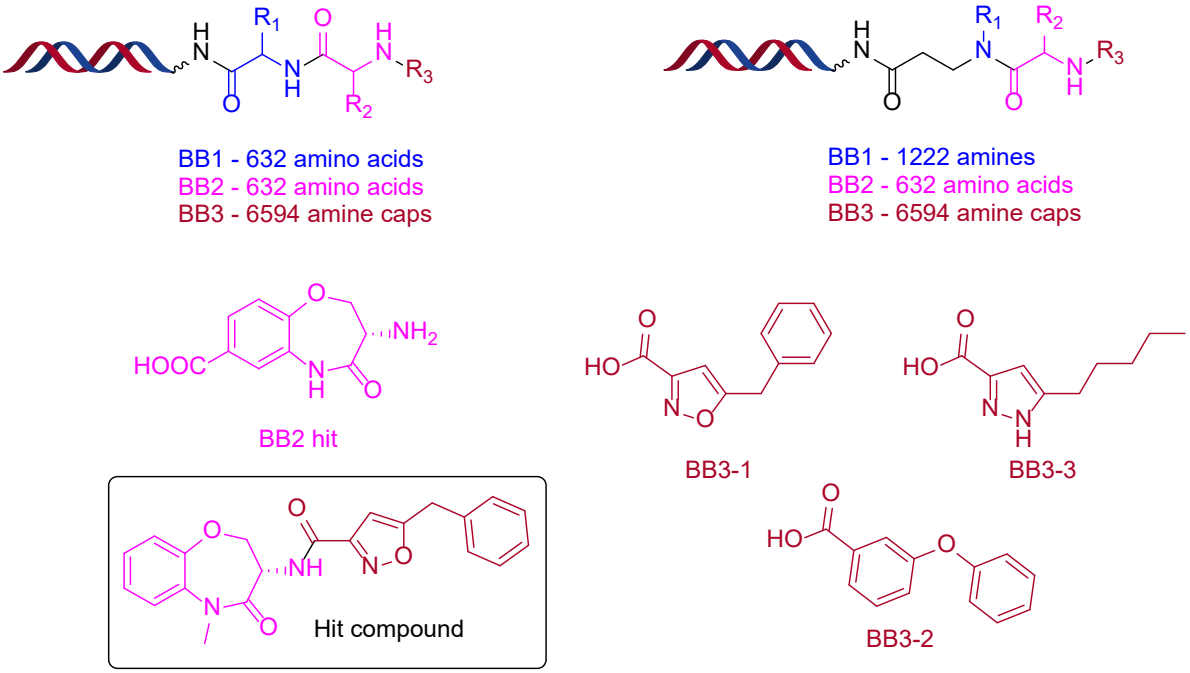


Figure 25 – Three cycle amino acid libraries used to screen against RIP1, the enriched building blocks found as a result of the library and the best hit compound from the library.[80]

To find novel RIP1 inhibitors, affinity selections of GSK's propriety collection of DNA encoded libraries were used against RIP1 kinase. Using three cycle amino acid libraries [Figure 25], a 7.7 billion member library was constructed.[17, 18] The resulting data gathered showed that benzo[b][1,4]-oxazepin-4-one was the only one of 632 BB2 amino acids to feature, and only 3 structurally similar BB3 analogues out of 6594 amines, showing how selective these libraries were.[80] BB1 however, had no difference in enrichment, showing that BB1 contributed very little to binding. The two building blocks were then coupled together removing the acid on BB1, to give 3 novel combinations that were used for further studies. It was found that all 3 were potent biochemically, with two analogues showing IC_{50} values of <100 nM, and a 5-10-fold decrease in potency was observed upon testing in human monocytic U937. The oxazole containing compound also showed good oral systemic exposure in rat, with an area under the curve (AUC) of 2.2 µg·h/mL, a maximum concentration (C_{max}) of 810 ng/mL, and a half-life ($t_{1/2}$) of 2.9 h at a dose of 2 mg/kg. These great enzymatic and cellular activities, as well as a low molecular weight and oral PK profiles, allowed this compound to be a great starting point for optimisation. The only change that improved this compound was to methylate the amide in benzoxazepinones which generated a compound with high potency inhibition of RIP1 and showed great selectivity against >450 off-target kinases. This project was a successful

application of DNA encoded libraries, generating a hit compound that contains a versatile template for further drug discovery programs.

1.8.2 - DDR1 Inhibitors

Fibrosis impacts several organs and is the common pathway to renal failure in chronic kidney disease. The Discoidin Domain Receptor 1 (DDR1) represents an attractive target as it is the only tyrosine kinase collagen receptor.[81, 82] However, efforts to generate a selective DDR1 small molecule inhibitor have been unsuccessful. Using a parallel DNA encoded library screen against both DDR1 and DDR2, a highly selective chemical series was obtained.[83]

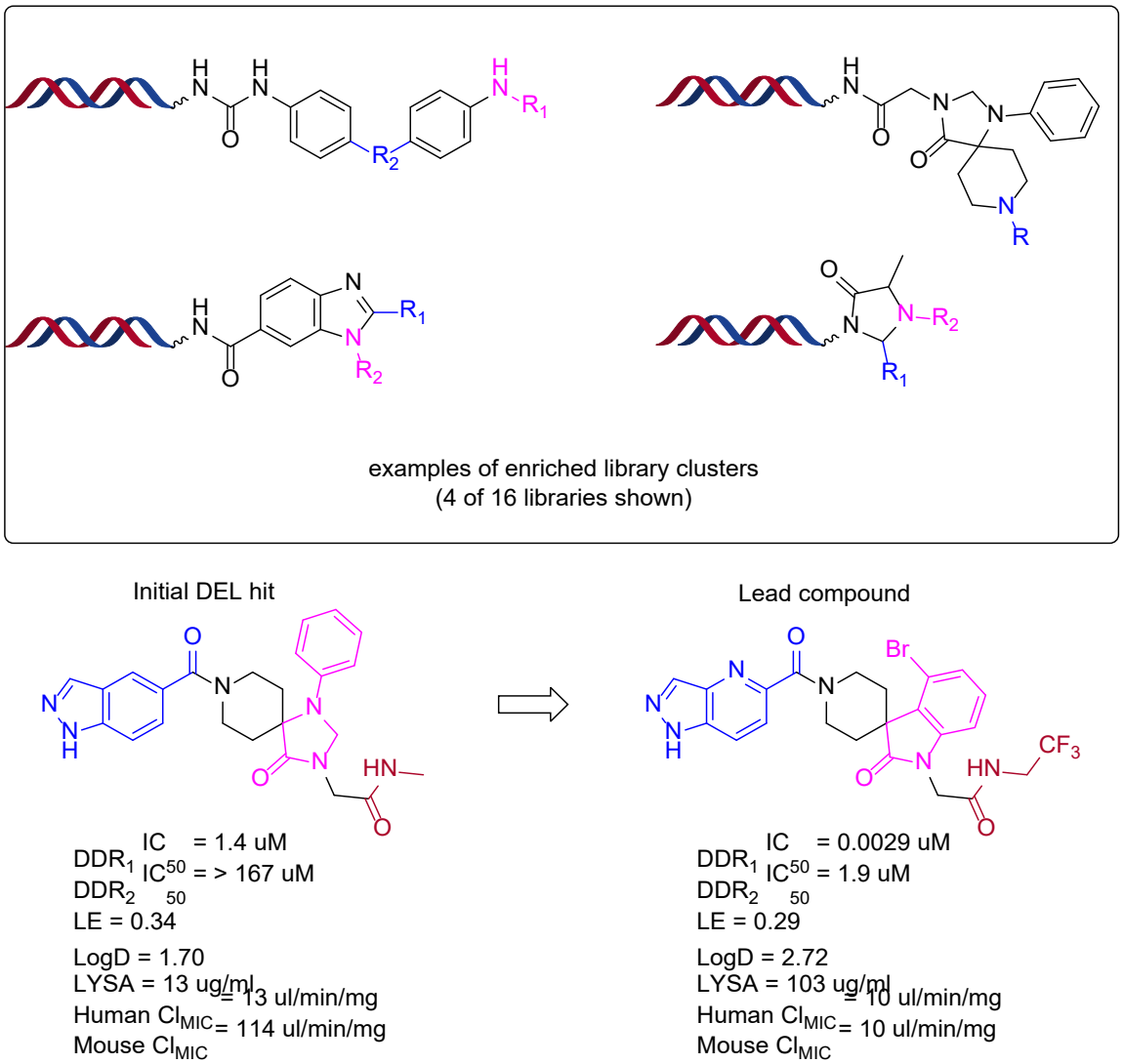


Figure 26 – 4 of the 16 libraries used in the DDR1 and 2 screen, with initial hit compound from the screen and the lead compound after SAR.[83]

After performing several screens utilizing 16 different libraries, 11 structurally distinct chemical clusters were found with confirmed DDR1 activity. The general library screen

contained two pools of ~85 billion compounds, and each library contained DNA encoded identifiers, allowing them to be screened together. The 11 clusters identified were from 5 of the individual libraries, showing that 11 screened libraries failed to contribute to DDR activity. Four clusters were found to have greater binding than others [Figure 26]. Off-DNA synthesis of compounds from each cluster identified 16 novel active compounds each with drug-like properties and clear vectors for optimisation. It was found that the cluster containing 1,3,8-triazaspiro[4.5]decanone spirocycle had great selectivity over DDR2, and rapid potency increases on SAR analysis. It was also attractive as the motif provided distinctive space-filling three dimensional properties which positively impacted the drug-like properties of the compounds.[29] Some SAR analysis quickly showed that introducing a trifluoromethyl instead of acetamide increased activity significantly. Introducing a nitrogen to form a pyrrolo-pyridine ring increased solubility and selectivity over DDR2. This compound showed metabolic instability in the spirocyclic core and so was replaced with spiro[indoline-3,4'-piperidine]-2-one, with further 4-bromo substituent increasing the potency. Overall, this compound had great binding and was produced from an interesting library design utilising a novel spirocyclic core. Using 16 different libraries in parallel to produce multiple novel hit molecules that are selective against DDR1, but not DDR2, is a great use of DNA encoded libraries. This shows the variety of applications these libraries can have and as a means of achieving selectivity, screening huge amounts of compounds in parallel against both target and those with close homology.

1.8.3 - Soluble Epoxide Hydrolase Inhibitor

Soluble epoxide hydrolase (sEH) is a member of the α-β-hydroxylase-fold family of epoxide hydrolases, with a dual function of a C-terminal hydrolase domain and a N-terminal phosphatase domain. The major role of sEH is to metabolize epoxyeicosatrienes which have been shown to be modulators of endothelial cell function in the control of vasomotor tone and in the proliferation and survival of endothelial cells.[84, 85] sEH inhibitors that can potentially block the metabolism are proposed to have a potential therapeutic use in a number of

diseases including diabetes, metabolic syndrome, cardiovascular disease, respiratory disease, and neuropathic pain.

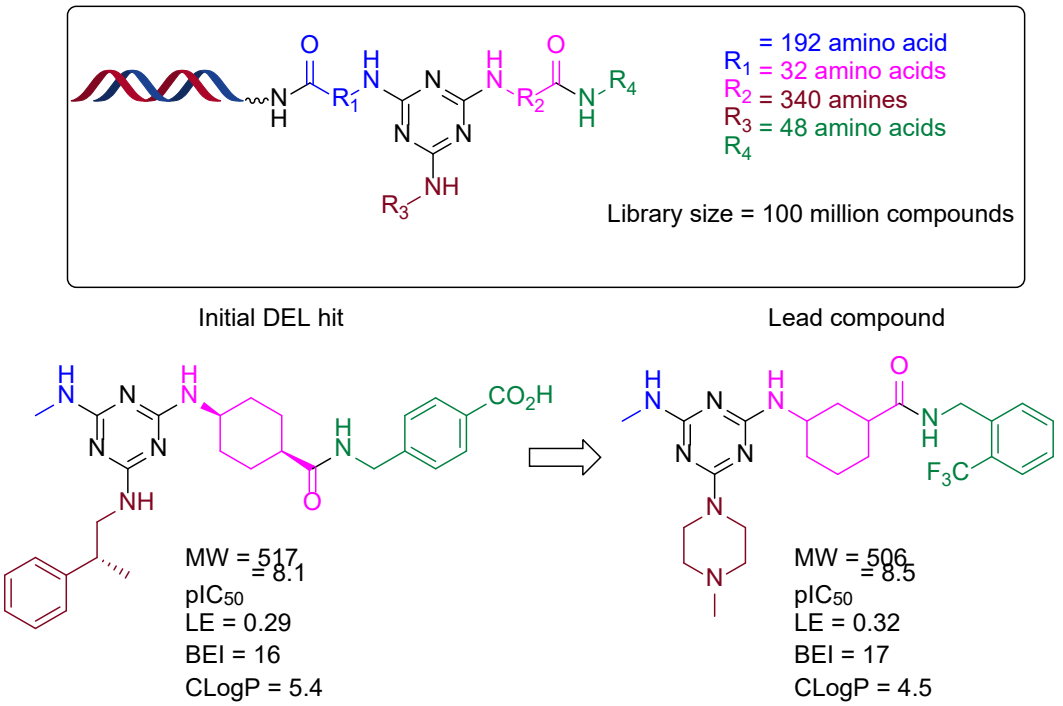


Figure 27 – Initial library design to synthesize a 10^{15} library using a 4 diversity point triazine library, the initial DEL hit and the lead compound after SAR.[86]

A triazine library of 100 million compounds, utilizing a four step diverse library were used in this screen.[86] Three rounds of selection were performed, each using 10^{15} molecules with 1 µM hsEH in the presence or absence of 1 µM AUDA (a potent sEH inhibitor) or a control with only buffer used. q-PCR was performed after each selection and the round 3 outputs were PCR amplified, sequenced and analysed. Many building block hits resulted from this screen, with too many available for resynthesis and follow-up. Therefore, additional selections were made with decreasing concentrations of sEH, at 100 nM and 10 nM, and focused on molecules which persisted at lower protein concentrations. A reduced set of compounds was then synthesised and tested for sEH activity, all of which had an IC_{50} of 8-50 nM. From this set of compounds the initial DEL hit [Figure 27] was discovered, where neither the first building block (blue) nor the third building block (red) were crucial for activity. This permitted modifications to be made at these positions to improve the molecular properties of the hit compound. Further optimisation led to the discovery of the lead compound, with enhanced potency and superior molecular properties. This compound was then taken into pre-clinical development.

1.9 – Future of DNA Encoded Libraries

There are many advantages to the DEL platform for screening compounds against target proteins. The cost of screening a DEL is incredibly low compared to HTS.[87] Screening a 1 million member library using the conventional HTS method could cost over half a million dollars. This results in a library where each member costs in the region of $1000. Due to this, HTS is often only available to large organisations with the capability and resources to deal with these figures as well as the infrastructure in place to store and run the library. This means that small companies and academia often need to collaborate with big pharmaceutical companies and CROs to operate assays. In comparison, DELs are a more cost-effective way of making and storing libraries. To create a DEL with 800 million members is thought to cost approximately $150,000, a fraction of the price of a HTS type library. Also, a library of this size could be stored in one vessel thus removing the need for infrastructure and resource to manage and deploy the compound library. Screening a DEL only requires one assay well and the whole library is then assessed against the target in parallel. This directly opens the possibility of small companies and academia having access to their own libraries which in turn should improve the hit rate of potential ligand binders.

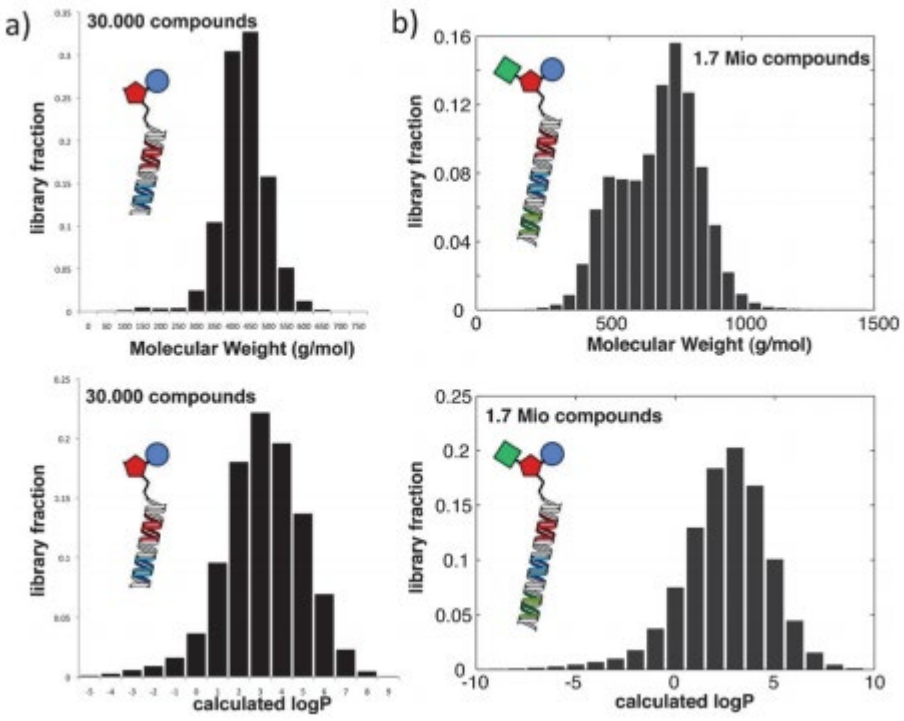


Figure 28 - The above tables show the estimated molecular weight and logP by the program Instant J. Chem. a) The estimated molecular weight and logP of a 30,000 member library with

2 points of variation; b) The estimated molecular weight and logP of a 1.7 million compound library with 3 diversity points. [88]

The current drawbacks of DELs are quite apparent. The main concerns are the variation of chemistry, the amount of chemical space that can be probed by such methods, and in some cases, poor physiochemical properties of the final compounds produced. Often a library of just 3 diversity points would contain a higher molecular weight than desired [Figure 28], therefore producing hits that are poor starting points for drug discovery.[88] As many drugs on the market contain a clogP of lower than 4 and a molecular weight of less than 400,[89] the calculated properties are less than ideal. Many of the problems lie in the design and synthesis of a DEL and could possibly be improved by more successful chemical methodology. This should expand as more DNA compatible chemistry is reported. Also, the need for more drug-like lead compounds is quite crucial, so a method that could be used to design much smaller molecular weight compounds with better pharmacokinetics would have a large impact.

The technology has advanced rapidly in recent years. From the first concept design in 1992 by Brenner and Lerner of a chemically synthesised library with individually encoded building blocks, to the present time where many novel ligands have been discovered. The advance of rapid and inexpensive sequencing has accelerated the technology in the last decade, and the inexpensive nature of building and screening a library has opened the opportunity for academia to progress the field. The future of this technology lies with expanding the scope of DNA compatible chemistry, and the development of drug-like hits.

Chapter 2 – Encoded Transformations

2.1 – Encoded Transformations Rational

The main aim of the project was to address some of the issues faced in the design and synthesis of DNA encoded libraries. The overall chemical diversity of a library is still limited by current methods of library synthesis and compatible chemistry. Although novel chemistry has been applied to DNA encoded systems, the resulting libraries are still limited and cover only a small portion of chemical space. This requires multiple libraries to be screened in parallel to access a larger pool of chemical matter,[83] something that can only be achieved once multiple different library syntheses have occurred. The use of innovative reaction types could be applied in parallel to novel library design and synthesis techniques to greater improve the diversity of individual libraries. One way to enable a greater degree of chemical diversity of DELs could be the use of a new paradigm, entitled "encoded transformations".

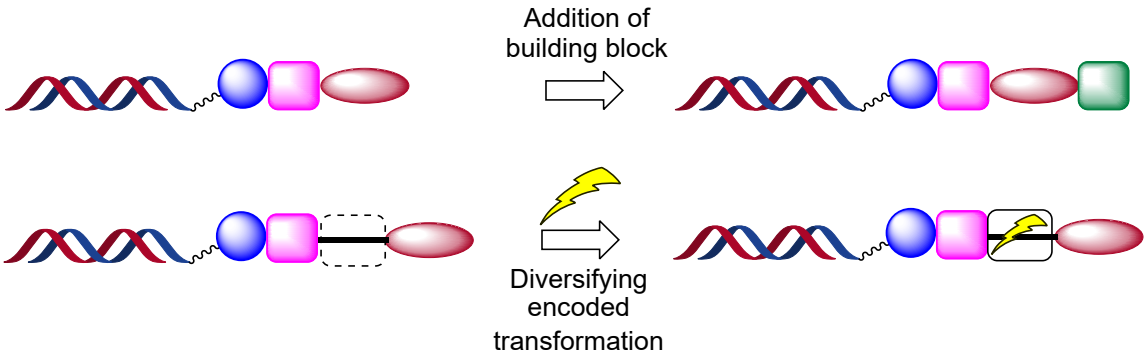


Figure 29 – DNA encoded library synthesis using traditional split and pool addition of building blocks compared to using encoded transformations to increase diversity.

In a traditional split and pool synthesised DNA encoded library, chemical diversity is improved through the addition of multiple building blocks at each stage of the synthesis. This means that the libraries are often overpopulated in certain areas of chemical space as the building blocks must all share a chemical similarity to undergo the same reaction, such as an amine and acid to produce an amide. Using this strategy, often molecules with large molecular weights are produced as building blocks are added in a linear fashion, meaning that the physiochemical property profile of the final libraries are often not ideal.

An encoded transformation strategy utilises a series of chemical reactions that can be performed on a single common core, chemically transforming that core and using a coding step to identify which reaction has taken place. This strategy would allow for greater diversity

in the library due to the variety of chemical entities that would be produced. Using this method, a series of encoded transformations would be performed on a common reactive functional core. Further chemical diversity could be achieved by varying the functionality of the groups flanking the central core. This strategy is expected to lead to libraries with lower molecular weight, greater core diversity, and improved lead-like properties over traditional split and pool paradigms.

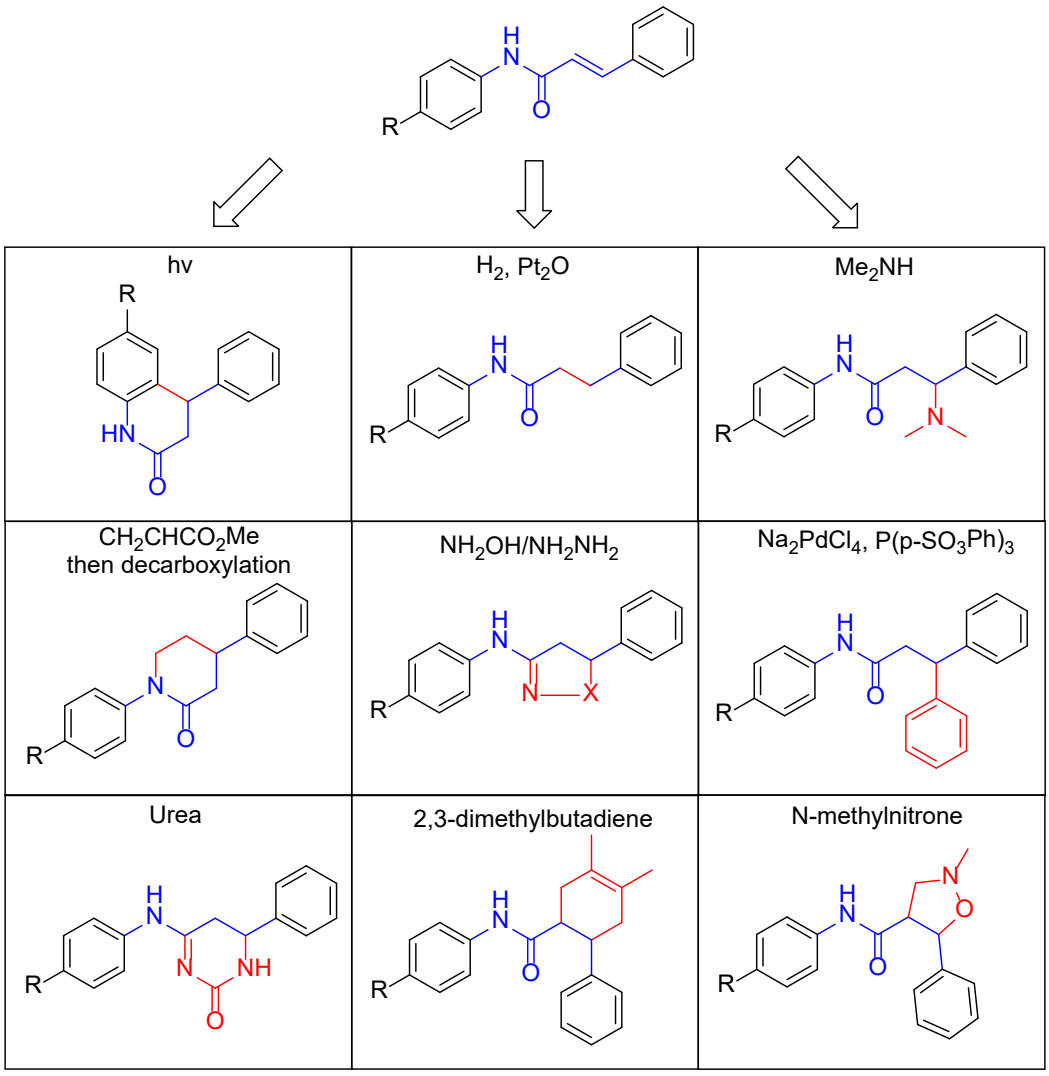


Figure 30 - Enamide reactive core and initial encoded transformations. Core structure shown in blue; chemical transformations in red.

Selection of the central core is an important factor for this new paradigm. A reactive core that can undergo a series of chemically robust reactions must be chosen as the template. A huge variety of reactions are available utilising a reactive group such as an alkene. Using an acylamide results in an electron deficient alkene that can be synthesised by amide coupling.

This enamide is also able to react with a large variety of functional groups [Figure 30] in many potentially DNA compatible reactions.

For reactions to be compatible with a DNA tag, the chemistry must be robust, proceed well in protic solvents, not require strongly acidic conditions, and not degrade the DNA tag.[90] Hence, only a limited number of reaction conditions can be utilised which exclude many common reagents and reactions. DNA tends to be soluble in aqueous solutions and is stable to highly basic reaction conditions. However, using large amounts of organic solvent can be detrimental due to rising solubility issues of DNA conjugates in organic solvents. Successful reactions tend to be run in a mixture of an aqueous buffered solution and a water-miscible organic solvent to aid dissolution and recovery of the DNA linked products and organic reactants.

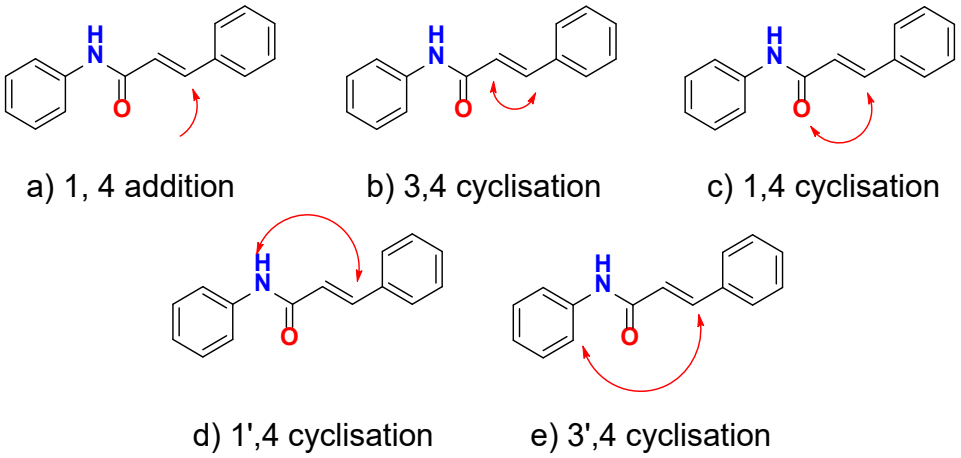


Figure 31 - Intended conceptual transformations of the enamide core.

The enamide core was chosen initially as there are a variety of transformations that exist in literature that could be adapted to be DNA compatible. The reactions that could be applied to this core show some degree of DNA compatibility in literature, such as the use of aqueous non-acidic conditions and a low to moderate temperature range. The products of these reactions also retain a high degree of structure diversity and drug-like properties. This is valuable as creating diversity using encoded transformations is vital to the project. These reactions can be separated into 5 main categories [Figure 31] each containing multiple different chemical transformations. These consist of 1,4 additions to maintain the acyclic core, along with cyclisation between the 4 position and either the 3, 1, 1’ or 3’ positions.

To begin this process, screening the reactions was performed on a non-DNA containing substrate. The reactions were developed in water using DNA compatible conditions and the products characterised off-DNA. With the reactions established on the model system, a DNA conjugated system was synthesised to further develop the reaction conditions. The end goal was to create a 3-step library, utilising 2 diversity points and the last reaction as an encoded transformation to generate a small proof of concept library that could be screened against target proteins.

2.2 – Enamide Synthesis and Transformations

Synthesis of the reactive core species begun off-DNA to easily characterise the reactions products and to trial which reactions would perform well in the study. It was also used as an initial test to see if the reactive core would be suitable for the study. An aromatic system, using an aniline and cinnamic acid, was used to introduce an enamide core. There are a huge number of aromatic building blocks available and therefore using an aromatic system would allow incorporation of a large degree of that diversity in the library. As the initial characterisation reactions must be run using DNA compatible conditions, solubility of the enamide would need to be considered. As the system is primarily aromatic and lipophilic in nature, a solubilising group would be effective at introducing some desired hydrophilic characteristics to the system. The system would also require a vector for DNA to be conjugated to the organic moiety, which should be inert and not interfere with any future reactions. For this reason, a polyethyleneglycol (PEG) linker was used as a solubilising group and to mimic the position of the DNA conjugate.

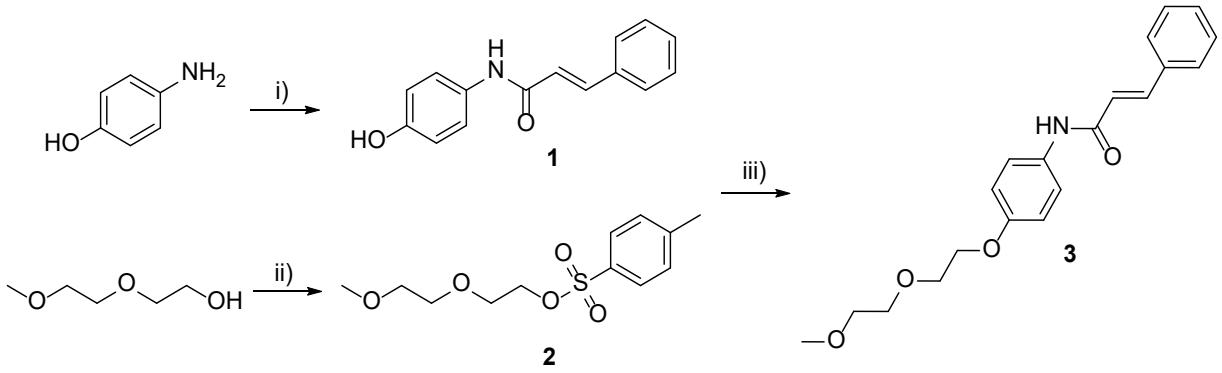


Figure 32 – PEG linked enamide core synthesis.; i) Cinnamic acid, EDC, triethylamine, DCM; ii) 4-toluenesulphonyl chloride, pyridine; iii) Potassium carbonate, acetonitrile, 80°C.

The reactive enamide core was introduced from 4-aminophenol and cinnamic acid by EDC mediated amide coupling. To introduce the solubilising PEG unit, diethylenegylcol methyl ether was reacted with *p*-toluenesulphonyl chloride to form the tosylated PEG species **2**, followed by alkylation with phenol **1** to produce the desired pegylated enamide **3**. For the enamide core to be successfully employed in an encoded transformation, it must undergo simple chemical modifications under mild conditions. The initial reactions chosen to test the core were the 1,4-cyclisations with urea, thiourea or hydroxylamine, and 1,4-addition of dimethylamine. When subjected to sodium hydroxide induced 1,4-cyclisations in THF and water, no product was formed using urea, thiourea or hydroxylamine, even at elevated temperatures of 80°C. Also, 1,4-addition reactions yielded no product, except when using neat dimethylamine, where only 11% conversion was seen at 80°C.

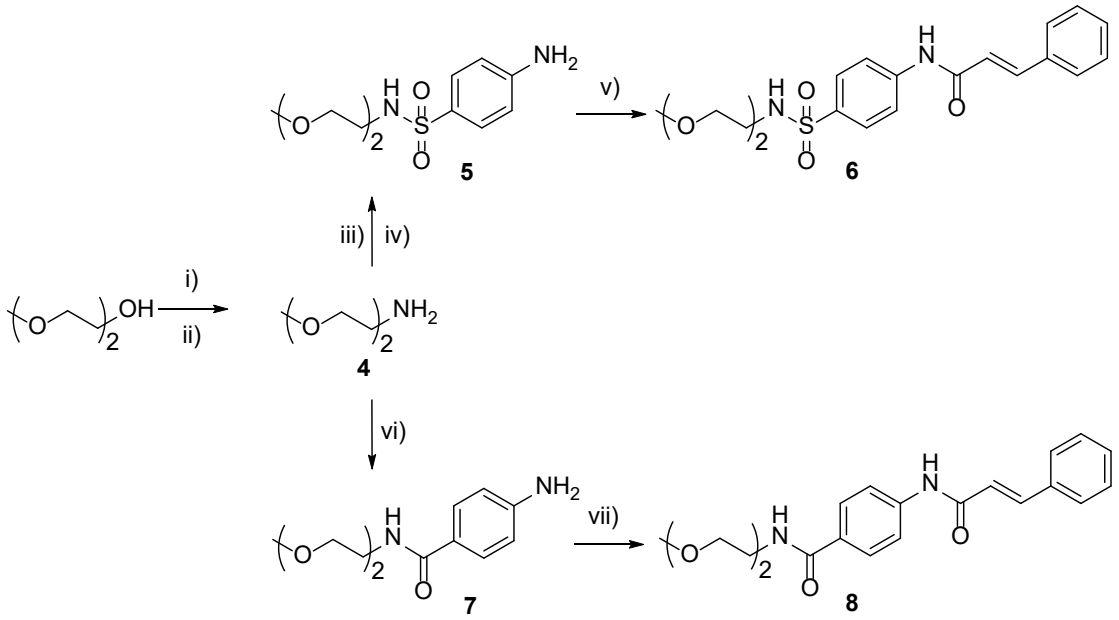


Figure 33 – Synthesis of amide and sulphonamide linked enamides; i) Phthalimide, DIAD, PPh_3, DCM; ii) Hydrazine hydrate, EtOH, 70°C; iii) 4-nitrobenzenesulphonyl chloride, TEA, DCM; iv) H_2, Pd/C, EtOH, H-Cube, 40°C; v) EDC, TEA, DCM; vi) 4-aminobenzoic acid, EDC, TEA, DCM; vii) Cinnamoyl chloride, DIPEA, THF.

The enamide used was thought to be unreactive due to the alkene not being sufficiently electron deficient, indicating that by using mild conditions the desired transformations would not be successful. The ether linkage at the 4-position also acts as an electron donating group. The introduction of an electron withdrawing PEG linker should have a positive effect on the reactivity by reducing the electron density on the enamide. A sulphonamide and amide linkage were proposed as electron withdrawing linkers, which are also easy to employ to DNA.

First synthesis of the amine capped PEG linker **4** was carried out using a Mitsunobu reaction between diethylenegylcol methyl ether and phthalimide. The amine was then deprotected using hydrazine to yield the amine capped PEG linker **4** for further reactions. At this point, two different routes were used to access the relevant sulphonamide and amide compounds. To access the sulphonamide, an acylation of the amino-PEG and 4-nitrobenzenesulphonyl chloride was carried out, followed by a palladium catalysed hydrogenation of the nitro afforded the aniline intermediate **5**. Finally, the same amide coupling conditions were used to introduce cinnamic acid to afford enamide compound **6**. The amide linked compound was synthesised by EDC coupling of compounds **4** and 4-aminobenzoic acid to yield compound **7**. Using EDC coupling conditions to introduce cinnamic acid failed, and so the acid was converted to the acid chloride using thionyl chloride, which was then reacted with compound **7** to yield the enamide product **8**.

The same transformations that were used initially on the ether linked enamide core **3** were then trialled on both cores. The Michael addition of dimethylamine was more successful, producing ~60% conversion by LCMS after 5 hours at 80°C. However, reactions with both urea and hydroxylamine were not successful on either core. The conclusion was that the enamide core was still unreactive even with the more electron deficient substrates. This core would be unsuccessful in providing enough reactions as a test substrate for DNA encoded transformations. Therefore a new more reactive core would need to be produced to gain the desired reactivity.

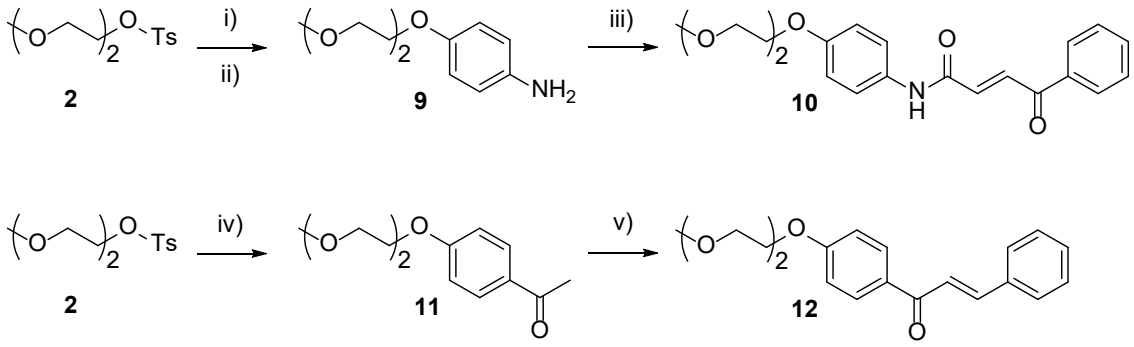


Figure 34 – Synthesis of keto-enamide 10 and chalcone 12 reactive intermediates. i) 4-nitrophenol, potassium carbonate, DMF, 100°C; ii) H_2, Pd/C, EtOH, H-Cube, 40°C; iii) trans-3-benzoylacrylic acid, HATU, DIPEA, DMF; iv) p-Acetophenol, potassium carbonate, acetonitrile, 80°C; v) benzaldehyde, 2M NaOH, EtOH.

2.3 - Ketoenamide Synthesis and Transformations

For these reactions to work it was understood that the core would need to be more reactive as a Michael acceptor, as well as for the carbonyl to be more accessible for condensation reactions. The amide was removed as it is not as accessible to condensation reactions as an aldehyde or ketone. A chalcone and a ketoenamide were chosen as a new reactive core due to the presence of both a better Michael acceptor and a condensable carbonyl. These two cores are also accessible by relatively simple chemistry and should allow for a multitude of diversity points to be added prior to transformations. Both were synthesised from the common tosylated PEG starting material **2** using alkylation of para-nitro or para-acetaldeyde phenols. The nitro group was then hydrogenated using a H-cube to the corresponding aniline **9**, which underwent a HATU mediated amide coupling with trans-3-benzoylacrylic acid to form ketoenamide **10**. To form the chalcone, an aldol condensation of ketone **11** with benzaldehyde, yielded PEGylated chalcone **12**.

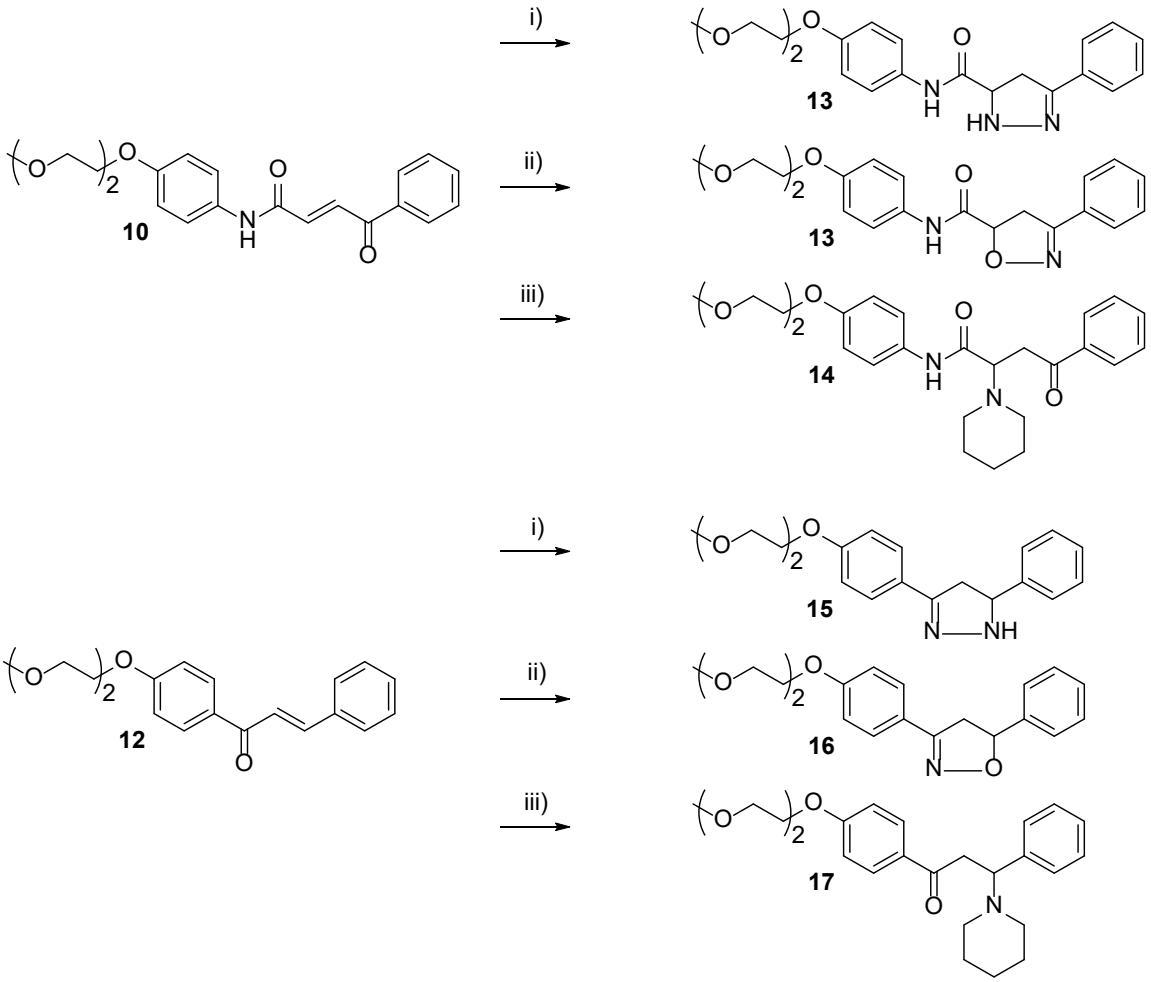


Figure 35 – Initial test reactions of both cores using hydrazine and hydroxylamine cyclisations, and piperidine Michael additions. i) hydrazine, H_2O, EtOH, 60°C; ii) hydroxylamine, NaOH, H_2O, EtOH, 60°C; iii) piperidine, H_2O, EtOH.

2.3.1 – Initial Transformations

Both chalcone and ketoenamide intermediates were trialled simultaneously with some simple transformations. Addition of hydrazine to form the pyrazoline was successful for both cores proceeding at 60°C and in 4 hours. Also, Michael addition with piperidine proceeded well. Using the ketoenamide core, reaction with hydroxylamine yielded the dihydroisoxazole product. However, when using chalcone intermediate **12**, reaction with hydroxylamine and a hydroxide base resulted in loss of the PEG linker. An S_NAr type reaction was thought to have occurred, promoted by the electron withdrawing group of the chalcone which result in the loss of the PEG linker. The reaction was then repeated using DMF and potassium carbonate and proceeded to form a majority of the dihydroisoxazolel product with no displacement of the PEG linker. This indicated that the chalcone was reacting with hydroxyl bases to displace

the linker. This core was therefore deprioritised since a large number of the selected transformations required hydroxide bases.

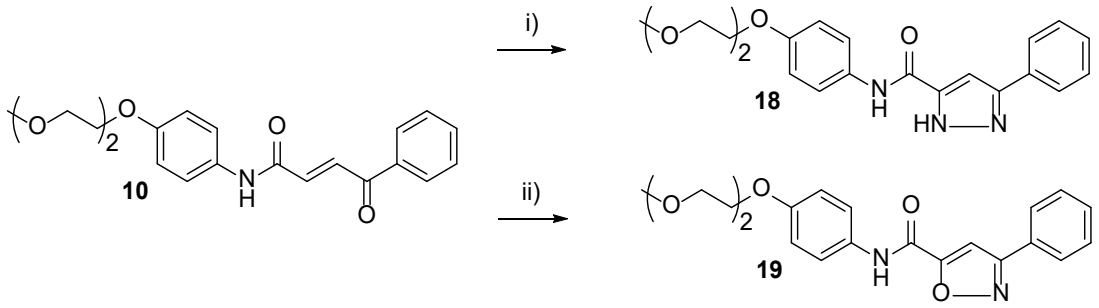


Figure 36 – Synthesis of pyrazol core. i) p-Toluenesulfonyl hydrazide, I_2, K_2CO_3, EtOH, H_2O, 65°C; ii) tosylhydroxylamine, EtOH, H_2O, 65°C.

2.3.2 – Isoxazole and Pyrazole

Access to unsaturated, aromatic heterocycles are also vital in this diversity orientated synthesis. Although using hydroxylamine or hydrazine gave access to the dihydro products, oxidation is required to synthesize pyrazoles and isoxazoles. Oxidising agents often destroy the integrity of DNA. Consequently, alternatives are required to synthesize these compounds. Synthesis of isoxazoles has been achieved using N-tosylhydroxylamine.[91] This system relies upon the release of the tosyl group after cyclisation with N-tosylhydroxylamine. This effectively oxidises the ring to yield an isoxazole. When used on the ketoenamide system, the reaction proceeded cleanly at 65°C even with the presence of water in the reaction mixture.

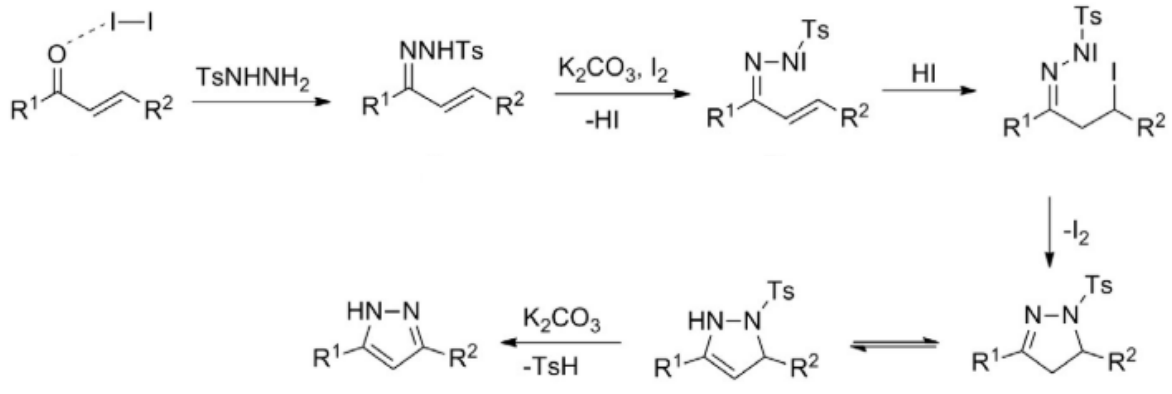


Figure 37 – Proposed mechanism for the iodine catalysed pyrazole formation.

Pyrazole synthesis was achieved in a similar manner. The use of phase transfer catalyst, tetrabutylammonium bromide, with sodium hydroxide was shown to synthesize pyrazoles in good yields.[92] Applying this technique to the ketoenamide system resulted in pyrazole

formation. However, the product was not produced cleanly and multiple side reactions seemed to have occurred. A second approach to the same transformation was proposed utilising catalytic quantities of iodine in ethanol. This reaction also used a similar tosylhydrazine starting material as the isoxazole reaction and pyrazoles were reported with modest yields using this technique.[93] These conditions successfully yielded the required pyrazole core in 2 hours at 75°C in aqueous medium. The reaction mechanism was proposed to form HI in situ which would catalyse the pyrazoline ring formation. This would then undergo an oxidation by loss of the tosyl group [Figure 37]. The reaction was repeated without iodine to see if it was promoting the reaction. It was found that the reaction progressed without iodine but was much slower and produced a profile containing many more impurities, including the dihydro pyrazoline compounds.

2.3.3 – Cycloaddition Reactions

One of the major benefits of using a ketoenamide core is that the alkene is highly electron deficient. This allows a wide range of cycloaddition reactions to occur on this reactive core. This could allow for some very interesting chemical structures to be prepared using a variety of simple cycloaddition reactions. Often, cycloadditions are also enhanced in reactivity by the use of water in the reaction[94, 95] which is required in DEL synthesis. This is thought to be due to the hydrophobic effect of water when mixed with organic compounds.[96] Water cannot hydrogen bond with organic parts of a molecule and when used in a reaction, water will often repel organic compounds leading to insolubility. The lowest energy state of many organic molecules in water is to have as little organic surface area in contact with the polar aqueous solution, forcing the organic moieties to form hydrophobic clusters. This has a beneficial effect on a cycloaddition reaction, as water will aid in driving the reactants together, thereby assisting reaction rates and reactivity. This makes these reactions ideal candidates for DNA encoded libraries as they are enhanced by the presence of water. Currently, this reaction type has been achieved on DNA[12, 38, 39, 65] and used to build libraries, making it a candidate for encoded transformations.

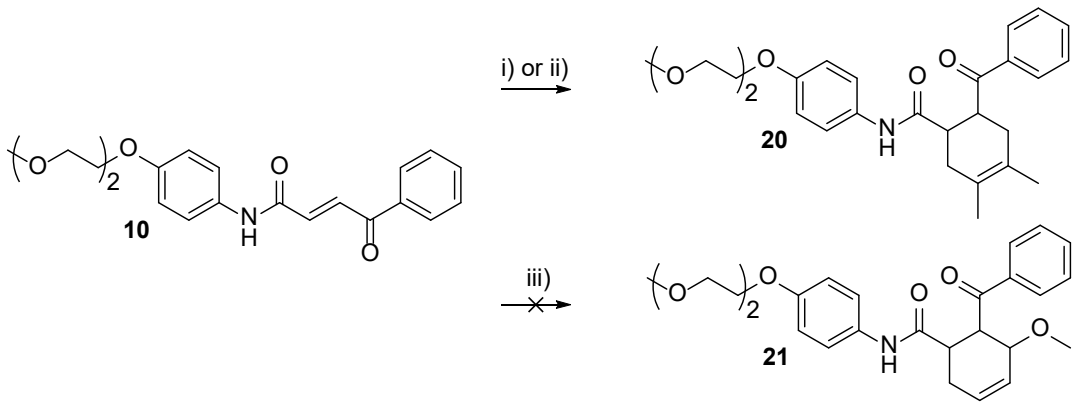


Figure 38 – Diels-Alder reactions. i) 2,3-Dimethyl butadiene, DMF, H_2O, 50°C; ii) 2,3-Dimethyl butadiene, 5% NOK in H_2O; iii) 1-Methoxy-1,3-butadiene, DMF, H_2O, 50°C.

The 4+2 cycloaddition is an extensively utilised reaction in organic chemistry, with the Diels-Alder reaction between a conjugated diene and a dienophile being perhaps the most common. The dienophile in this case is the electron deficient ketoenamide core which could react with a variety of dienes. Initial studies utilising 2,3-dimethyl butadiene showed that full conversion could be achieved in a DMF and water mixture at relatively low temperatures. However, when using 1-methoxy-1,3-butadiene, the reaction would stall at around 60% conversion. Micelles have been shown to increase reactivity of the Diels-Alder reaction by creating a hydrophobic core for the reaction to take place.[97] Utilising micellar technology and a designer surfactants NOK and TPGS-750-M created by the Lipshutz group,[98] the reaction was successful at room temperature. To gain knowledge on the stereochemistry of the product molecules, crystallographic data was required. However, after multiple attempts at growing crystals of this molecule, none were able to be formed, possibly due to the highly solubilising nature of the PEG chain. The 2,4-dinitrophenylhydrazine was prepared which gave extremely small crystals, but they were still unsuitable for crystallography. This means that although these reactions were successful, establishing the stereochemistry of the products was not possible. This reaction allows great scope into interesting 3-dimensional and sp^3 containing cores utilising simple cycloaddition reactions. Further scope for cycloadditions would be carried out on a DNA template to offer an overall reactant scope that could be used in library synthesis which would allow access to a multitude of different core structures.

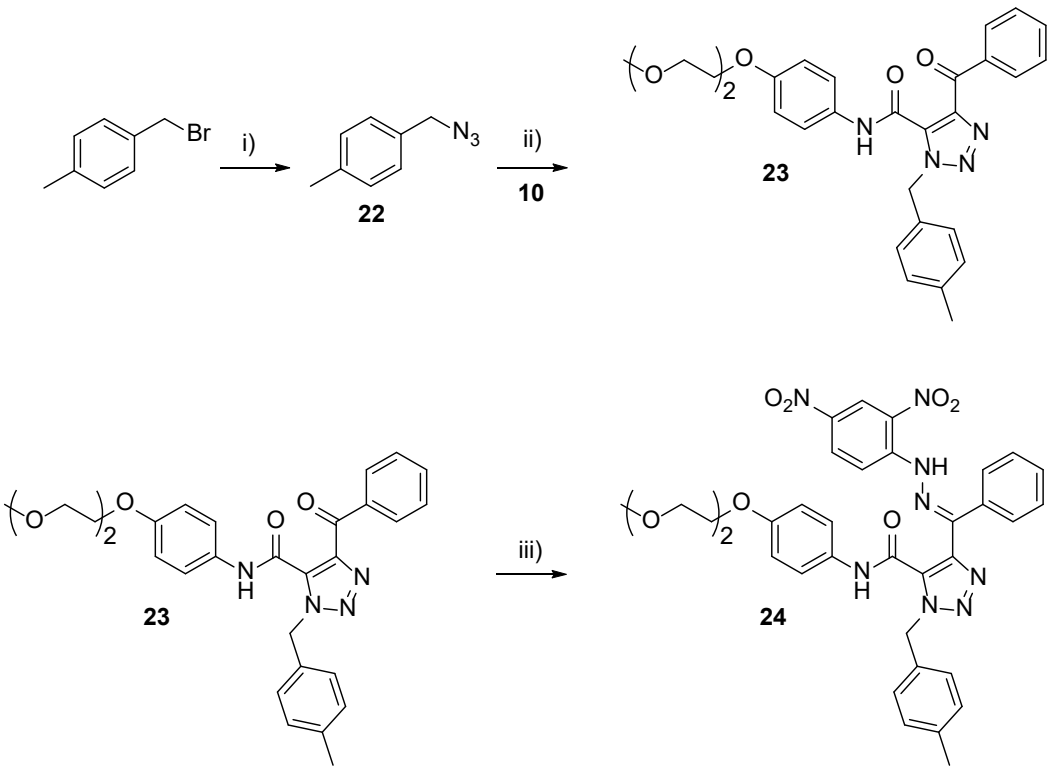


Figure 39 – Triazole **23** synthesis and synthesis of compound **24** to aid in X-ray crystal structure determination; i) Sodium azide, water, acetone, overnight at RT; ii) **10**, Potassium carbonate, water, dioxane, 60°C for 6 hours; iii) 2,4-dinitrophenylhydrazine, THF, overnight at RT.

An adaptation of the copperless "click" reaction was utilised analogous to a [3+2] cycloaddition reaction. This reaction had recently been developed using chalcones as the reactive core species[99] and would therefore be an ideal candidate for a ketoenamide transformation. This reaction was attractive due to the ability to increase diversity of the library by utilising a variety of azides in library synthesis. Azides are widely available, and alkyl azides are extremely easy to synthesize from alkyl halide moieties increasing the desirability of this reaction type. This reaction is also known to be compatible with aqueous solvents, for example using 1,4-Dioxane as a co-solvent, basic conditions, and a reaction temperature of 80 °C. All of which should make this reaction transferable to a DNA conjugated system. The electron deficient alkene in the core molecule could react with an azide using a concerted cyclisation reaction. This reaction differs from the click reaction as it does not require a catalyst such as copper and instead utilises the Michael accepting nature of the enone [Figure 40]. The mechanism is proposed to proceed via a stepwise 1,3-dipolar cycloaddition reaction. This produces a triaziline intermediate which undergoes a 1,3-hydride shift and subsequent oxidation to produce the desired triazole core.

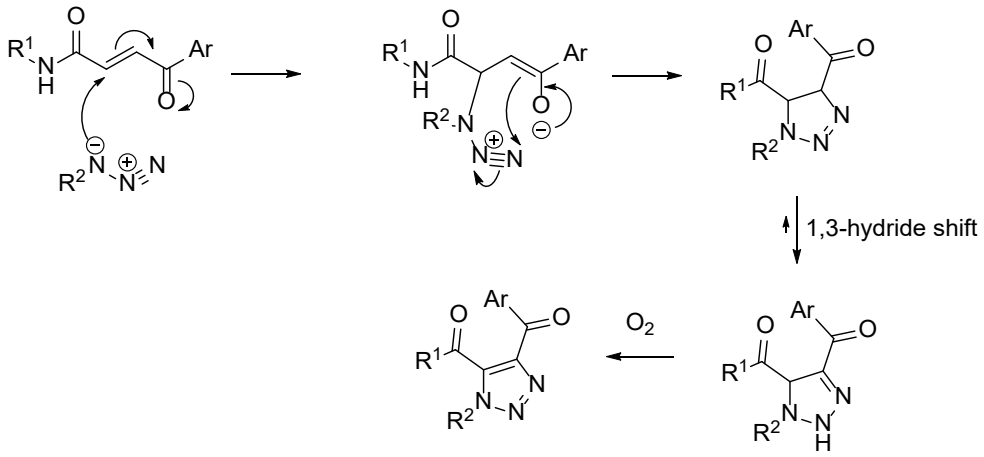


Figure 40 – Proposed reaction mechanism of azide with ketoenamide core.[99]

A 4-methylbenzyl azide was chosen as a test substrate for this reaction due to availability and simple synthesis of the starting material. To synthesize the azide, 4-methylbenzyl bromide was alkylated with sodium azide in a water and acetone solution, producing the corresponding benzyl azide in good yields [Figure 39]. Initial reaction screening utilising 9:1 water and 1,4-dioxane with 2 equivalents of potassium carbonate base were trialled on the substrate. It was discovered that these conditions led efficiently to the triazole. These conditions were extremely mild, using only potassium carbonate as a base, and a lower temperature was required than the literature suggested. This should permit this reaction as compatible with on-DNA synthesis, and hopefully little validation would be required. NMR analysis was inconclusive as to the orientation in which the azide had reacted with the alkene and NOE did not help identify the potential product synthesised.

The proposed reaction mechanism predicts that for a chalcone only one product would be obtained [Figure 40]. Only one product was observed in both HPLC and NMR traces from the chemistry carried out on the ketoenamide core. However, as it is a novel system it was desirable to show the absolute configuration of the product obtained in these reactions. It was proposed that the position α to the amide is the most electrophilic site due to the strongly electron withdrawing effect of the ketone. The suggested mechanism shows that the first step would be a 1,4-addition of the azide to the ketoenamide and as only one product is observed, the reaction was deemed to have a preference in the position the azide attacks. If this mechanism was correct, it would imply that the alkyl species attached to the azide would be adjacent to the amide in the product and crystallography could be used to prove that this was the expected isomer. Attempts at crystallising the initial triazole product were unsuccessful

in many differing conditions. This was likely due to the presence of the solubilising polyethylene glycol chain. However, a crystal structure was obtained from the corresponding 2,4-dinitrophenylhydrazone. The results of the crystal structure clearly show that the observed product is the expected triazole isomer [Figure 41].

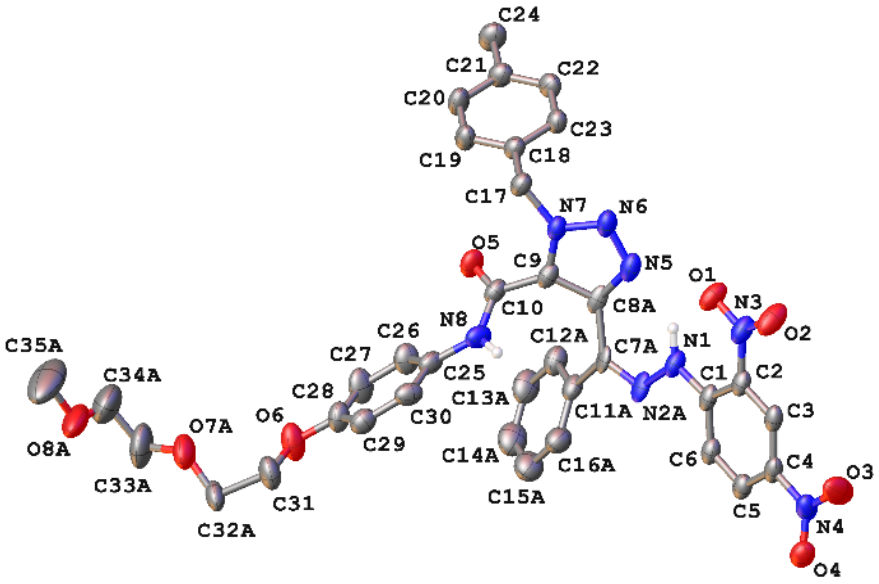


Figure 41 - Crystal structure of the 2,4-DNP derivative **24**.

Similar cyclisation reactions can also be carried out using N-methyl nitrones to form isoxazolidine cores,[100] and more recently they have been shown to work in the synthesis of DNA encoded libraries.[39] The reaction of a nitrone with alkenes, especially highly activated alkenes, are rapid by utilising mild conditions with no additives such as acids or base required. This suggests that this reaction should be compatible with DNA encoded technology. Like the above triazole synthesis, utilising this reaction can also increase chemical diversity by using different aldehydes and hydroxylamines in the synthesis of the nitrone.

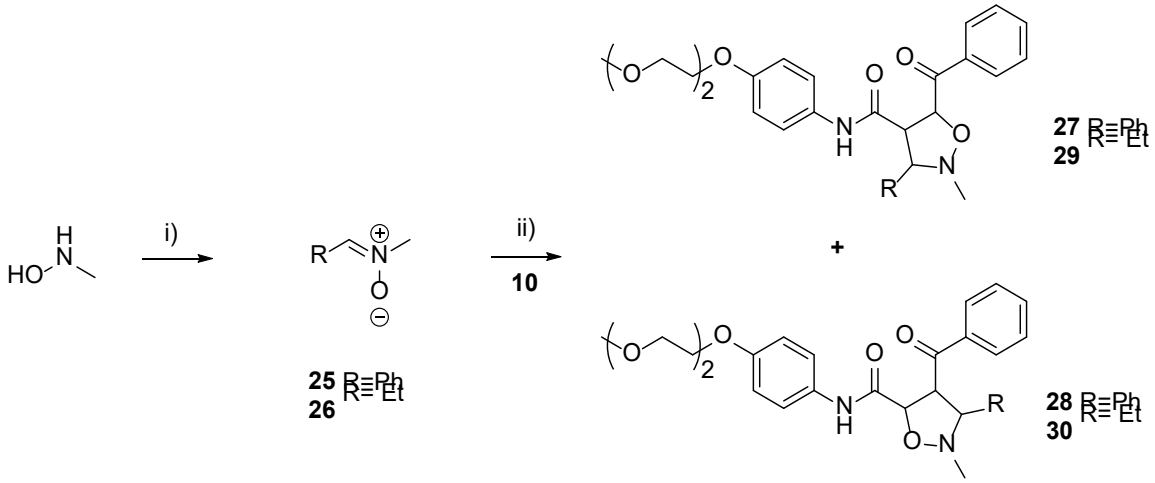


Figure 42 - Synthesis of the methylnitrone and isoxazolidine core; i) Acetaldehyde or benzaldehyde, EtOH, sodium acetate, 2 hours at RT; ii) Ketoenamide core, water, acetonitrile, overnight at 60°C or RT.

The benzylmethylnitrone was chosen for validation chemistry, but any simple aldehyde should work for this reaction. N-Methylhydroxylamine was reacted with benzaldehyde under basic conditions resulting in the pure methylnitrone after a simple filtration workup. It is worth noting that the product decomposes relatively quickly so needed to be used immediately. This could negatively affect the synthesis of a library. However, as many DNA compatible chemistries rely upon high concentrations of reactants, the rapid degradation was thought to be less of a concern as they react extremely quickly. The methylnitrone was then reacted with the ketoenamide core with acetonitrile and water as a solvent at 60°C overnight resulting in the formation of 2 products in a 1:1 mixture. This was not expected as, like the previous example, the 3 position of the ketoenamide is believed to be far more reactive, and so only 1 isomer should have formed. This could have occurred as the aromatic ring on the methylnitrone is positioned in proximity with the unsubstituted phenyl of the ketoenamide core. The aromatics may therefore repel as they come into proximity, forcing the nitrone away from the alkene. This would therefore disfavour the 1,3-addition of the oxygen. The steric hindrance of the aromatic ring must be so great that the 1,4-addition to the amide takes place as a competing reaction producing a 1:1 mixture of the two products. The propyl methylnitrone was synthesised to see what effect it would have on the stereochemistry of the reaction. The use of propyl methylnitrone significantly reduced the reaction temperature allowing the reaction to take place at room temperature. This is most likely due to the stereochemistry being more favoured to attack 1,4 into the ketone. Without the steric

hindrance of the aromatics the reaction can progress under milder conditions due to a drop in the energy required for 1,4-addition to take place. The product of this reaction showed a 19:1 reaction mixture in favour of the expected isoxazolidine. However, the NMR did show a mixture of diastereoisomers around the ethyl of the isoxazolidine which would need to be taken into consideration for any hit molecules that a library containing this compound produces. Due to the mild nature of this reaction, and the minimal amount of side products formed, it should be appropriate for use on a DNA scaffold.

Cycloaddition reactions have been found to be an extremely viable way to insert variety into the reactive ketoenamide core. The reactions shown here have all utilised extremely mild conditions, with few additives, under aqueous conditions to convert to the desired product. Using these reactions encompasses the added benefit of using a multitude of different easily accessible starting materials to increase chemical diversity of the entire library. Further chemical validation would be required for these reactions to be truly shown to work on a DNA template, but preliminary chemistry has shown to increase the likelihood of these reactions being tolerated.

2.3.4 – Aminoimidazole and Aminopyrazole 1,4 Cyclisations

The ketoenamide core is comparable to a chalcone in reactivity and scope of compatible chemistry. The main difference between the cores is that the ketoenamide seems to be more reactive towards 1,4-additions allowing for milder conditions to be used. As well as undergoing 1,4-addition, the ketone can undergo a sequential condensation resulting in a cyclised product. This reaction type performed previously when synthesising the 5-membered heterocyclic systems pyrazole and oxazole. Using a similar paradigm, aminoimidazole and aminopyrazole could potentially react using a 1,4-cyclisation. These reactions have the potential to add privileged heterocyclic structures to the library design which in turn add huge chemical diversity to the overall library.

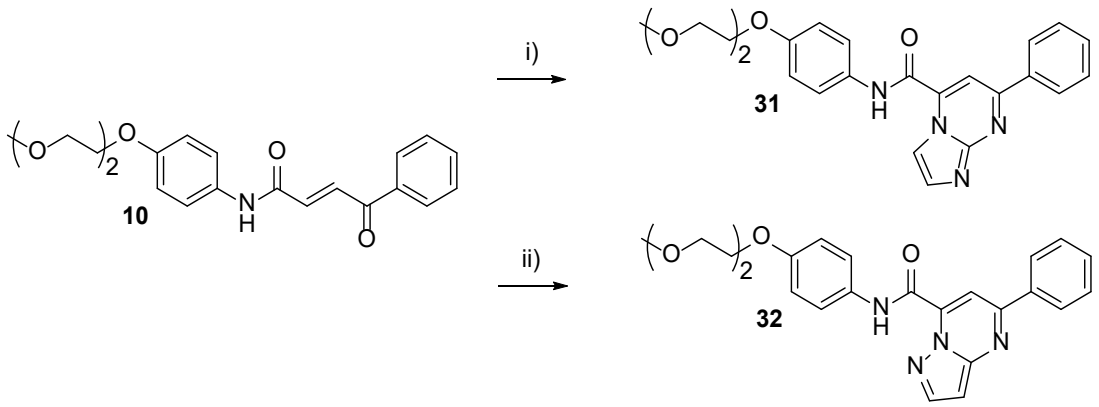


Figure 43 - Synthesis of the imidazopyrimidine and pyrazolopyrimidine cores; i) potassium carbonate, DMF, water, overnight at 60°C; ii) potassium hydroxide, water, THF, overnight at RT.

Each heterocyclic amine was enhanced using different conditions and developed to discover which solvents, base, temperature etc, gave the greatest yields and purity. Although this reaction could be either base or acid catalysed, base was chosen to enhance the compatibility of the reaction with DNA synthesis. Solvent and base were then trialled as well as varying temperatures. This resulted in the optimal conditions using potassium carbonate in DMF and water at 60°C overnight to synthesise the imidazopyrimidine, and potassium hydroxide in THF and water overnight at room temperature to synthesise the pyrazolopyrimidine.

These reactions proceeded well, yielding products in high purity with what should be DNA compatible chemistries. When characterising the imidazopyrimidine core by NMR it was found that the imidazole protons did show an appreciable coupling to each other, and COSY NMR showed little interaction between the protons. To fully characterise the product more information was required and a full analysis on a 700 MHz NMR was utilised. Proton NMR indicated that the imidazole protons appeared as tight doublets with coupling constants of 1.4 Hz, likely due to proximity of the three electron withdrawing nitrogen atoms. The COSY spectrum also showed that the 2 protons were coupling to each other. However, the NOE experiment did not show any interaction through space between NH 5 and the imidazole proton 14 [Figure 44 b)], but it did show an interaction between both protons 3 and 8. This was unexpected in the experiment as the proximity of 5 with 14 when the amide bond freely rotates around its axis should show an NOE interaction. A 2D ROESY experiment was used to promote the NH proton 5 which was overlaid with the proton NMR to show which protons are in proximity. The result of this clearly indicates that the amide proton is close to the

aromatic protons 3 and 8, but further from proton 14 of the imidazole. The data suggests the imidazole has reacted in the opposite configuration as expected, or that restricted rotation around the amide prevents the NH from coming into proximity with the imidazole ring. If the imidazole had reacted in the opposite conformation, a NOE would be expected between proton 14 and aromatic proton 11 which was not observed. This led to the conclusion that the amide bond must be restricted in rotation with a strong conformational preference.

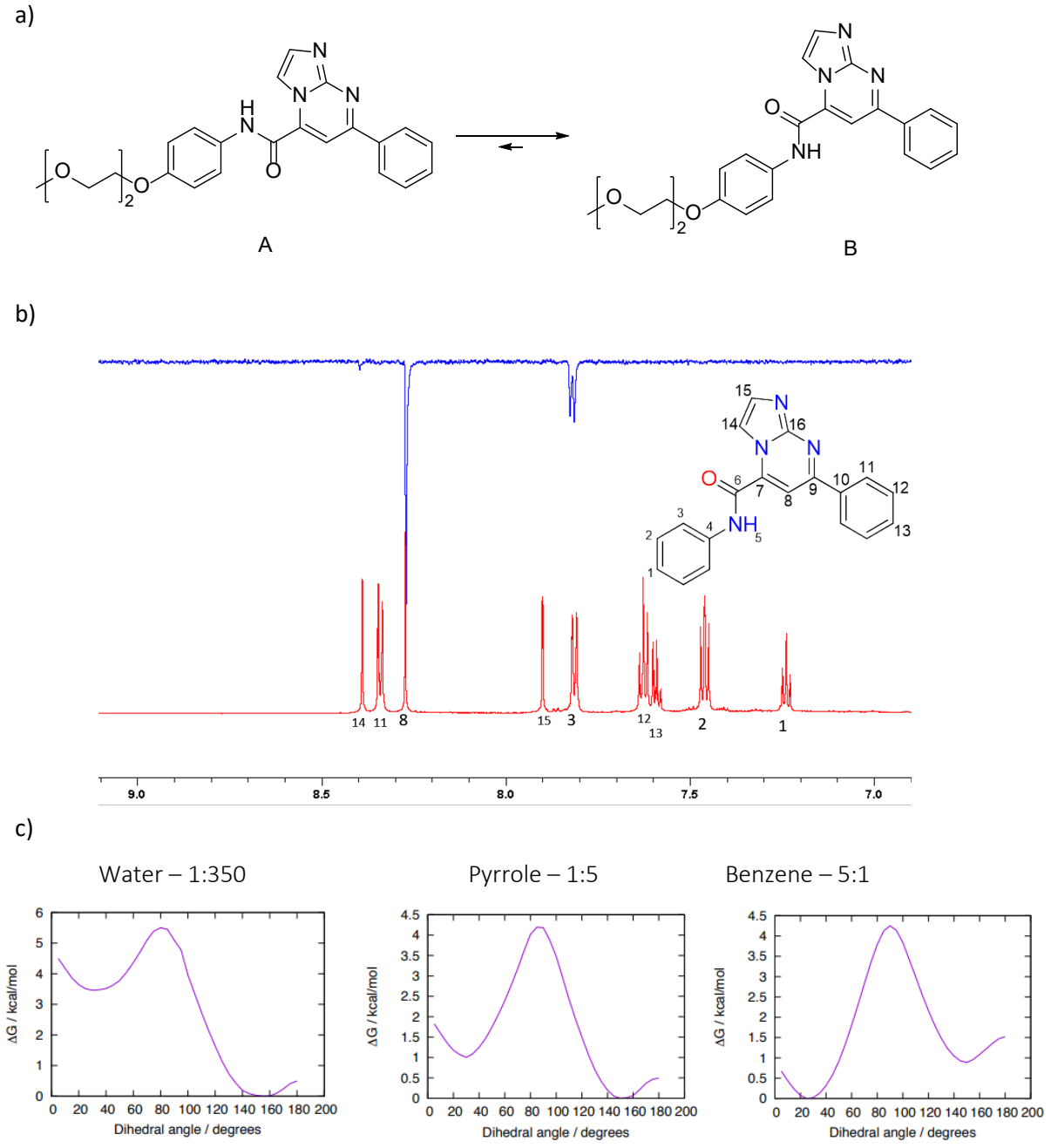


Figure 44 - a) Conformation of the amide bond. Shows that the bond has restricted rotation about its axis, leading to B being the prevalent confirmer; b) Blue - 2D ROESY experiment,

highlighting the NH proton; Red – 1H proton NMR; c) Computational analysis of the energy needed to rotate around the amide bond in water, pyrrole and benzene.

To further investigate the conformational preference, we sought to obtain a crystal structure of the compound. However, as the compound contained the solubilising PEG linker, this resulted in an oil and no crystals could be grown. Unlike previous examples there was no reactive handle to attach functionality to ease crystallisation. The compound was therefore resynthesized without the PEG linker and after many attempts crystals formed from a THF and petroleum ether mixture. The resulting crystal structure showed that the compound was the expected isomer and crystallized in the conformation predicted in the ROESY experiment [Figure 46]. There was a hydrogen bond present in the structure between the amide NH and one of the nitrogen lone pairs on the imidazole ring of a separate compound. This led to the belief that the restricted rotation was due to a hydrogen bonding interaction between the amide NH and the solvent in this example.

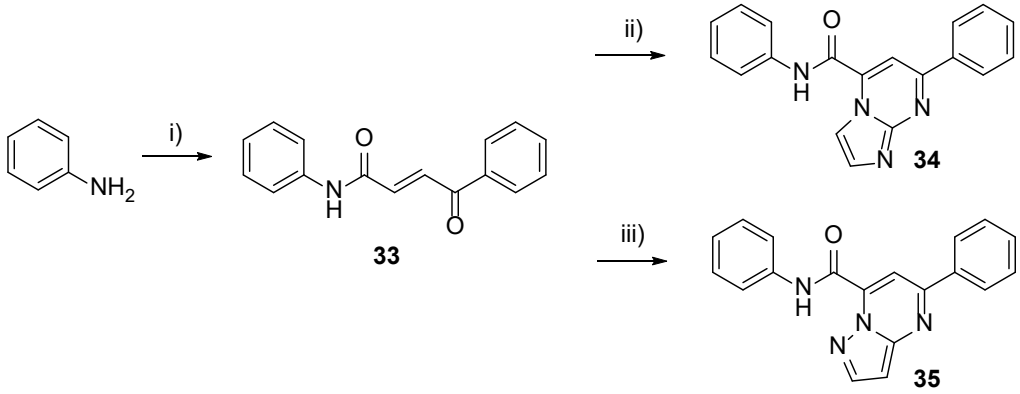


Figure 45 - Synthesis of non-PEGylated substrates; i) trans-3-benzoylacrylic acid, HATU, DIPEA, DMF; ii) aminoimidazole, potassium carbonate, DMF, water, overnight at 60°C; iii) aminopyrazole, potassium hydroxide, water, THF, overnight at RT.

To further explore the impact of this interaction a computation analysis was utilised to probe the impact of hydrogen bonding on restricted rotation. The degree of rotation and energy needed to adopt each conformation around the amide bond was analysed in different solvent conditions. It was discovered that in a solvent capable of forming hydrogen bonds, such as water or DMSO, the compound was 350 times more likely to rest with the amide bond as in the crystal structure with the NH situated away from the imidazole ring due to a hydrogen bonding interaction with the amide. This result implies that when the amide NH rotates towards the imidazole, a desolvation event would be required as the imidazole occupies the space of the hydrogen bonded solvent molecule. This desolvation penalty was shown to

increase the energy profile by nearly 4 kcal/mol of both conformations. Calculations suggested that the barrier to rotation was lower in apolar non-hydrogen bonding solvents. In pyrrole it was observed that the energy difference was only 5 times lower, but in benzene, it favoured the opposite configuration with a 5:1 preference. This interesting result could be used to great effect in an encoded transformation library as the core molecule should adopt this same restricted conformation in the library. This would permit greater overall diversity by utilising the 3D shape created in the library as well as utilising a multitude of different binding motifs on the core molecule.

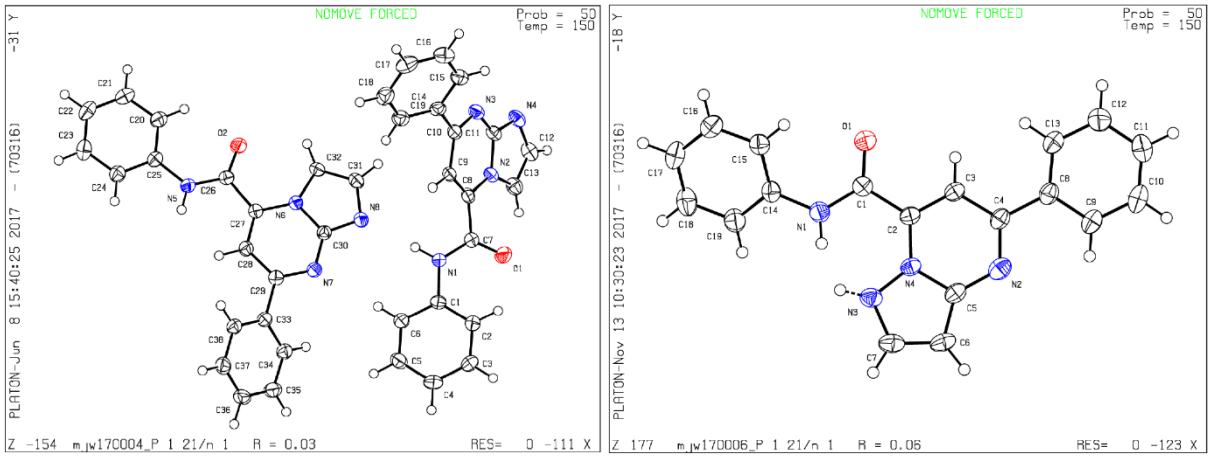


Figure 46 - Left - Crystal structure of imidazopyrimidine core; Right - Crystal structure of pyrazolopyrimidine core.

It was anticipated that by changing to a pyrazolopyrimidine core that the conformation may adopt a favourable interaction between the amide NH and the pyrazole. An intermolecular hydrogen bond could be formed between the amide NH and the pyrazole nitrogen that would hold the core in the opposite conformation to that of the imidazopyrimidine. Synthesis was carried out and characterisation of the product by NMR showed that the expected product had been produced. To further characterise the molecule an NOE experiment would not be as useful in this case due to the lack of close in space protons on the pyrazole moiety. A crystal structure of the compound would again be used to investigate the conformation of the product. As with the imidazopyrimidine, the PEGylated compound was an oil and so unable to form crystals, so it was synthesised without the PEG substituent. This resulted in a solid which was crystallised from THF and petroleum ether. The compound was shown to adopt the conformation in which the amide NH forms a hydrogen bond with the pyrazole nitrogen. The crystallography data suggests a proton that “shuttles” between the two nitrogen atoms, solidifying the secondary structure of the molecule. The proton was shown to lie directly

between the two nitrogen atoms, the presence of a bonding interaction with both nitrogen atoms. This is enough to hold the amide bond and the pyrazole in the opposite configuration to the imidazopyrimidine.

These results are encouraging for use as an encoded transformation in the proposed DNA encoded library. Both the imidazopyrimidine and pyrazolopyrimidine compounds adopt distinct three-dimensional shapes due to the restricted rotations around the amide bond, which in course provides not only chemical diversity but also diversity in shape and conformation in the resultant library. DNA encoded transformations would be used to create more diverse libraries using a core molecule that can undergo several transformations, producing not only structural diversity but diversity of shape. The two cores synthesised here would be highly beneficial in a DNA encoded library adding both structural diversity and modifying the overall shape of the compounds. One drawback of current encoded libraries is that only a small quantity of chemical space and diversity can be accessed through split and pool synthesis. This would mean multiple different libraries must be screened to gather hit molecules on a given target substrate. The ability to restrict the conformation of the core could have a large effect on the overall shape of the molecule. Having the matched pair of molecules that hold the conformation in both extremes is highly lucrative to change the overall shape of the library and increase the number of diverse hits the library could produce. These reactions will need to be further validated on DNA.

2.4 – DNA Conjugated Core Synthesis

To further identify which transformations could be utilised in the DNA encoded library, the trial reactions would need to be optimised in the presence of DNA. There are two ways in which to check the DNA compatibility. Usually the organic molecule is attached via a linker to a known DNA tag. The reactions are carried out and analysed via mass spectrometry and finally quantitative PCR. The validity of both the chemical reaction and the resulting DNA tag must be analysed before a chemical reaction can be applied across multiple substrates. In this case this would mean that the ketoenamide would need to be synthesised affixed to DNA tag, followed by a series of validation reactions analysed by mass spectrometry to follow the progress of the reactions.

The choice of linker is vital for this work. The linker originally used in DELs was double stranded, and the linker held both strands together covalently. The linker contained a free

amine that could be conjugated to an organic molecule. As it also used phosphate linkages, synthesis was carried out as a continuous modified DNA strand with the linker added as a modified base. A second type of linker has also been shown to successfully synthesise DELs as seen by Brunschweiger, in which an hexylamino linker is attached to the 5' end of the DNA strand.[41] This linker is also accessible from regular suppliers and can be synthesised in greater quantities. When used to synthesise a DNA encoded library, the first step would involve reacting the primary amine with the PEG linker followed by addition of the first building block. In these examples, a building block will not be added, but instead the ketoenamide reactive core was synthesised directly attached to the PEG linker. This also has the added benefit of extending the C6 linker using a PEG chain which allows for a greater distance between the DNA strand and the organic molecule and allows the resulting molecule to probe into deeper pockets of the protein. A truncated version of the Brunschweiger linker was chosen as a linker which created a 14mer double stranded DNA conjugated at the 5' end of one strand.

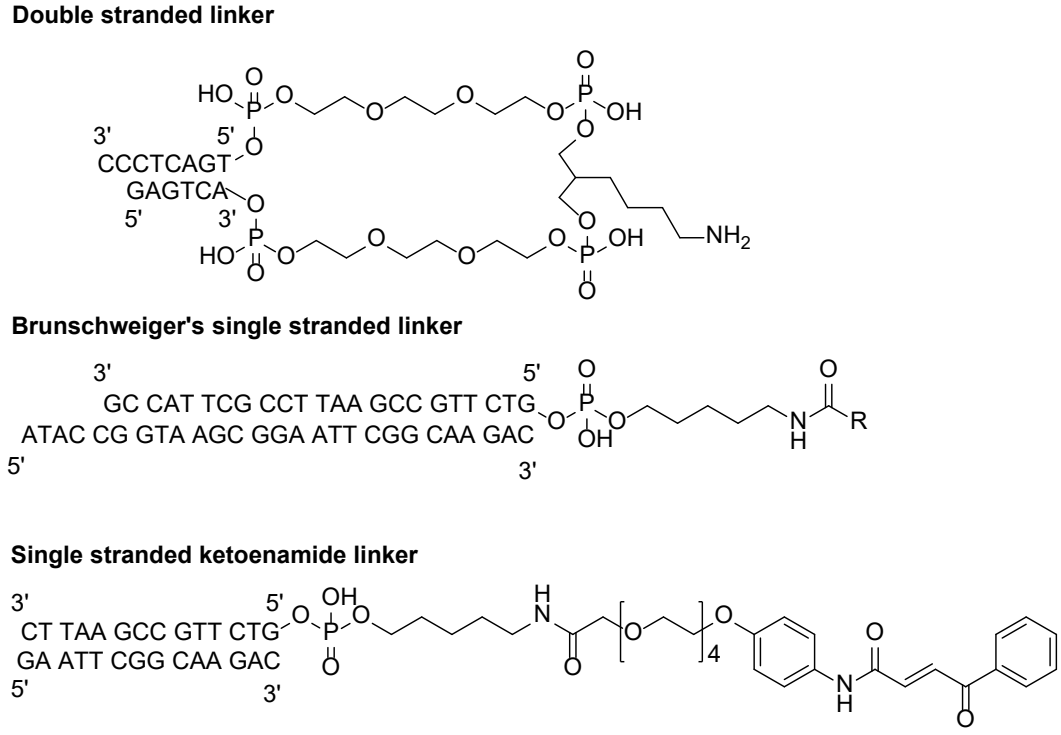


Figure 47 - The two main linkers used in DEL synthesis and modified single stranded one used here.

To synthesise the linker, nitrophenol was alkylated with tosyl terminated tetraethylene glycol under basic conditions. The terminal hydroxyl was then alkylated with *tert*-butyl bromoacetate to produce protected acid **38**. Hydrogenation with Pd/C using the H-cube,

followed by HATU coupling formed keto-enamide core compound **40**. This core was then deprotected with trifluoroacetic acid in DCM to leave the free carboxylic acid ready for coupling with DNA.

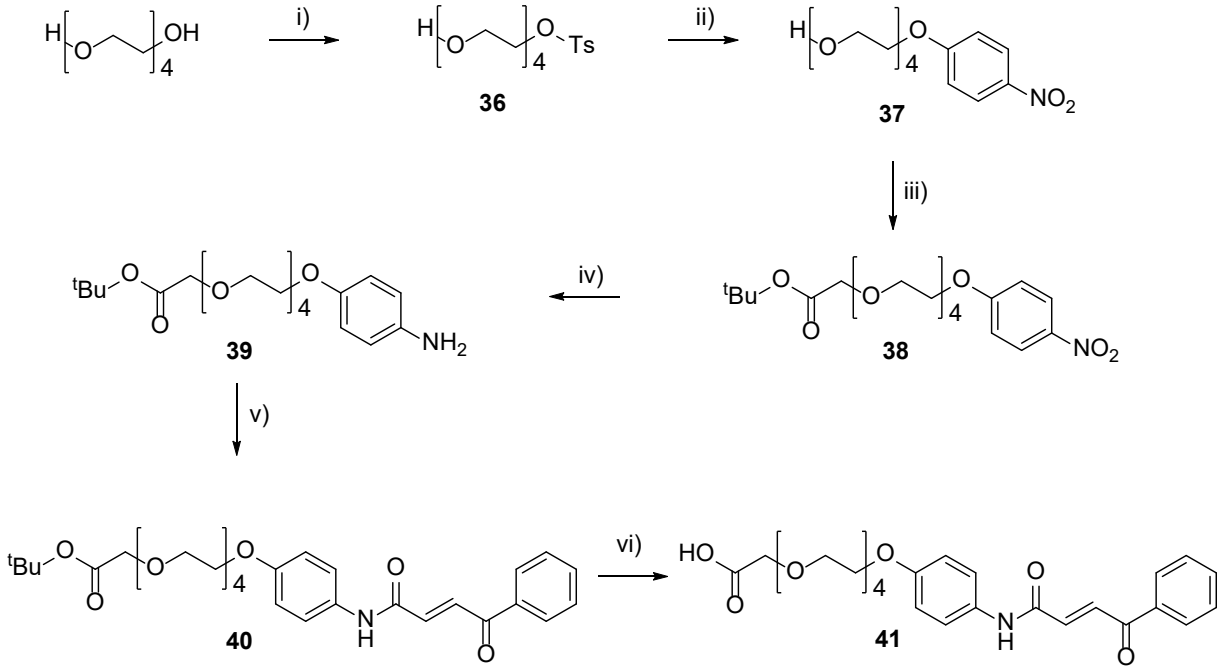


Figure 48 - Core synthesised for trial reactions attached to a DNA tag, the last steps are to be completed; i) 4-toluenesulphonyl chloride, chloroform, triethylamine, 3 hours at RT; ii) potassium carbonate, DMF, 6 hours at 80°C; iii) potassium *tert*-butoxide, *tert*-butanol, 6 hours at 45°C; iv) Hydrogen, Pd/C, EtOH, H-Cube, 40°C; v) trans-3-benzoylacrylic acid, HATU, DIPEA, DMF, overnight at RT; vi) TFA, DCM, 30 min at RT.

A solid supported approach was used to couple PEG linked ketoenamide **41** to the DNA strand. This approach has been previously reported in the literature[41] as a way to synthesise single stranded DNA conjugates utilising organic solvents. This has the added benefit of being capable of synthesising the DNA conjugate at higher scales than solution phase reactions. Utilising this strategy around 1 µmol of DNA can be synthesised in 1.5 ml microcentrifuge tubes. Solid support also simplifies the purification process after an organic reaction by using filtrations and washes with various solvents to solubilise and remove impurities present in the mixture. Once the coupling has taken place, the DNA conjugate is then cleaved from the solid support using aqueous ammonia and methylamine. In this case, as the ketoenamide is also a Michael acceptor, ammonia, methylamine, and water, underwent nucleophilic addition into the ketoenamide, yielding a mixture of products. This suggested that the design and synthesis of this linker had to be repeated to add the Michael acceptor after the DNA was cleaved from the solid support.

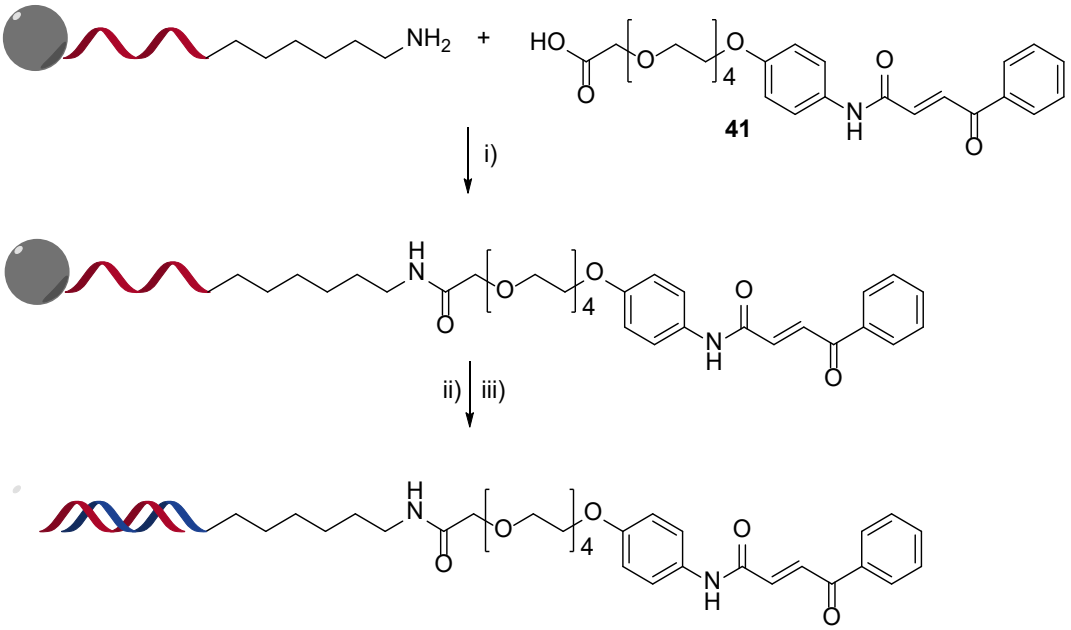


Figure 49 – Solid supported DNA conjugate synthesis; i) HATU, DMF, DIPEA; ii) Aqueous ammonia, aqueous methylamine; iii) 14mer single strand DNA, 40°C.

As the conjugate containing the Michael acceptor was problematic, the ketoenamide would need to be introduced after the DNA is cleaved from the solid support. To achieve this a protected aniline would require synthesis on solid support which, when treated with aqueous ammonia and methylamine, would produce the single stranded deprotected aniline. An Fmoc protected aniline was therefore proposed as the linker conjugate as this protecting group is cleaved under the same conditions as the solid support. Synthesis of the DNA linker was carried out in a similar way to previously but coupling of trans-3-benzoylacrylic acid was carried out after the DNA had been cleaved from the solid support.

The linker was prepared starting with a tosylation of tetraethylene glycol using *p*-toluenesulphonyl chloride in chloroform. The leaving group on the tetraethylene glycol chain was displaced with 4-nitrophenol in the following reaction. The free alcohol of the **37** was then alkylated with *tert*-butyl bromoacetate using tert-butoxide as a base in tert-butanol to yield tert-butyl ester **38** in modest yields, followed by hydrogenation of the nitro group. The ester was then subjected to acid deprotection using TFA to give carboxylic acid **42** in a quantitative yield. The final step was to Fmoc protect the aniline using 9-fluorenylmethoxycarbonyl chloride under slightly acidic conditions in dioxane. This yielded linker **43** as an Fmoc protected aniline and a free carboxylic acid to be directly coupled onto the modified DNA strand.

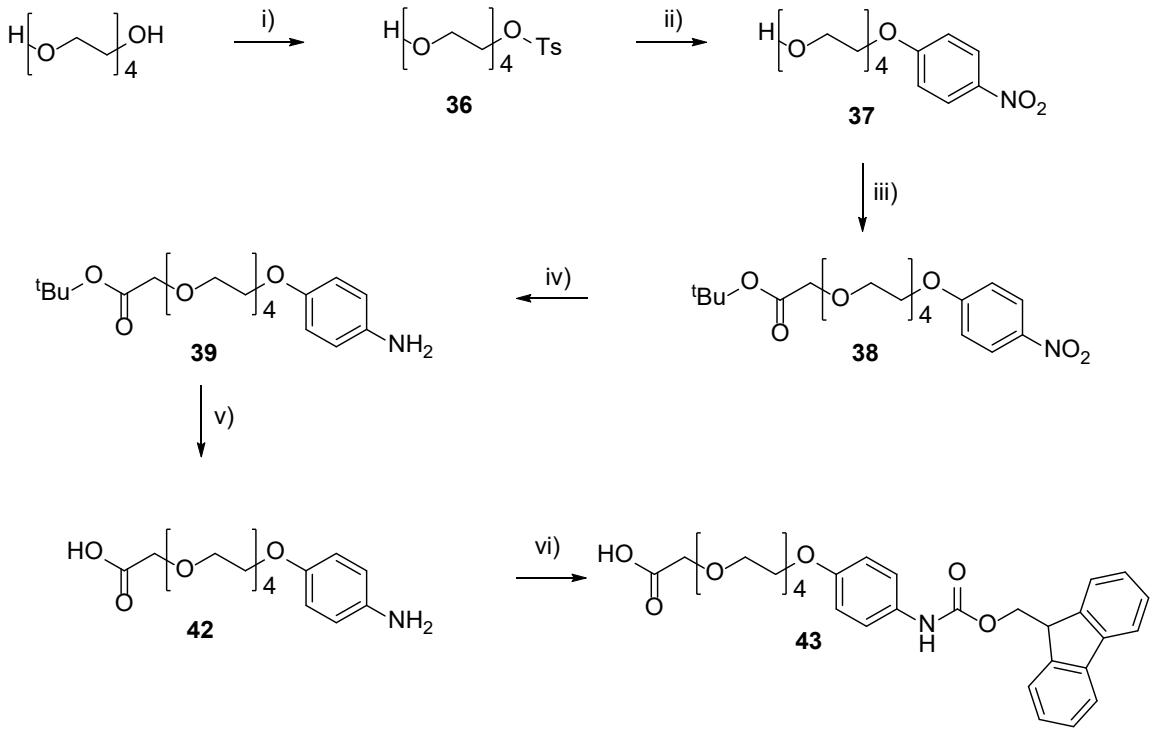


Figure 50 - Synthesis of the DNA linker compound 43; i) *p*-toluenesulphonyl chloride, chloroform, TEA, overnight at RT; ii) potassium carbonate, 4-nitrophenol, DMF, overnight at 60°C; iii) tert-butyl bromoacetate, *t*-BuOK, *t*-BuOH, overnight at 50°C; iv) Hydrogen, Pd/C, EtOH, MeOH, H-Cube, 40°C; v) TFA, DCM, 2 hours at RT; vi) 9-Fluorenylmethoxycarbonyl chloride, AcOH, water, dioxane, overnight at RT.

A HATU mediated coupling was used with solid supported DNA and compound **43** to attach the organic conjugate to the DNA strand. The Fmoc was then deprotected and the solid support cleaved simultaneously using aqueous ammonia and methylamine. Once complete, the single strand DNA was purified by preparative HPLC (using 200 mM HFIP and 20 mM TEA buffer and methanol). This removed any shortmers that were present from the solid supported synthesis as well as any starting material that was unreacted. Once purified the single stranded DNA **44** was subjected to a solution phase amide coupling using EDC and HOAT[32] and the product was further HPLC purified prior to the reverse 14mer DNA strand addition to produce double stranded DNA conjugate **45**.

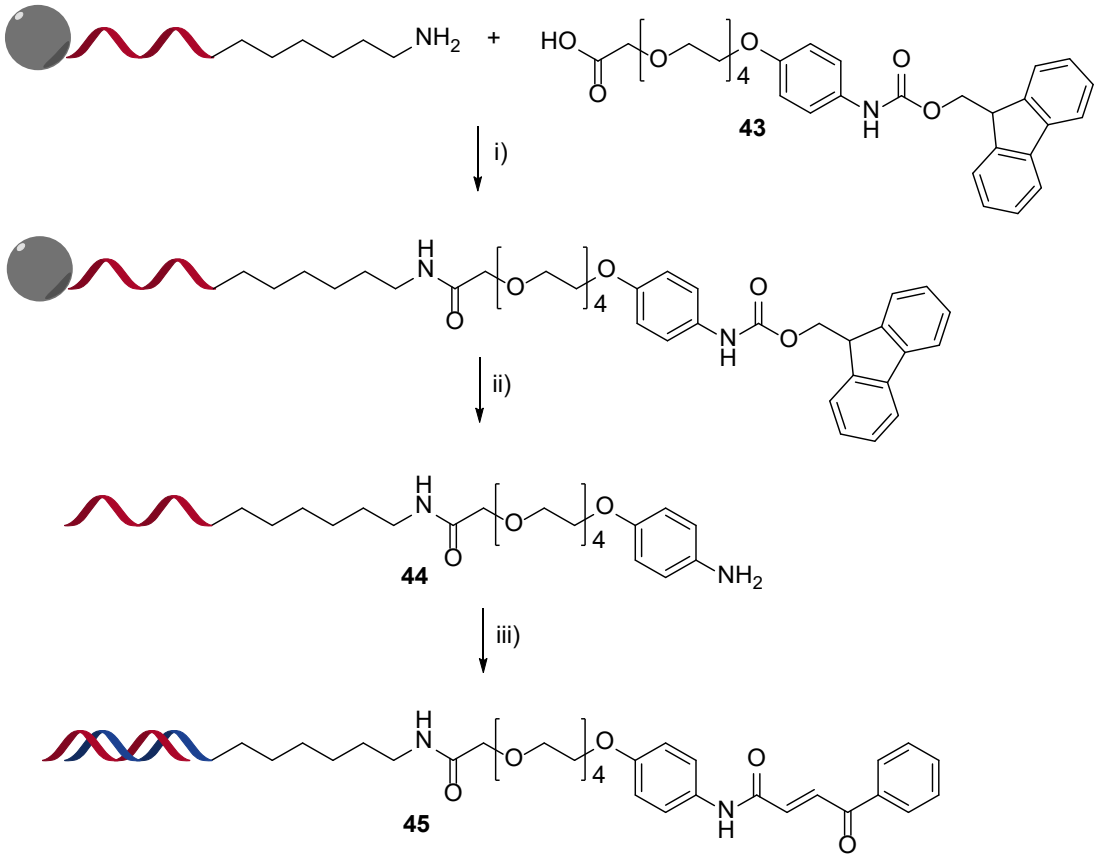


Figure 51 - Solid supported DNA conjugate 45 synthesis; i) HATU, DMF, DIPEA; ii) Aqueous ammonia, aqueous methylamine; iii) HATU, DIPEA, DMSO, water.

DNA conjugate **45** was dissolved to make a solution of 1 mM DNA and divided into 10 nmol reaction vessels. Proof of concept chemistry will need to be carried out for each transformation to see which, if any, are successfully employed with the DNA conjugate. Trial reactions were tested with hydrazine and hydroxylamine cyclisations, Diels-Alder with 2,3-dimethylbuta-1,3-diene, and aminoimidazole cyclisations. Unfortunately, none of these reactions were successful and showed no conversion to the desired product. It was thought that the difference between solution phase and on-DNA chemistry was vast and that in future, reactions should be identified and verified on-DNA rather than using conditions off-DNA first. This also would make it easier to screen reaction conditions as DNA reactions can be easily performed in 96-well plates and analysed quickly using mass spectrometry.

2.5 - Conclusions

The design and synthesis of DNA encoded libraries relies upon high-yielding, high-fidelity reactions resulting in chemically diverse libraries of compounds that can be screened against a potential biologically active target. The negatives of this approach currently include limited

available chemistry, preserving DNA fidelity, building block compatibility, and chemical diversity of the library. Utilising a new paradigm of encoded transformations could improve chemical diversity of potential libraries as well as including a variety of novel chemical transformations that could be used in the synthesis of other chemical libraries.

In this process a variety of successful chemical transformations have been examined on a common reactive core. The initial enamide core was modified to increase reactivity to improve the number of transformations that were successfully employed. Over the course of trialling chemical reactions, a variety of successful transformation were achieved using the more reactive ketoenamide core. The overall library produced from encoded transformations would contain a large degree of diversity, with core molecules with considerable three-dimensional space, sp3 character, and drug-like molecules. The larger the degree of chemical diversity in a library, increases the chances of acquiring a hit compound from a screen. Therefore, this approach could lead to pivotable changes in encoded library design.

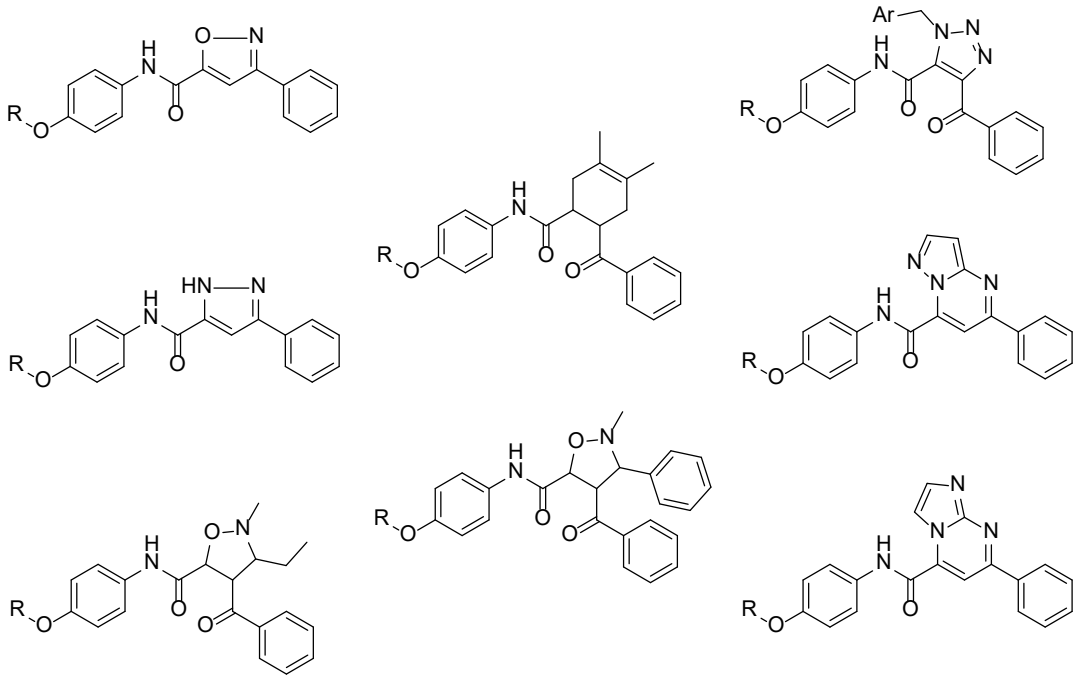


Figure 52 - Successful transformations applied to the off-DNA ketoenamide core.

Initial studies into chemical transformations on a DNA conjugate were not as successful as initially envisioned. However, with further on-DNA development these transformations are likely to be effective as development has taken place utilising DNA compatible chemistries. Synthesis of the DNA conjugate was successfully carried out on a large scale which would also allow for multiple reactions to be validated on this core. This synthesis allowed for initial

chemical and analytical techniques to be developed on DNA. This includes carrying out solid supported and solution phase DNA synthesis, purifications by both preparative HPLC and ethanol precipitations, analysis of DNA conjugates by mass spectrometry, and analysis of DNA concentrations using UV spectrometry. These techniques were all developed whilst carrying out initial DNA synthesis and trial transformations.

Chapter 3 – Micellar Technology

3.1 – Aqueous Catalysis

The need for "greener" synthetic methods has increased over the years. The reducing petroleum feedstocks will have a huge impact on the planet including chemistry. Many common solvents and reagents are derived from petroleum feedstocks which will eventually run out. Solvents account for about 80% of the total mass handled in chemical reactions.[101] Most of this solvent is not recycled and is incinerated instead to recover heat.[102] Organic solvents such as DMF and THF are still used extensively in organic chemistry due to their solubilising potential and polarity, whereas "green" polar solvents are not often considered as alternatives. This has meant that other methods utilising solvent-free reactions[103] and mechanochemistry[104] have been developed as green alternatives, but these conditions cannot be widely utilised.

Water accounts for nearly 71% of the Earth's surface area. It is in plentiful supply and is deemed the greenest of solvents having near to zero environmental impact. Water has also been the solvent of choice in nature for millions of years where a huge variety of biological and chemical reactions occur no matter the solubility of the substrates. Water is also economical, non-toxic, non-flammable, produces next to no greenhouse emissions, doesn't require synthesis, high polarity, large heat capacity, tuneable acidity, high boiling point, includes both hydrogen donor and acceptors, and an E-factor close to zero.[105] Yet it is often overlooked as a solvent. This is because water is often too polar, meaning that many organic substrates are insoluble and the strong hydrogen bonding gives rise to the hydrophobic effect [Figure 53] where water separates from apolar species rather than provide a solvation effect.

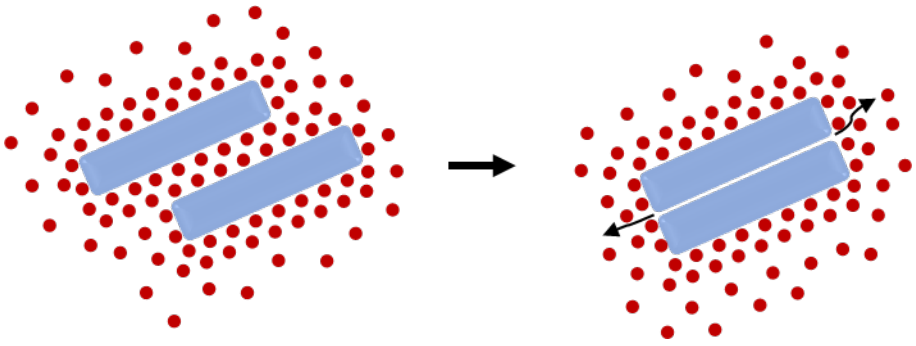


Figure 53 - The entropy driven hydrophobic effect on 2 lipophilic substrates (blue) being repelled from the aqueous phase (red) into proximity.

The hydrophobic effect is the exclusion of non-polar substrates from an aqueous solution. As this effect is driven primarily by entropy, water at the interface between 2 substrates is favourably returned to the bulk, compelling the substrates together [Figure 53]. Therefore, water as a solvent can be both attractive and repulsive dependent on the substrate in question. This means that hydrophobic substrates tend to agglomerate in water leading to the apparent reduction in solubility. However, this is not always a negative impact. In a Diels-Alder reaction,[106] water has been used to speed up the reaction process as it forces the substrates to be in closer proximity.[94] As a result the reaction rates can be increased by hundreds of times in aqueous media.

There are apparent concerns when using water as a green solvent. The requirement to extract the compound from the reaction mixture often utilises an organic solvent, and the purification often uses a larger volume of solvent than the reaction. This also does not account for further purification methods such as column chromatography. The resultant aqueous solution is also an organics contaminated stream and will need to be purified before being disposed of.[107] However, in catalysis the aqueous phase often contains the catalyst post work up which can mean that the reaction media can be reused.

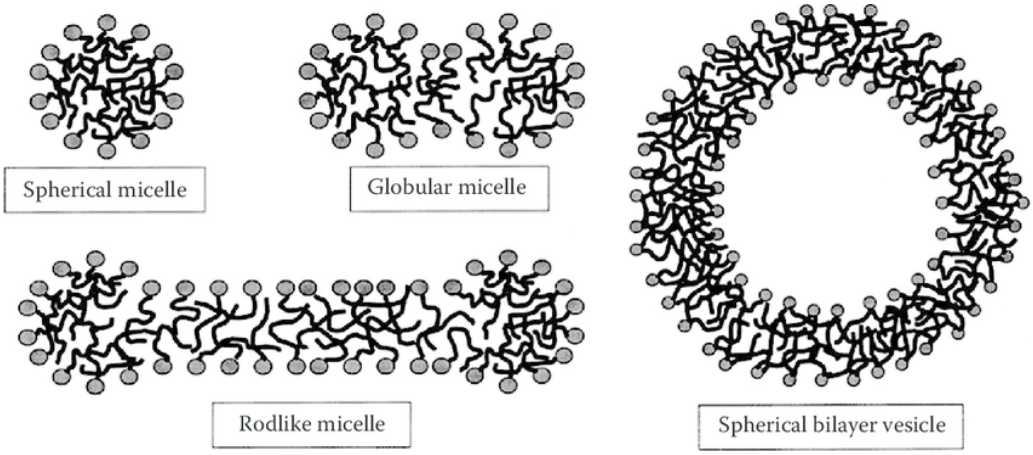


Figure 54 – Different shaped micelles and vesicles that can be formed in solution.

In nature, enzymes are often used to catalyse a reaction by providing a more lipophilic site in the centre of a water-soluble protein. These also contain sites where weak intermolecular interactions take place to improve substrate and product selectivity. An enzyme is therefore an ideal system to mimic in development of an artificial aqueously soluble reaction. A straightforward approach would be to mimic the features of enzymes, such as exploiting self-assembling units that are soluble in aqueous media and can surround a reactive catalyst. This

would form a nanoparticle that could accommodate the substrates, accelerate the reaction, and impart selectivity on the product. Small supramolecular macrocycles utilising glucose units, crown ethers and cyclodextrins can form hydrophobic cores for a substrate or catalyst to interact, whilst retaining a hydrophilic exterior allowing solubilisation.[108] Nanometre sized self-assembled substrates such as nanotubes, metal ligand capsules, vesicles and micelles can also be used. The objective of these systems is to mimic the environment of enzymatic reactions.[109]

3.2 – Properties of Micelles

Micelles are large nanoscale self-assembling structures made up of aggregates of over 100 monomers. These monomers rapidly exchange between aggregates meaning that the average lifetime of a micelle is only 0.001 seconds.[110] Each of these monomers is a surfactant molecule possessing both hydrophilic and hydrophobic regions. By utilising the hydrophobic effect these surfactants form different structures in aqueous media [Figure 54]. The hydrophobic effect forces the hydrophobic parts of the surfactant together whilst simultaneously forming positive interactions with the hydrophilic parts. In solution with low concentrations of surfactant, the molecules will sit at the water/air interface with the hydrophobic part away from the aqueous phase [Figure 55]. As the concentration of the surfactant rises, the surface tension decreases as more surfactant molecules arrange at the interface. Eventually the critical micelle concentration (CMC) is met where the surface of the liquid is saturated with surfactant. At this point the molecules start to arrange themselves in the lowest energy state, with the hydrophobic regions together in the centre and the polar hydrophilic parts facing into the aqueous medium.

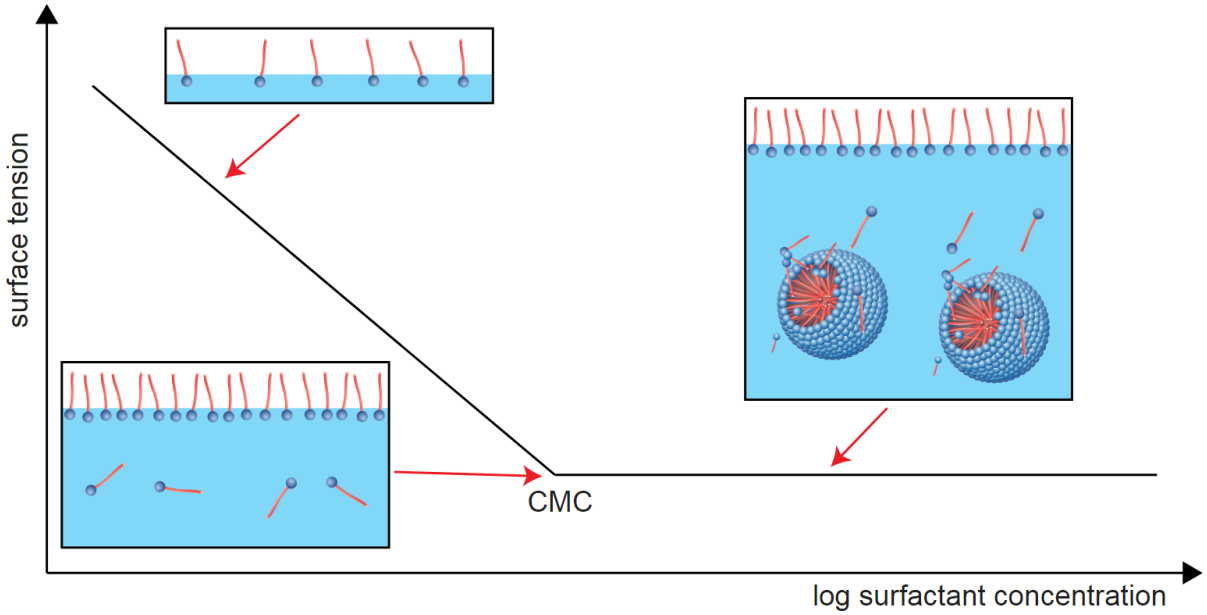


Figure 55 - Surface tension as a function of surfactant concentration.

Once the CMC has been met, micelles form in solution. The concentration of surfactant is vital in determining which structure predominates [Figure 54]. Initially spherical micelles are formed which on increasing concentration become ellipsoidal, rod-like and finally form spherical bilayer vesicles.[111] This forms a hydrophobic pocket at the centre of the aggregate where organic molecules and catalysts can react. Micelles can therefore be considered as a type of "nanoreactor" rather than just a solubilising detergent. Micelles can have a huge impact on the solubility of organic molecules in an aqueous media. The hydrophobic effect will help to "push" the organic substrates into the micelle, where solubilisation is enabled by the hydrophobic part of the surfactant molecules. This also has the effect of an enhanced local concentration and reactivity in the centre of the micelle aggregate. Micelles in theory should be beneficial for catalytic reactions where the substrates can be in high concentration around the catalytic site. However, there are some differences when working with micellar media compared to organic solvents. The size of the micelles matters greatly, too small and substrates cannot enter, too large and they form heterogeneous mixtures [Figure 56]. All ingredients in water are in constant exchange, including substrates, catalysts, and even the micelles themselves, allowing the micelles to rapidly exchange with other micelles and water.[112] There is also competition for solubilisation of the substrates in the micelle, which can depend on the lipophilicity of the substrates. The net concentration of the micelle is far greater than a normal organic reaction.

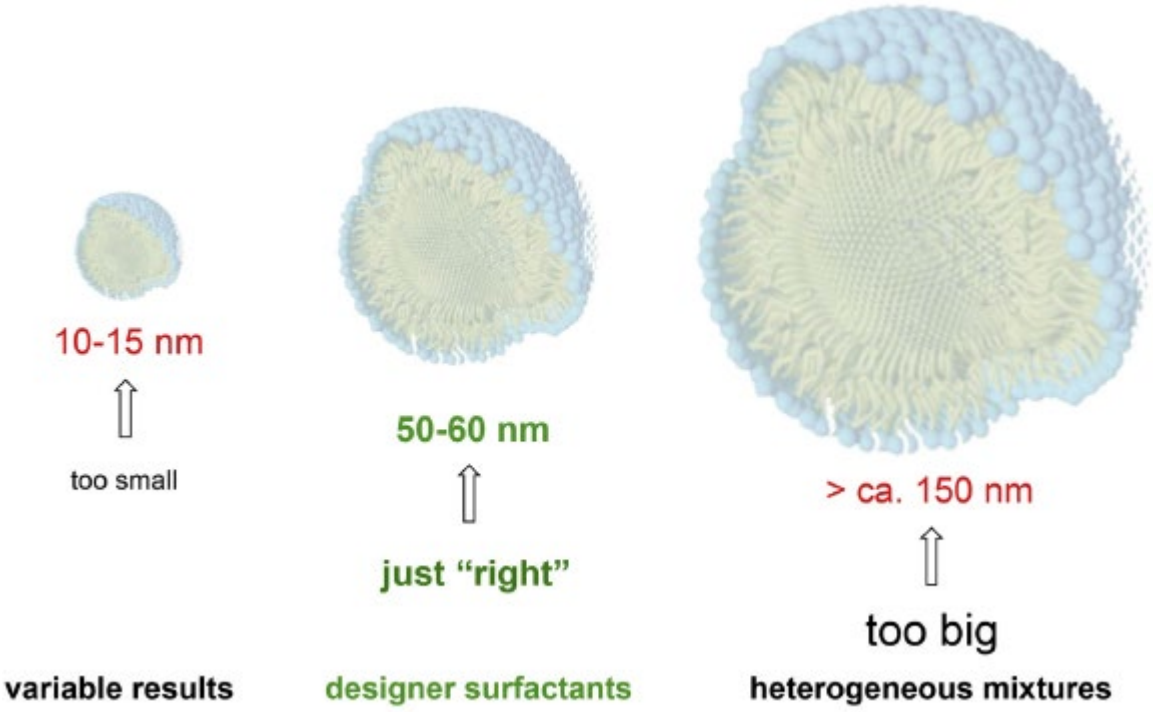


Figure 56 - The size of the micelle matters greatly and can range between 10 to >150 nm dependant on the surfactant used and the concentration of said surfactant.

Green chemistry is often defined by the E-Factor, or environmental factor, of the chemical reaction. Similar to atom economy in reactions,[113] E-factor is defined as the actual amount of waste produced in a process, defined as everything but the desired product.[101, 114] This was initially used as a vehicle to minimise chemical waste and maximise resource efficiency in industry, but it can also be used to describe how "green" a reaction is. In this case, water often used in the reaction is not considered, and much of the organic waste produced is organic solvent. Utilising micellar catalysis would greatly reduce the E-factor of the reaction as it not only utilises an aqueous system but also can be considered as a reusable reaction medium once the organic product has been extracted. It has been shown that E-factors of palladium catalysed reactions such as Heck, Suzuki-Miyaura and Sonogashira couplings can be reduced by 20 fold in micellar media compared to the standard organic conditions.[112] This takes into account the organic solvent used in work up, but when the aqueous is recycled the E-factor drops considerably with each subsequent reaction. This shows that micellar reactions are also a "green" alternative to the usual organic solvents.

3.3 – Commercially Available Designer Surfactants

There are multiple different classes of surfactant that can be used to form micelles. Anionic, cationic, and amphoteric molecules rely on charged residues to form the hydrophilic portion

of the surfactant. There are also non-ionic surfactants which use strong hydrophilic groups such as polyethylene glycol to form the hydrophilic section [Figure 57].[98, 115] The hydrophobic part is often made from fatty acid chains or lipophilic natural products. In most cases surfactants used to form micelles as nanoreactors would need to be inert to the reaction conditions. Therefore, the use of non-ionic surfactants is prevalent when forming micelles for use in organic chemistry. So-called "designer surfactants" have been developed by the Lipshutz group to carry out a variety of chemistry under micellar conditions.[116]

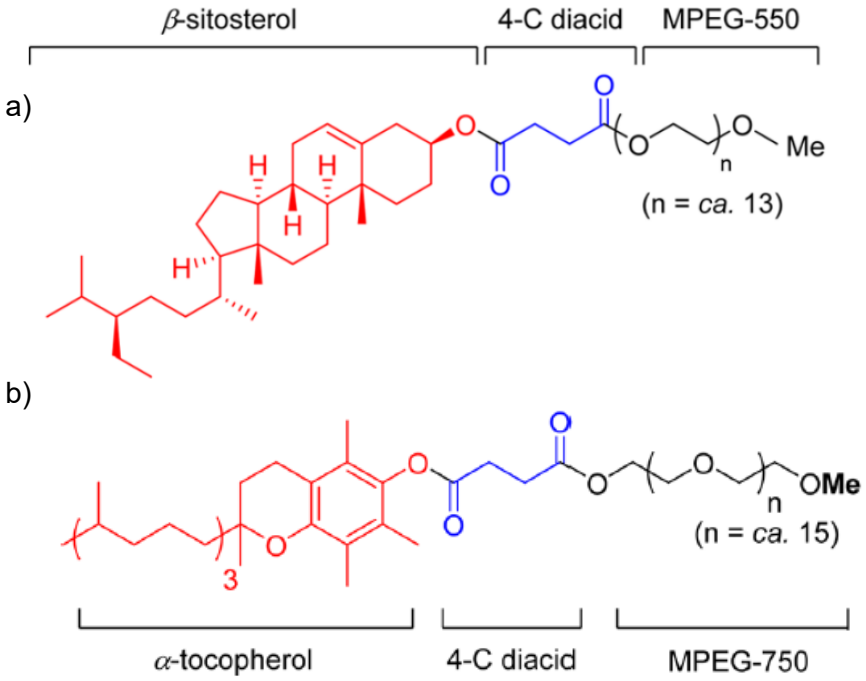


Figure 57 - a) 3rd generation designer surfactant, NOK, based upon beta-sitosterol as the hydrophobic part, and PEG-550 as the hydrophilic; b) Designer surfactant TPGS-750-M based upon vitamin D with a PEG-750-M hydrophilic part.

These designer surfactants have been studied extensively with a huge variety of successful reactions being developed. These include palladium catalysed Suzuki-Miyaura,[117, 118] Sonogashira,[119] Stille[120] and Heck couplings[121], copper catalysed cross-metathesis,[122] amide couplings,[123-125] S_NAr reactions,[126] and many others.[127] Enhanced reactivity has been shown to result from the use of designer surfactants compared not only to aqueous conditions but also to classic organic solvents. In development of these surfactants, many different hydrophobic and hydrophilic parts were tested. It was found that the length of the PEG chain had a vital role in the degree of conversion across several reactions. It was found that the "sweet spot" was to use methylated PEG-750 for many reactions. When the PEG chain was

lengthened or reduced, significant changes in conversion rates were observed.[115] This is thought to be due to the size and shape of the micelle that is produced where PEG-750-M yields the best overall micelle size and shape for organic chemistry to take place [Figure 58]. It is also noted that the micellar size indicates a "sweet spot" for chemistry to be successful which seems to be between 40 and 50 nm in diameter. Surfactants forming micelles in this range produce much greater chemical conversion rates. The hydrophobic part of the surfactant is made up from the readily available tocopherol, a vitamin E class of compound. One of the major benefits of using these surfactants is that the materials to synthesise them are cheap and readily available, requiring little chemical conversion in comparison to standard chemical solvents.

Surfactant	Average micelle diameter (nm)	Conversion (%)
none	-	30
TPGS-750-M	49	77
NOK (SPGS-550-M)	46	76
SPGS-600	59	71
SPGS-750-M	14	63
SPGS-1000	25	63
CPGS-750-M	25	44
PTS	23	71
PSS	10	62

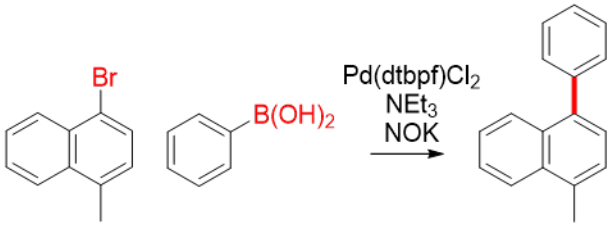
Figure 58 - Average diameter of surfactants in water and effect of these surfactants on the conversion of Grubbs-2 metathesis reactions.[98]

As well as being cheaper and more environmentally friendly, these solvents are also comparable, if not superior, to the usual organic solvents in terms of productivity. Many reactions that would usually require high temperatures, organic solvents and low dilutions can be carried out in high concentrations, low temperatures and at least 95% water. Palladium and metal catalysed reactions lend themselves perfectly to this technology. Utilising low catalyst loadings in high concentrations allows chemical reactions to take place at low

temperatures. This technology has been extensively exploited in palladium catalysis with great success. The interesting detail is that once the reaction has taken place, the aqueous phase can be extracted and reused with no loss in chemical reactivity. This further decreases the environmental impact of the chemistry as not only is the solvent used a green alternative but highly expensive and rare transition metal complexes can be used multiple times [Figure 59 a)].

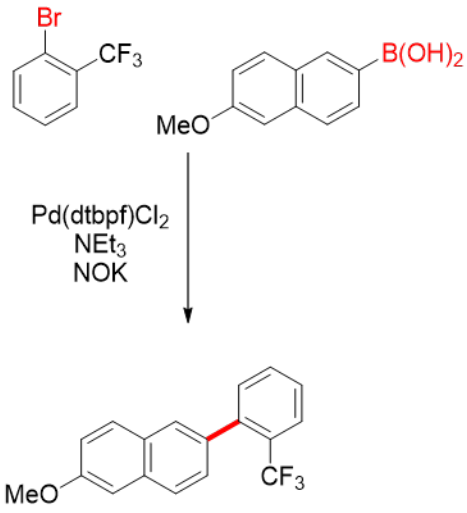
In addition to recycling the aqueous phase, the actual reaction has a significantly lower E-factor than that of standard solvent based organic chemistry [Figure 59 b)]. This reflects the high-yielding coupling reactions that take place, the low temperature required, and the minimum use of both aqueous and organic solvents in the overall process. If recycling were also considered, these E-factors would be even lower, making reactions in micellar solvents extremely green in terms of productivity and waste. A reaction Suzuki-Miyaura coupling run only once with no recycling would generate an E-factor of around 7.6 when organic solvents for extraction are accounted for [Figure 59 b)]. This would lower considerably with recycling of the reaction media. By comparison, typical E-factors in the pharmaceutical industry range between 25 and 100 for these types of reaction.[101]

a)



% Conversion after each reaction cycle					
1	2	3	4	5	6
>99	>99	>99	>99	>99	>99

b)



E Factor = 3.4 (based on organic solvent used in extraction)

E factor = 7.6 (based on both organic solvent used and water in reaction mixture)

Figure 59 - a) recycling of NOK in water determined by ^{1}H NMR spectroscopy; b) E-factor model reaction.

A multitude of reactions have been found to work in micelles, notably metal catalysed processes such as ring closing metathesis and palladium cross-couplings. Where metathesis was carried out, Grubbs-2 catalyst was used in a 2% weight surfactant in water.[128] These couplings were carried out at room temperature with yields of 70-100%. When carried out without the surfactant, the yields of these reactions dropped to 30% showing that the surfactant is vital in the product formation. Suzuki couplings were also carried out under similar conditions using 2% TPGS-750-M, $Pd(dtbpf)Cl_2$ and triethylamine at room temperature. Interestingly in all the examples shown, yields of over 90% were recovered at reduced temperatures compared to usual Suzuki couplings, where solvents such as DMF and dioxane are often required, and heating over 100°C is not a rare occurrence. These sorts of reactions were then extended to different palladium coupling reactions including, Buchwald, Heck, Sonagashira, Negishi and borylation reactions. All proceed at very low temperatures and in good yields.[98]

3.4 – DNA reactions in micelles

Carrying out organic reaction on DNA is inherently challenging as water is required to dissolve the DNA. Many organic solvents are not miscible with water, and certain organic solvents can even damage DNA when used in high concentrations. Many recent discoveries have often resulted in modifying the reactants to gain more solubility, such as adding sulphonate residues to ligands or even using solid supported materials to gain access to the huge variety of chemical reactions that are not compatible with aqueous reactions. However, many of these adaptations have specific requirements in terms of compound preparation or use of unusual reagents and are often more expensive or require specific synthesis to access. Utilising micelles could enable chemistry that is currently not accessible as well as allow the use of non-water-soluble reagents and increase DNA compatibility.

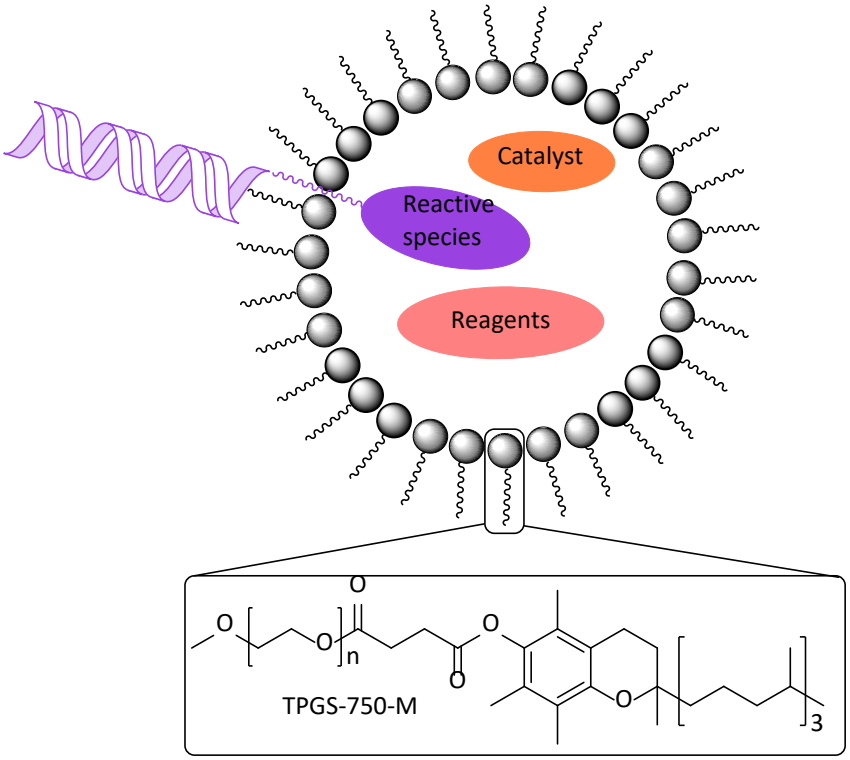


Figure 60 – Example micellar promoted reaction carried out on a DNA tagged substrate.

Micellar technology allows many chemical reactions to proceed in aqueous media and so most of the successful micellar reactions should be applicable to DNA substrates. The fact that these sorts of reactions can be accomplished at very low temperatures is also highly desirable. DNA can undergo degradation under harsh reaction conditions and elevated temperatures, many of which would be required to carry out usual synthesis, such as palladium coupling reactions. The insoluble reagents used in standard organic chemistry could be solubilised in the micellar core allowing them to be compatible with an aqueous synthesis [Figure 60]. This has the positive effect of allowing numerous reactants that would previously be insoluble and unreactive in aqueous chemistry to now be compatible for aqueous DNA conjugated chemistry. When used for classical chemical reactions, micelles often contain an elevated concentration of reactants in the core. This allows for a substantial increase in reactivity for the ongoing reaction as the rate is often highly affected by concentration. This is also apparent for catalytic reactions where a lower catalyst loading, and lower temperature can be used because reactant concentration is raised.

The benefit of having localised reactants is not simply raising reactivity in this example. When utilising DNA substrates, it is often found that reactants and substrates can react and deteriorate the DNA coding sequence. When using micellar conditions, the organic moieties

are in much higher concentration in the core and PEG region of the micelle than the bulk aqueous phase.[129] This would mean that the substrate and catalyst concentration in the aqueous phase will be extremely low. This in turn would help limit DNA degradation from reactants as it would not interact with the potential substrates which could degrade it. Having a lipophilic molecule on the terminal end of a PEG linker attached to DNA should permit the molecule to still enter the micelle and allow access to the highly concentrated reactants. As DNA concentration is extremely low in many reactions, this should allow rapid conversions to the desired products as well as protecting the DNA tag from the localised concentrated reactants. It should also allow any reaction to be carried out at low temperatures due to the high reactivity shown using micelles.

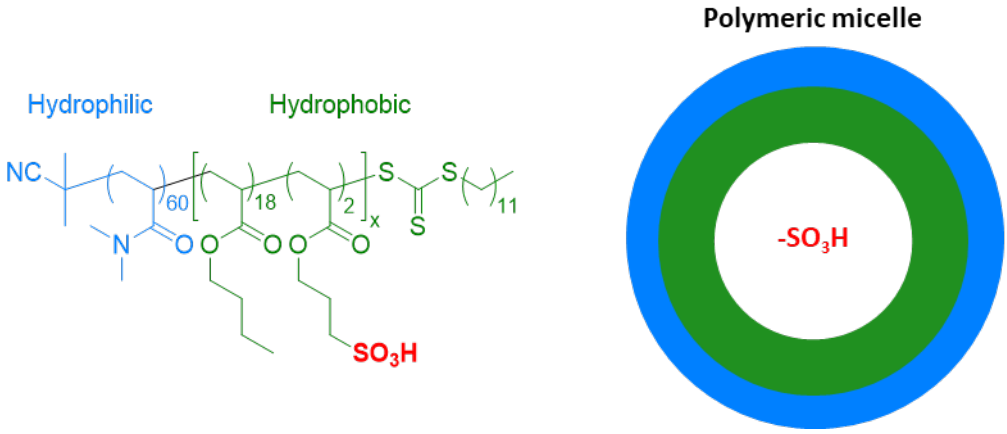


Figure 61 – Micellar forming polymeric material used for DNA compatible chemistry.

Simultaneously emerging research utilising a polymeric micelle has been conducted.[130] Here an amphiphilic block copolymer was used with an attached sulphonic acid residue to the hydrophobic interior. The polymer was shown to facilitate reactions such as the Povarov and Gröbke-Blackburn-Bienaymé multicomponent reactions and proceeded with no detectable DNA damage. The system was also modified using copper, bipyridine and TEMPO as catalysts for the oxidation of benzyl alcohol to the corresponding aldehyde with no detectable over oxidation or formation of side products. Using block copolymeric micelles shows that DNA is tolerated in these micellar reactions utilising components which would under normal circumstances damage or degrade the DNA strand.

Chapter 4 – Micellar Promoted Suzuki-Miyaura Reactions

4.1 – The Suzuki-Miyaura Coupling

Suzuki couplings[131] between a boronate and an aryl halide to form a biaryl system remain a vital tool in the construction of biological screening libraries, including DELs, as well as being extensively used more generally in medicinal chemistry [Figure 10].[29] Employing cross-coupling reactions permits the inclusion of diverse sets of building blocks to libraries utilising just one reaction type. There are thousands of commercially accessible halogenated aromatics and heteroaromatics and nearly as many boronic acids, esters and MIDA esters. In theory a high-fidelity cross-coupling reaction would permit large libraries to be rapidly produced using a variety of different building blocks. This allows a library to contain greater chemical diversity if several of these building blocks can be coupled efficiently. In combinatorial approaches, idea reaction conditions need to be efficient and applicable to the vast majority, if not all, of the building blocks used in a reaction step. For a Suzuki reaction, this requires reactions which proceed well for boronic acids and halogenated compounds of varying chemical characteristics and electronics. Reaction conditions that work for electron rich building blocks may be less reactive than electron deficient species or sterically hindered substrates, such as those with ortho substituents. Current methods using Suzuki couplings[31, 43, 44] carried out on DNA attached conjugates show that a large majority of building blocks react inefficiently and a large portion are unreactive under reported conditions. It is worth noting that these conditions are those that are currently being employed in the synthesis of publicised DNA encoded libraries.[31, 43, 44] In these examples it is often reported that there is low conversion and side products are formed such as dehalogenation products and DNA damage.

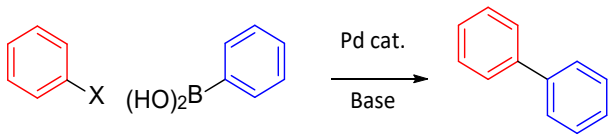


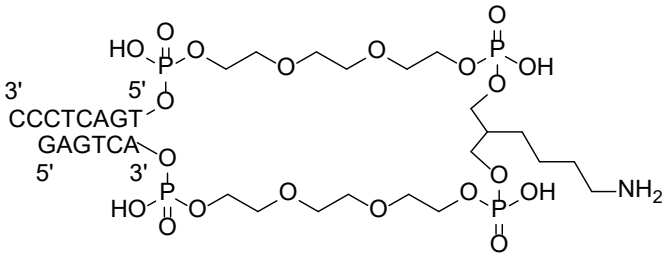
Figure 62 - Example Suzuki-Miyaura coupling.

There is a clear requirement for a more robust coupling strategy, and this also provides a great foundation for the development of micellar strategies for use with DNA encoded libraries. By utilising micelles it could be possible to further increase conversion rates, reactivity, and reduce side products formed in current palladium cross-coupling reactions. The Suzuki-

Miyaura coupling was chosen as an ideal candidate for analysing the effect of micelles with DNA conjugates due to the apparent need for development of this chemical reaction. Vast increases in reaction rates and reduced reaction temperatures were observed in solution phase chemistry when applying micelles to a catalytic system.

The use of micelles also allows the palladium and ligand complex to be tailored to the reaction. Currently many catalytic reactions require ligands incorporating solubilising groups for the reaction to take place. Using micellar technology, it is beneficial to have hydrophobic ligands as the catalyst. The catalyst would then tend to occupy the micellar core more readily than the aqueous phase. Utilising this reaction, it was expected to gain an understanding of how and if micellar catalysis is achievable when applied to a DNA encoded system. As using micelles is novel methodology, it is also vital to carry out in-depth analytical procedures to analyse not only the conversion rates and side products of the reaction, but also the integrity of the DNA tag.

Double stranded loop linker



Double stranded phenyl iodide linker

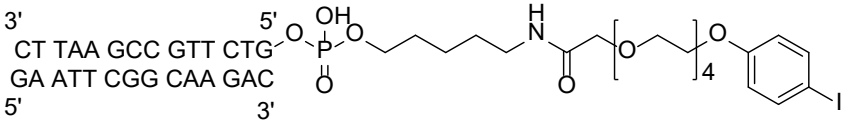


Figure 63 - Depiction of the looped linker used previously to study Suzuki-Miyaura couplings and the phenyl iodide linker proposed for this work.

The initial stage was to design a comparative micellar reaction with those used in current DNA encoded library synthesis. The standard hairpin linker used by preceding groups was replaced with a modified 5' linker which was then utilised as a position for conjugation of a reactive group for analysis of micellar catalysis. A simple phenyl iodide warhead was chosen as the initial functional group to study the effect of micellar catalysis, and this was also used to study current literature prevalent cross-coupling methodology. The initial concept was to compare

literature conditions of on-DNA catalysis with micellar conditions established to work in solution phase by the Lipshutz group.

4.2 – Design and synthesis of the linker

Synthesis of the phenyl iodide DNA conjugate begun with the off-DNA synthesis of the bespoke phenyl iodide linker. This was conducted off-DNA to limit the number of reactions, especially at a higher scale, that had to be carried out on-DNA. The concept was to synthesise the phenyl iodide with a PEG linker and carboxylic acid functionality with the ability to couple it to the C6-amino modified DNA strand [Figure 63]. This linker synthesis utilised simple high yielding chemistry to produce large quantities of substrate for conjugation with the accessible C6 amino modified DNA strand. The hairpin linker used by some other groups was extremely expensive and difficult to synthesise in-house.

The linker compound was prepared utilising uncomplicated chemistry and easily accessible starting materials [Figure 64]. PEG-4 was first mono tosylated with tosyl chloride using excess of the PEG-4 to mono tosylate the hydroxyl and was then displaced with 4-iodophenol in the following step with mild conditions. Both steps required no further purification save for an aqueous work-up. The PEG hydroxyl was then reacted with *tert*-butyl bromo acetate which was purified and then deprotected using TFA in the final step. This entire route was high yielding and straightforward. It also allowed direct access to a solid supported amide coupling reaction to synthesise the entire linker complex in a single on-DNA step. The alternative would be the synthesis of an amino capped PEG linker, attachment to DNA and then coupling to iodophenylcarboxylic acid. This route however would require an additional on-DNA reaction which would possibly add extra variables into the synthesis.

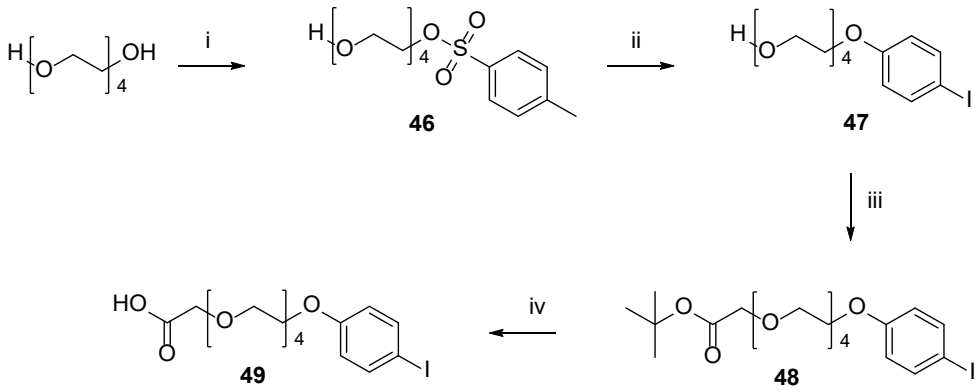


Figure 64 - synthesis of the linker compound; i) tosyl chloride, chloroform, TEA, RT overnight; ii) K_2CO_3, 4-iodophenol, DMF, overnight RT; iii) *t*-butylbromoacetate, *t*-BuOK, *t*-BuOH, 40°C overnight; iV) TFA, DCM, 4 hours.

To attach the linker to the DNA strand a solid supported method was used which was known to successfully couple linkers to a similar substrate.[63] Using this method, amide couplings could be carried out under mild conditions using organic solvents and reagents, allowing for an extremely efficient reaction and purification. To understand the quantity of DNA that was supported on a specific amount of solid supported material, trial cleavages of the DNA were carried out across known quantities of solid support [Table 1]. This appeared to show that on average 200 nmol of DNA could be cleaved from 10.3 mg of solid supported material. This was useful to determine expected yields from the solid supported reaction as well as to help calculate the amount of each reactant needed.

Table 1 – Average amount of DNA cleaved from the solid support as determined by nanodrop concentration.

Sample	mass of solid support (mgs)	mw of DNA	volume (ul)	DNA conc. (ng/ul)	nmol of DNA present
1	10.1	4408	500	1722.2	195.3
2	9.8	4408	500	1680.8	190.7
3	10.5	4408	500	1795.1	203.6
Average	10.1				196.5
Scaled	10.3				200

The DNA that was used was a solid supported modified 14mer (GTCTTGCCGAATTC) with a 5' C6 amino linker [Figure 65]. The first step was the removal of the MMT group using 3% TCA in DCM. This was achieved by flowing the solution over the solid supported material until a

colourless filtrate is visible, and then washing with DCM to remove excess acid. The free amine was reacted with the acid containing linker compound **49**. To couple the modified solid supported DNA, excess HATU, linker and DIPEA in DMF were used. After leaving at room temperature overnight, the resultant mixture was filtered and purified by washing with DMF, MeCN, MeOH and DCM sequentially to remove any organic impurities. The product was then cleaved from the solid support using aqueous methylamine and ammonia solution. Once cleavage was complete, the liquor was filtered to remove any fine solid support particulates and purified by preparative HPLC. This procedure yielded the pure single stranded DNA which was then dissolved in water and the concentration determined by NanoDrop. The complimentary 14mer was added and the mixture incubated [Appendix Figure 1]. This entire process was relatively high yielding, producing between 0.5-1 µmol of DNA synthesised from approximately 2 µmol of impure solid supported DNA. When larger quantities of DNA conjugate were required multiple reactions were performed in tandem.

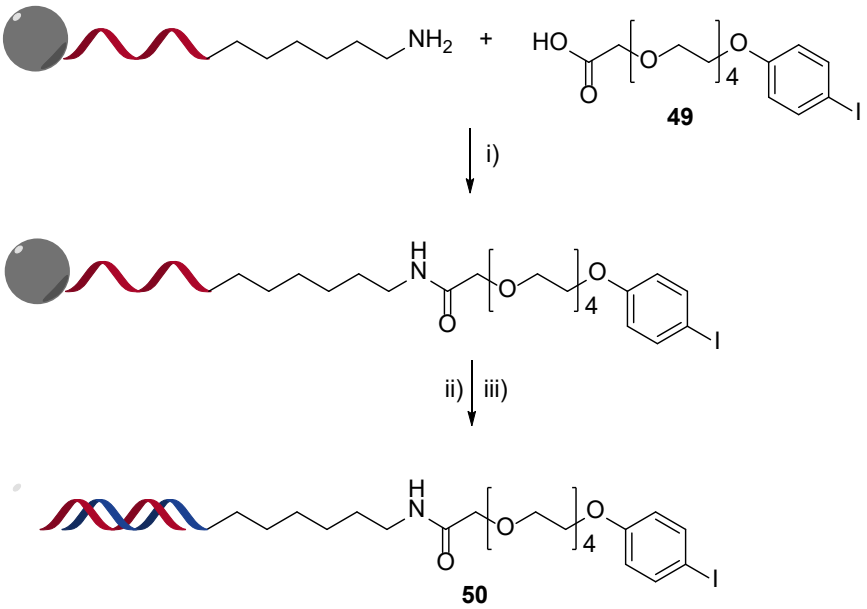


Figure 65 - Solid supported DNA conjugate synthesis; i) HATU, DMF, DIPEA; ii) aqueous ammonia, aqueous methylamine; iii) 14mer single strand DNA, 40°C.

4.3 – Initial Suzuki-Miyaura Reactions

Initial screening of test reactions could now be carried out on the DNA conjugated phenyl iodide compound. The first reaction to be trialled was a reported DNA encoded transformation from literature.[43] This reaction was discovered as one of the best conditions to carry out a Suzuki-Miyaura coupling at the time on a DNA substrate. 2 nmol of DNA was

added to a mixture of phenylboronic acid in acetonitrile and sodium carbonate in water. The tetrakis palladium was degassed using nitrogen and combined. The resulting mixture was shaken overnight at 80°C. These conditions resulted in the formation of 26% of the desired biaryl compound **51** with 39% starting material **50** remaining. These amounts were quantified by ion counts from the mass spectrometer and are representative of the bulk after a simple filtration and precipitation. A side product arising from of starting material **50** was observed at 34% of the bulk as well as some uncharacterised impurities observed throughout the spectrum. The complimentary strand can also be observed in this spectrum. This suggests that this reaction would not be suitable for use on wider substrates, even utilising simple boronic acids and halogenated reagents most of the reaction was dominated by starting material and dehalogenated side products.

Using the miceller forming surfactant NOK from literature,[98] a scaled down reaction was carried out on DNA conjugate **50**. The surfactant NOK is based on beta-sitosterol as the hydrophobic portion and MPEG-550 as the hydrophilic portion which form micelles on the addition of water. It was shown that high-yielding Suzuki cross-couplings could be achieved at room temperature in the presence of water and surfactant with no additional solvent. The conditions were slightly modified to be compatible with DNA chemistry by lowering the concentration of the halogenated substrates to 2 nmol and lowering the palladium loading to 0.1 mol% with respect of the boronic acid to limit the amount of palladium damage that could occur to the DNA strand. These initial conditions were a compromise between the organic micellar reactions and literature reactions shown to work on DNA.

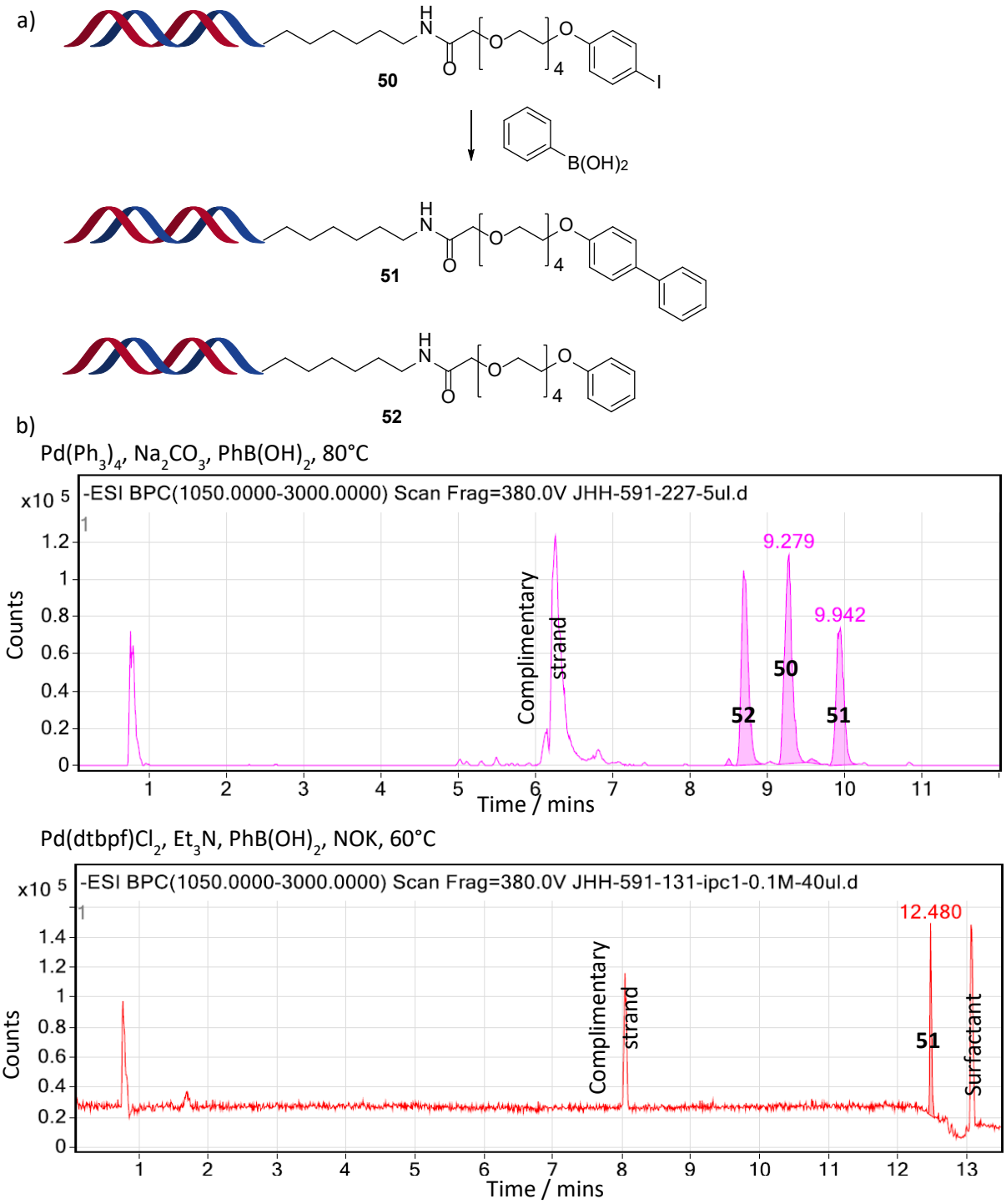
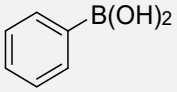
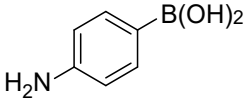
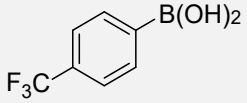
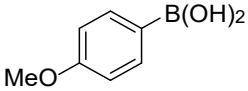
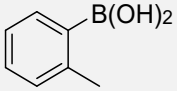
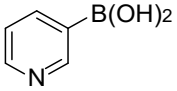
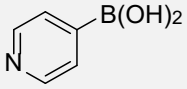
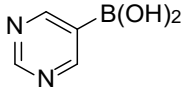
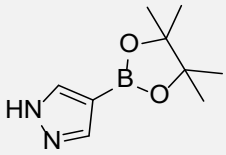


Figure 66 - a) Reaction scheme for the test iodophenyl DNA conjugate **50** and formation or desired biaryl compound **51** and dehalogenated product **52**; b) chromatograms of couplings using literature[43] and micellar conditions.

The reaction proceeded at 60°C overnight after the initial room temperature reaction failed to show anything other than starting material. Mass spectrometry analysis showed that there were 3 major peaks in the spectrum. Firstly, the complimentary strand, indicative of no loss in DNA integrity, and a peak for product, with no starting material or side products visible in

the spectrum. The third peak was found to be trace surfactant, proven with a blank water/surfactant run. These initial results were extremely encouraging that this simple modification of micellar conditions was a suitable alternative for use in DNA encoded libraries. However, phenyl boronic acid is a simple reactant and just an initial screening compound. For success as a library reaction it must work across a variety of boronic acids with varied properties.

Table 2 - Coupling of **50** with a range of boronic acids. Conditions: $Pd(dtbpf)Cl_2$ (1 eq.) in DMF, $ArB(OH)_2$ (1000 eq.), Et_3N, 1500 eq.) 2% Nok, 40 °C.

Aryl	% Product	% Starting Material, 50	% Dehalogenated, 52	% Other
	100	0	0	0
	0	100	0	0
	37	9	37	19
	57	0	36	7
	14	55	28	3
	0	92	8	0
	0	42	58	0
	23	63	14	0
	33	67	0	0

A range of substrates was then chosen to carry out a screen of boronic acids with DNA conjugate **50**. Boronic acids with different electronic properties were chosen to analyse the scope of the reaction. These contained varying functionality such as electron withdrawing 4-

trifluoromethylphenyl boronic acid and electron donating 4-methoxybenzene boronic acid. Also 2-methylbenzeneboronic acid was chosen due to the usual difficulty in coupling ortho substituents as well as an aminophenyl to discover if aniline is tolerated. Heteroaromatic substrates were also used in this trial, such as 3 and 4-pyridylboronic acid, 5-pyrimidylboronic acid and 4-pyrazoleboronic acid. These more privileged structures are vital in medicinal chemistry but frequently difficult to couple.

The reactions were completed at 40°C where phenyl boronic acid still yielded 100% conversion to product. However, these general conditions were not general across the entire range of boronic acids. As expected, the electron rich *para*-methoxyphenyl boronic acid yielded the highest conversion. With electron poor trifluoromethylphenyl boronic acid there was a significant drop in conversion. Across the board, the conversions achieved using these conditions was not ideal and exhibited large amounts of starting material and dehalogenated products as well as other impurities in some cases. Pyridyl compounds also revealed no conversion to product, but interestingly varying degrees of dehalogenation were seen across the different substrates. These results proved that the reaction would need to be improved considerably to be successfully employed in a library, but the general use of micellar forming surfactants leads to improved conversions compared to standard literature conditions.

4.3.1 – Dehalogenation

Dehalogenation events have occurred whilst using several different boronic acids in this Suzuki-Miyaura reaction. The mechanism of the Suzuki-Miyaura reaction has been studied in detail.[132] It proceeds via a complex catalytic cycle with possible different pathways. Many pathways or side reactions can take place which considerably affect how each individual system performs and, in many cases leading to the dehalogenated starting material. No mechanistic studies have been performed on the on-DNA couplings, but findings related to the mechanism of normal phase Suzuki-Miyaura couplings could be extrapolated to the on-DNA reactions.

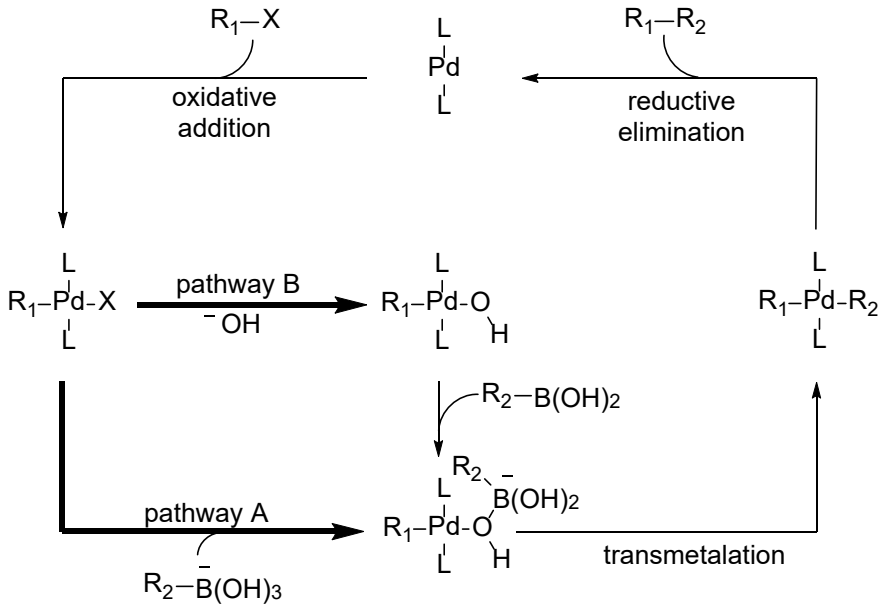


Figure 67 - Simplified generic mechanism for a Suzuki-Miyaura coupling. L= ligand or solvent, x=halide. The pre-transmetalation intermediate can either be generated through the boronate pathway A or the oxo-palladium pathway B.[132]

Transmetalation is often regarded as the rate determining step of a Suzuki cross-coupling which is implied in this case by the fact that dehalogenation occurs as a major side reaction. Often limited starting material remains in these reactions indicating that the palladium insertion occurs readily. To enhance the conversion of these reactions, it is vital to understand how dehalogenation occurs. There are two major pathways that commence the transmetalation sequence. Under either pathway this reaction [Figure 67] produces a pre-transmetalation intermediate which can undergo hydride elimination to generate the dehalogenated side product. Therefore, increasing the rate of addition of the boronic acid to the palladium intermediate should give the desired effect of reducing dehalogenation. However, the boronic acid can also degrade in the presence of base by a de-borylation reaction that proceeds through the boronate species. If too much of the boronate species is present, the faster the deborylation will occur. This is often significant when using heterocycles and electron deficient systems where increased rates of protodeborylation, oxidation and polymerization are visualised.[133, 134] The deborylation is often more apparent when using heterocycles and it has been shown that species such as MIDA esters and boronic esters hydroxylate more slowly.[135, 136] If pathway A is favoured, then this process will drastically reduce the rate of the reaction due to the instability of the boronate intermediate.

4.4 – Initial Optimisation

Although the initial results were encouraging, a significant improvement was required to create a reaction that is consistent across many substrates. The initial study did show that not only is it possible to carry out reactions in micellar media, but it may also be beneficial to do so. To further improve reactivity, a methodology that was implemented by Merck to optimise a standard organic Heck coupling was followed.[137] In this process, individual reaction factors are taken into account in a stepwise manner, sometimes screening a couple of factors simultaneously, to determine which factors are most important. 4-trifluoromethyphenyl boronic acid was chosen as the substrate to optimise the Suzuki conditions owing to the unreactive nature of the substrate and low conversion in the initial reaction. For the catalyst Pd(dtbpf)Cl_2 was used as it is known to work on a variety of Suzuki substrates and in micellar conditions.

Table 3 - Base screen with 4-trifluoromethylphenyl boronic acid, conditions: Pd(dtbpf)Cl_2 (1 eq.) in DMF, 4-trifluoromethylphenyl boronic acid (1000 eq.), base (1500 eq.) 2% Nok, 50 °C.

Base	% Product	% Starting Material 50	% Dehalogenated 52	% Other
DBU	0	100	0	0
$(i\text{-}Pr)_2EtN$	49	25	26	0
Bu_4NOAc	7	17	46	30
KOAc	6	78	16	0
Na_2CO_3	53	2	30	15
K_2CO_3	34	36	16	14
Cs_2CO_3	60	5	27	8
K_3PO_4	72	0	20	8
LiOH	88	2	4	8
NaOH	84	8	6	2
KOH	80	9	7	4

NaOH (NMP co-solvent)	86	8	6	0

The initial factor to be adjusted was the base, a highly important component in any catalytic reaction and one that is further complicated using micellar media. When using organic bases, the more hydrophobic bases tend to mix with the micellar reagents, whereas inorganic bases will often stay in the aqueous phase and are unable to enter the micelle. A variety of organic and inorganic bases were chosen [Table 3] of varied base strength. The initial screen demonstrated that DBU resulted in no product formation and Hunig's base was comparable to triethylamine. Tetrabutylammonium acetate was used to see if a phase transfer catalyst would have any effect by either helping to transfer the substrates into the micelle or allow a higher percentage of base to be present in the micelle. However, the yield was low and multiple peaks were seen in the spectra, maybe due to the phase transfer catalyst affecting micellar formation. Inorganic bases seemed to perform well in most aspects of the reaction. Carbonates were comparable to Hunig's base, but potassium phosphate increased the yield considerably. Potassium hydroxide was comparable to that of potassium phosphate with increasing conversion seen sodium and lithium hydroxide. It is possible that the counter ion stabilises the boronic acid intermediate at the trihydroxyboronate stage, and the nature of the counter ion could influence the increase of uptake into the micelle. It could also influence the overall ionic strength of the aqueous phase.

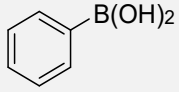
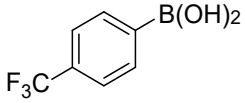
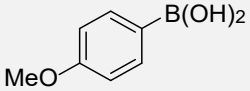
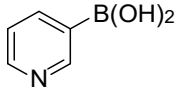
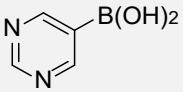
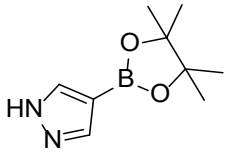
Table 4 – Screen of palladium equivalents, conditions: $Pd(dtbpf)Cl_2$ (X eq.) in NMP, 4-trifluoromethylphenyl boronic acid (1000 eq.), LiOH (1500 eq.) 2% Nok, 50 °C.

Palladium Equivalents	% Product	% Starting Material 50	% Dehalogenated 52	% Other
0.1	4	56	40	0
1	58	14	23	0
2	98	1	1	0
10	0	0	0	100

The dehalogenation could also be facilitated by ligand displacement with DMF,[138] which in turn could undergo beta-hydride elimination to yield the dehalogenated species. A reaction using NMP as the choice of co-solvent rather than DMF showed no considerable improvement, but also no deterioration and so future reactions were carried out using NMP to dissolve the catalyst to reduce the impact of DMF ligand displacement.

To study the effect of palladium on the reaction, the number of equivalents was varied with respect to the amount DNA. Increasing the amount of palladium resulted in a large impact on the reaction [Table 4] where using 2 equivalents yielded 98% conversion to desired product. Where 10 equivalents were used, multiple side products were formed in the reaction and when 0.1 equivalents were used the reaction progressed at a reduced rate. The use of 2 equivalents of palladium seemed to give a high yielding reaction for the trial electron deficient substrate. As these conditions had been optimised for just this substrate, they were screened using multiple different boronic acids.

Table 5 – Coupling using optimised conditions, conditions: $Pd(dtbpf)Cl_2$ (2 eq.) in NMP, boronic acid (1000 eq.), LiOH (1500 eq.) 2% Nok, 50 °C.

Boronate	% Product	% Starting Material, 50	% Dehalogenated, 52	% Other
	100	0	0	0
	100	0	0	0
	100	0	0	0
	21	79	0	0
	30	70	0	0
	20	80	0	0

It was uncovered that the phenyl boronic acids, both electron withdrawing and donating, proceeded in great yields as expected after the initial optimisation. However, when using heterocycles, the yield significantly decreased. Heterocycles are often less reactive than their phenyl counterparts due to the instability of the boronate complex leading to deborylation events occurring more efficiently. For an efficient library to be synthesised, these heterocycles must be included in the library design as many are privileged structures for drug discovery. The reaction conditions would therefore require further optimisation on these substrates.

4.5 – Switch to TPGS-750-M

Further optimisation of the micellar reaction was required, especially as the conditions need to support a wide variety of boronic acid substrates. For off-DNA micellar TPGS-750-M[115] is used much more often than Nok. To see the effect that the surfactant has on the overall reaction TPGS-750-M was trialled. Often different surfactants form different shaped and sized micelles, which in turn could allow for a more optimised reaction media for this novel on-DNA reaction system. Often in the literature the concentrations of both substrates and base were 0.5-1M.[98, 112, 118, 127, 139] In the previous experiments the overall concentration in respect to the boronic acid was only 0.1M. It would be near impossible to have this concentration of DNA without either massively increasing the scale of the reaction or reducing all other substrates and applying the reaction at tiny volumes, so the DNA concentration was left at 1mM. The higher concentration of boronic acid and base would increase the concentration of reactant inside the micelle allowing a greater chance of reaction when a DNA conjugate entered. When lower concentrations are used the overall rate would be expected to decrease as the concentrations of reactants inside the micelle would considerably reduce. This effect would only be the case for organic reactants, as if the micelle is packed with reactants the DNA conjugated iodide would have a high chance of reacting upon entering the micelle. As only a small mol% of palladium is used for this reaction in respect to the boronic acid, it is vital that when the palladium intermediate is formed by oxidative addition of the phenyliodide, the boronic acid is present to react with, therefore lowering the chance of dehalogenation. This technique should in theory improve reactivity and reduce dehalogenation simultaneously.

Table 6 - Base screen with 4-trifluoromethylphenyl boronic acid, conditions: $Pd(dtbpf)Cl_2$ (2 eq.), 4-trifluoromethylphenyl boronic acid (0.5M), base (0.75M), 2% surfactant, 50 °C.

Surfactant	Base	% Product	% Starting Material 50	% Dehalogenated 52
TPGS-750-M	LiOH	86	3	11
TPGS-750-M	K_3PO_4	86	0	14
TPGS-750-M / 15% THF	K_3PO_4	100	0	0

Both changing the concentration and switching to TPGS-750-M seemed to have no increase in reactivity for the reaction containing lithium hydroxide. Interestingly however an increase in reactivity was observed when using potassium phosphate as the base yielding similar conversions as hydroxide, and an increased miscibility of the entire reaction was seen. At these elevated concentrations miscibility could start becoming a problem, especially with less soluble substrates, and so potassium phosphate was chosen for future reactions. An investigation into co-solvents was then conducted as they can play a crucial role in micellar reactions.

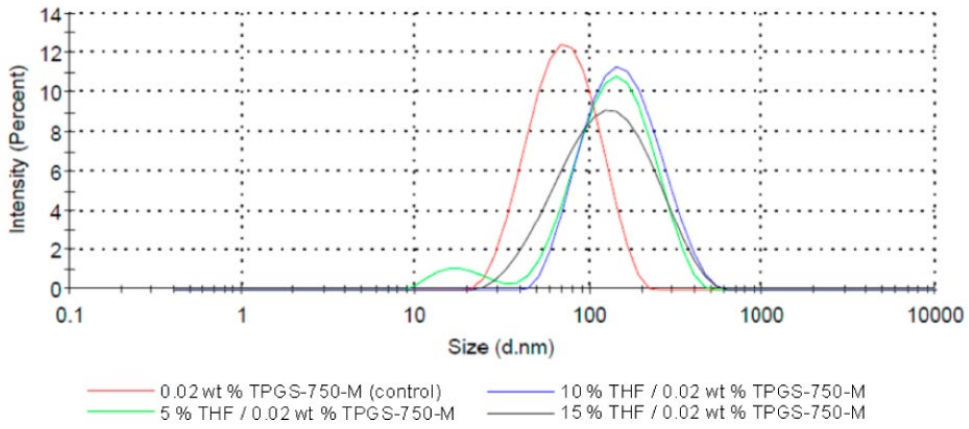
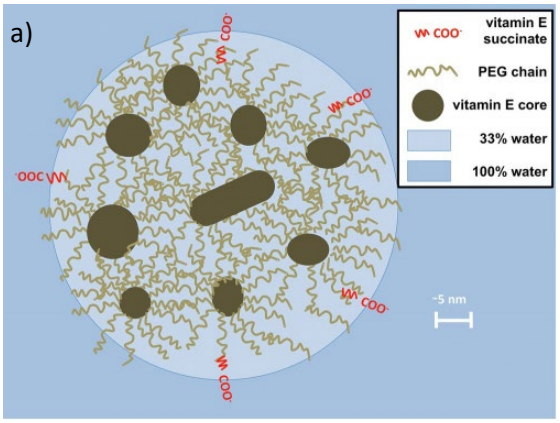


Figure 68 – Dynamic light scattering data (particle size): TPGS-750-M/H2O + THF.[139]

Using co-solvents in tandem with micellar forming surfactants can be beneficial on the overall conversion rates.[139] Co-solvents such as THF can be used to greatly enhance the solubility of specific species, especially catalysts, in micellar media. Under multiple examples with cross-coupling reactions, especially Suzuki-Miyaura, they found that 10-15% THF co-solvent

efficiently enhanced yields. It was also observed that the size of the micelles created in solution increased substantially when an organic solvent is added. The average particle size of TPGS-750M is 57 nm, but on addition of 10% THF this increases to 129 nm with populations becoming more diverse [Figure 68]. This is expected as the solvent molecules will organise themselves within the micelle framework causing growth of the particle up until a saturation point occurs. This data was captured using dynamic light scattering and provides an understanding into why these reactions are enhanced by the addition of THF as larger particle sizes can facilitate greater volumes and enhanced solubility of the substrate.

a)



b)

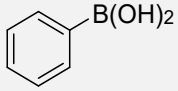
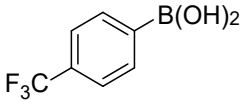
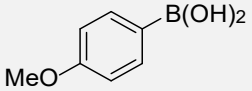
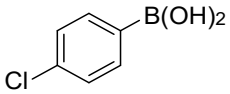
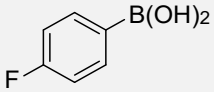
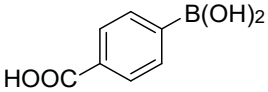
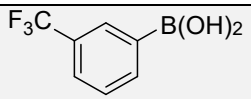
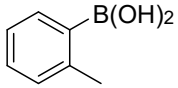
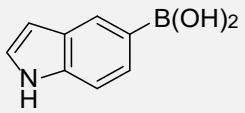
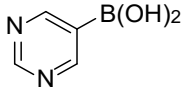
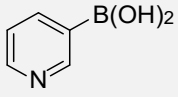
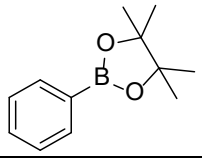
Molecule	Mole fraction phase 1– "PEG"	Mole fraction phase 2– "core"
water	0.632	0.0737
PEG region	0.044	0.00088
core	0.0032	0.529
vitamin E succinate (Rb)	0.00091	9_{E-7}
THF	0.319	0.37

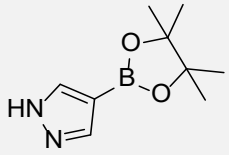
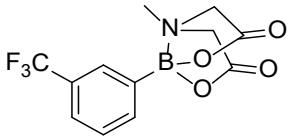
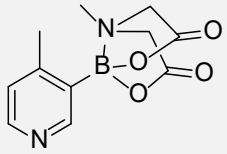
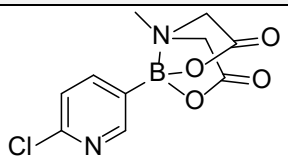
Figure 69 - a) The proposed structure of a larger nanoparticle comprised of micelles of around 50-60nm observed for TPGS surfactants. There are likely 30-40 micelles per nanoparticle and are expectant to be in flux between nanoparticles; b) COSMO-RS calculations for 2% wt TPGS-750M in water in the presence of 15% wt THF as a cosolvent.[129]

The shape and structure of micelles created using TPGS-750M has been studied in detail.[129] it was discovered that micelles were created in large nanoparticles containing multiple micellar units [Figure 69 a)]. Transmission electron microscopy imaging revealed the structure of these micelles were more closely related to a nanoparticle where multiple PEG regions interacted to form multiple greasy cores inside a larger particle. The distribution of each substrate was also calculated using a COSMO-RS model. On addition of THF much of the solvent was situated in the core of the micelle or the PEG region between the micelles. This gives an insight into how THF could improve the solubility and the conversion rates of the reactions if used as a co-solvent. It can be deliberated that the THF helps to solubilise the substrate outside the micelle and aided by the hydrophobic effect, can drive the reactants into the micellar region by providing a more solubilising environment. It could also aid in exchange of substrates between micelles and micelle mixing. The largest mol fraction of the PEG region was also taken up by water which would aid solubility of the DNA strand used in this on-DNA system.

It can be hypothesised that THF as a co-solvent could increase conversion rates in DNA micellar reactions due to the above phenomena. Using 15% THF in a test reaction yielded full conversion to the product with 4-trifluoromethylphenyl boronic acid [Table 6] showing that it may be applicable across multiple reactants.

Table 7 - Coupling of 1 with a range of boronic acids, conditions: $Pd(dtbpf)Cl_2$ (2 eq.), boronate (0.5M), K_3PO_4, (0.75M) 2% TPGS-750-M, 15% THF, 50 °C. *undetermined side product.

Boronate		% Product	% Starting Material, 50	% Dehalogenated, 52
a		100	0	0
b		100	0	0
c		100	0	0
d		100	0	0
e		100	0	0
f		69*	0	0
g		100	0	0
h		97	0	3
i		100	0	0
j		34	0	66
k		79	0	21
l		100	0	0

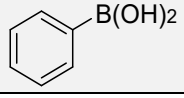
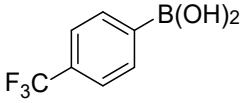
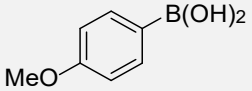
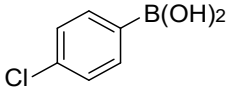
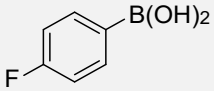
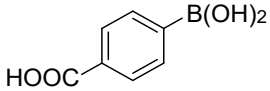
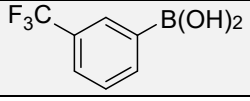
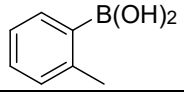
m		5	95	0
n		65	0	35
o		6	94	0
p		67	0	33

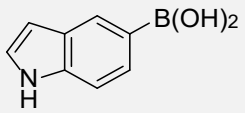
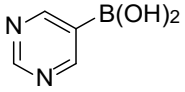
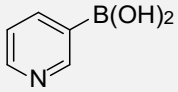
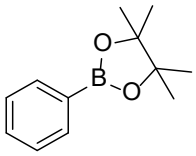
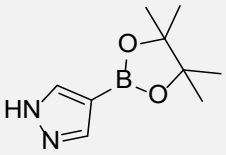
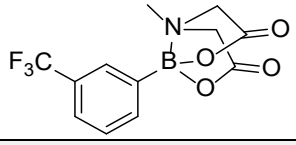
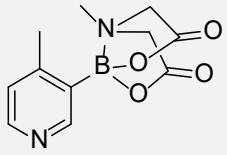
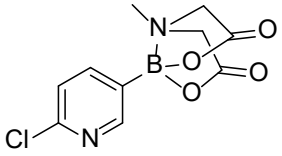
After performing these further optimised conditions across a broad range of boronic acids, it is clearly visible that the reaction performs well across a much wider range of substrates. These included multiple phenyl boronic acids with varied structures and functional groups as well as some heterocycles, boronic esters and heterocyclic MIDA esters. 100% conversion rates were seen for the trifluoromethylphenyl boronic acid species. The advantages of using a co-solvent were apparent as the substrates were visibly more miscible in the solvent. The same conditions when applied to phenyl, chlorophenyl, fluorophenyl and methoxyphenyl boronic acids also showed 100% conversion rates proving that these conditions were improved. Difficulties did arise when using the *para*-carboxyphenyl species, but the 31% bi product of the reaction could not be determined. A reaction was performed containing no palladium to test the *para*-carboxylphenyl substrate was not forming a by-product unrelated to the cross-coupling reaction. This resulted in no detrimental DNA damage. It was concluded that the reaction conditions using the carboxyl were creating an unknown side reaction when palladium was introduced. This was not followed up due to the conditions requiring optimisation for other substrates.

Heterocyclic compounds demonstrated increased conversions under these conditions. However, they still created large amounts of dehalogenated side products especially with pyrimidyl substrates. Pinacol esters appear to be tolerated but difficulties in reacting the pyrazole pinacol ester were observed. Finally, the MIDA esters worked to some degree. But

the substrates tested are difficult substrates for Suzuki couplings and so conditions that would increase the conversion of the heterocyclic boronic acids should also increase conversion of the MIDA esters. However, MIDA ester of meta-trifluoromethylphenyl did not covert as well as the boronic acid matched pair. This pattern could be due to the MIDA ester showing enhanced stability under these conditions on non-heterocyclic compounds and therefore not degrade into the reactive boronic acid species needed for reaction. It is currently up for debate how MIDA esters transmetalate to the palladium species although it is widely considered that a hydrolysation event must occur to activate the species as either the boronic acid or boronate. This would suggest that these reactions would likely be much slower when using micellar conditions as the MIDA ester would be within the micellar partition, slowing degradation by the aqueous phase base. However, this could also be beneficial due to the rapid decomposition of heterocycle boronates allowing enhanced reactions using heterocyclic MIDA esters.

Table 8 - Coupling of 1 with a range of boronic acids, conditions: $Pd(dtbpf)Cl_2$ (2 eq.), boronate (0.5M), K_3PO_4, (0.05M) 2% TPGS-750-M, 15% THF, 50 °C.

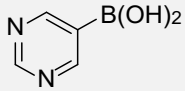
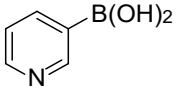
Boronate		% Product	% Starting Material, 50	% Dehalogenated, 52
a		74	0	22
b		59	0	0
c		70	8	15
d		65	0	0
e		74	0	18
f		60	0	0
g		13	72	0
h		41	20	30

i		74	0	26
j		100	0	0
k		94	0	6
l		52	0	0
m		18	82	0
n		Decomposition		
o		Decomposition		
p		12	83	0

As a large quantity of deborylation was still occurring in these reactions it was speculated that it could be due to the highly basic conditions creating a large excess of hydroxyl ions, which would aid in deborylation events of the boronate intermediate. Since heteroaromatics are electron poor, the negative charge induced when the boronate degrades can be stabilised over the system making these substrates prone to deborylation. However, depending on which route of transmetalation is dominant, these boronates could be essential intermediates. The concentration of base was reduced to prevent larger degrees of boronate formation to evaluate this concept. 5-pyrimidyl boronic acid was chosen as the current conditions yielded a majority of dehalogenated material. Consequently, conversion rates were significantly improved yielding 100% desired product with this substrate. This possibly proves that base induced deborylation could be why conversation was low in this example.

However, when these conditions were applied across multiple substrates [Table 8] the efficiency of the reactions deteriorated considerably, in some cases implying full decomposition of starting materials. This suggests that a larger quantity of base is required for optimal transmetalation to take place, but as the heterocycles degrade at such concentrations, an intermediate of the two would be required. It also suggested that the reaction is substrate dependent indicating that some substrates could be unreactive under conditions optimised for others.

Table 9 - Coupling of **50** with a range of boronic acids, conditions: $Pd(dtbpf)Cl_2$ (15 eq.), Boronate (0.5M), K_3PO_4 (0.75M), 2% TPGS-750-M, 15% THF, 50 °C.

Boronate	% Product	% Starting Material, 50	% Dehalogenated, 52
	60	40	0
	100	0	0

As an alternative to base induced degradation it is also know that substrates such as pyridine and other heterocycles can also ligate and poison palladium catalysts.[140] The excess of starting materials in this reaction meant that the palladium was at an overall concentration of 0.06 mol% relative to the boronic acids. This is extremely low loading and usual organic Suzuki-Miyaura couplings would be expected to be around 1-5 mol%. There was an increase up to 100% conversion to desired product for 3-pyridyl boronic acid when using 15 equivalents of palladium. This indicates that the quantity of palladium is an important factor in this reaction and heterocycles possibly negatively impact and poison the palladium species. A reduction in yield was seen for the pyrimidyl boronic acid substrate. Taking these results into account it was showing that there were co-dependencies between the reaction conditions and the substrates. For further optimisation, a factorial design experiment would be used to account for multiple dependencies at once.

4.6 – Factorial Experimental Design Experiment

Pyridyl and pyrimidyl boronic acids had clear differences in the optimal conditions for both substrates. These were chosen as the two boronic acids to be used in the FED experiment. It was assumed that if the reaction was to work well on the difficult to react substrates, then

the simpler substrates should also react to a similar degree. Whether there would be a reduction in conversion for these other substrates is not known, but it was believed that having a standard reaction that works well across a broad range of substrates has higher value than the current reaction conditions that only work with phenyl species.

Table 10 - Factorial experimental design assessing conversion for 3-pyridyl and 5-pyrimidinyl boronic acids.

Temp / °C	Pd (eq.)	Base (eq.)	Boronic Acid	% Conversion
40	20	100	pyrimidyl	67
40	2	100	pyrimidyl	19
60	2	1500	pyridyl	14
60	2	100	pyridyl	1
60	20	1500	pyridyl	100
50	11	800	pyridyl	100
60	20	100	pyrimidyl	1
40	20	1500	pyridyl	100
40	2	1500	pyrimidyl	25
40	2	100	pyridyl	1
50	11	800	pyrimidyl	100
40	20	1500	pyrimidyl	1
40	20	100	pyridyl	100
60	2	1500	pyrimidyl	30
40	2	1500	pyridyl	5
60	20	100	pyridyl	63
60	2	100	pyrimidyl	26
60	20	1500	pyrimidyl	1

The main parameters for the coupling were chosen prior to the FED being carried out. When choosing these parameters usually a high and a low point are chosen and using an FED with centre points helps to establish the values that are required for the reaction to be at its optimal. The equivalents of base were chosen with a maximum of 1500 and a minimum of 100, where low equivalents of base seem to favour the pyrimidyl substrate and high equivalents seem to favour the pyridyl. Following similar reasoning, palladium equivalents of 2 (low) and 20 (high) were used. The final variable optimised was the temperature of which the reaction was run, picking 40 to 60°C as the temperature range. Utilising JMP Pro 13, the minimum and maximum data points were added along with mid-points in the design, creating a 20 point experiment for the optimisation of the reaction with these two substrates [Table 10].

These reactions were carried out in a heated microreactor found to be highly efficient at both heat dispersion and of reducing solvent loss to an absolute minimum. Different temperatures were carried out on separate days, and it was observed that as the temperature rose the samples seemed to solubilise more readily. These experiments were all also carried out over a 5-hour period once reaching the required temperature with efforts to prevent overheating at the start of the reaction. The data that was gathered considered the percentage of product, starting material, dehalogenation, and any other side product formation. From initial analysis of the crude data set it is apparent that the temperature has a large effect on the outcome of the reaction where higher yielding reactions are observed as temperature rises. The experimental design was also shown to be comprehensive and the experimental conversion rates aligned well with the predicted data [Figure 70 b)]. There are multiple points at near to 100% conversion which are slight outliers due to the predicted model accounting for greater than 100% conversion. As the experimental data aligns well with the predicted results this indicated that the reaction process was a success.

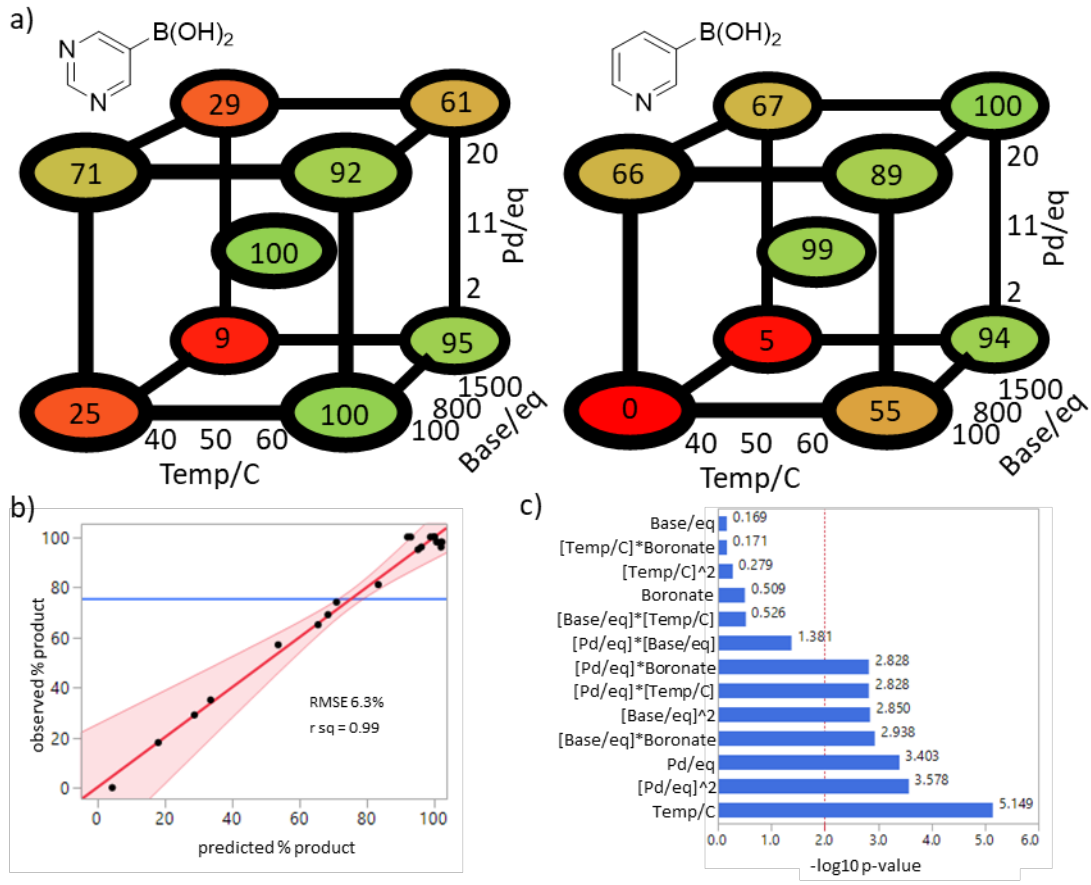


Figure 70 – Optimisation of coupling conditions leading to desired biaryl products using FED; a) Cube plots showing the percentage of desired product for 5-pyrimidinyl and 3-pyridyl

boronic acids showing the degree of conversion across the design space for the modelled data; b) Comparison of fitted data from the predictive model with the experimental data; c) Assessment of the most influential terms in the FED and profile of the effect of temperature, palladium equivalents and base equivalents.

Examining the cube plots of the data gathered [Figure 70 a)] reveals that there are considerable differences between the optimal conditions for both substrates. These differences also aligned well with the data gathered previously noting that palladium and base have significant effects on the reaction. Interestingly the centre points of the cube plots both show high-yielding reactions indicating that the maxima and minima data chosen were appropriate for this reaction as the optimal conditions should lie within the cube. As noted previously, temperature seems to have the greatest effect on the overall outcome of the reaction [Figure 70 c)]. Assessing which of the terms is most influential, temperature is deemed a large factor. This is however slightly influenced by the fact that low temperatures are unreactive. It is also noted that palladium and base, both with the second order effects, are vital factors in this experiment. This is further proof that a FED is required to find optimal conditions for this sort of reaction where multiple parameters with parabolic relationships are present.

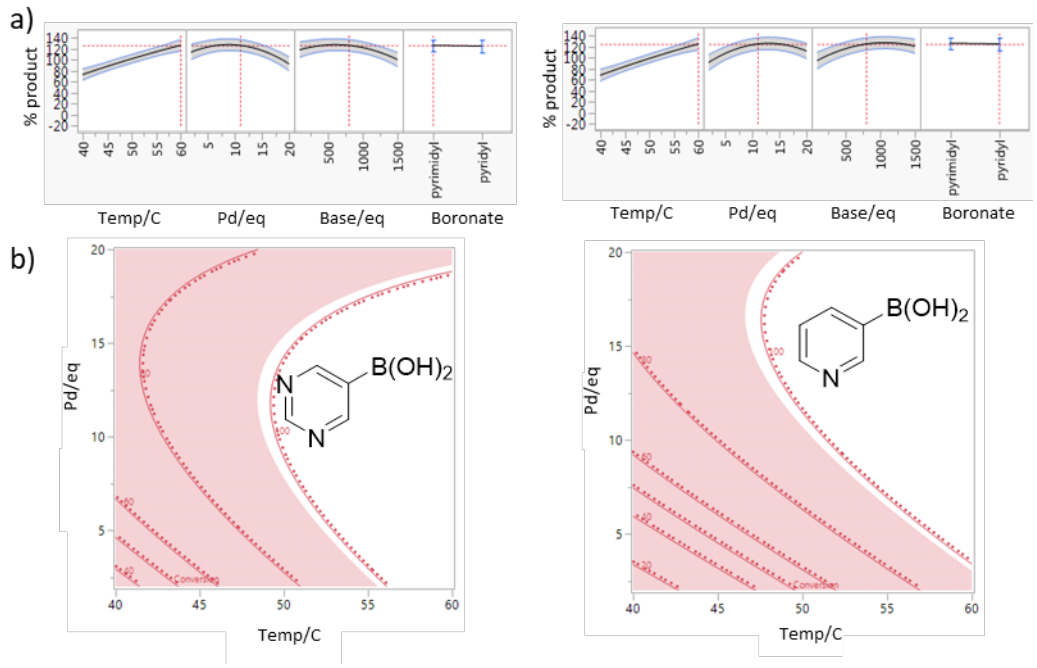


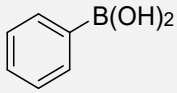
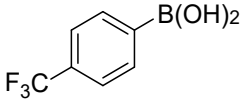
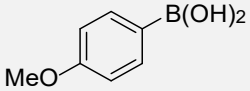
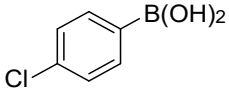
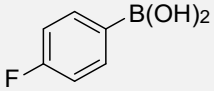
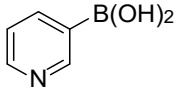
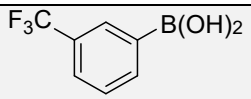
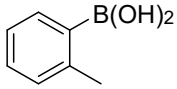
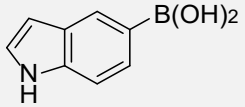
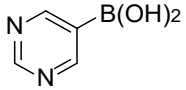
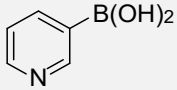
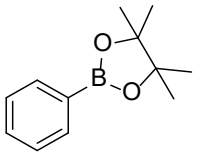
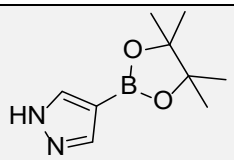
Figure 71 - a) Response curves showing relationship between % product formation and individual parameters for each boronic acid and robustness of response for optimised conditions (Temp 60 °C, 11 eq. Pd and 800 eq. base); b) Contour plots showing the combined effect of temperature and palladium equivalents for each substrate.

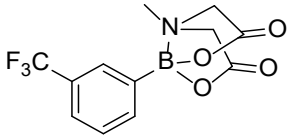
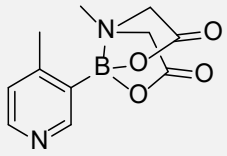
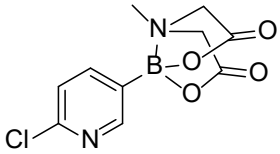
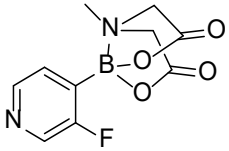
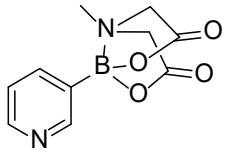
Analysis of the response curves and contour plots indicates that there is an overall relationship between both of the boronic acids [Figure 71]. It is evident that both reactants have distinct optimised conditions to yield full conversions for this reaction. However, a point at which both should be highly yielding is identifiable using this data set. As stated previously, temperature has an overall linear relationship with both boronic acids yet both palladium and base show parabolic relationships. However, these parabolic curves do have points at which they are both at near maxima for both boronic acids suggesting that optimal conditions for both could be achieved. It was observed that the optimal conditions were indeed close to the centre point for this reaction subset. Using 11 eq. palladium, 800 eq. base at 60 °C should be highly optimal. These results were hugely encouraging as two of the most challenging substrates previously are observed to react fully increasing from 0% observed in the initial screen. Applying these conditions, a further screen against the range of boronic acids was carried out to see if the optimised conditions for these two heterocycles was subsequently optimal for all substrates.

4.7 – Optimised System

The optimal conditions from the FED were then applied to the full range of boronic acid substrates [Table 11][Appendix Figures 2-19]. The results were highly encouraging showing conversion to desired product at >95% for all substrates and 12 out of the 18 only detecting product formation. This set of substrates included electron poor and electron rich phenyl derivatives, all of which progressed with 100% conversion rates. Sterically hindered ortho substituted partners also showed full conversion to product as well as indole. The technique was also highly applicable to heterocycles, boronic ester and MIDA esters. The only exception was 2-chloropyridin-3-yl MIDA boronate which showed further coupling with the chloro functionality of the product. Overall, there was a vast increase in product formation for these post FED optimised conditions compared to those used before. The least reactive substrates of 3-pyrazol boronic ester and 4-methylpyridin-3-yl MIDA boronate previously showed conversion of ~5% but now were achieved at >95%. All chromatograms of the optimised reaction condition substrates can be found in the appendix.

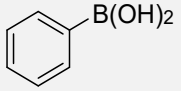
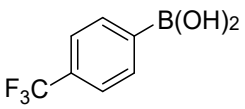
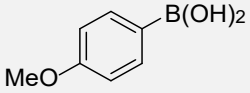
Table 11 - Coupling of **50** with a range of boronic acids, conditions: Pd(dtbpf)Cl_2 (11 eq.), boronate (750 eq.), K_3PO_4 (800 eq.), 2% TPGPS-750-M, 15% THF, 60 °C; (*18% further coupling with chloro functionality of the product).

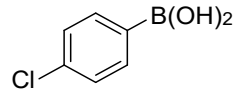
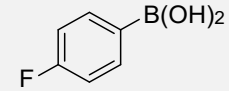
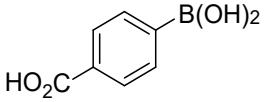
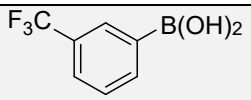
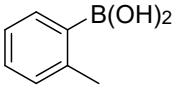
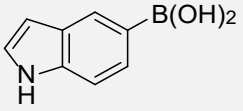
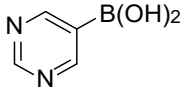
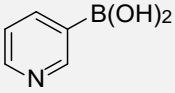
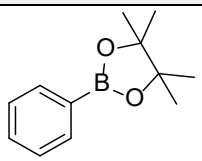
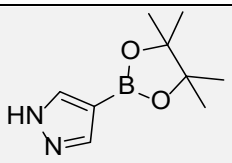
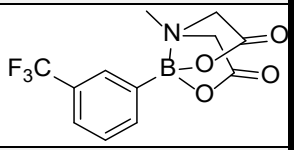
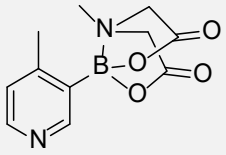
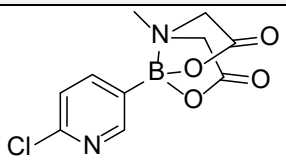
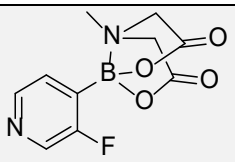
Boronate		% Product	% Starting Material, 50	% Dehalogenated, 52
a		97	0	0
b		100	0	0
c		100	0	0
d		100	0	0
e		100	0	0
f		100	0	0
g		100	0	0
h		100	0	0
i		100	0	0
j		100	0	0
k		100	0	0
l		96	0	0
m		97	3	0

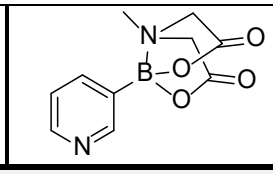
n		96	0	4
o		100	0	0
p		81*	0	1
q		98	0	2
r		100	0	0

To follow on from these highly encouraging results, the reactions would need to be repeated to see if the micellar substrate were fully required and beneficial to the reaction outcome. The repeated reactions would use identical methodology but removing either surfactant or THF or both partners. This would also permit a greater understanding of the role in which the co-solvent plays in these reactions.

Table 12 - Control experiments in the absence of THF co-solvent, TPGS-750-M or both relative to optimised micellar conditions; product (DP), starting material (SM) or dehalogenation (DH).

Boronate		No THF			No TPGS-750-M			No TPGS-750-M or THF		
		% DP	% SM 50	% DH 52	% DP	% SM 50	% DH 52	% DP	% SM 50	% DH 52
a		-	-	-	95	0	5	-	-	-
b		93	0	0	86	3	8	87	10	3
c		98	0	0	95	0	4	-	-	-

d		91	0	0	97	0	3	-	-	-
e		95	0	0	94	0	6	-	-	-
f		99	0	0	100	0	0	-	-	-
g		90	0	0	91	0	9	-	-	-
h		99	0	0	97	0	3	-	-	-
i		-	0	0	98	0	2	-	-	-
j		100	0	0	100	0	0	-	-	-
k		99	0	0	93	0	7	-	-	-
l		96	0	0	86	0	14	-	-	-
m		61	3	0	78	21	1	38	62	0
n		77	0	4	82	0	18	-	-	-
o		-	-	-	59	40	1	-	-	-
p		77	0	1	79	0	8	-	-	-
q		97	0	2	91	0	4	-	-	-

r		98	0	0	89	0	9	-	-	-
Average		91	3	5	89	4	6	63	36	2

The result of these screens revealed that there was a slight decline in reactivity especially with certain reactants. Overall, when removing any of the solvent components the reaction still proceeded to some degree. Most of the reactions involving the phenyl substrates observed a small reduction in conversion. However, when using MIDA and pinacol esters a larger reduction was observed. This factor was especially true for pyrazole substrate **m** where a reduction was observed from 97% to <80% in all cases and with only 38% product observed when neither solvent was used. The overall averages were significantly lower than the optimised conditions of 98% average conversion rate. The data is also visualised to considerably affect certain substrates [Figure 72] where it appears that heterocycles could be most affected by the addition of micellar media and these optimised conditions. This is not a surprising result as these substrates have always been difficult to react, possibly due to a mix of deborylation events, the unreactive nature of the substrate, and solubility issues in aqueous solvents.

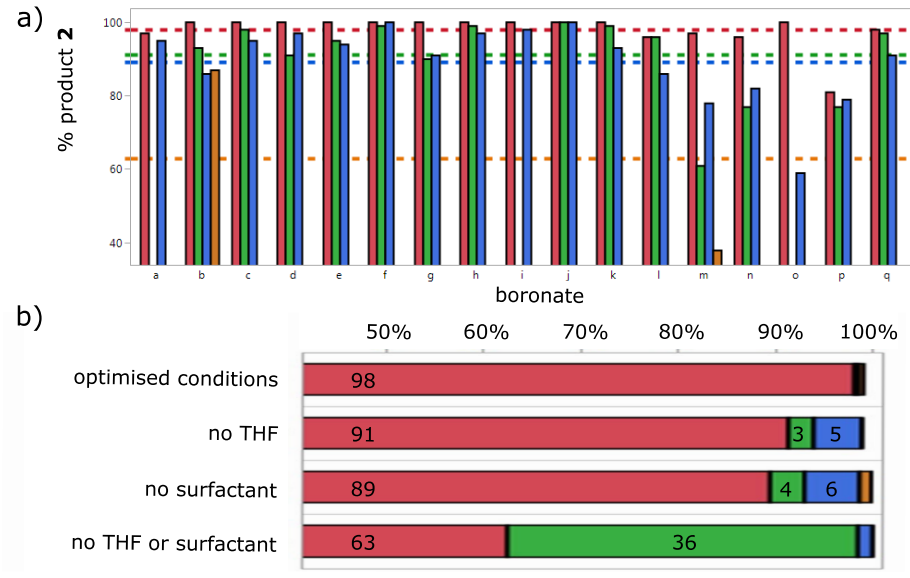


Figure 72 – Graphs depicting the promotion of the reaction by TPGS; a) formation of each substrate under optimised micellar conditions (red) compared to the reaction in the absence of THF (blue), TPGS (green) or both (orange); b) average formation of product (red) relative

to starting material (green), dehalogenated (blue) or other side products (orange) for the optimised micellar conditions compared to the absence of THF, TPGS or both.

Overall, these conditions performed extremely well on all substrates tested. It was also revealed that the micellar media has a great effect on the overall reaction and is required for the reaction to perform optimally. The conditions were optimised for two extremely difficult substrates which allowed conditions to be discovered that are compatible across a broad range of boronic acids, esters and MIDA esters. This also shows that the reaction is not limited in scope to specific types of boronic acid which greatly enhances its usability in a library format. The technique is also extremely simple to scale to a 96-well format using temperature-controlled microplates. This is useful when optimising a reaction used in library synthesis. Also, the work-up of these reactions is extremely simple using just an extraction with ethyl acetate, filtration, and ethanol precipitation. It is also anticipated that if used in a library build a further purification would be required such as molecular weight filtering or preparative HPLC to remove any organic impurities or excess palladium.

4.8 – Synthesis of 36 Member DEL

To further analyse the effect of micelles on DEL compatibility, a small library would be synthesised to evaluate the ability to PCR amplify and sequence the attached DNA code. To achieve this, a small two step test library was conceived utilising six building blocks in each stage. The library synthesis scheme consisted of an initial amide coupling of six arylhalide bearing carboxylic acids to a DNA conjugated amine, followed by Suzuki-Miyaura coupling with six different boronic acids. This test library would be synthesised using the same methods as a larger library, utilising full coding sequences for each building block, primers, and adapter sequences. The resulting library would then be subjected to a PCR amplification and next generation sequencing (NGS). The library would not be used to screen against a target protein as its sole function is as a proof of concept to identify whether micellar media, or the conditions used in the Suzuki-Miyaura reaction, compromise the ability to amplify and sequence the DNA. The resulting sequencing data would be expected to visualise 36 different DNA sequences for each individual library member.

4.8.1 – Synthesis of DNA Linker

The initial reaction step required a free amine conjugated to DNA. As a result, a linker was synthesised as this reactive handle for the initial building block addition. For this route to be

successful, the 6 arylhalide bearing carboxylic acids would first need to be coupled to the linker. An analogous PEG-4 amino linker was designed incorporating an acid at one terminus to couple with the C6 amino linker modifier on the DNA, and a protected amine at the other for coupling with the arylhalide bearing carboxylic acids as the diversifying first step.

Firstly PEG-4 was alkylated with *tert*-butyl bromoacetate to form the ester. The free hydroxyl was reacted with tosyl chloride to produce a leaving group for the displacement with sodium azide. The azide was reduced with triphenylphosphine to produce the amine which was hydrolysed under acidic conditions to remove the tertbutyl ester. Finally, an Fmoc protection was used to protect the amine. Fmoc was chosen as a protecting group in this case as it will readily deprotect under the same conditions used for cleavage of the DNA substrate after initial coupling to the solid supported DNA. This synthesis required just one column and one SCX silica purification and was high yielding throughout.

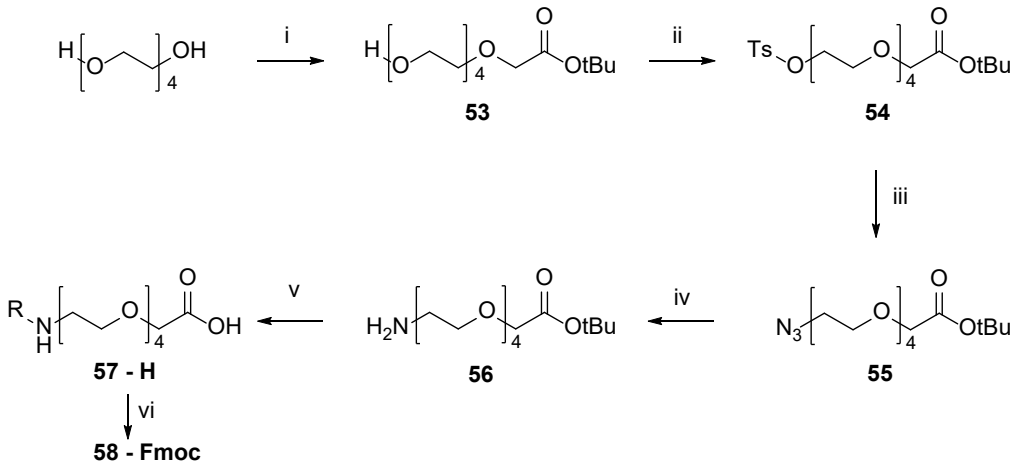


Figure 73 - synthesis of amino PEG linker. i) *t*Butylbromoacetate, K^{t}BuO, ^{t}BuOH, o/n 50°C; ii) Tosyl chloride, DCM TEA, o/n RT; iii) NaN_3, DMF, 80°C o/n; iv) PPh_3, THF, water, o/n RT; v) TFA, DCM, 6 hours RT; vi) FmocCl, $NaHCO_3$, dioxane, water, o/n RT.

The linker was then coupled with the solid supported DNA using the same conditions as previous examples. This method was reliable across all linkers conjugated, yielding 0.5-1 µmol DNA conjugate from 2 µmol of impure solid supported DNA. This method was also easily scalable allowing for larger scale reactions to be carried out.

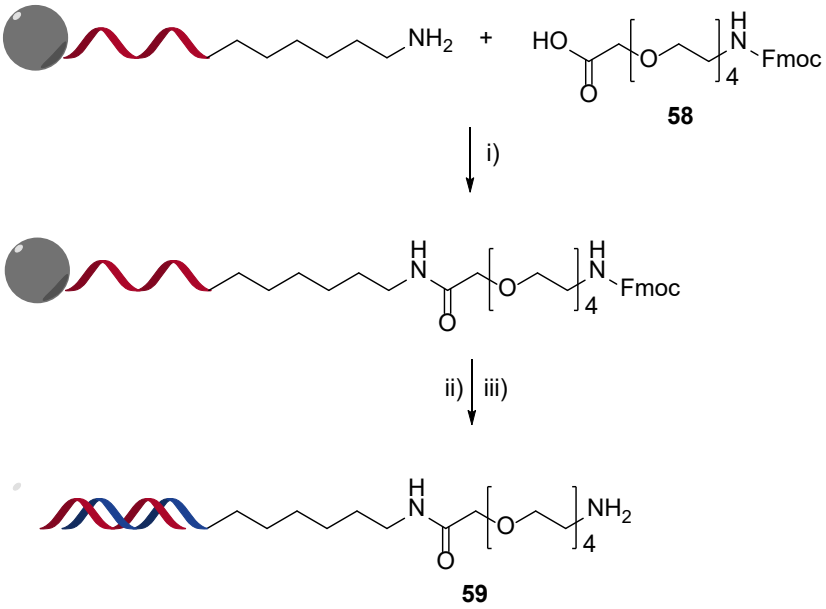


Figure 74 - Solid supported DNA conjugate synthesis; i) HATU, DMF, DIPEA; ii) Aqueous ammonia, aqueous methylamine; iii) 14mer single strand DNA, 40°C.

4.8.2 – Library Design

A vital part of the process is the design of the library. This often takes as long, if not longer, to carry out than the library build itself. As this library is the first one designed in house, a large amount of planning was required to design the library and carry out the appropriate DNA ligation to successfully synthesise the library. As this initial library was a concept for future library synthesis, it was important to use an encoding strategy that could be easily scaled for use in a much larger library.

As this library was a proof of concept, a variety of the boronates previously tested were chosen as substrates for the second reaction step. These included neutral, electron rich, and electron poor phenyl analogues, and three heterocycles including a MIDA boronate [Figure 75]. The library was designed to cover a diverse chemical space within this small library subset as a proof that no reagents or media would interfere with the amplification of the DNA or with the fidelity of the coding sequence. As well as six boronates, six arylhalide bearing carboxylic acids were chosen to also represent a range of likely coupling partners. These included phenyl, benzyl, and pyridyl scaffolds, with both bromo and iodo substituents.

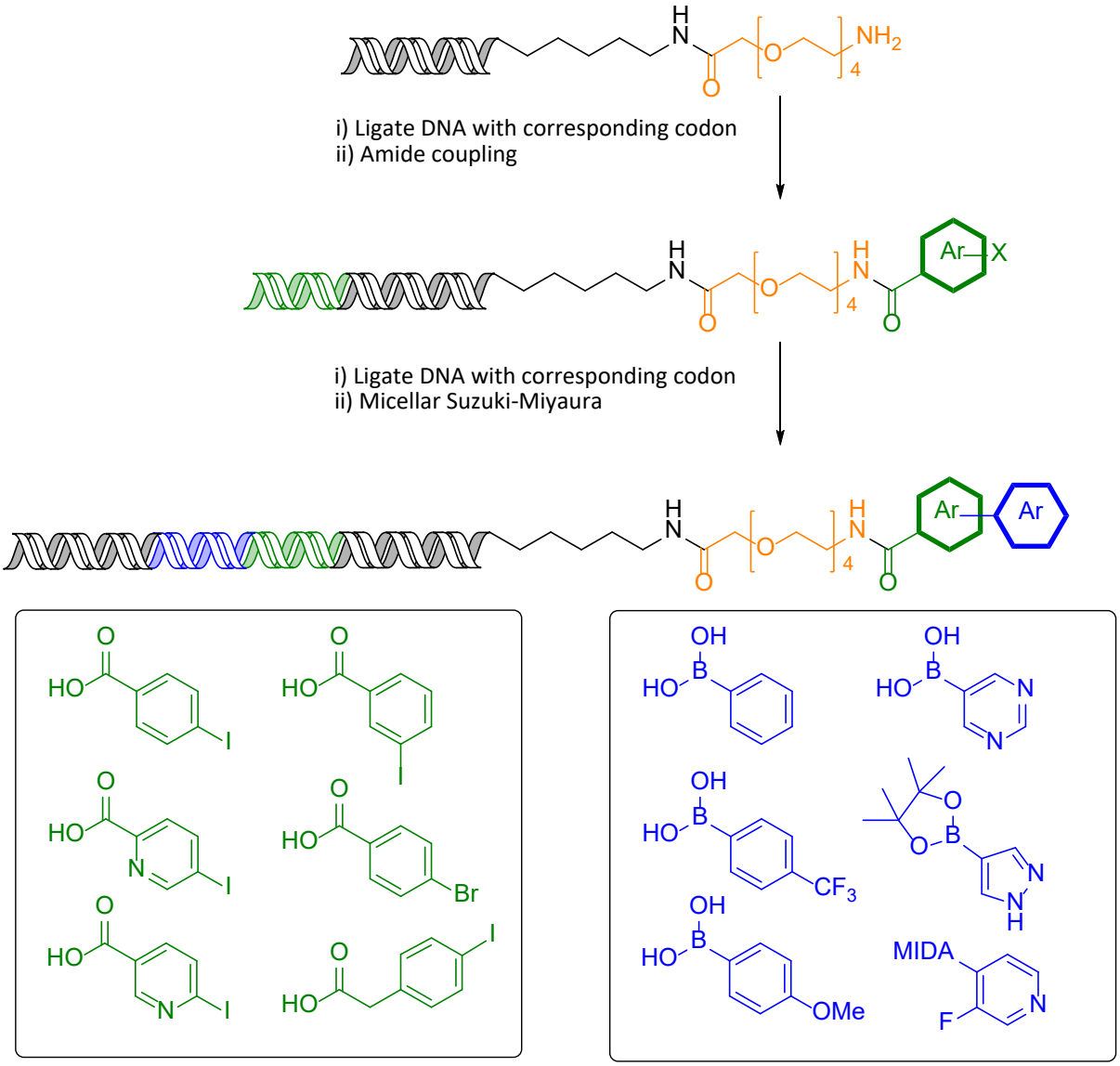


Figure 75 – Synthetic plan of the prototype 36-member library including depiction of coding and building blocks used.

The encoding strategy was based on one used previously and established to encode libraries of monolinked DNA strands similar to those used here.[41] The entire sequence is extended from the adapter sequence used in all chemistry to date. This requires the same solid supported method of building the linker onto the 14mer solid supported DNA strand and can be used to efficiently synthesise the initial linker and adapter sequence. Overall, 8 double stranded coding sequences would be used in this strategy to ligate and build the prototype library.

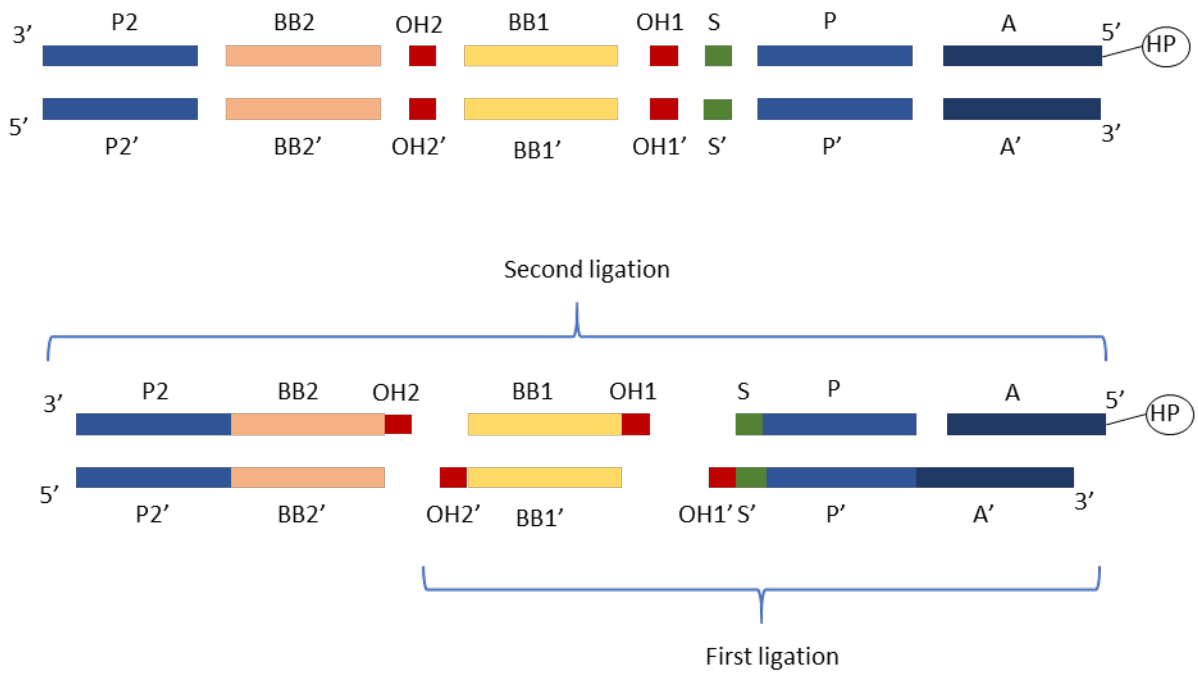


Figure 76 - Ligation and encoding strategy planned for use in the prototype 6x6 library.

The primer plays a highly important role in the entire library. During the PCR, a corresponding primer is added to the mixture which functions as a handle for the entire sequence to be duplicated and so primer choice is highly important. We chose to utilise a primer sequence used in literature for similar monolinked libraries.[41] These were implemented at each end of the encoding sequences and at one end would ligate to the adapter sequence to which the head piece is attached. A scaffold code is also often used as it contains an individual sequence that can be used to identify the library. This is highly important when screening multiple libraries at once as then the same building block codes can be used for all libraries, but each library has a clear identifier to then identify any hit compounds. In this example, a simple 3 base pair scaffold code was used with the sequence GCT as an identifier for the library.

Table 13 – Codes, functions, and corresponding sequence for each DNA section.

Code	Function	sequence 5'-3'
A	Adapter – 5’ aminolinked head piece	GTCTTGCCGAATTC
A'	Complimentary adapter	GAATTCGGCAAGAC
P	Primer	AGGTCGGTGTGAACGGATTTG
P'	Complementary primer	CAAATCCGTTCACACCGACCT
S	Scaffold code	GCT
S'	Complementary S	AGC
OH1	Ligation overhang 1	GTAT
OH1'	Complementary OH1	ATAC

BB1	Building block 1	xxxxxxxx
BB1'	Complementary BB1	xxxxxxxx
OH2	Ligation overhang 2	CCTA
OH2'	Complementary OH2	TAGG
BB2	Building block 2	xxxxxxxx
BB2'	Complementary BB2	xxxxxxxx
P2	Complementary to P2'	TGACCTCAACTACATGGTCTACA
P2'	Primer (reverse)	TGTAGACCATGTAGTTGAGGTCA

Building block codes are also highly important when designing libraries as these sequences are used to identify corresponding binding compounds. In this case, an 8 base pair sequence was used and attached to 4 base pair overhangs. The overhangs facilitate ligation reactions in the library synthesis and are called “sticky ends” as the two non-ligated sequences anneal via the overhanging sequence which decreases the chance of mismatched ligations. As the building block codes are the only part that distinguishes between each library member, the code requires careful consideration. An 8 base pair sequence for each building block codon was chosen in this example. This gives a huge variety of sequences to choose from as there are 65,536 unique sequences with 8 bases. However, when designing a DEL it is highly important not to contain mismatching of coding sequences as this would introduce false positive and negative results when screening the library.

Table 14 - Code, sequence and building blocks for respective building block sequences.

Code	BB1 Sequence 5’-3’	BB1' Sequence 5’-3’	Code	Building block
BB1a	ACTCTGGA	TCCAGAGT	BB1a'	4-Iodobenzoic acid
BB1b	GGTGTTAC	GTAACACC	BB1b'	3-Iodobenzoic acid
BB1c	AGTGTCGT	ACGACACT	BB1c'	5-Iodopicolinic acid
BB1d	GCTCTAAG	CTTAGAGC	BB1d'	4-Iodopyridine-2-carboxylic acid
BB1e	CATCGTGA	TCACGATG	BB1e'	4-Bromophenylacetic acid
BB1f	AATCGGTC	GACCGATT	BB1f'	4-Iodophenylacetic acid

Code	BB2 sequence	BB2' Sequence	Code	Building block
BB2a	ACTGGCTA	TAGCCAGT	BB2a'	Phenyl boronic acid
BB2b	GTTAGCCA	TGGCTAAC	BB2b'	4-Trifluormethylphenyl boronic acid
BB2c	TATCGCAG	CTGCGATA	BB2c'	4-Methoxyphenyl boronic acid
BB2d	CTTAGAGC	GCTCTAAG	BB2d'	5-pyrimidyl boronic acid
BB2e	GATCGACT	AGTCGATC	BB2e'	3-pyrazol boronic ester
BB2f	TCTGGAAC	GTTCCAGA	BB2f'	2-fluoro-4-pyridyl MIDA ester

Due to the importance of having no mismatching, the Hamming distance was used to assess the similarity between each coding sequence. For sequences of the same length, the Hamming distance measures the number of positions that are identical in the data set. It is anticipated that by choosing a Hamming distance of 4 across all coding sequences there will be no base pair mismatching in the library which significantly reduces the chance of false hits due to sequence mismatch. Generating DNA codes based on Hamming distance has been achieved using algorithms that produce lists of codes all of which satisfy the required Hamming distance and sequence length.[141] Using this information, 12 coding sequences were selected, all of which had 50% GC content to ensure similar annealing behaviour for each DNA strand in the library. Each of the building block sequences was also checked against its complimentary sequence to reduce any extra sequence pairing. As the library is small, this was a simple task. However, when increasing the size of the library further, it is an important part in sequence design to allow for minimal sequence mismatching. Therefore, this library was designed as though it was a larger library so that the sequences can be reused across multiple libraries. Finally, each of the building block codes were assigned a respective building block [Table 14] for the library.

4.8.3 – Test Ligations

For any library to be successful, ligations need to be highly resilient. Ligation techniques have been widely studied in biology, but when applying them to a DEL it is often found that the ligation is performed on a much larger scale. In this case the adapter sequence is separate from the primer sequence, either 2 or 3 ligations could be utilised to build the entire library. The adapter code could initially be ligated to the primer and scaffold sequences in a larger scale reaction. Then the first building block sequences could be ligated prior to the first building block addition. This same coding strategy could also be achieved by using a system where 5 individual oligonucleotide sequences are all ligated together in a single operation to build the entire sequence for the first building block in 1 step. The first method has the advantage because these types of ligations have been extremely well-documented. However, it requires an additional step of ligation, which not only takes longer, but also utilises more of the expensive enzymes required for phosphorylation and ligation reactions. This 5 sequence initial ligation is however documented on a 40 pmol scale in the literature.[41] This ligation sequence was used to trial reactions at both 40 pmol and 280 pmol scales, slightly modifying

it for the larger scale reaction. The reactions involved in the library build would be accomplished using 500 pmol of DNA in each well. Therefore, the ligations would need to be performed on a larger scale than those documented in current literature. When scaling these ligations, each reaction would be split over 2 x 280 nmol reactions and then pooled for subsequent work-up and chemical reaction.

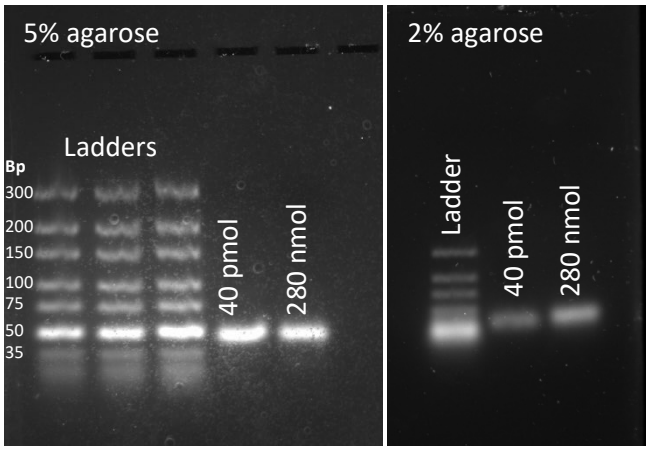


Figure 77 - Agarose gels representing first stage ligation of 40 pmol and 280 pmol reactions.

Initially, DNA requires phosphorylation of the 5' terminus of each sequence that is required in the ligation step. This was carried out using the previous methodology on a 40 pmol and 280 pmol scale. Phosphorylation was carried out by T4 polynucleotide kinase which catalyses the transfer of a gamma-phosphate from ATP to the free hydroxyl at the 5' end of DNA. This consumes ATP in the process which was therefore used in excess. This was a rapid process proceeding to completion at 37 °C in 20 mins, followed by temperature induced denaturing of the enzyme at 75 °C. The process was carried out on the strands coded as PS, OH1BB1, and OH1'S'P'A' as these sequences would be directly ligated at the 5' end. Once this stage was complete, the reaction mixture was transferred straight into the ligation step with no precipitation required, adding the non-phosphorylated adapter and OH2'BB1' sequences at the required concentrations. Ligations were then carried out the same concentrations as in literature but scaled up to 280 nmol as well as the presented 40 nmol. Ligations were carried out using T4 DNA ligase which catalyses the formation of a phosphodiester bond between a 5' phosphate and 3' hydroxyl termini of the DNA. The use of overhangs has a critical role here as it allows the strands to anneal prior to ligation taking place thus, reducing the chance of mismatches occurring and removing the need for more complex ligation systems. Once complete, a small sample of ligation mixture (0.1 µl diluted in 40 µl H_2O) was added to an

agarose gel [Figure 77] at both 2 and 5% agarose. As these DNA sequences are relatively short, 5% agarose was preferred as the method to run these samples as the bands in the ladder and product were easily separated. The base pair length of the adapter strand is 54 and the complimentary strand is 50. A major spot at around 50 bp in length was clearly visualised in the gels with a greater separation using 5% agarose. However, one major issue with using 5% agarose gels is that bubbles are prevalent when allowing the gel to set. This means that the bands are often rough, but this does not affect the overall quantification of the method. For subsequent work precast gels were used to overcome this issue.

4.8.4 – First library step

The initial library build started with a ligation step followed by the amide coupling reaction with six arylhalide bearing carboxylic acids and purification of the pooled library. The linker had been previously coupled to the modified adapter sequence at the 5' terminus at a large scale ready for the initial ligation. 616 pmol of each sequence that required 5' phosphorylation [Figure 78] were phosphorylated using the strategy outlined above. 280 pmol of each phosphorylated reaction mixture, adapter sequence and corresponding building block sequence were then added to PCR tubes for ligation.

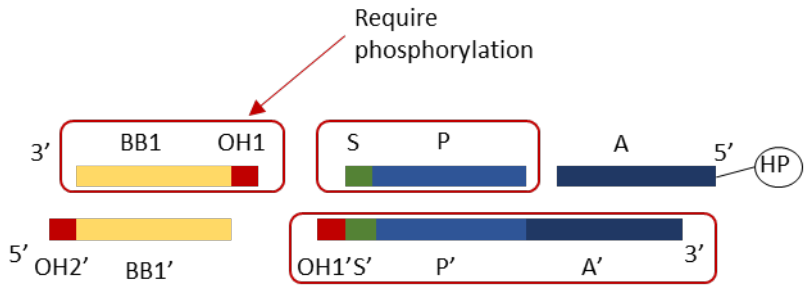


Figure 78 - Initial phosphorylation and ligation strategy for the 6x6 library.

Once the initial ligation stages were complete, a 5% agarose gel was run to determine the extent of ligated product [Figure 79]. This resulted in bands around 50 bp in length, with tiny amounts of shortmers present. This showed that the ligation was successful for each building block code. The reactions were ethanol precipitated using 10% volume 5M NaCl in water and 3 times the volume of cold ethanol and left at -78 °C for one hour. They were then centrifuged at 13500 rpm for 10 mins to form a pellet. Each of these pellets was dissolved in 72 µl of MOPS buffer ready for use in the initial amide coupling step.

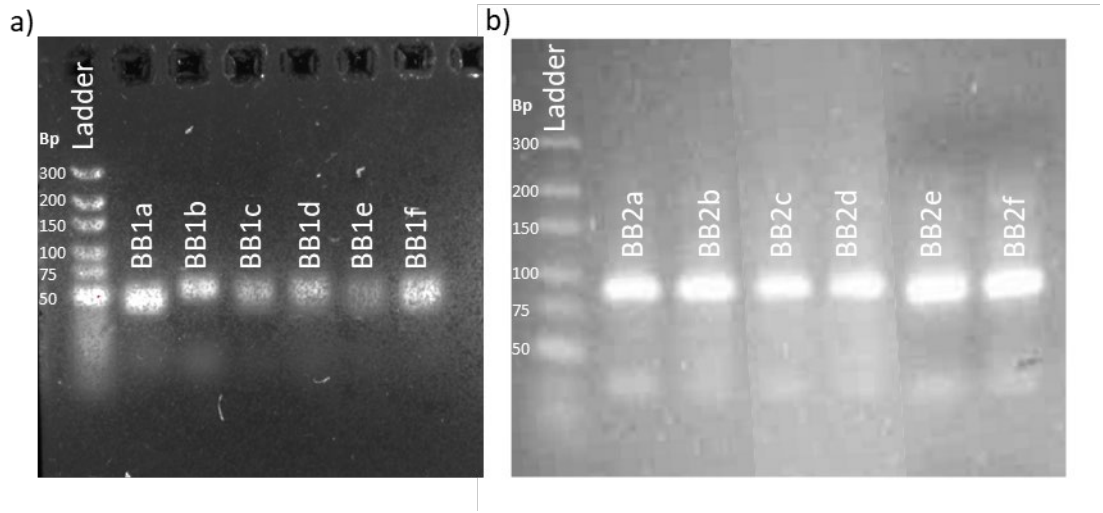


Figure 79 – a) 5% agarose gel indicating conversion for each ligation for BB1x; b) 5% agarose gel indicating conversion for each ligation for BB2x.

Amide couplings were performed using a general method from literature which was proven to work well with a broad range of acids.[32] In this study it was found that the conditions were much higher yielding than any used previously with coupling reagents such as DMT-MM which has good aqueously solubility. This technique utilised EDC.HCl and HOAT as the coupling reagents. The reaction was carried out in a mixture of MOPS buffer with DMSO as a co-solvent. Reagents were dissolved in DMSO prior to addition to the reaction which eases the process of split and pool chemistry as all reagents can be added simultaneously, promptly, and effortlessly with no weighing required. The reaction was carried out overnight at room temperature after which a further aliquot of EDC.HCl and HOAT was added and left for a further 6 hours. The reactions were then quenched with ammonium acetate and ethanol precipitated as described above.

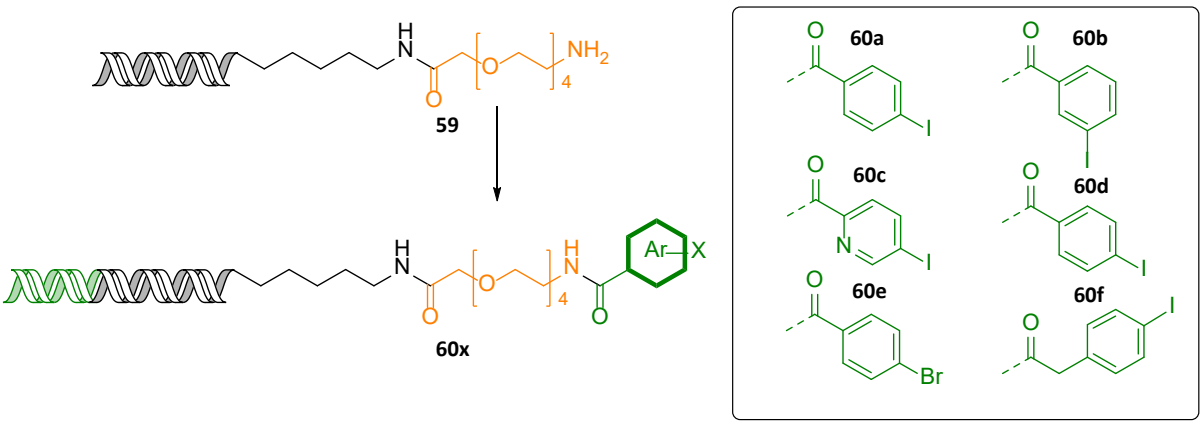


Figure 80 – Amide coupling products formed from the first step of the prototype library.

The pellets were then dissolved in water, combined and a second ethanol precipitation was carried out. The overall pooled library was dissolved in water ready for the second step of the library synthesis.

4.8.5 – Second library step

The entire pooled library was split over 6 wells for phosphorylation. The whole library needed to be phosphorylated in this way as the 5’ terminus of the complimentary strand is ligated in the second phase of the build. The 5’ terminus of the linker strand was already modified with the linker and so cannot undergo phosphorylation [Figure 81]. The same procedure was carried out as previously for both phosphorylation and ligation reactions. The resulting ligation products were precipitated and dissolved in water ready for the Suzuki-Miyaura coupling step. This procedure yields individual wells of 6 different encoded first stage DNA conjugates, ready sequenced with the second building block codes and primer sequences now in place.

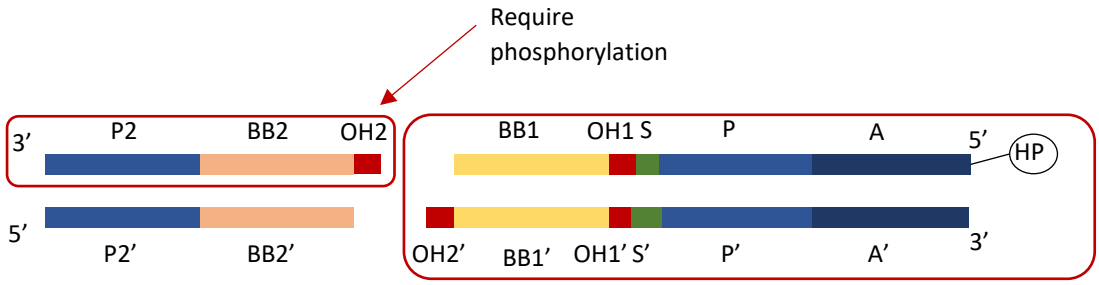


Figure 81 - Second phosphorylation and ligation strategy for the 6x6 library.

The final chemical synthesis step for the library build was to carry out the micellar Suzuki couplings previously optimised [Figure 82]. The 6 pellets from the ligation step were dissolved in water and split into a 96-well sealable microplate. The procedure used was analogous to those used in optimisation; the only difference was the amount of DNA used in each well. As the DNA is in a huge dearth compared to the reagents, lowering this concentration but not the other reagents was not considered detrimental to the overall reaction. This was further proven when optimising amide coupling conditions in a future chapter. These reactions were

carried out over a 5 hour period at 60 °C. Once complete the reactions were combined into one microcentrifuge tube, washed with chloroform three times, filtered, and precipitated.

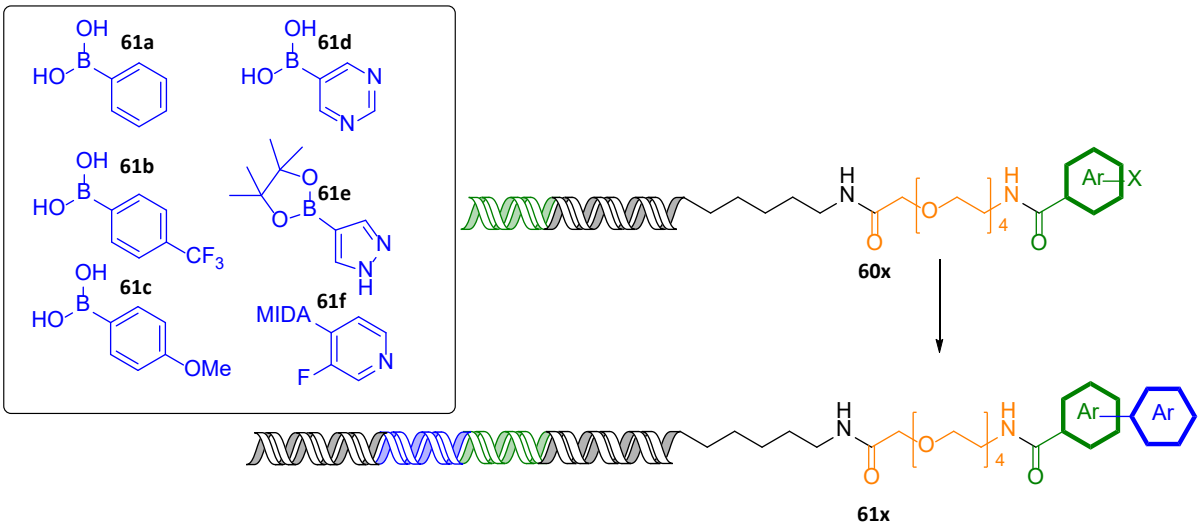


Figure 82 - Final ligation and Suzuki coupling step carried out in the formation of the 36-member library.

Once the library had been finished, the next stage was to PCR amplify and sequence the library. This would finalise this novel reaction media as DNA compatible and of value for synthesising a DNA encoded library.

4.8.6 – Amplification and Sequencing

To prove that the sequence was still intact, and the strands were able to be decoded via NGS, an aliquot of the synthesised library was PCR amplified. For this process to work successfully, elongation of the DNA strand is required. This elongation sequence is to anchor the individual DNA strand to a glass bead to initiate NGS sequencing. Elongation can be carried out when amplifying the library using PCR by the addition of a forward and reverse primer with the elongation sequence already coded at each end [Table 15]. The original library has an overall bp length of 85 and, when extended by these NGS elongation sequences, the PCR product would be ~140 BP in length.

Table 15 - Forward and reverse primers for PCR and NGS analysis. Primer sequence shown in blue, NGS elongation sequence shown in purple.

Primer	sequence 5'-3'	length
Forward primer	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTAGACCATGTAGTTGAGGTCA	56
Reverse primer	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGTCGGTGTGAACGGATTTG	53

PCR was carried out in 50 µl reaction mixtures using 100 ng of the library and 10 mM of each primer. Negative controls were carried out as distilled water in place of the template and single and double primer controls [Figure 83]. The gel showed a band at around 120 base pairs in length for the result of the PCR including both primers and library. In both negative controls there was no elongation or amplification present as expected. The band was then extracted from the gel and submitted for NGS.

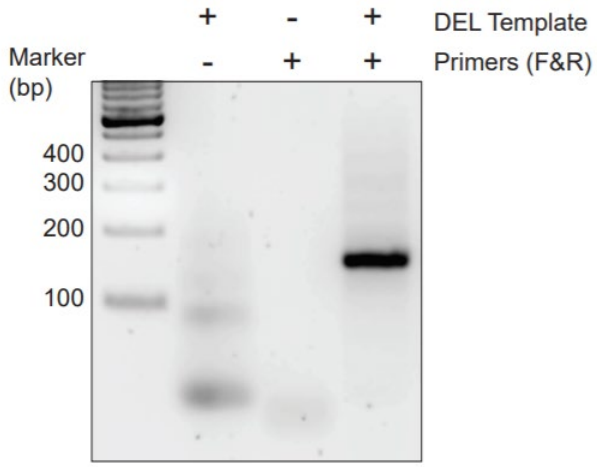


Figure 83 - Gel after PCR has been carried on the 6x6 concept library. The large band at ~140 bp that is consistent with the elongation sequence addition to the library.

Results from NGS showed that all expected sequences could be identified in the results [Table 16 and Appendix Table 1]. These results showed the linker strand DNA sequence after the NGS elongation sequence was removed. As can be seen, all 36 DNA sequences are present in the data with primers (blue), scaffold code (green), overhangs (red) and building block data sequences (black) all visible in the results. Realising all 6 of each round building block codes can be sequenced; this proves that this novel micellar methodology is compatible with the sequence interpretation of the encoded library. These results are highly encouraging and show that micellar media could be a highly important new tool for DNA encoded library synthesis.

Table 16 – Results of NGS of the amplified prototype library showing the linker strand.

Code (Linker strand) 5'-3'	Counts	BB1	BB2
AGGTCGGTGTGAACGGATTTGGCTGTATAGTGTCGTCCTATATCGCAGTGACCTCAACTACATGGTCTACA	8339	c	c
AGGTCGGTGTGAACGGATTTGGCTGTATAGTGTCGTCCTAGTTAGCCATGACCTCAACTACATGGTCTACA	7925	c	b
AGGTCGGTGTGAACGGATTTGGCTGTATCATCGTGACCTATATCGCAGTGACCTCAACTACATGGTCTACA	7220	e	c
AGGTCGGTGTGAACGGATTTGGCTGTATCATCGTGACCTAGTTAGCCATGACCTCAACTACATGGTCTACA	6548	e	b
AGGTCGGTGTGAACGGATTTGGCTGTATAATCGGTCCCTATATCGCAGTGACCTCAACTACATGGTCTACA	6192	f	c
AGGTCGGTGTGAACGGATTTGGCTGTATGGTGTTACCCTATATCGCAGTGACCTCAACTACATGGTCTACA	6011	b	c
AGGTCGGTGTGAACGGATTTGGCTGTATAATCGGTCCCTAGTTAGCCATGACCTCAACTACATGGTCTACA	5742	f	b
AGGTCGGTGTGAACGGATTTGGCTGTATGGTGTTACCCTAGTTAGCCATGACCTCAACTACATGGTCTACA	5041	b	b
AGGTCGGTGTGAACGGATTTGGCTGTATAGTGTCGTCCTAACTGGCTATGACCTCAACTACATGGTCTACA	4732	c	a
AGGTCGGTGTGAACGGATTTGGCTGTATACTCTGGACCTAGTTAGCCATGACCTCAACTACATGGTCTACA	4650	a	b
AGGTCGGTGTGAACGGATTTGGCTGTATACTCTGGACCTATATCGCAGTGACCTCAACTACATGGTCTACA	4361	a	c
AGGTCGGTGTGAACGGATTTGGCTGTATCATCGTGACCTAACTGGCTATGACCTCAACTACATGGTCTACA	4257	e	a
AGGTCGGTGTGAACGGATTTGGCTGTATGCTCTAAGCCTATATCGCAGTGACCTCAACTACATGGTCTACA	4197	d	c
AGGTCGGTGTGAACGGATTTGGCTGTATGCTCTAAGCCTAGTTAGCCATGACCTCAACTACATGGTCTACA	4169	d	b
AGGTCGGTGTGAACGGATTTGGCTGTATAATCGGTCCCTAACTGGCTATGACCTCAACTACATGGTCTACA	3554	f	a
AGGTCGGTGTGAACGGATTTGGCTGTATGGTGTTACCCTAACTGGCTATGACCTCAACTACATGGTCTACA	3189	b	a
AGGTCGGTGTGAACGGATTTGGCTGTATACTCTGGACCTAACTGGCTATGACCTCAACTACATGGTCTACA	2644	a	a
AGGTCGGTGTGAACGGATTTGGCTGTATGCTCTAAGCCTAACTGGCTATGACCTCAACTACATGGTCTACA	2403	d	a
AGGTCGGTGTGAACGGATTTGGCTGTATCATCGTGACCTAGATCGACTTGACCTCAACTACATGGTCTACA	1996	e	e
AGGTCGGTGTGAACGGATTTGGCTGTATAGTGTCGTCCTAGATCGACTTGACCTCAACTACATGGTCTACA	1660	c	e
AGGTCGGTGTGAACGGATTTGGCTGTATGGTGTTACCCTAGATCGACTTGACCTCAACTACATGGTCTACA	1377	b	e
AGGTCGGTGTGAACGGATTTGGCTGTATAATCGGTCCCTAGATCGACTTGACCTCAACTACATGGTCTACA	1249	f	e
AGGTCGGTGTGAACGGATTTGGCTGTATACTCTGGACCTAGATCGACTTGACCTCAACTACATGGTCTACA	969	a	e
AGGTCGGTGTGAACGGATTTGGCTGTATGCTCTAAGCCTAGATCGACTTGACCTCAACTACATGGTCTACA	944	d	e
AGGTCGGTGTGAACGGATTTGGCTGTATAGTGTCGTCCTATCTGGAACTGACCTCAACTACATGGTCTACA	448	c	f
AGGTCGGTGTGAACGGATTTGGCTGTATCATCGTGACCTATCTGGAACTGACCTCAACTACATGGTCTACA	400	e	f
AGGTCGGTGTGAACGGATTTGGCTGTATAGTGTCGTCCTACTTAGAGCTGACCTCAACTACATGGTCTACA	396	c	d
AGGTCGGTGTGAACGGATTTGGCTGTATCATCGTGACCTACTTAGAGCTGACCTCAACTACATGGTCTACA	381	e	d
AGGTCGGTGTGAACGGATTTGGCTGTATGGTGTTACCCTACTTAGAGCTGACCTCAACTACATGGTCTACA	356	b	d
AGGTCGGTGTGAACGGATTTGGCTGTATGGTGTTACCCTATCTGGAACTGACCTCAACTACATGGTCTACA	307	b	f
AGGTCGGTGTGAACGGATTTGGCTGTATAATCGGTCCCTATCTGGAACTGACCTCAACTACATGGTCTACA	295	f	f
AGGTCGGTGTGAACGGATTTGGCTGTATACTCTGGACCTATCTGGAACTGACCTCAACTACATGGTCTACA	226	a	f
AGGTCGGTGTGAACGGATTTGGCTGTATAATCGGTCCCTACTTAGAGCTGACCTCAACTACATGGTCTACA	224	f	d
AGGTCGGTGTGAACGGATTTGGCTGTATGCTCTAAGCCTATCTGGAACTGACCTCAACTACATGGTCTACA	206	d	f
AGGTCGGTGTGAACGGATTTGGCTGTATACTCTGGACCTACTTAGAGCTGACCTCAACTACATGGTCTACA	195	a	d
AGGTCGGTGTGAACGGATTTGGCTGTATGCTCTAAGCCTACTTAGAGCTGACCTCAACTACATGGTCTACA	194	d	d

4.9 – Expanded Reactant Scope

As this reaction has been successfully carried out on simple iodo headpieces, an expansion of the reaction scope was accomplished utilising some compounds synthesised in collaboration with Dr Andreas Brunschweiger. The Povarov and Castagnoli-Cushman multicomponent

reactions were utilised to synthesis a variety of different structures [Figure 84]. These were synthesised using a solid supported reaction and then cleaved from the bead and used as a single stranded reactant in the Suzuki coupling procedure.

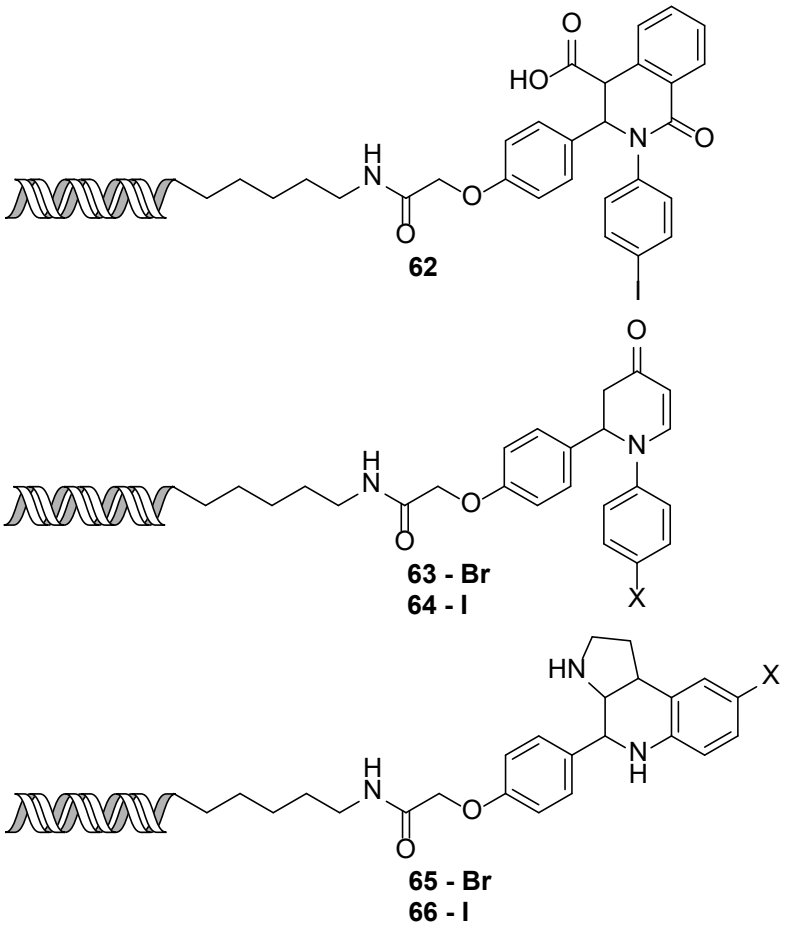


Figure 84 – Compounds providing chemically diverse halogenated components produced using Castagnoli–Cushman and Povarov multicomponent reactions.

These reactants also contained some interesting chemical diversity and contained different chemical moieties such as an acid, secondary amine, aniline, and a pseudo Michael acceptor. These diverse cores are interesting and privileged structures containing heterocyclic parts which could become interesting library members. These compounds also had bromo as well as iodo substituents further testing the scope of the Suzuki coupling procedure.

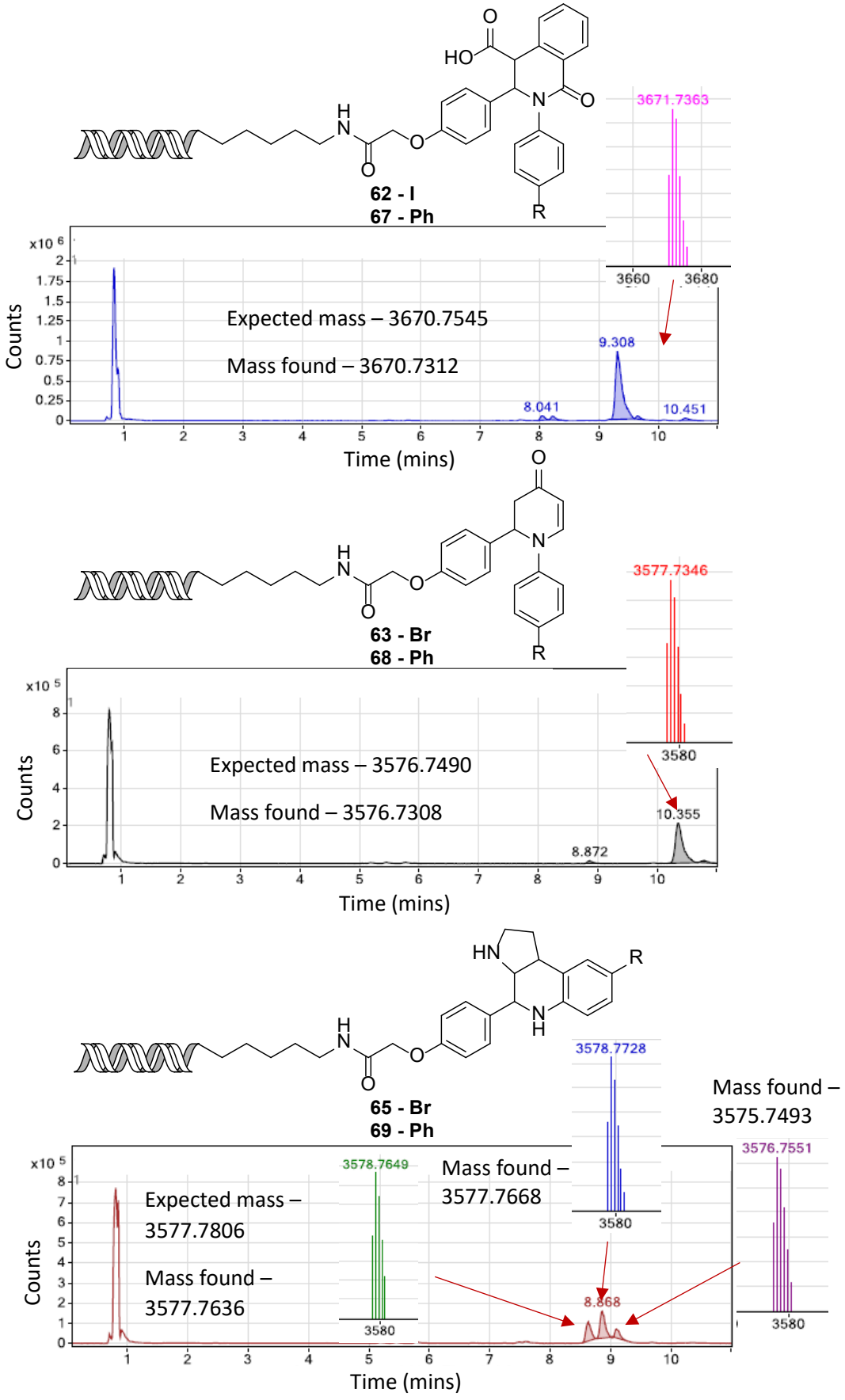


Figure 85 – Chromatograms depicting the products from chemically diverse starting materials.

These single stranded coupling partners were analysed by mass spectrometry prior to coupling [Appendix Figure 20]. It was found that compounds **62**-**64** were pure with no degradation of DNA or the attached organic compound visible. However, compounds **65** and **66** showed 3 peaks in the chromatograms. These were attributed to the stereoisomers of the central core, but the third peak showed a mass 2 daltons lower than expected. This peak was the smallest of the three and was attributed to an oxidation event occurring on the core at either one of the secondary amines present, either in the synthesis or storage of the compound.

The Suzuki-Miyaura couplings were carried out under the standard optimised conditions utilising 2 nmol of the single stranded coupling partners. The results of this study were highly encouraging once again showing full conversion to the desired product when phenyl boronic acid was used [Figure 85]. The reactions were also extremely efficient and clean at this scale, showing only product formation with no halogenated product or starting material present. There were a couple of tiny impurities but these were also present in the starting materials. The interesting case was that reactions with compound **65** and **66** also showed three distinct products in the chromatogram. These were again attributed to the stereoisomers and oxidised core of the compound. But it was highly encouraging to see that this oxidation was not worsened under the conditions of the Suzuki reaction showing how mild and chemically tolerated they are.

4.10 – Conclusions

Micellar media have been utilised in a significant number of chemical process and reactions to date. These reactions have often shown high yields, low temperatures, and encouragingly low E factors due to the recyclability and aqueous media used. The micelles used in many of these examples are commercially available which makes them highly accessible. These micelles have never been used for chemistry on a DNA conjugated molecule. As an initial proof of concept, a Suzuki-Miyaura cross-coupling reaction was chosen due to the lack of high-yielding conditions available. Transition metal catalysed reactions are also shown to be a highly desirable reaction in library build as they allow access to carbon-carbon bond forming reactions. Current conditions used in DEL synthesis for cross-coupling reactions are not high-yielding and often fail to couple the more privileged heteroaromatic partners.

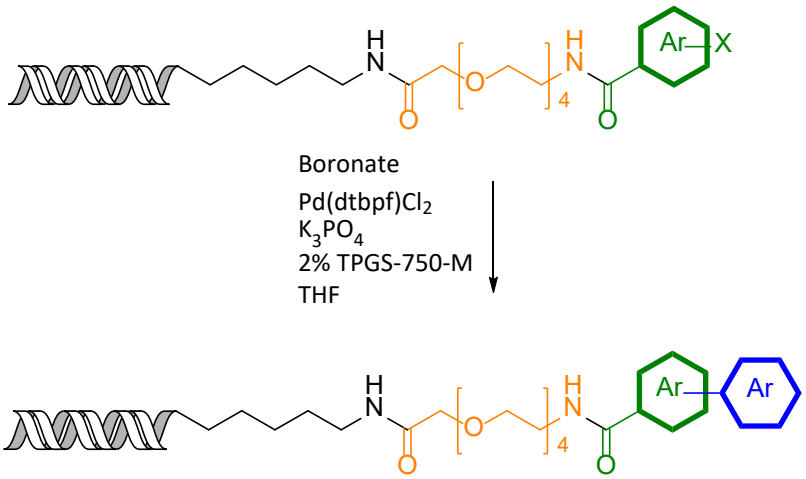


Figure 86 - optimised Suzuki coupling conditions using micellar media.

A study was performed that utilised an optimisation approach used in industry by Merck, followed by further optimisation by FED to produce an extremely high-yielding efficient reaction which tolerates a multitude of different functionality. The reaction was then performed on a small-scale prototype library to create a 36-member DEL. This library was then amplified by PCR and sequenced using NGS, successfully visualizing all 36 members and the complimentary DNA sequence. The reaction was a huge success allowing it to be used in future DNA encoded library builds, or for other DNA conjugate synthesis as it allows access to a huge variety of coupling partners whilst maintaining the fidelity of the attached DNA strand. It is proposed that this technology could be extended to many other organic reactions such as S_NAr, amide, and Buchwald, Heck, Sonogashira couplings thus further improving the scope of DNA compatible chemistry.

Chapter 5 – Micellar Promoted Amide Couplings

Micellar technology was successfully applied to Suzuki-Miyaura couplings, but this technique could also be applied to many other reactions increasing the scope of compatible DEL chemistry. Currently a huge variety of conditions are accessible for DNA compatible amide couplings. However, all are still inefficient and unreliable for coupling certain building blocks. Also, if the acid component is conjugated to a DNA strand instead of the amine the scope of the reaction is further reduced.[24] Solid supported couplings have also been proposed.[70] However, these techniques rely upon methods yet to be applied to a large scale split and pool library synthesis and are still unreliable at reacting certain classes of building block. One of the advantages of using micellar media is that the reaction is no different from any other solution phase chemistry and so does not require specialised equipment and can be carried out in standard reaction vessels.

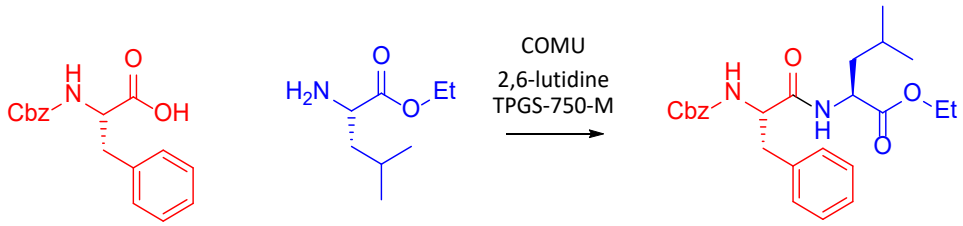


Figure 87 - Depiction of amide coupling conditions utilising micelles.

Amide bond formations are used at least once in a huge number of libraries. Amide bonds also form the backbone of many natural products including peptides, and thereby are highly sought after and utilised in DELs and medicinal chemistry. The initial building block addition in a DEL is often an amide coupling with the free amine conjugated to the DNA strand. Increasing the scope and reactivity of this reaction could have a large impact on the diversity of resulting encoded libraries. Currently peptides have been studied using micellar media and applied to standard organic chemistry.[123-125] In these examples it was discovered that using the new generation of coupling agent COMU, was highly successful in converting all reactions run under micellar conditions. COMU is derived from the peptide coupling additive oxyma [Figure 88] and a urea is formed from the coupling mechanism, as is the case for more commonly used coupling agents such as HATU. However, this amide coupling partner is less explosive and more soluble than its HATU counterpart. It was observed to be more reactive than HATU in standard organic reactions using micelles but its application to DNA conjugated chemistry has yet to be reported.

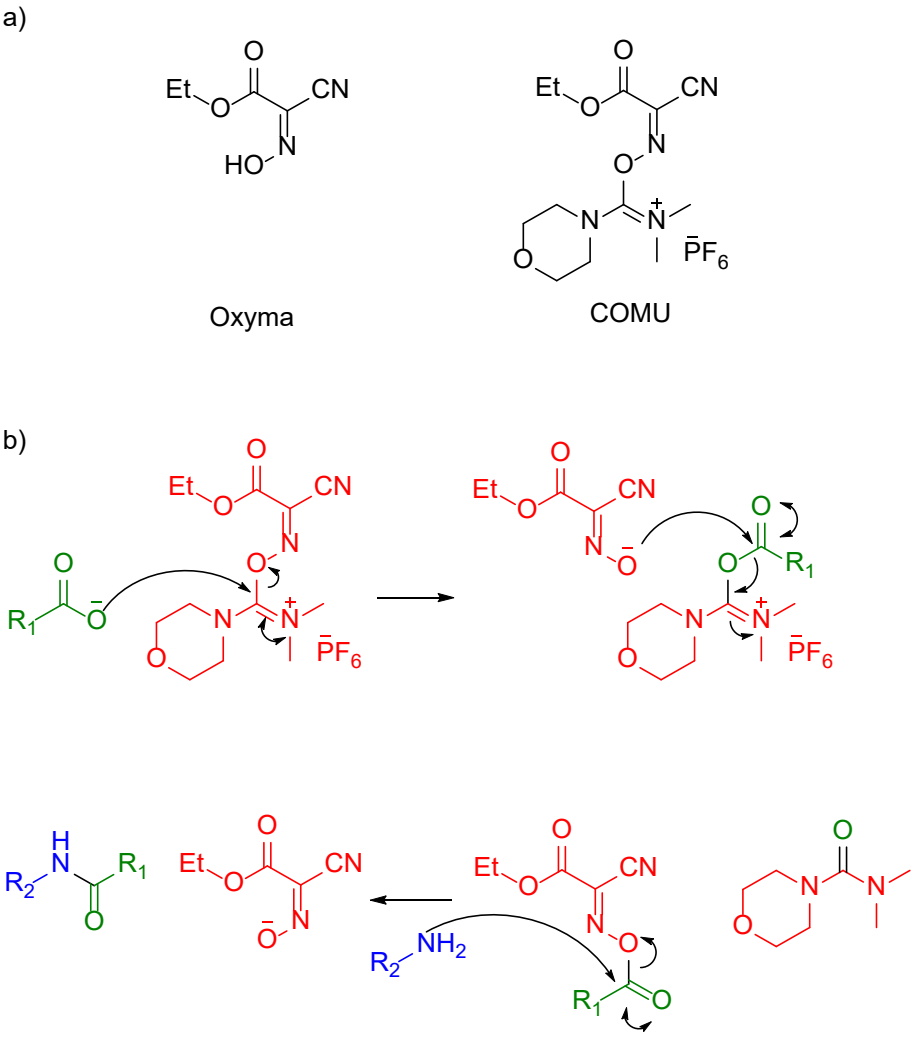


Figure 88 – a) Oxyma and COMU; b) Mechanism of amide coupling using COMU.

5.1 – Initial Couplings

An acid conjugated to a DNA strand was chosen for initial proof of concept as this amide coupling has been shown to be more challenging for DNA encoded reactions. This is possibly due to the large excess of water and hydroxide that is present in DNA compatible reactions. Once the coupling intermediate has been synthesised with any coupling partner it is highly reactive to nucleophiles. This allows water to react with the activated acid, reverting it to the acid starting material. This process would critically reduce the rate of reaction as the amine would have to compete with water to consistently react to form the desired product. This suggests that less reactive amines are highly unlikely to react at any extent, extremely limiting the reaction scope. It was envisioned that by using micelles the activated acid intermediate would form in the presence of the lipophilic part of the micelle and so would be slower to revert to the acid due to the lack of water present. Also, the amine would be at a high

concentration inside the micelle and so could react rapidly before this competing reaction takes place. It was also envisioned that any reaction that successfully converted this amide reaction, could also be applied to the converse reaction in which the amine is attached to the DNA strand.

Initial couplings were carried out using PEG linked substrate 70 [Figure 89] analogous to the linkers used for the Suzuki couplings. This linker was synthesised from commercially available 3,6,9,12,15-pentaoxaheptadecanedioic acid using the previously used successful solid supported method using HATU and DMF, followed by cleavage from the bead under basic conditions. When cleavage from the bead occurred, initially a methylamide or amide side product was visible. This was postulated to be due to the coupling partner forming the reactive intermediate on the terminal acid after coupling to DNA had occurred which then underwent a amide coupling to from these side products with methylamine and ammonia used in the bead cleavage stage. Prior to cleavage the solid support was shaken in water for 1 hour to hydrolyse the active ester which gave full conversion to the desired acid product. This allowed access to the acid attached to double stranded 14mer in preparation of amide coupling trials.

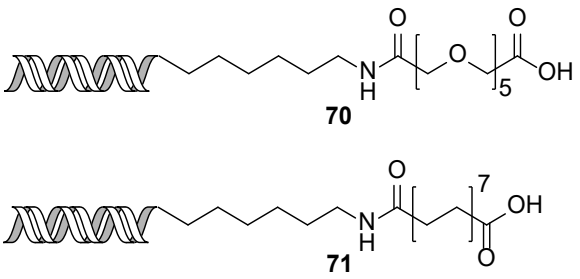


Figure 89 - Structures of PEG linked acid 70 and C14 linked acid 71.

Initial couplings were trialled using glycine ethyl ester, aniline, 2-aminothiazole and 2-aminoimidazole. However, when using this linker and both HATU and COMU in 2% TPGS-750-M there was no conversion to the desired product in any reaction. This was an unexpected outcome, as it was expected that these reactions would benefit from the increased concentration and reactivity of the reactants inside the micellar media, and some degree of conversion would be visible replicating modified literature micellar conditions. However, unlike the Suzuki couplings the reactive species in this case contains an acid directly conjugated to a PEG chain and DNA strand. It was envisioned that the acid was hydrophilic and so would be highly unlikely to enter the hydrophobic micellar core and therefore be less

likely to react under micellar conditions as it would rarely come into proximity of the reactive coupling species. Using a hydrophobic linker was hypothesised to increase the reactivity of the overall system by increasing the lipophilicity of the linker chain and increasing the probability of its access into the micelle. As the bulk aqueous phase is water, and the PEG chain is hydrophilic, any activated acid species formed in the reaction with HATU will revert to the acid rapidly due to the DNA conjugate lying majorly within the aqueous phase. So using a more hydrophobic linker would simultaneously enable mixing of the DNA with the micelle and halt the side reaction reverting the activated species back to the acid taking place.

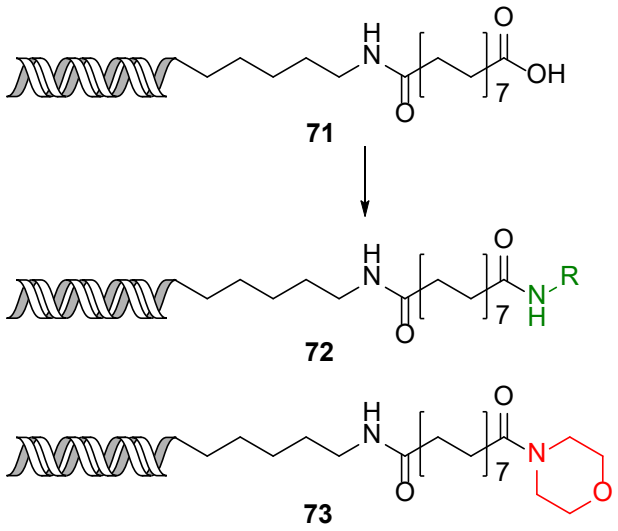
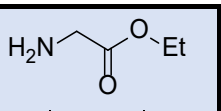
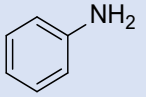
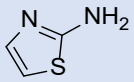
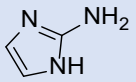


Figure 90 – Amide coupling of acid 71, producing desired product 72 and side product 73.

Initial results using the C14 linker were encouraging [Table 17][Appendix Figure 36]. When using micellar reactions it is important to consider the overall concentrations of the reaction media. Often reactions will be carried out in at least 0.5M concentrations of reactants with higher concentrations of base. In these examples the coupling agent and the amine were all at 0.5M and the base either at 1.2M or 1.5M to maintain an excess especially as many amines are HCl salts. Coupling with HATU showed a clear advantage over the other agents where high conversions were seen with glycine ethyl ester and aniline. Also, modest conversions were seen for both 2-aminothiozole and 2-aminoimidazole using HATU. Interestingly, COMU was a less successful coupling agent than HATU in these examples. A side product was being formed when using COMU, especially with less reactive heterocycles. This side product was discovered to be the morpholine amide **73**, which is possibly formed from morpholine either as a low-level impurity in the COMU, or an unknown side reaction occurring that incorporates morpholine from the COMU itself. If morpholine was present, or formed in this reaction, it

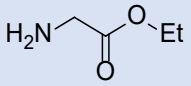
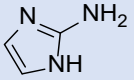
would likely outcompete the less reactive heterocycles suggesting why more of this side product was observed with the less reactive substrates.

Table 17 – Initial amide coupling conditions; **71** (4 nmol), DIPEA (1.2M) or 2,6-lutidine (1.5M), coupling agent (0.5M), 30 µL total volume, 16 hr, 40 °C. *other non-identified side product.

Coupling Agent	**Base**												
		% DP 72	**% SM 71**	**% SP 73**	**% DP 72**	**% SM 71**	**% SP 73**	**% DP 72**	**% SM 71**	**% SP 73**	**% DP 72**	**% SM 71**	**% SP 73**
HATU	DIPEA	90	10	0	10	90	0	90	10	0	10	90	0
EDC, HOAT	DIPEA	0	100	0	0	100	0	0	100	0	-	-	-
DMT-MM	DIPEA	0	100	0	0	100	0	0	100	0	-	-	-
HATU	2,6-lutidine	97	3	0	92	8	0	56	9	33 *	42	58	0
COMU	2,6-lutidine	61	40	0	9	79	12	0	4	96	0	31	69

As COMU was less successful, HATU was carried forwards as the favoured coupling reagent. Base appeared to make a difference in the test reactions showing that 2,6-lutidine was more successful than DIPEA in this reaction. Therefore, performing a base screen would be the first optimisation strategy for this reaction. There are several organic bases which can be chosen for this reaction and all were trialled at 1.5M concentrations. Glycine ethyl ester and aminoimidazole were both chosen as the coupling partners to represent reactive and less reactive amines. This provides a means to determine the most generally applicable coupling conditions across a range of amines. Interestingly pyridyl bases were most effective in this reaction and 2,6-lutidine was best for both glycine ethyl ester and 2-aminoimidazole. Higher conversion rates were seen for glycine across all bases with many showing conversions of greater than 90%. However, the less reactive heterocycle shows that 2,6-lutidine is by far the best base for this reaction.

Table 18 - Base screen using glycine ethyl ester and 2-aminoimidazole; conditions: **71** (4 nmol), amine (0.5M), base (1.5M), HATU (0.5M), 30 µL total volume, 16 hr, 40 °C.

Base						
	% DP	% SM 71	% SP	% DP	% SM 71	% SP
DIPEA	90	10	0	10	90	0
2,6-Lutidine	97	3	0	42	58	0
N-methylmorpholine	72	28	0	0	100	0
TEA	73	27	0	6	94	0
2,6-ditertbutyl-4-methylpyridine	93	1	6	2	98	0
NMI	87	9	4	27	41	32
Pyridine	95	1	4	21	79	0
DMAP	17	83	0	0	100	0
4-methoxypyridine	93	3	4	27	73	0
Quinoline	95	2	3	4	96	0
DABCO	64	31	5	0	100	0
4-cyanopyridine	48	51	0	0	100	0

Initial success using micelles and this hydrophobic linker has shown that considerable improvement is required for the reaction to be compatible with multiple reagents, especially the less reactive but medicinally desirable heterocycles. The change to the carbon linker improved the rection but showed significant substrate variation. It was anticipated that further optimisations could be made by using a FED. This technique was highly successful when used to improve the reaction scope of the micellar Suzuki-Miyaura reaction and was hoped to have a similar effect in improving the substrate scope of the amide coupling.

5.2 – FED Experiment

To further investigate the scope of the reaction an FED was designed to study the effect of different parameters on the reaction. An FED methodology has been previously used to optimise Suzuki coupling reactions which was found to contain second order effects between variables. These relationships would be difficulty to discover under more general optimisation techniques. The application of an FED could allow a significant understanding of the reaction and the effect of each parameter. The parameters chosen in this experiment were

temperature (40-60 °C), surfactant strength (2-5%) and base concentration (0.5 – 1.5 M) for glycine ethyl ester and 2-aminoimidazole.

Table 19 - Factorial experimental design assessing conversion for glycine ethyl ester and 2-aminoimidazole.

Amine	Temp. (°C)	TPGS-750-M (%)	2,6-lutidine (M)	Product 72	Starting material 71	Side product	Modelled conversion
Imidazole	40	2	1	43	54	4	52
Glycine	40	3.5	1.5	92	3	5	100
Imidazole	40	5	1.5	48	40	12	45
Imidazole	40	3.5	0.5	70	30	0	58
Glycine	40	2	0.5	92	3	5	97
Glycine	40	5	0.5	93	2	5	93
Imidazole	50	3.5	1	59	30	10	68
Glycine	50	3.5	0.5	92	4	4	76
Imidazole	50	2	1.5	63	24	12	63
Glycine	50	3.5	1	94	3	3	91
Glycine	50	2	1.5	94	3	3	92
Imidazole	50	5	0.5	38	59	3	53
Glycine	50	5	1	94	3	3	87
Imidazole	50	3.5	1	63	26	10	68
Imidazole	60	5	1	50	32	18	41
Glycine	60	5	0.5	2	1	97	12
Imidazole	60	3.5	1.5	61	28	11	57
Imidazole	60	2	0.5	2	4	96	0
Glycine	60	2	1	19	2	79	25
Glycine	60	5	1.5	69	8	23	60
Imidazole	45	2	1.5	66	24	10	55
Glycine	55	5	2.5	85	6	9	90
Imidazole	45	5	2.5	41	56	3	40
Glycine	55	3.5	1.5	71	3	26	80
Glycine	55	2	2.5	87	2	9	84
Imidazole	55	2	2.5	53	26	21	60
Glycine	45	2	2.5	93	3	4	87
Imidazole	55	5	1.5	63	27	10	64

Initial results showed that the base and temperature included a maximum within the design range. However, a linear relationship was observed with base concentration. The design was therefore augmented with additional data points including the addition of base concentration up to 2.5M and utilising a narrower temperature range of 45-55 °C. This allowed the design to consider higher base concentration, whilst simultaneously narrowing down the

temperature range. The resulting data showed that glycine couples more efficiently than aminoimidazole in all cases [Table 19]. However, the best conditions for glycine are different from those of aminoimidazole as can be seen in the cube plot [Figure 91]. Low base concentration and high temperature performs poorly for both substrates, but a compromise is likely needed to discover overall reaction conditions that are favourable for both substrates. The modelled data were in good agreement with the experimental values suggesting that the experiment is high quality and should perform well as a predictive model. Interestingly, the most influential terms seemed to be two-dimensional relationships. Apart from the amine itself, temperature and base concentration seemed to have an overall high influence over the expected outcome from the reaction.

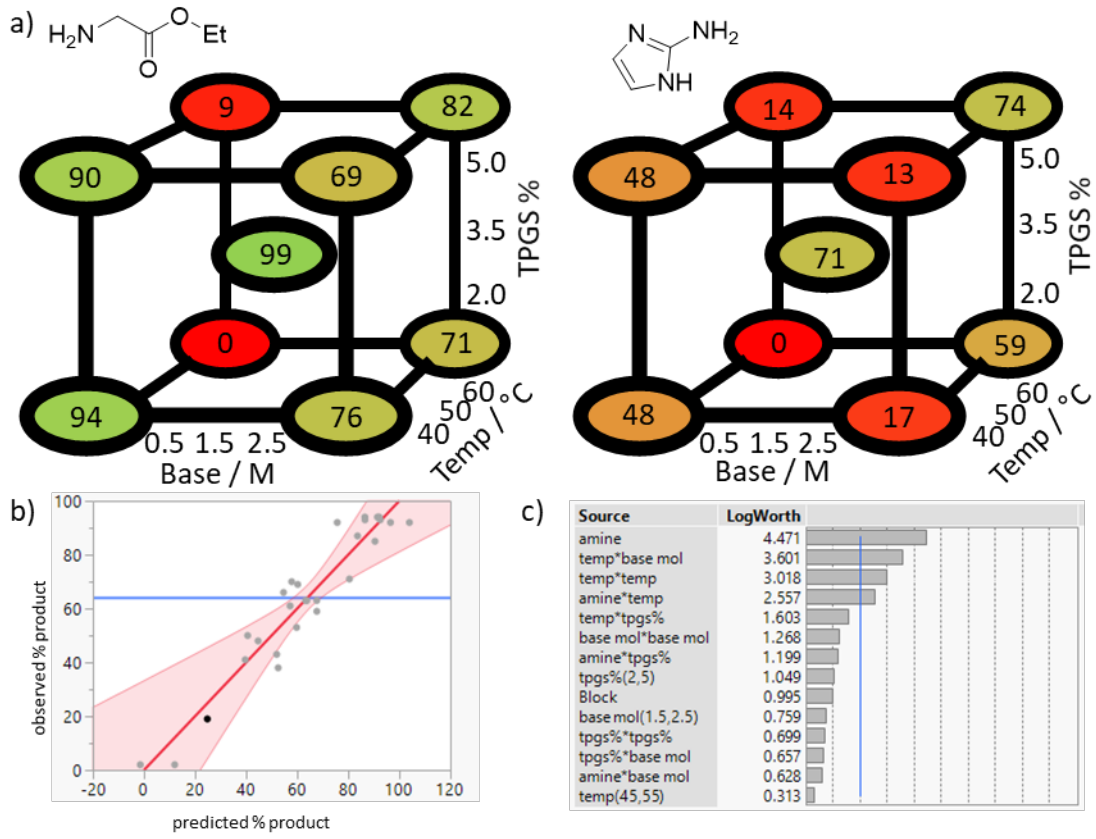


Figure 91 - Optimisation of coupling conditions leading to amide products using FED; a) Cube plots showing the percentage of desired product for glycine ethyl ester and aminoimidazole showing the degree of conversion across the design space for the modelled data; b) Comparison of fitted data from the predictive model with the experimental data; c) Assessment of the most influential terms in the FED and profile of the effect of temperature, surfactant percentage and base molarity.

High conversion was favoured at low temperature and low base concentration or at high temperature and high base concentration for both substrates. This introduces some

interesting two-dimensional relationships that would not typically be apparent in standard optimisations. The use of an FED early in the optimisation of this reaction allowed access to information about what effects the overall reaction conversions more rapidly than a standard optimisation procedure.

Temperature effects the 2 substrates differently, where lower temperatures favour high conversion for glycine but not for aminoimidazole. This could be due to a number of reasons. When raising the temperature it was noted that glycine containing reactions result in a higher percentage of side product formation, maybe showing that across a multitude of reagents a lower temperature may be more successful in producing optimised reaction conditions.

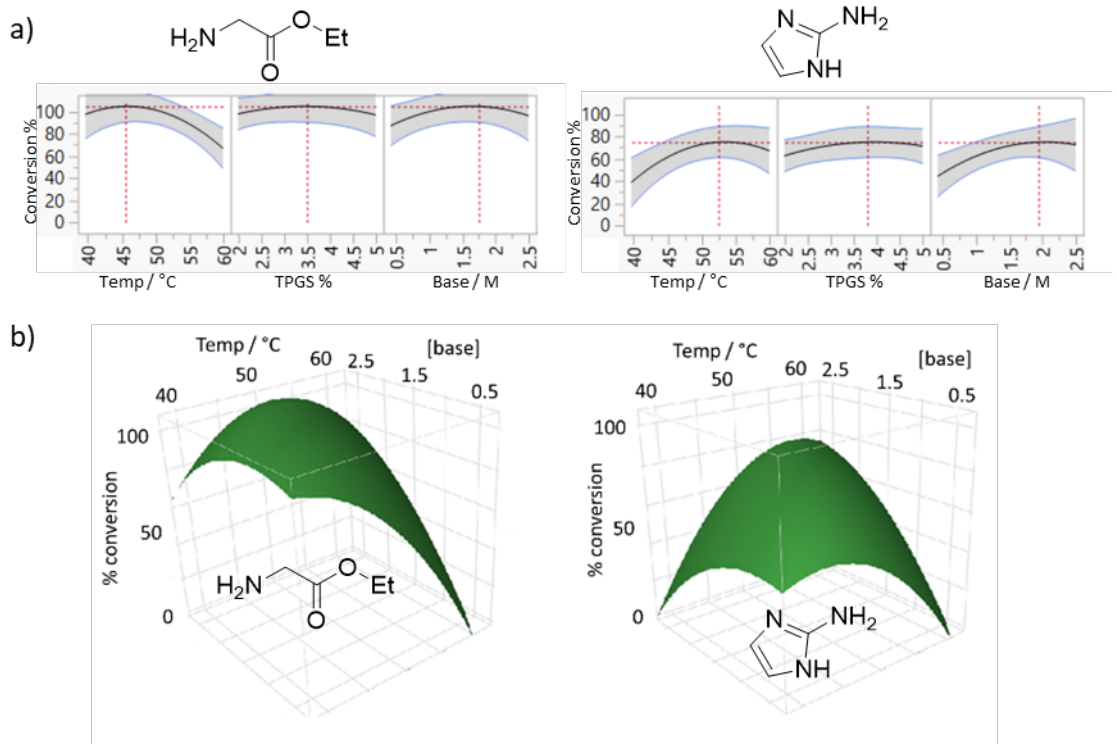
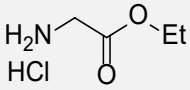
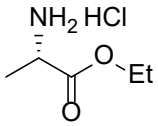
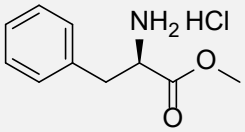
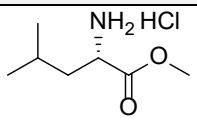
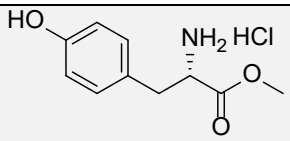
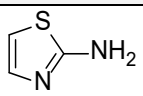
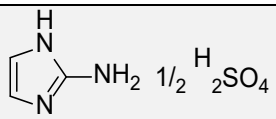
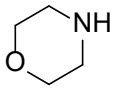
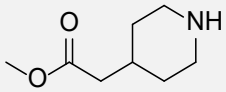
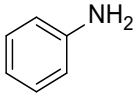
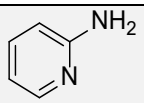


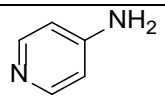
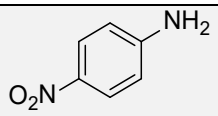
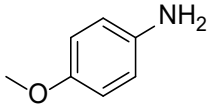
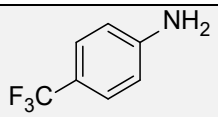
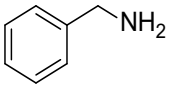
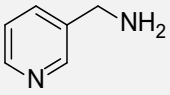
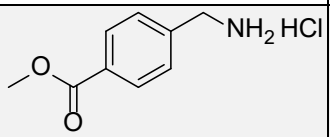
Figure 92 - a) Response curves showing relationship between % product formation and individual parameters for each amine; b) Contour plots showing the combined effect of temperature and base concentration in respect to conversion for each substrate.

Interestingly, some conditions yield no product at all and there is a sudden cut off between reaction conditions that work to those that are poorly performing. The data does therefore suggest that there may not be standard conditions that work for every substrate in this reaction. However, a compromise can be made and validation reactions may be important when considering building block choice. Viewing the response curves and contour plots shows that a parabolic relationship exists for all conditions. Increasing base concentration to 2M is beneficial to both amines where low concentrations are less desirable. Although the effect is

smaller, 3.5-4% surfactant is beneficial for both substrates. The temperature at which the reaction is optimal for both substrates differs. However, increasing side products are produced at higher temperatures, and so lower temperatures may be preferable. To lower the chances of overall side products for a majority of substrates 45 °C was chosen, to give overall post FED optimised conditions of amine (0.5M), 2,6-lutidine (2M), HATU, (0.5M), 3.5% TPGS-750-M, at 45 °C for 16 hours.

Table 20 – Results of screen using post FED optimised coupling conditions; Conditions: **72** (5 nmol), amine (0.5M), 2,6-lutidine (2M), HATU (0.5M), 3.5% TPGS, 30 µL total volume, 45 °C, 16 hr.

Amine		% Product 72x	% Starting Material 71	% Side Product
a		100	-	-
b		100	-	-
c		100	-	-
d		100	-	-
e		96	-	4
f		94	-	6
g		47	52	-
h		100	-	-
i		100	-	-
j		93	7	-
k		50	50	-

l		14	86	-
m		17	-	82
n		76	24	-
o		64	-	36
p		45	54	-
q		90	10	-
r		100	-	-
s		90	10	-

The FED optimised reaction conditions were screened against a variety of substrates. Once again these substrates were chosen to cover as large a chemical space, reactivity and functionality as possible. As natural and unnatural peptides are favoured building blocks in many syntheses due to the ability to have further functionality by reacting the protected acid, a few examples were chosen. Also, heterocycles, anilines with varying functionality and aliphatic secondary amines were chosen to represent many of the likely amines that would be chosen to build a library. This reaction could be useful as a capping reaction for the final library step and also in intermediate steps to incorporate handles for further diversification and so would ideally be compatible with further functionality. This screen shows that a high success rate was observed for all amino esters, including unnatural methyl 4-(aminomethyl)benzoate. This is extremely encouraging as it releases a method to build mixed natural and unnatural peptides on DNA. Many of the other substrates proceeded with high conversions. However, it was not universally successful across all amines. It was noticed that heterocycles were still reasonably unreactive producing only low conversions. It was also observed that electron poor amines such as para, nitro, and trifluoromethyl anilines were

particularly low yielding. Therefore, further optimisation would likely be required for the amide coupling with acid attached to the DNA strand.

It was believed that applying these conditions to the amine linked substrate should also work well as this reaction is less prone to side products and degradation of the activated acid intermediate matters less in this case as the acid is in a vast excess. Therefore, using the same optimised conditions for both would likely result in high conversion rates for this reaction.

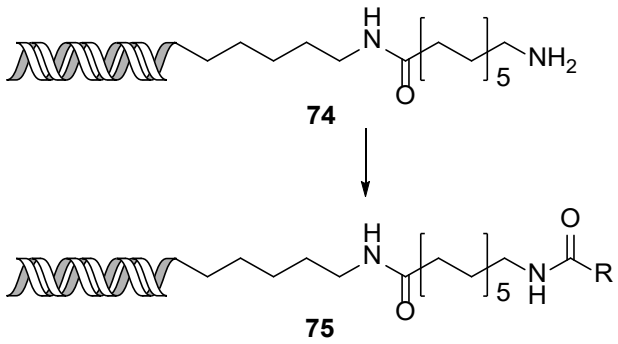
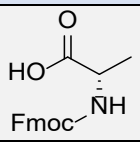
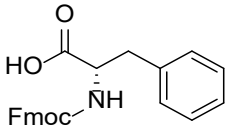
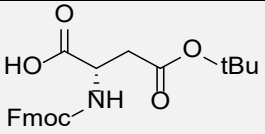
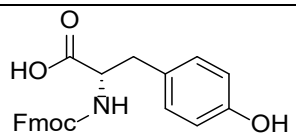
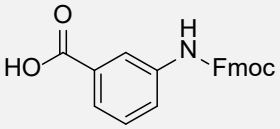
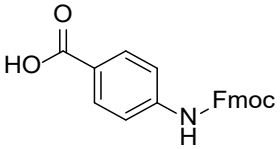
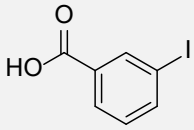
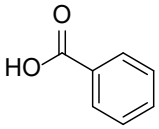
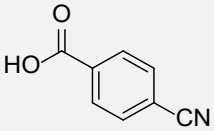
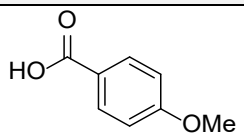
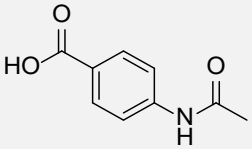
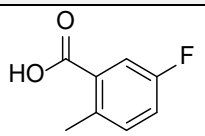
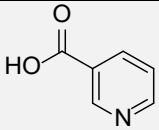
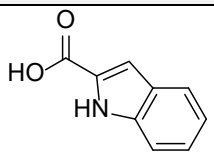


Figure 93 – Schematic representation of the amine conjugated DNA amide coupling.

To synthesise the amino linked compound the same solid supported method was used as previous examples. An Fmoc protected C14-linked acid was synthesised and coupled using the HATU-promoted solid-supported method. The single stranded DNA was then annealed to its complimentary strand and used in the amide coupling reactions. This synthesis was high yielding as with previous versions and could produce a large quantity of DNA conjugate ready for trial reactions [Figure 93][Appendix Figure 21].

Table 21 - Results of screen using post FED optimised coupling conditions; Conditions: **74** (5 nmol), acid (0.5M), 2,6-lutidine (2M), HATU (0.5M), 3.5% TPGS, 30 µL total volume, 45 °C, 16 hr.

	Acids	% Product 75x	% Starting Material 74	% Side Product
a		100	-	-
b		98	-	2
c		94	3	3

d		93	7	-
e		94	6	
f		89	-	11
g		97	-	3
h		98	-	2
i		98	-	2
j		97	-	3
k		100	2	-
l		100	-	-
m		100	5	-
n		100	-	-

These optimised conditions performed extremely well over all the substrates analysed, ranging from protected amino acids to aromatic acids with varying chemical properties [Table 21][Appendix Figures 22-35]. Natural and unnatural Fmoc protected amino acids appeared to be less soluble in the overall mixture. However, all showed high conversions. Asparagine is

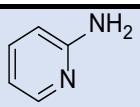
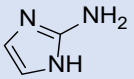
known to be a problematic reagent in coupling reactions and often undergoes side reactions such as deamidation in off of DNA reactions and in this case formed many side products when tested and only little product.[142, 143] All other acids subjected to these conditions performed exceedingly well producing conversion rates at 97% plus in every example or organic acids. This included electron rich and deficient systems, heterocycles, sterically hindered and halogenated substrates. The use of an iodinated aromatics could be extremely beneficial in the synthesis of a library containing tandem amide and Suzuki-Miyaura micelle promoted reactions.

After the DOE, it appeared that the reaction was improved relative to the conditions used previously. However, the conditions were not as successful as hoped across many heterocyclic amines when reacted with acid conjugate **71**. The reaction does work extremely well for froward amide coupling using amine conjugate **74** and a variety of acids, including difficult to couple amino acids. This method shows that micelles can be used to promote amide couplings performed on both DNA conjugates. It also indicates that even though the coupling was optimised for the conjugated acid the conjugated amine system is much simpler to react and therefore benefits from the same conditions. Further optimisation was carried out to further improve the conjugated acid **71** reaction.

5.3 – Optimising DNA Conjugated Acid Reaction

Although optimisation using FED was successful in improving the outcomes of the coupling, further optimisation was required due to the lack of reactivity especially amongst heterocyclic compounds. As stated previously, a problem that often occurs in this chemistry is that when the acid is attached to the DNA the excess water in the reaction can hydrolyse the activated species back to the acid. Using different coupling agents would possibly have an overall advantageous effect on the reaction by either creating an intermediate that is less reactive to hydrolysis, or more lipophilic to enter the micelle more readily increasing the conversion of less reactive substrates.

Table 22 - Results of coupling agent screen; Conditions: **71** (5 nmol), amine (0.5M), 2,6-lutidine (2M), HATU (0.5M), 3.5% TPGS, 30 µL total volume, 45 °C, 16 hr.

Coupling Agent						
	% DP	**% SM 71**	**% SP**	**% DP**	**% SM 71**	**% SP**
COMU	0	54	46	0	42	37
BOP	0	7	93	0	36	64
PBOP	0	100	0	0	100	0
DCC, HOAT	57	43	0	23	77	0
EDC, HOAT	0	100	0	0	100	0
DIC, HOAT	74	25	1	64	36	0

A variety of available coupling agents were screened for the coupling of two heterocyclic amines, 2-aminopyridine and 2-aminoimidazole [Figure 94]. COMU was once tested again as it was proven to be extremely effective in solution phase chemistry. However, even using the improved conditions the morpholine adduct was still formed to a significant degree [Table 22] and there was no desired product observed with either amine. BOP and PBOP were also trialled but again no product was observed in either reaction and surprisingly the reaction involving BOP produced a dimethylamide adduct.

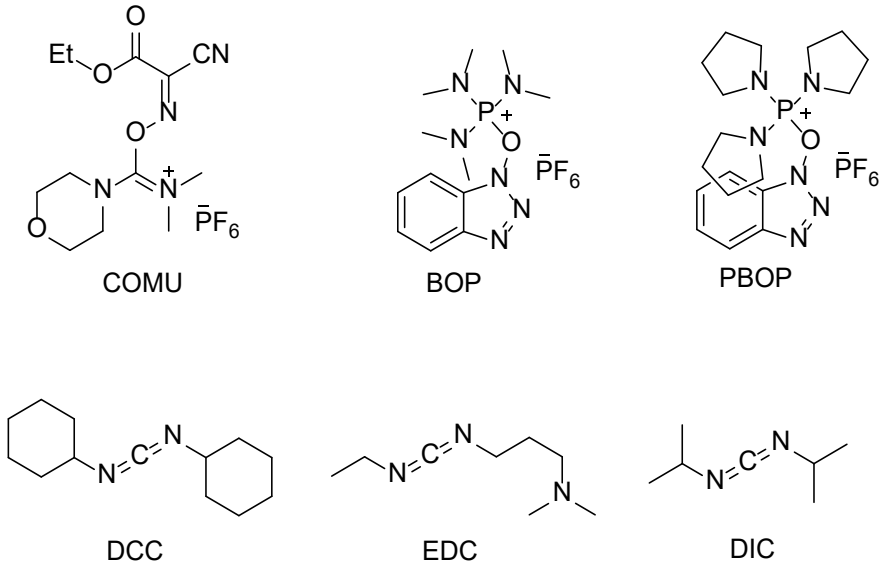


Figure 94 – Coupling reagents used in optimisation screen.

Utilising carbodiimide coupling agents would produce an active ester species with different reactive properties to HATU. HOAT is also added to these couplings to reduce the side product

formation. It was noticed that EDC did not proceed well in this reaction, possibly since EDC is highly water soluble due to the dimethylamino group. The resulting active ester may distribute to a larger degree to the aqueous phase and thus may hydrolyse back into the starting material rapidly. However, using both lipophilic DCC and DIC agents conversion rates were increased considerably. This is consistent with our hypothesis that these more lipophilic coupling agents and their resulting active esters would locate extensively to the hydrophobic interior of the micelle. When the acid encounters the micelle it would react quickly with the carbodiimide and then the amine that is also present in a high localised concentration. This also suggests that the lipophilic activated acid attached to the carbon chain would favour the presence of the lipophilic micelle core further reducing the rate of hydrolysis. Interestingly, DIC also was more successful in this reaction but it was noted that when using DCC there was some solid material in the reaction vial whereas with DIC the reaction mixture was fully homogenous. It is believed that DIC is more readily miscible with the micellar reaction media. Having established that DIC was the most effective coupling agent for these reactions, a repeat of the FED was carried out.

5.3.1 – Second FED Optimisation

As the coupling agents were changed in the process of optimisation for the reverse amide reaction an FED was designed using these new coupling partners. Similar parameters were used as with the HATU optimisation. A range of temperatures were chosen from 40-60 °C, surfactant percentage from 2-5%, and the concentration of the base from 0.5-1.5 M were assessed. The overall FED experiment was run simultaneously using glycine ethyl ester and aminoimidazole as the amines substrates as examples of reactive and deactivated amines and to maintain consistency with the previous FED [Table 23].

The overall conditions screened initially indicated that as temperature is increased, the amount of side product formed in these reactions also drastically increases. At 60 °C there was a large amount of side product in both sets of substrates. This is also consistent with the results from the previous FED. However, at lower temperatures starting material was still obtained in the mixture showing that an intermediate temperature was optimal. Glycine performed best at lower temperatures with much less side product formation, and a large degree of product formation. The less reactive aminoimidazole required elevated temperatures to react but this induced side product formation. This suggests that less reactive

substrates may not be able to proceed to full conversions under the screened conditions and some degree of either starting material or side product would always be present.

Table 23 - Factorial experimental design assessing conversion for glycine ethyl ester and 2-aminoimidazole.

Amine	Temp. (°C)	TPGS-750-M (%)	2,6-lutidine (M)	Product	Starting material 71	Side product
Imidazole	40	2	1	43	54	4
Glycine	40	3.5	1.5	92	3	5
Imidazole	40	5	1.5	48	40	12
Imidazole	40	3.5	0.5	70	30	0
Glycine	40	2	0.5	92	3	5
Glycine	40	5	0.5	93	2	5
Imidazole	50	3.5	1	59	30	10
Glycine	50	3.5	0.5	92	4	4
Imidazole	50	2	1.5	63	24	12
Glycine	50	3.5	1	94	3	3
Glycine	50	2	1.5	94	3	3
Imidazole	50	5	0.5	38	59	3
Glycine	50	5	1	94	3	3
Imidazole	50	3.5	1	63	26	10
Imidazole	60	5	1	50	32	18
Glycine	60	5	0.5	2	1	97
Imidazole	60	3.5	1.5	61	28	11
Imidazole	60	2	0.5	2	4	96
Glycine	60	2	1	19	2	79
Glycine	60	5	1.5	69	8	13

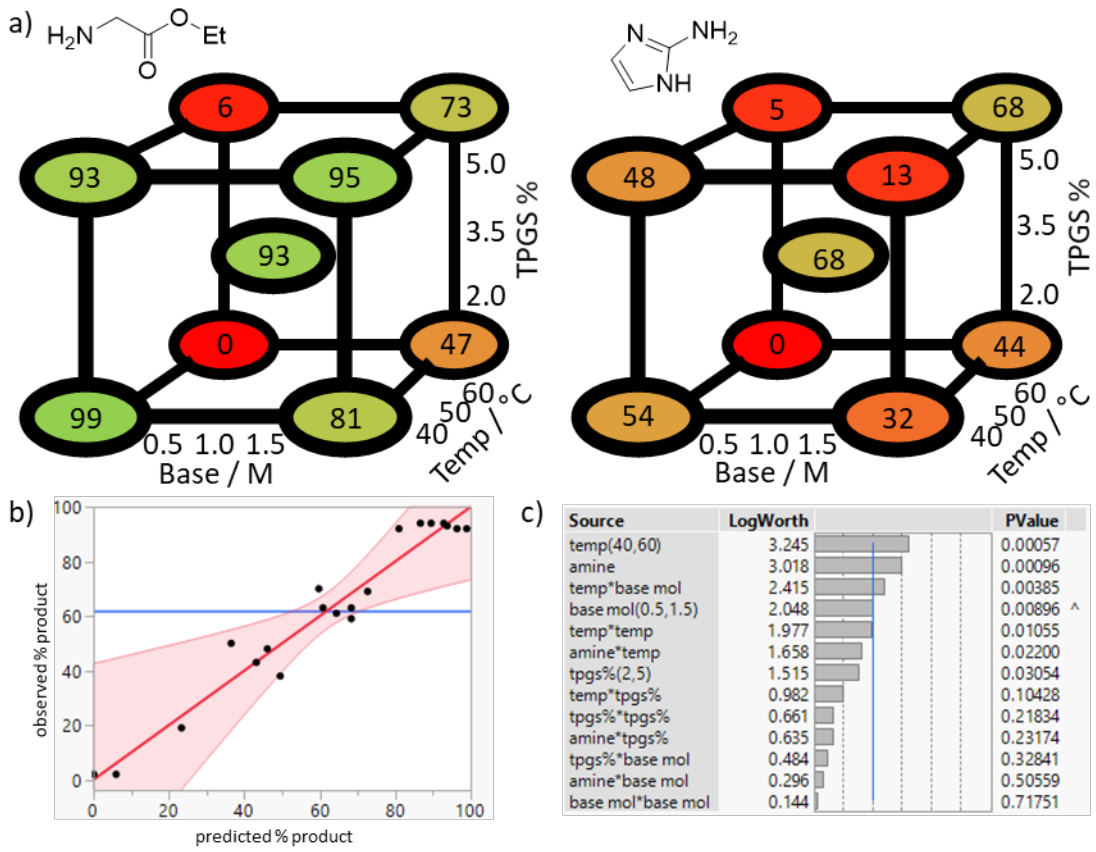


Figure 95 - Optimisation of coupling conditions leading to amide products using FED; a) Cube plots showing the percentage of desired product for glycine ethyl ester and aminoimidazole showing the degree of conversion across the design space for the modelled data; b) Comparison of fitted data from the predictive model with the experimental data; c) Assessment of the most influential terms in the FED and profile of the effect of temperature, surfactant percentage and base molarity.

A comparison of the experimental method with the predicted model was performed [Figure 95 b)] showing that the experimental values lie within the range of expected outcomes from the model system. There is also a cluster of data points visible near the 90-100% conversion region due to the glycine reaction having a larger success in reaction across the range of conditions used. The P-value terms were also calculated for this reaction [Figure 95 c)]. The most influential factor on the overall outcome of the reaction was once again the temperature. As stated previously, temperature has a high P-value because of its effect on both increasing yields and the increase of side products. The second largest factor was the choice of amine which was expected with the varying reactivity of the two amines chosen. Base concentration also has a large effect on the overall conversion rates and second order effects were also apparent for temperature and a combination of temperature and base concentration.

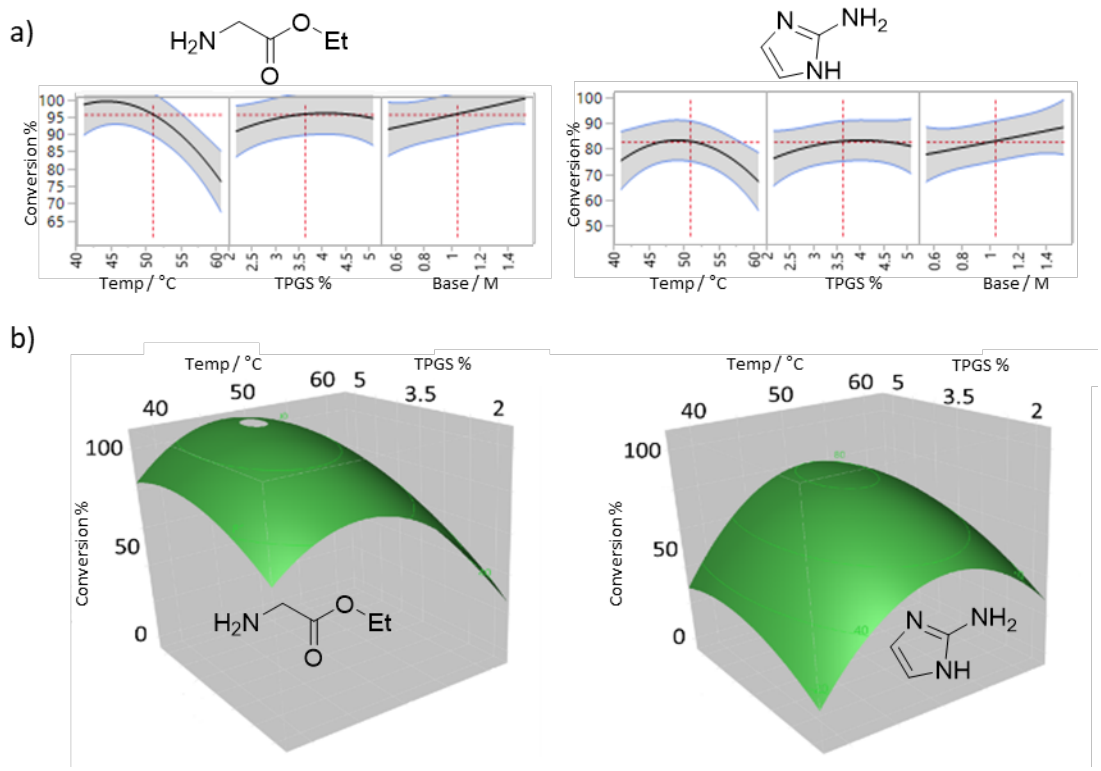


Figure 96 - a) Response curves showing relationship between % product formation and individual parameters for each amine; b) Contour plots showing the combined effect of temperature and surfactant percentage in respect to conversion for each substrate.

Both reagents showed a parabolic relationship overall with the factors tested [Figure 96 b)], but each had a maxima with different conditions. Both substrates seemed to react more favourably with a surfactant concentration of 4.5%, and a 2,6-lutidine concentration of 1.5M. However, the biggest difference between the two was the temperature used. For glycine, a temperature of 45-50 °C was optimal for the reaction to occur, whereas the optimum for aminoimidazole was 55 °C. However, above 50 °C an increase of side products is observed. If lower temperatures are used then less side products would be created in the library. However, if higher temperatures were used more amines may reach the validation cut off point at the detriment of the overall library. The decision to use a lower temperature affords a more balanced reaction where less side products would be produced increasing the overall library fidelity and reducing the chance of false positive and negative results due to side products in the library screening. However, the number of amines that would pass the validation process may also lower.

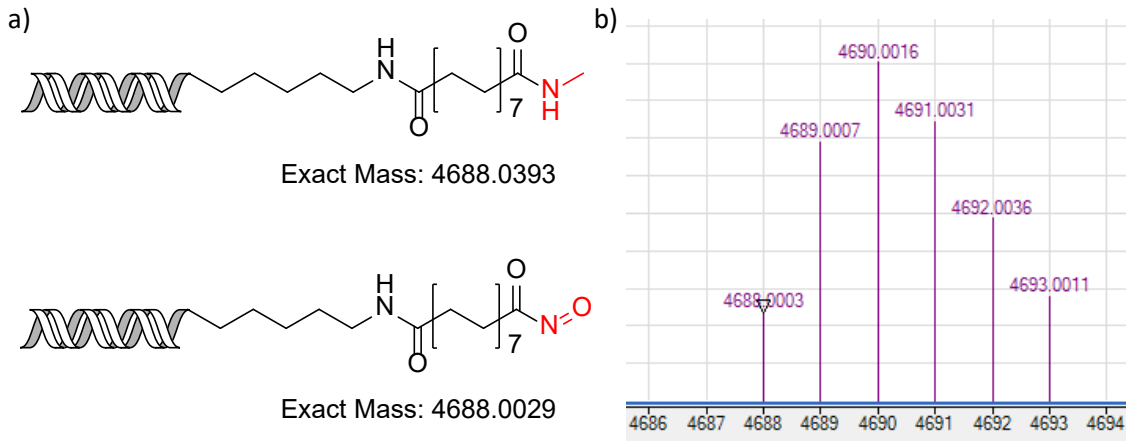


Figure 97 – a) Possible side products formed in reverse amide bond formation including predicted mass; b) observed molecular weight of major side product adduct formed in reverse amide reaction.

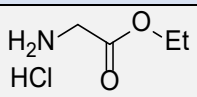
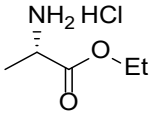
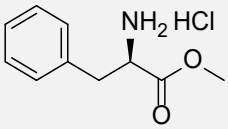
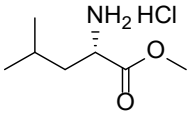
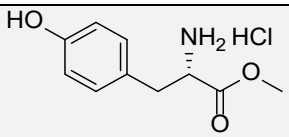
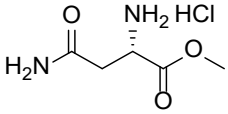
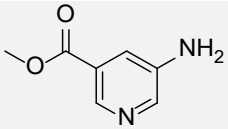
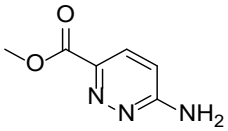
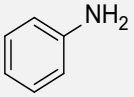
The major side product in the reactions appeared to be the same across both substrates and had the mass of 4688 Daltons [Figure 97 b)]. This side product was also seen when screening the reaction against other amines and its origin was unknown. The mass observed could be accounted for by either methylamine or a nitroxide like side product formed on reaction with the DNA conjugated acid [Figure 97 a)]. This side product is akin to those seen when using COMU as a coupling agent. However, neither methylamine nor any nitroxide containing compounds are used in these reactions. The side products are also the same across all amine substrates where they are observed suggesting that the product does not arise from impurities in the amines. Also the surfactant, base, DIC and HOAT used in these reactions are not synthesised using any chemical which could produce these side products, meaning that it is unlikely to be formed from an impurity in the reaction mixture. It also seems as though these side products are more likely to be produced when using unreactive amines such as heterocyclic amines, suggesting the side product formation may be due to the use of excess coupling agents in cases where the subsequent reaction with the amine is slow.

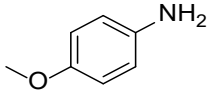
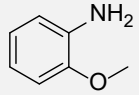
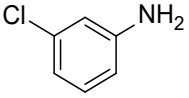
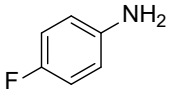
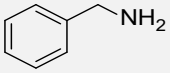
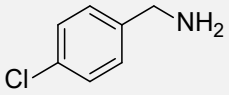
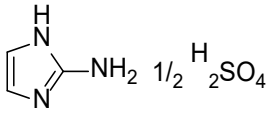
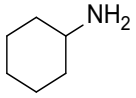
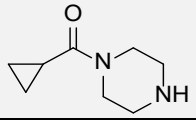
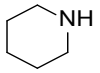
This reaction could possibly be further optimised, using a different surfactant, or screening of more conditions. However, the use of DIC, HOAT in TPGS-750-M and 2,6-lutidine resulted in high conversions for these compounds and would be applied to further amines to establish the scope of the reaction. This reaction has been modified greatly from those used in literature and is a huge improvement on the reverse amide scope previously seen using DMT-MM conditions.

5.4 – Optimised DIC Conditions

The optimised conditions were used to screen a range of amines. The objective was to find conditions that worked well across a broad range of substrates. As stated previously, this is no simple task as the activated intermediate is easily liberated back to the starting acid in this reaction. This means that unreactive amines are still extremely difficult to couple and some further modification to the conditions could be used. However, these conditions were used to screen against multiple amines, from simple amino esters to aromatic, heterocyclic and aliphatic amines.

Table 24 - Results of optimised DIC conditions; Conditions: 71 (5 nmol), amine (0.5M), 2,6-lutidine (2M), DIC (0.5M), HOAT (0.5M), 4.5% TPGS, 30 µL total volume, 45 °C, 4 hr.

Amine		% Product 72x	% Starting Material 71	% Side Product
a		100	-	-
b		100	-	-
c		100	-	-
d		96	-	4
e		100	-	-
f		98	-	2
g		35	-	65
h		58	-	42
i		100	-	-

j		95	-	5
k		85	-	15
l		84	1	15
m		100	-	-
n		100	-	-
o		100	-	-
p		58	-	42
q		100	-	-
r		100	-	-
s		100	-	-
t		100	-	-

These conditions were found to be extremely reliable on a large variety of substrates tested [Table 24][Appendix figures 37-56]. Amino esters are extremely valuable in tandem coupling reactions as they can be used to create polyamide libraries. All the canonical amino acid derived esters proceeded to product in 100% conversion. However, when using unnatural pyridine and pyridazine amino esters there is a large decline in conversion rate. All aliphatic and aromatic amines tested also gave extremely high conversions showing that across the board the scope is very high outside of heterocycles. However, under these conditions aminoimidazole was also not as high yielding as expected but was over 50% conversion rate improving the reaction in comparison to when HATU was used.

Overall, these reaction conditions are highly efficient and probably represent the best method for carrying out couplings of this type. However, there are still improvements that could be

made to improve reactivity of heterocycles. The reduction in rate of conversion led to the putative dimethylamide side product formation in all cases.

5.6 - Conclusions

The application of micellar media to the catalysis of amide couplings was demonstrated as a success. Two separate conditions were found that were optimal for both acid and amine conjugated DNA strands, both of which expand the current scope of amide couplings considerably [Figure 98]. The use of micellar media has a great effect on the reaction of a DNA conjugated acid as it lowers the rate of hydrolysis back to the starting material as well as increasing the concentration of the amine in the micelle. Overall, these conditions are highly applicable to DNA encoded library synthesis, showing no degradation of the DNA tag and the only detectable side product is believed to be the dimethylamide adduct. These reactions also show high conversion rates for natural and unnatural amino acid derivatives, meaning that this reaction could be extremely useful in creating large peptide libraries of a variety of different functionality.

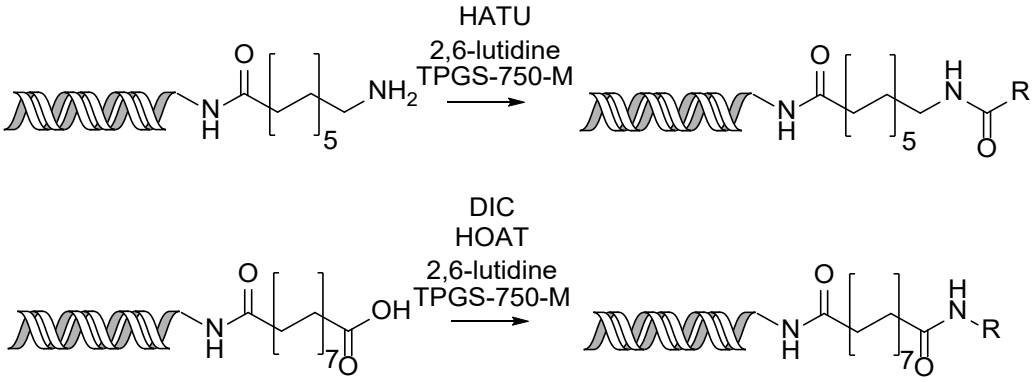


Figure 98 – Optimised conditions for reaction of both DNA conjugated amines and acids using micellar media.

Due to the side products formed when reacting a DNA conjugated acid with less reactive amines, this reaction could possibly be further optimised. It has been shown that a different surfactant MC-1 helps to produce high conversion rates for reactions proceeding off DNA.[125] MC-1 is a surfactant that contains a sulphone in the hydrophobic region. It is thought that this improves solubility and incorporation of more hydrophilic components into the micelle by offering a hydrogen bonding motif. This increases solubility and mixing of many of the components in the reaction allowing for improvements in reaction conversions. Using this solvent for on DNA amide couplings could be beneficial as the increased concentration in the

micelle of the amine could lead to higher conversion rates for the overall reaction. It may also help to alleviate the need for hydrophobic coupling agents like DIC as agents would be more solvated in the micelle. However, this system could be used as a follow up to the success of this original amide coupling procedure.

This use of micelles for amide coupling has been an overall success. When used with a lipophilic linker the reactions proceed in high conversions. The work up procedure is simple and the DNA tag remains intact. This overall system would increase the amount of chemical building blocks available for use with DNA encoded libraries, therefore increasing the chemical space the library probes and the chance of finding a hit compound.

Chapter 6 – Building a Library

The novel coronavirus, SARS-CoV-2, the causative agent of COVID-19 has had a significant impact as a worldwide pandemic. Currently there are no known treatments, even though research has also been carried out on closely related SARS-CoV and MERS-CoV, sources of the previous two coronavirus outbreaks.[144-146] SARS-CoV-2 is a single stranded RNA Betacoronavirus, in which the viral RNA encodes two open reading frames that generate the two polyproteins pp1a and pp1ab.[147] These two proteins generate many replicase-transcriptase proteins[148] and are produced by two viral cysteine proteases Papain-like protease (PL^{pro}) and the main protease (M^{pro}). These proteins are responsible for the cleavage of the polyproteins into functional units and are prime targets for antiviral drug development.[149]

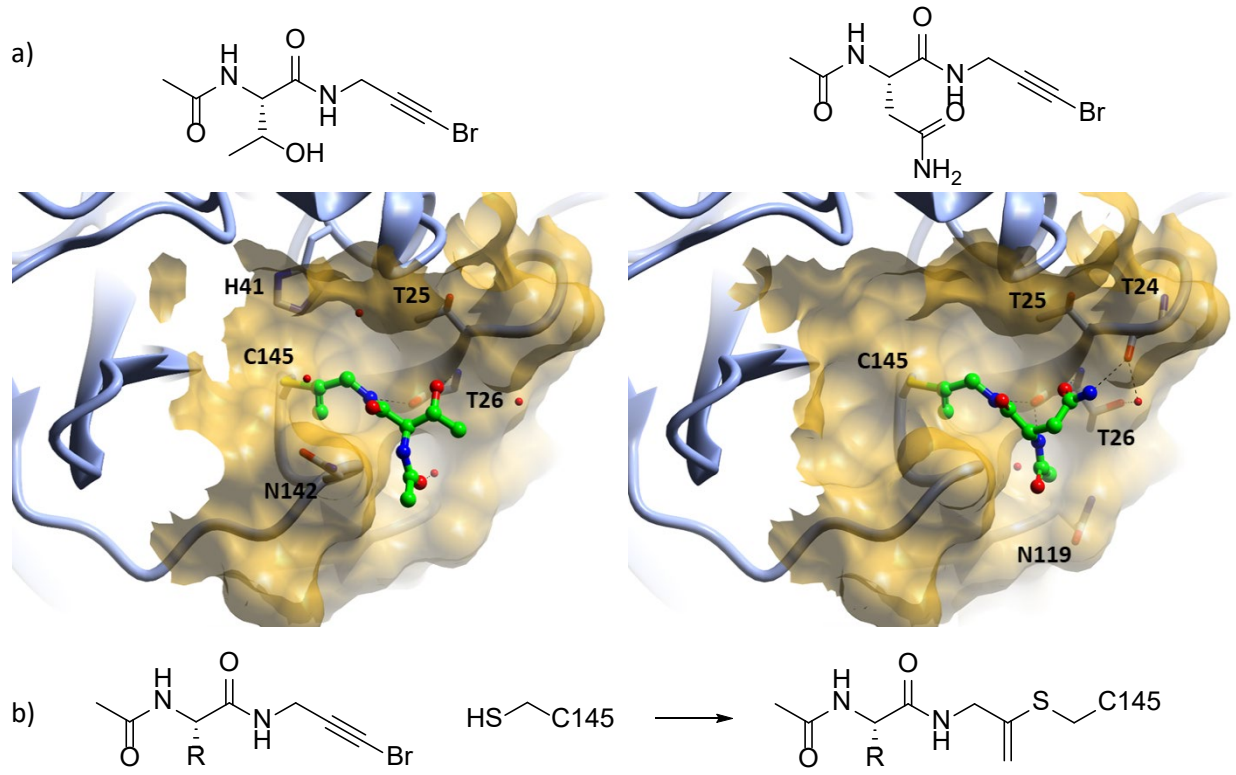


Figure 99 – a) crystal structure of 2 covalent fragments anchored and Cys145 of the orthosteric M^{pro} binding pocket; b) covalent reaction of bromoalkyne with Cys145.

Coronavirus M^{pro} active sites are highly conserved and are often structurally similar which has led to efforts of developments of broad-spectrum antivirals targeting this protease. Peptidomimetic α-ketoamides have been shown as successful classes of inhibitors for coronavirus replication[150] and the crystal structure of SARS-CoV-2 M^{pro} has been determined.[151] M^{pro} is therefore a highly attractive target for antiviral therapies. Accordingly,

our collaborators at Diamond Light Source carried out a crystallographic fragment screen against SARS-CoV-2 M^{Pro}. Interestingly from this screen 19 non-covalent and 39 covalent hits were discovered for the M^{pro} active site.[152]

Part of this screen included the novel fragment libraries known as FragLites and PepLites.[153] These libraries were designed to contain small compound fragments with bromo and iodo functionality. The Fraglites and PepLites identify drug-like interactions with a target protein and are identifiable by X-ray crystallography by exploiting the anomalous scattering of the halogen substituent. This allows very small fragments to be visualised when bound to a target protein. The PepLites contain an interesting bromoalkyne functionality to easily visualize the binding in normal crystallographic experiments. However, when used in the M^{pro} assay, crystallographic data showed that they bind via a covalent linkage with cysteine 145 of the orthosteric M^{pro} binding pocket [Figure 99]. This appears to have occurred with displacement of the bromo forming an alkene in the product by an unknown mechanism. The product is visualised by crystallography and the PepLite fragment also shows affinity in functional protease cleavage assays. The most active PepLites were those derived from threonine and asparagine which formed interactions with T24. It is believed that the acetyl functional group is directed towards the solvent away from the protein and takes part in no binding. Hence, this position of the molecule could be bound to a linker and the molecule could be extended from this position. If a linker is tolerated, a DNA encoded library could be synthesised as a targeted library for M^{pro} to quickly screen different both natural and unnatural peptides, different functional groups capable of forming covalent linkages with cysteine, and also what functionality can be tolerated out of the pocket. The result of this library is to help to find a new class of compounds that are active against the M^{pro} active site leading to future antiviral therapies as a treatment for COVID-19.

6.1 – Library building block design

The developed micellar amide coupling was used as a basis to design a targeted library for M^{pro} using the reverse amide coupling DIC conditions. As this library is a targeted library functionality that induces covalent linkages with the target protein is required. It would also include many other amines as a third step of the synthesis to probe other interactions with the active site and make the library more generally useful for hit finding on other targets. The other two building block additions would be a mixture of the same amino esters coupled

sequentially [Figure 100]. This would allow for a rapid synthesis of a large library with great chemical diversity. Amino esters bearing threonine and asparagine will also be included in the R_1 and R_2 building blocks to incorporate known active binding motifs into the library. The library will also include a variety of other natural and unnatural amino esters, especially those that were not included in the PepLites screen to further increase diversity and probing efficiency of the library.

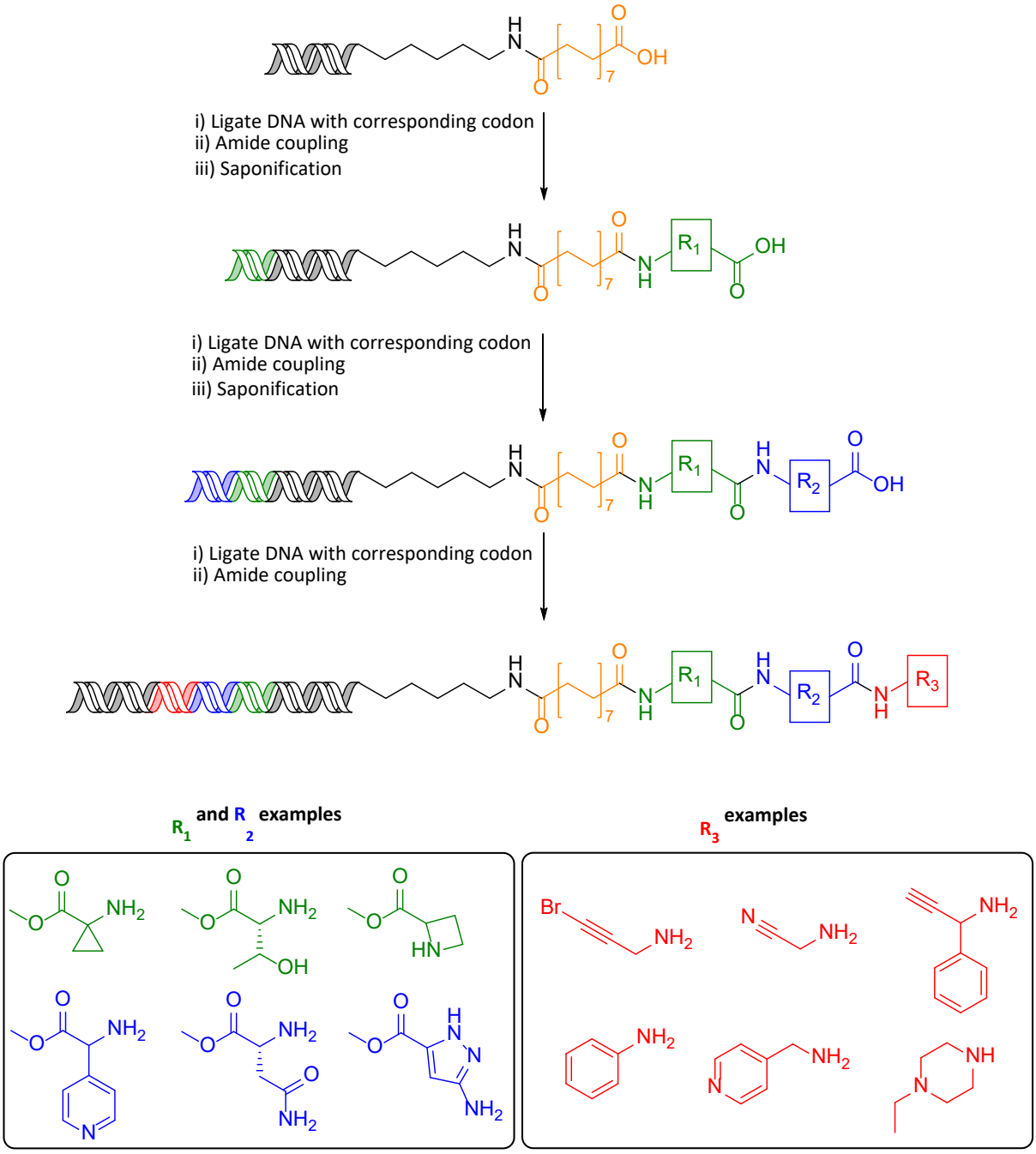


Figure 100 – SARS-CoV-2 M^{pro} targeted DNA encoded library synthesis and plan.

In this library design, a variety of amino esters were chosen as it allowed sequential additions of building blocks using the reverse amide coupling. The esters are easily cleaved in lithium hydroxide and can be deprotected once pooled together enabling an extremely simple method of building block addition. Amino esters were chosen to include threonine and asparagine, as well as a variety of other aromatic, aliphatic and heterocyclic esters. These esters were chosen and supplied by Enamine from their building block library. All amino esters were chosen to have a molecular weight less than 200 Daltons and were chosen to represent a broad chemical diversity as well including binding motifs similar to those seen in the active covalent fragments. It was also not known how much variation is tolerated in the R_1 position, so glycine was included in this set as a spacer to mimic the acetyl functional group on the original hit fragments.

The R_3 building blocks are all amine caps with varying chemical diversity. Firstly, several bromo alkynes were custom synthesised by Pharmaron including cyclopropyl, benzyl and geminal dimethyl functionality in the 2-position. This subset of compounds also included nitrile and alkynyl replacements of the bromo alkynes to probe the effect of different possible covalent binding molecules. The rest of this set of amines was chosen to contain a wide variety of functionality, including varieties of functional aromatic amines and benzylamines, heterocyclic amines, and aliphatic amines. The increased amine selection would help to not only identify compounds that are active against SARS-CoV-2 M^{pro}, but as a more general screening library. The overall size of the library was designed to contain 50 building blocks in each step, resulting in 125,000 individually coded compounds. This would include 22,500 potential covalent compounds for targeting cysteine 145 of M^{pro}.

6.2 – Test Compound Synthesis

A test 1x1x1 library was synthesised on a double stranded 14mer DNA conjugate. This compound was prepared to test the chemistry and ensure that the building blocks were tolerated on the core yielding high conversions. The process would couple a spacing glycine amino ester followed by a known active threonine amino ester and the active bromo alkyne substituted amine as a capping building block. This compound could be used as a control compound for selection experiments as it is a close analogue of one of the initial hits. Use of this combination would also establish that the covalent active amine is tolerated in the synthesis. The amide coupling procedure used involves nucleophilic components such as

HOAT which could in theory react with electrophilic species. The initial reaction was carried out on the double stranded C14 linked acid. The overall test compound was also synthesised at a higher quantity to allow for losses in purification at each step and prepare a sufficient amount of product for mass spectrometry and bioassay testing.

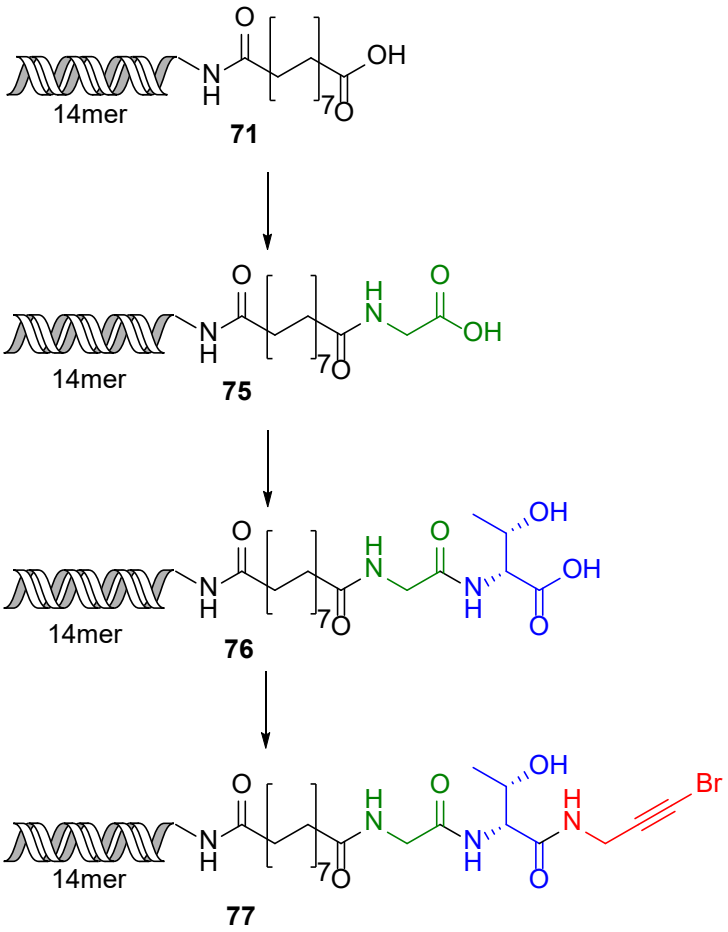


Figure 101 - synthesis of the SARS-CoV-2 tool compound.

The first coupling with glycine ethyl ester was achieved at a larger scale than those used in the test substrates at 24 nmol. The amount of other reagents used was not changed and only the DNA added was at a higher concentration. At this concentration, the reaction still proceeded with great efficiency at 100% conversion by mass spectrometry [Figure 102 a)]. To purify these compounds ethyl acetate was added and the organics extracted and discarded. They were then ethanol precipitated before saponification of the ethyl ester using lithium hydroxide. The DNA was further purified and extracted by ethanol precipitation to yield a pellet of 15 nmol of DNA. This pellet was used in the next amide coupling reaction involving threonine methyl ester where the process was repeated yielding full conversion to the desired product [Figure 102 b)]. After hydrolysis of the ester the amide coupling procedure was

repeated a third time with the addition of the bromo alkyne substituted amine. Again, the reaction proceeded with great conversions yielding 3 nmol of the desired product [Figure 102 c)]. On purification of this product a 3000 Dalton molecular weight spin filter was used to remove any extra organic materials, surfactants and salts from the DNA conjugate so that the compound was pure for SPR and MS analysis. This step was extremely efficient at removing any additional impurities from the reaction mixture and would be used as a purification strategy in between each stage of the library synthesis.

The successful synthesis of the 1x1x1 library showed that the micellar promoted amide coupling chemistry is compatible with DNA and sequential additions of amino esters, utilising a coupling, hydrolysis method. This also proved that the bromo-alkyne substituent is unaffected by the reaction conditions the DNA itself and purification process. The DNA also did not degrade under the conditions used here granting full conversions and pure reaction profiles at each stage of the reaction. This compound was used in bioassays and a MS binding assay for SARS-CoV-2 M^{pro}. However, no activity was found in the bioassay, and no covalent linked fragment mass was found by mass spectrometry. This does not mean that the compound does not bind as the bioassay is usually achieved at much higher concentrations of drug candidate. Also, the mass spectrometry analysis has never been performed before on a DNA conjugated covalent linked protein and so both methods need further development to give conclusive evidence for the binding of the compound.

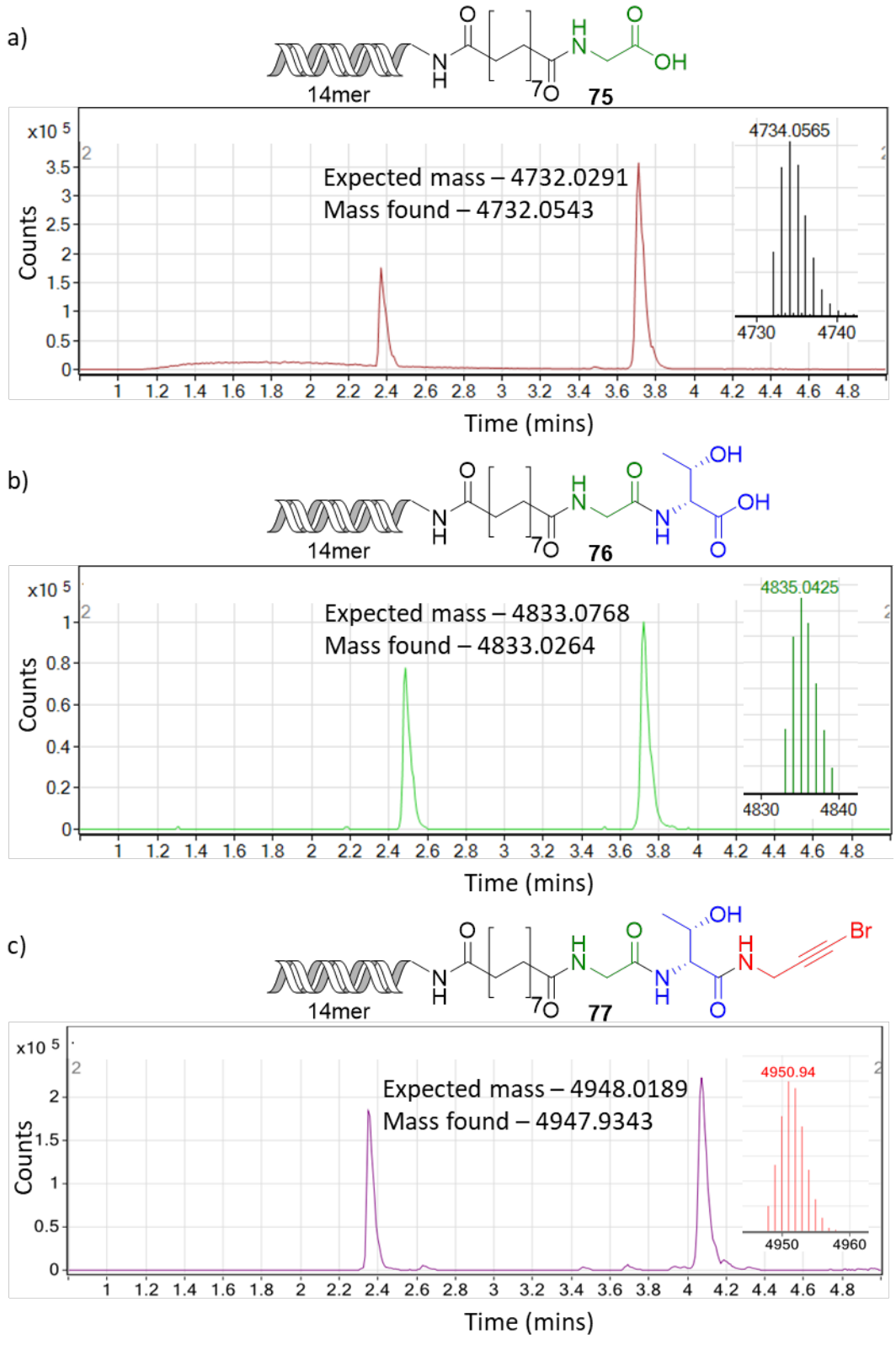
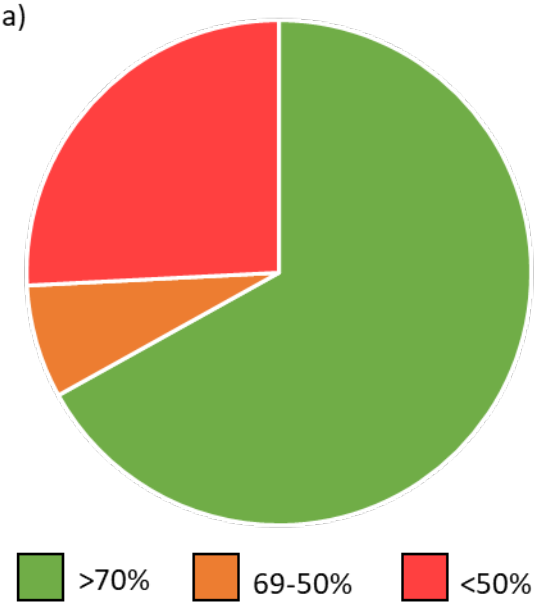


Figure 102 – Mass spectral analysis of the products synthesised in each phase of the 1x1x1 library showing the product and the complementary strand.

6.3 – Validation of building blocks

Although a 1x1x1 library was successfully synthesised utilising this method the building blocks that were planned to be used in the library build needed to be validated against a test system to see which building blocks react efficiently and if there are any impure building blocks or inefficient reactions. This had been done previously on a small number of building blocks to show that the chemistry worked on a range of different substrates. However, over 100 amino esters and amines would need to be validated to allow this library to be synthesised. These reactions were carried out in a 96-well plate format using Paradox micro plates and sealed and heated at 45°C for 3 hours. The reactions were then individually worked up and submitted for mass spectrometry analysis.

a)



b)

Conversion to DP	Number of substrates
>70%	83
50-69%	9
<50%	32
Total Substrates	124

Figure 103 – Validation of optimised reverse amide conditions using 124 amines and amino esters.

Mass spectrometry analysis was again used to analyse all substrates and showed no visible DNA damage on any substrate. All substrates were analysed for starting material, desired product and side product formation. When viewing the entire data set, 75% of the compounds converted at over 50% conversion from the starting material. Only 25% of reactants resulted in less than 50% conversion, many of which did not show any product maybe due to the degradation of the starting amines. The validation reaction was carried out on a mixture of amines and amino esters where it seemed that heterocyclic compounds were still often the more difficult substrates to react. This result shows that further optimisation

could be achieved on these substrates, but the reaction proceeds on a much larger substrate scope than seen in literature conditions. It is also worth noting that multiple different salts and salt ratios were included in these substrates, none of which had any visible detrimental impact on the resulting conversion rate.

The results of this screen were extremely encouraging and allowed 47 amino esters and 45 amines to pass the validation criterion of 50% conversion rate. This resulted in a library synthesis consisting of 47 amino esters in the first two building block additions, followed by 45 amine caps, resulting in a 99,405 membered DNA encoded library. This library also includes potential SARS-CoV-2 M^{pro} covalent inhibitors as well as a mixture of more general amine caps. To also study the ligation and chemical build of the library a 1x1x1 library, including coding steps, was produced alongside the library synthesis.

6.4 – Sequence Design

The DNA sequencing of a chemical library is paramount to its success as a screening library. To produce active hit compounds, a successful encoding strategy must be employed. This strategy was designed around the successful encoding strategy used for the 6x6 Suzuki library [Table 13] and was shown to successfully encode, PCR amplify and sequence the library. The primer sequences worked well in this example, allowing for good binding and amplification in PCR. The scaffold code successfully coded with a 3 base pair sequence giving enough sequences for 64 individual libraries. In future when more libraries are produced this sequence can be easily elongated to generate even more coding sequences. The major change in this sequencing strategy was to include more building block coding sequences as in total 139 different sequences would be required for the 139 individual building blocks that make up the library. These would maintain the 8 base pair coding strategy used previously and contain a Hamming distance of 4 throughout each coding sequence. The choice of these sequences was similar to that used previously where the codes were directly taken from literature[141] and used in the library design. However, these codes generated were single stranded and did not consider the reverse sequence of the complimentary strand. The complimentary sequences were created using excel functions. These complimentary strands were then analysed for matches against the original codes and any that matched were removed from the design. This generated a building block coding strategy that should be robust, include no base pair mismatches and remove the ability for mismatched ligation to

occur. This resulted in a DNA strategy in which false positives or negatives should not occur due to DNA sequencing or mismatching of base pairs. This design also included overall a GC content of between 40-60% for every library member.

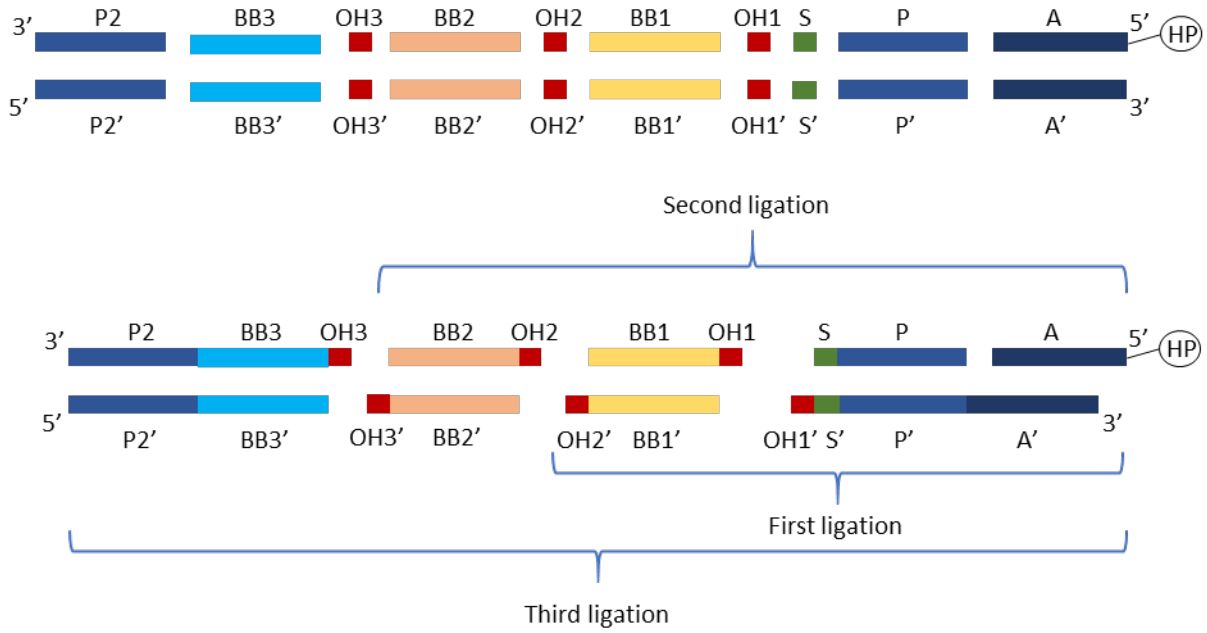
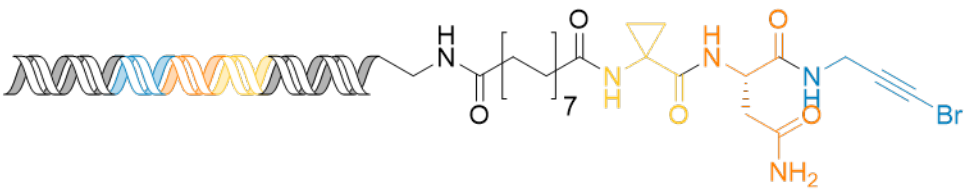


Figure 104 - Ligation and encoding strategy planned for use in the SARS-CoV-2 M^{pro} library.

With this encoding strategy in place, each member of the library will contain 3 individual codes for each building block used, a unique library scaffold code and two primer codes for PCR amplification [Figure 105]. The adapter sequence would be removed when PCR amplification takes place. Also, overhangs would be present and allow for simple recognition of each building block code. As this library utilises the same strategy that successfully coded the protocol Suzuki-Miyaura coupling library it was anticipated that it would also be successful when used to encode a larger library. To ensure that this is the case, ligations were once again tested on this substate at the concentrations that would be required in the library build, and a 1x1x1 library was simultaneously synthesised as a control. This 1x1x1 library would be used to follow the integrity of the overall library build to ensure that the products from the library could be sequenced and decoded to produce the codes for each building block used.



Linker strand (5'-3')
A P S OH1 BB1 OH2 BB2 OH3 BB3 P2
GTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCA ACTACATGGTCTACA
Complimentary strand (5'-3')
P2' BB3' OH3' BB2' OH2' BB1' OH1' S' P' A'
TGTAGACCATGTAGTTGAGGTCAGGTGCCGTCGTATGTGCGCGTAGGAGTCGATCATACAGCCAAATCCGTTCACACCGACC TGAATTCGGCAAGAC

Figure 105 – Graphical representation of compound and corresponding coding strategy used for a member of the library.

6.4.1 – Library Scale

As this library was much larger than any produced before, and would ideally be screened against multiple proteins, the scale of the library was considered. This would also need to be accounted for when synthesising the head piece and linker strand, and the scale of the ligation and the chemical reactions. These choices must consider losses between each building block addition, the overall required yield for library screening after purification at the final stage and the overall ease of synthesis of the library. Additionally, it was found that at least 1 million individual library members are required for a successful library screen to take place equivalent to 0.0016 fmol of compound.[154] Having this number of compounds reduces the chance of false negative hits, but a larger quantity of molecules is desirable. Other past DEL screens also have been shown to be successful at this scale, for example, 5 nmol of an 880 million member library where each library member was present at 0.006 fmol has led to validated hits.[17]

Table 25 – Total number of library members required to screen against a target protein.

47x47x45 library - 99405	Total Library Members		Scale
each library member	0.1	0.006	fmol
Total library amount per run	9.9	0.6	pmol
total library for 1000 runs	9.9	0.6	nmol

With the scale used in these examples, a 47x47x45 M^{pro} targeted amide library consisting of 99,405 members and screened at 0.006 fmol per member would require 0.6 pmol of the

overall library [Table 25]. If this library were to be used 1000 times it would only require a total amount of 0.6 nmol. For synthesis of the library that would suggest that only 0.012 nmol of DNA would be required from each split and pool step. However, due to the synthesis scale being reduced compared to previous validation reactions and as this being the first library of this size produced in our laboratory, it was decided to increase the synthesis to a more manageable scale.

Table 26 – Desired library synthesis scale using this 3-step amide coupling process.

47x47x45 library - 99405	**Amount of each Library Member**	**Scale**
last synthesis should yield	**101.25 (total)**	nmol
last step, 45 reactions at	2.25	nmol
second step, 47 reactions at	4.5	nmol
first step, 47 reactions at	9	nmol
total linker needed	**423**	nmol

The scale of the reaction was calculated assuming a 50% total yield from each chemical step, accounting for losses due to purification procedures. Using this scale, the first step would be carried out at 9 nmol requiring 423 nmol of the linker for 47 building blocks. This was synthesised using the solid supported method previously described. Accounting for 50% losses the next step would be at 4.5 nmol and the final at 2.25 nmol for each reaction. When pooled together this would result in a library of 99,405 compounds at a total quantity of 101 nmol. This much larger scale was used as scale up to this size has never been performed and many purification, analysis and handling methods would need to be trialled and applied at this scale. This scale of reaction was also known to perform well with the novel amide micellar chemistry involved in this reaction sequence.

6.4.2 – Ligation Strategy

Previously the ligation strategy used was for a 280 pmol reaction where multiple wells were used to scale the reaction to the preferred amount. When designing a larger member library, this is not ideal as it vastly increases the amount of work that needs to be carried out to synthesis and purify the library. If a 9 nmol reaction was required in the initial step, nearly 40 ligations would need to be carried out for every reaction that takes place. This would increase the amount of work needed considerably and increase the complexity of the overall reaction

process. To reduce the time and cost for this step, validations of ligation and phosphorylation reactions were carried out at increased concentrations.

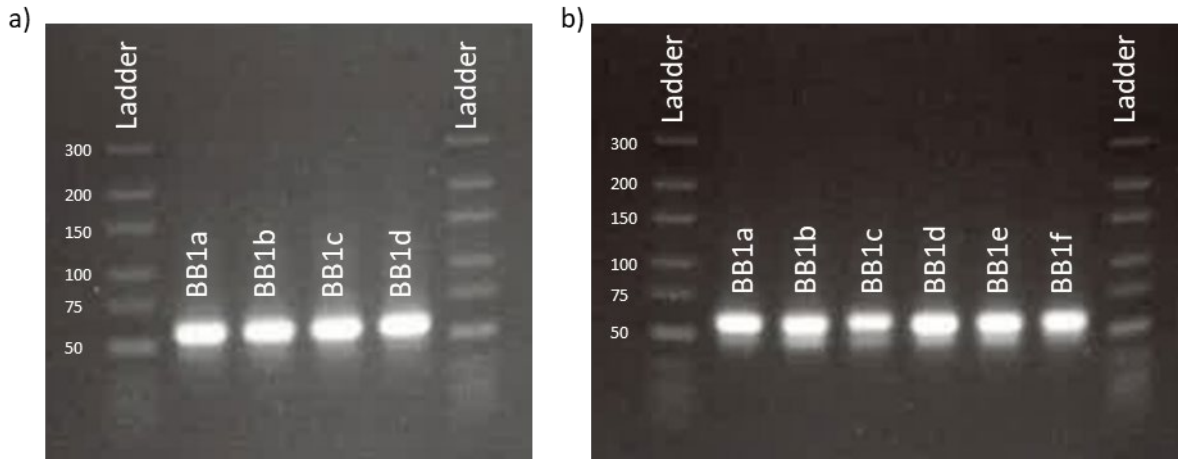


Figure 106 – Analysis of 5 nmol (a) and 10 nmol (b) ligations using 4% agarose gels.

Using similar conditions to those used previously it was found that 10 nmol of DNA could be successfully ligated. As these processes are enzymatic increasing the reaction time should also increase reaction progress. Phosphorylation reactions were carried out using the same amounts of reagents, in 20 µl reactions, but the DNA at a much higher concentration at 500 µM rather than 14 µM. These reactions were also left for 1 hour rather than 15 mins. The 20 µl reaction was then denatured at 75 °C prior to use in the ligation reaction. The ligations were carried out at 111 µM DNA in 90 µl solutions, using 3 µl of T4 ligase. These reactions were left for 16 hours and analysed by gel electrophoresis [Figure 106]. This concentration of DNA was well tolerated using these conditions and could be scaled to the 9 nmol reactions required for the library synthesis.

6.5 – Library Synthesis

The library synthesis was split into 3 stages with a ligation and chemical modification in each stage. The first stage would be to ligate the first building block and other sequences to the linker moiety, followed by a simple ethanol precipitation as a purification. The first amide coupling step would then take place all samples would be pooled and purified together. This would include a molecular weight filtration to remove any small DNA sequence starting materials from the ligations as well as any extra small molecule related matter. The entire library would then be hydrolysed using lithium hydroxide, followed by 2 ethanol precipitations to purify the library. The library would then be split and the process repeated

for the second building block. Once complete, the third building block would be added with corresponding sequence and capping primer sequence. No hydrolysis would be required at this stage as all amines are terminal. The library would then be purified by preparative HPLC ready for screening. As well as the library, a test 1x1x1 library would be synthesised in parallel to the larger library. This compound would be used to show that the resulting library can be PCR amplified and sequenced after the library synthesis steps have taken place.

6.5.1 – First Library Synthesis Step

The first step in the library synthesis was to introduce the initial sequences and first stage building block codes. The process was achieved using the same principles as the 6x6 prototype library synthesised as a proof of concept for the Suzuki-Miyaura coupling [Figure 107 a)]. The first ligation included sequences containing adapter, primer, scaffold codes, overhangs and building block 1 codes. Prior to ligations, the 5' terminus of any DNA strands that required ligating was phosphorylated. In this case it included sequences with identifiers PS, OH1BB1x and A'P'S'OH1', where the adapter sequence and BB1' sequence did not require phosphorylation. As codes PS and A'P'S'OH1' would be used in all ligation reactions they were phosphorylated in bulk in an Eppendorf tube rather than individually in PCR tubes as was done for the building block codes. The reactions were scaled to be the same concentration in all examples at 450 µM and then left to react for 1 hour.

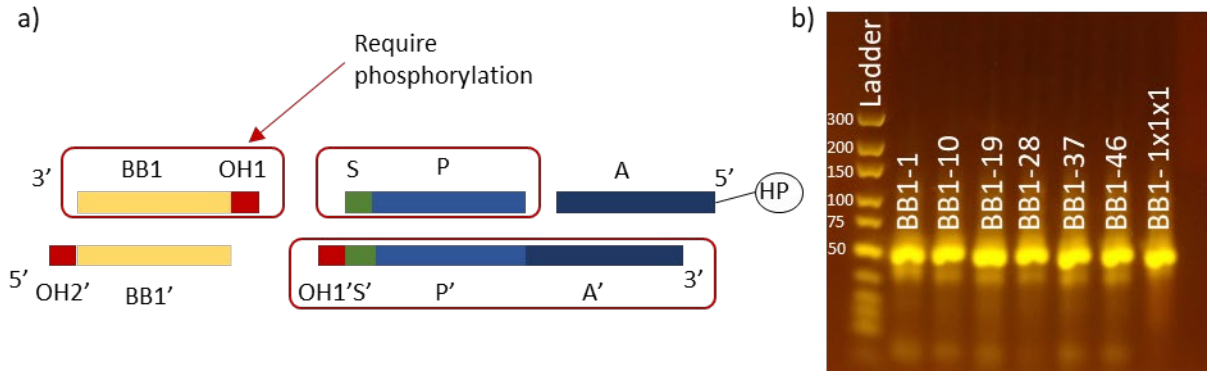


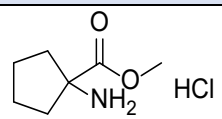
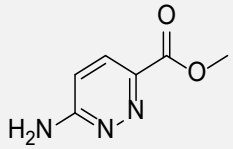
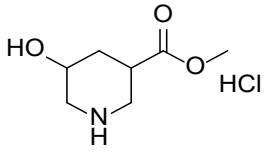
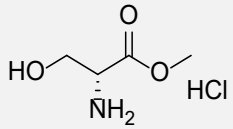
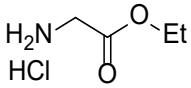
Figure 107 - a) Ligation plan used in the first library stage; b) 4% agarose gel electrophoresis results of first library ligations.

The ligations were then carried out in PCR tubes using 20 µl of each phosphorylation reaction mixture plus 9 µl of 1 mM solutions of non-phosphorylated A and OH2'BB1x'. These reactions were made up to a total of 90 µl at 100 µM of DNA and left for 16 hours before denaturing for 10 mins at 75 °C. Aliquots were taken across the diagonal of the PCR plate to analyse by

gel electrophoresis [Figure 107 b)]. The expected DNA sequence length was 50 and 54 base pairs, and it was clear that these ligations had been a success showing a distinct band at around the 50 base pair mark of the ladder. The 1x1x1 library was also analysed here showing similar results. Small amounts of excess DNA sequences are also visible, but these would be removed later using molecular weight filtrations. Each DNA sequence was then purified by ethanol precipitation and the individual pellets were dissolved in 4.5% TPGS-750-M in water.

Each DNA sequence was assigned to a corresponding amino ester building [Table 27][Appendix Table 3]. An eight-base pair sequence with a hamming distance of four was provided for each building block for identification in screening of the library. Each building block was dissolved in a 0.25 M NMP solution prior to reaction and 60 µl of each was added to the corresponding reaction well, along with 20 µl HOAT at 1 mg per 10 µl and the samples reduced to dryness using a Genevac centrifugal evaporator. This allows easy access to each building block with no need to weigh individual samples ready for reaction.

Table 27 – Representative amino esters, the consequent reaction wells and BB1 codes, as well as corresponding building block code sequences (yellow) and overhangs (red).

Amino ester	Reaction well	BB1 code	OH1BB1-x sequence (5'-3')	OH2'BB1-x' sequence (5'-3')
	A1	BB1-1	GTATTTATAACG	TAGGCGTTATAA
	B1	BB1-9	GTATTCACAGCA	TAGGTGCTGTGA
	B7	BB1-15	GTATCGAGATTC	TAGGGAATCTCG
	E5	BB1-37	GTATAGTCAGCT	TAGGAGCTGACT
	1x1x1	BB1-50	GTATGATCGACT	TAGGAGTCGATC

Once the building blocks had been dried, the corresponding DNA ligation products were added in 30 µl of 4.5% TPGS-750-M, along with 2,6-lutidine and DIC and the reactions heated to 40 °C for 3 hours in a sealed Paradox plate. After this time, the reactions were cooled to room temperature and combined, washing each reaction tube with 30 µl of water. The combined reactions were then diluted to 4 ml total volume and washed with DCM and ethyl acetate; centrifugation was required to remove the emulsion formed in this step. The resulting aqueous phase was then precipitated using ethanol. The pellet was dissolved in water and further purified using a 3000 Da spin filtration washing with water several times to remove any further organics and small ligation products. The 1x1x1 reaction product was purified separately in the same way as the bulk library.

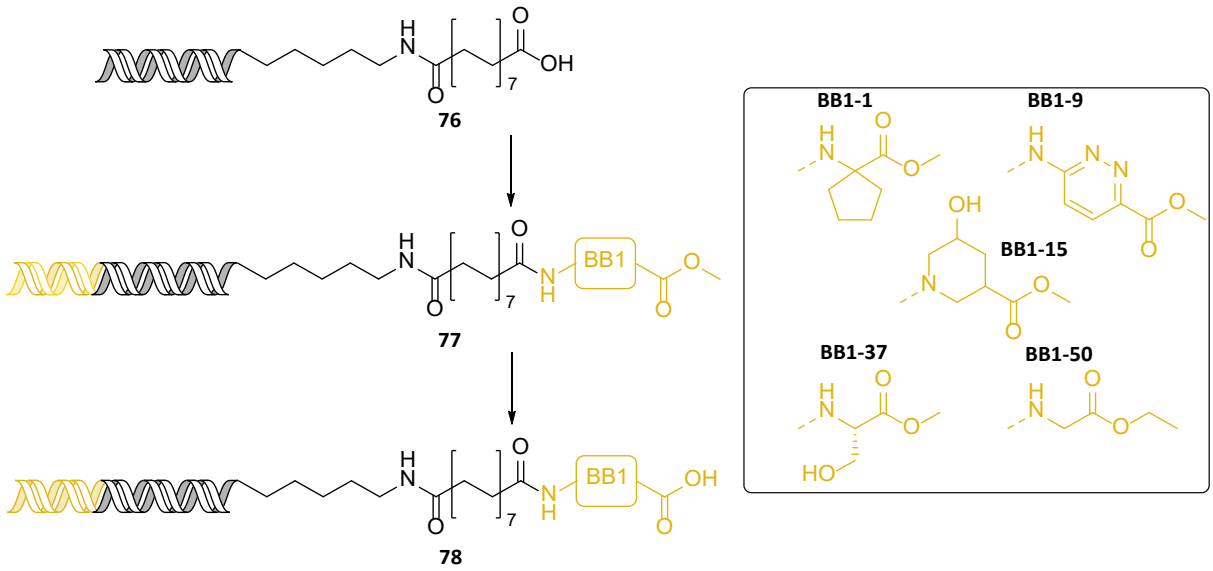


Figure 108 – Depiction of the first building block addition and example building blocks.

Once the library had been purified, a Nanodrop concentration was determined for both the library and the 1x1x1 test compound. It was discovered that the library was 325 nmol total, and the 1x1x1 test compound was 8.5 nmol. This showed that an overall library yield from this step of 72% and 94% for the 1x1x1 sample. This result was highly encouraging due to this being the first scaled library build performed using micellar technology.

To hydrolyse the esters, the entire library was treated with 0.25 M lithium hydroxide for one hour, prior to ethanol precipitation. This resulted in the library with building block 1 and corresponding coding sequence in place for further building block addition. A second library concentration was recorded at 311 nmol of library and 8.5 nmol of the 1x1x1 test compound.

The overall yield from this first coupling was highly encouraging and the library was ready to be split for the use in the second building block addition.

6.5.2 – Second Library Synthesis Step

The library was at 311 nmol total quantity after the first building block addition and was dissolved to produce a 1 mM solution. The second addition of building blocks was a repeat of the 47 amino esters used in the first addition, using different codes as identifiers. The library would be ligated to the 47 complimentary coding sequences using the overhangs present. Firstly, the entire library was phosphorylated in a large-scale reaction keeping the concentration of DNA at 325 µM, using an Eppendorf tube. The 5' terminus of the BB2OH2 sequence also required phosphorylation which was performed in PCR tubes in strips at 325 µM and 20 µl reactions. The reactions were again performed for 1 hour and denatured prior to addition into the ligation reaction.

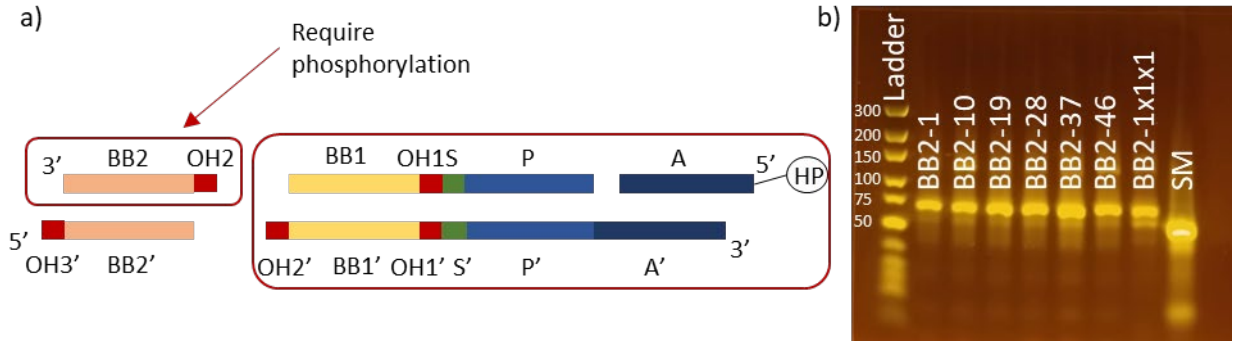
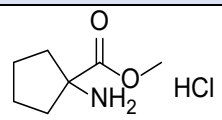
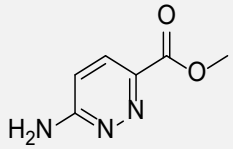
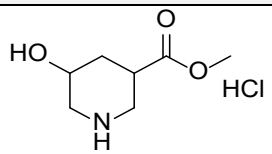
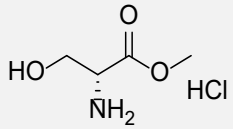
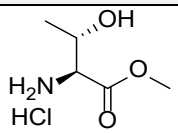


Figure 109 - a) Ligation plan used in the second library stage; b) 4% agarose gel electrophoresis results of second library ligations showing bands at ~65 BP

The ligations were then carried out in PCR tubes using 20 µl of each phosphorylation reaction including the library plus 9 µl of 1 mM solutions of the non-phosphorylated OH3'BB2x' sequence. The overall DNA concentration as 100 µM, using 65 µl overall reaction media and were carried out for 16 hours and then denatured at 75 °C as previous. This was also repeated for the 1x1x1 test compound. Gel electrophoresis analysis of the products showed that there was a significant band at approximately 65 base pairs in length [Figure 109 b)] with no visible starting material in any products. The DNA strands were then individually precipitated using ethanol and dissolved in 4.5% TPGS-750-M ready for the second amide coupling building block addition.

As with the first building block addition, each amino ester used was assigned an identifying 8 base pair sequence [Appendix Table 4][Table 28] and corresponding code. These once again were unique for each individual building block and were different from those used in the first addition. The building block solutions in NMP were added to the respective reaction wells along with HOAT and dried in a Genevac as previously. The solutions containing DNA were then added to the amino esters, along with DIC and 2,6-lutidine, sealed in a Paradox microplate and heated at 40 °C for 3 hours.

Table 28 - Representative amino esters, the consequent reaction wells and BB2 codes, as well as corresponding building block code sequences (orange) and overhangs (red).

Amino ester	Reaction well	BB2 code	OH2BB2-x sequence (5'-3')	OH3'BB2-x' sequence (5'-3')
	A1	BB2-1	CCTACAATCGGT	CGTAACCGATTG
	B1	BB2-9	CCTATCATGATA	CGTATATCATGA
	B7	BB2-15	CCTACCATGGCG	CGTACGCCATGG
	E5	BB2-37	CCTACCCGAAGC	CGTAGCTTCGGG
	1x1x1	BB2-50	CCTACGCGCACA	CGTATGTGCGCG

Reactions were combined and purified using the same procedures as the previous example, by washing with organic solvents, ethanol precipitation and molecular weight spin filtration. The 1x1x1 library was purified in the same manner. The two samples were dissolved in water and analysed by Nanodrop, indicating a yield of 186 nmol for the library and 5.6 nmol of the 1x1x1 compound. To hydrolyse the products, they were shaken with 0.25 M lithium hydroxide for 1 hour followed by an ethanol precipitation. The pellets were dissolved in water and

analysed by Nanodrop. Final yields were 151 nmol for the library and 4.3 nmol for the 1x1x1 compound. An overall yield of 51% was found for the library in this step. However, some losses were also seen in the hydrolysis maybe due to inefficient ethanol precipitation at -20 °C. However, there was still a full yield of over 50%, which is highly encouraging for a library synthesis. This second stage was therefore successful and the DEL was ready for the final capping amine addition.

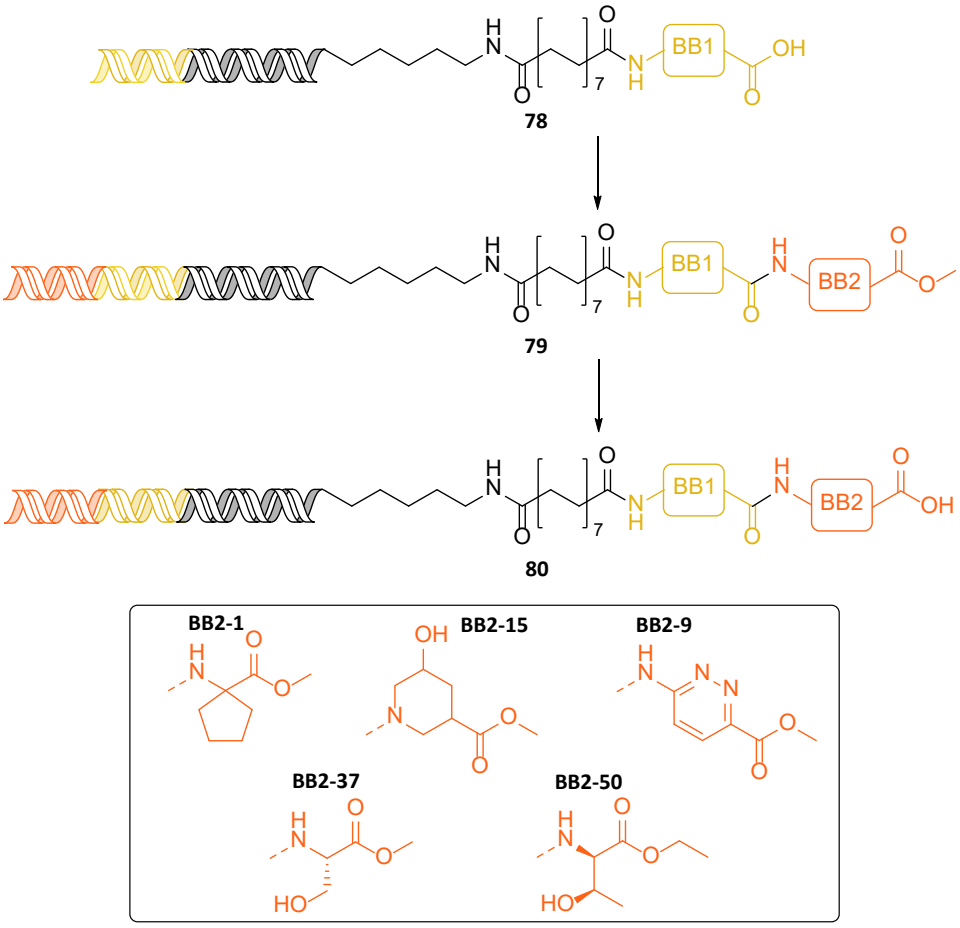


Figure 110 – Depiction of second building block addition, including amino ester examples.

6.5.3 – Third Library Synthesis Step

The total library was 151.3 nmol after the second library synthesis step had taken place. The library was dissolved in water to create a 1 mM solution of DNA. In the third library step an amine cap would be used to complete the library synthesis, which included a variety of amines and also some bespoke potential covalent binders such as bromoalkynes and nitriles [Appendix Table 5] to aid in the creation of the targeted SARS-Cov-2 M^{pro} library. When validating the chemistry, 45 amines were chosen to be utilised in this library step and so the library would be split over 45 well with 45 individual identifying sequences. As this step is also

the last step of the library synthesis, a longer sequence was ligated incorporating the individual sequence identifiers as well as an extension of the capping primer sequence for use in sequencing. These sequences at 31 and 35 base pairs in length. The P2BB3xOH3 sequence was phosphorylated at 165 µM total concentration of DNA in individual wells. The library was also phosphorylated once again in bulk ready to be divided into the individual ligation reactions. The 1x1x1 compound was also phosphorylated using building block code 50 as its identifier.

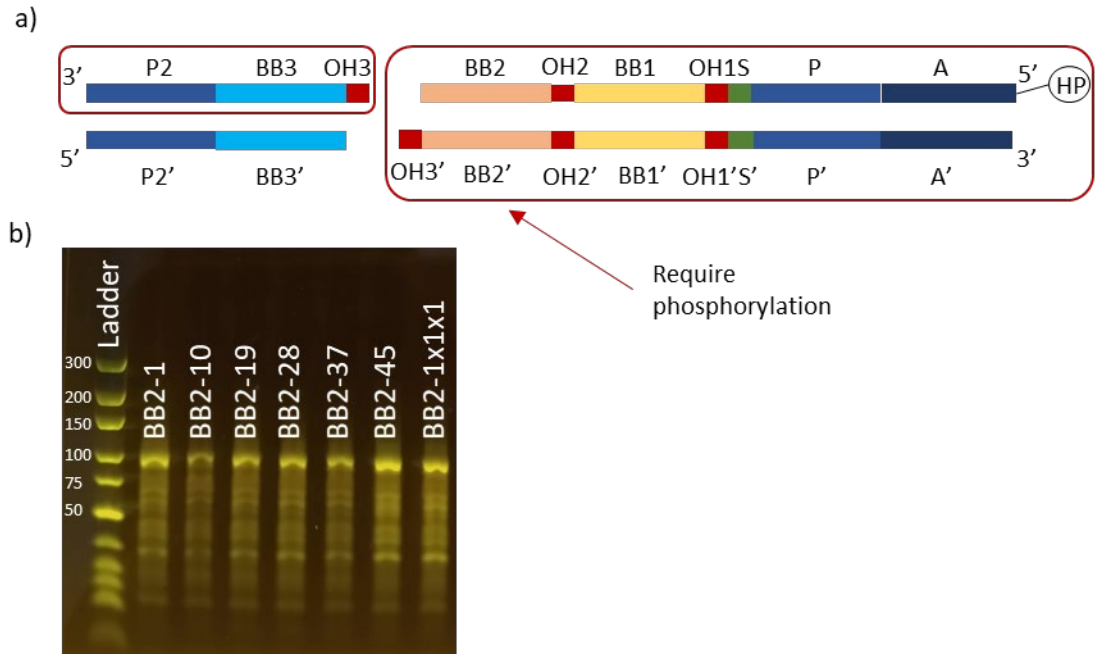
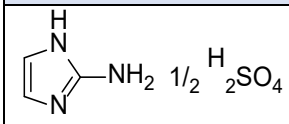
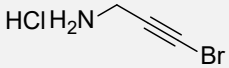
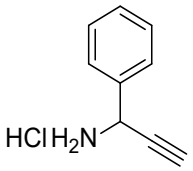
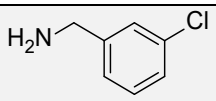
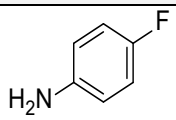


Figure 111 - a) Ligation plan used in the third library stage; b) 4% agarose gel electrophoresis results of third library ligations showing bands at ~100 BP.

Once phosphorylation had taken place the DNA sequences were transferred to the corresponding PCR tubes at a concentration of 55 µM and ligated for 16 hours using T4 ligase. Gel electrophoresis was used to analyse the products [Figure 111 b)] where bright bands were seen at ~100 BP (expected 97 BP). Some other bands were also visualised for shorter sequences, including a faint band at 35 BP corresponding to unreacted BB3 sequences, however no starting material bands were visible at 70 BP length. The shortmers would be removed later in the synthesis by molecular weight filtration and a final preparative HPLC purification. The DNA strands were individually precipitated using ethanol and dissolved in 4.5% TPGS-750-M ready for the final amide coupling building block addition.

Table 29 – Representative amines, the consequent reaction wells and BB3 codes, as well as corresponding building block code sequences (light blue), primers (navy blue) and overhangs (red).

Amine	Reaction well	BB3 code	OH3BB3-xP2 sequence (5'-3')	P2'BB3-x' sequence (5'-3')
	A1	BB3-1	TACGACCCCAGGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACCTGGGGT
	B6	BB3-14	TACGTCCAGAGTTGACCT CAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAACTCTGGA
	C3	BB3-19	TACGGTCGGCTGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACAGCCGAC
	D7	BB3-31	TACGGACGTCGCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGCGACGTC
	1x1x1	BB3-50	TACGACGGCACCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGGTGCCGT

Each amine was assigned a corresponding 8 base pair sequence and individual identifier code [Table 29][Appendix Table 5]. The 1x1x1 sample again utilised a building block represented in the library but used building block code 50 as the identifier. This allows the 1x1x1 to have a unique sequence to any of those in the library. Before the reaction could take place, each amine was dissolved in a 0.25 M solution of NMP, 60 µl removed and dried in the Genevac along with HOAT. The TPGS-750-M solutions containing DNA were then added to the corresponding amines along with DIC and 2,6-lutidine, sealed in a Paradox microplate and reacted at 40 °C for 3 hours.

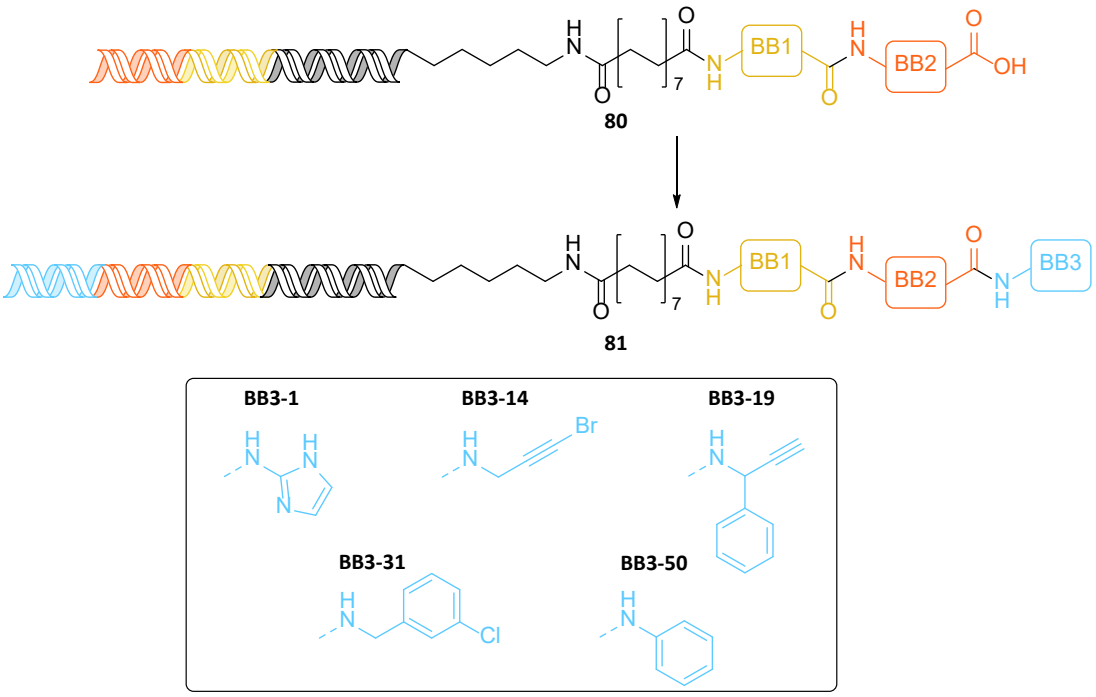


Figure 112 - Depiction of third building block addition, including amine examples.

Once complete the library containing wells were combined, washed with DCM and ethyl acetate, ethanol precipitated and redissolved in water. Further purification as achieved using a 10000 Dalton molecular weight spin filter to remove any shortmers and organics from the previous reactions. A 10000 Dalton spin filter was chosen as the sequences added in this step were over 3000 Daltons, to remove unreacted shortmer DNA sequences. The same purification process was carried out for the 1x1x1 library.

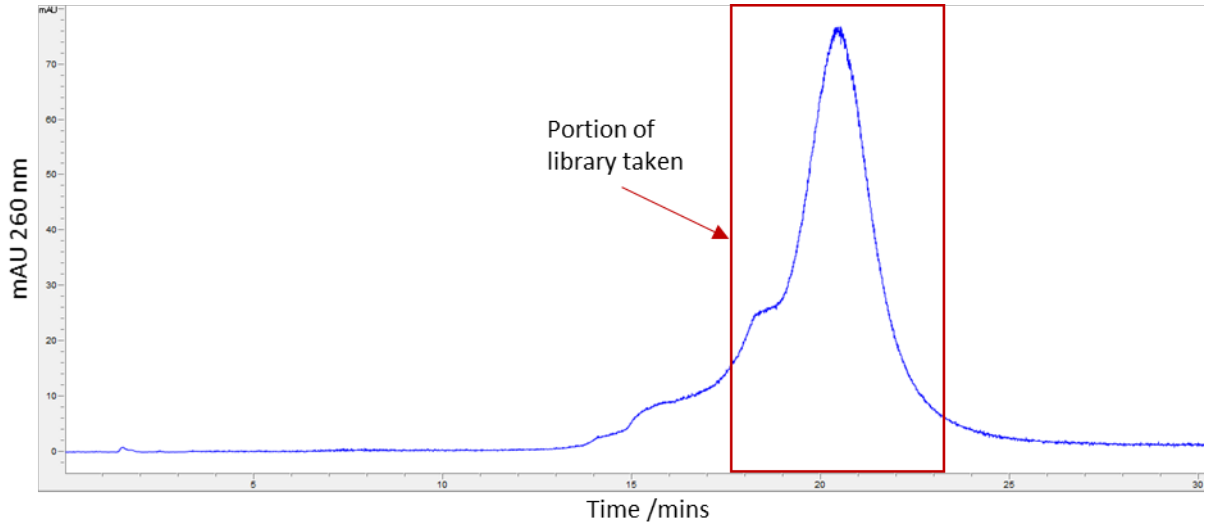
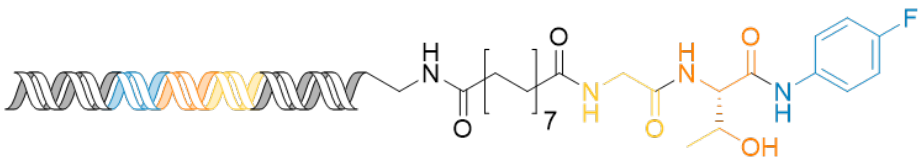


Figure 113 - HPLC trace of overall library using 16.3 mM TEA and 400 mM HFIP in water buffer with methanol, showing portion of the library separated.

A final purification was carried out on the entire library using preparative HPLC. To purify DNA strands of ~100 base pairs in length it was found that using buffers containing 16.3 mM TEA and 400 mM HFIP gave clear separation and ideal retention times on C_{18} columns.[155] This is a slightly different system to that used for analysis of the shorter 14mer compounds. Ion pairing HFIP and TEA buffers have been used extensively for analysis and separation of DNA,[156] however concentrations must be modified depending on the oligonucleotide length that is being analysed. It is thought that the TEA concentration rather than HFIP that plays the most important role in separation of oligonucleotides due to the impact on ion pairing at larger oligonucleotide lengths.[157] Trial separations on a small amount of the library showed that using a buffer solution of 16.3 mM TEA and 400 mM HFIP in water with methanol gave great purification if the library sample. The HPLC trace shows a broad peak as there are nearly 100,000 different sequences and organic moieties in the library. After purification, the part of the trace containing the library was dried in the genevac and redissolved in water, yielding 79 nmol of overall library. The overall percentage yield was 19% over the 3 steps including purifications. The 1x1x1 library was not purified by preparative HPLC as it was used straight in a PCR reaction as a proof of concept that the amplifiable sequence was still in place.

6.6 – PCR and Sequencing of 1x1x1 Compound

As this is the first library to be synthesised using micellar technology, the sequence must be proven to still exist undamaged in the product for the library to be successfully used in a screen. It is not easy to sequence an entire 100,000-member library to test for DNA fidelity but the 1x1x1 library allows a more facile assessment of whether the DNA sequence is intact and analysable. Before sequencing can take place PCR amplification of the compound is used as a test of sequence fidelity. If the compound does not amplify using PCR then the sequence has likely degraded in the library synthesis.



Linker strand (5'-3')
A P S OH1 BB1 OH2 BB2 OH3 BB3 P2
GTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCA ACTACATGGTCTACA
Complimentary strand (5'-3')
P2' BB3' OH3' BB2' OH2' BB1' OH1' S' P' A'
TGTAGACCATGTAGTTGAGGTCAGGTGCCGTCGTATGTGCGCGTAGGAGTCGATCATACAGCCAAATCCGTTCACACCGACC TGAATTCGGCAAGAC

Figure 114 – the 1x1x1 tool compound, with corresponding 97mer coding sequence.

The same primers were used in the PCR as for the prototype Suzuki-Miyaura 6x6 library created previously [Table 15]. These primers were known to successfully amplify the library, and so should amplify the 1x1x1 tool compound. The process of PCR amplification was carried out using 1 µg (0.5 µl) of the DNA template in all examples and either 0.2 µM, 2 µM primers concentration along with a negative control with the primers omitted [Figure 115 a)]. In total 40 denature, anneal and extension PCR cycles were carried out on these substrates prior to analysis by gel electrophoresis on a 4% agarose gel [Figure 115 b)].

a)

	0.2 µM Primers	2 µM Primers	No Primers
amplitaq gold master mix	25 µl	25 µl	25 µl
FSP-P2'	0.1 µl	1 µl	0 µl
RSP-P	0.1 µl	1 µl	0 µl
water	24.3 µl	22.5 µl	24.5 µl
DNA template	0.5 µl	0.5 µl	0.5 µl

b)

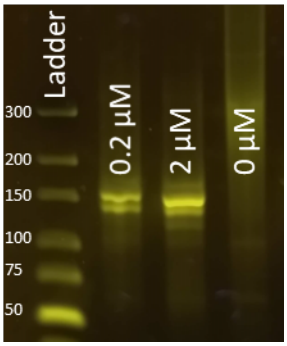


Figure 115 – a) PCR method used including negative control with no primer; b) 4% agarose gel electrophoresis results of PCR amplification showing bands at ~140 BP.

Analysis by agarose gel electrophoresis showed that there was a clear band around 140-150 base pairs in length, but as expected none in the negative control. The expected length of the DNA strand post amplification using the 33-base pair NGS extensions was 148-base pairs. In this example the 14mer adapter sequence is lost in amplification as the code is not represented in the primers. This means that the amplified DNA runs from primer 1 to primer

2, including the NGS extensions at each terminus. This result shows that the 1x1x1 library can be successfully amplified by PCR, showing that there should not be any damage to the primer sequences and no damage to the coding sequences.

As further proof that no damage has occurred to this example library sanger sequencing was used to identify the bases in the coding sequence. Initially this was submitted with the longer NGS primers, but the results returned as negatives, so shorter 19mer primers were created for sequencing. These results were also inconclusive as the first 50-100 bases when analysed by Sanger sequencing are often not recognisable. The PCR product was therefore sent for analysis by NGS [appendix table 2]. This resulted in >80% of the 77000 reads corresponding to the expected sequence, with most of the others being small amounts of shortmers of non-ligated material. This shows that the ligations were successful, and the chemistry is compatible with a DNA encoded library synthesis allowing for sequencing of the hit compounds from selection experiments.

6.7 - Conclusions

Production of a DEL is the epitomal final step for proof that any DNA compatible reaction is truly applicable to DEL synthesis. The 3-step synthesis of a DEL using only novel micellar reactions has been a success showing high yields and the ability to amplify the library product. A new sequencing strategy has been used in the synthesis of this DEL along with a hydrophobic, non-PEGylated linker to promote the association of the substrate with the micelles.

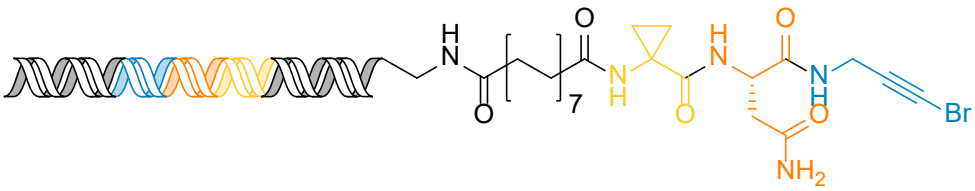


Figure 116 – Example SARS-CoV-2 M^{pro} targeted DEL substituent.

The novel protein target M^{pro} of SARS-CoV-2 has been identified as a target for antiviral activity. This novel coronavirus has had a devastating impact worldwide in 2020 and represents a vital target for new antiviral drug developments. The design of a novel targeted M^{pro} library incorporating covalent amine end caps and a variety of amino esters, was used with the aim of increasing the chance of finding novel M^{pro} inhibitors. After validation reactions, 47 amino esters and 45 capping amines were chosen to produce this targeted

library. The library also contained non-electrophilic amines so the library could be utilised in normal protein screening vastly increasing the total library efficacy. The library produced utilised a new method of DEL synthesis using micellar reactions which required little purification and high yielding reactions. This library could not be produced without the high yielding efficient micellar promoted reverse amide conditions discovered previously. The hydrophobic linker aided in reaction conversion rates for the library build, increasing the overall building block diversity and selection. The library was successfully synthesised in 3 steps using known ligation strategies on novel sequence designs and novel micellar reaction conditions. The result was a library of 99,405 members all with attached corresponding DNA sequences as identifiers. As a further proof a concept, a 1x1x1 compound was synthesised in parallel to the library and used to prove that the DNA still retained fidelity for amplification and sequencing. In future, this library will be screened against SARS-CoV-2 M^{pro} and will be made available for screening against other protein targets.

Chapter 7 – Conclusions and Future Work

The development of DNA compatible chemistry is perhaps the most significant way to expand the applications of DEL technology. There are an increasing number of reactions that can be utilised in a DNA compatible manner. Many of these reactions are based on modifications of existing reactions to operate under aqueous conditions or to modify the DNA; for example, by appending it to solid support to allow reactions using typical organic conditions. This emerging field has led to a large number of chemical reactions being successfully applied to DNA conjugates. However, there are still clear problems with the approach. Individual libraries offer limited chemical diversity. This is often overcome by screening multiple libraries simultaneously. Within a library, diversity can be limited due to reagents with different chemical properties not being tolerated in the validation reactions thus reducing the chemical variation for each building block addition. Both of these problems can be addressed by either generating new reactions capable of utilising extensive numbers of building blocks or improving the current reactions to expand the scope of the building blocks used allowing the use of more highly decorated and chemically diverse reagent sets.

One such approach improves the chemical diversity of a library was the use of the novel paradigm “encoded transformations”. New and existing DNA compatible reactions were developed to perform a transformation on a reactive core of a DNA conjugated compound. This chemistry was demonstrated to introduce a variety of different chemical structures to the core of the molecule, allowing a novel method for developing chemical diversity in an encoded library. These modifications included a variety of different shapes, electronics, and functionality showing that a diverse set of molecules could be synthesised and transformed using this method. This technique has yet to be applied to DNA conjugates, but the reactions were proven to proceed under DNA compatible conditions and would likely require little further modification to be utilised in DELs.

Modifying existing chemical methods has been a successful technique to create useful DNA compatible reactions. However, the scope of these reactions is often limited to few building blocks, severely reducing the diversity of the resulting libraries. Reactions such as amide coupling, S_NAr, Suzuki-Miyaura, reductive aminations and Buchwald couplings are used in most chemical syntheses in medicinal chemistry. All of these reactions are limited either by the number of compatible building blocks or specific substrates or scaffolds. For example,

S_NAr reactions have been widely employed but efficient reactions with a wide range of amines have typically only been observed with reactive substrates such as triazines. New chemical reactions can be developed showing DNA compatibility, but the majority of libraries produced using these new chemistries rely upon a subsequent chemical reaction to add diversity such as the amide coupling or Suzuki-Miyaura reaction. Therefore, increasing the efficiency of these reactions across a wider substrate scope is vital for producing high fidelity diverse chemical libraries.

Micellar chemistry was shown to promote reactions off DNA and was envisaged to have a positive impact on the reactions of DNA conjugates. Using micelles was believed to improve the reaction rate as well as have a positive impact of on the integrity of the overall DNA strand due to the reactants becoming encapsulated in the micelle, minimising the chance of DNA damage. As these micelles act as solvent for the reaction to take place, some of the constraints of using water soluble reactants and reagents are removed. This allows the catalyst to be tuned to the reaction system rather than having to be tuned to solubility and water compatibility. It allows transition metal chemistry to be modified using a variety of different ligands which in turn vastly increases the number of substrates and building blocks that react efficiently. This is vital when creating a reaction that can be applied to a large number and variety of chemical entities.

A Suzuki-Miyaura reaction was chosen as the foundation for testing micellar promoted reactions and their application on DNA. This reaction is used extensively in medicinal chemistry but was shown to have small substrate scope and low reported conversions when used in DELs. This reaction was therefore an ideal candidate for improvement by micellar chemistry. To begin the reaction was run using modified literature conditions and the designer surfactant NOK which showed significant advantages over the literature conditions on a simple iodophenyl headpiece. The conditions were then modified with the use of a FED experiment to significantly improve the scope and reactivity, expanding the scope to a variety of different boronic acid, ester and MIDA substrates. The FED was a vital tool to improve the reaction conditions as it identified and helped to improve multiple second order and squared interactions that would not be simply improved by other means. The resultant chemistry was extremely reliable showing conversions of 90% + for all substrates tested and showing no DNA damage or side products such as dehalogenation. The reactions also did not proceed as well

without the surfactant, showing that the surfactants promote the reaction. This reaction was used to build a proof of concept 2 step library using 6 building blocks in each stage, producing a 36-member DEL. The library was designed with a complete DEL coding strategy that is applicable to larger libraries. Each building block had an individual codon, and the overall sequence incorporated a scaffold code and primer sequences. The resulting library was successfully amplified by PCR and sequenced using NGS, proving that the chemistry is compatible with encoded library synthesis. This novel reaction system using commercially available surfactants and reagents is a vital step forwards in the discovery of novel DNA compatible chemistries and provides the most efficient biaryl coupling procedure described to date.

The micelle promoted methodology was further expanded to amide formation. Amide couplings are the most commonly used reactions in both DEL synthesis and medicinal chemistry more generally. There is a huge number of diverse building blocks with acid and amine functionality that can be utilised in library design. Also, many preliminary building block additions are carried out on an DNA conjugated amine linker utilising an amide coupling as the first reaction. Current amide coupling conditions perform poorly on many substrates, resulting in low conversion and side product formation. This is even more challenging in the reverse amide coupling, in which the acid component is attached to the DNA conjugate. The use of micelles and a hydrophobic linker was shown to improve the reactivity of the reverse and forward amide coupling reaction considerably. The reverse amide reaction proceeded with vastly improved conversions showing nearly 75% of reagents tested coupled at over 50% efficiency. The forward amide coupling was even more successful showing high conversion rates over every reactant tested. This improved reaction opens a variety of new options for the design of chemical libraries. Using DNA conjugated acids allows the synthesis of chemical libraries that were previously inaccessible and improving the forward amide coupling reaction allows for more expansive building block selection.

The reverse amide coupling was successfully used in the design of a targeted SARS-CoV-2 M^{pro} library. The library was designed to incorporate covalent binding motifs such as bromoalkyne, alkyne and nitrile capping amines as well as introducing more variety with the addition of non-covalent binding amines. A 3-step synthesis was designed using sequential reverse amide couplings to create a library of 99,405 individual members. This library was encoded with 8

base pair sequences for each building block. Each building block was validated on a test substrate prior to the library build and these consisted of 47 amino esters, which were deprotected after each amide coupling and 45 capping amines passing the validation reaction. The library synthesis was successfully completed using simple purification procedures between each library synthesis step with a HPLC purification of the final library. A 1x1x1 proof of concept DNA conjugate was synthesised in parallel to the library and successfully PCR amplified. This library will be used as a targeted screening tool to find hit compounds against SARS-CoV-2 M^{pro}, as well as being a useful library for screening other protein targets.

Micellar promoted reactions could become a vital tool in the design and synthesis of new DELs. The application of this new technique however is not limited to Suzuki-Miyaura or amide couplings. Micellar promotion has been utilised significantly in other reactions in organic chemistry and these applications could potentially be extended to DNA conjugates. Micellar conditions perform extremely well for transition metal catalysed processes, suggesting that this approach could be applied to Buchwald, Sonogoshira, Heck and Ullmann reactions on DNA. These reactions would follow on from the Suzuki-Miyaura coupling in vastly increasing the diversity of libraries with access to previously unreactive building blocks. Other reactions such as S_NAr and reductive aminations could also be promoted by these conditions. Finally, to further improve the amide coupling reaction, new surfactants such as MC1 could be explored.

DELs are an extremely efficient method of screening protein targets and are increasingly used by big pharma, biotechs and academia. The technology relies upon quick and relatively simple synthesis of huge libraries, and their rapid screening in one vessel. Multiple libraries can be screened simultaneously providing a highly efficient method of screening huge numbers of compounds. The synthesis of these libraries has been difficult due to the limited number of high yielding reactions available. The development of more efficient and broadly applicable reactions could provide access to more diverse, higher fidelity libraries, and therefore have a large impact on the future of drug discovery.

Chapter 8 – Experimental

8.0 – Techniques

8.0.1 – Safety

All procedures were carried out according to the Newcastle University Chemical Safety Policy. COSHH risk assessments were completed before starting any practical work. Overnight forms were completed when required.

8.0.2 – Solvents and Reagents

Chemicals were purchased from Fluorochem, Sigma-Aldrich, Acros and other standard suppliers, and were used without further purification. Bottles of anhydrous solvents using SureSeal™ or Acroseal™ were purchased from Sigma-Aldrich or Acros, respectively. Deuterated solvents for NMR spectra determination were purchased from Sigma-Aldrich.

Amino esters used in the library build were purchased from Enamine and used without further purification. Bespoke amines for use with the covalent SARS-CoV-2 M^{pro} library were synthesised by Pharmaron.

Solid supported 14mer DNA and corresponding strand were custom synthesised by Sigma-Aldrich. All other DNA building blocks were purchased in plates or tubes directly from IDT and provided as 100 µM solutions in water, or dry pellets.

8.0.3 – Analytical Techniques

LC-MS analyses were conducted using a Waters Acquity UPLC system with PDA and ELSD. When a 2 min gradient was used, the sample was eluted on an Acquity UPLC BEH C18, 1.7µm, 2.1 x 50mm, with a flow rate of 0.6 ml/min using 5-95% 0.1% HCOOH in MeCN. HRMS analyses were conducted using an Agilent 6550 iFunnel QTOF LC-MS in either positive or negative mode with an Agilent 1260 Infinity UPLC system. The sample was eluted on Acquity UPLC BEH C18 (1.7 µm, 2.1 x 50 mm) with a flow rate of 0.7 ml/min, at a gradient of 1.2 min 5-95% 0.1% HCOOH in MeCN with 0.1% aq. HCOOH. Calculated exact masses were quoted from ChemDraw Professional 15.0. FTIR spectra were measured using an Agilent Cary 630 FTIR as a neat sample.

DNA mass spectrometry was conducted on an Agilent 6550 QTOF in negative mode, using a standard 3200 m/z maximum and a 2GHz extended dynamic range. Drying gas temperature

was 260°C at 12l/min, sheath gas temperature was 400°C at 12 l/min, nebulizer at 45 psig, VCap voltage of 4000V and nozzle voltage of 2000V. The LC was carried out on an Agilent 1260 infinity 2 on an Agilent Advancedbio oligonucleotides column, 2.1x150 mm where the gradient was run at 0.4 ml/min from 10% MeOH to 40% MeOH over 8 mins against a 200 mM HFIP:8 mM TEA buffer solution. A 3 min flush at 95% MeOH preceded each run; or an Agilent Advancedbio oligonucleotides column, 2.1x100 mm where the gradient was run at 0.8 ml/min from 10% MeOH to 50% MeOH over 4 mins against a 50 mM HFIP:15 mM DIPEA buffer solution. A 1 min flush at 95% MeOH preceded each run. Analysis of data was carried out by Agilent Qualitative Analysis version 7; or an Agilent Advancedbio oligonucleotides column, 2.1x100 mm where the gradient was run at 0.8 ml/min from 10% MeOH to 70% MeOH over 5 mins against a 50 mM HFIP:8 mM DIPEA buffer solution. A 1 min flush at 95% MeOH preceded each run.

1H NMR spectra were obtained using a Bruker Avance III 500 spectrometer using a frequency of 500 MHz. 13C and 19F NMR spectra were acquired using the Bruker Avance III 500 spectrometer operating at a frequency of 125 MHz, and 470 MHz, respectively. The abbreviations for spin multiplicity are as follows: s = singlet; d = doublet; t = triplet; q = quartet, quin = quintet, sept = septet and m = multiplet. Combinations of these abbreviations are employed to describe more complex splitting patterns (e.g. dd = doublet of doublets) and where broadening of the peak is observed, spin multiplicity is accompanied by the prefix br = broad.

Gel electrophoresis were carried out using either 4% Agarose gels cast in house and run in 1x TBE buffer, or using prepacked 4% E-Gel™ EX Agarose Gels on a Invitrogen E-Gel Power Snap Electrophoresis System, using Invitrogen™ Ultra Low Range DNA Ladders.

DNA concentrations were calculated using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, pipetting 1 µl of sample on the loading plate.

8.0.4 – Chromatography and Equipment

Flash column chromatography purifications were carried out using Biotage SP4 and Isolera automated flash system with UV monitoring at 278 nm and collection at 254 nm. Grace Resolve pre-packed flash cartridges were used for normal phase separations.

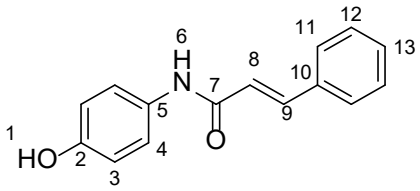
Preparative HPLC purification was carried on an Agilent 1260 infinity system using a Phenomenex Clarity 5 um Oligo-RP column, 10x150 mm. The gradient was run at 5 ml/min from 10% MeOH to 90% MeOH over 22 mins against a 200 mM HFIP:8 mM TEA buffer solution. Fractions were analysed at 260 nm wavelength; or using a Phenomenex Clarity 5 um Oligo-RP column, 21.2x250 mm, the gradient was run at 20 ml/min from 10% MeOH to 60% MeOH over 11 mins against a 200 mM HFIP:8 mM TEA buffer solution. Fractions were analysed at 260 nm wavelength.

Preparative HPLC purification of the 100mer library was carried on an Agilent 1260 infinity system using a Phenomenex Clarity 5 um Oligo-RP column, 10x150 mm. The gradient was run at 5 ml/min from 10% MeOH to 50% MeOH over 40 mins against a 400 mM HFIP:16.3 mM TEA buffer solution. Fractions were analysed at 260 nm wavelength.

PCR, ligation and phosphorylation were carried out using a Bio-Rad MJ Mini Personal Thermal Cycler.

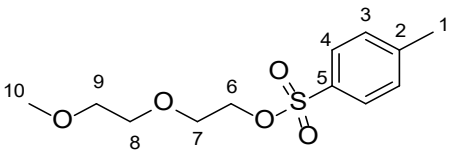
8.1 – Organic Procedures

8.1.1 – Synthesis of N-(4-hydroxyphenyl)cinnamamide



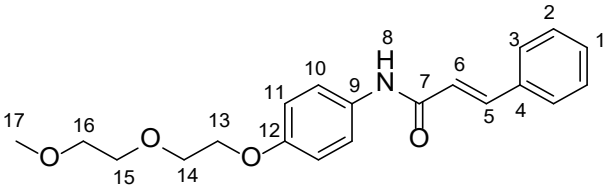
EDC.HCl (1.3 g, 6.8 mmol) was added to a stirred solution of cinnamic acid (1.0 g, 5.6 mmol) and trimethylamine (91 µl, 12 mmol) in DCM (20 ml) at 0°C. After 20 minutes 4-aminophenol (610 mg, 5.6 mmol) was added. The reaction was allowed to warm to room temperature and stirred overnight. The reaction mixture was diluted with water (30 ml) and extracted with 1:4 IPA:DCM (3x30 ml). The organics were combined, dried over magnesium sulphate and concentrated. (932 mg, 70% yield). Rf = 0.24 (DCM/MeOH 19:1); IR: νmax/cm-1 3020 (br, N-H stretch), 1650 (C=O stretch), 1616 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.66 – 7.57 (m, 3H, H-9 and H-11), 7.48 – 7.37 (m, 5H, H-4, H-12 and H-13), 6.74 - 6.79 (m, 3H, H-8 and H-3); 13C NMR (126 MHz, Methanol-d4) δ 164.93 (C-7), 154.11 (C-2), 140.81 (C-9), 134.94 (C-10), 130.46(C-5), 129.49 (C-13), 128.58 (C-11), 127.47 (C-12), 121.72 (C-4), 120.96 (C-8), 114.89 (C-3); LCMS m/z 240.1 ([M+H]+); HRMS (ESI): m/z calcd for $C_{15}H_{13}NO_2$: 240.1019 ([M+H]+); found: 240.1098 ([M+H]+)

8.1.2 – Synthesis of 2-(2-methoxyethoxy)ethyl 4-methylbenzenesulfonate



2-(2-Methoxyethoxy)ethan-1-ol (620 µl, 5.0 mmol) was stirred in pyridine (10 ml) at 0°C. Toluene sulphonyl chloride (1.1 g, 6.0 mmol) was added portion wise, and the mixture stirred at room temperature for 2 hours. The reaction mixture was diluted with water (30 ml) and extracted with ethyl acetate (3x30 ml). The organics were combined, dried over magnesium sulphate and concentrated. The crude product was then purified by column chromatography (24g, 0-100% ethyl acetate in petroleum ether) yielding the product as a colourless oil (797 mg, 58%). R_f = 0.61 (DCM/MeOH 19:1); IR: νmax/cm-1 2800 (C-H stretch), 1401 (S=O stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.82 – 7.78 (m, 2H, H-4), 7.36 – 7.32 (m, 2H, H-3), 4.19 – 4.15 (m, 2H, H-6), 3.71 – 3.67 (m, 2H, PEG-H), 3.60 – 3.56 (m, 2H, PEG-H), 3.50 – 3.46 (m, 2H, PEG-H), 3.35 (s, 3H, H-10), 2.45 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 143.65 (C-5), 133.76 (C-2), 129.67 (C-3), 128.21 (C-4), 71.45 (PEG), 70.75 (PEG), 70.02 (PEG), 67.87 (PEG), 59.31 (C-10), 21.46 (C-1); LCMS m/z 275.1 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{12}H_{18}O_5S$: 275.0948 ($[M+H]^+$); found: 275.0982 ($[M+H]^+$)

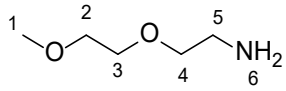
8.1.3 – Synthesis of N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)cinnamamide



N-(4-Hydroxyphenyl)cinnamamide (500 mg, 2.1 mmol) and potassium carbonate (1.1 g, 7.6 mmol) were suspended in acetonitrile (8 ml). Then 2-(2-methoxyethoxy)ethyl 4-methylbenzenesulfonate (520 mg, 1.9 mmol) in acetonitrile (2 ml) was added dropwise. The reaction mixture was stirred overnight at room temperature, at which point the reaction was filtered and the filter cake washed with acetonitrile. The solvent was removed under vacuum and purified by column chromatography (0-5% methanol in DCM) to yield a pale yellow solid (594 mg, 92%). R_f = 0.36 (DCM/MeOH 19:1); IR: νmax/cm-1 3061 (br, N-H stretch), 1648 (C=O

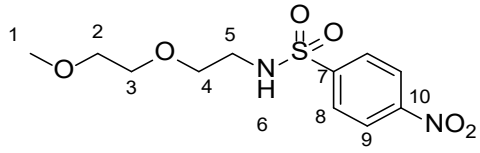
stretch), 1608 (C=C stretch); 1H NMR (400 MHz, Chloroform-d) δ 7.74 (d, J = 15.5 Hz, 1H, H-5), 7.56 – 7.48 (m, 4H, H-3 and H-10), 7.42 – 7.35 (m, 3H, H-1 and H-2), 6.90 (d, J = 9.0 Hz, 2H, H-11), 6.52 (d, J = 15.5 Hz, 1H, H-6), 4.13 (t, J = 5.2 Hz, 2H, H-13), 3.86 (t, J = 4.8 Hz, 2H, H-14), 3.75 – 3.71 (m, 2H, H-15), 3.61 – 3.57 (m, 2H, H-16), 3.40 (s, 3H, H-17); 13C NMR (126 MHz, Chloroform-d) δ 163.82 (C-7), 155.63 (C-12), 141.91 (C-5), 134.74 (C-4), 131.42 (C-9), 129.85 (C-1), 128.86 (C-3), 127.91 (C-2), 121.60 (C-10), 120.98 (C-6), 115.02 (C-11), 71.97 (PEG), 70.77 (PEG), 69.79 (PEG), 67.70 (C-13), 59.08 (C-17); LCMS m/z 342.3 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{20}H_{23}NO_4$: 342.1700 ($[M+H]^+$); found: 342.1459 ($[M+H]^+$)

8.1.4 – Synthesis of 2-(2-methoxyethoxy)ethan-1-amine



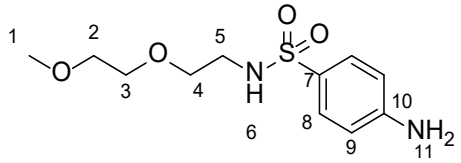
Phthalimide (3.0 g, 20 mmol) was added to a solution of 2-(2-methoxyethoxy)ethan-1-ol (2.1 ml, 16 mmol) and triphenylphosphine (5.4 g, 20 mmol) in THF (70 ml). The mixture was stirred at 0°C for 15 minutes, followed by addition of DIAD (4.2 ml, 20 mmol) dropwise and allowed to warm to room temperature and stirred overnight. The reaction mixture was reduced under vacuum and portioned between diethylether (50 ml) and saturated sodium bicarbonate (50 ml). The organics were separated, washing the aqueous with ether (3 x 30 ml). The organics were combined, dried over sodium sulphate, and reduced by 2/3. The organic layer was cooled to 0°C, filtered, concentrated, and dissolved in ethanol (50 ml). Hydrazine hydrate (1.7 ml, 34 mmol) was added and the reaction heated at reflux overnight. The reaction mixture was then reduced under vacuum and ethanol (20 ml) was added. The suspension was filtered and the filter cake washed with ethanol. The reaction mixture was then diluted with saturated sodium bicarbonate (20 ml) and extracted with 1:4 IPA:DCM (3x 20 ml). The aqueous was then reduced under vacuum, and the solid product loaded onto a silica plug and washed with 20% ammonia in methanol solution. The fractions were combined and concentrated giving a colourless oil. (348 mg, 17%) 1H NMR (500 MHz, Methanol-d4) δ 3.61 (m, 2H, H-4), 3.57 – 3.54 (m, 2H, H-3), 3.51 (t, J = 5.3 Hz, 2H, H-2), 3.37 (s, 3H, H-1), 2.79 (t, J = 5.3 Hz, 2H, H-5); 13C NMR (126 MHz, Chloroform-d) δ 70.02 (PEG), 69.85 (PEG), 69.12 (PEG) 43.52 (C-1), 40.15 (C-5); LCMS m/z 120.2 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_5H_{13}NO_2$: 120.1019 ($[M+H]^+$); found: 120.1009 ($[M+H]^+$)

8.1.5 – Synthesis of N-(2-(2-methoxyethoxy)ethyl)-4-nitrobenzenesulfonamide



4-Nitrobenzenesulfonyl chloride (300 mg, 1.3 mmol) was added to a solution of 2-(2-methoxyethoxy)ethan-1-amine (160 mg, 1.3 mmol) and triethylamine (150 µl, 2.0 mmol) in DCM (5 ml). Th reactions was stirred at room temperature for 2 hours and then quenched with saturated sodium bicarbonate (10 ml). The aqueous was extracted with DCM (3x10 ml), the organics combined, dried over sodium sulphate and concentrated. Further purification was achieved by column chromatography (0-100% ethyl acetate in petroleum ether) to give a colourless oil (232 mg, 57%). R_f = 0.48 (DCM/MeOH 19:1); R: νmax/cm-1 1547 (N-O stretch), 1353 (S=O stretch); 1H NMR (500 MHz, Methanol-d4) δ 8.45 – 8.41 (m, 2H, H-9), 8.14 – 8.10 (m, 2H, H-8), 3.52 – 3.46 (m, 6H, H-2, H-3 and H-4), 3.36 (s, 3H, H-1), 3.14 (t, J = 5.3 Hz, 2H, H-5); 13C NMR (126 MHz, Chloroform-d) δ 149.30 (C-10), 137.56 (C-7), 132.87 (C-8), 126.34 (C-9), 71.56 (PEG), 70.41 (PEG), 69.32 (PEG), 59.02 (C-5), 43.81 (C-1); LCMS *m/z* 305.1 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{11}H_{16}N_2O_6S$: 305.0802 ($[M+H]^+$); found: 305.0814 ($[M+H]^+$)

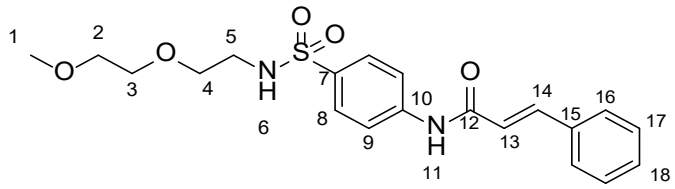
8.1.6 – Synthesis of 4-amino-N-(2-(2-methoxyethoxy)ethyl)benzenesulfonamide



N-(2-(2-Methoxyethoxy)ethyl)-4-nitrobenzenesulfonamide (230 mg, 0.8 mmol) was dissolved in ethanol (20 ml) and subjected to hydrogenation on the H-Cube using a 20% Pd/C cartridge at 40°C. After 2 hours the reaction mixture was concentrated to dryness to give a colourless oil (208 mg, 99%); R_f = 0.41 (DCM/MeOH 19:1); IR: νmax/cm-1 3410 (N-H stretch), 1364 (S=O stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.65 (m, 2H, H-9), 7.43 (m, 2H, H-8), 5.25 (t, J = 6.0 Hz, 1H, H-6), 3.54 – 3.47 (m, 6H, H-2, H-3 and H-4), 3.37 (s, 3H, H-1), 3.15 (q, J = 5.3 Hz, 2H, H-5); 13C NMR (126 MHz, Chloroform-d) δ 144.62 (C-10), 135.02 (C-7), 129.31 (C-8), 120.04 (C-9), 71.24 (PEG), 70.28 (PEG), 69.59 (PEG), 58.86 (C-5), 43.46 (C-1); LCMS m/z 275.1

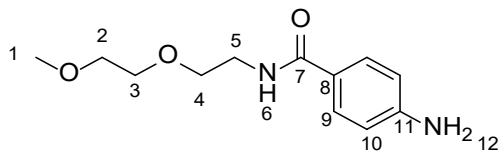
([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{11}H_{18}N_2O_4S$: 275.1060 ([M+H]$^+$); found: 275.1088 ([M+H]$^+$)

8.1.7 – Synthesis of N-(4-(N-(2-(2-methoxyethoxy)ethyl)sulfamoyl)phenyl)cinnamamide



EDC (150 µl, 0.9 mmol) was added to a solution of cinnamic acid (140 mg, 0.9 mmol), triethylamine (85 µl, 1.1 mmol) and HOAT (120 mg, 0.9 mmol) in DCM (5 ml). After 10 minutes 4-amino-N-(2-(2-methoxyethoxy)ethyl)benzenesulfonamide (210 mg, 0.8 mmol) was added. The reaction was stirred overnight at room temperature followed by 40°C for 6 hours. The reaction was then diluted with saturated sodium bicarbonate (10 ml) and extracted with DCM (3x10 ml). The organics were combined, washed with brine, dried over sodium sulphate and concentrated. The crude product was further purified by column chromatography (0-100% ethyl acetate in petroleum ether) to yield a pale yellow solid (150 mg, 49%). R_f = 0.32 (DCM/MeOH 19:1); IR: νmax/cm-1 3181 (br, N-H stretch), 1672 (C=O stretch), 1625 (C=C stretch), 1358 (S=O stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.86 – 7.76 (m, 5H, H-14, H-9 and H-16), 7.62 (s, 1H, H-11), 7.56 (dd, J = 6.7, 2.9 Hz, 2H, H-8), 7.41 (q, J = 2.9 Hz, 3H, H-17 and H-18), 6.57 (d, J = 15.5 Hz, 1H, H-13), 5.25 (t, J = 6.0 Hz, 1H, H-6), 3.57 – 3.48 (m, 6H, H-2, H-3 and H-4), 3.38 (s, 3H, H-1), 3.14 (q, J = 5.4 Hz, 2H, H-5); 13C NMR (126 MHz, Chloroform-d) δ 164.41 (C-12), 143.50 (C-14), 142.24 (C-10), 134.54 (C-7), 134.35 (C-15), 130.33 (C-18), 128.96 (C-8), 128.32 (C-16), 128.11 (C-17), 120.23 (C-13), 119.63 (C-9), 71.77 (PEG), 70.33 (PEG), 69.35 (PEG), 58.99 (C-5), 43.01 (C-1); LCMS m/z 405.3 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{20}H_{24}N_2O_5S$: 405.1479 ([M+H]$^+$); found: 405.1396 ([M+H]$^+$)

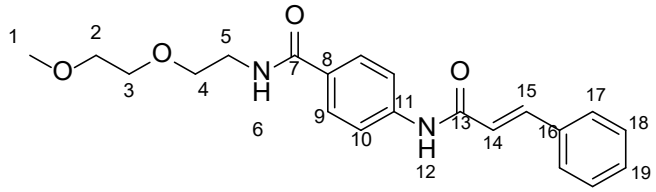
8.1.8 – Synthesis of 4-amino-N-(2-(2-methoxyethoxy)ethyl)benzamide



EDC (219 µl, 1.61 mmol) was added to a solution of 4-aminobenzoic acid (300 µl, 1.6 mmol), 2-(2-methoxyethoxy)ethan-1-amine (160 mg, 1.3 mmol) and triethylamine (150 µl, 2.0 mmol)

in DCM (5 ml) and stirred at room temperature overnight. The reaction was diluted with saturated sodium bicarbonate (10 ml) and extracted with DCM (3x10 ml), the organics were combined, dried over sodium sulphate and reduced. The reaction mixture was further purified by column chromatography (0-100% ethyl acetate in petroleum ether) to yield a colourless oil (142 mg, 54%). R_f = 0.43 (DCM/MeOH 19:1); 1H NMR (500 MHz, Methanol-d4) δ 7.64 – 7.59 (m, 2H, H-9), 6.70 – 6.66 (m, 2H, H-10), 3.66 – 3.62 (m, 4H, H-2 and H-3), 3.59 – 3.53 (m, 4H, H-4 and H-5), 3.39 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 166.88 (C-7), 145.25 (C-11), 123.36 (C-8), 118.89 (C-10), 117.52 (C-9), 71.25 (PEG), 69.95 (PEG), 69.02 (PEG), 59.43 (C-1), 39.54 (C-5); LCMS m/z 239.2 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{12}H_{18}N_2O_3$: 239.1390 ($[M+H]^+$); found: 239.1402 ($[M+H]^+$)

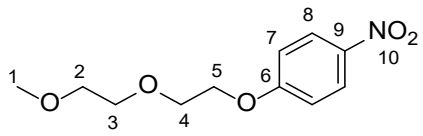
8.1.9 – Synthesis of 4-cinnamamido-N-(2-(2-methoxyethoxy)ethyl)benzamide



4-Amino-N-(2-(2-methoxyethoxy)ethyl)benzamide (140 mg, 0.6 mmol) and DIPEA (160 µl, 1.2 mmol) were dissolved in THF (4 ml). Cinnamoyl chloride (170 mg, 1.0 mmol) in THF (1 ml) was added dropwise and the solution stirred at room temperature overnight. The reaction mixture was diluted with water (10 ml) and extracted with ethyl acetate (3x10 ml), the combined organics were dried over sodium sulphate and concentrated. The crude product was further purified by column chromatography (0-100% ethyl acetate in petroleum ether, followed by a 0-25% methanol flush) to yield a pale yellow solid (71.6 mg, 33%). R_f = 0.25 (DCM/MeOH 19:1); IR: νmax/cm-1 3241 (br, N-H stretch), 2869 (br, N-H stretch), 1620 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.83 – 7.76 (m, 3H, H-9 and H-15), 7.70 (d, J = 8.3 Hz, 2H, H-10), 7.55 (m, 2H, H-17), 7.48 (s, 1H, H-6), 7.40 (m, 3H, A-H-18 and H-19), 6.76 (s, 1H, H-12), 6.55 (d, J = 15.5 Hz, 1H, H-14), 3.70 – 3.65 (m, 6H, H-2, H-3 and H-4), 3.59 – 3.55 (m, 2H, H-5), 3.40 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 166.88 (C-7), 164.13 (C-13), 143.02 (C-15), 141.06 (C-11), 134.48 (C-16), 130.18 (C-8), 128.93 (C-18), 128.15 (C-17), 128.02 (C-19), 120.49 (C-10), 119.31 (C-9), 71.88 (PEG), 70.19 (PEG), 69.84 (PEG), 59.05 (C-1),

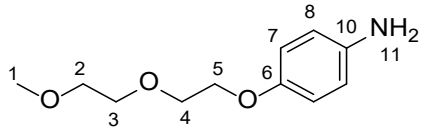
39.72 (C-5); LCMS m/z 369.3 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{21}H_{24}N_2O_4$: 369.1809 ([M+H]$^+$); found: 369.1722 ([M+H]$^+$)

8.1.10 – Synthesis of 1-(2-(2-methoxyethoxy)ethoxy)-4-nitrobenzene



4-nitrophenol (2.2 g, 16 mmol) and potassium carbonate (8.6 g, 63 mmol) were suspended in acetonitrile (70 ml). 2-(2-methoxyethoxy)ethyl 4-methylbenzenesulfonate (4.3 g, 16 mmol) was added and the solution heated at 80°C for overnight then room temperature for 2 days. 4-nitrophenol (2.2 g, 16 mmol) was added and heated at 80°C for 3 hours. The reaction was then concentrated, DMF (70 ml) was added and heated to 100°C overnight. Reaction was diluted with water (150 ml) and extracted with ethyl acetate (3x100 ml). The organics were combined, washed with brine, dried over sodium sulphate and concentrated. Purified by column chromatography (0-100% ethyl acetate in petroleum ether) to yield a pale brown oil (3.23 g, 86%). R_f = 0.2 (DCM); 1H NMR (500 MHz, Chloroform-d) δ 8.22 – 8.17 (m, 2H, H-8), 7.00 – 6.95 (m, 2H, H-7), 4.25 – 4.21 (m, 2H, H-5), 3.92 – 3.87 (m, 2H, H-4), 3.75 – 3.70 (m, 2H, H-3), 3.60 – 3.56 (m, 2H, H-2), 3.39 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 161.21 (C-6), 140.03 (C-9), 123.67 (C-8), 117.38 (C-7), 72.45 (PEG), 70.91 (PEG), 70.26 (PEG), 69.11 (PEG), 56.26 (C-1); LCMS m/z 242.1 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{11}H_{15}NO_5$: 242.1023 ([M+H]$^+$); found: 242.1089 ([M+H]$^+$)

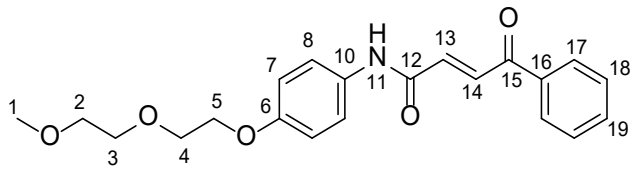
8.1.11 – Synthesis of 4-(2-(2-methoxyethoxy)ethoxy)aniline



1-(2-(2-Methoxyethoxy)ethoxy)-4-nitrobenzene (3.2 g, 13 mmol) was dissolved in MeOH (80 mL), and DCM (20 mL) and reacted using the H-cube using a Pd/C (10% wt) cartridge. The solution was allowed to circulate through the system for 21 h at 40 °C. The reaction mixture was concentrated in vacuo to give the desired compound (2.82 g, 99%) as a brown oil. The compound was used without any further purification. R_f = 0.1 (10% MeOH in DCM); λmax = 300.6 nm; IR: νmax/cm-1 3432 (br NH), 2900 (m, C-H aromatic), 2875 (C-H alkyl), 1509 (C=C aromatic); 1H NMR (500 MHz, Chloroform-d): δ 6.78 (d, J = 6.60 Hz, 2H, H-8), 6.66 (d, J = 6.6

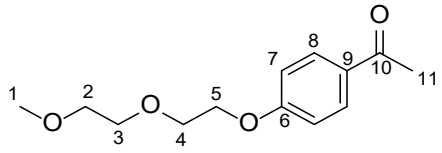
Hz, 2H, H-7), 4.09 (t, J = 4.8 Hz, 2H, H-5), 3.84 (t, J = 5.2 Hz, 2H, H-4), 3.73 (t, J = 4.5 Hz, 2H, H-3), 3.60 (t, J = 3.1 Hz, 2H, H-2), 3.41 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 152.0 (C-6), 139.9 (C-9), 116.5 (C-7), 115.9 (C-8), 72.0 (PEG), 70.7 (PEG), 70.0 (PEG), 68.1 (PEG), 59.1 (C-1); LRMS (ES+) m/z 211.0 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{11}H_{17}NO_3$: 212.1281 ([M+H]$^+$); found: 212.1263 ([M+H]$^+$)

8.1.12 – Synthesis of (E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide



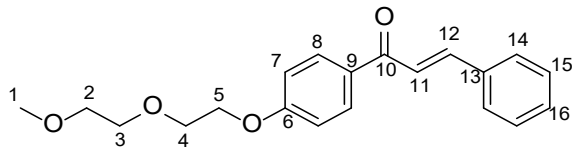
(E)-4-Oxo-4-phenylbut-2-enoic acid (420 mg, 2.4 mmol) and HATU (990 mg, 2.6 mmol) were dissolved in DMF (15 ml) and DIPEA (830 µl, 4.7 mmol). 4-(2-(2-methoxyethoxy)ethoxy)aniline (500 mg, 2.4 mmol) was added after 10 minutes and the reaction stirred at room temperature for 5 hours. The reaction mixture was quenched with water (40 ml) and extracted with ethyl acetate (3x 50 ml), organics combined, dried over sodium sulphate and concentrated. The crude was purified by column chromatography (0-100% ethyl acetate in petroleum ether) to yield a yellow solid (713 mg, 81%). R_f = 0.61 (DCM/MeOH 9:1); IR: νmax/cm-1 2981 (br, N-H stretch), 1673 (C=O stretch), 1622 (C=O stretch), 1599 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 8.11 – 8.04 (m, 3H, H-17 and H-13), 7.66 – 7.58 (m, 2H, H-8), 7.58 – 7.50 (m, 4H, H-18, H-19 and H-11), 7.12 (d, J = 14.8 Hz, 1H, H-14), 6.94 – 6.89 (m, 2H, H-7), 4.17 – 4.13 (m, 2H, H-5), 3.88 – 3.85 (m, 2H, PEG-H), 3.75 – 3.72 (m, 2H, PEG-H), 3.61 – 3.58 (m, 2H, PEG-H), 3.40 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 189.67 (C-16), 161.72 (C-10), 156.13 (C-6), 136.87 (C-12), 135.83 (C-17), 133.89 (C-14), 133.73 (C-13), 130.83 (C-9), 128.91-128.90 (C-18 AND C-19), 121.67 (C-8), 115.11 (C-7), 71.97 (PEG), 70.78 (PEG), 69.77 (PEG), 67.71 (PEG), 59.09 (C-1); LCMS m/z 370.2 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{21}H_{23}NO_5$: 370.1649 ([M+H]$^+$); found: 370.1595 ([M+H]$^+$).

8.1.13 - Synthesis of 1-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)ethan-1-one



2-(2-Methoxyethoxy)ethyl 4-methylbenzenesulfonate (410 mg, 1.5 mmol) in acetonitrile (2 ml) was added dropwise to a solution of 4-hydroxyacetophenone (230 mg, 1.7 mmol) and potassium carbonate (830 mg, 6.0 mmol) in acetonitrile (8 ml). The reaction was stirred at 80°C overnight. 2-(2-methoxyethoxy)ethyl 4-methylbenzenesulfonate (180 mg, 0.6 mmol) was added and stirred at 80°C overnight. The reaction was filtered, the filter cake washed with acetonitrile, and concentrated. The crude was purified by column chromatography (0-80% ethyl acetate in petroleum ether). The product eluted with a small impurity in the last 2 fractions, which were combined separately to yield the product with 2 different purities (220 mg, 100% purity; 150 mg, 75% purity). R_f = 0.46 (DCM/MeOH 19:1); IR: νmax/cm-1 2879 (C-H stretch), 1671 (C=O stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.94 – 7.90 (m, 2H, H-8), 6.97 – 6.92 (m, 2H, H-7), 4.23 – 4.19 (m, 2H, H-5), 3.91 – 3.87 (m, 2H, H-4), 3.74 – 3.70 (m, 2H, H-3), 3.61 – 3.56 (m, 2H, H-2), 3.39 (s, 3H, H-1), 2.55 (s, 3H, H-11); 13C NMR (126 MHz, Chloroform-d) δ 196.12 (C-10), 163.01 (C-6), 129.89 (C-8), 129.26 (C-7), 125.47 (C-9), 72.19 (PEG), 71.42 (PEG), 69.87 (PEG), 68.91 (PEG), 58.52 (C-1), 28.65 (C-11); LCMS *m/z* 239.2 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{13}H_{18}O_4$: 239.1278 ($[M+H]^+$); found: 239.1364 ($[M+H]^+$)

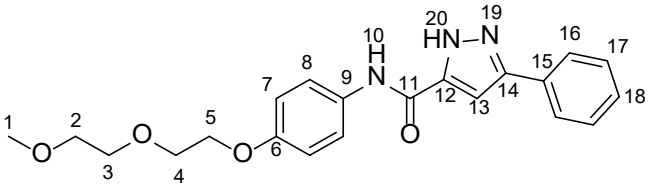
8.1.14 - Synthesis of (E)-1-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-3-phenylprop-2-en-1-one



1-(4-(2-(2-Methoxyethoxy)ethoxy)phenyl)ethan-1-one (220 mg, 0.9 mmol) and benzaldehyde (94 µl, 0.9 mmol) were dissolved in ethanol (2 ml) and 2M sodium hydroxide (1.4 ml). The reaction was stirred overnight at room temperature. The reaction mixture was concentrated to 1 ml, neutralised with 1M HCl and extracted with DCM (3x 10 ml). The organics were combined, dried over sodium sulphate and concentrated to yield a white solid (279 mg, 92%). R_f = 0.43 (DCM/MeOH 19:1); IR: νmax/cm-1 2875 (C-H stretch), 1676 (C=O stretch), 1595 (C=C

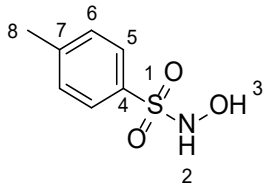
stretch); 1H NMR (500 MHz, Chloroform-d) δ 8.05 – 8.01 (m, 2H, H-8), 7.80 (d, J = 15.6 Hz, 1H, H-12), 7.67 – 7.63 (m, 2H, H-14), 7.55 (d, J = 15.7 Hz, 1H, H-11), 7.44 – 7.39 (m, 3H, H-15 and H-16), 7.02 – 6.98 (m, 2H, H-7), 4.25 – 4.22 (m, 2H, H-5), 3.92 – 3.88 (m, 2H, PEG-H), 3.76 – 3.72 (m, 2H, PEG-H), 3.61 – 3.57 (m, 2H, PEG-H), 3.40 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 188.74 (C-10), 162.66 (C-6), 143.99 (C-12), 135.10 (C-13), 131.22 (C-9), 130.79 (C-8), 130.33 (C-16), 128.92 (C-15), 128.37 (C-14), 121.90 (C-11), 114.45 (C-7), 71.96 (PEG), 70.85 (PEG), 69.58 (PEG), 67.63 (PEG), 59.11 (C-1); LCMS *m/z* 327.2 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{20}H_{22}O_4$: 327.1591 ($[M+H]^+$); found: 327.1474 ($[M+H]^+$)

8.1.15 - Synthesis of N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-3-phenyl-1H-pyrazole-5-carboxamide



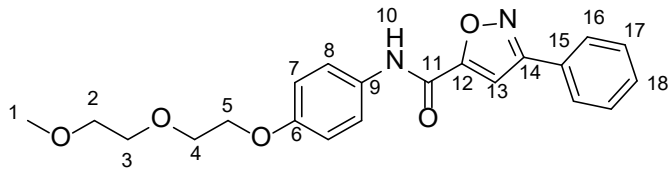
(E)-N-(4-(2-(2-Methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (20 mg, 0.06 mmol), p-Toluenesulfonyl hydrazide (13 mg, 0.07 mmol)and iodine (0.4 mg, 0.006 mmol), were dissolved in ethanol (0.3 ml) and water (0.1 ml). The reaction was stirred at 65°C for 20 minutes, then potassium carbonate (12 mg) was added, and the reaction stirred for 4 hours at 65°C. The reaction mixture was concentrated and purified by column chromatography (0-100% ethyl acetate in petroleum ether, with 2% triethylamine) to yield a white solid (5 mg, 22%). R_f = 0.41 (Heptanes/Ethyl Acetate, 1:1); IR: νmax/cm-1 3482 (N-H stretch), 1661 (C=O stretch), 1582 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.66 – 7.59 (m, 4H, H-8 and H-16), 7.52 – 7.47 (m, 2H, H-17), 7.46 – 7.41 (m, 1H, H-18), 7.17 (s, 1H, H-13), 6.97 – 6.93 (m, 2H, H-7), 4.19 – 4.15 (m, 2H, H-5), 3.91 – 3.87 (m, 2H, PEG-H), 3.77 – 3.73 (m, 2H, PEG-H), 3.63 – 3.59 (m, 2H, PEG-H), 3.42 (s, 3H, H-1). 13C NMR (126 MHz, Chloroform-d) δ 129.3 (C-17), 129.2 (C-18), 125.6 (C-16), 121.4 (C-8), 115.1 (C-7), 72.0 (PEG), 70.8 (PEG), 69.9 (PEG), 67.8 (PEG), 59.1 (C-1); HRMS (ESI): m/z calcd for $C_{21}H_{23}N_3O_4$: 382.1761 ($[M+H]^+$); found: 382.1785 ($[M+H]^+$)

8.1.16 - Synthesis of N-hydroxy-4-methylbenzenesulfonamide



Hydroxylamine hydrochloride (73 mg, 1.05 mmol) was dissolved in methanol (0.35 ml) and water (0.25 ml). Magnesium oxide (63 mg, 1.57 mmol) was then added to the mixture and stirred for 10 minutes. Toluenesulphonyl chloride (100 mg, 0.52 mmol) dissolved in THF (3.5 ml) was then added dropwise. The reaction was stirred at room temperature for 3 hours. The reaction was filtered through celite and the filter cake washed with THF (5 ml). The mother liquor was then concentrated and purified by column chromatography (0-100% ethyl acetate in petroleum ether) to give a white solid (58 mg, 62%). R_f = 0.32 (petroleum ether/Ethyl Acetate, 1:1); IR: νmax/cm-1 3371 (N-H stretch), 3250 (O-H stretch), 1344 (S=O stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.84 (d, J = 8.3 Hz, 2H, H-5), 7.37 (d, J = 7.8 Hz, 2H, H-6), 6.65 (d, J = 3.6 Hz, 1H, H-2), 6.03 (d, J = 4.3 Hz, 1H, H-3), 2.46 (s, 3H, H-8). 13C NMR (126 MHz, Chloroform-d) δ 129.81 (C-6), 128.75 (C-5), 53.41 (C-8). LCMS *m/z* 188.0 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_7H_9NO_3S$: 188.0376 ($[M+H]^+$); found: 188.0337 ($[M+H]^+$)

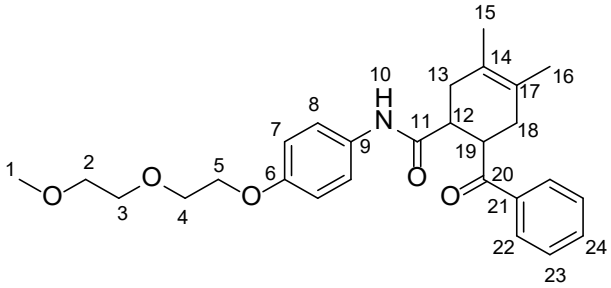
8.1.17 - Synthesis of N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-3-phenylisoxazole-5-carboxamide



(E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (20 mg, 0.06 mmol), N-hydroxy-4-methylbenzenesulfonamide (36 mg, 0.19 mmol) and potassium carbonate (41 mg, 0.30 mmol) were dissolved in DMF (0.3 ml) and water (0.1 ml) and stirred at 65°C overnight. The reaction mixture was concentrated under vacuum and purified by column chromatography (0-100% ethyl acetate in petroleum ether) to yield a white solid (14 mg, 61%). R_f = 0.28 (petroleum ether/Ethyl Acetate, 1:1); IR: νmax/cm-1 3487 (N-H stretch), 1669 (C=O stretch), 1576 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 8.48 (s, 1H, H-10), 7.83 (dd, J = 7.7, 2.0 Hz, 2H, H-16), 7.60 – 7.56 (m, 2H, H-17), 7.50 (dd, J = 6.4, 1.4 Hz, 3H, H-8 and H-18), 7.04 (s, 1H, H-13), 6.94 (d, J = 9.0 Hz, 2H, H-7), 4.17 – 4.14 (m, 2H, H-5), 3.87

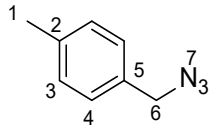
(dd, J = 5.7, 4.1 Hz, 2H, PEG-H), 3.74 – 3.71 (m, 2H, PEG-H), 3.61 – 3.57 (m, 2H, PEG-H), 3.40 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 160.1 (C-6), 129.2 (C-9), 128.5 (C-17), 127.9 (C-18), 126.1 (C-16), 122.6 (C-8), 116.4 (C-7), 106.2 (C-13) 71.8 (PEG-C), 70.2 (PEG-C), 68.1(PEG-C), 67.2 (PEG-C), 59.4 (C-1); LCMS m/z 383.5 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{21}H_{22}N_2O_5$: 383.1601 ($[M+H]^+$); found: 383.1597 ($[M+H]^+$).

8.1.18 - Synthesis of 6-benzoyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-3,4-dimethylcyclohex-3-ene-1-carboxamide



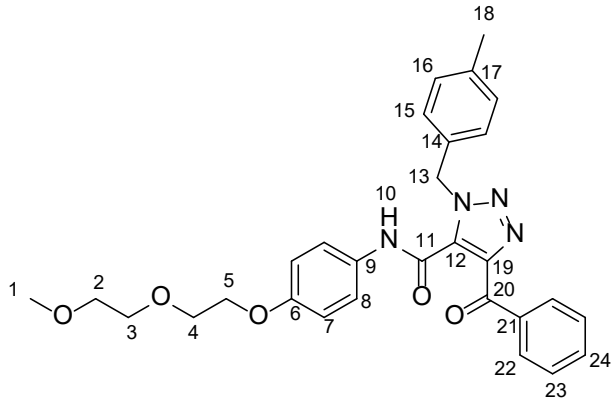
(E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (20 mg, 0.06 mmol) and dimethyl butadiene (7 mg, 0.09 mmol) were dissolved in DMF (0.4 ml) and water (0.1 ml) and stirred at 50°C overnight. The reaction mixture was evaporated to dryness and purified by column chromatography (0-100 % ethyl acetate in heptanes) to yield a white solid (12 mg, 45%). R_f = 0.43 (Heptanes/Ethyl Acetate, 1:1); IR: νmax/cm-1 3313 (N-H stretch), 2920 (C-H stretch), 1675 (C=O stretch), 1518 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.99 (d, J = 7.1 Hz, 2H, H-22), 7.57 – 7.52 (m, 1H, H-24), 7.46 – 7.41 (m, 2H, H-23), 7.36 (s, 1H, H-10), 7.33 – 7.29 (m, 2H, H-8), 6.80 (d, J = 9.0 Hz, 2H, H-7), 4.10 – 4.05 (m, 2H, H-5), 3.96 (td, J = 11.5, 5.5 Hz, 1H, H-19), 3.84 – 3.78 (m, 2H, H-4), 3.71 – 3.67 (m, 2H, H-3), 3.58 – 3.54 (m, 2H, H-2), 3.38 (s, 3H, H-1), 2.98 – 2.91 (m, 1H, H-12), 2.60 – 2.49 (m, 1H, H-13), 2.32 – 2.21 (m, 2H, H-13 and H-18), 2.12 – 2.02 (m, 1H, H-18), 1.69 (s, 3H, H15), 1.63 (s, 3H, H-16). 13C NMR (126 MHz, Chloroform-d) δ 203.86 (C-20), 172.99 (C-11), 155.45 (C-6), 136. 28 (C-21), 133.23 (C-9), 131.21(C-24), 128.64 (C-22/23), 128.59 (C-22/23), 124.51 (C-14/17), 123.93 (C-14/17), 121.63 (C-8), 114.85 (C-7), 71.94 (PEG), 70.74 (PEG), 69.76 (PEG), 67.70 (PEG), 59.08 (C-1), 44.67 (C-12), 44.59 (C-19), 35.93 (C-13/18), 35.04 (C-13/18), 18.73 (C-15/16), 18.67 (C-15/16); HRMS (ESI): m/z calcd for $C_{27}H_{33}NO_5$: 452.2431 ($[M+H]^+$); found: 452.2394 ($[M+H]^+$)

8.1.19 - Synthesis of 1-(azidomethyl)-4-methylbenzene



Benzyl bromide (541 mg, 2.92 mmol) was dissolved in water (1 ml) and acetone (4 ml), followed by the addition of sodium azide (285 mg, 4.39 mmol). The reaction was stirred overnight at 40°C. DCM (10 ml) and water (10 ml) were added and the organic layer extracted. The aqueous was extracted 4 times with DCM (10 ml). The organics were combined, dried over magnesium sulphate and concentrated under vacuum to give a pale yellow oil (337 mg, 79%). R_f = 0.65 (petroleum ether/Ethyl Acetate, 1:1); IR: νmax/cm-1 2923, 2090 (N_3 stretch); 1H NMR (500 MHz, Chloroform-d) δ 7.26 – 7.20 (m, 4H, H-2 and H-3), 4.32 (s, 2H, H-5), 2.39 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 138.15 (C-2), 132.30 (C-5), 129.51(C-3/4), 128.27 (C-3/4), 54.64 (C-6), 21.18 (C-1); HRMS (ESI): m/z calcd for $C_8H_9N_3$: 148.0830 ($[M+H]^+$); no mass found.

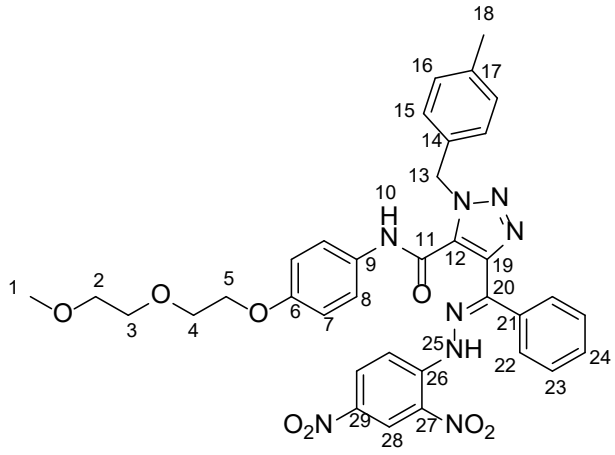
8.1.20 - Synthesis of 4-benzoyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-1-(4-methylbenzyl)-1H-1,2,3-triazole-5-carboxamide



(E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (20 mg, 0.06 mmol), 1-(azidomethyl)-4-methylbenzene (18 mg, 0.12 mmol) and potassium carbonate (17 mg, 0.12 mmol) were dissolved in water (0.34 ml) and dioxane (0.06 ml). The reaction was stirred overnight at 60°C. Water (3 ml) was then added and the aqueous extracted with ethyl acetate (3x5 ml). The organics were combined, dried over magnesium sulphate and reduced under vacuum. The compound was further purified by column chromatography (0-100% ethyl acetate in petroleum ether) to yield a pale yellow tacky solid. R_f = 0.48 (Heptanes/Ethyl

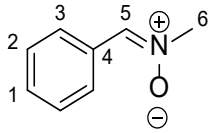
Acetate, 1:1); IR: νmax/cm-1 3496 (N-H stretch), 1673 (C=O stretch), 1626 (C=O), 1571 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 11.86 (s, 1H, H-10), 8.14 – 8.10 (m, 2H, H-22), 7.64 – 7.57 (m, 3H, H-24 and H-8), 7.49 – 7.43 (m, 2H, H-23), 7.33 – 7.30 (m, 2H, H-15), 7.07 (d, J = 7.8 Hz, 2H, H-16), 6.89 – 6.84 (m, 2H, H-7), 6.09 (s, 2H, H-13), 4.11 – 4.06 (m, 2H, H-5), 3.80 (dd, J = 5.6, 4.1 Hz, 2H, H-4), 3.68 – 3.64 (m, 2H, H-3), 3.55 – 3.50 (m, 2H, H-2), 3.33 (s, 3H, H-1), 2.24 (s, 3H, H-18); 13C NMR (126 MHz, Chloroform-d) δ 190.48 (C-20), 159.17 (C-11), 156.17 (C-6), 154.31 (C-12), 142.73 (C-19), 138.40 (C-21), 134.09 (C-17), 132.09 (C-24/14), 131.47 (C-22), 130.77 (C-9), 129.46 (C-16), 128.68 (C-23/15), 128.36 (C-23/15), 122.08 (C-8), 115.07 (C-7), 71.97 (PEG), 70.79 (PEG), 69.77 (PEG), 67.72 (PEG), 59.10 (C-1), 54.04 (C-13), 21.17 (C-18); LCMS *m/z* 515.6 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{29}H_{30}N_4O_5$: 515.2289 ([M+H]$^+$); found: 515.2272 ([M+H]$^+$)

8.1.21 - Synthesis of (E)-4-((2-(2,4-dinitrophenyl)hydrazono)(phenyl)methyl)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-1-(4-methylbenzyl)-1H-1,2,3-triazole-5-carboxamide



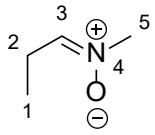
4-benzoyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-1-(4-methylbenzyl)-1H-1,2,3-triazole-5-carboxamide (4 mg, 0.008 mmol) was dissolved in THF (0.3 ml). 2,4-Dinitrophenylhydrazine (1.7 mg, 0.008 mmol) was added and stirred overnight at room temperature. The reaction mixture was filtered and the solid was recrystallized from petroleum ether and THF. Crystal structure seen in [Figure 41].

8.1.22 - Synthesis of (Z)-N-methyl-1-phenylmethanimine oxide



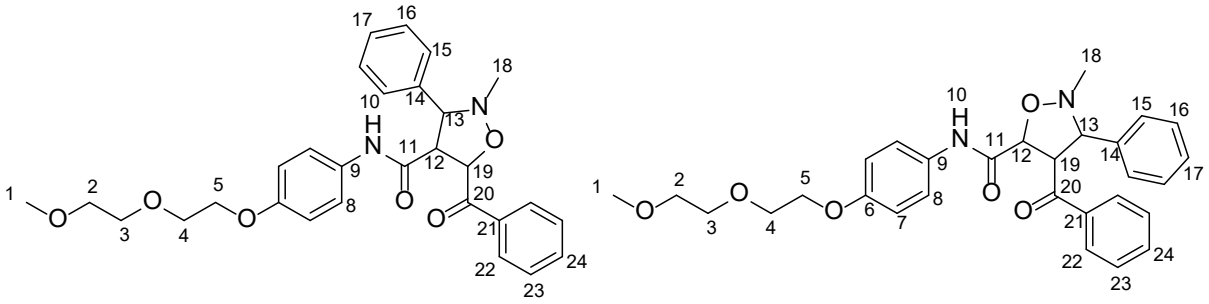
N-Methylhydroxylamine hydrochloride (102 mg, 1.23 mmol), benzaldehyde (96 µl, 0.94 mmol) and sodium hydrogencarbonate (197 mg, 2.35 mmol) were dissolved in DCM (1 ml). The reaction mixture was stirred at 50°C for 3 hours. The mixture was then cooled, filtered through celite, washing with DCM. The organics were concentrated under vacuum to give a colourless oil (110 mg, 94%). R_f = 0.84 (petroleum ether/Ethyl Acetate, 1:1); IR: υmax/cm-1 3182 (O-H stretch), 3056 (C-H stretch), 2872 (C-H), 1659 (N-O stretch); 1H NMR (500 MHz, Chloroform-d) δ 8.23 – 8.19 (m, 2H, H-2), 7.44 – 7.40 (m, 3H, H-1 and H-3), 7.37 (s, 1H, H-5), 3.89 (s, 3H, H-6); 13C NMR (126 MHz, Chloroform-d) δ 135.17 (C-5), 130.44 (C-4), 128.50 (C-2/3), 128.41 (C-2/3), 54.43 (C-6); LCMS *m/z* 136.2 ($[M+H]^+$); HRMS (ESI): m/z calcd for C_8H_9NO: 136.0684 ($[M+H]^+$); no mass found.

8.1.23 - Synthesis of (Z)-N-methylpropan-1-imine oxide



N-Methylhydroxylamine hydrochloride (144 mg, 1.72 mmol), propanal (100 mg, 1.72 mmol) and sodium methoxide (93 mg, 1.72 mmol) were dissolved in ethanol (2 ml). The reaction mixture was stirred at RT for 2 hours. The mixture was then concentrated, suspended in DCM and filtered through celite, washing with DCM. The organics were concentrated under vacuum to give a colourless oil (121 mg, 81%). R_f = 0.46 (petroleum ether/Ethyl Acetate, 3:1, with potassium permanganate stain); 1H NMR (500 MHz, Chloroform-d) δ 6.64 (t, J = 5.7 Hz, 1H, H-3), 3.63 (s, 3H, H-5), 2.50 – 2.41 (m, 2H, H-2), 1.05 (t, J = 7.7 Hz, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 52.20 (C-3), 20.37 (C-5), 10.66 (C-2), 9.80 (C-1); HRMS (ESI): m/z calcd for C_4H_9NO: 88.0718 ($[M+H]^+$); no mass found.

8.1.24 - Synthesis of 5-benzoyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-2-methyl-3-phenylisoxazolidine-4-carboxamide and 4-benzoyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-2-methyl-3-phenylisoxazolidine-5-carboxamide

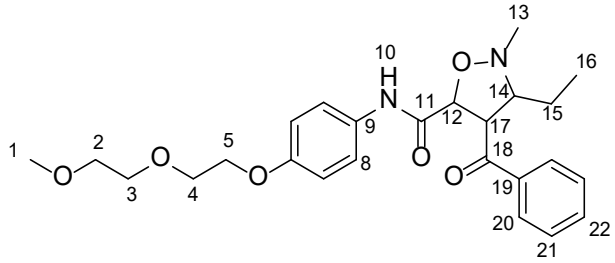


(E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (20 mg, 0.06 mmol) and (Z)-N-methyl-1-phenylmethanimine oxide (11 mg, 0.08 mmol) a were dissolved in water (0.2 ml) and acetonitrile (0.2 ml) and was stirred overnight at 60°C. The reaction mixture was cooled and concentrated. The crude was purified by column chromatography (0-100% DCM in petroleum ether). This yielded the 2 isomers, and to separate these further purifications by reverse phase column chromatography (5-95% acetonitrile [0.1% formic acid] in water [0.1% formic acid]) were carried out to yield the 2 isomers as pale yellow tacky solids.

5-benzoyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-2-methyl-3-phenylisoxazolidine-4-carboxamide: R_f = 0.43 (DCM/petroleum ether, 1:1); IR: υmax/cm-1 3272 (N-H stretch), 2891 (C-H), 1656 (C=O stretch), 1514 (C=O stretch); 1H NMR (500 MHz, Chloroform-d) δ 8.19 (dt, J = 8.5, 1.8 Hz, 2H, H-22), 7.66 – 7.62 (m, 1H, H-24), 7.53 (d, J = 7.6 Hz, 2H, H-23), 7.47 – 7.44 (m, 2H, H-8), 7.42 – 7.32 (m, 5H, H-15, H-16 and H-17), 6.86 (d, J = 9.0 Hz, 2H, H-7), 5.53 (d, J = 5.3 Hz, 1H, H-19), 4.11 (q, J = 4.7 Hz, 3H, H-5 and H-13), 4.04 (s, 1H, H-12), 3.85 (dd, J = 5.7, 4.1 Hz, 2H, H-4), 3.77 – 3.69 (m, 2H, H-3), 3.61 – 3.57 (m, 2H, H-2), 3.40 (s, 3H,H-1), 2.63 (s, 3H, H-18); 13C NMR (126 MHz, Chloroform-d) δ 199.01 (C-20), 167.70 (C-11), 155.75 (C-6), 134.84 (C-14), 133.74 (C-21), 130.77 (C-24), 129.87 (C-22), 129.05 (C-16), 128.70 (C-15), 128.51 (C-23), 128.14 (C-17), 121.49 (C-8), 114.95 (C-7), 71.95 (PEG), 70.76 (PEG), 69.76 (PEG), 67.72 (PEG), 60.49 (C-12), 59.07 (C-1), 42.88 (C-18); LCMS m/z 505.3 ($[M+H]^+$); HRMS (ESI): m/z calcd for C29H32N2O6: 505.2333 ($[M+H]^+$); found: 505.2329 ($[M+H]^+$)

4-benzoyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-2-methyl-3-phenylisoxazolidine-5-carboxamide: R_f = 0.43 (DCM/petroleum ether, 1:1); 1H NMR (500 MHz, Chloroform-d) δ 8.72 (s, 1H, H-10), 7.97 (dq, J = 8.5, 1.6 Hz, 2H, H-22), 7.61 – 7.55 (m, 2H, H-23), 7.52 (tt, J = 7.2, 1.3 Hz, 1H, H-24), 7.40 – 7.31 (m, 3H, H-8 and H-17), 7.22-7.30 (m, 4H, H-15 and H-16), 6.97 (dt, J = 8.8, 2.3 Hz, 2H, H-7), 4.89 (dd, J = 3.6, 1.4 Hz, 1H, H-12), 4.86 (ddd, J = 7.7, 3.5, 1.4 Hz, 1H, H-13), 4.18 (td, J = 4.9, 1.4 Hz, 2H, H-5), 4.08 (dd, J = 7.9, 1.4 Hz, 1H,H-19), 3.89 (ddd, J = 6.1, 4.2, 1.3 Hz, 2H, H-4), 3.75 (td, J = 4.6, 2.3 Hz, 2H, H-3), 3.61 (ddd, J = 6.0, 3.1, 1.4 Hz, 2H, H-2), 3.42 (d, J = 1.4 Hz, 3H, H-1), 2.77 (d, J = 1.4 Hz, 3H, H-18); 13C NMR (126 MHz, Chloroform-d) δ 196.93 (C-20), 169.97 (C-11), 155.85 (C-6), 136.78 (C-14), 135.74 (C-21), 133.62 (C-24), 130.63 (C-9), 129.30 (C-22), 128.94 (C-23), 128.59 (C-15), 128.54 (C-24), 128.10 (C-16), 121.57 (C-8), 115.12 (C-7), 79.39 (C-13), 76.47 (C-12), 71.98 (PEG), 70.79 (PEG), 69.80 (PEG), 67.79 (PEG), 62.83 (C-1), 59.10 (C-19), 43.12 (C-18); LCMS *m/z* 505.3 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{29}H_{32}N_2O_6$: 505.2333 ($[M+H]^+$); found: 505.2351 ($[M+H]^+$)

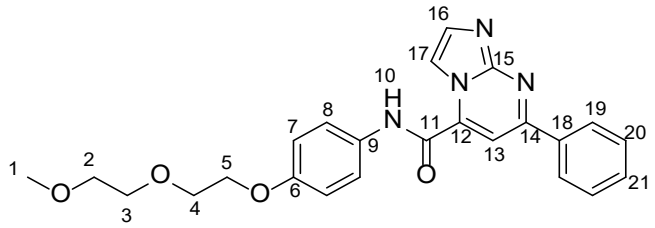
8.1.25 - Synthesis of 4-benzoyl-5-ethyl-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-1-methylpyrrolidine-3-carboxamide



(E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (30 mg, 0.08 mmol) and (Z)-N-methylpropan-1-imine oxide (13 mg, 0.15 mmol) were dissolved in water (0.3 ml) and acetonitrile (0.4 ml). The reaction was stirred overnight at RT. The reaction mixture was concentrated and purified by column chromatography (0-20% MeOH in DCM). Further purification by reverse phase column chromatography (5-95% acetonitrile [0.1% formic acid] in water [0.1% formic acid]) to yielded the product as pale yellow tacky solid (16 mg, 47%). R_f = 0.32 (MeOH/DCM, 1:9); IR: νmax/cm-1 3280 (N-H stretch), 2926 (C-H stretch), 2877 (C-H), 1666 (C=O stretch), 1510 (C=O stretch); 1H NMR (500 MHz, Chloroform-d) δ 8.13 – 8.08 (m, 1H, H-20), 8.05 – 8.01 (m, 1H, H-20), 7.62 – 7.58 (m, 1H, H-22), 7.52 – 7.42 (m, 4H, H-8 and H-21), 6.93 – 6.85 (m, 2H, H-7), 5.56 (d, J = 2.5 Hz, 0.5H, H-12), 5.27 (d, J = 6.5 Hz, 0.5H, H-12), 4.15 – 4.09 (m, 2H, H-5), 3.88 – 3.82 (m, 2H, H-4), 3.76 (dd, J = 5.7, 2.6 Hz, 1H, H-

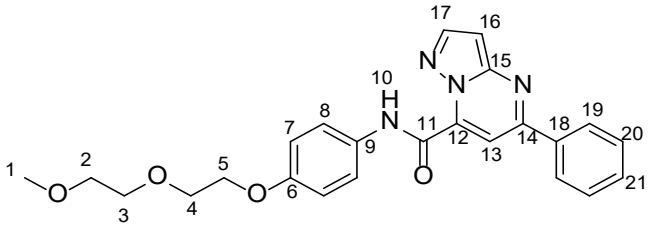
17), 3.74 – 3.69 (m, 2H, H-3), 3.66 (m, 1H, H-14), 3.61 – 3.55 (m, 2H, H-2), 3.40 – 3.38 (m, 3H, Me-1), 2.84 (m, 3H, H-13), 1.73 – 1.58 (m, 2H, H-15), 1.05 – 0.97 (m, 3H, H-16); 13C NMR (126 MHz, Chloroform-d) δ 195.18 (C-18), 168.32 (C-11), 155.71 (C-6), 133.95 (C-19), 129.96 (C-22), 129.14 (C-9), 128.85 (C20/21), 128.50 (C-20/21), 121.39 (C-8), 115.03 (C-7), 71.96 (C-12), 70.89 (C-14/PEG-C), 70.77 (C-14/PEG), 69.78 (PEG), 67.75 (PEG), 59.08 (C-13), 20.98 (C-15), 11.43 (C-16); LCMS *m/z* 457.4 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{25}H_{32}N_2O_6$: 457.2333 ([M+H]$^+$); found: 457.2318 ([M+H]$^+$)

8.1.26 - Synthesis of N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-7-phenylimidazo[1,2-a]pyrimidine-5-carboxamide



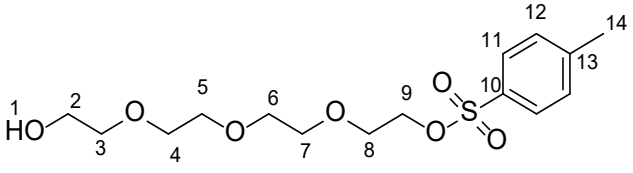
(E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (30 mg, 0.08 mmol), amino imidazole hemisulphate (72 mg, 0.25 mmol) and potassium carbonate (39 mg, 0.25 mmol) were dissolved in DMF (0.6 ml) and water (0.2 ml) and stirred at 60°C overnight. The reaction mixture was concentrated under vacuum and purified by column chromatography (0-30% methanol in DCM) to yield a white solid (18 mg, 52%). R_f = 0.21 (methanol/DCM, 1:9); IR: νmax/cm-1 3486 (N-H stretch), 1668 (C=O stretch), 1576 (C=C stretch); 1H NMR (700 MHz, DMSO-d6) δ 10.95 (s, 1H, H-10), 8.40 (d, J = 1.4 Hz, 1H, H-17), 8.37 – 8.33 (m, 2H, H-19), 8.29 (s, 1H, H-8), 7.91 (s, 1H, H-16), 7.74 – 7.70 (m, 2H, H-8), 7.65 – 7.61 (m, 2H, H-20), 7.61 – 7.57 (m, 1H, H-21), 7.06 – 7.01 (m, 2H, H-7), 4.15 – 4.09 (m, 2H, H-5), 3.78 – 3.73 (m, 2H, H-4), 3.63 – 3.59 (m, 2H, H-3), 3.50 – 3.47 (m, 2H, H-2), 3.27 (s, 3H, H-1); 13C NMR (176 MHz, DMSO-d6) δ 159.75 (C-14), 156.05 (C-11), 155.35 (C-12), 137.97 (C-18), 136.95 (C-16), 136.49 (C-9), 131.37 (C-21), 131.25 (C-17), 129.52 (C-20), 127.67 (C-19), 122.87 (C-8), 115.06 (C-7), 106.17 (C-13), 71.76 (PEG), 70.20 (PEG), 69.41 (PEG), 67.79 (PEG), 58.55 (C-1); LCMS *m/z* 433.4 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{24}H_{24}N_4O_4$: 433.1870 ([M+H]$^+$); found: 433.1812 ([M+H]$^+$)

8.1.27 - Synthesis of N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-5-phenylpyrazolo[1,5-a]pyrimidine-7-carboxamide



(E)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-4-oxo-4-phenylbut-2-enamide (30 mg, 0.08 mmol), aminopyrazole (10 mg, 0.12 mmol) and potassium hydroxide (7 mg, 0.12 mmol) were dissolved in THF (0.6 ml) and water (0.2 ml). stirred at RT overnight. The reaction mixture was concentrated under vacuum and purified by column chromatography (0-100% ethylacetate in petroleum ether) to yield a white solid (18 mg, 52%). R_f = 0.56 (ethylacetate/petroleum ether, 1:1); IR: υmax/cm-1 3479 (N-H stretch), 1674 (C=O stretch), 1581 (C=C stretch), 1562 (C=C stretch); 1H NMR (500 MHz, Chloroform-d) δ 12.69 (s, 1H, H-10), 8.36 (s, 1H, H-13), 8.20 (d, J = 2.5 Hz, 1H, H-17), 8.18 – 8.14 (m, 2H, H-19), 7.83 – 7.78 (m, 2H, H-8), 7.52 – 7.47 (m, 3H, H-20 and H-21), 7.12 (m, 2H, H-7), 6.86 (d, J = 2.5 Hz, 1H, H-16), 4.17 – 4.11 (m, 2H, H-5), 3.81 – 3.79 (m, 2H, H-4), 3.68 – 3.63 (m, 2H, H-3), 3.57 – 3.53 (m, 2H, H-2), 3.32 (s, 3H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 162.66 (C-14), 159.80 (C-11), 156.76 (C-12), 146.59 (C-17), 141.18 (C-9), 136.51 (C-18), 128.98 (C-21), 127.19 (C-20), 124.73 (C-19), 121.92 (C-8), 117.21 (C-7), 112.74 (C-13), 103.65 (C-16), 72.01 (PEG), 70.93 (PEG), 69.32 (PEG), 67.53 (PEG), 58.97 (C-1); LCMS *m/z* 433.4 ([M+H]$^+$); HRMS (ESI): m/z calcd for $C_{24}H_{24}N_4O_4$: 433.1870 ([M+H]$^+$); found: 433.1895 ([M+H]$^+$)

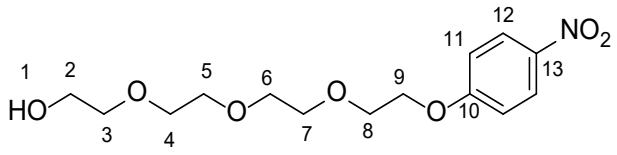
8.1.28 - Synthesis of 2-(2-(2-(2-hydroxyethoxy)ethoxy)ethoxy)ethyl 4-methylbenzenesulfonate



2,2'-((oxybis(ethane-2,1-diyl))bis(oxy))bis(ethan-1-ol) (2.26 ml, 13.1 mmol) and potassium hydroxide (224 mg, 3.93 mmol) were suspended in chloroform (20 ml). Toluene sulphonyl chloride (500 mg, 2.62 mmol) was added at 0°C and the reaction stirred at room temperature for 3 hours. The reaction was diluted with water (20 ml), extracted with DCM (2x30 ml),

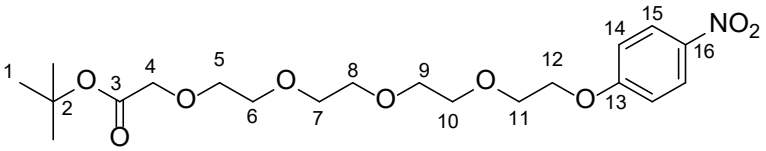
organics combined, washed with 10% citric acid solution (2x50 ml), dried over sodium sulphate and concentrated to yield a colourless oil (764 mg, 84%). R_f = 0.21 (ethyl acetate); 1H NMR (500 MHz, Chloroform-d) δ 7.80 (d, J = 8.3 Hz, 2H, H-11), 7.36 – 7.32 (m, 2H, H-12), 4.19 – 4.14 (m, 2H, H-9), 3.74 – 3.58 (m, 14H, PEG-H), 2.47 (t, J = 6.1 Hz, 1H, H-1), 2.44 (s, 3H, H-14); 13C NMR (125 MHz, Chloroform-d) δ 143.82 (C-10), 132.95 (C-13), 129.75 (C-12), 128.10 (C-11), 72.45 (PEG), 70.83 (PEG), 70.76 (PEG), 70.54 (PEG), 70.43 (PEG), 69.86 (PEG), 68.01 (PEG), 62.12 (PEG), 22.69 (C-14); LRMS *m/z* 350.1 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{15}H_{24}O_7S$: 349.1316 ($[M+H]^+$); found: 349.1351 ($[M+H]^+$).

8.1.29 - Synthesis of 2-(2-(2-(2-(4-nitrophenoxy)ethoxy)ethoxy)ethoxy)ethan-1-ol



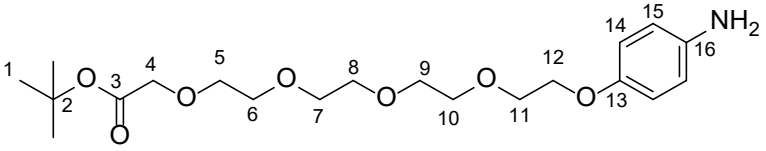
2-(2-(2-(2-hydroxyethoxy)ethoxy)ethoxy)ethyl 4-methylbenzenesulfonate (764 mg, 2.19 mmol) was dissolved in acetonitrile (15 ml). Potassium carbonate (303 mg, 2.19 mmol) and 4-nitrophenol (457 mg, 3.28 mmol) were then added and the reaction stirred at 70°C overnight. The reaction was concentrated, DMF (15 ml) was added and stirred at 80°C for 6 hours. Reaction was quenched with water (30 ml), extracted with ethyl acetate (4x30 ml), dried over magnesium sulphate and concentrated. The crude was then purified by column chromatography (100% ethyl acetate, followed by a 20% methanol flush) to yield a colourless glassy solid (530 mg, 77%). R_f = 0.15 (ethyl acetate); 1H NMR (500 MHz, Chloroform-d) δ 8.22 (d, J = 9.2 Hz, 2H, H-12), 7.01 (d, J = 9.3 Hz, 2H, H-11), 4.28 – 4.22 (m, 2H, H-9), 3.94 – 3.89 (m, 2H, H-2), 3.78 – 3.61 (m, 12H, PEG-H); 13C NMR (125 MHz, Chloroform-d) δ 153.67 (C-10), 142.67 (C-13), 126.13 (C-12), 115.02 (C-11), 72.89 (PEG), 71.26 (PEG), 70.94 (PEG), 70.42 (PEG), 70.19 (PEG), 69.41 (PEG), 68.46 (PEG), 62.83 (PEG); LRMS *m/z* 316.3 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{14}H_{21}NO_7$: 316.1391 ($[M+H]^+$); found: 316.1427 ($[M+H]^+$)

8.1.30 - Synthesis of tert-butyl 14-(4-nitrophenoxy)-3,6,9,12-tetraoxatetradecanoate



2-(2-(2-(2-(4-nitrophenoxy)ethoxy)ethoxy)ethoxy)ethan-1-ol (490 mg, 4.55 mmol) was added to a solution of potassium *tert*-butoxide (262 mg, 2.33 mmol) in *tert*-butanol (15 m) and stirred at 40°C for 30 minutes. *Tert*-butyl bromoacetate (253 µl, 1.71 mmol) was added slowly and the reaction stirred at 45°C for 6 hours. The reaction mixture was diluted with water (30 ml) and extracted with ethyl acetate (5x30 ml). Organics were combined, washed with brine, dried over sodium sulphate and concentrated. The crude was purified further by column chromatography (0-5% methanol in DCM) to yield a colourless oil (204 mg, 31%). R_f = 0.60 (ethyl acetate); 1H NMR (500 MHz, Chloroform-d) δ 8.20 (d, J = 9.2 Hz, 2H, H-15), 6.98 (d, J = 9.3 Hz, 2H, H-14), 4.24 – 4.20 (m, 2H, H-12), 4.01 (s, 2H, H-4), 3.92 – 3.87 (m, 2H, H-11), 3.75 – 3.65 (m, 12H, PEG-H), 1.47 (s, 9H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 163.87 (C-2), 125.87 (C-15), 114.60 (C-14), 81.59 (C-2), 70.93 (PEG), 70.73 (PEG), 70.65 (PEG), 70.63 (PEG), 70.62 (PEG), 70.59 (PEG), 69.39 (PEG), 69.03 (PEG), 68.23 (PEG), 28.12 (C-1); LRMS *m/z* 274.2 ($[M+H]^+$); LRMS *m/z* 430.5 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{20}H_{31}NO_9$: 430.2072 ($[M+H]^+$); found: 430.1927 ($[M+H]^+$).

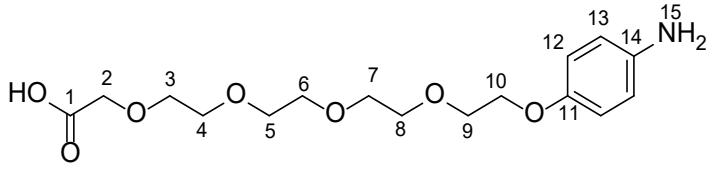
8.1.31 - Synthesis of tert-butyl 14-(4-aminophenoxy)-3,6,9,12-tetraoxatetradecanoate



tert-butyl 14-(4-nitrophenoxy)-3,6,9,12-tetraoxatetradecanoate (204 mg, 0.48 mmol) was dissolved in ethanol (15 ml) and methanol (15 ml) and subjected to hydrogenation on the H-Cube using a 20% Pd/C cartridge at 40°C. After 16 hours the reaction mixture was concentrated to dryness to yield a colourless oil (165 mg, 87%). R_f = 0.43 (DCM/MeOH 19:1); 1H NMR (500 MHz, Chloroform-d) δ 6.77 – 6.74 (m, 2H, H-14), 6.64 – 6.61 (m, 2H, H-15), 4.08

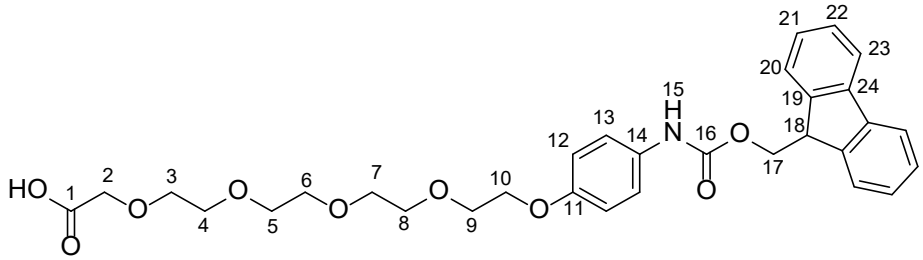
– 4.04 (m, 2H, H-12), 4.02 (s, 2H, H-4), 3.83 – 3.79 (m, 2H, H-11), 3.73 – 3.66 (m, 14H, PEG-H), 1.47 (s, 9H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 169.69 (C-3), 151.99 (C-13), 140.13 (C-16), 116.36 (C-14), 115.91 (C-15), 81.53 (C-2), 70.78 (PEG), 70.73 (PEG), 70.64 (PEG), 70.61 (PEG), 69.92 (PEG), 69.06 (PEG), 68.18 (PEG), 28.13 (C-1). LRMS *m/z* 400.5 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{20}H_{33}NO_7$: 400.2330 ($[M+H]^+$); found: 400.2232 ($[M+H]^+$).

8.1.32 - Synthesis of 14-(4-aminophenoxy)-3,6,9,12-tetraoxatetradecanoic acid



Tert-butyl 14-(4-aminophenoxy)-3,6,9,12-tetraoxatetradecanoate (263 mg, 0.66 mmol) was dissolved in TFA (3 ml) and DCM (3 ml) and stirred for 3 hours at RT. The reaction mixture was then concentrated under vacuum to yield an orange oil (225 mg, 99%) R_f = 0.72 (DCM/MeOH 19:1); IR: νmax/cm-1 3374 (N-H stretch), 2920 (C-H stretch), 1703 (C=O); 1H NMR (500 MHz, Chloroform-d) δ 7.13 (d, J = 8.3 Hz, 2H, H-13), 6.50 (d, J = 8.0 Hz, 2H, H-12), 4.00 (s, 2H, H-2), 3.88 – 3.40 (m, 16H, PEG-H); 13C NMR (126 MHz, Chloroform-d) δ 158.33 (C-1), 124.39 (C-11), 123.19 (C-14), 115.23 (C-12/13), 70.63-67.44 (PEG). LCMS *m/z* 344.4 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{16}H_{25}NO_7$: 344.1704 ($[M+H]^+$); found: 344.1824 ($[M+H]^+$)

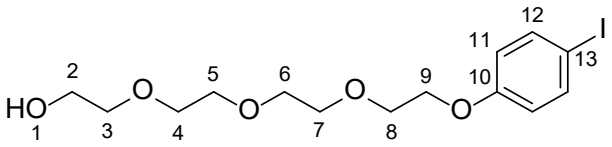
8.1.33 - Synthesis of 14-(4-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)phenoxy)-3,6,9,12-tetraoxatetradecanoic acid



14-(4-aminophenoxy)-3,6,9,12-tetraoxatetradecanoic acid (43 mg, 0.125 mmol) was dissolved in 10% aq. AcOH (1 ml) and dioxane (1 ml). FmocCl (34 mg, 0.131 mmol) was then added and the mixture stirred for 4 hours at RT. The reaction mixture was then concentrated and purified by column chromatography (0-20% MeOH in DCM) to yield a white solid (44 mg,

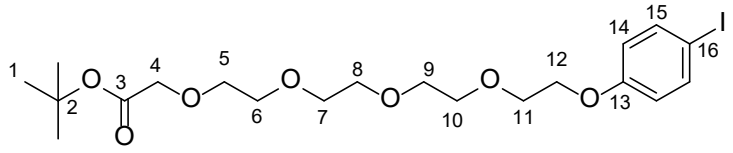
63%). R_f = 0.36 (DCM/MeOH 19:1); IR: νmax/cm-1 3321 (N-H stretch), 2879 (C-H stretch), 1695 (C=O), 1518 (C=O); 1H NMR (500 MHz, Chloroform-d) δ 7.70 (d, J = 7.6 Hz, 2H, H-23), 7.54 (s, 2H, H-20), 7.33 (t, J = 7.5 Hz, 2H, H-22), 7.28 – 7.19 (m, 4H, H-21 and H-13), 6.77 (d, J = 8.4 Hz, 2H, H-12), 4.47 – 4.42 (m, 2H, H-17), 4.21 – 4.16 (m, 1H, H-18), 4.04 – 3.99 (m, 4H, H-2 and H-10), 3.78 – 3.74 (m, 2H, H-9), 3.67 – 3.51 (m, 12H, PEG-H); 13C NMR (126 MHz, Chloroform-d) δ 143.81 (C-19), 141.36 (C-24), 127.75 (C-22/23), 127.11 (C-21), 124.99 (C-12), 120.02 (C-20), 115.12 (C-12), 70.82 (PEG), 70.58 (PEG), 70.45 (PEG), 70.33 (PEG), 70.19 (PEG), 69.74 (PEG), 67.70 (C-17), 47.18 (C-18). LCMS *m/z* 566.4 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{31}H_{35}NO_9$: 564.2185 ($[M-H]^-$); found: 564.2195 ($[M-H]^-$).

8.1.34 - Synthesis of 2-(2-(2-(2-(4-iodophenoxy)ethoxy)ethoxy)ethoxy)ethan-1-ol



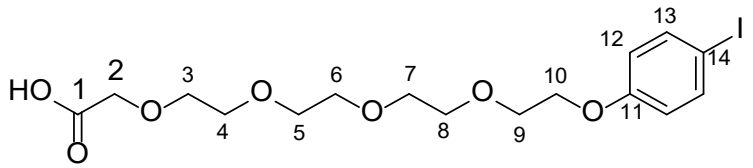
2-(2-(2-(2-hydroxyethoxy)ethoxy)ethoxy)ethyl 4-methylbenzenesulfonate (609 mg, 1.74 mmol) was dissolved in DMF (20 ml). Potassium carbonate (346 mg, 1.74 mmol) and 4-iodophenol (574 mg, 2.61 mmol) were then added and the reaction mixture stirred at 75°C for 5 hours. The reaction was quenched with water (30 ml), extracted with ethyl acetate (4 x 30 ml), dried over magnesium sulphate and concentrated to yield a colourless oil (681 mg, 98%). R_f = 0.44 (ethyl acetate); IR ν_{max}/cm^{-1} 3428 (OH), 1585 (Ar C=C); UV/Vis λ_{max} (Ethanol)/nm 280.8, 233.6; 1H NMR (500 MHz, Chloroform-d) δ 7.57 – 7.50 (m, 2H, H-12), 6.73 – 6.66 (m, 2H, H-11), 4.12 – 4.06 (m, 2H, H-9), 3.87 – 3.81 (m, 2H, H-8), 3.75 – 3.63 (m, 10H, PEG-H), 3.63 – 3.57 (m, 2H, PEG-H), 2.52 (t, J = 6.3 Hz, 1H, H-1).; 13C NMR (125 MHz, Chloroform-d) δ 158.82 (C-10), 138.33 (C-12), 117.22 (C-11), 83.10 (C-13), 72.64 (PEG), 70.98 (PEG), 70.82 (PEG), 70.74 (PEG), 70.50 (PEG), 69.78 (PEG), 67.69 (PEG), 61.92 (PEG); LRMS m/z 397.2 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{14}H_{21}IO_5$: 397.0506 ($[M+H]^+$); found: 397.0510 ($[M+H]^+$)

8.1.35 - Synthesis of tert-butyl 14-(4-iodophenoxy)-3,6,9,12-tetraoxatetradecanoate



2-(2-(2-(2-(4-iodophenoxy)ethoxy)ethoxy)ethoxy)ethan-1-ol (402 mg, 1.01 mmol) was added to a solution of potassium *tert*-butoxide (237 mg, 1.70 mmol) in *tert*-butanol (20 ml). The reaction mixture was stirred at 40°C for 30 minutes, followed by dropwise addition of *tert*-butylbromoacetate (237 mg, 1.22 mmol). The reaction mixture was stirred at 40°C overnight. The reaction mixture was then diluted with water (30 ml) and extracted with ethyl acetate (5 x 30 ml). The organics were combined, washed with brine, dried over sodium sulphate and concentrated. The crude was purified further by column chromatography (0-5% methanol in DCM) to yield a colourless oil (301.5 mg, 60%). R_f = 0.45 (1:1 Petroleum ether:ethyl acetate); IR ν_{max}/cm^{-1} 1744 (C=O), 1586 (Ar C=C); UV/Vis λ_{max} (Ethanol)/nm 280.6, 233.6; 1H NMR (500 MHz, Chloroform-d) δ 7.57 – 7.50 (m, 2H, H-15), 6.72 – 6.66 (m, 2H, H-14), 4.12 – 4.06 (m, 2H, H-12), 4.01 (s, 2H, H-4), 3.87 – 3.81 (m, 2H, H-11), 3.74 – 3.64 (m, 12H, PEG-H), 1.47 (s, 9H, H-1); 13C NMR (125 MHz, Chloroform-d) δ 169.79 (C-3), 158.86 (C-13), 138.32 (C-15), 117.23 (C-14), 83.05 (C-16), 81.66 (C-2), 71.01 (PEG), 70.89 (PEG), 70.79 (PEG), 70.77 (PEG), 70.76 (PEG), 69.76 (PEG), 69.21 (PEG), 67.72 (PEG), 28.27 (C-1); LRMS m/z 455.3 ([M-^{t}Bu+H]$^+$); HRMS (ESI): m/z calcd for $C_{20}H_{31}IO_7$: 511.1187 ([M+H]$^+$); found: 528.1469 ([M+NH4]$^+$), 533.1021 ([M+Na]$^+$)

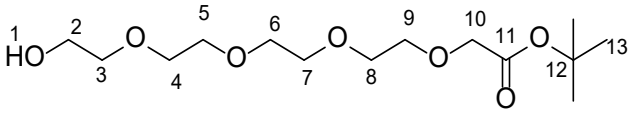
8.1.36 - Synthesis of 14-(4-iodophenoxy)-3,6,9,12-tetraoxatetradecanoic acid



Tert-butyl 14-(4-iodoophenoxy)-3,6,9,12-tetraoxatetradecanoate (328 mg, 0.62 mmol) was dissolved in TFA (10 ml) and DCM (10 ml) and was stirred for 3 hours at RT. The reaction mixture was then concentrated under vacuum to yield an orange oil (280 mg, 99%) R_f = 0.14 (Ethyl acetate); IR ν_{max}/cm^{-1} 3443 (COOH), 1735 (C=O); UV/Vis λ_{max} (Ethanol)/nm 280.2, 233.6; 1H NMR (500 MHz, Chloroform-d) δ 7.58 – 7.51 (m, 2H, H-13), 6.72 – 6.65 (m, 2H, H-12), 5.80

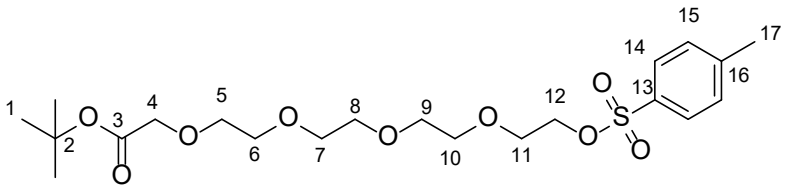
(s, 4H), 4.15 (s, 2H, H-10), 4.13 – 4.07 (m, 2H, H-9), 3.90 – 3.84 (m, 2H, PEG-H), 3.79 – 3.73 (m, 4H, PEG-H), 3.73 – 3.64 (m, 8H, PEG-H); 13C NMR (125 MHz, Chloroform-d) δ 173.68 (C-1), 158.48 (C-11), 138.23 (C-13), 117.01 (C-12), 83.12 (C-14), 71.45 (PEG), 70.75 (PEG), 70.50 (PEG), 70.31 (PEG), 70.26 (PEG), 70.20 (PEG), 69.64 (PEG), 68.85 (PEG), 67.27 (PEG); LRMS m/z 455.3 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{16}H_{23}IO_7$: 455.0561 ($[M+H]^+$); found: 455.0551 ($[M+H]^+$)

8.1.37 - Synthesis of tert-butyl 14-hydroxy-3,6,9,12-tetraoxatetradecanoate



PEG-4 (26.2 g, 135 mmol) was added to a solution of potassium tert-butoxide (1.8 g, 16.2 mmol) in tert-butanol (64 ml) and stirred at 50°C for 30 minutes, followed by slow addition of *tert*-butylbromoacetate (2 ml, 13.5 mmol). The reaction was then stirred at 50°C overnight. The reaction was diluted with water (100 ml), extracted with DCM (3x80 ml), organics combined, washed with 10% citric acid solution (2x50 ml), dried over sodium sulphate and concentrated to yield a colourless oil (2.08 g, 52%). 1H NMR (500 MHz, Chloroform-d) δ 3.96 (s, 2H, H-10), 3.68 – 3.53 (m, 16H, PEG-H), 1.41 (s, 9H, H-13); 13C NMR (126 MHz, Chloroform-d) δ 169.67 (C-11), 81.56 (C-12), 72.52 (PEG), 70.73 (PEG), 70.62 (PEG), 70.58 (PEG), 70.55 (PEG), 70.35 (PEG), 69.04 (PEG), 61.76 (C-2), 28.12 (C-13); HRMS (ESI): m/z calcd for $C_{14}H_{28}O_7$: 308.1908 ($[M+H]^+$); found: 308.1896 ($[M+H]^+$)

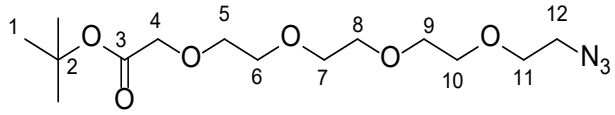
8.1.38 - Synthesis of tert-butyl 14-(tosyloxy)-3,6,9,12-tetraoxatetradecanoate



Tert-butyl 14-hydroxy-3,6,9,12-tetraoxatetradecanoate (2.08 g, 6.75 mmol) and TEA (7.85 ml, 13.51 mmol) were dissolved in DCM (20 ml). Toluene sulphonyl chloride (1.42 g, 7.43 mmol) was added at 0°C, and the reaction was stirred at room temperature overnight. The reaction was diluted with water (20 ml), extracted with DCM (2x30 ml), organics combined, washed

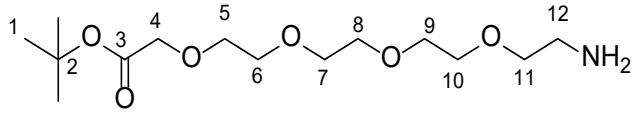
with saturated sodium bicarbonate solution (2x50 ml), dried over sodium sulphate and concentrated to yield a colourless oil (2.59 mg, 83%). R_f = 0.49 (1:1 Petroleum ether:ethyl acetate); 1H NMR (500 MHz, Chloroform-d) δ 7.73 (d, J = 8.3 Hz, 2H, H-14), 7.28 (d, J = 8.0 Hz, 2H, H-15), 4.11 – 4.06 (m, 2H, H-12), 3.97 – 3.92 (m, 2H, H-4), 3.68 – 3.53 (m, 14H, PEG-H), 2.38 (s, 3H, H-17), 1.40 (d, J = 1.3 Hz, 9H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 169.65 (C-3), 144.77 (C-13), 133.04 (C-16), 129.81 (C-15), 127.99 (C-14), 81.54 (C-2), 70.76 (PEG), 70.72 (PEG), 70.59 (PEG), 70.53 (PEG), 69.24 (PEG), 69.04 (PEG), 68.69 (PEG), 28.12 (C-1), 21.65 (C-17); LRMS m/z 462.2 ($[M+H]^+$); HRMS (ESI): m/z calcd for $C_{21}H_{34}O_9S$: 463.1996 ($[M+H]^+$); found: 463.1986 ($[M+H]^+$)

8.1.39 - Synthesis of tert-butyl 14-azido-3,6,9,12-tetraoxatetradecanoate



Tert-butyl 14-(tosyloxy)-3,6,9,12-tetraoxatetradecanoate (2.59 g, 5.61 mmol) was dissolved in DMF (100 ml). Sodium azide (0.55 g, 8.42 mmol) was added and the reaction stirred at 80°C overnight. The reaction was diluted with water (200 ml), extracted with ETOAc (3x200 ml), organics combined, washed with brine (3x100 ml), dried over sodium sulphate and concentrated to yield a colourless oil (1.81 g, 97%). IR: νmax/cm-1 2892, 2123 (N_3 stretch); 1H NMR (500 MHz, Chloroform-d) δ 3.96 (s, 2H, H-4), 3.67 – 3.58 (m, 14H, PEG-H), 3.33 (t, J = 5.1 Hz, 2H, H-12), 1.41 (s, 9H, H-1); 13C NMR (126 MHz, Chloroform-d) δ 169.67 (C-3), 81.55 (C-2), 70.72 (PEG), 70.69 (PEG), 70.65 (PEG), 70.64 (PEG), 70.61 (PEG), 70.59 (PEG), 70.03 (PEG), 69.04 (PEG), 50.70 (C-12), 28.12 (C-1); HRMS (ESI): m/z calcd for $C_{14}H_{27}N_3O_6$: 334.1973 ($[M+H]^+$); no mass found.

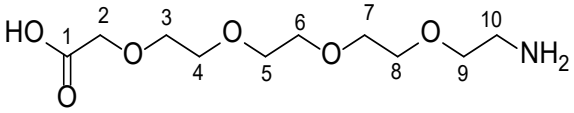
8.1.40 - Synthesis of tert-butyl 14-amino-3,6,9,12-tetraoxatetradecanoate



Tert-butyl 14-azido-3,6,9,12-tetraoxatetradecanoate (1.81 g, 5.42 mmol) and triphenylphosphine (1.7 g, 6.50 mmol) were dissolved in THF (54 ml) and stirred at room temperature for 2 hours. Water (0.2ml) was added and the reaction mixture stirred at room

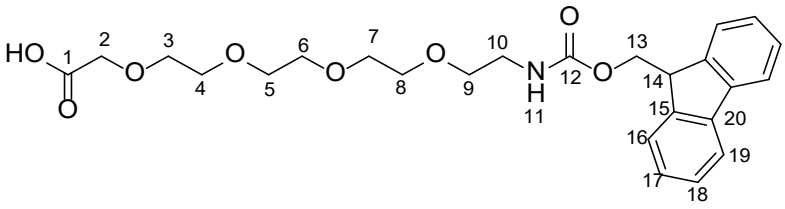
temperature overnight. The reaction was concentrated to dryness, redissolved in MeOH and purified by SCX-2 silica (10g), using 2M ammonia in MeOH to elute the compound as a colourless oil (1.49 g, 89%). 1H NMR (500 MHz, Chloroform-d) δ 3.95 (s, 2H, H-4), 3.66 – 3.55 (m, 14H, PEG-H), 2.87 – 2.83 (m, 2H, H-12), 1.40 (s, 9H, H-1).; 13C NMR (126 MHz, Chloroform-d) δ 169.77 (C-3), 81.73 (C-2), 72.51 (C-11), 70.66 (PEG), 70.54 (PEG), 70.52 (PEG), 70.45 (PEG), 70.16 (PEG), 69.01 (PEG), 41.57 (C-12), 28.11 (C-1); HRMS (ESI): m/z calcd for $C_{14}H_{29}NO_6$: 308.2068 ($[M+H]^+$); found: 308.1977 ($[M+H]^+$)

8.1.41 - Synthesis of 14-amino-3,6,9,12-tetraoxatetradecanoic acid



Tert-butyl 14-amino-3,6,9,12-tetraoxatetradecanoate (1.49 mg, 4.82 mmol) was dissolved in TFA (20 ml) and DCM (20 ml) and stirred for 6 hours at RT. The reaction mixture was then concentrated under vacuum to yield an orange oil (1.20 mg, 99%) 1H NMR (500 MHz, Chloroform-d) δ 4.10 (s, 2H, H-2), 3.77 – 3.72 (m, 2H, H-9), 3.69 – 3.56 (m, 14H, PEG-H), 3.22 – 3.13 (m, 2H, H-10); 13C NMR (126 MHz, Chloroform-d) δ 173.74 (C-1), 70.30 (PEG), 70.05 (PEG), 69.73 (PEG), 69.63 (PEG), 69.48 (PEG), 68.01 (PEG), 66.66 (PEG), 39.83 (C-10); HRMS (ESI): m/z calcd for $C_{10}H_{21}NO_6$: 252.1442 ($[M+H]^+$); found: 252.1461 ($[M+H]^+$)

8.1.42 - Synthesis of 1-(9H-fluoren-9-yl)-3-oxo-2,7,10,13,16-pentaoxa-4-azaoctadecan-18-oic acid

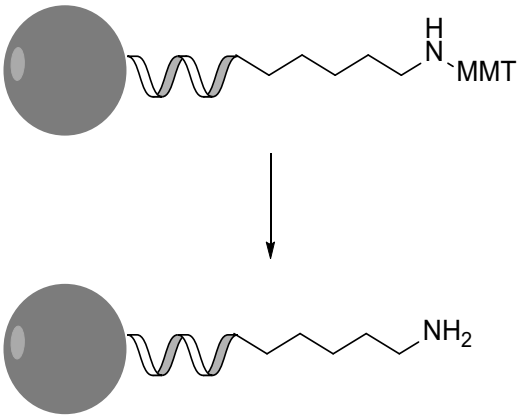


14-amino-3,6,9,12-tetraoxatetradecanoic acid (426 mg, 1.69 mmol) was dissolved in dioxane (20 ml) and water (15 ml). The pH was adjusted to 8 with $NaHCO_3$ before the addition of FmocCl (524 mg, 2.03 mmol) at 0°C portion wise. The reaction was stirred overnight at room temperature. The reaction mixture was then concentrated under vacuum, using toluene to azeotrope the water. It was then purified by column chromatography (0-15% MeOH in DCM) to yield a colourless oil (329 mg, 41%); R_f = 0.53 (1:1 Petroleum ether:ethyl acetate); IR: νmax/cm-1 3343 (N-H stretch), 2870 (C-H stretch), 1704 (C=O); 1H NMR (500 MHz,

Chloroform-d) δ 7.69 (dt, J = 7.6, 1.0 Hz, 2H, H-19), 7.54 (d, J = 7.5 Hz, 2H, H-16), 7.33 (tt, J = 7.5, 0.9 Hz, 2H, H-18), 7.24 (td, J = 7.5, 1.2 Hz, 2H, H-17), 5.37 (s, 1H, NH-11), 4.33 (d, J = 7.0 Hz, 2H, H-13), 4.15 (t, J = 6.9 Hz, 1H, H-14), 4.07 (s, 2H, H-2), 3.68 – 3.46 (m, 14H, PEG-H), 3.36 – 3.29 (m, 2H, H-10); 13C NMR (126 MHz, Chloroform-d) δ 170.85 (C-1), 156.53 (C-12), 144.02 (C-15/20), 141.32 (C-15/20), 127.65 (C-18/19), 127.03(C-18/19), 125.07(C-17), 119.94 (C-16), 70.92 (PEG), 70.62 (PEG), 70.57 (PEG), 70.36 (PEG), 70.07 (PEG), 68.59 (C-13), 66.57 (C-12), 51.77 (C-14), 47.31 (C-10); HRMS (ESI): m/z calcd for $C_{25}H_{31}NO_8$: 472.1977 ($[M-H]^-$); found: 472.2003 ($[M-H]^-$).

8.2 - General Suzuki-Miyaura DNA Methods

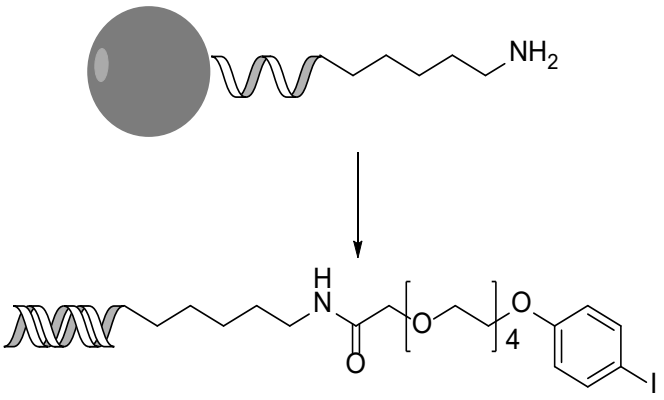
8.2.1 - MMT deprotection



The average loading of single stranded DNA attached to solid support was discovered by cleavage from the solid support using the below method and repeating 3 times. Nanodrop concentration of cleaved DNA showed that 103 mg gave 2 µmol of impure DNA.

The single stranded DNA used was a 14mer (GTCTTGCCGAATTC) with a 5' MMT amino C6 linker bound to solid support at the 3' end. The solid supported DNA (103 mg, ca. 2 µmol) was washed with 5% trichloroacetic acid in DCM (6 x 500 µl). A yellow colour indicated that the deprotection was in progress. Once this colour subsides the solid supported DNA was washed with DCM (3 x 500 µl) and left to air dry for 20 mins before coupling to the headpiece.

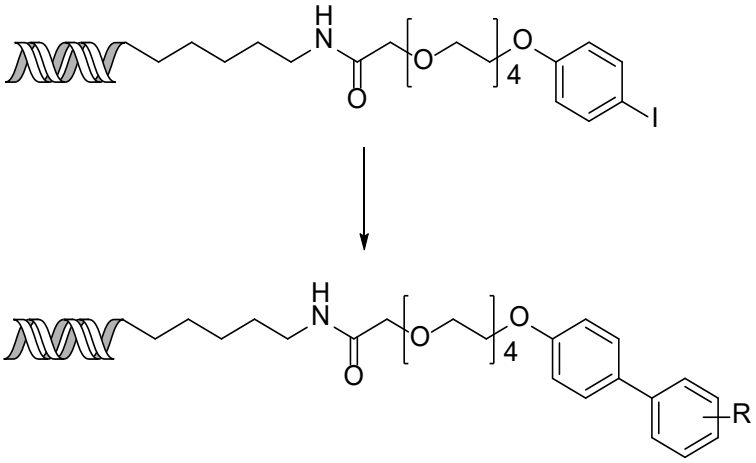
8.2.2 - Iodo-headpiece coupling and cleavage to give 50



In a 1.5 ml micro centrifuge tube was added HATU (17 mg, 44 µmol), DIPEA (17 µl, 100 µmol) and DMF (1 ml). To this was added 14-(4-iodophenoxy)-3,6,9,12-tetraoxatetradecanoic acid (18 mg, 40 µmol) and the mixture shaken for 20 mins at room temperature. The deprotected solid supported DNA (ca. 2 µmol) was added and the mixture shaken at room temperature

overnight. The mixture was then filtered and washed with DMF (3 x 500 µl), MeCN (3 x 500 µl), MeOH (3 x 500 µl) and DCM (3 x 500 µl) and allowed to air dry for 20 mins. 40% methylamine in water (500 µl) and 33% ammonia in water (500 µl) were mixed in a 1.5 ml centrifuge tube. The solid supported DNA was then added and the mixture shaken overnight at room temperature. The mixture was then filtered and washed with water (3 x 500 µl) and concentrated to ~0.5 ml using a Genevac at 40°C. The crude product was then purified by HPLC, fractions concentrated using a Genevac at 40°C and dissolved in water (1 ml). The concentration of the samples was then quantified by UV using a NanoDrop1 by Thermofisher. The usual amount was ca. 1 µmol of DNA after HPLC purification. The exact amount of the complimentary 14mer (GAATTCGGCAAGAC) was then added in water, and the mixture heated to 80°C for 1 hour, allowing to cool slowly. The double stranded DNA was then concentrated using a Genevac at 40°C until dry and dissolved in either water or 2% TPGS-750-M in water to form a 1 mM solution of **50**. The solution was then plated into wells of 20 µl with 20 nmol of DNA per well and frozen at -20°C. The samples were defrosted at room temperature for 30 mins prior to use in reactions.

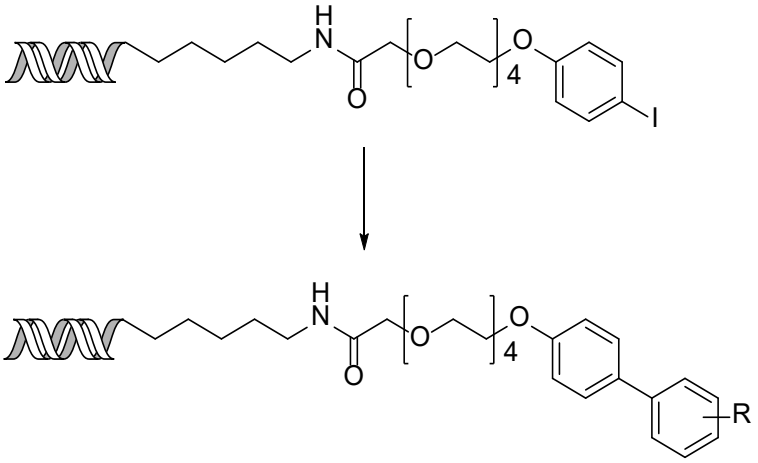
8.2.3 - Final Suzuki coupling conditions



An aliquot of boronic acid solution (20 µl, 0.75 M in DMF) was added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The DMF was then removed at 55°C in a Genevac for 30 mins. To this solution was added 5% TPGS-750-M (4 µl), potassium phosphate (6 µl, 113.23 mg in 200 µl water) and **50** (20 µl, 1 mM in 2% TPGS-750-M in water). The vials were vortexed for 30 seconds each to enhance mixing. $Pd(dtbpf)Cl_2$ (4.5 µl, 6.37mg in 200 µl THF) was then added and the samples vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at

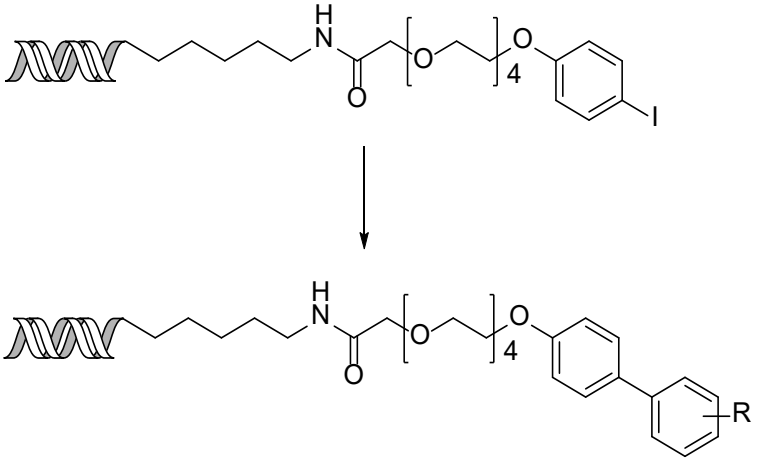
60°C for 5 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter. Chloroform (3 x 40 µl) was added to each and the vial vortexed. The organics were removed, aqueous sodium chloride (4 µl, 4M) and ethanol (70 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. Ethanol 70% (70 µl) was added and the process repeated. The pellet of DNA was then dissolved in water to give a 1mM solution.

8.2.4 - Literature Suzuki coupling conditions



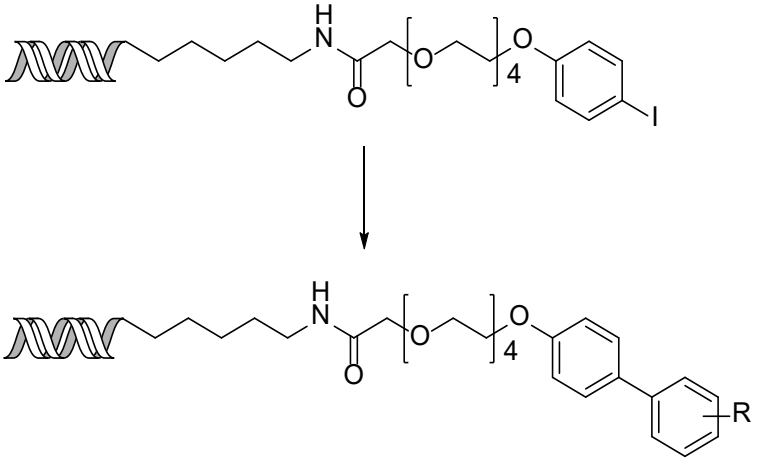
To DNA starting material **50** in water (1 mM, 20 nmol, 20 µl) was added Na_2CO_3 (0.5 µl, 1.6 M in water), and boronic acids (1 µl, 400 mM in 1:1 ACN:water). To this solution was then added degassed palladium tetrakistriphenylphosphine (0.5 µl, 20 mM in 1:2:2 DCM:tol:ACN) and the solution allowed to heat overnight at 80°C. Analysis was carried out by mass spectrometry using a sample of reaction media (5 µl) diluted into water (595 µl) and filtered through a PTFE filter. Reactions were not purified.

8.2.5 – Initial Suzuki coupling conditions (Table 2)



To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate the boronic acid (2 µMol) was added along with 5% NOK (4 µl), water (6 µl), **50** (20 µl, 0.1 mM in 2% NOK in water) and TEA (0.38 µl, 3 µMol). The vials were vortexed for 30 seconds each to enhance mixing. Pd(dtbpf)Cl2 (1 µl, 0.26 mg in 200 µl DMF) was then added and the samples vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 40°C for 16 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

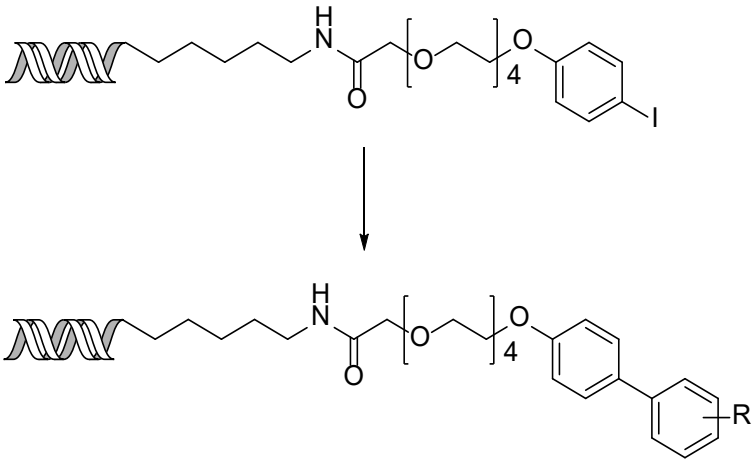
8.2.6 – Suzuki base screen conditions (Table 3)



To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate 4-trifluoromethylphenyl boronic acid (0.38 mg, 2 µMol) was added along with 5% NOK (4 µl), water (6 µl), **50** (20 µl, 0.1 mM in 2% NOK in water) and base (3 µMol). The vials were vortexed for 30 seconds each to enhance mixing. Pd(dtbpf)Cl2 (1 µl, 0.26 mg in 200 µl DMF) was then

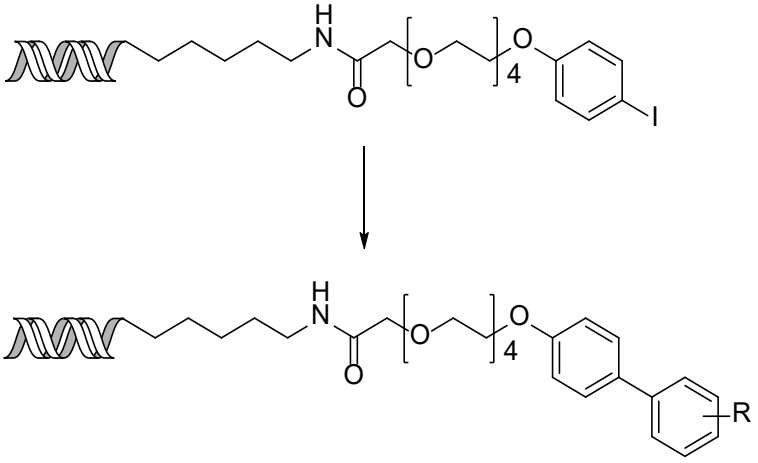
added and the samples vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 40°C for 16 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

8.2.7 – Suzuki palladium screen conditions (Table 4 and 5)



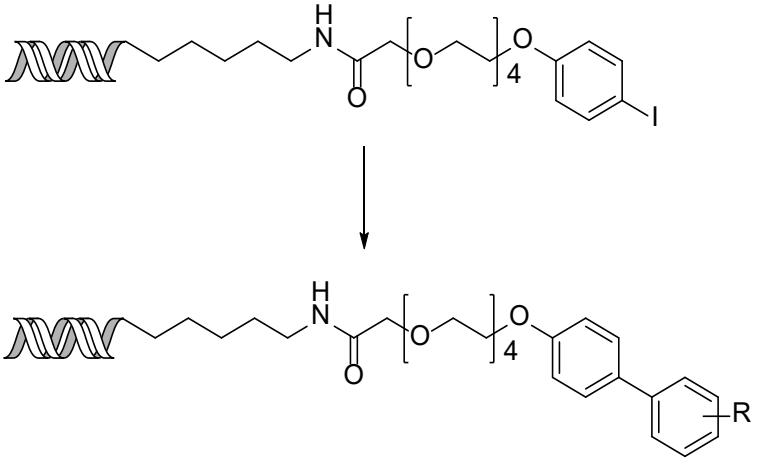
To a 50 µl glass insert for a Paradox™ 96-well micro photoredox/optimisation Plate 4-trifluoromethylphenyl boronic acid (0.38 mg, 2 µMol) was added along with 5% NOK (4 µl), water (6 µl), **50** (20 µl, 0.1 mM in 2% NOK in water) and LiOH (0.07 mg, 3 µMol). The vials were vortexed for 30 seconds each to enhance mixing. Where 0.1 eq was used, Pd(dtbpf)Cl2 (1 µl, 0.03 mg in 200 µl DMF) was then added; here 1 eq was used, Pd(dtbpf)Cl2 (1 µl, 0.26 mg in 200 µl DMF) was then added; where 2 eq was used, Pd(dtbpf)Cl2 (1 µl, 0.52 mg in 200 µl DMF) was then added; where 10 eq was used, Pd(dtbpf)Cl2 (1 µl, 2.6 mg in 200 µl DMF) was then added; and the samples vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 50°C for 16 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

8.2.8 – initial TPGS-750-M Suzuki coupling conditions (Table 6 and 7)



To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate the boronic acid (15 µMol) was added along with 5% TPGS-750-M (4 µl), water (6 µl), base (22.5 µMol) and **50** (20 µl, 0.1 mM in 2% TPGS-750-M in water). The vials were vortexed for 30 seconds each to enhance mixing. Pd(dtbpf)Cl_2 (4.5 µl, 0.58 mg in 200 µl DMF or THF) was then added and the samples vortexed again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 50°C for 16 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

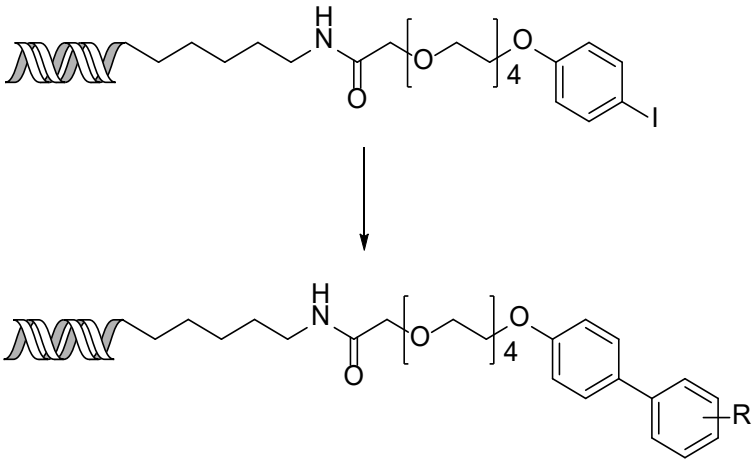
8.2.9 – Lower base concentration conditions (Table 8)



To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate the boronic acid (15 µMol) was added along with 5% TPGS-750-M (4 µl), water (6 µl), Potassium phosphate (0.32 mg, 1.5 µMol) and **50** (20 µl, 0.1 mM in 2% TPGS-750-M in water). The vials were vortexed for 30 seconds each to enhance mixing. Pd(dtbpf)Cl_2 (4.5 µl, 0.58 mg in 200 µl

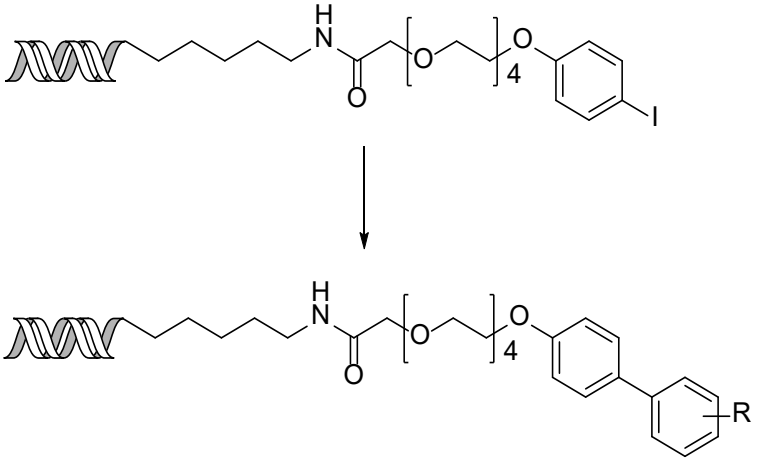
THF) was then added and the samples vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 50°C for 16 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

8.2.10 – higher palladium concentration conditions (Table 9)



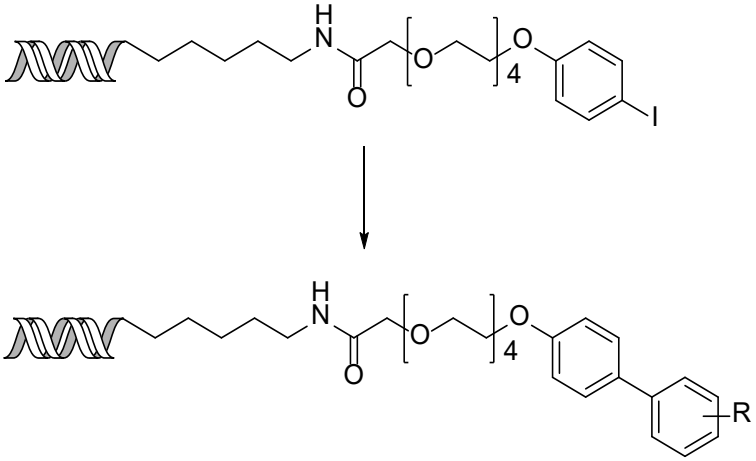
To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate the boronic acid (15 µMol) was added along with 5% TPGS-750-M (4 µl), water (6 µl), Potassium phosphate (4.8 mg, 22.5 µMol) and **50** (20 µl, 0.1 mM in 2% TPGS-750-M in water). The vials were vortexed for 30 seconds each to enhance mixing. Pd(dtbpf)Cl_2 (4.5 µl, 4.35 mg in 200 µl THF) was then added and the samples vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 50°C for 16 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

8.2.11 – FED conditions (Table 10)



An aliquot of boronic acid solution (20 µl, 0.75 M in DMF) was added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The DMF was then removed at 55 °C in a Genevac for 30 mins. To this solution was added 5% TPGS-750-M (4 µl) and **50** (20 µl, 0.1 mM in 2% TPGS-750-M in water). Potassium phosphate (6 µl in water) was added. Where 100 eq was used (7.08 mg in 100 µl of water); where 800 eq was used (56.61 mg in 100 µl of water); where 1500 eq was used (106.17 mg in 100 µl of water). The samples were then vortexed for 30 seconds to enhance mixing. Pd(dtbpf)Cl2 (4.5 µl in THF) was added. Where 2 eq was used (0.58 mg in 100 µl of THF); where 11 eq was used (3.19 mg in 100 µl of THF); where 20 eq was used (5.79 mg in 100 µl of THF). The samples were vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 40 °C, 50 °C or 60 °C for 5 hours. Mass spectrometry was used to analyse the DOE. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

8.2.12 – Control Suzuki coupling conditions (Table 12)



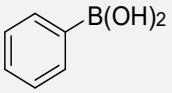
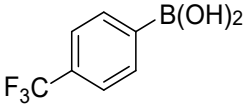
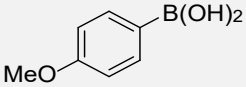
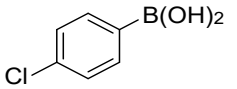
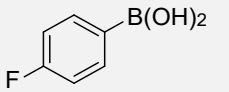
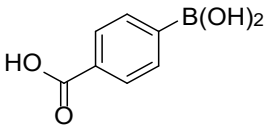
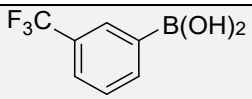
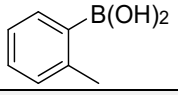
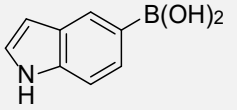
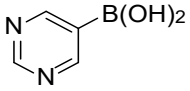
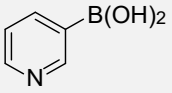
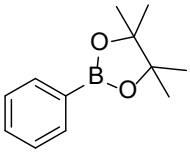
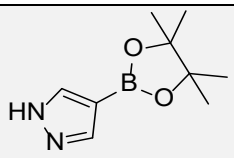
An aliquot of boronic acid solution (20 µl, 0.75 M in DMF) and Pd(dtbpf)Cl2 (4.5 µl, 6.37mg in 200 µl DMF) were added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The DMF was then removed at 55°C in a Genevac for 30 mins. To this solution was added either 5% TPGS-750-M (4 µl) or water (4 µl) or THF (4.5 µl) as required. Potassium phosphate (6 µl, 113.23 mg in 200 µl water) and **50** (20 µl, 0.1 mM in 2% TPGS-750-M in water or 0.1 mM in water) were then added and vials vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 60°C for 5 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (5 µl) to water (595 µl) and filtered through a hydrophilic PTFE filter.

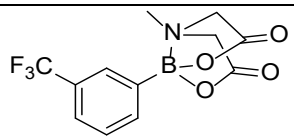
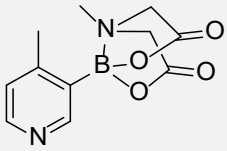
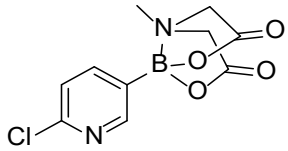
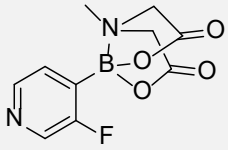
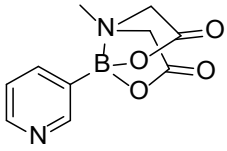
8.3 – Suzuki coupling Mass spectrometry results

Experimental Table 1 - Calculated and found masses for starting material and dehalogenated side product in the Suzuki-Miyaura coupling. Calculated masses from ChemDraw. Analytical procedure used - Agilent advancedbio oligonucleotides column, 2.1x150 mm where the gradient was ran at 0.4 ml/min from 10% MeOH to 40% MeOH over 8 mins against a 200 mM HFIP:8 mM TEA buffer solution. A 3 min flush at 95% MeOH preceded each run.

Compound		RT	Calc. Mass	Mass Found
50	Linker strand	9.758	4842.8421	4842.8713
	Complimentary strand	7.058	4294.7735	4294.8035
52	Dehalogenated side product	8.824	4716.9454	4716.9263

Experimental Table 2 – Calculated and found masses for each product formed in the Suzuki-Miyaura coupling. Calculated masses from ChemDraw. Analytical procedure used - Agilent advancedbio oligonucleotides column, 2.1x150 mm where the gradient was ran at 0.4 ml/min from 10% MeOH to 40% MeOH over 8 mins against a 200 mM HFIP:8 mM TEA buffer solution. A 3 min flush at 95% MeOH preceded each run.

Boronate used to produce 51x		% Prod	% SM, 50	% Dehal, 52	RT	Calc. Mass	Mass Found
a		97	0	0	9.963	4792.9751	4792.9648
b		100	0	0	10.969	4860.9641	4860.9630
c		100	0	0	9.979	4822.9873	4822.9619
d		100	0	0	10.470	4826.9378	4826.9064
e		100	0	0	10.016	4810.9673	4810.9475
f		100	0	0	8.523	4836.9666	4836.9517
g		100	0	0	10.922	4860.9641	4860.9326
h		100	0	0	10.452	4806.9924	4806.96533
i		100	0	0	9.235	4831.9876	4831.9643
j		100	0	0	8.657	4794.9672	4794.9413
k		100	0	0	9.277	4793.9720	4793.9623
l		96	0	0	9.977	4792.9767	4792.9537
m		97	3	0	8.436	4782.9672	4782.9389

n		96	0	4	10.941	4860.9641	4861.9321
o		100	0	0	9.748	4807.9876	4807.9564
p		81*	0	1	9.519	4827.9330	4827.8990
					10.215	4904.9596	4904.9217
q		98	0	2	9.322	4811.9626	4811.9371
r		100	0	0	9.285	4793.9720	4793.9407

Experimental Table 3 - Calculated and found masses for each product formed in the Suzuki-Miyaura coupling on diverse scaffolds. Calculated masses from ChemDraw. Analytical procedure used - Agilent advancedbio oligonucleotides column, 2.1x150 mm where the gradient was ran at 0.4 ml/min from 10% MeOH to 40% MeOH over 8 mins against a 200 mM HFIP:8 mM TEA buffer solution. A 3 min flush at 95% MeOH preceded each run.

Compound	RT	Calc. Mass	Mass Found
62	8.712	3720.6198	3720.5964
63	9.094	3578.6282	3578.6113
64	9.612	3626.6143	3626.5970
65	7.623	3579.6598	3579.6482
	7.789	3579.6598	3579.6487
	7.896	3577.6442	3577.6359
66	8.102	3627.6460	3627.6194
	8.210	3627.6460	3627.6244
	8.386	3625.6304	3625.6037
67	9.308	3670.7545	3670.7312
68 – from 63	10.448	3576.7490	3576.7287
68 – from 64	10.355	3576.7490	3576.7308
69 – from 65	8.654	3577.7806	3577.7636
	8.868	3577.7806	3577.7668
	9.095	3575.7650	3575.7493
69 – from 66	8.984	3577.7806	3577.7584
	9.179	3577.7806	3577.7549
	9.456	3575.7650	3575.7370

8.4 - 6x6 protocol Suzuki library synthesis

8.4.1 - General ethanol precipitation method

To the reaction mixture was added 10% volume NaCl (5 M in water), and 3x volume cold ethanol. The mixture was allowed to sit for 1 hour at -78 °C, or overnight at -20 °C. The mixture was then centrifuged at 13500 rpm for 30 mins. The supernatant decanted, and 70% cold EtOH added. The mixture was allowed to sit at -78 °C for 1 hour, before being centrifuged at 13500 rpm for 30 mins and the supernatant decanted. The pellet was then allowed to air dry before being redissolved in the appropriate solvent.

8.4.2 - Amide coupling reactions

Amide couplings were performed using the general procedure from literature[32]

To a 150 µl PCR tube was added 4-iodobenzoic acid or 4-iodophenylacetic acid (45 µl, 60 mM in DMSO), EDC.HCl (4 µl, 300 mM in DMSO), HOAT (4 µl, 60 mM in DMSO) and DIPEA (4 µl, 300 mM in DMSO). To this mixture was added PEG4-NH_2 DNA conjugate formed in the pervious step (0.5 µl, 500 pmol, 1mM) diluted in MOPS buffer (72 µl, pH 8, 50 mM, 0,5M NaCl). The mixture was shaken at room temperature overnight. EDC.HCl (4 µl, 300 mM in DMSO) and HOAT (4 µl, 60 mM in DMSO) were then added to the reaction mixture and shaken for 6 hours. The reaction was quenched for 30 mins with NH_4OAc (500 mM, 25 µl). The mixture was then purified using the general ethanol precipitation procedure.

8.4.3 - Suzuki library reaction

An aliquot of boronic acid solution (20 µl, 0.75 M in DMF) was added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The DMF was then removed at 55°C in a Genevac for 30 mins. To this solution was added 5% TPGS-750-M (4 µl), potassium phosphate (6 µl, 113.23 mg in 200 µl water) and split library of DNA conjugates (20 µl, 500 pmol in 2% TPGS-750-M in water). The vials were vortexed for 30 seconds each to enhance mixing. Pd(dtbpf)Cl_2 (4.5 µl, 6.37mg in 200 µl THF) was then added and the samples vortexted again for 10 seconds each. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 60°C for 5 hours. The reactions were cooled, and then pooled together in a 1.5 ml centrifuge tube, washing each well with water (2 x 20 µl). Chloroform (3 x 200 µl) was added to each, the vial vortexed and the organics removed. The mixture was then purified using the general ethanol precipitation procedure.

8.4.4 - General ligation protocol

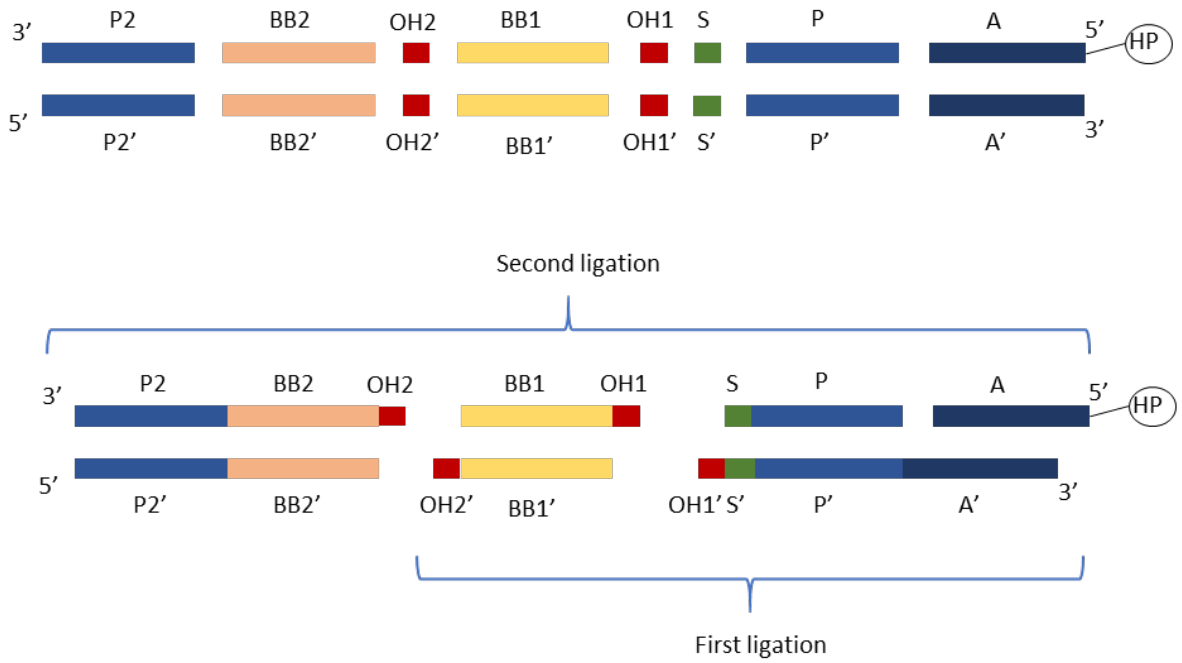


Figure 117 – Encoding strategy to be used in the 6x6 prototype library.

Experimental Table 4 - Codes, functions, and corresponding sequence for each DNA section

Code	Function	sequence 5'-3'
A	Adapter – 5' aminolinked head piece	GTCTTGCCGAATTC
A'	Complimentary adapter	GAATTCGGCAAGAC
P	Primer	AGGTCGGTGTGAACGGATTTG
P'	Complementary primer	CAAATCCGTTCACACCGACCT
S	Scaffold code	GCT
S'	Complementary S	AGC
OH1	Ligation overhang 1	GTAT
OH1'	Complementary OH1	ATAC
BB1	Building block 1	xxxxxxxx
BB1'	Complementary BB1	xxxxxxxx
OH2	Ligation overhang 2	CCTA
OH2'	Complementary OH2	TAGG
BB2	Building block 2	xxxxxxxx
BB2'	Complementary BB2	xxxxxxxx
P2	Complementary to P2'	TGACCTCAACTACATGGTCTACA
P2'	Primer (reverse)	TGTAGACCATGTAGTTGAGGTCA

Phosphorylation

Prior to ligation, the 5' terminus of each strand was phosphorylated. DNA strands (14 µM, 616 pmol in overall reaction media of 44 µl) were added PNK reaction buffer (4.4 µl, 500mM Tris-HCl [pH 7.6 at 25°C], 100mM $MgCl_2$, 50 mM DTT, 1mM spermidine), ATP (0.5 µl, 100 mM, Thermo Scientific), T4 Polynucleotide Kinase (10U/µl, Thermo Scientific) and nuclease free water (up to 44 µl). The reaction was carried out at 37 °C for 20 mins, followed by heating to 75 °C for 10 mins. DNA was used in the ligation steps without purification or precipitation.

Ligation

Ligations contained DNA (2 µM, 280 pmol in overall reaction media of 140 µl), 10X T4 DNA ligase buffer (14 µl, 400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP9), water (up to 140 µl) and T4 DNA Ligase (7 µl, 30 Weiss U/µL). The ligations were carried out at 25°C for 6 hours, followed by heating to 75°C for 10 mins. Each ligation was purified by ethanol precipitation prior to the subsequent organic reaction taking place.

8.4.5 - PCR and NGS

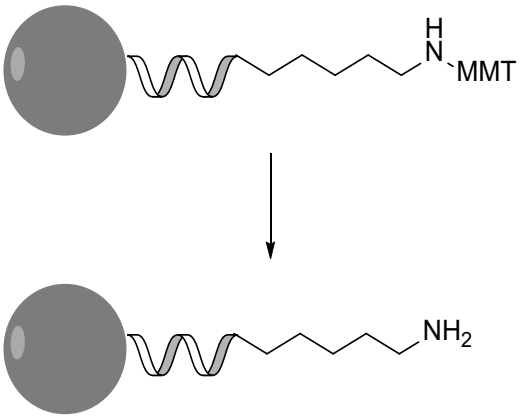
Forward (JHH-FSP-P) and reverse (JHH-FSP-P2) primers were designed to amplify the DEL library, flanked by 5' Illumina adapter sequences to enable downstream sequence analysis. Each PCR was performed in a 50µl reaction mixture containing AmpliTaq Gold® 360 Master Mix (Thermo Fisher) and 100ng of 6x6 DEL library (at 3.6 µM). The final concentration of each primer used was 10mM. PCRs were carried out using a Bio-Rad T100 Thermal cycler. The thermal cycling conditions were as follows: 10 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 30 s at 70°C, and 30 s at 72°C, with a final extension time of 420s at 72°C. Negative controls (distilled water in place of template, single and double primer controls) were included in each run. Following the PCR reactions the samples were run on a 2.5% NuSieve Agarose gel in TAE, gel extracted and sequenced using Next Generation Sequencing (GeneWiz, South Plainfield, NJ, USA). FastQC (version 0.11.8) was used to assess the quality of sequenced reads and to extract overrepresented sequences which make up more than 0.1% of the total (n=122298).

Experimental Table 5 - Forward and reverse primers for PCR and NGS. Primer sequence (in blue) and NGS elongation sequence (in purple).

Primer	sequence 5'-3'	length
Forward primer	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTAGAC CATGTAGTTGAGGTCA	56
Reverse primer	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGTCGG TGTGAACGGATTTG	53

8.5 - General Amide DNA Methods

8.5.1 - MMT deprotection

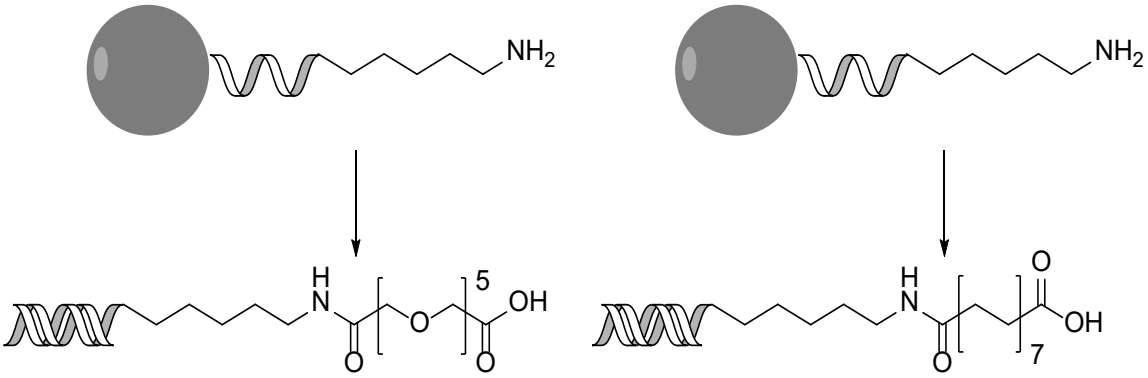


The average loading of single stranded DNA attached to solid support was found by cleavage from the solid support using the below method and repeating 3 times. Nanodrop concentration of cleaved DNA showed that 103 mg gave 2 µmol of DNA.

The single stranded DNA used was a 14mer (GTCTTGCCGAATTC) with a 5' MMT amino C6 linker bound to solid support at the 3' end. The solid

supported DNA (103 mg, ca. 2 µmol) was washed with 5% trichloroacetic acid in DCM (6 x 500 µl). A yellow colour indicates that the deprotection is in progress. Once this colour subsides the solid supported DNA was washed with DCM (3 x 500 µl) and left to air dry for 20 mins before coupling to the headpiece.

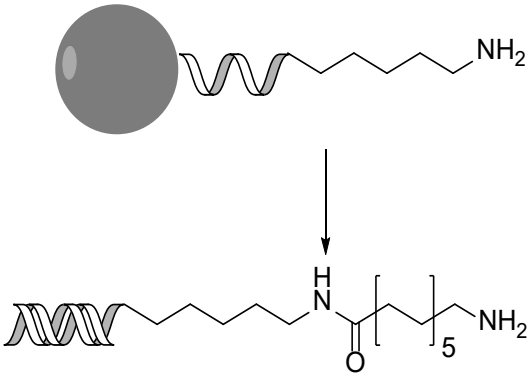
8.5.2 - General acid linker coupling and cleavage to give 70 and 71



In a 1.5 ml micro centrifuge tube was added HATU (17 mg, 44 µmol), DIPEA (17 µl, 100 µmol) and DMF (1 ml). To this was added either hexadecanedioic acid (17 mg, 40 µmol) or 3,6,9,12,15-pentaoxaheptadecanedioic acid (18 mg, 40 µmol) and the mixture shaken for 20

mins at room temperature. The deprotected solid supported DNA (ca. 2 µmol) was added and the mixture shaken at room temperature overnight. The mixture was then filtered and washed with DMF (3 x 500 µl), MeCN (3 x 500 µl), MeOH (3 x 500 µl) and DCM (3 x 500 µl) and allowed to air dry for 20 mins. Water (1.5 ml) was added and the mixture shaken for 1 hour at room temperature, before filtration and washing with water (3 x 500 µl). 40% methylamine in water (500 µl) and 33% ammonia in water (500 µl) were mixed in a 1.5 ml centrifuge tube. The solid supported DNA was then added and the mixture shaken for 1 hour at room temperature. The mixture was then filtered and washed with water (3 x 500 µl) and concentrated to ~0.5 ml using a Genevac at 40°C. The crude product was then purified by HPLC, fractions concentrated using a Genevac at 40°C and dissolved in water (1 ml). The concentration of the samples was then quantified by UV using a NanoDrop1 by Thermofisher. The usual amount was ca. 0.5-1 µmol of DNA after HPLC purification. The exact amount of the complimentary 14mer (GAATTCGGCAAGAC) was then added in water, and the mixture heated to 80°C for 1 hour, allowing to cool slowly. The double stranded DNA was then concentrated using a Genevac at 40°C until dry and dissolved in either water or 2% TPGS-750-M in water to form a 1 mM solution of **70** or **71**. The solution was then plated into wells of 20 µl with 20 nmol of DNA per well and frozen at -20°C. The samples were defrosted at room temperature for 30 mins prior to use in reactions.

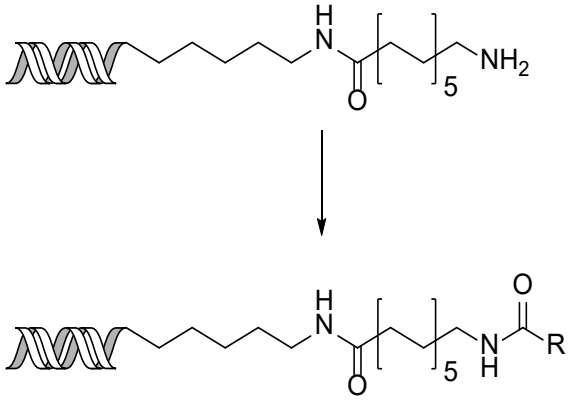
8.5.3 - General acid linker coupling and cleavage to give 74



In a 1.5 ml micro centrifuge tube was added HATU (17 mg, 44 µmol), DIPEA (17 µl, 100 µmol) and DMF (1 ml). To this was added 12-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)dodecanoic acid (26 mg, 40 µmol) and the mixture shaken for 20 mins at room temperature. The deprotected solid supported DNA (ca. 2 µmol) was added and the mixture shaken at room temperature overnight. The mixture was then filtered and

washed with DMF (3 x 500 µl), MeCN (3 x 500 µl), MeOH (3 x 500 µl) and DCM (3 x 500 µl) and allowed to air dry for 20 mins. 40% methylamine in water (500 µl) and 33% ammonia in water (500 µl) were mixed in a 1.5 ml centrifuge tube. The solid supported DNA was then added and the mixture shaken for 1 hour at room temperature. The mixture was then filtered and washed with water (3 x 500 µl) and concentrated to ~0.5 ml using a Genevac at 40°C. The crude product was then purified by HPLC, fractions concentrated using a Genevac at 40°C and dissolved in water (1 ml). The concentration of the samples was then quantified by UV using a NanoDrop1 by Thermofisher. The usual amount was ca. 0.5-1 µmol of DNA after HPLC purification. The exact amount of the complimentary 14mer (GAATTCGGCAAGAC) was then added in water, and the mixture heated to 80°C for 1 hour, allowing to cool slowly. The double stranded DNA was then concentrated using a Genevac at 40°C until dry and dissolved in either water or 2% TPGS-750-M in water to form a 1 mM solution of **74**. The solution was then plated into wells of 20 µl with 20 nmol of DNA per well and frozen at -20°C. The samples were defrosted at room temperature for 30 mins prior to use in reactions.

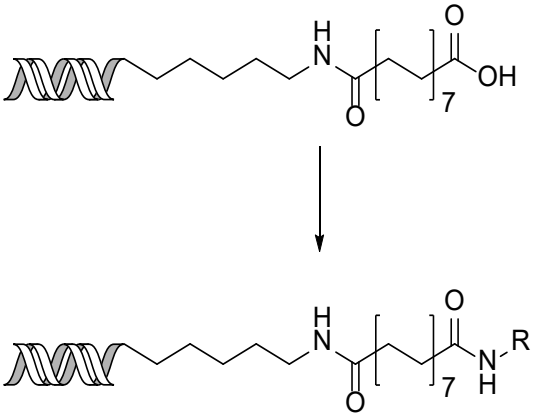
8.5.4 - Final forward amide coupling conditions



An aliquot of acid solution (60 µl, 0.25 M in NMP) was added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was then removed at 55°C in a Genevac for 60 mins. To this solution was added **74** (30 µl, 5 nmol in 3.5% TPGS-750-M in water), 2,6-lutidine (6.92 µl, 0.06 mmol) and HATU (5.7 mg, 0.015 mmol)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 45°C overnight. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter. To purify each sample they were diluted with water (50 µl), DCM (2 x 100 µl) was added to each and the vial vortexed. If an emulsion

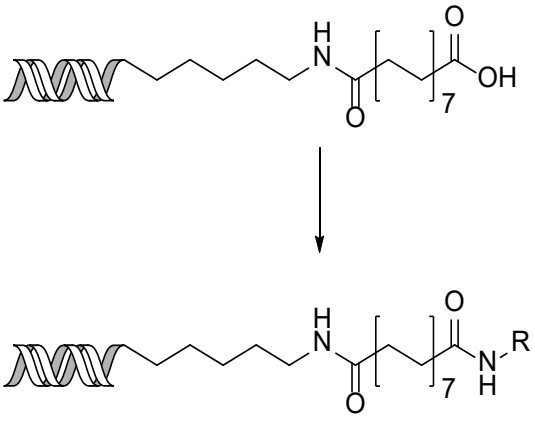
remained, the sample was centrifuged to aid separation. The organics were removed, and aqueous washed with ethyl acetate (2 x 100 µl). Aqueous sodium chloride (8 µl, 4M) and ethanol (264 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. The pellet of DNA was then dissolved in water to give a 1mM solution.

8.5.5 - Final reverse amide coupling conditions



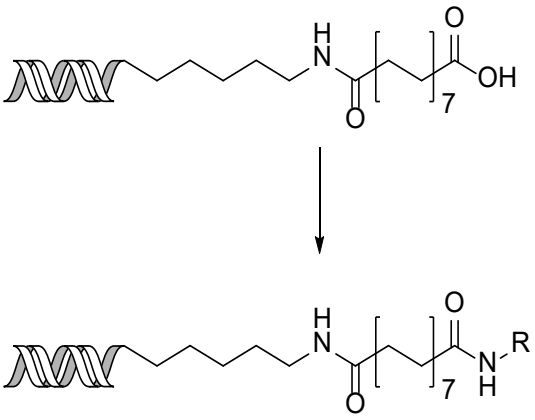
An aliquot of amine solution (60 µl, 0.25 M in NMP) and HOAT (20 µl, 10 mg per 100 µl in NMP) were added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was then removed at 55°C in a Genevac for 60 mins. To this solution was added **71** (30 µl, 5 nmol in 4.5% TPGS-750-M in water), 2,6-lutidine (5.2 µl, 0.045 mmol) and DIC (2.2 µl, 0.015 mmol)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 45°C for 5 hours. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter. To purify each sample they were diluted with water (50 µl), DCM (2 x 100 µl) was added to each and the vial vortexed. If an emulsion remained, the sample was centrifuged to aid separation. The organics were removed, and aqueous washed with ethyl acetate (2 x 100 µl). Aqueous sodium chloride (8 µl, 4M) and ethanol (264 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. The pellet of DNA was then dissolved in water to give a 1mM solution.

8.5.6 - Initial coupling agent screen conditions



To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate was added glycine ethyl ester hydrochloride (2.1 mg, 0.015 mmol), aniline (1.37 µl, 0.015 mmol), 2-aminothiazole (1.5 mg, 0.015 mmol) or 2-Aminoimidazol hemisulphate (2 mg, 0.015 mmol), **71** (30 µl, 5 nmol in 3.5% TPGS-750-M in water), 2,6-lutidine (5.2 µl, 0.045 mmol) or DIPEA (6.2 µl, 0.036 mmol) and HATU (5.7 mg, 0.015 mmol), or EDC.HCl (2.9 mg) or DMT-MM (4.2 mg)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at room temperature or 40 °C overnight. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter.

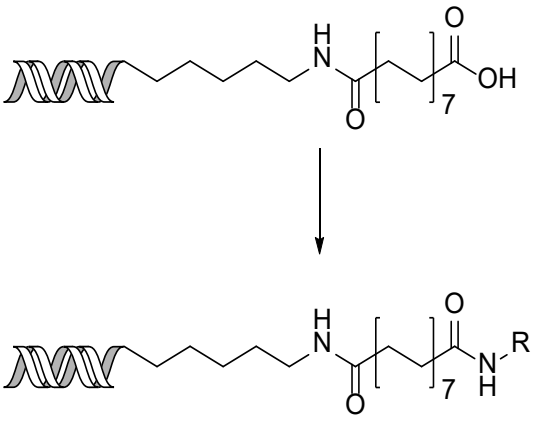
8.5.7 - Base screen conditions



To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate was added glycine ethyl ester hydrochloride (2.1 mg, 0.015 mmol) or 2-Aminoimidazol hemisulphate (2 mg, 0.015 mmol), **71** (30 µl, 5 nmol in 3.5% TPGS-750-M in water), base (0.045 mmol) and HATU (5.7 mg, 0.015 mmol)**.** The vials were vortexed for 30 seconds each

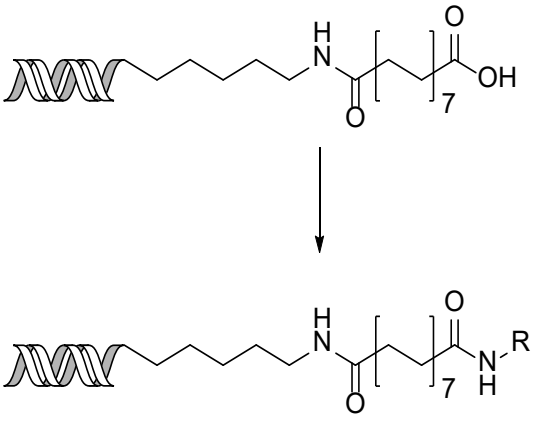
to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at room temperature or 40 °C overnight. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter.

8.5.8 - First FED experimental



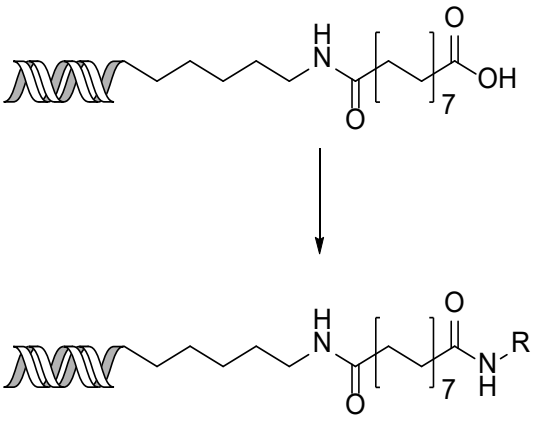
An aliquot of either imidazolamine hemisulphate or glycine ethyl ester hydrochloride solution (60 µl, 0.25 M in NMP) was added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was then removed at 55°C in a Genevac for 60 mins. To this solution was added **71** (1 µl, 5 mM in water), either 2%, 3.5% or 5% TPGS-750-M (29 µl) and HATU (5.7 mg, 0.015 mmol). 2,6-lutidine was added, where 0.5M was used (1.73 µl); where 1M was used (3.46 µl); where 1.5M was used (5.2 µl); where 2.5M was used (8.66 µl). The samples were then vortexed for 30 seconds to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 40 °C, 45 °C, 50 °C, 55 °C or 60 °C overnight. Mass spectrometry was used to analyse the DOE. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter.

8.5.9 - Coupling agent screen conditions



To a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate was added either 2-aminopyridine (1.4 mg, 0.015 mmol) or 2-Aminoimidazol hemisulphate (2 mg, 0.015 mmol), **71** (30 µl, 5 nmol in 3.5% TPGS-750-M in water), 2,6-lutidine (6.92 µl, 0.06 mmol) and COMU (6.4 mg, 0.015 mmol) or BOP (6.6 mg, 0.015 mmol) or PyBOP (7.8 mg, 0.015 mmol) or EDC.HCl (2.9 mg, 0.015 mmol) or DCC (3.1 mg, 0.015 mmol) or DIC (2.2 µl, 0.015 mmol), and where carbodiimides were used HOAT (2 mg, 0.015 mmol)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at room temperature or 40 °C overnight. Mass spectrometry was used to analyse reactions. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter.

8.5.10 - Second FED experimental



An aliquot of either 2-Aminoimidazol hemisulphate or glycine ethyl ester hydrochloride solution (60 µl, 0.25 M in NMP) and HOAT (20 µl, 10 mg per 100 µl in NMP) were added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was

then removed at 55°C in a Genevac for 60 mins. To this solution was added **71** (1 µl, 5 mM in water), either 2%, 3.5% or 5% TPGS-750-M (29 µl) and DIC (2.2 µl, 0.015 mmol). 2,6-lutidine was added, where 0.5M was used (1.73 µl); where 1M was used (3.46 µl); where 1.5M was used (5.2 µl). The samples were then vortexed for 30 seconds to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 40 °C, 50 °C, or 60 °C overnight. Mass spectrometry was used to analyse the DOE. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter.

8.6 – Amide coupling Mass spectrometry results

Experimental Table 6 - Calculated and found masses for each product formed in the amide using capping amines and final reverse amide coupling conditions. Calculated masses from ChemDraw. Analytical procedure used - Agilent advancedbio oligonucleotides column, 2.1x100 mm where the gradient was ran at 0.8 ml/min from 10% MeOH to 50% MeOH over 4 mins against a 50 mM HFIP:8 mM DIPEA buffer solution.

Amine	SM	DP	SP	RT	Expected mass	found mass
Anthanilamide		27	73	4.26	4793.0607	4793.0074
3-Chloro-4-fluoroaniline		45	55	4.682	4802.0065	4801.9504
2-Aminothiazole		11	89	4.234	4757.0066	4756.9513
2-Aminoimidazole hemisulphate	0	58	42	4.01	4740.0454	4739.9886
2-Aminopyridine		0	100	-	4751.0502	-
4-aminopyridine	75		25		4751.0502	
p-Anisidine		95		4.329	4780.0655	4780.0152
4-Chloroaniline		50	50	4.572	4784.016	4783.9683
4-Cyanobenzylamine hydorchloride		100		4.175	4789.0658	4789.0276
Cyclopropanemethylamine hydrochloride		100		4.191	4728.0706	4728.0269
3,4-Difluorobenzylamine		100		4.403	4800.0517	4800.0086
3,5-Dimethoxyaniline		50	50	4.585	4810.0761	4810.0274
4-Dimethylaminopiperidine dihydrochloride		100		4.277	4785.1284	4785.0835
4-Hydroxypiperidine	44	56		4.078	4758.0812	4758.0349
1-Methyl-1H-indole-5-amine	0	61	39	4.443	4403.0815	4802.994
4-(Methylamino)pyridine	80		20		4765.0658	
3-Methylsulphonylaniline hydrochloride			100		4828.0325	
3-Aminobenzamide		71	29	4.064	4793.0607	4793.0159
4-Amino-2-chloropyridine		0	100		4785.0112	
3-amino-5-fluoropyridine		0	100		4769.0408	
6-Aminoindazole		87	13	4.191	4790.0611	4790.0217
3-Amino-1-N-methyl-azetidine dihydrochloride		100		3.953	4743.0815	4743.0423
5-Amino-3-methylisoxazole			100		4755.0451	
4-Aminophenol		100		4.04	4766.0499	4766.0122

3-bromoprop-2-yn-1-amine hydrochloride		100		4.378	4789.9498	4789.9856
prop-2-yn-1-amine		100		3.976	4712.0393	4712.0012
2-aminoacetonitrile hydrochloride		100		3.856	4713.0345	4712.9694
1-aminocyclopropane-1-carbonitrile hydrochloride	10	30	60	3.956	4739.0592	4739.0114
1-(bromoethynyl)cyclopropan-1-amine hydrochloride	7	76	17	4.28	4815.9654	4815.9288
4-bromo-2-methylbut-3-yn-2-amine hydrochloride		21	73	4.991	4817.9811	4817.9465
2-amino-4-bromobut-3-yn-1-ol hydrochloride		95		4.5	4819.9694	4820.0035
2-amino-2-phenylacetonitrile hydrochloride		30	70	4.288	4789.0658	4789.0202
1-phenylprop-2-yn-1-amine hydrochloride		100		4.444	4788.0796	4788.0364
m-anisidine		100		4.394	4780.0655	4780.0396
o-anisidine	1	85	14	4.498	4780.0655	4780.027
aniline		100		4.329	4750.0549	4750.0226
3-Amino-1-methyl-1H-pyrazole	2	84	13	4.069	4754.0611	4754.0287
5-Amino-2-methoxypyridine		95		4.239	4781.0607	4781.0301
N-Benzylmethylamine	13	87		4.726	4778.0882	4778.057
1-Benzylpiperidine	5	95		4.86	4833.1284	4833.0959
(R)-4-Chloro-alpha-methylbenzylamine		100		4.624	4812.0473	4812.0154
3-Chloroaniline	1	84	15	4.604	4784.016	4783.9866
2-Chloroaniline		10	90	4.531	4784.016	4783.9837
2-Chlorobenzylamine		95		4.523	4798.0316	4798.001
4-Chlorobenzylamine		100		4.567	4798.0316	4798.0126
3-Chlorobenzylamine		100		4.518	4798.0316	4798.0122
Cycloheptylamine	65	35		4.665	4770.1175	4770.0898
Cyclohexanemethylamine		100		4.731	4770.1175	4770.0927
Cyclopentylamine	5	95		4.33	4742.0862	4742.0653
1-(Cyclopropylcarbonyl)piperazine		100		4.224	4811.1077	4811.0839
Diethylamine		10	90	4.505	4730.0862	4730.059
N,N-Diethylethylenediamine	68	32		4.213	4773.1284	4773.1027
2,6-Difluoroaniline			100		4786.0361	
3,4-Difluoroaniline		75	25	4.511	4786.0361	4786.0181
cis-2,6-Dimethylmorpholine		95	5	4.42	4772.0968	4772.0704
1-Ethylpiperazine	40	60		4.358	4771.1128	4771.0889
4-Fluoroaniline		100		4.386	4768.0455	4768.0224
Indoline	0	75	0	4.791	4776.0796	4776.0173
morpholine	28	72		4.106	4744.0655	4744.0384
Piperidine		100		4.538	4742.0862	4742.0647
4-(Aminomethyl)pyridine	5	95		4.014	4765.0658	4765.0464
4-Aminobenzotrifluoride			100		4818.0423	
3-Picolylamine		76	24	4.035	4765.0658	4765.0374
Benzylamine		100		4.329	4764.0796	4764.048
azetidine		100		4.188	4714.0549	4714.0206

Experimental Table 7 - Calculated and found masses for each product formed in the amide using amino esters and final reverse amide coupling conditions. Calculated masses from ChemDraw. Analytical procedure used - Agilent advancedbio oligonucleotides column, 2.1x100 mm where the gradient was ran at 0.8 ml/min from 10% MeOH to 50% MeOH over 4 mins against a 50 mM HFIP:8 mM DIPEA buffer solution.

Amino ester	SM	DP	SP	RT	Expected mass	found mass
methyl 1-aminocyclopentane-1-carboxylate hydrochloride		100		4.366	4800.0917	4800.0772
methyl 1-aminocyclopropane-1-carboxylate hydrochloride		100		4.093	4772.0604	4772.0522
methyl 2-aminonicotinate		0	100		4809.0557	
methyl 5-aminonicotinate hydrochloride		35	53	4.368	4809.0557	4809.0421
ethyl 4,5,6,7-tetrahydro-1H-pyrazolo[4,3-c]pyridine-3-carboxylate hydrochloride	100				4852.0979	
methyl 2-aminopyrimidine-5-carboxylate		0	100		4810.0509	
methyl 2-amino-2-methylpropanoate hydrochloride		100		4.207	4774.0761	4774.0583
methyl 6-aminonicotinate		0	100		4809.0557	
methyl 6-aminopicolinate		0	100		4809.0557	
methyl D-tryptophanate hydrochloride		100		4.468	4875.1026	4875.0878
methyl 1-aminocyclobutane-1-carboxylate hydrochloride		100		4.227	4786.0761	4786.0596
methyl asparaginate hydrochloride		100		3.094	4803.0662	4803.0548
methyl (1r,4r)-4-aminocyclohexane-1-carboxylate hydrochloride		100		4.401	4814.1074	4814.094
methyl D-histidinate dihydrochloride		100		3.996	4826.0822	4826.0719
methyl 6-aminopyridazine-3-carboxylate		58	42	4.361	4810.0509	4810.0358
ethyl 2-aminooxazole-4-carboxylate		0	100		4813.0506	
methyl indoline-2-carboxylate hydrochloride		0	100		4834.0761	
methyl tyrosinate hydrochloride		100		4.167	4852.0866	4852.0763
methyl 4-aminotetrahydro-2H-pyran-4-carboxylate hydrochloride		100		4.108	4816.0866	4816.0727
ethyl 5-amino-1,3,4-oxadiazole-2-carboxylate		0	100		4814.0458	
methyl 2-amino-2-(thiophen-2-yl)acetate hydrochloride		100		4.391	4828.0325	4828.0144
methyl D-valinate hydrochloride		100		4.391	4788.0917	4788.0771
methyl morpholine-2-carboxylate hydrochloride		68	30	4.195	4802.071	4802.0554
methyl glutaminate hydrochloride		26	74	3.92	4817.0819	4817.071
ethyl 2-aminothiazole-5-carboxylate		0	100		4829.0277	
methyl 5-hydroxypiperidine-3-carboxylate hydrochloride		100		4.181	4816.0866	4816.0693
methyl 5-amino-1H-1,2,4-triazole-3-carboxylate hydrochloride		0	100		4798.0509	

methyl (2S,4R)-4-hydroxypyrrolidine-2-carboxylate hydrochloride		100		4.049	4802.071	4802.0552
methyl (1R,3S)-3-aminocyclopentane-1-carboxylate hydrochloride		95	5	4.343	4800.0917	4800.0748
ethyl azetidine-3-carboxylate hydrochloride		59	41	4.406	4786.0761	4786.0593
methyl (S)-2-amino-3-cyclohexylpropanoate hydrochloride		100		4.925	4842.1387	4842.1236
methyl L-allothreoninate hydrochloride		100		4.021	4790.071	4790.056
methyl L-threoninate hydrochloride		100		4.029	4790.071	4790.0575
methyl 2-amino-3-methoxy-2-methylpropanoate		80	20	4.242	4804.0866	4804.0665
methyl (R)-2-amino-2-(4-hydroxyphenyl)acetate		100		4.158	4838.071	4838.0622
methyl D-alaninate hydrochloride		100		4.101	4760.0604	4760.05
ethyl 5-imino-2,5-dihydroisoxazole-4-carboxylate		0	100		4813.0506	
methyl (S)-2-aminopentanoate		100		4.41	4788.0917	4788.0778
methyl methioninate hydrochloride		100		4.308	4820.0638	4820.0545
methyl 2-amino-2-(tetrahydro-2H-pyran-4-yl)acetate		100		4.186	4830.1023	4830.096
methyl Nt-methyl-L-histidinate dihydrochloride		100		4.082	4840.0979	4840.0935
ethyl 5-amino-1H-pyrazole-3-carboxylate		87	13	4.345	4812.0713	4812.064
methyl 2-(3-oxopiperazin-2-yl)acetate		100		4.058	4829.0819	4829.0802
methyl (2R,4S)-4-methoxypyrrolidine-2-carboxylate hydrochloride		100		4.299	4816.0866	4816.0862
ethyl 3-amino-3-methylbutanoate hydrochloride		100		4.497	4802.1074	4802.1032
methyl D-leucinate hydrochloride		100		4.565	4802.1074	4802.103
methyl (S)-piperidine-2-carboxylate hydrochloride		100		4.621	4800.0917	4800.0925
ethyl 2-amino-3-(pyridin-4-yl)propanoate dihydrochloride		100		4.27	4851.1026	4851.1038
ethyl (S)-3-amino-3-(pyridin-3-yl)propanoate		100		4.275	4851.1026	4851.1062
methyl D-serinate hydrochloride		100		3.951	4776.0553	4776.0566
methyl azetidine-2-carboxylate hydrochloride		100		4.254	4772.0604	4772.0561
ethyl (R)-piperidine-3-carboxylate		60	40	4.683	4814.1074	4814.1097
ethyl L-prolinate hydrochloride		100		4.524	4800.0917	4800.0953
methyl 2-amino-2-cyclopropylacetate hydrochloride		100		4.254	4786.0761	4786.083
methyl (R)-4,4-difluoropyrrolidine-2-carboxylate		100		4.386	4822.0572	4822.0621
Glycine Ethyl Ester Hydrochloride		100		4.126	4760.0604	4760.0681
Alanine Ethyl Ester Hydrochloride		100		4.25	4774.0761	4774.0845
D-Phenylalanine Methyl Ester Hydrochloride		100		4.712	4850.1074	4850.1184

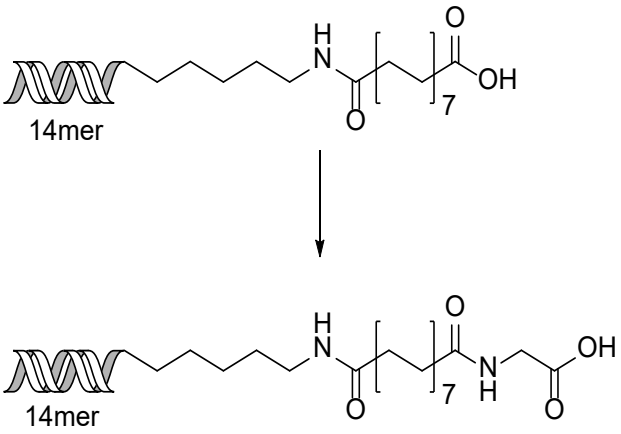
L-Leucine Methyl Ester Hydrochloride		100		4.551	4802.1074	4802.1183
L-Tyrosine methyl ester hydrochloride		100		4.162	4852.0866	4852.0984

Experimental Table 8 - Calculated and found masses for each product formed in the amide using acids and final forward amide coupling conditions. Calculated masses from ChemDraw. Analytical procedure used - Agilent advancedbio oligonucleotides column, 2.1x100 mm where the gradient was ran at 0.8 ml/min from 10% MeOH to 70% MeOH over 5 mins against a 50 mM HFIP:8 mM DIPEA buffer solution.

Acid		SM	DP	SP	RT	Expected mass	found mass
a	(((9H-fluoren-9-yl)methoxy)carbonyl)-L-alanine		100		4.490	4897.0870	7897.0915
b	(((9H-fluoren-9-yl)methoxy)carbonyl)-L-phenylalanine		98	2	4.825	4973.1183	4973.1182
c	(((9H-fluoren-9-yl)methoxy)carbonyl)-aspartic acid α-t-butyl ester	3	94	3	4.886	4997.1394	4997.0964
d	(((9H-fluoren-9-yl)methoxy)carbonyl)-L-tyrosine	7	93		4.451	4989.1132	4989.0749
e	3-(9H-fluoren-9-yl)amino benzoic acid	6	94		4.790	4945.0870	4945.0599
f	4-(9H-fluoren-9-yl)aminomethyl benzoic acid		89	11	4.651	4959.1026	4959.0759
g	3-iodobenzoic acid		97	3	4.016	4833.9046	4833.8541
h	benzoic acid		98	2	3.716	4708.0080	4707.9660
i	4-cyanobenzoic acid		98	2	3.671	4733.0032	3732.9707
j	4-methoxybenzoic acid		97	3	3.779	4738.0185	4737.9839
k	4-acetamidobenzoic acid		100		3.562	4765.0294	4764.9875
l	5-fluoro-2-methylbenzoic acid		100		3.836	4740.0142	4739.9758
m	picolinic acid		100		3.808	4709.0032	4708.9429
n	1H-indole-2-carboxylic acid		100		3.857	4747.0189	4746.9738

8.7 – 14mer Tool compound synthesis

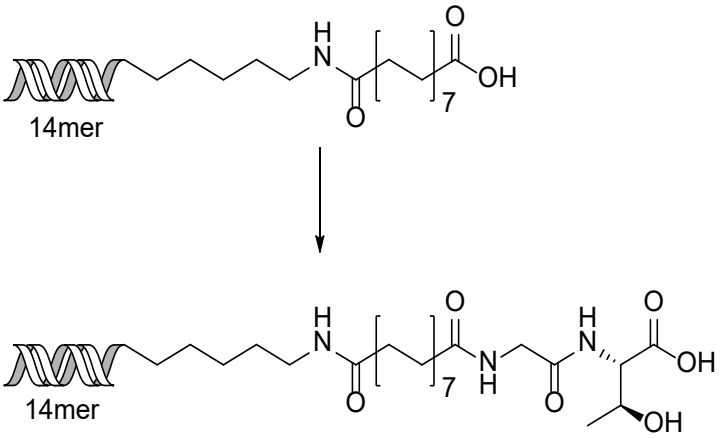
8.7.1 – Synthesis of 14mer DNA conjugate 75



An aliquot of glycine ethyl ester hydrochloride (60 µl, 0.25 M in NMP) and HOAT (20 µl, 10 mg per 100 µl in NMP) were added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was then removed at 55°C in a Genevac for 60 mins. To this solution was added **71** (30 µl, 24 nmol in 4.5% TPGS-750-M in water), 2,6-lutidine (5.2 µl, 0.045 mmol) and DIC (2.2 µl, 0.015 mmol)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 45°C for 5 hours. Mass spectrometry was used to analyse reaction. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter. HRMS (ESI): exact mass calcd : 4760.0604; found : 4760.0629.

The reaction was diluted with water (50 µl), DCM (2 x 100 µl) was added to each and the vial vortexed. If an emulsion remained, the sample was centrifuged to aid separation. The organics were removed, and aqueous washed with ethyl acetate (2 x 100 µl). Aqueous sodium chloride (8 µl, 4M) and ethanol (264 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. The pellet of DNA was then dissolved in 0.25M LiOH (50 µl) and shacked for 1 hour. Aqueous sodium chloride (5 µl, 4M) and ethanol (165 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. The pellet of DNA was then dissolved in water (100 µl) and purified by a 3000 Dalton molecular weight spin filter, washing with water (3 x 100 µl). The product was eluted and mass spectrometry was used to analyse reaction. HRMS (ESI): exact mass calcd : 4732.0291; found : 4732.0543.

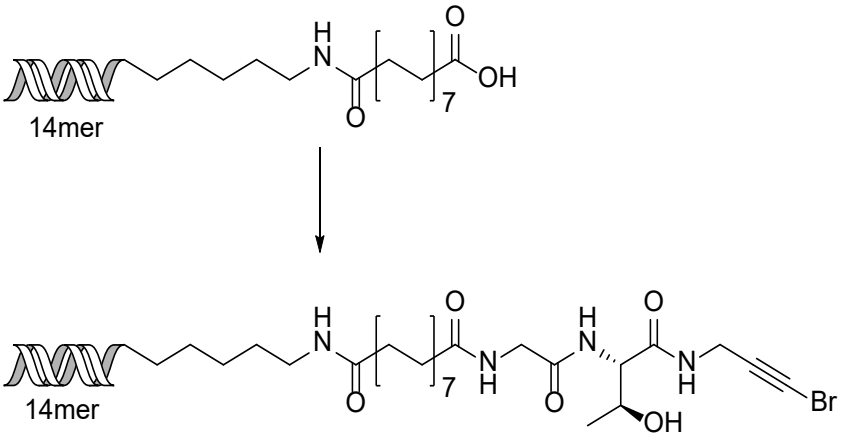
8.7.2 – Synthesis of 14mer DNA conjugate 76



An aliquot of methyl L-threoninate hydrochloride (60 µl, 0.25 M in NMP) and HOAT (20 µl, 10 mg per 100 µl in NMP) were added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was then removed at 55°C in a Genevac for 60 mins. To this solution was added **75** (30 µl, 15 nmol in 4.5% TPGS-750-M in water), 2,6-lutidine (5.2 µl, 0.045 mmol) and DIC (2.2 µl, 0.015 mmol)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 45°C for 5 hours. Mass spectrometry was used to analyse reaction. Samples prepared by adding reaction mixture (1 µl) to water (20 µl) and filtered through a hydrophilic PTFE filter. HRMS (ESI): exact mass calcd : 4847.0924; found : 4847.0302.

The reaction was diluted with water (50 µl), DCM (2 x 100 µl) was added to each and the vial vortexed. If an emulsion remained, the sample was centrifuged to aid separation. The organics were removed, and aqueous washed with ethyl acetate (2 x 100 µl). Aqueous sodium chloride (8 µl, 4M) and ethanol (264 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. The pellet of DNA was then dissolved in 0.25M LiOH (50 µl) and shacked for 1 hour. Aqueous sodium chloride (5 µl, 4M) and ethanol (165 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. The pellet of DNA was then dissolved in water (100 µl) and purified by a 3000 Dalton molecular weight spin filter, washing with water (3 x 100 µl). The product was eluted and mass spectrometry was used to analyse reaction. HRMS (ESI): exact mass calcd : 4833.0768; found : 4833.0264

8.7.3 – Synthesis of 14mer DNA conjugate 77



An aliquot of 3-bromoprop-2-yn-1-amine hydrochloride (60 µl, 0.25 M in NMP) and HOAT (20 µl, 10 mg per 100 µl in NMP) were added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was then removed at 55°C in a Genevac for 60 mins. To this solution was added **76** (30 µl, 4 nmol in 4.5% TPGS-750-M in water), 2,6-lutidine (5.2 µl, 0.045 mmol) and DIC (2.2 µl, 0.015 mmol)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 45°C for 5 hours. The reaction was diluted with water (50 µl), DCM (2 x 100 µl) was added to each and the vial vortexed. If an emulsion remained, the sample was centrifuged to aid separation. The organics were removed, and aqueous washed with ethyl acetate (2 x 100 µl). Aqueous sodium chloride (8 µl, 4M) and ethanol (264 µl) were added and the mixture incubated at -78°C for 1 hour. The mixture was then centrifuged and the ethanol layer removed. The pellet of DNA was then dissolved in water (100 µl) and purified by a 3000 Dalton molecular weight spin filter, washing with water (3 x 100 µl). The product was eluted and mass spectrometry was used to analyse reaction. HRMS (ESI): exact mass calcd : 4948.0189; found : 4947.9343.

8.8 – Library synthesis

8.8.1 - General ethanol precipitation method

To the reaction mixture was added 10% volume NaCl (5 M in water), and 3x volume cold ethanol. The mixture was allowed to sit for 1 hour at -78 °C, or overnight at -20 °C. The mixture was then centrifuged at 13500 rpm for 30 mins. The supernatant decanted, and 70% cold EtOH added. The mixture was allowed to sit at -78 °C for 1 hour, before being centrifuged at

13500 rpm for 30 mins and the supernatant decanted. The pellet was then allowed to air dry before being redisolved in the appropriate solvent.

8.8.2 – General amide coupling procedure

An aliquot of amine or amino ester (60 µl, 0.25 M in NMP) and HOAT (20 µl, 10 mg per 100 µl in NMP) were added to a 50 µl glass insert for a Para-dox™ 96-well micro photoredox/optimisation Plate. The NMP was then removed at 55°C in a Genevac for 60 mins. To this solution was added the corresponding ligated library aliquot (30 µl in 4.5% TPGS-750-M in water), 2,6-lutidine (5.2 µl, 0.045 mmol) and DIC (2.2 µl, 0.015 mmol)**.** The vials were vortexed for 30 seconds each to enhance mixing. The samples were then heated in a Para-dox™ 96-well micro photoredox/optimisation plate at 45°C for 5 hours. The reaction was diluted with water (50 µl) and all reaction aliquots combined. DCM (2 x 4000 µl) was added and the vial vortexed. If an emulsion remained, the sample was centrifuged to aid separation. The organics were removed, and aqueous washed with ethyl acetate (2 x 4000 µl). The DNA was precipitated using the general ethanol precipitation method. The pellet of DNA was then dissolved in water (400 µl) and purified by a 3000 Dalton molecular weight spin filter, washing with water (3 x 100 µl). The product was eluted and the overall library concentration analysed.

8.8.3 – General hydrolysis procedure

The library after the amide coupling step was dissolved to a concentration of 0.25 M LiOH (1 ml) and shaken for 1 hour. The DNA was precipitated using the general ethanol precipitation method. The pellet of DNA was then dissolved in water (400 µl) and purified by a 3000 Dalton molecular weight spin filter, washing with water (3 x 100 µl). The product was eluted and the overall library concentration analysed.

8.8.4 – General ligation protocol

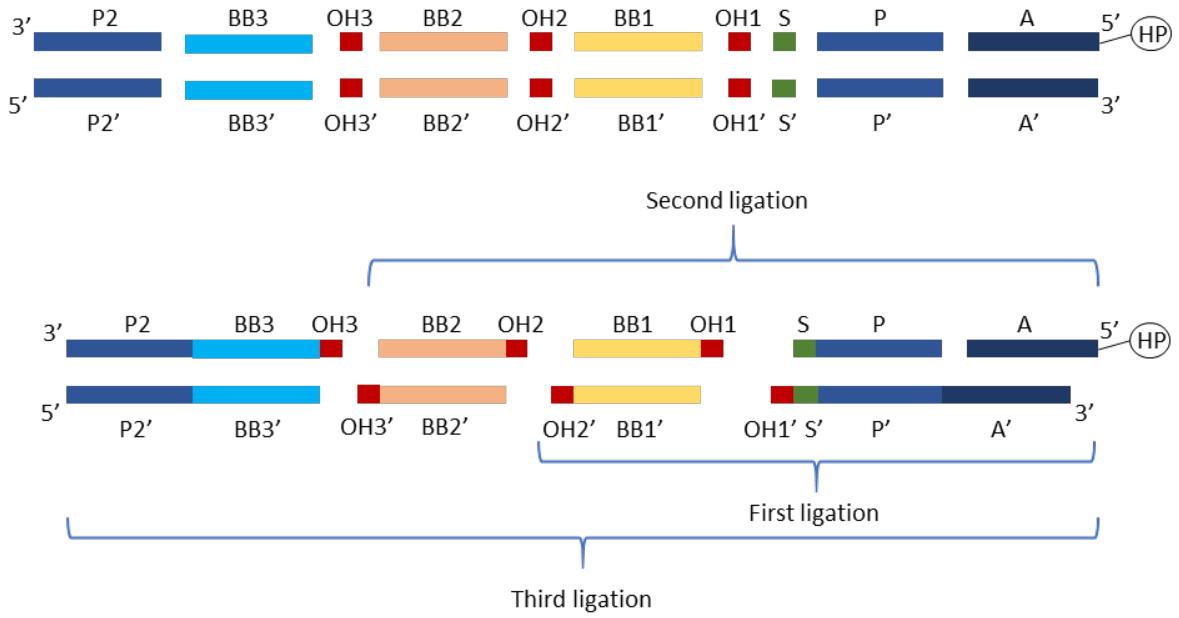


Figure 118 – Encoding strategy used for the 99,405 member library.

Experimental Table 9 - Codes, functions, and corresponding sequence for each DNA section

Code	Function	sequence 5'-3'
A	Adapter – 5’ aminolinked head piece	GTCTTGCCGAATTC
A'	Complimentary adapter	GAATTCGGCAAGAC
P	Primer	AGGTCGGTGTGAACGGATTTG
P'	Complementary primer	CAAATCCGTTCACACCGACCT
S	Scaffold code	GCT
S'	Complementary S	AGC
OH1	Ligation overhang 1	GTAT
OH1'	Complementary OH1	ATAC
BB1	Building block 1	xxxxxxxx
BB1'	Complementary BB1	xxxxxxxx
OH2	Ligation overhang 2	CCTA
OH2'	Complementary OH2	TAGG
BB2	Building block 2	xxxxxxxx
BB2'	Complementary BB2	xxxxxxxx
OH3	Ligation overhang 3	TACG
OH3'	Complementary OH3	CGTA
BB3	Building block 3	xxxxxxxx
BB3'	Complementary BB3	xxxxxxxx
P2	Complementary to P2'	TGACCTCAACTACATGGTCTACA
P2'	Primer (reverse)	TGTAGACCATGTAGTTGAGGTCA

Phosphorylation

Prior to ligation, the 5' terminus of each strand was phosphorylated. DNA strands (450 µM, 9000 pmol in overall reaction media of 20 µl) were added PNK reaction buffer (2 µl, 500mM Tris-HCl [pH 7.6 at 25°C], 100mM $MgCl_2$, 50 mM DTT, 1mM spermidine), ATP (2 µl, 10 mM, Thermo Scientific), T4 Polynucleotide Kinase (1 µl, 10U/µl, Thermo Scientific) and nuclease free water (up to 20 µl). The reaction was carried out at 37 °C for 1 hour, followed by heating to 75 °C for 10 mins. DNA was used in the ligation steps without purification or precipitation.

Ligation

Ligations contained DNA (100 µM, 9000 pmol in overall reaction media of 90 µl), phosphorylated DNA strands (20 µl, 9000 pmol), 10X T4 DNA ligase buffer (9 µl, 400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP9), water (up to 90 µl) and T4 DNA Ligase (3 µl, 30 Weiss U/µL). The ligations were carried out at 25°C for 16 hours, followed by heating to 75°C for 10 mins. Each ligation was purified by ethanol precipitation prior to the subsequent organic reaction taking place.

8.8.5 – Library Amounts

Experimental Table 10 – Nucleic acid concentrations and resulting library amount after each reaction stage. Analysis calculated using a NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, pipetting 1 µl of sample on the sample loading plate.

Sample ID	Nucleic Acid Conc. (ng / µl)	mw	mM	ml	nmol
Library 1st step – pre hydrolysis	23596	32642	0.723	0.45	325.3
1x1x1 1st step – pre hydrolysis	27819	32642	0.085	0.1	8.5
Library 1st step – post hydrolysis	25368	32620	0.778	0.4	311.1
1x1x1 1st step – post hydrolysis	27652	32620	0.848	0.01	8.5
Library 2nd step – pre hydrolysis	11778	39929	0.295	0.63	185.8
1x1x1 2nd step – pre hydrolysis	2042	39929	0.051	0.11	5.6
Library 2nd step – post hydrolysis	30202	39904	0.756	0.2	151.3
1x1x1 2nd step – post hydrolysis	16974	39904	0.425	0.01	4.3
Library 3rd step – Pre prep HPLC	31265	60599	0.516	0.2	103.2
Library 3rd step – Post prep HPLC	15923	60599	0.263	0.3	78.8
1x1x1 3rd step	16546	60599	0.273	0.01	2.7

8.8.6 - PCR and NGS

Forward (JHH-FSP-P) and reverse (JHH-FSP-P2) primers were designed to amplify the DEL library, flanked by 5' Illumina adapter sequences to enable downstream sequence analysis. Each PCR was performed in a 50µl reaction mixture containing AmpliTaq Gold® 360 Master Mix (Thermo Fisher) and 100 ng of 1x1x1 prototype library (at 4.2 µM). The final concentration of each primer used was 10 mM. PCRs were carried out using a Bio-Rad MJ Mini Personal Thermal Cycler. The thermal cycling conditions were as follows: 10 min at 95 °C, followed by 40 cycles of 30 s at 95 °C, 30 s at 55°C, and 1 min at 72°C, with a final extension time of 420 s at 72°C. A negative control (distilled water in place of primers) was included in each run. Following the PCR reactions the samples were run on a 4% Agarose E-gel, the gel extracted and sequenced using sanger (both ~53mer and 19mer primers) or using Next Generation Sequencing (GeneWiz, South Plainfield, NJ, USA).

Experimental Table 11 - Forward and reverse primers for PCR and NGS. Primer sequence (in blue) and NGS elongation sequence (in purple).

Primer	sequence 5'-3'	length
Forward primer	ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGTAGAC CATGTAGTTGAGGTCA	56
Reverse primer	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGTCGG TGTGAACGGATTTG	53
Forward short primer	ACACTCTTTCCCTACACGA	19
Reverse short primer	GACTGGAGTTCAGACGTGT	19

8.9 - Crystal data

8.9.1 - Crystal data for N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-7-phenylimidazo[1,2-a]pyrimidine-5-carboxamide

Empirical formula	C19H14N4O
Formula weight	314.34
Temperature/K	150.0(2)
Crystal system	monoclinic
Space group	P21/n
a/Å	7.5837(3)
b/Å	18.3526(6)
c/Å	23.4427(8)
α/°	90
β/°	99.932(3)
γ/°	90
Volume/Å^3	3213.9(2)
Z	8
ρcalcg/cm^3	1.299
μ/mm^{-1}	0.675
F(000)	1312.0
Crystal size/mm^3	0.22 × 0.14 × 0.06
Radiation	CuKα (λ = 1.54184)
2Θ range for data collection/°	7.658 to 133.64
Index ranges	-8 ≤ h ≤ 9, -21 ≤ k ≤ 21, -27 ≤ l ≤ 15
Reflections collected	22508
Independent reflections	5649 [Rint = 0.0279, Rsigma = 0.0217]
Data/restraints/parameters	5649/0/439
Goodness-of-fit on F2	1.039
Final R indexes [I>=2σ (I)]	R1 = 0.0325, wR2 = 0.0789
Final R indexes [all data]	R1 = 0.0384, wR2 = 0.0829
Largest diff. peak/hole	e Å^{-3} 0.14/-0.20

Experimental

Crystal structure data for mjw170004_off was collected on a Xcalibur, Atlas, Gemini ultra diffractometer equipped with an fine-focus sealed X-ray tube (λ CuKα = 1.54184 Å) and an Oxford Cryosystems CryostreamPlus open-flow N2 cooling device. Cell refinement, data collection and data reduction were undertaken via software CrysAlisPro 1.171.38.42b (Rigaku OD, 2015). Intensities were corrected for absorption using CrysAlisPro 1.171.38.42b (Rigaku Oxford Diffraction, 2015) Empirical absorption correction using spherical harmonics, implemented in SCALE3 ABSPACK scaling algorithm.

8.9.2 - Crystal data for N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-5-phenylpyrazolo[1,5-a]pyrimidine-7-carboxamide

Empirical formula	C19H14N4O
Formula weight	314.34
Temperature/K	150.0(2)
Crystal system	monoclinic
Space group	P21/n
a/Å	12.6275(4)
b/Å	8.2824(3)
c/Å	15.5885(5)
α/°	90
β/°	110.885(4)
γ/°	90
Volume/Å^3	1523.22(9)
Z	4
ρcalcg/cm^3	1.371
μ/mm^{-1}	0.713
F(000)	656.0
Crystal size/mm^3	0.39 × 0.11 × 0.02
Radiation	CuKα (λ = 1.54184)
2Θ range for data collection/°	7.782 to 133.532
Index ranges	-15 ≤ h ≤ 14, -9 ≤ k ≤ 7, -18 ≤ l ≤ 18
Reflections collected	10969
Independent reflections	2680 [Rint = 0.0460, Rsigma = 0.0336]
Data/restraints/parameters	2680/0/218
Goodness-of-fit on F2	1.083
Final R indexes [I>=2σ (I)]	R1 = 0.0583, wR2 = 0.1507
Final R indexes [all data]	R1 = 0.0718, wR2 = 0.1598
Largest diff. peak/hole	e Å^{-3} 0.77/-0.22

Experimental

Crystal structure data for mjw170006_fa was collected on a Xcalibur, Atlas, Gemini ultra diffractometer equipped with an fine-focus sealed X-ray tube (λ CuKα = 1.54184 Å) and an Oxford Cryosystems CryostreamPlus open-flow N2 cooling device. Cell refinement, data collection and data reduction were undertaken via software CrysAlisPro 1.171.38.42b (Rigaku OD, 2015). Intensities were corrected for absorption using CrysAlisPro 1.171.38.42b (Rigaku Oxford Diffraction, 2015) Analytical numeric absorption correction using a multifaceted crystal model based on expressions derived by R.C. Clark & J.S. Reid. (Clark, R. C. & Reid, J. S. (1995). Acta Cryst. A51, 887-897) Empirical absorption correction using spherical harmonics, implemented in SCALE3 ABSPACK scaling algorithm.

8.9.3 - Crystal data for (E)-4-((2-(2,4-dinitrophenyl)hydrazono)(phenyl)methyl)-N-(4-(2-(2-methoxyethoxy)ethoxy)phenyl)-1-(4-methylbenzyl)-1H-1,2,3-triazole-5-carboxamide

Empirical formula	C35H34N8O8
Formula weight	694.70
Temperature/K	150.0(2)
Crystal system	triclinic
Space group	P-1
a/Å	7.7440(2)
b/Å	12.0569(4)
c/Å	18.7331(7)
α/°	97.719(3)
β/°	96.262(3)
γ/°	95.625(3)
Volume/Å^3	1711.58(10)
Z	2
ρcalcg/cm^3	1.348
μ/mm^{-1}	0.816
F(000)	728.0
Crystal size/mm^3	0.47 × 0.08 × 0.03
Radiation	CuKα (λ = 1.54184)
2Θ range for data collection/°	7.448 to 133.898
Index ranges	-7 ≤ h ≤ 9, -14 ≤ k ≤ 14, -22 ≤ l ≤ 22
Reflections collected	23897
Independent reflections	5966 [Rint = 0.0543, Rsigma = 0.0422]
Data/restraints/parameters	5966/1394/719
Goodness-of-fit on F2	1.036
Final R indexes [I>=2σ (I)]	R1 = 0.0599, wR2 = 0.1631
Final R indexes [all data]	R1 = 0.0797, wR2 = 0.1852
Largest diff. peak/hole	e Å^{-3} 0.53/-0.29

Experimental

Crystal structure data for mjw170005_fa was collected on a Xcalibur, Atlas, Gemini ultra diffractometer equipped with an fine-focus sealed X-ray tube (λ CuKα = 1.54184 Å) and an Oxford Cryosystems CryostreamPlus open-flow N2 cooling device. Cell refinement, data collection and data reduction were undertaken via software CrysAlisPro 1.171.38.42b (Rigaku OD, 2015). Intensities were corrected for absorption using CrysAlisPro 1.171.38.42b (Rigaku Oxford Diffraction, 2015) Analytical numeric absorption correction using a multifaceted crystal model based on expressions derived by R.C. Clark & J.S. Reid. (Clark, R. C. & Reid, J. S. (1995). Acta Cryst. A51, 887-897) Empirical absorption correction using spherical harmonics, implemented in SCALE3 ABSPACK scaling algorithm.

9.0 - References

1. J. P. Hughes, S. Rees, S. B. Kalindjian, K. L. Philpott, Principles of early drug discovery. *Br. J. Pharmacol.*, 2011, **162** (6), 1239-1249.
2. C. A. Lipinski , F. Lombardo, B. W. Dominy, P. J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv Rev.* 2001, **46**, 3-26.
3. S. Brenner, R. Lerner, Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA,* 1992, **89**, 5381-5383.
4. J. Nielsen, S. Brenner, K. D. Janda, Synthetic Methods for the Implementation of Encoded Combinatorial Chemistry, *J. Am. Chem. Soc.,* 1993, **115**, 9812-9813.
5. Y. Kinoshita, K. Nishigaki, Enzymatic synthesis of code regions for encoded combinatorial chemistry (ECC). *Nucleic Acids Symp Ser*., 1995, **34**, 201-202.
6. Z. J. Gartner, D. R. Liu, The Generality of DNA-Templated Synthesis as a Basis for Evolving Non-Natural Small Molecules. *J. Am. Chem. Soc.,* 2001, **123** (28), 6961-6963.
7. Z. J. Gartner, M. W. Kanan, D. R. Liu, Expanding the Reaction Scope of DNA-Templated Synthesis. *Angew. Chem. Int. Ed.,* 2002, **41** (10), 1796.
8. S. Melkko, J. Scheuermann, C. E. Dumelin, D. Neri, Encoded self-assembling chemical libraries. *Nat Biotechnol.,* 2004, **22** (5), 568-74.
9. D. R. Halpin, P. B. Harbury, DNA display I. Sequence-encoded routing of DNA populations. *PLoS Biol.,* 2004, **2** (7), 1015-1021.
10. D. R. Halpin, P. B. Harbury, DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution*. PLoS Biol.,* 2004, **2** (7), 1022-1030.
11. D. R. Halpin, J. A. Lee, S. J. Wrenn, P. B. Harbury, DNA display III. Solid-phase organic synthesis on unprotected DNA. *PLoS Biol.,* 2004, **2** (7), 1031-1038.
12. F. Buller, L. Mannocci, Y. Zhang, C. E. Dumelin, J. Scheuermann, D. Neri, Design and synthesis of a novel DNA-encoded chemical library using Diels-Alder cycloadditions*. Bioorg Med Chem Lett.,* 2008, **18** (22), 5926-31.
13. B. Morgan, Methods for synthesis of encoded libraries. US 2011/0136697 A1, 2011.
14. T. Franch, Enzymatic encoding methods for efficient synthesis of large libraries. US 2009/0264300 A1, 2009.
15. L. Mannocci, Y. Zhang, J. Scheuermann, M. Leimbacher, G. De Bellis, E. Rizzi, C. Dumelin, S. Melkko, D. Neri, High-throughput sequencing allows the identification of binding molecules isolated from DNA-encoded chemical libraries. *Proc. Natl. Acad. Sci. USA,* 2008, **105** (46), 17670-5.
16. A. Litovchick, C. E. Dumelin, S. Habeshian, D. Gikunju, M. A. Guie, P. Centrella, Y. Zhang, E. A. Sigel, J. W. Cuozzo, A. D. Keefe, M. A. Clark, Encoded Library Synthesis Using Chemical Ligation and the Discovery of sEH Inhibitors from a 334-Million Member Library. *Scientific reports,* 2015, **5**, 10916.
17. M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuozzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Gefter, S. P. Hale, N. J. Hansen, D. I. Israel, J. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O'Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L.

Tang, K. van Vloten, R. W. Wagner, G. Yao, B. Zhao, B. A. Morgan, Design, synthesis and selection of DNA-encoded small-molecule libraries. *Nature chemical biology,* 2009, **5** (9), 647-54.
18. H. Deng, J. Zhou, F. S. Sundersingh, J. Summerfield, D. Somers, J. A. Messer, A. L. Satz, N. Ancellin, C. C. Arico-Muendel, K. L. Sargent Bedard, A. Beljean, S. L. Belyanskaya, R. Bingham, S. E. Smith, E. Boursier, P. Carter, P. A. Centrella, M. A. Clark, C. W. Chung, C. P. Davie, J. L. Delorey, Y. Ding, G. J. Franklin, L. C. Grady, K. Herry, C. Hobbs, C. S. Kollmann, B. A. Morgan, L. J. Pothier Kaushansky, Q. Zhou, Discovery, SAR, and X-ray Binding Mode Study of BCATm Inhibitors from a Novel DNA-Encoded Library. *ACS Med. Chem. Lett.* 2015, **6** (8), 919-24.
19. M. L. Malone, B. M. Paegel, What is a "DNA-Compatible" Reaction? *ACS combinatorial science,* 2016, **18** (4), 182-7.
20. T. Kalliokoski, Price-Focused Analysis of Commercially Available Building Blocks for Combinatorial Library Synthesis. *ACS Comb. Sci,* 2015, **17** (10), 600-7.
21. R. A. Goodnow, C. E. Dumelin, A. D. Keefe, DNA-encoded chemistry: enabling the deeper sampling of chemical space. *Nat Rev Drug Discov.,* 2017, **16** (2), 131-147.
22. R. Franzini, F. Samain, M. Abd Elrahman, G. Mikutis, A. Nauer, M. Zimmermann, J. Scheuermann, J. Hall, D. Neri, Systematic evaluation and optimization of modification reactions of oligonucleotides with amines and carboxylic acids for the synthesis of DNA-encoded chemical libraries. *Bioconjug Chem.* 2014, **25** (8), 1453-61.
23. R. A. Goodnow, A Handbook for DNA-encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery. *John Wiley & Sons: Hoboken*, New Jersey, 2014.
24. Y. He, D. R. Liu, A sequential strand-displacement strategy enables efficient six-step DNA-templated synthesis. *J. Am. Chem. Soc.* 2011, **133** (26), 9972-5.
25. A. H. Gouliaev, T. Franch, M. A. Godskesen, K. B. Jensen. Bi-functional complexes and methods for making and using such complexes. WO 2011/127933, 2011
26. Y. Chen, A. S. Kamlet, J. B. Steinman, D. R. Liu, A biomolecule-compatible visible-light-induced azide reduction from a DNA-encoded reaction-discovery system. *Nat Chem.* 2011, **3** (2), 146-53.
27. F. Buller, M. Steiner, K. Frey, D. Mircsof, J. Scheuermann, M. Kalisch, P. Buhlmann, C. T. Supuran, D. Neri, Selection of Carbonic Anhydrase IX Inhibitors from One Million DNA-Encoded Compounds. *ACS Chem Biol.* 2011, **6** (4), 336-44.
28. H. Deng, H. O'Keefe, C. P. Davie, K. E. Lind, R. A. Acharya, G. J. Franklin, J. Larkin, R. Matico, M. Neeb, M. M. Thompson, T. Lohr, J. W. Gross, P. A. Centrella, G. K. O'Donovan, K. L. Bedard, K. van Vloten, S. Mataruse, S. R. Skinner, S. L. Belyanskaya, T. Y. Carpenter, T. W. Shearer, M. A. Clark, J. W. Cuozzo, C. C. Arico-Muendel, B. A. Morgan, Discovery of highly potent and selective small molecule ADAMTS-5 inhibitors that inhibit human cartilage degradation via encoded library technology (ELT). *J. Med Chem.* 2012, **55** (16), 7061-79.
29. D. G. Brown, J. Bostrom, Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med Chem*. 2016, **59** (10), 4443-58.
30. Z. Zhu, A. Shaginian, L. C. Grady, T. O'Keeffe, X. E. Shi, C. P. Davie, G. L. Simpson, J. A. Messer, G. Evindar, R. N. Bream, P. P. Thansandote, N. R. Prentice, A. M. Mason, S. Pal,

Design and Application of a DNA-Encoded Macrocyclic Peptide Library. *ACS Chem Biol.* 2018, **13** (1), 53-59.
31. Y. Ding, G. J. Franklin, J. L. DeLorey, P. A. Centrella, S. Mataruse, M. A. Clark, S. R. Skinner, S. Belyanskaya, Design and Synthesis of Biaryl DNA-Encoded Libraries*. ACS Comb. Sci*. 2016, **18** (10), 625-629.
32. Y. Li, E. Gabriele, F. Samain, N. Favalli, F. Sladojevich, J. Scheuermann, D. Neri, Optimized Reaction Conditions for Amide Bond Formation in DNA-Encoded Combinatorial Libraries. *ACS Comb. Sci*. 2016, **18** (8), 438-43.
33. H. Deng, J. Zhou, F. Sundersingh, J. A. Messer, D. O. Somers, M. Ajakane, C. C. Arico-Muendel, A. Beljean, S. L. Belyanskaya, R. Bingham, E. Blazensky, A. B. Boullay, E. Boursier, J. Chai, P. Carter, C. W. Chung, A. Daugan, Y. Ding, K. Herry, C. Hobbs, E. Humphries, C. Kollmann, V. L. Nguyen, E. Nicodeme, S. E. Smith, N. Dodic, N. Ancellin, Discovery and Optimization of Potent, Selective, and in Vivo Efficacious 2-Aryl Benzimidazole BCATm Inhibitors. *ACS Med Chem Lett.* 2016, **7** (4), 379-84.
34. Y. Ding, J. Chai, P. A. Centrella, C. Gondo, J. L. DeLorey, M. A. Clark, Development and Synthesis of DNA-Encoded Benzimidazole Library. *ACS Comb. Sci.* 2018, **20** (5), 251-255.
35. A. L. Satz, J. Cai, Y. Chen, R. Goodnow, F. Gruber, A. Kowalczyk, A. Petersen, G. Naderi-Oboodi, L. Orzechowski, Q. Strebel, DNA Compatible Multistep Synthesis and Applications to DNA Encoded Libraries. *Bioconjug Chem.* 2015, **26** (8), 1623-32.
36. H. C. Du, H. Huang, DNA-Compatible Nitro Reduction and Synthesis of Benzimidazoles. *Bioconjug Chem.* 2017, **28** (10), 2575-2580.
37. H. C. Du, N. Simmons, J. C. Faver, Z. Yu, M. Palaniappan, K. Riehle, M. M. Matzuk, A Mild, DNA-Compatible Nitro Reduction Using $B_2(OH)_4$. *Org Lett.* 2019, **21** (7), 2194-2199.
38. H. Li, Z. Sun, W. Wu, X. Wang, M. Zhang, X. Lu, W. Zhong, D. Dai, Inverse-Electron-Demand Diels-Alder Reactions for the Synthesis of Pyridazines on DNA. *Org Lett.* 2018, **20** (22), 7186-7191.
39. C. J. Gerry, Z. Yang, M. Stasi, S. L. Schreiber, DNA-Compatible [3 + 2] Nitrone-Olefin Cycloaddition Suitable for DEL Syntheses. *Org Lett.* 2019, **21** (5), 1325-1330.
40. Z. J. Gartner, M. W. Kanan, D. R. Liu, Expanding the Reaction Scope of DNA-Templated Synthesis. *Angew. Chem. Int. Ed.* 2002, **41**, 1796-1800.
41. M. K. Škopić, O. Bugain, K. Jung, S. Onstein, S. Brandherm, T. Kalliokoski, A. Brunschweiger, Design and synthesis of DNA-encoded libraries based on a benzodiazepine and a pyrazolopyrimidine scaffold. *Med. Chem. Commun.* 2016, **7** (10), 1957-1965.
42. C. Cao, P. Zhao, Z. Li, Z. Chen, Y. Huang, Y. Bai, X. Li, A DNA-templated synthesis of encoded small molecules by DNA self-assembly. *Chem. Commun*. 2014, **50** (75), 10997-9.
43. Y. Ding, M. A. Clark, Robust Suzuki-Miyaura cross-coupling on DNA-linked substrates. *ACS Comb. Sci*. 2015, **17** (1), 1-4.
44. Y. Ding, J. L. DeLorey, M. A. Clark, Novel Catalyst System for Suzuki-Miyaura Coupling of Challenging DNA-Linked Aryl Chlorides. *Bioconjug Chem*. 2016, **27** (11), 2597-2600.
45. J. Y. Li, H. Huang, Development of DNA-Compatible Suzuki-Miyaura Reaction in Aqueous Media. *Bioconjug Chem*. 2018, **29** (11), 3841-3846.
46. N. Favalli, G. Bassi, T. Zanetti, J. Scheuermann, D. Neri, Screening of Three Transition Metal-Mediated Reactions Compatible with DNA-Encoded Chemical Libraries. *Helv. Chim. Acta* 2019, **102** (4).

47. H. Xu, F. Ma, N. Wang, W. Hou, H. Xiong, F. Lu, J. Li, S. Wang, P. Ma, G. Yang, R. A. Lerner, DNA-Encoded Libraries: Aryl Fluorosulfonates as Versatile Electrophiles Enabling Facile On-DNA Suzuki, Sonogashira, and Buchwald Reactions. *Adv. Sci.* 2019, **6** (23), 1901551.
48. J. Dong, L. Krasnova, M. G. Finn, K. B. Sharpless, Sulfur(VI) fluoride exchange (SuFEx): another good reaction for click chemistry. *Angew. Chem. Int. Ed.* 2014, **53** (36), 9430-48.
49. X. Wang, H. Sun, J. Liu, W. Zhong, M. Zhang, H. Zhou, D. Dai, X. Lu, Palladium-Promoted DNA-Compatible Heck Reaction. *Org Lett*. 2019, **21** (3), 719-723.
50. X. Lu, S. E. Roberts, G. J. Franklin, C. P. Davie, On-DNA Pd and Cu promoted C-N cross-coupling reactions. *Med. Chem. Comm*. 2017, **8** (8), 1614-1617.
51. Y. Ruff, F. Berst, Efficient copper-catalyzed amination of DNA-conjugated aryl iodides under mild aqueous conditions. *Med. Chem. Comm.* 2018, 9 (7), 1188-1193.
52. E. de Pedro Beato, J. Priego, A. Gironda-Martinez, F. Gonzalez, J. Benavides, J. Blas, M. D. Martin-Ortega, M. A. Toledo, J. Ezquerra, A. Torrado, Mild and Efficient Palladium-Mediated C-N Cross-Coupling Reaction between DNA-Conjugated Aryl Bromides and Aromatic Amines. *ACS Comb. Sci.* 2019, **21** (2), 69-74.
53. S. A. Cochrane, Z. Huang, J. C. Vederas, Investigation of the ring-closing metathesis of peptides in water. *Org Biomol Chem.* 2013, **11** (4), 630-9.
54. X. Lu, L. Fan, C. B. Phelps, C. P. Davie, C. P. Donahue, Ruthenium Promoted On-DNA Ring-Closing Metathesis and Cross-Metathesis. *Bioconjug. Chem.* 2017, **28** (6), 1625-1629.
55. X. Wang, H. Sun, J. Liu, D. Dai, M. Zhang, H. Zhou, W. Zhong, X. Lu, Ruthenium-Promoted C-H Activation Reactions between DNA-Conjugated Acrylamide and Aromatic Acids. *Org Lett*. 2018, **20** (16), 4764-4768.
56. L. Fan, C. P. Davie, Zirconium(IV)-Catalyzed Ring Opening of on-DNA Epoxides in Water. *Chembiochem*. 2017, **18** (9), 843-847.
57. J. Wang, H. Lundberg, S. Asai, P. Martin-Acosta, J. S. Chen, S. Brown, W. Farrell, R. G. Dushin, C. J. O'Donnell, A. S. Ratnayake, P. Richardson, Z. Liu, T. Qin, D. G. Blackmond, P. S. Baran, Kinetically guided radical-based synthesis of C(sp(3))-C(sp(3)) linkages on DNA. *Proc. Natl. Acad. Sci. USA.* 2018, **115** (28), 6404-6410.
58. D. K. Kolmel, R. P. Loach, T. Knauber, M. E. Flanagan, Employing Photoredox Catalysis for DNA-Encoded Chemistry: Decarboxylative Alkylation of alpha-Amino Acids. *ChemMedChem.* 2018, **13** (20), 2159-2165.
59. J. P. Phelan, S. B. Lang, J. Sim, S. Berritt, A. J. Peat, K. Billings, L. Fan, G. A. Molander, Open-Air Alkylation Reactions in Photoredox-Catalyzed DNA-Encoded Library Synthesis. *J. Am. Chem. Soc.* 2019, **141** (8), 3723-3732.
60. C. Leriche, X. He, C.T. Chang, H. Liu, Reversal of the Apparent Regiospecificity of NAD(P)H-Dependent Hydride Transfer: The Properties of the Difluoromethylene Group, A Carbonyl Mimic. *J. Am. Chem. Soc.* 2003, **125**, 6348-6349.
61. S. O. Badir, J. Sim, K. Billings, A. Csakai, X. Zhang, W. Dong, G. A. Molander, Multifunctional Building Blocks Compatible with Photoredox-Mediated Alkylation for DNA-Encoded Library Synthesis. *Org Lett.* 2020, **22** (3), 1046-1051.
62. D. K. Kolmel, J. Meng, M. H. Tsai, J. Que, R. P. Loach, T. Knauber, J. Wan, M. E. Flanagan, On-DNA Decarboxylative Arylation: Merging Photoredox with Nickel Catalysis in Water. *ACS Comb. Sci.* 2019, **21** (8), 588-597.

63. M. K. Skopic, O. Bugain, K. Jung, A. Gohla, L. J. Doetsch, D. dos Santos, A. Bhat, B. Wagner, A. Brunschweiger, Acid- and Au(I)-mediated synthesis of hexathymidine-DNA-heterocycle chimeras, an efficient entry to DNA-encoded libraries inspired by drug structures. *Chem. Sci.* 2017, **8**, 3356-3361.
64. M. Potowski, V. B. K. Kunig, F. Losch, A. Brunschweiger, Synthesis of DNA-coupled isoquinolones and pyrrolidines by solid phase ytterbium- and silver-mediated imine chemistry. *Med. Chem. Comm.* 2019, **10** (7), 1082-1093.
65. M. Potowski, F. Losch, E. Wunnemann, J. K. Dahmen, S. Chines, A. Brunschweiger, Screening of metal ions and organocatalysts on solid support-coupled DNA oligonucleotides guides design of DNA-encoded reactions. *Chem. Sci.* 2019, **10** (45), 10481-10492.
66. V. B. K. Kunig, C. Ehrt, A. Domling, A. Brunschweiger, Isocyanide Multicomponent Reactions on Solid-Phase-Coupled DNA Oligonucleotides for Encoded Library Synthesis. *Org. Lett.* 2019, **21** (18), 7238-7243.
67. P. A. Cistrone, P. E. Dawson, Click-Based Libraries of SFTI-1 Peptides: New Methods Using Reversed-Phase Silica. *ACS Comb. Sci*. 2016, **18** (3), 139-43.
68. D. R. Halpin, P. B. Harbury, DNA display I. Sequence-encoded routing of DNA populations. *PLoS Biol*. 2004, **2** (7), E173.
69. D. R. Halpin, P. B. Harbury, DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biol*. 2004, **2** (7), E174.
70. D. T. Flood, S. Asai, X. Zhang, J. Wang, L. Yoon, Z. C. Adams, B. C. Dillingham, B. B. Sanchez, J. C. Vantourout, M. E. Flanagan, D. W. Piotrowski, P. Richardson, S. A. Green, R. A. Shenvi, J. S. Chen, P. S. Baran, P. E. Dawson, Expanding Reactivity in DNA-Encoded Library Synthesis via Reversible Binding of DNA to an Inert Quaternary Ammonium Support. *J. Am. Chem. Soc.* 2019, **141** (25), 9998-10006.
71. J. Scheuermann, S. Melkko, Y. Zhang, L. Mannocci, M. Jaggi, J. Sobek, D. Neri, DNA-Encoded Chemical Libraries for the Discovery of MMP-3 Inhibitors. *Bioconjug Chem.* 2008, **19**, 778-785.
72. W. Decurtins, M. Wichert, R. M. Franzini, F. Buller, M. A. Stravs, Y. Zhang, D. Neri, J. Scheuermann, Automated screening for small organic ligands using DNA-encoded chemical libraries. *Nature protocols.* 2016, **11** (4), 764-80.
73. C. Zambaldo, J. P. Daguer, J. Saarbach, S. Barluenga, N. Winssinger, Screening for covalent inhibitors using DNA-display of small molecule libraries functionalized with cysteine reactive moieties. *Med. Chem. Commun.* 2016, **7** (7), 1340-1351.
74. C. E. Dumelin, S. Trussel, F. Buller, E. Trachsel, F. Bootz, Y. Zhang, L. Mannocci, S. C. Beck, M. Drumea-Mirancea, M. W. Seeliger, C. Baltes, T. Muggler, F. Kranz, M. Rudin, S. Melkko, J. Scheuermann, D. Neri, A portable albumin binder from a DNA-encoded chemical library. *Angew. Chem. Int. Ed.* 2008, **47** (17), 3196-201.
75. L. Encinas, H. O'Keefe, M. Neu, M. J. Remuinan, A. M. Patel, A. Guardia, C. P. Davie, N. Perez-Macias, H. Yang, M. A. Convery, J. A. Messer, E. Perez-Herran, P. A. Centrella, D. Alvarez-Gomez, M. A. Clark, S. Huss, G. K. O'Donovan, F. Ortega-Muro, W. McDowell, P. Castaneda, C. C. Arico-Muendel, S. Pajk, J. Rullas, I. Angulo-Barturen, E. Alvarez-Ruiz, A. Mendoza-Losana, L. Ballell Pages, J. Castro-Pichel, G. Evindar, Encoded library technology as a source of hits for the discovery and lead optimization of a potent and selective class of

bactericidal direct inhibitors of Mycobacterium tuberculosis InhA. *J. Med. Chem*. 2014, **57** (4), 1276-88.
76. F. Samain, T. Ekblad, G. Mikutis, N. Zhong, M. Zimmermann, A. Nauer, D. Bajic, W. Decurtins, J. Scheuermann, P. J. Brown, J. Hall, S. Graslund, H. Schuler, D. Neri, R. M. Franzini, Tankyrase 1 Inhibitors with Drug-like Properties Identified by Screening a DNA-Encoded Chemical Library. *J. Med. Chem*. 2015, **58** (12), 5143-9.
77. Z. Wu, T. L. Graybill, X. Zeng, M. Platchek, J. Zhang, V. Q. Bodmer, D. D. Wisnoski, J. Deng, F. T. Coppo, G. Yao, A. Tamburino, G. Scavello, G. J. Franklin, S. Mataruse, K. L. Bedard, Y. Ding, J. Chai, J. Summerfield, P. A. Centrella, J. A. Messer, A. Pope, D. I. Israel, Cell-Based Selection Expands the Utility of DNA-Encoded Small-Molecule Library Technology to Cell Surface Drug Targets: Identification of Novel Antagonists of the NK3 Tachykinin Receptor. *ACS Comb. Sci*. 2015, **17** (12), 722-31.
78. J. Silke, J. A. Rickard, M. Gerlic, The diverse role of RIP kinases in necroptosis and inflammation*. Nat Immunol.* 2015, **16** (7), 689-97.
79. D. Ofengeim, J. Yuan, Regulation of RIP1 kinase signalling at the crossroads of inflammation and cell death. *Nat. Rev. Mol. Cell Biol*. 2013, **14** (11), 727-36.
80. P. A. Harris, B. W. King, D. Bandyopadhyay, S. B. Berger, N. Campobasso, C. A. Capriotti, J. A. Cox, L. Dare, X. Dong, J. N. Finger, L. C. Grady, S. J. Hoffman, J. U. Jeong, J. Kang, V. Kasparcova, A. S. Lakdawala, R. Lehr, D. E. McNulty, R. Nagilla, M. T. Ouellette, C. S. Pao, A. R. Rendina, M. C. Schaeffer, J. D. Summerfield, B. A. Swift, R. D. Totoritis, P. Ward, A. Zhang, D. Zhang, R. W. Marquis, J. Bertin, P. J. Gough, DNA-Encoded Library Screening Identifies Benzo[b][1,4]oxazepin-4-ones as Highly Potent and Monoselective Receptor Interacting Protein 1 Kinase Inhibitors. *J. Med. Chem*. 2016, **59** (5), 2163-78.
81. H. L. Fu, R. R. Valiathan, R. Arkwright, A. Sohail, C. Mihai, M. Kumarasiri, K. V. Mahasenan, S. Mobashery, P. Huang, G. Agarwal, R. Fridman, Discoidin domain receptors: unique receptor tyrosine kinases in collagen-mediated signaling*. J. Biol. Chem.* 2013, **288** (11), 7430-7.
82. C. M. Borza, A. Pozzi, Discoidin domain receptors in disease. *Matrix Biol*. 2014, **34**, 185-92.
83. H. Richter, A. L. Satz, M. Bedoucha, B. Buettelmann, A. C. Petersen, A. Harmeier, R. Hermosilla, R. Hochstrasser, D. Burger, B. Gsell, R. Gasser, S. Huber, M. N. Hug, B. Kocer, B. Kuhn, M. Ritter, M. G. Rudolph, F. Weibel, J. Molina-David, J. J. Kim, J. V. Santos, M. Stihle, G. J. Georges, R. D. Bonfil, R. Fridman, S. Uhles, S. Moll, C. Faul, A. Fornoni, M. Prunotto, DNA-Encoded Library-Derived DDR1 Inhibitor Prevents Fibrosis and Renal Function Loss in a Genetic Mouse Model of Alport Syndrome. *ACS Chem. Biol.* 2019, **14** (1), 37-49.
84. C. Morisseau, B. D. Hammock, Epoxide hydrolases: mechanisms, inhibitor designs, and biological roles. *Annu Rev Pharmacol Toxicol.* 2005, **45**, 311-33.
85. C. Morisseau, B. D. Hammock, Impact of soluble epoxide hydrolase and epoxyeicosanoids on human health. *Annu Rev Pharmacol Toxicol*. 2013, **53**, 37-58.
86. S. L. Belyanskaya, Y. Ding, J. F. Callahan, A. L. Lazaar, D. I. Israel, Discovering Drugs with DNA-Encoded Library Technology: From Concept to Clinic with an Inhibitor of Soluble Epoxide Hydrolase. *Chembiochem.* 2017, **18** (9), 837-842.

87. R. A. Goodnow, The Changing Feasibility and Economics of Chemical Diversity Exploration. In A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space, *Ed. Wiley & Sons: Hoboken*, New Jersey, 2014, 417-426.
88. L. Mannocci, M. Leimbacher, M. Wichert, J. Scheuermann, D. Neri, 20 years of DNA-encoded chemical libraries. *Chem Commun*. 2011, **47** (48), 12747-53.
89. J. P. Hughes, S. Rees, S. B. Kalindjian, K. L. Philpott, Principles of early drug discovery. *Br J Pharmacol.* 2011, **162** (6), 1239-49.
90. K. C. Luk, A. Satz, DNA-compatible chemistry. In A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space*, Ed. Wiley & Sons: Hoboken,* New Jersey, 2014, 67-98.
91. J. H. Shibing Tang, S. Yongquan, H. Liuer, S. Xuegong, Efficient and Regioselective One-Pot Synthesis of 3-Substituted and 3,5-Disubstituted Isoxazoles. *Org. Lett.* 2009, **11** (17), 3982-3985.
92. J. Wen, Y. Fu, R. Y. Zhang, J. Zhang, S. Y. Chen, X. Q. Yu, A simple and efficient synthesis of pyrazoles in water. *Tetrahedron.* 2011, **67** (49), 9618-9621.
93. H. Zhang, Z. Guodong, Q. Jingping, W. Baomin, A facile and expeditious approach to substituted 1H-pyrazoles catalyzed by iodine. *Tetrahedron Lett.* 2016, **57**, 2633-2637.
94. J. F. Blake, D. Lim, W. L. Jorgensen, Enhanced Hydrogen Bonding of Water to Diels-Alder Transition States. Ab Initio Evidence*. J. Org. Chem.* 1994, **59**, 803-805.
95. P. A. Grieco, Z. He, "Micellar" catalysis in the aqueous intermolecular Diels-Alder reaction: Rate acceleration and enhanced selectivity. *Tetrahedron Lett*. 1983, 24 (18), 1897-1300.
96. P. S. Pandey, I. K. Pandey, Hydrophobic Effect on 1,3-Dipolar Cycloaddition Reactions. *Tetrahedron Lett*. 1997, **38** (41), 7237-7240.
97. T. Rispens, J. Engberts, Micellar Catalysis of Diels-Alder Reactions: Substrate Positioning in the Micelle. *J. Org. Chem.* 2002, **67**, 7369-7377.
98. P. Klumphu, B. H. Lipshutz, "Nok": a phytosterol-based amphiphile enabling transition-metal-catalyzed couplings in water at room temperature. *J. Org. Chem.* 2014, **79** (3), 888-900.
99. Y. Wenchao, L. Pinhua, W. Lei, Regioselective synthesis of triazoles via basepromoted oxidative cycloaddition of chalcones with azides in aqueous solution. *RSC Adv.* 2015, **5**, 95833-95839.
100. F. M. Cordero, B. B. Khairnar, A. Brandi, Modular Access to Highly Functionalised Tetrahydroquinolines. *Eur. J. Org. Chem.* 2014, **32**, 7122-7133.
101. R. A. Sheldon, The E Factor: fifteen years on. *Green Chem.* 2007, **9** (12).
102. D. J. C. Constable, A. D. Curzons, V. L. Cunningham, Metrics to 'green' chemistry—which are the best? *Green Chem*. 2002, **4** (6), 521-527.
103. K. Tanaka, F. Toda, Solvent-Free Organic Synthesis*. Chem. Rev*., 2000, **100**, 1025-1074.
104. J. L. Howard, Q. Cao, D. L. Browne, Mechanochemistry as an emerging tool for molecular synthesis: what can it offer? *Chem. Sci.* 2018, **9** (12), 3080-3094.
105. R. A. Sheldon, Green solvents for sustainable organic synthesis: state of the art. *Green Chem.* 2005, **7** (5).

106. D. C. Rideout, R. Breslow, Hydrophobic Acceleration of Diels-AIder Reactions. *J. Am. Chem. Soc.* 1980, **102**, 7817-7818.
107. D. G. Blackmond, A. Armstrong, V. Coombe, A. Wells, Water in organocatalytic processes: debunking the myths. *Angew. Chem. Int. Ed*. 2007, **46** (21), 3798-800.
108. L. Marchetti, M. Levine, Biomimetic Catalysis. *ACS Catalysis*. 2011, **1** (9), 1090-1118.
109. Z. Dong, Q. Luo, J. Liu, Artificial enzymes based on supramolecular scaffolds. *Chem. Soc. Rev.* 2012, **41** (23), 7890-908.
110. J. H. Clint, Surfactant Aggregation. *Chapman and Hall: New York*, 1992.
111. B. Bagchi, Water Dynamics in the Hydration Layer around Proteins and Micelles. *Chem. Rev.* 2005, **105** (9), 3197-3219.
112. B. H. Lipshutz, N. A. Isley, J. C. Fennewald, E. D. Slack, On the way towards greener transition-metal-catalyzed processes as quantified by E factors. *Angew. Chem. Int. Ed.* 2013, **52** (42), 10952-8.
113. B. M. Trost, The Atom Economy A Search for Synthetic Efficiency. *Science.* 1991, **254**, 1471-1477.
114. R. A. Sheldon, The E factor 25 years on: the rise of green chemistry and sustainability. *Green Chem.* 2017, **19** (1), 18-43.
115. B. H. Lipshutz, S. Ghorai, A. R. Abela, R. Moser, T. Nishikata, C. Duplais, A. Krasovskiy, R. D. Gaston, R. C. Gadwood, TPGS-750-M: a second-generation amphiphile for metal-catalyzed cross-couplings in water at room temperature. *J. Org. Chem*. 2011, **76** (11), 4379-91.
116. S. Ghorai, B. H. Lipshutz, "Designer"-Surfactant-Enabled Cross-Couplings in Water at Room Temperature. *Aldrichimica Acta.* 2012, **45** (1), 3-18.
117. B. S. Takale, R. R. Thakore, S. Handa, F. Gallou, J. Reilly, B. H. Lipshutz, A new, substituted palladacycle for ppm level Pd-catalyzed Suzuki-Miyaura cross couplings in water. *Chem. Sci*. 2019, **10** (38), 8825-8831.
118. B. H. Lipshutz, T. B. Peterson, A, R. Abela, Room-Temperature Suzuki−Miyaura Couplings in Water Facilitated by Nonionic Amphiphiles. *Org. Lett*. 2007, **10** (7), 1333-1336.
119. S. Handa, J. D. Smith, Y. Zhang, B. S. Takale, F. Gallou, B. H. Lipshutz, Sustainable HandaPhos-ppm Palladium Technology for Copper-Free Sonogashira Couplings in Water under Mild Conditions. *Org. Lett*. 2018, **20** (3), 542-545.
120. G. P. Lu, C. Cai, B. H. Lipshutz, Stille couplings in water at room temperature. *Green Chem.* 2013, **15** (1), 105-109.
121. B. H. Lipshutz, B. R. Taft, Heck Couplings at Room Temperature in Nanometer Aqueous Micelles. *Org. Lett.* 2008, **10** (7), 1329-1332.
122. K. Voigtritter, S. Ghorai, B. H. Lipshutz, Rate enhanced olefin cross-metathesis reactions: the copper iodide effect. *J. Org. Chem*. 2011, **76** (11), 4697-702.
123. C. M. Gabriel, M. Keener, F. Gallou, B. H. Lipshutz, Amide and Peptide Bond Formation in Water at Room Temperature. *Org. Lett*. 2015, **17** (16), 3968-71.
124. M. Cortes-Clerget, J. Y. Berthon, I. Krolikiewicz-Renimel, L. Chaisemartin, B. H. Lipshutz, Tandem deprotection/coupling for peptide synthesis in water at room temperature. *Green Chem.* 2017, **19** (18), 4263-4267.

125. M. Cortes-Clerget, S. E. Spink, G. P. Gallagher, L. Chaisemartin, E. Filaire, J. Y. Berthon, B. H. Lipshutz, MC-1. A “designer” surfactant engineered for peptide synthesis in water at room temperature. *Green Chem*. 2019, **21** (10), 2610-2614.
126. E. B. Landstrom, M. Nichol, B. H. Lipshutz, M. J. Gainer, Discovery-Based SNAr Experiment in Water Using Micellar Catalysis. *J. Chem. Educ.* 2019, **96** (11), 2668-2671.
127. B. H. Lipshutz, Synthetic chemistry in a water world. New rules ripe for discovery. *Curr Opin Green Sust.* 2018, **11**, 1-8.
128. B. H. Lipshutz, G. T. Aguinaldo, S. Ghorai, K. Voigtritter, Olefin Cross-Metathesis Reactions at Room Temperature Using the Nonionic Amphiphile “PTS”: Just Add Water. *Org Lett*. 2008, **10** (7), 1325-1328.
129. M. P. Andersson, F. Gallou, P. Klumphu, B. S. Takale, B. H. Lipshutz, Structure of Nanoparticles Derived from Designer Surfactant TPGS-750-M in Water, As Used in Organic Synthesis. *Chemistry.* 2018, **24** (26), 6778-6786.
130. M. K. Skopic, K. Gotte, C. Gramse, M. Dieter, S. Pospich, S. Raunser, R. Weberskirch, A. Brunschweiger, Micellar Bronsted Acid Mediated Synthesis of DNA-Tagged Heterocycles. *J. Am. Chem. Soc.* 2019, **141** (26), 10546-10555.
131. N. T. S. Phan, M. Van Der Sluys, C. W. Jones, On the Nature of the Active Species in Palladium Catalyzed Mizoroki–Heck and Suzuki–Miyaura Couplings – Homogeneous or Heterogeneous Catalysis, *ACS A Critical Review*. 2006, **348** (6), 609-679.
132. A. J. Lennox, G. C. Lloyd-Jones, Transmetalation in the Suzuki-Miyaura coupling: the fork in the trail. *Angew Chem Int Ed.* 2013, **52** (29), 7362-70.
133. K. Billingsley, S. L. Buchwald, Highly Efficient Monophosphine-Based Catalyst for the Palladium-Catalyzed Suzuki-Miyaura Reaction of Heteroaryl Halides and Heteroaryl Boronic Acids and Esters. *J. Am. Chem. Soc*. 2007, **129**, 3358-3366.
134. A. J. Lennox, G. C. Lloyd-Jones, Selection of boron reagents for Suzuki-Miyaura coupling. *Chem Soc Rev*. 2014, **43** (1), 412-43.
135. A. J. Lennox, G. C. Lloyd-Jones, The Slow-Release Strategy in Suzuki-Miyaura Coupling. *Isr. J. Chem.* 2010, **50** (5-6), 664-674.
136. D. M. Knapp, E. P. Gillis, M. D. Burke, A General Solution for Unstable Boronic Acids: Slow-Release Cross-Coupling from Air-Stable MIDA Boronates. *J. Am. Chem. Soc.* 2009, **131**, 6961-6963.
137. M. Shevlin, Practical High-Throughput Experimentation for Chemists. *ACS Med. Chem. Lett*. 2017, **8** (6), 601-607.
138. M. M. Heravi, M. Ghavidel, L. Mohammadkhani, Beyond a solvent: triple roles of dimethylformamide in organic chemistry. *RSC Adv.* 2018, **8** (49), 27832-27862.
139. C. M. Gabriel, N. R. Lee, F. Bigorne, P. Klumphu, M. Parmentier, F. Gallou, B. H. Lipshutz, Effects of Co-solvents on Reactions Run under Micellar Catalysis Conditions. *Org Lett.* 2017, **19** (1), 194-197.
140. K. M. Engle, J. Q. Yu, Developing ligands for palladium(II)-catalyzed C-H functionalization: intimate dialogue between ligand and substrate. *J Org Chem.* 2013, **78** (18), 8927-55.
141. L. V. Bystrykh, Generalized DNA barcode design based on Hamming codes. *PLoS One*. 2012, **7** (5), e36852.

142. L. J. Kempkes, J. Martens, J. Grzetic, G. Berden, J. Oomens, Deamidation Reactions of Asparagine- and Glutamine-Containing Dipeptides Investigated by Ion Spectroscopy. *J Am Soc Mass Spectr*. 2016, **27** (11), 1855-1869.
143. H. Yang, R. A. Zubarev, Mass spectrometric analysis of asparagine deamidation and aspartate isomerization in polypeptides. *Electrophoresis.* 2010, **31** (11), 1764-72.
144. A. Bermingham, M. A. Chand, C. S. Brown, E. Aarons, C. Tong, C. Langrish, K. Hoschler, K. Brown, M. Galiano, R. Myers, R. G. Pebody, H. K. Green, N. L. Boddington, R. Gopal, N. Price, W. Newsholme, C. Drosten, R. A. Fouchier, M. Zambon, Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro. Surveill.* 2012, **17** (20290), 1 - 5.
145. A. M. Zaki, S. van Boheemen, T. M. Bestebroer, A. D. Osterhaus, R. A. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med.* 2012, **367** (19), 1814-20.
146. T. Kuiken, R. A. M. Fouchier, M. Schutten, G. F. Rimmelzwaan, G. van Amerongen, D. van Riel, J. D. Laman, T. de Jong, G. van Doornum, W. Lim, A. E. Ling, P. K. S. Chan, J. S. Tam, M. C. Zambon, R. Gopal, C. Drosten, S. van der Werf, N. Escriou, J. C. Manuguerra, K. Stöhr, J. S. M. Peiris, A. D. M. E. Osterhaus, Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *The Lancet.* 2003, **362** (9380), 263-270.
147. P. J. Bredenbeek, C. J. Pachuk, A. F. H. Noten, J. Charite, W. Luytjes, S. R. Weiss and W. J. M. Spaan, The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Research*. 1990, **18** (7), 1825-1832.
148. V. Thiel, K. A. Ivanov, A. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, J. Ziebuhr, Mechanisms and enzymes involved in SARS coronavirus genome expression. *J Gen Virol.* 2003, **84**, 2305-2315.
149. R. Hilgenfeld, From SARS to MERS: crystallographic studies on coronaviral proteases enable antiviral drug design. *FEBS J.* 2014, **281** (18), 4085-96.
150. L. Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E. J. Snijder, H. Liu, R. Hilgenfeld, alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J Med Chem.* 2020, **63** (9), 4562-4578.
151. L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science.* 2020, **368**, 409-412.
152. A. Douangamath, D. Fearon, P. Gehrtz, T. Krojer, P. Lukacik, C. D. Owen, E. Resnick, C. Strain-Damerell, A. Aimon, P. Abranyi-Balogh, J. Brandao-Neto, A. Carbery, G. Davison, A. Dias, T. D. Downes, L. Dunnett, M. Fairhead, J. D. Firth, S. P. Jones, A. Keeley, G. M. Keseru, H. F. Klein, M. P. Martin, M. E. Noble, P. O'Brien, A. Powell, R. N. Reddi, R. Skyner, M. Snee, M. J. Waring, C. Wild, N. London, F. von Delft, M. A. Walsh, Crystallographic and electrophilic fragment screening of the SARS-CoV-2 main protease. *Nature Comm*. 2020, **11** (1), 5047.
153. D. J. Wood, J. D. Lopez-Fernandez, L. E. Knight, I. Al-Khawaldeh, C. Gai, S. Lin, M. P. Martin, D. C. Miller, C. Cano, J. A. Endicott, I. R. Hardcastle, M. E. M. Noble, M. J. Waring, FragLites-Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An

Efficient Approach to Druggability Assessment and Hit Generation. *J. Med Chem.* 2019, **62** (7), 3741-3752.
154. A. L. Satz, R. Hochstrasser, A. C. Petersen, Analysis of Current DNA Encoded Library Screening Data Indicates Higher False Negative Rates for Numerically Larger Libraries. *ACS Comb Sci.* 2017, **19** (4), 234-238.
155. K. J. Fountain, M. Gilar, J. C. Gebler, Analysis of native and chemically modified oligonucleotides by tandem ion-pair reversed-phase high-performance liquid chromatography/electrospray ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* 2003, **17** (7), 646-53.
156. A. C. J. Apffel, S. Fischer, K. Lichtenwalter, W. S. Hancock, Analysis of Oligonucleotides by HPLC-Electrospray Ionization Mass Spectrometry. *Anal Chem.* 1997, **69**, 1320-1325.
157. M. Gilar, Analysis and purification of synthetic oligonucleotides by reversed-phase high-performance liquid chromatography with photodiode array and mass spectrometry detection. *Anal Biochem*. 2001, **298** (2), 196-206.

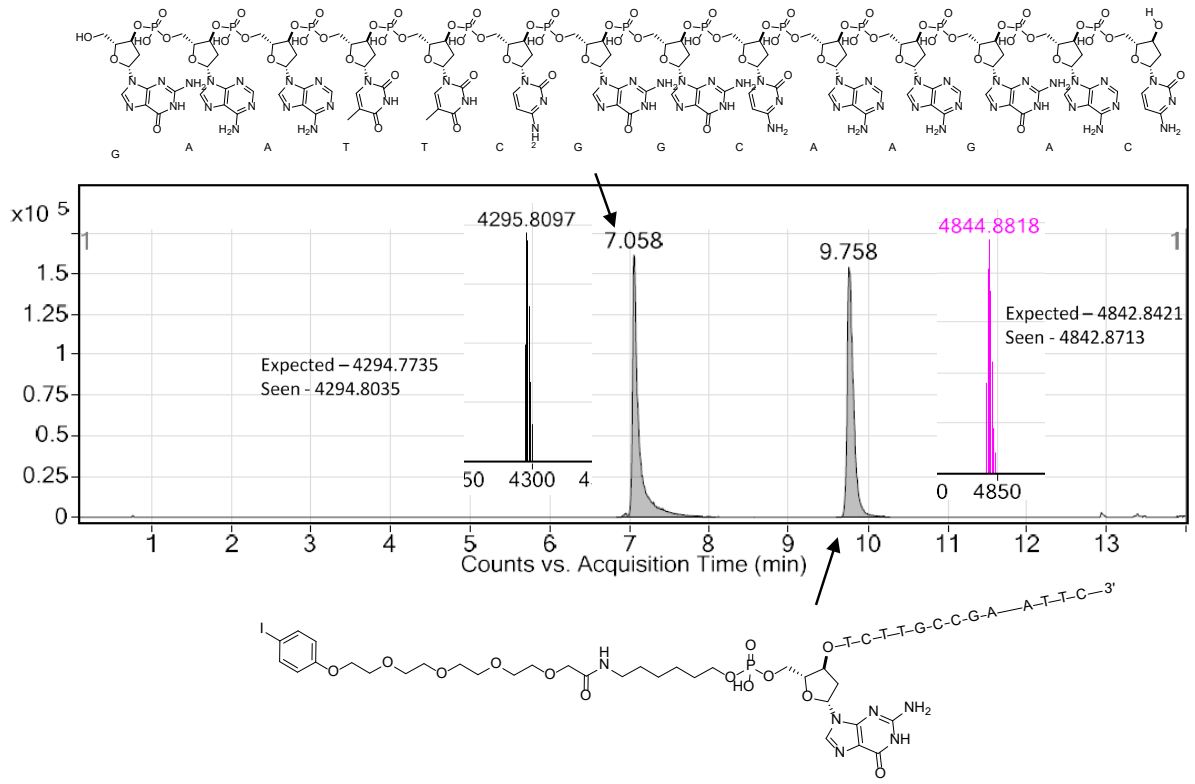
Appendix

Appendix Table 1 - Results of NGS of the amplified prototype library showing the complimentary strand.

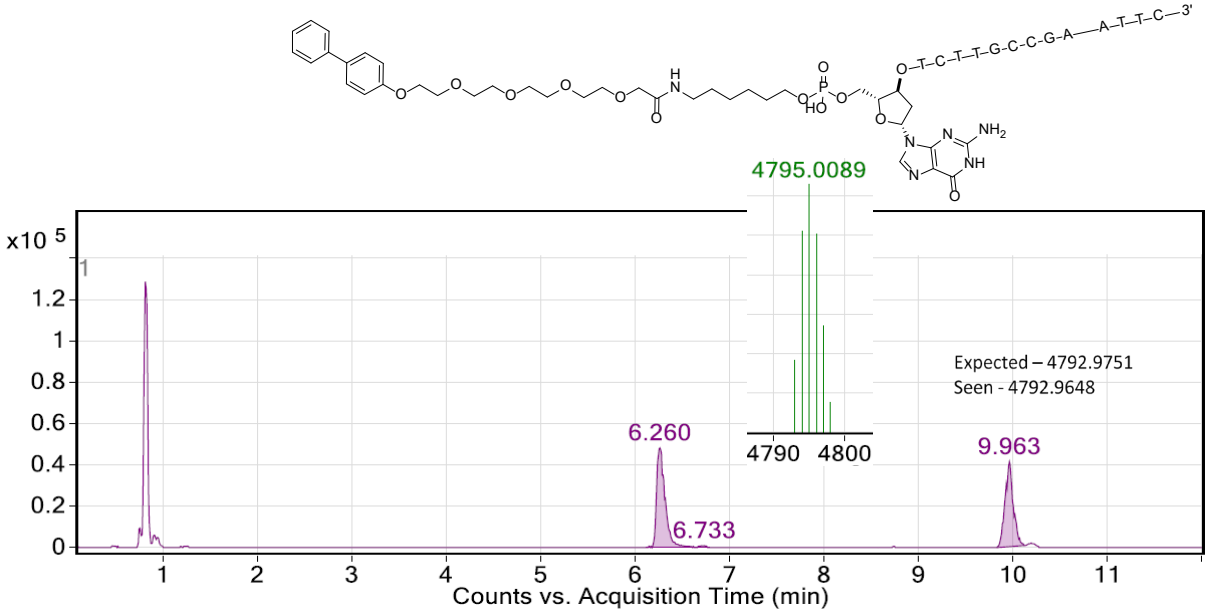
Code (Linker strand) 5'-3'	Counts	BB1	BB2
TGTAGACCATGTAGTTGAGGTCACTGCGATATAGGACGACACTATACAGCCAAATCCGTTCACACCGACCT	7377	c	c
TGTAGACCATGTAGTTGAGGTCATGGCTAACTAGGACGACACTATACAGCCAAATCCGTTCACACCGACCT	7171	c	b
TGTAGACCATGTAGTTGAGGTCACTGCGATATAGGTCACGATGATACAGCCAAATCCGTTCACACCGACCT	6487	e	c
TGTAGACCATGTAGTTGAGGTCATGGCTAACTAGGTCACGATGATACAGCCAAATCCGTTCACACCGACCT	5830	e	b
TGTAGACCATGTAGTTGAGGTCACTGCGATATAGGGACCGATTATACAGCCAAATCCGTTCACACCGACCT	5486	f	c
TGTAGACCATGTAGTTGAGGTCACTGCGATATAGGGTAACACCATACAGCCAAATCCGTTCACACCGACCT	5356	b	c
TGTAGACCATGTAGTTGAGGTCATGGCTAACTAGGGACCGATTATACAGCCAAATCCGTTCACACCGACCT	5083	f	b
TGTAGACCATGTAGTTGAGGTCATGGCTAACTAGGGTAACACCATACAGCCAAATCCGTTCACACCGACCT	4465	b	b
TGTAGACCATGTAGTTGAGGTCATAGCCAGTTAGGACGACACTATACAGCCAAATCCGTTCACACCGACCT	4192	c	a
TGTAGACCATGTAGTTGAGGTCATGGCTAACTAGGTCCAGAGTATACAGCCAAATCCGTTCACACCGACCT	4161	a	b
TGTAGACCATGTAGTTGAGGTCACTGCGATATAGGTCCAGAGTATACAGCCAAATCCGTTCACACCGACCT	3851	a	c
TGTAGACCATGTAGTTGAGGTCATAGCCAGTTAGGTCACGATGATACAGCCAAATCCGTTCACACCGACCT	3784	e	a
TGTAGACCATGTAGTTGAGGTCACTGCGATATAGGCTTAGAGCATACAGCCAAATCCGTTCACACCGACCT	3745	d	c
TGTAGACCATGTAGTTGAGGTCATGGCTAACTAGGCTTAGAGCATACAGCCAAATCCGTTCACACCGACCT	3723	d	b
TGTAGACCATGTAGTTGAGGTCATAGCCAGTTAGGGACCGATTATACAGCCAAATCCGTTCACACCGACCT	3137	f	a
TGTAGACCATGTAGTTGAGGTCATAGCCAGTTAGGGTAACACCATACAGCCAAATCCGTTCACACCGACCT	2859	b	a
TGTAGACCATGTAGTTGAGGTCATAGCCAGTTAGGTCCAGAGTATACAGCCAAATCCGTTCACACCGACCT	2332	a	a
TGTAGACCATGTAGTTGAGGTCATAGCCAGTTAGGCTTAGAGCATACAGCCAAATCCGTTCACACCGACCT	2158	d	a
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCACGATGATACAGCCAAATCCGTTCACACCGACCT	1765	e	e
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGACGACACTATACAGCCAAATCCGTTCACACCGACCT	1482	c	e
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGTAACACCATACAGCCAAATCCGTTCACACCGACCT	1218	b	e
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGGACCGATTATACAGCCAAATCCGTTCACACCGACCT	1125	f	e
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGTCCAGAGTATACAGCCAAATCCGTTCACACCGACCT	870	a	e
TGTAGACCATGTAGTTGAGGTCAAGTCGATCTAGGCTTAGAGCATACAGCCAAATCCGTTCACACCGACCT	829	d	e
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGACGACACTATACAGCCAAATCCGTTCACACCGACCT	382	c	f
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGACGACACTATACAGCCAAATCCGTTCACACCGACCT	364	c	d
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCACGATGATACAGCCAAATCCGTTCACACCGACCT	351	e	f
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCACGATGATACAGCCAAATCCGTTCACACCGACCT	343	e	d
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGTAACACCATACAGCCAAATCCGTTCACACCGACCT	321	b	d
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGACCGATTATACAGCCAAATCCGTTCACACCGACCT	264	f	f
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGGTAACACCATACAGCCAAATCCGTTCACACCGACCT	263	b	f
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGTCCAGAGTATACAGCCAAATCCGTTCACACCGACCT	205	a	f
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGGACCGATTATACAGCCAAATCCGTTCACACCGACCT	191	f	d
TGTAGACCATGTAGTTGAGGTCAGTTCCAGATAGGCTTAGAGCATACAGCCAAATCCGTTCACACCGACCT	188	d	f
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGTCCAGAGTATACAGCCAAATCCGTTCACACCGACCT	183	a	d
TGTAGACCATGTAGTTGAGGTCAGCTCTAAGTAGGCTTAGAGCATACAGCCAAATCCGTTCACACCGACCT	170	d	d

Appendix Table 2 - Results of NGS of the amplified 1x1x1 SARS-COV-2 library showing the complimentary strand.

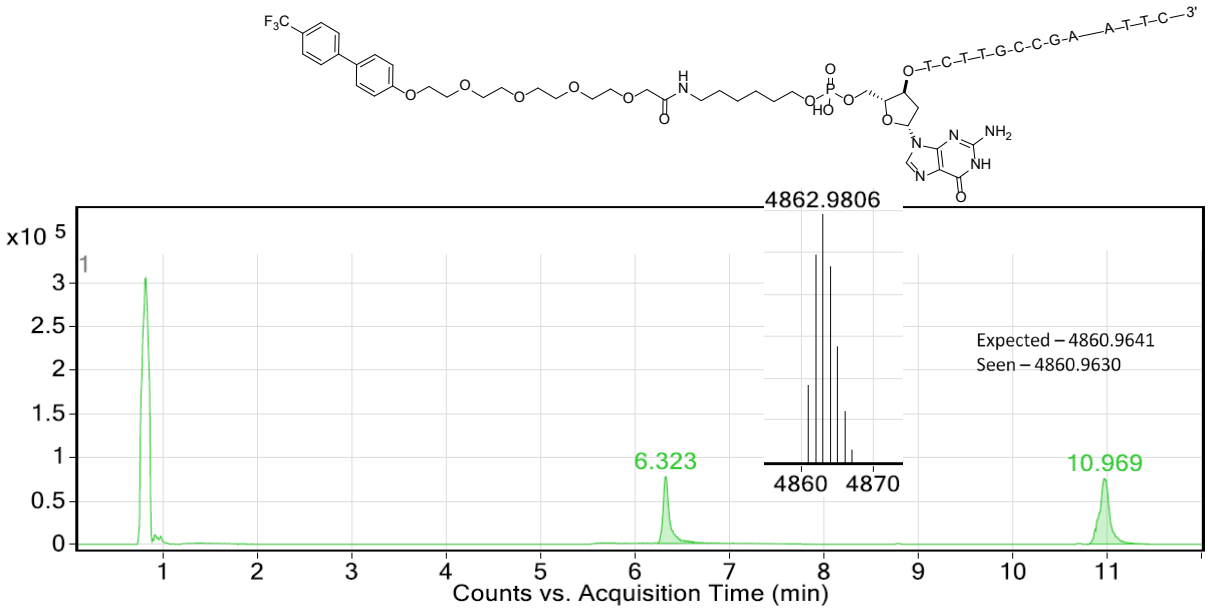
Level	Count	Prob
Total	**76902**	**1**
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	62909	0.81804
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTGTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	2785	0.03621
AGGTCGGTGTGAACGGATTTGGCTGTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	1937	0.02519
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACAGTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	143	0.00186
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACAACGGCACCTGACCTCAACTACATGGTCTACA	104	0.00135
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	104	0.00135
AGGTCGGTGTGAACGGATTTGGTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	102	0.00133
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	101	0.00131
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	92	0.0012
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCGTCTTGCCGAATTCAGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACTACATGGTCTACA	90	0.00117
AGGTCGGTGTGAACGGATTTGGCTGTATGATCGACTCCTACGCGCACATACGACGGCACCTGACCTCAACACATGGTCTACA	84	0.00109



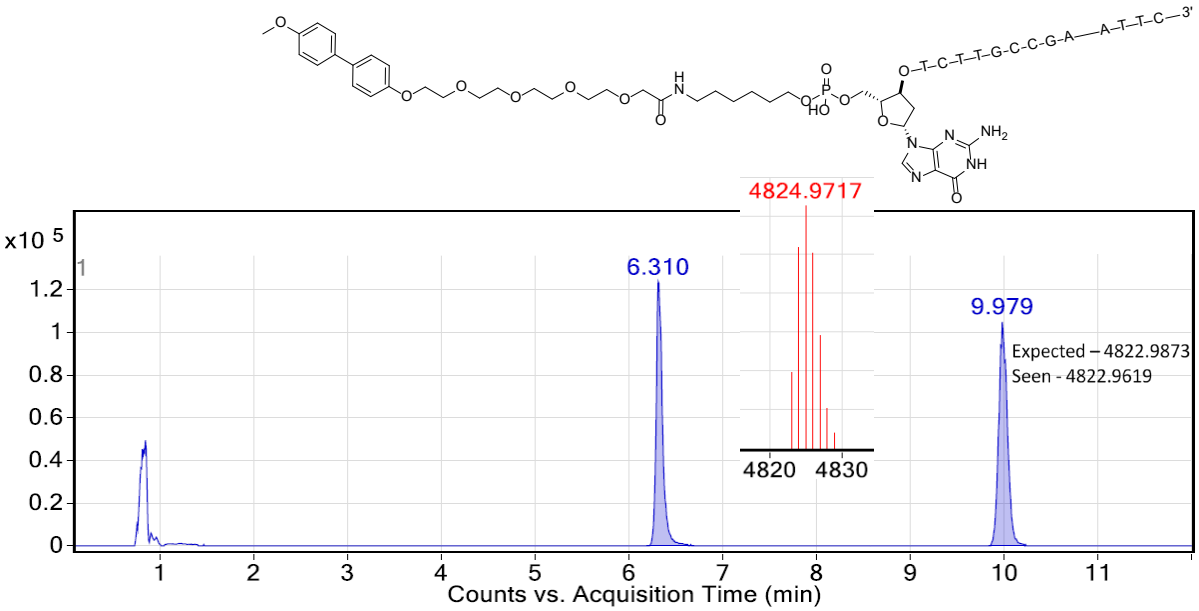
Appendix Figure 1 - Chromatogram of double stranded starting material 50 used for Suzuki-Miyaura couplings.



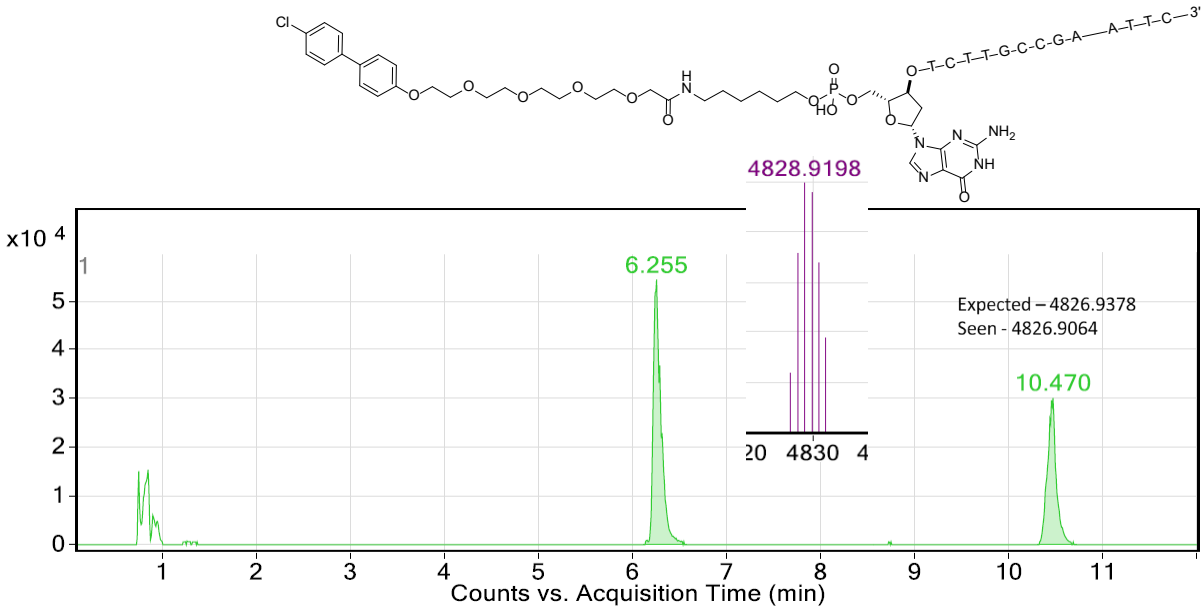
Appendix Figure 2 - Chromatogram of double stranded Suzuki coupling product 51a.



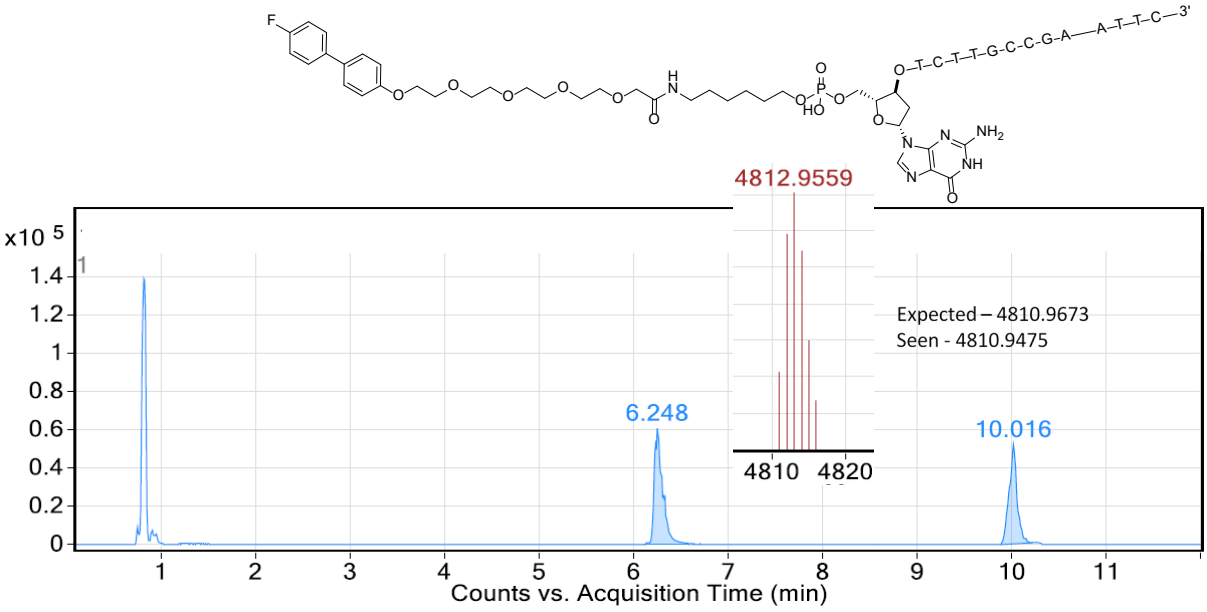
Appendix Figure 3 - Chromatogram of double stranded Suzuki coupling product 51b.



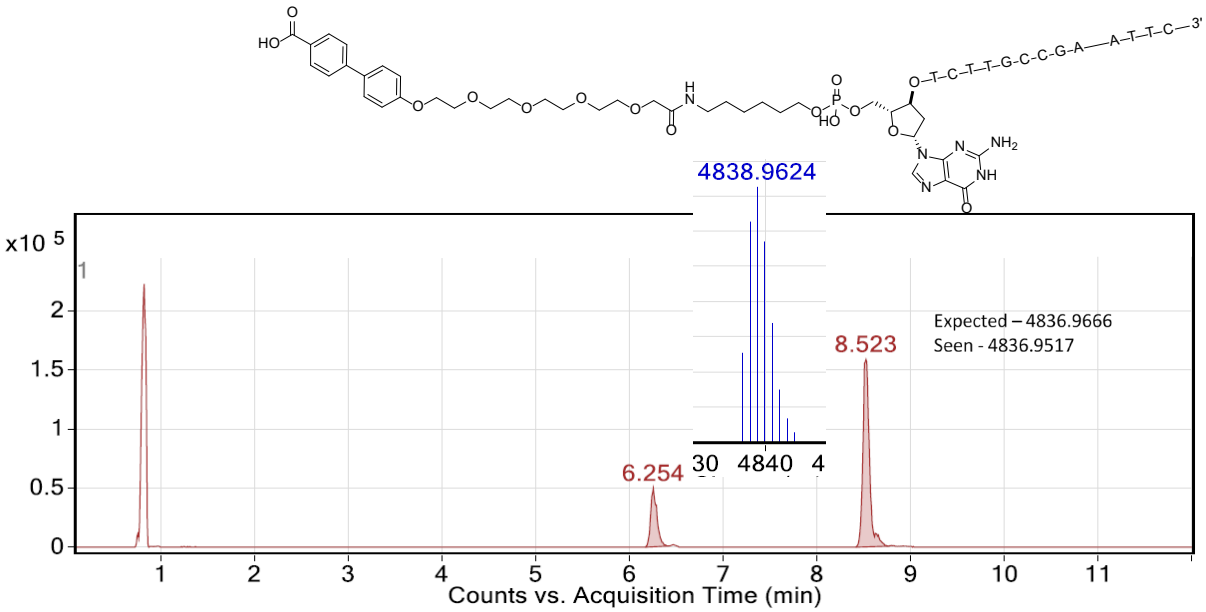
Appendix Figure 4 - Chromatogram of double stranded Suzuki coupling product 51c.



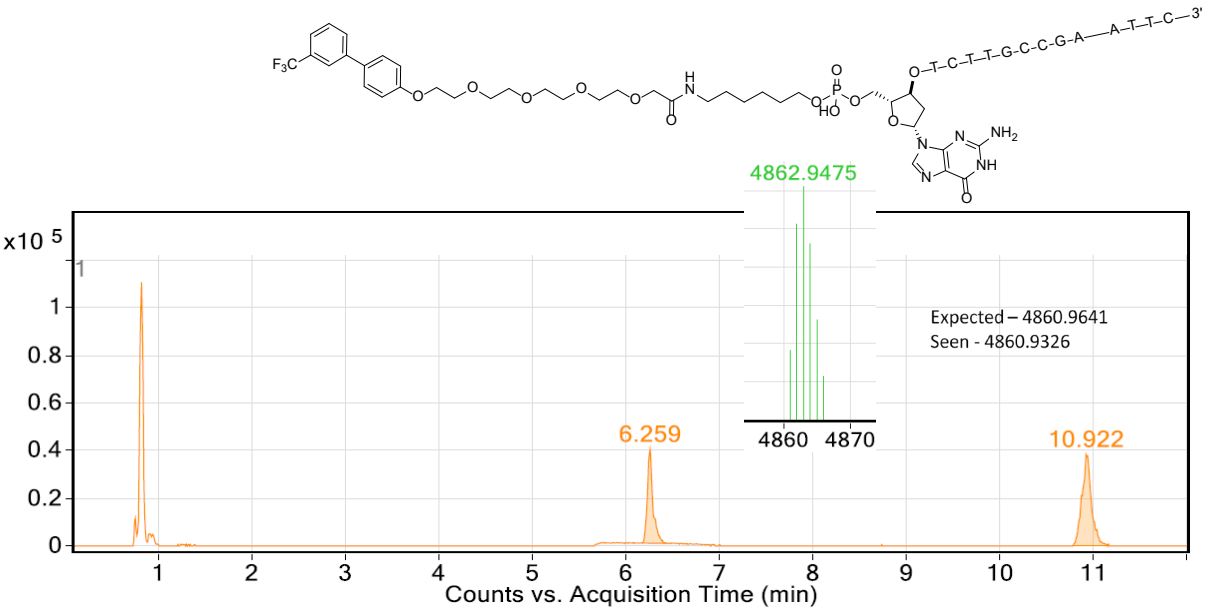
Appendix Figure 5 - Chromatogram of double stranded Suzuki coupling product 51d.



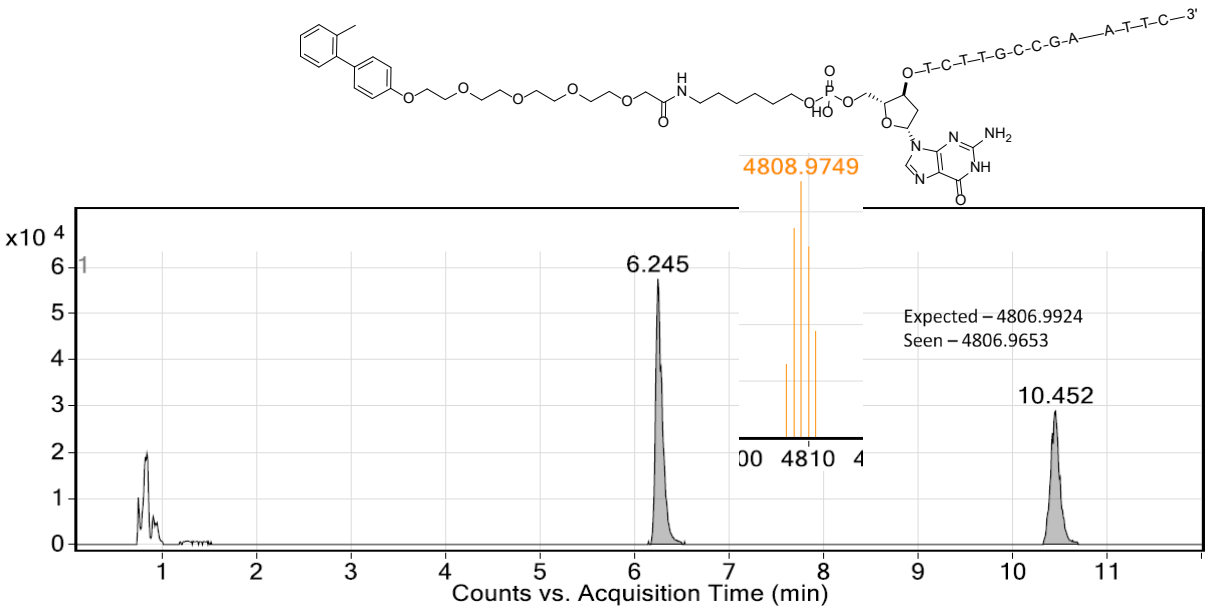
Appendix Figure 6 - Chromatogram of double stranded Suzuki coupling product 51e.



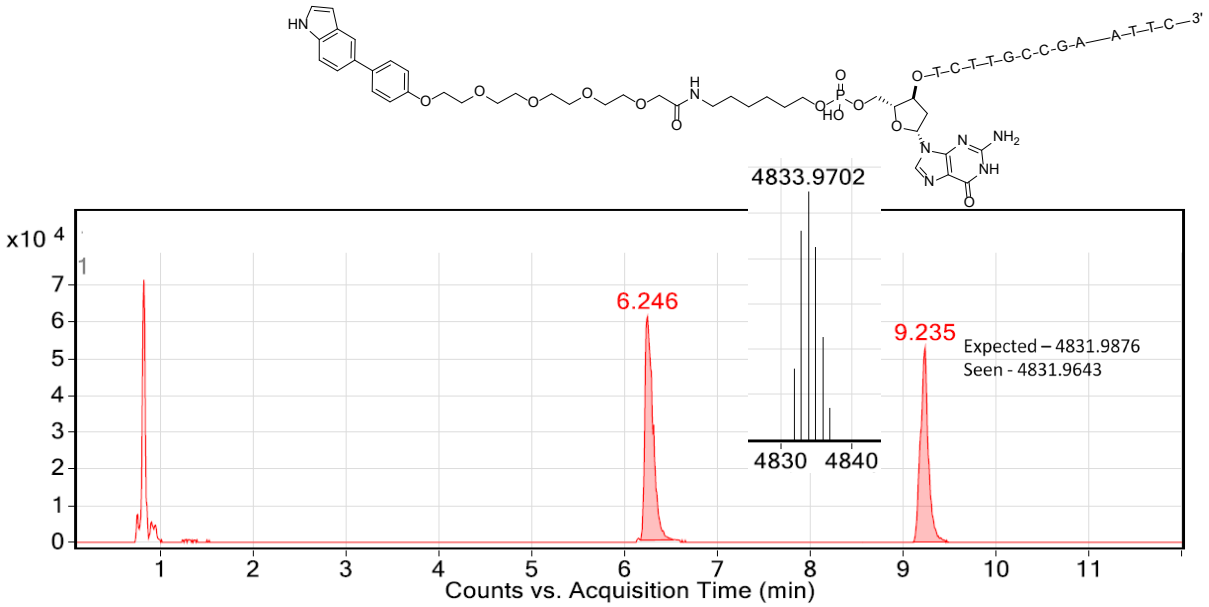
Appendix Figure 7 - Chromatogram of double stranded Suzuki coupling product 51f.



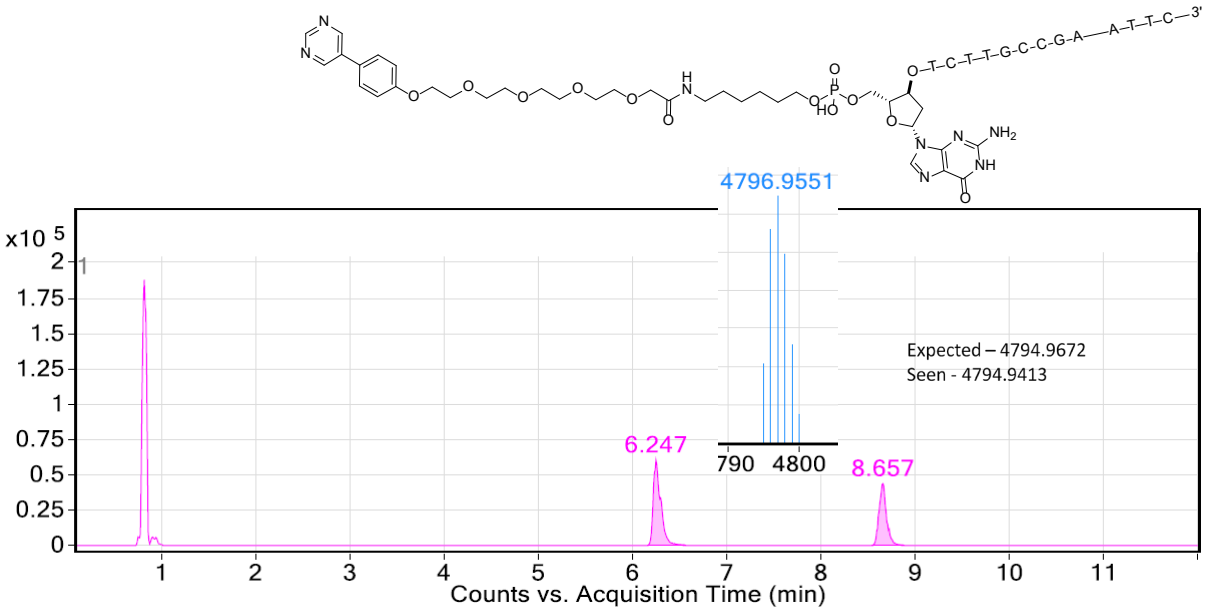
Appendix Figure 8 - Chromatogram of double stranded Suzuki coupling product 51g.



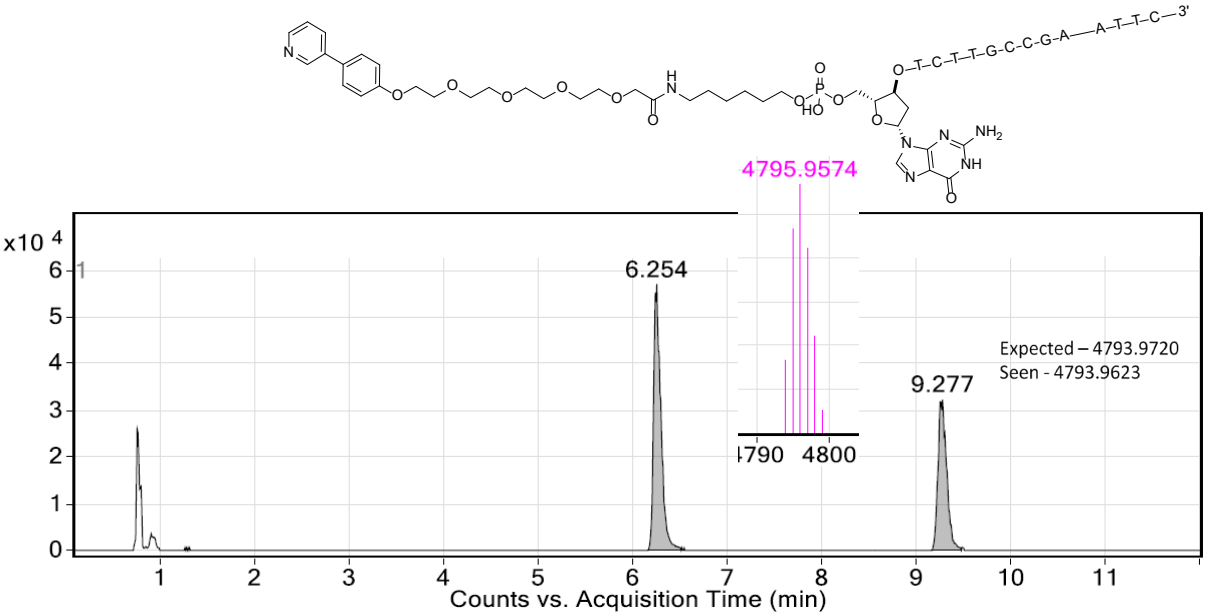
Appendix Figure 9 - Chromatogram of double stranded Suzuki coupling product 51h.



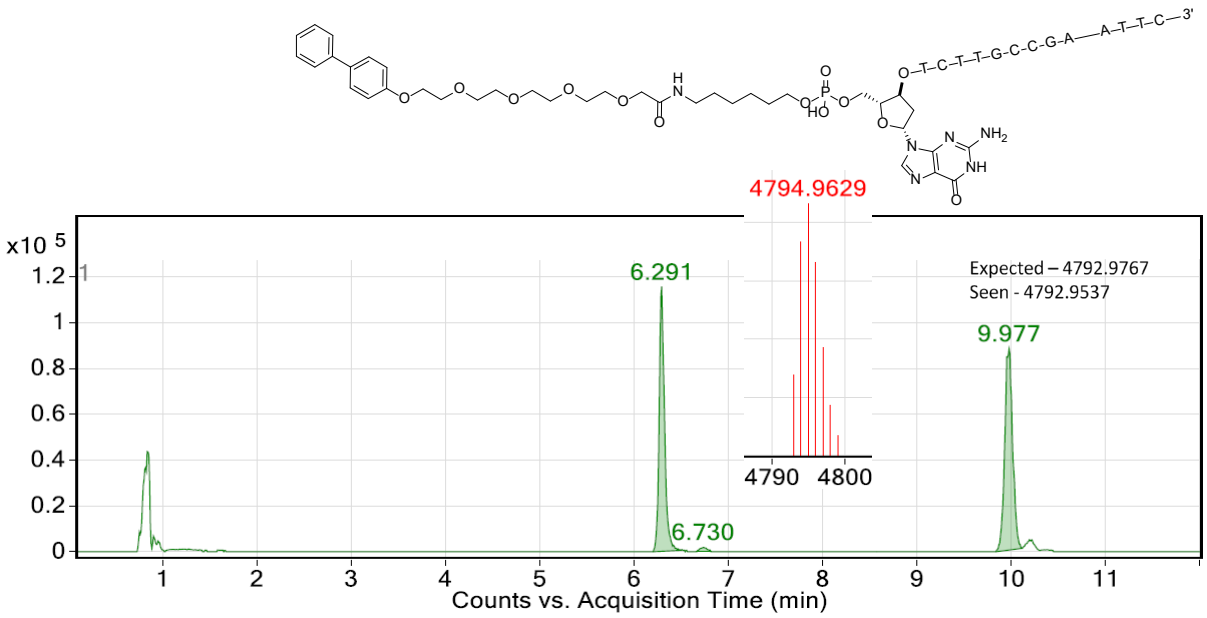
Appendix Figure 10 - Chromatogram of double stranded Suzuki coupling product 51i.



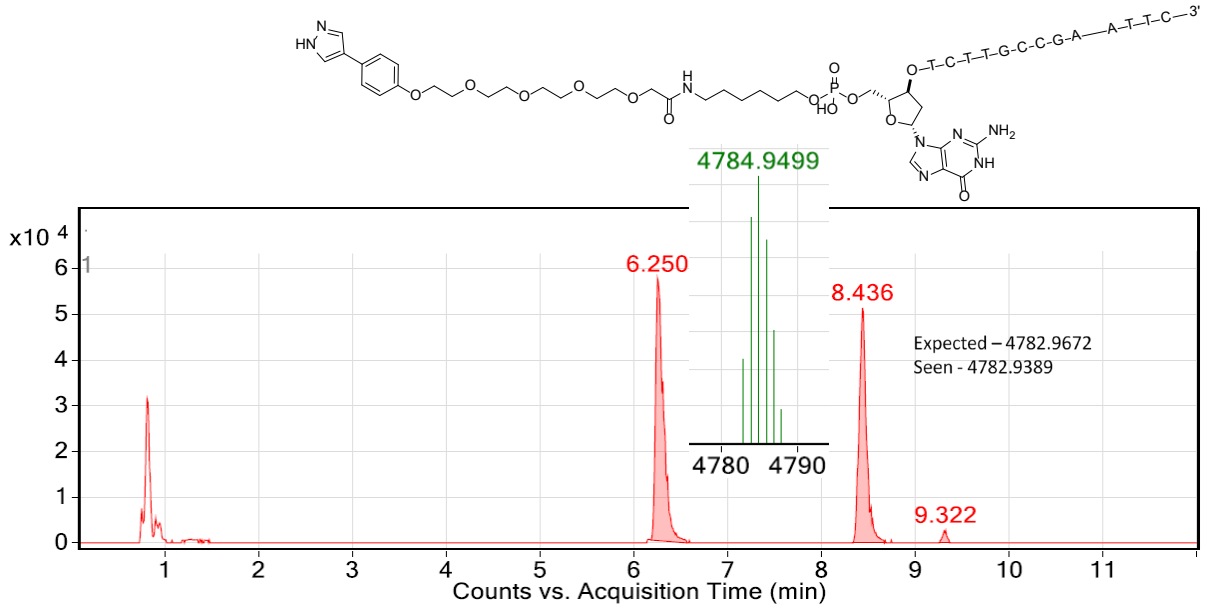
Appendix Figure 11 - Chromatogram of double stranded Suzuki coupling product 51j.



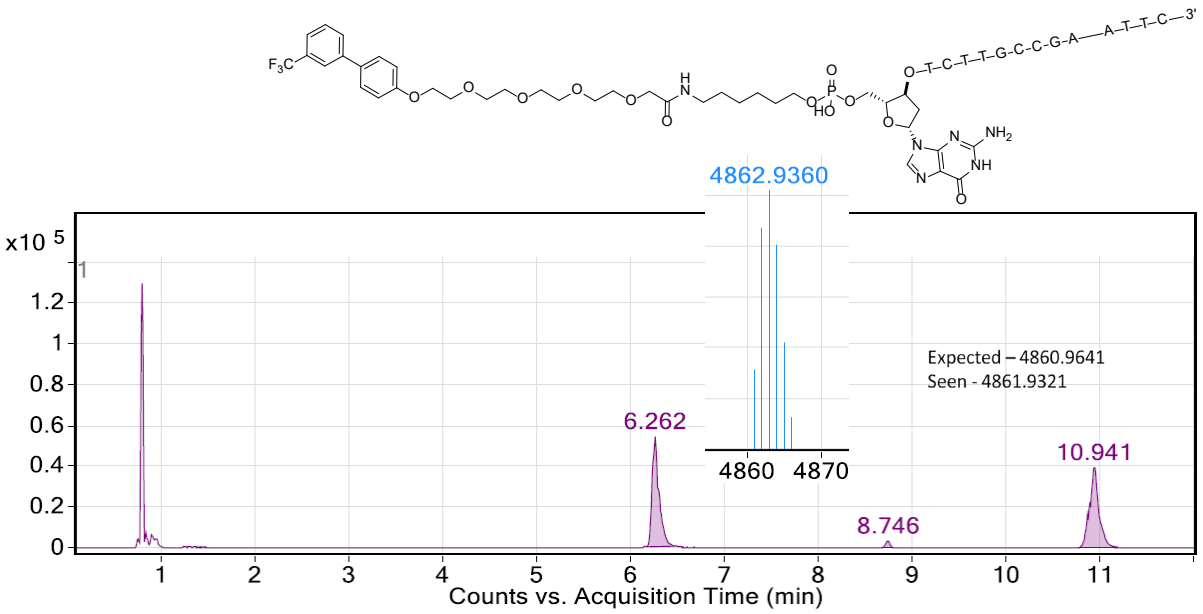
Appendix Figure 12 - Chromatogram of double stranded Suzuki coupling product 51k.



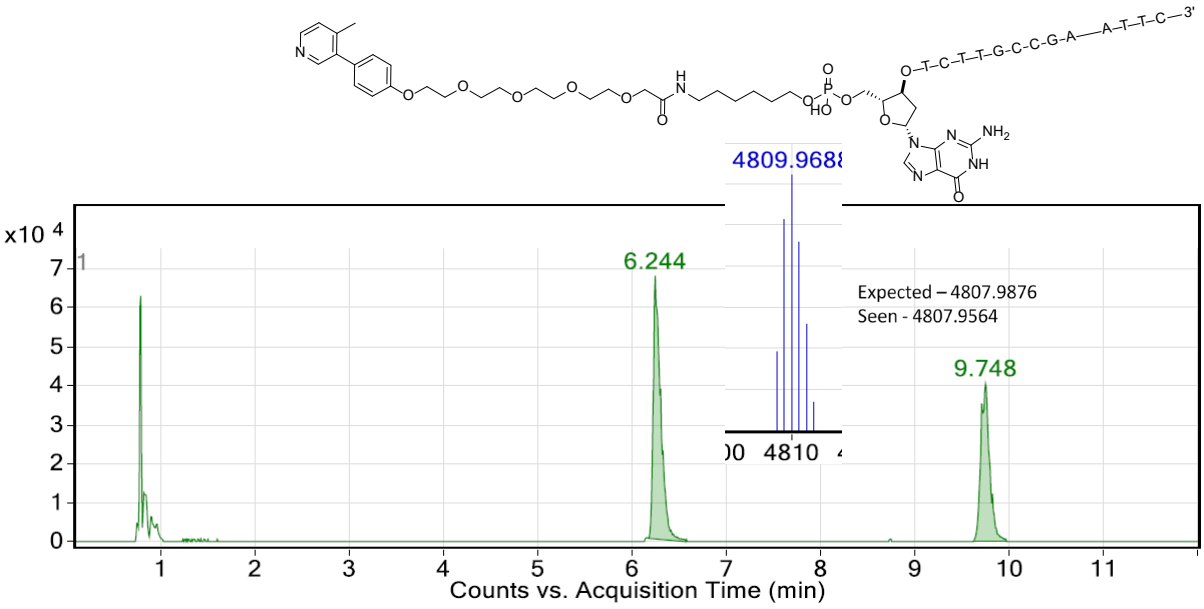
Appendix Figure 13 - Chromatogram of double stranded Suzuki coupling product 51l.



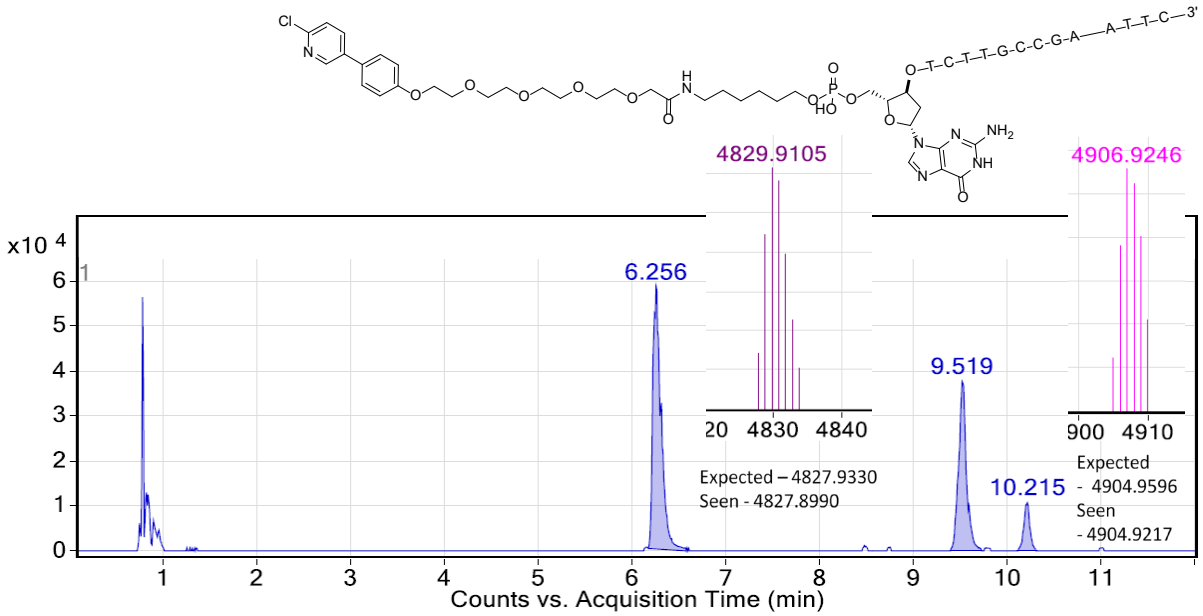
Appendix Figure 14 - Chromatogram of double stranded Suzuki coupling product 51m.



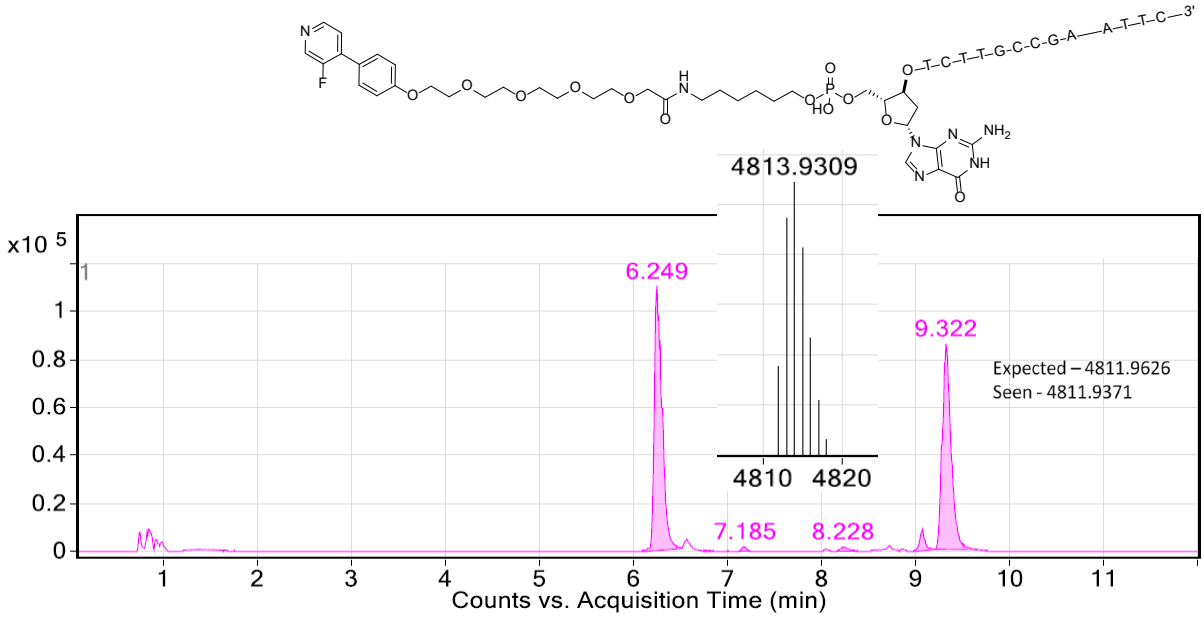
Appendix Figure 15 - Chromatogram of double stranded Suzuki coupling product 51n.



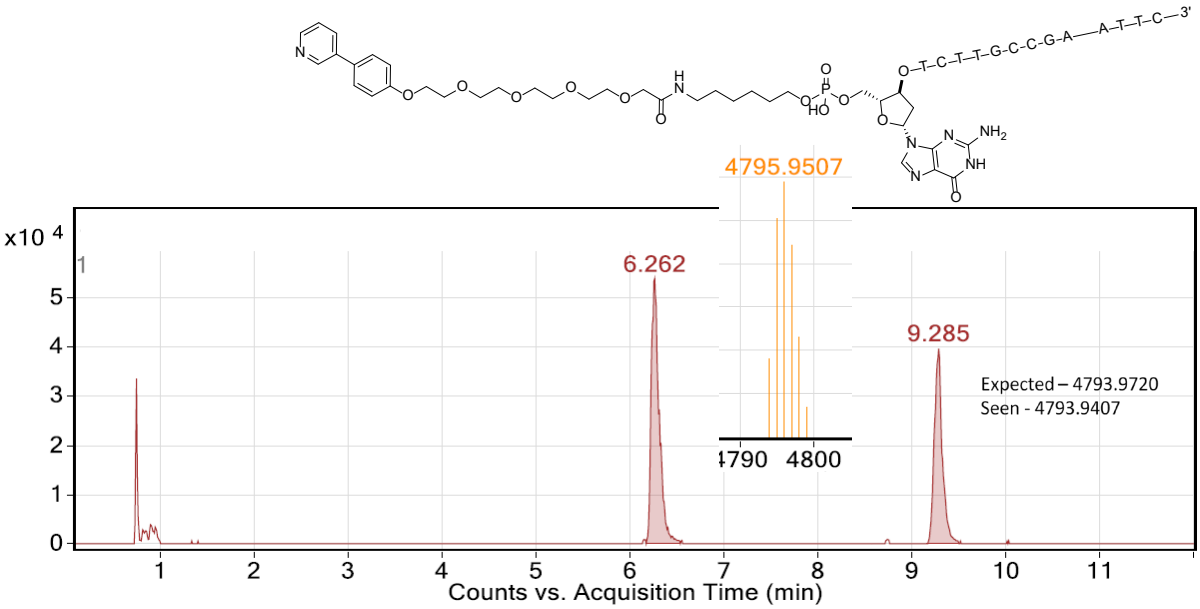
Appendix Figure 16 - Chromatogram of double stranded Suzuki coupling product 51o.



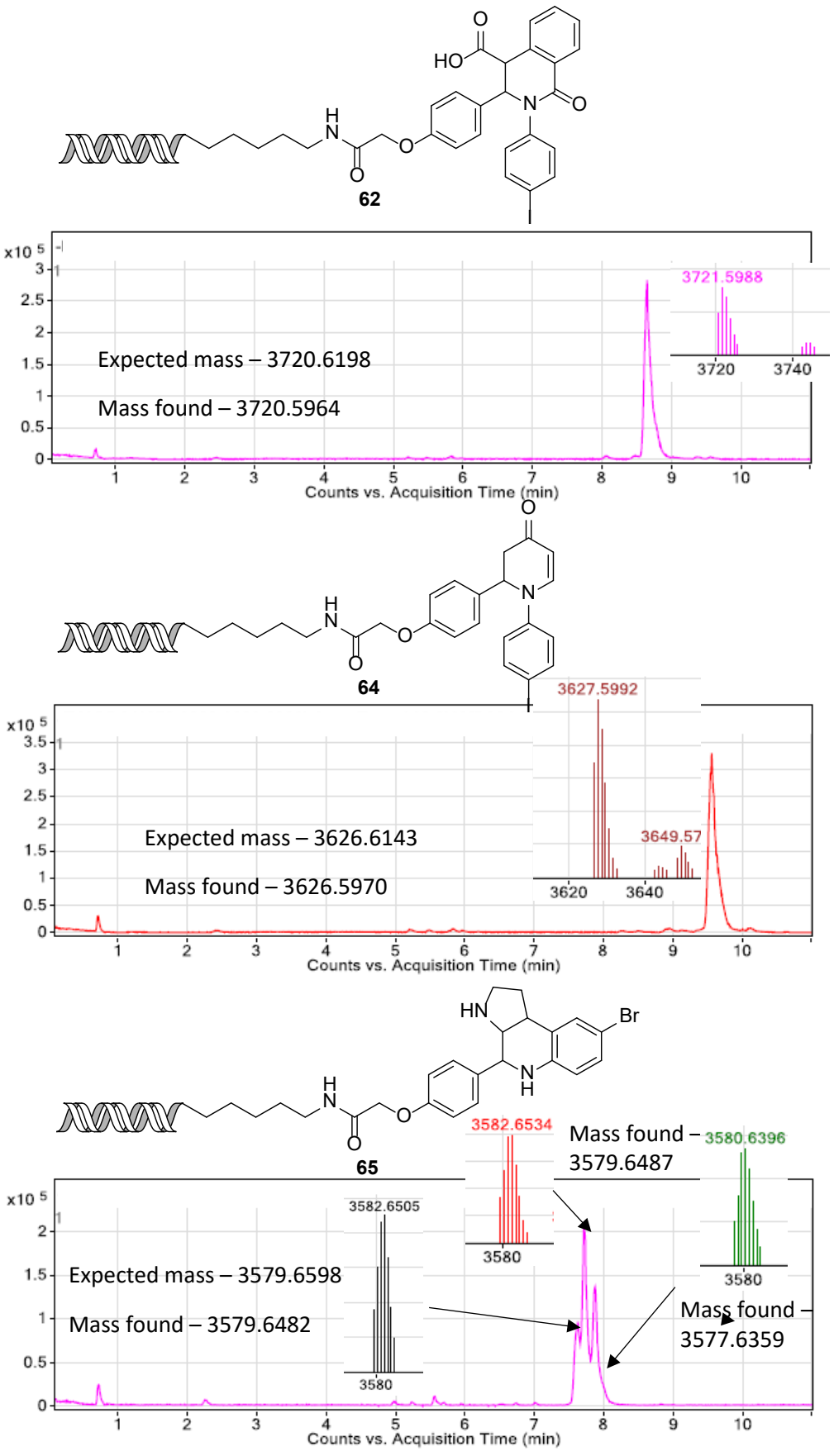
Appendix Figure 17 - Chromatogram of double stranded Suzuki coupling product 51p.



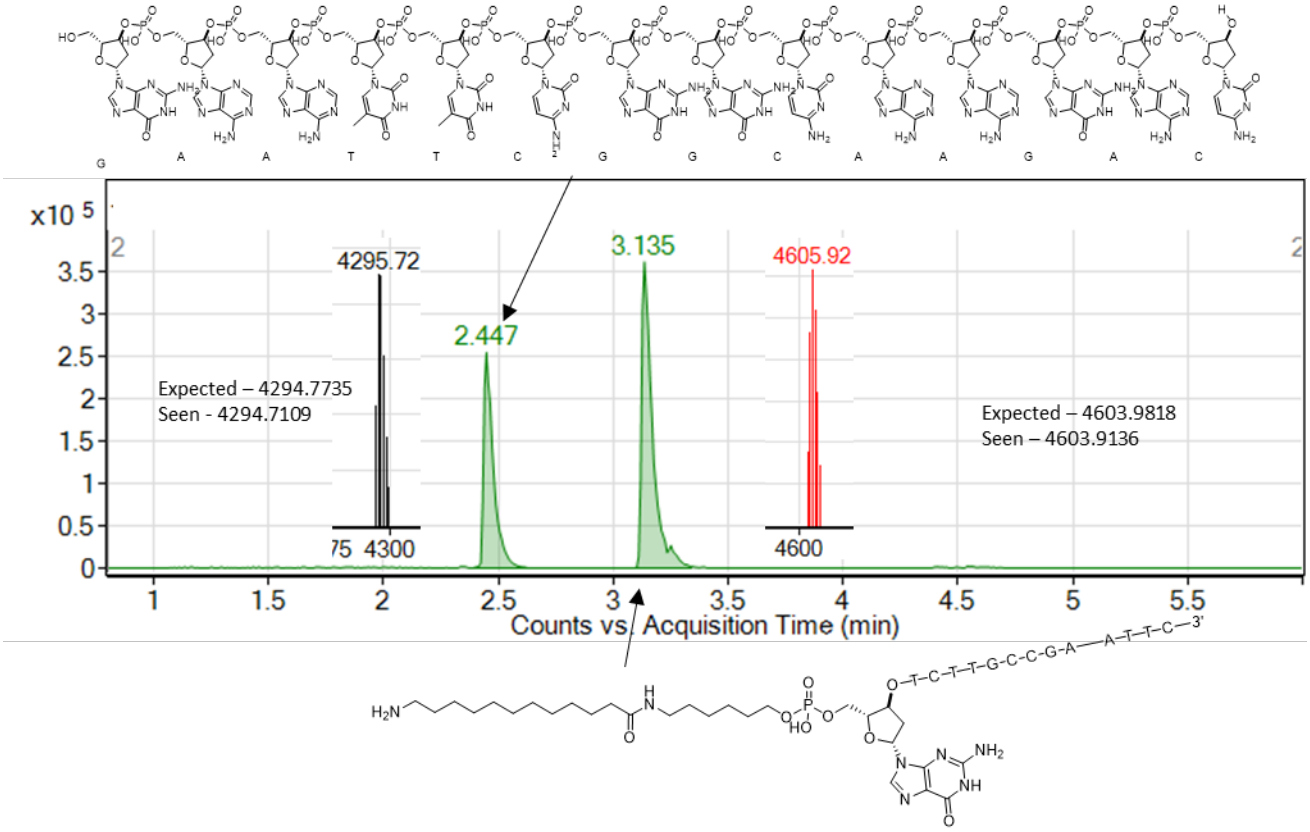
Appendix Figure 18 - Chromatogram of double stranded Suzuki coupling product 51q.



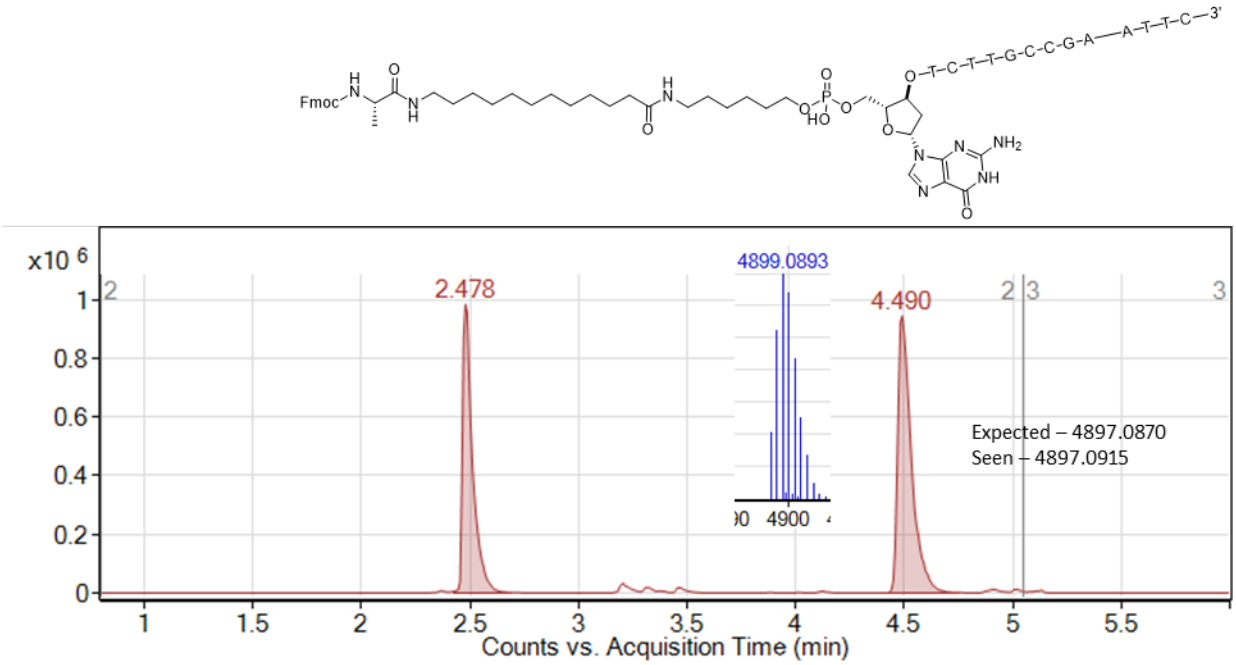
Appendix Figure 19 - Chromatogram of double stranded Suzuki coupling product 51r.



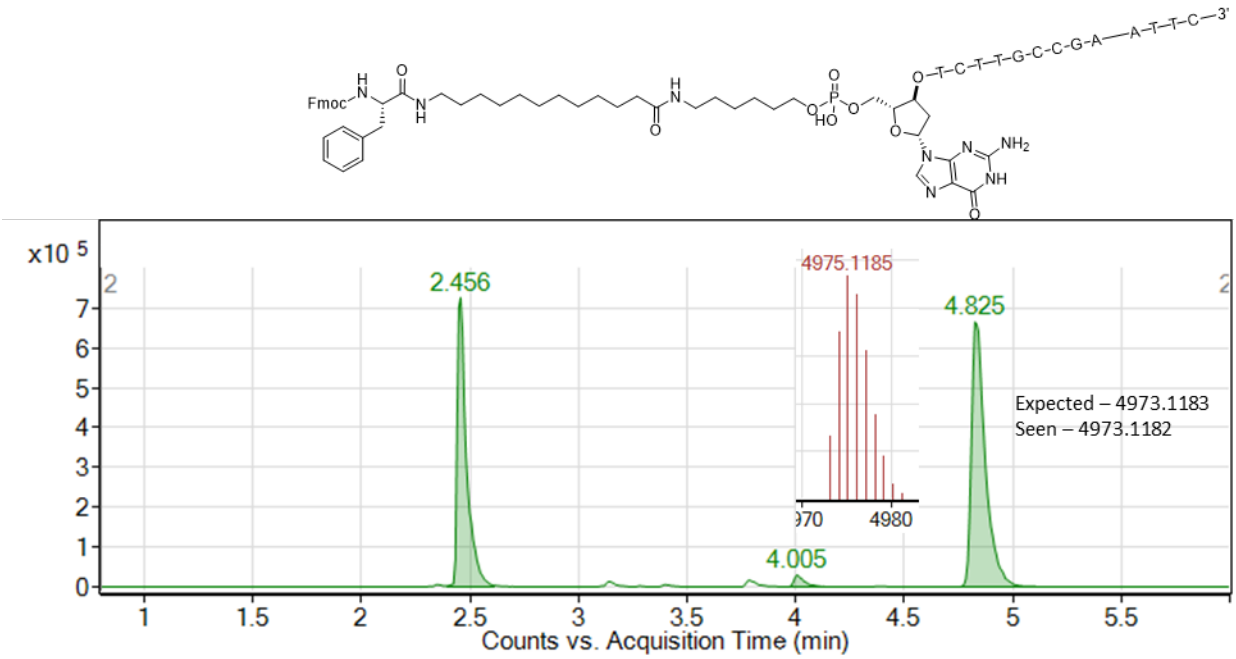
Appendix Figure 20 – Chromatograms of products from multicomponent reactions.



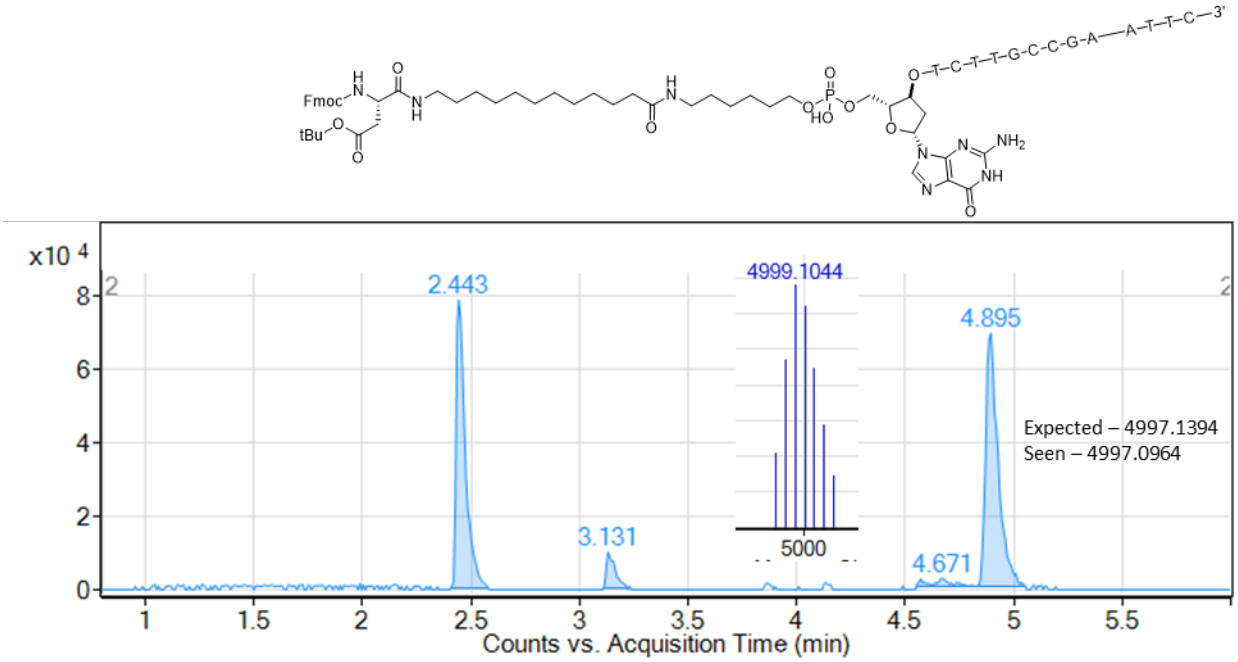
Appendix Figure 21 - Chromatogram of double stranded starting material 74 used Amide couplings.



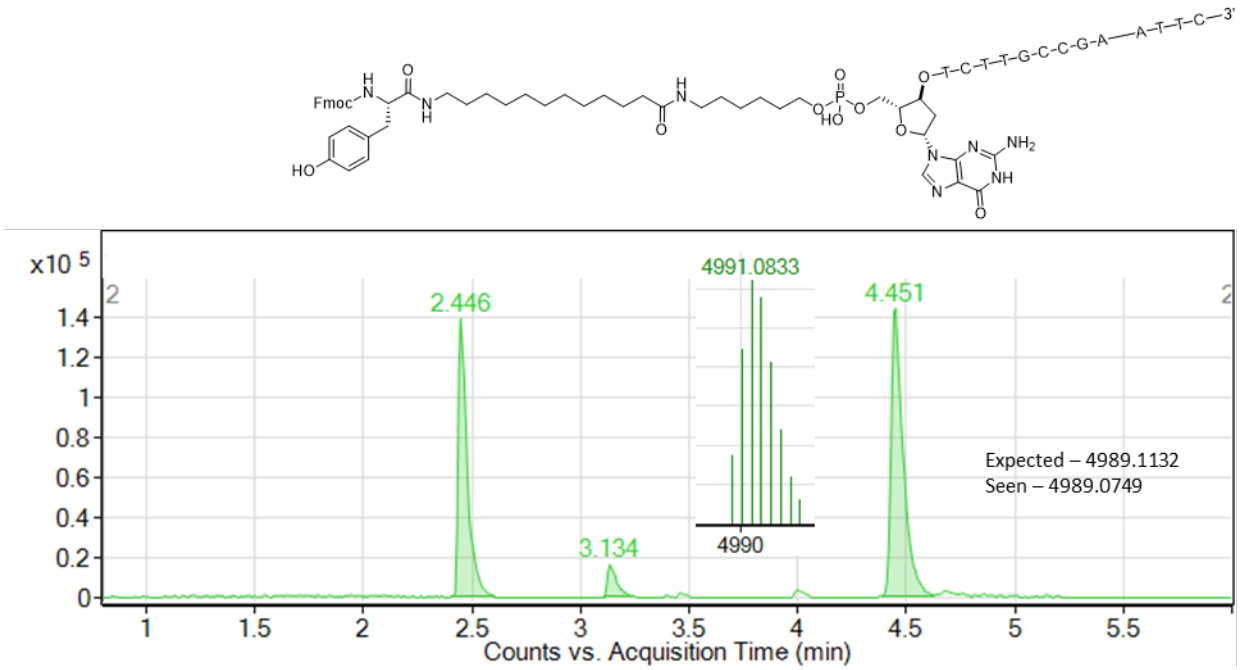
Appendix Figure 22 - Chromatogram of double stranded amide coupling product 75a.



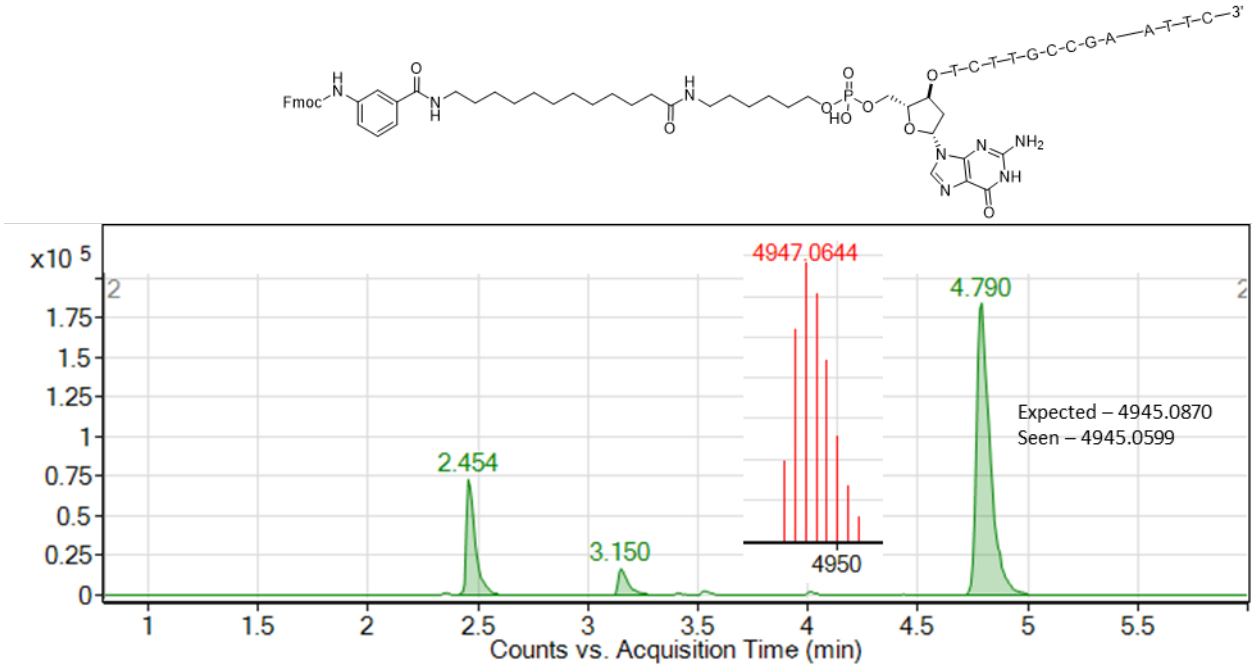
Appendix Figure 23 - Chromatogram of double stranded amide coupling product 75b.



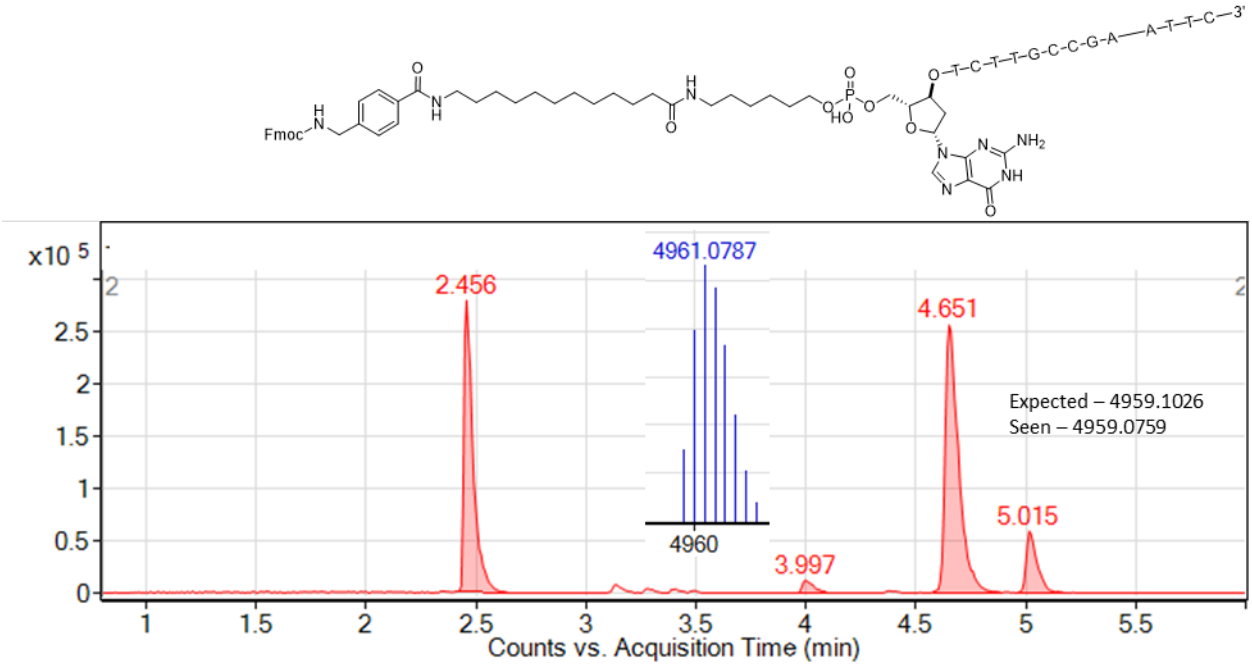
Appendix Figure 24 - Chromatogram of double stranded amide coupling product 75c.



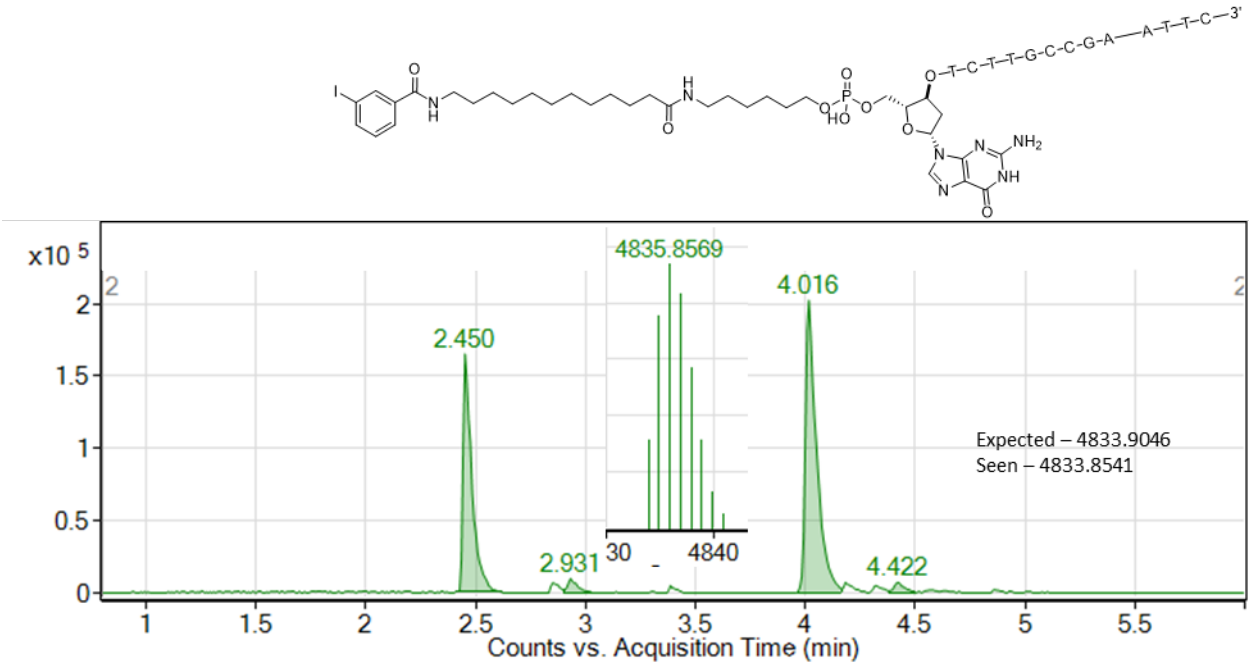
Appendix Figure 25 - Chromatogram of double stranded amide coupling product 75d.



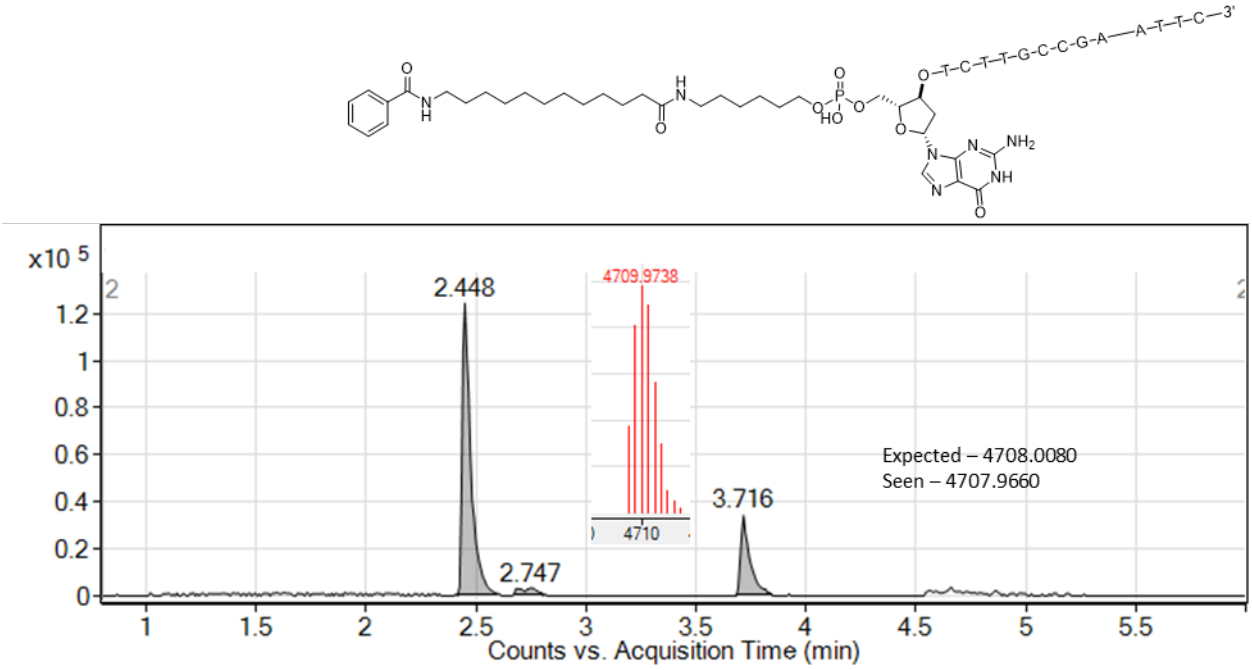
Appendix Figure 26 - Chromatogram of double stranded amide coupling product 75e.



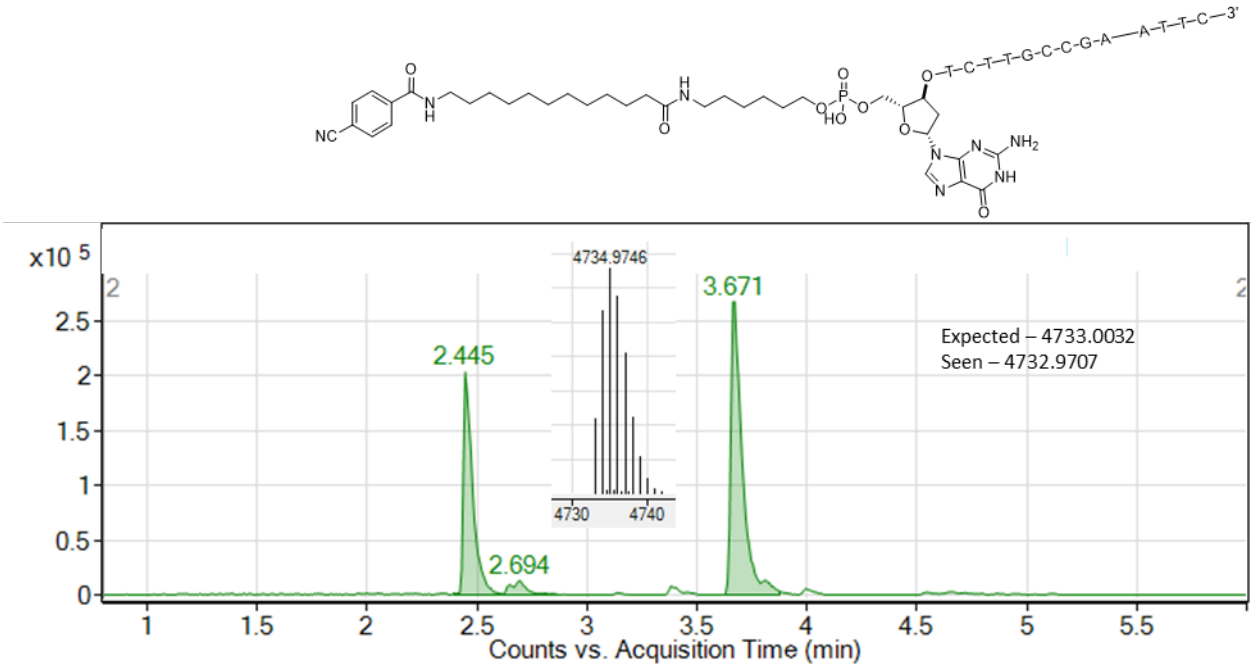
Appendix Figure 27 - Chromatogram of double stranded amide coupling product 75f.



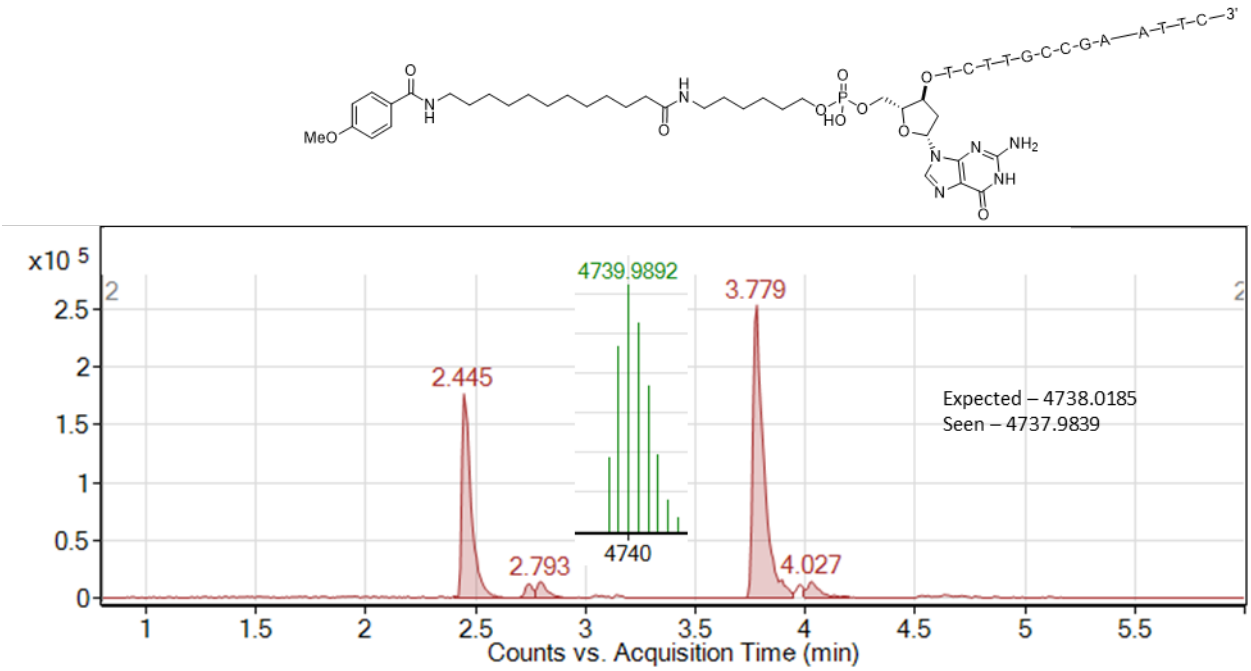
Appendix Figure 28 - Chromatogram of double stranded amide coupling product 75g.



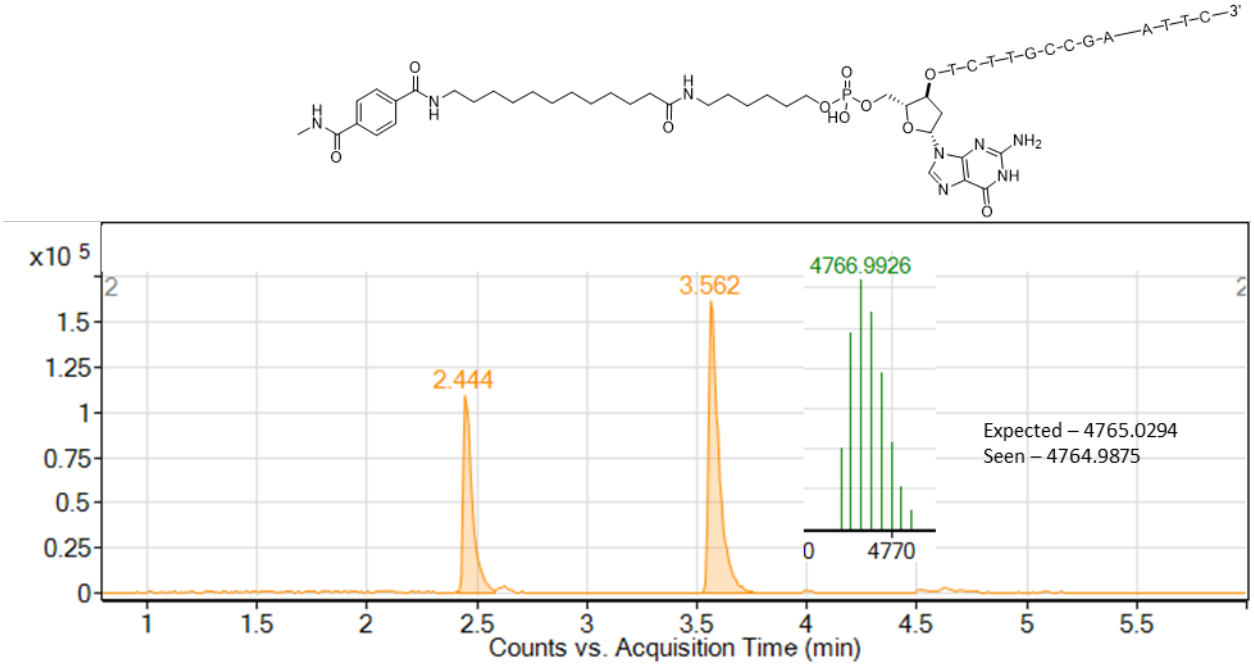
Appendix Figure 29 - Chromatogram of double stranded amide coupling product 75h.



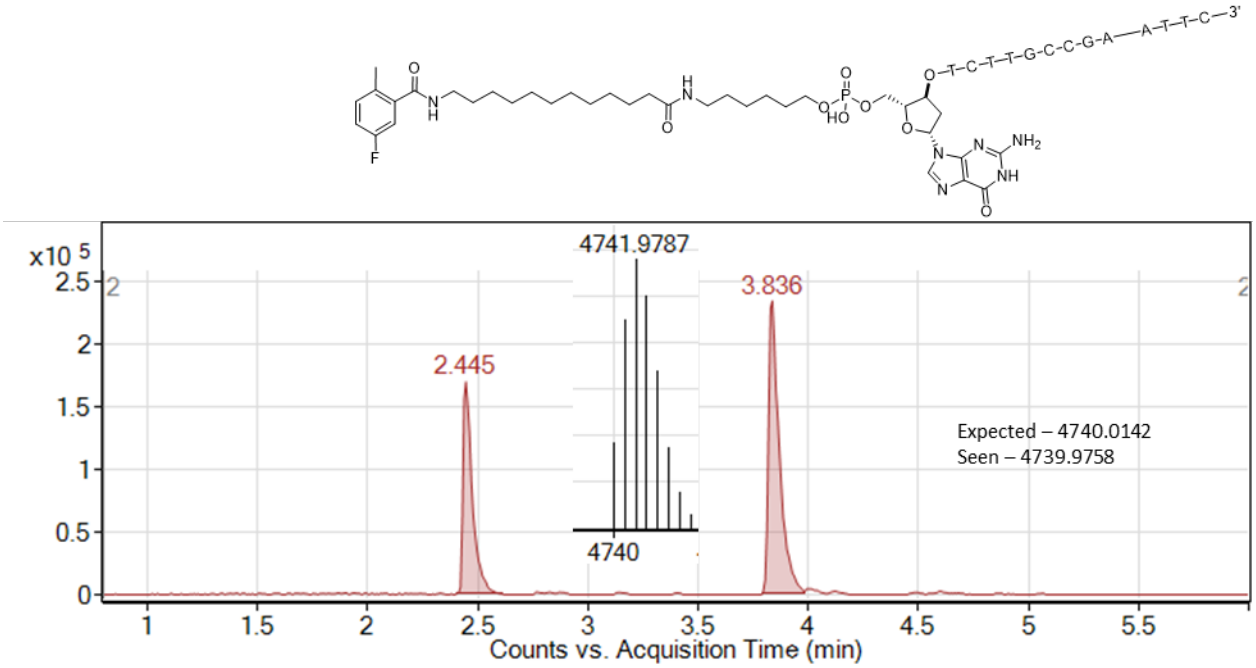
Appendix Figure 30 - Chromatogram of double stranded amide coupling product 75i.



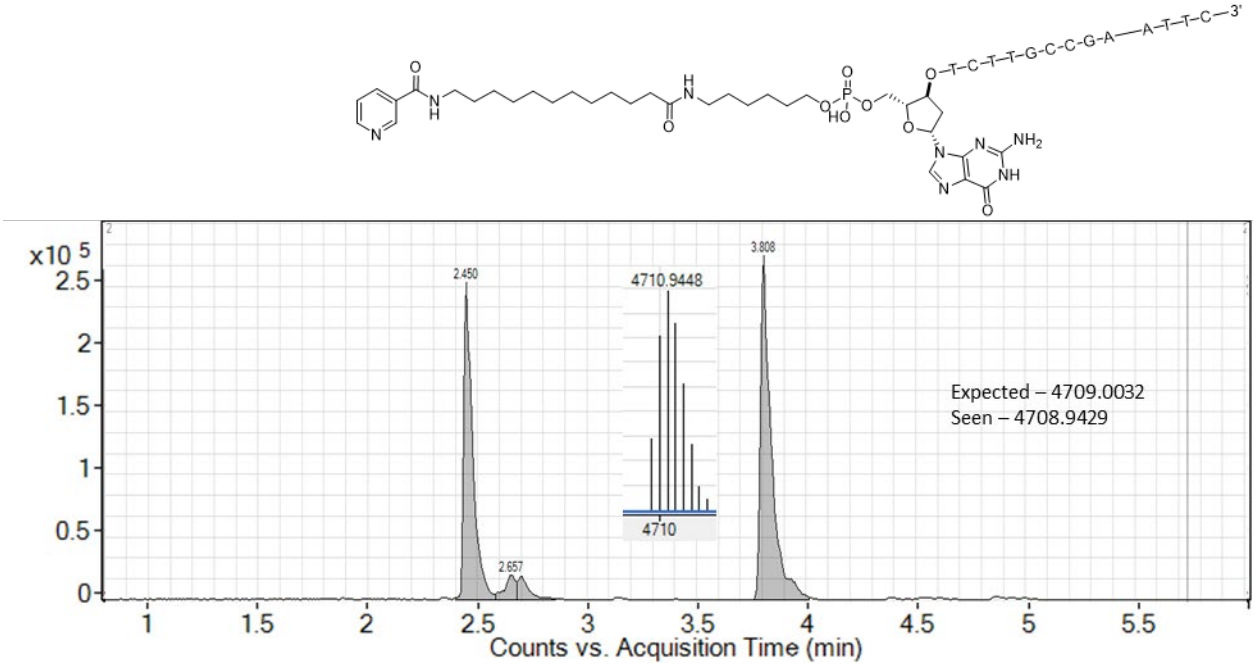
Appendix Figure 31 - Chromatogram of double stranded amide coupling product 75j.



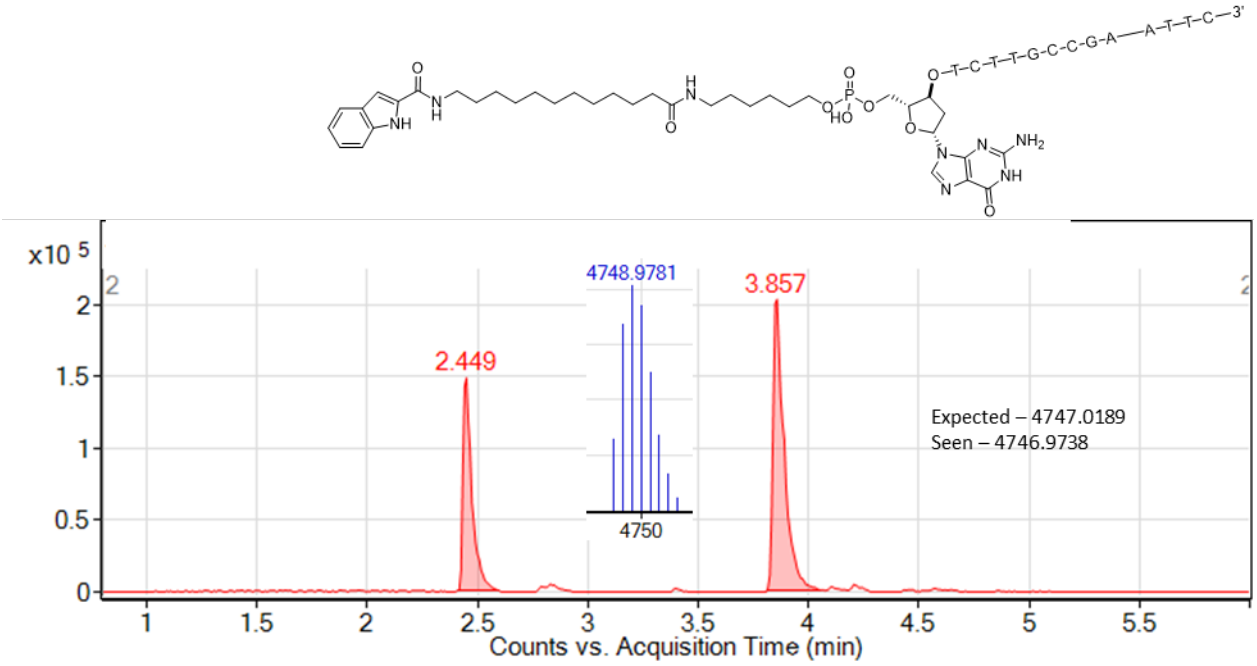
Appendix Figure 32 - Chromatogram of double stranded amide coupling product 75k.



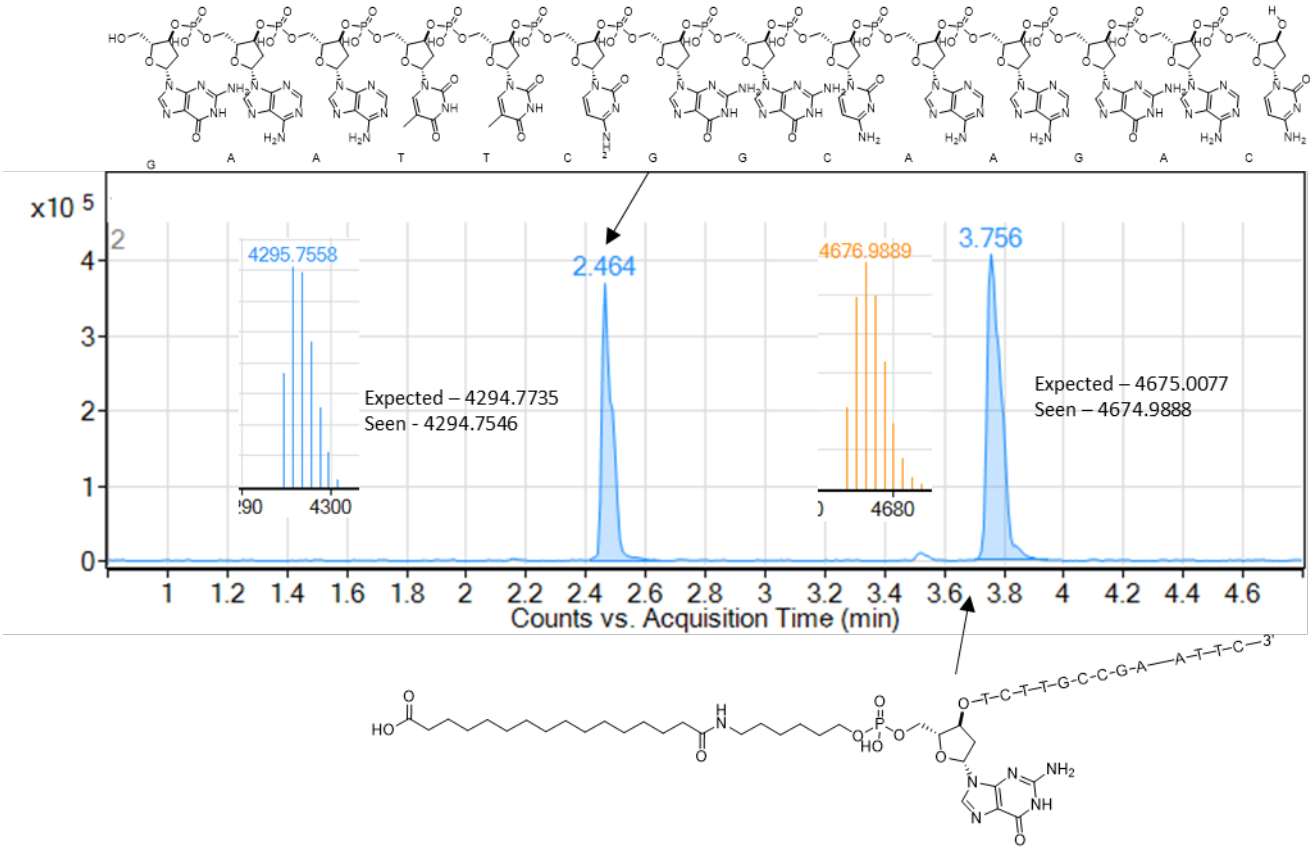
Appendix Figure 33 - Chromatogram of double stranded amide coupling product 75l.



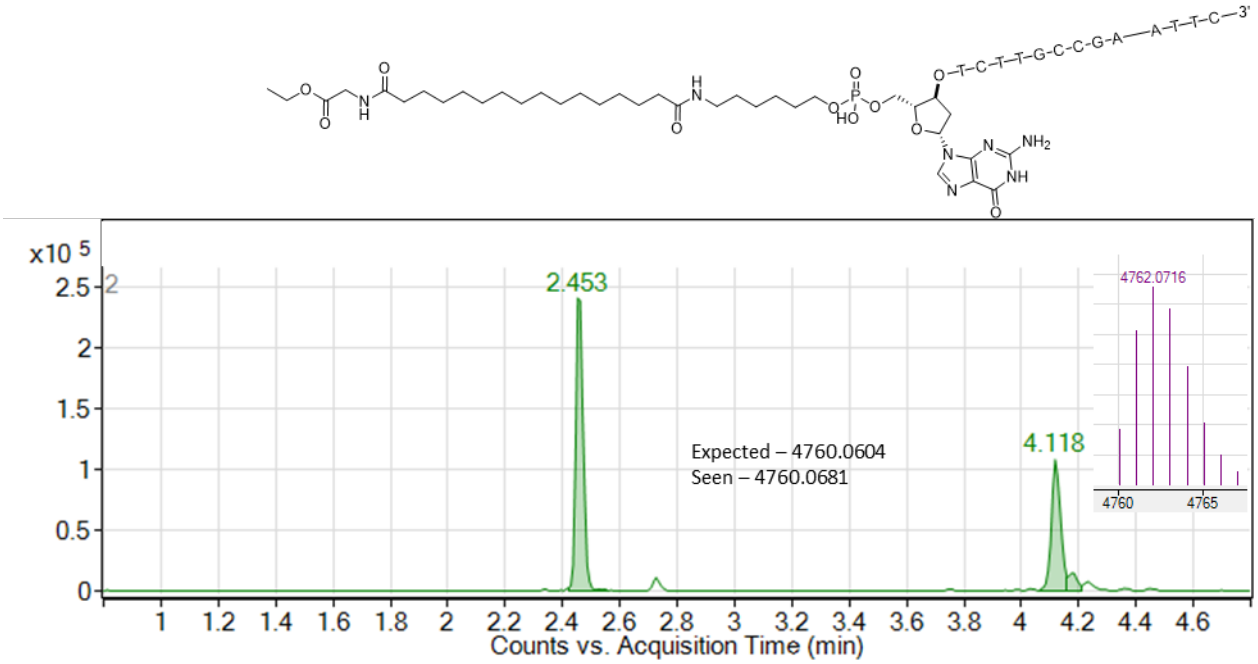
Appendix Figure 34 - Chromatogram of double stranded amide coupling product 75m.



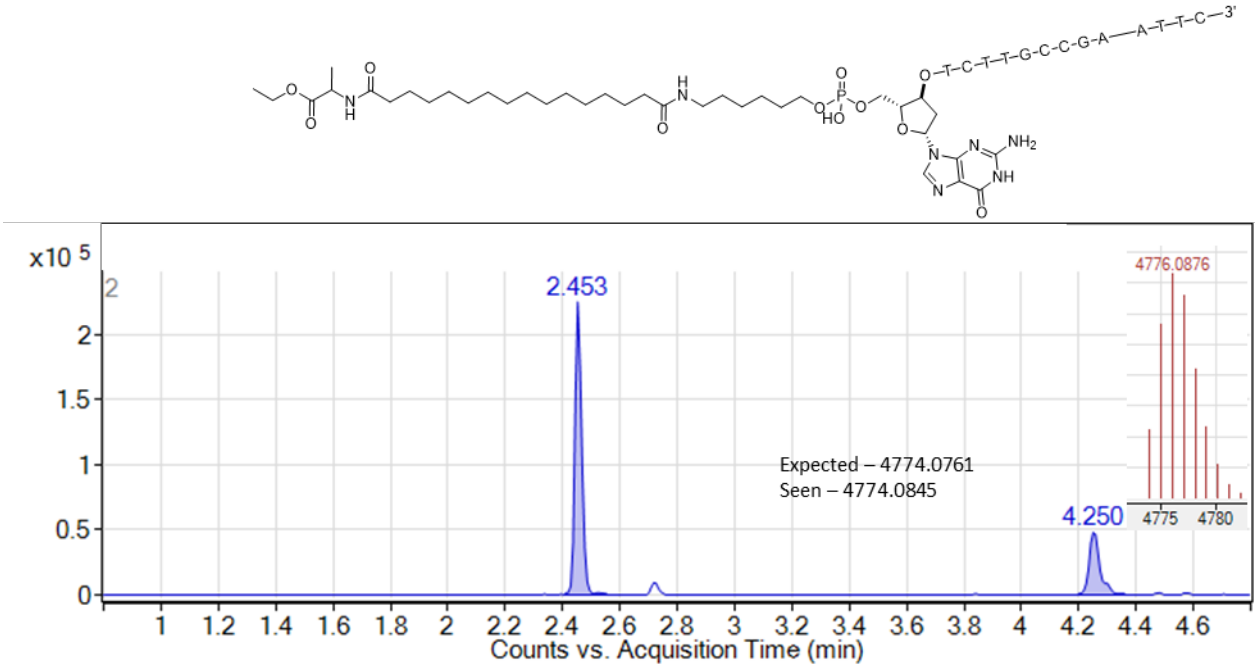
Appendix Figure 35 - Chromatogram of double stranded amide coupling product 75n.



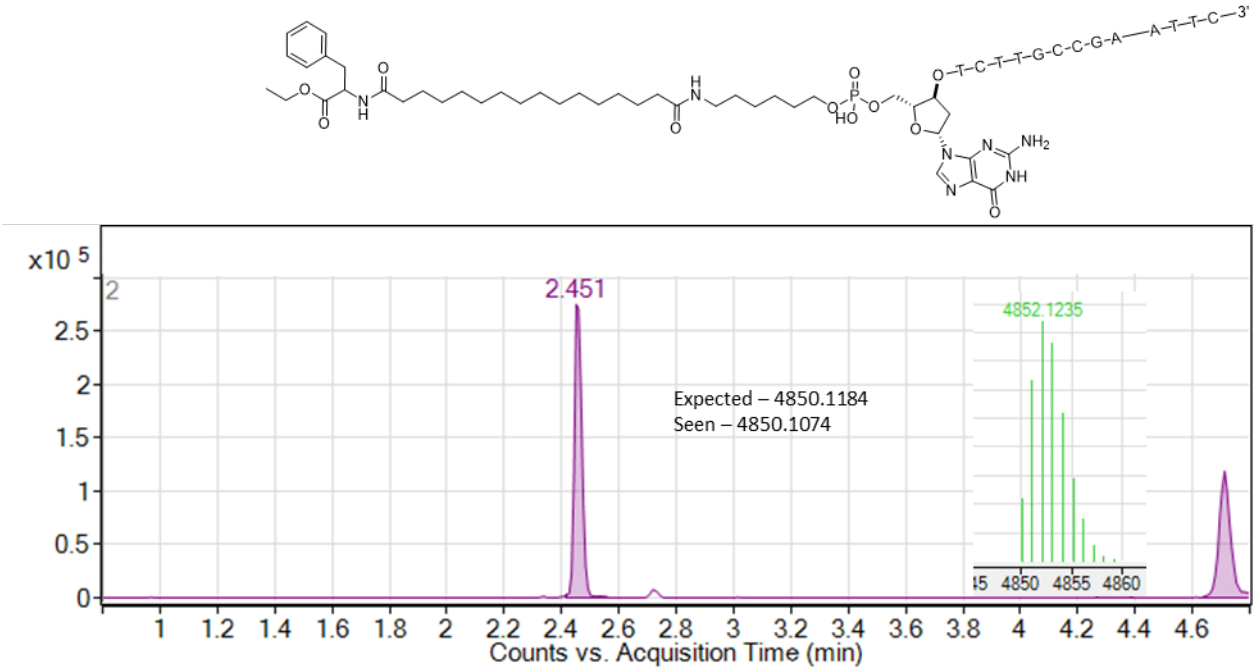
Appendix Figure 36 - Chromatogram of double stranded starting material 71 used Amide couplings.



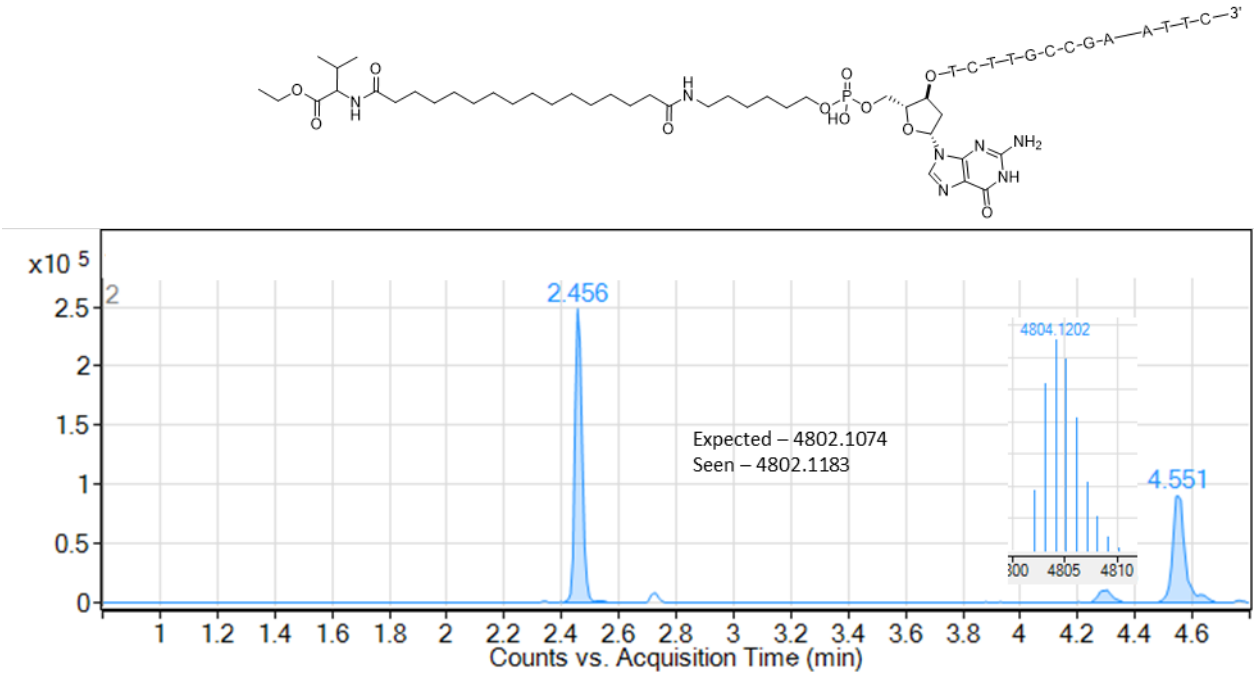
Appendix Figure 37 - Chromatogram of double stranded reverse amide coupling product 72a.



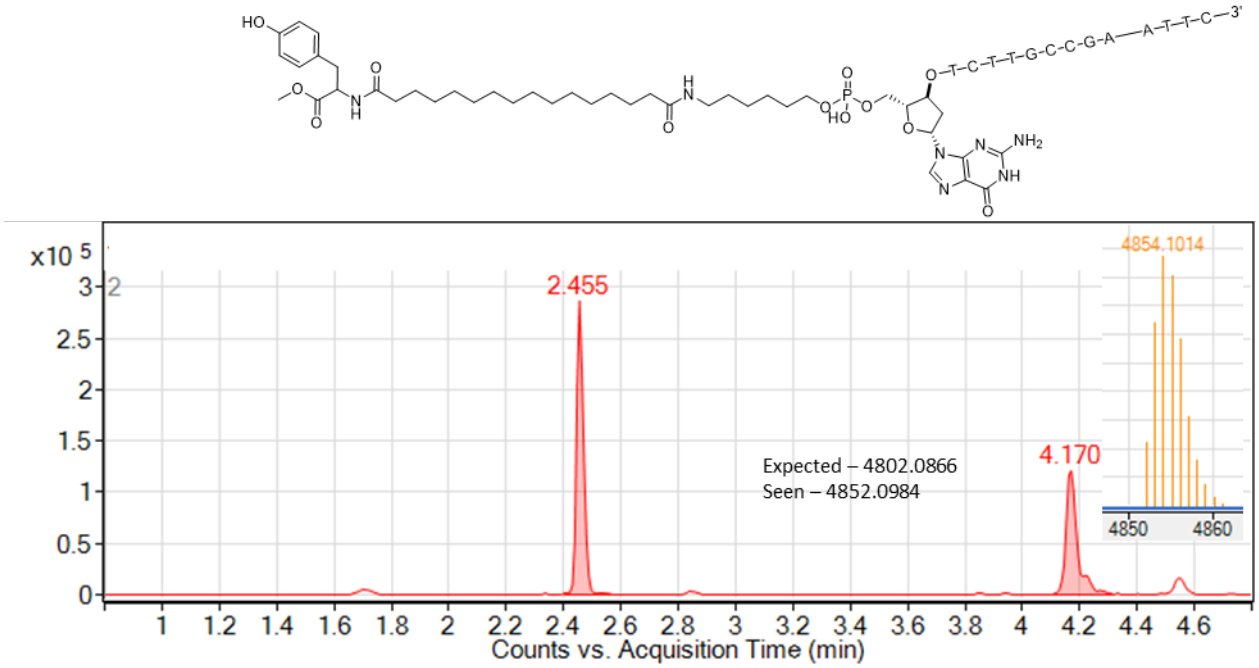
Appendix Figure 38 - Chromatogram of double stranded reverse amide coupling product 72b.



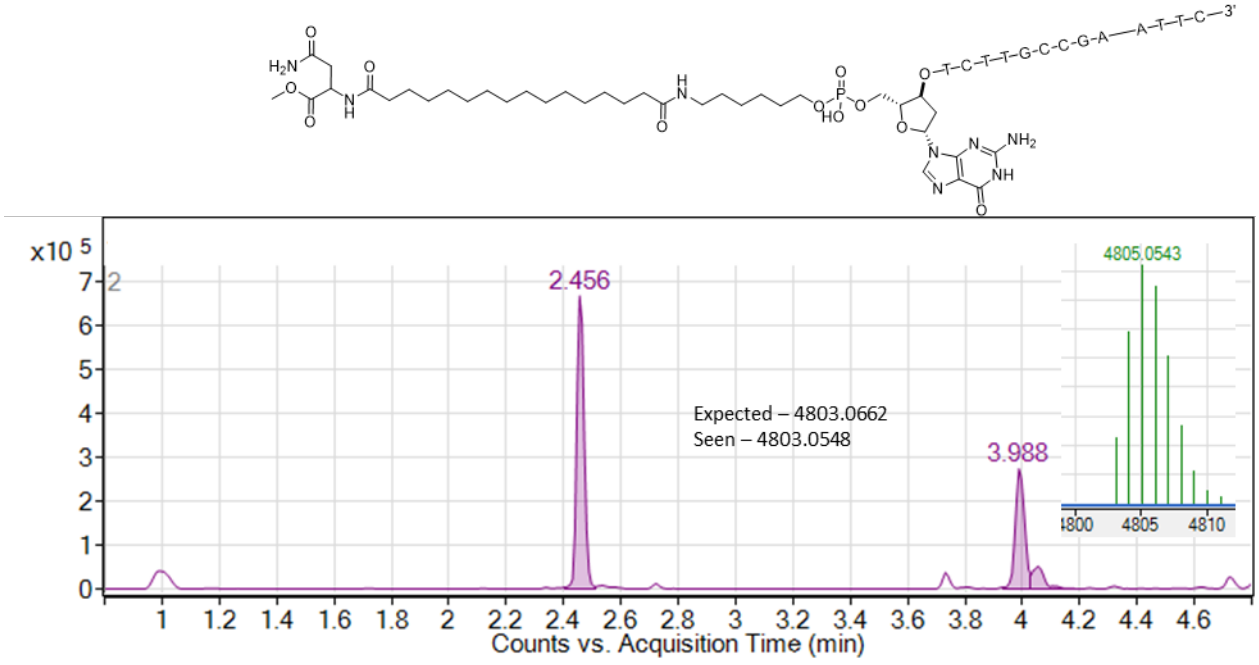
Appendix Figure 39 - Chromatogram of double stranded reverse amide coupling product 72c.



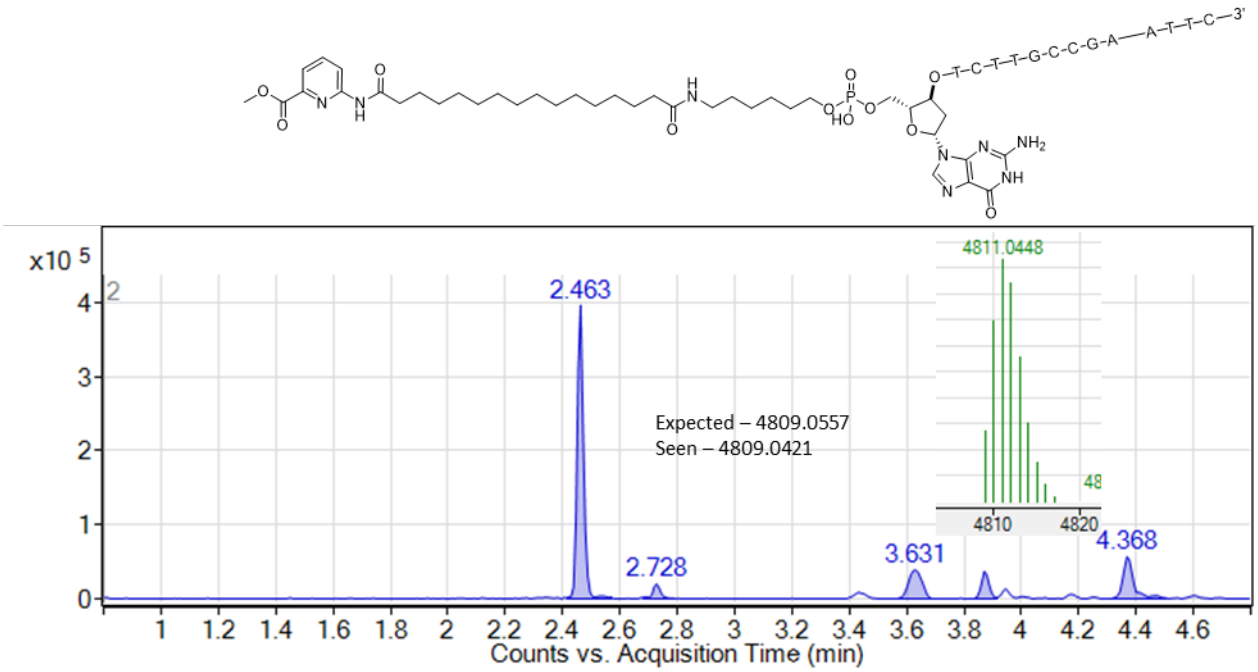
Appendix Figure 40 - Chromatogram of double stranded reverse amide coupling product 72d.



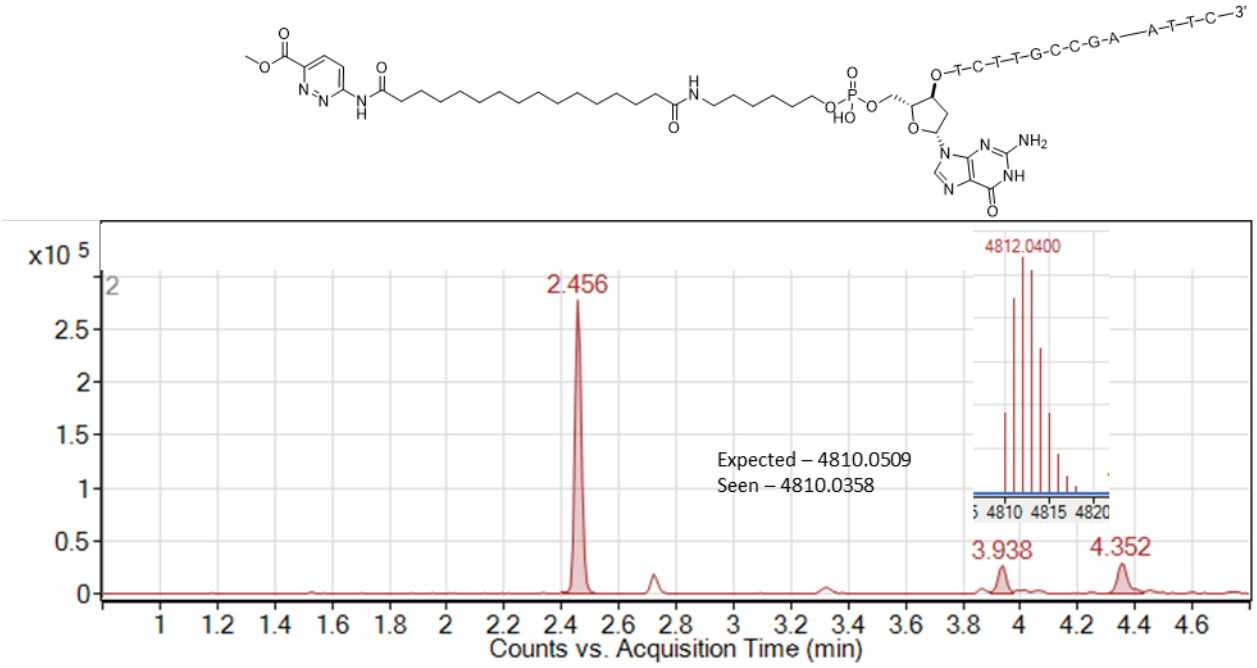
Appendix Figure 41 - Chromatogram of double stranded reverse amide coupling product 72e.



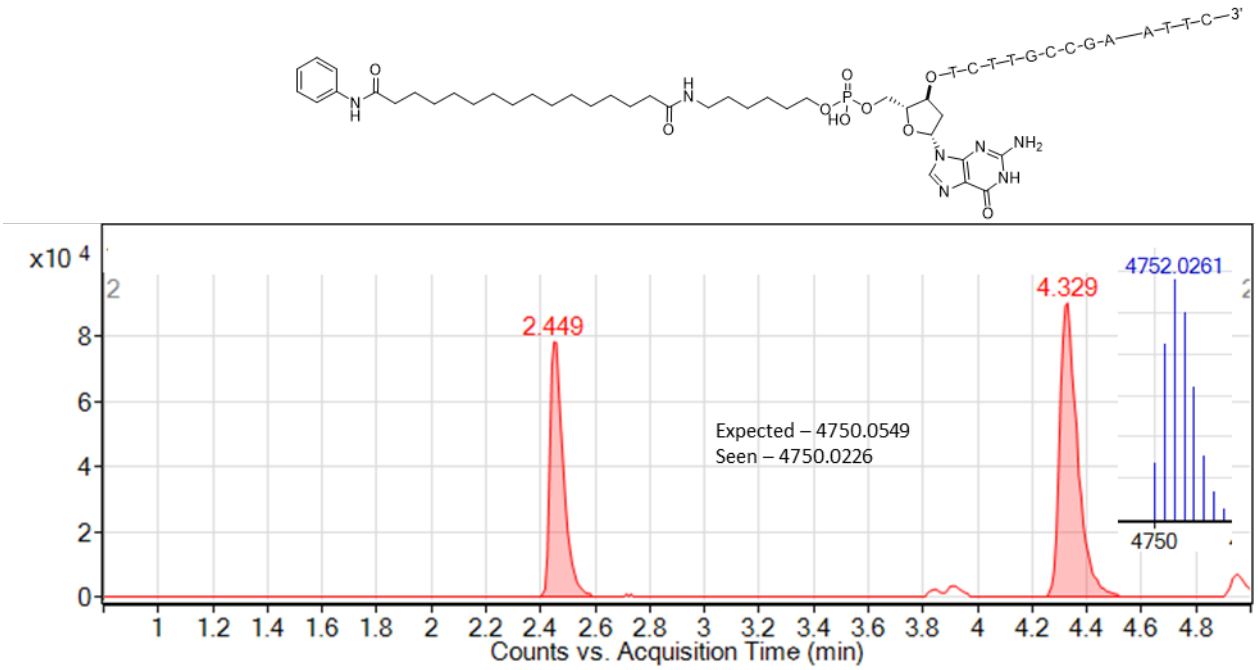
Appendix Figure 42 - Chromatogram of double stranded reverse amide coupling product 72f (shoulder same mass as product).



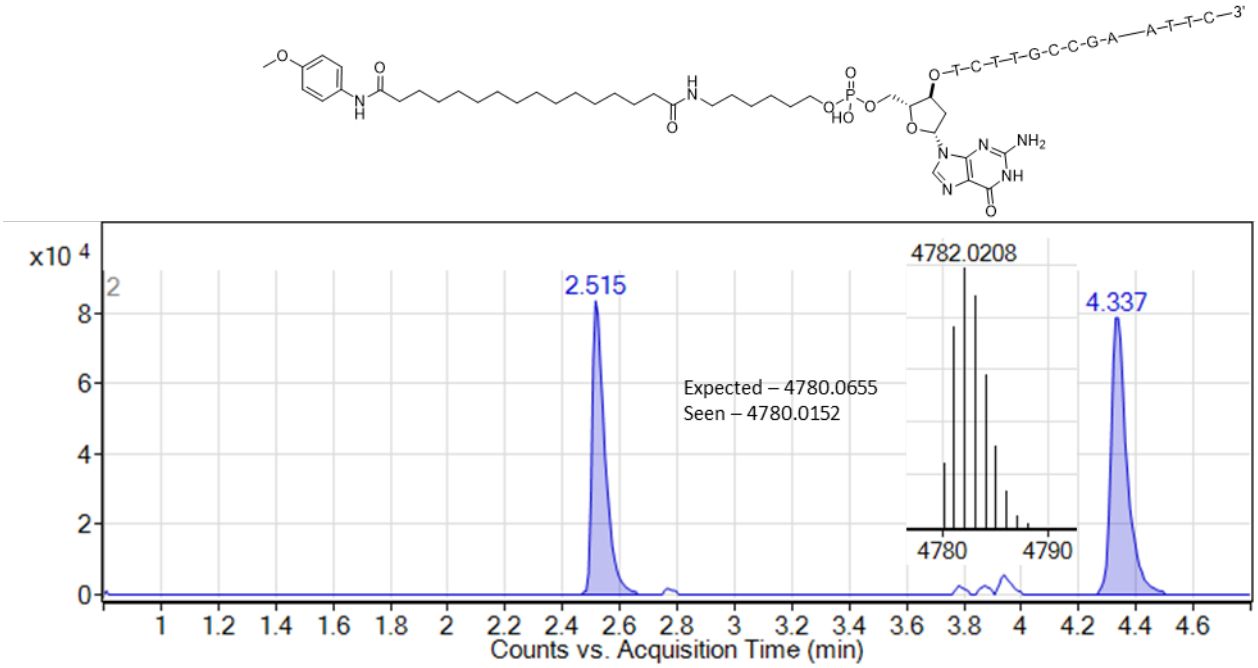
Appendix Figure 43 - Chromatogram of double stranded reverse amide coupling product 72g.



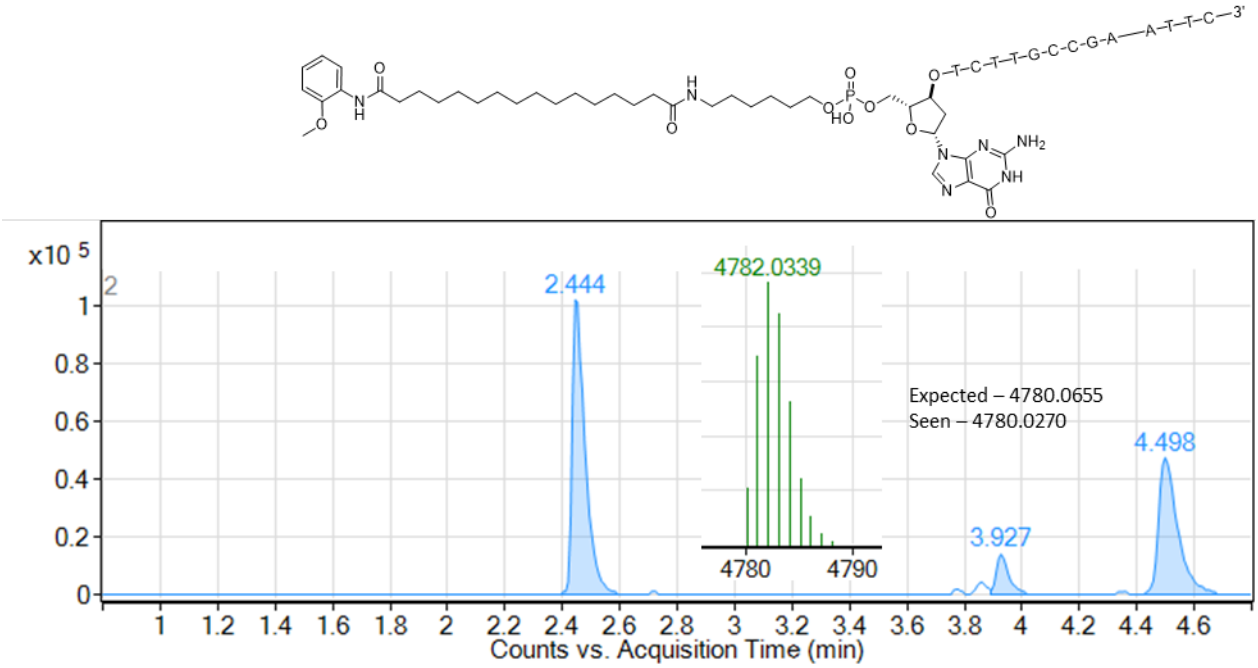
Appendix Figure 44 - Chromatogram of double stranded reverse amide coupling product 72h.



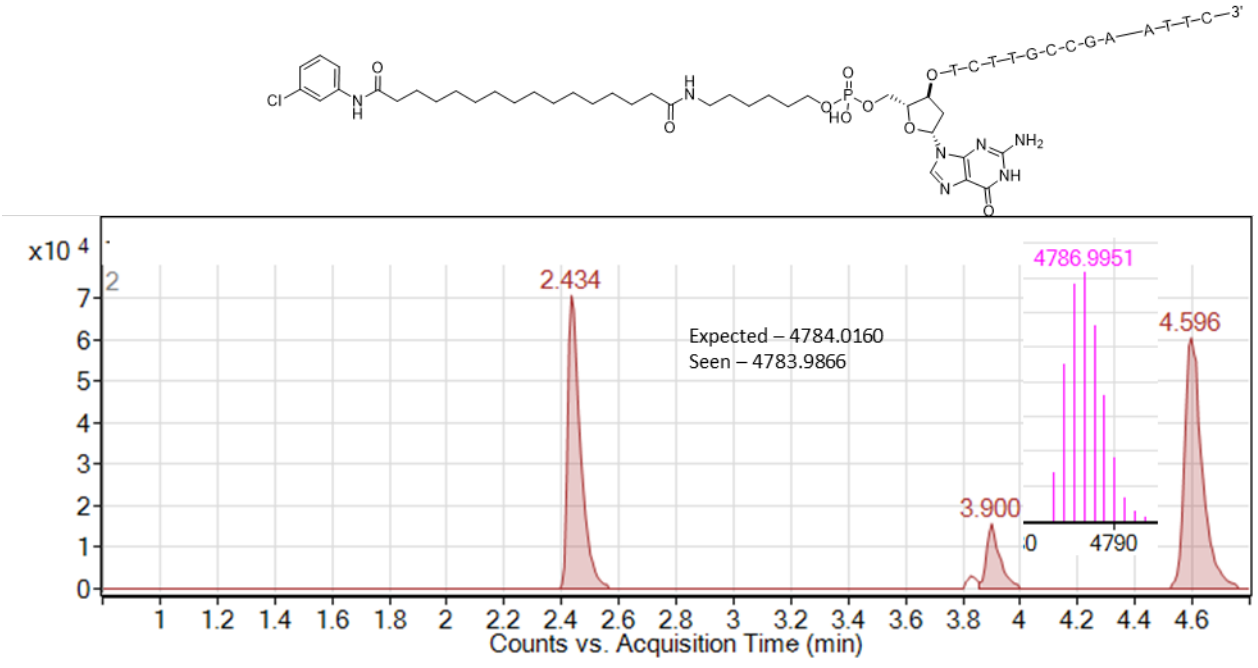
Appendix Figure 45 - Chromatogram of double stranded reverse amide coupling product 72i.



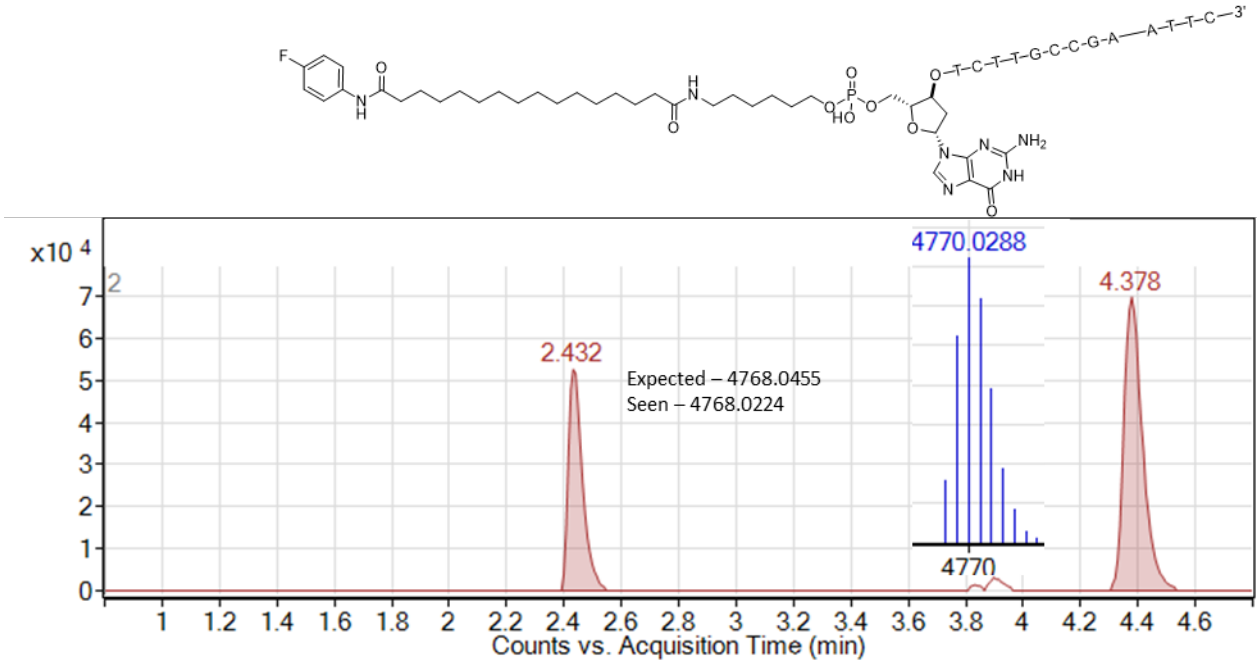
Appendix Figure 46 - Chromatogram of double stranded reverse amide coupling product 72j.



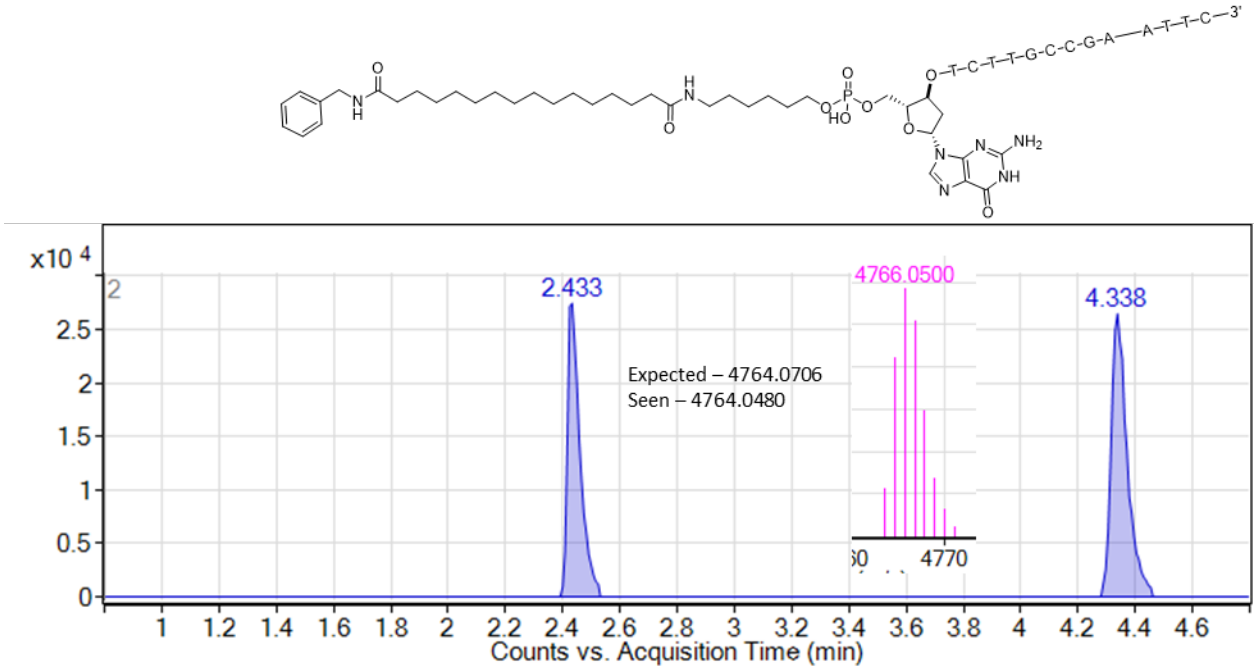
Appendix Figure 47 - Chromatogram of double stranded reverse amide coupling product 72k.



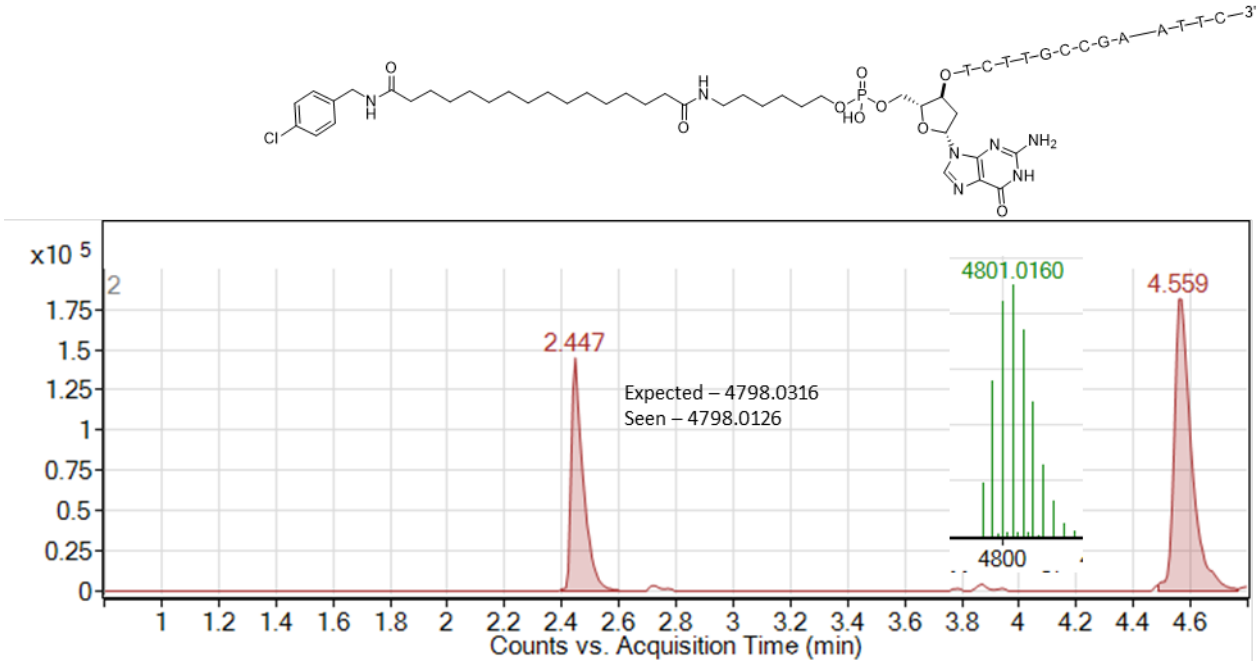
Appendix Figure 48- Chromatogram of double stranded reverse amide coupling product 72l.



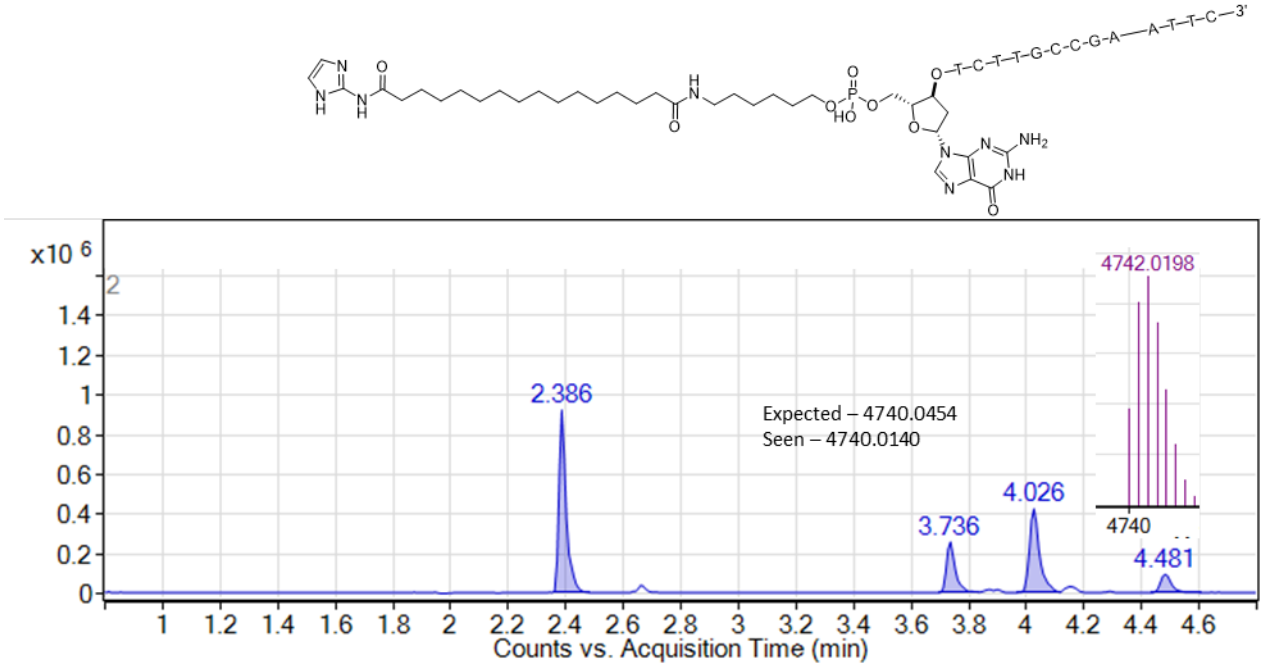
Appendix Figure 49 - Chromatogram of double stranded reverse amide coupling product 72m.



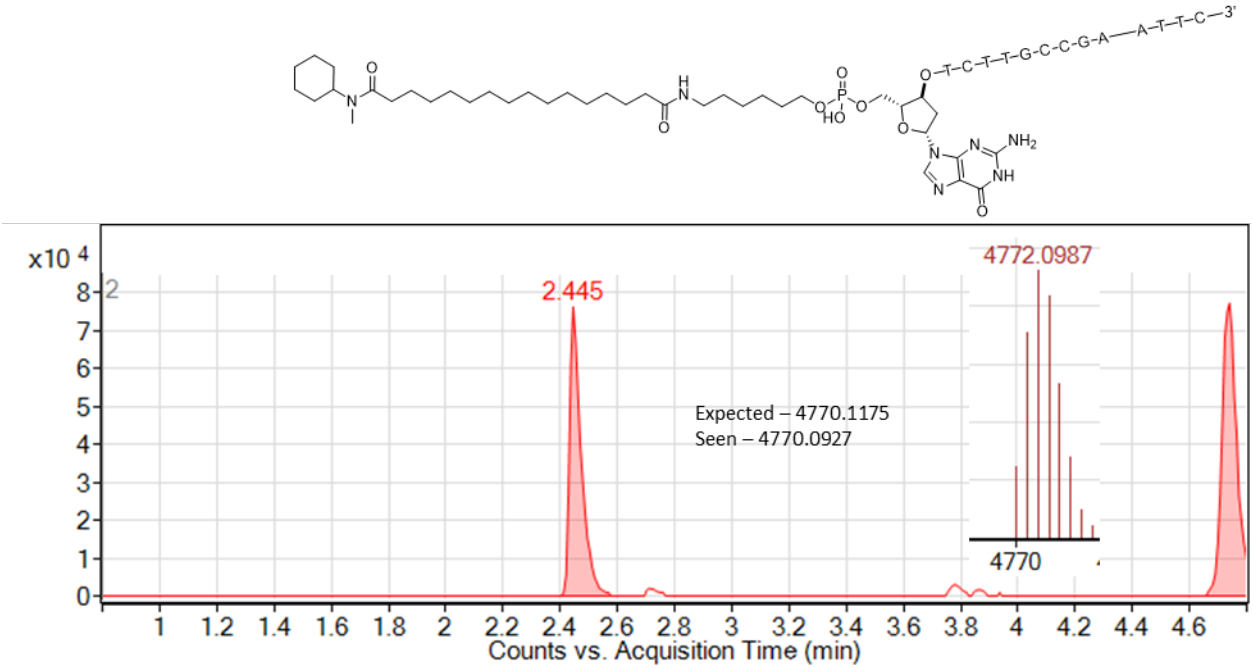
Appendix Figure 50 - Chromatogram of double stranded reverse amide coupling product 72n.



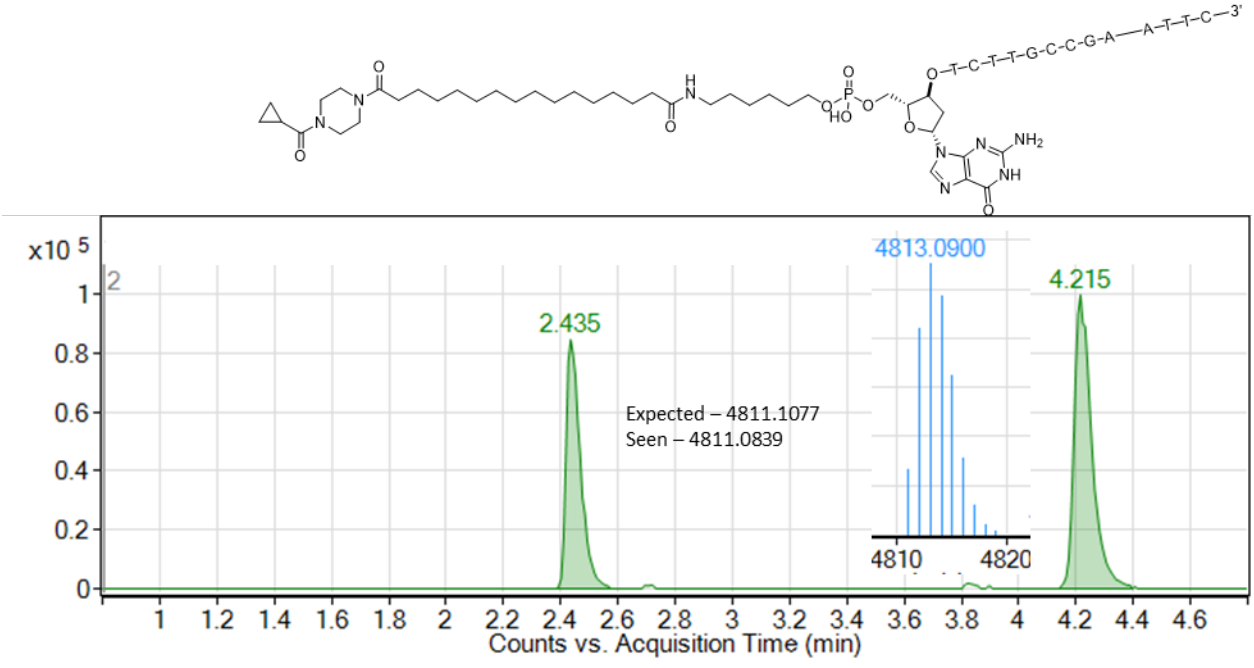
Appendix Figure 51 - Chromatogram of double stranded reverse amide coupling product 72o.



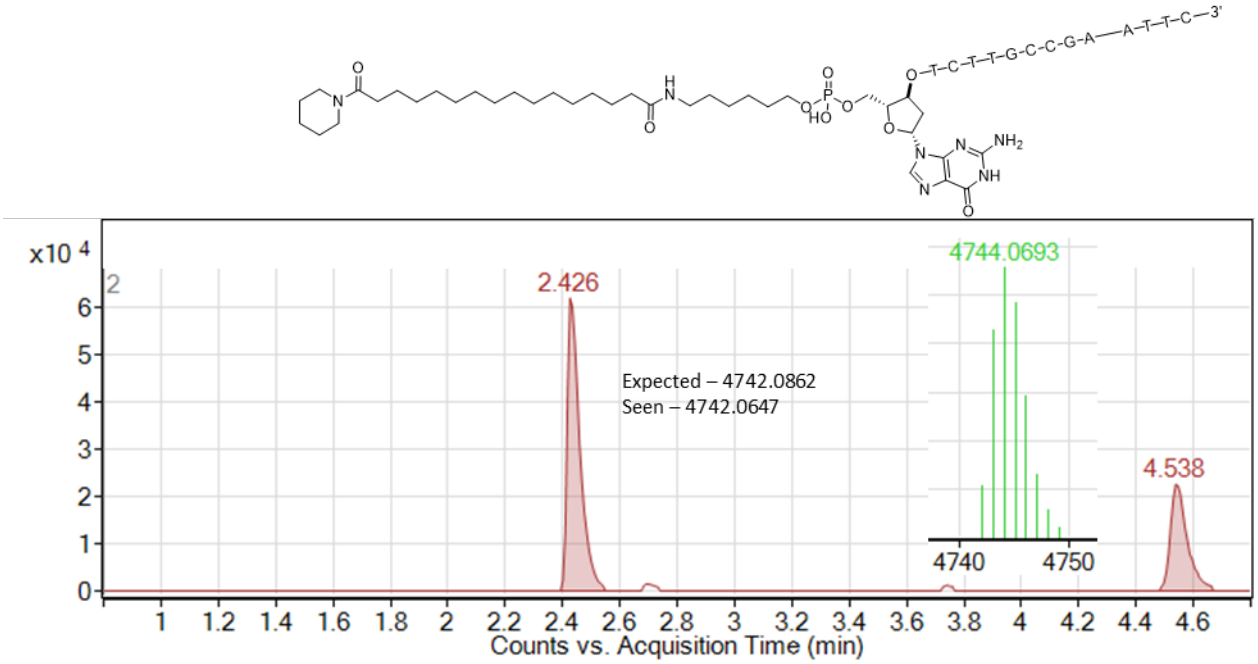
Appendix Figure 52 - Chromatogram of double stranded reverse amide coupling product 72p.



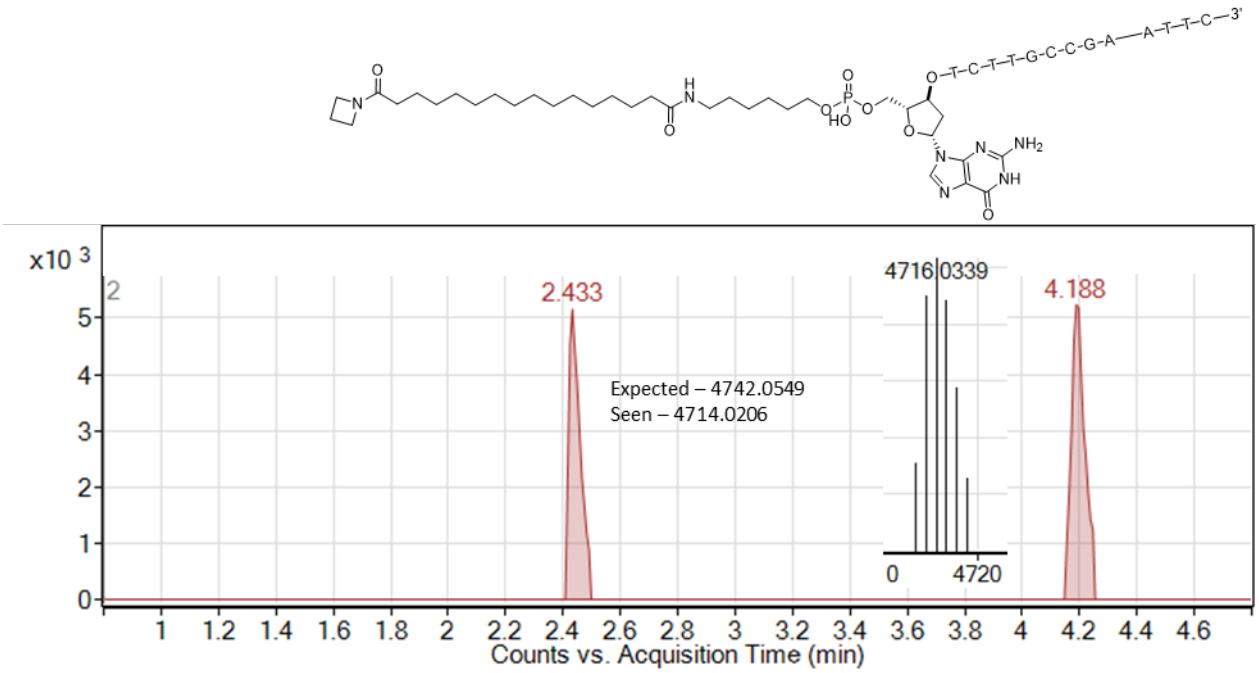
Appendix Figure 53 - Chromatogram of double stranded reverse amide coupling product 72q.



Appendix Figure 54 - Chromatogram of double stranded reverse amide coupling product 72r.



Appendix Figure 55 - Chromatogram of double stranded reverse amide coupling product 72s.



Appendix Figure 56 - Chromatogram of double stranded reverse amide coupling product 72t.

Appendix Table 3 – All amino esters, the consequent reaction wells and BB1 codes, as well as corresponding building block code sequences and overhangs.

Amino ester (smiles)	Reaction well	BB1 code	OH1BB1-x sequence (5'-3')	OH2'BB1-x' sequence (5'-3')
methyl 1-aminocyclopentane-1-carboxylate hydrochloride	A1	BB1-1	GTATTTATAACG	TAGGCGTTATAA
methyl 1-aminocyclopropane-1-carboxylate hydrochloride	A2	BB1-2	GTATGGAGAAGA	TAGGTCTTCTCC
methyl 2-amino-2-methylpropanoate hydrochloride	A3	BB1-3	GTATCCACAATG	TAGGCATTGTGG
methyl D-tryptophanate hydrochloride	A4	BB1-4	GTATATATACAC	TAGGGTGTATAT
methyl 1-aminocyclobutane-1-carboxylate hydrochloride	A5	BB1-5	GTATTGAGACCT	TAGGAGGTCTCA
methyl asparaginate hydrochloride	A6	BB1-6	GTATGCACACGC	TAGGGCGTGTGC
methyl (1r,4r)-4-aminocyclohexane-1-carboxylate hydrochloride	A7	BB1-7	GTATCAAAACTT	TAGGAAGTTTTG
methyl D-histidinate dihydrochloride	A8	BB1-8	GTATAGAGAGAG	TAGGCTCTCTCT
methyl 6-aminopyridazine-3-carboxylate	B1	BB1-9	GTATTCACAGCA	TAGGTGCTGTGA
methyl tyrosinate hydrochloride	B2	BB1-10	GTATGAAAAGGG	TAGGCCCTTTTC
methyl 4-aminotetrahydro-2H-pyran-4-carboxylate hydrochloride	B3	BB1-11	GTATCTATAGTA	TAGGTACTATAG
methyl 2-amino-2-(thiophen-2-yl)acetate hydrochloride	B4	BB1-12	GTATACACATAT	TAGGATATGTGT
methyl D-valinate hydrochloride	B5	BB1-13	GTATTAAAATCC	TAGGGGATTTTA
methyl morpholine-2-carboxylate hydrochloride	B6	BB1-14	GTATGTATATGT	TAGGACATATAC
methyl 5-hydroxypiperidine-3-carboxylate hydrochloride	B7	BB1-15	GTATCGAGATTC	TAGGGAATCTCG

methyl (2S,4R)-4-hydroxypyrrolidine-2-carboxylate hydrochloride	B8	BB1-16	GTATTAATCAAC	TAGGGTTGATTA
methyl (1R,3S)-3-aminocyclopentane-1-carboxylate hydrochloride	C1	BB1-17	GTATGTAGCACT	TAGGAGTGCTAC
ethyl azetidine-3-carboxylate hydrochloride	C2	BB1-18	GTATCGACCAGC	TAGGGCTGGTCG
methyl (S)-2-amino-3-cyclohexylpropanoate hydrochloride	C3	BB1-19	GTATACAACATT	TAGGAATGTTGT
methyl L-allothreoninate hydrochloride	C4	BB1-20	GTATTTAGCCAG	TAGGCTGGCTAA
methyl L-threoninate hydrochloride	C5	BB1-21	GTATGGACCCCA	TAGGTGGGGTCC
methyl 2-amino-3-methoxy-2-methylpropanoate	C6	BB1-22	GTATCCAACCGG	TAGGCCGGTTGG
methyl (R)-2-amino-2-(4-hydroxyphenyl)acetate	C7	BB1-23	GTATAAATCCTA	TAGGTAGGATTT
methyl D-alaninate hydrochloride	C8	BB1-24	GTATTGACCGAT	TAGGATCGGTCA
methyl (S)-2-aminopentanoate	D1	BB1-25	GTATGCAACGCC	TAGGGGCGTTGC
methyl methioninate hydrochloride	D2	BB1-26	GTATCTGGTCGG	TAGGCCGACCAG
methyl 2-amino-2-(tetrahydro-2H-pyran-4-yl)acetate	D3	BB1-27	GTATGTGGTGCC	TAGGGGCACCAC
methyl Nt-methyl-L-histidinate dihydrochloride	D4	BB1-28	GTATACGATGTC	TAGGGACATCGT
ethyl 5-amino-1H-pyrazole-3-carboxylate	D5	BB1-29	GTATTTGGTTAA	TAGGTTAACCAA
methyl 2-(3-oxopiperazin-2-yl)acetate	D6	BB1-30	GTATCCGATTGA	TAGGTCAATCGG
methyl (2R,4S)-4-methoxypyrrolidine-2-carboxylate hydrochloride	D7	BB1-31	GTATCCTAAAAT	TAGGATTTTAGG
ethyl 3-amino-3-methylbutanoate hydrochloride	D8	BB1-32	GTATAATTAACC	TAGGGGTTAATT
methyl D-leucinate hydrochloride	E1	BB1-33	GTATTTTGAAGT	TAGGACTTCAAA

methyl (S)-piperidine-2-carboxylate hydrochloride	E2	BB1-34	GTATGGTCAATC	TAGGGATTGACC
ethyl 2-amino-3-(pyridin-4-yl)propanoate dihydrochloride	E3	BB1-35	GTATCATTACAA	TAGGTTGTAATG
ethyl (S)-3-amino-3-(pyridin-3-yl)propanoate	E4	BB1-36	GTATTGTCACGA	TAGGTCGTGACA
methyl D-serinate hydrochloride	E5	BB1-37	GTATAGTCAGCT	TAGGAGCTGACT
methyl azetidine-2-carboxylate hydrochloride	E6	BB1-38	GTATGATTAGTT	TAGGAACTAATC
ethyl (R)-piperidine-3-carboxylate	E7	BB1-39	GTATCGTCATAG	TAGGCTATGACG
ethyl L-prolinate hydrochloride	E8	BB1-40	GTATTATTATGG	TAGGCCATAATA
methyl 2-amino-2-cyclopropylacetate hydrochloride	F1	BB1-41	GTATTATGCACG	TAGGCGTGCATA
methyl (R)-4,4-difluoropyrrolidine-2-carboxylate	F2	BB1-42	GTATGTTCCAGA	TAGGTCTGGAAC
Glycine Ethyl Ester Hydrochloride	F3	BB1-43	GTATCGTACATG	TAGGCATGTACG
Alanine Ethyl Ester Hydrochloride	F4	BB1-44	GTATAATGCCAC	TAGGGTGGCATT
D-Phenylalanine Methyl Ester Hydrochloride	F5	BB1-45	GTATGGTACCGC	TAGGGCGGTACC
L-Leucine Methyl Ester Hydrochloride	F6	BB1-46	GTATTGTACGCA	TAGGTGCGTACA
L-Tyrosine methyl ester hydrochloride	F7	BB1-47	GTATCATGCGTA	TAGGTACGCATG
Glycine Ethyl Ester Hydrochloride	1x1x1	BB1-50	GTATGATCGACT	TAGGAGTCGATC

Appendix Table 4 - All amino esters, the consequent reaction wells and BB2 codes, as well as corresponding building block code sequences and overhangs.

Amino ester (smiles)	Reaction well	BB2 code	OH2BB2-x sequence (5'-3')	OH3'BB2-x' sequence (5'-3')
methyl 1-aminocyclopentane-1-carboxylate hydrochloride	A1	BB2-1	CCTACAATCGGT	CGTAACCGATTG
methyl 1-aminocyclopropane-1-carboxylate hydrochloride	A2	BB2-2	CCTAATAGCGTC	CGTAGACGCTAT
methyl 2-amino-2-methylpropanoate hydrochloride	A3	BB2-3	CCTATCAACTAA	CGTATTAGTTGA

methyl D-tryptophanate hydrochloride	A4	BB2-4	CCTACTAGCTGA	CGTATCAGCTAG
methyl 1-aminocyclobutane-1-carboxylate hydrochloride	A5	BB2-5	CCTAAGACCTTG	CGTACAAGGTCT
methyl asparaginate hydrochloride	A6	BB2-6	CCTAGAAGGAAG	CGTACTTCCTTC
methyl (1r,4r)-4-aminocyclohexane-1-carboxylate hydrochloride	A7	BB2-7	CCTACTACGACA	CGTATGTCGTAG
methyl D-histidinate dihydrochloride	A8	BB2-8	CCTAAGAAGAGG	CGTACCTCTTCT
methyl 6-aminopyridazine-3-carboxylate	B1	BB2-9	CCTATCATGATA	CGTATATCATGA
methyl tyrosinate hydrochloride	B2	BB2-10	CCTAGTACGCAT	CGTAATGCGTAC
methyl 4-aminotetrahydro-2H-pyran-4-carboxylate hydrochloride	B3	BB2-11	CCTACGAAGCCC	CGTAGGGCTTCG
methyl 2-amino-2-(thiophen-2-yl)acetate hydrochloride	B4	BB2-12	CCTAACATGCGT	CGTAACGCATGT
methyl D-valinate hydrochloride	B5	BB2-13	CCTATAAGGCTC	CGTAGAGCCTTA
methyl morpholine-2-carboxylate hydrochloride	B6	BB2-14	CCTAGGAAGGAA	CGTATTCCTTCC
methyl 5-hydroxypiperidine-3-carboxylate hydrochloride	B7	BB2-15	CCTACCATGGCG	CGTACGCCATGG
methyl (2S,4R)-4-hydroxypyrrolidine-2-carboxylate hydrochloride	B8	BB2-16	CCTAAAAGGGGA	CGTATCCCCTTT
methyl (1R,3S)-3-aminocyclopentane-1-carboxylate hydrochloride	C1	BB2-17	CCTATTACGGTG	CGTACACCGTAA
ethyl azetidine-3-carboxylate hydrochloride	C2	BB2-18	CCTAGCATGTAC	CGTAGTACATGC
methyl (S)-2-amino-3-cyclohexylpropanoate hydrochloride	C3	BB2-19	CCTAATACGTGC	CGTAGCACGTAT
methyl L-allothreoninate hydrochloride	C4	BB2-20	CCTATGAAGTTT	CGTAAAACTTCA

methyl L-threoninate hydrochloride	C5	BB2-21	CCTACAACTAAT	CGTAATTAGTTG
methyl 2-amino-3-methoxy-2-methylpropanoate	C6	BB2-22	CCTAATAATACC	CGTAGGTATTAT
methyl (R)-2-amino-2-(4-hydroxyphenyl)acetate	C7	BB2-23	CCTATGATTAGT	CGTAACTAATCA
methyl D-alaninate hydrochloride	C8	BB2-24	CCTAGCAGTATC	CGTAGATACTGC
methyl (S)-2-aminopentanoate	D1	BB2-25	CCTACTAATCAA	CGTATTGATTAG
methyl methioninate hydrochloride	D2	BB2-26	CCTAAGATTCCG	CGTACGGAATCT
methyl 2-amino-2-(tetrahydro-2H-pyran-4-yl)acetate	D3	BB2-27	CCTATCAGTCGA	CGTATCGACTGA
methyl Nt-methyl-L-histidinate dihydrochloride	D4	BB2-28	CCTAGAACTCTG	CGTACAGAGTTC
ethyl 5-amino-1H-pyrazole-3-carboxylate	D5	BB2-29	CCTACGATTGAC	CGTAGTCAATCG
methyl 2-(3-oxopiperazin-2-yl)acetate	D6	BB2-30	CCTAACAGTGCT	CGTAAGCACTGT
methyl (2R,4S)-4-methoxypyrrolidine-2-carboxylate hydrochloride	D7	BB2-31	CCTATAACTGGC	CGTAGCCAGTTA
ethyl 3-amino-3-methylbutanoate hydrochloride	D8	BB2-32	CCTAGTAATGTT	CGTAAACATTAC
methyl D-leucinate hydrochloride	E1	BB2-33	CCTACCAGTTAG	CGTACTAACTGG
methyl (S)-piperidine-2-carboxylate hydrochloride	E2	BB2-34	CCTATTAATTGG	CGTACCAATTAA
ethyl 2-amino-3-(pyridin-4-yl)propanoate dihydrochloride	E3	BB2-35	CCTATTCAAAAC	CGTAGTTTTGAA
ethyl (S)-3-amino-3-(pyridin-3-yl)propanoate	E4	BB2-36	CCTAGGCTAACT	CGTAAGTTAGCC
methyl D-serinate hydrochloride	E5	BB2-37	CCTACCCGAAGC	CGTAGCTTCGGG
methyl azetidine-2-carboxylate hydrochloride	E6	BB2-38	CCTAAACCAATT	CGTAAATTGGTT
ethyl (R)-piperidine-3-carboxylate	E7	BB2-39	CCTATGCTACAG	CGTACTGTAGCA
ethyl L-prolinate hydrochloride	E8	BB2-40	CCTAGCCGACCA	CGTATGGTCGGC

methyl 2-amino-2-cyclopropylacetate hydrochloride	F1	BB2-41	CCTACACCACGG	CGTACCGTGGTG
methyl (R)-4,4-difluoropyrrolidine-2-carboxylate	F2	BB2-42	CCTAATCAACTA	CGTATAGTTGAT
Glycine Ethyl Ester Hydrochloride	F3	BB2-43	CCTATCCGAGAT	CGTAATCTCGGA
Alanine Ethyl Ester Hydrochloride	F4	BB2-44	CCTAGACCAGCC	CGTAGGCTGGTC
D-Phenylalanine Methyl Ester Hydrochloride	F5	BB2-45	CCTACTCAAGGT	CGTAACCTTGAG
L-Leucine Methyl Ester Hydrochloride	F6	BB2-46	CCTAAGCTAGTC	CGTAGACTAGCT
L-Tyrosine methyl ester hydrochloride	F7	BB2-47	CCTATACCATAA	CGTATTATGGTA
methyl (2S,3R)-2-amino-3-hydroxybutanoate hydrochloride	1x1x1	BB2-50	CCTACGCGCACA	CGTATGTGCGCG

Appendix Table 5 - All amines, the consequent reaction wells and BB3 codes, as well as corresponding building block code sequences, primer and overhangs.

Amine	Reaction well	BB3 code	OH3BB3-xP2 sequence (5'-3')	P2'BB3-x' sequence (5'-3')
2-Aminoimidazole hemisulphate	A1	BB3-1	TACGACCCCAGGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACCTGGGGT
p-Anisidine	A2	BB3-2	TACGTACACATATGACCT CAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCATATGTGTA
4-Chloroaniline	A3	BB3-3	TACGGGCGCCATTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAATGGCGCC
4-Cyanobenzylamine hydorchloride	A4	BB3-4	TACGAATATTGCTGACCT CAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGCAATATT
Cyclopropanemethyl amine hydrochloride	A5	BB3-5	TACGAACACCGTTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAACGGTGTT
3,4-Difluorobenzylamine	A6	BB3-6	TACGTTCTCCTCTGACCT CAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGAGGAGAA
3,5-Dimethoxyaniline	A7	BB3-7	TACGGCCCCGAATGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCATTCGGGGC
4-Dimethylaminopiperidine dihydrochloride	A8	BB3-8	TACGCACACGCGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACGCGTGTG

4-Hydroxypiperidine	B1	BB3-9	TACGTGCGCGTGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACACGCGCA
3-Aminobenzamide	B2	BB3-10	TACGGACACTACTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGTAGTGTC
6-Aminoindazole	B3	BB3-11	TACGAGCGCTGCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGCAGCGCT
3-Amino-1-N-methyl-azetidine dihydrochloride	B4	BB3-12	TACGCTCGGAATTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAATTCCGAG
4-Aminophenol	B5	BB3-13	TACGAGCCGACCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGGTCGGCT
3-bromoprop-2-yn-1-amine hydrochloride	B6	BB3-14	TACGTCCAGAGTTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAACTCTGGA
prop-2-yn-1-amine	B7	BB3-15	TACGGACTGATCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGATCAGTC
2-aminoacetonitrile hydrochloride	B8	BB3-16	TACGCGCCGCAATGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCATTGCGGCG
1-(bromoethynyl)cyclopropan-1-amine hydrochloride	C1	BB3-17	TACGACCAGCCGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACGGCTGGT
2-amino-4-bromobut-3-yn-1-ol hydrochloride	C2	BB3-18	TACGTACTGCGATGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCATCGCAGTA
1-phenylprop-2-yn-1-amine hydrochloride	C3	BB3-19	TACGGTCGGCTGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACAGCCGAC
m-anisidine	C4	BB3-20	TACGCCCAGGACTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGTCCTGGG
o-anisidine	C5	BB3-21	TACGAACTGGCTTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAAGCCAGTT
aniline	C6	BB3-22	TACGGGCCGGTTTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAAACCGGCC
3-Amino-1-methyl-1H-pyrazole	C7	BB3-23	TACGCACTGTAGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACTACAGTG
5-Amino-2-methoxypyridine	C8	BB3-24	TACGTGCCGTGGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACCACGGCA

N-Benzylmethylamine	D1	BB3-25	TACGATCCTAAATGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCATTTAGGAT
1-Benzylpiperidine	D2	BB3-26	TACGTGCATACGTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCACGTATGCA
(R)-4-Chloro-alpha-methylbenzylamine	D3	BB3-27	TACGGCCTTAGATGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCATCTAAGGC
3-Chloroaniline	D4	BB3-28	TACGCACGTATGTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCACATACGTG
2-Chlorobenzylamine	D5	BB3-29	TACGAGCATCACTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCAGTGATGCT
4-Chlorobenzylamine	D6	BB3-30	TACGTCCTTCCTTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCAAGGAAGGA
3-Chlorobenzylamine	D7	BB3-31	TACGGACGTCGCTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCAGCGACGTC
Cyclohexanemethylamine	D8	BB3-32	TACGCTCCTCTTTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCAAAGAGGAG
Cyclopentylamine	E1	BB3-33	TACGTACGTGCATGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCATGCACGTA
1-(Cyclopropylcarbonyl)piperazine	E2	BB3-34	TACGCGCATGTATGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCATACATGCG
3,4-Difluoroaniline	E3	BB3-35	TACGAACGTTATTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCAATAACGTT
cis-2,6-Dimethylmorpholine	E4	BB3-36	TACGGGCATTGTTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCAACAATGCC
1-Ethylpiperazine	E5	BB3-37	TACGGGGAAAAGTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCACTTTTCCC
4-Fluoroaniline	E6	BB3-38	TACGCCGTAACATGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCATGTTACGG
Indoline	E7	BB3-39	TACGAAGGAAGGTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCACCTTCCTT
morpholine	E8	BB3-40	TACGTTGCAATATGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCATATTGCAA
Piperidine	F1	BB3-41	TACGGCGTACATTGACCTCAACTACATGGTCTACA	TGTAGACCATGTAGTTGAGGTCAATGTACGC

4-(Aminomethyl)pyridine	F2	BB3-42	TACGCAGGACCCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGGGTCCTG
3-Picolylamine	F3	BB3-43	TACGATGCACGTTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAACGTGCAT
Benzylamine	F4	BB3-44	TACGTGGAACTCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGAGTTCCA
azetidine	F5	BB3-45	TACGCTGCAGCGTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCACGCTGCAG
4-Fluoroaniline	1x1x1	BB3-50	TACGACGGCACCTGACC TCAACTACATGGTCTACA	TGTAGACCATGTAGTTG AGGTCAGGTGCCGT