Silane Mediated Amidation and Difluoroethylation Reactions Using Carboxylic Acids

Oska T Pugh

MChem, AMRSC

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

December 2023

"There is a way out of every box, a solution to every puzzle. It's just a matter of finding it."

Captain Jean-Luc Picard
(Sir Patrick Stuart)

Abstract

Initial discussion centres firstly around silicon and its use within synthetic organic chemistry with a particular focus on organosilicon reagents as reductants. Secondly there is a brief discussion on the importance of metal free chemistry.

Chapter One discusses the development of an efficient silane mediated amidation reaction. To this end two methods were attempted. The first method made use of a silane immobilised on a polystyrene support, this immobilised silane was tested and optimised in an amidation reaction. The second method re-examined the amidation with phenylsilane with the intent reduce the amount of silane without compromising the reaction outcome. This reaction was applied to the synthesis of active pharmaceutical ingredients and natural products. Lastly in this chapter is described the near total synthesis of the natural product rehmagluamide.

Chapter Two discusses a practical one pot reductive amination of difluoroacetic acid for the synthesis of difluoroethylamines. This process is flexible in that it can be used for the synthesis of just difluoroacetamides in the first step, or difluoroethylamines in the full one-pot process. After the development of a Boc deprotection using difluoroacetic acid, the amide reduction reaction was also applied to the synthesis of unsymmetrical piperazines in a one pot procedure starting with N-Boc-piperazine.

Chapter Three discusses the development of a machine learning model to predict the pKa values for Brønsted acids in varying organic solvents. This model was iterated upon over 4 successive generations and did produce a viable workflow for predicting pKa values for individual acids. The most effective generation of the model was able to produce predicted pKa values for Brønsted acids known to the model but in solvents previously unknown to the model. The machine learning model is currently unable to reliably differentiate between solvents and work continues in this area.

Acknowledgments

I would like to thank the University of Nottingham for the opportunity to conduct my postgraduate studies at the Carbon Neutral Laboratory.

I would also like to thank the Engineering and Physical Sciences Research Council and Science Foundation Ireland for the funding of my postgraduate studies as part of the Centre for Doctoral Training in sustainable chemistry.

I would like to thank the staff of the centre for doctoral training for the support, training and events run by the CDT over the years of my postgraduate studies. especially during the first year and during the pandemic. I would also like to thank my fellow member of cohort 6 of the CDT for all the good time we had during our first year.

I would like to give special thanks to my principal supervisor Professor Ross Denton for the opportunity to work on these projects and within the group. I would also like to thank Ross for all the support, advice and help given during my postgraduate studies. With his help I have grown and developed significantly as a chemist, I will always be grateful for the support and insight he has given me. I would also to thank my secondary supervisors Professor Ender Özcan and Dr Grazziela Figueredo for all their help and support in teaching me the fundamentals of machine learning.

I would like to thank the technical, analytical, and support staff of the chemistry department for all the support given over the last 4 years.

I would like to thank the past and present members of the Denton group for all the support and good times had over the last 4 years. I would like to give particular thanks to Alex Uner and Connor Nolan for their support in the difluoroethylation project. Special thanks go to Amelia Stoneley, Isabel Wood, Laura Blair, Mia Bonfield and Tom Barber for all the good times both in the lab and in the office.

I would like to thank the members of my D&D group for the welcome escape they have provided over the course of my PhD in particular Alex Parkinson and Joseph Mccurry for running them games. The chance to get away from chemistry for a few hours a week gave me time to rest my mind and perform better.

My journey through education hasn't always been straight forward but the help and support of the leaning support offices at Thomas Alleyne School, North Hertfordshire Collage and the disability support offices at Manchester Metropolitan University and the University of Nottingham was always there to lend a hand at each stage of my education.

I would like to give thanks to my family, especially my mum Nicola Pugh, and my sister Tamara Pearce for the support they have given me over the course of not only my PhD but also my journey through education. With out them I would not have had the level of success I now have.

I would like to give very special thank my partner Jonathan Fundament for all the help, support, and tolerance throughout my PhD. Thank you standing with me through the hardest parts of my PhD.

There are many others who have come and gone throughout my long journey through education, more than could be recorded here properly. I would like to thank you all as each of you has in your own way helped get me to this point.

Lastly, I would like to recognise my own progress. My journey through education has not been typical, I have at time been my own worst enemy. But as I have developed, I have grown to become my greatest advocate. I am proud of the work I have done, and I will always be grateful for the help people have given me over the years.

Abbreviations

2c-2e – two centre two electron bond

3c-4e – three centre four electron bond

AI – artificial intelligence

Ar – aryl group

Cy – cyclohexyl group

DCC – N,N-dicyclohexylcarbodiimide

DCM – dichloromethane

DIPEA – diisopropyl ethylamine

DFT – density functional theory

DMA – dimethyl acetamide

DMC – dimethyl carbonate

DMF – dimethyl formamide

DMSO – dimethyl sulfoxide

Dppe – 1,2-bis(diphenylphosphino)ethane

Et – ethyl group

ee – enantiomeric excess

EEDQ – 2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline

EDC/EDCl 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride

EMPA – ethylmethylphosphinic anhydride

IR – infrared

InChi – International Chemical Identifier

^{i}Pr – isopropyl group

LASSO – least absolute shrinkage and selection operator

LDA – lithium diisopropylamine

Me – methyl group

ML – machine learning

NAD – Nicotinamide adenine dinucleotide

NADH – Nicotinamide adenine dinucleotide + hydrogen

NMA – N-Methylaniline

NMM – *N*-methyl morpholine

NMPi – N-methylpyrrolidine

NMR – nuclear magnetic resonance spectroscopy

PCA – principal component analysis

PFAS – perfluorinated alkyl substances

PMHS – polymethylhydrosiloxane

PTSA – *para*-toluene sulfonic acid

Ph – Phenyl group

PhMe – toluene

SAR – Structure activity relationship

SMILES – Simplified molecular-input line-entry system

SVM – support vector machines

T3P – N-propanephosphonic acid anhydride

TBAF – tetrabutyl ammonium fluoride

^{t}Bu – tertbutyl group

TEMPO - 2,2,6,6-Tetramethylpiperidine 1-oxyl

TFA – trifluoroacetic acid

THF – tetrahydrofuran

TMS – tetramethyl silyl group

TBS – tertbutydimethyl silyl group

TBDPS – terbutyldiphenyl silyl group

TIPS – triisopropyl silyl group

TTMSS - Tris(trimethylsilyl)silane

Table of Contents

Abstract 3

Acknowledgments 4

Abbreviations 6

General Introduction 13

1.1 Fundamentals and Reactivity of Organosilanes 14

1.1.1 Silicon 14

1.1.2 Organosilicon Compounds: Principals and Reactivity 15

1.1.2.1 General Chemical Properties 15

1.1.2.2 Organosilicon vs Carbon 17

1.1.2.3 Hypervalence vs Hypercoordination 25

1.1.3 Examples in Synthesis 29

1.1.4 Silyl Ethers 29

1.1.5 Silyl Enol Ethers 30

1.1.6 Silanes as Reductants 31

1.2 Importance of Metal Free Chemistry 38

1.2.1 Limitations of Metals in Synthesis 39

1.2.2 Advantages and Examples of Metal Free Chemistry 40

1.2.2.1 Synthetic Examples of Metal Free Chemistry. 40

1.3 References 44

Chapter One 48

2.1 Introduction 49

2.1.1 Fundamentals of Amides 49

2.1.2 Classical Amidation Reactions 50

2.1.3 Silane Mediated Amidations 61

2.1.4 Immobilized Reagents and Flow Chemistry Fundamentals 67

2.2 Immobilised Silanes for Amidation Reactions 69

2.2.1 Aims and Objectives 69

2.2.2 Results and Discussion 70

2.3 Efficient Phenylsilane Mediated Amidation Reactions 77

2.3.1 Aims and Objectives 77

2.3.2 Results and Discussion 78

2.3.2.1 Re-optimisation 78

2.3.2.2 Calculation of green metrics 80

2.3.2.3 API and natural product synthesis 83

2.4 Conclusions 93

2.5 Experimental 94

2.5.1 Immobilised Silanes for Amidation Reactions 95

2.5.2 Efficient Phenylsilane Mediated Amidation Reactions 98

2.6 References 106

Chapter Two 108

3.1 Introduction 109

3.1.1 Fluorinated Amines 109

3.1.1.1 Tri-fluorinated Ethylamines 109

3.1.1.2 Mono-fluorinated Ethylamines 114

3.1.1.3 Di-fluorinated Ethylamines 115

3.1.2 Reductive Aminations and Amide Reductions 117

3.1.2.1 Reductive Aminations 117

3.1.2.2 Amide Reductions 122

3.1.3 Silane Mediated Amide Reductions 123

3.1.4 Prior Work by the Denton Group 125

3.1.5 Aims and Objectives 129

3.2 Results and Discussion 129

3.2.1 Difluoroacetamides 129

3.2.2 Tertiary Difluoroethylamines 131

3.2.3 *In situ* Boc Deprotections and Unsymmetrical Piperazines 136

3.2.4 Secondary Difluoroethylamines 139

3.3 Conclusions 141

3.4 Future Work 142

3.5 Experimental 142

3.5.1 Difluoroacetamides 143

3.5.2 Tertiary Difluoroethylamines 148

3.5.3 Unsymmetrical Piperazines 154

3.5.4 Secondary Difluoroethylamines 159

3.6 References 160

Chapter Three 163

Prediction of Brønsted Acid pK_a Values by Machine Learning 163

4.1 Introduction 164

4.1.1 Acid-Base theory and pK_a 164

4.1.1.1 Fundamentals 164

4.1.1.2 Brønsted Acidity vs Lewis Acidity 165

4.1.1.3 pK_a 166

4.1.2 Machine Learning for Organic Chemistry 168

4.1.2.1 Model Types 168

4.1.2.2 Typical ML Workflow 173

4.1.3 Machine Learning and pKa Prediction 176

4.2 Aims and Objectives 178

4.3 Data set generation and curation 178

4.4 Model Development 180

4.4.1 1st Generation Model 180

4.4.2 2nd Generation Model 182

4.4.3 3rd Generation Model 183

4.4.4 4th Generation Model 186

4.5 Conclusions 188

4.6 Future Work 188

4.7 Experimental 188

4.8 References 192

General Introduction

An introduction to organosilicon chemistry and metal free synthesis. Detailed herein are the uses of organosilicon reagents within synthesis. More specific uses will of organosilanes be detailed at the beginning of each chapter.

1.1 Fundamentals and Reactivity of Organosilanes

1.1.1 Silicon

Silicon, a group 14 element and one of the metalloids, in its elemental form it is a hard brittle crystalline solid with a blue/grey metallic colour. It is the eighth most abundant element on earth and the second most abundant in the earth's crust after oxygen. First identified in 1787 by Antoine Lavoisier, who concluded that silica must be the oxide of a previously unknown element. Pure elemental silicon was discovered in 1823 by Jöns Jacob Berzelius, who prepared a pure amorphous sample of the element by reducing potassium fluorosilicate with molten potassium (**Figure 1**).[1] Crystalline silicon was prepared later by Henri-Étienne Sainte-Claire Deville, who isolated from aluminium melts where it had been previously identified as an impurity.[2]

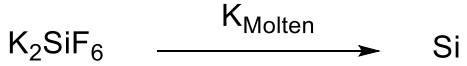


Figure 1: Original synthesis of elemental silicon by Berzelius

It has been used extensively since antiquity as silicates in mortar and in various early examples of glassware. More contemporary uses of silicon include silicates in Portland cement, high quality glass for optical systems, or in the form of siloxanes for silicone polymers, elastomers, and resins.[3] The most famous use for silicon today is in the form of crystalline silicon wafers for semiconductors used in nearly all computers, leading to the "Silicon Age" within computing.[4] While silicon has many industrial and commercial uses there has been significant attention given to its uses within chemistry and in particular chemical synthesis.

1.1.2 Organosilicon Compounds: Principals and Reactivity

1.1.2.1 General Chemical Properties

Much like other group 14 elements silicon can readily form tetravalent compounds and often form sp^3 hybrid species.[5] Electronically silicon's first four ionisation energies (786.3, 1576.5, 3228.3, 4354.4 kJ mol^{-1}) are too high to permit most simple cationic chemistry.[6] However, it should be noted in more contemporary literature cationic silylium and silyliumylidene species do exist and demonstrate a wide variety of chemistry.[7,8]

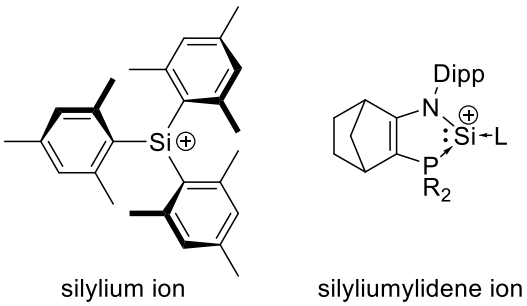


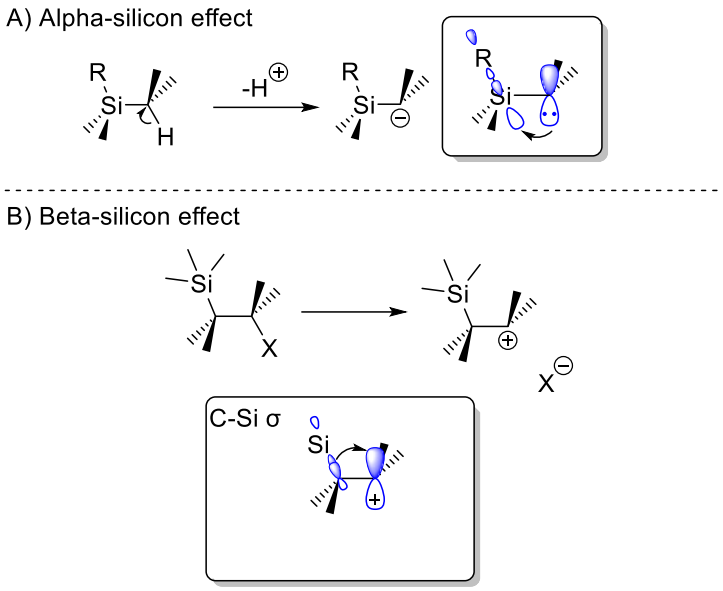
Figure 2: examples of silylium and silyliumylidene species

In its elemental form silicon forms face centred cubic crystals which are largely inert save for the formation of a protective thin layer of silicon dioxide. Further oxidation in air is not measurable below 900 °C, and at 1400 °C silicon readily reacts with atmospheric nitrogen. Against halogens the protective silicon dioxide layer offers only a limited defence, silicon readily reacts with fluorine forming various silicon fluoride species at room temperature. For the remaining halides, silicon reacts with chlorine at around 300 °C and bromine and iodine require temperatures over 500 °C.[3]

Conjugation and Negative hyperconjugation of silicon compounds

Negative hyperconjugation is the interaction between a filled Π or *p* and an adjacent σ* orbital, this type of resonance has the ability to stabilise molecules and transition states.[9] For

elements in the 3rd period particularly silicon negative hyperconjugation is that cause of a number of observable effects.



Scheme 1: Alpha and beta silicon effects

The Alpha-silicon effect (**Scheme 1A**) can be characterised as silicon's ability to stabilise α-carbanions and conversely destabilise α-carbocations, this effect was first encountered by Sommer and coworkers in 1946. Mechanistically for carbanions the alpha silicon effect works by electron donation from either a filled *p* orbital or a sp^3 orbital into the Si-C σ* orbital. The authors found that the hydrolysis of C-Cl bonds beta and gamma to silicon could readily occur but when silicon is alpha to the C-Cl bond hydrolysis was not observed.[10] Thermodynamic data produced by Yong *et al.* showed that negative hyperconjugation plays a role in the destabilisation of carbocations alpha to silicon.[11]

For carbocations beta to silicon a different effect is observed, known as the beta-silicon effect (**Scheme 1**B). This term describes how though conjugation by silicon in an antiperiplanar relationship can stabilise a beta carbocation.[10] The exact mechanism of this effect is the donation of electrons from the Si-C σ orbital into the C-X σ* orbital thereby lowering the energy barrier and weakening the bond and allowing for easier bond cleavage.[12]

These effects have a profound impact on how organosilicon compounds can behave and how they perform in reactions. To examine in greater depth how silicon behaves in organic synthesis it is useful to draw comparisons on how silicon bonds with other heteroatoms to one of its fellow group 14 elements, carbon.

1.1.2.2 Organosilicon vs Carbon

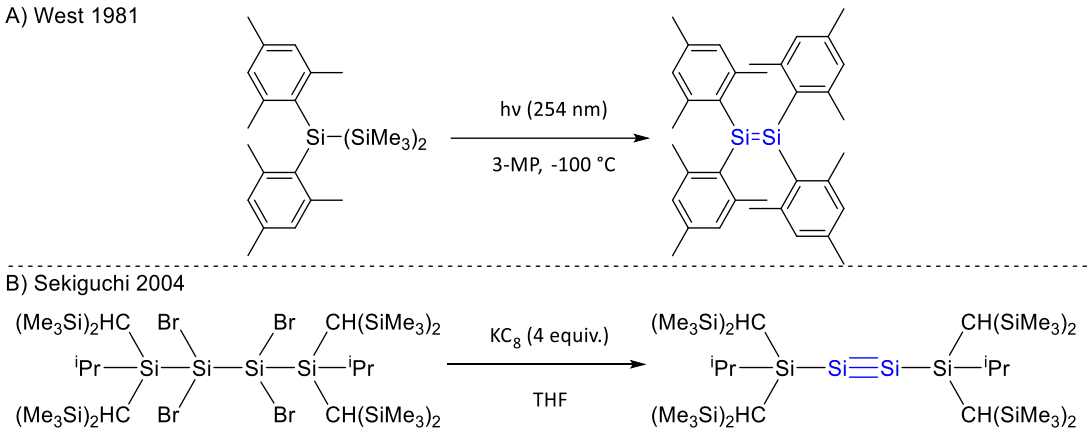
Silicon-Silicon bonds vs Carbon-Carbon bonds

Silicon is often compared with carbon owing to their relative positions in group 14 of the periodic table. However, despite this comparison silicon has several clear differences, firstly carbon has a higher electronegativity 2.5 for C vs 1.8 for silicon. Silicon can attain higher coordination numbers as shown in **Table 1**, carbon can readily make double and triple bond systems whereas silicon requires highly specific conditions. Beyond fundamental properties the principal difference is that carbon has a significantly larger range of reactivity in the field of organic chemistry, a range that silicon simply cannot match. Unlike carbon, long chains of silicon hydrides (polysilanes) are rare and are typically unstable. Polysilanes with the formula Si_nH_{2n+2} typically exist as gases or as highly volatile liquids that are pyrophoric or explosive.[3] A prime example of this can be seen in disilane (Si_2H_6) which slowly decomposes over time (approx. 2.5% per 8 months) and is pyrophoric.[3] Trisilanes and tetrasilanes decompose more rapidly at room temperature, only silane (SiH_4) is thermally stable yet it is still pyrophoric.[13] In direct comparison carbon readily form long highly stable alkyl chains, making them suitable for use in bulk materials such as plastics.

	C	Si
Electronic configuration	$1s^2\ 2s^2\ 2p^2$	$1s^2\ 2s^2\ 2p^6\ 3s^2\ 3p^2$
Covalent radius (pm)	77	117
Stable coordination numbers	1,2,3,4	3,4,5,6
Pauling electronegativity	2.5	1.8
Double or triple bonds	Stable	Typically unstable
Hypercoordinate systems	unstable	Stable

Table 1: Overview of fundamental properties of silicon and carbon[14]

Silicon-silicon double bonds known as disilenes are rare and typically less stable than their alkene equivalents.[15] early attempts at producing disilenes often proved unsuccessful and Kipping concluded that attempts at multiple bonds at silicon was not "a realistic objective".[16] One of the first stable disilenes synthesised and isolated was reported by West *et al.* in 1981 as shown in **Scheme 2A**. The authors reported that Tetramesityldisilene could be prepared by UV irradiation of 2,2-bis(mesityl)hexamethyltrisilane to afford a yellow crystalline solid.[17] Silicon-silicon triple bonds or disilynes (Si≡Si) are significantly less stable and common than polysilenes, the first reported synthesis of a polysilynes was reported independently by Sekiguchi and Wiberg in 2004.[18–20] In the work by Sekiguchi they made use of highly sterically bulky silyl groups to produce stable Polysilynes as shown in **Scheme 2B**. in this reaction an excess of potassium graphite (KC_8) in THF was used to allow for the reduction of the tetrabrominated precursor with a yield of 73%.



Scheme 2: Synthesis of a stable polysilene and a polysilyne

From an electronic standpoint silicon has a much larger atomic radius than carbon as well as silicon has much larger and more diffuse series of orbitals than carbon.[21] This leads to much weaker bonds between silicon and carbon or silicon and carbon in comparison carbon-carbon bonds (**Table 2**). however silicon bonds with heteroatoms such as oxygen and fluorine are significantly stronger than comparable bonds to carbon.[14]

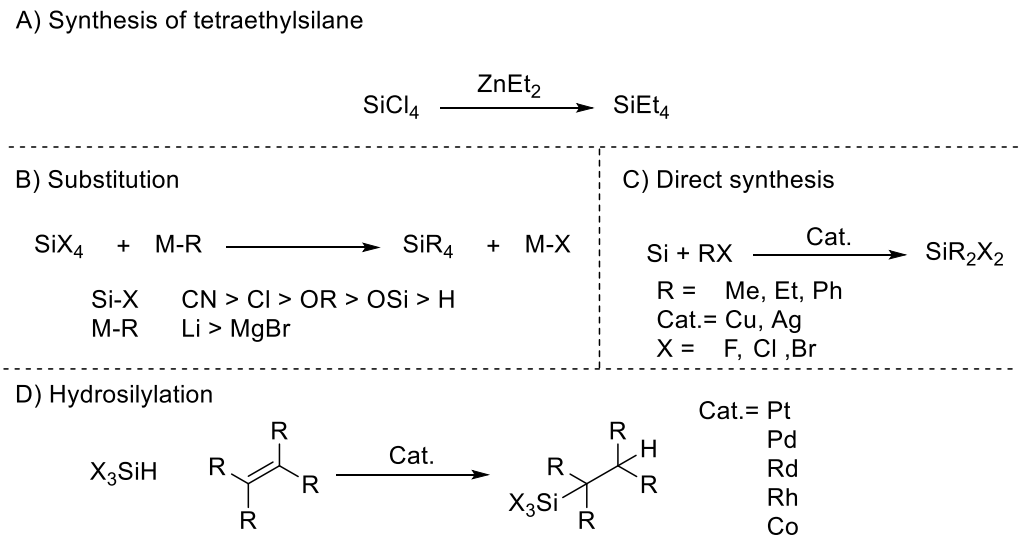
	C	Si	H	N	O	F	Cl	Br	S
C-X	347	301	413	305	358	447	339	284	271
Si-X	301	314	317	435	452	565	380	310	293

Table 2: Bond strengths for C-X and Si-X, in kJ mol-1

Silicon-carbon bonds vs carbon-carbon bonds

As discussed in the section above carbon readily form stable hydrocarbon chains, in fact all alkanes up to dodecane have autoignition temperatures above 200 °C.[22] In comparing silicon to carbon the carbon- carbon bond (C-C) and the silicon-carbon bond (Si-C) bonds show many key differences. The prime difference is that carbon-silicon bonds are longer and weaker than carbon-carbon bonds. The silicon-carbon bond is polarised towards carbon owing to carbons notably greater electronegativity.[23]

The earliest example of a synthetic silicon-carbon bond is also the earliest example of a synthetic organosilicon compound. Produced by C. Friedel and J. Crafts in 1863 tetraethylsilane was prepared through the reaction of diethyl zinc and silicon tetrachloride (**Scheme 3A**).[24] Since this initial synthesis a great diversity of organosilicon reagents now exist and it is possible to access a number of more bespoke reagents through three routes.[25] The first route is through substitution of a organosilicon compound with ether an organolithium reagent or a Grignard reagent (**Scheme 3**B), these methods in particular the Grignard methods offer the ability to control the amount of substitution through the stoichiometric control of the reagent.



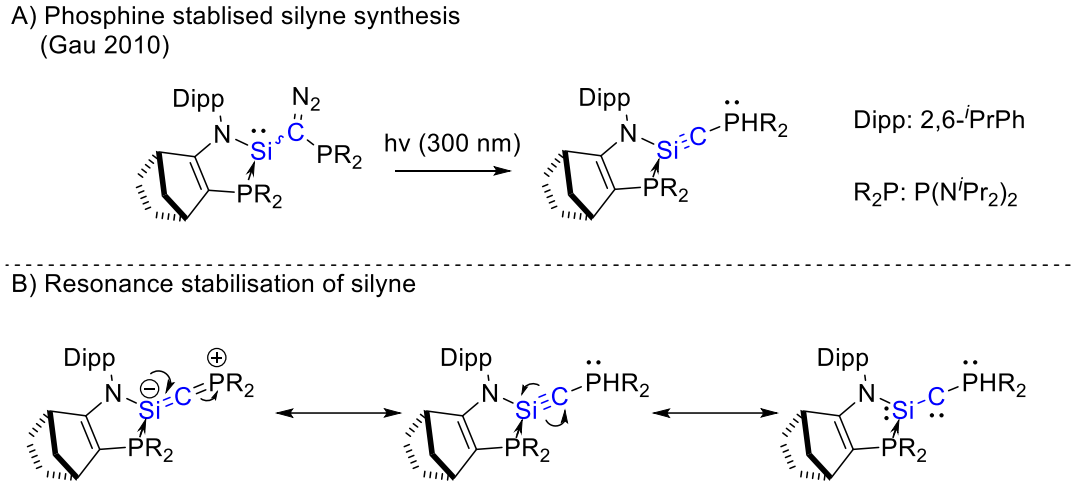
Scheme 3: Common methods for Silicon-C bond formation

The second method is direct synthesis from elemental silicon with alkyl halide and a metal catalyst (**Scheme 3C**), this process was initially developed independently by Rochow and Müller in the 1940's.[26] This process is the current standard reaction for the large scale production of methylchloro and phenylchlorosilanes, though in theory it is possible to use other organohalides, it has been shown that these give poor yields necessitating the use of other methods.[27,28] The third method is a metal catalysed hydrosilylation of olefins (**Scheme**

3D), these process are typically employed for the synthesis of functionalised silanes, alkyl silanes and crosslinked silicone polymers.

The general strategy of substitution employed by Friedel and Crafts was expanded on by Kipping and was generalised for many halides and pseudohalide species. Further development has proved that substitution is often the most practical method to produce organosilanes, through careful control of reactant stoichiometry. Despite Kipping's extensive research in this area and publishing over 50 papers on the subject he declared that "the prospect of any immediate and important advances in this section of organic chemistry does not seem to be very hopeful".[16] Works described in later sections has since proven this statement untrue.

Compounds with carbon silicon double bonds known as Silenes (Si=C), these compounds much like their parent disilene compounds are typically unstable. The number of accessible synthetic routes to these compounds are very limited. The earliest reported synthesis by Brook *et al.* in 1982 produced a silene through the photolysis of α-acylsilanes.[29] More contemporary methods involve a sila-Peterson type reaction (Apeloig 1992) or by the reaction of a 1,1-dilithiosiliol with a sterically bulky ketone (West 2004).[30,31] In all these methods the silene is kept stable by the use of extremely bulky substituents and as a result the range of silenes that can be produced is limited. Silynes (Si≡C) are in comparison to alkynes extremely unstable and difficult to synthesise. The first reported synthesis of a silyne was reported in 2010 by Gau *et al.* where they describe a silyne stabilised with a phosphine ligands.[32] The key synthetic step is the photolysis of a phosphino(silyl)diazomethane (**Scheme 4A**) to afford the silyne product. The author verified the presence of the silyne by x-ray crystallography compared to DFT calculations. The silyne is stabilised through the formation a phosphonium sila-ylide through resonance as shown in **Scheme 4B**.



Scheme 4: Silyne synthesis and stabilisation

Silicon-hydrogen bond vs Carbon-hydrogen bond

The silicon hydrogen bond (Si-H) is noticeably weaker and longer than that of a carbon hydrogen bond (C-H). The relationship of the Si-H bond can be seen as inverse to that of the C-H bond (**Figure 3**). This can be attributed to the electronegativities of the atoms involved. As carbon has a greater electronegativity relative to hydrogen the C-H bond is slightly polarised in favour of Carbon. For Si-H bond the opposite is seen, Hydrogen being more electronegative that silicon means that the Si-H bond is polarized to in favour of Hydrogen. In synthetic terms many organosilanes species are used either as source of hydride (used in reductions), or in hydrosilylation.

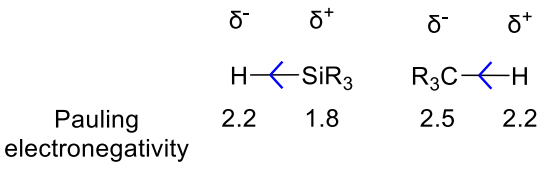
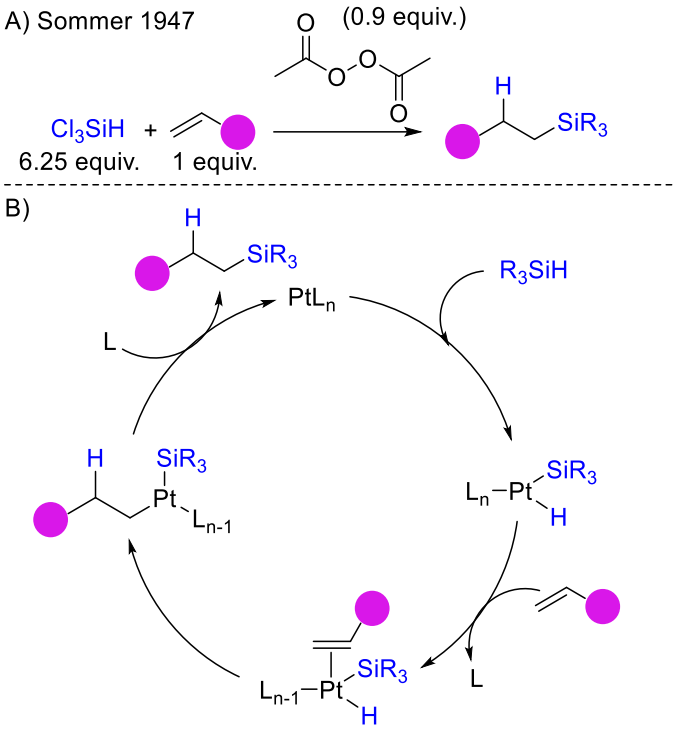


Figure 3: Difference in bond polarity between silicon and carbon

Hydrosilylation of alkenes and alkynes

Hydrosilylation is the addition of a silicon-hydrogen bond across an alkene or alkyne to produce the corresponding organosilicon compounds. These reactions have wide application within industry as a key process in the formation crosslinked silicone polymers, and

organosilicon reagents for fine chemical synthesis.[33] The earliest reported hydrosilylation was reported by Sommer *et al*. in 1947 (**Scheme 5A**).[34] The authors prepared n-octyltrichlorosilane from trichlorosilane and 1-octene using a diacetyl peroxide, though this process showed poor selectivity.[34] Modern methodology typically employs metal catalysts like platinum, ruthenium, or iron though metal free alternatives exist. Platinum catalysed hydrosilylations are the most used methods, proceeding via a Chalk-Harrod mechanism and in the case of terminal akenes will adopt an anti-Markovnikov arrangement shown below in **Scheme 5B**.[35] Asymmetric hydrosilylations are possible and commonly make use of chiral ligands and a platinum catalyst.[36]



Scheme 5: A) Hydrosilylation by Sommer et al.. B) Chalk-Harrod mechanism for metal catalysed hydrosilylation

Silicon-Oxygen Bond vs Carbon-Oxygen

In inorganic chemistry silicon-oxygen bonds are extremely common with silicates and in silicone polymers, in organic chemistry silicon is renowned for its oxophilicity.[3] The silicon-oxygen bond is longer (1.63 Å) and more polar than the corresponding carbon-oxygen bond (1.43 Å), but it is around 0.2 Å shorter than models predict.[37] The cause of the strength of the

Si-O bond has been highly debated, the most accepted model involves negative hyperconjugation of the O lone pair with an adjacent Si-R σ* orbital (**Figure 4**).[38] This effect is maximised by the adjacent Si-R bond has a higher p character. The effect adjacent hybridisation has, can be seen in the bond angles of dimethyl ether and disilyl ether. Dimethyl ether has a C-O-C bond angle of 111.2° from this the oxygen lone pairs are calculated to have a p character of 78.3%. While disilyl ether has a larger bond angle of 151.2° which correlates to the oxygen lone pairs having a p character of 95.2%.

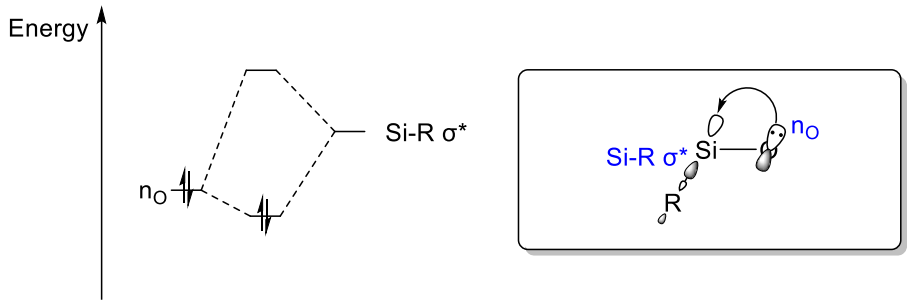


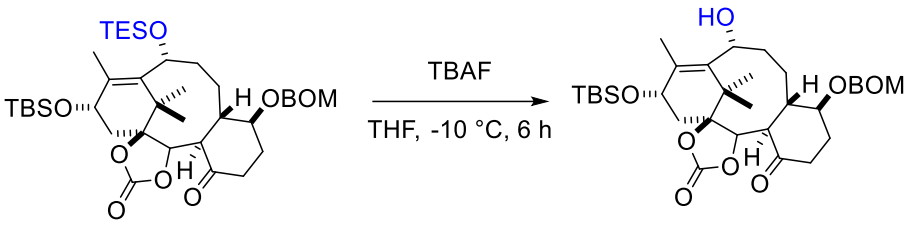
Figure 4: Molecular orbital diagram for negative hyperconjugation of heteroatom lone pair electrons with adjacent Si-R σ* orbital.[39]

The silicon-oxygen double bond sometimes referred to as a silanone, is much weaker than its carbon-based equivalent. Silanones have a typical bond strength of 590 KJ mol^{-1} vs 715 KJ mol^{-1} for C=O bonds, The cause for this is a poor overlap between of the p orbitals between oxygen and silicon.[40] The comparable weakness of the Si=O bond means that most silanones are unstable and prone to oligomerisation or polymerisation.[41] For this reason it is typical to only silanones as intermediates, however, in 2014 Filippou and coworkers developed a stable silanone from silylidyne compound, though this compound readily reacts with water to give a dihydroxysilyl species.[42]

Silicon-Fluorine bond vs Carbon-Fluorine

The key characteristic of the silicon fluorine bond is its strength, at 565 kJ mol^{-1} it is the strongest single bond silicon can form and the second strongest single bond in chemistry behind HF kJ mol^{-1} at 569.[43] Due to the strength of this bond silicon is notably fluorophilic

surpassing even its oxophilicity. The earliest example of a synthesised Si-F bond was before the discovery of silicon itself, in 1771 Carl Wilhelm Scheele produced silicon tetrafluoride by dissolving silica in hydrofluoric acid.[44] A contemporary application of Si-F bond formation is the use fluoride sources for the deprotection of organosilicon protecting groups, classical fluorine sources include TBAF, HF·Pyridine and KF·2H_2O.[45,46] These deprotection methods can even be engineered to be selective for specific silyl protecting groups. During Holton's synthesis of Taxol they demonstrated a selective deprotection of triethyl silane using TBAF in the presence of other silyl ether protecting groups (**Scheme 6**).[47]



Scheme 6: Selective deprotection of a TES group

1.1.2.3 Hypervalence vs Hypercoordination

When comparing carbon to silicon there is one property that carbon does not possess, that property is hypercoordination. To understand hypercoordination it is best to examine how the term and concept was developed and that starts with another term, hypervalence.

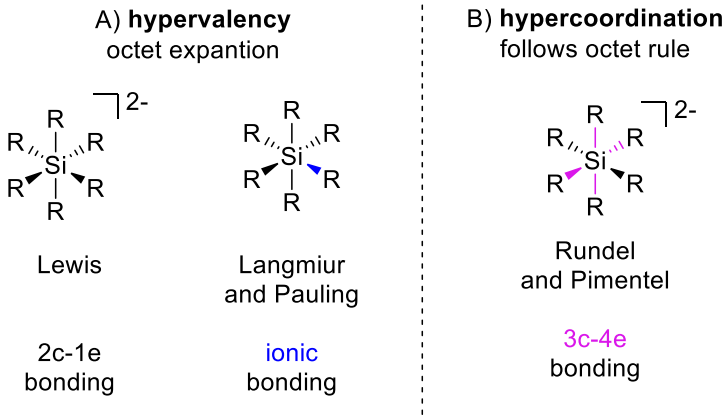


Figure 5: Hypervalency vs hypercoordination

Hypervalence as defined by Musher in 1968 is:

"Molecules and ions formed by elements in groups V-VIII of the periodic table in any of their valences other than their lowest stable chemical valance of 3,2,1, and 0 respectively."[48]

Simply put hypervalent compounds "appear to disobey" the octet rule by having more than eight electrons in their valence shell. Hypervalence as a concept was first discussed in the 1920's by G. Lewis and I. Langmuir during the development of their theory on atomic bonding.[49] Lewis favoured the concept of octet expansion and that all bonds within the molecule followed the 2 centre 2 electron (2c-2e) bond model (**Figure 5**).[50] Langmuir's explanation preserved the octet rule and assumed that bonding in a hypervalent molecule was ionic as opposed to covalent.[51] In the late 1920's and early 30's S. Sugden proposed that the presence of two centre one electron (2c-1e) bonds in conjunction with 2c-2e bonds could be used to rationalise Hypervalence (**Figure 5**).[52] This idea was expanded upon by Rundle and Pimentel who proposed a three centre 4 electron bond (3c-4e) model where only two of the electrons are involved in bonding. With the remaining electrons occupying either nonbonding or weak antibonding orbitals on the periphery of the molecule as shown in **Figure 6**.[53,54] When Musher defined the term hypervalent he rejected the concept of octet expansion in favour of the Rundel-Pimentel 3c-4e model, later computational models have also validated the 3c-4e model with precautions about the inclusion of d-orbitals.

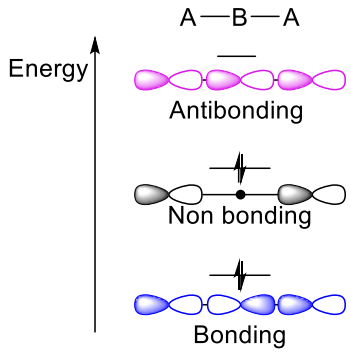


Figure 6: Schematic depiction of a MO for a 3 centre 4 electron bond[55]

Many main block elements readily form hypervalent species with sulphur, iodine, chlorine and even some noble gases like xenon having a diverse array of examples. Examples of hypervalent species include Dess-Martin periodinane, pentavalent phosphorus (PCl_5, PF_5), sulfuranes, xenon hexafluoride.[56]

Hypercoordination described by Paul von Ragué Schleyer in 1986 is an alternative term to hypervalence. The key difference distinction the two terms is that hypercoordination does not make assumptions on any possible bonding modes within the molecule and can avoid the issue altogether.[49] Formally hypercoordination is defined as "a property of main-group atoms in molecular entities to acquire coordination numbers greater than four".[57] Hypercoordination is a much broader term as it covers the entirety of main group elements as opposed to Hypervalence witch only covers groups 15 to 18. Early attempts to explain hypercoordination in detail invoked d orbital interactions that could form bonds in a similar way to metal complexes.[58] Later studies have rejected this explanation in favour of the 3c-4e bonding models that were used to describe hypervalancy.[59] General examples of hypercoordinate compounds include pentacoordinate carbon, phosphorus, sulfur, and boron species even main group metals such as aluminium can adopt hypercoordinate species.[60,61]

In terms of silicon, penta and hexacoordinate silicon are the most common hypercoordinate species observed. Silicon's ability to expand its valance and coordination sphere is critical to substitution of organosilicon reagents.[62] The way in which silicon is able to do so is not because silicon can adopt a higher oxidation state like in the case of sulfur. Instead silicon makes use of its Lewis acidity allows for explanation of its coordination sphere.[62] In this expanded coordination sphere silicon can form either weak dative bonds where silicon act as the acceptor, or it can form weak anionic species that are largely unstable. Stable hypercoordinate silicon species are uncommon are highly dependent on the ligands coordinated to the silicon centre.[63]

Hypercoordinate silicon species are known to be significantly more reactive than tetracoordinate silane species. This increase in reactivity has a significant impact on what mechanistic pathways are observed during reactions. It is typical to observe reactivity through the axil bonds of a hypercoordinate as they are far more susceptible to heterolytic cleavage than the equatorial bonds.[62] This can be seen in nucleophilic substitution reactions around silicon, these reactions typically procced by a pentacoordinate complex. Computational work by Bento and Bickelhaupt showed that the energy barrier for silicon can change significantly depending of the substituents bound to silicon (**Figure 7**).[64] They also show that the axial bonds in the pentacoordinate transition state are longer than the groups in the equatorial position.

An example of a stable pentacoordinate silicon species has been shown by Tacke *et al.* in 2000 where they produced two pentacoordinate compound based around an SiO_5 core. this X-ray crystallography of these compounds clearly shows the elongated axial bonds.

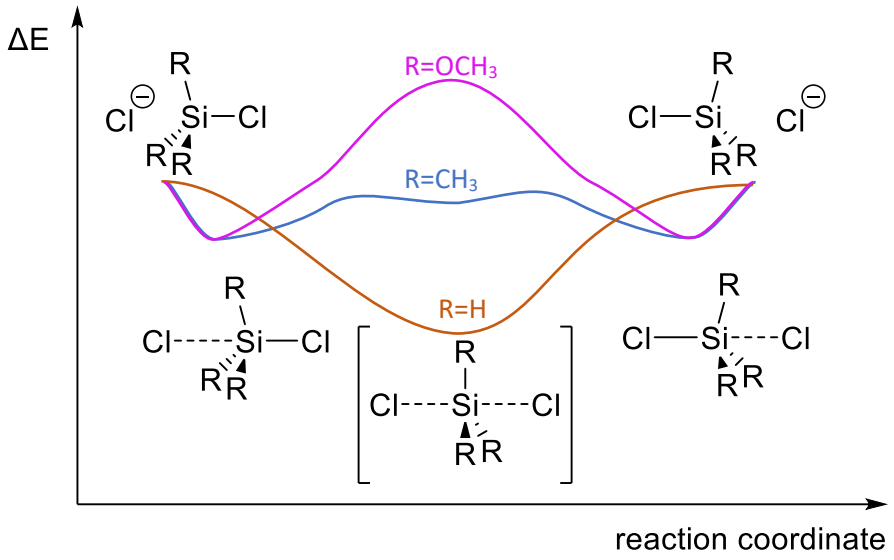


Figure 7: Computed potential energy surfaces for nucleophilic substitution at silicon

1.1.3 Examples in Synthesis

For organosilicon reagents may come in the form of oils, elastomers, or resins often with high thermal stability and chemical inertness. These properties allow for a low barrier of both skill and equipment for the use of organosilicon reagents.

In organic chemistry silicon is often seen in a wide variety of reagents from silyl ethers, silyl enol ethers, silyl esters, and a multitude of other species. These organosilicon reagents have a wide verity of uses from protecting groups to reducing agents.[65,66] Two main species of organosilicon reagent are commonly employed silyl halides (Si-X) and hydrosilanes (Si-H).

1.1.4 Silyl Ethers

A primary use of silyl ethers is as protecting groups, developed by E. J. Corey in 1972, silyl ethers are a common method to protect alcohols.[66] Examples include trimethyl silane (TMS), tertbutyl dimethyl silane (TBS), tertbutyl diphenyl silane (TBDPS), and triisopropyl silane (TIPS). Silyl ethers offer excellent protection as they can resist a diverse array of conditions though it should be noted that TMS groups are still quite labile and are easy to hydrolyse.[67] Silylation of alcohols is typically performed using a silyl chloride and imidazole, this method is by far the most widely used due to its extremely wide scope and ease of use.[67] Deprotection can vary depending on the steric and electronics of the protecting group, smaller groups such as TMS can be deprotected with strong aqueous acids while more bulky groups will require the use of fluoride reagents such as tertbutyl ammonium fluoride (TBAF). While silyl groups are most employed for alcohols, they can also be used to protect carboxylic acids, amines, alkynes, and thiols.[67]

Outside the role of protecting groups in synthesis silyl ethers also can be employed in prodrugs allowing for *in vivo* deprotection and controlled release of the active pharmaceutical. Work

published by Parrott *et al.* showed how an asymmetric bifunctional silyl ether could act as a method of delayed release for the drug molecule.[68]

1.1.5 Silyl Enol Ethers

The synthesis of silyl enol ethers as shown in **Figure 8** is typically performed through an enolisable ketone formed with a base such as LDA or triethylamine and then the addition of a silyl chloride.[69] Though there have been many modifications to this process depending on substrate allowing for the formation of both the kinetic and thermodynamic products.[70]

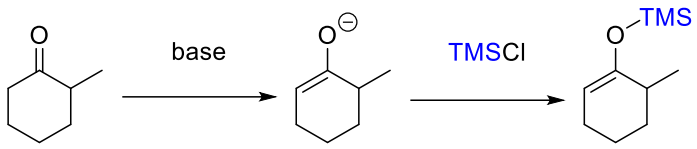
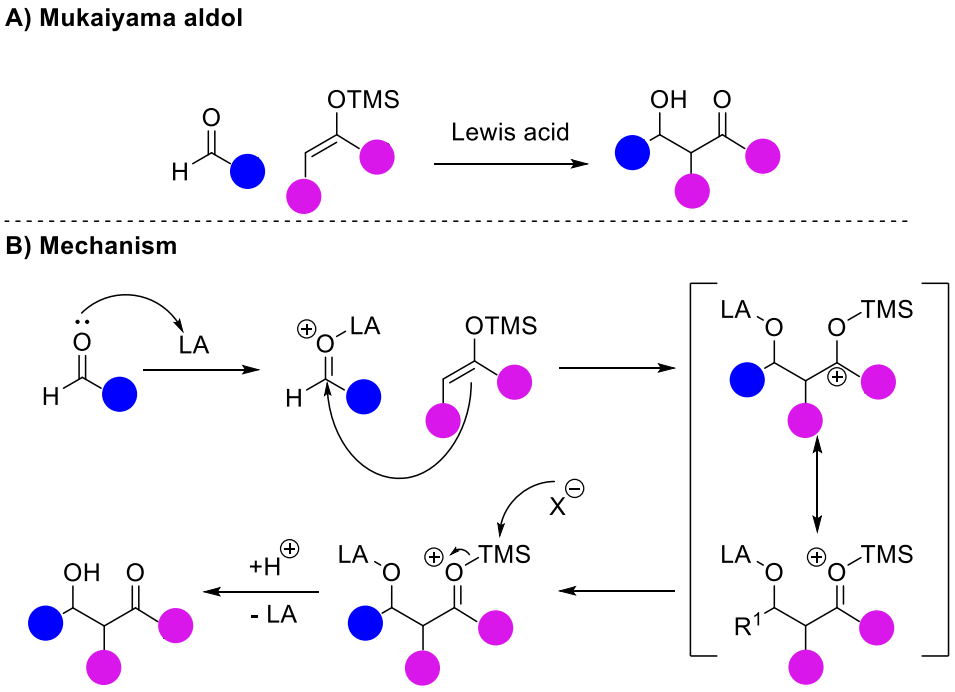


Figure 8: formation of silyl enol ethers

Originally employed as precursors to produce specific enolates, silyl enol ethers have since been further explored in synthisis.[71] They have since become, well-studied and well used surpassing many other enolate derivatives. Some of the key advantages are the ease of formation, mild conditions to desilylate, and high selectivity.



Scheme 7: Mukaiyama aldol addition

An example of a silyl enol ether in synthesis is in the Mukaiyama Aldol addition (**Scheme 7A**). A variant of the traditional aldol reaction this metal catalysed process uses a silyl enol ether in the presence of a Lewis acid to form a 1,3-keto alcohol.

Mechanistically the Mukaiyama aldol is similar to the conventional aldol as shown in **Scheme 7B**. The reaction begins with the attack of the silyl enol ether double bond into the Lewis acid activated aldehyde to afford the aldol product. This reaction is very well represented in the literature with may Lewis acid species such as $Tl^{(III)}$, $Fe^{(III)}$, $Fe^{(II)}$, $Cu^{(II)}$ being employed.[72]

1.1.6 Silanes as Reductants

Hydrosilanes are often employed in various ways as reducing agents, to the extent that this area is the most extensively studied use for hydrosilanes. The reason for this is that hydrosilanes are mild, air/water stable, and relatively inexpensive source of hydrides, though these silanes often require activation.[73] As previously discussed in section 1.1.2.2 the Si-H bond is polarised so that the hydrogen is mildly hydridic in character. The different substituents on silicon allow for differing steric and electronic effects that may impact reaction

outcomes. It should also be noted the diverse array of hydrosilanes available allows for comprehensive reaction screening to achieve optimal results.[74] Early work by Anderson in 1958 showed that diethyl and triethyl silane can reduce simple inorganic metal salts back to the free metals.[75] From this early result the field has greatly expanded and with that the conditions these reductions may be performed under has expanded. Conditions include acidic, basic, radical, transition metal catalysed and as part of organocatalysed processes as shown in **Figure 9**.

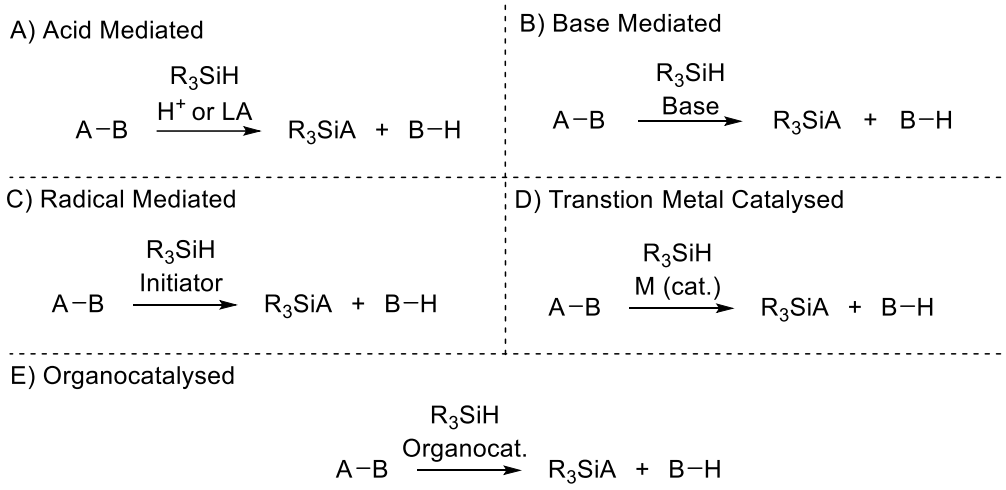
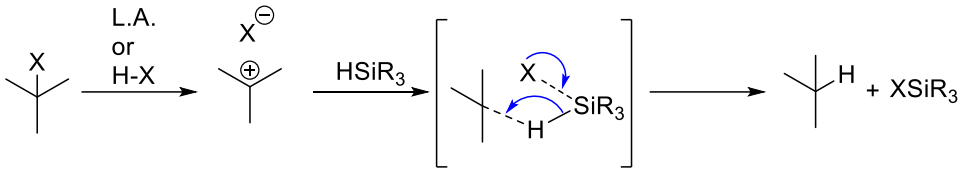


Figure 9: typical conditions for reductions with silanes

Acidic mediated reductions

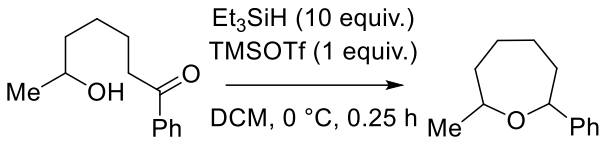
For many substrates silanes are not strong enough reducing agents, in most cases only substrates with high cationic character can be reduced. In these cases, acids can act as an activator to improve reactivity, allowing for the reductions of groups typically to unreactive for hydrosilanes alone.[76] Groups such as aldehydes, ketones and alkenes can be reduced in acid mediated conditions.[76] Mechanistically these reactions proceed *via* the generation of a carbocation of the substrate that can act as acceptor in a hydride transfer step as shown in **Scheme 8**. Detail mechanistic investigations have shown that the hydride transfer is energetically favourable by 33 kJ mol^{-1}.[77] However, due to silicon's inability to stabilise a positive charge trivalent silylium ions are unlikely to be involved in this reaction.[78]



Scheme 8: General pathway for acid mediated silane reduction

A classical example of this type of reduction can be seen by West *et.al.* in 1973, where they demonstrated a selective reduction of aryl aldehydes and ketones using triethylsilane and trifluoracetic acid.[79] This methods is advantageous as it showed a convenient pathway to take the product of a Friedel-Crafts acylation and reduce to the corresponding methylene or methyl group. The weaknesses of this early process are, variable reaction times (0.5-336 h), and side reactions resulting in either γ-lactone formation or a possible Friedel-Crafts alkylation.[79] A different example by Fry *et.al.* use a boron based organic Lewis acid to catalyse the reduction of alcohols to the corresponding alkanes.[80] this work has been greatly expanded upon with later work allowing for the reduction of aldehydes, and ketones via hydrosilylation.[81]

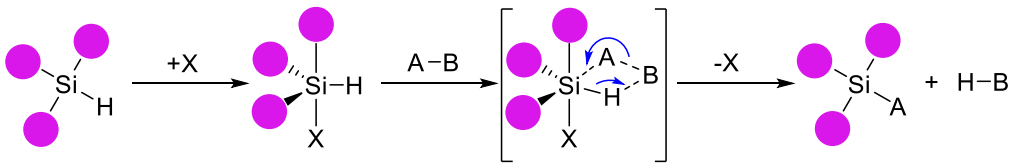
Under acidic conditions it is also possible to perform reductive esterifications and etherifcations using organosilanes. An example by Doyle *et al.* used TFA and triethylsilane to produce esters from aldehydes, while this appears to be an advantageous process is not fully optimised to exclude ether side products.[82] As for etherification both aldehydes and ketones in the presence of silanes and Brønsted acids readily give ether products in good yields.[73] These reactions are very well studied with conditions highly optimised for many different substrates. A notable example is by Nicolaou *et al.* where in their synthesis of brevetoxin B they used diphenyl methyl silane and TMS triflate in the formation of an oxepane as shown in **Scheme 9**.[83]



Scheme 9: Oxepane synthesis with silanes by Nicolaou et al.

Base Mediated Reductions

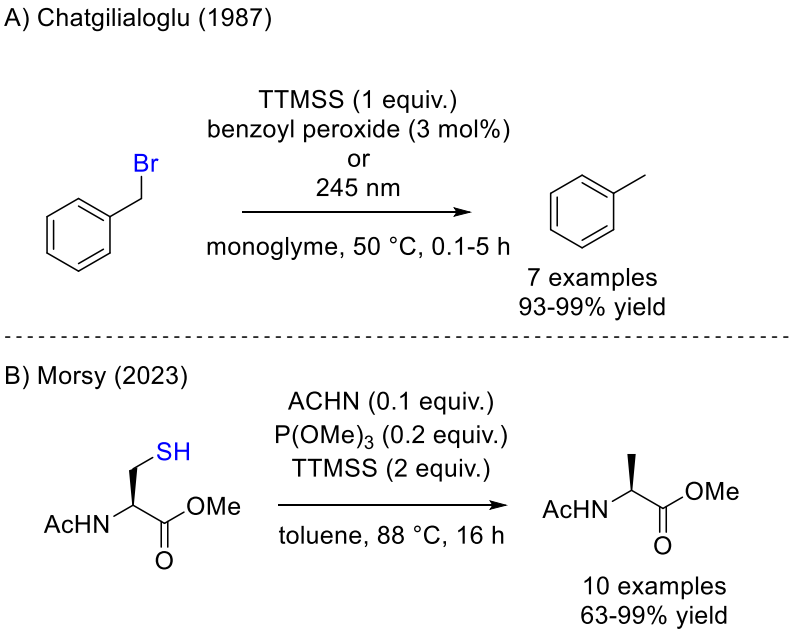
Lewis bases are capable of activating silanes to produces a more reactive source of hydride; this typically occurs through a pentacoordinate silicon species increasing its nucleophilicity. Reactivity can be increased by changing the substituents on silicon, Lewis acidic electronegative substituents are able to support much greater electron density in the hypercoordinate intermediate. From this alkoxy silanes have been shown to be the best performing reducing agents in comparison to the alkyl silanes.[84] Environmental factors can also impact reaction efficiency, solvents in particular Lewis basic solvents like DMF or HMPA can significantly improve reaction rates.[85] For many base mediated reductions the separation of charge means that groups like aldehydes and ketones are greatly favoured over other reductions sensitive groups like nitro groups or amides. The general mechanism as shown in **Scheme 10** for this reduction begins with the formation of a pentacoordinate silicon species through the addition of the base onto silicon, this species then coordinates to the target molecule in a hexavalent transition state. Finally simultaneous exchange occurs to give the reduced compound. An example of this method is shown by Fedorov *et al.* they describe using triethylsilane and K^{t}BuO to reductively cleave diaryl and aryl-alkyl ethers into alcohols.[86] The author applied this method to the selective ring opening of benzofuran groups in lignin based natural products.



Scheme 10: general mechanism for base mediated silane reduction

Radical Mediated Reductions

Si-H bonds in most trialkyl silanes are typically too strong (398 kJ mol^{-1}) for homolysis to occur. However, if the other substituents on silicon are replace with trimethylsilyl groups there is a decrease in Si-H bond strength of around 17 kJ mol^{-1}.[87] $(TMS)_3SiH$ or TTMSS is a mild, low toxicity reagent that can readily undergo homolysis as the Si-H bond with a bond dissociation energy of around 351 kJ mol^{-1}.[87] These properties have allowed TTMSS to be identified as a replacement for tin-based reagents in a number of radical based reactions.[88] On its own TTMSS does not readily react with groups such as acid chlorides, alkenes, and aldehydes but the addition of a radical initiator such as ABIN or photochemical excitation open most reaction pathways.[87,89] A classical example of a radical mediated reduction with silanes was described by Chatgilialoglu *et.al.* who reduced a series of organohalides using TTMSS by ether a photolysis or by initiation with benzoyl peroxide (**Scheme 11A**).[90] This method represents the first use of TTMSS as a reducing agent. It is an advantageous process due to simple reaction conditions and highly convenient reaction times (0.1-2.5 h) the principal limitation of it is the narrow scope demonstrated. A more contemporary method by Morsy *et al.* where they report the metal free reduction of thiols with TTMSS in conjunction with a phosphine catalyst (**Scheme 11B**). In this example the TTMSS in involved in the initiation step and in the termination step that gives the CH_3 group in the product. This process is advantageous as it has is compatible with primary, secondary, and tertiary thiols, has wide functional group tolerance, and short reaction times.

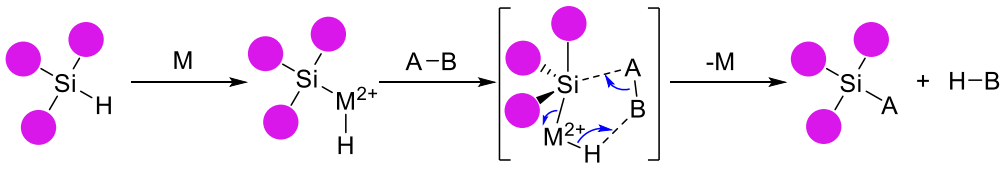


Scheme 11: Radical mediated reductions with TTMSS

Transition metal catalysed reductions

Organosilanes are often employed as alternatives to direct hydrogenation in several different processes. As previously discussed organosilanes are particularly advantageous as they offer milder and safer processes compared to molecular hydrogen or metal hydrides. This makes organosilanes advantageous as a hydride source in metal catalysed reduction reactions, simple copper, zinc and rhodium salts are the most commonly used metal catalysts for these reductions.[74] As for the silanes, there is a diverse array of silanes with differing steric properties available like n-octyl silane, diphenylsilane, and triethyl silane, though each of these has different costs. Polymethylhydrosiloxane (PMHS) is a lower cost polymeric hydridic siloxane that is commonly used in the reduction reactions. Though it can make sample purification more complex as PMHS can be difficult to remove. This class of reduction reactions are well represented in the literature and cover an extremely large area of chemical space from alkenes to more complex functionality imines and lactones.[91,92] Within these reduction reactions the silane is typically involved in the formation of the metal hydride -the active form of the catalyst. This active catalyst then allows for hydride transfer onto the

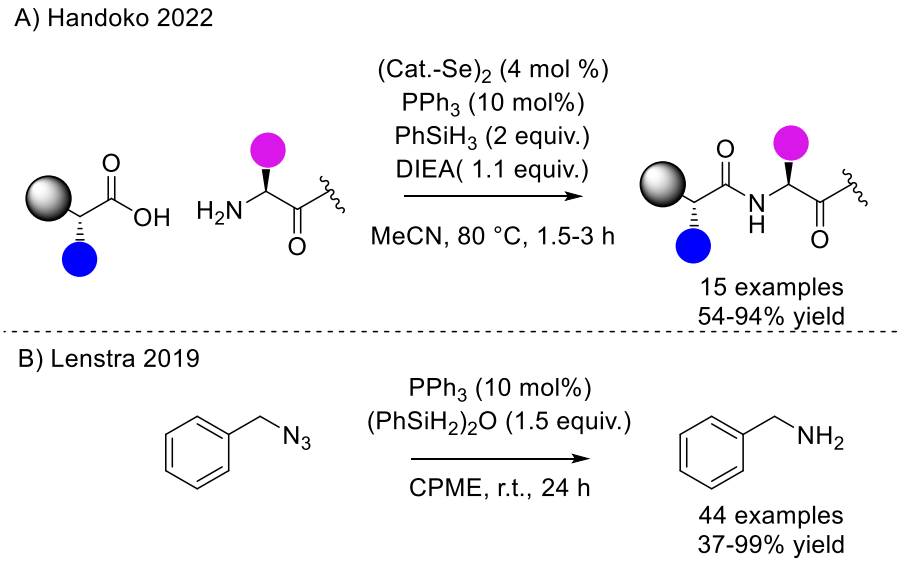
substrate to give the reduced product as shown in **Scheme 12**. A classical example of a metal catalysed reduction with silanes is by Buchwald *et al.* in 1994 where they describe the enantioselective reduction of a ketone with a chiral titanocene catalyst and PMHS, n-butyl lithium and TBAF.[93] A contemporary example is demonstrated by Denton *et.al.* in 2022 in which they reduced a carboxylic acid to a primary alcohol with phenylsilane and zinc acetate. This method is advantageous as it offers a convenient, accessible way to convert carboxylic acids to alcohols in good yields with good functional group tolerance.



Scheme 12: general mechanism for transition metal catalysed reductions with silanes

Reductions in Organocatalysed processes

Since 2000 the development of organocatalysis has allowed for a diverse array of reactions that either, were performed by precious metal catalysis, or are entirely new processes. A contemporary example of an organocatalysed reaction with a hydrosilane based reduction can be seen in the work of Handoko *et al.* in 2022 (**Scheme 13**A). The work shows a hydrosilane as the hydride source for the reduction of a phosphine oxide in a 2-component organocatalyzed redox peptide coupling.[94]



Scheme 13: Organocatalysed reactions featuring silane mediated phosphine oxide reduction.

This use of silane as a reductant for phosphine is a common strategy to reduce the amount of phosphine required in a process, this strategy has been applied to reactions such as the Wittig, Mitsunobu, and Appel reactions.[95–97] Much like in the Handoko example shown above Lenstra and coworkers applied this strategy to the Staudinger reduction, using diphenyl siloxane as the hydride source (**Scheme 13B**).[98] Mechanistically (**Figure 10**) these reductions begins with the hydrosilylation of the phosphine oxide. The silylated intermediate then undergoes elimination of a silanol likely through a hydride transfer from phosphorus onto the oxygen to give the reduced phosphine.[99]

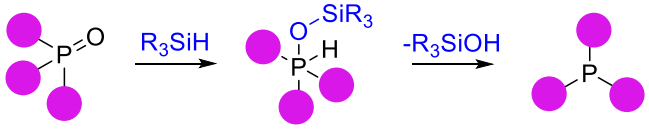


Figure 10: Silane reduction of a phosphine oxide

1.2 Importance of Metal Free Chemistry

The availability of several metals critical to chemical synthesis is increasingly limited and in the case of some endangered. The European Union has identified that Silver, Zinc, Gallium and other to be critically endangered with disruption to supply expected within a century.[100] Other

metals such as platinum, iridium, and ruthenium are at high risk due to increased demand.[100] To reduce the pressure on these elements newer strategies are being employed to make the most of the resources available.[101] This can be done by the following strategies:

1) Reducing the number and loading of metal-based catalysts and reagents used within chemical processes.
2) Improve metal catalyst durability and turnover.
3) Replacing low abundance metals in a chemical process with either high abundance metals
4) Eliminate the use of metal catalysts and reagents in favour of nonmetal catalysts and reagents.

For the purpose of this section, we will explore strategy four.

1.2.1 Limitations of Metals in Synthesis

The use of metals within synthesis is not entirely benign and issues can arise in their use. Some of the issues that can be encountered are varying degrees of air and water sensitivity, high reactivity, toxicity issues with the metal itself.[102] For the last few decades, the levels of metal impurities permitted within pharmaceuticals has been heavily regulated.[102] Palladium catalysis in particular used in cross couplings are a well-known and well understood group of reactions that hold privileged status in organic synthesis. However, upon scaleup to industrial quantities these reactions are often disfavoured due to the risk palladium toxicity, and considerable effort is taken to minimise the levels of palladium contaminants in the final product.[103,104] Another key limitation of metals catalysts are their cost, many metals are subject to economic forces that put price pressures on any catalysts developed from these metals.[105]

To counter some of these issues there has been a push in research to develop process using more earth abundant mentals such as iron and copper. Copper for example has an LD_{50} of 300

mg/kg compared to 3 mg/kg for palladium.[106,107] While the use of these more common metals is an improvement on the use of precious metals, they do not full account for these limitations. By fully removing the metals from a reaction it is possible to bypass their limitations.

1.2.2 Advantages and Examples of Metal Free Chemistry

Key advantages of metal free chemistry are:

- Selectivity
- Recyclability
- Environmental compatibility

In addition to the points above the most important advantage of metal free catalysts is the opening of new processes and new modes of activation.

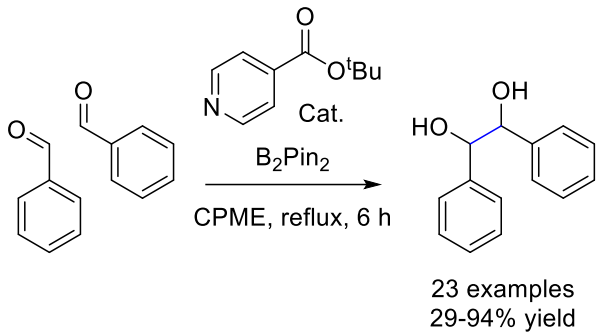
1.2.2.1 Synthetic Examples of Metal Free Chemistry

Significant attention has been given to metal free chemistry in recent years, as such an increasing portion of publish literature demonstrates various metal free reactions. Below are described examples of how metal free chemistry can eliminate metal catalysts and reagents.

Catalysis Processes

In traditional synthesis metals have been used extensively in a multitude of roles, one of the principal roles is as a catalyst as such it can be desirable to replace these metals with an alternative. Of the many methods at replacing metal catalysts, two fields have received extensive attention from research groups globally, organocatalysis and biocatalysis. Firstly organocatalysis i.e. the use of discrete organic molecules as catalysts offers an excellent pathway to remove metals from many reactions. An example of this can be seen by Yasui *et al.* where they demonstrate a metal free thermal pinacol coupling of arylaldehydes using B_2Pin_2 and an isonicotinate catalyst (**Scheme 14**).[108] This gave a series of diol produces in fair to excellent yields with good functional group tolerance. This work offers a convenient metal

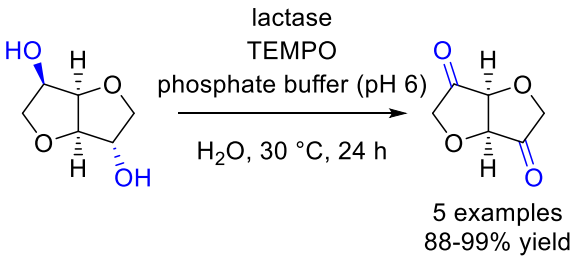
free approach in contrast to more traditional pinacol couplings can make use of many different metal catalysts such as copper, aluminium, iridium etc.[109]



Scheme 14: metal free pinacol coupling by Yasui *et al.*

Biocatalysis make use of enzymes as the catalyst, mimicking a biological process. They key advantage of using biocatalysts are highly tuneable to a given process though the use of directed evolution.[110] Directed evolution is a process that allows for the editing of the enzyme structure though modification of the genetics of the organism used to produce the enzyme.

An example of a metal free biocatalytic process was shown by Gross *et al.* who described a biocatalytic approach to the oxidation of alcohols.[111] This process makes use of a lactase enzyme and TEMPO in pH 5 buffer to perform the reduction. The advantage of this reaction is the simple work up and purification process, as there is no need for chromatography to isolate the products. The disadvantage of this reaction is the very narrow scope described in the paper; it is difficult to ascertain what types of functional groups could be tolerated in this reaction.

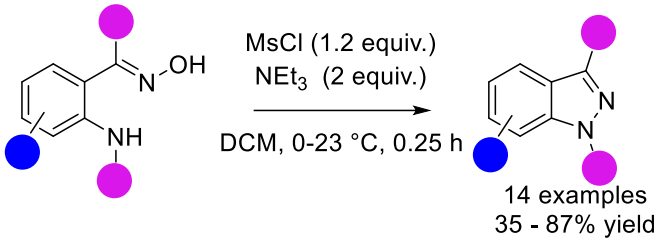


Scheme 15: biocatalytic oxidation of alcohols by Gross *et al.*

Non-Catalytic Processes

Aside from catalysis metal reagents have been used extensively in non-catalytic processes, from $LiAlH_4$ reductions, Magnesium in Grignard reagents, and Zinc in the Simmons-Smith reaction. These reactions while powerful and give synthetically useful products all employ large quantities of metals. While it may not be possible to eliminate all metal reagents from synthesis may there have been developments in which metal-based reagents have been successfully eliminated.

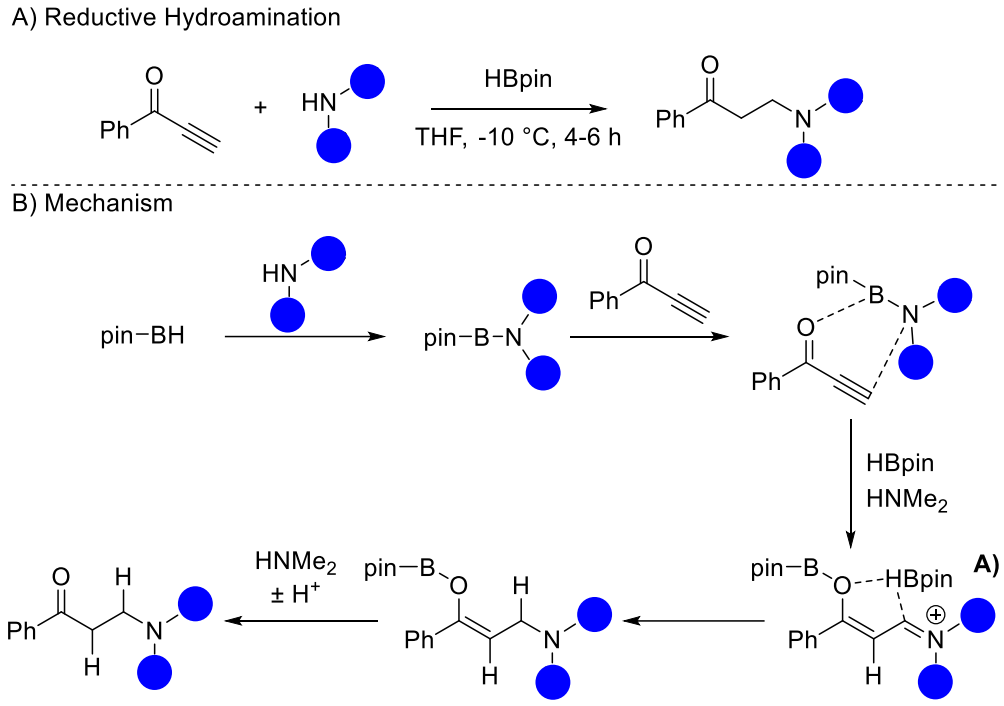
An example of a metal free non catalytic process is from Counceller *et al.* in 2008 where they describe the metal free synthesis of 1H-Indazoles from o-aminobenzoximes.[112] In this process the authors used a slight excess of methanesulfonyl chloride and triethylamine. The reaction proceeds via mesylation of the oxime allowing for the nucleophilic attack of the ortho-amine into the oxime to give the 1H-indazole product. Other synthetic routes to 1H-indazoles involve metal catalysis such as silver or copper to achieve the same effect.[113] The authors successful use inexpensive methanesulfonyl chloride to activate the oxime as an alternative to metal catalysis served as inspiration for other metal free heterocycle synthesises.[114,115]



Scheme 16: metal free synthesis of 1H-indazoles

A more contemporary example of a metal free non catalytic process can be seen by Fu *et al.* in 2022, where they demonstrated a metal free synthesis of β-aminoketones through the reductive hydroamination ynones **Scheme 17A**.[116] In the method developed by the authors the employ HBpin as the reductant. As shown in **Scheme 17B** the reaction proceeds *via* an aminoborane species that coordinates to ynone and reduces the alkyne to the complex **A).**

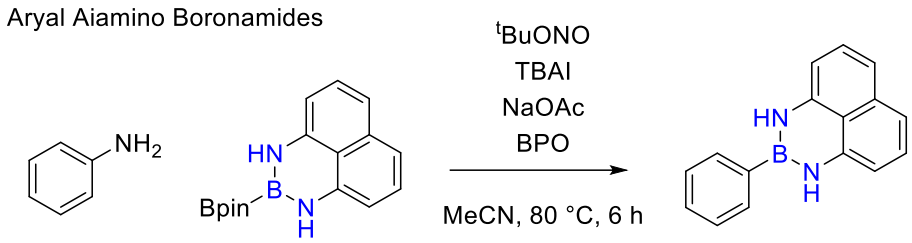
from this complex the ynone is reduced to a iminium enol which coordinates to a second equivalent of HBpin. From this second complex the imine undergoes reduction to the amine. The Bpin-enol intermediate undergoes protonolysis and then tautomerizes to form the final β-aminoketone product.



Scheme 17: metal free reductive hydroamination ynones

The classical method for reductive hydroaminations of ynones typically involve Rhodium, Gold, or Palladium catalysis.[117] This method takes inspiration from hydroboration chemistry to remove the need for a metal catalyst.

A third example has been shown in work by Ding *et al.* in 2019 where they describe the formation of aryl naphthalene- boronamides in a metal free Sandmeyer-type reaction.[118] In this reaction they make use of an unsymmetrical diboron species composed of a B(pin) moiety and a naphthalene-- boronamide (B(dan)) moiety.



Scheme 18: synthesis of aryl naphthalene-1,8-diamino boronamides

This process compares favourably with the metal catalysed process which is a multistep synthesis starting with a Miyaura borylation then the Fe^{II} catalysed addition of 1,8-diaminonaptahline. In reducing the number of steps in a reaction down to a single step the authors have shown how moving to a metal free process.

Building off of the concepts discussed in this introduction subsequent chapters will explore specific cases of metal free organosilicon chemistry.

1.3 References

1 J. J. Berzelius, *Ann. Phys.*, 1824, **77**, 169–230.
2 D. Seyferth, *Organometallics*, 2001, **20**, 4978–4992.
3 N. N. Greenwood and A. Earnshaw, eds. N. N. GREENWOOD and A. B. T.-C. of the E. (Second E. EARNSHAW, Butterworth-Heinemann, Oxford, 1997, pp. 328–366.
4 S. Wang, X. Liu and P. Zhou, *Adv. Mater.*, 2022, **34**, 2106886.
5 R. B. King, *Inorganic Chemistry of Main Group Elements*, VCH, 1995.
6 A. Kramida, Y. Ralchenko, J. Reader and A. T. NIST, NIST Atomic Spectra Database (version 5.10).
7 E. Fritz-Langhals, *Org. Process Res. Dev.*, 2019, **23**, 2369–2377.
8 R. Nougué, S. Takahashi, A. Dajnak, E. Maerten, A. Baceiredo, N. Saffon-Merceron, V. Branchadell and T. Kato, *Chem. – A Eur. J.*, 2022, **28**, e202202037.
9 J. Mullins, *J. Chem. Educ.*, 2012, **89**, 834–836.
10 L. H. Sommer, E. Dorfman, G. M. Goldberg and F. C. Whitmore, *J. Am. Chem. Soc.*, 1946, **68**, 488–489.
11 M. J. Bausch and Y. Gong, *J. Am. Chem. Soc.*, 1994, **116**, 5963–5964.
12 J. B. Lambert, Y. Zhao, R. W. Emblidge, L. A. Salvador, X. Liu, J.-H. So and E. C. Chelius, *Acc. Chem. Res.*, 1999, **32**, 183–190.
13 S. Koda, *Prog. Energy Combust. Sci.*, 1992, **18**, 513–528.
14 J. J. Petkowski, W. Bains and S. Seager, *Life*, 2020, 10.
15 H. Ottosson and P. G. Steel, *Chem. – A Eur. J.*, 2006, **12**, 1576–1585.
16 F. S. Kipping, *Proc. R. Soc. London. Ser. A - Math. Phys. Sci.*, 1997, **159**, 139–148.
17 R. West, M. J. Fink and J. Michl, *Science (80-.).*, 1981, **214**, 1343–1344.
18 A. Sekiguchi, R. Kinjo and M. Ichinohe, *Science (80-.).*, 2004, **305**, 1755–1757.
19 N. Wiberg, S. K. Vasisht, G. Fischer and P. Mayer, *Zeitschrift für Anorg. und Allg. Chemie*, 2004, **630**, 1823–1828.
20 Y. Ding, Y. Li, J. Zhang and C. Cui, *Angew. Chemie Int. Ed.*, 2022, **61**, e202205785.
21 M. Kaupp, *J. Comput. Chem.*, 2007, **28**, 320–325.
22 Hydrocarbons - Autoignition Temperatures and Flash Points.,

https://www.engineeringtoolbox.com/flash-point-autoignition-temperature-kindling-hydrocarbons-alkane-alkene-d_1941.html, (accessed 30 July 2023).
23 M. Janeta and S. Szafert, *J. Organomet. Chem.*, 2017, **847**, 173–183.
24 C. Friedel and J. M. Crafts, *Justus Liebigs Ann. Chem.*
25 L. Rösch, P. John and R. Reitmeier, in *Ullmann's Encyclopedia of Industrial Chemistry*, 2000.
26 B. Pachaly and J. Weis, in *Organosilicon Chemistry III*, 1997, pp. 478–483.
27 G. Fritz, *Angew. Chemie Int. Ed.*, 1987, **26**, 1111–1132.
28 R. Müller, *Zeitschrift für Chemie*, 1985, **25**, 309.
29 A. G. Brook, S. C. Nyburg, F. Abdesaken, B. Gutekunst, G. Gutekunst, R. Krishna, M. R. Kallury, Y. C. Poon, Y. M. Chang and W. N. Winnie, *J. Am. Chem. Soc.*, 1982, **104**, 5667–5672.
30 D. Bravo-Zhivotovskii, V. Braude, A. Stanger, M. Kapon and Y. Apeloig, *Organometallics*, 1992, **11**, 2326–2328.
31 I. S. Toulokhonova, I. A. Guzei and R. West, *J. Am. Chem. Soc.*, 2004, **126**, 5336–5337.
32 D. Gau, T. Kato, N. Saffon-Merceron, A. De Cózar, F. P. Cossío and A. Baceiredo, *Angew. Chemie Int. Ed.*, 2010, **49**, 6585–6588.
33 D. Troegel and J. Stohrer, *Coord. Chem. Rev.*, 2011, **255**, 1440–1459.
34 L. H. Sommer, E. W. Pietrusza and F. C. Whitmore, *J. Am. Chem. Soc.*, 1947, **69**, 188.
35 Y. Nakajima and S. Shimada, *RSC Adv.*, 2015, **5**, 20603–20616.
36 T. Hayashi, S. Hirate, K. Kitayama, H. Tsuji, A. Torii and Y. Uozumi, *J. Org. Chem.*, 2001, **66**, 1441–1449.
37 M. Kaftory, M. Kapon and M. Botoshansky, in *The Chemistry of Organic Silicon Compounds*, 1998, pp. 181–265.
38 F. Weinhold and R. West, *Organometallics*, 2011, **30**, 5815–5824.
39 I. Fleming, in *Molecular Orbitals and Organic Chemical Reactions*, 2010, pp. 69–125.
40 N. N, *Periodicity and the s- and p-Block Elements*, oxford university press, 1997.
41 S. S. Sen, *Angew. Chemie Int. Ed.*, 2014, **53**, 8820–8822.
42 A. C. Filippou, B. Baars, O. Chernov, Y. N. Lebedev and G. Schnakenburg, *Angew. Chemie Int. Ed.*, 2014, **53**, 565–570.
43 S. J. Blanksby and G. B. Ellison, *Acc. Chem. Res.*, 2003, **36**, 255–263.
44 J. Davy and H. Davy, *Philos. Trans. R. Soc. London*, 1997, **102**, 352–369.
45 Y. Peng and W. D. Z. Li, *Synlett*, 2006, 1165–1168.
46 A. M. DiLauro, W. Seo and S. T. Phillips, *J. Org. Chem.*, 2011, **76**, 7352–7358.
47 R. A. Holton, H. B. Kim, C. Somoza, F. Liang, R. J. Biediger, P. D. Boatman, M. Shindo, C. C. Smith and S. Kim, *J. Am. Chem. Soc.*, 1994, **116**, 1599–1600.
48 J. I. Musher, *Angew. Chemie Int. Ed. English*, 1969, **8**, 54–68.
49 W. B. Jensen, *J. Chem. Educ.*, 2006, **83**, 1751.
50 G. N. Lewis, *Valence and the Structure of Atoms and Molecules*, Chemical Catalog Company, Incorporated, 1923.
51 I. Langmuir, *Science (80-.).*, 1921, **54**, 59–67.
52 S. Sugden, *The Parachor and Valency*, G. Routledge, 1930.
53 R. E. Rundle, *J. Am. Chem. Soc.*, 1947, **69**, 1327–1331.
54 G. C. Pimentel, *J. Chem. Phys.*, 1951, **19**, 446–448.
55 C. R. Landis and F. Weinhold, *J. Comput. Chem.*, 2016, **37**, 237–241.
56 J. C. Martin, R. J. Arhart, J. A. Franz, E. F. Perozzi and L. J. Kaplan, *Org. Synth.*, 1977, **57**, 22.
57 V. I. Minkin, 1999, **71**, 1919–1981.
58 A. Rauk, L. C. Allen and K. Mislow, *J. Am. Chem. Soc.*, 1972, **94**, 3035–3040.
59 E. Magnusson, *J. Am. Chem. Soc.*, 1990, **112**, 7940–7951.
60 T. Ooi, D. Uraguchi, N. Kagoshima and K. Maruoka, *J. Am. Chem. Soc.*, 1998, **120**, 5327–5328.
61 M. Yamashita, Y. Yamamoto, K. Akiba, D. Hashizume, F. Iwasaki, N. Takagi and S. Nagase, *J. Am. Chem. Soc.*, 2005, **127**, 4354–4371.
62 E. P. A. Couzijn, A. W. Ehlers, M. Schakel and K. Lammertsma, *J. Am. Chem. Soc.*, 2006, **128**, 13634–13639.
63 S. N. Tandura, M. G. Voronkov and N. V Alekseev, Springer Berlin Heidelberg, Berlin, Heidelberg, 1986, pp. 99–189.
64 A. P. Bento and F. M. Bickelhaupt, *J. Org. Chem.*, 2007, **72**, 2201–2207.
65 Y. Nagai, *Org. Prep. Proced. Int.*, 1980, **12**, 13–48.
66 E. J. Corey and A. Venkateswarlu, *J. Am. Chem. Soc.*, 1972, **94**, 6190–6191.

67 P. Wuts and T. Greene, 2006, pp. 1–15.
68 M. C. Parrott, M. Finniss, J. C. Luft, A. Pandya, A. Gullapalli, M. E. Napier and J. M. DeSimone, *J. Am. Chem. Soc.*, 2012, **134**, 7978–7982.
69 G. Stork and P. F. Hudrlik, *J. Am. Chem. Soc.*, 1968, **90**, 4462–4464.
70 N. C. Dwulet, V. Ramella and C. D. Vanderwal, *Org. Lett.*, 2021, **23**, 9616–9619.
71 P. Brownbridge, *Synth.*, 1983, **1983**, 1–28.
72 J. Matsuo and M. Murakami, *Angew. Chemie Int. Ed.*, 2013, **52**, 9109–9118.
73 G. L. Larson and J. L. Fry, in *Organic Reactions*, 2008, pp. 1–737.
74 G. L. Larson and R. J. Liberatore, *Org. Process Res. Dev.*, 2021, **25**, 1719–1787.
75 H. H. Anderson, *J. Am. Chem. Soc.*, 1958, **80**, 5083–5085.
76 M. P. Doyle, D. J. DeBruyn and D. A. Kooistra, *J. Am. Chem. Soc.*, 1972, **94**, 3659–3661.
77 G. G. Hess, F. W. Lampe and L. H. Sommer, *J. Am. Chem. Soc.*, 1965, **87**, 5327–5333.
78 G. Bertrand, *Science (80-.).*, 2004, **305**, 783–785.
79 C. T. West, S. J. Donnelly, D. A. Kooistra and M. P. Doyle, *J. Org. Chem.*, 1973, **38**, 2675–2681.
80 M. G. Adlington, M. Orfanopoulos and J. L. Fry, *Tetrahedron Lett.*, 1976, **17**, 2955–2958.
81 M. P. Doyle, C. T. West, S. J. Donnelly and C. C. McOsker, *J. Organomet. Chem.*, 1976, **117**, 129–140.
82 M. P. Doyle, D. J. DeBruyn, S. J. Donnelly, D. A. Kooistra, A. A. Odubela, C. T. West and S. M. Zonnebelt, *J. Org. Chem.*, 1974, **39**, 2740–2747.
83 K. C. Nicolaou, C. K. Hwang and D. A. Nugiel, *J. Am. Chem. Soc.*, 1989, **111**, 4136–4137.
84 L. Horner and J. Mathias, *J. Organomet. Chem.*, 1985, **282**, 155–174.
85 Y. Goldberg, E. AbĀbele, M. Shymanska and e. Lukevics, *J. Organomet. Chem.*, 1989, **372**, C9–C11.
86 A. Fedorov, A. A. Toutov, N. A. Swisher and R. H. Grubbs, *Chem. Sci.*, 2013, **4**, 1640–1645.
87 C. Chatgilialoglu, *Chem. – A Eur. J.*, 2008, **14**, 2310–2320.
88 M. P. Sibi, Y.-H. Yang and S. Lee, *Org. Lett.*, 2008, **10**, 5349–5352.
89 C. Chatgilialoglu, *Acc. Chem. Res.*, 1992, **25**, 188–194.
90 C. Chatgilialoglu, D. Griller and M. Lesage, *J. Org. Chem.*, 1988, **53**, 3641–3642.
91 K. A. Nolin, R. W. Ahn and F. D. Toste, *J. Am. Chem. Soc.*, 2005, **127**, 12462–12463.
92 L. Hu, Y. Zhang, G.-Q. Chen, B.-J. Lin, Q.-W. Zhang, Q. Yin and X. Zhang, *Org. Lett.*, 2019, **21**, 5575–5580.
93 M. B. Carter, B. Schiott, A. Gutierrez and S. L. Buchwald, *J. Am. Chem. Soc.*, 1994, **116**, 11667–11670.
94 Handoko, N. R. Panigrahi and P. S. Arora, *J. Am. Chem. Soc.*, 2022, **144**, 3637–3643.
95 C. J. O'Brien, F. Lavigne, E. E. Coyle, A. J. Holohan and B. J. Doonan, *Chem. – A Eur. J.*, 2013, **19**, 5854–5858.
96 H. A. van Kalkeren, S. H. A. M. Leenders, C. R. A. Hommersom, F. P. J. T. Rutjes and F. L. van Delft, *Chem. - A Eur. J.*, 2011, **17**, 11290–11295.
97 J. A. Buonomo and C. C. Aldrich, *Angew. Chemie Int. Ed.*, 2015, **54**, 13041–13044.
98 D. C. Lenstra, J. J. Wolf and J. Mecinović, *J. Org. Chem.*, 2019, **84**, 6536–6545.
99 A. M. Kirk, C. J. O'Brien and E. H. Krenske, *Chem. Commun.*, 2020, **56**, 1227–1230.
100 A. J. Hunt, A. S. Matharu, A. H. King and J. H. Clark, *Green Chem.*, 2015, **17**, 1949–1950.
101 M. Monai, M. Melchionna and P. Fornasiero, ed. C. B. T.-A. in C. Song, Academic Press, 2018, vol. 63, pp. 1–73.
102 V. Balaram, *TrAC Trends Anal. Chem.*, 2016, **80**, 83–95.
103 M. Economidou, N. Mistry, K. M. P. Wheelhouse and D. M. Lindsay, *Org. Process Res. Dev.*, 2023, **27**, 1585–1615.
104 K. S. Egorova and V. P. Ananikov, *Organometallics*, 2017, **36**, 4071–4090.
105 *Nat. Catal.*, 2019, **2**, 735.
106 *Patty's toxicology*, 2013, vol. 50.
107 S. Hajimohammadi, S. Gharibi, V. Pourbarkhordar, S. R. Mousavi and H. Salmani Izadi, *Egypt. J. Intern. Med.*, 2022, **34**, 84.
108 M. Yasui, K. Hanaya, T. Sugai and S. Higashibayashi, *RSC Adv.*, 2021, **11**, 24652–24655.
109 M. Nakajima, E. Fava, S. Loescher, Z. Jiang and M. Rueping, *Angew. Chemie Int. Ed.*, 2015, **54**, 8828–8832.
110 E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner and S. L. Flitsch, *Nat. Rev. Methods Prim.*,

2021, **1**, 46.
111 J. Gross, K. Tauber, M. Fuchs, N. G. Schmidt, A. Rajagopalan, K. Faber, W. M. F. Fabian, J. Pfeffer, T. Haas and W. Kroutil, *Green Chem.*, 2014, **16**, 2117–2121.
112 C. M. Counceller, C. C. Eichman, B. C. Wray and J. P. Stambuli, *Org. Lett.*, 2008, **10**, 1021–1023.
113 A. Park, K.-S. Jeong, H. Lee and H. Kim, *ACS Omega*, 2021, **6**, 6498–6508.
114 M. Tsujii, M. Sonoda and S. Tanimori, *J. Org. Chem.*, 2016, **81**, 6766–6773.
115 C. M. Counceller, C. C. Eichman, B. C. Wray, E. R. Welin and J. P. Stambuli, in *Organic Syntheses*, 2012, pp. 33–41.
116 R. Fu, Y. Liu, T. Wu, X. Zhang, Y. Zhu, J. Luo, Z. Zhang and Y. Jiang, *Chem. Commun.*, 2022, **58**, 3525–3528.
117 S. Streiff and F. Jérôme, *Chem. Soc. Rev.*, 2021, **50**, 1512–1521.
118 S. Ding, Q. Ma, M. Zhu, H. Ren, S. Tian, Y. Zhao and Z. Miao, *Molecules*, 2019, 24.

Chapter One

Efficient Silane Mediated

Amidations using Carboxylic Acids

Amides are an important class of compounds with a diverse array of approaches to their synthesis. however, many of these approaches have significant drawbacks. Herein is described the development of an efficient phenylsilane mediated amidation and its application to the synthesis of natural products and an API.

2.1 Introduction

2.1.1 Fundamentals of Amides

The amide functional group is one of the most common in organic chemistry. It is a privileged group within the pharmaceutical industry with 25% of pharmaceuticals containing amides.[1] Amides are also highly prevalent in nature and are present in may natural products. In addition, amides form the basis of proteins though the linkage of amino acids. The synthesis of amides has over the last century been significant and due to the prevalence of amides within fine chemicals this focus has continued on into contemporary research.[2]

Amides have several key properties, first is that the nitrogen of the amide has almost no basic character when compared to amines, the pK_a of a primary amide is approximately 18 and it's pKaH is approximately -1.[3,4] Secondly, amides are far less susceptible to hydrolysis when compared to esters. The principal reason for this is the resonance stability inherent to amides as the lone pair of electrons nitrogen can delocalise, forming a conjugated system. As a result of this amides require much harsher hydrolysis conditions.[5] Another important property of amides relates to hydrogen bonding, much like esters the highly polarised C=O bond means that the oxygen is a H-bond acceptor while the N-H acts as the H-bond donor.

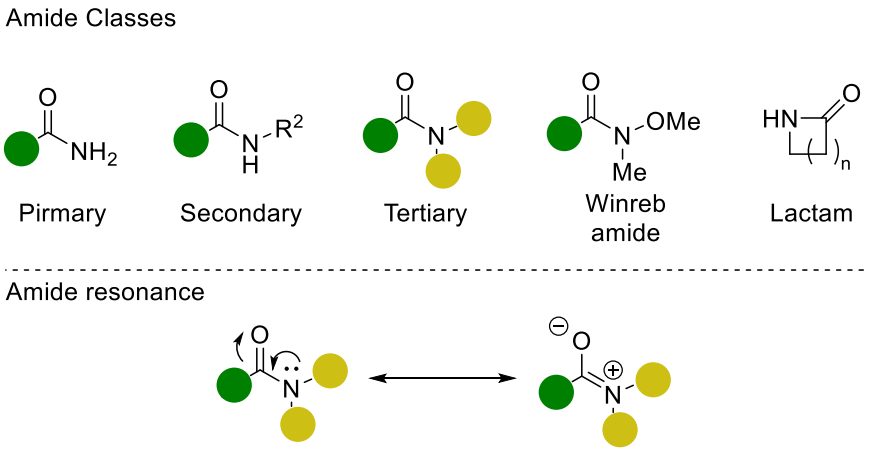
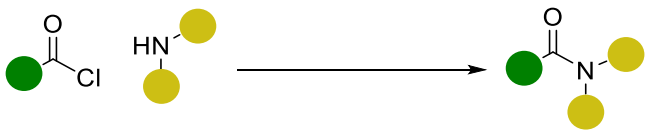


Figure 11: amide classes and amide resonance

An amide can take several forms based mostly around the substitution at nitrogen well known to synthetic chemists (**Figure 11**). Outside conventional amides there are two special classes, Weinreb amides are the first of the two, originally developed in 1981 by Nahm and Weinreb as the key reagent for the Weinreb ketone synthesis.[6] A Weinreb amide differs from others as the it is significantly more susceptible to nucleophilic attack, without over addition that is typical of other amides. The second special class of amides to examine is the lactam, these heterocyclic amides have unique properties that differ from conventional amides. Lactams have varying degrees of stability that are dependent on ring size. A study by Irmming *et al.* showed that lactams with either 4 or 6 membered rings are more sensitive to hydrolysis due to ring strain preventing resonance.[7] Other lactams including 5 membered or 7 membered and larger have greater resistance to hydrolysis.

2.1.2 Classical Amidation Reactions

The most common occurrence of an amide bond in nature is in polypeptides where amide bonds constitute the major linkages between all amino acid subunits. Nature has developed many sophisticated ways of making and breaking amides for protein synthesis. In chemical synthesis this is very much an ongoing process. There is an extremely diverse array of methods available for the synthesis of amide. A common theme within many early strategies is to have a readily reactive electrophile for the nucleophilic amine to attack.

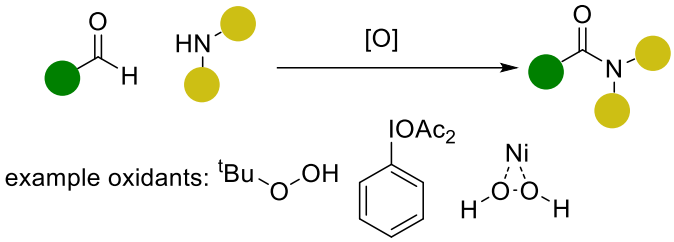


Scheme 19: Schotten-Baumann reaction

The first strategy to examine is the Schotten-Baumann Reaction (**Scheme 19**), it is the oldest amidation reaction and is a common way to develop amides and peptides. Developed in 1884 the reaction makes use of an acid chloride in the presence of a base, the chloride makes an

excellent leaving group so no further activation of the carbonyl is needed.[8,9] This reaction gives excellent yields and a large scope. The Schotten-Baumann Reaction was developed further and applied to amino acids by Fisher in the 1900's to produce peptide chains in the Fisher peptide synthesis.[10]

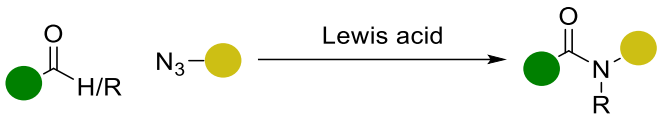
The next strategy is an oxidative amidation reaction as shown in **Scheme 20**, the first example is from the 1960's where Nakagawa and colleagues demonstrated an oxidative amidation using an aldehyde and amine with nickel peroxide as the oxidant.[11] This method has been extensively studied in the literature with many approaches based on metal catalysis, organocatalysis and photocatalysis. However, the key to oxidative amidation reactions is that it relies on activation of the aldehyde by either a hemiaminal or radical intermediate then oxidation to the amide.[12,13] The key issue with these approaches is they often employ difficult to handle materials such as nickel hydrogen peroxide or require an extra synthetic step in the production of the starting material which its typically produced from a carboxylic acid.



Scheme 20: Oxidative amidation reaction

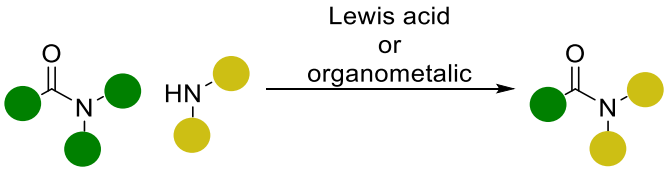
The Schmidt reaction (**Scheme 21**) developed in the 1920's uses ketones or aldehydes as the electrophile with azides in the presence of a Lewis acid to produce amides.[14,15] This reaction is well studied with a broad scope and good functional group tolerance. The mechanism for this reaction is disputed as dependant on conditions and azide source used in the reaction can proceed by either a Beckmann type pathway or by a Bayer- Villiger pathway.[16] The Schmidt reaction is also a very common method of producing lactams, an example by Aube in 1991

shows an intramolecular Schmidt reaction of an alkyl azide using titanium tetrachloride as the Lewis acid to generate lactams in excellent yield.[14]



Scheme 21: Schmidt Reaction

Another strategy is transamidation reactions using a preexisting amide as the electrophile in the presence of either a Lewis acid or an organometallic reagent. This reaction is classically slow due to the inherent stability of the amide group, but the additives used can modulate the rate of reactivity.[17] There are a diverse array of Lewis acids and organometallics that can mediate this reaction, there are even transition metal catalysed examples.[18,19] A recent example by Ali *et al.* showed a transamidation reaction of a secondary amide using inexpensive aluminium(III) oxide as a catalyst alongside triethylamine.[20] This process is advantageous as it takes advantage of the amphoteric nature of the catalyst, allowing for the simultaneous activation of both the amide and the amine to produce the new amide.



Scheme 22: transamidation reaction

Other amidation reactions include the Beckmann rearrangement, the Ugi Reaction, the Ritter reaction, the Chapman rearrangement, and the Leuckhart amide synthesis.[21–25] The diversity of amidation methods shows how important the amide functional group is and the efforts researchers will go to improve the methods used.

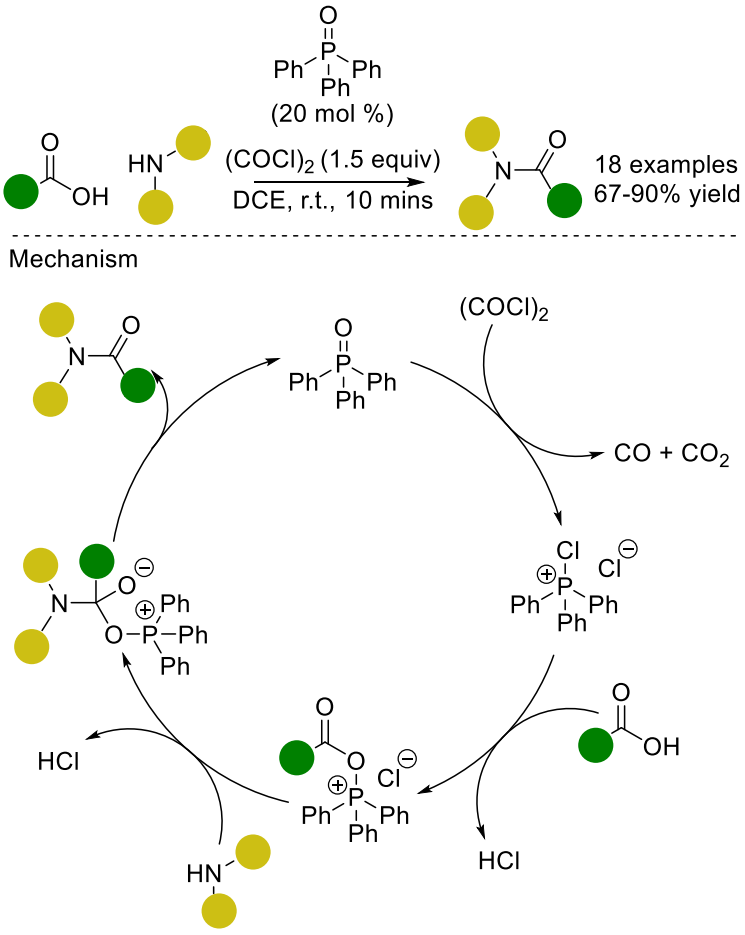
Amidations using Carboxylic Acids

While many of these early amidation reactions have excellent performance, they are hampered by the electrophiles used, many need to be prepared from existing carboxylic acids. This can add to both time and resource costs making these processes less desirable. Switching to a carboxylic acid as the nominal electrophile can be advantageous as they are relatively inexpensive and in high abundance. There much larger and more diverse pool of carboxylic acids to draw from compared to esters, ketones, aldehydes, and alcohols. The ideal amidation reaction would be the direct coupling of a carboxylic acid to an amine in a condensation reaction. However, while this reaction does exist, it suffers from poor yield and a very narrow scope. This is due to the ease with which an ammonium carboxylate salt can form, and the simple fact that a carboxylic acid itself is a poor electrophile. To overcome this limitation, various strategies have been employed. The key to all the strategies is the activation of the carboxylic acid as the electrophile using a "coupling agent". These coupling agents can be broken down into families based on the mechanism of activation with respect to the carboxylic acid.

Chlorinating agents

The first of these are the chlorinating agents such as thionyl chloride, oxalyl chloride and phosphorus oxychloride.[26] This typically allows for the *in-situ* generation of the acid chloride and then proceeds as the Schotten-Baumann Reaction as previously discussed. The principal advantage of this as opposed to the original Schotten-Baumann Reaction conditions is that it occurs as a one pot procedure. The principal disadvantages are firstly that many of these chlorinating agents are highly toxic and great care must be taken when handling them. Secondly these reaction conditions can generate HCl as a side product, this can interfere with acid sensitive functional groups and impair the reaction.

A contemporary example of this type of amidation can be seen by Ren *et al*. in 2021 (**Scheme 23**).[27] Here the authors showed how amidations using oxalyl chloride in the presence of triphenylphosphine oxide catalyst could produce amides without racemisation. This process was used to produce amides and dipeptides in excellent yields with high enantiomeric excess and was shown to work on a large (100 mmol) scale. This reaction proceeds *via* the generation of a chlorotriphenyl phosphonium chloride species. Instead of forming the acid chloride an acyl phosphonium chloride is formed. This highly active species that undergoes nucleophilic attack in very convenient reaction times (10 minutes) without racemisation of the amine.

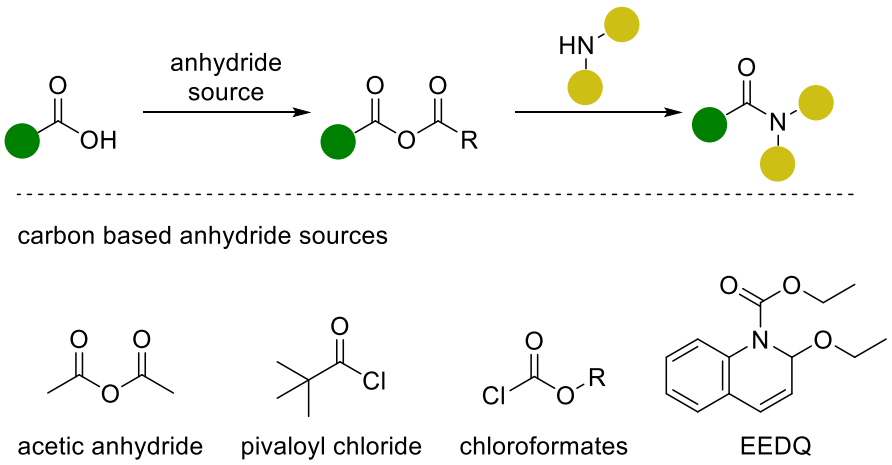


Scheme 23: amidation by Ren *et al.*

Anhydrides

The next family of coupling agents are the anhydrides, these can be generated as either mixed or symmetrical anhydrides. Mixed anhydrides are by far the most common of the two forms for this approach with an extremely diverse array of reagents applicable to the method.[28] The largest family of mixed anhydride are the carbon-based anhydrides. Within the family of mixed anhydrides are two classes.

First are the mixed carboxylic acid anhydrides (**Scheme 24**). The typical way to generate this type of anhydride is through the addition of the reagents such as acetic acid anhydride or pivaloyl chloride to a carboxylic acid in the presence of a non-nucleophilic base such as *N*-methyl morpholine (NMM).[28,29] This method has two disadvantages, one is the lack of control as to which half of the anhydride forms the final amide. Though it should be noted that this can be overcome through a more sterically bulky reagent. The other is the likelihood of the symmetrical anhydride forming as a competing side product.[28]

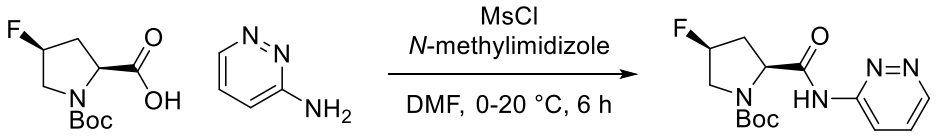


Scheme 24: amidation reactions using mixed anhydrides

The second class are the mixed carbonic acid anhydrides. These are produced from reacting a carboxylic acid with reagents such as chloroformates, or 2-ethoxy-1-ethoxycarbonyl-1,2-dihydroquinoline (EEDQ).[30,31] These reagents are relatively benign but in the case of the

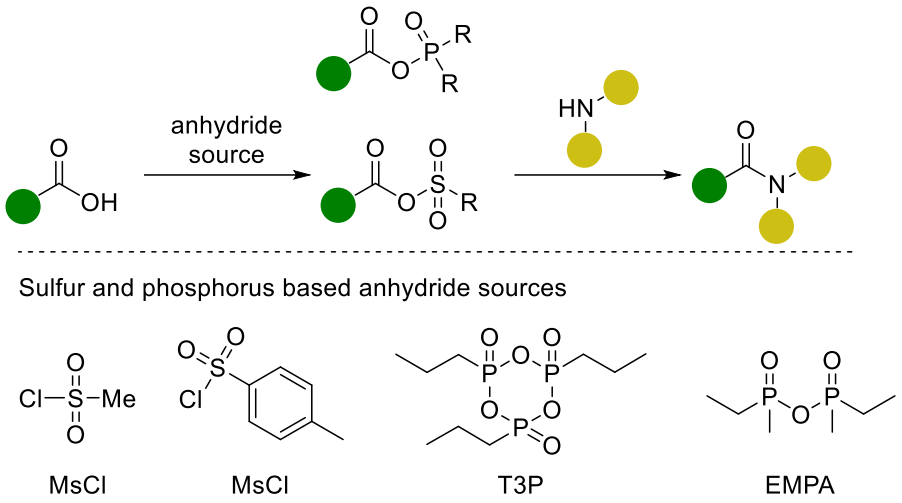
chloroformates can readily give off large quantities of CO_2 and reactions must be properly quenched to prevent pressurisation.[32]

Beyond carbon-based anhydrides, sulfur and phosphorus-based anhydrides are simple and highly effective amidating agents. Examining sulfonate-based anhydrides, this is a well-represented approach in the literature. The two principal reagents used are methane sulfonyl chloride (MsCl) and p-toluenesulfonyl chloride (TsCl). Mechanistically these reagents show excellent selectivity for the nucleophilic attack of the carboxylate over the sulphonate ester. The advantages of this method are firstly they are widely available and inexpensive and secondly, they are highly effective in preventing epimerisation when using a chiral carboxylic acid.[33] The disadvantages of these reagents are that MsCl is a highly corrosive, hygroscopic highly toxic and a lacrimator.[34] An example of this type of amidation can be seen by Liu *et al.* where they produced a peptide deformylase inhibitor, key to this synthesis was the formation of a secondary amide using MsCl (**Scheme 25**).[35]



Scheme 25: anhydride based amidation by Liu *et al.*

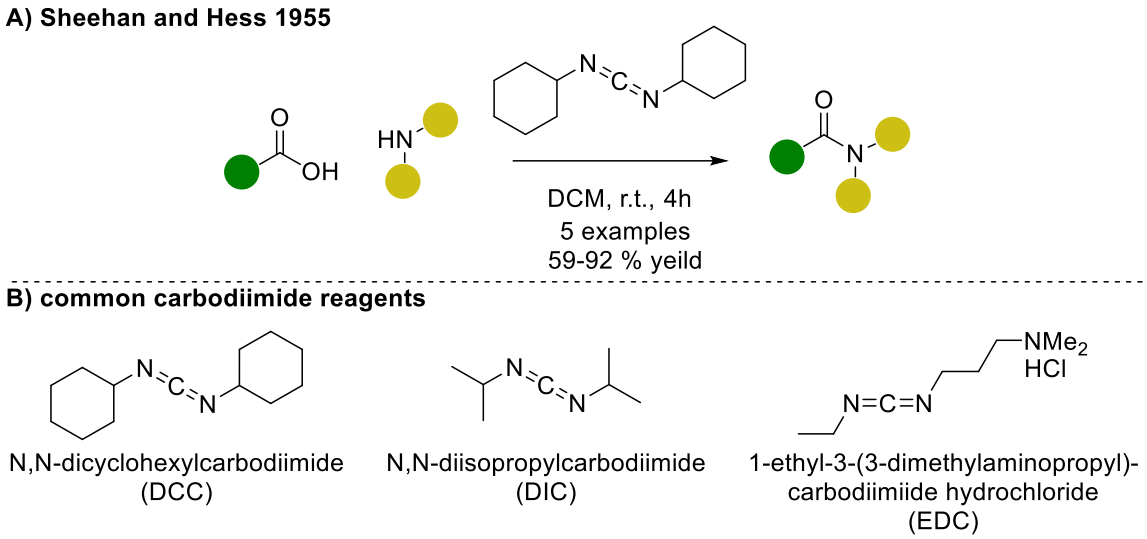
A third class of mixed anhydrides can be derived from phosphorus (**Scheme 26**). Reagents like ethylmethylphosphinic anhydride (EMPA) and *N*-propanephosphonic acid anhydride (T3P) can be used to form carboxylic-phosphoric anhydrides. T3P was first used for amidation in 1980 as described by Wissmann *et.al.* who developed this methods as an alternative to carbodiimides (which will be discussed in a later section).[36] Using T3P is desirable as it has low toxicity, high shelf stability, and is readily available in solution form. synthetically it is highly effective at suppressing epimerisation of acids with an α chiral centre.[37] Its principal disadvantage is the water-soluble phosphonate waste that requires specialist disposal.[28]



Scheme 26: amidations using sulfur and phosphorus-based anhydride sources.

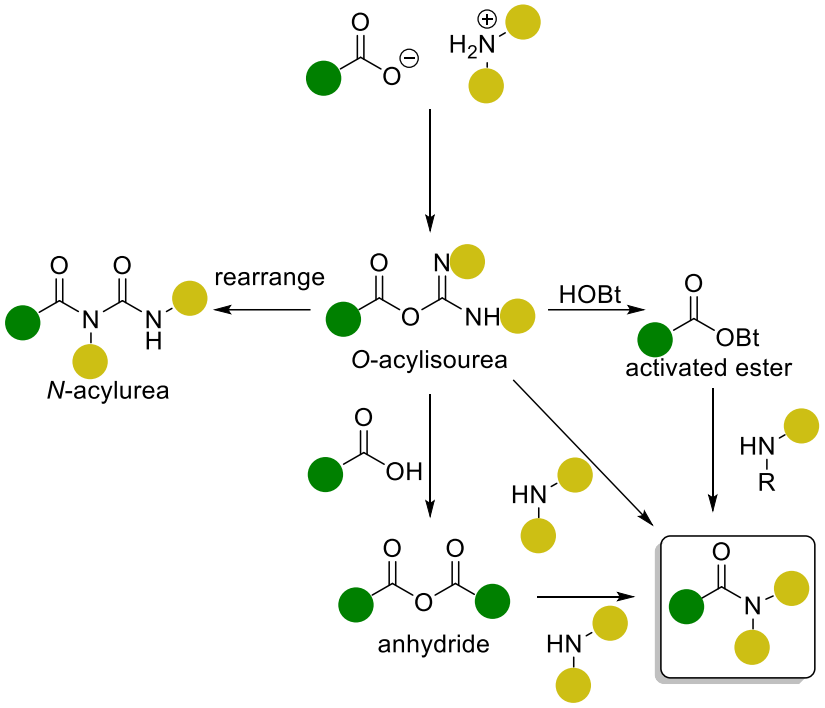
carbodiimides

A popular class of amidation agents are the carbodiimides, the first reported use of a diimide was by Sheehan and Hess in 1955 where they used *N,N*-dicyclohexylcarbodiimide (DCC) the synthesis of dipeptides.[38] Carbodiimides as shown in **Scheme 27**, are well studied with a broad scope and three members of this class have risen to prominence as amidating agents DCC, *N,N*-diisopropylcarbodiimide (DIC) and 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC or EDCI). The advantages of carbodiimides are firstly they are relatively inexpensive, with good solubility except for DCC the byproduct of the reaction is a water-soluble urea which is simple to remove. The disadvantages of this class of amidating agents are that they are severe skin sensitisers and must be handled carefully.



Scheme 27: carbodiimide mediated amidations and common carbodiimides

The mechanism for the carbodiimide mediated amidation (**Scheme 28**) begins with the formation of an ammonium carboxylate salt. The carboxylate ion will react with the carbodiimide to from a O-acylisourea, this intermediate is a highly reactive acylating agent. From this point there are several pathways that the reaction can follow. The first is nucleophilic attack of the amine *via* aminolysis to give the desired product. The second occurs when there is an excess of carboxylic acid, the *O*-acylisourea can react with the carboxylic acid to form the symmetrical anhydride. This anhydride is also susceptible to nucleophilic attack by the amine to give the desired product. The third is the rearrangement of the *O*-acylisourea to an *N*-acylurea, this pathway is irreversible and prevents the formation of the desired product.



Scheme 28: carbodiimide mediated amidation pathways

Diimide mediated amidation can also be modified to resist the rearrangement, this is done *via* the addition of an auxiliary nucleophile such as 1-hydroxybenzotriazole (HOBt) or 1-hydroxy-7-azabenzotriazole (HOAt) (**Figure 12**).[39] Mechanistically this creates a new pathway for the reaction as the *O*-acylisourea can react with HOBt to form an activated ester intermediate. The formation of the activated ester is significantly faster than the formation of the *N*-acylurea. The disadvantages of using HOBt and its derivatives is that its very shock sensitive, as such there are transport restriction on it.[40]

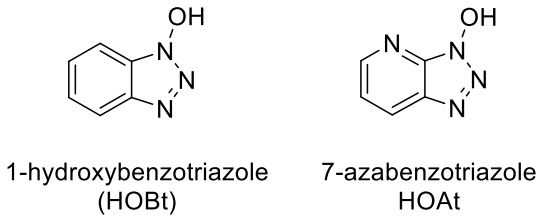


Figure 12: benzotriazole additives

Guanidinium and Uronium Salts

These triazole based salts are fast acting and convenient agents for amidations, in particular they are useful in the activation of sterically hindered carboxylic acids (**Figure 13**).[28] There is a diverse array of guanidinium and uronium salts capable of performing and amidation reaction, but the commonly used regents of this family are HATU ((*N*-[(dimethylamino)-1H-1,2,3-triazolo[4,5-b]pyridin-1-ylmethylene]-N-methylmethanaminium hexafluorophosphate N-oxide)) and HBTU.[41]

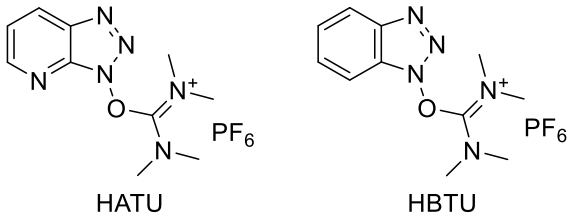
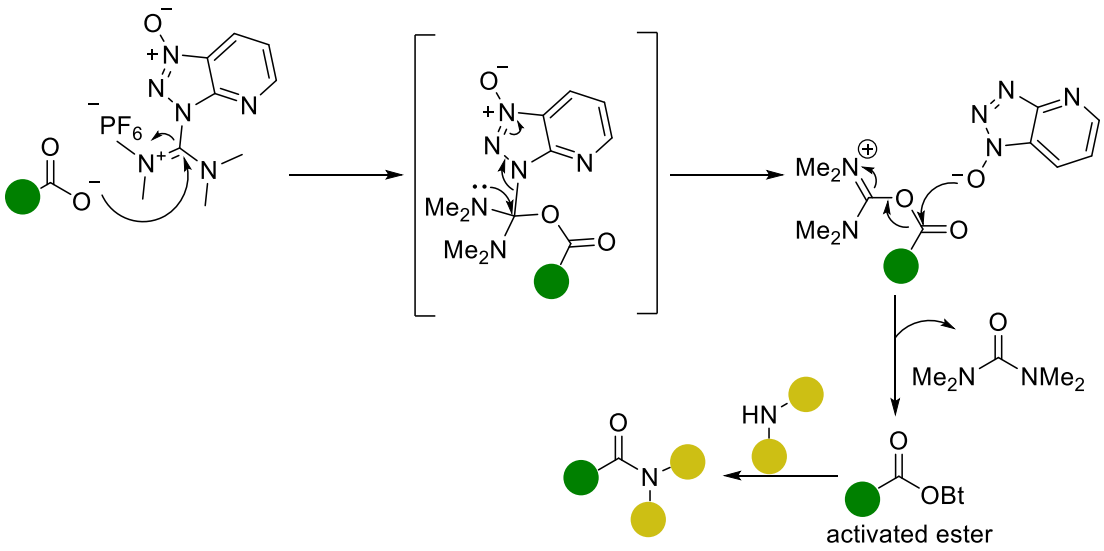


Figure 13: guanidinium and uronium salts

The mechanism for this amidation (**Scheme 29**) begins with the formation of an isourea species from carboxylate, the isourea is then attacked by the oxytirazoleate to form the activated ester. This ester is readily attacked by the amine to give the desired product.

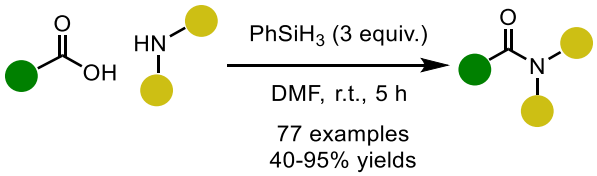


Scheme 29: HATU/HBTU mediated amidation mechanism

The key advantage of this method is that it produces the desired products in expedient reaction times and has a simple work up and purification process. The key disadvantage is the high molecular weight of these reagents, as such it leads to low efficiency reactions as the coupling agent mass is lost. This is also true of other coupling agents, however in the case of guanidinium and uronium salts it is more significant and directly translates to a high cost per mole.[42]

2.1.3 Silane Mediated Amidations

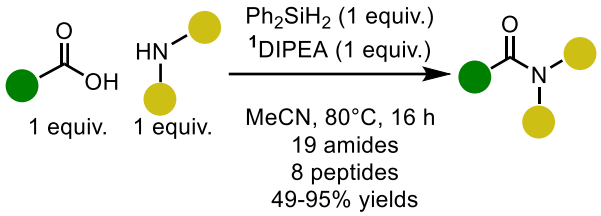
A series of papers by the Mukaiyama group in the mid-2000s demonstrate the use of organosilanes in etheric solvents in direct amidations showing good scope across major classes of acid-amine coupling.[43–45] While these reagents performed well and produced silicon dioxide as opposed to water, each method was limited as they required complex workup and purification. Since this early work there has been significant attention given to organosilanes as amidating agents. There are now many examples of silane-based reagents being used to produce amides shown below are several key examples.



Scheme 30: Ruan et al. amidation approach

More recent work in this area has looked at methoxysilanes and hydrosilanes as the sole coupling reagent. The first reported use of phenyl silane as a coupling agent was by Ruan *et al* (**Scheme 30**). Which showed excellent performance under inert conditions from a matrix of 77 amides derived from 7 carboxylic acids and 11 amines, with yields ranging from 40% to 95%.[46] It should be noted that the authors did find that sterically hindered amines and anilines did not produce amides. The authors proceeded to demonstrate the amidation on solid

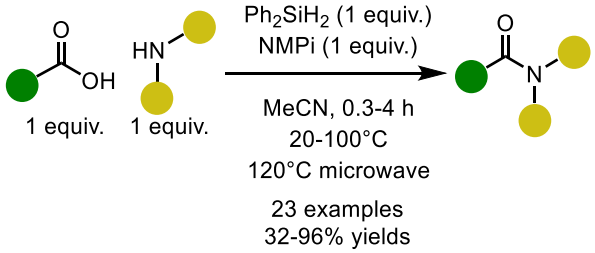
supported carboxylic acids using large excesses of silane (20 mol equiv.) and amine (10 mol equiv.).[46]



Scheme 31: Charette amidation approach

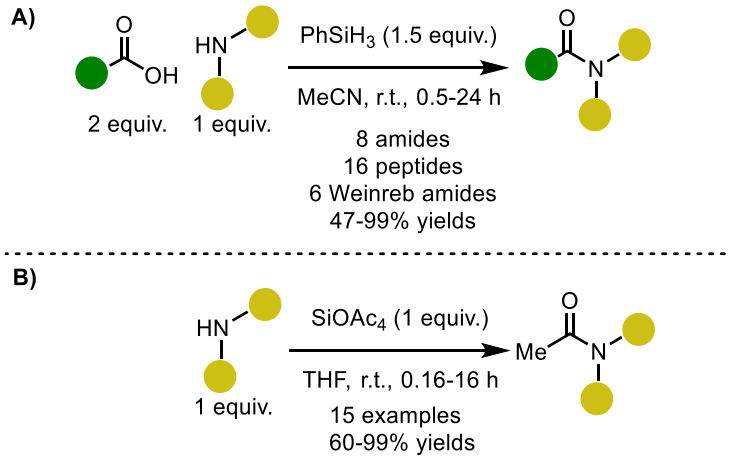
[1]for peptide synthesis only

Within the last few years silanes have received increased attention and as a result many approaches to silane amidations now exist within the literature. The Charette group developed a diphenyl silane mediated approach (**Scheme 31**). They used several mono, di, and tri hydrosilanes to gain mechanistic insights. From these experiments the authors postulated two competing pathways based of a chemical ligation pathway or a carboxylic acid activation pathway having obtained evidence for both. The authors also noted that the aromatic silanes were better able to stabilise intermediates than alkyl silanes. the authors reported a good scope for this reaction with aromatic, benzylic and aliphatic amides with decent functional group tolerance. The authors adapted the amidation to peptide synthesis by the induction of a base additive and producing 4 dipeptides and 4 tripeptides giving moderate to good yields, they noted however that yields drop significantly for tripeptide synthesis and sterically hindered amino acids.[47]



Scheme 32: Adler amidation approach

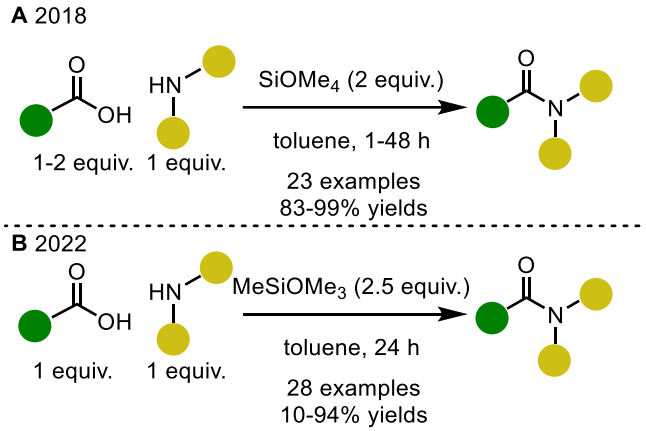
The Adler groups demonstrated a diphenylsilane mediated amidation that does not require anhydrous conditions and can be rapidly performed under microwave irradiation (**Scheme 32**). This was done through the addition of a extra base, from optimisation N-methylpyrrolidine (NMPi) they found that without the need for anhydrous conditions and showed comparable performance under both conventional heating and microwave irradiation.[48] From the substrate scope the authors noted that bulkier substrates were not as effective in the amidation and produced lower yields compared to their less hindered alternative, this reduction in yields was also seen for anilines. The authors performed a Glorious robustness study to examine functional group tolerance toward both the reaction intermediates and the NMPi additive, the results from this showed that aliphatic alcohols and benzaldehyde were incompatible with the reaction as they both reduced the yield and consumed the additive.[48]



Scheme 33: Blanchet amidation approaches

The Blanchet group showed how phenylsilane can act as a coupling reagent with a more benign solvent and expedient reaction times (**Scheme 33A**). Initially screening several phenyl silane analogues, they found phenyl silane had the highest performance and carried that forward to the scope. Producing 8 amides, the authors demonstrated their protocol's ability to make di and tri peptides with 19 examples and the synthesis of Weinreb amides with 6 examples. The authors noted that during the peptide synthesis the amidation with amine

hydrochloride salts significantly reduced yields. Also noted was the influence of peptide chain length which matched results from other studies. The authors additionally used silyl acetates as a as an acylating agent (**Scheme 33**). Initially intended as a surrogate of a potential intermediate for the amidation. This gave insights into how higher order silyl esters *i.e* the bis, tris or tetrakis silyl esters influence both reaction speed and conversion, the higher the order of silyl ester the faster the reaction becomes. The acylation reaction was explored with 15 examples including anilines which required more forcing conditions [49]



Scheme 34: Braddock polymethoxy silane approach

In place of hydridic silanes as coupling agents the Braddock group have demonstrated how methoxy silanes can be used in direct amidations (**Scheme 34**). In 2018 the Braddock group demonstrated how tetramethyl orthosilicate as a coupling agent for amidations, From the scope the authors were able to acquire excellent yields even for extremely bulky carboxylic acids. Functional group tolerance was also of note as the authors were able to use ethanolamine (yield 80%) without a detrimental effect on yields, this shows a possible way to bypass the sensitivity to alcohols that other methods shown here have faced. During their mechanistic study the authors were able to isolate a trimethoxy silyl ester intermediate. More recently in 2022 the Braddock group demonstrated Methyltrimethoxysilane's potential as an alternative reagent in direct amidations. This reagent overcomes tetramethyl orthosilicates toxicity at the cost of higher silane loadings. In the substrate scope the amidation performed

well with fair to excellent yields overall. However, the authors noted several poorly performing amidations including poorly tolerated functional groups such as a benzyl alcohol (yield 14%) in contrast tetramethyl orthosilicates performance with alcohols. The authors noted that the silanol by product from the reaction can undergo a condensation reaction with an unreacted methyltrimethoxysilane to form a siloxane dimer the explaining the need for higher loadings of the silane.[50,51]

Phenylsilane mediated amidation and its mechanism

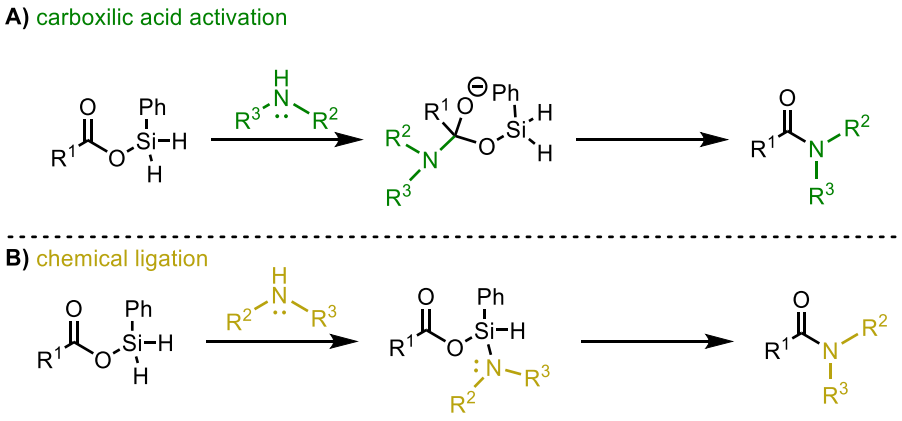


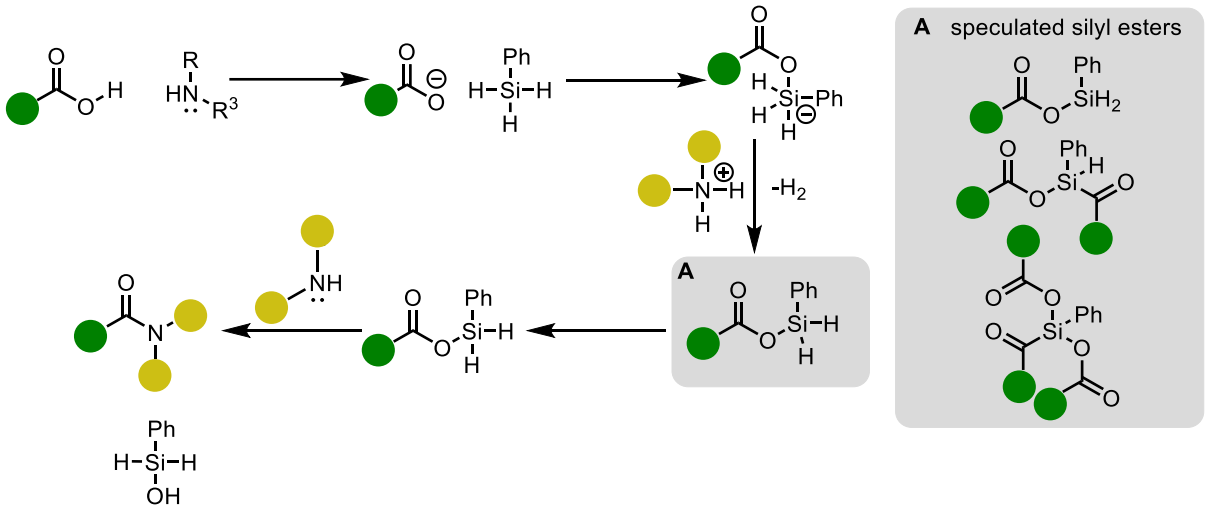
Figure 14: proposed mechanistic pathways for silane amidation

The mechanism for this amidation was the subject of much debate within the published literature, as discussed earlier two different pathways taking prominence (**Figure 14**), the first based of a chemical ligation type mechanism and the second based off an activated carboxylic acid mechanism. Both mechanistic pathways were first proposed by Chan and Wong in their study of silicon tetrachloride as an amidating agent.[52]

The carboxylic acid activation pathway (**Figure 13** section A, **Scheme 35**) can be broken down in to two stages, stage one is the activation of the carboxylic acid *via* a silyl ester species and the second is the nucleophilic attack of the amine. In the first stage after deprotonation by the amine the carboxylic acid will attack onto silane, the resulting anionic species will be quenched by the protonated ammonium species. After forming the silyl ester, the amine will attack the ester carbonyl for simplicity **Scheme 35** only shows the mono ester. On reformation

of the now amide carbonyl the silane will deprotonate the amide nitrogen to form a silanol. It has been shown experimentally that the silanol reacts further to form a polysiloxane. Key to the mechanism is the formation of the silyl ester (**Scheme 35** section A), however experimental work by the Blanchet group uncovered the possibility that either the mono, bis, or tris silyl ester may be responsible for producing the activated carboxylic acid species.[49] The mono silyl ester is the fastest to from but is the least susceptible to nucleophilic attack. The bis silyl ester second to form and has a higher reactivity than the mono but less than the tris. The tris silyl ester is the slowest to form but the most reactive.[53]

The chemical ligation pathway (**Figure 14** section B) begins in a similar manner to the carboxylic acid activation pathway with the formation of a silyl ester species. The next stage in this pathway is the addition of the amine onto the silicon atom. This amino silyl ester then undergoes transfer of the amino group onto the carbonyl of the silyl ester eliminating the silicon to afford the amide product.



Scheme 35: Proposed activated carboxylic acid mechanism for phenylsilane mediated amidation

The chemical ligation pathway was initially the preferred mechanism over the carboxylic acid activation pathway from experimental data from a reaction using salicylic acid and aniline. After hydrolysis only the salicylic acid was recovered and not the amide, this was due to salicylic acid forming a stable chelate with silicon. More recently computational studies show

a preference for the intermolecular carboxylic acid activation pathway, shown in **Scheme 35**. This greater body of evidence has led the carboxylic acid activation pathway to be the preferred mechanism for this reaction.

2.1.4 Immobilized Reagents and Flow Chemistry Fundamentals

Flow chemistry is based on the concept that it is possible for a chemical reaction to proceed in a flowing stream of reagents.[54] This area of chemistry has seen significant attention over recent years and as such there is now a diverse array of chemistry that can be accomplished in a flow reactor. This has greatly expanded the tools available to synthetic chemists. Photochemistry, electrochemistry, and more conventional synthesis can all be done inside a flow reactor cell.

Packed bed/packed column flow reactors are one popular method due to low barrier to entry i.e. simple to set up and run. Fundamentally a packed bed reactor uses a stationary phase to immobilise a reagent or catalyst packed into a short column connected to the flow system. The remaining reagents are in the liquid phase which is pumped though the column. These systems can be made to be recirculating so the regents may be passed though the column repeatedly as a method to improved yield. An example of such a system can be shown in **Figure 15**.[55]

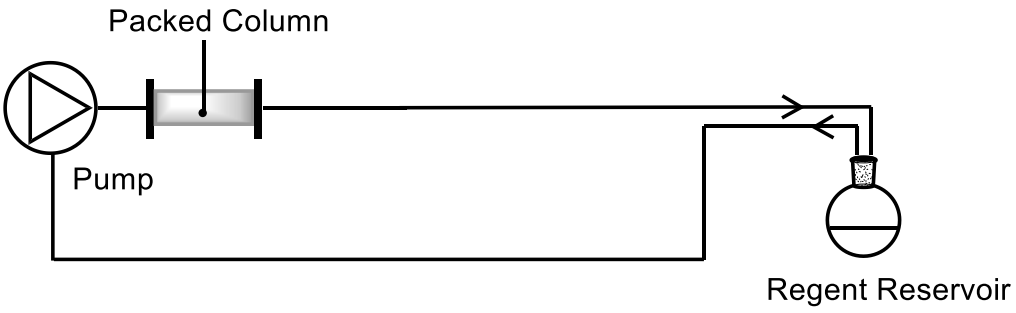
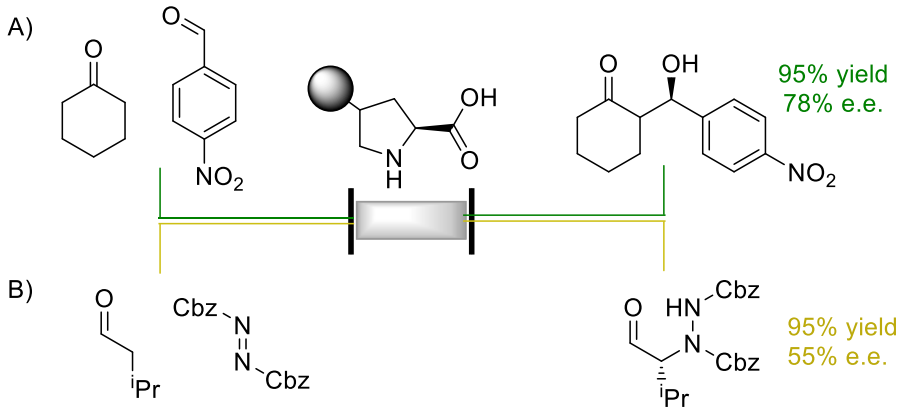


Figure 15: example of a recirculating packed column flow reactor

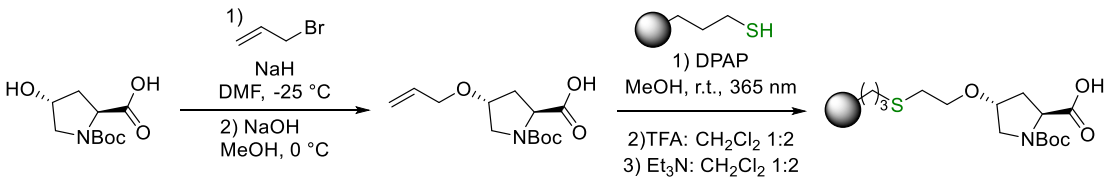
For packed bed reactors the key is the immobilisation of the reagent or catalyst. There are several approaches for this the first being the functionalisation of a porous support or the bespoke polymerisation of a functionalised monomer.

An example reagent immobilisation was demonstrated by Massi *et al.* where they developed a continuous organocatalyzed aldol and α-amination process.[56] This process makes use of a proline derivative bound to silica as the active immobilised catalyst. This process can handle classical aldol conditions (**Scheme 36A**) or with a diazo reagent (**Scheme 36B)**.



Scheme 36: packed column flow process by Massi *et al.*

The hydroxy proline was bound though assembly of an ether/thioether bridge between the mercaptopropyl silica and the hydroxy proline (**Scheme 37**). The advantages of this reaction are firstly the dual functionality in that this process works for both an aldol reaction and an α-amination process. Secondly the short reaction times, with residence times of 10 minutes for the aldol reaction and 30 minutes for the α-amination reaction. The disadvantage of this process is that the scope has not been explored as such functional group tolerance for this reaction cannot be verified.

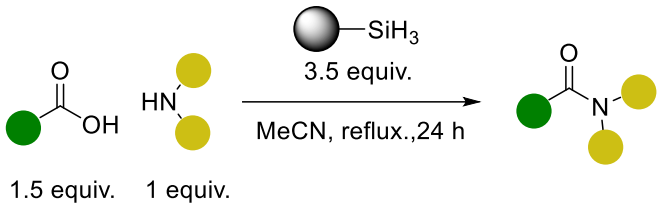


Scheme 37: immobilisation of the proline catalyst

2.2 Immobilised Silanes for Amidation Reactions

2.2.1 Aims and Objectives

Aims



The aims of this work were to develop an immobilised derivative of phenylsilane from ethier a commercially available polymer scaffold or as a bespoke reagent. This functionalised polymer would then be utilised as a reagent in an amidation reaction in both batch and flow procedures.

This strategy is advantageous for three principal reasons. Firstly, it would allow for easy removal of the immobilised reagent thereby simplifying. This would compare favourably with competing processes as the work up and purification processes could be simplified. In addition, it would allow for recovery of the spent silicon regent for later regeneration. Secondly, this work would allow for both batch and flow modality making it a flexible reagent, this could open new . Thirdly, by taking advantage of silicon's low toxicity and ease of handling.

Objectives

The objectives for this work were:

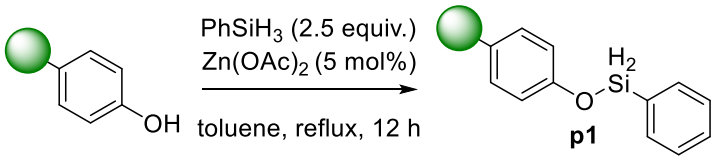
- The development of an immobilised phenylsilane reagent for use in an amidation reaction for in both a batch and flow reactor.
- Development of the batch process silane mediated amidation and its application to the synthesis of natural products and APIs.

- Development of the flow process silane mediated amidation to allow for greater flexibility in reaction set up and access to a continuous synthesis method for amides.

2.2.2 Results and Discussion

Functionalization of commercial polymers

The development of the polymeric system began by examining functionalising commercially available polymers. The polymers selected all contained alcohol to allow for the formation of a dihydridic silyl ether. The functionalisation reaction was a simple procedure involving linking phenylsilane to the polymer support in the presence of zinc acetate, an example of this can be seen in **Scheme 38**. This method is based on a prior work within the Denton group where zinc acetate and phenyl silane were used to reduce an amide. These conditions applied to an alcohol offered a simple way to functionalise polymers with a terminal alcohol.



Scheme 38: silane functionalisation of 4-hydroxy polystyrene using zinc acetate

It was decided that IR spectroscopy would be used to characterise the polymers as the principal IR stretches for the various silicon hydride species are highly diagnostic.[57] The strength of this technique is simple immediate empirical information on the functional groups at silicon. However, there is a critical limitation in that silicon IR spectroscopy gives no information on the polymer e.g. porosity, particle size, or quantification of how much silicon has been bound.

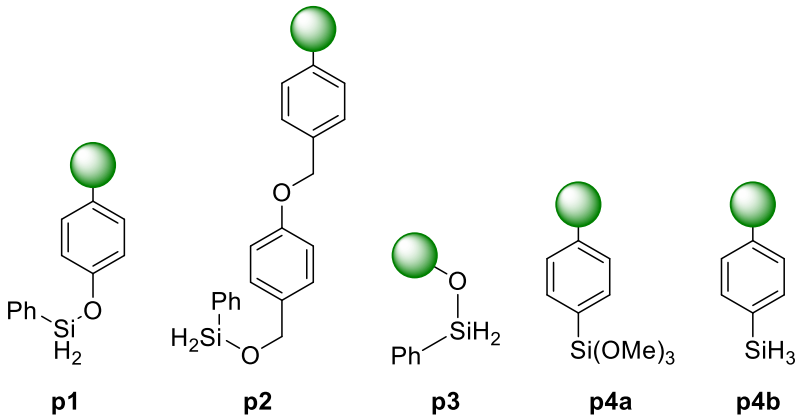
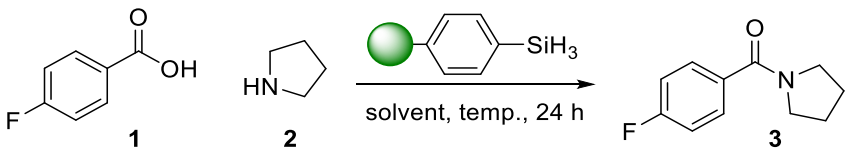


Figure 16: silane functionalised polymers

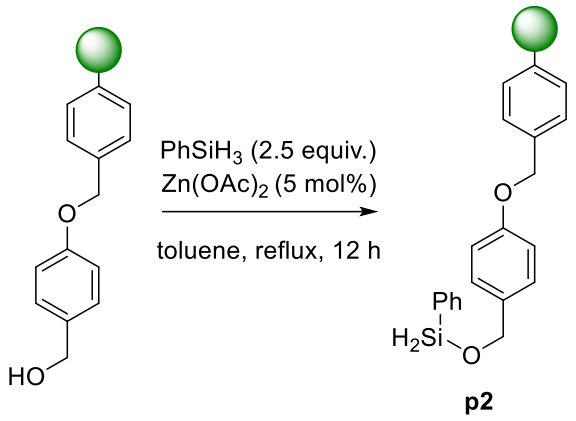
The first example (**Figure 16**) of these was functionalised from (4-hydroxymethyl)polystyrene (**Table 3**, entry 1), the polymer (4-hydroxy)polystyrene was functionalised by covalent attachment of the silane directly onto the alcohol of the polymer to form a silyl ether linkage. IR analysis of the product **p1** showed that while the silane did bond to the alcohol as shown by the Si-O stretch ~1240 cm^{-1} , the polymer was unviable as the Si-H stretch ~2280 cm^{-1} was not present. Test amidation reactions were attempted however no reaction occurred and only the starting materials were detected upon analysis.

Table 3: polymeric silane mediated direct amidation.



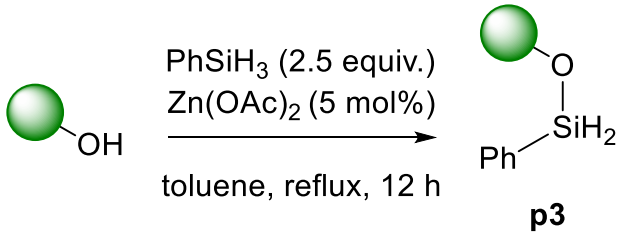
entry	polymer	Polymer loading (mmol)	4-fluorobenzoic Acid (mmol)	Pyrrolidine (mmol)	solvent	Yield (%)
1	p1	1.2	1.6	0.8	acetonitrile	0[a]
2	p1	1.2	1.6	0.8	toluene	0[a]
3	p2	1.2	1.6	0.8	acetonitrile	0[a, b]
4	p4a	1.2	0.4	0.4	toluene	50[c]
5	p4b	2.4	1.6	0.8	toluene	29[c]

[a] yield determined by ^{1}H NMR spectroscopy. [b] commercial variant used. [c] isolated yield.



Scheme 39: synthesis of p2

The second functionalised polymeric system explored was derived from Wang resin **p2** (**Table 3**, entry 3). The resin **p2** was synthesised with the same zinc catalysed reaction as **p1** (**Scheme 39**). A commercially available form of wang resin was first tried but that performed identically to the (hydroxymethyl)polystyrene system. Another route to Wang resin was made starting from Merrifield resin with the intention that the larger pore size may help in polymer performance. However, this polymer did not work as the Si-H ~2280 cm^{-1} stretch was not detected upon analysis. A commercial version of Wang resin was procured, and functionalisation attempted, this polymer was used in a direct amidation, but no product was formed.



Scheme 40: synthesis of p3

The third system explored was based on polyvinyl alcohol as the chosen support, this polymer **p3 (Scheme 40**) was not based on a phenol like in previous polymers and it was hoped that this different structure would be able to produce an active silane species. However, upon

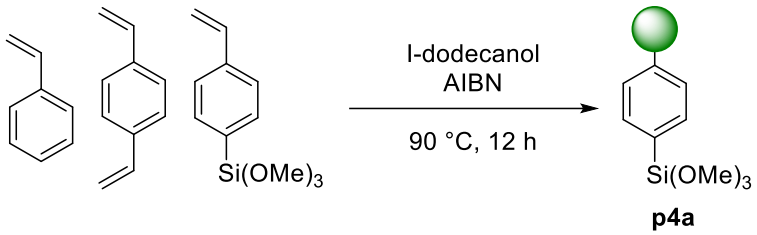
functionalisation IR analysis showed that it had functionalised in a similar manner to previous polymers and was not useable in the amidation process.

The issue with these polymer systems is likely due to the fact that when the alcohol is functionalised the remaining hydrides on the silane are quenched resulting in an inactive product. Future work in this project would be to more accurately study the degree of functionalisation. A simple test would be to thoroughly wash the polymer and extract the remaining silane to quantify the result. The hydrides are critical for the amidation as discussed in section 2.1.3 they allow for the formation of a silyl ester intermediate with the carboxylic acid. Without the hydrides it is not possible to form the key intermediate, and the reaction cannot continue.

Considering the issues found in the system described above, a different approach was needed. This led to the concept of producing a bespoke polymer featuring silane based functional group that could withstand the polymerisation conditions without becoming inactive. To this end trimethoxy(4-vinylphenyl)silane was selected as the functional monomer. This reagent was chosen over a trihydridic silane as it is likely that during polymerisation the hydrides would have been lost and the silicon bound to the polymer chain. The methoxy groups acted as a protecting group for the silane allowing for this process to proceed as planned. Under bulk free radical polymerisation conditions, polymerisation was successful and was possible to produce the silane functionalised polymer **p4a**. Repeat experiments showed that this was a reliable process that could give the desired product with little issue.

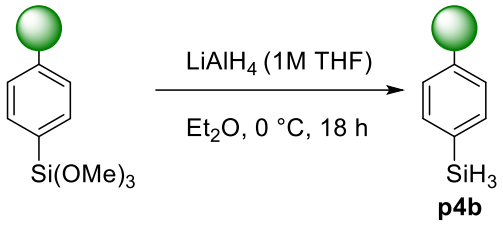
Bespoke polystyrene polymer

The trimethoxyphenysilane polystyrene **p4a** (**Table 3**, entry 5, **Scheme 41**) was used as a rough ground powder for a direct amidations with both pyrrolidine and morpholine and performed well with good yields for an initial unoptimized reaction. This led to an optimisation process that did result in improved yields however fell short of matching phenylsilane's performance.



Scheme 41: synthesis of polymer p4a

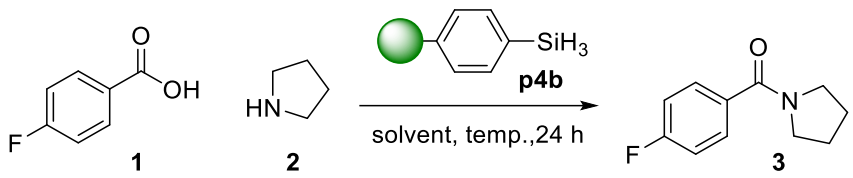
A second viable polymer **p4b** (**Table 3**, entry 6, **Scheme 42**) was created when trimethoxyphenysilane polystyrene **p4a** was reduced with $LiAlH_4$ the reduction was straightforward and yielded a polymer bound variant of phenyl silane. IR spectroscopy showed the loss of the methoxy group, and the presence of the desired hydride stretches.



Scheme 42: reduction of p4a to p4b

A comparison between phenylsilane polystyrene and phenylsilane using IR spectroscopy showed that the critical Si-H stretch was present. This analysis was a strong indicator that the polymer was viable for the amidation reaction. An initial test reaction with 4-fluorobenzoic acid and pyrrolidine gave a yield of 30%. From this initial proof of concept, the polymer was carried forward into the test reaction and underwent optimisation as shown below in **Table 4**.

Table 4: optimisation of polymeric silane mediated amidation



Entry	Solvent	acid (mmol)	amine (mmol)	Silane equiv.	Temp (°C)	Time (h)	Isolated Yield (%)
1	toluene	1.6	0.8	3	110	22	29
2	acetonitrile	1.0	0.5	4	82	18	36
3	acetonitrile	1.0	0.5	4	82	24	58
4	acetonitrile	1.0	0.5	4	82	24	57
5	acetonitrile	1.0	0.5	5.5	82	24	47
6	acetonitrile	1.0	0.5	3.5	82	24	62
7	acetonitrile	1.0	0.5	3	82	24	36
8	acetonitrile	1.0	0.5	10	82	24	0

Changing the solvent to acetonitrile afforded a moderate improvement in yield, as was increasing the reaction time to 24 hours. From there the amount of silane used in each reaction was adjusted to find the optimal loading, the best yield came from 3.5 equivalents of silane. While there was a significant improvement in yields the performance compared to phenylsilane was not sufficient. It should be noted that the silane equivalence in Table 2 refers to the amount of silane monomer used in each batch of polymer. There is a significant probability that some or all of the performance shown above arises from free monomer, and in future work quantifying and isolating any unreacted monomer would be valuable. It is possible that due to the bulk polymerisation process that there could be significant variations

in the polymer structure as such, a more controlled polymerisation may assist in improving reaction performance.

Flow process

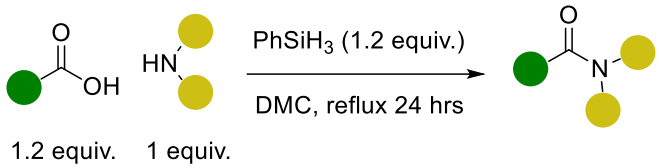
Examining the flow process, a test reaction was performed on a 1 mmol scale with a flow rate of 1 ml per minute in a packed bed flow reactor system based on the most optimal conditions from **Table 4**. However, on completion of the reaction ^{19}F NMR analysis showed that only traces of product had formed. This outcome was less than desirable and meant that significant optimisation work would be required to improve reaction performance, and while it did prove the concept that it is possible for this reaction to work in a flow process.

Overall, the performance of the immobilised reagent was not equal to that of phenylsilane. Based on analysis from the optimisation Tables it is unlikely that the underlying organosilicon chemistry is the cause for this lack of performance. Instead, the likely cause of this is due to the properties of the polymer support in particular the degree and regularity of its porosity. The porosity of this polymer is not formed under any form of control other than the addition of the porogen dodecanol. The irregularity of pores in the polymer could impact the formation of the silyl ester intermediates by adding too much steric interference. Full structural analysis of the polymer would prove valuable in determining how the nature of the polymer porosity. The success of this work can be shown in that an immobilised silane can be used in an amidation reaction, but for this to become a viable reagent the polymer support must be re-examined. For these reasons, work on this reaction was suspended until a more optimal support could be developed. Instead, the reaction with phenylsilane was re-examined.

2.3 Efficient Phenylsilane Mediated Amidation Reactions

2.3.1 Aims and Objectives

Aims



Scheme 43: amidation conditions

The aims of this work were to reexamine the phenylsilane mediated amidation reaction in order to maximise the efficiency of the reaction. In particular the amount of phenylsilane used in the reaction should be minimised to improve the overall mass efficiency as phenylsilane is not present in the final product. Once the reaction was reoptimized the reaction was applied to the synthesis of the API Lacosamide and the natural products Piperine, Capsaicin and Rehmagluamide. In addition, green metrics were calculated and assed to demonstrate how efficient this reaction is.

This work is advantageous due to the fact that it is maximising the amount of amount of product while limiting the amount of silane used.

Objectives

To achieve the aims stated above these objectives were developed.

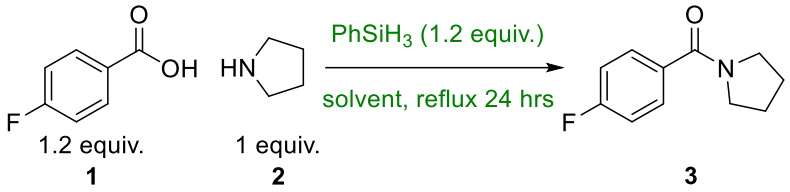
- Re-optimisation of phenylsilane mediated amidation reaction
- Calculation of sustainability metrics
- Application to synthesis of API and natural products

2.3.2 Results and Discussion

2.3.2.1 Re-optimisation of the batch amidation

Re-examining the optimisation of the batch process using phenylsilane with a goal of making the process as mass efficient as possible. Key areas for screening were silane loading, solvent choice, and temperature. The model reaction optimised was the amidation of 4-fluorobenzoic acid (**1)** and pyrrolidine (**2)** owing to the ease of monitoring the reaction by ^{19}F NMR spectroscopy. The optimal conditions from previous work from within the Denton group were used as a starting point to move forward and was used as a control experiment for comparison (**Table 5** Entry 1).[53] While in previous reported work acetonitrile did show good performance those conditions used higher loadings of carboxylic acid and phenylsilane(1.5 equiv.), when using lower loadings of these reagents (**Table 5** Entry 2) performance was reduced significantly. 2-methyl THF was the first of the sustainable alternative solvents to be assessed and showed extremally poor performance (**Table 5** Entry 3) and was not utilised again. Another sustainable alternative solvent cyclopentyl methyl ether (CPME) showed excellent yields (**Table 5**, Entry 4) this led to attempting to further reducing the silane loading to encourage re-use of the silane (**Table 3** Entries 5,6) while these showed moderate yields it was a decrease from the previous result.

Table 5: optimisation of silane mediated amidation



Entry	Solvent	Silane (mmol)	Acid (mmol)	Amine (mmol)	Temperature (°C)	Yield (%)
1	toluene (control)	1.2	1.2	1	110 (reflux)	99%
2	acetonitrile	1.2	1.2	1	82 (reflux)	57%
3	2Me-THF	1.2	1.2	1	80 (reflux)	11%
4	CPME	1.2	1.2	1	106 (reflux)	95%
5	CPME	0.334	1.2	1	106 (reflux)	56%
6	CPME	0.6	1.2	1	106 (reflux)	61%
7	MTBE	1.2	1.2	1	55 (reflux)	96%
8	DMC	1.2	1.2	1	90 (reflux)	99%
9	TPGS-750-M (2 wt. % H_2O)	1.2	1.2	1	r.t.	0%

It should be noted that when using CPME at 50°C the reaction precipitated into a white solid, this solid melted back to a solution at 75°C. Methyl tertbutyl ether (MTBE) showed comparable performance to CPME (**Table 5**, entry 7) and showed the congealing phenomena at 50°C. Dimethyl carbonate (DMC) also a sustainable alternative solvent showed excellent performance identical to the control experiment (**Table 5**, Entry 8) with the added advantage of a lower reflux temperature, DMC at reflux showed the best performance of the solvents tested. As such these conditions were carried forward into natural product and API synthesis. An attempt at performing this reaction in a surfactant TPGS-750-M was attempted (**Table 5**, entry 9) though this did not show any yield and rendered the surfactant unusable. The rational

for using the surfactant was to take advantage of the micelles to overcome phenylsilane's incompatibility with water.

With regards to keeping the reaction temperature high it was found that there was a trade-off between high reaction temperatures and reaction times. To achieve convenient reaction times higher temperatures were required.

2.3.2.2 Calculation of green metrics

To examine fully how efficient and sustainable this process is, a comparison with previous work was required. To this end the work by Ruan and Blanchet was examined alongside the work shown above looking at both reaction mass efficiency and mass intensity.

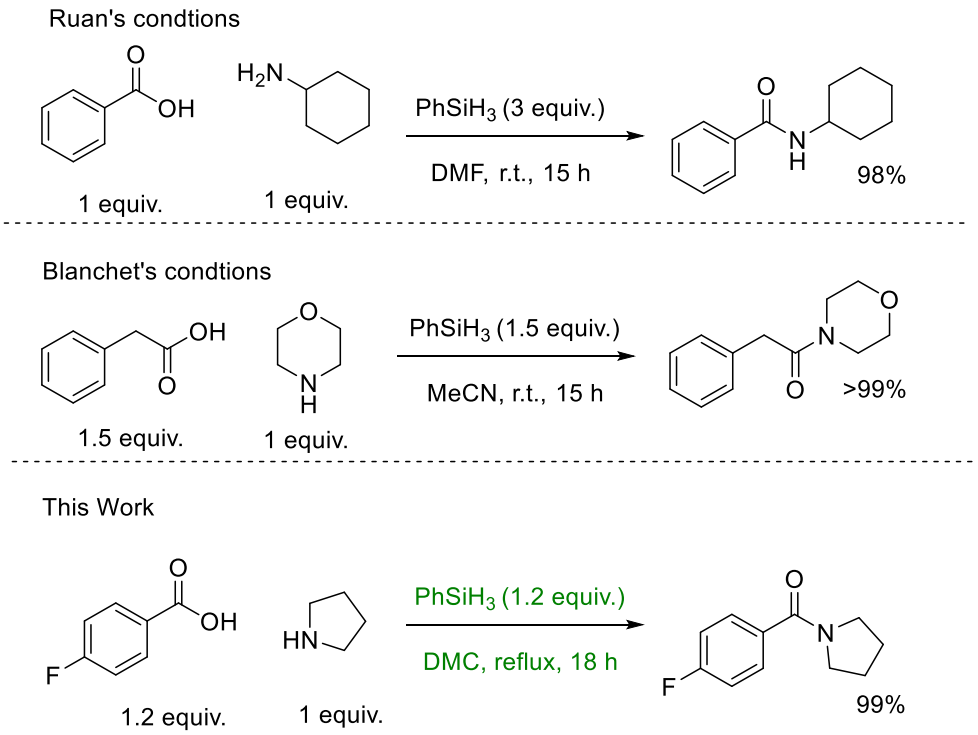


Figure 17: silane amidation conditions from Ruan, Blanchet, and this work

Though it could be considered a superficial aspect to examine solvent selection does have a significant impact on sustainability. The GSK solvent selection guide offers a detailed life cycle assessment of various common solvents examining key metrics on how the solvent is handled. This selection guide is not a perfect measure of sustainability by it does offer a framework on

how best to replace a solvent for a given process. Using this guide, we can examine how sustainable the solvent choices in the three reactions are. For Ruan's work the solvent use is DMF, which is according to the GSK solvent selection guide categorised as a "red" solvent. This means that unless absolutely necessary this solvent should not be used.

For Blanchet's work the solvent used, was acetonitrile, which according to the GSK solvent selection guide is categorised as a "yellow" solvent. This means that, if possible, an alternative should be used. For the work presented in this section dimethyl carbonate was used, which according to the GSK solvent selection guide is categorised as a "green" solvent. This means that it is a preferred solvent for use.

CONDITIONS	SOLVENT	MASS INTENSITY (G/G)	REACTION MASS EFFICIENCY* (%)
Ruan	DMF	24.34	36
Blanchet	MeCN	18.85	39
This Work	DMC	7.53	52

Table 6: green metrics assessment for silane amidations *including amount of silane

Reaction mass efficiency (RME) (**Equation 1**) is a metric that describes the amount of reactant mass that is present in the final product, expressed as a percentage. Excluding the silane from the equation shows verry little difference from the three sets of conditions. This represents an issue with this measure of reaction efficiency, as there are often more reagents/catalysts present in a reaction than just the principal reactants. For the purpose of this assessment the amount of silane has been included as this is the principal difference between the three sets of conditions. Assessing the reaction mass efficiency (**Table 6**) for the three sets of conditions shows only a small between Ruan's work and Blanchet's. Ruan's conditions were shown to be slightly less efficient due to the high silane loading. For the work described in this chapter the reaction mass efficiency is 52%, the reason for this in the lower loading of the carboxylic acid and the silane.

$$\left(\frac{mass\ of\ desired\ product}{mass\ of\ reactants}\right) \times 100 = RME\ (\%)$$

Equation 1: reaction mass efficiency

Mass intensity shown in **Equation 2** is a metric that assess the amount of material used in a reaction to produce 1 g of product. A more detailed variant of this metric exists call process mass intensity which also account for all the mass used in the entire process, work up and purification steps included. How ever for these three methods not all the required information is available so only mass intensity was used. For Ruan's conditions (**Table 6** entry 1) the mass intensity is high due to the amount of solvent and silane (3 equivalents) used in the reaction. The mass intensity for Blanchet's conditions was lower at 18.85 g/g this can be attributed to much lower loading of silane and less solvent used, though it is still a high value. For the work described in this chapter the mass intensity is significantly lower at just 7.53 g/g the decrease in intensity is from the lower solvent, carboxylic acid and silane loadings which represent a significant mass saving.

Equation 2: Mass Intensity

$$\frac{mass\ of\ reaction\ (g)}{mass\ of\ product\ (g)} = mass\ intensity$$

The analysis this that the three reactions selected are limited due to the fact that they are not the same reaction. The limitation this fact imposes is that you cannot make exact comparisons between the three different methods. The reason why experimental data was used over a purely theoretical dataset is that it had experimental yield values, which gave insight into the reactions performance.

2.3.2.3 API and natural product synthesis

Lacosamide

Lacosamide is an antiepileptic used in the treatment of partial onset seizures developed in 1996 by Choi *et al.* and approved for medical use in 2007 by the FDA.[58] Structurally its core scaffold is O-Me-serine functionalised with a benzylamide group and an acetamide group.

Lacosamide was chosen as the target API owing to the two amide groups present in the final molecule, this example is advantageous as it allows for a demonstration of the reaction's performance in a "back to back" synthesis in a small but highly functionalised molecule.

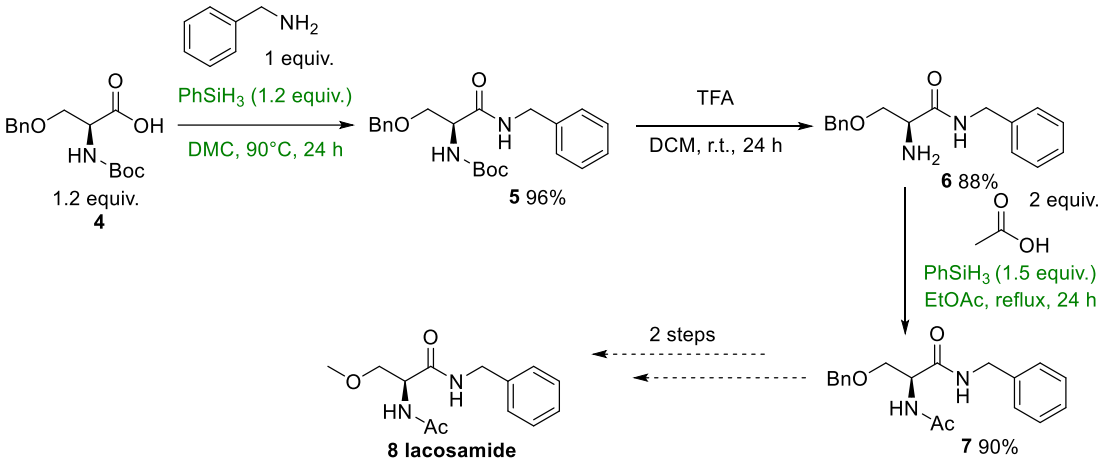


Figure 18: synthesis of lacosamide

The synthesis of Lacosamide (**Figure 18**) began with the amidation of Boc benzyl protected serine using the optimal amidation conditions described previously in **Table 5**. This reaction produced the benzylamide in excellent yields. Moving forward to the deprotection of **5**, this was performed according to prior literature using TFA and gave good yields.[59] The second amidation step required slightly modified conditions in that a greater amount of phenyl silane was required and ethyl acetate was used as the solvent, this produced the diacetamide product in excellent yields. The final two steps in the synthesis of lacosamide were not performed as these two steps are readily found in the literature and with the diamide

precursor now produced with the chemistry developed in this work it was of little scientific value to retrace these steps.[58]

Piperine

Piperine is a natural product derived from both black and white pepper, it was discovered in 1819 by Hans Christian Ørsted who extracted the compound from said plants. In food piperine gives the distinct pungency and flavour pepper is known for. Piperine has long been used in traditional medicine and in recent years attention has been given to its pharmacological properties. In particular its usefulness as an anti-inflammatory, and has been shown to be beneficial for conditions such as inflammatory bowel disease.[60]

Piperine synthetically offers a chance to explore how a bulky doubly unsaturated carboxylic acids perform under silane mediated conditions.

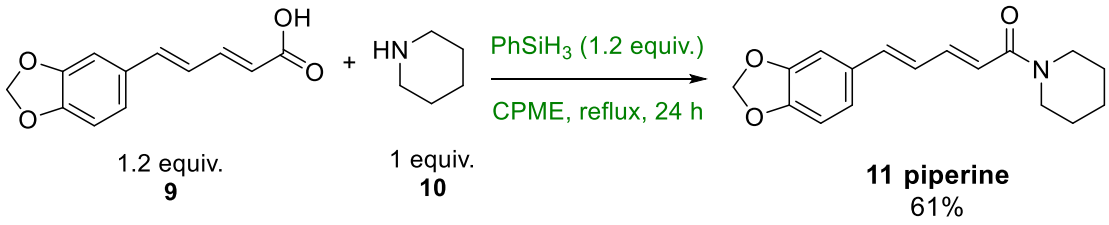


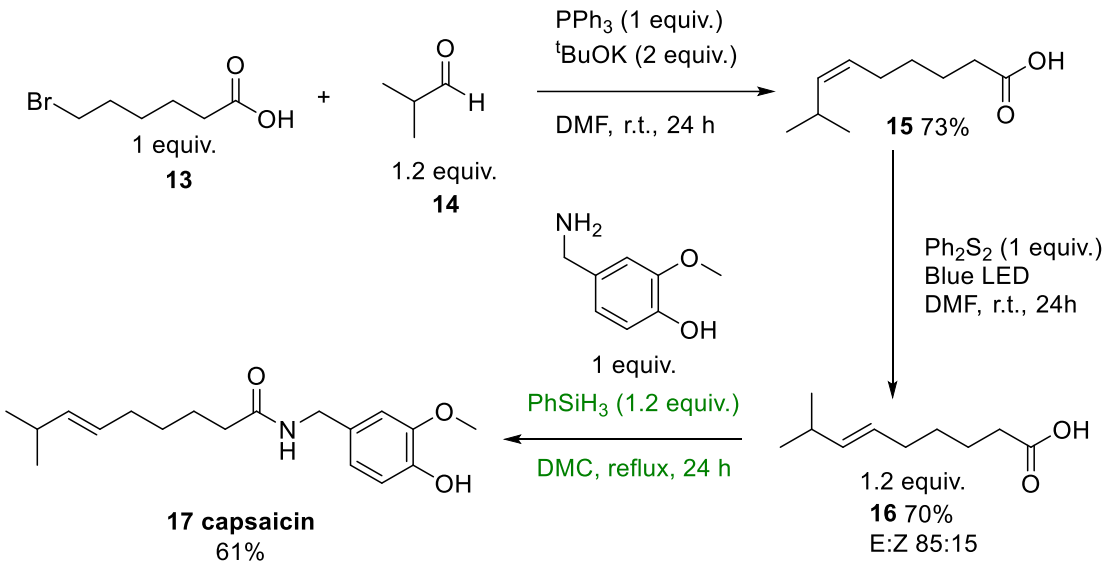
Figure 19: Synthesis of piperine

Due to solubility of the carboxylic acid the synthesis of Piperine **11** (**Figure 19**) was done with CPME in place of DMC this gave the product in fair yield though it is likely that the steric bulk of the silyl ester intermediate contributed to the lower than anticipated yields.

Capsaicin

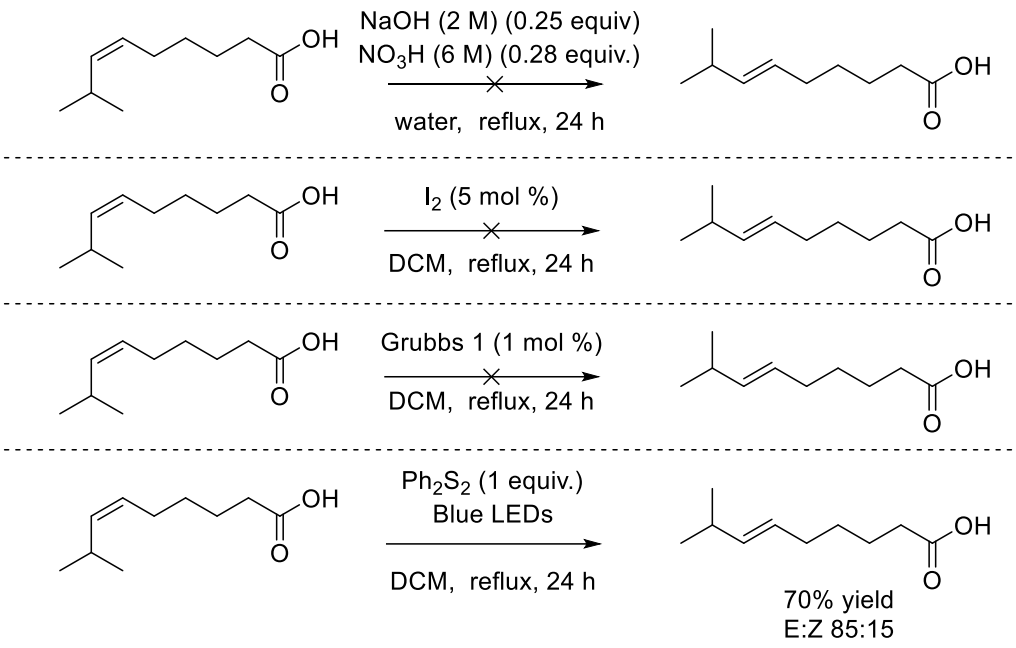
Capsaicin is a natural alkaloid product that comes from capsicum family of plants most notably from chilli peppers. It gives it name to the capsacinoid family of natural products which all share similar structure to capsaicin, a vanillamide bound to a long alkyl chain. Capsaicin is particularly famous as it is the source of the hot sensation chilli peppers give to many foods. It has a long use within traditional medicine, in contemporary medicine it is used most

commonly as an analgesic in topical creams to provide temporary pain relief for joint pain.[61] Synthetically capsaicin gives the opportunity to examine how effective the silane amidation would be as a replacement for the Schotten-Baumann Reaction, which was previously used in the synthesis of capsaicin.[62]



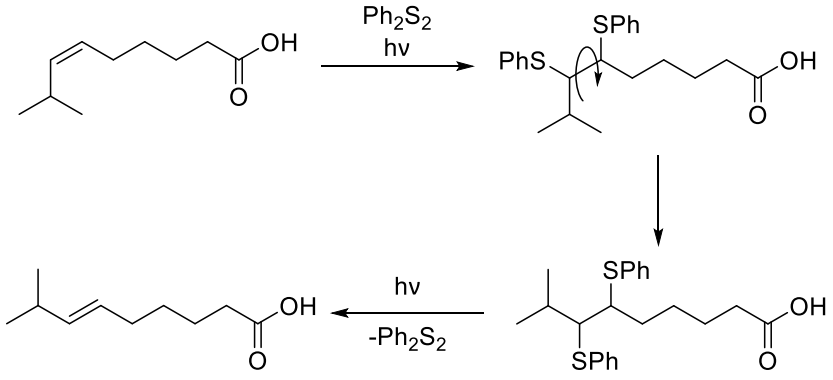
Scheme 44: Synthesis of Capsaicin

In the synthesis of capsaicin (**Scheme 44**), the first step was a Wittig olefination reaction taken from the available literature.[62] This product gave the cis alkene in good yield but necessitated isomerisation. Initial attempts at isomerising the alkene were also taken from the literature.[62] However, this did not give the desired trans product as expected and so other isomerisation methods were tried. Methods used included using elemental iodine, and an attempted Grubb's isomerisation but these also did not yield the desired trans isomer (**Scheme 45**).



Scheme 45: attempted isomerisations

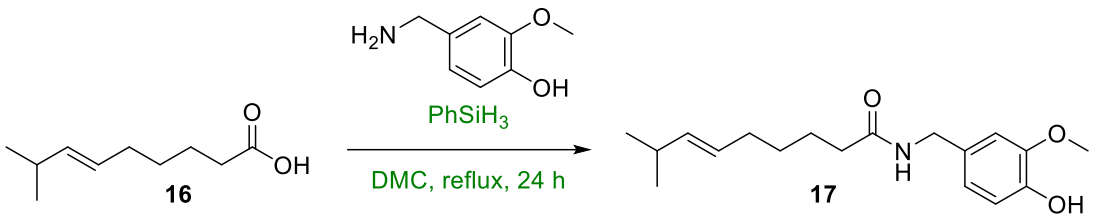
Taking a different approach, a method recently developed by the Silvi group proved to effectively produce the desired trans isomer in good yields.[63] This isomerisation uses diphenyl disulfide in a photochemical isomerisation (**Scheme 45**).



Scheme 46: proposed isomerisation mechanism

This isomerisation (**Scheme 46**) likely proceeds *via* the homolysis of the disulfide into a thiyl radical. This radical will react with the akene to give a dithioether thus breaking the alkene double bond allowing for free rotation between the carbon atoms. The two thiophenyl ether groups provides the necessary steric clash to produce the correct conformation. Under photochemical conditions the dithioether homolyse back into the alkene now in the *trans*

conformation. This reaction gave a good yield of 70%, ^{1}H NMR analysis of the acid showed that it has a E:Z ratio of 85:15 which matched favourably with a commercially available analytical standard. This reaction gave an overall good yield of 70%, with the purified acid used in the next step.



Scheme 47: amidation step

The amidation (**Scheme 47**) with the *trans* carboxylic acid gave a moderate yield of 61%, much like with piperine the likely cause of the lower yield is the high steric bulk of the carboxylic acid making the attack of the vanillamine more difficult.

Rehmagluamide

Rehmagluamide (**Figure 20**) is a natural product that is found in the roots of *Rehmannia glutinosa* it was discovered by Liu *et al.* in 2020.[64] The plant it is derived from and has a long history of use within traditional Chinese medicine. Biological testing done by Liu *et al.* shows that Rehmagluamide shows renoprotective effects. Structurally rehmagluamide is a doubly conjugated keto amide with the amide being a derivative of ethanolamine. To date there has been no total synthesis of rehmagluamide despite is possible use within medicine. To that end any total synthesis would advantageous as it could act as the launch point for further biological studies and medicinal chemistry.

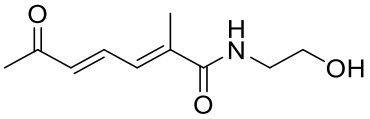


Figure 20: structure of rehmagluamide

Examining rehmagluamide gave two retrosynthetic routes that could potentially give the final product, the first is a largely linear route the second is convergent strategy.

Retrosynthes route A

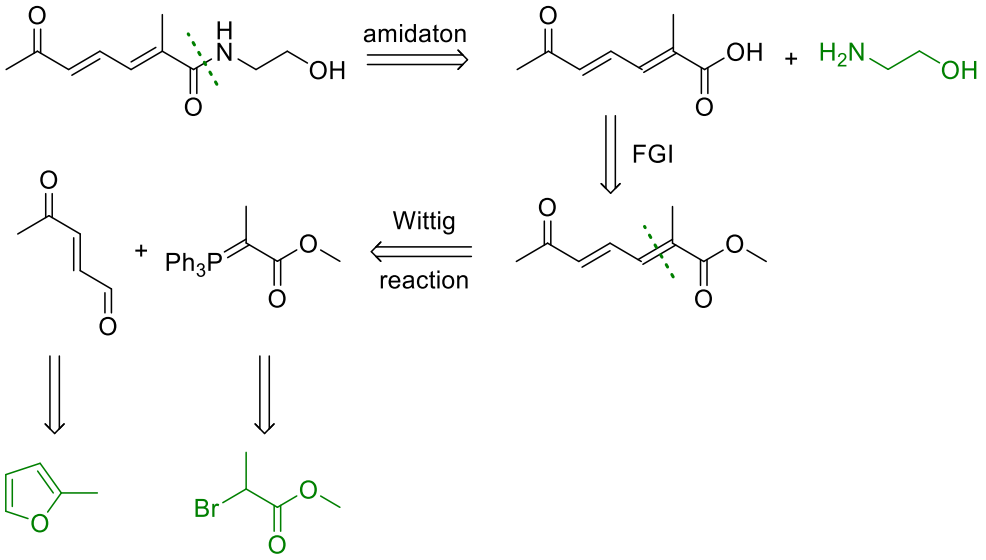


Figure 21: retrosynthetic analysis A (desired starting materials in green)

The linear retrosynthesis of Rehmagluamide (**Figure 21**) starts with the simple disconnection of the amide bond to give a doubly conjugated keto carboxylic acid and ethanolamine. The second disconnection produces the challenge of producing this particular keto carboxylic acid, there is no literature examples for it. There is however examples of the methyl ester of this carboxylic acid as such after functional group interconversion now provides a route for the next disconnection. The double bond closest to the ester is the next choice as it sets up a Wittig reaction though the halide of the methyl ester and a short keto aldehyde. The last step is to form the keto aldehyde. Within the literature are several examples of the synthesis of the compound from 2 methyl furan in an Achmatowicz type reaction.

Retrosynthesis route B

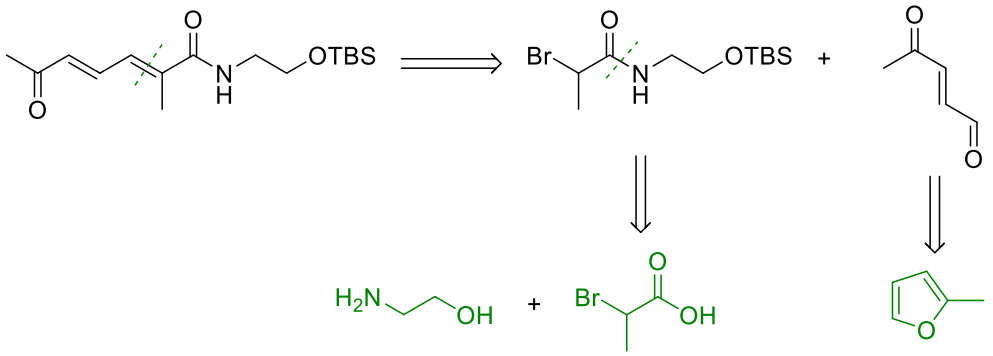
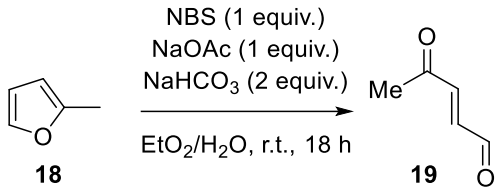


Figure 22: retrosynthetic analysis B

The convergent retrosynthesis of Rehmagluamide (**Figure 22**) share many of the same steps as route A but tries a different order of reactions to achieve the same outcome. Sarting with the first disconnection at the alkene to set the Wittig reaction provides the halo amide and the keto aldehyde, for the halo amide the next disconnection is the amide to afford ethanolamine and an the bromo ester. Finally, the keto aldehyde is produced in the same manner as in route A.

Both retrosynthetic routes A and B were attempted, however only route A showed significant progress and only the outcomes of route A are detailed below.

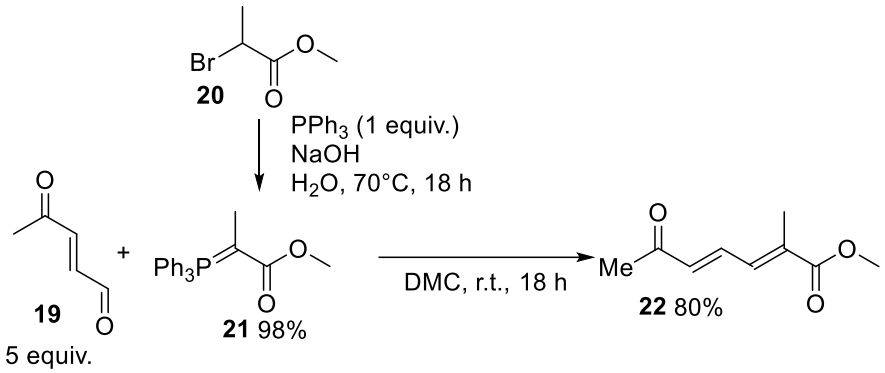
Synthesis



Scheme 48: Achmatowicz type ring opening

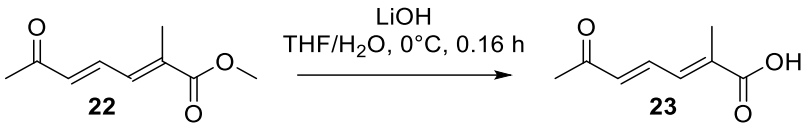
Synthesis route A started with the Achmatowicz type ring opening of 2-methyl furan **18** (**Scheme 48**). This proved to be a challenging reaction due to the high reactivity of the product. various conditions were tried to obtain the product **19**. The examples from existing literature only gave the product in yields of 20% or less which is unsuitable for this work. Instead, a

buffered reaction using NBS, sodium acetate and sodium hydrogen carbonate gave afforded the product. Unfortunately, due to the high volatility and reactivity of *trans*-oxo-pentanal isolation and full characterisation proved difficult, as such only ^{1}H NMR spectroscopy could be acquired. The data that could be acquired is in good agreement with literature.[65] Handling this compound also proved difficult as its high reactivity meant that it could not be stored even at -20 °C and at room temperature would undergo various undesired reactions within one hour. To counter this the reaction was telescoped into the subsequent Wittig olefination (**Scheme 49**).



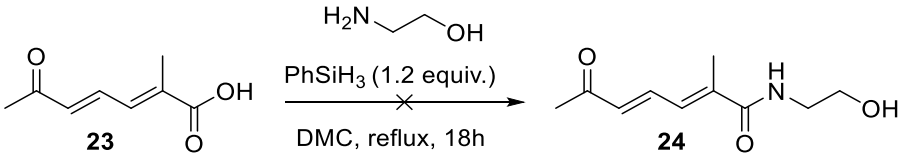
Scheme 49: ylide formation and Wittig olefination

For the Wittig olefination reaction, it was it was advantageous to produce the stabilised phosphonium ylide prior to the Wittig reaction itself. Following a simple literature procedure, the ylide **21** was produced in excellent yield after purification.[66] The Wittig prove more difficult than initially theorised owing the high reactivity of the aldehyde **19** and the low solubility of the ylide. After a brief solvent screen, it was found that DMC was the best solvent for the ylide. To solve the issue of the aldehyde reactivity it was necessary to use an excess of five equivalents of aldehyde to ensure conversion to the desired product **22**. The overall yield for the was good (80%) based on the starting amount of alkyl bromide **20**, this was more than sufficient for the next step.



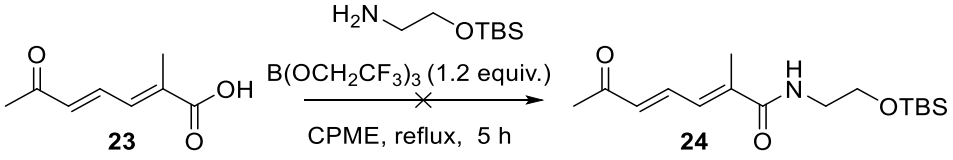
Scheme 50: ester hydrolysis step

The ester hydrolysis step (**Scheme 49**) gave the carboxylic acid **23** in excellent yield. At the time of writing this doubly conjugated keto acid **23** is not reported in any literature, as such this is the first reported synthesis for this compound.



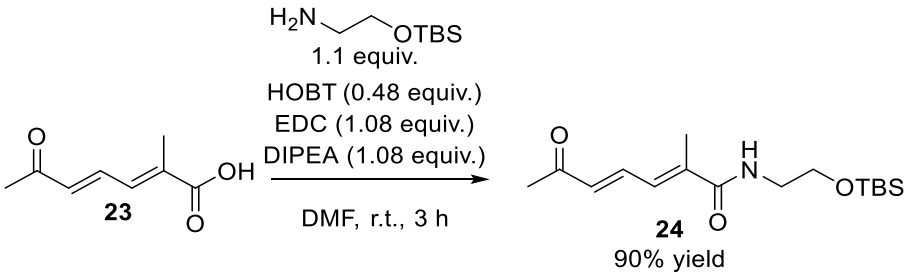
Scheme 51: proposed silane amidation

The final amidation step proved to be very challenging, initial attempts using the phenylsilane method described previously failed (**Scheme 51**). Instead of the final amide product instead it was observed that the carboxylic acid had degraded leaving a complex mixture. ^{13}C NMR spectroscopy clearly showed the loss of the ketone and the adjacent alkene environments. This indicates that this region of the molecule has undergone undesirable side reactions that prevents formation of the final molecule. Modifications to the reaction were attempted such as TBS protecting the ethanolamine, lower reaction temperature, shorter reaction times. But none of these methods yield any desired product. This result proved to be a complication; it was intended that this natural product be produced with a sustainable amidation method. Once we had the synthesis fully proven the earlier steps would have been re optimised for efficiency.



Scheme 52: proposed borate ester mediated amidation

At this stage it was deemed necessary to try a different efficient amidation method as a strategy to produce rehmagluamide. A method developed by the Sheppard group (**Scheme 52**) was selected owing to good functional group tolerance.[67] This method uses a borate ester $B(OCH_2CF_3)_3$ as the amidation agent.[68] This amidation works though the formation of a activated boron ester species that readily undergoes nucleophilic attack.[67] Following the literature procedure the amidation was attempted but again the desired product was not detected.



Scheme 53: HOBt/EDC amidation

Moving to a more conventional amidation method did yield the desired amide product. A HOBt/EDC coupling with DIPEA (**Scheme 53**) of the carboxylic acid and TBS-ethanolamine afforded TBS-rehmagluamide in excellent yield. Despite the addition of the TBS protecting group a rough comparison with the literature data for rehmagluamide can still be made. ^{1}H, and ^{13}C NMR spectroscopy were in good agreement with the natural product despite the protected alcohol. This method was reattempted without the TBS protecting group; however, this did not afford the desired product.

The need for the TBS protecting group necessitates a deprotection step, this proved challenging as attempts with classical deprotection agents such as TBAF degraded the starting material. Other TBS deprotection

During the characterisation of TBS-rehmagluamide the mass spectroscopy analysis showed a serendipitous result. Instead of the expected mass the protected form, what was instead observed was the mass (including adducts) of the deprotected final product.

2.4 Conclusions

In this work it has been demonstrated that a silane immobilised on a polystyrene support can successfully act as a coupling agent in a silane mediated amidation reaction. This process has been shown for both a trimethoxy silane and a trihydridic silane. However, the performance of this supported silane is not at the level as the non-immobilised silanes. As such, significant work remains to improve the polymer support to such a degree that it matches phenyl silane's level of performance.

In addition, the amidation mediated with phenylsilane has been reoptimised with enhanced efficiency as assessed by the calculation of reaction mass efficiency and mass intensity. This process was then applied to the successful synthesis of the benzyl precursor to lacosamide, and the synthesis of natural products piperine and capsaicin.

Also detailed is they synthesis of TBS-rehmagluamide. The last stage was intended to use the phenylsilane amidation process. This was attempted however, this proved unsuccessful. Several different amidation methods were tried for rehmagluamide but HOBt/EDC coupling with TBS protected ethanolamine proved successful, efforts afford the deprotected product are ongoing.

2.5 Future work

With regards to the polymer supported silanes the future work is based primarily around the analysis of the polymer supports and the quantification of the amount of silane bound to the polymer. In addition, the exploration of alternative bespoke polymer supports should be undertaken.

For the phenylsilane amidation in the green metrics analysis. A more accurate analysis could be determined through a series of control reactions using the same starting material to gain the necessary experimental data.

2.6 Experimental

General information

All regents were purchased from commercial suppliers and used without further purification. Dry 2-MeTHF and DMC was purchased from Sigma Aldrich and acetonitrile and toluene were dried over sodium wire and obtained from a solvent tower, where degasses solvent was passed though columns of activated alumina and a 7-micron filter under 4-bar of pressure. All water was deionised before use, all experiments were carried out in oven dried glassware with an argon balloon atmosphere. Flow reaction performed on a vapourtec R-series flow platform with a packed column reactor set.

Analysis and characterisation

Analytical Thin Layer Chromatography (TLC) was performed on Merck aluminium-backed silica-gel plates 60 F254 plates and visualized by ultraviolet (UV) irradiation (254 nm) or by staining with a solution of potassium permanganate or vanillin. Fourier Transform Infrared Spectrometry (IR) was carried out using a Bruker Tensor 27 using an Attenuated Total Reflection (ATR) attachment and peaks are reported in terms of frequency of absorption (cm^{-1}). High Resolution Mass Spectrometry (HRMS) were measured on a Bruker microTOF II with Electron Spray Ionisation (ESI).

^{1}H NMR spectra were recorded on either a Bruker AV 400, AV(III) 400HD or AV(III) 500HD in $CDCl_3$. ^{1}H NMR chemical shifts (δ) were reported in parts per million (ppm) and coupling constants (J) are given in Hertz (Hz), with residual protic solvent as the internal reference ($CDCl_3$ δ = 7.26 ppm). The proton spectra are reported as follows: δ (multiplicity, coupling constant J, number of protons). Abbreviations used include s – singlet, d – doublet, t – triplet, q – quartet, sept – septet, m – multiplet, br – broad, app. – apparent. 13C NMR were recorded on a 400 MHz spectrometer, chemical shifts (δ) were reported in ppm relative to the ^{13}C

signals in the solvent (central peak of $CDCl_3$ δ = 77.16 ppm) and coupling constants (J) are given in Hertz (Hz). All 13C NMR are reported as proton decoupled spectra. ^{19}F NMR were recorded on a 376 MHz spectrometer, chemical shifts (δ) were reported in ppm relative to $CFCl_3$ at 0.00 ppm and are reported as proton decoupled spectra.

2.6.1 Immobilised Silanes for Amidation Reactions

Procedure 1 (Direct amidation 1)

To a 20°C solution of 4-flourobenzoic acid (1.6 mmol, 2.0 equiv.) in acetonitrile (4 mL), was added pyrrolidine (0.8 mmol, 1.0 equiv.) then phenylsilane (1.2 mmol, 1.5 equiv) or polymeric silane (2.4 mmol, 3.0 equiv.). After 24 hours reaction was filtered then diluted with CH_2Cl_2. The product was extracted with HCl (10 mL of a 1 M aqueous solution), NaOH (10 mL of a 1 M aqueous solution) and saturated brine (10 mL). The combined organic layers were dried with $MgSO_4$, and the solvent removed under reduced pressure to afford **3** *N*-(4-fluorobenzoyl)pyrrolidine as a white crystalline solid.

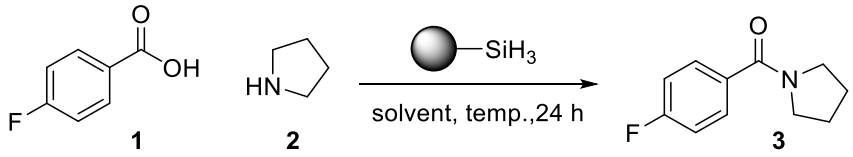
Procedure 2 (direct amidation 2)

To a room temperature solution of 4-flourobenzoic acid (0.4 mmol 1.0 equiv.) in DMC (4 mL), was added pyrrolidine (0.4 mmol 1.0 equiv.) then polymeric silane (1.2 mmol 4.0 equiv)., The reaction was stirred for 12 hours and then filtered then diluted with CH_2Cl_2. The product was extracted with HCL (10 mL of a 1 M aqueous solution), NaOH (10 mL of a 1 M aqueous solution) and saturated brine (10 mL). The combined organic layers were dried with MgSO4 and the solvent removed under reduced pressure to afford N-(4-fluorobenzoyl)pyrrolidine as a white solid.

Flow amidation procedure

The **p4a** was ground into a powder in a mortar and pestle. Then using the packed column kit provided with a vapourtec R series flow kit, the powder was poured into an empty column and compressed until no free volume remained. The system was then primed with acetonitrile.

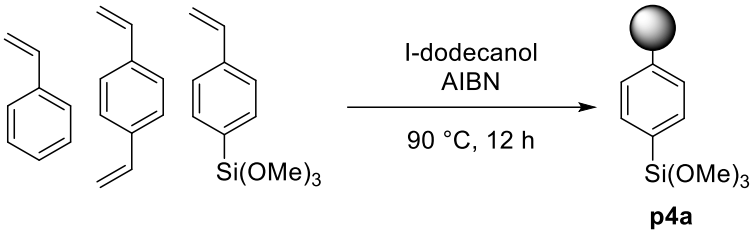
A solution of 4-fluorobenzoic and pyrrolidine in acetonitrile was pumped though a packed column of polymer supported silane. After 24 hours . The product was extracted with HCl (10 mL of a 1 M aqueous solution), NaOH (10 mL of a 1 M aqueous solution) and saturated brine (10 mL). The combined organic layers were dried with MgSO4 and the solvent removed under reduced pressure.



3) *N*-(4-fluorobenzoyl)pyrrolidine[69]

The following reagents were combined in the amounts given below according to the procedure 1. $PhSiH_3$ (130 mg, 1.20 mmol), 4-fluorobenzoic acid (234 mg, 1.60 mmol) pyrrolidine (57.0 mg, 0.80 mmol) purification by flash column chromatography (50% ethyl acetate 50% cyclohexane). **^{1}H NMR**: (400 MHz $CDCl_3$) δ 7.57-7.52 (m, 2H), 7.11-7.06 (m, 2H), 3.64 (t, *J* = 6.9, t Hz 2H), 3.48 (t, *J* = 6.5 Hz, 2H) 1.86-2.01 (m, 4H). **^{13}C NMR**: (400 MHz $CDCl_3$) δ 169.5, 164.8, 131.3, 129.5, 115.8, 45.8, 40.6, 25.8, 22.7. **^{19}F NMR**: (376 MHz $CDCl_3$) δ -110.41 (s 1F). **HRMS** [ESI (M+H)] m/z calculated for C_{11} H_{13} O N F 194.0976, found 194.0976.

Silane polymer **p4a** ((trimethoxyphenysilane)polystyrene)

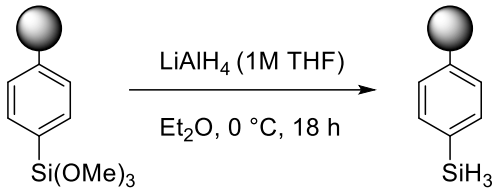


p4a

To trimethoxy(4-vynylphenyl)silane (496 mg, 2.21 mmol) was added styrene (89 µL, 0.77 mmol), divinylbenzene (430 µL, 3.10 mmol) and 1-dodecanol (864.00 mg, 4.64 mmol). The mixture was then heated to 50 °C until homogenous. To the reaction mixture was then added azobisisobutyronitrile (9 mg, 0.06 mmol) and was stirred at 50 °C for 5 min. The reaction was then transferred to a sealed 5 mL syringe barrel and heated to 85 °C for 12 hours. The reaction was cooled to room temperature and a monolith of 1.3 ml volume was ground up to a course powder (1mm approx. particle size) and washed with CH_2Cl_2 (20 mL) to give polymer as a translucent white solid.

IR ν_{max} cm^{-1}: 3024, 2920, 1695, 1602, 1558, 1509, 1492, 1239, 1125, 1046, 917, 819, 788, 699, 551, 479.

Silane polymer **p4b** ((trihydrophenysilane)polystyrene)



p4b

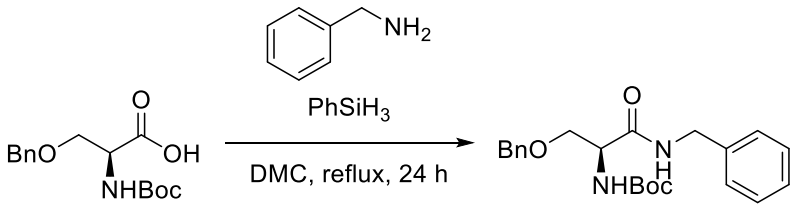
To a 0°C suspension of (trimethoxyphenysilane)polystyrene 1.30 g 2.21 mmol) in Et_2O, $LiAlH_4$ (4.42 ml, 4.42 mmol of a 1.00 M solution in THF) was added. The reaction was left to warm to room temp and run for 18 hours. Solvent was removed *via* canular filtration, and polymer was washed *via* canular with Et_2O (3x10ml). Polymer was filtered and collected as a translucent white solid.

IR vmax cm-1: 3024, 2920, 2150, 1695, 1602, 1558, 1509, 1492, 1402, 1125, 917, 899, 788, 699, 551, 479.

2.6.2 Efficient Phenylsilane Mediated Amidation Reactions

(S)-Lacosamide

5) *tert*-butyl (*S*)-(1-(benzylamino)-3-(benzyloxy)-1-oxopropan-2-yl)carbamate[70]

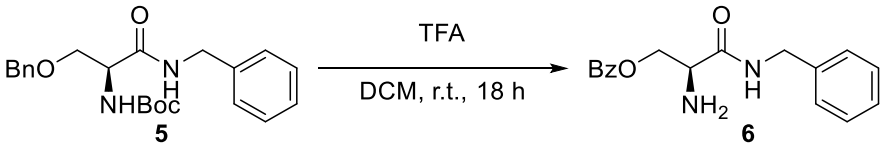


5

To a solution boc-o-benzyl-l-serine (**4**) of (354 mg, 1.20 mmol) in DMC (5 ml) was added benzylamine (107 mg, 1.00 mmol) and phenylsilane (125 mg, 1.2 mmol), reaction was left at reflux for 24 hours. Reaction was diluted with EtOAc (10 mL) then washed with HCl (10 ml of a 1 M solution), NaOH (10 mL of a 1 M solution), and 10 ml brine (10 mL of a saturated solution), organic layer dried with $NaSO_4$, and solvent removed *in vacuo*. To afford the product as a yellow oil.

Isolated Yield 96% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.35-7.25 (m, 10H), 6.87 (br. s 1H), 4.55 (br. s, 1H), 4.60-4.50 (m, 4H), 4.35 (br.s, 1H), 3.99 (dd, J=9.2, 3.8 Hz, 1H), 3.63 (dd, j= 9.2, 6.3 Hz), 1.46 (s, 9H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 170.2, 137.9, 137.4, 128.7, 128.5, 127.9, 127.8, 127.6, 127.5, 73.5, 70.0, 54.9, 43.5. **HRMS (ESI+):** Exact mass calcd for $C_{21}H_{28}N_2O_4Na$ (M+Na) 407.1947, found 407.1940. **$[\alpha]_D^{20}$** +30.0 (c. 1.0, $CHCl_3$).

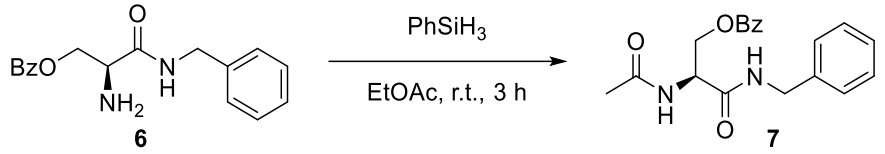
6) (*S*)-2-amino-*N*-benzyl-4-(benzyloxy)butanamide[71]



To a solution of **5** in DCM (408 mg, 1.00 mmol) was added an excess of TFA (5 mL), reaction stirred for 18 hours. The solvent removed *in vacuo* then oily crude product suspended $NaCO_3$ (10 ml of a saturated solution) and washed with DCM (3x 10 ml). combined organic layers were dried with $MgSO_4$ and solvent removed under reduced pressure. Crude product purified by silica plug (90:10 cyclohexane: ethyl acetate 10 mL, then 80:20 DCM: methanol 10 mL), solvent removed *in vacuo* to afford a yellow oil.

Isolated Yield: 88% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.76 (br. s, 1H), 7.36- 7.26 (m, 10H), 4.58-4.51 (m, 2H), 4.47-4.41 (m, 2H) 3.79-3.72 (m, 2H), 3.65-3.63 (m, 1H), 1.76 (s, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 172.5, 138.3, 137.8, 128.6, 128.5, 127.8, 127.7, 127.6, 127.4, 73.4, 72.3, 55.0, 43.1. **HRMS (ESI+):** Exact mass calcd for $C_{17}H_{21}N_2O_2$ (M+H) 285.1597 found 285.1598. **$[\alpha]_D^{20}$** +4 (c. 1.0, $CHCl_3$).

7) (*S*)-2-acetamido-*N*-benzyl-4-(benzyloxy)butanamide[71]



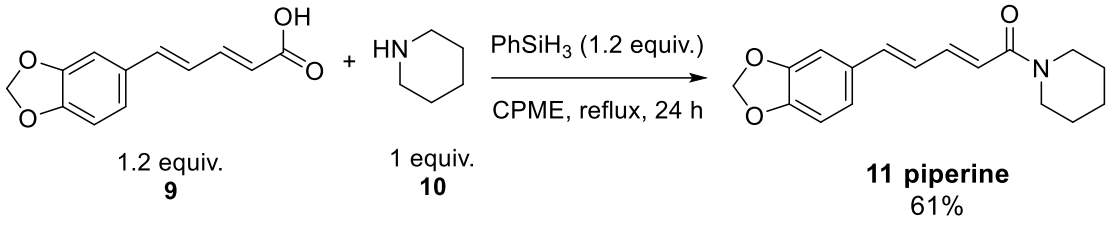
To a solution of **6** (348 mg, 1.00 mmol) in ethyl acetate (5 mL) was added acetic acid (120 mg, 2.00 mmol) and phenylsilane (167 mg, 1.50 mmol). The reaction mixture was stirred at room temperature for 3 h. once a white precipitate had formed. Silica was added (5 g) and the reaction heated at reflux for 10 minutes. The crude reaction mixture was concentrated *in vacuo* and the residue loaded onto a bed of silica. The silica pad was washed with 20:80 EtOAc/cyclohexane (50 mL) 50:50 EtOAc/cyclohexane (50 mL) then 100% EtOAc (50mL). The EtOAc wash was concentrated *in vacuo* and taken up in ethyl acetate, washed with aqueous

sodium carbonate (10ml of a saturated solution) then dried with $NaSO_4$ and concentrated to *in vacuo* to give a white solid.

isolated yield 90% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.33-7.21 (m, 10H) 6.76 (br. s, 1H), 6.45 (d, J=7.2 Hz, 1H), 4.61-4.59 (m, 2H) 4.45 (m, 3H) 3.91 (dd, J=9.2, 4.1 Hz, 1H), 3.52 (dd, J=9.2, 7.7 Hz, 1H) 2.01 (s, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 170.3, 169.9, 137.7, 137.2, 128.7, 128.6, 128.1, 127.9, 127.6, 127.6, 73.6, 69.5, 43.7, 23.25. **HRMS (ESI+):** Exact mass calcd for $C_{20}H_{24}N_2O_3Na$ (M+Na) 349.1523 found 349.1525. **$[\alpha]_D^{20}$** +29.0 (c. 1.0 $CHCl_3$).

Piperine

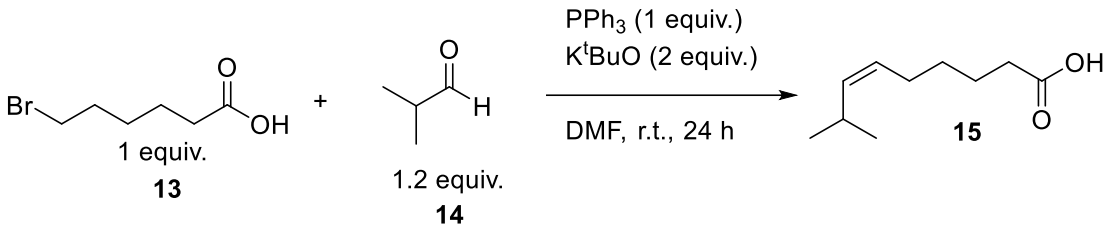
11) (2*E*,4*E*)-5-(benzo[d][1,3]dioxol-5-yl)-1-(piperidin-1-yl)penta-2,4-dien-1-one (piperine)[72]



To a solution (**9**) of (436 mg, 1.20 mmol) in DMC (8 ml) was added piperidine (**10**) (136 mg, 1.00 mmol) and phenylsilane (125 mg, 1.20 mmol), reaction was left at reflux for 24 hours. Reaction was diluted with EtOAc (10 mL) then washed with HCl (10 ml of a 1 M solution), NaOH (10 mL of a 1 M solution), and 10 ml brine (10 mL of a saturated solution), organic layer dried with $NaSO_4$, and solvent removed *in vacuo*. Crude product purified by column chromatography (50% EtOAc 50% cyclohexane) **Isolated yield**: 61% **^{1}H NMR** (400 MHz, $CDCl_3$) 7.47 – 7.32 (1 H, m), 6.97 (1H, s), 6.88 (1 H, d, J 8.0), 6.79 – 6.70 (3 H, m), 6.43 (1 H, d, *J* 14.7), 5.97 (2 H, s), 3.85 – 3.19 (5 H, m), 1.70 – 1.55 (6 H, m). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 165.4, 148.2, 142.5, 138.2, 131.0, 125.4, 122.5, 120.0, 108.4, 105.6, 101.2, 46.9, 43.2, 26.7, 25.6, 24.6. **HRMS (ESI+):** Exact mass calcd. for $C_{17}H_{20}NO_3$ (M+H) 286.1438, found 286.1738. **R.f.** (50% EtOAc 50% cyclohexane), 0.36.

Capsaicin

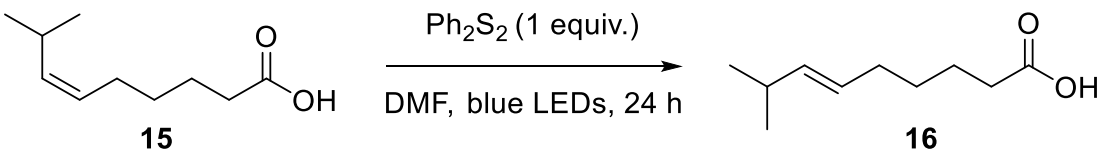
15) (*cis*)-8-methyl-6-nonenoic acid[73]



6-bromohexananoic acid (**13**) (780 mg, 4.00 mmol) and triphenylphosphine (1.05g, 4.00 mmol) were melted together at 145°C for 3 hours. The resulting salt was triturated in $CHCl_3$ and extracted with Et_2O (20 mL) then the suspension was filtered to produce the phosphonium Salt. To a solution of the phosphonium salt in DMF (20 mL) was added isobutyraldehyde (346 mg, 4.80 mmol) and potassium tertbutoxide (898 mg, 8 mmol) was added. Reaction stirred for 24hrs at room temperature. Reaction mixture poured over ice water (30 ml) and filtered then washed with toluene (2x 30ml). Aqueous layer acidified with HCl (15 mL of a 1 M solution), product extracted with Et_2O the dried with $NaSO_4$ and solvent removed *in vacuo*. Crude product purified by short path distillation.

Isolated yield 73%, **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.31-5.20 (2 H, m), 2.64 – 2.51 (1 H, m), 2.36 (2 H, t, J 7.5), 2.11 – 2.00 (1 H, m), 1.65 (2 H, m), 1.46 – 1.36 (2 H, m), 0.94 (5 H, d, J 6.7).**^{13}C NMR** (101 MHz, $CDCl_3$) δ 179.4, 138.1, 126.6, 33.9, 29.3, 26.9, 26.5, 23.2. **HRMS** (ESI+): Exact mass calcd. for $C_{10}H_{18}NaO_2$ (M+Na) 193.1199, found 193.1184.

16 (*trans*)-8-methyl-6-nonenoic acid[62]

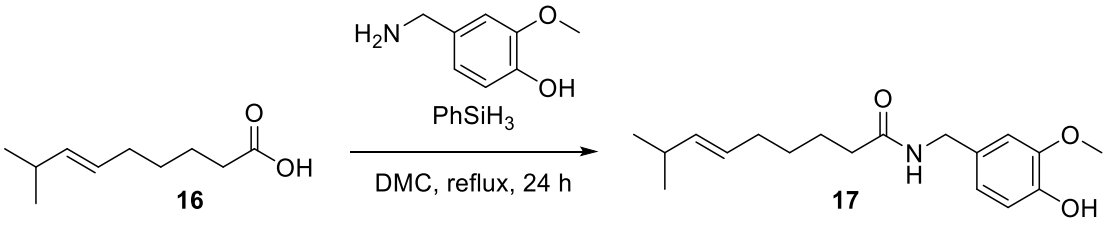


To a solution of (*cis*)-8-methyl-6-nonenoic acid (**15**) (420 mg, 2.4 mmol) in DMF (4 mL) was added diphenyl disulphide (524 mg, 2.4 mmol), reaction was irradiated with blue LEDs at room

temperature for 24 hours. Solvent was removed *in vacuo* and crude product purified by column chromatography (90% cyclohexane 10% ethyl acetate) to afford a light-yellow oil.

isolated yield 70% *E*:*Z* (85/15) **1H NMR** (500 MHz, CDCl3) δ, 5.39 – 5.17 (2 H, m) 2.21-2.13 (m 1H), 2.04 (2H, t, *j* 7.5), 1.86-1.82 (2H, m), 1.48-1.42 (2H, m), 1.21-1.14 (1H, m), 0.96 (6H d, J = 6.7 Hz). **13C NMR** (126 MHz, CDCl3) δ 180.1, 137.8, 126.5, 33.6, 32.1, 31.1, 28.8, 24.0, 22.5. **HRMS** (ESI+): Exact mass calcd. for $C_{10}H_{18}NaO_2$ (M+Na) 193.1199, found 193.1197. R.f. (90% cyclohexane 10% ethyl acetate) 0.53.

17 (*trans*)-*N*-(4-hydroxy-3-methoxybenzyl)-8-methyl-6-nonenamide (capsaicin)[74]

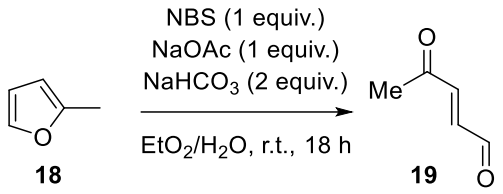


To a solution of (*trans*)-8-methyl-6-nonenoic acid (204 mg, 1.20 mmol), in DMC (4 mL) was added, vanilamine (153 mg, 1.00 mmol) then phenyl silane (125 mg, 1.20 mmol), reaction was left for 16 hours at 90 °C. Reaction was diluted with EtOAc then washed successively with HCl (10 ml of a 1 M solution), NaOH (10ml of a 1 M solution), and brine (10 ml of a saturated solution), organic layer dried with $MgSO_4$, and solvent removed *in vacuo*. Crude product purified by flash column chromatography (40% EtOAc, 60% cyclohexane) to afford a white solid.

isolated yield 61% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 6.85 (1 H, d, J 8.0) , 6.78 (s, 1H), 6.57 (1 H, d, J 8.0), 5.81 (1H, Br. S), 5.39-5.26 (2H, m), 4.34 (1H, m), 3.85 (1H, s), 2.19 (1H, m), 2.00-1.95 (1H, m), 1.85 (2H, t, J 7.5 Hz), 1.68-1.60 (2H, m), 1.39-1.33 (3H, m), 0.97 (5 H, d, J 6.7 Hz). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 172.9, 146.8, 145.2, 138.1, 130.3, 126.5, 120.8, 114.4, 110.7, 55.9, 43.5, 36.7, 32.2, 30.9, 25.3, 22.7. **HRMS** (ESI+): Exact mass calcd. for $C_{18}H_{28}NO_3$ (M+H) 306.2064. Found 306.2064. **R.F.** (40% EtOAc, 60% cyclohexane) 0.44.

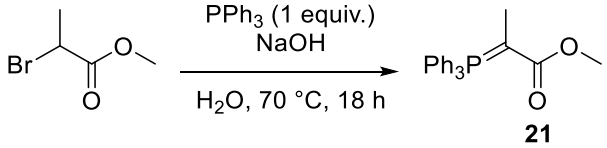
Rehmagluamide

19 trans-4-oxo-2-pentenal[65]



To a solution of 2-methylfuran (**18**) (575 mg, 7.00 mmol) in Et_2O (24 mL) and water (8 mL) (3:1) was added NaOAc·H_2O (952 mg, 7.00 mmol), $NaHCO_3$ (1.18 g, 14.00 mmol) and NBS (1.24 g, 7.00 mmol), reaction left for 2.5 hours at -20 °C. Reaction quenched with saturated $NaHCO_3$ solution then product extracted with Et_2O (2x20 mL), combined organic layers dried with $NaSO_4$ and solvent removed *in vacuo*. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 9.79 (1H, d, J = 7.0 Hz), 6.84 (1H, d, J = 16.3 Hz), 6.75 (1H, dd, J = 16.3, 7.0 Hz, 1H), 2.41 (3H, s). further data could not be collected due to instability.

21 Methyl-2-(triphenylphosphoranylidene)propionate[66]

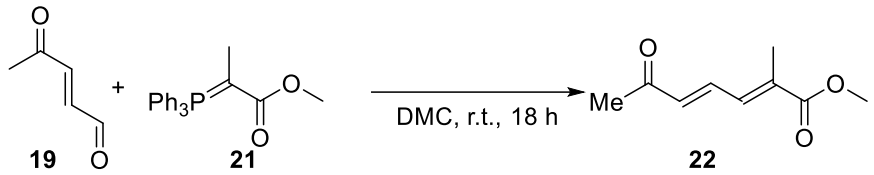


To a solution of triphenylphosphine (835 mg, 5.00 mmol) in water (10 mL) was added 2-bromomethyl propanoate (1.30 g, 5.00 mmol), reaction was left overnight at 70 °C. reaction was cooled and NaOH (20ml of a 1 M solution) was added, and reaction was diluted with DCM (10 mL). Aqueous layer was extracted with DCM (2x10 mL), dried with $NaSO_4$ and solvent removed.

Isolated Yield: 98%, **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.61-7.30 (m, 15H), 3.61 (42%) +3.14 (58%) (s, 3H) 1.66-1.60 (m, 3H). **^{31}P NMR** (162 MHz, $CDCl_3$) δ 22.64, **HRMS (ESI+):** Exact mass calcd

for C_{22} H_{22} O_2 P_1 (M+H) 349.1352. Found 349.1352 (M+H). **IR** (cm^{-1}) 3054, 1598, 1481, 1434, 1306, 1087, 748, 737, 712, 690, 530, 521, 507.

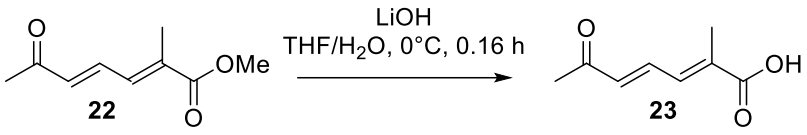
22 Methyl-2-methyl-6-oxo-hepta-2*E*,4*E*-diene-1-oate[75]



To a solution of methyl-2-(triphenylphosphoranylidene)propionate (**21**) (1.74 g, 5.00 mmol) in 2Me-THF was added trans-4-oxo-2-pentenal (**19**) (2.45 g, 25.00 mmol), reaction was stirred at room temperature for 16 hours. Solvent was removed *in vacuo.* Residue was taken up in cyclohexane (20 mL), triphenyl phosphine precipitate filtered and solvent removed *in vacuo*. Crude product purified by column chromatography (60% EtOAc, 40% cyclohexane).

Isolated yield: 80% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.38 (1H, dd, *J* = 15.3, 11.6 Hz), 7.20 (1H, d, *j* 11.6), 6.38 (1H, d, *J* = 15.3 Hz), 3.77 (3H, s), 2.31 (3H, s), 2.06 (3H, br. s). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 198.05, 167.81, 136.54, 135.71, 135.00, 52.26, 28.06, 13.42. **HRMS (ESI+):** Exact mass calcd for C_9 H_{12} O_3 Na (M+Na) 191.0679, found 191.0673. **R.f.** (60% EtOAc, 40% cyclohexane) 0.63

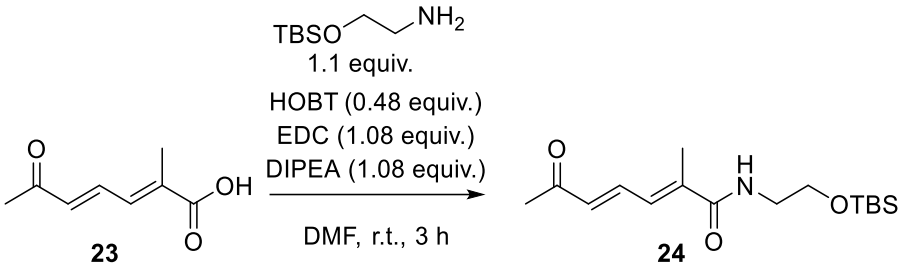
23 (2*E*,4*E*)-2-methyl-6-oxo-hepta-2,4-dienoic acid



To a solution of methyl 2-methyl-6-oxo-hepta-2E,4E-diene-1-oate in a 3:1:1 mix of THF, methanol and water was added LiOH, reaction was stirred at room temp for 4 hours. Product was extracted with ethyl acetate (2x10 mL) and dried with $NaSO_4$ then solvent removed *in vacuo*.

Isolated yield: 94%. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.42 (1H, dd, *J* = 14.9, 11.7 Hz), 7.33 (1H, dd, *J* = 11.7 Hz), 6.45 (1H, d, *J* = 14.9 Hz), 2.35 (1H, s), 2.09 (1H, s). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 198.1, 172.4, 136.2, 135.6, 135.0, 30.9, 28.2, 13.1. **HRMS (ESI+):** Exact mass calcd for $C_8H_{10}O_3Na$ (M+Na) 177.0522 found 177.506. **M.p** 90-95 °C **IR** (cm^{-1}) 2926, 1691, 1640, 1621, 1589, 1196, 986, 976, 787. **RF** (60% EtOAc, 40% cyclohexane) 0.26.

24 (2*E*,4*E*)-*N*-(2-((tert-butyldimethylsilyl)oxy)ethyl)-2-methyl-6-oxohepta-2,4-dienamide (TBS-Rehmagluamide)



To a solution of (2*E*,4*E*)-2-methyl-6-oxo-hepta-2,4-dienoic acid (94 mg, 0.61 mmol), HOBT (39 mg 0.29 mmol), and EDC (126 mg 0.65 mmol) in DMF (2 mL) was added DIPEA (84 mg, 0.65 mmol) and TBS-ethanolamine (118 mg, 0.68 mmol). After 3 hours solvent was removed *in vacuo* and crude product purified by column chromatography (40% EtOAc, 60% cyclohexane)

Isolated yield: 90%. **^{1}H NMR** (500 MHz, MeOD) δ 7.60 (dd, J = 15.4, 11.4 Hz, 1H), 6.96 (d, J = 11.4 Hz, 1H), 6.41 (d, J = 15.4 Hz, 1H), 3.79 (t, J = 6.0 Hz, 2H), 3.44 (t, J = 6.0 Hz, 2H), 2.38 (s, 3H), 2.15 (d, J = 1.6 Hz, 3H), 0.96 (s, 8H), 0.12 (s, 6H).**^{13}C NMR** (126 MHz, MeOD) δ 200.7, 170.9, 141.2, 138.9, 134.9, 131.6, 62.5, 43.3, 27.7, 26.4, 13.8, -5.2. (**HRMS (ESI+):** under mass spectrometry conditions TBS group was deprotected, data observed: Exact mass calcd for $C_{10}H_{15}NO_3$ (M-TBS+Na) 220.0945 found 220.0459. **IR** (cm^{-1}) 3359, 2930, 2857, 2497, 2069, 1647, 1536, 1441, 1360, 1253, 1061, 976, 867, 834, 774, 667. **RF** (40% EtOAc, 60% cyclohexane) 0.62.

2.7 References

1 R. M. Lanigan and T. D. Sheppard, *European J. Org. Chem.*, 2013, **2013**, 7453–7465.
2 E. Massolo, M. Pirola and M. Benaglia, *European J. Org. Chem.*, 2020, **2020**, 4641–4651.
3 A. Bagno, G. Lovato and G. Scorrano, *J. Chem. Soc. Perkin Trans. 2*, 1993, 1091–1098.
4 A. Oliva, B. Henry and M. F. Ruiz-López, *Chem. Phys. Lett.*, 2013, **561**–**562**, 153–158.
5 C. R. Kemnitz and M. J. Loewen, *J. Am. Chem. Soc.*, 2007, **129**, 2521–2528.
6 S. Nahm and S. M. Weinreb, *Tetrahedron Lett.*, 1981, **22**, 3815–3818.
7 P. Imming, B. Klar and D. Dix, *J. Med. Chem.*, 2000, **43**, 4328–4331.
8 C. Schotten, *Berichte der Dtsch. Chem. Gesellschaft*, 1884, **17**, 2544–2547.
9 E. Baumann, *Berichte der Dtsch. Chem. Gesellschaft*, 1886, **19**, 3218–3222.
10 E. Fischer, *Berichte der Dtsch. Chem. Gesellschaft*, 1903, **36**, 2982–2992.
11 K. Nakagawa, H. Onoue and K. Minami, *Chem. Commun.*, 1966, 17–18.
12 A. Tillack, I. Rudloff and M. Beller, *European J. Org. Chem.*, 2001, **2001**, 523–528.
13 S. Gaspa, A. Porcheddu and L. De Luca, *Tetrahedron Lett.*, 2016, **57**, 3433–3440.
14 J. Aube and G. L. Milligan, *J. Am. Chem. Soc.*, 1991, **113**, 8965–8966.
15 J. Aube, G. L. Milligan and C. J. Mossman, *J. Org. Chem.*, 1992, **57**, 1635–1637.
16 P. A. S. Smith, *J. Am. Chem. Soc.*, 1948, **70**, 320–323.
17 V. Kumar, S. Dhawan, R. Bala, P. S. Girase, P. Singh and R. Karpoormath, *Chem. Pap.*, 2023, **77**, 4057–4084.
18 D. Yang, T. Shin, H. Kim and S. Lee, *Org. Biomol. Chem.*, 2020, **18**, 6053–6057.
19 S. E. Eldred, D. A. Stone, S. H. Gellman and S. S. Stahl, *J. Am. Chem. Soc.*, 2003, **125**, 3422–3423.
20 M. A. Ali, A. Nath, M. M. Islam, S. B. Shaheed and I. N. Dibbo, *RSC Adv.*, 2022, **12**, 11255–11261.
21 A. H. Blatt, *Chem. Rev.*, 1933, **12**, 215–260.
22 M. A. Fouad, H. Abdel-Hamid and M. S. Ayoup, *RSC Adv.*, 2020, **10**, 42644–42681.
23 M.-E. Chen, X.-W. Chen, Y.-H. Hu, R. Ye, J.-W. Lv, B. Li and F.-M. Zhang, *Org. Chem. Front.*, 2021, **8**, 4623–4664.
24 J. W. Schulenberg and S. Archer, in *Organic Reactions*, 2011, pp. 1–51.
25 R. M. de Figueiredo, J.-S. Suppo and J.-M. Campagne, *Chem. Rev.*, 2016, **116**, 12029–12122.
26 A. Leggio, E. L. Belsito, G. De Luca, M. L. Di Gioia, V. Leotta, E. Romio, C. Siciliano and A. Liguori, *RSC Adv.*, 2016, **6**, 34468–34475.
27 J.-W. Ren, M.-N. Tong, Y.-F. Zhao and F. Ni, *Org. Lett.*, 2021, **23**, 7497–7502.
28 J. R. Dunetz, J. Magano and G. A. Weisenburger, *Org. Process Res. Dev.*, 2016, **20**, 140–177.
29 M. J. Martinelli, *J. Org. Chem.*, 1990, **55**, 5065–5073.
30 B. Belleau and G. Malek, *J. Am. Chem. Soc.*, 1968, **90**, 1651–1652.
31 Y. Hirokawa, T. Horikawa, H. Noguchi, K. Yamamoto and S. Kato, *Org. Process Res. Dev.*, 2002, **6**, 28–35.
32 A. Berkessel, W. Harnying, N. Duangdee, J.-M. Neudörfl and H. Gröger, *Org. Process Res. Dev.*, 2012, **16**, 123–128.
33 T. Norris, J. VanAlsten, S. Hubbs, M. Ewing, W. Cai, M. L. Jorgensen, J. Bordner and G. O. Jensen, *Org. Process Res. Dev.*, 2008, **12**, 447–455.
34 L.-C. Campeau, S. J. Dolman, D. Gauvreau, E. Corley, J. Liu, E. N. Guidry, S. G. Ouellet, D. Steinhuebel, M. Weisel and P. D. O'Shea, *Org. Process Res. Dev.*, 2011, **15**, 1138–1148.
35 Y. Liu, M. Prashad, L. Ciszewski, K. Vargas, O. Repič and T. J. Blacklock, *Org. Process Res. Dev.*, 2008, **12**, 183–191.
36 H. Wissmann and H.-J. Kleiner, *Angew. Chemie Int. Ed. English*, 1980, **19**, 133–134.
37 D. E. Patterson, J. D. Powers, M. LeBlanc, T. Sharkey, E. Boehler, E. Irdam and M. H. Osterhout, *Org. Process Res. Dev.*, 2009, **13**, 900–906.
38 J. C. Sheehan and G. P. Hess, *J. Am. Chem. Soc.*, 1955, **77**, 1067–1068.
39 W. König and R. Geiger, *Chem. Ber.*, 1970, **103**, 788–798.
40 K. D. Wehrstedt, P. A. Wandrey and D. Heitkamp, *J. Hazard. Mater.*, 2005, **126**, 1–7.
41 L. A. Carpino, H. Imazumi, A. El-Faham, F. J. Ferrer, C. Zhang, Y. Lee, B. M. Foxman, P. Henklein, C. Hanay, C. Mügge, H. Wenschuh, J. Klose, M. Beyermann and M. Bienert, *Angew. Chemie Int. Ed.*, 2002, **41**, 441–445.
42 N. G. Anderson, *Practical Process Research and Development: A Guide for Organic Chemists*,

Elsevier Science, 2012.
43 T. Tozawa, Y. Yamane and T. Mukaiyama, *Chem. Lett.*, 2005, **34**, 1586–1587.
44 T. Tozawa, Y. Yamane and T. Mukaiyama, *Chem. Lett.*, 2005, **34**, 1334–1335.
45 T. Tozawa, Y. Yamane and T. Mukaiyama, *Chem. Lett.*, 2005, **34**, 734–735.
46 Z. Ruan, R. M. Lawrence and C. B. Cooper, *Tetrahedron Lett.*, 2006, **47**, 7649–7651.
47 M. Sayes and A. B. Charette, *Green Chem.*, 2017, **19**, 5060–5064.
48 M. C. D'Amaral, N. Jamkhou and M. J. Adler, *Green Chem.*, 2021, **23**, 288–295.
49 E. Morisset, A. Chardon, J. Rouden and J. Blanchet, *European J. Org. Chem.*, 2020, **2020**, 388–392.
50 D. C. Braddock, P. D. Lickiss, B. C. Rowley, D. Pugh, T. Purnomo, G. Santhakumar and S. J. Fussell, *Org. Lett.*, 2018, **20**, 950–953.
51 D. C. Braddock, J. J. Davies and P. D. Lickiss, *Org. Lett.*, 2022, **24**, 1175–1179.
52 T.-H. Chan and L. T. L. Wong, *J. Org. Chem.*, 1969, **34**, 2766–2767.
53 K. G. Andrews, University of Nottingham, 2017.
54 M. Guidi, P. H. Seeberger and K. Gilmore, *Chem. Soc. Rev.*, 2020, **49**, 8910–8932.
55 M. B. Plutschack, B. Pieber, K. Gilmore and P. H. Seeberger, *Chem. Rev.*, 2017, **117**, 11796–11893.
56 A. Massi, A. Cavazzini, L. Del Zoppo, O. Pandoli, V. Costa, L. Pasti and P. P. Giovannini, *Tetrahedron Lett.*, 2011, **52**, 619–622.
57 Barry Arkles and Gerald Larson, *Silicon Compounds: Silanes and Silicones (2nd edtion)*, Gelest, 2008.
58 D. Choi, J. P. Stables and H. Kohn, *J. Med. Chem.*, 1996, **39**, 1907–1916.
59 N. Stuhr-Hansen, S. Padrah and K. Strømgaard, *Tetrahedron Lett.*, 2014, **55**, 4149–4151.
60 G. Derosa, P. Maffioli and A. Sahebkar, eds. S. C. Gupta, S. Prasad and B. B. Aggarwal, Springer International Publishing, Cham, 2016, pp. 173–184.
61 P. Anand and K. Bley, *Br. J. Anaesth.*, 2011, **107**, 490–502.
62 H. Kaga, M. Miura and K. Orito, *J. Org. Chem.*, 1989, **54**, 3477–3478.
63 D. Filippini and M. Silvi, *Nat. Chem.*, 2022, **14**, 66–70.
64 Y.-L. Liu, Y.-G. Cao, Y.-X. Kan, Y.-J. Ren, M.-N. Wang, X.-L. Fan, X.-K. Zheng and W.-S. Feng, *J. Asian Nat. Prod. Res.*, 2022, **24**, 163–169.
65 M. He, J. R. Struble and J. W. Bode, *J. Am. Chem. Soc.*, 2006, **128**, 8418–8420.
66 M. Liniger, C. M. Neuhaus and K.-H. Altmann, *Mol.* , 2020, 25.
67 P. Starkov and T. D. Sheppard, *Org. Biomol. Chem.*, 2011, **9**, 1320–1323.
68 V. Karaluka, R. M. Lanigan, P. M. Murray, M. Badland and T. D. Sheppard, *Org. Biomol. Chem.*, 2015, **13**, 10888–10894.
69 X.-F. Wang, S.-S. Yu, C. Wang, D. Xue and J. Xiao, *Org. Biomol. Chem.*, 2016, **14**, 7028–7037.
70 S. Fuse, N. Tanabe and T. Takahashi, *Chem. Commun.*, 2011, **47**, 12661–12663.
71 WO2011039781A1, 2011, 68.
72 P. S.-W. Leung, Y. Teng and P. H. Toy, *Org. Lett.*, 2010, **12**, 4996–4999.
73 A. K. Migglautsch, M. Willim, B. Schweda, A. Glieder, R. Breinbauer and M. Winkler, *Tetrahedron*, 2018, **74**, 6199–6204.
74 B. Wang, F. Yang, Y.-F. Shan, W.-W. Qiu and J. Tang, *Tetrahedron*, 2009, **65**, 5409–5412.
75 D. H. Dethe, V. Kumar, N. C. Beeralingappa, K. B. Mishra and A. K. Nirpal, *Org. Lett.*, 2022, **24**, 2203–2207.

Chapter Two

Silane Mediated Reductive Amination Reactions of Difluoroacetic Acid

Difluoroethylamines are an underexplored functional group, with useful properties centred on the weakly basic amine. Herein is described a convenient approach to producing difluoroethylamines starting with difluoroacetic acid. This approach uses phenyl silane to produce the difluoroacetamide which is reduced by phenylsilane and benzene sulfonic acid to give the desired product. This approach was applied to the synthesis of unsymmetrical piperazines.

3.1 Introduction

3.1.1 Fluorinated Amines

General properties

Beta fluorinated alkanes have a noticeable impact upon the pK_a of an adjoining amine. In comparison to the non-fluorinated amine each additional beta fluorine reduces the pK_aH of the amine as shown in **Figure 23**.[1] The difference in pK_aH between the mono, di and tri fluoroethyl amine is significant, the trifluoroethylamine having a pK_aH of just 5.7 means that the amine is significantly less basic. The difluoroethylamine retains some weak basicity and the mono fluoroethyl amine can be considered a basic amine.[2] This pattern likely arises from the electronegativity of fluorine which pulls electron density away from the amine limiting its typical basicity.[3,4]

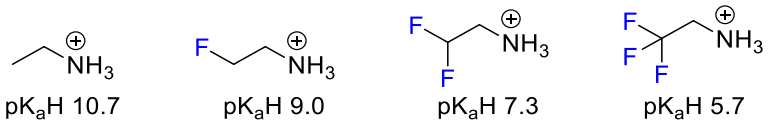


Figure 23: influence of beta fluorines on ethylamine basicity

3.1.1.1 Tri-fluorinated Ethylamines

Within the literature there is a disparity in which fluorinated ethylamines have been examined.[5] After the non-fluorinated ethylamines there has been significant attention given to trifluoroethylamines while mono and difluoro ethylamines have received less attention as shown in **Figure 24**.

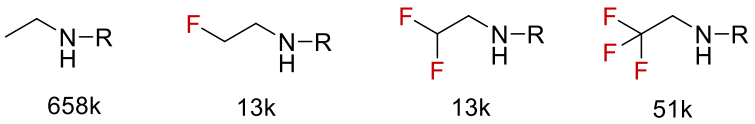
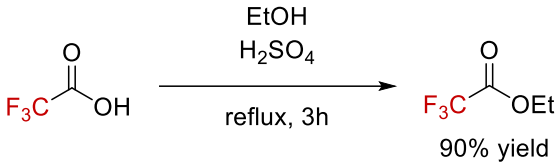


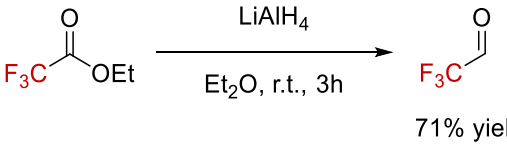
Figure 24: number of fluorinated ethylamine examples

There are many approaches to producing a fluorinated amine the most typical approach is the alkylation of amines. Trifluoroethylamines are often produced from derivatives of inexpensive trifluoracetic acid (TFA). The derived aldehyde, alcohol, and amine are the most common reagents through TFA itself can also be used as the starting material.[6–8] The route to producing these derivatives starts with the synthesis of ethyl 2,2,2-trifluoroacetate as shown in **Scheme 54**.[9]



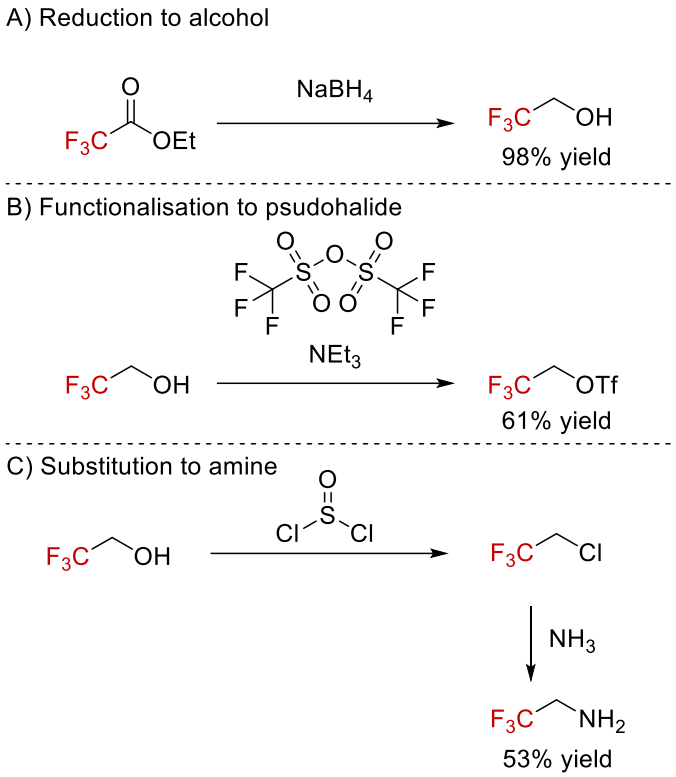
Scheme 54: esterification of TFA

2,2,2-Trifluoroacetaldehyde is produced through the reduction of ethyl 2,2,2-trifluoroacetate with $LiAlH_4$. This reduction method is the most common route to produce this aldehyde.[10,11]



Scheme 55: synthesis of 2,2,2-trifluoroacetalaldehyde

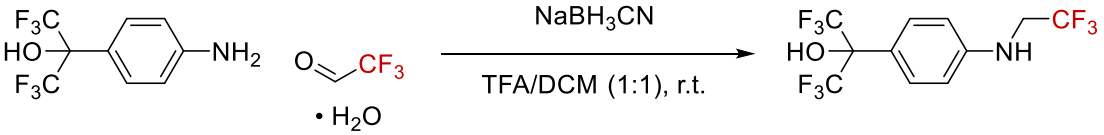
2,2,2-Trifluoroethanol is produced by the reduction of ethyl 2,2,2-trifluoroacetate with $NaBH_4$ (**Scheme 56A**).[12] 2,2,2-trifluoroethanol can be further developed into the pseudohalide such as the triflate or tosylate. This is done through the typical pathways for pseudohalide synthesis. The example in **Scheme 56B** shows the synthesis of 2,2,2-trifluoroethyl triflate which is the most connonly used psudohalide derviative of trifluoroacetic acid.[13] 2,2,2-tifluoroethylamine typically is produced from 2,2,2-trifluoroethanol *via* the substitution of the alcohol to 2,2-difluoroethyl chloride as shown in **Scheme 56C**.[14,15]



Scheme 56: synthesis of 2,2,2-trifluoroethanol, 2,2,2-trifluoroethyl triflate, and 2,2,2-trifluoroethylamine

These derivatives of trifluoroacetic acid (**Scheme 56**) are applied to the synthesis of trifluoroethylamines in the approaches shown below.

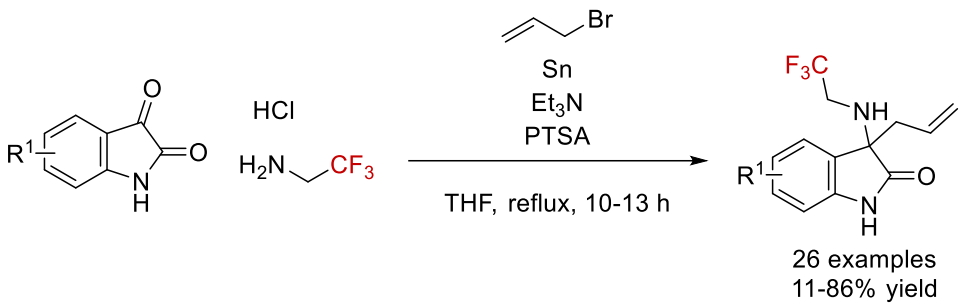
The first approach is the reductive amination of the 2,2,2-trifluroacetylaldeyde to an amine to give the alkylated product. An example of this can be seen in work by Fauber *et al.* (**Scheme 57**).[16] Where they describe the reductive amination of trifluoroacetaldehyde hydrate with 2-(4-aminophenyl)-hexafluoropropan-2-ol.



Scheme 57: synthesis of a trifluoroethylamine by reductive amination

This process is advantageous due to the use of the trifluoroacetaldehyde hydrate in place of the non-hydrated aldehyde. The non hydrated aldehyde is typically a gas which requires specialist equipment and handling while the hydrate is a liquid.

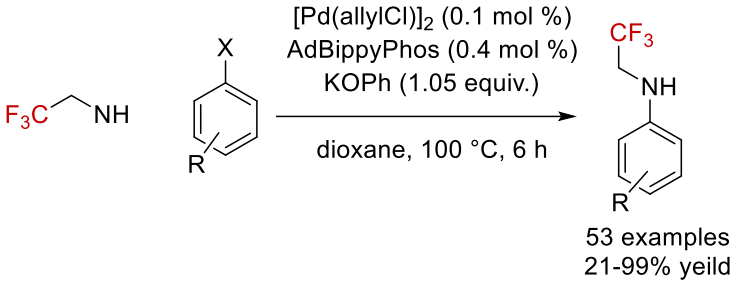
Another approach can be seen by Wang *et al.* in which they reported the functionalisation of trifluoroethylamines bound to a disubsituted oxaindole (**Scheme 58**).[17] In this process trifluoroethylamine in the presence of tin powder, triethylamine, and para-toluene sulfonic acid (PTSA) is reacted with isatins and allylbromide to give the desired amine product. This reaction gives the disubsituted oxaindole products in fair to excellent yields. Mechanistically this reaction proceeds *via* the formation of an imine which is attacked by the allyl tin(II) bromide to give the final product. The advantage of this reaction is the large amount of complexity developed during the reaction offing a convenient way to substitute the 3-position of the oxaindole. The disadvantage of this process is the stoichiometric amount of tin powder required; this complicates the scalability of the reaction as large amounts of tin would be required.



Scheme 58: synthesis of 3,3-disubsituted oxaindoles

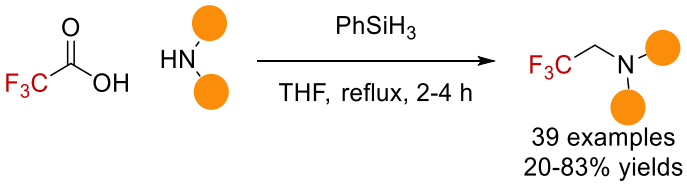
A catalytic approach to trifluoroethylamines was shown by Hartwig et al., in this approach they used Buchwald-Hartwig coupling to produce many fluorinated amines including trifluoronated examples (**Scheme 59**).[18] The reaction has a large scope with good to excellent yield. This reaction follows the typical mechanism for a Buchwald-Hartwig coupling reaction. This method is advantageous as it has a wide scope with excellent functional group tolerance including nitriles, Boc protected amines, methyl esters and ethers. This method is readily compatible with the formation of both mono- and di- fluoroethyl amines. The principal disadvantage of this method is that the $[Pd(allylCl)]_2$ catalyst, in addition to the issues

palladium has as described in the general introduction the catalyst is expensive (£35,113 per mol).



Scheme 59: trifluoroethylation of an aryl halide by Hartwig *et al.*

One more contemporary approach to trifluoroethylamines can be seen by Denton *et al.* in 2017 where they demonstrated a convenient way to produce a trifluoroethylamine directly from trifluoroacetic acid (**Scheme 60**).[6] This silane mediated method proceeds via the formation of a silyl acetyl that can be attacked by the amine to form an iminium ion that is readily reduced to the free amine. This reaction has a good scope and is tolerant of functional groups sensitive to reduction such as nitriles, methyl esters and nitro groups.



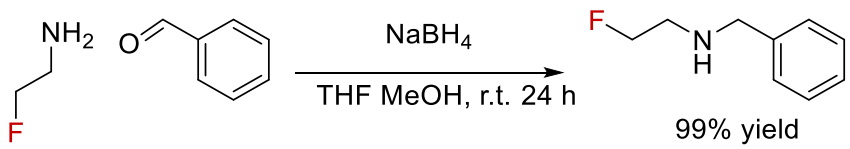
Scheme 60: synthesis of trifluoroethylamines from TFA

The future of TFA in industry has recently been called into question by efforts to curtail the use of perfluorinated alkyl substances (PFAS). The European Union have launched efforts to restrict many of the precursors to PFAS including TFA.[19,20] While the proposal do include exemptions for research, any large scale process chemistry may need to find alternatives to TFA.

3.1.1.2 Mono-fluorinated Ethylamines

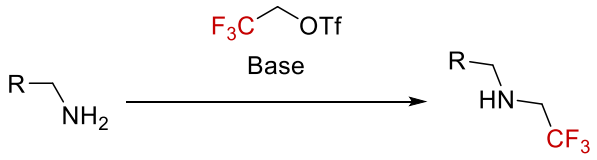
Monofluoro ethylamines have seen very little attention, the principal reason behind this is that the routes to synthesise these amines often use 2-fluoroacetic acid or its derivatives. 2-fluoroacetic acid is known to be highly toxic as it can interfere with metabolic process in particular the enzyme aconitase.[21] As such, it more difficult to handle but can still be desirable in synthetic chemistry. Its derivatives are produced in a similar way to the trifluoroinated examples described previously.

The classical process for the synthesis of monofluoroethyl amine is a reductive amination. In an identical way as shown above in the section on trifluoronated ethylamines. An example of this can be seen in work by Dasgupta *et al.* where they produced a series of fluorinated ethylamines including monofluoroethylamine as "stoppers" for [2]rotaxane (**Scheme 61**).[22]



Scheme 61: reductive amination by Dasgupta *et al.*

An alternative is through the formation of 2,2,2-trifluoroethyl trifluoromethanesulfonate from the parent alcohol. This triflate is then used in an *N*-alkylation reaction to form the fluorinated amine product (**Scheme 62**).

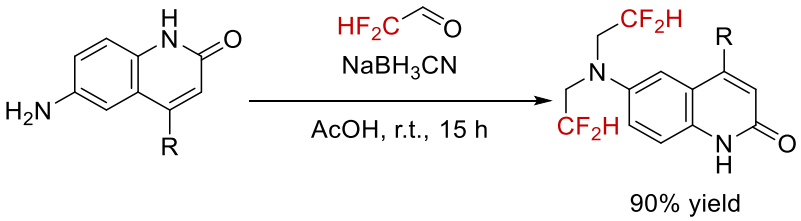


Scheme 62: *N*-trifluoroalkylation using triflates

3.1.1.3 Di-fluorinated Ethylamines

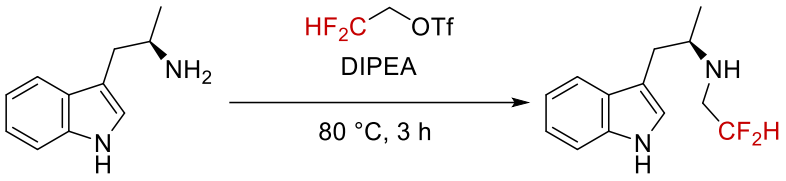
Much like monofluoro ethylamines, difluoro ethylamines have not been as well studied as their trifluorinated counterpart.[5] Classical approaches to the synthesis of difluoroethylamines typically involve the use of derivatives of difluoroacetic acid. Much like with the mono and trifluoronated examples shown previously difluoronated examples are produced in a similar way.

Methods include using 2,2-difluoroacetaldehyde in a reductive amination in a similar fashion to examples seen for mono and trifluorinated ethylamines. An example of this can be shown by van Oeveren *et al.* in 2007 (**Scheme 63**).[7] in the authors work to produce a series of selective androgen receptor modulators they produced a bis-difluoroethylated 6-amino-2-quinolinone as example using 2,2-difluoroacetaldehyde in a reductive amination reaction. Reductive aminations are a highly reliable reaction that will be explored in more depth in section 3.1.2



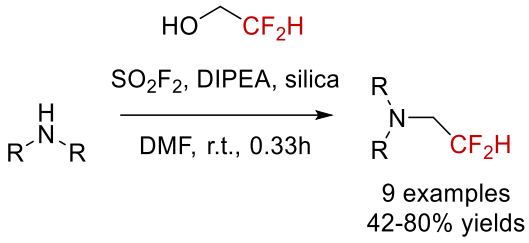
Scheme 63: difluoroethylamine synthesis by reductive amination

An alternative to the aldehyde is 2,2-difluoroethyl trifluoromethanesulfonate which can be used in an *N*-alkylation reaction. An example of this can be seen by Liang *et al.* in 2021 where they employed an *N*-alkylation in the synthesis of selective estrogen receptor antagonists (**Scheme 64**).[23] The goal of the reaction was to add 2 difluoroethyl group to an aromatic amine. Much like reductive amination, N-alkylation is also a well-studied and convenient reaction.



Scheme 64: N-alkylation using 2,2-difuoroethyl trifluoromethanesufonate

A different approach can be seen by Sammis *et al.* where they use 2,2-difluoroethan-1-ol in the presence of SO_2F_2, DIPEA, and silica (**Scheme 65**).[24] This one-pot reaction proceeds *via* formation of 2,2-difluoroethyl fluorosulfuonate as the active source of the fluoroethyl group that goes on to perform an *N*-alkylation reaction. This reaction is advantageous as it offers an expedient way to produce the pseudohalide *in situ* reducing the total number of synthetic steps. The disadvantage of this reaction is the SO_2F_2. This regent is an extremely potent greenhouse gas (approximately 4000 time more potent than CO_2) and must be handled with exceptional care as to prevent it leaking into the atmosphere.[25,26] This severely limits the usefulness of this reaction as specialist equipment is needed to contain SO_2F_2, in this case the use of a 2-chamber reactor.



Scheme 65: N-alkylation by Sammis *et al.*

The fact the authors chose to use such a reagent is a demonstration of the lengths taken by researchers to access this type of functionality. There is significant space within which to explore a difluoroethylation while reducing the need for hazardous reagent.

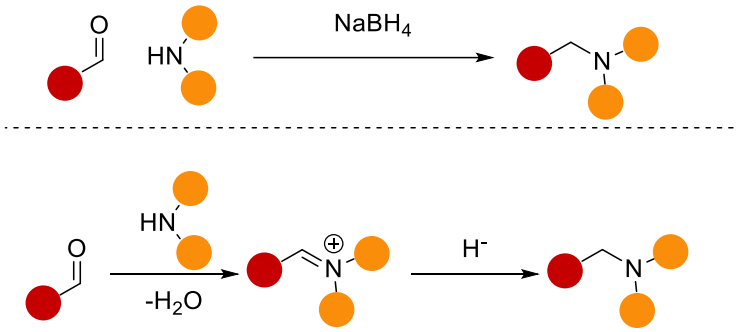
As many of these difluoroethyl reagents shown above are themselves derived from difluoroacetic acid. It would be more efficient and expedient to use difluoroacetic acid and cut out the steps needed to make the derivatives.

3.1.2 Reductive Aminations and Amide Reductions

3.1.2.1 Reductive Aminations

Classical Methods

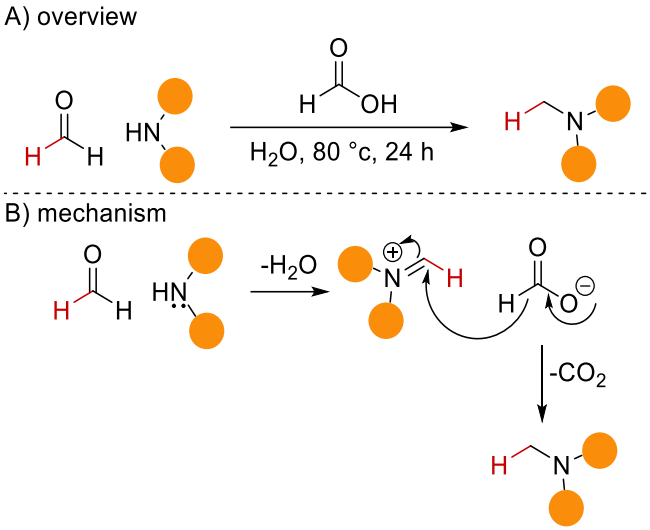
The classical method to perform a reductive amination involves the reaction of an aldehyde in the presence of a hydride source such as sodium borohydride ($NaBH_4$) or sodium cyanoborohydride ($NaBH_3CN$) (**Scheme 66**).[27] This reaction is exceptionally well studied and favoured in organic synthesis, and is the principal way to alkylate an amine.



Scheme 66: classical reductive amidation reaction

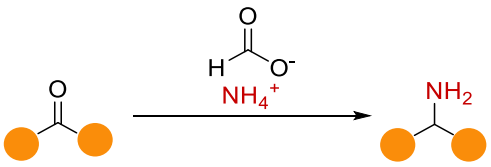
Mechanistically this reaction proceeds *via* a condensation reaction with the amine to give an imine or iminium depending on the substitution at nitrogen. This imine/iminium is readily reduced by the borohydride to give the desired amine product.

Closely related to the standard reductive amination reaction is the Eschweiler–Clarke reaction (**Scheme 67**).[28,29] This reaction traditionally makes use of formaldehyde and an excess of formic acid while more modern versions exchange this for $NaBH_3CN$ to methylate an amine.[30] The mechanism for this reaction proceeds much like the standard reductive amination, however, in this case the imine undergoes reduction from the deprotonated formate to give the reduced amine and CO_2.



Scheme 67: Eschweiler-Clarke reaction

Another variant on the standard reductive amination is the Leuckart reaction (**Scheme 68**). This variant is used to synthesise primary amines from ketones or aldehydes using ammonium formate and heat to reduce to the desired amine.[31]

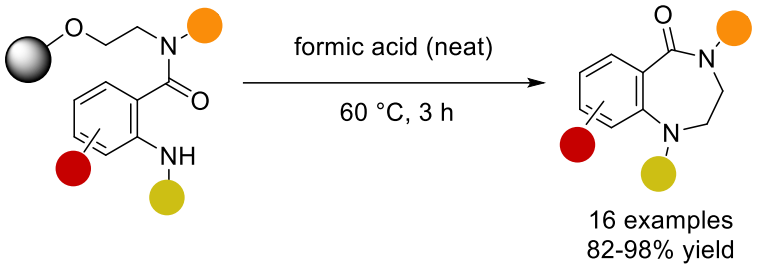


Scheme 68: Leuckart reaction

Mechanistically this reaction works by the formation of the imine that then undergoes reduction in the same pathway as the Eschweiler-Clarke reaction to give the primary amine product.

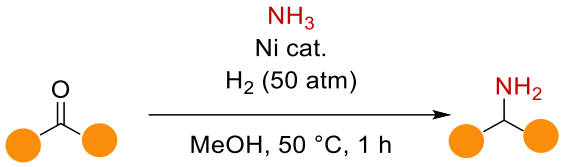
An example of the Leuckart reaction was demonstrated by Lee et al. where they synthesised a series of cyclic amines (**Scheme 69**).[32] The key of this work is the intramolecular Leuckart reaction forming the main 7-membred ring. A unique aspect of this example is that in place of the conventional aldehyde there is a resin bound ether, this ether is readily cleaved to form the key imine intermediate. The advantage of this reaction is that the immobilised starting material favours the intramolecular reaction over the unfavourable intermolecular reaction.

A second advantage is the simplicity of the work up and purification, a simple filtration followed by freeze drying allows for efficient recovery of the desired products.



Scheme 69: Leuckart reduction by Lee *et al.*

Other variations of reductive aminations include Mignonac Reaction (**Scheme 70**). First developed in 1924 this reaction is the synthesis of a primary amine from and aldehyde or ketone with ammonia in the presence of a nickel catalyst.[33]



Scheme 70: general conditions for the Mignonac reaction

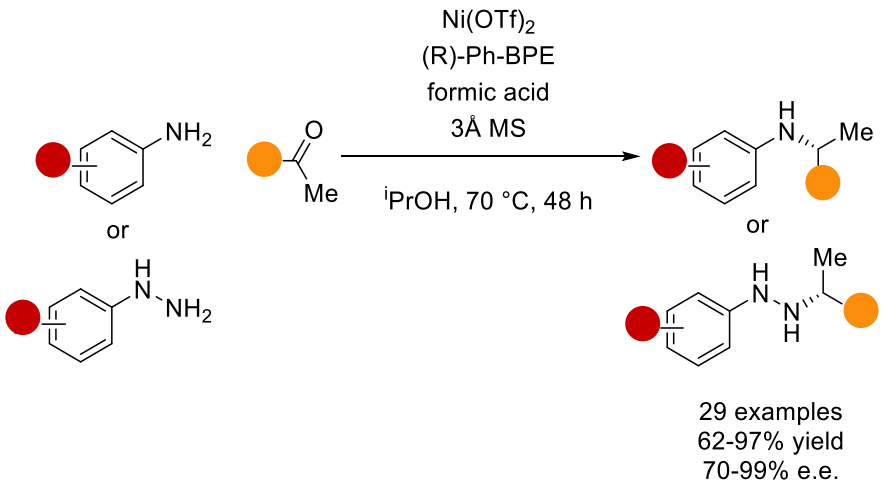
This reaction has received several modifications to improve its effectiveness and overall efficiency. Use of Raney nickel offer much higher efficiency over more traditional elemental nickel catalysts at the cost of higher pressures of ammonia reqired.[34,35] Platinum catalysts with ammonium chloride can also be used as an alternative to nickel in gentler conditions.[36]

Contemporary catalytic methods

Modern catalytic methods for reductive amination are increasingly the focus of many research programs, as they offer a way to reduce the need for sensitive/hazardous reagents. Many metals such as nickel, aluminium, or cobalt can be used as the catalysts.[37–39]

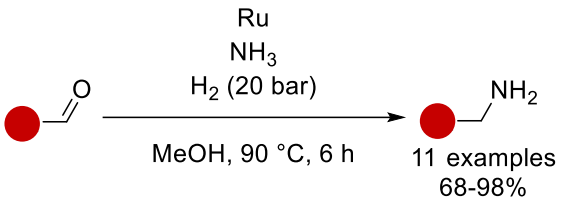
A nickel catalysed example can be seen by Yang *et al.* (**Scheme 71**) where they reported the enantioselective reductive amination of aryl amines and benzhydrazides using ketones .[40] This

method uses $Ni(OTf)_2$ and a chiral biphosphine ligand in the presence of formic acid to attain the desired product. The advantage of the reaction is the applicability of the reaction not only to amines but also to the benzhydrazides, this gives a much wider applicability for this process. The key disadvantage of this reaction is that it must be performed in a glovebox, this is a significant restriction on how convenient the reaction can be and represents a high barrier to accessibility.



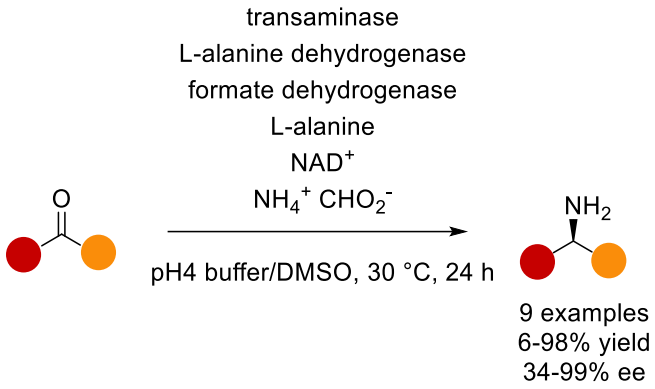
Scheme 71: enantioselective reductive amination of aryl amines and benzhydrazides using ketones by Yang *et al.*

Ruthenium can also be used as a catalyst in reductive aminations, an example was reported by Chandra *et al.*(**Scheme 72**).[41] In this paper the authors describe using ruthenium nanoparticles as a catalyst for the reductive amination of cyclic aldehydes with ammonia to give primary amines in excellent yields. The scope of this reaction is tolerant of heterocycles and halides. The main advantage of this process is the simplicity of the catalyst, the Ru nanoparticles are easy to produce and handle with a high reaction turnover. The main disadvantage of this process is the use of high pressures of hydrogen, this requires specialist equipment and safety protocols and limits the ease of access.



Scheme 72: ruthenium catalysed reductive amination by Chandra *et al.*

A biocatalytic approach was developed by Koszelewski *et al.* (**Scheme 73**) where they show a transaminase catalysed reductive amination.[42] This process works through a cascade reaction first starting with the tranamination using the transaminase from L-alanine to give the desired product. The second part is the regeneration of L-alanine through a reductive amination using L-alanine dehydrogenase and NADH. The last part of the cascade is the formation of NADH from NAD^+ through the transfer of hydrogen from the formate to L-alanine using the formate dehydrogenase.

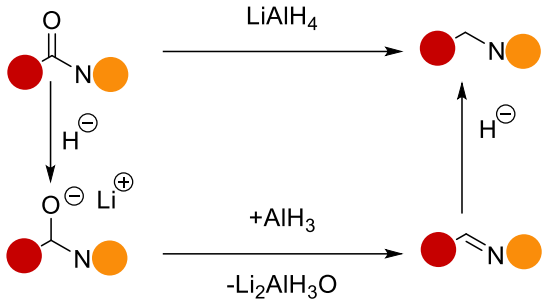


Scheme 73: biocatalysed reductive amination by Koszelewski *et al.*

This process is advantageous as it is a metal free approach to reductive animation using relatively simple source of amine in gentle conditions. The disadvantage of this process is the relatively narrow scope while the authors demonstrated esters and ether examples it is unclear if other functionality such as nitriles, halides or alcohols may be possible.

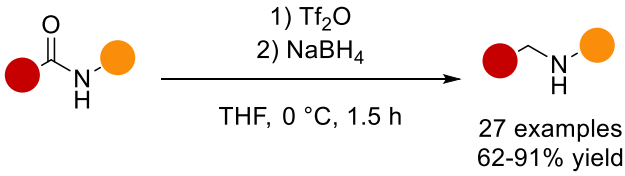
3.1.2.2 Amide Reductions

Amide reduction reactions typically involve removing the carbonyl oxygen from an amide to produce the corresponding amine. This approach has a number of advantages, the first and most importantly is the many routes to producing an amide, amidation reaction were explored in detail in chapter one. The classical approach to reducing an amide is with $LiAlH_4$, this works though the formation of an imine/iminium which is in turn reduced to form the amine (**Scheme 74**).[43]



Scheme 74: amide reduction with $LiAlH_4$

A more recent example of an amide reduction can be seen by Huang *et al.* (**Scheme 75**) where they described a triflate and sodium borohydride reduction of secondary amides.[44] This reaction has good functional group tolerance and a wide scope. The authors also applied the reduction to lactams with good yields. This reaction is advantageous as it is compatible with many functional groups that are sensitive to reduction such as esters and nitro groups. The second advantage is that this is a one-pot procedure which offer a simple way to a complex chemical transformation.



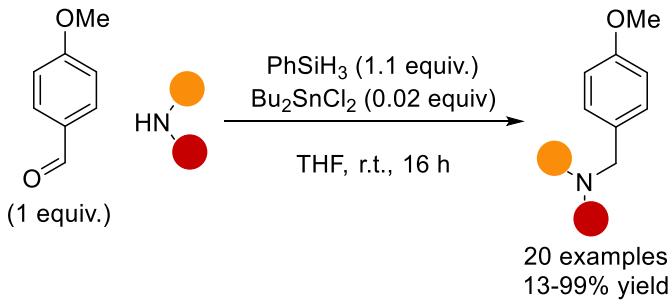
Scheme 75: amide reduction by Huang *et al.*

3.1.3 Silane Mediated Amide Reductions

General Methods

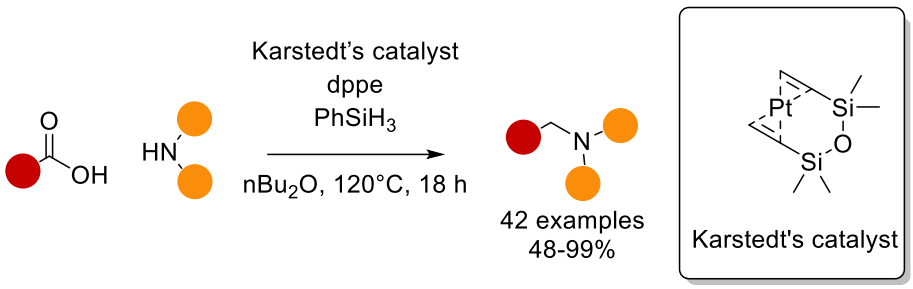
As described in the general introduction silanes make for effective reductants and their effectiveness can be modified through the use of catalysts and additives. A more recent approaches to amide reduction involves using a different reducing agents, typically modern amide reduction reactions use catalytic systems as they removed the need for sensitive reducing agents.[45,46] silanes are often used as the hydride source in these reduction reactions due to their stability, low toxicity and ease of handling.[45]

A tin catalysed example was demonstrated by Apodaca *et al.* (**Scheme 76**). In this work the authors showed how phenylsilane and dibutyltin dichloride as the catalyst can be used in a reductive amination.[47] Mechanically this reaction proceeds *via* imine formation in a similar pathway to the conventional reductive amination. This imine is then reduced by a tin hydride species with phenylsilane as the hydride source for the tin. The key advantage of this reaction is the broad scope as functional groups such as esters, nitro groups, and halides are well supported by the reported reaction conditions. The main disadvantage of this reaction is the dibutyltin dichloride which is a very toxic reagent and requires careful handling.



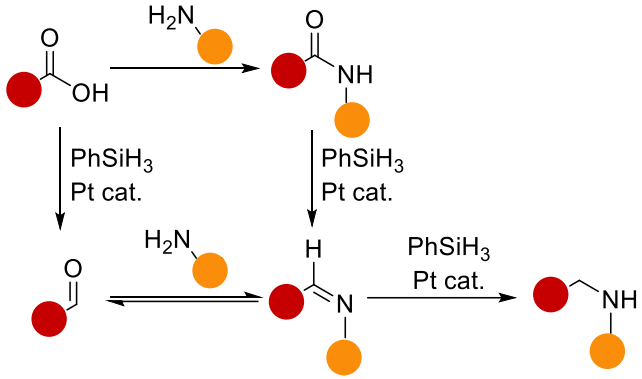
Scheme 76: reductive amination with phenyl silane by Apodaca *et al.*

An example by Beller *et al.* in 2014 (**Scheme 77**) showed that a carboxylic acid with Karstedt's catalyst, dppe, and phenylsilane can effectively perform a reductive amination in fair to excellent yields.[48]



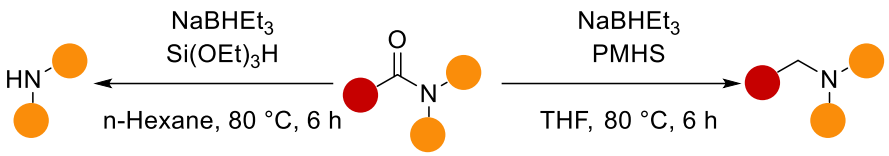
Scheme 77: reductive amination by Beller *et al.*

For this process the authors proposed two methods (**Scheme 78**), the first is a typical reductive amination pathway through the *in-situ* reduction of the carboxylic acid to an aldehyde, that goes on to form the imine that is finally reduced to give the desired product. The second pathway involves the formation of an amide which undergoes reduction to the imine and on to the desired product.



Scheme 78: proposed reaction pathways

Another method was developed by Yao *et al*. (**Scheme 79**) where they report the use of a silane and sodium triethyl borohydride catalyst in an amide reduction.[49] This process is bimodal as depending on the silane used can result in either reduction of the amide to the amine or cleavage of the C-N bond to form a different amine. The key to this reaction is the modes of activation from the silanes, if triethoxysilane is used then the amide reduction pathway is activated. If PMHS is instead used, then the C-N cleavage pathway is activated. The author shows a broad scope with excellent yields. The advantage of this reaction is its bimodality, this gives broad applicability for the production of many amines.

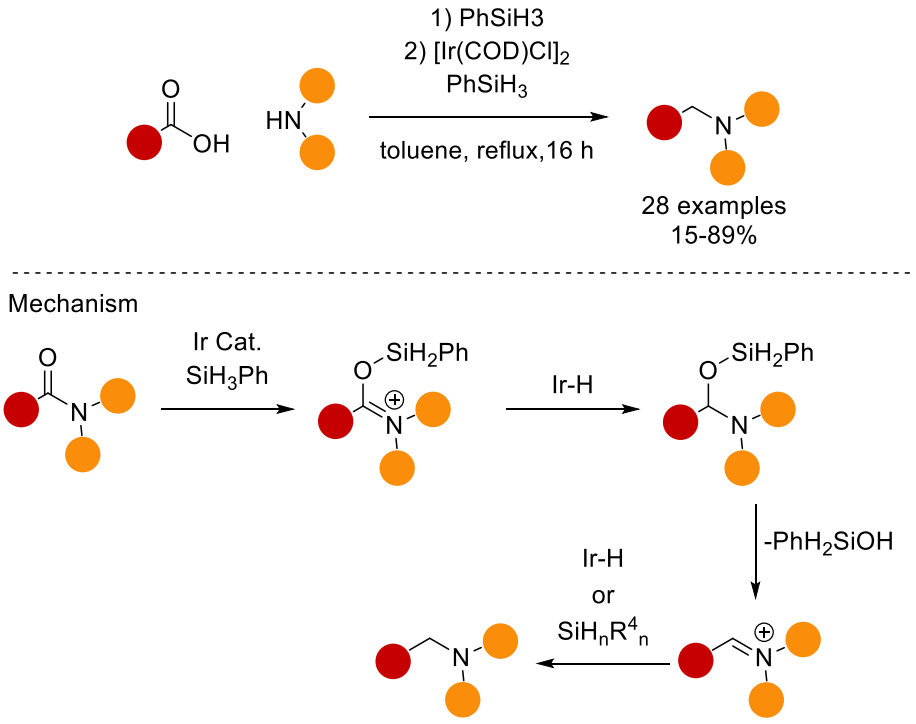


Scheme 79: amide reduction and amide cleavage by Yao et al.

3.1.4 Prior Work by the Denton Group

Work by the Denton group has generated three successive generations of one-pot amidation and amide reduction reactions. Each successive generation has improved upon the reaction performance and ease of use.

The "1st generation silane amide reduction" by Keith Andrews made use of an iridium catalyst with phenylsilane or diethylsilane in a one-pot amidation and amide reduction process (**Scheme 80**).[50]

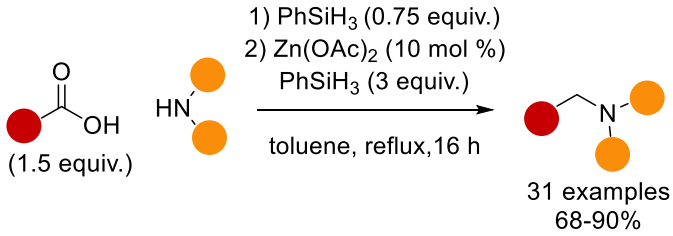


Scheme 80: iridium catalysed one-pot amide reduction of carboxylic acids.

Mechanistically this reaction proceeds first *via* the amidation described in chapter one. The reduce amide is reduced by the activation of the amide *via* first a *O*-silylation then a reduction

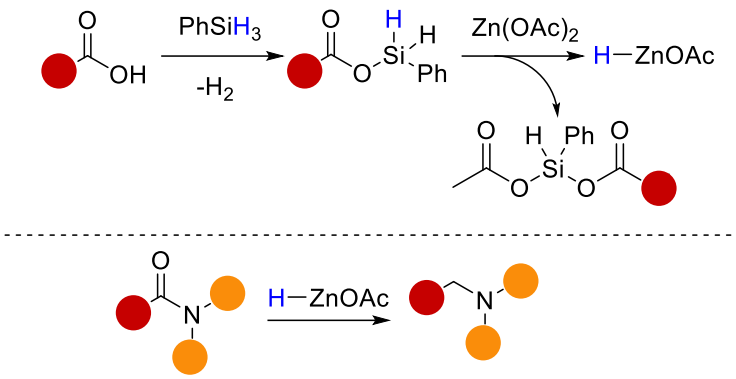
of the iminium to form the key *O*-silyl hemiaminal. The silanol group is then eliminated to form the iminium ion that undergoes to reduction to give the reduced amine product. This process is advantageous as it offers a convenient way to produce functionalised amines from inexpensive carboxylic acids. The principal disadvantage is the use of a comparably expensive $[Ir(COD)Cl]_2$ catalyst (£100,000 per mol), a catalyst from a more common metal would be more sustainable.

The "2nd generation silane amide reduction" was developed by Emma Stoll, this process exchanges the expensive iridium catalyst for a more inexpensive zinc acetate catalyst (£538 per mol) (**Scheme 81**).[51]



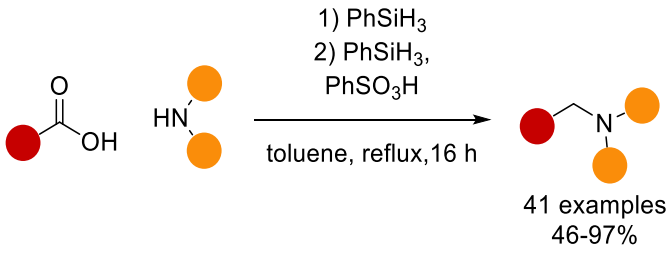
Scheme 81: zinc catalysed one-pot amide reduction of carboxylic acids.

Mechanistically this process works though the formation of a formation of a silyl ester as described in chapter one, these silyl esters act as the hydride source for the generation of zinc hydride as shown in **Scheme 82**. The zinc hydride species acts as the principal reductant to reduce the amide to the amine. This process is advantageous as the zinc catalyst is significantly less expensive that the iridium catalyst used in the 1st generation process. The main disadvantage of this process is that zinc acetate is highly moisture sensitive and therefore this process requires careful setup to eliminate water.



Scheme 82: formation of a zinc hydride species for the amide reduction reaction

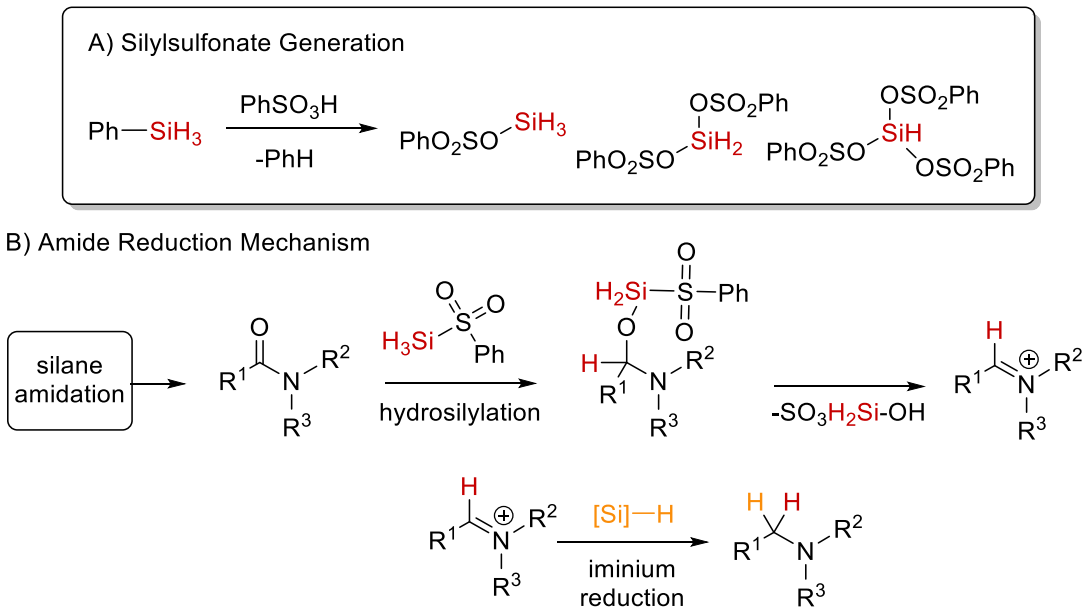
Work by Amelia Stoneley within Denton group developed the "3rd generation silane amide reduction" this metal free approach exchanged the metal catalyst for benzene sulfonic acid.[52] This method is advantageous as it removes the need for a metal catalyst as such less precautions are needed when setting up the reaction, and while benzene sulfonic acid is hygroscopic, the hydrate form is still able to partake in the reaction with minimal loss of yield. This process has a very broad scope of secondary and tertiary amines and is tolerant of functional groups sensitive to reduction such as nitriles, methyl esters and nitro groups.



Scheme 83: metal free reductive amidation using benzene sulfonic acid as a silane activator.

This reaction is advantageous as it is a truly metal free reaction with easy setup and a simple workup/purification. The main disadvantage is that anilines will not work and will not give the desired product. This is due to the silane amidation step as anilines are not nucleophilic enough to attack the silyl ester thereby preventing the amidation from proceeding as expected.

This method proceeds through the *in-situ* generation of a silylsulfonate species as the principal reductant. There are four possible silylsulfonate species possible in this reaction but only three are relevant to the reduction. The mono, bis and tris, silylsulfonate as shown in **Scheme 84A** are the active reductants, the tetrasilyl sulfonate (not shown) is incapable of performing the reduction as it has no available hydride. The first step of the reduction (**Scheme 84B**) is the hydrosilylation of the amide carbonyl. The rate of hydrosilylation depends on the strength of the reductant. the mono silylsufonate is the fastest to form but the slowest to undergo hydrosilylation. Conversely the tris silylsufonate is the slowest to form but the fastest to undergo hydrosilylation. The bis represents the middle ground. After the hydrosilylation step is the elimination of the oxysilylsufonate to form an iminium species. In the final step the iminium is readily reduced with another equivalent of the silylsufonate reductant to give the desired product.

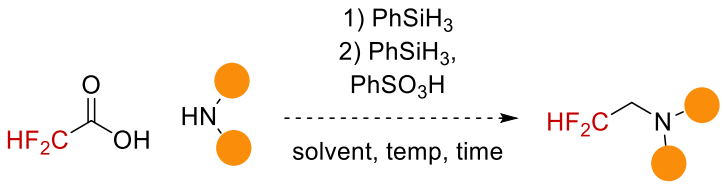


Scheme 84: one-pot metal free amide reduction mechanism

As a one pot process this mechanism is a continuation on the amidation described in chapter one. Though it should be noted that it is possible to do this reaction as a standalone step.

3.1.5 Aims and Objectives

Aims



Scheme 85: proposed difluoroethylation reaction

The principal aim of this work was to develop a convenient route to synthesise terminal difluoroethylamines utilising inexpensive difluoroacetic acid. This work was broken up into first optimising the reaction then examining the reaction scope with respect to tertiary and secondary difluoroethylamines. The secondary amine was to apply the process to the synthesis of unsymmetrical *N*-difluoro ethyl piperazines. The novelty of this process arises from the fact it is a one pot reaction from the carboxylic acid, this removes several synthetic steps and makes a more approachable reaction.

Objectives

The objectives for this work were:

1) Develop the scope of tertiary difluoroethylamines
2) Optimise and develop the scope of secondary difluoroethylamines
3) Apply method to the synthesis of unsymmetrical piperazines

3.2 Results and Discussion

3.2.1 Difluoroacetamides

The first step of this work was a short scope of difluoroacetamides using the amidation method described in chapter one. The intent behind this is to examine how difluoroacetic acid performs under these conditions and to establish the mass balance for the amidation step in

the one pot process described in section 3.1.4. In total seven difluoroacetamides were produced in fair to high yields as shown in **Figure 25**. Of note are compounds **27, 29** and **30** which have desirable functional groups. This was a good result as it established that the amidation step is reliable and tolerant different functional groups such as halides, esters and alcohols. The inclusion of alcohol is important as they are not typically compatible with silanes at high temperature as they could form the silyl ether. In this case the phenylsilane derived silyl ether would likely be cleaved off by the excess of difluoroacetic acid present in the reaction.

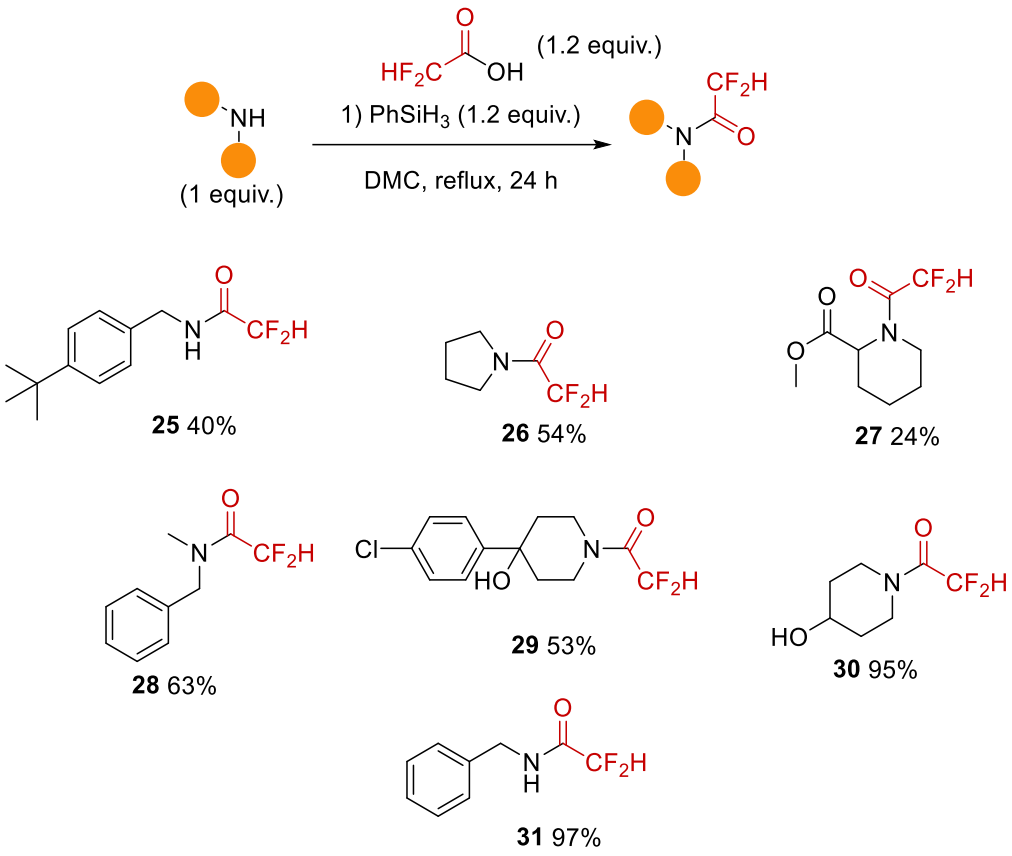


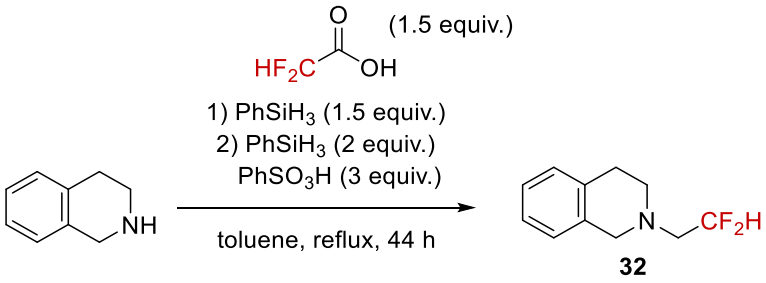
Figure 25: difluoroacetamide scope

3.2.2 Tertiary Difluoroethylamines

Optimisation

The brief optimisation for this process was centred on whether there was a need for inert conditions or could the reaction be performed open to the air. As such two reactions were performed one performed under Schlenk conditions the other completely open to the air.

Table 7: optimisation of difluoroethylation reaction



32

Entry	Conditions	Yield[a] (%)
1	inert	95
2	open	92

[a]isolated yield

The results of these tests show only a very minor difference in yield across the two reactions. The implication of this is that there is little need for inert conditions for this reaction. As such all subsequent work was done using the conditions shown in **Table 7** entry 2. This is advantageous as it represents a much lower barrier to perform this reaction as Schlenk apparatus are not required. This makes the reaction a much simpler process as it lowers the equipment barrier to perform this reaction, this represents an attractive advantage for this process.

Scope

The scope for the tertiary difluroroethylamines was done alongside supporting work by coworkers and produced 22 compounds with yields ranging from 26% to 95%.

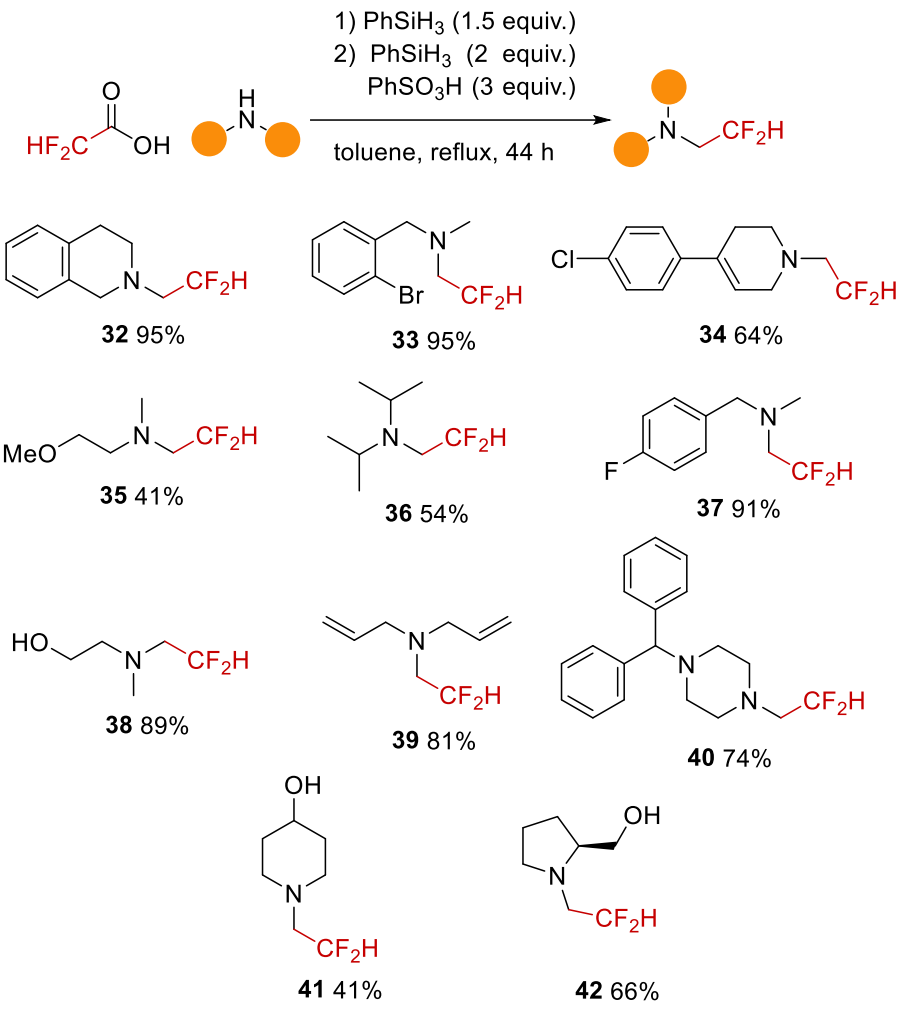
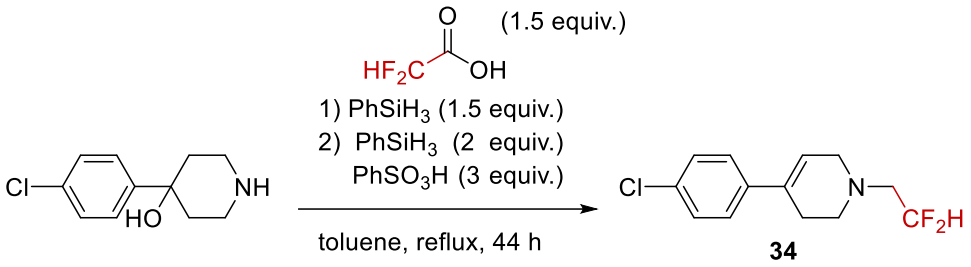


Figure 26: tertiary difluoroethylamine scope

One of the most important findings in this scope are key functional groups that would typically not be compatible with silanes, primary and secondary alcohols would be expected form a silyl ether with phenylsilane however, this group may be labile in highly acidic conditions and decompose back to the alcohol. Tertiary alcohols do not survive and instead undergo the expected elimination to an alkene as shown in **Scheme 86**. This compound **34** however is a useful fragment due to the position of the alkene, as such this represents a convenient synthetic route for this compound.



Scheme 86: elimination of a tertiary alcohol

In addition, large bulky amines as shown in compound **40** are also well supported giving good yields. Compound **36** is a difluorinated analogue of DIPEA and the difluoroethyl group will likely impact the basicity of nitrogen as explained in section 3.1.1. alkenes are also well tolerated as shown by the diallyl example (compound **39**) giving good a good yield. Compound **33** is of note not only for its excellent yield, but it is a fragment common in many drug candidates found in patent literature.

In supporting work by coworkers denoted with "s", additional examples were produced as shown in **Figure 27**. Within these examples by Connor Nolan it was discovered that functional groups sensitive to reduction such as nitriles shown in compound **s1** and methyl esters (compounds **s2** and **s3)** can endure the reaction conditions and afford good to excellent yields.[53]

Large bulky amines appear to impair the reaction, in work shown by Alex Uner dioctylamine **s11** showed very poor reduction likely due to the steric bulk preventing hydrosilylation leading to a yield of 26%.[54] A similar but lesser effect can be seen with dibenzyl amine **s6**.

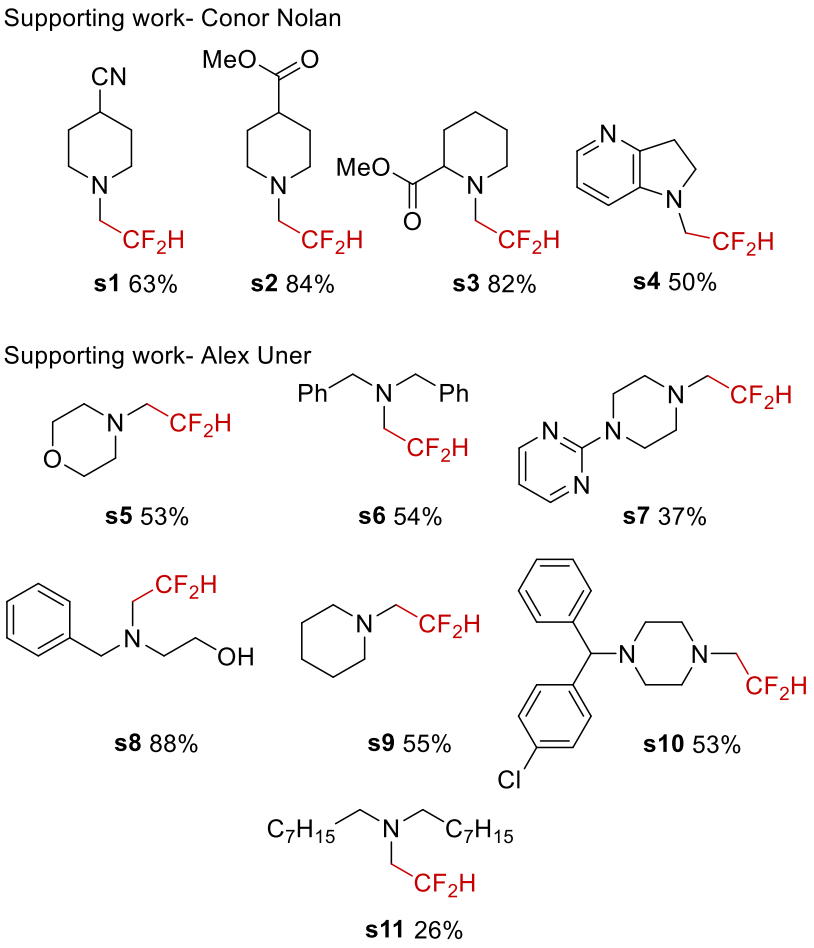
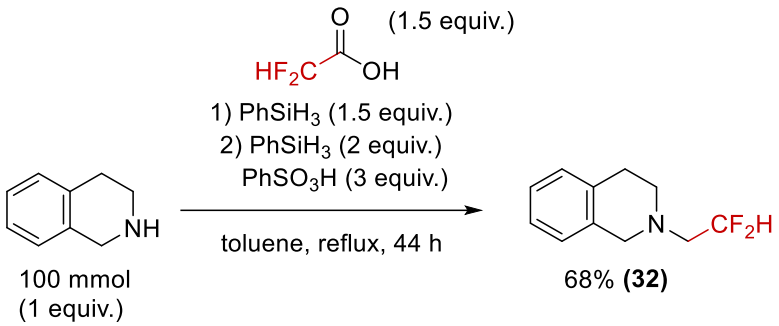


Figure 27: supporting work by co-workers

From analysis of the crude product from the reaction it was uncovered that the unreduced amide is the principal impurity, this accounts for most of the mass balance for this reaction. Part of this is due to some of the amines are not fully soluble in the reaction mixture which can have a negative impact on yield, this can be overcome by switching the solvent to acetonitrile.

Large scale synthesis



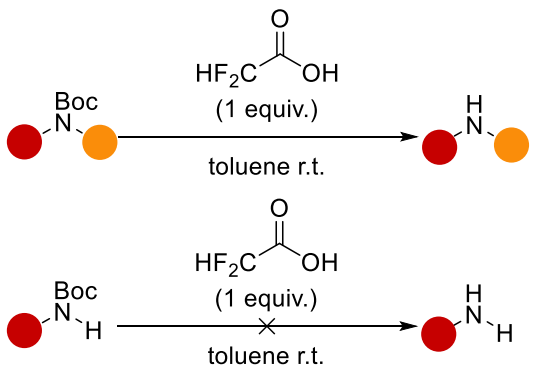
Scheme 87: large scale synthesis of a difluoroethylamine

To prove the scalability and suitability for process chemistry of this method a large-scale variant at 100 mmol was performed (**Scheme 87**, **Figure 28**). Compound **32** was the target molecule for this reaction. The principal challenge of converting this reaction to a large scale is controlling the evolution of hydrogen gas produced in the reaction. In particular the addition of the benzene sulfonic acid can energetically release hydrogen gas as such the benzene sulfonic acid was added portionwise to mitigate this. The outcome of this reaction was favourable the yield dropped comparison to the standard reaction, but this is likely due to mixing within the reactor. It is possible the longer reaction times or more vigorous mixing would improve yield.

Figure 28: large scale reaction in a jacked vessel

3.2.3 *In situ* Boc Deprotections and Unsymmetrical Piperazines

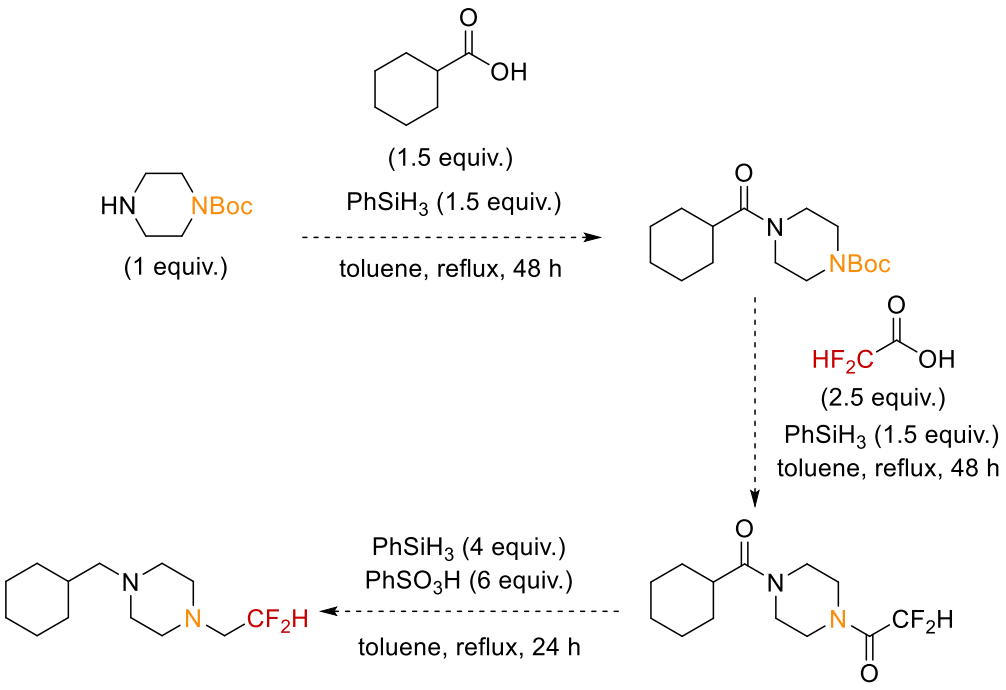
During the course of the tertiary difluoroethylamine scope it was uncovered that difluoroacetic acid can deprotect a secondary Boc-protected amine. In conjunction with a co-worker two test reaction to deprotect both a primary and secondary amine uncovered that difluoroacetic acid is sufficient to deprotect a secondary amine but not a primary (**Scheme 88**).



Scheme 88: use of difluoroacetic acid for the deprotection of a Boc group

A limitation from this test it that it has only been performed once, to make a more robust analysis replicates should be performed.

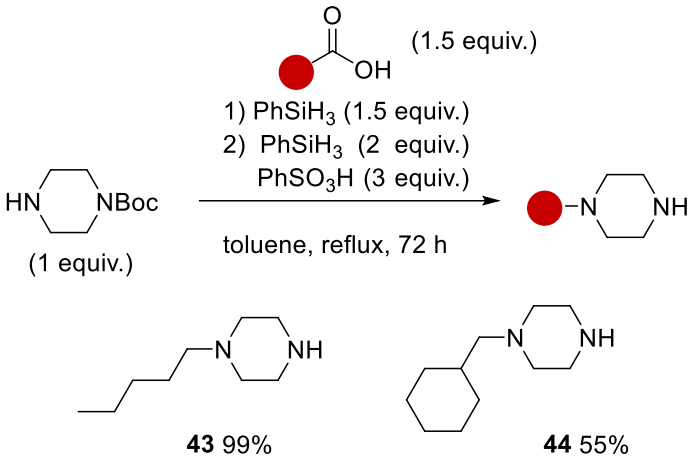
The discovery of this deprotection led to the development of a process to produce unsymmetrical piperazines. The planned synthesis involved a one pot process in which the two amidations would be performed sequentially and then reduced together to give the desired product. Previous attempts to produce unsymmetrical piperazines using this reaction have proved unsuccessful due to the need for a separate deprotection step where the deprotecting agent may interfere. So, it would be advantageous to develop a reaction that could.



Scheme 89: sequence for one pot synthesis of unsymmetrical piperazines

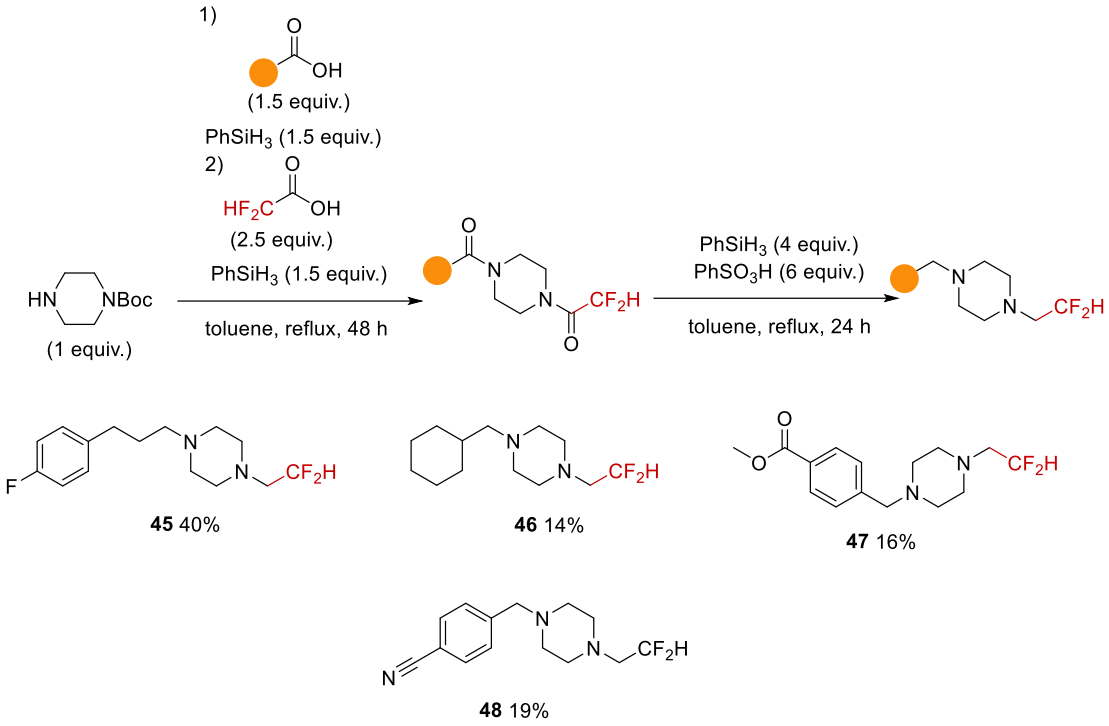
Using this sequence, a series of unsymmetrical piperazines was produced is varied yields (**Scheme 91**). The apparent low yields are due to the complexity of the procedure, the principal impurities were a mixture of partially reduced amides as observed by ^{1}H NMR spectroscopy.

To validate the sequence shown above two control reactions were performed, in both cases the Boc group did not survive to give the monoalkylated products **43** and **44** as shown in **Scheme 90**. With this validation in place the scope for the unsymmetrical piperazines was undertaken.



Scheme 90: monoalkylated piperazines

The scope for the unsymmetrical piperazines as shown above in **Scheme 90**, offer poor yields. This was due to the overall complexity of the process. An *in situ* Boc deprotection followed by two successive rounds of amidation coupled with the simultaneous reduction of both amides. This series of reactions in one pot leads to a very complex environment with may impair the final reduction steps. The two notable examples compounds, as both contain groups that would typically be reduced, though this has been seen in the previous section it was valuable to re-validate as this version of the reaction uses significantly more of the reducing agent. Analysis of the crude reaction mixture showed that the principal impurities were the unreduced amides. This implies that it is the reduction step that is not going to completion.



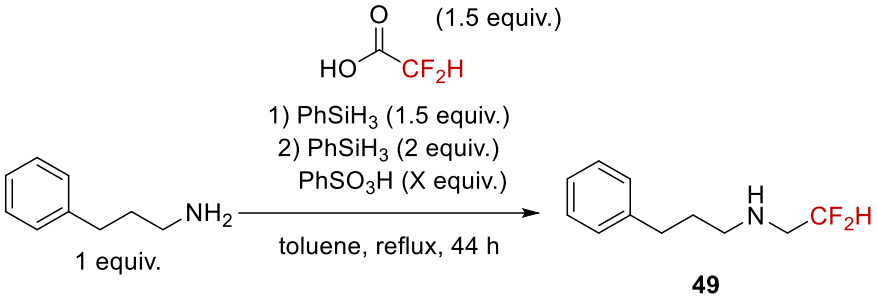
Scheme 91: unsymmetrical piperazine scope

The results shown in **Scheme 91** are an excellent proof of concept that this reaction is a viable process. This represents a convenient but not efficient pathway to developing a complex molecule in one pot process. The remaining work needed is to optimise the final step to maximize the yield.

3.2.4 Secondary Difluoroethylamines

Secondary difluoroethylamines are far more challenging to produce as secondary amides are less reactive than tertiary amides. With this in mind an optimisation screen was undertaken to convert the reaction to secondary difluoroethylamines. The first part of the screen was to examine the loading of benzene sulfonic acid. This proved unsuccessful as the yield never exceeded the yield of the control experiment.

Table 8: benzene sulfonic acid loading screening



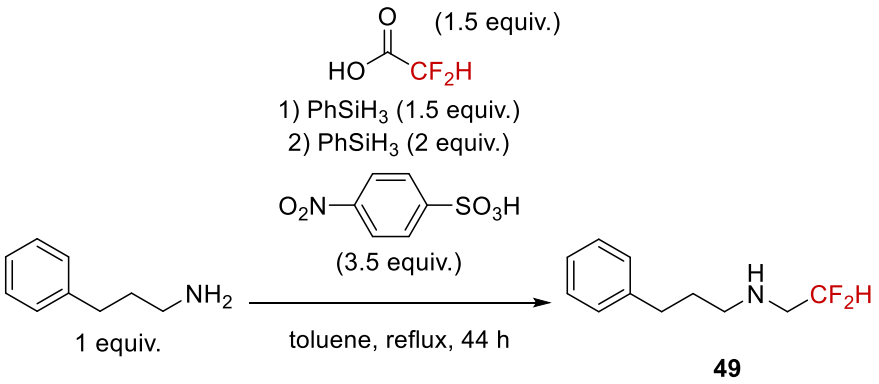
49

Entry	Benzene sulfonic acid (equiv.)	Yield[a]
1	6	15
2	5.5	17
3	5	24
4	4.5	10
5	4	23
6	3.5	7
7	3	29

[a]isolated yield

Instead, a different stronger sulfonic acid was tested. *para*-nitro benzene sulfonic acid, this acid was produced from the hydrolysis of the commercially available acid chloride. 4-nitro benzene sulfonic acid proved to be completely insoluble in toluene, so a solvent screen was performed to counter this. The screen proved successful with a noticeable improvement in yield when acetonitrile was used. 2-MeTHF also had an improvement in yield but proved lesser than acetonitrile. CPME and DMC proved highly unsuccessful and did not yield any product. Analysis of the crude reaction mixture showed only the unreacted amide being present.

Table 9: solvent screen



Entry	Solvent	Yield[A]
1	toluene	0
2	acetonitrile	90
3	2MeTHF	55
4	DMC	0
5	CPME	0

[a]isolated yield

The scope for secondary difluoroethylamines was attempted however this proved unsuccessful as only trace amounts of product could be seen with the unreacted amide. The development of a general method for a secondary difluoroethylamine has proven to be very challenging and elusive. A more detailed examination of the reaction and the reagents used will be required to ensure a general method.

3.3 Conclusions

In this work it has been demonstrated that using difluoroacetic acid is a viable starting material to synthesise difluoroacetamides and tertiary difluoroethylamines in a one pot amidation and amide reduction reaction. A short examination of the amidation step offers the difluoroacetamides in good yields. The full one pot reaction offers good functional group tolerance with esters, nitriles and alcohols being readily compatible, and in good yields. . In addition, it has been shown that difluoroacetic acid can deprotect a secondary Boc protected

amine but under current conditions will not deprotect a primary amine. This was used in a one-pot synthesis of a series of unsymmetrical piperazines from N-Boc-piperazine. The development of a general procedure for secondary difluoroethylamines is still on going.

3.4 Future Work

The remain work on this reaction is centred on the optimisation of the reaction for secondary difluoroethylamines. So far has proven challenging and may require the thorough screening of different silanes and sulfonic acids to overcome the lower reactivity of the secondary difluoroacetamide. Once the general procedure is developed the scope for secondary difluoroethylamine can then be examined.

Additionally, for the unsymmetrical piperazines the final reduction step needs optimisation to improve the yields. As discussed in section 3.2.3 the overall yield for the reaction was low due to incomplete reduction of the amides, a re-examination of this reaction step may prove to be beneficial in enhancing the final yield.

It may also be advantageous to apply both the difluoroethylation and the one pot unsymmetrical piperazines to the synthesis of either a natural product or APIs. This can establish how effective these processes can be as part of a more complex multi-step synthesis.

3.5 Experimental

General Information

All regents were purchased from commercial suppliers and used without further purification. Dry 2-MeTHF and DMC was purchased from Sigma Aldrich and acetonitrile and toluene were dried over sodium wire and obtained from a solvent tower, where degasses solvent was passed though columns of activated alumina and a 7-micron filter under 4-bar of pressure. All water was deionised before use, all experiments were carried out in oven dried glassware with an argon balloon atmosphere.

Analysis and characterisation

Analytical Thin Layer Chromatography (TLC) was performed on Merck aluminium-backed silica-gel plates 60 F254 plates and visualized by ultraviolet (UV) irradiation (254 nm) or by staining with a solution of potassium permanganate or vanillin. Fourier Transform Infrared Spectrometry (IR) was carried out using a Bruker Tensor 27 using an Attenuated Total Reflection (ATR) attachment and peaks are reported in terms of frequency of absorption (cm^{-1}). High Resolution Mass Spectrometry (HRMS) were measured on a Bruker microTOF II with Electron Spray Ionisation (ESI).

^{1}H NMR spectra were recorded on either a Bruker AV 400, AV(III) 400HD or AV(III) 500HD in CDCl3. ^{1}H NMR chemical shifts (δ) were reported in parts per million (ppm) and coupling constants (*J*) are given in Hertz (Hz), with residual protic solvent as the internal reference ($CDCl_3$ δ = 7.26 ppm). The proton spectra are reported as follows: δ (multiplicity, coupling constant J, number of protons). Abbreviations used include s – singlet, d – doublet, t – triplet, q – quartet, sept – septet, m – multiplet, br – broad, app. – apparent. ^{13}C NMR were recorded on a 400 MHz spectrometer, chemical shifts (δ) were reported in ppm relative to the ^{13}C signals in the solvent (central peak of $CDCl_3$ δ = 77.16 ppm) and coupling constants (J) are given in Hertz (Hz). All ^{13}C NMR are reported as proton decoupled spectra. ^{19}F NMR were recorded on a 376 MHz spectrometer, chemical shifts (δ) were reported in ppm relative to $CFCl_3$ at 0.00 ppm and are reported as proton decoupled spectra.

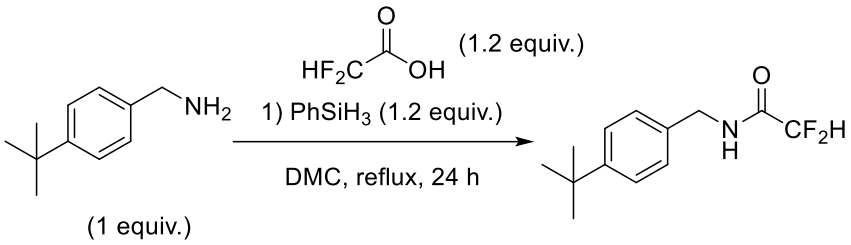
3.5.1 Difluoroacetamides

General procedure for the synthesis of difluoroacetamides

To a room temperature solution of difluoro acetic acid (115.2 mg, 1.20 mmol) in dimethyl carbonate (2 mL), was added amine (1.00 mmol, 1.00 equiv.) then phenylsilane (129.8 mg, 1.20 mmol) or After 24 hours reaction was filtered then diluted with EtOAc. The product was extracted with HCL (10 mL of a 1 M aqueous solution), NaOH (10 mL of a 1 M aqueous solution)

and saturated brine (10 mL). The combined organic layers were dried with $MgSO_4$ and the solvent removed under reduced pressure to afford the product stated below.

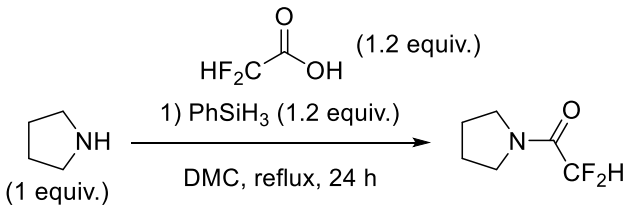
25) *N*-(4-(tert-butyl)benzyl)-2,2-difluoroacetamides



Produced under the general procedure for difluoroacetamides, 163.2 mg, 1.00 mmol of 4-tertbutyl benzylamine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 40%. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.48 – 7.42 (m, 2H), 7.31 – 7.26 (m, 2H), 5.91 (t, J = 54.4 Hz, 1H), 4.50 (d, J = 5.8 Hz, 2H), 1.39 (s, 9H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 162.62 (t, *J* = 24.6 Hz), 151.1, 133.7, 127.8, 125.9, 108.6 (t, *J* = 252.2 Hz), 43.1, 34.6, 31.34. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -126.14. **IR (cm^{-1}):** 3312, 2955, 1679, 1559, 1513, 1377, 1361, 1285, 1107, 1082, 1046, 1026, 899, 851, 833, 818, 792, 674, 574, 474. **HRMS (ESI+):** Exact mass calcd for $C_{13}H_{17}F_2NONa$, 264.1170 found 264.1171 (M+Na).

26) 2,2-difluoro-1-(pyrrolidin-1-yl)ethan-1-one

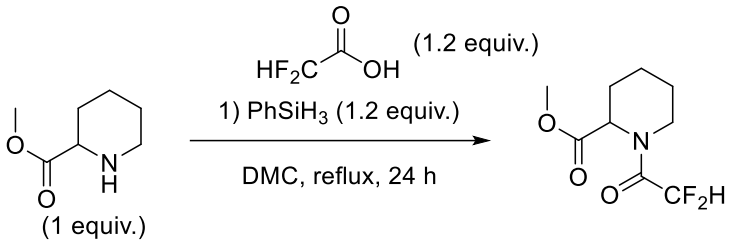


Produced under the general procedure for difluoroacetamides, 71.2 mg, 1.00 mmol of proline used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 54%. **^{1}H NMR** (400 MHz, CDCl3) δ 5.99 (t, *J* = 53.8 Hz, 1H), 3.51 (dt, *J* = 32.2, 6.8 Hz, 4H), 1.95 – 1.69 (m, 4H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 160.9 (t, *J* = 25.6 Hz), 109.9 (t, *J* =

252.0 Hz), 46.7, 23.3. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -123.33. **Ir (cm^{-1}):** 3073, 1661, 1593, 1458, 1430, 1346, 1058, 912, 850, 726, 695, 585, 480. **HRMS (ESI+):** Exact mass calcd for $C_6H_9NOF_2Na$, 172.0544, found 172.0533 (M+Na).

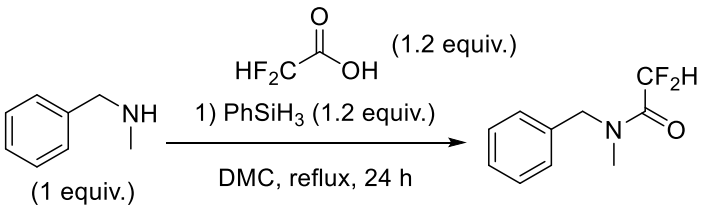
27) methyl 1-(2,2-difluoroacetyl)piperidine-2-carboxylate



Produced under the general procedure for difluoroacetamides, 143.2 mg, 1.00 mmol of methyl piperidine-2 carboxylate used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 24%. **^{1}H NMR** 6.07 – 5.72 (m, 1H), 4.12-4.15 (m, 1H), 2.06 (s, 2H), 1.83 – 1.60 (m, 2H), 1.28 (m, 3H), 1.07 – 0.86 (m, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 171.3, 161.9 (t, *J* = 24.3 Hz), 109.1 (t, *J* = 252.8 Hz), 60.5, 48.8, 33.6, 32.6, 29.8, 21.0. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -125.99 **Ir (cm^{-1}):** 2952, 1742, 1666, 1593, 1430, 1327, 1213, 1129, 1016, 920, 848, 737, 696, 579, 481. **HRMS (ESI+):** Exact mass calcd for $C_9H_{13}F_2NO_3Na$, 244.0756 found 244.0755 (M+Na).

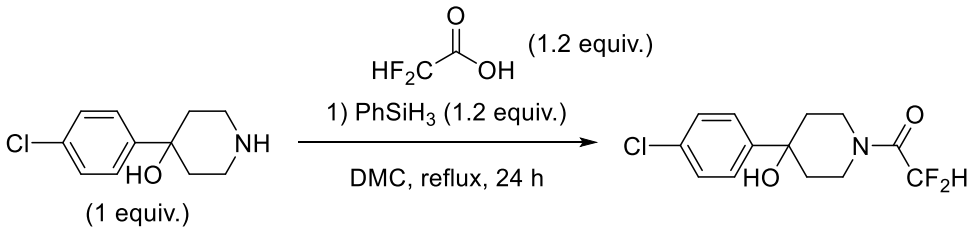
28) *N*-benzyl-2,2-difluoro-*N*-methylacetamide



Produced under the general procedure for difluoroacetamides, 121.1 mg, 1.00 mmol of N-methyl benzylamine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy. NMR Spectroscopy shows a mixture of rotamers, minor rotamer peaks shown in italics.

Isolated yield: 63%. **^{1}H NMR** (400 MHz, CDCl3) δ 7.47 – 7.34 (m, 5H), 6.21 (t, *J* = 53.2 Hz, 1H), (*6.25 (t, J = 53.2 Hz, 1H)*, 4.63 (s, 2H), (*4.67* (*s, 2H)*) 2.93 (s, 1H), (*3.05 (s, 3H)*). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 162.4 (t, J = 24.9 Hz), , 135.7, 129.0, 128.9, 128.2, 127.9, 127.2, 112.8, 110.4, 110.5, 107.7, 51.6, 33.6. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -122.16. **Ir (cm^{-1}):** 1670, 1594, 1496, 1454, 1430, 1365, 1091, 1027, 865, 727, 695, 486 **HRMS (ESI+):** Exact mass calcd for $C_{10}H_{11}NOF_2Na$, 222.0701, found 222.7072(M+Na).

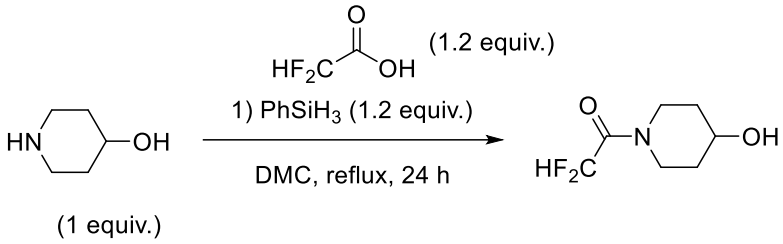
29) 1-(4-(4-chlorophenyl)-4-hydroxypiperidin-1-yl)-2,2-difluoroethan-1-one



Produced under the general procedure for difluoroacetamides, 211.6 mg, 1.00 mmol of 4-(4-chlorophenyl)piperidin-4-ol used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 53%. **^{1}H NMR** (101 MHz, $CDCl_3$) δ 7.51 – 7.38 (m, 2H), 7.29 – 7.19 (m, 2H), 6.16 (t, *J* = 53.7 Hz, 1H), 2.20 – 1.65 (m, 8H). **^{13}C NMR** 160.7 (t, *J* = 25.0 Hz), 146.0, 129.1, 128.6, 126.1, 110.4 (t, *J* = 253.3 Hz), 70.8, 39.0, 38.7, 38.2, 37.5. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -121.26. **Ir (cm^{-1}):** 3432, 2954, 1665, 1594, 1492, 1462, 1430, 1374, 1270, 1130, 1091, 1027, 1009, 825, 728, 695, 544, 482. **HRMS (ESI+):** Exact mass calcd for $C_{13}H_{14}ClF_2NO_2Na$, 312.0573 found 312.0573 (M+Na).

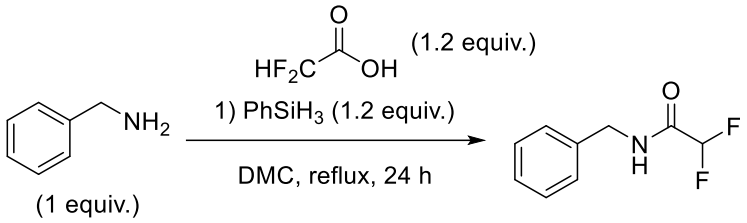
30) 2,2-difluoro-1-(4-hydroxypiperidin-1-yl)ethan-1-one



Produced under the general procedure for difluoroacetamides, 101.1 mg, 1.00 mmol of 4-hydroxypiperidine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 95%. **^{1}H NMR** 6.07 (t, J = 53.5 Hz, 1H), 3.92 – 3.62 (m, 3H), 3.27 – 3.13 (m, 2H), 1.85 – 1.35 (m, 4H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 160.6 (t, *J* = 25.3 Hz), 109.9 (t, *J* = 252.5 Hz), 66.0, 39.9, 33.1. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -121.77 **Ir (cm^{-1}):** 2932, 1660, 1464, 1430, 1374, 1266, 1129, 1038, 915, 874, 847, 730, 696, 480. **HRMS (ESI+):** Exact mass calcd for $C_7H_{12}F_2NO_2$, 180.0827 found 180.0831 (M+H).

31) *N*-benzyl-2,2-difluoroacetamide



Produced under the general procedure for difluoroacetamides, 107 mg, 1.00 mmol of benzylamine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

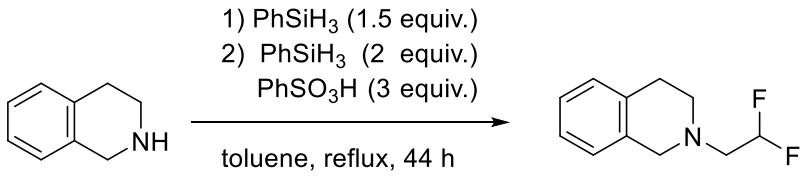
Isolated yield: 97%. **^{1}H NMR** (400 MHz, CDCl3) δ 7.46 – 7.22 (m, 6H), 5.91 (t, J = 54.4 Hz, 1H), 4.50 (d, J = 5.9 Hz, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 162.65 (t, *J* = 24.8 Hz), 136.63, 128.91, 128.04, 127.91, 108.53 (t, *J* = 252.2 Hz), 43.32. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -126.26. **Ir (cm -1):** 3295, 1681, 1547, 1497, 1454, 1343, 1054, 878, 732, 698, 602, 482. **HRMS (ESI+):** Exact mass calcd for $C_9H_9F_2NONa$, 208.0545 found 208.0540 (M+Na).

3.5.2 Tertiary Difluoroethylamines

General procedure for reductive difluoroethylation of secondary amines

To a refluxing solution of difluoroacetic acid (144 mg, 1.50 mmol) in toluene (1 mL) was added phenylsilane (162 mg, 1.50 mmol) and amine (1.00 mmol, 1.00 equiv.). After 24 hours additional phenylsilane (216 mg, 2.00 mmol) and benzene sulfonic acid (474 mg, 3.00 mmol). After a further 18 hours reaction was diluted with EtOAc (10 mL) and washed with HCl (10 mL of a 3M aq. solution). Aqueous solution was made to pH 14 with NaOH (10 mL of a 6M aq. solution), crude product extracted from aqueous solution with DCM (3x 10 mL) to give the products show below.

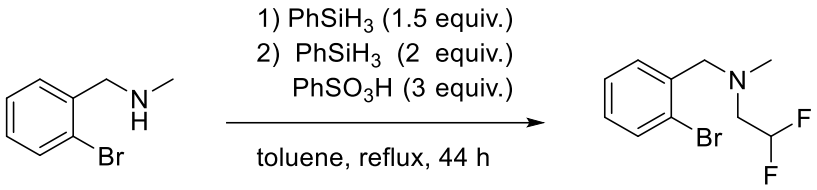
32) 2-(2,2-difluoroethyl)-1,2,3,4-tetrahydroisoquinoline[55]



Produced according to general procedure for difluoroethylation of secondary amines, 133.2 mg, 1.00 mmol of 1,2,3,4-tetrahydroisoquinoline used, crude product purified via a pad of silica (EtOAc 100%)

Isolated yield: 95%. **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.22 – 7.01 (m, 4H), 5.99 (tt, J = 56.1, 4.4 Hz, 1H), 3.81 (s, 2H), 3.01 – 2.87 (m, 6H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 134.1, 133.8, 128.8, 126.6, 126.4, 125.8, 118.5, 116.1 (t, *J* = 241.1 Hz), 113.68, 60.08, 59.83 (t, *J* = 24.9 Hz), 59.59, 56.53, 51.77. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -118.11. **HRMS (ESI+):** Exact mass calcd for $C_{11}H_{14}F_2N$, 198.1089 found 198.1085 (M+H).

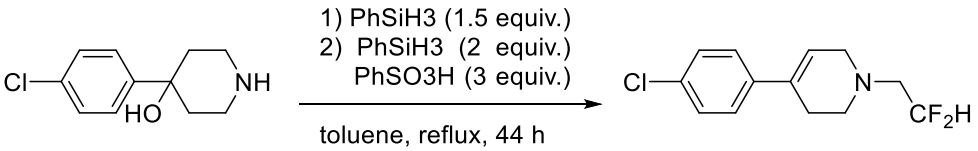
33) N-(2-bromobenzyl)-2,2-difluoro-N-methylethan-1-amine



produced according to general procedure for difluoroethylation of secondary amines, 200 mg, 1.00 mmol of 2-bromo benzyl methylamine. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 95% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.62 – 7.42 (m, 2H), 7.36 – 7.04 (m, 2H), 5.89 (tt, *J* = 56.0, 4.4 Hz, 1H), 3.73 (s, 2H), 2.87 (td, *J* = 14.9, 4.4 Hz, 2H), 2.40 (s, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 137.6, 132.9, 130.8, 128.7, 127.4, 124.5, 116.0 (t, *J* = 241.4 Hz), 61.8, 59.1 (t, *J* = 24.9 Hz), 43.1. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -118.72. . **Ir (cm^{-1}):** 2794, 1567, 1439, 1366, 1276, 1123, 1022, 888, 748, 689, 661, 500, 437. **HRMS (ESI+):** Exact mass calcd for $C_{10}H_{13}F_2NBr$, 264.0194, found 264.1090 (M+H).

34) 4-(4-chlorophenyl)-1-(2,2-difluoroethyl)-1,2,3,6-tetrahydropyridine

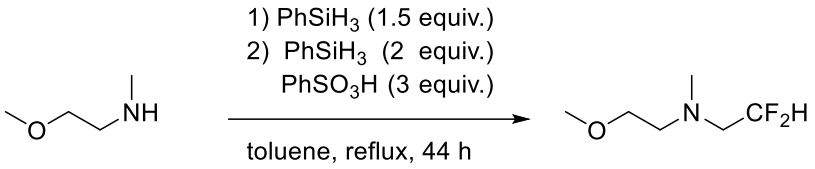


produced according to general procedure for difluoroethylation of secondary amines, 212 mg, 1.00 mmol of 4-(4-chlorophenyl)piperidin-4-ol used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 63% as an off white crystalline solid **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.34 – 7.20 (m, 2H), 5.97 (tt, *J* = 56.1, 4.4 Hz, 1H), 3.34 – 3.18 (m, 1H), 2.96 – 2.74 (m, 2H), 2.59 – 2.45 (m, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 138.9, 133.9, 132.8, 128.4, 126.1, 121.8, 115.9 (t, *J* = 241.0 Hz), 59.7 (t, *J* = 24.9 Hz), 53.6, 50.9, 27.7. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -118.28. **Ir (cm^{-1}):** 2924, 1644,

1495, 1465, 1404, 1370, 1320, 1147, 1123, 1042, 1008, 964, 800, 772, 730, 617. **HRMS (ESI+):** Exact mass calcd for $C_{13}H_{14}F_2NCl$, 257.0783 found 257.0857 (M+H). **Melting Point:** 80 °C

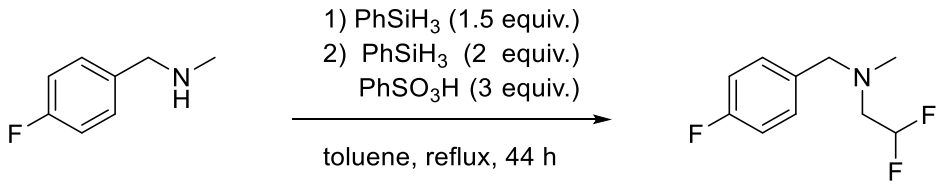
35) 2,2-difluoro-*N*-(2-methoxyethyl)-*N*-methylethan-1-amine



produced according to general procedure for difluoroethylation of secondary amines, 89.1 mg, 1.00 mmol of *N*-(2-methoxyethyl) methylamine. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 41% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.79 (tt, *J* = 55.9, 4.4 Hz, 1H), 3.42 (t, *J* = 5.5 Hz, 2H), 3.29 (s, 3H), 2.77 (td, *J* = 15.0, 4.4 Hz, 2H), 2.65 (t, *J* = 5.5 Hz, 2H), 2.35 (s, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 115.9 (t, *J* = 240.9 Hz), 70.6, 59.4 (t, *J* = 24.7 Hz), 58.7, 57.6, 43.6. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -119.20. **Ir (cm^{-1}):** 2922, 1594, 1430, 1090, 1026, 727, 694, 585, 483, 420. **HRMS (ESI+):** Exact mass calcd for 154.1038, found 154.1036 (M+H).

37) 2,2-difluoro-*N*-(4-fluorobenzyl)-*N*-methylethan-1-amine

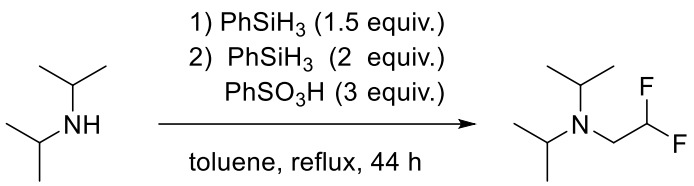


produced according to general procedure for difluoroethylation of secondary amines, 138.2 mg, 1.00 mmol of 4-fluoro-*N*-methyl benzylamine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 91% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.38 – 7.28 (m, 2H), 7.04 (t, *J* = 8.6 Hz, 2H), 5.86 (tt, *J* = 56.2, 4.4 Hz, 1H), 3.61 (s, 2H), 2.80 (td, *J* = 14.9, 4.4 Hz, 2H), 2.37 (s, 3H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 160.9, 134.2, 130.4, 130.4, 115.7 (t, *J* = 241.4 Hz), 61.9, 58.5 (t, *J* = 24.7 Hz),

43.12. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -115.51, -118.82. **Ir (cm^{-1}):** 2796, 1637, 1508, 1457, 1368, 1220, 1155, 1123, 1093, 1032, 851, 820, 769, 554, 503. **HRMS (ESI+):** Exact mass calcd for $C_{10}H_{13}F_3N$, 204.0995, found 204.0994 (M+H).

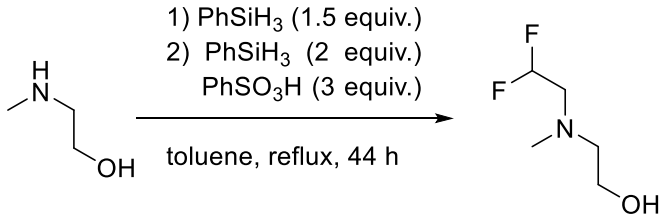
36) N-(2,2-difluoroethyl)-N-isopropylpropan-2-amine



produced according to general procedure for difluoroethylation of secondary amines, 101.2 mg, 1.00 mmol of diisopropylamine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 54% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.89 – 5.42 (m, 1H), 3.05 (h, *J* = 6.6 Hz, 2H), 2.80 (td, *J* = 14.7, 4.5 Hz, 2H), 1.03 (d, *J* = 6.8 Hz, 13H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 117.8 (t, *J* = 242.1 Hz), 49.7, 48.2 (t, *J* = 25.7 Hz), 20.8. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -119.60. **Ir (cm -1):** 2252, 1375, 1331, 1216, 1105, 917, 859, 765, 740, 663, 558. **HRMS (ESI+):** Exact mass calcd for $C_8H_{18}NF_2$, 166.1402, found 166.1402 (M+H).

38) 2-((2,2-difluoroethyl)(methyl)amino)ethan-1-ol

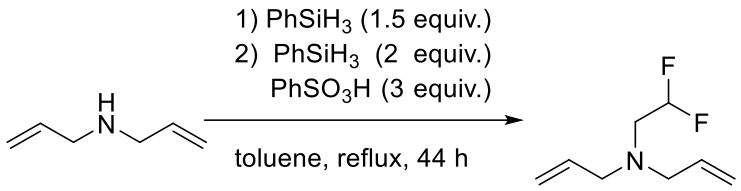


produced according to general procedure for difluoroethylation of secondary amines, 75 mg, 1.00 of *N*-methyl ethanolamine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy

Isolated yield: 89% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.83 (tt, *J* = 55.8, 4.1 Hz, 1H), 3.57 (t, *J* = 5.4 Hz, 2H), 2.79 (td, *J* = 15.0, 4.1 Hz, 2H), 2.63 (t, *J* = 5.4 Hz, 2H), 2.37 (s, 3H).**^{13}C NMR** δ (101 MHz,

$CDCl_3$) δ 115.6 (t, *J* = 241.2 Hz), 59.7, 59.3 (t, *J* = 24.5 Hz), 58.7, 42.9. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -119.76. **Ir (cm^{-1}):** 3383, 2860, 1458, 1410, 1369, 1295, 1124, 1030, 957, 885, 804, 498. **HRMS (ESI+):** Exact mass calcd for $C_5H_{12}F_2NO$, 140.0881 found 140.0877 (M+H).

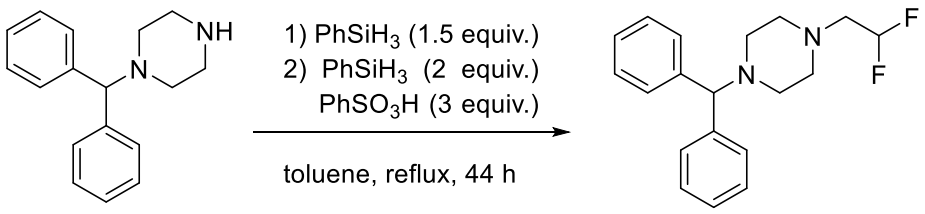
39) N-allyl-N-(2,2-difluoroethyl)prop-2-en-1-amine



produced according to general procedure difluoroethylation of secondary amines, 97 mg, 1.00 mmol of diallylamine used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 81% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 6.04 – 5.64 (m, 1H), 5.28 – 5.17 (m, 1H), 3.24 (d, *J* = 6.6 Hz, 1H), 2.85 (td, *J* = 14.9, 4.4 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 135.0, 118.1, 116.3 (t, *J* = 241.4 Hz), 58.0, 54.8 (t, *J* = 24.9 Hz). **^{19}F NMR** (376 MHz, $CDCl_3$) δ -119.42. **Ir (cm^{-1}):** 2923, 1594, 1429, 1093, 1025, 726, 694, 585, 482. **HRMS (ESI+):** Exact mass calcd for $C_8H_{14}F_2N$, 162.1089 found 162.1099 (M+H).

40) 1-benzhydryl-4-(2,2-difluoroethyl)piperazine

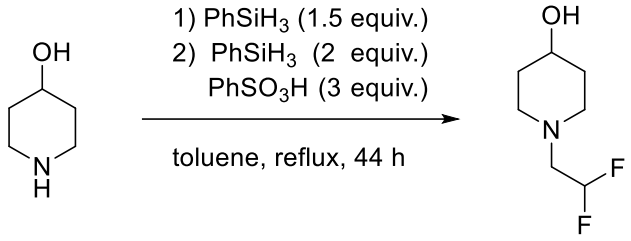


produced according to general procedure 252.3 mg, 1.00 mmol of 1-benzhydrylpiperazine used. Purified by addition of silica and DCM (10mL) then filtration, solvent removed *in vacuo.*

Isolated yield: 74% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.57 – 7.52 (m, 4H), 7.41 – 7.35 (m, 4H), 7.31 – 7.25 (m, 2H), 5.96 (tt, *J* = 55.9, 4.3 Hz, 1H), 2.82 (td, J = 15.1, 4.4 Hz, 2H), 2.71 (s, 4H), 2.55 (s, 4H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 142.7, 128.6, 128.0, 127.1, 115.9 (t, *J* = 240.8 Hz), 76.2, 60.3

(t, *J* = 24.6 Hz), 54.3, 51.9. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -117.94. **Ir (cm^{-1}):** 2810, 1671, 1491, 1450, 1303, 1132, 1044, 1007, 967, 923, 745, 703, 644, 609. **HRMS (ESI+):** Exact mass calcd for $C_{19}H_{23}F_2N_2$, 317.1824 found 317.1816 (M+H).

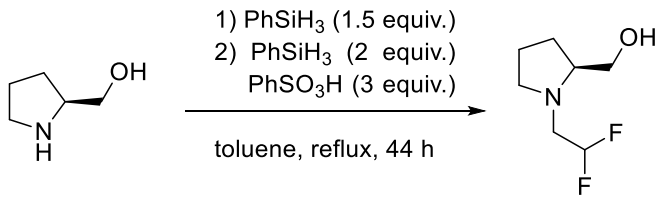
41) 1-(2,2-difluoroethyl)piperidin-4-ol



Reaction performed in 1 mL of acetonitrile all other conditions as per the general procedure for difluoroethylation of secondary amines, 101.1 mg, 1.00 mmol of 4-hydroxy piperidine. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 41% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.85 (tt, *J* = 56.1, 4.4 Hz, 1H), 3.65 (tt, *J* = 9.0, 4.2 Hz, 1H), 2.84 – 2.76 (m, 2H), 2.69 (td, *J* = 15.1, 4.3 Hz, 2H), 2.37 – 2.25 (m, 2H), 1.88 – 1.78 (m, 2H), 1.63 – 1.46 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 115.76 (t, *J* = 241.0 Hz), 67.12, 59.99 (t, *J* = 24.7 Hz), 51.81, 34.23. **19F NMR** δ -118.39. (376 MHz, $CDCl_3$) **Ir (cm -1):** 3334, 2943, 1429, 1368, 1285, 1118, 1041, 982, 887, 789, 689, 490. **HRMS (ESI+):** Exact mass calcd for $C_7H_{14}F_2NO$, 166.1038, found 166.1040 (M+H).

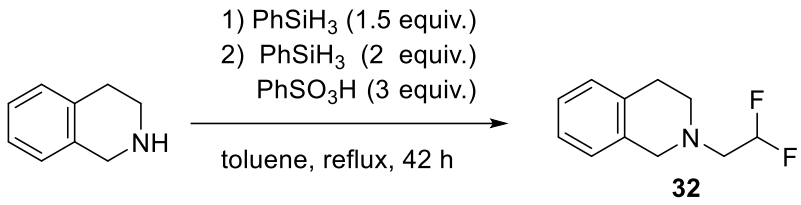
42) (*S*)-(1-(2,2-difluoroethyl)pyrrolidin-2-yl)methanol



produced according to general procedure for difluoroethylation of secondary amines, 101.2 mg, 1.00 mmol of *S*-prolinol used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 66% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.84 (tt, *J* = 56.1, 4.2 Hz, 1H), 3.65 – 3.35 (m, 2H), 3.32 – 2.98 (m, 2H), 2.87 – 2.36 (m, 4H), 2.08 – 1.52 (m, 4H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 115.8 (t, *J* = 240.7 Hz), 65.7, 62.7, 56.8 (t, *J* = 24.6 Hz), 55.9, 27.1, 24.0. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -118.85. **Ir (cm^{-1}):** 2922, 2594, 1430, 1090, 1026, 727, 694, 585, 483, 420. **HRMS (ESI+):** Exact mass calcd for $C_7H_{13}F_2NO$, 166.1038, found 166.1038 (M+H).

Multigram synthesis of **32)** 2-(2,2-difluoroethyl)-1,2,3,4-tetrahydroisoquinoline



32

To a solution of difluoroacetic acid (14 g, 150 mmol) in toluene (100 mL) was added phenylsilane (16 g, 150 mmol) and 1,2,3,4-tetrahydroisoquinoline (13 g, 100 mmol). After 24 hours reaction was cooled to 100 °C then phenylsilane (22 g, 200 mmol) and benzenesulphonic acid (74 g 300 mmol) added portionwise (9x5 g over 15 minutes) reaction heated to 100 °C. After a further 18 hours reaction was diluted with EtOAc (100 mL) and washed with HCl (2x100 mL of a 3M aq. solution). Aqueous solution was made to pH 14 with NaOH (150 mL of a 6M aq. solution), crude product extracted from aqueous solution with DCM (3x 70 mL). crude product purified by pad of silica (100% DCM) to give 12.17 g of 2-(2,2-difluoroethyl)-1,2,3,4-tetrahydroisoquinoline with an isolated yield of 62%

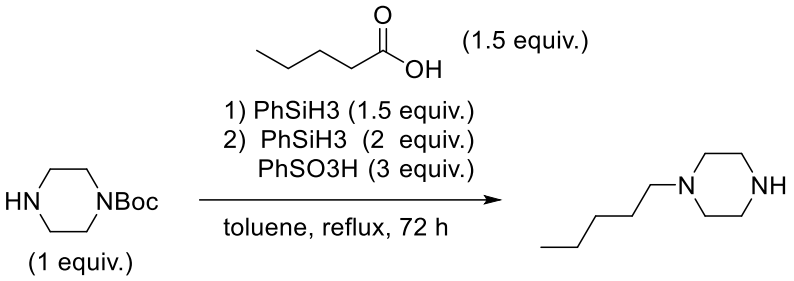
See compound **32** for characterisation data

3.5.3 Unsymmetrical Piperazines

General procedure for the synthesis of monofunctionalized piperazines

To a refluxing solution of **carboxylic acid** (1.5 mmol, 1.50 equiv.) in toluene (1 mL) was added phenylsilane (162 mg, 1.50 mmol) and N-Boc-piperazine (186 mg, 1.00 mmol). After 24 hours additional phenylsilane (216 mg, 2.00 mmol.) and benzenesulfonic acid (474 mg, 3.00 mmol). after a further 18 hours reaction was diluted with EtOAc (10 mL) and washed with HCl (10 mL of a 3M aq. solution). Aqueous solution was made to pH 14 with NaOH (10 mL of a 6M aq. solution), crude product extracted from aqueous solution with DCM (3x 10 mL) to give the products show below.

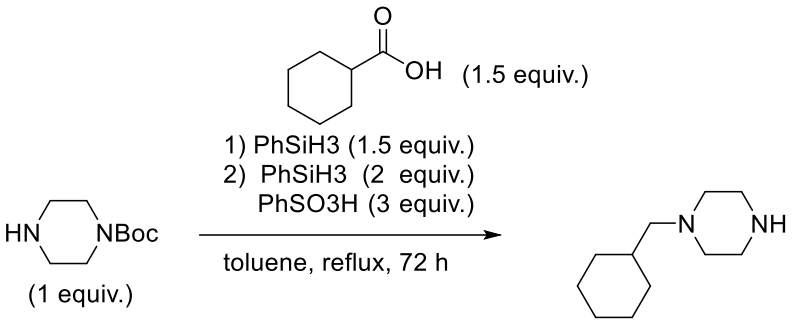
43) 1-pentylpiperazine



Produced under general procedure for monofunctionalized piperazines, 184.3 mg, 1.50 mmol of valeric acid used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 99% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 2.89 – 2.71 (m, 1H), 2.37 – 2.07 (m, 2H), 1.45 – 1.26 (m, 1H), 1.22 – 1.04 (m, 2H), 0.74 (t, *J* = 7.0 Hz, 1H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 59.23, 58.69, 54.22, 53.09, 45.72, 29.69, 26.42, 26.16, 22.47 **Ir (cm^{-1}):** 2929, 2859, 2806, 1605, 1570, 1457, 1311, 1268, 1223, 1143, 1061, 1004, 896, 816, 791, 767, 727, 597, 490. **HRMS (ESI+):** Exact mass calcd for $C_9H_{21}N_2$, 157.1660, found 157.1685 (M+H).

44) 1-(cyclohexylmethyl)piperazine



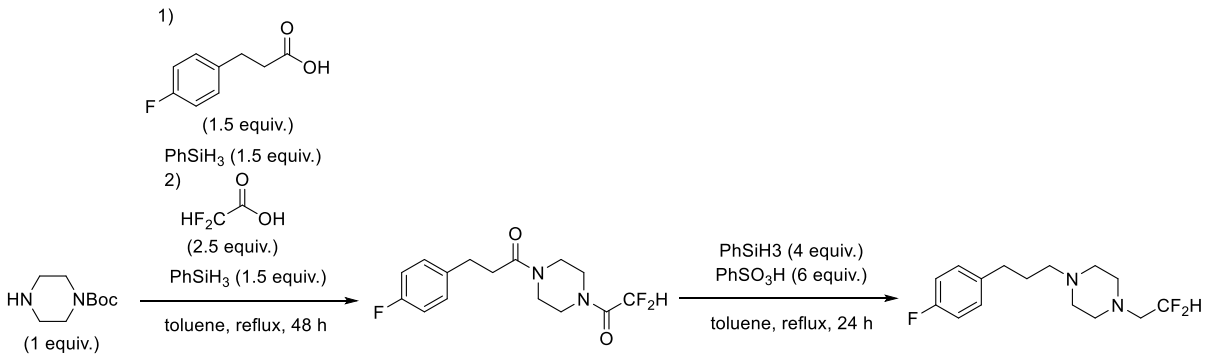
Produced according to the general procedure for monofunctionalized piperazines. 192 mg, 1.50 mmol of cyclohexane carboxylic acid used. No further purification required, purity assessed By ^{1}H NMR spectroscopy.

Isolated yield: 55% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 2.87 (s, 3H), 2.58 (s, 2H), 2.34 (s, 3H), 2.08 (d, *J* = 7.1 Hz, 2H), 1.82 – 1.61 (m, 5H), 1.54 – 1.39 (m, 1H), 1.30 – 1.07 (m, 3H), 0.96 – 0.76 (m, 2H). **^{13}C NMR** 66.3, 54.9, 45.9, 34.8, 31.9, 26.8, 26.2. **Ir (cm^{-1}):** 3276, 2920, 2844, 1639, 1523, 1403, 1327, 1263, 1118, 998, 802, 589, 421. **HRMS (ESI+):** Exact mass calcd for $C_{11}H_{23}N_2$, 183.1856, found 183.1851 (M+H). **melting point:** sublimes at 70 °C

General procedure for the synthesis of difluoro ethyl piperazines

To a refluxing solution of carboxylic acid (1.50 mmol, 1.50 equiv.) in toluene (1 mL) was added phenylsilane (162 mg, 1.50 mmol) and 1-boc-piperazine (182 mg, 1 mmol). After 24 hours additional phenylsilane (162 mg, 1.5 mmol) and difluoroacetic acid (240 mg, 2.5 mmol). After a further 24 hours additional phenylsilane (432 mg, 4 mmol) and benzenesulfonic acid (948 mg, 6 mmol) was added. After a further 24 hours reaction was diluted with EtOAc (10 mL) and washed with HCl (10 mL of a 3M aq. solution). Aqueous solution was made to pH 14 with NaOH (6M aq. solution), crude product extracted from aqueous solution with DCM (3x 10 mL). Crude products purified by column cromatography to give the products show below.

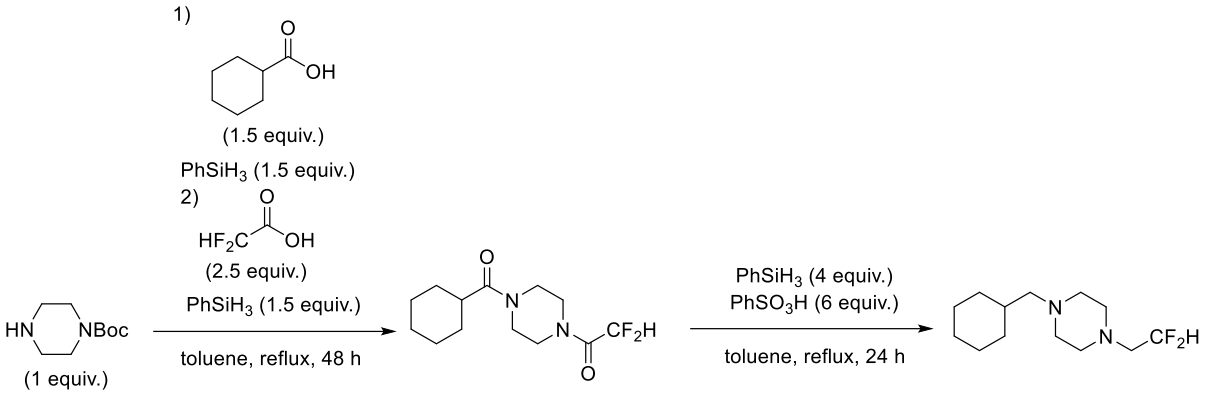
45) 1-(2,2-difluoroethyl)-4-(3-(4-fluorophenyl)propyl)piperazine



Produced according to the general procedure. 252 mg, 1.50 mmol of 3-(4-fluorophenyl)propanoic acid used. Crude purified by column chromatography (0-50% ethyl acetate in cyclohexane)

Isolated yield: 40% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.14 – 7.07 (m, 2H), 6.98 – 6.89 (m, 2H), 5.86 (tt, *J* = 55.9, 4.3 Hz, 1H), 2.72 (td, *J* = 15.0, 4.4 Hz, 2H), 2.64 – 2.55 (m, 6H), 2.45 (s, 4H), 2.37 – 2.28 (m, 2H), 1.83 – 1.69 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 160.0, 137.7, 129.7, 115.7 (t, *J* = 240.6 Hz), 115.1, 114.9, 60.2 (t, *J* = 24.8 Hz), 53.94, 53.00, 32.79, 28.67. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -117.85, -118.25. **Ir (cm^{-1}):** 2924, 2810, 1672, 1600, 1508, 1459, 1291, 1219, 1157, 1118, 1043, 1014, 955, 821, 757, 701, 550. **HRMS (ESI+):** Exact mass calcd for: $C_{15}H_{22}N_2F_3$, 287.1730 found 287.1735 (M+H). **R.F.** 0.4 (50% cyclohexane 50% ethyl acetate).

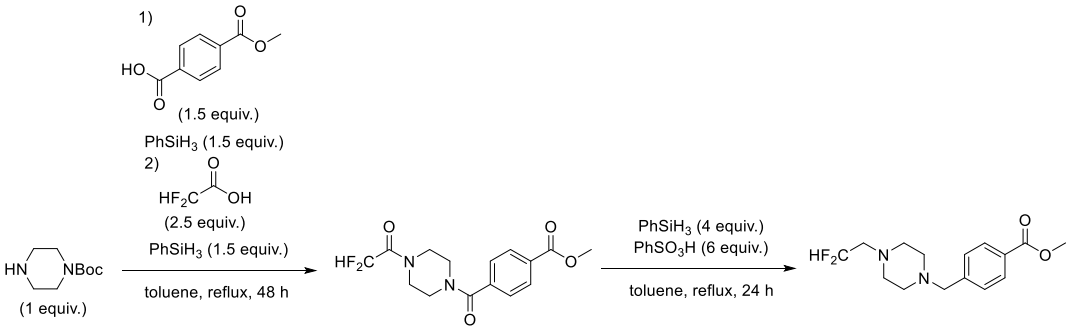
46) 1-(cyclohexylmethyl)-4-(2,2-difluoroethyl)piperazine



Produced according to general procedure, 192 mg, 1.50 mmol of cyclohexane carboxylic acid used. Crude purified by column chromatography (0-50% ethyl acetate in cyclohexane)

Isolated yield: 14% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 5.86 (tt, *J* = 56.1, 4.7 Hz, 1H), 2.70 (td, *J* = 15.0, 4.4 Hz, 2H), 2.58 (s, 4H), 2.40 (s, 4H), 2.09 (d, *J* = 7.1 Hz, 2H), 1.82 – 1.60 (m, 5H), 1.51 – 1.39 (m, 1H), 1.30 – 1.06 (m, 3H), 0.97 – 0.74 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 115.7 (t, *J* = 240.9 Hz), 65.6, 60.2 (t, *J* = 24.7 Hz), 53.9, 53.4, 34.9, 31.9, 26.8, 26.2. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -118.18. **Ir (cm^{-1}):** 2922, 2812, 1458, 1364, 1297, 1123, 1042, 1013, 958, 889, 835, 731, 689. **HRMS (ESI+):** Exact mass calcd for: C12H25F2N2, 247.1980, found 247.1984 (M+H). **R.F.** 0.41 (50% ethyl acetate, 50% cyclohexane).

47) methyl 4-((4-(2,2-difluoroethyl)piperazin-1-yl)methyl)benzoate

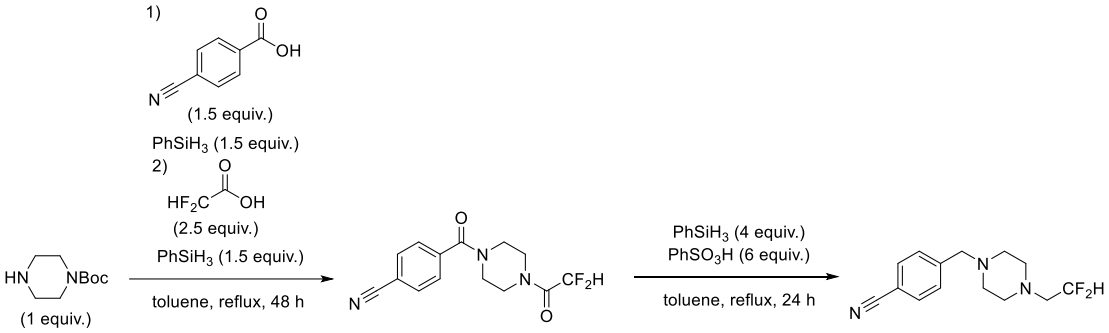


Produced according to general procedure, 270 mg, 1.50 mmol of 4-(methoxycarbonyl)benzoic acid used. Purified by column chromatography (0-40% ethyl acetate in cyclohexane)

Isolated yield: 16% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.97 (d, *J* = 8.0 Hz, 2H), 7.38 (d, *J* = 8.0 Hz, 2H), 5.84 (tt, *J* = 55.9, 4.6 Hz, 1H), 3.89 (s, 2H), 3.53 (s, 2H), 2.72 (td, *J* = 15.1, 4.2 Hz, 2H), 2.63 –

2.56 (m, 4H), 2.50 – 2.42 (m, 4H). **^{13}C NMR** (101 MHz, $CDCl_3$) 167.0, 143.6, 129.6, 129.0, 128.9, 115.7 (t, *J* = 241.0 Hz), 62.5, 60.1 (t, *J* = 24.7 Hz), 53.9, 52.9, 52.0. **^{19}F NMR** (376 MHz, $CDCl_3$) δ 118.29. **Ir (cm^{-1}): HRMS (ESI+):** Exact mass calcd for: $C_{15}H_{21}F_2N_2O_2$, 299.1566, found 299.1572 (M+H). **R.F.** 0.35 (40% ethyl acetate 60% cyclohexane).

48) 4-((4-(2,2-difluoroethyl)piperazin-1-yl)methyl)benzonitrile

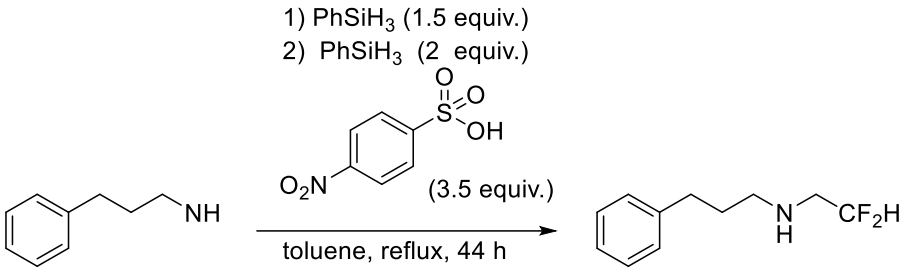


Produced according to the general procedure, 221 mg, 1.50 mmol of 4-cyanobenzoic acid used. Crude product purified by column chromatography 40% ethyl acetate 60% cyclohexane.

Isolated yield: 19% **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.58 (d, *J* = 8.0 Hz, 1H), 7.43 (d, *J* = 7.9 Hz, 1H), 5.85 (tt, *J* = 55.9, 4.4 Hz, 1H), 3.53 (s, 1H), 2.72 (td, *J* = 15.1, 4.4 Hz, 1H), 2.62 – 2.56 (m, 2H), 2.49 – 2.43 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 144.1, 132.1, 129.5, 118.9, 115.8 (t, *J* = 241.0 Hz), 110.9, 62.3, 60.1 (t, *J* = 24.7 Hz), 53.8, 52.9. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -118.31. **Ir (cm^{-1}):** 2926, 2813, 1632, 1454, 1430, 1293, 1252, 1130, 1043, 1006, 941, 886, 829, 741, 698, 612. **HRMS (ESI+):**, Exact mass calcd for: $C_{14}H_{18}N_3F_2$ 266.1464, found 266.1468 (M+H). **R.F.** 0.46 (40% ethyl acetate 60% cyclohexane).

3.5.4 Secondary Difluoroethylamines

49) N-(2,2-difluoroethyl)-3-phenylpropan-1-amine



To a solution of difluoroacetic acid (144 mg, 1.5 mmol) in acetonitrile (2 mL) was added phenylsilane (162 mg, 1.5 mmol) and phenyl propylamine (135 mg, 1 mmol). After 24 hours additional phenylsilane (216 mg, 2 mmol) and 4-nitro benzenesulphonic acid (711 mg, 3.5 mmol). after a further 18 hours reaction was diluted with EtOAc (10 mL) and washed with HCl (10 mL of a 3 M aq. solution). Aqueous solution was made to pH 14 with NaOH (10 mL of a 6 M aq. solution), crude product extracted from aqueous solution with DCM (3x 10 mL) to give the product as an orange oil.

Isolated yield: 90% **^{1}H NMR** (400 MHz, $CDCl_3$) δ(400 MHz, $CDCl_3$) δ 7.36 – 7.15 (m, 5H), 5.85 (tt, *J* = 56.7, 4.1 Hz, 1H), 2.95 (td, *J* = 15.3, 4.3 Hz, 2H), 2.78 – 2.60 (m, 4H), 2.02 – 1.71 (m, 2H). **^{13}C NMR** (101 MHz, $CDCl_3$) δ 142.09, 125.81, 115.96 (t, *J* = 240.3 Hz), 51.55 (t, *J* = 24.4 Hz), 49.28, 35.24, 33.37. **^{19}F NMR** (376 MHz, $CDCl_3$) δ -121.89. **HRMS (ESI+):** Exact mass calcd for $C_{11}H_{16}F_2N$ 200.1245, found 200.1244 (M+H).

3.6 References

1 L. Settimo, K. Bellman and R. M. A. Knegtel, *Pharm. Res.*, 2014, **31**, 1082–1095.
2 K. P. Melnykov, O. Tavlui, A. Skreminskiy, Y. O. Kuchkovska and O. O. Grygorenko, *Chem. – A Eur. J.*, 2022, **28**, e202201601.
3 Y. Pan, *ACS Med. Chem. Lett.*, 2019, **10**, 1016–1019.
4 K. Müller, in *Progress in Fluorine Science*, eds. G. Haufe and F. R. B. T.-F. in L. S. P. Leroux Medicinal Diagnostics, and Agrochemicals, Academic Press, 2019, pp. 91–139.
5 CAS, Scifinder, https://scifinder-n.cas.org/, (accessed 14 January 2023).
6 K. G. Andrews, R. Faizova and R. M. Denton, *Nat. Commun.*, 2017, **8**, 15913.
7 A. van Oeveren, M. Motamedi, E. Martinborough, S. Zhao, Y. Shen, S. West, W. Chang, A. Kallel, K. B. Marschke, F. J. López, A. Negro-Vilar and L. Zhi, *Bioorg. Med. Chem. Lett.*, 2007, **17**, 1527–1531.
8 D. A. Griffith, D. M. Hargrove, T. S. Maurer, C. A. Blum, S. De Lombaert, J. K. Inthavongsay, L. E. Klade, C. M. Mack, C. R. Rose, M. J. Sanders and P. A. Carpino, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 2641–2645.
9 E. Bergman, *J. Org. Chem.*, 1958, **23**, 476–477.
10 J. M. Kang, S. O. Kwon, J. Ann, S. Lee, C. Kim, N. Do, J. J. Jeong, P. M. Blumberg, H. Ha, T. N. L.

Vu, S. Yoon, S. Choi, R. Frank-Foltyn, B. Lesch, G. Bahrenberg, H. Stockhausen, T. Christoph and J. Lee, *Bioorg. Med. Chem. Lett.*, 2021, **48**, 128266.
11 S. Lee, C. Kim, J. Ann, S. A. Thorat, E. Kim, J. Park, S. Choi, P. M. Blumberg, R. Frank-Foltyn, G. Bahrenberg, H. Stockhausen, T. Christoph and J. Lee, *Bioorg. Med. Chem. Lett.*, 2017, **27**, 4383–4388.
12 WO2022239587, 2022.
13 S.-M. Wang, J.-B. Han, C.-P. Zhang and H.-L. Qin, *Tetrahedron Lett.*, 2015, **56**, 6219–6222.
14 J. B. Dickey, E. B. Towne, M. S. Bloom, G. J. Taylor, D. J. Wallace, J. Sagal, M. A. McCall and D. G. Hedberg, *Ind. Eng. Chem.*, 1956, **48**, 209–213.
15 WO2013/180782, 2013, 48.
16 B. P. Fauber, G. de Leon Boenig, B. Burton, C. Eidenschenk, C. Everett, A. Gobbi, S. G. Hymowitz, A. R. Johnson, M. Liimatta, P. Lockey, M. Norman, W. Ouyang, O. René and H. Wong, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 6604–6609.
17 X. Wang, D. Huang, K.-H. Wang, J. Liu, W. Zong, J. Wang, Y. Su and Y. Hu, *Appl. Organomet. Chem.*, 2019, **33**, e4995.
18 A. T. Brusoe and J. F. Hartwig, *J. Am. Chem. Soc.*, 2015, **137**, 8460–8468.
19 E. Union, Per- and polyfluoroalkyl substances (PFAS), https://echa.europa.eu/hot-topics/perfluoroalkyl-chemicals-pfas, (accessed 28 November 2023).
20 N. D. Tyrrell, *Org. Process Res. Dev.*, 2023, **27**, 1422–1426.
21 N. V Goncharov, R. O. Jenkins and A. S. Radilov, *J. Appl. Toxicol.*, 2006, **26**, 148–161.
22 S. Dasgupta, K.-W. Huang and J. Wu, *Chem. Commun.*, 2012, **48**, 4821–4823.
23 J. Liang, J. R. Zbieg, R. A. Blake, J. H. Chang, S. Daly, A. G. DiPasquale, L. S. Friedman, T. Gelzleichter, M. Gill, J. M. Giltnane, S. Goodacre, J. Guan, S. J. Hartman, E. R. Ingalla, L. Kategaya, J. R. Kiefer, T. Kleinheinz, S. S. Labadie, T. Lai, J. Li, J. Liao, Z. Liu, V. Mody, N. McLean, C. Metcalfe, M. A. Nannini, J. Oeh, M. G. O'Rourke, D. F. Ortwine, Y. Ran, N. C. Ray, F. Roussel, A. Sambrone, D. Sampath, L. K. Schutt, M. Vinogradova, J. Wai, T. Wang, I. E. Wertz, J. R. White, S. K. Yeap, A. Young, B. Zhang, X. Zheng, W. Zhou, Y. Zhong and X. Wang, *J. Med. Chem.*, 2021, **64**, 11841–11856.
24 M. Epifanov, P. J. Foth, F. Gu, C. Barrillon, S. S. Kanani, C. S. Higman, J. E. Hein and G. M. Sammis, *J. Am. Chem. Soc.*, 2018, **140**, 16464–16468.
25 J. Mühle, J. Huang, R. F. Weiss, R. G. Prinn, B. R. Miller, P. K. Salameh, C. M. Harth, P. J. Fraser, L. W. Porter, B. R. Greally, S. O'Doherty and P. G. Simmonds, *J. Geophys. Res. Atmos.*, , DOI:https://doi.org/10.1029/2008JD011162.
26 M. P. Sulbaek Andersen, D. R. Blake, F. S. Rowland, M. D. Hurley and T. J. Wallington, *Environ. Sci. Technol.*, 2009, **43**, 1067–1070.
27 S. Murov, *Organic Chemistry: Structure and Function*, 2007, vol. 84.
28 W. Eschweiler, *Berichte der Dtsch. Chem. Gesellschaft*, 1905, **38**, 880–882.
29 H. T. Clarke, H. B. Gillespie and S. Z. Weisshaus, *J. Am. Chem. Soc.*, 1933, **55**, 4571–4587.
30 S. H. Pine and B. L. Sanchez, *J. Org. Chem.*, 1971, **36**, 829–832.
31 E. R. Alexander and R. B. Wildman, *J. Am. Chem. Soc.*, 1948, **70**, 1187–1189.
32 S.-C. Lee and S. B. Park, *Chem. Commun.*, 2007, 3714–3716.
33 Z. Wang, in *Comprehensive Organic Name Reactions and Reagents*, 2010, pp. 1945–1947.
34 E. J. Schwoegler and H. Adkins, *J. Am. Chem. Soc.*, 1939, **61**, 3499–3502.
35 C. F. Winans, *J. Am. Chem. Soc.*, 1939, **61**, 3566–3567.
36 E. R. Alexander and A. L. Misegades, *J. Am. Chem. Soc.*, 1948, **70**, 1315–1316.
37 Z. Wei, Y. Cheng, K. Zhou, Y. Zeng, E. Yao, Q. Li, Y. Liu and Y. Sun, *ChemSusChem*, 2021, **14**, 2308–2312.
38 Q. Hu, S. Jiang, Y. Wu, H. Xu, G. Li, Y. Zhou and J. Wang, *ChemSusChem*, 2022, **15**, e202200192.
39 H. Yuan, J.-P. Li, F. Su, Z. Yan, B. T. Kusema, S. Streiff, Y. Huang, M. Pera-Titus and F. Shi, *ACS Omega*, 2019, **4**, 2510–2516.
40 P. Yang, L. H. Lim, P. Chuanprasit, H. Hirao and J. (Steve) Zhou, *Angew. Chemie Int. Ed.*, 2016, **55**, 12083–12087.
41 D. Chandra, Y. Inoue, M. Sasase, M. Kitano, A. Bhaumik, K. Kamata, H. Hosono and M. Hara, *Chem. Sci.*, 2018, **9**, 5949–5956.
42 D. Koszelewski, I. Lavandera, D. Clay, G. M. Guebitz, D. Rozzell and W. Kroutil, *Angew. Chemie Int. Ed.*, 2008, **47**, 9337–9340.
43 C. V. Wilson and J. F. Stenberg, *Org. Synth.*, 1956, **36**, 48.

44 P.-Q. Huang and H. Geng, *Org. Chem. Front.*, 2015, **2**, 150–158.
45 C. W. Lee and H. M. Ko, *Asian J. Org. Chem.*, 2023, **12**, e202300098.
46 A. Chardon, E. Morisset, J. Rouden and J. Blanchet, *Synthesis (Stuttg).*, 2018, **50**, 984–997.
47 R. Apodaca and W. Xiao, *Org. Lett.*, 2001, **3**, 1745–1748.
48 I. Sorribes, K. Junge and M. Beller, *J. Am. Chem. Soc.*, 2014, **136**, 14314–14319.
49 W. Yao, L. He, D. Han and A. Zhong, *J. Org. Chem.*, 2019, **84**, 14627–14635.
50 K. G. Andrews, D. M. Summers, L. J. Donnelly and R. M. Denton, *Chem. Commun.*, 2016, **52**, 1855–1858.
51 E. L. Stoll, T. Tongue, K. G. Andrews, D. Valette, D. J. Hirst and R. M. Denton, *Chem. Sci.*, 2020, **11**, 9494–9500.
52 A. Stoneley, Univeristy of Nottingham, 2023.
53 C. Nolan, Univerisity of Nottingham, 2022.
54 A. Uner, Univeristy of Nottingham, 2023.
55 R. Ábrahámi, S. Fustero, F. Fülöp and L. Kiss, *Synlett*, 2018, **29**, 2066–2070.

Chapter Three

Prediction of Brønsted Acid pK_a Values by Machine Learning

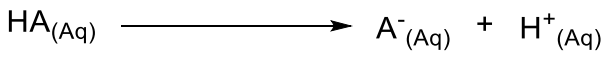
Brønsted acids hold significance to organic chemists, pK_a is used to quantify the potency of these compounds. pK_a as a property is dependent on the environment a compound exists in. Within the literature there are many gaps and missing data for pK_a in most organic solvents. Herein is discussed a workflow to develop a machine learning model to predict the pKa values for Brønsted acids in organic solvents.

4.1 Introduction

4.1.1 Acid-Base theory and pK_a

4.1.1.1 Fundamentals

Acid-base theory is of paramount importance to synthetic chemists, the concepts it describes have significant influence over many chemical processes. The oldest understanding of acidity was developed far back in antiquity, the first formal theory by Lavoisier in 1776 was centred around oxygen as the source of acidity.[1] This was later disproven by Davy in 1810 when the composition of the hydrohalic acids were discovered.[1] The second more refined acid theory was developed by Justus von Liebig in 1838 in which an acid was defined as a "substance containing a hydrogen that can be exchanged by a metal".[2] Liebig's theory would remain the prevailing acid theory until Arrhenius in 1884 developed a more comprehensive acid theory. Within this theory acid is defined as a compound that increases the concentration of H^+ ions in solution. This theory can be expressed in the following scheme:



Scheme 92: Arrhenius acidity

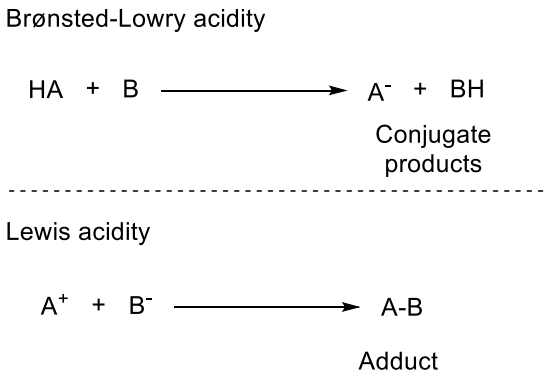
From **Scheme 92** it can be shown that for an Arrhenius acid in an aqueous solution will dissociate into an anion (A^-) and a proton (H^+). This theory did offer a foundational explanation for acidity it did not however give a complete picture. Efforts by other researchers over the last 150 years have worked to develop a more complete idea of acidity.

Alongside acids, theories explaining basicity were developed in parallel, the oldest defined in 1754 by Guillaume-François Rouelle simply defined a base as a " substance that reacts with an acid to form a solid.[3] Arrhenius later expanded the idea of bases and defined them as a "

substance that dissociates to form a hydroxide in solution".[4] Later theories unify concepts of both acids and bases.

4.1.1.2 Brønsted Acidity vs Lewis Acidity

Brønsted-Lowry acidity, often shortened to Brønsted acidity was developed in 1923 as a further development of Arrhenius' concepts on acidity.[5,6] This theory developed independently by Brønsted and Lowry focused on the concept of proton donation and acceptance (**Scheme 93**). A Brønsted acid can be defined as a proton donor, conversely a Brønsted base is defined as a proton acceptor.[7] Brønsted acids once they donate the proton form a conjugate base and a Brønsted base forms a conjugate acid after acceptance of a proton. Brønsted-Lowry acidity while not a complete understanding of acidity is effective at explaining the behaviour of many organic acids.



Scheme 93: Brønsted and Lewis acidity

Developed at the same time as Brønsted-lowry's theory, Lewis acids and bases are an alternate broader concept defined by the acceptance/donation of an electron pair (**Scheme 93**).[8] A Lewis acid serves as the electron acceptor, while a Lewis base acts as the electron donor. In comparison to Brønsted acidity as Lewis acidity is not dependant on any element it can be used to explain how many inorganic compounds can behave as an acid. Simple examples of a Lewis acid include boric acid, and aluminium trichloride.

For the purposes of this chapter further discussion will be using Brønsted-Lowry acid-base theory.

4.1.1.3 pK_a

The acid dissociation equilibrium constant or pK_a is a quantitative measure of the strength of an acid in solution. It is a key measure of properties for compounds with variable protonation.[9] pK_a is of paramount importance for drug discovery as it partially informs, solubility, membrane permeability and protein binding affinity.[10] This measure is not truly a constant as it highly dependent on the environment an acidic compound exists in.[11] Environmental factors such as temperature, ionic strength and solvent dielectric constant have significant impact on the pK_a value of a given compound.[12]

Each solvent has its own range of accessible/measurable pK_a value ranges, this range is largely dependent on the autoprotolysis constant of the solvent.[13] for some solvents like water or methanol protonation is reversible and many acids will be in a disassociated state in these solvents. But other solvents like DCE or MeCN protonation can be irreversible as it decomposes the solvent.[14] The pKa limit for these solvents is determined by solvent stability. For low polarity solvents autoprotolysis becomes irrelevant. This can be seen with purely alkyl solvents like hexane or pentane where any autoprotolysis is likely due to traces of water.[12]

Over the last century the process of acquiring pKa values has received extensive attention and refinement to produce accurate data.[11] The oldest and simplest method first developed in 1908 is potentiometric titration. This method is a direct comparison of a known quantity of reagent to an analyte, two electrodes placed in the analyte measure the change in electrical potential as the reagent is added these measurements are compared to a reference material.[15] More modern versions of this experiment make use of an automated pH probe to make the measurement and the resulting pH measurement can be converted to pK_a. The advantages of this method are the low costs/ease of access to the necessary equipment and

the large body of literature on method valadation.[16,17] The disadvantage of potentiometric titration is the difficulty in precise measurement of the titration curve, though modern digital titration systems have largely overcome this limitation.[18]

Another common method in determining pK_a is UV/vis spectroscopy, first developed in 1925 by Holmes and Snyder in the measurement of dye compounds.[19] Central to this is the presence of a chromophore in an indicator compound.[20] The difference in spectra between the ionised and non-ionised compound can be measured. A plot between the ratio of ionised and non-ionised indicator vs pH can be used to determine pK_a. In a similar fashion to the potentiometric method the UV/vis spectroscopy method can be readily automated.[21] The main advantage of this method is that simple experimental set up and common equipment. The disadvantage of this method is the need for prior knowledge of experimental data such as absorption coefficient of both the neutral and ionised compound.

A more modern approach is through the use of HPLC as it was discovered that elution times could be changed by the adjustment of pH.[22] A general formula was developed by Horváth for acid components in reverse phase HPLC:[23]

$$k = (1 - \alpha) \times k_0 \times k_{-1}$$

Equation 3: equation for pKa determination by HPLC

From **Equation 3**, K_0 and k_{-1} represent the capacity factors for the non-ionic and ionic component and α is the degree of dissociation to give "k" as the capacity factor for the system. A plot of k vs pH gives a curve with an inflection point of α=0.5 where pH=pK_a. This method is advantageous as it can assess pKa for a wide verity of samples including impure samples, however α values from 0 to unity cannot always be ascertained.

4.1.2 Machine Learning for Organic Chemistry

Machine learning (ML) is a computational process commonly associated with artificial intelligence that has evolved from statistical modelling.[24] These models can, to varying degrees mimic human cogitation and these models can be used to either predict or solve problems with little to no input from a human. In broad terms ML models can be broken down into three categories, supervised learning, unsupervised learning, and reinforcement learning. supervised learning uses labelled datasets that "supervise" the model during training, these labels can be used to measure accuracy. Unsupervised learning uses unlabelled data to uncover hidden patters in the dataset without human intervention. Reinforcement learning uses no predefined data and relies on environmental interaction. For the purposes of this discussion, we will focus only on supervised learning methods.

Within organic chemistry, machine learning is used to make predictions linked to given chemical phenomena. The entry of machine learning into chemistry has been significant and will only increase as development continues,

4.1.2.1 Model Types

Linear regression

The simplest method for machine learning is linear regression. This is the linear relationship between a single descriptor and the desired variable. This method uses the standard equation for a linear relationship.

$$y = mx + c$$

Equation 4: equation for a linear relationship

This method works well for simple systems but is incompatible for complex multidimensional problems typically found in chemistry. To overcome this multilinear regression (MLR) expands **Equation 4** to factor in additional variables arrayed in "hyperplanes" of "*x*" to give **Equation**

5.[25] For a two-dimensional problem there would be two hyperplanes x_1 and x_2. The equation for this regression would be:

$$y = ax_1 + bx_2 + c$$

Equation 5: multilinear regression for a 2-dimensional problem

Coefficients *a* and *b* represent the overall positive or negative contribution each hyperplane gives to the prediction. The fitted coefficients make the workings of the model transparent and easy to interpret this makes MLR a popular choice for simple problem such as prediction of enantioselectivity.[26] The downside of this method is that it cannot handle codependent variables or large data sets very well and is often outperformed by more complex models.

Support Vector Machines

Support vector machines or SVM is a ML method used to predict the classification of a desired phenomenon. This works by creating a hyperplane through the dataset, this hyperplane is created by forming two margins called support vector classifiers. Support vectors arise from classification some the data that forms a soft margin as shown in **Figure 29**. Between the classifier line and the support vector is a variable margin of error the allows for some data to be misclassified. This may seem counter intuitive however it often leads to more accurate predictions.[27]

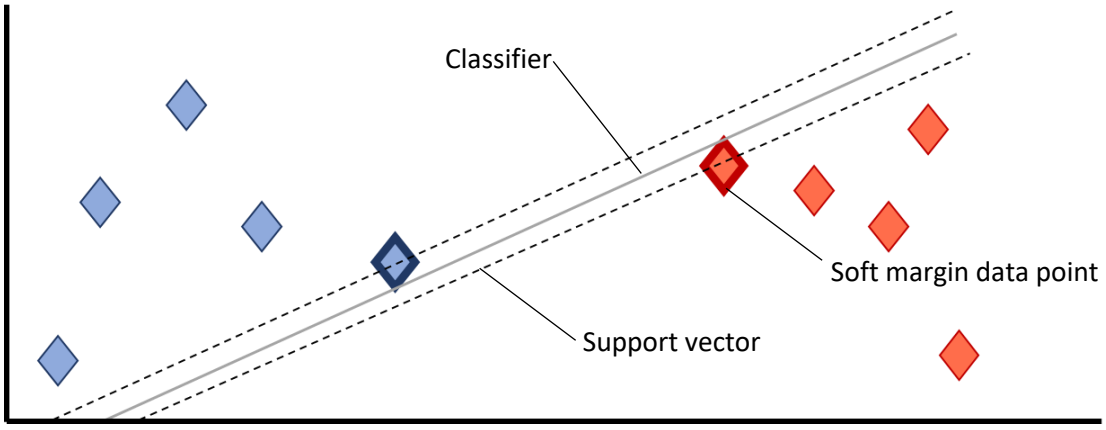


Figure 29: support vector classifier for linear dataset

The “hardness” of the margins can vary depending on how easily the data can be separated, a completely hard margin allows for no misclassification but is more sensitive to bias. Conversely if the margins are too wide the model cannot make predictions. The main classified line exists at one dimension lower than the dataset. For the general example shown in **Figure 29** which is in two-dimensional space the classifier is a one-dimensional line, this can scale up to n-dimensions with ease.[28]

For nonlinear data a kernel function is used to convert the data so a linear relationship can be found.[29] Kernels are mathematical functions that can compare different instances to each other in a higher dimension than should be possible for the data.

The advantages of this method is that it is very good at handling small to medium datasets, these datasets are commonly found in chemistry.[30] The main disadvantage of SVM is that the computational time requires scales exponentially with the size of the dataset, for large datasets this can lead to either unrealistic calculation times or the need for high performance computer time.[31]

Decision Trees

Decision trees are a ML method that uses a series of binary choices classify and make choices that sort inputs in groups based on similarity.[32] The tree is made of series of branching nodes and within each node is a binary question that gives an outcome. There are three types of node, the starting or root node, the internal nodes which breaks down and classify the data, and finally the leaf node that give the outputs and form the prediction.[33] The questions used in each node can be arbitrary and complex so long as the answers can be easily computed. The tree is grown by the addition of questions incrementally from a given starting point. An ideal question can sort a heterogeneous data set in to two homogeneous outputs.

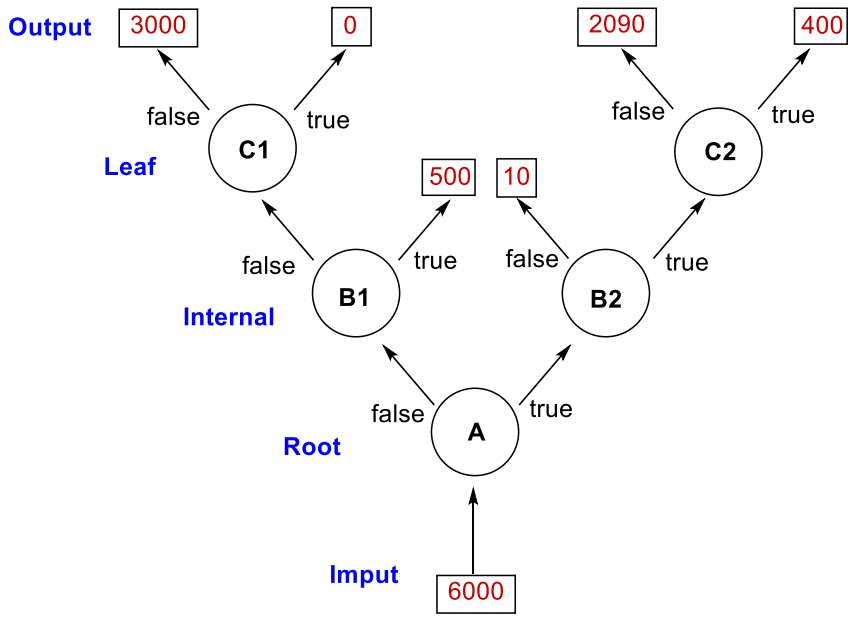


Figure 30: layout of a simple classification decision tree of a data set of 6000 entities

As shown in **Figure 30** the root node of the tree comes from the descriptor that gives the greatest variability to the target variable. this is determined by calculating the probability of a classification being incorrect.[34] This calculation is also used to form both the internal nodes and the leaf nodes sequentially.

Decision trees are advantageous as they are easy to interpret owing to the transparency of the questions used. Each question used, and output generated within the tree can be identified and interpreted.[32] Single decision trees are not typically used within chemical problems as the error is too high, but they serve as the basis for a much more powerful model type.[28]

Random Forest

A variation of decision trees is random forest this ensemble method uses many decision trees and uses an aggregated average of the output. each tree within the forest is develop as described above. The "randomness" of the model arises from how the forest is developed; a subset of the descriptors is used to grow the trees. The trees are developed with a varying level of predictive power, these steps are repeated hundreds of times to form the forest of

predictions.[35] This process will be repeated in random number of instances of the forest, the predictions from each tree in each instance of the forest will be aggregated. The aggregated result can be averaged using many different methods, averaging examples include voted choice or simple averaging.[28] The averaged prediction is often the best due to the "wisdom of crowds" though this can vary depending on the data set.[35] This model is easily interpretable as the weighting of the different descriptors can give insight into both the model operation and the nature of the problem being explored.

This method is advantageous as it can for several reasons. Firstly, it can mitigate overfitting of a model through the averaging the output of multiple decision trees. The second reason is the ease of handling for large descriptor sets. Lastly the model can estimate missing data as a way to maintain accuracy.[36] These reasons make random forest a popular method for ML research into chemical problems.[37,38]

Neural Networks

Neural networks are a ML model type that tries to mimic the structure of neural structures found in an organic brain.[39] In practical terms a neural network is composed of a series of decision layers connected by switching functions as shown in **Figure 31**. Each layer is composed of nodes or neurons, each neuron is connected to every node in the layer in front of it.

There are three types of decision layer, the first is the input layer, this layer is composed of a neuron for each descriptor. The second is the hidden layer, these layers are complex and not easily described, though there can be multiple of these layers. The third is the output layer that takes the information from the hidden layer and normalises it to give the full output.

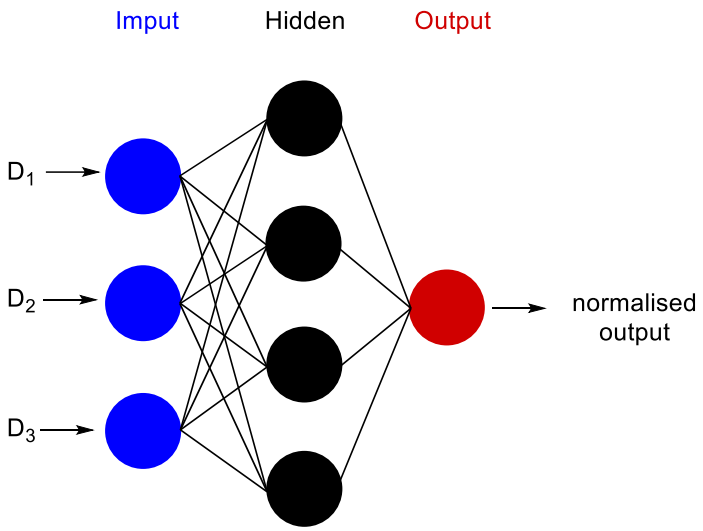


Figure 31: layout of a simple neural network

The nature of the hidden layer makes optimisation of the number of layers and nodes difficult though there are many examples of working procedures and model architecture in the literature.[40–42] The advantage of this method is its ability to process many complex non-linear problems. The key disadvantage is the high computational expense of neural networks, this can require specialist computer hardware to attain convenient completion times.[28] In addition, there is a large diversity of variants to neural networks that offer their own strengths and weaknesses. This can be overwhelming for researchers entering the field but there is literature to aid with startup.[43–45]

4.1.2.2 Typical ML Workflow

Development or workflow (**Figure 32**) of a machine learning model starts with the creation of a data set relevant to the chemical problem/phenomena being explored. When creating this data set the integrity of the data set must be carefully studied as the quality and diversity of the data will set an upper limit for the model performance.[24] The most prevalent and disruptive influence in the data set is bias both in data set selection and innate bias in the data.[46] Owing to data reporting in synthetic chemistry many data sets can be highly biased, lack of negative data in particular is the single largest bias in chemical data.[47] Dataset curation

is paramount in insuring dataset integrity and limiting bias, methods for curation are highly varied and may be unique to the chemical problem being explored.

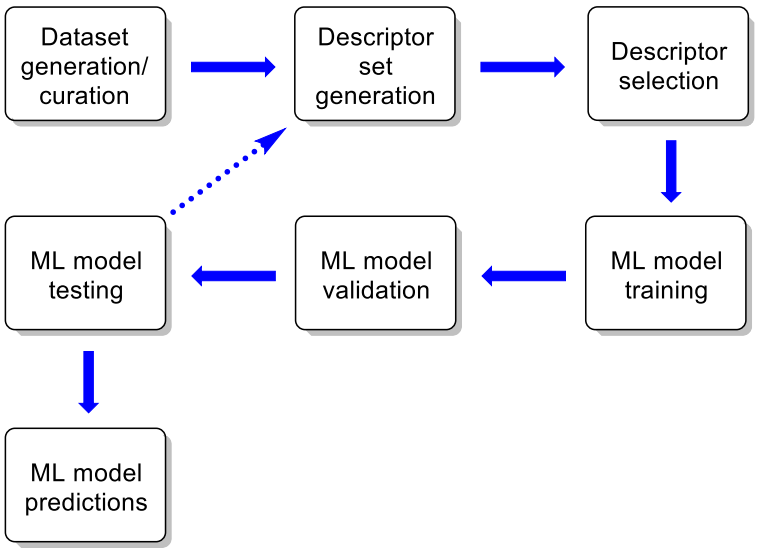


Figure 32: general overview of a ML model development workflow

With the dataset formed, the next step in model development is feature/descriptor generation (for simplicity they will be referred to as descriptors from this point). Descriptors are representations of chemical information in a form understandable mathematically so they can be used by ML models.[48] There are many different approaches to generate descriptor sets often in the form of dedicated software packages that allow for description of chemical structures often in the form of a SMILES string or InChi key.[49,50] SMILES strings and InChi keys are linear representations of chemical structures that are machine readable.[51,52]

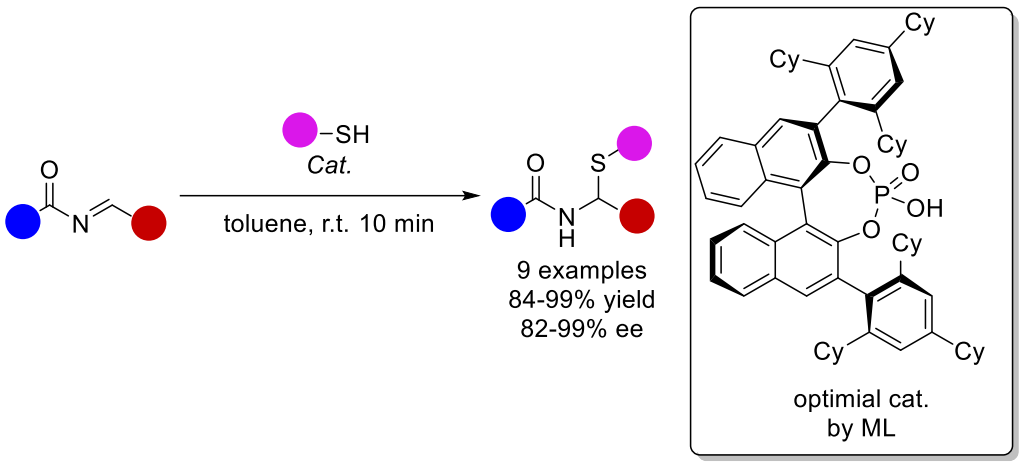
With a full set of descriptors, it can be useful to find out which descriptors have the most impact on the model performance.[53] This can be done through descriptor selection methods, there are two ways to perform feature selection. The first is through manual removal of irrelevant descriptors, this is largely done for descriptors that have no information in them i.e. have a total value of 0.[28] The second is through calculation of a variance threshold. This is where the variance of a given descriptor is calculated, if the variance does not exceed the preset threshold, then it is discounted.[54] This is often an effective way to reduce the side of a

descriptor set with compromising the relevant patterns it contains. The rationale behind descriptor selection is that if a ML model is subjected to too many descriptors it can increase calculation times to the point of impracticality.[28] Also, excessive descriptors can promote overfitting of the model.[55] Overfitting is a general term for performance loss in a trained model that prevents the model from being used for tasks outside its initial training set.[56]

The next step in a typical ML workflow is the training and validation of a model. the training of the model can be performed on many standard computer packages using the model types describe above in section 4.1.2.1.[28] In order to confirm whether the model performs adequately and issues like over fitting are minimised the model must be validated, the most common approach is by cross validation. There are two types of cross validation commonly used. The first is *k*-fold cross validation this is where the data is split up into *k* number of subsets.[57] from this number of subsets *k*-1 number of subsets is used for training of the model while the remaining subset is used for testing. This process is repeated *k* number of times to give the validated model.[57] The second commonly used validation model is leave one out cross validation, this is broadly similar to the *k*-fold method. In this case *k* equals the total number of entries within the data set.[57] leave one out cross validation is more robust than *k* fold however is significantly more computationally expensive.[57]

The final stage of a typical workflow is to test the model and generate predicted data. Test datasets can be generated in many different ways. The most common approach is through the holdout method, this is where a percentage of the total data set is partitioned and not used in training and validation.[58] This partitioned data is typically sourced by random sampling to minimise bias. The test set is used to estimate the overall accuracy of the model.[58]

An example of this workflow applied to a chemical problem can be seen by Denmark *et al.*. in this paper the authors describe the development of a ML model to predict novel catalysts for chiral phosphoric acid–catalysed thiol additions to N-acylimines (**Scheme 94**).[59]



Scheme 94: chiral phosphoric acid–catalysed thiol additions to N-acylimines by Denmark *et al.*

For this model the authors developed a library of 1075 reactions (both experimental and theoretical) split into 600 for the training set and 475 for the test set. 16,384 descriptors per reaction was generated and this was reduced using a variance threshold method to reduce this to a more manageable level of 2000 descriptors per reaction. The authors trained the model on a number of methods including SVM, neural networks and random forest each using 5-fold cross validation. SVM was shown to be the most precise method, and this was applied to the test data set. The output of the model was the calculation of *ee* values base on the use of different chiral phosphoric acids in a model reaction. The model outputs were then screened in physical reactions and the optimal catalyst used in a short substrate scope (**Scheme 94**). From this scope the theoretical and experimental *ee* values were compared and showed good agreement.

4.1.3 Machine Learning and pKa Prediction

pK_a is receiving increasing attention from researchers interested in using ML to predict pK_a values for various compounds. There have been several attempts at producing a practical ML model for predicting pKa values for organic Brønsted acids however all attempts face the same limitations. In a review by Hou et al. describes the fundamental challenges in developing a model for pKa prediction.[60] The largest challenge is data scarcity, the total number of well-

defined pKa values is less than 40,000 in total and if this total is separated into basic and acidic groups that total is reduced by half.[60] The other major challenge is the complexity of factors that affect ionisation of an acidic compound. Factors like conformational flexibility, tautomerism, charge transfer systems all make acidic systems subject to a level of multidimensionality that is difficult for ML models to compute.[60]

Several examples of ML have been developed over the last few years; a typical example can be seen by Jang *et al.* in 2021. The authors report ML workflow in the prediction of pK_a values using DFT to expand the starting data set.[61] The authors intended to predict the pKa values of acids relevant to polymer electric membranes. The data set was formed of 80 acids including phosphoric, sulfonic, and carboxylic acid examples. The workflow described differs little from the general formula (**Figure 32**), the notable variation is the use of DFT calculate additional pKa values to build the test set of 37 novel acids. The authors compared the performance of neural networks, Kernel Ridge Regression and process regression models, the Kernel Ridge Regression model proved the most precise with a mean square error (MSE) of 0.6 pK_a units. This model was carried forward and was used for the test set. Interpretation of the model uncovered that number of acidic hydrogens and oxidation level of the acid was critical to the model performance. the authors note that the key limitation of this model is its ability to predict for acids whose pK_a is less than 0.

4.2 Aims and Objectives

Aims

The aim of this work was to develop a machine learning model that could predict the pK_a values for Brønsted acids in varying organic solvents. Within this a data set of sufficient size covering a diverse area of chemical space will be required. Novelty arises from the application to sustainable solvents which have so far been left out of prior work.

Objectives

The objectives for this work were:

- Generation and curation a dataset of pK_a values for relevant compounds.
- Development of a ML model to predict PK_a values for a test set of acids and solvents.
- Analysis and interpretation of the model outputs.
- Validation of predictions by comparison with experimental data

4.3 Data set generation and curation

The training data set was generated from an online database of pKa values, from this database pKa values for sulfonic acids, benzoic acid and phenyl acetic acids were selected. There were two iterations of the data set based on the size of the number of acids used. The first and smallest data set contained a total of 312 pK_a values derived from 61 acids. This first data set was intended primarily as a workflow demonstration and validation set. The second was composed of 836 pK_a values from 160 acids this would be the main working set for the model. In the generation of this data set the information taken from the iBond data base was curated to exclude estimated values as they proved unreliable and increased errors in the model. In the case of acids with multiple reported pKa values only the value from the most recent source was used as a way to reduce conflicting errors that can arise from multiple values. While this

can introduce some bias into the data set, all data curation methods carry the risk of adding both implicit and unconscious bias.

Alongside the pK_a values in a given solvent, the solvent dielectric constant and the solvent dipole moment, these data points serve as environmental component of the descriptor set.

From each acid was generated two sets of descriptors, the first set were produced on program called Dragon. The descriptors generated from this program are largely mathematical and highly abstract, but they are based on different ways of calculating structure activity relationships. While this can give the model insights into structure activity relationship (SAR) for each acid it makes model interpretation more difficult as any key descriptor will contain very little practical chemical information. This descriptor set will be known a dragon descriptors from this point going forward.[62] As the dragon descriptor sets is so large there is a risk of long training times and higher computational expense a method to overcome this is "least absolute shrinkage and selection operator" (LASSO). LASSO is a regression model that in this case can sort and select the most important descriptors in the data set cutting out the superfluous data.[63] This method reduced the number of descriptors down from 1987 per acid down to 16 per acid.

The second descriptor set generated are signature molecular fragment descriptors these can be generated in several different programs.[64] The key to these this that they break down a molecule into smaller fragments expressed as a SMILES string and give a count as to how many of these fragment's a given molecule has.[65] As such they may contain significant amount chemical information and are much easier to interpret results. Moving forward this descriptor set will be called fragment descriptors.

A major challenge was uncovered during the generation of the training data set. This challenge was the large gaps that were uncovered in the raw data, solvents like NMA or DMA had very few pKa values. Even solvents like water and DMSO did not have complete data. This

demonstrates the need for a predictive model and the challenge in developing said predictive model. The challenge arises from the ML models needing large, unbiased data sets in order to make predictions, any gaps in the data set may increase errors in model predictions.

Alongside the training data set was the test data set, this data set is the one that will be used to make predicted pKa values from. It is primarily focused on compounds familiar to the model with solvents unfamiliar to the trained model. This test set contained eight acids and 3 solvents.

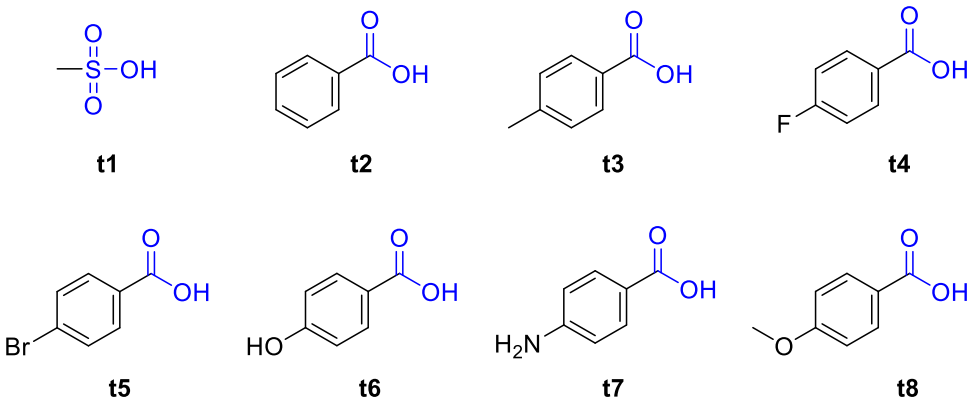


Figure 33: test set compounds

4.4 Model Development

4.4.1 1st Generation Model

The first generation of the model was developed as a proof of concept/initial proof of workflow in the prediction the pKa values of Bronsted acids using the initial smaller data set of 312 pKa values derived from 61 acids . To that end a screen of different model types was performed and compared against each other using root mean square error (RMSE) of each model.

What was uncovered in this initial screen was that methods like linear regression, SVM, and neural networks showed poor performance compared to a decision tree method. With this

initial screen completed the decision tree trained model was applied to test set 1 of compounds known to the model but with solvents previously unknown to the model.

Table 10: predictions with first generation trained model with the fragment descriptors

Entry	Compound	toluene	THF	DCM
1	t1	4.4300	4.4300	4.4300
2	t2	20.2675	20.2675	20.2675
3	t3	20.2675	20.2675	20.2675
4	t4	20.2675	20.2675	20.2675
5	t5	19.8800	19.8800	19.8800
6	t6	20.2675	20.2675	20.2675
7	t7	20.2675	20.2675	20.2675
8	t8	19.8800	19.8800	19.8800

Table 11: predictions with first generation trained model with the dragon descriptors

Entry	Compound	toluene	THF	DCM
1	t1	7.2000	7.2000	7.2000
2	t2	20.1233	20.1233	20.1233
3	t3	20.5200	20.5200	20.5200
4	t4	16.9700	16.9700	16.9700
5	t5	19.3700	19.3700	19.3700
6	t6	20.1233	20.1233	20.1233
7	t7	20.1233	20.1233	20.1233
8	t8	20.5200	20.5200	20.5200

This first set of predictions **(Table 10** and **Table 11**) showed apparently reasonable output for these acids, however several issues can be clearly identified. The first was no difference in pK_a values between solvents this is clearly shown across **Table 10** which all have the same pK_a values predicted. The second issue was identical predicted values for compounds with similar

functional groups this can be seen in **Table 11** entries 2, 3, 4, 6, and 7 which all have the same pK_a values. This problem was consistent for both sets of descriptors and represented a fundamental issue with the workflow. The underlying cause is likely due to the limitations in the data sets coverage of chemical space reducing the model's ability to differentiate between functional groups thereby reducing model effectiveness. As this was a proof of workflow experiment this represents encouraging initial results, through significant development is needed to attain more accurate predictions.

4.4.2 2nd Generation Model

To combat the repetitions in predictions several methods to redevelop and refine the model were tried. This eventually led to a second generation of the model. The first change was to create sets of descriptors for each of the solvents used in the training set, descriptors for the test set also featured a solvent descriptor set based of the dragon descriptors package for each of the test solvents. This unfortunately had no positive impact on model performance and was in fact deleterious to the model performance. All models produce with full solvent descriptors showed significantly worse errors than the first model. This may have been due shear number of solvent descriptors having a negative impact and producing overfitting of the model. Efforts to reduce the number of descriptors to just those most relevant to the model also failed as methods such as LASSO would completely remove the full solvent descriptor set, this indicates that these descriptors had no real influence on the model. This was a less than desirable outcome as this mean that methods to describe the environmental effects were of little importance to the model. As such a simpler method to describe the solvent was employed.

Table 12: comparison of error performance for model with and without "one hot" descriptors

Descriptor set	method	Leaf size	RMSE	R^2	MSE	MAE
fragment descriptors	Decision tree	8	1.5855	0.94	2.5138	0.9619
fragment + one hot	Decision tree	8	1.6221	0.94	2.6313	0.9956
Fragment + one hot	Bagged trees (ensemble)	4	1.5041	0.85	2.2623	0.9223

Another approach taken involved using the solvent identity as a constant descriptor, for example if a particular pKa value for a compound was taken in methanol, its "methanol" descriptor would be set to "1" and all remaining solvent descriptors would be set to "0", this is a simplified version of the "one hot" descriptor method.[66] The output from this approach (**Table 12**) using default parameters was interesting as not only was there an improvement in the model error measurements but also an entirely different model proved to be more precise. Previously a decision tree method was shown to be the most precise however with the "one hot" descriptor method was a random forest method was shown to be the most precise. When compared to the decision tree method that had in the prior generation been the most precise there was a significant improvement in RMSE, MSE, and MAE values. This was an interesting result and offered the 1st example in this work of a viable ensemble method. The trained model with the one hot descriptor set added was then tested using the test data set. The results were much poorer than expected with no improvement on the 1st gen model.

4.4.3 3rd Generation Model

The issues surrounding the model's ability to predict in different solvents has proved challenging thus far and after the attempts previously described showed that the principal

data set was too small and to account for this the data set was expanded to cover a much larger area of chemical space. One issue that was apparent in curating this new expanded data is that there are significant gaps in the known pK_a values. Not one of the compounds in the data set had a complete set of pK_a values for all test set solvents, and in the case of newer more sustainable solvents there is so little data available. The distribution of the pKa values is also uneven so some solvents are better represented in the data set than others.

This successful model was further optimised by changing a key hyperparameter- the leaf size, that means the number of observations in the terminal decision in the model. By progressively reducing the leaf size showed a small but noticeable improvement in model improvement performance. Examining the difference in true and predicted values in **Figure 34**, aside from a several outliers the model can clearly show regions such as sulfonic acids in water (**Figure 34**, region 4) or benzoic acids in aprotic solvents (**Figure 34**, region 1).

This iteration of the model shown in **Figure 34** shows a great number of regions that can be attributed to the different solvent conditions in which the pK_a values were obtained. Region 1 shows benzoic acids in polar aprotic sol

vents, this region also breaks down into two sub-regions, the higher sub- region is for data in acetonitrile and dichloroethane, and the lower sub-region is data in acetone and nitromethane. Region 2 shows data from benzoic acids in protic solvents and DMSO. Region three is the overlap between benzoic acid in water and sulfonic acids in organic solvents. Region four shows sulfonic acids in water and should be noted that for the extreme low pK_a values the model struggles to train with any predictions, this is likely due to have few data points that are available to train the model with indicating a limitation that will prove difficult to overcome. Several outliers can also be seen outside the defined regions, the exact cause of these have been difficult to determine.

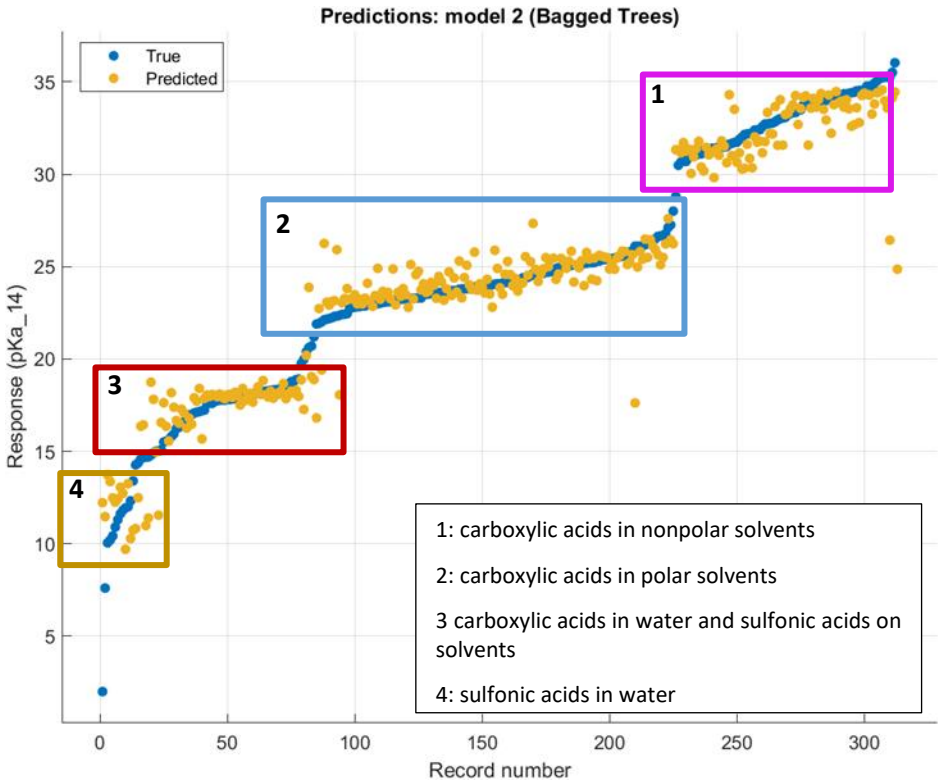


Figure 34: experimental vs predicted values for 3rd gen working model

With the refined and optimised model, predictions were attempted but with a change in the test solvents, the second test solvent set exchanged THF for DMC, a greener alternative solvent used in both chapters one and two. The reason behind this is that there is no pKa data for sustainable alternative solvents as such it would be advantageous to develop predicted values for such solvents.

Table 13: pKa predictions made from refined and optimised 3rd generation model.

Entry	Acid	DCM	Toluene	DMC
1	t1	4.2542	4.2542	4.2542
2	t2	15.8636	15.8636	15.8636
3	t3	16.0569	16.0569	16.0569
4	t4	14.4227	14.4227	14.4227
5	t5	14.1318	14.1318	14.1318
6	t6	15.0313	15.0313	15.0313
7	t7	14.8122	14.8122	14.8122
8	t8	14.7228	14.7228	14.7228

This model results (**Table 13**) has shown some improvement in that it has produced unique pKa values for each individual acid however the duplication across solvent environment persists. The limitations in the data set size may be the cause of this duplication and necessitates expanding the data set to cover more chemical space.

4.4.4 4th Generation Model

The fourth generation of the model exchanges the first smaller data set from the work shown in the 3rd generation model (section 4.4.3) for the larger data set described in section 4.3. this larger data set retains many of the characteristics of the previous set use in the 3rd generation model. Using the dragon descriptors, the model was trained on the full verity of model types and the training data produced gave uncovered a consequence of a larger dataset. In this it was uncovered that this larger data set made training the model much more computationally expensive. The average training times rose from 1-2 seconds for all previous models to two minutes for simple models to one hour for complex models like neural networks. For this generation of the model was a Boosted random forest with optimised parameters was shown

to be the best model, however the error metrics for this model is proved to be only slightly worse than for the third-generation model (**Table 14**).

Table 14: post validation error metrics for 4th generation model

method	RMSE	R^2	MSE	MAE
Boosted random forest	1.7801	0.91	3.1867	0.8364

This was disadvantageous as despite improving the area of chemical space covered by the trained model, it had not yielded a more precise model. When the trained model was used to predict pKa values for the test set once again it gave predictions that repeated across solvents with the notable exception of compound **t1** in DCM and DMC which gave an encouraging result (**Table 15** entry 1).

Table 15: predictions from 4th generation model

Entry	Compound	toluene	DMC	DCM
1	t1	19.6767	19.6767	18.1423
2	t2	33.843	33.843	33.843
3	t3	34.3185	34.3185	34.3185
4	t4	32.547	32.547	32.547
5	t5	32.5742	32.5742	32.5742
6	t6	34.2913	34.2913	34.2913
7	t7	32.2254	32.2254	32.2254
8	t8	35.2366	35.2366	35.2366

This is the first result that has differed across some solvents. While this is an encouraging result, significant work still remains to generate a more robust working model.

4.5 Conclusions

In this work it has been demonstrated that a machine learning model based on a random forest method can be used to predict the pKa values for a series of Brønsted acids. This method was developed through a series of model generations that were design to improve the model's overall performance while fully accounting for the chemical problem. Each generation did represent some iterative improvements in different respects, but issues remain. significant work remains to improve the model's performance relating to predicting the pKa values for the test set in different solvents.

4.6 Future Work

Looking to the future of this work should focus on the production of different descriptor sets outside the ones described previously. Once a suitable set is attained it should be passed though the ML workflow to assess its potential viability. Once a viable model is developed the results should be analysed and interpreted to gain understanding of how the model is achieving the results. In addition, any chemical information gain grom model interpretation should also be studied. Finaly the model should be applied to compounds completely unknown to the model to test the limits of its prediction capacity.

4.7 Experimental

Equipment

All model training and predictions were performed on a HP pavilion 15-eh0xxx laptop, with a AMD Ryzen 3 4300U 2.70 GHz (L2 Cache: 2 MB, L3 Cache: 4 MB), 8.00 GB DDR3 RAM, Microsoft Windows 11 64 bit operating system.

Dataset

PKa values were collected from the iBond database (http://ibond.nankai.edu.cn/), values were curated using the following rules.

1. Calculated or estimated data was excluded.
2. If multiple values for identical compounds in the same solvent were shown, raw data from the original source were examined and the data from the most recently published sourced was used.

The first full data set is composed of 22 sulfonic acids and 39 benzoic acids. Each acid has pKa data for up to 11 different solvents leading to 312 unique data entries. The curated data set was ordered by pKa value as this way shows the different solvents as pKa increases.

Chemical Class	Number of Molecules	Number of pKa values
Sulfonic acids	22	48
Benzoic acids	105	654
Aliphatic carboxylic acids	33	134
Total	160	836

Table 16: Data Set Composition

The second data set was significantly larger and contain a much more diverse array of compounds covering a much larger area of chemical space, including categories not previously used such as aliphatic carboxylic acids. This dataset was curated and arrayed in the same way as the first to ensure consistency when making comparisons.

Descriptors sets and descriptor selection.

Descriptors sets were generated from the data set using the dragon descriptor set package with 1987 descriptors per compound. For the signature fragment descriptors, 2240 descriptors per compound. These descriptor sets in conjunction with the solvent dipole moment and solvent dielectric constant were used to train the ML models.

Training of the models

Model was trained on Matlab (version 2022a through to version 2023b) using the regression learner tool using standard parameters. several methods were applied to the data set using fivefold cross validation to obtain a working model. The methods include:

- Linear regression
- Support vector machines (SVM)
- Decision tress
- Bagged decision trees (ensemble method)
- Neural networks

For the decision tree-based methods leaf size was used as a variable hyperparameter to optimise model performance. All reported model errors are post validation.

Predictions

Predictions using the trained model were performed on MATLAB version 2022a and version 2023b by making use of stock code from MATLAB's online library. Predictions were made with eight compounds extracted from the data set and given new solvent data for a solvent not in the training set.

Training set solvents:

water, acetone, acetonitrile, ethanol, methanol, pyridine, DCE, DMF, DMSO, DMA, NMA,

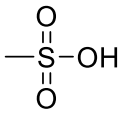
Test set one solvents:

Toluene, DCM, THF

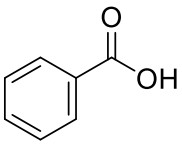
Test set two solvents:

toluene, DCM, DMC

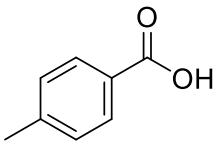
Test set compounds:



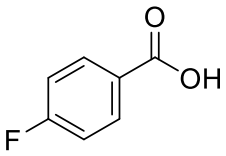
t1



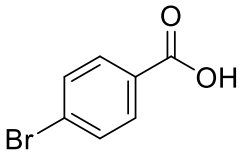
t2



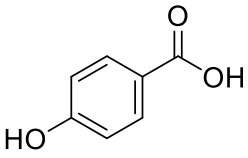
t3



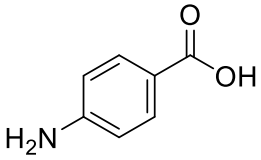
t4



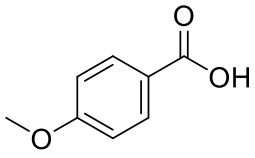
t5



t6



t7



t8

In the optimisation of the model several of the hyperparameter were changed below are the optimised parameters for the 3rd gen model.

- Random forest (bagged trees)
- Parameters changed- max leaf size, number of learners
- Default parameters- max leaf size: 8, number of learners 30
- Optimised parameters- max leaf size: 4, number of learners 30

In the optimisation of the model several of the hyperparameter were changed below are the optimised parameters for the 4th generation model.

- Method: boosted decision trees
- Minimum leaf size: 75
- Number of learners: 483
- Leaning rate: 0.91313
- Number of predictors to sample: 522

4.8 References

1 N. F. Hall, *J. Chem. Educ.*, 1940, **17**, 124.
2 G. Meyer, *Zeitschrift für Anorg. und Allg. Chemie*, 2008, **634**, 201–222.
3 W. B. Jensen, *J. Chem. Educ.*, 2006, **83**, 1130.
4 J. Clark, THEORIES OF ACIDS AND BASES, https://www.chemguide.co.uk/physical/acidbaseeqia/theories.html, (accessed 13 December 2023).
5 J. N. Brönsted, *Recl. des Trav. Chim. des Pays-Bas*, 1923, **42**, 718–728.
6 T. M. Lowry, *J. Soc. Chem. Ind.*, 1923, **42**, 43–47.
7 R. H. Morris, *Chem. Rev.*, 2016, **116**, 8588–8654.
8 G. N. Lewis, *Valence and the Structure of Atoms and Molecules*, Chemical Catalog Company, Incorporated, 1923.
9 E. Rossini, A. D. Bochevarov and E. W. Knapp, *ACS Omega*, 2018, **3**, 1653–1662.
10 D. T. Manallack, R. J. Prankerd, E. Yuriev, T. I. Oprea and D. K. Chalmers, *Chem. Soc. Rev.*, 2013, **42**, 485–496.
11 J. Reijenga, A. van Hoof, A. van Loon and B. Teunissen, *Anal. Chem. Insights*, 2013, **8**, ACI.S12304.
12 A. Kütt, S. Selberg, I. Kaljurand, S. Tshepelevitsh, A. Heering, A. Darnell, K. Kaupmees, M. Piirsalu and I. Leito, *Tetrahedron Lett.*, 2018, **59**, 3738–3748.
13 E. Paenurk, K. Kaupmees, D. Himmel, A. Kütt, I. Kaljurand, I. A. Koppel, I. Krossing and I. Leito, *Chem. Sci.*, 2017, **8**, 6964–6973.
14 I. Kaljurand, A. Kütt, L. Sooväli, T. Rodima, V. Mäemets, I. Leito and I. A. Koppel, *J. Org. Chem.*, 2005, **70**, 1019–1028.
15 H. G. Denham, *J. Chem. Soc. Trans.*, 1908, **93**, 41–63.
16 L. Z. Benet and J. E. Goyan, *J. Pharm. Sci.*, 1967, **56**, 665–680.
17 W. C. Gardiner and H. L. Sanders, *Ind. Eng. Chem. Anal. Ed.*, 1937, **9**, 274–278.
18 A. Avdeef, J. E. A. Comer and S. J. Thomson, *Anal. Chem.*, 1993, **65**, 42–49.
19 C. A. S. Bergström, K. Luthman and P. Artursson, *Eur. J. Pharm. Sci.*, 2004, **22**, 387–398.
20 L. A. Flexser, L. P. Hammett and A. Dingwall, *J. Am. Chem. Soc.*, 1935, **57**, 2103–2115.
21 K. Y. Tam and K. Takács-Novák, *Anal. Chim. Acta*, 2001, **434**, 157–167.
22 R. P. Singhal and W. E. Cohn, *Biochemistry*, 1973, **12**, 1532–1537.
23 C. Horváth, W. Melander and I. Molnar, *Anal. Chem.*, 1977, **49**, 142–154.
24 N. Artrith, K. T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain and A. Walsh, *Nat. Chem.*, 2021, **13**, 505–508.
25 D. J. Olive, ed. D. J. Olive, Springer International Publishing, Cham, 2017, pp. 17–83.
26 J. P. Reid and M. S. Sigman, *Nature*, 2019, **571**, 343–348.
27 J. Luts, F. Ojeda, R. Van de Plas, B. De Moor, S. Van Huffel and J. A. K. Suykens, *Anal. Chim. Acta*, 2010, **665**, 129–145.
28 J. D. Hirst, S. Boobier, J. Coughlan, J. Streets, P. L. Jacob, O. Pugh, E. Özcan and S. Woodward, *Artif. Intell. Chem.*, 2023, **1**, 100006.
29 A. J. Smola and B. Schölkopf, *Stat. Comput.*, 2004, **14**, 199–222.
30 K. Heikamp and J. Bajorath, *Expert Opin. Drug Discov.*, 2014, **9**, 93–104.
31 P. Birzhandi, K. T. Kim and H. Y. Youn, *Soft Comput.*, 2022, **26**, 3729–3742.
32 C. Kingsford and S. L. Salzberg, *Nat. Biotechnol.*, 2008, **26**, 1011–1013.
33 S. Yao, A. Kronenburg, A. Shamooni, O. T. Stein and W. Zhang, *Appl. Energy Combust. Sci.*, 2022, **11**, 100077.
34 L. Breiman, *Classification and regression trees*, Routledge, 2017.
35 L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32.
36 P. Probst, M. N. Wright and A. Boulesteix, *Wiley Interdiscip. Rev. data Min. Knowl. Discov.*, 2019, **9**, e1301.
37 S. M. Maley, D.-H. Kwon, N. Rollins, J. C. Stanley, O. L. Sydora, S. M. Bischof and D. H. Ess, *Chem. Sci.*, 2020, **11**, 9665–9674.
38 D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher and A. G. Doyle, *Science (80-.).*, 2018, **360**, 186–190.

39 A. K. Jain, J. Mao and K. M. Mohiuddin, *Computer (Long. Beach. Calif).*, 1996, **29**, 31–44.
40 S. Curteanu and H. Cartwright, *J. Chemom.*, 2011, **25**, 527–549.
41 M. Uzair and N. Jamil, in *2020 IEEE 23rd International Multitopic Conference (INMIC)*, 2020, pp. 1–6.
42 J. E. Dayhoff, *Neural network architectures: an introduction*, Van Nostrand Reinhold Co., 1990.
43 M. H. S. Segler, T. Kogej, C. Tyrchan and M. P. Waller, *ACS Cent. Sci.*, 2018, **4**, 120–131.
44 Z. Tu and C. W. Coley, *J. Chem. Inf. Model.*, 2022, **62**, 3503–3513.
45 Z. Wu, J. Wang, H. Du, D. Jiang, Y. Kang, D. Li, P. Pan, Y. Deng, D. Cao and C.-Y. Hsieh, *Nat. Commun.*, 2023, **14**, 2585.
46 X. Jia, A. Lynch, Y. Huang, M. Danielson, I. Lang'at, A. Milder, A. E. Ruby, H. Wang, S. A. Friedler, A. J. Norquist and J. Schrier, *Nature*, 2019, **573**, 251–255.
47 *Nature*, 2019, **573**, 164–164.
48 A. Seko, A. Togo and I. Tanaka, ed. I. Tanaka, Springer Singapore, Singapore, 2018, pp. 3–23.
49 F. Grisoni, D. Ballabio, R. Todeschini and V. Consonni, ed. O. Nicolotti, Springer New York, New York, NY, 2018, pp. 3–53.
50 M. Karelson, V. S. Lobanov and A. R. Katritzky, *Chem. Rev.*, 1996, **96**, 1027–1044.
51 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
52 S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi and I. Pletnev, *J. Cheminform.*, 2013, **5**, 1–9.
53 J. Miao and L. Niu, *Procedia Comput. Sci.*, 2016, **91**, 919–926.
54 I. Ponzoni, V. Sebastián-Pérez, C. Requena-Triguero, C. Roca, M. J. Martínez, F. Cravero, M. F. Díaz, J. A. Páez, R. G. Arrayás and J. Adrio, *Sci. Rep.*, 2017, **7**, 2403.
55 D. M. Hawkins, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 1–12.
56 X. Ying, in *Journal of physics: Conference series*, IOP Publishing, 2019, vol. 1168, p. 22022.
57 D. Berrar, 2019.
58 G. C. Cawley and N. L. C. Talbot, *J. Mach. Learn. Res.*, 2010, **11**, 2079–2107.
59 A. F. Zahrt, J. J. Henle, B. T. Rose, Y. Wang, W. T. Darrow and S. E. Denmark, *Science (80-.).*, 2019, **363**, eaau5631.
60 J. Wu, Y. Kang, P. Pan and T. Hou, *Drug Discov. Today*, 2022, **27**, 103372.
61 R. Lawler, Y.-H. Liu, N. Majaya, O. Allam, H. Ju, J. Y. Kim and S. S. Jang, *J. Phys. Chem. A*, 2021, **125**, 8712–8722.
62 T. S.r.l., Dragon, https://www.talete.mi.it/products/dragon_description.htm, (accessed 12 December 2023).
63 F. Santosa and W. W. Symes, *SIAM J. Sci. Stat. Comput.*, 1986, **7**, 1307–1330.
64 D. C. Weis and D. P. Visco, *Comput. Chem. Eng.*, 2010, **34**, 1018–1029.
65 D. P. Visco, R. S. Pophale, M. D. Rintoul and J.-L. Faulon, *J. Mol. Graph. Model.*, 2002, **20**, 429–438.
66 H. Kaneko, *ACS Omega*, 2023, **8**, 21781–21786.


University of
Nottingham
UK | CHINA | MALAYSIA