Multi-perspective analysis of the applicability of bio-based solvents as sustainable options for biocatalysis

Giulia Paggiola

Doctor of Philosophy

University of York

Chemistry

September 2016

Abstract

Biocatalysis and sustainable solvents are two increasingly popular instruments of the green synthetic toolkit. This thesis explores the use of bio-based solvents in reactions catalysed by *Candida Antarctica* lipase B, and discusses their *greenness* and wider applicability.

Firstly, to clarify the mechanism of solvent influence on enzyme catalysis, a systematic multi-variable approach is applied for the first time. The results challenge the established theory which relied solely on partition coefficient, logP, revealing that catalytic performance is governed instead by the solvent's ability to engage in hydrogen bonding. In addition, thermodynamic insight is given, with evidence of isokinetic effect and of genuine enthalpy-entropy compensation. Both effects were previously undocumented for such systems. Secondly, a strong case is made for citrus waste-derived solvents, D-limonene and *p*-cymene, as effective alternatives to typical petroleum-derived counterparts used in such processes. Their first use as media for biocatalysis is herein reported, in particular applied to the chemo-enzymatic synthesis of the pharmaceutical compound (S,S)-Reboxetine.

Finally, novel methods from socio-economic sciences are employed to reveal the main barriers faced by users in the uptake of green solvents. As a result of a survey of stakeholder perception, the most pressing priorities to be addressed appear 1) *cost*, 2) *lack of data*, and 3) *availability & supply*. The third of these aspects is later examined in detail as part of a dedicated case-study, which compares D-limonene potential supply against toluene demand. While complete global substitution appears unlikely, the greatest potential is shown for citrus-growing countries such as Brazil, Spain, India and South Africa which could feasibly implement the substitution as exemplary models of bio-economy.

Contents

List of Figures

List of Tables

Acknowledgements

My biggest thanks go to Professor James Clark, for supporting my interests and development in countless ways and occasions. I will always treasure the learnings, advice and motivation I received from him, the ever inspiring discussions about the world of Green Chemistry, and all the doors he opened for me to become part of it.

Being part of the CHEM21 project was an invaluable experience for me as a student, to be confronted with industrial points of view from such early stage. I am grateful to Dr Andy Hunt and Dr Rob McElroy for offering me to work on such a novel project and stimulating context, and for the great encouragement and guidance they provided me throughout my studies at York. Alongside them, I would like to thank all colleagues and friends at the Green Chemistry Centre of Excellence (GCCE) and within the CHEM21 project. In particular, pivotal has been the help of Dr Vitaliy Budarin, Dr James Sherwood, Dr Roger Howard, Dr Jonathan Moseley, Dr John Hayler, Dr Anthony Green, and that of the students who contributed to my work, Tilia Dahou, Ceren Ulger, Sytze van Stempvoort.

What I most enjoyed in my PhD has been the opportunity to engage with different disciplines and departments, and thus being able to explore a different dimension to my research. For this, I am very appreciative of the care and support I received from everyone at Stockholm Environment Institute at York (SEI-Y), Centre for Natural Agricultural Products (CNAP), York Environmental Sustainability Institute (YESI), Escola Brasileira de Quimica Verde (EBQV), and greenSTEMS.

Ultimately, it is to the love of my five biggest supporters that I dedicate this thesis and achievement: Giselle, Toni, Antonio, Lilian, and Aurelio.

Declaration

The work presented herein was conducted at the University of York (UK) between October 2012 and September 2015, during which time I was supervised by Professor James Clark, Dr Andrew Hunt and Dr C. Rob McElroy. I declare that this is original work of which I am the sole author. All additional contributions to the research by collaborators and co-workers are fully acknowledged below.

Contribution	Chapter	Collaborator	Institution
Telescoping the synthesis from Reboxetine diol to the zwitterion and additional solvent screenings	3	Tilia Dahou (laboratory work)	National Institute of Applied Science, Rouen (France)
Regioselectivity screening for Reboxetine acylation and additional solvent screenings	3 & 4	Ceren Ulger (laboratory work)	University of Ankara (Turkey)
Linearisation of reaction profiles	3 & 4	Vitalyi Budarin	GCCE, UoY (UK)
Planning of survey methodology	5	Roberto Rinaldi James Sherwood	Stockholm Environment Institute, SEI-York (UK) GCCE, UoY (UK)
Researching relevant sources for oil and citrus data and developing preliminary framework for the calculations	6	Sytze Van Stempvoort (research work)	University of Amsterdam (The Netherlands)
Screenings of *Pseudomonas Stutzeri* lipase	App. C	Jonathan Moseley Nolwenn Derrien	CatSci Ltd. (UK)

Materials, techniques and equipment support provided by:

- James Sherwood, training on regression analysis and *in silico* solvent property determination;
- Vitalyi Budarin, OriginLab, training on UV and IR spectroscopy, statistics and thermodynamics analyses;
- Rachael Castle and Tabitha Petchey, GC support;
- Heather Fisher, NMR support;
- Neil Bruce's group (CNAP), use of laboratories and HPLC support;
- Paul Elliot, safety training and general laboratory support;
- Allison Edmonds, Sophie Palmer, Katy Brooke, admin support;
- Roger Howard (Pfizer), laboratory samples, protocols and analytical material for Reboxetine project;
- Jonathan Moseley (CatSci), training on automated high-throughput screening kits and LCMS support.

Useful discussions and intellectual input provided by:

- Vitaliy Budarin, Duncan MacQuarrie, James Sherwood, Julen Bustamante (GCCE);
- John Hayler, Joe Adams (GSK);
- Gideon Grogan (CNAP, UoY);
- Roberto Rinaldi, Corrado Topi, Jose Vega Barbero, Murray Rudd (SEI-Y);
- Anthony Green, Sabine Flitsch, Nicholas Turner (University of Manchester);
- Roger Howard (Pfizer);
- Dominique Depre, Cyril Benhamin (Johnson & Johnson);
- Denis Prat (Sanofi);
- Tony Duncan (Circa Group);
- Peter Seidl (Universidade Federal do Rio de Janeiro).

The research leading to these results has received funding from the Innovative Medicines Initiative (www.imi.europa.eu) Joint Undertaking under grant agreement n115360, resources of which are composed of financial contribution from the European Unions Seventh Framework Programme (FP7/2007-2013) and EFPIA companies in kind contribution.

Some of the results discussed in this thesis have been published in academic journals and presented at conferences in the form of oral and poster presentations.

Publications and contribution by the author	Chapter
Tools and techniques for solvent selection: green solvent selection guides FP Byrne, S Jin, G Paggiola, THM Petchey, JH Clark, TJ Farmer, AJ Hunt, CR McElroy, J Sherwood *Sustainable Chemical Processes*, **2016**, 4 (1), 1-24	1
Biocatalysis in bio-derived solvents: an improved approach for medium optimisation G Paggiola, AJ Hunt, CR McElroy, J Sherwood, JH Clark *Green Chemistry*, **2014**, 16 (4), 2107-2110	2
What are the top priorities to accelerate the adoption of green solvents? G Paggiola, J Sherwood, R Rinaldi, AJ Hunt, CR McElroy, HF Sneddon, JH Clark **2016**, manuscript in preparation	5
Can bio-based chemicals meet demand? Global and regional case-study around citrus waste-derived limonene as a solvent for cleaning applications G Paggiola, S Van Stempvoort, J Bustamante, JM Vega Barbero, AJ Hunt, J Sherwood, JH Clark *Biofuels, Bioproducts & Biorefining*, **2016**, 10(6), 686-698	6
Sustainable solvents: Perspectives from research, business and international policy JH Clark, AJ Hunt, G Paggiola, J Sherwood *RSC Book Series*, **2016**, book in preparation	5 & 6
Pseudomonas Stutzeri Lipase (PSL) in bio-based solvents for the synthesis of model pharmaceutical amides G Paggiola, N Derrien, JD Moseley, A Green, AJ Hunt, CR McElroy, JH Clark **2016**, manuscript in preparation	App.C
Tunable solvents: Shades of green S Abou-Shehada, JH Clark, G Paggiola, J Sherwood *Chemical Engineering and Processing: Process Intensification*, **2016**, 99, 88-96	-

Presentations at conferences and external meetings
International Workshop on Green Initiatives 2013 – New Delhi, 2-3 Dec 2013 Poster title: Biocatalysis in bio-solvents: an improved approach for medium optimisation *Winner of poster presentation award*
Industrial Green Chemistry World 2013 - Mumbai, 4-5 Dec 2013 Poster title: Chemical Manufacturing Methods for the 21^{st} Century Pharmaceutical Industries
ACS Green Chemistry & Engineering 2014 – Washington D.C., 16-19 June 2014 Oral presentation title: Bio-based solvents in enzymatic catalysis
ACS Pharmaceutical Roundtable Reception 2014 – Washington D.C., 18 June 2014 Poster title: A multi-perspective on solvent substitution - Case-study on biocatalysis in the pharmaceutical industry
CHEM21 Annual General Meeting 2014 – Graz, 22-23 Sept 2014 Oral presentation title: Solvent substitution case-study: the synthesis of (S,S)-Reboxetine
Cafè Scientifique 2015 (Science Outreach) – York, 2 June 2015 Oral presentation title: The making of greener medicines
University of York KMS competition 2015 – York, 25 June 2015 Oral presentation title: Investigating applicability of bio-based solvents in enzyme-catalysed pharmaceutical processes *Finalist for KMS award*
CHEM21 Annual General Meeting 2015 – Paris, 7-8 Sept 2015 Poster title: Understanding solvent requirements across manufacturing operations
BIO Latin America 2015 – Rio de Janeiro, 14-16 Oct 2015 Poster title: Barriers & drivers to the adoption of green solvents in industry
Escola Brasileira Quimica Verde 2015 – Campinas, 19-20 Oct 2015 Oral presentation title: Application of green solvents in pharmaceutical industry *Finalist for presentation award*
Stockholm Environment Institute Annual Seminar – York, 11 Dec 2015 Poster title: Greener solvents in medicines production
RSC Brazil Day – London, 14 Dec 2015 Oral presentation title: Brazilian bio-resources and the 21^{st} century pharmaceutical industry

Chapter 1

Introduction

1.1 Green chemistry and the green economy

1.1.1 Sustainability

In today's society, sustainability expresses the critical need for a more ethical and balanced approach to human development that is socially and environmentally responsible while favouring economic development and long-term quality of life (Figure 1.1).[1] This expansion of the sustainability agenda to go beyond the merely environmental issues is captured by the term 'triple bottom line' coined by John Elkington in 1994.[2]

Figure 1.1: The three pillars of sustainable development.[2]

Embedding this philosophy into current economic models faces vast challenges and would require a dramatic paradigm shift. From an environmental and resource perspective, this would fundamentally require moving away from the reliance on fossil sources of fuel, energy and materials, towards a responsible use of renewable resources. This would in turn require establishing sustainable production systems, cleaner transport and industrial practices, and designing products that do not harm the environment or human life at any stage of their life-cycle.[3,4] Social sustainability was defined to include broadly social equity and justice, thus touching upon innumerous aspects such as education, health, urban development, employment, consumer behaviour community, amongst others.[5,6] Finally, the economic dimension of sustainability would most critically require a transition

from a linear economy to a circular one, characterised by the full reuse of goods, to the extent where possible, and the concept of waste as a resource.[7] In particular, strategic sustainability planning as part of business practices is promoted as a way to drive competitive advantage, cooperation, transparency and resilience.[8] In substance, the sustainability framework is all-encompassing, and engages all fields of human development bringing a strong drive for innovation. Ultimately, the pursuit of these targets would ideally generate a 'green economy', defined according to the United Nations Environment Programme (UNEP) as quoted below.[9]

"UNEP defines a green economy as one that results in improved human well-being and social equity, while significantly reducing environmental risks and ecological scarcities. In its simplest expression, a green economy is low-carbon, resource efficient, and socially inclusive. In a green economy, growth in income and employment are driven by public and private investments that reduce carbon emissions and pollution, enhance energy and resource efficiency, and prevent the loss of biodiversity and ecosystem services."[9]

Also noteworthy are the United Nations' Sustainable Development Goals (SDGs) devised in 2015 for the 2030 global sustainability agenda, encompassing 17 categories of targets across the three pillars of environment, society and economic wealth.[10]

1.1.2 The 12 Principles of Green Chemistry

'Green Chemistry' reflects the efforts undertaken by chemistry research, manufacturing and policy to advance sustainability within the field and contribute towards a green economy. The concept of Green Chemistry was first envisaged by Paul Anastas and John Warner in 1998, when they devised the '12 Principles of Green Chemistry' giving guidance for the application of sustainable-thinking to chemistry practice.[11,12]

1. **Prevention** *"It is better to prevent waste than to treat or clean up waste after it has been created."*

2. **Atom economy** *"Synthetic methods should be designed to maximize the incorporation of all materials used in the process into the final product."*

3. **Less Hazardous Chemical Syntheses** *"Wherever practicable, synthetic methods should be designed to use and generate substances that possess little or no toxicity to human health and the environment."*

4. **Designing Safer Chemicals** *"Chemical products should be designed to effect their desired function while minimizing their toxicity."*

5. **Safer Solvents and Auxiliaries** *"The use of auxiliary substances (e.g. solvents, separation agents, etc.) should be made unnecessary wherever possible and innocuous when used."*

6. **Design for Energy Efficiency** *"Energy requirements of chemical processes should be recognized for their environmental and economic impacts and should be minimized. If possible, synthetic methods should be conducted at ambient temperature and pressure."*

7. **Use of Renewable Feedstocks** *"A raw material or feedstock should be renewable rather than depleting whenever technically and economically practicable."*

8. **Reduce Derivatives** *"Unnecessary derivatisation (use of blocking groups, protection/deprotection, temporary modification of physical/chemical processes) should be minimized or avoided if possible, because such steps require additional reagents and can generate waste."*

9. **Catalysis** *"Catalytic reagents (as selective as possible) are superior to stoichiometric reagents."*

10. **Design for Degradation** *"Chemical products should be designed so that at the end of their function they break down into innocuous degradation products and do not persist in the environment."*

11. **Real-time analysis for Pollution Prevention** *"Analytical methodologies need to be further developed to allow for realtime, in-process monitoring and control prior to the formation of hazardous substances."*

12. **Inherently Safer Chemistry for Accident Prevention** *"Substances and the form of a substance used in a chemical process should be chosen to minimize the potential for chemical accidents, including releases, explosions, and fires."*

These principles have greatly influenced the way chemistry and chemical processes are being designed, carried out and monitored.[13]

1.2 Solvents

1.2.1 Definition

Solvents in chemistry are predominantly used as media for dissolving and mixing chemical reactants and reagents, extracting and purifying products of interest, cleaning vessels and equipment, amongst other uses.[14] An authoritative definition of solvents was published in 1999 by the European Commission on Solvents Emissions Directive, which describes a solvent as any volatile organic compound (VOC) - *i.e.* bearing vapour pressure higher than 0.01 kPa under the conditions of use - *"which is used alone or in combination with other agents, and without undergoing a chemical change, to dissolve raw materials, products or waste materials, or is used as a cleaning agent to dissolve contaminants, or as a dissolver, or as a dispersion medium, or as a viscosity adjuster, or as a surface tension adjuster, or a plasticiser, or as a preservative"*.[15]

The above lines describe solvents with a special focus on their use rather than by detailing physicochemical parameters to identify them. This itself is an indication of the extremely large range of properties represented within this class of chemicals, making it difficult to frame. Moreover, new solvents are constantly being designed and tailored for specific user criteria and for filling property gaps within the existing portfolio.

1.2.2 Solvent properties

No property alone is able to define what a solvent is without being simplistic or exclusive. Even the volatility-based definition, such as the VOC label, in practice phases out all high boiling solvents, deep-eutectic solvents, ionic liquids and supercritical or sub-critical

fluids.[15] Typically, solvent descriptors have trouble to account for the dual nature of solvents, simultaneously molecular species and continuum media. Arguably, the best solvent property spanning across both aspects is polarity, representing 1) dipole moment at molecular level, 2) 'likeness' between solvents and in relation to solutes.[16] However, partly due to its broad scope, polarity is not universally defined and different parameters are typically employed to describe its diverse facets. Some of the most used solvent parameters to account for polarity and other key properties are illustrated below.[16]

Dielectric constant or relative electric permittivity

Dielectric constant or permittivity (ϵ_r or κ) measures the polar susceptibility of a medium under the effect of an electric field. As shown in equation 1.2.1, the measure is given against the permittivity of vacuum (ϵ_0). Permittivity is often predicted computationally and used as an indication of polarity. Furthermore, it is a useful tool to estimate the potential for accumulation of electrostatic charge by a solvent, which is a critical safety factor in process chemistry. However, in interpreting solvent effects it has the limitation of representing only a bulk property and not effects at molecular level.[16]

$$\epsilon_r = \frac{\epsilon}{\epsilon_0} \tag{1.2.1}$$

Dipole moment

Molecular dipole moment accounts for the distribution of electric charge across a molecule, and is typically expressed in Debye units (D) or in Couloumb·meter. Dipole moment is indirectly measured by multiplying the dielectric constant (ϵ) by the strength of the electric field applied (E), as illustrated in equation 1.2.2.[16]

$$D = \epsilon \cdot E \tag{1.2.2}$$

Refractive index

Refractive index (n) represents the way a medium alters the speed of a ray of light that passes through it. This effect is expressed as a ratio of the speed of light (standard wavelength λ = 589 nm) in vacuum (c_0) against the speed observed in the solvent (c), as in equation 1.2.3. The molar polarisation (P_m) can also be derived from the refractive index, and takes into account the relative molar mass (M_r) and the density (ρ) of the solvent to yield an indirect measure of its polarisability (equation 1.2.4).[16]

$$n = \frac{c_0}{c} \tag{1.2.3}$$

$$P_m = \frac{(n^2 - 1)}{(n^2 + 1)} \cdot \frac{M_r}{\rho} \tag{1.2.4}$$

Molar volume

Molar volume (V_m) represents the volume occupied by a mole unit, similarly to how molecular weight represents the mass of a mole unit. V_m is most easily calculated via equation 1.2.5 as the ratio between molecular weight (M_w) and density (ρ) of the solvent.[16]

$$V_m = \frac{M_w}{\rho} \tag{1.2.5}$$

Reichardt's E_T(30) parameter

This is an empirical parameter of solvent polarity determined via a Ultraviolet (UV) spectroscopic method based on the solvatochromism of a dye, pyridinium-*N*-phenolate betaine (Figure 1.2). The distribution of electronic charge across the molecule can be stabilised by the solvent's polarity to different extents, which causes the dye to change colour. The maximum wavelength of the transition energy of the dye in a solvent defines its E_T(30) parameter (equation 1.2.6). This value is often used in a normalised form against the interval between tetramethylsilane (low polar) and water (high polar), as shown in equation 1.2.7.[16] In this expression, h is Planck's constant, c is speed of light, v_{max} and λ_{max} are characteristic

of the UV absorption peak of the solvent, and N_A is Avogadro's constant.

$$E_T(30) = hc\nu_{max} \cdot N_A = \frac{28,591}{\lambda_{max}} \tag{1.2.6}$$

$$E_T^N = \frac{E_{T(solvent)} - E_{T(TMS)}}{E_{T(water)} - E_{T(TMS)}} = \frac{E_{T(solvent)} - 30.7}{32.4} \tag{1.2.7}$$

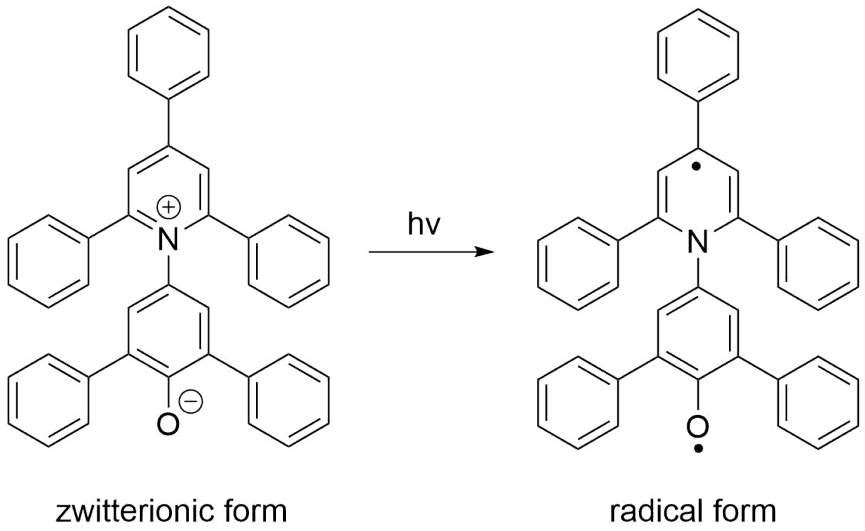


Figure 1.2: Reichardt's dye, pyridinium-N-phenolate betaine.[16]

Kamlet-Taft solvatochromic parameters

Similarly to Reichardt's polarity parameter, Kamlet and Taft developed a solvatochromic method which allows the evaluation of independent contribution of three factors in solvent polarity. Kamlet-Taft parameters are often considered appropriate for mechanistic studies of solvent effects for this reason. They outlined below are measured via UV spectroscopy taking the wavelength of maximum absorption, λ_{max}, of three dyes in the relevant solvent. The structures of the corresponding dyes are illustrated in Figure 1.3.[17]

- Polarisability, π^*, using *N*,*N*-diethyl-4-aniline (NN).

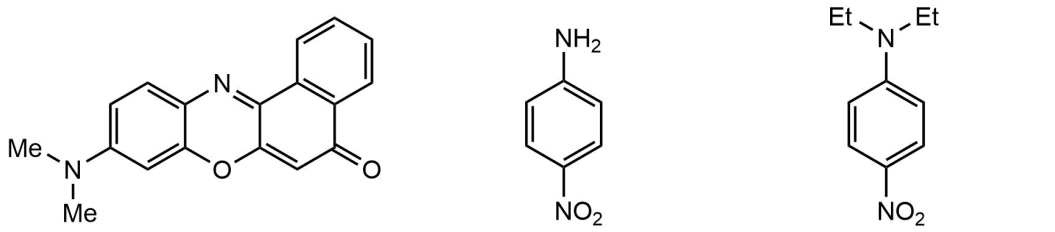
$$\pi^* = 14.57 - \frac{4,270}{\lambda_{max}} \tag{1.2.8}$$

- Hydrogen-bond accepting ability or 'basicity', β, using 4-nitroanisole (NA);

$$\beta = 11.134 - \frac{3,580}{\lambda_{max}} - 1.125 \cdot \pi^* \tag{1.2.9}$$

- Hydrogen-bond donating ability or 'acidity', α, using Nile Red (NR);

$$\alpha = 0.0485 \cdot \frac{28,591}{\lambda_{max}} - 0.46 \cdot \pi^* - 2.75 \tag{1.2.10}$$



Nile Red (NR) | 4-Nitroaniline (NA) | N,N-Diethyl-4-nitroaniline (NN)

Figure 1.3: Solvatochromic dyes used for measurement of Kamlet-Taft parameters.[16]

Overall, it was demonstrated that the combined contribution of Kamlet-Taft parameters to describe a solvent-dependent phenomenon can be represented through a Linear Solvation Energy Relationship (LSER) such as in equation 1.2.11. This is often applied to define variations in reaction rate under the effect of different solvents.[16]

$$ln(rate) = A_0 + a\alpha + b\beta + p\pi^* \tag{1.2.11}$$

Partition coefficient

A chemical's partition coefficient ($\log P_{o/w}$) represents its tendency to partition between an aqueous and an organic phase, typically water and 1-octanol. The position of the partitioning equilibrium is a measure of its hydrophilicity *versus* hydrophobicity, given by the logarithm of the molar concentration ([S]) found in the two phases (equation 1.2.12).[16] As shown in equation 1.2.13, it was demonstrated that this parameter can be defined in the

form of LSER from the combination of other variables: molar volume (V_m), Kamlet-Taft's hydrogen-bond accepting ability (β) and polarisability (π^*).[18]

$$logP_{o/w} = log\left(\frac{[S]_{octanol}}{[S]_{water}}\right) \tag{1.2.12}$$

$$logP_{o/w} = 0.24 - 3.38\beta + 0.0266V_m - 0.96\pi^* \tag{1.2.13}$$

Hildebrand's solubility parameter

Hildebrand's solubility is a thermodynamic parameter that measures the cohesive intermolecular energy of a solvent. This is obtained as in Equation 1.2.14 from molar enthalpy of vapourisation of the solvent (ΔH_v) at standard temperature (T = 298 K), molar volume (V_m) and gas constant (R).[16]

$$\delta = \sqrt{\frac{\Delta H_v - RT}{V_m}} \tag{1.2.14}$$

Hansen solubility parameters

Hansen's solubility parameters (HSP) were developed as to provide more detailed insight into the factors contributing to Hildebrand's solubility parameter. The purpose of this set of parameters is typically in understanding and interpreting solvent behaviour in relation to solubility, diffusion, and permeation amongst other effects. Similar to Kamlet-Taft parameters, Hansen's parameters are also seen in correlation to reaction properties in LSER-type equations, however, unlike Kamlet-Taft's, Hansen's parameters are computationally derived and not empirical properties. They consist in:

- Dipolar intermolecular forces, δ_p,
- Dispersion (Van der Waals) forces, δ_d,
- Hydrogen-bond forces, δ_h.

The combination of these parameters yields the total cohesive energy density (δ) according to equation 1.2.15. In nature, this value is equivalent to Hildeband's parameter, but its determination through the individual contributions reportedly achieves more accurate results.[16]

$$\delta = \sqrt{\delta_d^2 + \delta_p^2 + \delta_h^2} \tag{1.2.15}$$

Volatility

Volatility is often used to class solvents into categories, and is typically a determining feature in their application. It can be estimated through multiple parameters, such as boiling point (b.p.), vapour pressure (P_v), and heat of vapourisation (ΔH_v). Moreover, flash point (f.p.) and autoignition temperature (AIT) are critical aspects considered in process' safety which generally correlate to volatility.[19]

1.2.3 Solvent use

Worldwide, solvents make up a market of more than 20 million tonnes per year, with an estimated value of 30-40 billion USD.[20,21] Their application is ubiquitous across sectors as shown by the breakdown of consumption in Figure 1.4.[22] Notably, paint products make up the biggest portion of solvent usage, followed by contributions in the formulation of personal care and household products, inks, adhesives and cleaning products. Many solvents also find application as dispersing media for agrochemical and pharmaceutical products as well as in the manufacturing thereof. In some cases, solvents are employed as precursors for polymers and other specialised chemicals. Other secondary applications are grouped under the category 'Other' and encompass asphalt compounding, biotechnology, coating, metal and wood industry, food industry, medical applications, natural product extraction, amongst others.[23] The proportion of solvent used as media for industrial chemical synthesis, remains unclear, however. In Figure 1.4 this contribution may be distributed across a number of relevant sectors.

Of special interest to the work presented in this thesis is pharmaceutical manufacture,

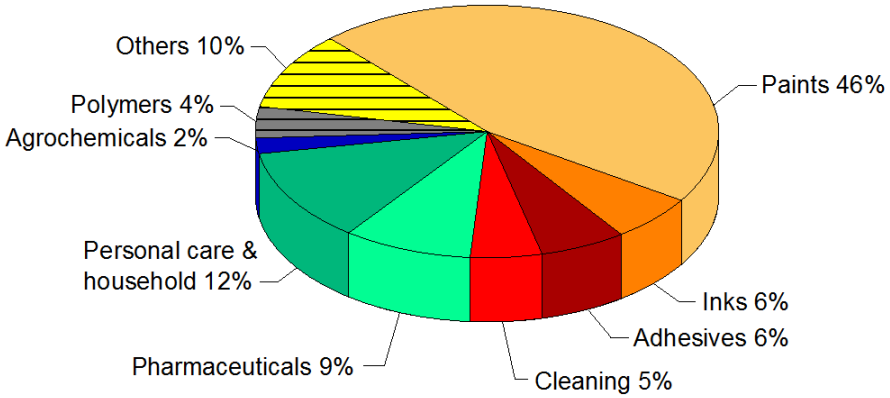


Figure 1.4: Breakdown of global solvent use by sector. Adapted from reference.[22]

which is highly solvent-intensive and, therefore, an important target for the green transition. Typically, the synthesis of an active pharmaceutical ingredient (API) consumes approximately 80-90 kilograms of solvents per kilogram of product - not including water.[24] As detailed later in this chapter, the large solvent consumption and the nature thereof, have become a matter of major concern for issues such as safety and regulation. In particular, the European legislation concerning the Registration, Evaluation, Authorisation and restriction of CHemicals (REACH) has been leading the way globally in terms monitoring trade and use of chemicals and enforcing standards.[25] This is also reflected in recent trends of solvent use in pharmaceutical manufacture. In fact, a number of publications documented the efforts towards lowering the impact of solvents and for promoting the adoption of greener solvents in the industry.[24,26–28]

1.2.4 Green and bio-based solvents

To date, the solvent sector is mostly dominated by petroleum-derived products.[29] In view of the transition to a green economy, the implementation of more benign and sustainable solvents has the potential to yield a significant impact across the industrial landscape. Ideally, in line with the 12 Green Chemistry Principles, a green solvent is one that is sustainable and safe throughout its life-cycle, *i.e.* clean manufacture from renewable starting materials, ease of recovery and recycling, and benign environmental fate.[30]

Arguably, in chemical processes, the *greenest* choice may be to avoid the use of a solvent

in the first place.[31] However, this solution often turns out to be impractical, and requiring additional use of solvents downstream.[31] Alternatively, the use of bio-derived solvents is considered a way of addressing *greenness* from at least one angle, which focuses on sustainability upstream by sourcing solvents from renewable natural feedstock.[32] By 2020, the bio-based solvent market is estimated to worth 8 billion USD, thus making up approximately 1/5 of the total sector's value.[33]

Nevertheless, although biomass is highly abundant and constantly generated on the planet, there are sustainability risks and concerns over its exploitation that must not be ignored. One highly debated example is the potential for dangerous competition with food production, both in terms of land use and market price, as occurred in the infamous case of biodiesel.[34] Another concern is the fact that, in the current state-of-art, end-to-end production of biomass is itself hardly free of fossil carbon contribution due to the use of fertilisers, pesticides, and fuels in agriculture. A way to minimise these issues has been demonstrated in the exploitation of food-waste and agricultural/forestry waste, as a valuable yet currently underutilised source of bio-derived chemicals.[35]

According to European Committee for Standardisation (CEN) the general definition of a bio-based product is one that is *"wholly or partly composed of bio-based constituent(s)"* meaning those *"derived from biomass undergone physical, chemical or biological treatment(s)"*. Biomass is hereby defined as *"material of biological origin excluding material embedded in geological formations and/or fossilized"*.[36] According to this definition, Figure 1.5 presents a range of commercially available solvents for which a production route from biomass has been identified or is already in place. The solvents falling under this definition will be referred to as bio-based solvents from this point onwards.

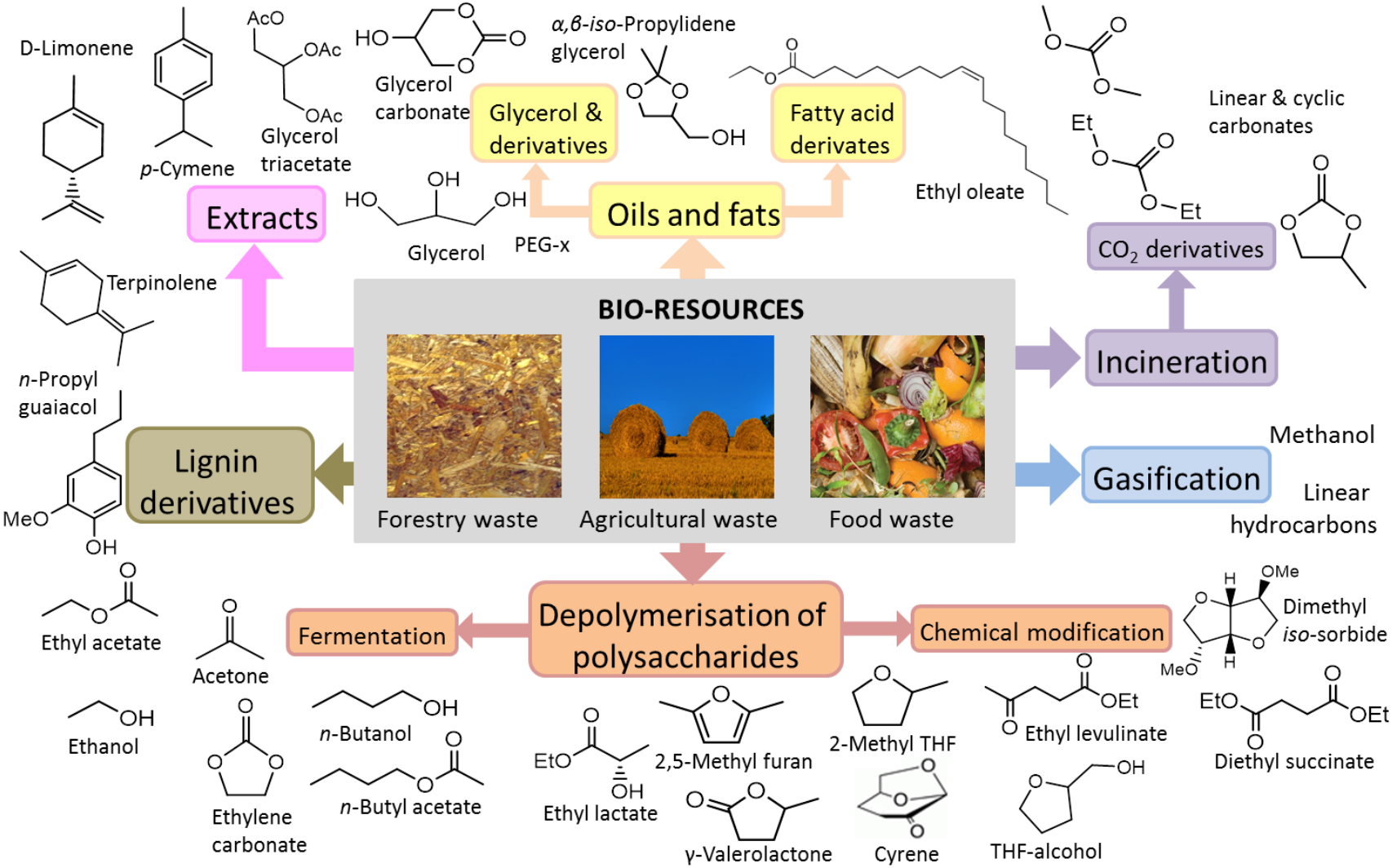


Figure 1.5: Summary of commercially available bio-based or potentially bio-based solvents. Scheme developed from information in reference.[14]

1.2.5 Case example: Citrus-derived solvents *versus* Toluene

A comprehensive analysis of most commercially available bio-based solvents, their manufacture, uses, and impact, can be found in various reviews and manuals.[14,31,37–41]

As demonstrated in the literature and in Chapters to follow, citrus-derived D-limonene and its derivative *p*-cymene can be valid solvent substitutes for traditional petrochemical solvents such as toluene in a number of chemical applications (Figure 1.6). As these solvents are of particular interest within this thesis, herein is presented an overview of their general properties and impact, alongside a comparison of the typical production routes employed for their manufacture.

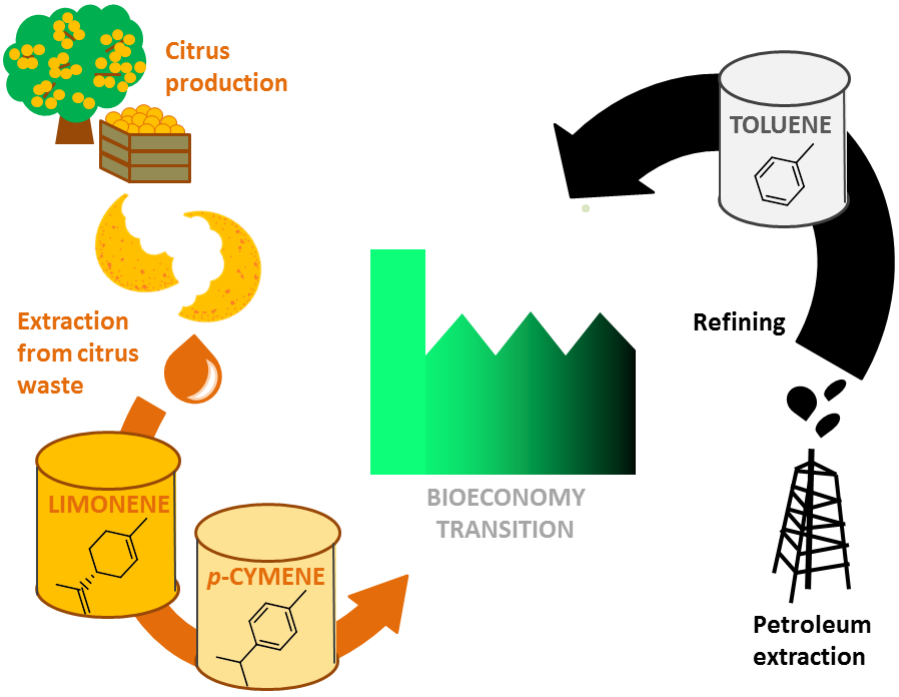


Figure 1.6: Scheme of the potential for substitution of toluene with citrus derived solvents in the chemical industry.

Manufacture

While multiple pathways for the production of D-limonene, *p*-cymene and toluene have been reported, including via biotechnology[47], the most common used ones are those summarised in Figure 1.7.[42–46]

The manufacture of toluene, right-hand side of Figure 1.7, is predominantly carried out

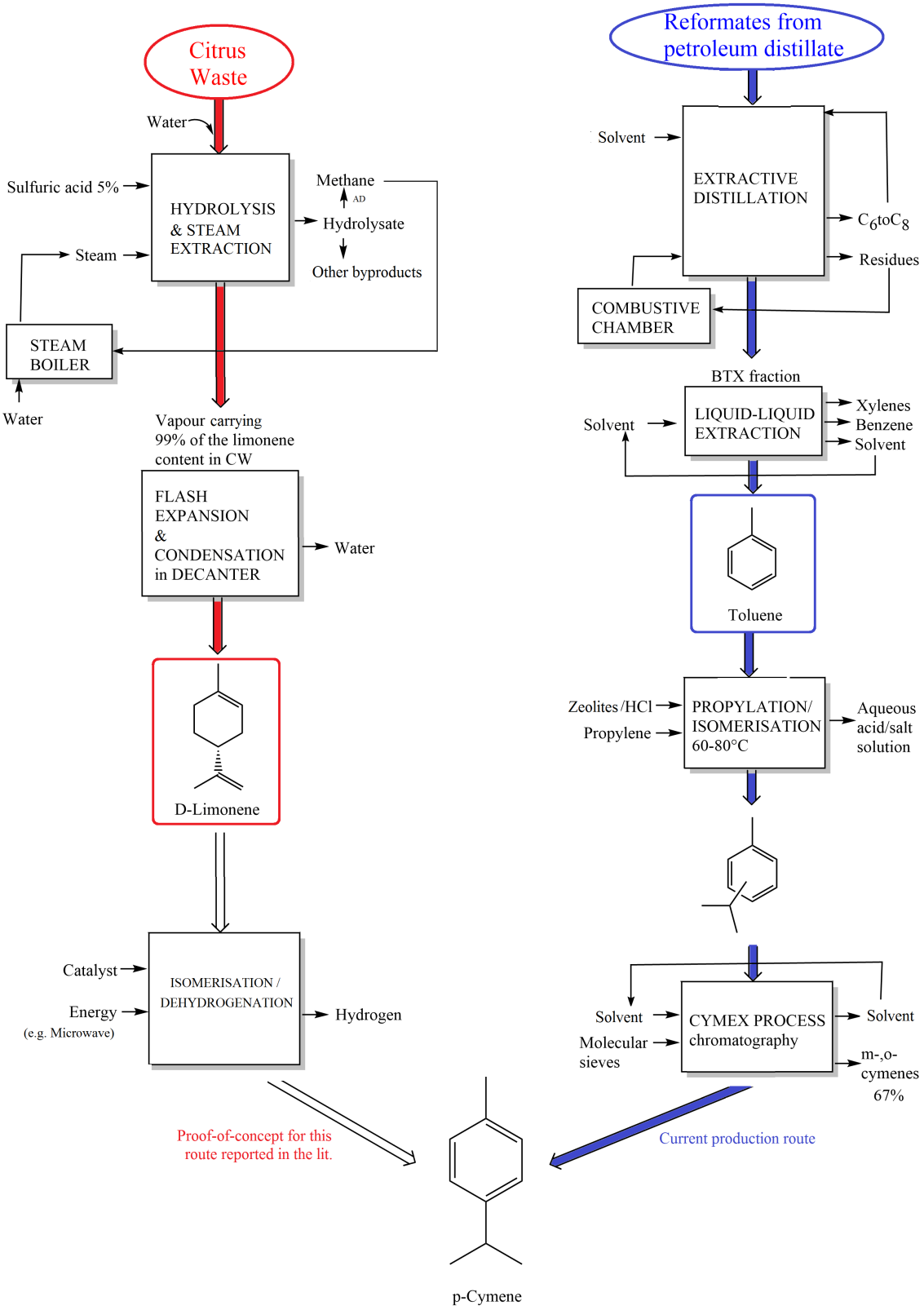


Figure 1.7: Production pathways of toluene and D-limonene as precursors to *p*-cymene. Scheme developed from information in references.[42–46]

from the naphtha fraction of crude oil distillate. Naphtha undergoes catalytic reforming to generate so-called reformates which are high in aromatic content, and include benzene, toluene, xylenes, ethylbenzenes amongst other aliphatic hydrocarbons.[48] The isolation of toluene from this mixture is typically achieved via two consecutive extractions:[42]

1. Extractive distillation with the aid of a solvent - often *N*-methyl-2-pyrrolidone (NMP). This process isolates the so-called BTX fraction, containing benzene, toluene and xylenes, from the lower and higher boiling fractions.

2. Liquid-liquid extraction using a solvent eluent. This allows further separation of the components of the BTX fraction and thus the purification of toluene.

Once isolated through the above method, pure toluene can be employed as a solvent or as a precursor for the synthesis of other chemicals. One of the derivative chemicals that can be manufactured from toluene is *p*-cymene, obtained via Friedel-Crafts propylation as part of the so-called Cymex process.[49,50] This synthesis leads to a range of cymene regioisomers, of which the desired *para*-isomer accounts for approximately 1/3 of the mixture. Chromatography is then necessary for the isolation of desired *p*-cymene.[49,50]

Although this is currently the most commonly employed route for the production of *p*-cymene, an emerging alternative is gaining attention, as it opens the opportunity for accessing this solvent from a renewable resource, *i.e.* citrus peel waste. Citrus waste feedstock is generated in large volumes from the juice industry and both household and commercial food waste. Most easily accessible within juicing facilities, the residual peel and pulp is typically fed into the following, potentially overlapping, paths:[51]

1. Direct disposal, *i.e.* landfill;

2. Drying, mostly leading to animal feed or energy generation;

3. Cold-pressing to drain citrus oil;

4. Hydrolysis and steam extraction of citrus oil.

The latter two options were demonstrated to offer the highest potential for valorisation of this waste biomass through the extraction not only of citrus oil and useful solvents, but of several valuable chemicals - *e.g.* pectin, hesperidin, sugars, cellulosic fibre, amongst others - as reviewed comprehensively elsewhere.[51–53] Citrus-derived limonene is a chiral diterpene that constitutes more than 90% of citrus oil and is most abundant in nature in the form of (R)-stereoisomer (also referred to as D-limonene).[a] It is a valuable natural product used in a number of applications - *e.g.* fragrance, flavouring, additive in food products, personal care and cosmetics, cleaning agent, amongst others - and is a proven precursor for other chemicals among which is the already mentioned *p*-cymene.[54,55]

Limonene is commercially extracted mostly via the left-hand side route described in Figure 1.7, which finds extensive reference in the literature with small variations.[46,51,52,56] This method involves acid hydrolysis of the peel and pulp matrix under steam extraction conditions, which separates the oil-rich fraction and condenses it through an expansion tank into a decanter. Ultimately, removal of the residual water leads to a reported yield of 0.75 wt% of limonene based on the initial wet peel mass.[53,57] The quality and purity of the limonene obtained may vary depending on the source and other factors, and therefore further processing may be employed to eliminate residual impurities, although such procedures are not typically reported.

Precedent in the literature shows proof-of-concept and pilot production of *p*-cymene from limonene extracted from citrus peel. The reported procedures involve dehydrogenation of limonene under heat in catalytic conditions, which directly delivers *p*-cymene with hydrogen as a byproduct.[44–46] In a technoeconomic assessment carried out by Davila *et al.*, citrus oil was directly extracted and converted as part of a continuous flow process making 9.22 kg/h of *p*-cymene with 97% purity and at an estimated cost of 5.27 USD/kg.[46] However, to date, no evidence of commercial production of *p*-cymene through this route has been found.

Currently, there are limitations and barriers affecting the citrus route becoming global such as 1) being geographically bound to citrus production or juicing and facing challenges

[a] In this thesis it will be simply referred to as limonene.

of feedstock collection, 2) being less developed and less widely spread than oil refining, and 3) lacking integration within an holistic biorefinery model. Nevertheless, the fast development of the sector and great scientific interest gathered in the field in recent years, are proving encouraging and may soon provide solutions for these present challenges.[35]

Environment, Health & Safety (EHS) impact

Table 1.1 presents a parallel between toluene, limonene and *p*-cymene in relation to a number of physicochemical properties of interest to synthetic chemistry, where a high level of similarity between these solvents can be seen across several criteria.

Table 1.1: Physicochemical properties of toluene, limonene, *p*-cymene.

Property	Unit	Toluene	Limonene	*p*-Cymene	Ref.
Molecular weight (MW)	g/mol	92.14	136.24	134.21	[58]
Density (d)	g/cm^3	0.87	0.84	0.86	[58]
Molar volume (V_m)	cm^3/mol	106.3	161.9	156.6	[58]
Viscosity (η)	c_P (25°C)	0.56	0.90	0.87	[58]
Boiling point (b.p.)	°C	110	176	177	[58]
Melting point (m.p.)	°C	-93	-74	-73	[58]
Flash point (f.p.)	°C	4	50	47	[58]
Heat of vaporisation (ΔH_v)	cal/mol (25°C)	7,933	10,508	10,516	[16]
Dielectric constant (ϵ)	(23°C)	2.38	2.30	2.30	[58]
Dipole moment (D)	D	0.31	0.29	0.0	[16]
Vapour pressure (P_v)	mmHg (20°C)	22	2.0	1.5	[58]
Hildebrand solubility (δ^2)	MPa	331	318	299	[59]
Hansen H-bonding (δ_h)	$MPa^{\frac{1}{2}}$	2.0	0.0	2.4	[59]
Hansen polarity (δ_p)	$MPa^{\frac{1}{2}}$	1.4	0.6	2.4	[59]
Hansen dispersion (δ_d)	$MPa^{\frac{1}{2}}$	18	16.5	17.3	[59]
Solubility in water	mg/l	526	insoluble	insoluble	[58]

In addition, Table 1.2 reports the GHS statements for toluene in comparison to citrus-derived solvents. Although none of the three solvents is listed as a substance of high concern within REACH, toluene presents a limitation in its commercialisation:"*Shall not be*

placed on the market, or used, as a substance or in mixtures in a concentration equal to or greater than 0.1% by weight where the substance or mixture is used in adhesives or spray paints intended for supply to the general public."[25] In fact, robust evidence indicates that toluene carries risks of reprotoxicity and causing long-term damage to organs under prolonged exposure. Some references suggest it being a hormone-disrupting substance and affecting vision and hearing (ototoxic).[60–62] As can be seen from Table 1.2, limonene and *p*-cymene show lower flammability hazard and overall lower health impact, indicating their superiority as safer alternatives to toluene. However, the risk of skin sensitisation (H317) and the higher aquatic toxicity (H410) carried by limonene may constitute factors of concern to consider in its overall profile in comparison to *p*-cymene.[63]

Table 1.2: Comparison of GHS phrases related to Environment Health & Safety profile.[63]

Toluene	Limonene	*p*-Cymene
H225 Highly flammable liquid and vapour	**H226** Flammable liquid and vapour	**H226** Flammable liquid and vapour.
H304 May be fatal if swallowed and enters airways	**H304** May be fatal if swallowed and enters airways	**H304** May be fatal if swallowed and enters airways
H315 Causes skin irritation	**H315** Causes skin irritation	
H336 May cause drowsiness or dizziness	**H317** May cause allergic skin reaction	
H362d Suspected of damaging the unborn child		
H373 May cause damage to organs through prolonged exposure		
	H410 Very toxic to aquatic life with long-lasting effects	**H411** Toxic to aquatic life with long lasting effects

Life-cycle assessment (LCA)

As for the overall life-cycle impact, a study was carried out by an industry association to measure the impact of citrus-derived oils in comparison to traditional petrochemical counterparts.[64] Table 1.3 reports the data obtained for toluene and limonene across 13 categories. Data for citrus-derived *p*-cymene is not yet available. From these results, it appears that the potential for causing climate change is 8-times higher for toluene than

for limonene. Overall, limonene shows a more favourable profile across all categories but three, where the biggest cause behind the negative result is allegedly the impact derived from the use of agrochemicals in growing the citrus fruits.[64] While the present results of this comparison are very encouraging, further life-cycle improvements for limonene could be achieved through the adoption of greener solutions for crop defence and soil treatment in citrus farming. Alternative ways of producing limonene and toluene may also be assessed in the future, potentially giving a more favourable profile.

Table 1.3: Comparison of life-cycle impact data for toluene and limonene from citrus peel (no data available for *p*-cymene either from petroleum or citrus peel).[64]

Criteria	Unit	Toluene	Limonene
Climate change	kg CO_2 eq	1654	207
Water depletion	m^3	1.03	0.75
Metal depletion	kg Fe eq	0.892	0.371
Fossil depletion	kg oil eq	1526	52
Terrestrial acidification	kg SO_2 eq	4.25	0.83
Fresh water eutrophication	kg P eq	6.87E-03	4.84E-02
Marine eutrophication	kg N eq	1.22	0.28
Terrestrial ecotoxicity	kg 1,4-DB eq	0.016	0.006
Fresh water ecotoxicity	kg 1,4-DB eq	0.86	0.468
Marine ecotoxicity	kg 1,4-DB eq	0.99	0.54
Human toxicity	kg 1,4-DB eq	90.23	49.97
Ozone depletion	kg CFC-11 eq	8.28E-08	9.42E-06
Photochemical oxidant formation	NMVOC	5.57	0.81

1.3 Biocatalysis in organic solvents

In 2002, Straathof *et al.* published a survey of industrial biocatalytic landscape reporting an exponential increase in adoption of biocatalytic technology across sectors. They found that over 50% of all industrial biotransformations at the time were carried out in pharmaceutical manufacturing, most frequently involving the use of hydrolases and oxidative biocatalysts (as shown in Figure 1.8 adapted from the study).[65] This early interest and engagement by the pharmaceutical industry is not surprising considering the importance of chirality in drug molecules and the ability of enzymes to install and preserve chiral selectivity.[66–68]

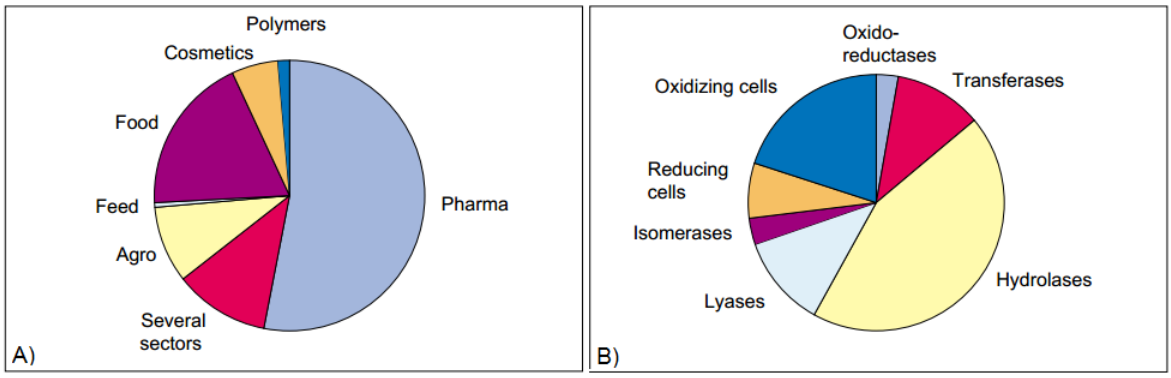


Figure 1.8: Breakdown of industrial sectors (A) using bio-processes at commercial scale and of and enzyme types (B) employed in such processes. Results based on 134 industrial processes reported by Straathof *et al.*[65]

This trend (later confirmed by other studies[69]) has recently progressed with a great expansion of the biocatalytic toolbox made possible by significant breakthroughs in biotechnology and protein engineering.[70,71] Nevertheless, while the knowledge, selection and availability of enzymes for specific transformations are vast, there is a pressing need to integrate these developments within synthetic practices and as part of routine retrosynthetic thinking. Moreover, making these catalysts practical, off-the-shelf options would highly facilitate their integration and adoption in chemistry laboratories.[68]

Nevertheless, several examples of pharmaceutical syntheses where enzymes brought significant advantages in alternative to or in combination with a chemical route are comprehensively described in the literature.[69,72,73] Overall, on the quest for *greenness*, Wenda

et al. presented a check-list of factors that must be taken into account for defining a green biocatalytic process:[74]

- Sustainability considerations of catalyst and other reaction components,
- Availability of substrates and scope,
- Availability & cost of the catalyst,
- Stability & reusability of the catalyst,
- Atom economy,
- Conversion & yield $> 95\%$,
- Enantioselectivity $> 99\%$,
- Volumetric productivity > 100g/l - not applicable when using living organisms,
- Specific activity of enzyme > 1U/mg.

In biocatalytic processes as much as in chemical ones, the solvent medium constitutes the highest impact in terms of process mass, as discussed earlier. Therefore, the choice of solvent is a very important component in making a biocatalytic route sustainable. Although predominantly applied in water, the use of organic solvents in biocatalysis is typically a means for addressing solvation and isolation difficulties found with traditional aqueous processes.

The first published work on the use of an enzyme in a neat organic solvent dates back to the 1930's, when Sym (1935) reported a transesterification in chloroform catalysed by an esterase.[75] It was only 50 years later that the field of non-aqueous biocatalysis started attracting wider interest. In the early days, numerous attempts were made to unravel the complex matrix of solvent effects to try and model solvent behaviour in biocatalysed systems.[76–78]

This Section reviews these efforts and provides a summary of the state-of-the-art with special focus on lipase catalysed reactions and the use of bio-based organic solvents. In the most recent years, the focus appears that have mostly shifted towards testing enzymatic activity in unconventional media - also termed *neoteric* solvents - seeking to unveil new behaviours and/or to improve performance, practicality or sustainability of the catalytic tool.[79,80] Ultimately, focusing on solvents that can be derived from renewable resources is

seen as a strategy bringing benefits to the process without compromising on its safety and sustainability.

1.3.1 Lipases and other esterification methods

A special focus in the enzymology literature is found on lipases. Lipases are members of the hydrolase family and in nature catalyse the formation and cleavage of fatty esters, as well as numerous metabolic pathways.[81] Unlike most enzymes, they are cofactor- and metal-free, they activate when in contact with a hydrophobic surface, and they tend to have a broad substrate scope. Most notably, lipases have outstanding performance in hydrophobic systems and are highly versatile tools in chemical and biochemical synthesis, catalysing esterifications, transesterifications, amidations, and hydrolyses, amongst other interesting transformations.[71,81]

According to numerous pharmaceutical industry publications, the formation and hydrolysis of esters and amides, also as protecting groups, are invariably top transformations in drug syntheses.[82–84] Most often, the techniques used to build acyl bonds rely on old-fashioned chemistries that are prevalently acid catalysed in solution - *e.g.* Fischer esterification (1895).[85] Carrying out these reactions in milder conditions often requires the use of activated groups and coupling agents, at the expense of the atom efficiency of the process and often its safety.[86] Overall, several methods are available for the synthesis of esters which can be found reviewed elsewhere.[87,88] A number of these methods have been assessed from a green chemistry perpective as part of a dedicated reagent guide developed by GlaxoSmithKline in 2013. Table 1.4 reports the results of this assessment as a classification of preferred and unpreferred methods using a red/amber/green ranking.[84] Out of this assessment, the use of enzymes for catalysing ester bond formation is amongst the *greenest* options.[84]

The synthesis of Pregabalin illustrated in Figure 1.9 is a notable example where adopting an enzymatic strategy brought great savings and green improvements. This compound for the treatment of neuropathic pain, was originally prepared via a chemical route yielding 20% overall yield. The introduction of a stereospecific lipase allowed resolving the

racemic intermediate via dynamic kinetic resolution (DKR), thus doubling the overall yield and bringing down significantly the volume of waste generated by the process.[72,89]

Table 1.4: GSK's green reagent guide for common use esterification methods.[84]

Desirable methods	Some issues	Undesirable methods
Enzyme - lipase	Mukaiama reagent	Steglich (DCC, DMAP)
Solid acid catalyst	I_2 catalytic	Mitsunobu (DIAD, Ph_3P)
HCl	COMU®	Triphosgene
H_2SO_4	TMSCl	TMS-diazomethane
Dimesitylammonium pentafluorobenzenesulfonate	Et_3N, $MgBr_2$	Diazomethane
Diphenylammonium triflate	Thionyl chloride	Methyl iodide
Acetyl chloride	Oxalyl chloride	Boric acid
		Dimethyl sulphate
		EDCI, DMAP

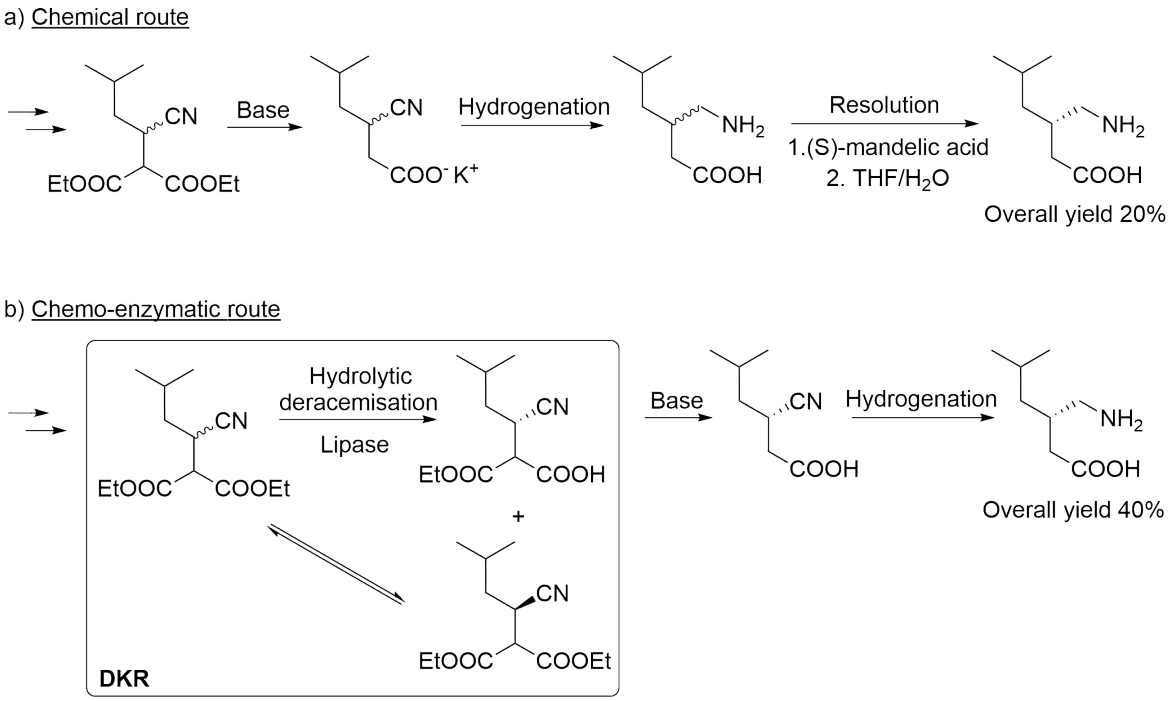


Figure 1.9: Parallel between chemical and chemo-enzymatic routes to Pregabalin. Adapted from reference.[72]

1.3.2 *Candida Antarctica* Lipase

While a wide range of lipases are capable of catalysing ester bond formation, not all result competitive against traditional chemical methods in terms of 1) commercial availability and price, 2) recoverability and reusability, 3) substrate scope, 4) throughputs.[74] Popular enzymes that are well known to deliver these requirements are lipases from *Candida Antarctica*, type A (CALA) and type B (CALB). As suggested by the name, the original strain of *Candida Antarctica* was first isolated from deep Antarctic ice, where the harsh environmental conditions forced microorganisms to develop sturdy enzymes with remarkable stability.[90] For this reason, their use as biocatalysts for organic synthesis has been extensively investigated, particularly in organic solvents where they show particularly enhanced performance.[91,92]

Once the substrates enter the active site, the catalysis is known to follow classic Michaelis-Menten kinetics, characterised by first-order reaction kinetics shown in equations 1.3.1. The rate-limiting step is considered to be the formation of the enzyme-substrate complex [EA], also called Michaelis-Menten complex. The rate of formation of the final product will depend on the measured maximum rate of consumption of the substrate, $rate_{max}[A]$, and the so-called Michealis constant, K_M, representing at the concentration of substrate A at $\frac{rate_{max}}{2}$ (equations 1.3.1). Figure 1.10 shows the accepted mechanism of lipase catalysis, involving the catalytic triad made of three non-sequential aminoacid residues $[Ser...His...Asp/Glu]$ and the oxyanion hole which hosts the carboxyl head.[77,92]

$$\mathrm{E + A \rightleftharpoons [EA] \longrightarrow E + P}$$

$$rate = \frac{rate_{max}[A]}{K_M + [A]} = \frac{k_{cat}[A]}{K_M + [A]} \tag{1.3.1}$$

1.3.3 Organic *versus* Aqueous media

In nature, enzymes tend to perform predominantly within aqueous environments, and their secondary and tertiary structures have evolved accordingly.[77] Typically the active

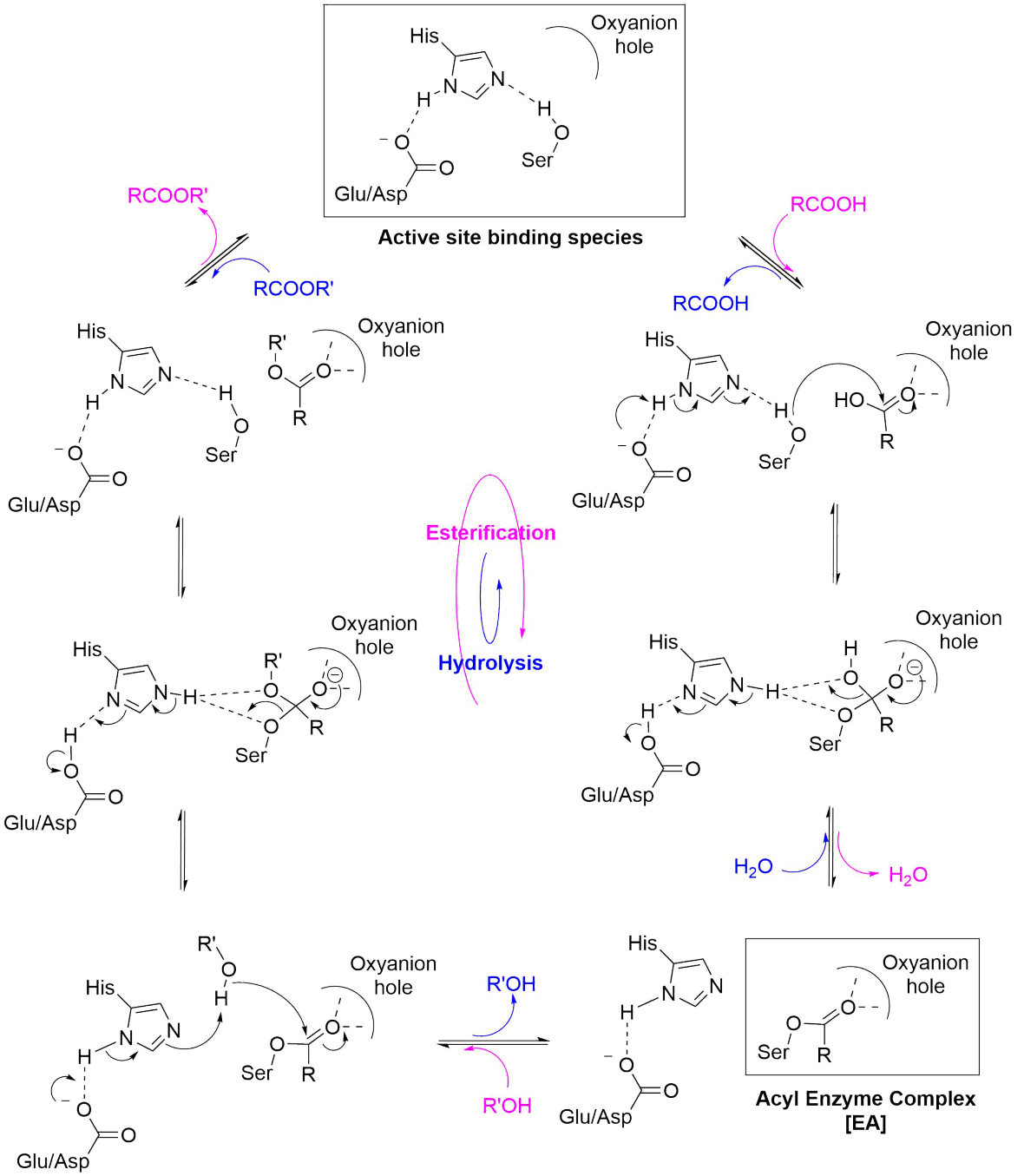


Figure 1.10: Mechanism of lipase's catalytic cycle. Clockwise (pink) is the mechanism for ester formation, anticlockwise (blue) is the mechanism for ester hydrolysis. Adapted from reference.[92,93]

conformation exhibits the most hydrophilic aminoacid residues at the outer surface of the protein, whilst shielding the most hydrophobic ones internally, away from the water. As a consequence, there is a wide spread belief that, once in an organic hydrophobic solvent, an enzyme would inevitably unfold and denaturate.[94] Chen argues that such rooted perception has dramatically held back the application of biocatalysis in organic media.[95] However, additional causes are also seen in the lack of understanding of how solvents actually influence biocatalysis, and the challenges faced so far in developing an all-encompassing model to facilitate solvent selection in this application.[95] Typical advantages of using enzymes in organic solvents over the traditional aqueous biocatalysis are summarised below.[94,96]

- Reducing dissolution and leaching of enzymes in solution thus favouring recovery and reuse;
- Increasing enzymes structural rigidity;
- Enhancing thermostability and pH-resistance;
- Altering and/or tuning enzyme specificity;
- Improving stereoselectivity.

1.3.4 Solvent effects

It was demonstrated that the principal factors influencing enzymatic activity in organic solvents are: 1) the solvent-tolerant nature of the enzyme, 2) the water content of the system, 3) the temperature/pressure applied, and 4) the medium properties. In this work, major attention was paid to the effects of solvent medium on enzymatic activity. With respect to this category of effects, a recent perspective article by Dyson and Jessop delineated and rationalised some of the common traits of solvent effects that are seen consistently across catalysis.[97] By drawing from their suggested classification of solvent effects, a review is herein proposed of those that have proven to be critical in biocatalysis.

Solvent interaction with the catalyst

Although enzymes can be successfully used in non-aqueous media, water remains an indispensable component in the structure of a functioning enzyme. Water molecules are crucial to allow the right tertiary conformation of the enzyme, impacting directly on the

shape of the active-site pocket and its flexibility to bind substrates and intermediates. As Schmitke pointed out, *"water acts as a molecular lubricant, increasing the conformational mobility of the enzyme"*.[98] Studies on the influence of water content in enzyme-catalysed processes have concluded that 100% dry conditions typically lead to deactivation of the enzyme. However, a gradual reactivation of the enzyme is observed upon increasing the water content and the optimum water content appears to be both enzyme- and solvent-dependent.[99,100]

Notably, solvent polarity was found to be critical in determining the amount of water that is effectively available to the enzyme's surroundings and, therefore, its activity. The more polar the solvent, the more it will disperse any water present and so deprive the enzyme of it. Thus, the best catalytic activity for an enzyme is typically achieved in non-polar solvents even at very low water content; in contrast, when polar solvents are employed, they are predominantly seen as co-solvents in aqueous systems where water remains in great excess.[101] This direct relationship between dehydration and deactivation was supported by various computer modelling studies, such as by Micaelo *et al.* and by Lousa *et al.*[102,103] Empirical confirmation of these results is also found, although in such studies accurate quantification of the water content in molecular terms proves challenging.[77]

Overall, the solvent parameters that account for solvent-water interaction proved to be most critical in describing observed changes in activity. A pioneering study carried out by Laane *et al.* showed that the water partition coefficient, logP, could be used as a useful indication for medium optimisation in biocatalysis.[104] Although this work did not demonstrate a linear correlation between logP and enzyme activity, it showed that solvents having logP < 2 (*i.e.* good water-solubility) are more likely to cause disruption to the enzyme conformation. Conversely, solvents with logP > 4 (*i.e.* hydrophobic) tend to favour and enhance enzymatic activity.[104]

Following Laane's work, other solvent descriptors were investigated in a quest to produce a good model for solvent effects.[95,105–108] A summary of the key studies is reported in Table 1.5. It can be derived that these single variable approaches were mostly unsuccessful to provide a definite answer and appeared to be highly enzyme-specific. A study

published by Schneider *et al.* attempted to correlate the activity of yeast alcohol dehydrogenase against three solvent parameters as part of a descriptive 3D-model, showing higher significance of Hansen's hydrogen-bonding parameter (δ_h) over Hildebrand's parameter (δ) and solvent hydrophobicity (logP).[109] However, so far, very little research has investigated multivariable approaches.[78]

Table 1.5: Past attempts to correlate enzymatic activity to solvent properties.

Solvent property	Symbol	References
Water activity	a_w	106,110
Partition coefficient	log P	104,110,111
Water solubility in the solvent	log $S_{w/o}$	105,109,112
Solvent solubility in water	log $S_{o/w}$	105
Dielectric constant	ε	104,113
Dipole moment	D	104
Hansen hydrogen bonding	δ_h	104,109,114
Polarisability	π^*	104
Hildebrand solubility	δ	109,115
Reichardt's polarity	$E_N^T(30)$	116

Solvent interaction with solutes

Typically organic solvents are employed to improve solubility of a substrate that proves problematic in aqueous systems. Hence, traditionally, medium optimisation has been typically focused on the optimisation of substrate solubility. Laane *et al.* argued that the partition coefficient was also an indication of the extent of solutes diffusion between the organic continuum and the more polar environment surrounding the enzyme.[104] Effectively, according to this theory, the active site's environment should provide high compatibility with the substrate, thus attracting it from the solution and favouring binding.[104] Meanwhile, the solvent should ideally have higher affinity for the product than for the substrate, so that it would disperse the product as soon as it is formed and thus shift the reaction

equilibrium towards conversion.

This realisation led to a dramatic change in the way medium optimisation is pursued in such systems, shifting the focus from optimising solubility of the substrate to improving the stabilisation of the product.[77]

It is worth noting another diffusion-related effect which is determined by changes in temperature and pressure. Less viscous media typically favours mobility of solutes, implying that high temperatures and low pressures should lead to higher throughputs.[117,118] This consideration is especially relevant when the solvent is a supercritical or near-critical fluid, and is often coupled with favourable thermodynamic contributions in the case of endothermic reactions.[119]

Solvent-induced activating effects

The presence of a micro-biphasic environment surrounding enzymes in organic media is a very peculiar feature of these catalytic systems and, to a large extent, it dictates their behaviour. In the case of lipases, this has an additional characteristic effect which enhances performance and broadens applicability of this technology.[120]

Uniquely, lipases tend to undergo a special type of activation when in contact with any sort of hydrophobic-hydrophilic interface, which may be typically 1) liquid-liquid interface, 2) a solid hydrophobic surface, or 3) simply the solvent coating of the superficial moisture of the enzyme's structure.[77] In such conditions the active site was found to open up to the non-polar phase, thus welcoming and hosting the catalytic process.

Figure 1.11 presents a scheme of the interface activation which triggers the active site pocket to open up wide once in contact with a non-polar surface - being it an organic interface or an hydrophobic solid support. Crystallographic evidence showed a proteic bulge similar to a lid covering and locking the active site when the medium is aqueous or highly polar.[121] In contact with an interface, the lid curls up to expose its inner hydrophobic residues to the like hydrophobic surface, thus making the active site available to receive the substrates.

Due to this effect, lipase activity appears to be less dependent on solute solubility and

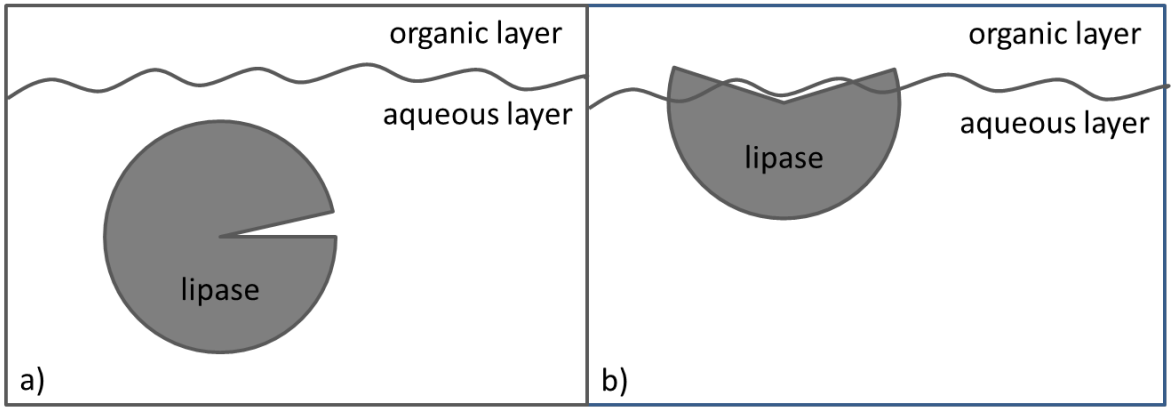


Figure 1.11: Scheme of typical lipase activation at an organic-aqueous interface.[77]

diffusion, as access to the active site is so favoured. Indeed, the relative apolar character of the active site and the direct exposure of the binding sites to the organic phase tend to minimise the impact of partitioning effects and to speed up the critical binding process.[91,122,123]

1.4 Examples of non-aqueous biocatalysis

Historically, the solvents in which enzymatic reactions have been screened and reported in the literature mirrored the range of commonly used solvents in general laboratory practices. A survey of over 50 studies amongst the most influential in the period from 1980 to 2015 yielded the indicative frequency ranking illustrated in Figure 1.12.

Overall, analysis showed the most frequently used solvents to be: toluene, hexane, tetrahydrofuran (THF), dioxane, acetonitrile, methanol, ethanol, benzene, acetone, dimethylsulfoxide (DMSO), and chloroform. In this list a striking diversity in terms of chemical nature, functionality, polarity and water-affinity of the solvents is apparent. This fact may have two interpretations:

1. Each enzyme and reaction is highly specific in its medium requirements, spanning across a wide range of properties. No unique solvent could serve as optimal medium for every biocatalytic process.

2. A conceivably random approach to solvent selection for screening may have been broadly employed due to a lack of mechanistic understanding of such systems.

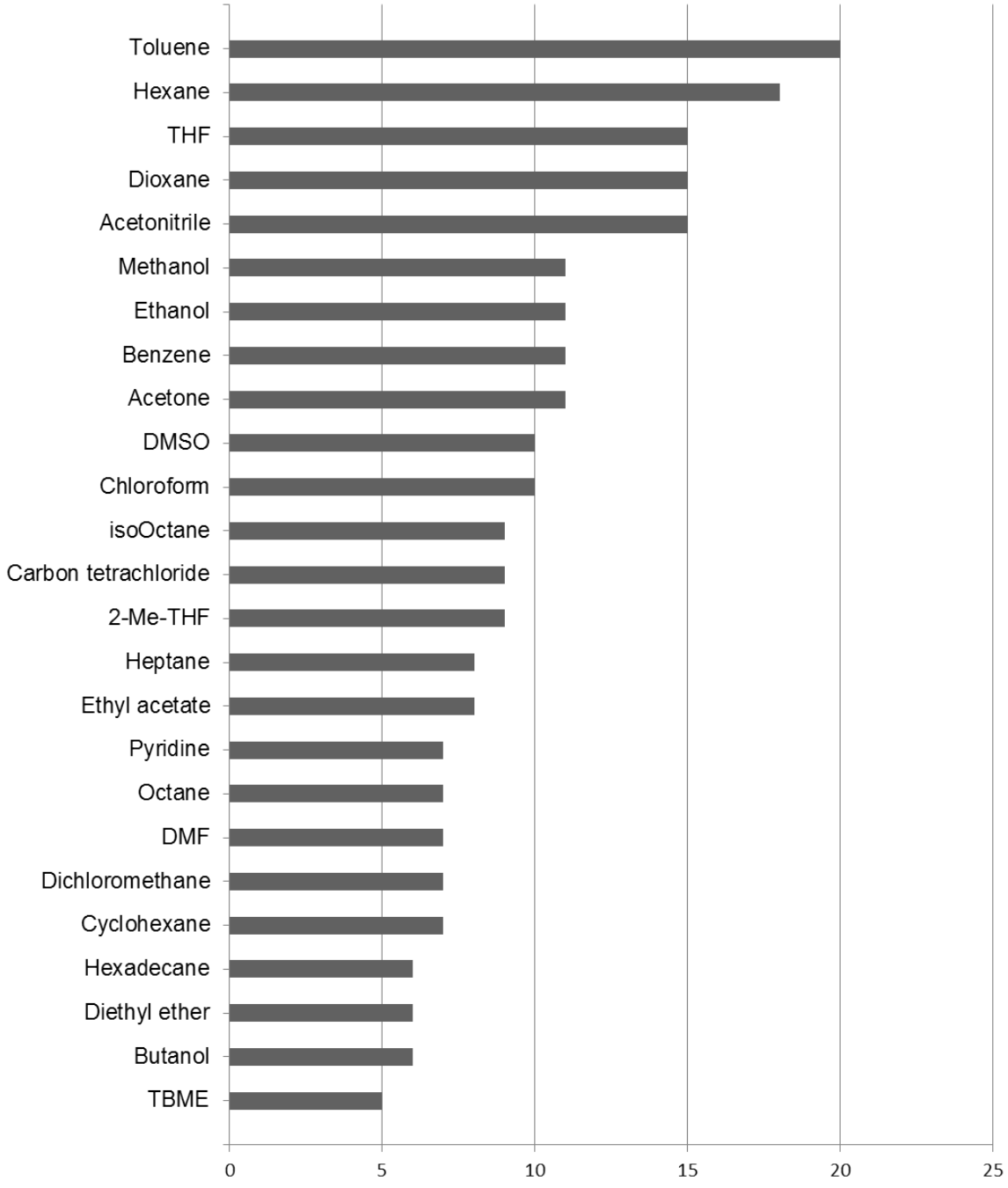


Figure 1.12: Frequency of occurrence of common organic solvents used in biocatalytic processes and screenings as gathered from 54 literature articles from 1980 to 2015.

From a green chemistry point of view, most of these solvents are derived from fossil resources and carry significant concerns as to their toxicity, safety and environmental impact.[26] In the most recent publications, a clear sign of growing awareness and concern about the risks and sustainability of solvents is seen, with an increased interest in screening enzymes in neoteric or bio-derived solvents.[79,80] These offer opportunities not only for improved safety and *greenness*, but also for enhancing performance, tuning selectivity, easing purification and unveiling novel effects. Figure 1.13 shows some of the classes of neoteric solvents that are being investigated as media for biocatalysis. In this modified schematic representation of polarity *versus* volatility proposed by Lozano *et al.* it appears that there is wide scope for sustainable organic solvents to fill gaps in the solvent space which are not currently satisfied by aqueous media or other neoterics.[124] In the sections below, a review of literature examples reporting the use of biocatalysis in bio-based solvents or in pharmaceutical synthesis is presented, with special attention towards transformations catalysed by enzymes of the hydrolase family .

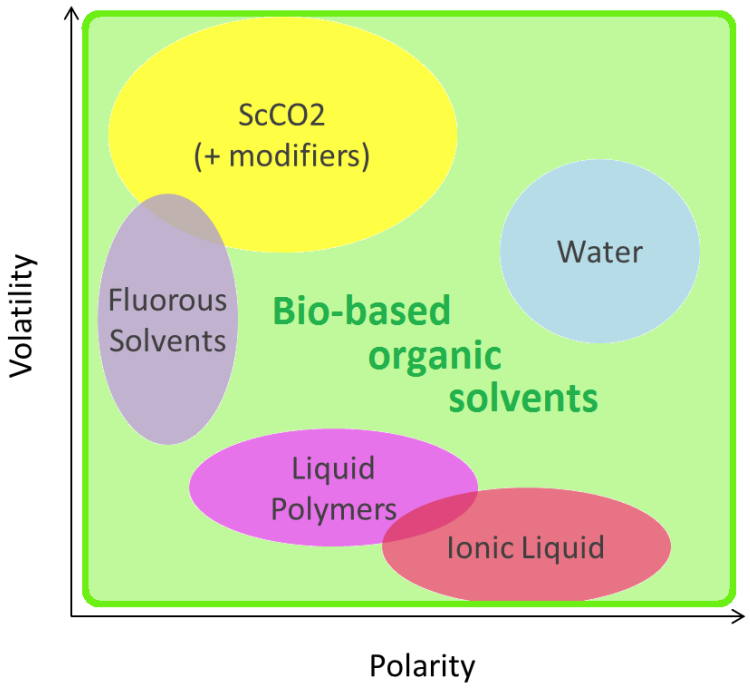


Figure 1.13: Neoteric solvents in the polarity *versus* volatility solvent space. Adapted from reference.[124]

1.4.1 Applications in bio-based solvents

Over the last 10 years, a growing but still limited number of studies have investigated the use of bio-based organic solvents. Recent review articles have comprehensively reviewed the state-of-the-art in this field.[79,80,125] Large screenings including several bio-based solvents appear to be limited in the recent literature, as studies tend to mainly focus on one specific solvent or family of solvents at a time. One of the most comprehensive assays investigating bio-based solvents was by Perez *et al.* giving proof-of-concept of successful galactosidase catalysed reactions in the presence of various glycerol-based solvents and *N,N*-dimethylformide derivatives as illustrated in Figure 1.14.[79] The best conditions were found to be with solvents 5-hydroxy-1,3-dioxane (glycerol formal) and 2-hydroxy-*N,N*-dimethylpropanamide - framed by a dashed line in Figure 1.14.

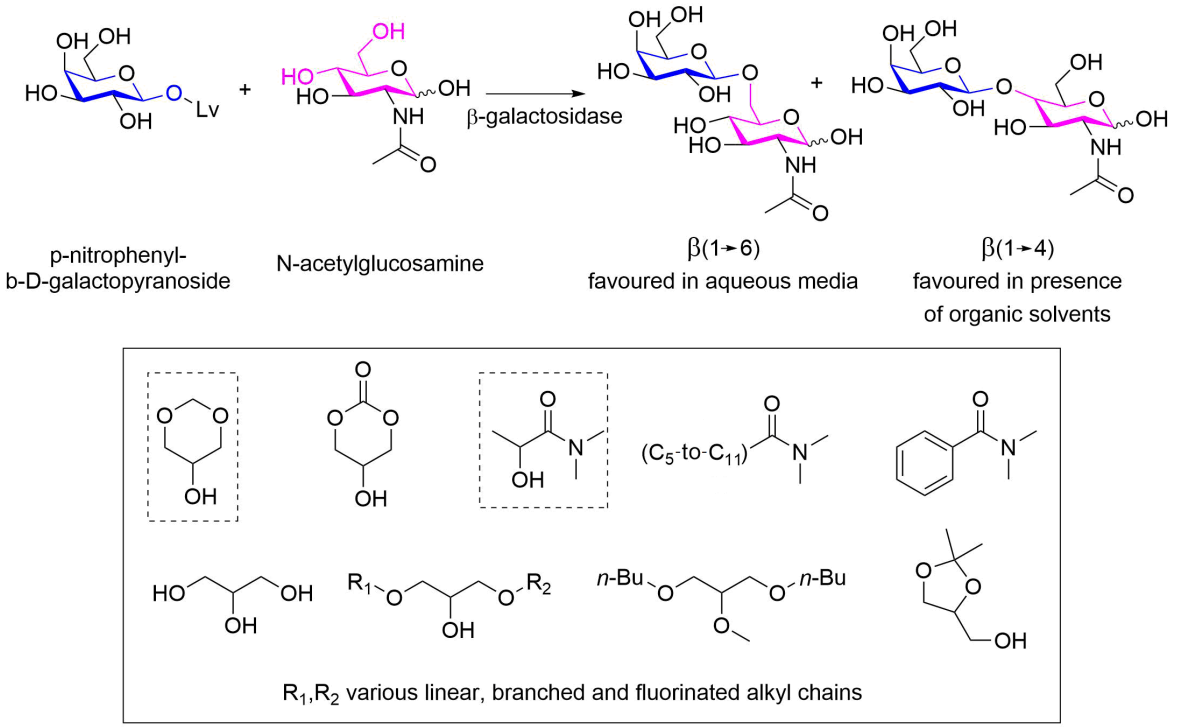
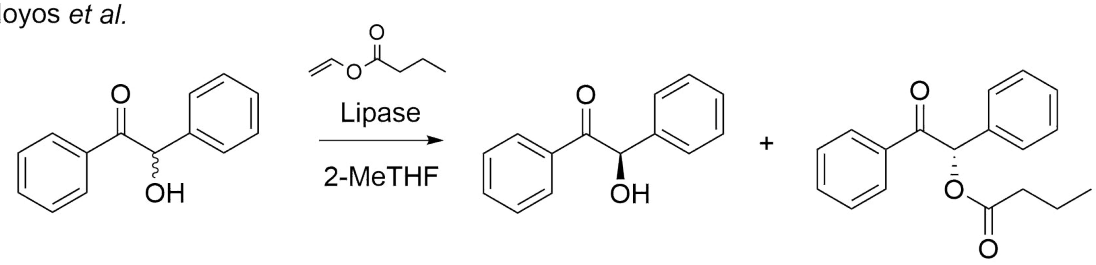


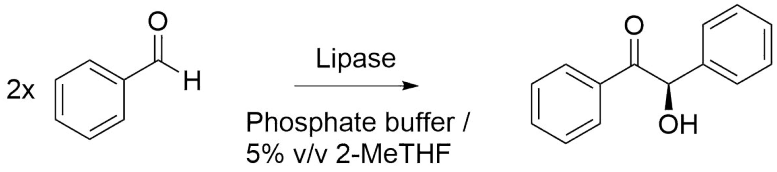
Figure 1.14: Bio-based solvents screened by Perez *et al.* in galactosidase catalysed reactions. Adapted from reference.[79]

By far, the bio-based solvent that has been attracting the most attention and proof-of-concept application is 2-methyltetrahydrofuran (2-MeTHF), a sugar-derived solvent. Figure 1.15 shows various reported examples of enzyme catalysed reactions carried out in 2-MeTHF and other bio-based solvents.[126,130] For example, Hoyos *et al.* proved 2-MeTHF

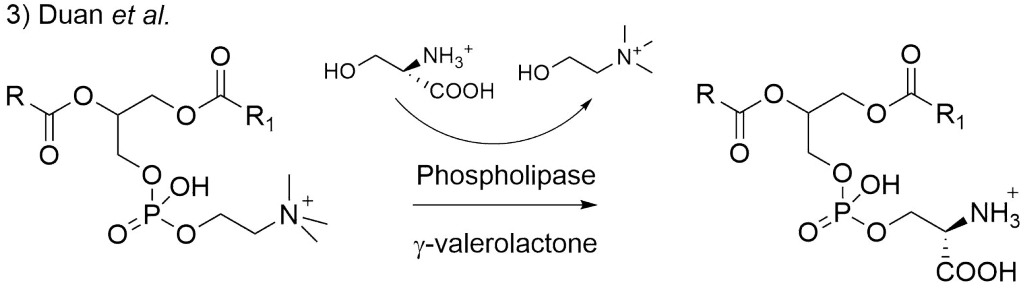
1) Hoyos *et al.*



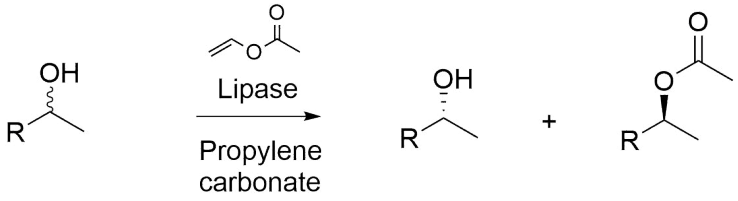
2) Shanmuganathan *et al.*



3) Duan *et al.*



4) Wu *et al.*



5) Bavaro *et al.*

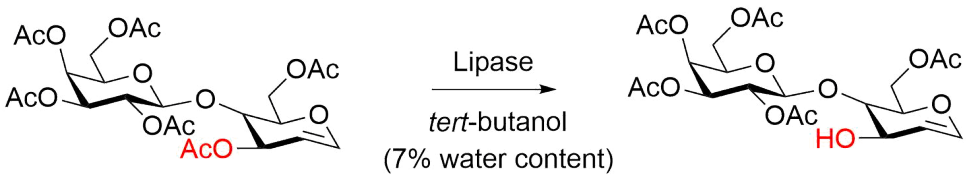


Figure 1.15: Literature precedent for the use of 2-MeTHF and other bio-based in enzymatic transformations.[125–129]

to be superior to THF in selective acetylation of benzoin substrates.[128] In another example, even the addition of only 5% of 2-MeTHF as co-solvent led to great improvements over purely aqueous systems.[127]

In a publication by Duan *et al.*, another sugar-derived solvent, γ-valerolactone, was successfully used in the lipase catalysed synthesis of phosphatidylserine, providing a first example of using this novel solvent as a greener reaction medium in biotransformations.[125] Propylene carbonate, another potentially bio-derived solvent, has also been reported for the kinetic resolution of various secondary alcohols via enantioselective acylation.[129] Alcohols have also been used as solvents in biocatalysis, often as excess reagents. Nonetheless, Bavaro *et al.* showed an example of using neat *tert*-butanol in the regioselective hydrolysis of a disaccharide by a lipase, where the water necessary for the hydrolysis was being provided through the natural moisture of the undried solvent.[131]

Overall, this field of research is finding scope and is rapidly growing. The range of novel bio-platform molecules and bio-based solvents is also expanding, thus opening further opportunity for exploring applications.

1.4.2 Applications in pharmaceutical synthesis

A number of reviews have summarised various biocatalytic strategies used to achieve pharmaceutically active compounds at large scale.[72,132–134] However, only few show application in organic solvents; some examples are herein described and illustrated in Figures 1.16 and 1.17.

In one very early industrial example of enzymatic process in organic solvents an HIV-inhibitor was synthesised from castanospermine employing lipases in different media optimised for each step: 1) pyridine, 2) THF and 3) an aqueous buffer solution.[135]

In general, most of the reported applications are found in the chiral resolution of racemic intermediates, mostly achieved via protection/deprotection strategies. As reported in Figure 1.17, Chen *et al.* screened various lipases in six solvents aiming to optimise the yield and regioselectivity in the acetylation of a pharmaceutically active compound.[130] Against DMSO, *N*,*N*-dimethylformamide (DMF), acetone, THF and toluene, only 2-MeTHF

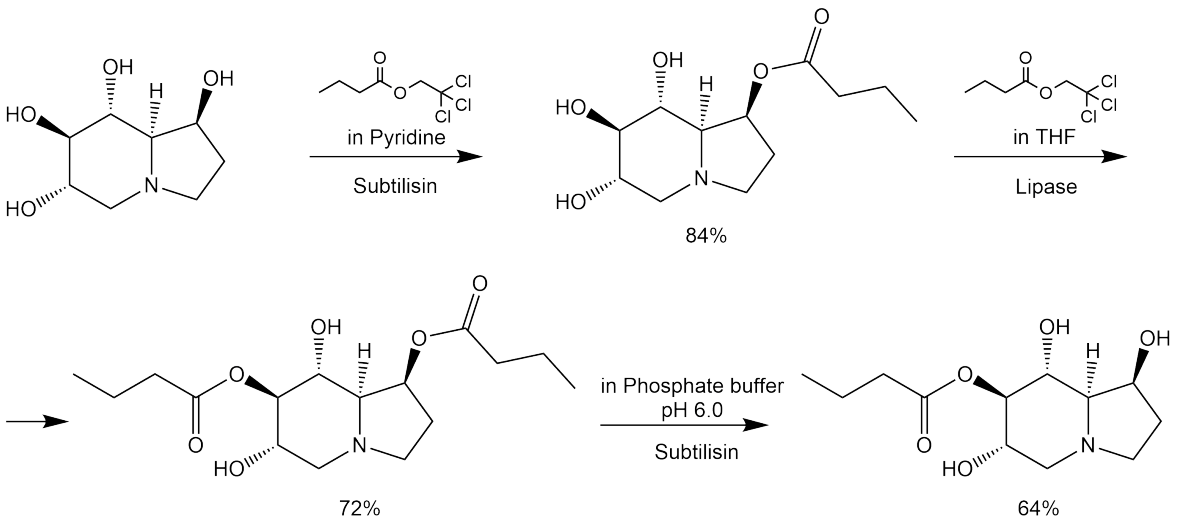


Figure 1.16: First example of large scale industrial process using enzymatic catalysis in organic media. Synthesis of HIV-inhibitor via chemoenzymatic route.[135]

gave excellent yield (96%) and total selectivity for the desired primary alcohol. Ghanem *et al.* investigated various carboxylic acids of pharmaceutical interest such as Ibuprofen and other analogues. Esterification reactions catalysed by CALB in neat organic solvents, methyl-*tert*-butyl ether (MTBE), toluene, ethyl acetate, as well as in solvent systems with an aqueous buffer, obtaining the best results with a biphasic mixture of phosphate buffer with MTBE as a co-solvent.[136] Numerous other studies have explored the late stage chiral resolution of Ibuprofen, and other solvents were also considered.[137–139] D'Antona *et al.* showed a similar approach to the resolution of Ketoprofen were 18 organic solvents were screened and dichloropropane was selected as the optimum medium for the reaction.[140]

Furthermore, another interesting example consists in the chemoenzymatic route to (S,S)-Reboxetine, which employed CALB for the acylation of a diol intermediate. While in this case the purpose was not chiral resolution, the lipase showed the highest regioselectivity in comparison to other methods, with optimum yield achieved in toluene.[141] This work is described in further detail in Section 1.4.3.

Finally, while most of the resolution strategies based on hydrolytic and deprotection reactions were carried out in water or aqueous buffer systems,[142,143] in some cases the addition of organic co-solvents was found beneficial. For instance, Murtagh *et al.* (2011) used a different lipase in a dynamic kinetic resolution (DKR) approach towards the synthesis of

a protein binder of medicinal interest. They achieved the stereoselective hydrolysis of the ethyl ester of the desired enantiomer by carrying out the reaction in water with 20 wt% of dichloromethane as co-solvent.[144]

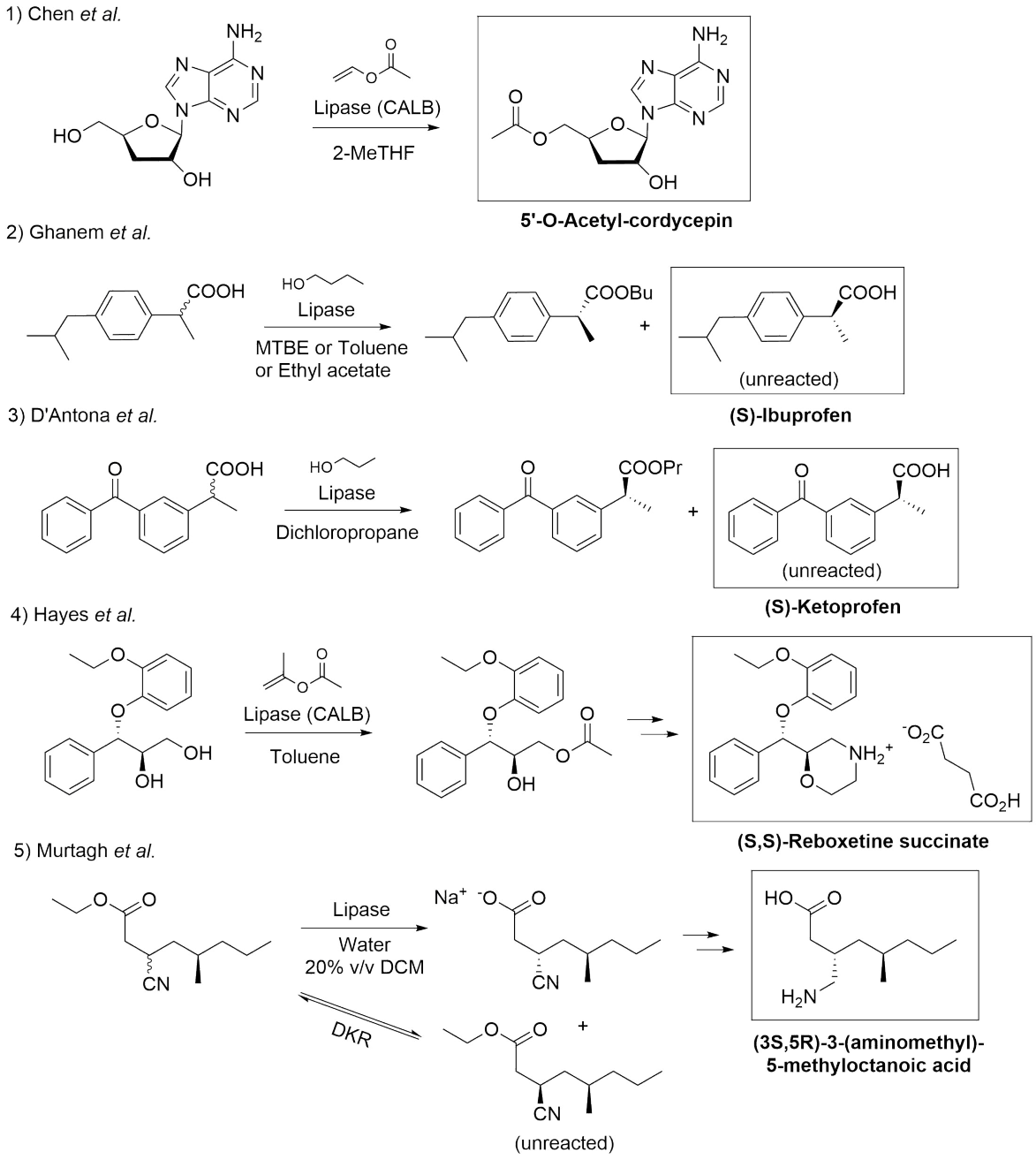


Figure 1.17: Some examples of biotransformations in non-aqueous media for the synthesis of pharmaceutically active compounds.[130,136,140,141,144]

1.4.3 Reboxetine case study

Among other examples, the synthesis of Reboxetine became of special interest to the work presented in this thesis, due to a collaboration established with Pfizer as part of the CHEM21 initiative. This recent and well-documented commercial pharmaceutical process employed an enzyme in toluene as part of a telescoped sequence of four chemo-enzymatic steps. A study of solvent substitution with bio-derived solvents and investigation over solvent effects are presented in Chapters 3 and 4, thus background to the project is herein provided.

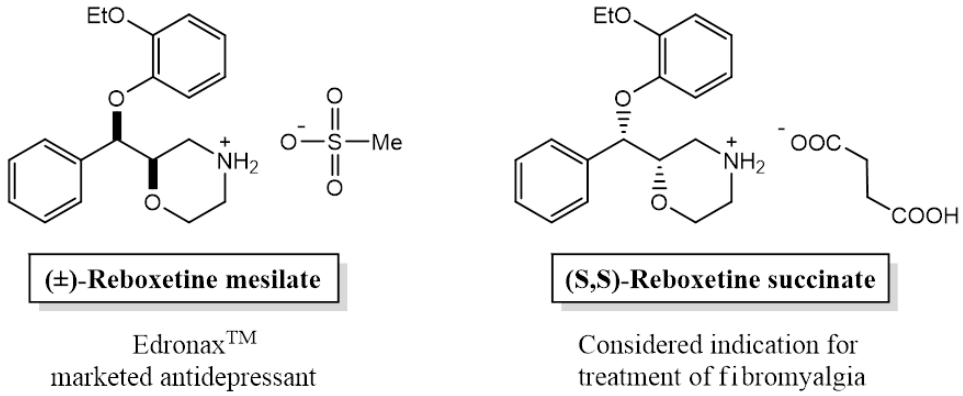


Figure 1.18: Pharmaceutically active compounds based on the Reboxetine core structure.[141]

Discovered in the first half of 1980s, the chiral compound Reboxetine (Figure 1.18) opened the way to an innovative class of antidepressants called Norepinephrine Reuptake Inhibitors (NRIs), and it has been marketed worldwide under the trade name of Edronax^{TM}. Over its patent-lifetime, the compound underwent various changes of ownership and market strategy as summarised in the timeline of Figure 1.19. Initially developed by FarmItalia Carlo Erba, the compound was then taken-over by Pharmacia in early nineties, and commercialised globally since.[145] In 1999, a failed attempt to access the United States market left Pharmacia with a high volume excess of racemic Reboxetine, which prompted to initiate the (S,S)-Reboxetine project. The project aimed to investigate the potential for expanding the scope of the enantiopure drug as a new indication for the treatment of fibromyalgia.[141]

As described in a publication by Assaf *et al.*, the project progressed successfully for

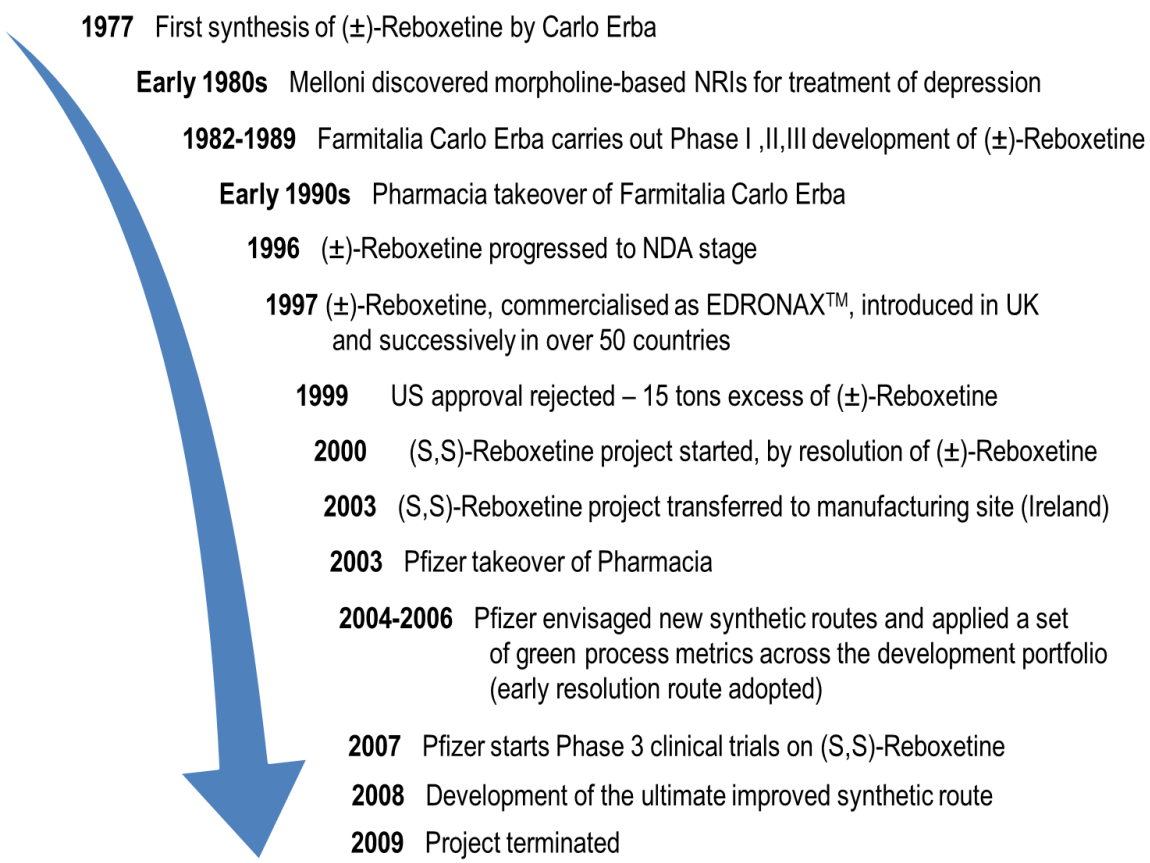


Figure 1.19: Historical timeline of Reboxetine, from the commercialisation of the racemate to the development of the single enantiomer.[141,145,146]

ten years, persisting despite acquisition of Pharmacia by Pfizer. At Phase III clinical trials in humans, the drug candidate proved enhanced activity and better profile than the corresponding (±)-racemate. From a synthetic perspective, the route was optimised and monitored via green chemistry metrics, leading to what was ultimately intended as route for manufacture (Figure 1.20). Noteworthy achievements relevant to the research described in this thesis include:[146]

- The stereochemistry of the product was introduced early stage in the synthesis and was retained throughout the preparation;

- The use of CALB to mono-acetylate the diol intermediate made for an efficient, catalytic, heterogeneous, regioselective and mild alternative to the previously protection strategy which introduced a trimethysilyl group in cryogenic conditions. The new method also eliminated the need for the use of low melting chlorinated solvents;

- Toluene was found as a suitable medium for the lipase-catalysed step, and also allowed telescoping of four consequent reactions with extraction of the final intermediate as water-soluble zwitterionic salt. Reportedly, this technique alone allowed reducing the solvent consumption of the route overall by 20%.

Despite the thorough optimisation and promising results from the clinical trials, (S,S)-Reboxetine was never taken to market. Subsequently, details of the project were published by Pfizer once the project was discontinued.[141,146–148] In conclusion, while the project was commercially abandoned, it represents a valuable example of sustainability-minded optimisation where the use of non-aqueous enzymatic catalysis was successfully employed.

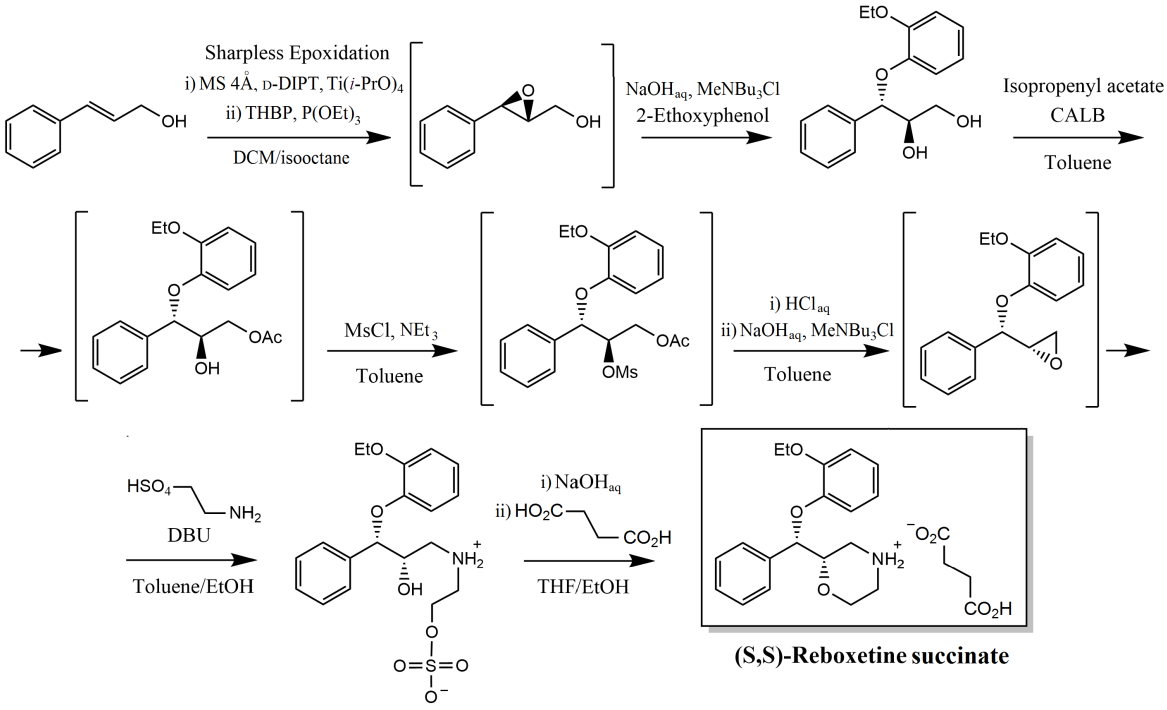


Figure 1.20: Synthetic route to (S,S)-Reboxetine intended for commercialisation.[141,147]

1.5 Scope of the work

The aim of the work presented in this thesis is to evaluate the potential for applicability of bio-based solvents as sustainable media for non-aqueous biocatalysed reactions. A main focus was given to esterifications catalysed by *Candida Antarctica* lipase type B (CALB).

In Chapter 2 is given an account of scoping experiments carried out in the first year of this project. These aimed to develop an initial understanding of the factors influencing lipase catalysis, and to help devise and test a new approach to evaluating the effect of bio-based solvents in enzymatic processes. This preliminary work set the direction for subsequent efforts which focused further into the solvent perspective.

Chapter 3 and 4 outline a project prompted by a CHEM21 collaboration with Pfizer aimed to provide proof-of-concept for the replacement of toluene with a bio-based alternative in the route towards a commercial pharmaceutical ingredient, (S,S)-Reboxetine. In addition, an in-depth kinetic and thermodynamic screening of solvent effects in a CALB-catalysed process was carried out to clarify the mechanism of solvent influence in such systems.

In a second phase, the research acquired an interest in the challenges that may hinder the implementation of bio-based solvents in industrial applications. Chapter 5 describes a survey of stakeholder perception on this topic carried out in the form of a participatory workshop. This involved experts from academia, industry and other organisations, and sought to reveal the key priorities that need to be tackled for allowing the wider uptake of green solvents in the chemical sector. Based on one of the factors emerged most prominently from this analysis, Chapter 6 investigated in detail a case-study surrounding the substitution of toluene with citrus waste-derived limonene from a supply and demand angle.

Finally, in Chapter 7 are summarised the key findings gathered from this work, and are given recommendations for future research in the area of bio-based solvents and especially in their study in biocatalysis. In the Appendixes additional details are provided, as well as a brief account of the results obtained from side projects with other enzymes.

Overall, the work presented in this thesis adopted a green chemistry stance, and sought to advance research in the field by aligning particularly with the following green chemistry principles:[11]

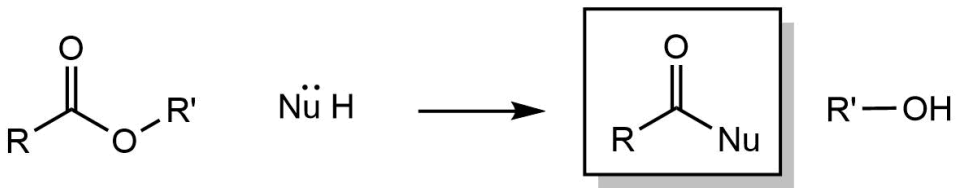
- Atom economy (Principle 2), by designing processes that avoid use of stoichiometrics additives, coupling partners, or activating agents, but instead employ a reusable and recoverable catalyst.
- Safer solvents and auxiliaries (Principle 5), by exploring alternatives to solvents that carry health and safety concerns;
- Use of renewable resources (Principle 7), by focusing on solvents that can - or could potentially - be derived from renewable resources and waste biomass;
- Catalysis (Principle 9), by investigating enzymatic catalysis in combination with bio-based solvents for the syntheses of chemicals.

Chapter 2

Scoping solvent effects in CALB catalysis

2.1 Introduction

At present, CALB is considered one of the most versatile commercially available enzymes and has been highly popular in chemical processing for decades.[149] As mentioned in Chapter 1, CALB is well known to catalyse acylations and hydrolytic reactions with a wide substrate scope. Figure 2.1 presents a general scheme of the transformations considered in this work. Herein, various esterifications and ester hydrolyses are described, whereas a summary of additional work on amide formation can be found in Appendix C.



R' = H or Alkyl chain
Nu = OH (Hydrolysis)
OR" (Esterification)
NR" (Amidation)

Figure 2.1: General scheme of typical transformations catalysed by CALB.

The commercial enzyme used in this thesis work was Novozyme435, an immobilised preparation where the enzyme is supported onto solid beads of acrylic resin. Novozyme435 is arguably the most widely studied, and most abundantly produced and used enzyme in chemical synthesis.[150] In order to confirm some of the fundamental literature regarding the use of this enzyme, this Chapter presents a number of experiments that were carried out to evaluate the scope of CALB in relation to its application in green solvents.

Furthermore, these screenings were used to investigate a new approach to the study of solvent effects. So far, as previously discussed, the use of single-variable linear regression methods has prevailed, providing some insight yet not fully clarifying the relationship between enzyme activity and various medium properties.[104,106,110,111,113,116] Herein, a multivariate approach was trialled aiming to develop a model able to explain and predict solvent effects in CALB-catalysed systems.

2.2 Influence of catalyst loading

The investigation started by evaluating the role of the catalyst in the context of a model esterification reaction for the formation of a fatty ester, hexyl dodecanoate, also called hexyl laurate (Figure 2.2).

Figure 2.2: Model esterification reaction used for the preliminary investigation into CALB Novozyme435 catalysis.

Six experiments were performed employing different amounts of Novozyme435 ranging from 0% to approximately 12.5 wt% of the substrate loading. The reactions were run at 2 mmol scale in heptane (500 mM), a solvent that had precedent for being employed in similar CALB catalysed esterifications.[90]

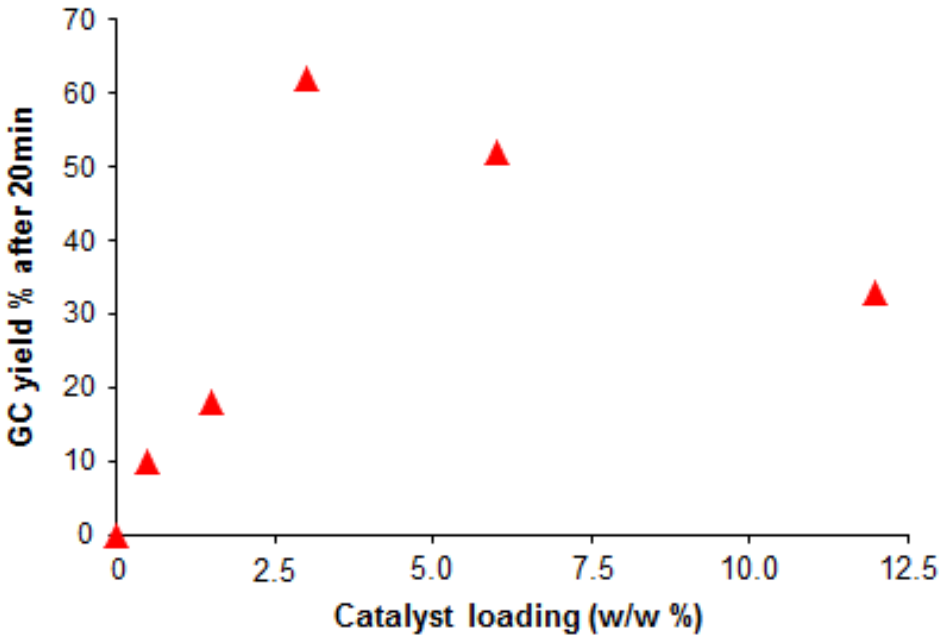


Figure 2.3: Effect of catalyst loading in the formation of hexyl laurate. Conditions: 500 mM lauric acid, 1 eq 1-hexanol, 4 ml heptane, CALB Novozyme435 in varying amounts, 40°C.

Figure 2.3 shows results after 20 minutes where yields increased drastically up to 2.5 wt% catalyst loading, peaking at 61% yield. Further increase in catalyst loading was found

to be counter-productive under these conditions, potentially due to the surplus of free active-sites more likely to bind the newly formed ester product and hydrolysing it straight-away.

In absence of the enzyme there was no sign of conversion after 20 minutes. After 8 hours, approximately 3% product yield was achieved, as shown in the background reaction profile of Figure 2.4a, proving that the enzyme is indeed crucial for allowing fast and significant conversions.

An additional important control test in biocatalytic studies consists in assessing whether the catalysis is purely enzymatic, or for instance contributed by the presence of active metal centres contained in the protein or other background catalysis. This is typically undertaken by purposely denaturating the protein so that it loses tridimensional architecture crucial for its mechanism of binding. In this state the enzyme is expected to be totally inactive and thus the reaction profile should resemble the one of the uncatalysed background reaction. After refluxing 20 mg of Novozyme435 in water for 30 minutes, the 'dead' catalyst was then dried and added to a solution of lauric acid and hexanol and monitored for activity. Figure 2.4b shows no evidence of residual catalytic activity and a profile equivalent to the uncatalysed reaction (Figure 2.4a). This result demonstrates that the ability of the CALB Novozyme435 to efficiently catalyse the formation of esters is purely dependent on its tertiary conformation and not contributed in any way by its acrylic resin support or other potential chemical components.

2.3 Influence of solvent

Secondly, the study focused on evaluating the effect of solvents on various aspects of the catalysed process, such as on:

- Reaction rate;
- Reaction equilibrium;
- Reusability of the catalyst;
- Inverse hydrolysis rate;
- Substrate scope.

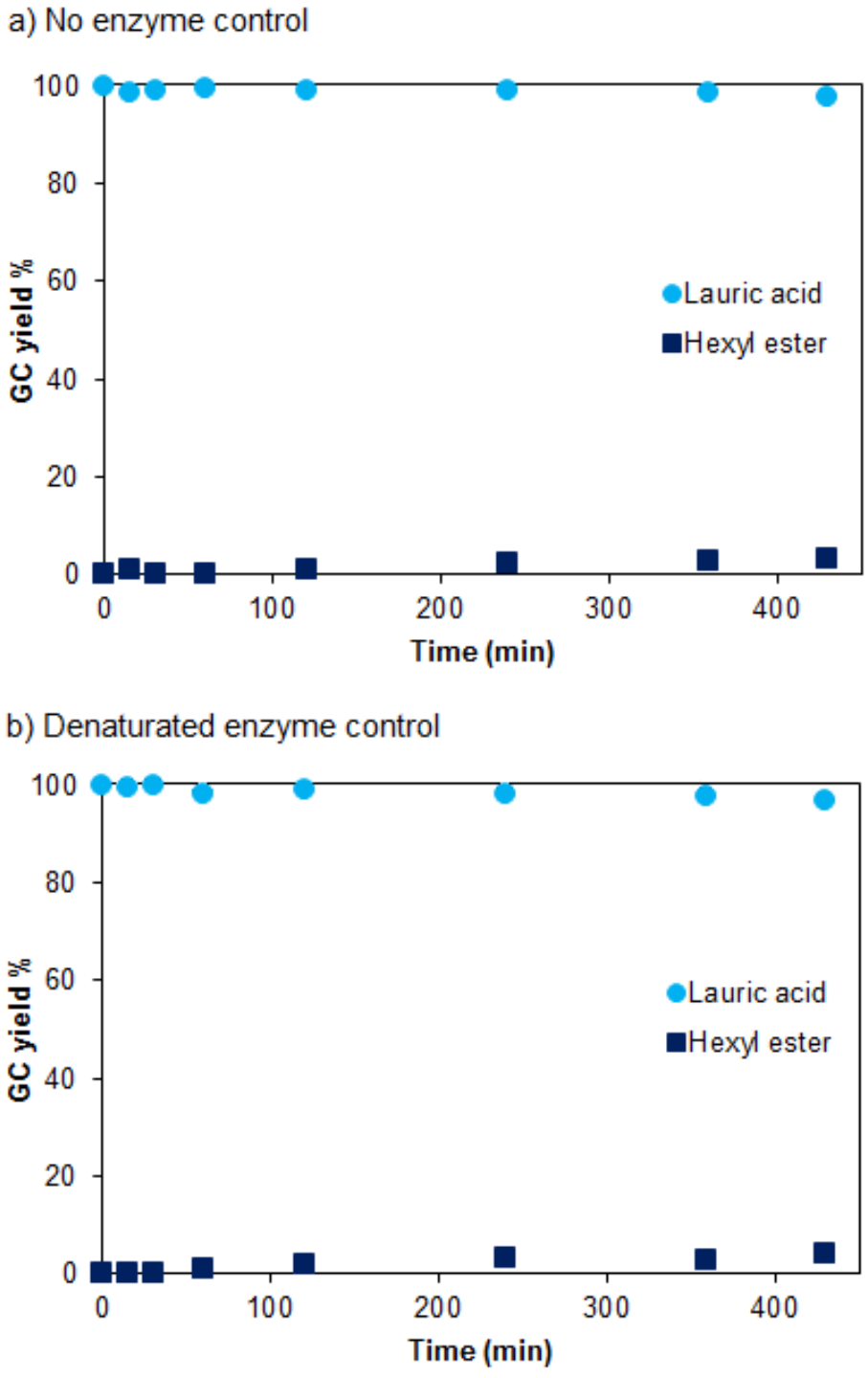


Figure 2.4: Control tests for background catalysis: a) in absence of catalysts, and b) in presence of denaturated enzyme (20mg). Conditions: 500 mM lauric acid, 1 eq 1-hexanol, 4 ml heptane, 40°C.

2.3.1 Reaction profiles in different solvents

The model esterification of lauric acid with optimised catalyst loading was screened in water and in three aprotic solvents of different nature. These were selected to include potentially bio-derived solvents with significantly different logP values, according to literature guidelines:[104]

- *p*-Cymene, logP = 3.47, potentially derived from citrus waste or directly extracted from eucaliptus leaves,[44]
- MTBE, logP = 1.02, potentially derived from bio-isobutene and bio-methanol yet carrying high concerns over reprotoxicity,[151,152]
- Acetone, logP = -0.24, potentially obtained via the Acetone-Butanol-Ethanol (ABE) fermentative process.[153]

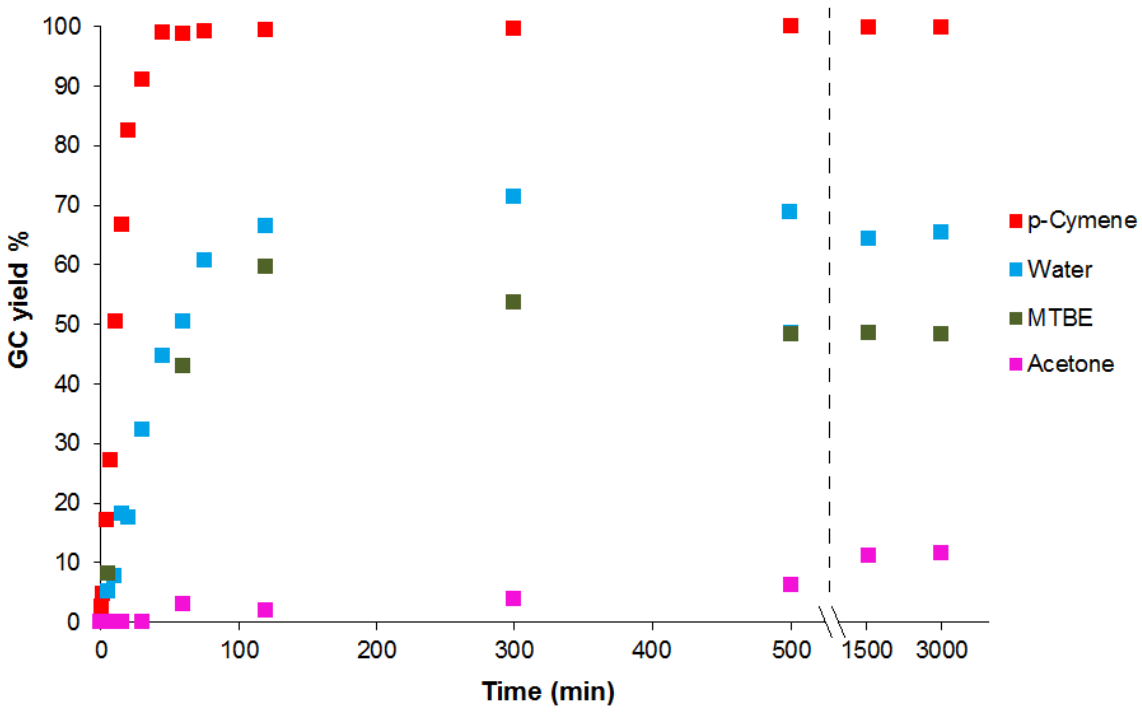


Figure 2.5: Solvent effects on the reaction course and equilibrium monitored over a period of 2 days. Conditions: 500 mM lauric acid, 1 eq 1-hexanol, 4 ml solvent, 20 mg CALB Novozyme435, 40°C.

Analytical samples were collected at short intervals in the first two hours of reaction to capture differences in initial rate of reaction. Reactions were then left to proceed for

50 hours in order to reveal solvent effects imposed on the final equilibrium. Overall, as shown in Figure 2.5, dramatic changes in both the initial rate of reaction and final equilibrium point were observed depending on the solvent utilised. The most hydrophobic solvent, *p*-cymene, gave 100% yield in under 1 hour, whereas the same reaction in the most hydrophilic solvent, acetone, resulted in only 13% yield, leaving starting materials mostly unreacted. This trend is in good correlation with literature precedent.[104]

According to this trend, water was expected to perform poorly, readily reverting to the more favoured hydrolytic product under the large excess of water. Instead, ester formation was observed in surprisingly good yields. It is unclear whether this effect is genuine or is a consequence of the low reagent and product solubility in water skewing the quantification of the samples. In fact, as it is noticeable from Figure 2.6, the reaction system in water is visibly highly heterogeneous as opposed to the clear and homogeneous solution obtained in *p*-cymene.

Figure 2.6: Picture of the reaction system of study in water (left), and *p*-cymene (right), exemplifying the solubility issue.

The reaction profile observed in MTBE showed moderate performance, in keeping with its polarity as compared to the other two organic solvents. In this case, product yield peaked at 120 minutes with 60% yield, then exhibited a slow decline that stabilised at approximately 50% yield. This fact suggests hydrolytic consumption of the product taking

place, although at a slower rate than the ester formation.

2.3.2 Effect of water

In order to gain further insight into the effects seen in Figure 2.5, two hypotheses were considered and tested. The first hypothesis contemplated the possibility of product-induced inhibition on the catalyst. In this case, if the solvent reached its capacity to extract and disperse the ester away from the active site, the reaction would stop regardless of any further addition of reactants. On the other hand, the second hypothesis considered the presence of an equilibrium between esterification and the inverse hydrolytic reaction. If this equilibrium were solvent-dependent, it could be expected that a solvent favouring the esterification would be highly unfavourable for the hydrolysis. In order to demonstrate which interpretation best explained the system, two sets of experiments were designed.

First, the screening of Figure 2.5 was repeated in the three organic solvents with 1 equivalent of lauric acid and a large excess of 1-hexanol, 10 equivalents. Once a stable equilibrium point had been achieved (at 100 minutes) an additional equivalent of lauric acid was added to the mixture. Reaction profiles pre- and post-addition of fresh reactants are presented in Figure 2.7. The two profiles appear unchanged, both in terms of initial rate and of relative final yield. This result demonstrates that 1) the observed equilibrium effects are not driven by product-inhibition, and 2) the enzyme can be effectively reused *in situ* with no observable loss of activity.

Second, screening of inverse hydrolysis reactions in the three solvents were then carried out to verify the second hypothesis. Hexyl laurate was employed as a reactant in the presence of 2 equivalents of water in pre-dried *p*-cymene, MTBE, and acetone using molecular sieves (except water). Figure 2.7 shows an inverse trend from that observed in the formation of the ester. In this instance, acetone gave the fastest hydrolysis with nearly complete conversion, while *p*-cymene was low. Acetone, being more hydrophilic, is likely to carry and diffuse water molecules more efficiently than *p*-cymene, and thus make water molecules more readily available in the active-site. Conversely, *p*-cymene may phase water out of the reaction system, potentially to the surface, this way hindering hydrolysis of the

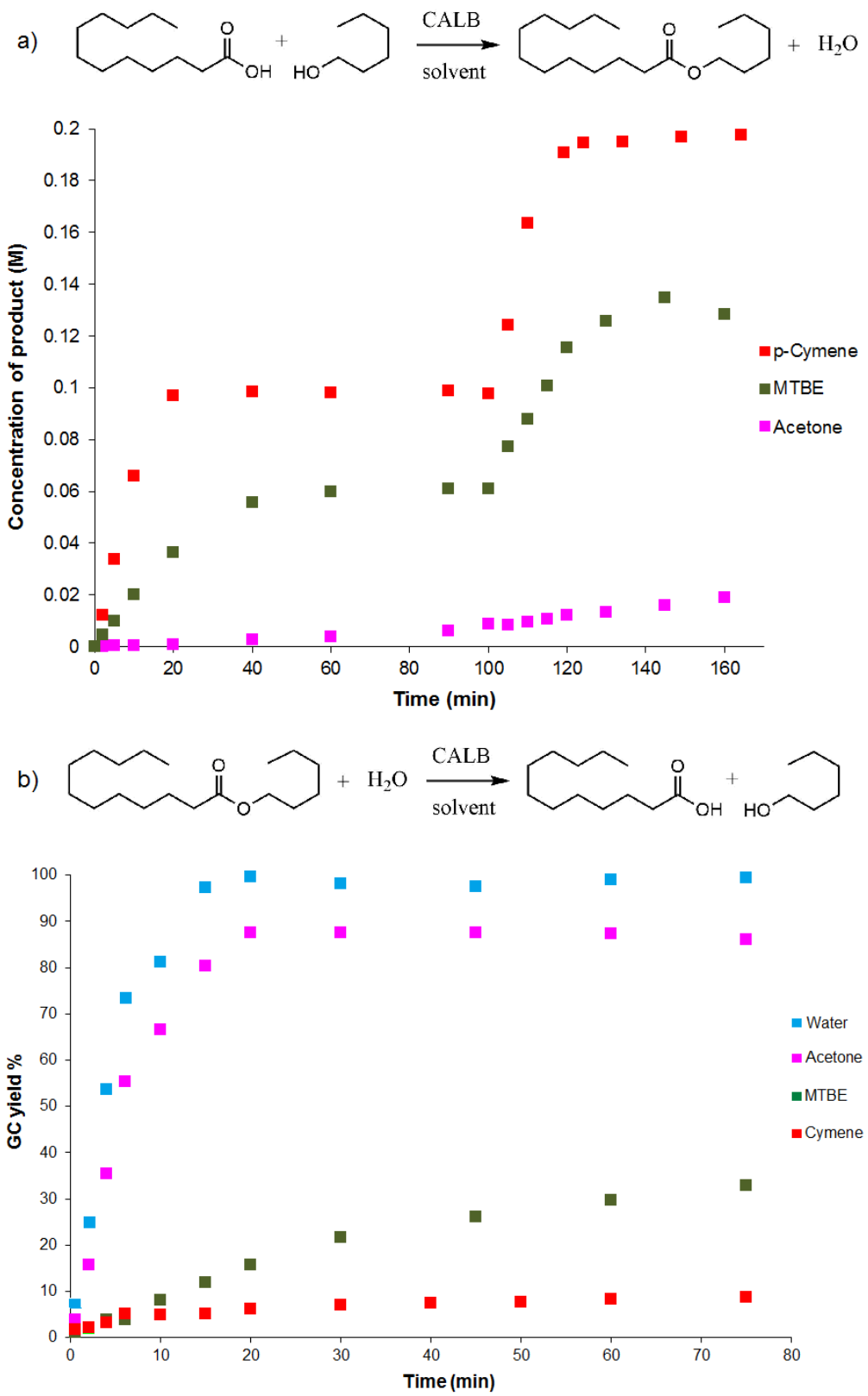


Figure 2.7: Testing hypotheses to explain solvent effects on equilibria: a) Activity retention after addition of fresh reactant rules out the hypothesis of product inhibition; b) Inverse solvent profiles in hydrolysis reaction suggest solvent-induced equilibria driven by water availability.

ester to occur. A comparison of the MTBE profile between the esterification and hydrolysis plots reveals that both transformations are significant. This suggests that the ester is produced faster while hydrolysis reaction becomes significant only in a later stage of the reaction, thus explaining the decline of yield at longer reaction times.

2.3.3 Solvent effects on different substrates

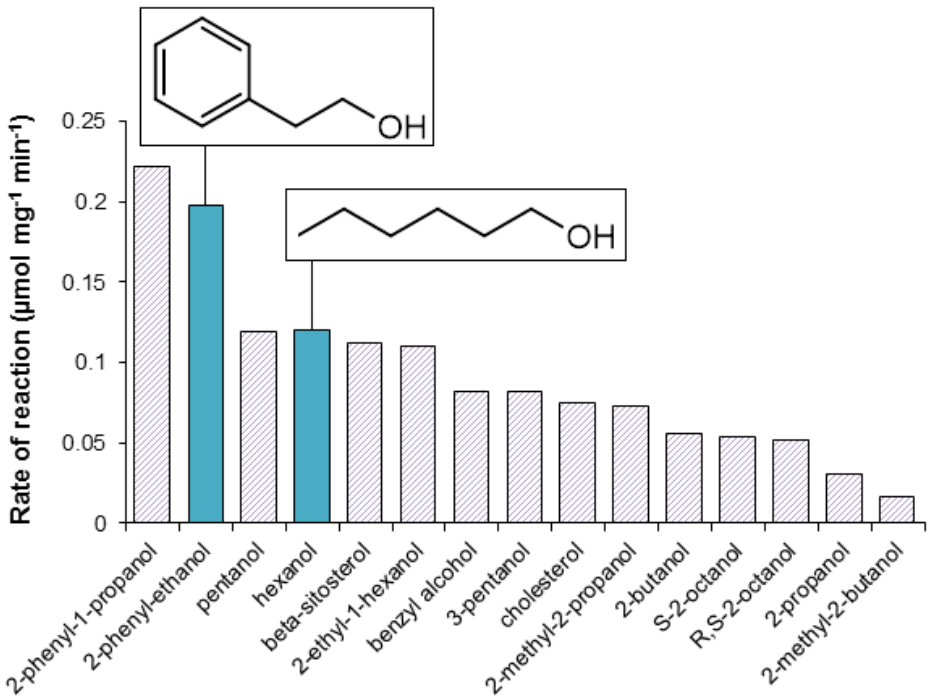


Figure 2.8: Literature precedent of substrate scoping with immobilised *Candida Antarctica* lipase in the formation of lauric esters in heptane. In blue those selected for screening. Figure adapted from reference.[90]

A study by Kirk and Christensen reported a broad investigation into the substrate specificity of lipases from *Candida Antarctica*.[90] An adapted extract from this publication is shown in Figure 2.8, where the initial rates of reaction normalised to the mass of catalyst are compared across a varied set of alcohols used in the preparation of lauric esters.[90] According to this work, reacting lauric acid with 2-phenyl-ethanol - also called phenethyl alcohol - would be significantly faster than the same reaction with 1-hexanol. Both phenethyl and hexyl esters are considered valuable ingredients in personal care and cosmetic applications.[154,155] With an interest to test these findings and to assess the solvent influence on them, the esterification of phenethyl and hexyl alcohols with lauric acid were screened in

ten organic solvents and water, and followed by GC. Figure 2.9 stacks the two sets of results gathered in the first 40 minutes of reaction.

In both cases, the highest yields and fastest rates of phenethyl laurate were obtained in hydrocarbon solvents. Notably, citrus-derivable *p*-cymene exhibited 100% yield in both reactions. Across the set, reactions appeared faster with phenethyl alcohol than with hexanol, in keeping with the literature.[90] Moderate performance was registered in water for both alcohols - although the same quantification issue described earlier was faced. In the same yield range were found ethers such as cyclopentylmethylether (CPME) and MTBE. From this category of solvents, 2-MeTHF gave the lowest performance, with a reaction profile that resembled that of the polar aprotic solvents such as chloroform and dimethylcarbonate (DMC). Finally, ethyl acetate showed dominant formation of ethyl laurate byproduct and no evidence of the desired ester.

In conclusion, a differentiation between the three major classes of solvents tested - *i.e.* hydrocarbons, ethers, and polar aprotic solvents - was observed with both substrates. This appears accentuated with hexanol as a substrate, potentially due to the lower initial rates of reaction allowing for greater diversification of effects. Overall, the solvent effects and trends appear consistent between the two screenings with regards to most solvents. The results broadly support the observations made by Kirk and Christensen with regards to the higher rate of formation of phenethyl ester,[90] yet indicate some level of variability in relation to some solvents. In particular, the largest relative variations were observed with 2-MeTHF, and limonene. Further research efforts were focused on the interpretation of solvent effects as opposed to the study of substrate-dependency.

2.4 Analysis of solvent effects on rate of reaction

The results from the previous Section prompted further investigation into the primary origin of the observed solvent effects. With this aim, the results of Figure 2.9b were further populated with additional solvents and initial rates calculated and analysed by means of linear regression. Since the formation of hexyl laurate occurred more slowly in the condi-

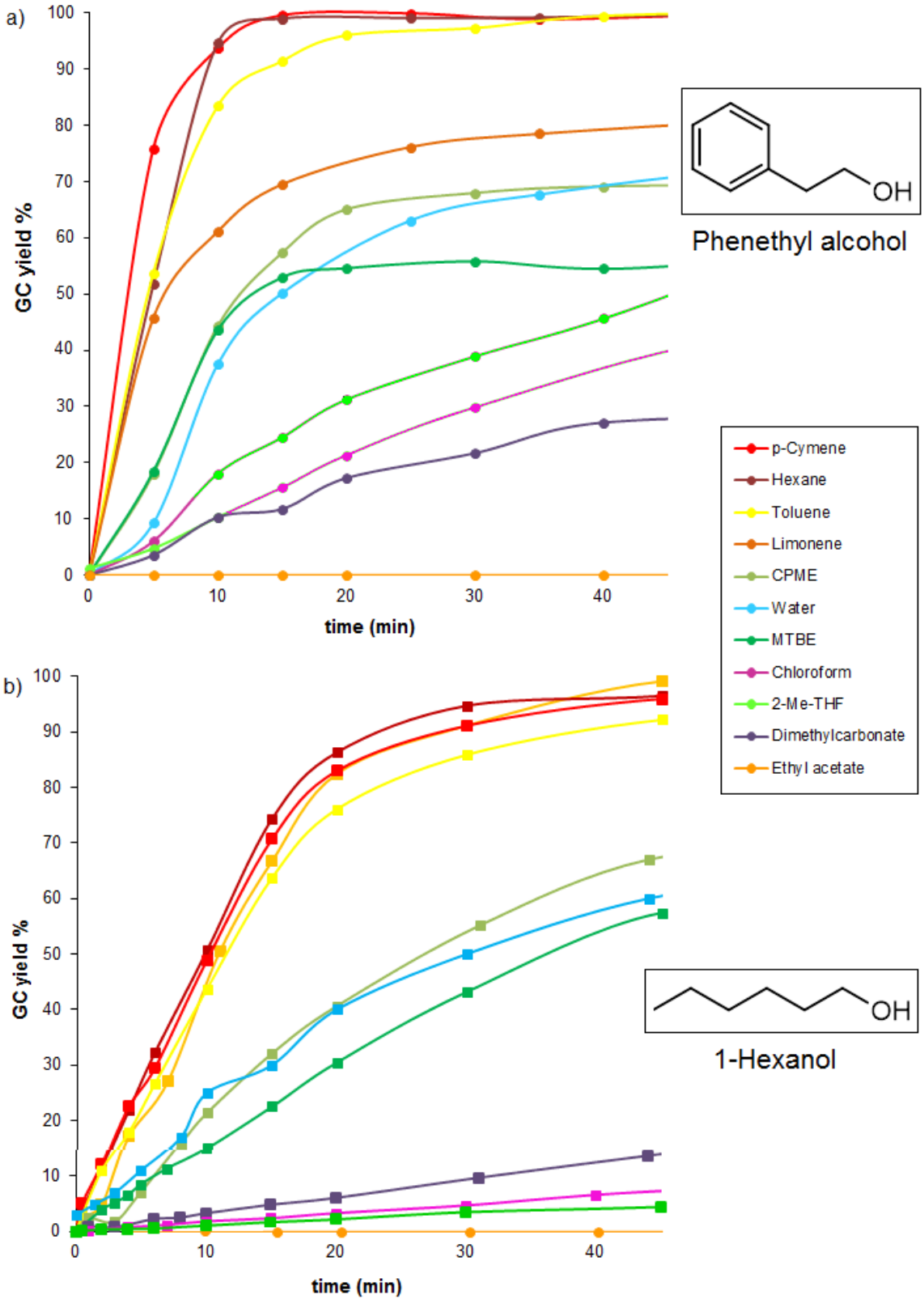


Figure 2.9: Solvent screening in the formation of lauric esters: a) Yield of phenethyl laurate, b) Yield of hexyl laurate. Conditions: 100 mM lauric acid, 1 eq 1-hexanol, 10 ml solvent, 10 mg CALB Novozyme435, 40°C.

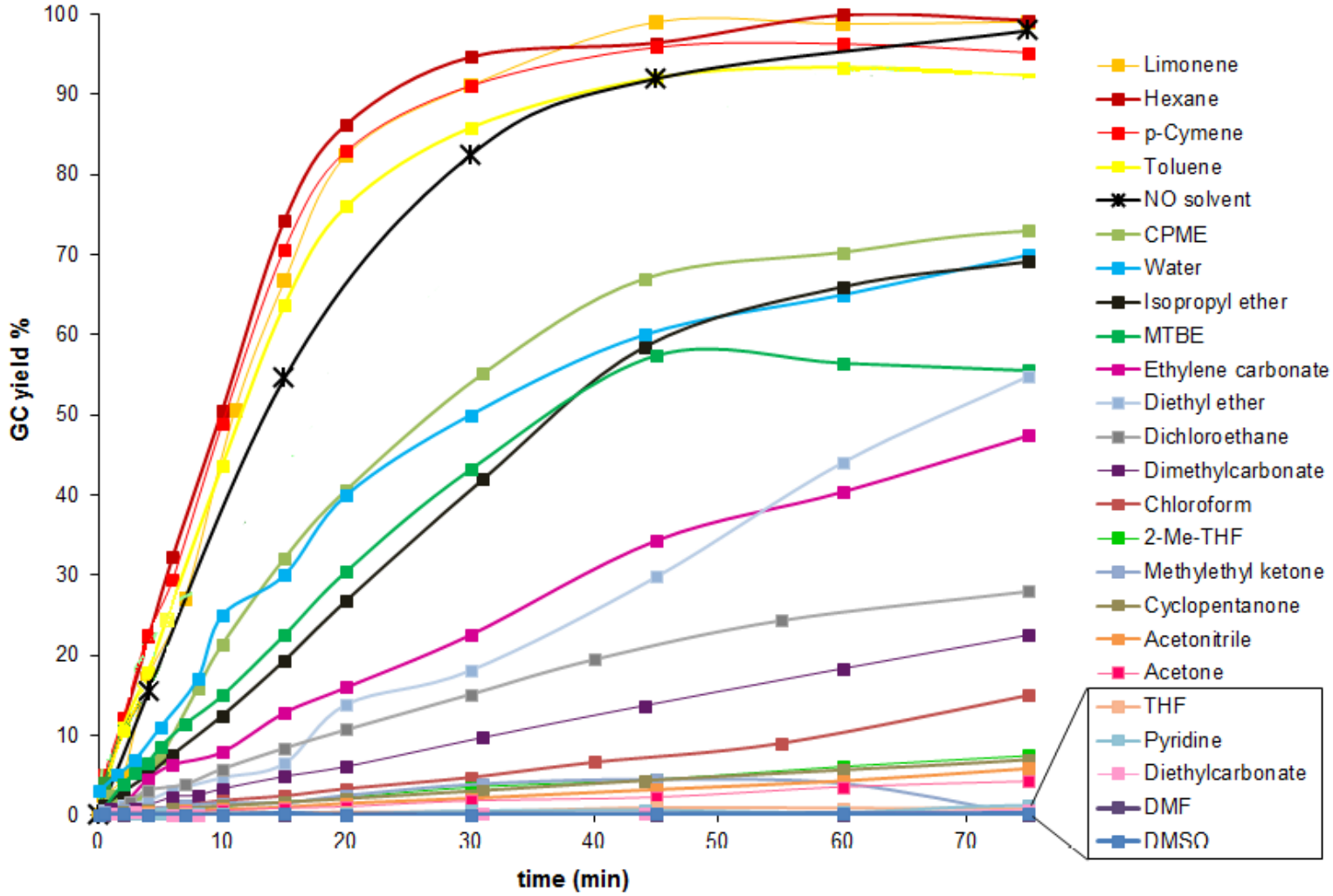


Figure 2.10: Broad kinetic screening for solvent effects in the formation of hexyl laurate. Conditions: 100 mM lauric acid, 1 eq 1-hexanol, 4 ml solvent, 20 mg CALB Novozyme435, 40°C.

tions of study, it was seen as better suited for accurate determination of initial reaction rates. Efforts focused on filling the gaps between solvent groups identified in the early screening (Figure 2.9b) by employing bio-based solvents where possible. In order to achieve this, logP was used as a guide, based on the correlation described by Laane *et al.*[104] Figure 2.10 shows the full set of reaction profiles obtained by screening a total of 23 solvents.

This broader screening emphasised the diverse range of effects previously described. Notably, a number of solvents allowed to reach full conversion while others seemed to plateau at an earlier equilibrium point, *e.g.* water, isopropylether, MTBE, CPME, and dichloroethane (DCE). The reaction was also run in solvent-free conditions applying an excess of hexanol. Arguably, this example may be the greenest option for the synthesis of hexyl laurate, as it exhibits complete conversion after 75 minutes without the need of an additional auxiliary - the solvent. However, a large excess of alcohol substrate was required to ensure the reaction mixture was both homogeneous and of sufficient volume to allow agitation. A quantitative and qualitative green metrics comparison of this option compared the reactions in solvents is presented at the end of this Chapter.

2.4.1 Initial rate *versus* log P

In the analysis of the kinetic screening shown in Figure 2.10, the measurement of initial reaction rates was approached via a gradient method applied on the concentration of starting material. The declining plot of lauric acid concentration against time was fitted with a line and the resulting x coefficient was used as an indication of the initial rate, with units of mmol min^{-1}. The method is exemplified in Figure 2.11 for the case of *p*-cymene. All initial rates values thereby obtained can be found in Table 2.1.

As mentioned above, solvents were selected with the aim to provide a good coverage of the logP space, with values ranging from -1.45 for DMSO up to +4.57 for limonene. In Figure 2.12 are illustrated the initial rates of reaction in relation to each logP value. The obtained distribution resembles the one reported by Laane *et al.*: S-shaped, with a clear cut-off point around logP = 2. So far, this kind of representation have constituted one of the best and most used criteria for solvent selection in biocatalysed processes. While it

Table 2.1: Results from kinetic screening of solvent effects in hexyl laurate formation. LogP used as a guiding parameter in solvent selection.

Solvent	Yield %	Initial rate (mM min^{-1})	ln(rate)	logP
Hexane	99.8	4.2	1.44	3.9
p-Cymene	97.3	4.3	1.46	3.47
Limonene	98.7	3.7	1.31	4.57
CPME	72.9	3.7	1.31	1.22
Water	69.9	3.6	1.28	n.d.
Toluene	93.7	3.2	1.16	2.58
Diisopropyl ether	67.5	1.5	0.41	1.52
MTBE	53.2	1.2	0.18	1.02
Dichloroethane	27.9	0.7	-0.36	1.58
Diethylether	50.0	0.6	-0.51	0.89
2-MeTHF	8.6	0.4	-0.92	1.25
Dimethyl carbonate	20.8	0.4	-0.92	0.24
Ethylene carbonate	61.5	0.4	-0.92	-0.23
Chloroform	33.5	0.2	-1.61	1.83
Methylethylketone	4.0	0.1	-2.30	0.3
Cyclopentanone	10.2	0.1	-2.30	0.89
Acetonitrile	5.9	0.1	-2.30	-0.33
Pyridine	1.2	<0.1	-2.66	0.68
Acetone	4.3	<0.1	-2.66	-0.24
DMF	0.1	<0.1	-3.22	-0.97
THF	0.8	<0.1	-3.51	0.45
Diethyl carbonate	0.6	<0.1	-3.91	1.21
DMSO	0.2	<0.1	-4.61	-1.45

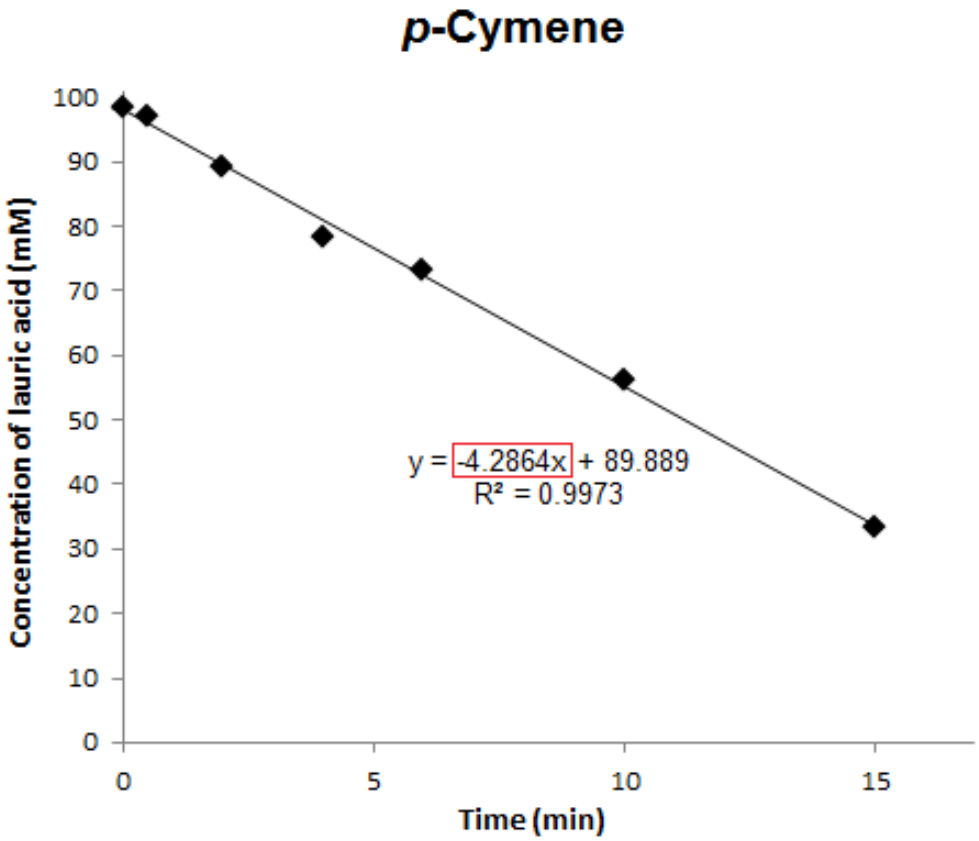


Figure 2.11: Method used for determination of initial rate constant.

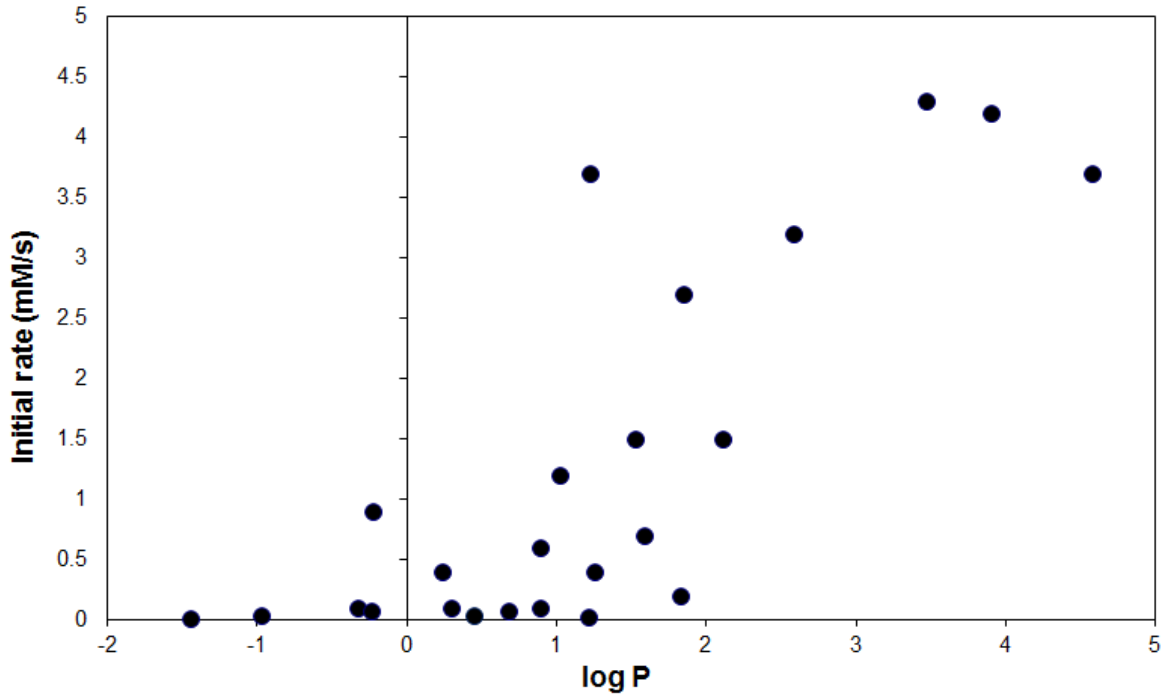


Figure 2.12: Distribution of initial reaction rates against logP.

clearly is a valuable, simple and quick method for discerning likely suitable solvents (logP > 2) from poor ones (logP < 2), it lacks a real correlation. In particular, logP appears to be a poor descriptor for medium-to-low performing systems.

2.4.2 Correlation with solvent parameters

Aiming to clarify the apparent significance of logP in describing the system and potentially unveil better descriptors, a multivariate regression analysis was pursued where the following parameters were considered:

1. Partitioning coefficient (logP);
2. Kamlet-Taft solvatochromic parameters (α, β, π^*);
3. Hildebrand's solubility parameter (δ^2);
4. Molar volume (V_m);
5. Water content (ln[H_2O]).

The natural logarithm of the measured initial rate of reaction ($ln(rate)$) was used as dependent variable (y) in the analysis. The rates from 17 out of the 23 solvents screened were employed, thus excluding water - not being an organic solvent - and 5 organic solvents with initial rates below 0.1 mmol min^{-1} from Table 2.1, as their inclusion was found to strongly distort the dataset trends.

The ability of each parameter to describe changes in ln(rate) was first evaluated through individual plots as presented in Figure 2.13 and 2.14. Figure 2.13 illustrates the relationship between ln(rate) and logP discussed earlier, from which a roughly linear trend emerged (R^2 = 0.6109).

Similarly, were traced the other six relationships with individual solvent parameters as represented in Figure 2.14. From this analysis based on the Partial Least Squares method, the least correlated variables of the set, with R^2 < 0.30, appear to be Kamlet-Taft parameter α and Hildebrand's parameter δ^2. This suggests they have no significant effect on the system. On the other hand, the most promising variables appear to be Kamlet-Taft parameter β and π^*, as well as molar volume V_m, with R^2 > 0.50. The latter alone provided an equivalent degree of correlation to logP. Moving into the multivariable regression from

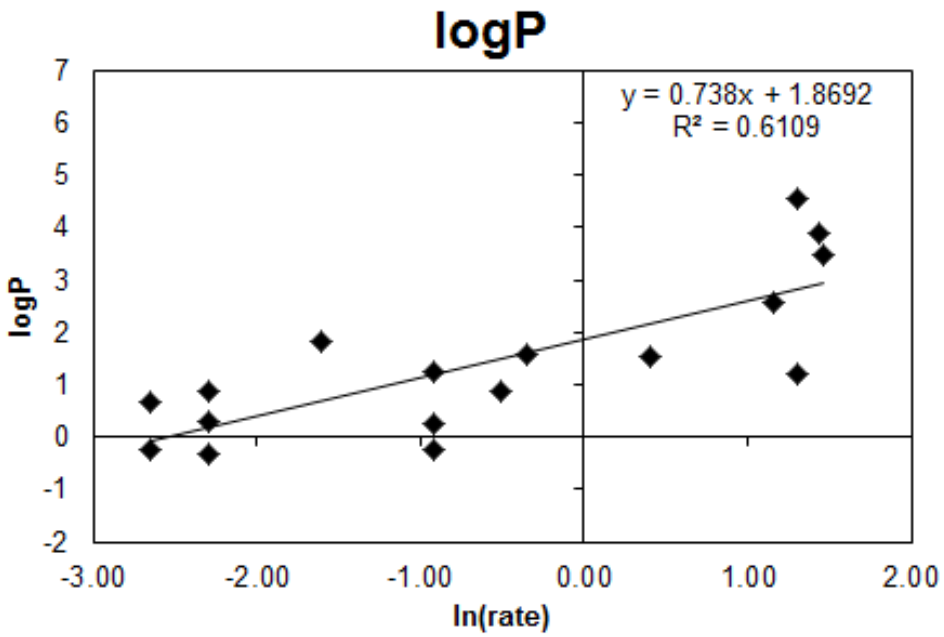


Figure 2.13: Testing for linear correlation between ln(rate) and logP.

this preliminary test, it can be expected that β, π^* and V_m will appear in the descriptive regression equation.

2.4.3 Multivariable regression

By using Excel data regression feature (ANOVA), it was possible to run a number of iterative analyses incorporating different combinations of variables. Each output was judged on the basis of the following statistical parameters:

- Adjusted R^2, accounting for least squares across multiple variables - ideally $R^2 = 1$;
- $\frac{F_{value}}{F_{statistic}}$, representing the quality of the multivariate correlation in describing variance across the set - in a good iteration F value is significantly higher than F statistic;
- P-value, describing the probability of false correlation associated with each variable - ideally below 0.05, corresponding to 95% confidence in effective correlation.

Statistical parameters produced in the analysis at each iteration are reported in Table 2.2.

The first iteration encompassed all six variables considered in the study - excluding logP used as benchmark. While R^2 and F-test gave favourable results, all individual P-values of each variable failed the F-test threshold thus indicating high risk of random

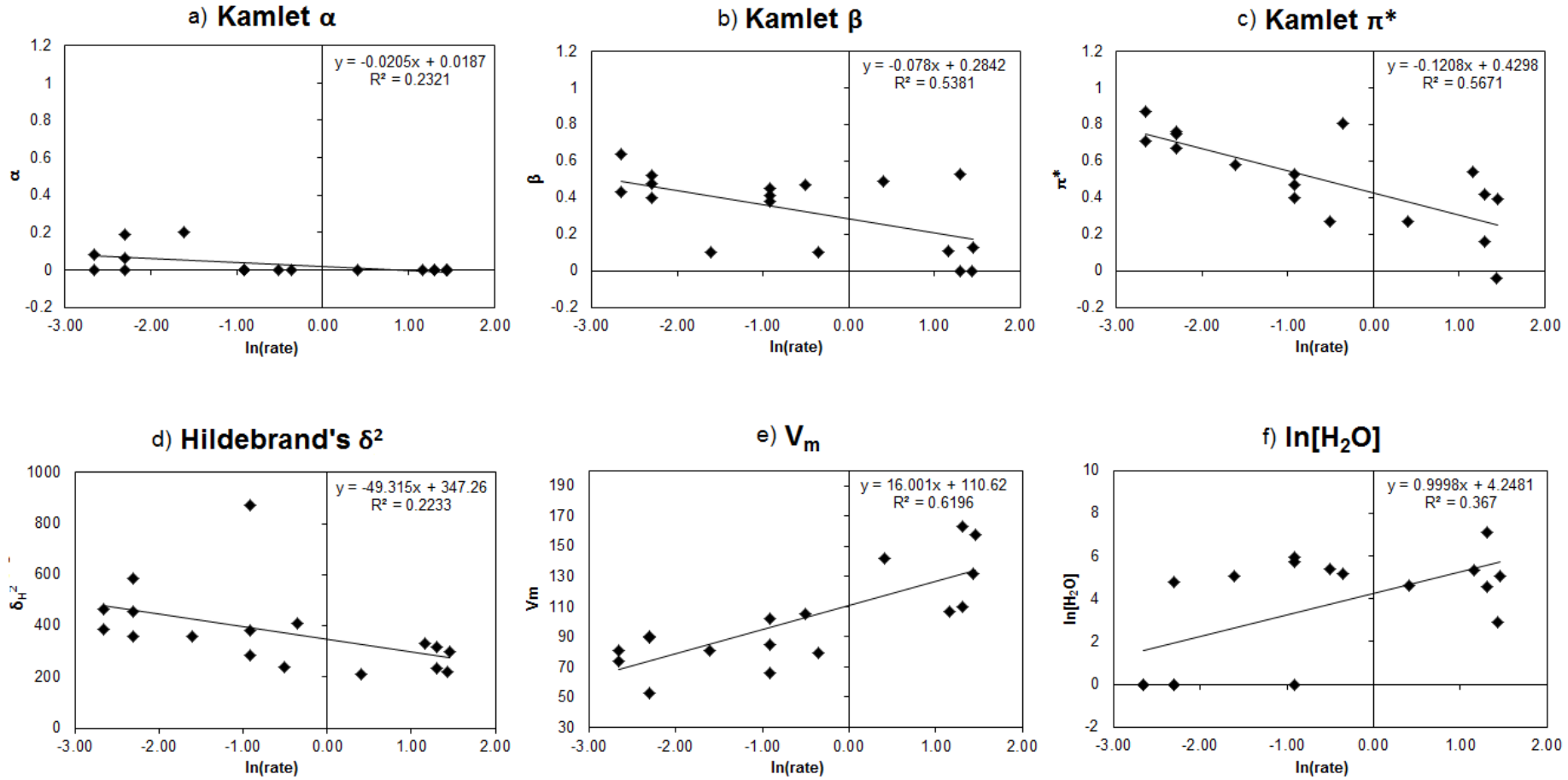


Figure 2.14: Testing for linear correlation between ln(rate) and each individual solvent parameter considered in the study.

Table 2.2: Regression statistics obtained at each iteration encompassing different combinations of variables. Iteration 6 based on a set of 14 solvent entries out of the 17 contemplated in the other iterations.

	P-values						F-test	Least squares
Iter.	α	β	π^*	V_m	δ^2	$ln[H_2O]$	F/F_{stat}	Adj. R^2
1	0.137	0.065	0.278	0.317	0.884	0.470	1.66E+03	0.8052
2	0.130	0.051	0.186	0.135	-	-	2.41E+04	0.7915
3	0.124	0.031	-	0.022	-	-	3.61E+04	0.7468
4	-	**0.086**	-	**0.001**	-	-	**6.36E+04**	**0.6942**
5	-	-	-	0.031	-	-	5.12E+04	0.6196
6	-	0.0001	-	0.0002	-	-	2.34E+05	0.8929

association - here likely due to over-correlation. By using P-value as the discriminative criteria, the variables with the highest P-value were discarded and precluded from the following iterations - *i.e.* δ^2 and ln[H_2O]. Iteration 2 was then run with the four remaining variables and the same rationale was pursued until all statistical parameters were found satisfactory. This procedure identified β and V_m in iteration 4 (Table 2.2) as the most representative parameters for this set, as expressed in equation 2.4.1. Further withdrawal of β from the regression showed to be detrimental on the statistic parameters.

$$ln(rate) = -3.13 - 2.25\beta + 0.032V_m \tag{2.4.1}$$

$$R^2 = 0.6942$$

Seeking to improve the model, the predicted rates - as calculated using equation 2.4.1 - were plotted against experimentally observed values with the aim to identify individual data-points that were potentially misrepresented in the model. Figure 2.15 shows the correlation between the two sets and highlights the 95% confidence band applied to the fitting line. As can be noticed from Figure 2.15, three points were found to lie outside of the 95% confidence band. This fact prompted for the exclusion of the corresponding solvent entries from the model - *i.e.* hexane, CPME, ethylene carbonate - as part of a new iteration of the analysis. Iteration 6 (Table 2.2) resulted in significant improvement over previous

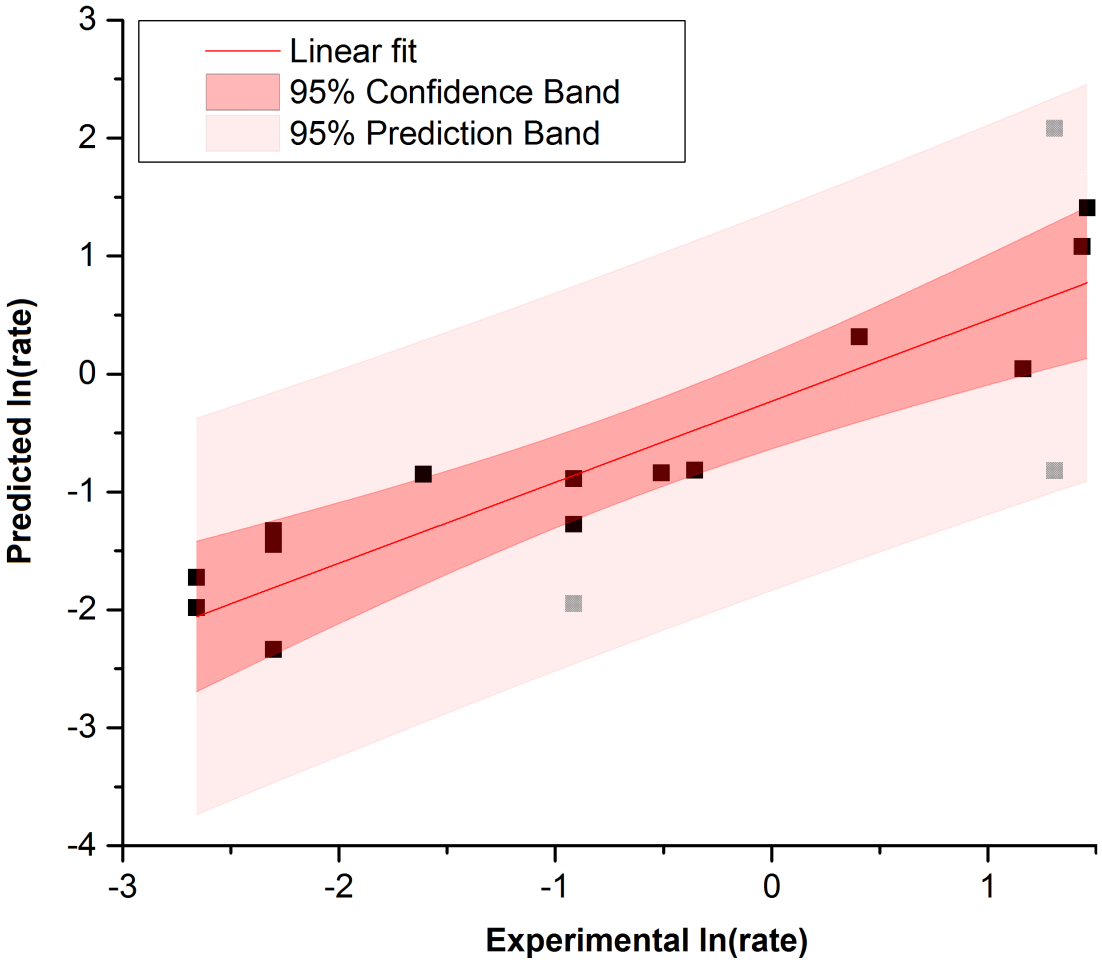


Figure 2.15: Confidence test used to identify data points that poorly fit with the model.

calculations: P-values well below 0.05, while F-value and F-statistic as well as R^2 values were the highest of all iterations. The coefficients produced through iteration 6 gave model equation 2.4.2. Similarly to equation 2.4.1, β and V_m display opposite contributions in sign to ln(rate).

$$ln(rate) = -3.70 - 4.53\beta + 0.044V_m \tag{2.4.2}$$

$$R^2 = 0.8929$$

Finally, the ability of this improved model to describe and predict reaction rates was compared against the traditional logP-based model. Figure 2.16 presents the overlap of the two models and the corresponding fitting statistics. Pleasingly, the newly devised model shows superiority over the one based solely on logP, with approximately 10% increase in

R^2. Moreover, the slope and intercept of the fitting line present values closer to the ideal values of 1 and 0 respectively, which would identify the perfect bisect. This indicates that β and V_m are better descriptors than logP for this system, thus representing an improvement from the state of the art in the interpretation of solvent effects in biocatalysis.

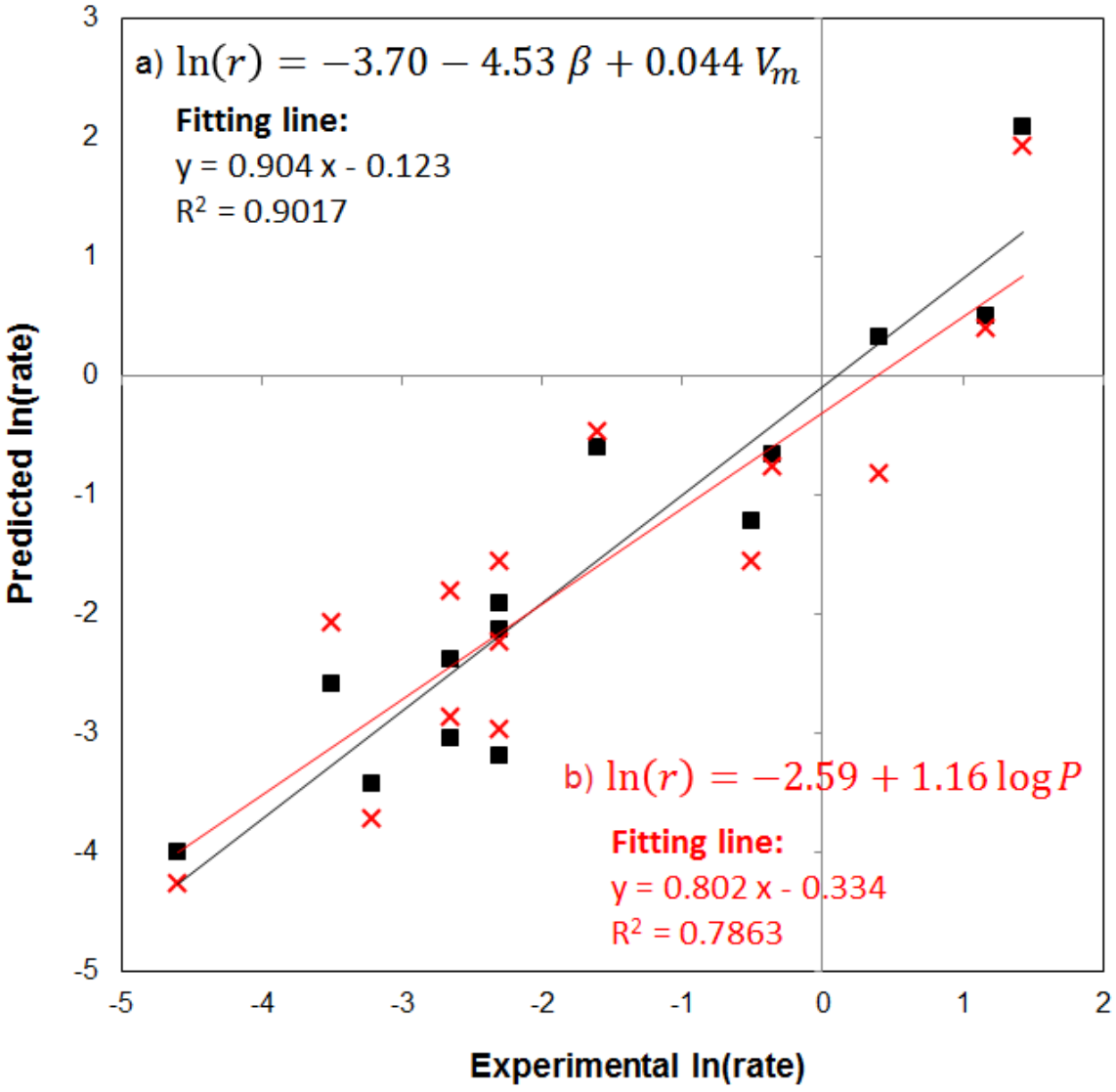


Figure 2.16: Model validation comparing experimental rates against predicted rates for the two model equations: a) iteration 6 based on β and V_m (×), and b) benchmark logP correlation (■).

2.5 Green considerations

In a recent publication produced by the CHEM21 consortium, a methodology was developed to combine the outcomes of multiple independent solvent assessments into one, and to score emerging green solvents previously neglected.[26] This Section aims to evaluate the results of the solvent screening described in this Chapter against such green process criteria devise by the CHEM21 consortium.[156] Table 2.3 provides an account of the CHEM21 EHS scoring and green ranking of the solvents employed in the screening - the lower the scoring the greener the solvent. Besides, the performance of each solvent in the reaction of study is reported and also ranked according to the CHEM21 process metrics - above 90% considered ideal.[26,156]

By stacking greenness and performance near one another, it becomes evident that, disappointingly, all solvents ranked as 'Recommended' had given unsatisfactory yields, thus negating the benefit of their use. Out of the overall set, only four solvents achieved a green yield flag, *i.e.* hexane (99.3% yield), limonene (99.1% yield), *p*-cymene (95.2% yield), toluene (92.5% yield). In order to facilitate the selection of the best solvent candidate according to the process metrics and aid discussion over their individual sustainability impact, all EHS scores were plotted in Figure 2.17 in the form of stacked bars.

At the top of Figure 2.17 are highlighted the four solvents mentioned above. It is evident from this representation that hexane is highly undesirable on all aspects of its green assessment, thus ranked as 'red' solvent. On the other hand, the remaining four solvents from the short-list show better overall scoring with no significant difference on the overall score - 'amber' ranking.

In terms of environmental impact, hexane is amongst the highest actoss the whole set due to 1) its harm to organs and suspected reprotoxicity, 2) its high vapour pressure increasing the risks for health and safety and atmospheric release, and 3) its high toxicity to aquatic life.[58] In particular, the high volatility of this solvent increases the chances of human exposure to this toxic substance as well as the risks of fires and vapour explosion because of the inherent very high flammability and low flash point (-26°C).[58] As for the other solvents,

Table 2.3: Solvent greenness scores according to CHEM21 assessment.[26,156] (*) Solvents not ranked within the CHEM21 guide for which the EHS scores were assigned based on a near neighbour solvent.

Solvent	Safety	Health	Envir.	CHEM21 ranking	Yield% (75min)
Hexane	8	7	7	Hazardous	99.3
p-Cymene	4	5	5	Problematic	95.2
Limonene	4	2	7	Problematic	99.1
Toluene	5	6	3	Problematic	92.5
CPME	7	2	5	Problematic	73.0
Diisopropylether	9	3	5	Hazardous	69.2
MTBE	8	3	5	Hazardous	53.3
DCE	4	10	3	Highly hazardous	28.0
Diethylether	10	3	7	Hazardous	54.8
2-MeTHF	6	5	3	Problematic	7.4
Dimethyl carb.	4	1	3	Recommended	22.5
Ethylene carb.	1	2	7	Problematic	47.4
Chloroform	2	7	5	Highly hazardous	15.0
MEK	5	3	3	Recommended	4.0
Cyclopentanone*	3*	2*	5*	Problematic*	7.0
Acetonitrile	4	3	3	Problematic	5.9
Pyridine	3	2	3	Hazardous	1.3
Acetone	5	3	5	Recommended	4.3
DMF	3	9	5	Hazardous	0.1
THF	6	7	5	Problematic	0.8
DMSO	1	1	5	Problematic	0.3
Diethyl carb.*	4*	1*	3*	Recommended*	0.7

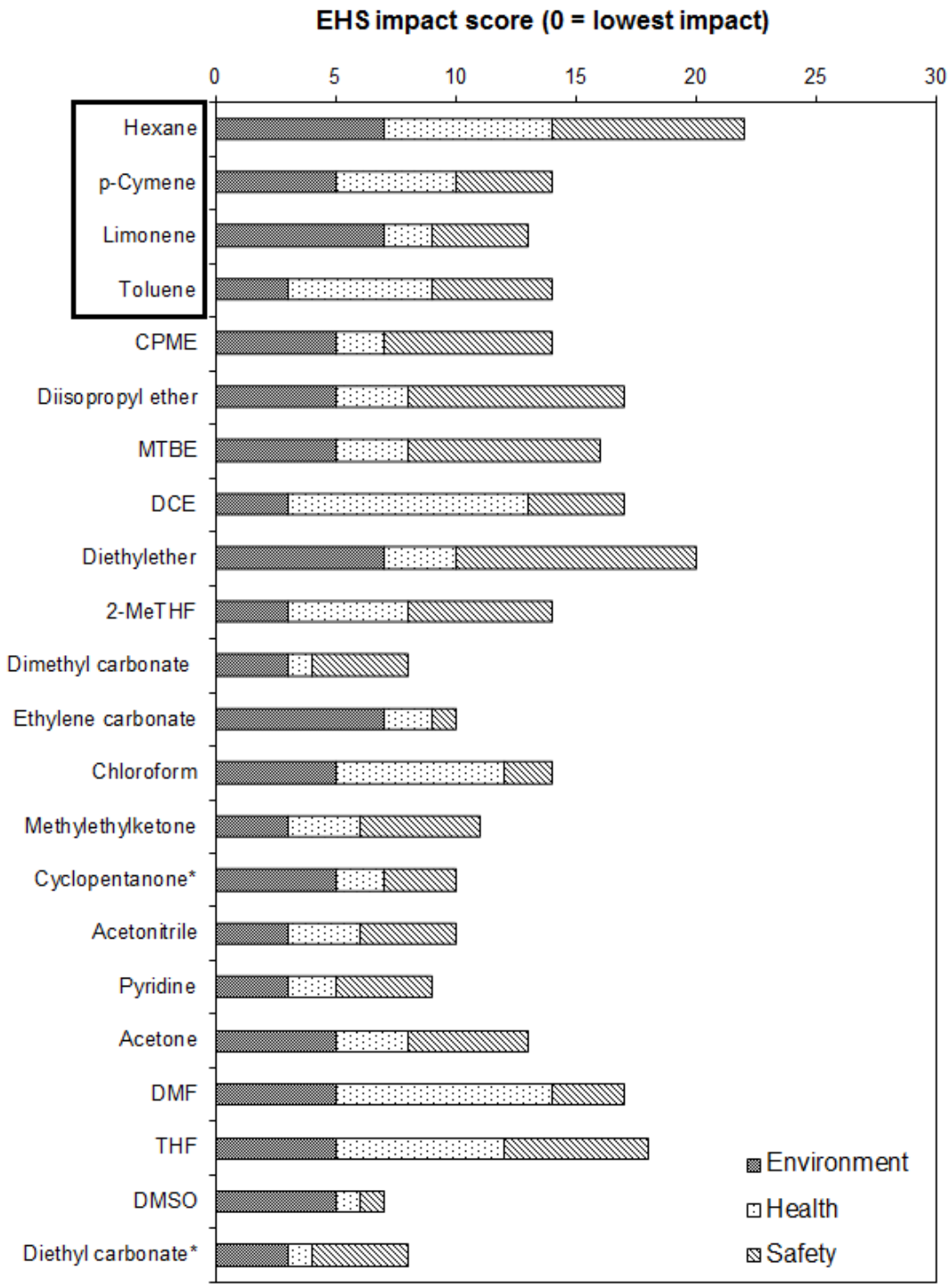


Figure 2.17: Bar-chart comparison of greenness scores of solvents used in screening according to CHEM21 assessment.[26]

in spite of their high molecular similarity they show differences in their individual EHS contributions. As for environmental impact, limonene is the only one to bear a GHS H410 statement ('Very toxic to aquatic life with long lasting effects') which penalises its scoring to 4.[26] Due to a lack of data at the time when CHEM21 solvent guide assessment was published, *p*-cymene was precautionary assigned a medium score of 5, although it is now confirmed to present the same GHS phrases as toluene (scored 3).[58] Arguably, this method of scoring does not give a full account of a solvent's environmental impact, as it does not consider upstream manufacturing, sustainability of starting materials, and overall life-cycle assessment (LCA). An assessment accounting for LCA impact and crediting renewable sourcing may result in a very different relative scoring of these solvents.

By examining the impact on health, toluene appears to carry the greatest concerns due to GHS phrases H361d ('Suspected of damaging the unborn child') and H373 ('May cause damage to organs through prolonged exposure'), which deserve it a score of 6.[58] Also in this case, information on *p*-cymene was not available at the time when the scoring methodology was developed and was therefore assigned a middle score on the health category. Based on the data presently available from its updated safety data sheet, a new assessment would more likely assign a score of 2 based on statement H304 ('May be fatal if swallowed and enters airways') - equivalent to limonene.[58]

Finally, in terms of safety, the solvents were scored predominantly on the basis of their flammability, with toluene as the most flammable (score 5). Limonene and *p*-cymene scored 4, penalised by one unit in the methodology for their potential risk to form peroxides.

2.6 Chapter summary

In conclusion, based on both EHS and performance considerations, limonene and *p*-cymene appear to be the most promising bio-based solvents for CALB-catalysed esterifications. The work described in this Chapter provides preliminary insight into the major effects of solvents in the enzyme catalysed reaction.

First, the effect of catalyst loading in the reaction was assessed as well as the the varia-

tion of solvent effects depending on the substrate.

The solvent screenings carried out revealed a combined effect on initial rate and final equilibrium of reaction. Some further investigation into the causes of the different equilibria ruled out the hypothesis of product-inhibition or enzyme deactivation/denaturation. It is concluded that the effect is more likely due to different inclinations of each solvent to favour either esterification or hydrolysis, possibly attributed to the ability of the solvent to dissolve water and make water molecules available to the active-site. Highly hydrophilic solvents allow diffusion of water through the medium and favour hydrolysis, whereas hydrophobic solvents may effectively phase water out of the reaction and promote the esterification reaction.

From the kinetic screening and regression analysis, it was possible to identify the most relevant solvent parameters describing the observed trends. From this, it emerged that Kamlet-Taft's parameter β - which accounts for the solvent's hydrogen-bond accepting ability - and molar volume V_m - which identifies the molar density of the solvent - were the most significant variables dictating solvent behaviour. In other words, bulky solvents with low ability to engage in hydrogen bonding gave the most favourable results in this reactions of study.

However, this preliminary investigation carries some limitations. First, an approximation was made in the estimation of initial rates of reaction. Secondly, the development of the model to achieve good statistical standards imposed the need to discriminate a considerable number of solvents. While excluding the lowest performing entries from the analysis is common practice, the confidence test required us to exclude additional solvents from the set for which behaviour was not sufficiently described by the model. From the experience gained with this initial work, a more advanced, thorough and systematic study was pursued and applied to the esterification of a pharmaceutical relevant compound. Chapters 3 and 4 illustrate the developments of these efforts as part of a case-study of pharmaceutical relevance, where a systematic account of kinetic and thermodynamic factors is given to clarify the mechanism of solvent influence in lipase catalysis.

Chapter 3

Solvent effects on kinetic properties: (S,S)-Reboxetine synthesis

3.1 Introduction

The study presented herein was developed in collaboration with industrial partners at Pfizer as part of a CHEM21 collaboration. Initially, it was envisaged to provide proof-of-principle for the substitution of toluene within a pharmaceutical synthesis of industrial relevance, the commercial synthesis of (S,S)-Reboxetine. Alongside the interest in finding the best bio-based candidates for substitution, the research presented an opportunity for in-depth investigation over the role of the solvent in the enzymatically catalysed step illustrated in Figure 3.1 which is herein described in detail.

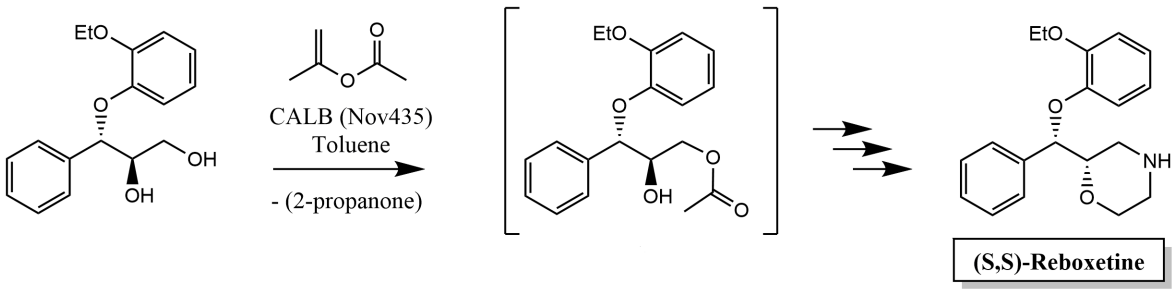


Figure 3.1: CALB catalysed regioselective acetylation of the primary alcohol position of Reboxetine diol intermediate within the synthesis of (S,S)-Reboxetine, active pharmaceutical ingredient (API).[141]

Later, the project developed far beyond this goal, exploring solvent influence on the catalysis at a mechanistic level. The complexity of the system and the number of species and variables involved make it challenging to give a definite interpretation of the experimental observations, especially since the very nature of solvent effects in biocatalysis is still highly disputed in the academic literature.[78,105] As already introduced, it appears that a relatively under-explored way of investigating solvent effects in biocatalysis is through multivariate analysis, and especially considering both kinetic and thermodynamic approaches towards improving mechanistic understanding of such systems. Within the study presented in this Chapter, a holistic approach was adopted and medium effects were investigated from multiple perspectives by looking at the relative contribution of key solvent properties, their influence on the reaction regioselectivity, rate, and activation energy. A theory was proposed based on the experimental observations, which attempts to combine

aspects of the dual nature of solvents - bulk medium or molecular entity?

3.2 Solvent selection

Table 3.1: Properties of the selected solvents. Details and relative references can be found in Appendix A.

	Hansen[59]			Kamlet-Taft[157]			S_{diol}	$[H_2O]$	V_m	logP
Solvent	δ_d	δ_p	δ_h	α	β	π^*	[mM]	[ppm]	[ml/mol]	[-]
Toluene	18.0	1.4	2.0	0.00	0.12	0.50	70	212	106.3	2.58
p-Cymene	17.3	2.4	2.4	0.00	0.13	0.39	10	159	156.6	3.47
Limonene	1.8	4.3	4.0	0.00	0.00	0.16	15	1256	161.9	4.57
2-MeTHF	16.9	5.0	4.3	0.00	0.57	0.51	307	311	100.9	1.25
Prop. carb.	20.0	18.0	4.1	0.00	0.38	0.90	198	nd	85.1	-0.41
Cyrene	18.8	10.6	6.9	0.00	0.61	0.93	198	nd	102.5	0.35
Acetone	15.5	10.4	7.0	0.08	0.43	0.71	379	nd	74.0	-0.24
DMI	17.6	7.1	7.5	0.00	0.43	0.84	197	nd	151.2	nd
2-MeFuran	17.3	2.8	7.4	nd	nd	nd	155	nd	90.2	1.85
MEK	16.0	9.0	5.1	0.06	0.48	0.67	371	nd	90.1	0.30
DiMe carb.	15.5	8.6	9.7	0.00	0.38	0.47	nd	377	84.7	0.24
DMEU	18.2	10.0	8.1	0.00	0.80	0.92	570	nd	108.1	-0.30
DCM	17.0	7.3	7.1	0.13	0.10	0.82	274	180	63.9	1.58
Chloroform	17.8	3.1	5.7	0.20	0.10	0.58	266	158	80.1	1.83

With the aim to capture a breadth of solvent effects, 14 solvents were selected considering a broad range of solvent properties and chemical diversity. Some limitations imposed by the reaction chemistry entailed that no alcoholic solvent could be included, despite it being the richest category of bio-based solvents. The risk was that protic nucleophilic solvents could act as competitive reagents in the reaction of study. In an attempt to circumvent this issue, two non-bio-derivable chlorinated solvents were included in the screening. While this was not preferred from a green perspective, it was deemed necessary to provide an account for some of the typical properties of such solvents (electron-accepting or -donating effects) whilst benefiting from their inert nature. Table 3.1 presents the solvent set selected

for this study, and reports values for the key descriptors considered in later analysis.

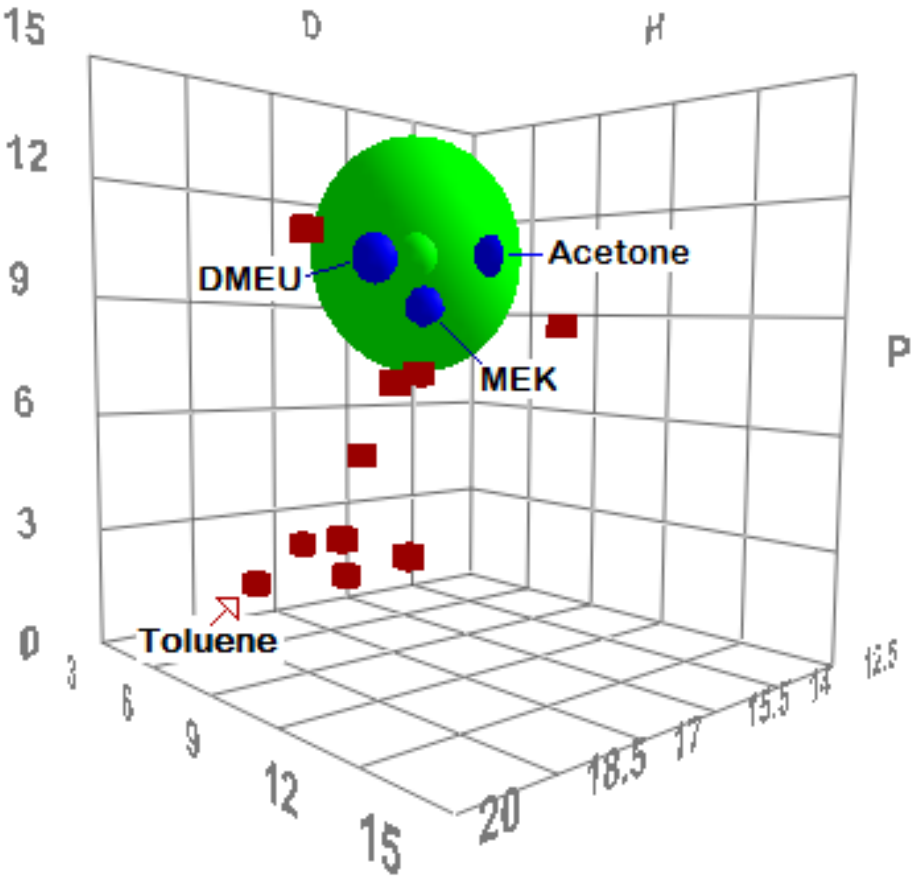


Figure 3.2: 3D Hansen solubility plot (HSPiP) of the selected bio-based solvents.[59] The sphere of highest dissolving ability in respect to the diol substrate incorporates solvents with saturation point over 350mM.

The dataset from Table 3.1 was used to develop graphic representations of the solvent space intended for screening, such as Figure 3.2 and Figure 3.3, in order to proove coverage of a broad set of properties. Figure 3.2 displays the Hansen's space developed with HSPiP software,[59] where the three-dimensional axes represent the dispersive (δ_d), polar (δ_p) and hydrogen-bonding (δ_h) contributions to the total cohesive energy of each solvent.[158] The plot combines a solvent's Hansen parameters with the experimental solubility of the substrate in the solvent in order to computationally identify the sphere of highest dissolving ability, showed in green. The sphere incorporates three solvents bearing a saturation point over 350 mM - acetone, *N*,*N*′-dimethylethyleneurea (DMEU), and methylethylketone (MEK). It appears evident that the position of toluene in the plot, as benchmark solvent

for the reaction, is far outside the sphere, being able to solubilise just 70 mmol per cm^3 - further details can be found in Appendix A.

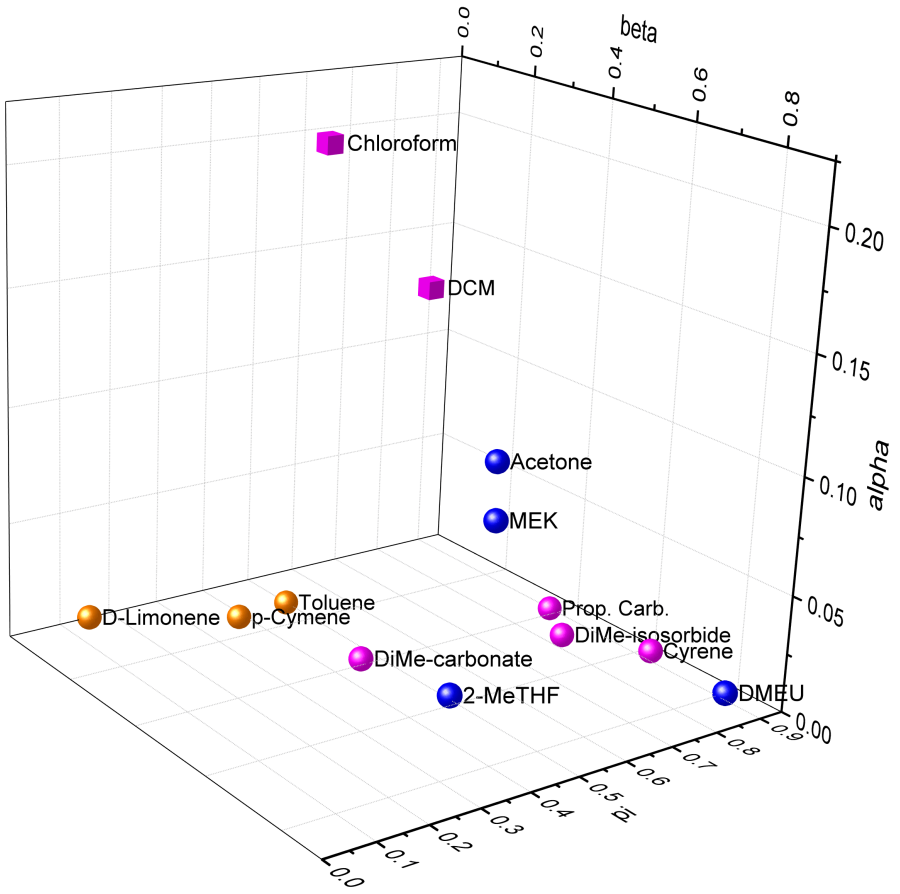


Figure 3.3: 3D plot of Kamlet-Taft parameters for the solvents of study. Colour-code represents measured substrate solubility: *orange* below 100mM, *pink* from 100mM to 200mM, *blue* above 200mM.

Similar observations can be drawn from Figure 3.3, a 3D-plot of the Kamlet-Taft parameters of the selected solvents: hydrogen-donating ability (α), hydrogen-accepting ability (β) and polarisability (π^*). The difference between these two sets of parameters is predominantly in the way they are generated: Hansen's parameters are derived computationally via software modelling and are typically used in combination with empirical solubility data, whereas Kamlet-Taft parameters are obtained experimentally by means of a spectroscopic method and use of solvatochromic dyes. In essence, the properties that Hansen and Kamlet-Taft parameters describe are strongly related, with Hansen's δ_h incorporating the combined meaning of Kamlet-Taft's H-donating and H-accepting properties (α, β), and Hansen's δ_p and δ_h splitting the polarisability effect of Kamlet-Taft's π^* into its polar and

non-polar nature.[16]

In Figure 3.3 it is evident that only few solvents have a significant α value, in keeping with the requirement for non-protic solvents imposed by the chemistry of study. However, four solvents had been selected to account for $\alpha \neq 0$, and these were methylethylketone (MEK), acetone, DCM and chloroform. A solubility dimension was also added to the figure by means of colour-coding solubility information, where orange represents the lowest ability to dissolve Reboxetine diol, and blue the highest. This indicates that with increasing values of β and π^*, the solvent's ability to dissolve the reaction substrate was also enhanced. Likewise in the Hansen's plot, toluene is here found in the region of lowest solubility (low β and π^*) together with *p*-cymene and limonene - as could be expected based on structure similarity.

3.3 Regioselectivity

As highlighted in Hayes *et al.*, the Reboxetine project team at Pfizer encountered a number of challenges when screening the acetylation of Reboxetine diol in different conditions.[141,159] In the study, they considered and compared chemical and enzymatic catalysis and various protecting groups aiming to functionalise the primary alcohol position with optimal regioselectivity.[159]

As illustrated in Figure 3.4, there are two transformation pathways effectively catalysed by the enzyme in this system: i) synthetic, attaching acyl group to free OH positions; and ii) hydrolytic, using water molecules present in the system to cleave formed ester bonds. Reportedly, a heat-induced equilibrium takes part in shifting the equilibrium between the two mono-acylated isomers, leading isomer ratios to converge at approximately 85:15 desired/undesired product.[141] By building on such knowledge of the process gathered from published and unpublished reports,[141,146,159,160] it was possible to replicate experiments under the same conditions, with the aim to investigate a wider range of bio-based solvent options.

Figure 3.5 illustrates the results of an HPLC screening carried out on the bio-based

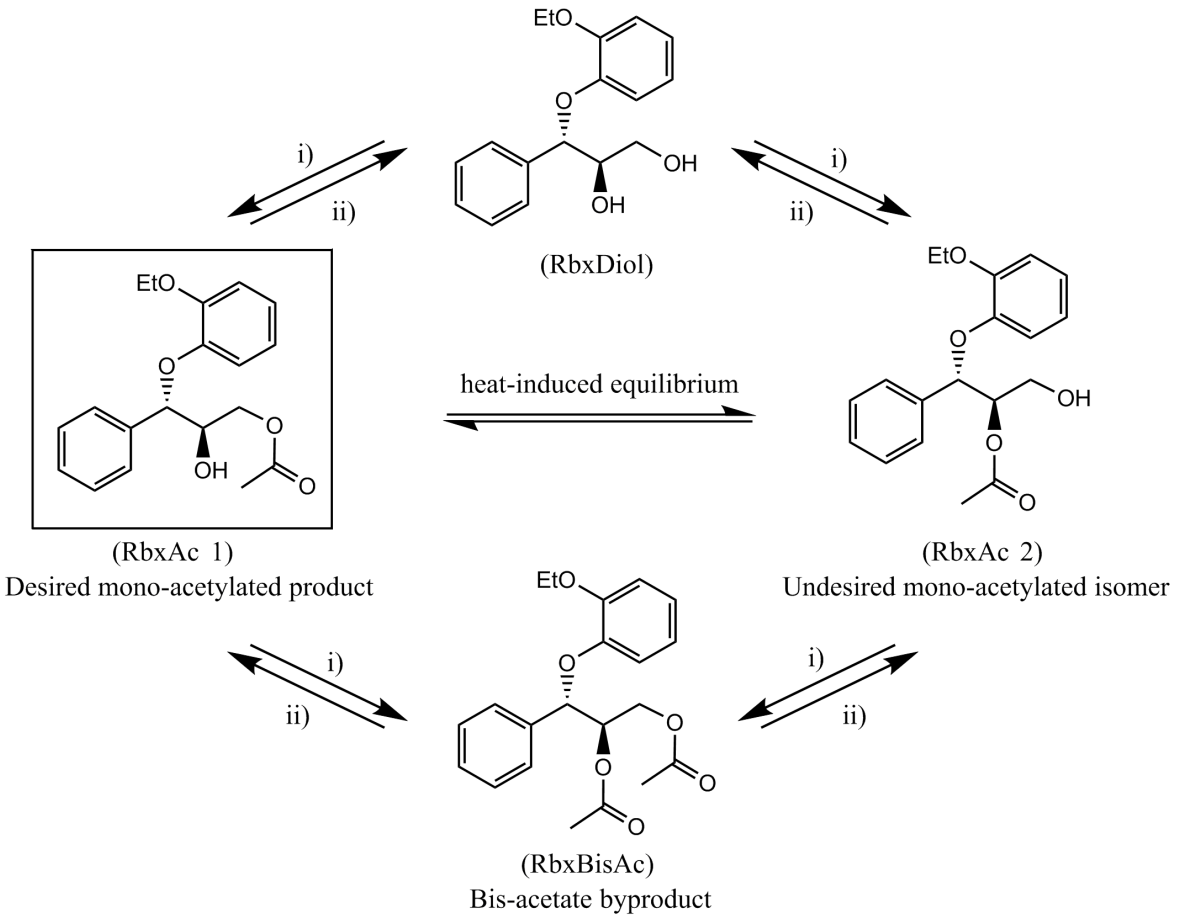


Figure 3.4: Equilibria influencing regioselectivity of the CALB catalysed system of study: i) Esterification with isopropenyl acetate; ii) Hydrolysis with water.[141]

solvents of interest (here not including the chlorinated solvent). Solvent performance was evaluated as relative yield of the preferred mono-acetylated product against that of the undesired mono-acetylated isomer, and was as recorded by HPLC analysis after 6 hours of reaction. Formation of the bis-acetate product was in some cases detected yet with intensities below the quantification threshold.

Within the distribution presented in Figure 3.5 an ideal solvent candidate would place itself in the top left corner of the graph, with the highest selectivity towards the desired product. Toluene - benchmark solvent for this reaction - is found in this area with a regioselectivity of 98.9% (average of two experiments). Notably, among the most promising bio-based solvents also falling in this area there are:

- 2-MeTHF (99.7%);

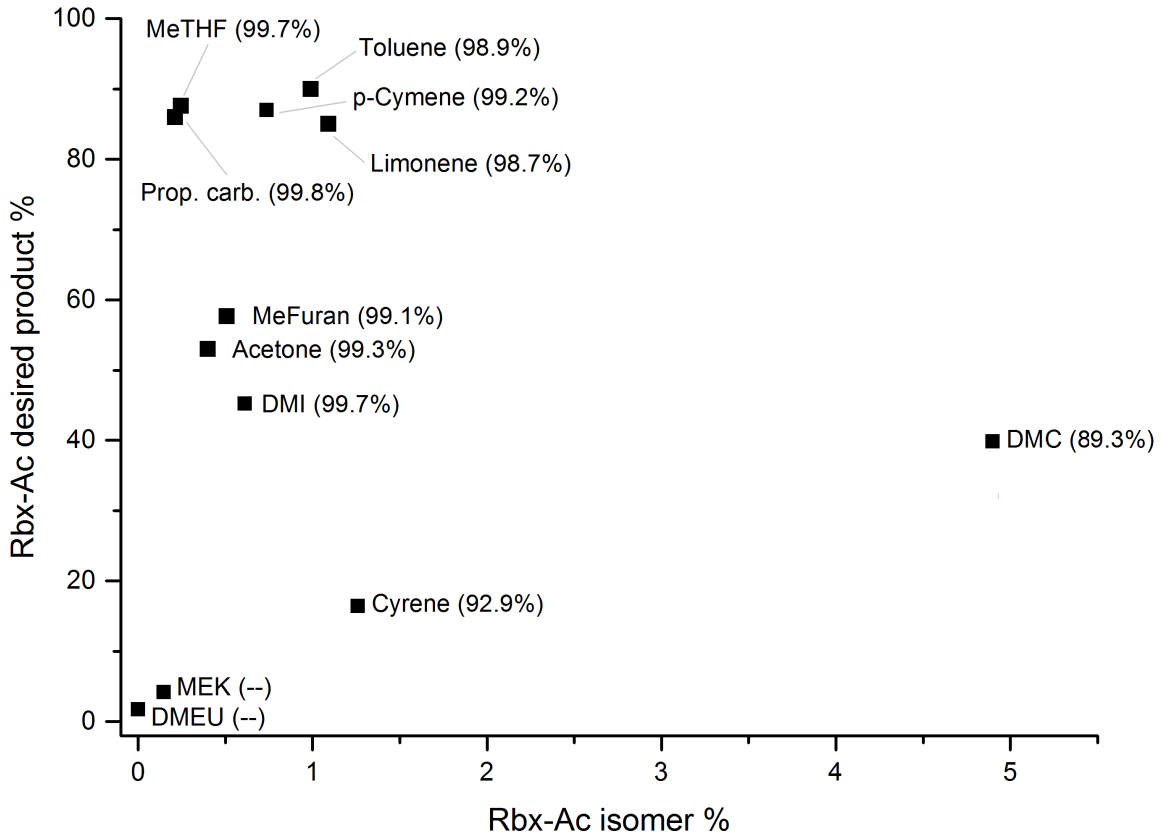


Figure 3.5: Solvent influence on regioselectivity of CALB catalyzed acetylation of reboxetine diol - percent regioselectivities reported within brackets. Conditions: 347 mM Reboxetine diol, 2 eq isopropenyl acetate, 2 mg CALB Novozyme435, 1 ml solvent, 50°C. Samples were collected at 6-hour time-point and analysed via calibrated HPLC.

- Propylene carbonate (99.7%);
- *p*-Cymene (99.2%);
- Limonene (98.7%).

These bio-based solvents appear to be ideal candidates for the substitution of toluene in the enzyme-catalysed acylation of Reboxetine diol. This result was taken further through the downstream synthetic steps up to product isolation.

3.4 Proof-of-concept telescoped sequence

In order to retain the benefits introduced through the telescoped process, the substitution of toluene in the enzymatic step would have the highest relevance if it still allowed telescoping the downstream chemistry successfully in the alternative solvent of choice (Figure 3.6). For the purpose of providing proof-of-concept substitution of toluene, two bio-

based solvents emerged from the regioselectivity screening. Limitations in sample availability allowed us to carry out the synthetic sequence in 2-MeTHF and in *p*-cymene, with toluene run as a benchmark. The chemo-enzymatic sequence was carried out up to isolation of the zwitterion salt with intermediate isolation of Reboxetine acetate and mesylate for the purpose of characterisation and quantification.

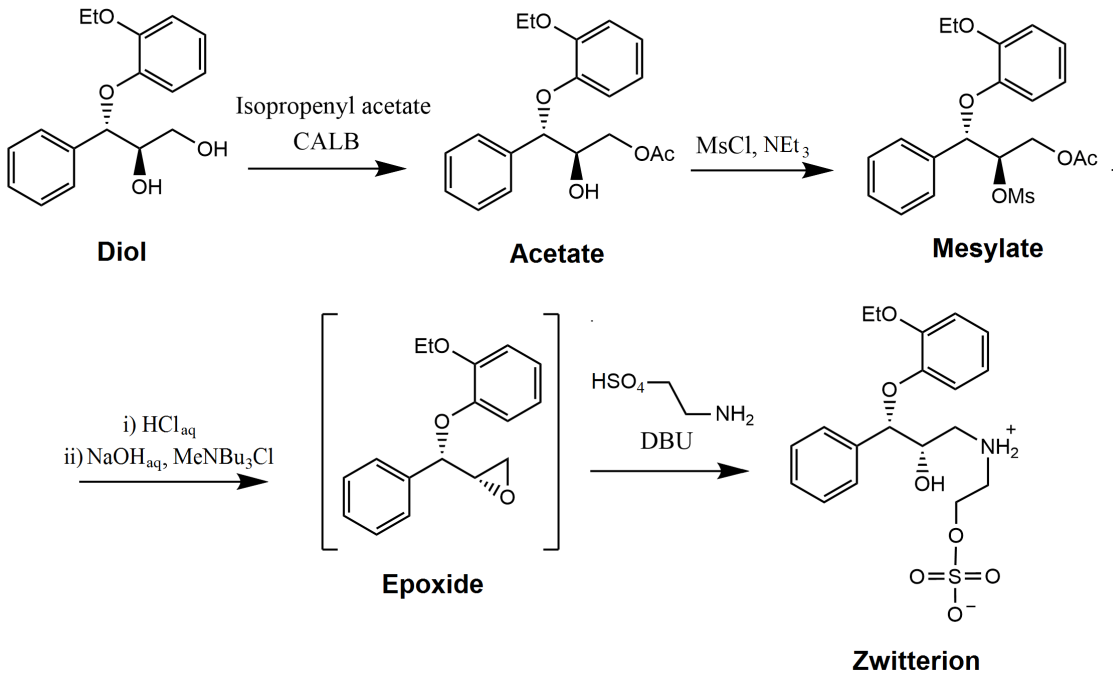


Figure 3.6: Telescoped synthesis from Reboxetine diol to Reboxetine zwitterion in one solvent.

Figure 3.7 represents the reaction profiles observed via GC for the first two steps of the synthesis, the acetylation of the primary alcohol followed by mesylation of the secondary one. The profiles show *p*-cymene and toluene as be rather equivalent, both leading to high conversions in both steps. In the first step, 2-MeTHF shows the highest initial reaction rate, however the yield hits a maximum at 70% and then drops suggesting the hydrolytic started to interfere. Also in the second reaction 2-MeTHF resulted in the lowest performance.

The synthesis was then progressed through epoxide ring formation, first by cleavage of the acetyl moiety in acid conditions followed by base-catalysed ring closure with elimination of the mesyl activating group. It became apparent that the latter step could not progress in 2-MeTHF, where the higher miscibility of the solvent with the alkaline aque-

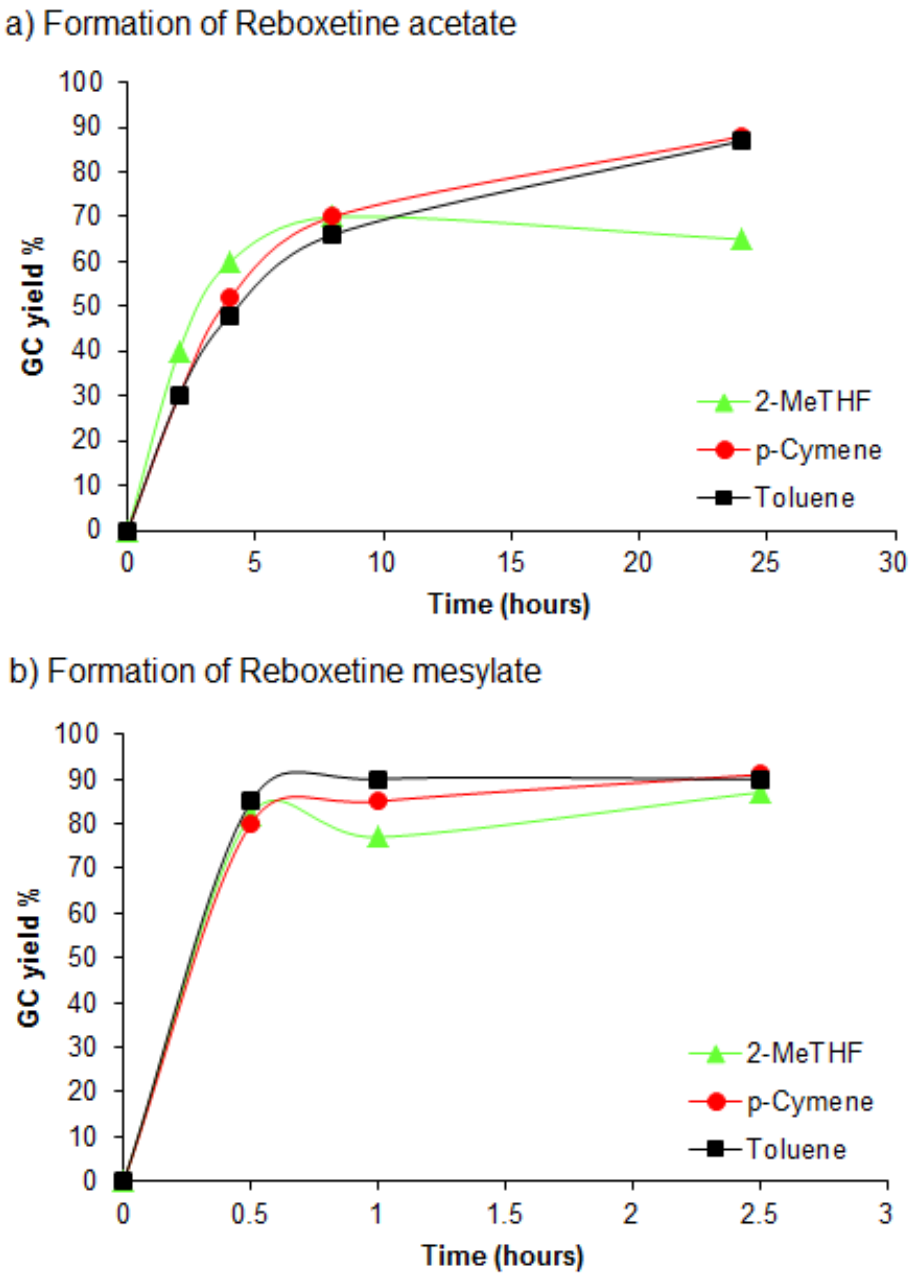


Figure 3.7: Comparison of three solvents over two steps of the telescoped synthesis, a) esterification, b) mesylation.

ous phase affected the stability of the epoxide intermediate. In 2-MeTHF, by GC analysis showed that the mesylate product converted back to the diol starting material. It is believed that, if at all formed, the epoxide would have readily opened under attack of the soluble OH^- nucleophile and re-delivered the diol. Nevertheless, it was possible to progress the reactions in *p*-cymene and toluene up to the final zwitterion intermediate - which is a direct precursor to the API (Figure 3.8).

Table 3.2: Comparison of yields of zwitterion product obtained by carrying out the telescoped sequence in three solvents. Benchmark optimised value reported from Hayes *et al.*[141]

	Isolated yield % in respect to diol starting material		
Solvent	Acetate	Mesylate	Zwitterion
2-MeTHF	55%	43%	not formed
p-Cymene	78%	69%	40%
Toluene	81%	75%	45%
Toluene (Hayes *et al.*)	not isolated	not isolated	86%

Table 3.2 summarises the results obtained and compares them against the benchmark yield reported in the literature by Pfizer for the optimised kg-scale synthesis (86%). In this proof-of-concept work, the yields isolated from toluene and *p*-cymene were significantly lower than the literature value, largely due to: 1) the considerably smaller scale of the synthesis (500 mg), 2) the additional isolation steps introduced for characterisation, and 3) the under-optimised conditions. Yet the yields from the two solvents appear comparable, indicating scope for improvement and promising potential for the substitution of toluene with *p*-cymene throughout the telescoped series.

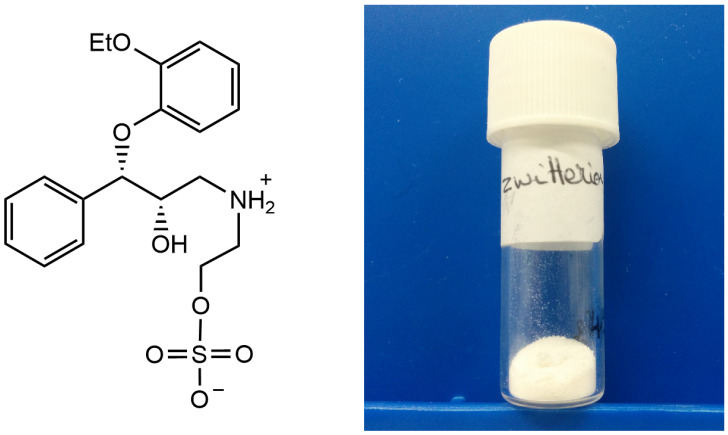


Figure 3.8: Synthesised and isolated Reboxetine zwitterion from telescoped process run in *p*-cymene.

3.5 Kinetic profiles in bio-based solvents

In order to build on considerations made above with regards to yield and regioselectivity in the enzyme-catalysed step, a full kinetic screening was performed with the aim to investigate the factors behind observed solvent effects. Unlike the work of Chapter 2, OriginLab advanced statistical and regression functions were employed in place of Excel's. Rate constants were calculated via linearisation of conversion profiles, and accompanied by an analysis of the reaction order across the solvent set.

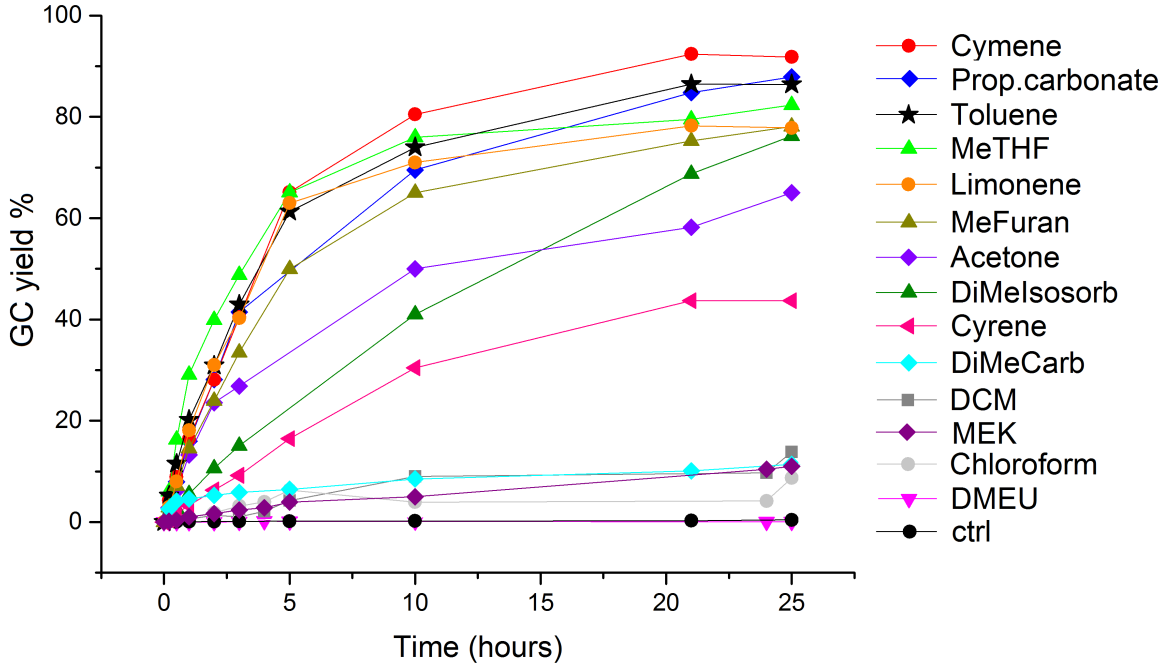


Figure 3.9: Kinetic screening of enzymatic catalysed acetylation of reboxetine diol representing the conversion to the desired regioisomer Conditions: 115 mM diol, 1.5 eq isopropenyl acetate, 2 mg enzyme CALB Novozyme435, 3 ml solvent, 40°C. Yields as by calibrated GC area of desired acetylated product.

The reaction in each solvent was monitored over a period of 25 hours, followed via calibrated GC-FID. Conversion curves are displayed in Figure 3.9. Under the reaction conditions a broad spread of solvent effects was observed. *p*-Cymene emerged as the highest performing solvent in terms of product yield providing higher yields than toluene - reference solvent. In terms of rate, 2-MeTHF gave the fastest initial reaction, yet plateaued at an equilibrium at least 10% lower than the one achieved in *p*-cymene. Propylene carbonate, limonene and 2-methylfuran were also shown to favour the reaction, providing high initial

rates and conversions. Conversely, dimethyl carbonate, MEK, DMEU and the chlorinated solvents showed to strongly disfavour the reaction and therefore appear to be the least suited solvents for this application - reasons clarified at the end of the Chapter.

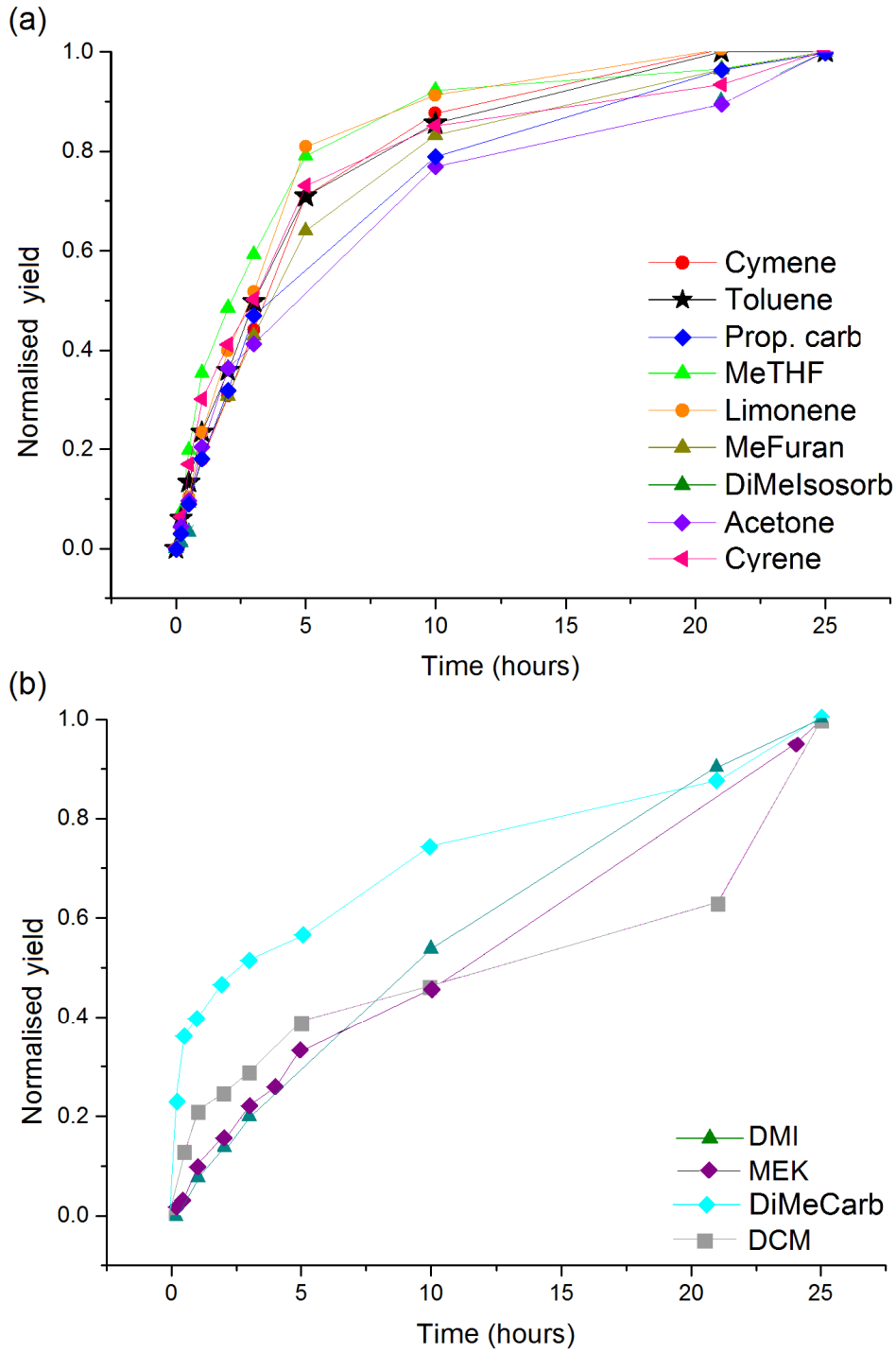


Figure 3.10: Normalised conversion curves up to a top value of 1, sorted in two classes - a) and b) - based on profile similarity.

Further to this analysis, additional insights were obtained by applying normalisation to the above curves. Normalisation is a technique that can be used to compare reaction profiles in a way that is independent of their performance. By assigning to all final yields the normalised value of 1, and proportionally converting the values of intermediate time-points, a new version of the reaction curves was obtained. Figure 3.10 illustrates the normalised results, split into two categories - (a) and (b) - that showed significant differences in shape and behaviour.

The solvents grouped in graph (a) of Figure 3.10 show a remarkable overlap in their normalised profiles. On the other hand, dimethyl carbonate (DMC), dimethylisosorbide (DMI), MEK and DCM distanced themselves from all other solvents behaviours, suggesting perhaps a different mechanism of influence taking place (Figure 3.10b). Normalised conversions for the reactions in chloroform and DMEU were found to be highly affected by the relative error associated with their low conversions and were not included - see Appendix A. A speculative interpretation of the outcome of this analysis may suggest the presence of two broad classes of solvent behaviour in the process: 1) those favouring the reaction to reach its natural equilibrium following a similar course; 2) those imposing a low yield (< 10%) and a different course of reaction. However, this observation may be highly affected by the inherently higher experimental error associated with the lower performing solvents of the set, which correspond to those drifting from the main trend.

3.5.1 Rate constants

The analysis proceeded to investigate the order of each reaction in each solvent and to calculate corresponding rate constants. Figure 3.11 presents the processed experimental data for the reaction in toluene as a representative case example - other solvent plots are provided in Appendix A.

Figure 3.11a shows the initial consumption of the starting material in the first 5 hours of the reaction. By representing the natural logarithm of the starting material's concentration against time, as done in plot (b), the result is a straight line with a negative slope suggesting first order kinetics. According to the definition of first order kinetics reported in

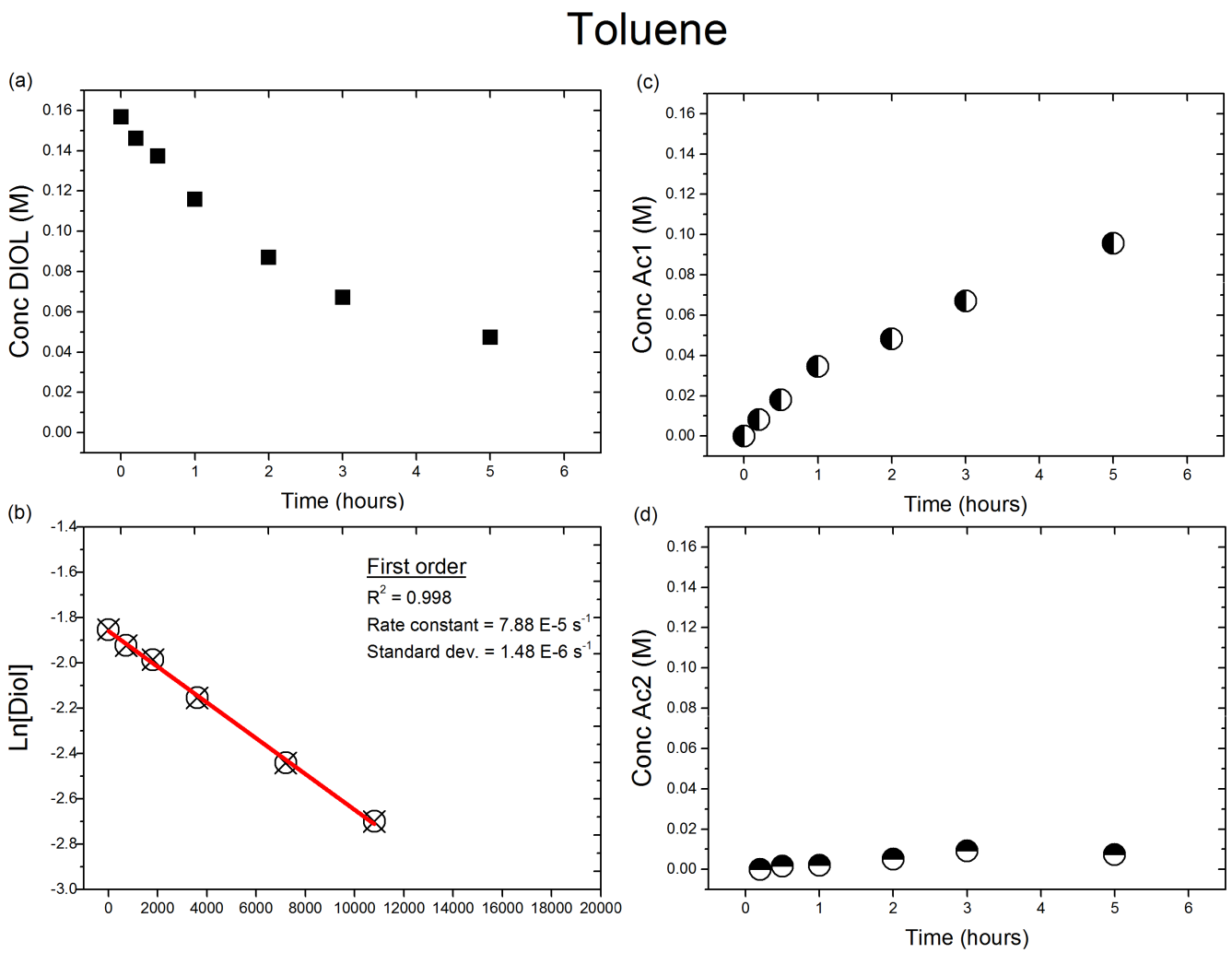


Figure 3.11: Example of determination of reaction rate constant for reaction in toluene. a) Consumption of starting material, b) Yield of desired product, c) Yield of undesired regioisomer, d) Linearisation process used to extract the order and rate constant of the reaction.

equation 3.5.1, the rate of the reaction - *i.e.* the variation of the starting material's concentration over time - is linearly dependent on the concentration of the reagent by a constant factor r, the rate constant.[161]

$$Rate = -\frac{d[SM]}{dt} = r[SM] \tag{3.5.1}$$

$$[SM] = [SM]_0 e^{-rt} \tag{3.5.2}$$

$$ln[SM] = ln[SM]_0 - rt \tag{3.5.3}$$

The integrated and linearised forms of the first order formula (equations 3.5.2 and 3.5.3) provide the basis for the experimental determination of reaction order. Since the linearity was experimentally satisfied across all solvents tested (see Appendix A), the reaction of study can be said to follow a first order mechanism, dependent on the concentration of starting material. Through equation 3.5.3 it becomes possible to derive the rate constants of each reaction from the gradient of the linearised fits. The rate constant for each reaction was extracted as absolute value of the gradient of the fitting line, $\mid r \mid$. All obtained rate constants are reported in Table 3.3 alongside corresponding standard deviation (σ_r) and natural logarithm form.

3.6 Multivariate regression

Thus far, the results proved that solvents can critically influence the yield, rate and regioselectivity of the biocatalytic reaction of study. In order to better understand the origin of the observed solvent-induced effects, a multivariate analysis was pursued as a statistical methodology for assessing the influence of a set of independent parameters on a representative dependent variable of the system, the reaction rate constant.

Table 3.3: Results from kinetic screening of solvent effects in Reboxetine acetate formation.

Solvent	rate const (r)	std error (σ_r)	$ln(r)$
Toluene	1.10E-04	7.22E-06	-9.12
p-Cymene	1.04E-04	4.14E-06	-9.17
Limonene	8.23E-05	3.67E-06	-9.41
2-MeTHF	1.17E-04	5.39E-06	-9.05
Prop. Carb.	5.57E-05	6.65E-07	-9.80
Cyrene	1.71E-05	1.43E-06	-10.98
Acetone	3.44E-05	6.95E-07	-10.28
DMI	2.50E-05	7.22E-07	-10.60
2-MeFuran	4.19E-05	1.20E-06	-10.08
MEK	9.01E-06	3.56E-07	-11.62
DiMe-Carb.	3.85E-06	2.85E-06	-12.47
DMEU	1.40E-06	3.96E-07	-13.48
DCM	5.72E-06	2.63E-06	-12.07
Chloroform	5.97E-06	4.37E-07	-12.03
(ctrl no enzyme)	5.76E-06	7.04E-07	-12.06

3.6.1 Selection of regression variables

The natural logarithm of the rate constant was adopted as dependent variable (y). As for the set of independent variables (x_n), eight medium-related parameters were employed:

- Kamlet-Taft solvatochromic parameters:
  - α, hydrogen bond donating ability;
  - β, hydrogen bond accepting ability;
  - π^*, polarisability;
- Hansen's solubility parameters:
  - δ_d, dispersion forces;
  - δ_p, polar-polar forces;
  - δ_h, hydrogen bond forces;
- Solubility of starting material - expressed as saturate concentration of the starting material in each solvent ($ln[SM]_{sat}$);
- Molar volume of the solvent (V_m).

Here, logP was here excluded from the regression, for being intrinsically dependent on some of the variables selected for the study. By definition, logP presents contributions of Kamlet-Taft parameters β and π^* as well as a factor accounting for the solvent's molar volume, as by equation 3.6.1.[18]

$$logP = 0.24 - 3.38\beta + 0.0266V_m - 0.96\pi^* \tag{3.6.1}$$

3.6.2 Partial Least Squares analysis

Once the variables of interest were selected, a series of tests were performed in order to test the significance of the independent variables and avoid the issue of multicollinearity. As part of the Partial Least Squares (PLS) function in OriginLab software, the Singular Value Decomposition (SVD) algorithm was applied for performing the tests detailed below. The results of such tests are illustrated in Figure 3.12.

(a) **Cross Validation** Produces the root mean of Predicted REsidual Sum of Squares (PRESS) of the set of data, which suggests the ideal number of descriptors to adopt in the model - given by the number of factors which minimises PRESS.

(b) **Variable Importance Plot (VIP)** Quantifies the relative importance and independence of variables considered for the model. When variables in the plot are associated with values below 0.8, they are most likely to carry a high risk of covariance and ultimately will have low impact on the system.

(c) **Coefficients plot** Calculates the coefficients for a descriptive model equation based on the set of variables under study. High coefficients - in absolute terms - indicate a significant role of the variable in predicting changes in the system. If a variable shows low coefficient and low importance value, then it can be confidently excluded from further analysis.

(d) **Residuals spread** Shows the spread of residuals generated from the preliminary model. When variance is *normally distributed*, residuals tend to fall evenly around the zero line proving there is no drift or bias in the analysis.

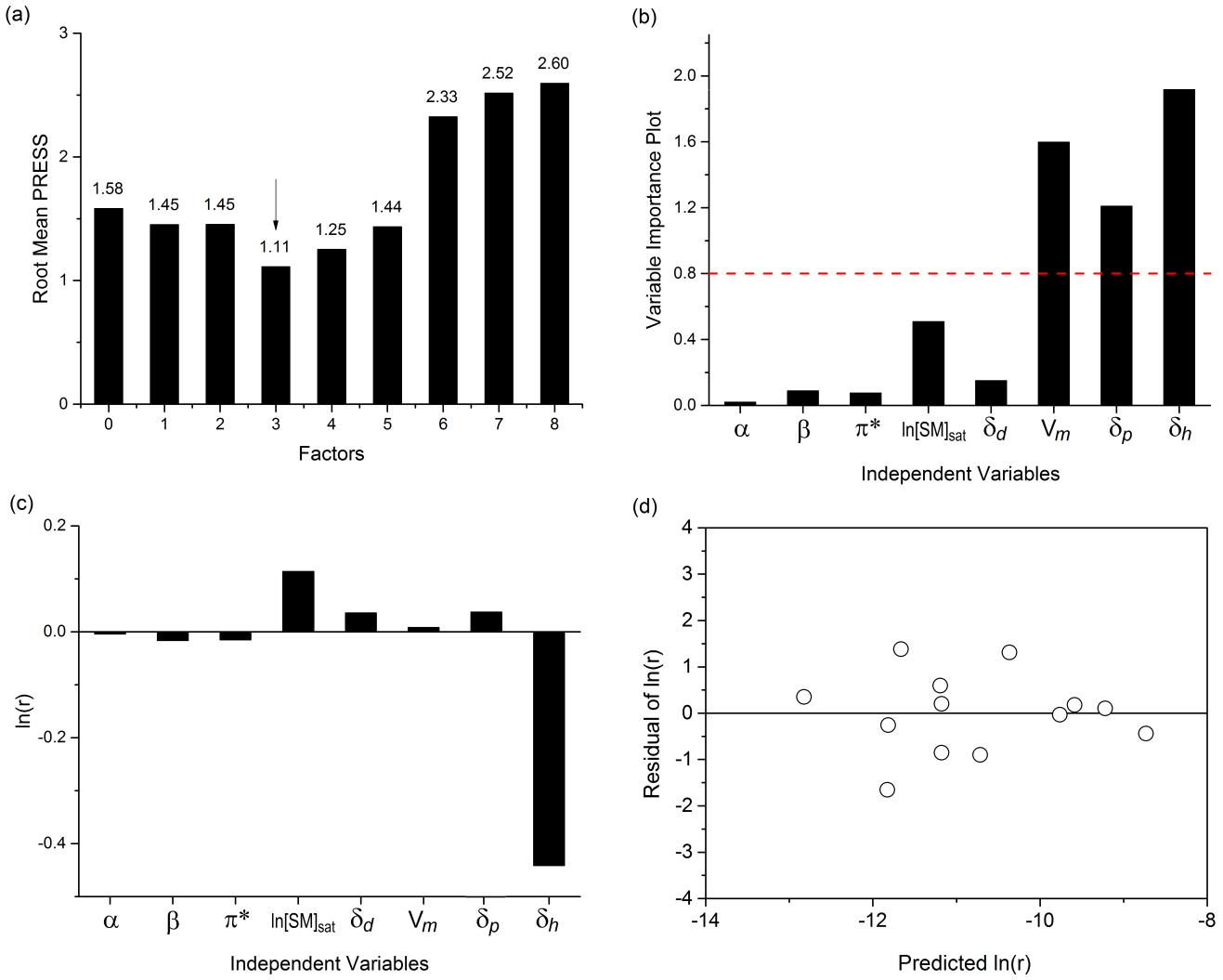


Figure 3.12: Outcomes from Partial Least Squares analysis on the data-set, where natural logarithm of rate constant was employed as dependent variable against eight properties of the solvent.

The cross validation plot (a) and the variable importance plot (b) in Figure 3.12 provide indications that three factors should be employed to describe the system of study: molar volume (V_m), Hansen's polar forces (δ_p) and hydrogen bonding forces (δ_h). Among these, the hydrogen bonding factor is dominant in determining the solvent effects across the dataset. Its pronounced negative coefficient from plot (c) indicates that the higher its value - and so the solvent's ability to engage in hydrogen bonding - the lower the rate of reaction. The other two variables which passed the variable importance test, V_m and δ_p, show to have positive coefficients of much lower magnitude suggesting a lower impact countering the effect of δ_h. Finally, graph (d) provides reassurance that the distribution of the predicted data through this preliminary analysis is normal and balanced.

3.7 Regression model

Once the most representative variables were short-listed, it became possible to iterate a regression analysis using OriginLab ANOVA algorithm by feeding only the variables of interest. Details of regression coefficients and predicted $ln(r)$ values for each solvent are presented in Table 3.4. Firstly, the regression was run on the full set of solvents - only excluding DMEU which rate constant of nearly zero made its logarithm value highly error-sensitive and skewing the results. Model equation 3.7.1 was produced based on the obtained regression coefficients.

$$ln(r) = -10.4 + 0.0152V_m + 0.0568\delta_p - 0.397\delta_h \tag{3.7.1}$$

Figure 3.13a gives a representation of the predicted ln(r) values obtained through equation 3.7.1 against the actual ln(r) values experimentally measured. The linear fitting obtained for this set was poor, with a R^2 of 0.789 and poor symmetry between predicted and actual values. This fact prompted restriction of the number of entries in the model by excluding the poorest performing solvents according to the two classes of behaviour identified above through the normalisation of the kinetic profiles (Figure 3.10). This selection criteria makes for a more rigorous strategy for the exclusion of data that may be distorting

Table 3.4: Selected parameters employed in the descriptive model and corresponding predicted values for reaction rate according to the two model equations obtained for all solvents (excluding DMEU) and for best solvents only - *i.e.* those in which reaction yield was above 40% after 25 hours (*).

	Regression variables				Predicted ln(r)	
Solvents	$y = ln(r)$	$x_1 = \delta_h$	$x_2 = \delta_p$	$x_3 = V_m$	All solv.	Best solv.*
2-MeTHF	-9.05	4.3	5.0	100.9	-10.10	-9.40
Toluene	-9.12	2.0	1.4	106.3	-9.32	-8.97
p-Cymene	-9.17	2.4	2.4	156.6	-8.83	-9.10
Limonene	-9.41	4.3	1.8	161.9	-9.42	-9.59
Prop.Carb.	-9.80	4.1	18.0	85.1	-10.04	-9.95
2-MeFuran	-10.08	7.4	2.8	88.6	-11.34	-10.17
Acetone	-10.28	7.0	10.4	74.0	-11.27	-10.34
DiMe-isosorb.	-10.60	7.5	7.1	151.2	-10.55	-10.61
Cyrene	-10.98	6.9	10.6	102.5	-10.89	-10.73
MEK	-11.62	5.1	9.0	90.1	-10.45	-
Chloroform	-12.03	5.7	3.1	80.1	-10.87	-
DCM	-12.07	7.1	7.3	63.9	-11.47	-
DiMe-carb.	-12.47	9.7	8.6	84.7	-12.08	-
DMEU	-13.48	8.1	10.0	108.1	-	-

the model, as opposed to the simple elimination of outliers. Under this rationale, the following solvents were excluded from the analysis: MEK, chloroform, DCM, DMC, DMEU. The regression analysis was reiterated on this restricted set, yielding model equation 3.7.2.

$$ln(r) = -7.80 + 0.00366V_m + 0.0430\delta_p - 0.261\delta_h \tag{3.7.2}$$

In this case, as shown in Figure 3.13b, a stronger match was found between predicted and experimental ln(r) values. The fitting line bisects the plot with good parameters, showing significantly reduced standard deviation, σ_D, and increased R^2 above 0.90 which is an encouraging result in systems of such complexity.

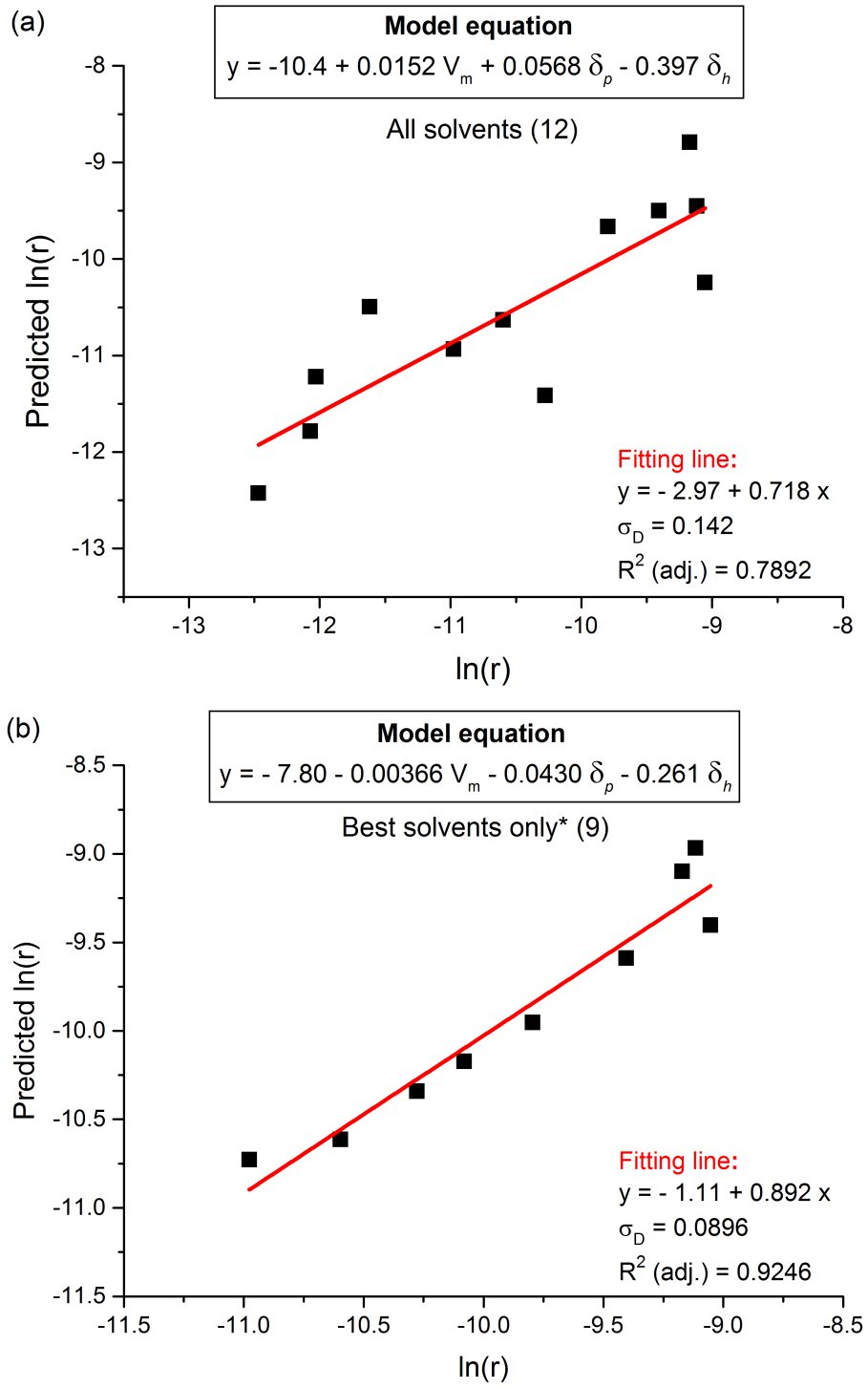


Figure 3.13: Validation of the two best regression equations a) and b). Plotting the values of $ln(r)$ predicted through the model (y axis) against those experimentally observed (x axis), ideally seeking the symmetry of a $y = x$ fitting equation.

3.8 Chapter summary

The scope of developing a descriptive equation lies not only in the given ability to predict reaction rate constants based on solvent properties, but more importantly in allowing the interpretation of solvent effects by means of a systematic and robust statistical analysis.

When considering the solvent set as a whole the analysis produced model equation 3.8.1; by excluding low performing solvents from the model yielded equation 3.8.2. In both models, Hansen's hydrogen bonding factor shows to be a consistently dominant factor.

$$ln(r) = -10.4 + 0.0152V_m + 0.0568\delta_p - 0.397\delta_h \tag{3.8.1}$$

$$ln(r) = -7.80 + 0.00366V_m + 0.0430\delta_p - 0.261\delta_h \tag{3.8.2}$$

Based on this observation and for the purpose of discussion, a plot of experimental ln(r) against pure δ_h is proposed in Figure 3.14 which confirms a clear covariance - if not strict correlation - between two variables. The distribution of data points is somewhat linear for ln(r) values especially above -11, fitting the darker best-fit line with relatively good statistical parameters - acknowledging the limitation imposed here by using only one descriptor. When considering the lower end of rate performance, and thus including the solvents represented in lighter grey, the earlier tight distribution appears to broaden and worsen the fitting statistics. While this direct correlation yields poorer statistical outcomes than the three-variable model developed, it confirms that Hansen's δ_h is a cardinal descriptor of the system. It is reasonable to conclude that Hansen's hydrogen bonding parameter δ_h alone can be used to roughly predict a solvent's suitability for a CALB-catalysed esterification such as the one of study. Moreover it can be used to interpret solvent effects from a mechanistic perspective as proposed in the discussion of Chapter 4.

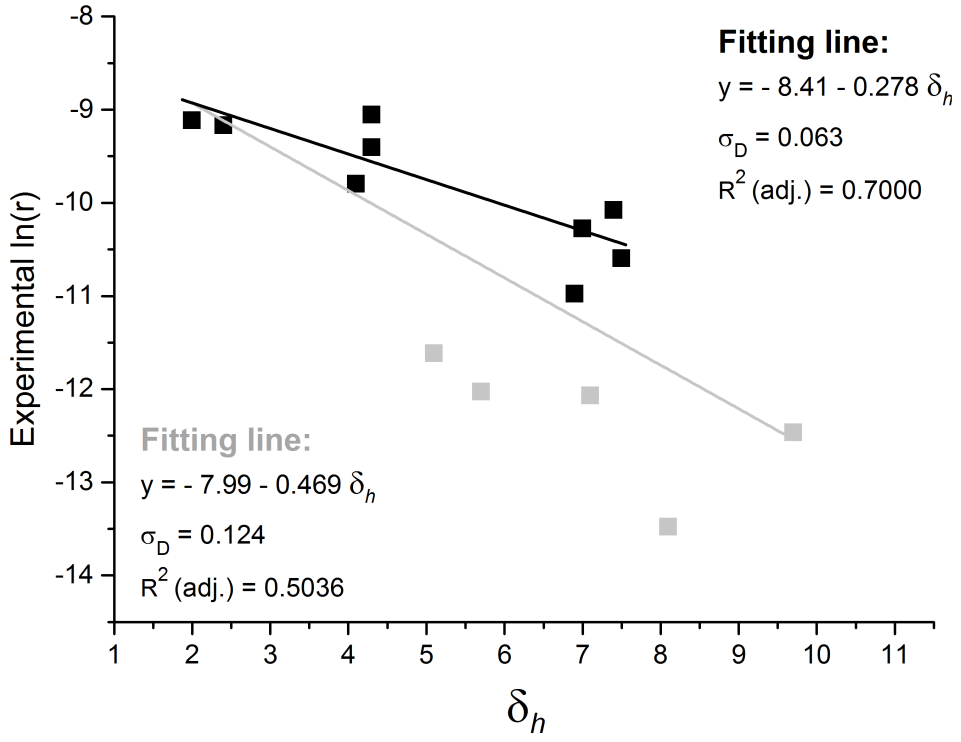


Figure 3.14: Direct correlation attempts between observed logarithm of rate constant in different solvents and the corresponding Hansen's hydrogen bonding parameter.

Chapter 4

Solvent effects on thermodynamic properties: (S,S)-Reboxetine synthesis

4.1 Introduction

The kinetic screening presented in Chapter 3 provided further insight into the dependency of reaction rate on solvent properties, where the solvent was the only variable component in the system. A further level of analysis can be achieved by introducing an additional variation and studying the solvent effects across different temperatures. The results presented in this Chapter allow an evaluation of the influence that solvents have on the activation energy of the system, and allow for for an holistic interpretation of the mechanistic role of solvents in CALB-catalysed esterification.

4.2 Temperature profiles in bio-based solvents

Five solvents from the previous kinetic screening were selected for the study of temperature effects on the acylation of Reboxetine diol: toluene (as benchmark), *p*-cymene, acetone, 2-methylfuran, and propylene carbonate. Reactions were run at 25°C, 32.5°C, and 40°C and monitored via GC over 8 hours. The screening was also performed at 55°C for the purpose of validation but not used in the determination of trends.

Figure 4.1 illustrates a representative example of the data collected in the case of toluene across the four temperatures. Plots a) and b) show the yield of formation of the desired mono-acetylated product and corresponding consumption of the starting material Reboxetine diol respectively. As expected, increase in temperature led to faster formation of the product across the four solvents, characteristic of endothermic processes. In toluene, the final equilibria of reactions at different temperatures appear mainly to converge at approximately 85% yield, where complete conversion of the starting material is achieved (with *ca.* 15% regioisomer formation). However, at the lowest temperature (25°C) the reaction appears to plateau at an early equilibrium point, and not proceed beyond 40% conversion. This phenomenon was observed consistently for the reaction in toluene and *p*-cymene but not in the other solvents, and thus may be attributed to a solubility effect affecting the reaction at low temperature. This was not investigated further as attention was focused on

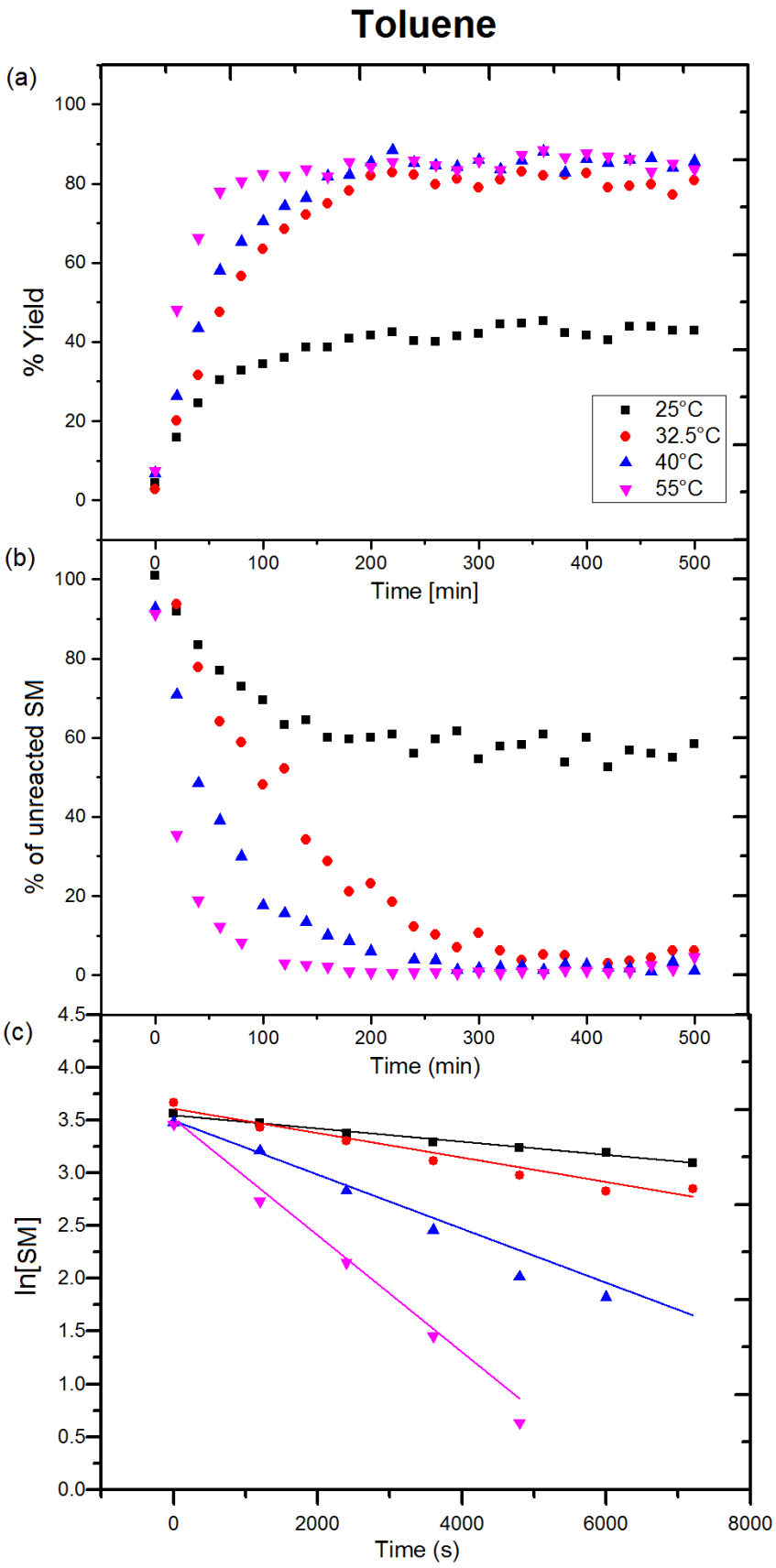


Figure 4.1: Kinetic plots for the reaction in toluene at four temperatures. a) Yield % of desired product, b) Conversion % of starting material , c) linearised starting material concentration. Conditions: 58 mM substrate, 2 eq isopropenyl acetate, 2 mg CALB Novozyme435, 3 ml solvent. Followed via calibrated GC.

initial rates of reaction and into investigating the activation energies associated with the system.

Figure 4.1c presents the consumption of the starting material in the first two hours of reaction in the form of natural logarithm of its molar concentration against time expressed in seconds. The linearity of the data obtained through this representation confirms the first order kinetics observed as discussed in Chapter 3 and prove that the mechanism does not vary with temperature. The obtained values, characteristic of each reaction temperature and each solvent, are reported in Table 4.1 - individual linear profiles can be found in Appendix A.

Table 4.1: Rate constants from kinetic screening of solvent effects at different temperatures.

	ln(rate)		
Solvent	at 25°C	at 32.5°C	at 40°C
Toluene	-9.62	-8.72	-8.31
p-Cymene	-9.82	-8.91	-8.18
Acetone	-11.52	-9.96	-9.26
2-Mefuran	-13.25	-11.31	-10.27
Propcarb	-11.38	-10.18	-9.33

4.3 Arrhenius plot

Thermic effects on reaction kinetics are typically studied and explained through the Arrhenius equation, a fundamental empirical relationship in chemical thermodynamics.[161]

$$r = Ae^{-\frac{E_a}{RT}} \tag{4.3.1}$$

$$ln(r) = ln(A) - \frac{E_a}{R}\left(\frac{1}{T}\right) \tag{4.3.2}$$

In the Arrhenius law illustrated in equations 4.3.1 and 4.3.2, r is the rate constant, A

a pre-exponential factor, E_a the empirical activation energy, R the gas constant, and T the temperature in kelvin degrees. In the context of the Arrhenius equation, E_a is the energy barrier that reagents have to overcome in order to transform into the product(s); ln(A) represents the probability of successful reactive collisions between the reaction components, encompassing factors such as geometry, diffusion and entropy.[162] Both E_a and ln(A) are evidently dependent on the reaction conditions, but most importantly in the case of a catalysed process they are tightly bound to the catalytic activity and mechanism. With a biocatalyst such as the lipase of study, the preliminary binding of substrates in the active site will favour effective interaction between reagents, thus influencing ln(A); as a consequence, the activation energy barrier will decrease, due to catalytic activation of the reagents and/or stabilisation of the transition state.

Based on the linearised Arrhenius equation 4.3.2, the experimental rate constant can be represented as ln(r) and plotted against the reciprocal of temperature as in Figure 4.2. Such representation is referred to as Arrhenius plot and each line is represented by the general equation 4.3.3 where the parameters vary depending on the solvent, i. From the gradient of each Arrhenius line it is possible to derive the empirical activation energy of the reaction E_a, while the intercept provides the pre-exponential term ln(A) in the Arrhenius equation. The obtained values are reported in Table 4.2 for each solvent, alongside the corresponding statistical errors resulted from the analysis.

$$\therefore y_i = ln(A)_i - \frac{E_{a,i}}{R}x \tag{4.3.3}$$

Table 4.2: Parameters derived from Arrhenius plot.

Solvent	Slope	σ_D	$E_a[Jmol^{-1}]$	Intercept	σ_D	$ln(A)$
Toluene	-8,121	1,631	67,518	17.70	5.34	17.70
p-Cymene	-10,217	506	84,944	24.47	1.66	24.47
Acetone	-14,066	2,278	116,945	35.79	7.42	35.79
2-Mefuran	-18,563	2,959	154,333	48.15	9.69	48.15
Prop.carb	-12,786	1,099	106,303	31.55	3.60	31.55

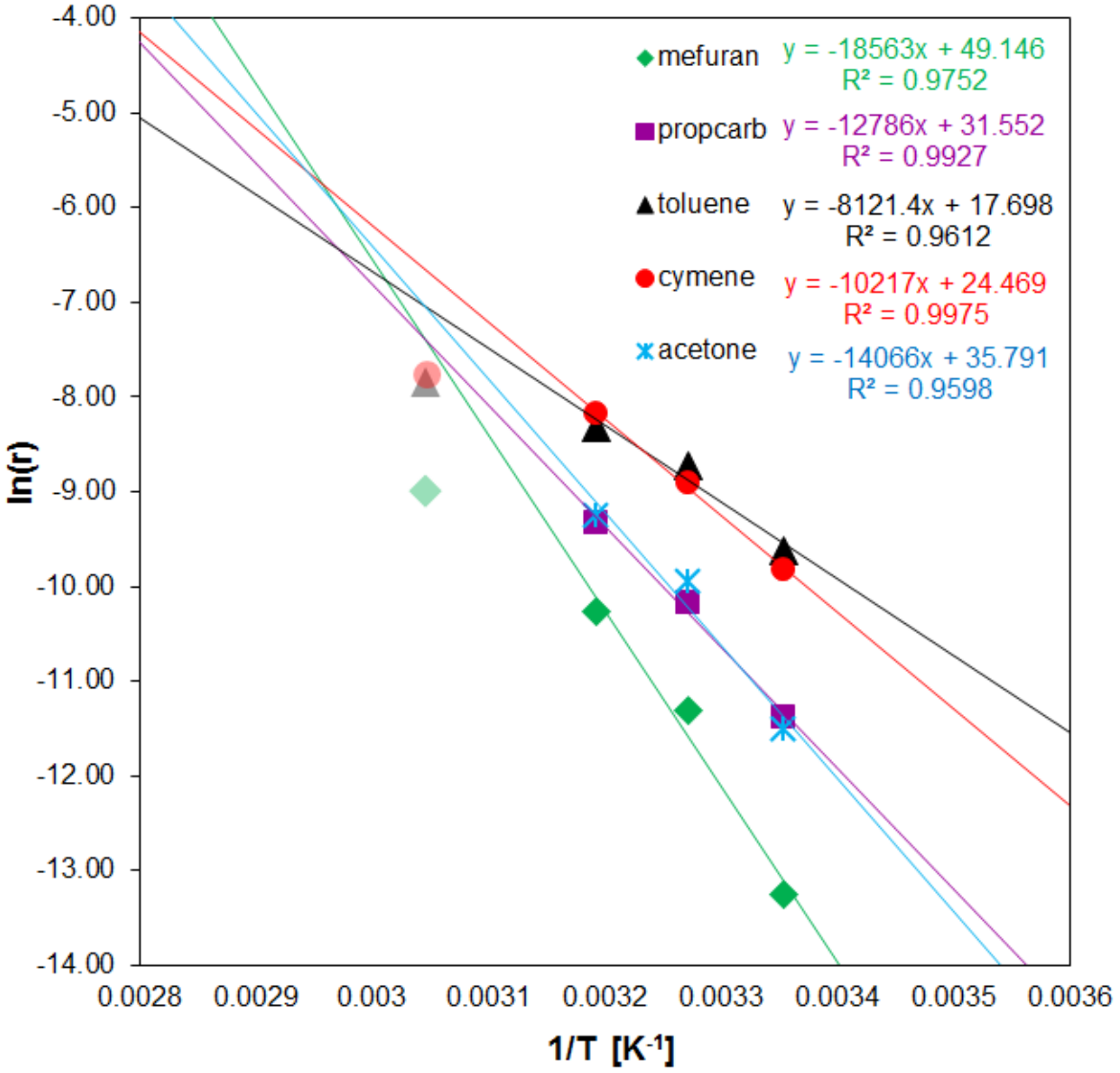


Figure 4.2: Arrhenius-type plot representing the relationship between logarithm of rate constant and the reciprocal of temperature. The intersection of the five solvent lines identifies the isokinetic point.

The above experimentally derived terms are associated with one solvent and inherently vary depending on it. With poorer performing solvents, such as 2-methylfuran and acetone, even small variations in temperature have a strong impact on the reaction rate. The effect is less pronounced on solvents that exhibit good performance starting from ambient temperature. This difference implies that the lines extrapolated from the data-points converge towards one common point as seen in Figure 4.2, at approximately 0.00295 K^{-1}.

$$T_{iso} \approx 339K = 66^{\circ}C \tag{4.3.4}$$

The presence of a common intersection among a set of Arrhenius plots is not an unusual phenomenon, and it implies that there is a characteristic temperature, defined as isokinetic temperature (T_{iso}), at which the rate of the reaction is independent from any change in the variable on which the plots were screened - in this case the solvent.[163] In order to experimentally validate the isokinetic temperature to confirm the described trends, additional reactions were run at 55°C in three solvents, as represented in Figure 4.2 in lighter colour shade. The newly added data-points were found to lie at significantly lower reaction rates than expected from the extrapolation. This drift may have been anticipated since at this temperature the enzyme is forced to the limit of its activity range - *ca.* 60°C. At 55°C the thermodynamic benefit from raising the temperature of reaction may be countered by degradation of catalytic activity, typically caused by changes in the enzyme's conformation through gradual unfolding of the protein.[90] Since the thermostability of the enzyme hindered the direct confirmation of the isokinetic effect outside of the narrow temperature range already tested, the phenomenon was validated by repeating the screening on a different model compound - see Section 4.6.

4.4 Thermodynamic and mechanistic considerations

A theoretical extension of the Arrhenius equation 4.3.1 can be carried out on the basis of transition state theory, expressed by the following Arrhenius-like relationship called Eyring equation.[161]

$$r = c \cdot \frac{k_B T}{h} \cdot e^{-\frac{\Delta G‡}{RT}} \tag{4.4.1}$$

$$\therefore ln(r) = ln\left[c \cdot \frac{k_B T}{h}\right] - \frac{\Delta G^‡}{RT} \tag{4.4.2}$$

In equations 4.4.1 and 4.4.2 where $\Delta G^‡$ is the free energy of activation representing the thermodynamic spontaneity of the process, k_B is Boltzmann's constant, h is Planck's

constant, and c is the concentration of the activated catalytic complex. The presence of $\Delta G^{\ddagger}$ in the exponential factor in place of the experimental E_a changes the overall thermodynamic meaning of the expression from empirical to mechanistic.[161]

According to the transition state theory, for a reaction to occur it is necessary to go via an intermediate stage, called the transition state. The transition state is a high-energy activated-complex which is in equilibrium with the starting materials while also close to the product conformation. This condition is called a 'quasi-equilibrium'.[161] This theory can be applied to the Arrhenius trends based on the assumption that the formation of the activated complex is the rate-limiting step of the process and determines the experimentally observed rate constant. This assumption finds support with Michaelis-Menten theory of enzyme kinetics. Under this theory, the rate-determining step is the formation of the activated complex (EA) between the enzyme (E) and the binding reagent (A). The scheme presented in Figure 4.3 provides a representation of the widely accepted step-by-step mechanism of lipase catalysis, highlighting the rate-determining step in the formation of Michaelis-Menten complex, EA.[93,164]

Overall, the thermodynamic definition of $\Delta G^{\ddagger}$ encompasses contributions of activation enthalpy, $\Delta H^{\ddagger}$, and activation entropy, $\Delta S^{\ddagger}$ - as described in equation 4.4.3 - which can be used to further detail the relationship. As a consequence, the Eyring equation 4.4.2 can be updated to include this definition.[161] Equation 4.4.5 presents the resulting relationship where 1/T is isolated.

$$\Delta G^{\ddagger} = \Delta H^{\ddagger} - T\Delta S^{\ddagger} \tag{4.4.3}$$

$$ln(r) = ln\left[c \cdot \frac{k_B T}{h}\right] - \frac{\Delta H^{\ddagger} - T\Delta S^{\ddagger}}{RT} \tag{4.4.4}$$

$$ln(r) = \left\{ln\left[c \cdot \frac{k_B T}{h}\right] + \frac{\Delta S^{\ddagger}}{R}\right\} - \frac{\Delta H^{\ddagger}}{R}\left(\frac{1}{T}\right) \tag{4.4.5}$$

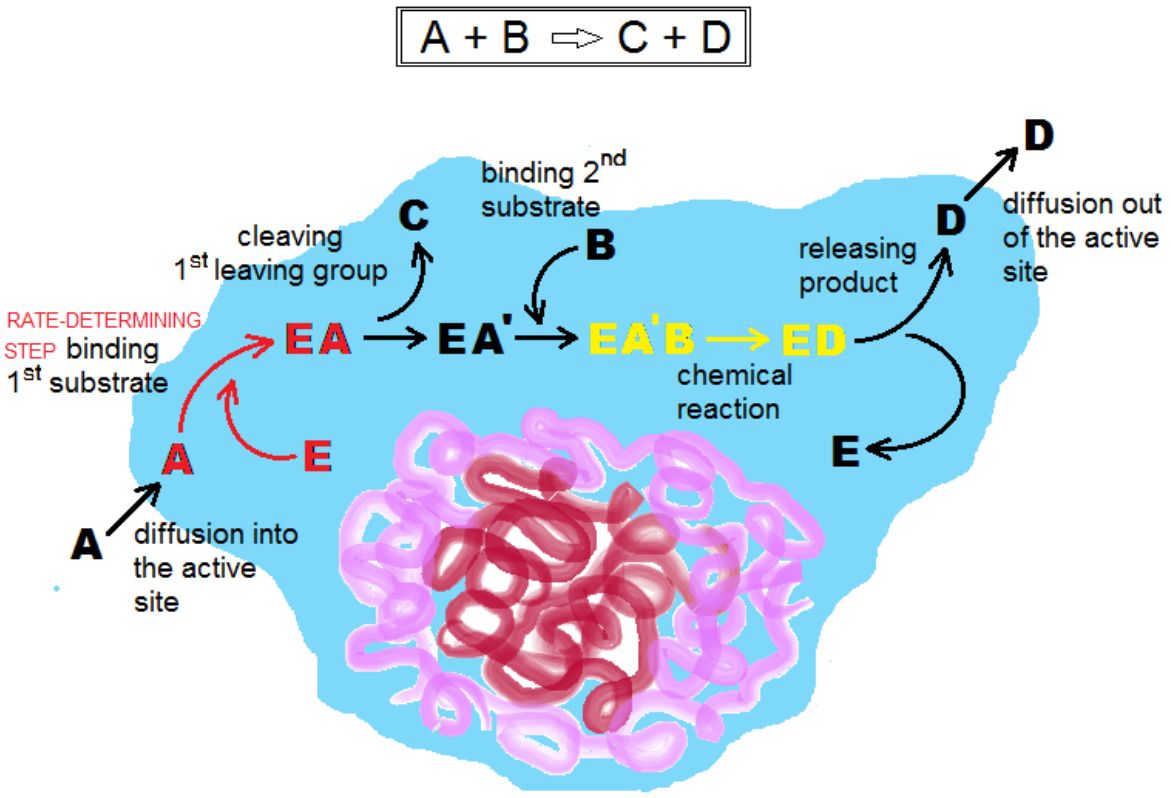


Figure 4.3: Scheme of mechanism of lipase catalysis according to Michaelis-Menten kinetics. Scheme adapted from reference.[164]

Through this elaboration it is possible to bring further insight into the meaning of the variables derived from the Arrhenius plot. In this manner, the slope and intercept obtained for each solvent (Table 4.2) can be re-interpreted in terms of thermodynamic energy with added mechanistic value: the slope of the Arrhenius plots become a measure of the activation enthalpy of the reaction in each solvent (equations 4.4.6 and 4.4.7); while the intercept provides an account of the activation entropy of reaction (equations 4.4.8 and 4.4.9).

$$Slope_i = -\frac{\Delta H_i^{\ddagger}}{R} \tag{4.4.6}$$

$$\therefore \Delta H_i^{\ddagger} = -R \cdot slope_i \tag{4.4.7}$$

$$Intercept_i = ln\left[c \cdot \frac{k_B T}{h}\right] + \frac{\Delta S_i^{\ddagger}}{R} \tag{4.4.8}$$

$$\therefore \Delta S_i^{\ddagger} = (constant) + R \cdot intercept_i \tag{4.4.9}$$

4.5 Compensation effect

Following this analysis for each solvent, the Arrhenius-derived data can be re-labeled according to the above definitions as presented in Table 4.3. Of note is that solvents bringing higher activation enthalpy to the process are also those associated with higher entropy figures. Likewise, the lower $\Delta H^{\ddagger}$ the lower the corresponding $\Delta S^{\ddagger}$, resulting in overall free activation energies $\Delta G^{\ddagger}$ that are relatively constant across the set of solvents. In order to investigate this relationship further the two energy variables were plotted against one another in Figure 4.4.

Table 4.3: Energy parameters derived from Arrhenius plot re-elaborated according to the transition state theory and Free Gibbs Energy relationship.

Solvent	$\Delta H^{\ddagger}$ [kJ mol^{-1}]	σ_D	$\Delta S^{\ddagger}$ [kJ mol^{-1}K^{-1}]	σ_D	$\Delta G^{\ddagger}$ at T_{iso} [kJ mol^{-1}]
Toluene	67.52	13.56	0.1516	0.0444	16.89
p-Cymene	84.94	4.21	0.2034	0.0138	14.96
Acetone	116.90	18.94	0.2975	0.0617	14.58
2-Mefuran	154.30	24.60	0.4003	0.0806	16.62
Prop.carb.	106.30	9.14	0.2624	0.0299	16.07

The resulting graph shows a strong linear relationship between the two activation energies, which is often referred to as enthalpy-entropy compensation effect. This phenomenon denotes that imposed changes in the reaction conditions - *i.e.* the solvent - impact evenly on the two energy-related properties such that they vary under the same rule, one directly compensating variations in the other.[165] As a consequence, the nature of this energy compensation implies that no difference in free energy occurs when changing solvent in the reaction. $\Delta G^{\ddagger}$ values calculated for each solvent in Table 4.3 show negligible difference between any two solvent i and j, in line with what expected from the compensation cri-

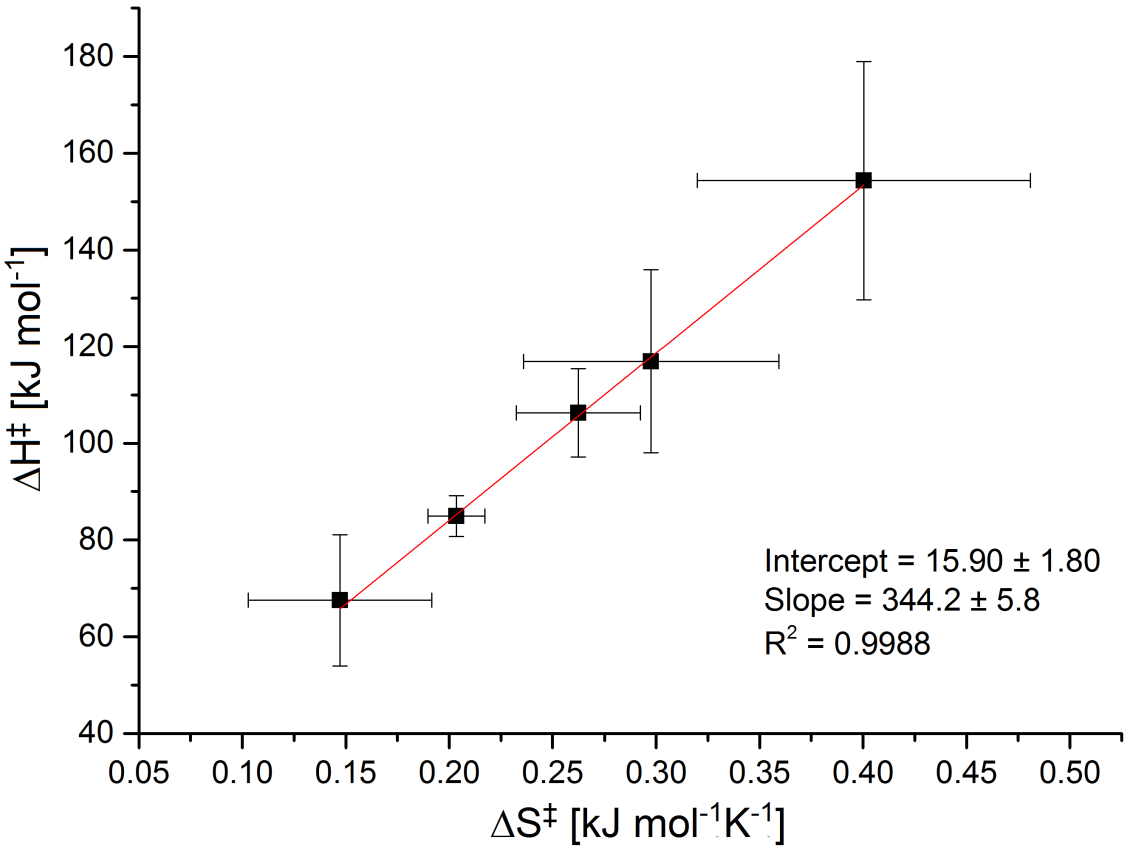


Figure 4.4: Enthalpy of activation ($\Delta H^{\ddagger}$) and entropy of activation ($\Delta S^{\ddagger}$) determined for 5 solvents show a clear compensation effect. Linear fitting carried out with direct weighing of each data-point's standard deviation.

teria of equation 4.5.1).[165] This peculiar phenomenon was reported in a number of fields, including the following relevant to this study: solvation thermodynamics, protein binding, and various enzyme catalysed reactions.[166]

$$\Delta\Delta G^{\ddagger}_{i,j} = \Delta\Delta H^{\ddagger}_{i,j} - T\Delta\Delta S^{\ddagger}_{i,j} = 0 \tag{4.5.1}$$

4.5.1 Linear Free-Energy Relationship

Finally, the common rule under which both ΔH and ΔS vary depending on the solvent may be represented as a Linear Free-Energy Relationship (LFER) based on the generalised equation 4.5.2. This expression can be used to describe the energy compensation relationship of Figure 4.4, where α is the intercept of the compensation line crossing the enthalpy axis at $\Delta S^{\ddagger} = 0$, and β is the slope of the line.[165] By substituting the real values, it can be noticed that β bears temperature units, and its value corresponds to the so-called compen-

sation temperature of the process, T_c, expressed in equation 4.5.4.

$$\Delta H = \alpha - \beta \Delta S \tag{4.5.2}$$

$$\alpha = +15.90 kJ mol^{-1} \tag{4.5.3}$$

$$\beta = T_c = (344.2 \pm 5.8) K = (71.1 \pm 5.8)^{\circ} C \tag{4.5.4}$$

$$\Delta H_i^{\ddagger} = 15.90 kJ mol^{-1} - 344.2 K \cdot \Delta S_i^{\ddagger} \tag{4.5.5}$$

The fact that this temperature is not significantly dissimilar from the isokinetic temperature of the process ($T_{iso} \approx 66^{\circ}$C, from equation 4.3.4) is no coincidence. In fact in a perfect system - *i.e.* free of experimental and statistical error - compensation temperature and isokinetic temperature would be expected to coincide. While isokinetic behaviour does not imply compensation and viceversa, in cases when both effects are observed it is not unusual to find close relationship between T_{iso} and T_c.[163,165]

In conclusion, as a consequence of the linear free energy relationship, it can be demonstrated that a solvent screening carried out at temperature T = β (*i.e.* equivalent to T_{iso} and T_c) would yield no significant variation in rate of reaction across the different solvents. To show this, the LFER relationship 4.5.2 can be introduced in the Eyring law as illustrated in equations 4.5.6 and 4.5.7:

$$ln(r) = ln\left[c \cdot \frac{k_B T}{h}\right] + \frac{\Delta S^{\ddagger}}{R} - \frac{\Delta H^{\ddagger}}{RT} \tag{4.5.6}$$

$$ln(r) = ln\left[c \cdot \frac{k_B T}{h}\right] + \frac{\Delta S^{\ddagger}}{R} - \frac{(\alpha + \beta \Delta S^{\ddagger})}{RT} \tag{4.5.7}$$

By introducing temperature T = β it becomes possible to simplify the expression as shown below in equation 4.7.1. Strikingly, equation 4.7.1 contains all constant values that

are independent of the solvent. By substituting the known constants as shown in equation 4.5.9, the unique ln(r) value is found which is common to all solvents at T = 344.2 K (equation 4.5.10). By confronting this result with the Arrhenius plot in Figure 4.2, it can be seen that at T = 344.2 K, the Arrhenius coordinates identify indeed the crossing point of all solvent lines which is in the range of ln(r) = -6. This result supports the validity of this rationale and the transition state theory for describing the reaction system.

$$ln(r_{T=\beta}) = ln\left[c \cdot \frac{k_B\beta}{h}\right] + \frac{\Delta S^{\ddagger}}{R} - \frac{(\alpha + \beta\Delta S^{\ddagger})}{R\beta} = ln\left[c\frac{k_B\beta}{h}\right] - \frac{\alpha}{R\beta} \tag{4.5.8}$$

$$ln(r_{344.2K}) = ln\left[0.058 \cdot \frac{1.38 \cdot 10^{-23} \cdot 344.2}{6.62 \cdot 10^{-34}}\right] - \frac{15.90 \cdot 10^{3}}{8.314 \cdot 344.2} \tag{4.5.9}$$

$$\therefore ln(r_{344.2K}) = ln(r_{T=\beta}) - 6.97 \tag{4.5.10}$$

4.6 Validation with model compound

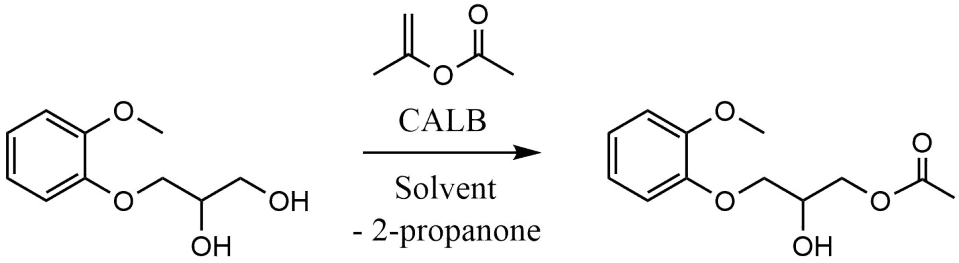


Figure 4.5: Acetylation of guaiacol glyceryl ether (GGE) catalysed by CALB in organic solvents. Conditions: 58 mM substrate, 2 eq isopropenyl acetate, 2 mg CALB Novozyme435, 3 ml solvent.

In order to confirm the results and effects presented above, validation was sought on a model compound, which was subjected to the same screening experiments in four solvents: toluene, *p*-cymene, 2-MeTHF, and DMI. The model substrate guaiacol glyceryl ether (GGE) was chosen for its similarity with the Reboxetine core and its diol functionality. The transformation of study and reaction conditions were equivalent to the those with Reboxetine diol, as shown in Figure 4.5. Herein is presented a summary of the outcomes obtained

replicating the thermodynamic methodology with this substrate. Additional information can be found in Appendix A.

Table 4.4: Collection of data from temperature profiles in individual solvents for the model substrate GGE.

	ln(rate)		
Solvent	at 30°C	at 40°C	at 50°C
Toluene	-8.39	-7.57	-7.25
p-Cymene	-8.71	-8.08	-7.36
2-MeTHF	-8.19	-7.45	-7.01
DMI	-10.96	-9.93	-8.72

In order to develop the Arrhenius' plot for this process, four solvents were screened at three temperatures: 30°C, 40°C and 50°C. Individual rate constants of each reaction were gathered following the linearisation method described earlier and are reported in Table 4.4. Overall, esterification of GGE appeared slightly slower than Reboxetine. The processed conversion curves showed good statistical regressions in the first two hours of reaction across the range, and confirmed both the endothermic behaviour and the first order kinetics also for this set.

In Figure 4.6 the results of this analysis are presented in the form of an Arrhenius plot. Similarly to the Reboxetine case, each solvent line showed a characteristic gradient, with DMI displaying the steepest line, and 2-MeTHF outperforming all other solvents at all temperatures screened. All lines converged at a common intersection point at approximately x = 0.0028 K^{-1} which confirms the isokinetic effect in this case study and identifies the isokinetic temperature of the process to be slightly higher than that obtained with Reboxetine (equation 4.6.1).

$$T_{iso,GGE} \approx 357K = 84^{\circ}C \tag{4.6.1}$$

Moving straight from this representation to the application of the transition state theory, it was possible to obtain the enthalpy and entropy of activation of the transformation.

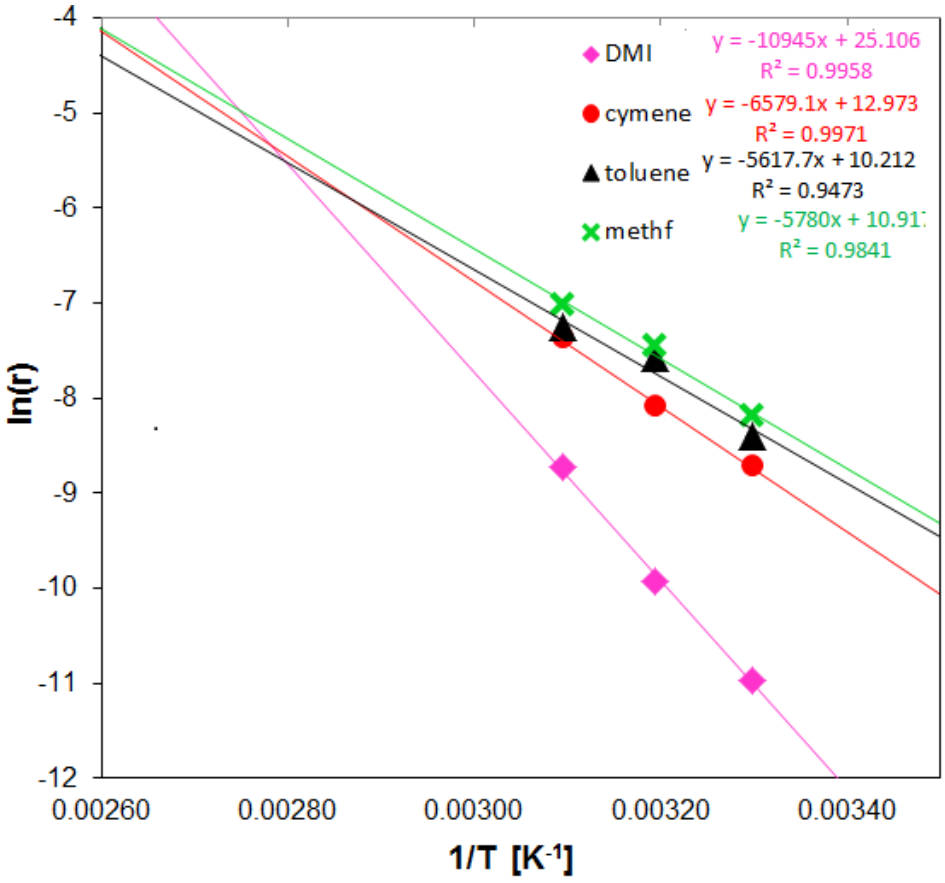


Figure 4.6: Arrhenius' plot for GGE acetylation showing the effect of changes in temperature on the reaction rate.

Table 4.5 reports these results and corresponding statistical errors. The calculation of $\Delta G^{\ddagger}$ for T = T_{iso} yielded values that were remarkably constant across the solvent set, as well as very close to those obtained with Reboxetine. In both case-studies, $\Delta G^{\ddagger}$ was positive and approximately 16 kJ mol^{-1}, indicating that the process is endoergonic. Once again, enthalpy provides the positive contribution which determines the positive $\Delta G^{\ddagger}$. On the other hand, entropy provides a countering effect, which changes proportionally with all changes in enthalpy. This proved once again the energy compensation behaviour, as found with Reboxetine. Figure 4.7 shows the strong linear correlation between $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$.

As discussed above, the compensation temperature T_c can be derived from the gradient of the compensation line, and can demonstrated to be statistically equivalent to the isokinetic temperature identified above - within standard deviation error.

$$T_{c,GGE} = (359.5 \pm 4.3)K = (86.4 \pm 4.3)^{\circ}C \approx T_{iso,GGE} = 84^{\circ}C \tag{4.6.2}$$

Table 4.5: Enthalpy-entropy compensation effect confirmed on model substrate.

Solvent	$\Delta H^{\ddagger}$ [kJ mol^{-1}]	σ_D	$\Delta S^{\ddagger}$ [kJ mol^{-1}K^{-1}]	σ_D	$\Delta G^{\ddagger}$ at $T_{iso,GGE}$ [kJ mol^{-1}]
Toluene	46.71	11.01	0.0849	0.0353	16.40
p-Cymene	54.70	2.97	0.1078	0.0095	16.20
MeTHF	48.05	6.10	0.0908	0.0195	15.64
DMI	90.99	5.88	0.2088	0.0188	16.47

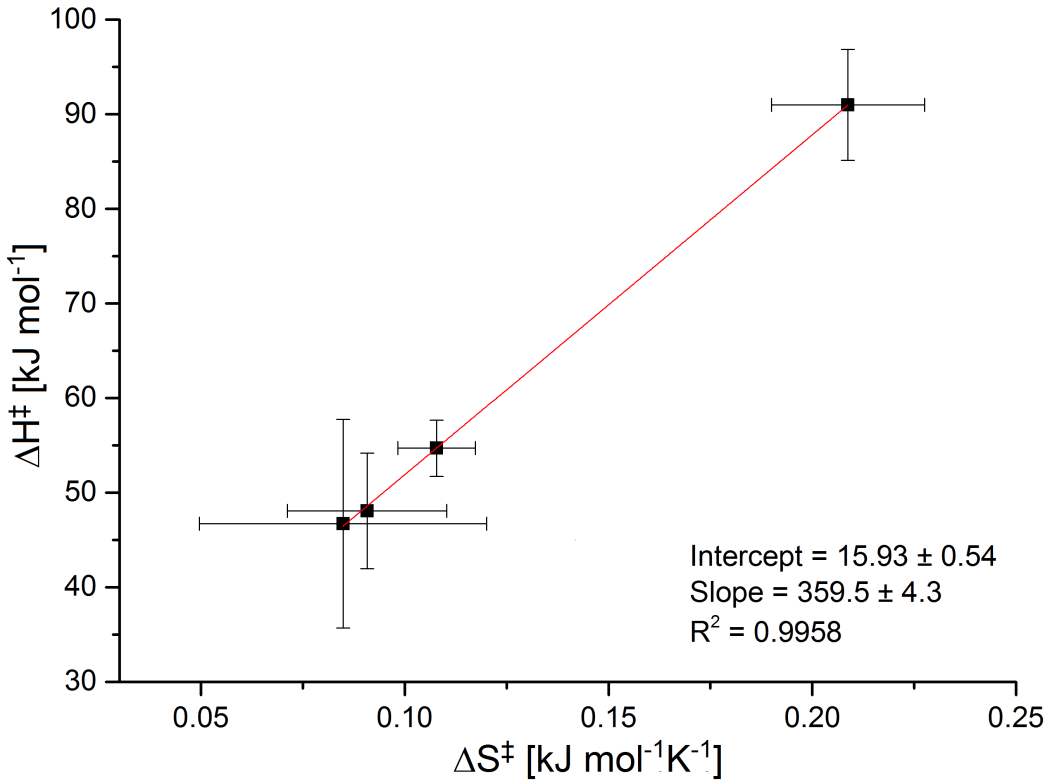


Figure 4.7: Relationship between enthalpy and entropy of activation for the reaction on model compound GGE. Linear fitting carried out with direct weighing of each datapoint's standard deviation.

The compensation temperature detailed in equation 4.6.2 is 15.3°C higher than the one identified for Reboxetine (see equation 4.5.4). This substantiates the inherent substrate dependency of the figures obtained, yet proving the overall consistency of the behaviours and trends described.

4.7 Discussion of effects

In both the Reboxetine and GGE case studies, the investigation over solvent-related thermodynamic effects led to the identification and confirmation of two peculiar phenomena. The observed effects can be confidently attributed to the interaction of the medium with the chemical components and active reaction complexes. In the Section herein, are discussed the meaning and theoretical grounds of the isokinetic and compensation effects, including the different criteria typically applied to judge their validity and significance. In conclusion, an interpretation of solvent effects is proposed based on the trends gathered through a combination of both studies.

4.7.1 Conditions for real isokinetic effect

The isokinetic effect is the product of overlaying multiple Arrhenius' lines related to a set of closely-related chemical reactions. The natural variation throughout the set can be found in terms of a) gradient, inevitably identifying at least one intersection point, or b) intercept only, where lines are distributed in a parallel fashion. It is important to highlight that both behaviours can lead to energy compensation effect, but only the first one can define isokinetic effect, *i.e.* when all best-fit lines cross at one common intersection.[165]

In both case-studies discussed in this Chapter, an isokinetic point was identified in a temperature region outside of the linear working range of the enzyme and higher than the range of screening applied. This satisfies one of the criteria for real isokinetic relationship, which states that the isokinetic temperature of the system must be significantly different from the average of the temperature range tested. While the definition of what can be considered 'significantly different' is loose and can be hard to delimit, especially for chemical process that can only be studied within a narrow window of activity - *e.g.* enzyme catalysed reactions. This requirement appears to be satisfied, and the isokinetic effect to be a valid observation, as shown by values reported in Table 4.6.[166]

Figure 4.8 provides a graphical summary of the overall trends gathered from both Arrhenius plots of two case-studies, herein combined in a simplified fashion for the purpose

Table 4.6: Comparison of characteristic temperatures of the process of study.

Case-study	Average T of range	T_{iso}
Reboxetine	33°C	66°C
GGE	40°C	84°C

of this analysis. As discussed and demonstrated earlier by applying the linear free-energy relationship, at temperature T = β = T_{iso} all reactions are expected to exhibit the same rate constant regardless of the solvent employed, by satisfying the equation below.

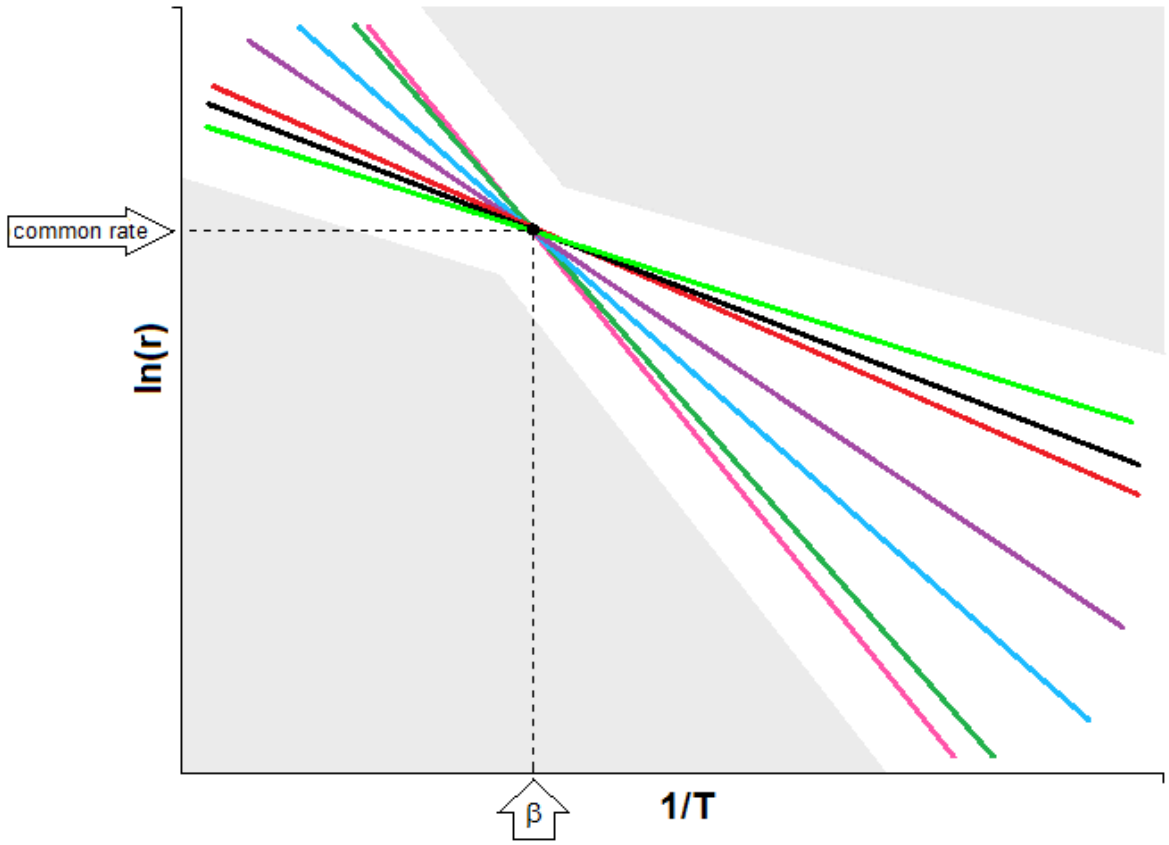


Figure 4.8: Combined trends of isokinetic effect observed with Reboxetine and GGE.

$$ln(r_\beta) = ln\left[c \cdot \frac{k_B \beta}{h}\right] - \frac{\alpha}{R\beta} \tag{4.7.1}$$

The existence of such isokinetic effect suggests that the impact of solvent effects decreases gradually with increasing temperatures, giving in to the greater impact of the thermic effect. Higher diversification of effects can be observed at lower temperatures, and, in theory, also at temperatures above the isokinetic point. However, as the latter is not com-

patible with the nature of this process, it could not be physically achieved and assessed. In addition, it should be noticed that high rate solvents - such as 2-MeTHF, toluene, *p*-cymene - present the lowest gradient lines on the Arrhenius' plot, thus suggesting that fast reactions are less subject to variations induced by temperature.

4.7.2 Conditions for real compensation effect

As already demonstrated in this Chapter, if the Arrhenius' law and transition state theory hold, it is reasonable to extract free energy factors from the Arrhenius plot. In both Reboxetine and GGE examples, a clear compensation effect was observed between activation enthalpy and activation entropy of the reaction.

In the scientific literature, energy compensation is a highly disputed topic and there appears to be a high level of scepticism over its scientific meaning and legitimacy.[167] Liu *et al.* presented a review of the topic, highlighting theoretical and factual support for both sides of the argument.[165] In general, concerns around claims of energy compensation surround the risk of random association and artefactual linearity affecting the data. Notably, in 1961, Petersen *et al.* demonstrated that an existing correlation between the experimental error associated with $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ can often be misinterpreted as energy compensation. It was proposed that a simple criteria to prevent this risk is to verify that the slope of the compensation line (T_c) is significantly different from the result of equation 4.7.2, representing a relationship driven simply by statistical error.[168] It appears, as reported in Table 4.7, that Petersen's criterion was satisfied in both case-studies, considering that the 'temperature of false compensation', $T_{non-compensation}$, falls far from the actual compensation temperature measured, notably at the opposite end of the screening range.

$$T_{non-compensation} = \frac{2T_{min}T_{max}}{3T_{min} - T_{max}} \tag{4.7.2}$$

Another test was recommended by Liu *et al.* as a general rule. It consists in representing error bars associated with experimental energy values, for they should never be larger than the distance between the two farthest data-points. As the authors put it, *"explicitly drawing*

Table 4.7: Confirmation of genuine compensation against artefactual error association.

Case-study	T_{min}	T_{max}	$T_{non-compensation}$	T_c
Reboxetine	25°C	40°C	17°C	66°C
GGE	30°C	50°C	21°C	86°C

the error bars (or confidence region) in the correlation diagram is a simple, clear, and correct test for the existence of the compensation effect".[165]

Compliance with both criteria appears encouraging for the proposed interpretation of the results, as shown in Figures 4.4 and 4.7. Nevertheless, it is important to be aware of the possibility of the results not being a significant representation of the broader thermodynamic behaviour of the system due to the limitations in the temperature range available for testing.

In conclusion, an overall analysis of the observed compensation behaviours from a thermodynamic perspective highlights that the solvents with the highest enthalpy/entropy values were those yielding the lowest reaction performance. Typically, high and positive enthalpy values represent a barrier to a chemical reaction, indicating that the balance between formation and cleavage of chemical bonds through the reaction is energetically disfavoured. In contrast, high and positive entropy contributions tend to benefit a process, as they are representative of more degrees of freedom and higher disorder in the reaction environment. These trends are illustrated in Figure 4.9 as arrows along the axes. In the scheme, are represented the solvents screened in both case-studies in relative position to one other according to the trends observed. As indicated by the central arrow, the most favourable conditions were created by those solvents lying in the lower left corner of the graph, associated with low enthalpy and low entropy changes in the system. From this observation, it is possible to gather that the process of study is enthalpy-driven. Therefore, the best solvents for the reaction are those providing the lowest activation enthalpy.

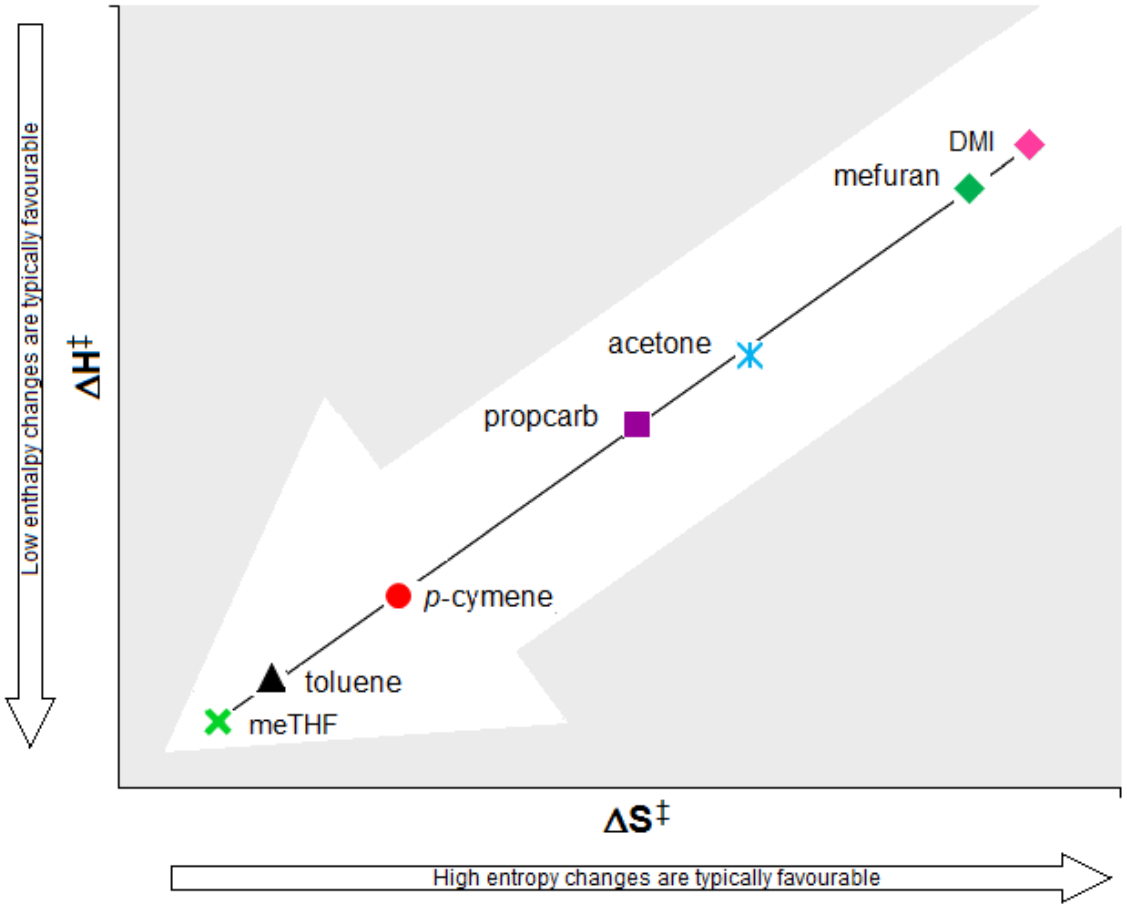


Figure 4.9: Combined trends of compensation effect observed with Reboxetine and GGE.

4.8 Chapter summary

Among the solvents screened in this study, 2-MeTHF, toluene and *p*-cymene confirmed best performance. From this study, it emerged that this favourable behaviour may be linked to the lower enthalpic barrier imposed by these solvents on the overall energy of reaction, whilst the overall free energy of activation remains unchanged across solvents. However, the origin and nature of how this is achieved is not fully understood.

Figure 4.10 illustrates the thermodynamic meaning of the results. In order to maintain $\Delta G^{\ddagger}$ constant across different solvents, the energy difference between reagents and transition state must remain constant, as represented in the scheme. This would mean that the interaction between solvent-substrate, solvent-enzyme, and solvent-transition state is not a discriminating factor in the catalytic performance and does not alter the energy barrier of the reaction. This is justified by the fact that the best solvents for this process were the least

polar and unable engage in hydrogen bonding, indicating that their interaction with the protic substrates and charged transition complex is unlikely.

However, the experiments showed that the driver for the reaction was the ability of some solvents to lower $\Delta H^{\ddagger}$ of the reaction at the expense of $\Delta H^{\ddagger}$. $\Delta H^{\ddagger}$ is represented in Figure 4.10 as the difference in energy between the product and the starting materials. In this context, a lower $\Delta H^{\ddagger}$ of the reaction would only be possible if the solvent exercised a stabilising effect on the product. This conclusion is in line with early observations made by Laane *et al.* on the importance of prioritising product rather than substrate solubility in medium optimisation.[104] This theory is now substantiated by thermodynamic evidence. Further work would be needed in order to validate the proposed mechanism of solvent influence on the process. Moreover, future research should extend this work to other biocatalysed transformations, especially those where the application of greener organic solvents is required.

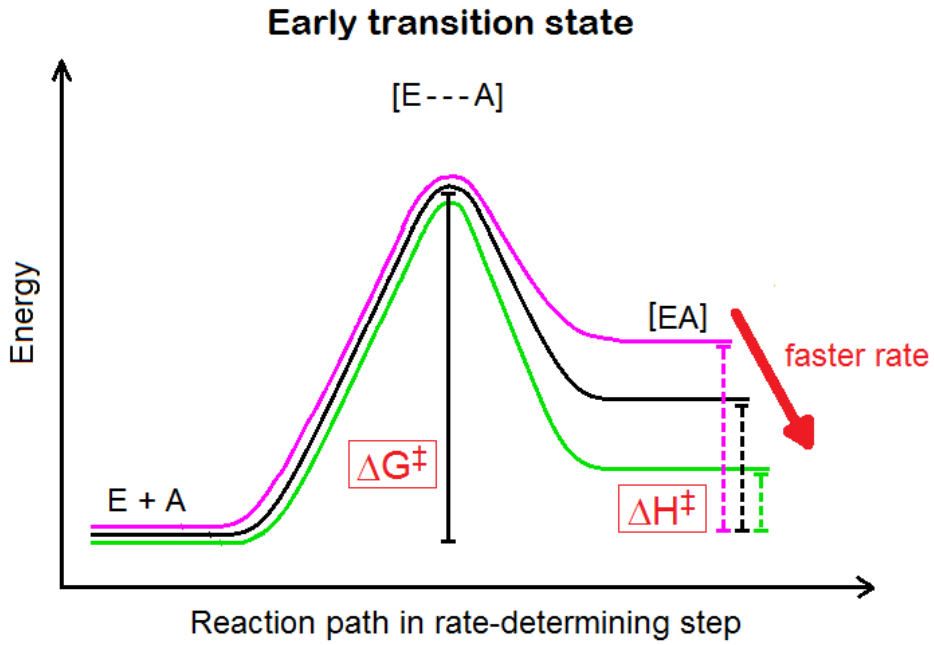


Figure 4.10: Scheme of mechanistic interpretation of observed solvent effects reflecting the observed trends of energy compensation. In green a representative a high performing solvent, such as 2-MeTHF, in pink a low performing solvent, such as DMI.

Chapter 5

Survey of barriers and priorities for the adoption of green solvents

5.1 Introduction

The work discussed in previous chapters provided encouraging proof-of-concept for the use of bio-based solvents in biocatalytic reactions. Citrus-derived solvents such as limonene and *p*-cymene performed best in the reactions of study, both in terms of yield and from the sustainability assessment. Nonetheless, the wider adoption of these solvents for application in industry may face many obstacles, especially in relation to reported difficulties in downstream processing, issues with quality specifications, higher cost of purchase or, arguably, simple inertia to change.[169] Despite these challenges being widely acknoweledged, there appears to be little research aimed at investigating and documenting them. It appears that only one study has addressed this topic in an all-encompassing manner, yet from a general chemicals perspective rather without a specific focus on solvents. In 2010, the UK Chemical Stakeholder Forum reported some of the barriers and challenges faced by their industry members in trying to shift from petrochemicals resources and products to biochemicals.[169]

Sutherland *et al.* described various examples where collaborative approaches were used towards the identification of barriers and priorities in science and policy.[170] For instance, in the field of environmental science, Boxall *et al.* used a 'key questions' approach to identify the top research questions perceived by experts on the issue of pharmaceuticals and personal care products in the environment. They designed a workshop to gather initial inputs from academics, government and businesses, which were later categorised and ranked following the workshop. The work was aimed at directing future research efforts in the field towards areas of most pressing need.[171] Similarly, Fleishman *et al.* and Rudd *et al.* targeted priorities for conservation policy through 'horizon scanning' exercises.[172,173] In another publication, it was stressed the strong need for more studies of this sort in a call for more concerted efforts towards addressing environmental and sustainability challenges.[174]

This goal and approach could be extended to help address green chemistry challenges. Most notably, it could help: 1) spot early warning signs of forthcoming issues in the field, 2) identify threats and opportunities for new green chemistry technologies, 3) join research efforts to speed up the development of solutions, and 4) engage with policymakers to en-

able change on a scientific evidence basis.[175] In this Chapter is presented a first example of a participatory methodology applied to the identification of the top barriers and perceived priorities within one critical area of green chemistry: the adoption of green solvents.

5.2 Survey methodology

For this study, a 1-hour workshop was hosted as part of a solvent-themed industrial engagement event at the Green Chemistry Centre of Excellence (GCCE) in collaboration with partners from the Stockholm Environment Institute (SEI-Y). The participatory workshop aimed to collect both qualitative and quantitative data on barriers perceived by participants for the adoption of green solvents.

5.2.1 Sample population

The audience of participants included 34 diverse stakeholders from the solvent sector, including:

- 21 academic researchers from UK Universities;
- 13 non-academic experts, of which 10 representatives from European industries and 3 from other organisations such as consultancies and government bodies.

Although the population sampled in this survey was limited and shall not be considered 100% representative of the wider opinion, it provided insight into the perception of a diverse and substantial pool of experts, and consisted of a novel opinion-gathering opportunity on this topic.

5.2.2 Roles

On the day, the workshop was run by the organisers covering specific roles. One person was responsible for leading the various activities and communicating with the audience throughout the workshop. A time-keeper was assigned to ensure the planned structure was respected and the methodology was followed consistently. In addition, six facilitators

helped with the group discussions, with the role to 1) encourage interaction amongst the groups, 2) maintain the conversation on track, and 3) report the outcomes of each group discussion to all participants.

5.2.3 Structure of the workshop

The workshop followed the structure and timings presented in Table 5.1 and detailed in individual sections below. Key to the design of the structure and methodology was to make it applicable to standardisation, in order to allow reproducibility and future comparisons with other similar exercises.

Table 5.1: Structure of the workshop.

#	Session	Time (min)
1	Introduction	4
2	Preliminary individual brainstorming	1
3	Facilitated group brainstorming	20
4	Clustering	10
5	Voting on priorities	10
6	Conclusion	5

Phase 1: Introduction

Firstly, the lead organiser (*i.e.* author) gave an initial presentation to the audience, providing context and introduction to the planned activities. In this phase, the participants were presented with the main objective of the workshop.

Phase 2: Individual brainstorming

Participants were then given the question statement which kicked-off the brainstorming session.

"The aim of the exercise is to identify the key barriers for green solvents becoming common use

across sectors. In other words, what are the biggest barriers facing you and your sector in the transition to green solvents?"

It is critical in such studies that scope and research question under discussion are structured and communicated in a fully comprehensive and unambiguous manner, especially considering the diversified background of the audience. For this reason, a selection process was carried out, where various wordings and terminologies were considered and discussed amongst the organisers prior to the event. For instance, it was decided to avoid the phrasing 'barriers to the *use* of green solvents', in order to not limit the target to users only. Furthermore, it was decided to propose the statement in a first person form, for the audience to be more inclined to input responses based on their personal experience rather than based on general knowledge or popular challenges discussed in the field.

Participants were first invited to brainstorm on the given subject individually. They were provided sticky notes and allowed one minute to write down their three most immediate answers to the given research statement. All note entries were anonymous. This activity was used as warm-up, but also with the aim to gather the most genuine responses from participants, limiting the influence of any conscious censoring or social bias.

Phase 3: Facilitated group discussion

This phase was core to the workshop. Participants were invited to share and discuss their thoughts as part of heterogeneous break-out groups which were prearranged. They were provided additional sticky notes and clear poster boards for noting down the main topics of each group discussion. One facilitator was present in each group in order to stimulate discussions and make sure that emerging topics were captured onto the poster board. This activity lasted 20 minutes in order to allow exchange of views among participants, thus building from the topics from individual brainstorming with further level of detail.

Phase 4: Clustering of entries

At the end of the group activity, facilitators brought together the responses from each break-out group and worked together to identify overarching themes. In this phase, participants were welcome to feedback on the clustering activity, which engaged them in: 1) identifying common themes from their post-it entries, 2) allocating individual entries to each theme category, 3) defining each category, *i.e.* title and content. In a few cases, the definition of categories changed and developed during the activity, adapting flexibly according to the discussion and inputs. Later processing of the comments revealed some involuntary mis-assignments, an inevitable result of the limited time available for clustering phase. While overall categories and voting outcomes from the original clustering were maintained, reassignment of 5 entries was required. Overall, this activity built the basis for the later voting activity, as shown in Figure 5.1.

Phase 5: Voting on priorities

This phase constituted the second key data-gathering activity within the workshop. Participants were asked to vote on what they perceived as the most pressing priorities to be addressed for favouring the wider adoption of green solvents. To do this, they were provided with five stickers each, colour-coded depending on their type of affiliation: red for academia, silver for industry, green for all others. Voters were allowed to allocate stickers freely in correspondence to the listed categories on the poster boards. The purpose of this activity was to quantify the sense of urgency and priority perceived by participants in relation to the wider barriers to adoption of green solvents.

Phase 6: Conclusion

In the final stage, the number of stickers associated with each category was counted by the organisers and the preliminary results were fed-back to the audience. In this, the scope and wider interest of the survey was re-emphasised, and a summary of the future developments from this work was provided. The workshop was then followed by a debriefing

Figure 5.1: Representative view of a section of the voting board.

amongst the organisers, to highlight main outcomes and lessons learnt and to report any formal and informal feedback received from participants.

5.3 Results from the survey

The data produced during Phases 2, 3, and 5 were analysed in three ways:

1. Qualitative analysis: perception on barriers captured during brainstorming. Participants were encouraged to submit entries based on the barriers to green solvents that affected their professional activity (Phases 2 and 3). This information was analysed for the purpose of a qualitative discussion, based on the topics captured in the entries.

2. Quantitative analysis: perception of priorities captured through voting activity. Participants were provided with 5 stickers each and were asked to vote on what they believed were the wider priorities to be addressed for the benefit of the green solvents sector (Phase 5). This information was used for a quantitative analysis, based on the distribution of voting stickers across the different barrier categories as identified by the clustering of written entries.

3. Semi-quantitative analysis: cross-validation of the two above methods used for capturing the audience perception. The entries assigned to each category during the clustering phase were counted and this number was compared against the number of votes assigned to the same categories during the voting activity. This enabled comparison of the conclusions obtained from the two types of approaches.

5.3.1 Qualitative analysis of perceived barriers

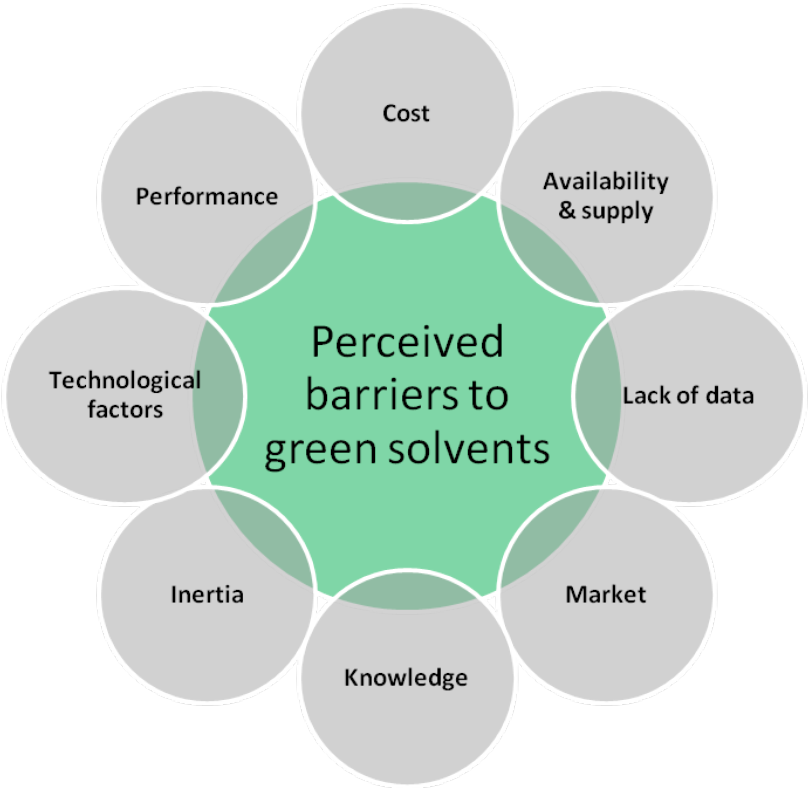


Figure 5.2: Main barrier categories identified through the clustering activity.

From the clustering phase of the workshop eight categories of barriers emerged, each of which grouped several entries produced during the brainstorming activity (Figure 5.2).

Cost

The 'Cost' issue gathered significant attention among participants yielding the highest number of entries. Table 5.2 reports all individual entries submitted in the brainstorming phase, which point a broad definition of 'Cost' as intended by participants. The relatively higher purchase price of green solvents emerged as a barrier, seen as a direct consequence of higher upstream costs (feedstock and processing), often penalised by the smaller scale production. Besides, less evident costs were highlighted, such as investment for development - whether of a solvent or its new application - as well as training of specialists. Both these aspects significantly impact on time and efficiency in the short term. Overall, the financial risk resulted as a point of high concern, especially where it may impact on the chances of regulatory approval and, therefore, on the life-time of a product on the market. These risks are perceived most negatively affecting small and medium enterprises (SMEs).

Table 5.2: List of entries which were clustered into the 'Cost' category.

COST CATEGORY (Total = 27)	
Cost (x8)	Expense
Overall cost on product	Capex of net raw materials
Resource - i.e. time, money, labour to explore potential alternatives	Cost aspect compared with existing solutions
Cost / risk of development of a new solvent	Regulation hurdle = cost (higher for SMEs)
Re-training cost	Time for approval
Cost of new plants	Time for product development
Capital	Economic feasibility
Cost of feedstock	Cost of green solvents
Cheap production	Approval
Finance	Core of production

Lack of data

Under the 'Lack of data' category were grouped all entries pointing at a lack of physico-chemical, toxicological, and environmental impact information for green solvents (Table 5.3). This set of entries captured a special concern with regards to any hazardous long-term health effects on exposure to less known solvents - *i.e.* carcinogenicity, mutagenicity, reprotoxicity (CMR).

Table 5.3: List of entries which were clustered into the 'Lack of data' category.

LACK OF DATA CATEGORY (Total = 15)	
Not enough research	Genotoxicity
Lack of data for regulatory approval	Information to ensure clean CMR profile
Lack of technical information to support claims	Full life-cycle data
Lack of data on substances	Health risk data
Low knowledge	Lack of toxicological data
Data	Safety
Lack of data	Lack of physical chemical property data
Lack of development	

Availability & supply

The issue of 'Availability & supply' encompassed aspects such as availability of bio-feedstocks, production throughput, reliability of supply, diversification of solvent products, and quality thereof. Notably, the risk of single-sourcing solvents that are supplied only by one manufacturer can be alarming especially for highly regulated industries that have to undergo tight compliance and approval procedures. For example, in the pharmaceutical industry there is a necessity to ensure consistency of impurity profiles and therefore of choosing solvents and suppliers responsibly.[176] These factors - reported in Table 5.4 - showed to be of great importance to participants, garnering 26 votes and thus constituting the second most critical barrier perceived by the group.

Table 5.4: List of entries which were clustered into the 'Availability & supply' category.

AVAILABILITY & SUPPLY CATEGORY (Total = 26)	
Availability (x7)	Supplies
Availability - volume	Volume of production
Quality	Reliable availability and quality
Some solvent producers may not be interested to develop alternatives	Not all biobased solvents can be locally supplied/grown
Availability at commercial scale	Clean production from biomass
What about low boiling low polarity?	Consistent quality from variable feedstocks
Availability of green solvents	Availability - risk of single-sourcing
Impurity profile	Sources
Chemical space	Lack of feedstock
Quality + stability issues	Useful applicable properties

Market

The 'Market' category clustered all entries that referred to external forces affecting or threatening the green solvents sector. As presented in Table 5.5 these included trends in competitor sectors (*e.g.* petroleum, shale gas, agriculture), but also encompassed other external factors such as regulation, globalisation, green economy, amongst others.

Performance

Under the realm of 'Performance' were discussed the expectations on green solvents used as substitutes for conventional problematic solvents. Users are not willing to compromise on performance and quality simply on the basis of a greener profile or outcome. Whether in product formulations or used in chemical processes or other applications, green solvents need to stand up to comparison with conventional solvents. However, it was discussed that ticking all the boxes of sustainability, safety and performance proves very challenging, especially with regards to some key solvents for chemistry (Table 5.6).

Table 5.5: List of entries which were clustered into the 'Market' category.

MARKET CATEGORY (Total = 16)	
Market reluctant conservatism	Uncertainty in regulation
Consumers are not ready to pay more	Pushing it to market (scaling)
Shale gas availability	Range of applications
Incumbency (current non-bio manufacturers)	To make the bio-innovation in the right timeframe
Oil subsidies	Market risks
Market forces	Competition for food
Future proofing	Competition for land use
Local bio-based market is more likely to work in developing countries	Understanding the needs for the solvent in the final application

Table 5.6: List of entries which were clustered into the 'Performance' category.

PERFORMANCE CATEGORY (Total = 10)	
Performance (x2)	Solvent as inert material
Process suitability	Replacement for chlorinated solvents
Replacements don't work as well	Properties
Impact of substitution on chemical and organolectic quality	Performance compared to currently used solvent
Recovered solvents not as good as virgin	Difficulty to find which green solvent could replace the actual one in an existing formulation or process
Useful applicable properties	

Technological factors

From discussion with participants, some 'Technological' barriers emerged that were believed to be separate from 'Performance'. These involved mainly re-usability and scale-up, and the establishments of business-to-business networks for recycling waste solvents or biomass into resources, as shown in Table 5.7.

Table 5.7: List of entries which were clustered into the 'Technological factors' category.

TECHNOLOGICAL FACTORS CATEGORY (Total = 7)	
How to recycle them?	Easy to remove
Volatility	Ability to recycle
Complications on work up compared to traditional solvents	If not recyclable should be biodegradable in a standard bio-treatment plant
Uptake by industry (B2B)	

Knowledge

During the participatory clustering activity, it was agreed to split human factors into two categories depending on whether they were related to 1) barriers determined by lack of knowledge, or 2) behavioural inertia driven by deliberate opposition and preconception against green solvents. Although it became clear during the exercise that these categories were likely to show significant overlap, it was decided to maintain the separate distinctions.

In Table 5.8 are listed all entries that were identified as part of the 'Knowledge' category (12 entries). Among general comments of lack of awareness, dissemination and familiarity, were highlighted issues regarding the definition of *green* and the classification of green solvents.

Table 5.8: List of entries which were clustered into the 'Knowledge' category.

KNOWLEDGE CATEGORY (Total = 10)	
Lack of knowledge of alternatives	Low knowledge
What is green? (x2)	General unawareness of green alternatives
Knowledge (what is available, what are the benefits)	For IL a broad spectrum is available but people are not aware
Classificability	Complexity
Knowledge of usability	

Inertia

To complement the above category, clues expressing any intentional resistance or negative perception against the use of green solvents were categorised under 'Inertia'. As captured in Table 5.9, participants pointed out perception, preconception and inertia as barriers to change. Furthermore, it was discussed that symbolic actions are often taken in order to build a green reputation yet are not aimed to bring any significant or wider long-term benefit. This sort of behaviour was talked about as being influenced by a global green movement, which itself is increasingly fuelled by this sort of approach. Further dialogue also pointed towards barriers of unfamiliarity and uneasiness to using new solvents as part of the 'Inertia' category since they may often be independent from personal knowledge and awareness.

Table 5.9: List of entries which were clustered into the 'Inertia' category.

INERTIA CATEGORY (Total = 10)	
Inertia	Preconception
Perception (x2)	Green washing - reputation
Familiarity (x2)	People attitude and behaviour
Unwillingness to move out of comfort zone - *i.e.* happy using conventional solvents	People are not at ease using it

5.3.2 Quantitative analysis of priorities

The results of the later voting exercise carried out in these categories are illustrated in Figure 5.3 and detailed in Table 5.10.

Figure 5.3a presents the distribution of results based on the number of votes counted from the voting activity - absolute figures; Figure 5.3b provides a split view of the same data once normalised based on the number of participants per type of affiliation. In both graphs, votes from academic and non-academic participants are represented respectively in red and in blue. Moreover, a further differentiation is provided within the non-academic affiliation,

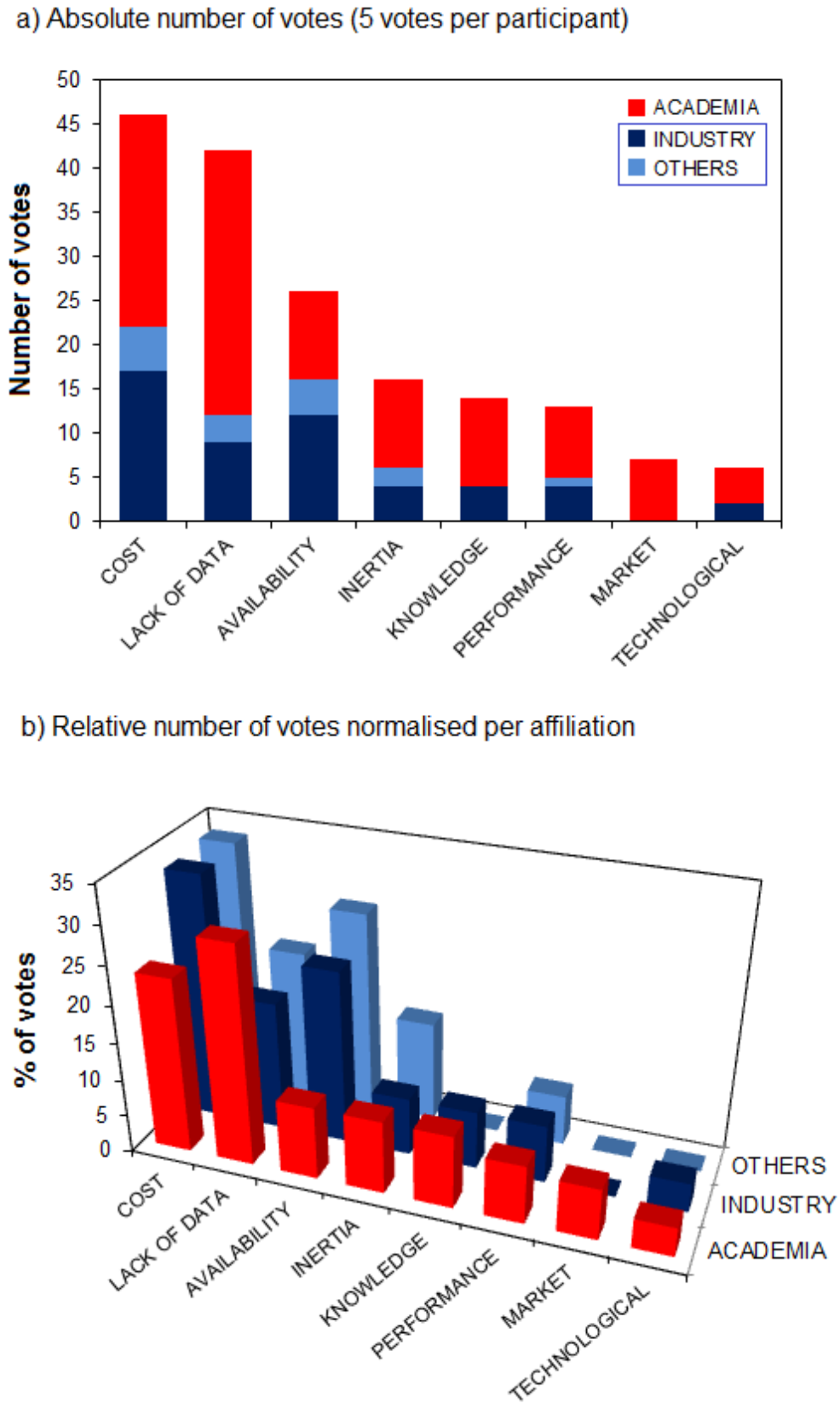


Figure 5.3: Results from voting activity over the priority barriers to be addressed.

Table 5.10: Summary of votes gathered per category sorted between academic and non-academic affiliation of voters. Each voter could assign 5 votes to any category of choice. Numbers indicate the absolute number of votes, while in brackets are reported the corresponding percentages of total for each column.

Categories	Academia	Industry+Others	TOTAL per category
Number of participants	21	10+3	34
Cost	24 (23%)	17+5 (33%)	46 (27%)
Lack of data	30 (29%)	9+3 (18%)	42 (25%)
Availability & supply	10 (10%)	12+4 (24%)	26 (15%)
Inertia	10 (10%)	4+2 (9%)	16 (9%)
Knowledge	10 (10%)	4+0 (6%)	14 (8%)
Performance	8 (8%)	4+1 (7%)	13 (8%)
Market	7 (7%)	0+0 (0%)	7 (4%)
Technological	4 (4%)	2+0 (3%)	6 (4%)
TOTAL per affiliation	103 (100%)	52+15 (100%)	170 (100%)

between members from 'Industry' and from 'Other' institutions and organisations. The latter was reported as a matter of transparency, although due to the very low number of members is not considered representative.

From both charts, it can be seen that the categories 'Cost', 'Lack of data' and 'Availability & supply' emerged as the most pressing issues perceived by the surveyed stakeholders, respectively accounting for 27%, 25% and 15% of total votes. The attention given to each of these categories from members of the three affiliation types was somewhat consistent. While academic participants seem to be most concerned with 'Lack of data', the non-academics assigned higher priority to 'Cost' and 'Availability & supply'. All the lower priority categories, received consistently fewer votes across participants groups. Higher granularity around these may be achieved with a bigger sample population.

Overall, despite the slight differentiation within the top three priorities, the general distribution of votes from academic and non-academic participants was relatively even, suggesting no obvious correlation between the type of affiliation and differences in the perceived priorities. The outcomes emphasise the need for cheaper and more accessible green solvents, accompanied by a more consolidated knowledge on their toxicological and

environmental impact.

5.3.3 Semi-quantitative validation

In order to provide validation of the results, the two types of data collected in the survey were compared against one another in a semi-quantitative fashion. To achieve this, it was considered that both the number of votes received in each category in the quantitative study, as well as the number of entries per category that were used for the qualitative analysis, could be a measure of the significance of a category in the perception of participants. Therefore, normalised percentage values of both indicators for the quantitative and qualitative analysis were calculated as reported in Table 5.11. In Figure 5.4 the resulting numbers are compared graphically.

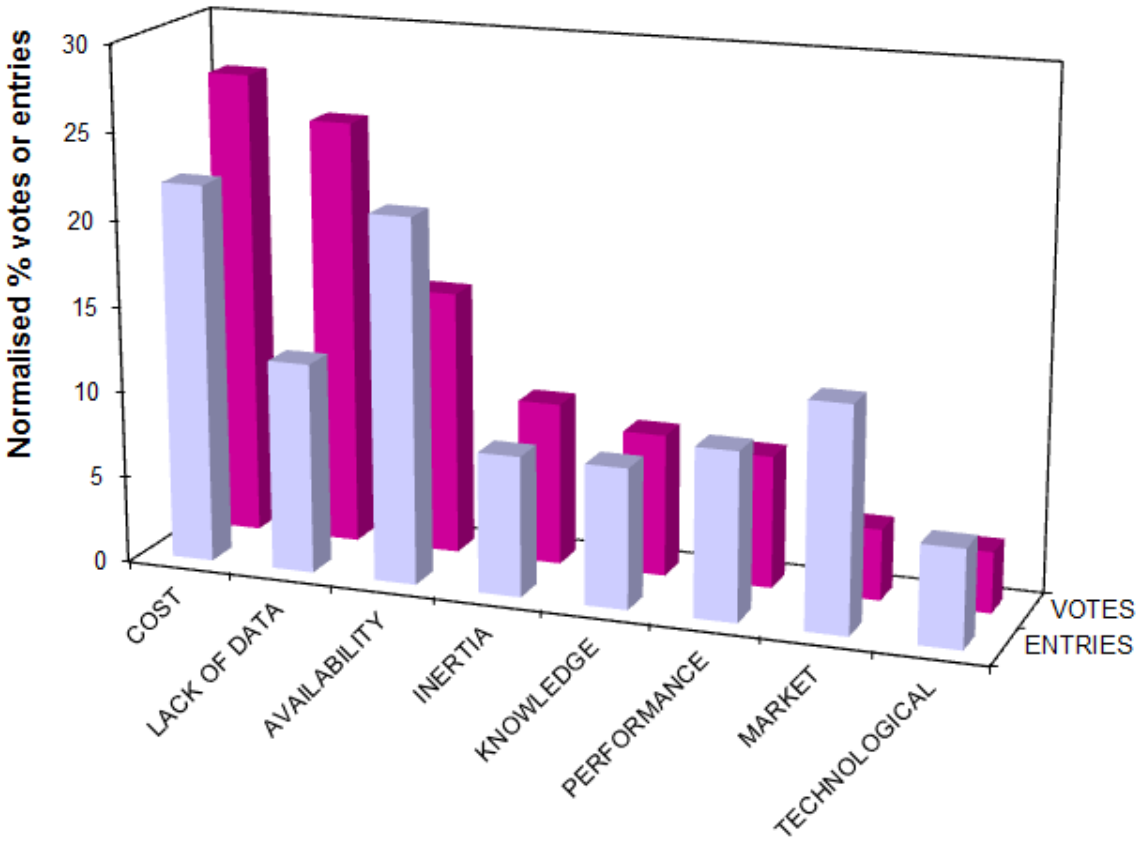


Figure 5.4: Semi-quantitative analysis based on percentage comparison between the number of qualitative entries per category and later voting activity, divided between academic and non-academic participants.

Table 5.11: Results of semi-quantitative analysis. Comparison of normalised percentages from the quantitative and qualitative results.

	Normalised percentage based on total number	
Categories	Votes % (from quantitative)	Entries % (from qualitative)
Cost	27%	22%
Lack of data	25%	12%
Availability & supply	15%	21%
Inertia	9%	8%
Knowledge	8%	8%
Performance	8%	10%
Market	4%	13%
Technological	4%	6%
TOTAL	100%	100%

The purpose of this analysis was to confront the relevance of each category in the perception of barriers against the perception of priority to be addressed, and to point out any discrepancy between the two data-gathering methods. Figure 5.4 shows that overall the outcomes were mostly aligned. It may be commented that issues related to 'Lack of data' emerged most prominently from the voting on priorities than from the barrier entries, suggesting it may be perceived as a strong need of relatively simple solution. Conversely, 'Market' is largely acknowledged as a barrier, but lesser as a priority to be addressed. In this case, the difference may be interpreted as good confidence in the market's ability to change without need for intervention. However, these conclusions ought to be considered relatively to the approximation of quantifying qualitative data as well as to the scale of the study. Overall, the good resemblance of two trends is an indication of the validity of the two methods applied and of the reliability of the data collected.

5.4 Chapter summary

The participatory workshop carried out at the Green Chemistry Centre of Excellence gave the opportunity to gather novel insight into the stakeholder perception of barriers for the wider adoption of green solvents across sectors. Overall, the activity was well received; participants were engaged in the workshop and later provided positive feedback.

Examples of similar survey methodologies are found in other fields of scientific research with the aim to identify priorities for research and policy. This work represents a first example of a participatory research applied in the field of green chemistry, and especially with a focus on barriers and priorities to the adoption of bio-based solvents. The study successfully produced qualitative and quantitative data, highlighting eight key categories of barriers for the uptake of green solvents: cost, lack of data, availability & supply, inertia, knowledge, performance, market, technological. The definition of these categories was made with participants during the workshop with an analysis of the clues produced during individual and group brainstorming and are reported. The following voting activity on these categories indicated that the top three issues perceived by participants were: cost, lack of data, and availability & supply.

Although this preliminary survey presents limitations associated with the size of the pool of experts and the time-constraints of the data-gathering workshop, the results provide bases for future studies of larger scope in this area that is considered a top priority in the world of green chemistry. Based on the experience from this proof-of-concept study and in relation to the guidelines for participatory studies published by Sutherland *et al.*, future research efforts in this area should target and survey a larger community of stakeholders and experts. Most preferably submission of inputs should be solicited prior the workshop, so they can be discussed and developed with a higher level of detail during the group exercise. Ideally, multiple workshops should be organised in order to ensure robustness of the methodology, which may also be carried out at regular intervals with different participants in order to track changes in wider perception with time.[170]

In conclusion, the results from this workshop will hopefully serve as an indication of

the most critical aspects to be addressed by research and policy for the uptake of green solvents: cost, lack of EHS data, availability & supply. While extensive research has tried to address the issues of cost and lack of data with regards to green solvents, a gap was noticed in the area of availability and supply. In fact, to date, no research is found to have investigated the issue of resource availability and of supply capacity for bio-based solvent (and chemicals), especially against their oil counterparts. Hence in Chapter 6, presents a novel in-depth analysis of such potential based on a case-study substitution of toluene with bio-based limonene.

Chapter 6

Can bio-based solvents meet demand?

A case-study

6.1 Introduction

In Chapter 5 it was highlighted how the issues of cost, lack of EHS data and supply availability are important in the opinion of stakeholders. Herein, one of these priority areas was considered for further investigation, as part of a case study aimed at quantifying the potential supply of a bio-based solvent against demand in a specialised application. Despite the evident significance of supply considerations for any biomass-derived chemical and solvent, only a few studies have targeted this issue.[177,178]

In general, the issue of securing bio-based feedstocks to meet the growing demand for agricultural products is well represented in the literature.[177] Recently, Piotrowski *et al.* raised the argument about the missing balance between demand and supply of biomass crops, which could jeopardise the creation of a stable bio-based market.[179] In another example, Kircher *et al.* compared biomass crops and fossil feedstocks in terms of their potential to either fix or release carbon, by quantifying the corresponding net carbon balance in relation to consumption volumes.[180] Although these and other studies provided a picture of the global landscape and in some cases discussed global forecasts for common agricultural products against predicted oil supply,[181] they did not extend this assessment to exemplify chemical products downstream. Furthermore, the topic appears to be mainly discussed at either global macro-level[179] or at national level, yet no exchange between the two models is discussed.[180] A broader consideration of the geographical dimension would allow identification of key regional players in the bio-chemical transition and quantify their potential within the global arena.

6.1.1 Design of the research

In order to target this gap, it was necessary to:

1. Identify a bio-chemical of interest and a petroleum-derived counterpart, for which similarity is known and substitution is proven feasible;

2. Select a specific sector of study on which to investigate the potential for substitution

from the point of view of resource availability and supply capacity;

3. Devise a method by which to estimate the potential for solvent substitution from the point of view of availability of the corresponding feedstocks;

4. Choose relevant countries or regions of interest to assess regional transition potential.

The development of each of these points is detailed below.

Chemical target

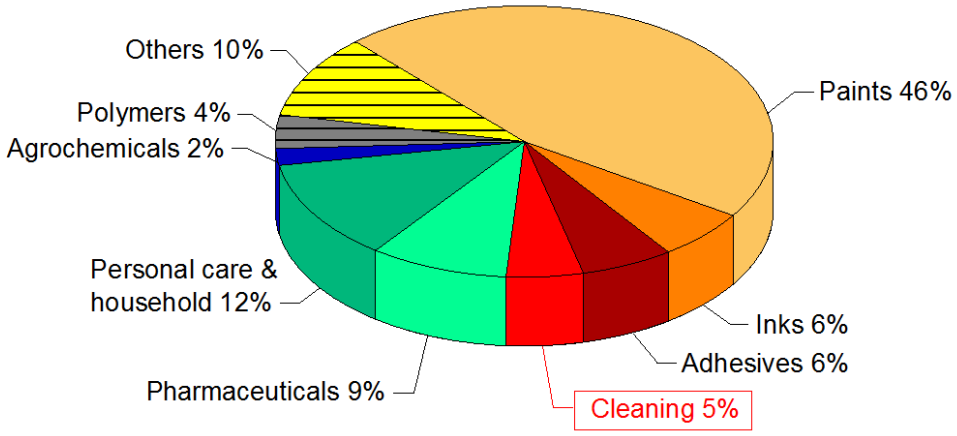
Citrus-derived limonene was selected as a case example for the development of this study, although it may be further applied to other bio-based chemicals and solvents. In the work discussed in previous Chapters, it was highlighted that citrus-derived solvents have high potential for replacing toluene in synthetic applications of interest to the pharmaceutical industry. The physico-chemical similarity between these two solvents is well recognised, and examples of drop-in substitution are seen in several applications - *e.g.* in process chemistry, electronics, paints, printing, cleaning and aerospace.[39,182–184]

Application of interest

Here, the cleaning sector was adopted as a case-study for the substitution of toluene with limonene. Cleaning practices are ubiquitous across the chemical sector, including the pharmaceutical industry, and substitution is well attested in cleaning applications.[184,185] Among multiple examples, limonene was notably used as a clean-up agent during the oil-spill of Deepwater Horizon in 2010.[186]

In Figure 6.1 is given a graphical representation of solvent use breakdown, based on the latest publicly available data - where possible the key remains consistent between the two piecharts.[22,187] Figure 6.1a shows the breakdown relevant to the broad solvents category. Overall, more than 80% of total use (represented by non-striped colours) appears dedicated to applications were they are used as solvents, *i.e.* as dispersing media, formulation additives, or cleaners.

a) Breakdown of solvents use per sector



b) Breakdown of toluene use per sector

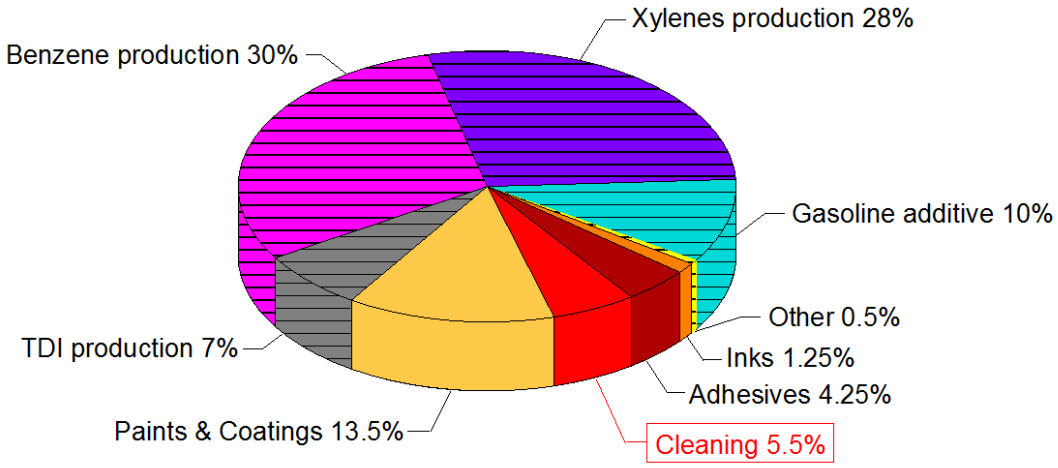


Figure 6.1: Piecharts representing the available data on the breakdown of solvents consumption across sectors: a) solvents category as a whole, based on 2009 data;[187] b) specific of toluene use, based on 2015 data.[22]

Comparing this with the specific breakdown for toluene use shown in Figure 6.1b, reveals how diverse the application streams can be for individual solvents. Evidently, toluene represents a chemical precursor of great importance in the chemical sector, to a much larger extent than for other solvents. In fact, its use as a solvent appears limited to 25% of total output, while in 75% of cases (represented in striped colours) it is employed as a precursor to other chemicals, such as benzene, xylenes and toluene diisocyanate, or it is used as a fuel additive.[188]

Selecting a sector of interest for a substitution case-study was restricted to the fields covered in Figure 6.1b, for which the percentage of toluene consumption was known. According to Figure 6.1 the cleaning sector accounts for approximately 5% of solvent use in general as well as in the specific case of toluene, suggesting it would make for a robust case-study in the analysis.

Cleaning processes typically involve the use of neat solvents, or a rich formulation thereof, for the de-greasing de-soiling and decontamination of surfaces, including containers, equipment and mechanical parts; this is typically achieved by application of the solvent through spraying, vapourising, or brushing, as well as by directly dipping the material into a batch of solvent, a practice called cold-cleaning.[189] As pointed out by Durkee (2013), a number of bio-based or potentially bio-based solvents are already used as part of common practice in the cleaning industry, *e.g.* ethanol, propanol, lactic acid, acetic acid, acetone, ethyl acetate, fatty esters, supercritical CO_2 and limonene.[188]

Methodology framework

A methodology was developed according to the amount and type of data available for the research. The aim was to compare toluene consumption in the sector, representative of demand, against figures of limonene production. In order to do so, historical market trends were analysed on the basis of publicly accessible databases and archives. Unfortunately, direct market data on the production or consumption of the two solvents was limited and largely unavailable free-of-charge. However, information was available on their precursors, *i.e.* naphtha for toluene and citrus fruits and juices for limonene. Information on volumes and trends for these goods were utilised to estimate their derivatives.

Overall, the time-frame considered was between 1990 to 2013, with the trends being extrapolated up to 2030 via linear regression. The final outcomes were benchmarked against the United Nations' Sustainable Development Goal (SDG) aiming to turn 30% of the world's economy to renewables by 2030,[10] thus indicating that 30% of toluene may be hoped to be replaced with bio-based limonene by that year.

In this methodology, a number of assumptions had to be made in order to limit the

effect of unpredictable factors in the forecasts generated.

1. The regulatory status of the two solvents was assumed to remain unchanged within the time-frame of consideration, with no disruptive alteration expected to occur;

2. The breakdown of solvent use per sector and within the cleaning applications were considered to remain constant within the timeframe;

3. The technology for the production of both solvents was assumed to have reached maximum potential and not expected to undergo any disruptive change or advance, *i.e.* limonene extraction yield, naphtha refining, petroleum-fields discovery;

4. Extraction of limonene was estimated from citrus peel derived from juiced fruits and unprocessed fruits for whole consumption. It was assumed that other citrus peel processing activities would have a negligible impact on the consumption of this bio-resource, *e.g.* the production of marmalade and jams, candies, or cosmetic or food ingredients.

Geographical scope

Finally, the study analysed the transition from a geographical perspective at different levels. First, a global scenario was developed by analysing historical trends and extrapolating them up to 2030. This allowed discussion on the potential for substitution long-term, and highlighted some of the key influential factors driving the trends. Following from this, the analysis was taken to a regional level, where relevant data was compared across regions based on data from year 2013. In this, nine case-study regions were selected in an effort to highlight some of the key players in the global citrus and chemical economies emerged from a preliminary assessment, as listed below:

1. United States
2. Florida, USA
3. European Union (EU27)
4. Germany
5. France

6. Brazil
7. India
8. China
9. South Africa

The overall selection intended to provide a broad representation of the diversity of key markets across continents in terms of their cultural, economic and environmental uniqueness. Moreover, different levels of depth where considered within the regional analyses. For example, the continental level addressed with the case-study on the European region was then narrowed down to highlight individual contributions by two countries with very different profiles. Similarly, the case-study on United States was developed further by analysing the situation at state level within State of Florida.

6.2 Toluene from naphtha

6.2.1 Global growth trends

The global consumption of toluene in 2014 reached 14.79 million metric tonnes (MT). Although various manufacturing routes are possible, including production from biomass,[190,191] the most economic and dominant practice remains the extraction from liquid hydrocarbon-rich fractions of petroleum reformates, *i.e.* naphtha.[42] As described by Fabri *et al.* (2005), the industrial production of toluene is tightly linked to the naphtha market, with oscillations dependent on sector-specific chemical demands and on the gasoline market price.[42]

Table 6.1 reports expected values of growth for relevant product categories as estimated in a recent review by Organization of the Petroleum Exporting Countries (OPEC).[192] The measure of growth is expressed as Compound Annual Growth Rate (CAGR) which finds definition in equation 6.2.1 - defined for a general product in the time-frame starting from year A to year B. The forecast of Table 6.1 may be more easily evaluated as part of a visual representation of the historical trends detailed in Figure 6.2.

$$CAGR\% = \left(\frac{ProductVolume_{yearB}}{ProductVolume_{yearA}} \right)^{\left(\frac{1}{B-A}-1\right)} \tag{6.2.1}$$

Table 6.1: Petrochemical growth per category reported by OPEC.[192]

	by 2020	by 2030
Global category	CAGR 2014-2020	CAGR 2020-2030
Global GDP	3.6%	3.5%
Petrochemicals	0.8%	1.2%
Naphtha	1.3%	1.4%
Gasoline	1.0%	0.4%

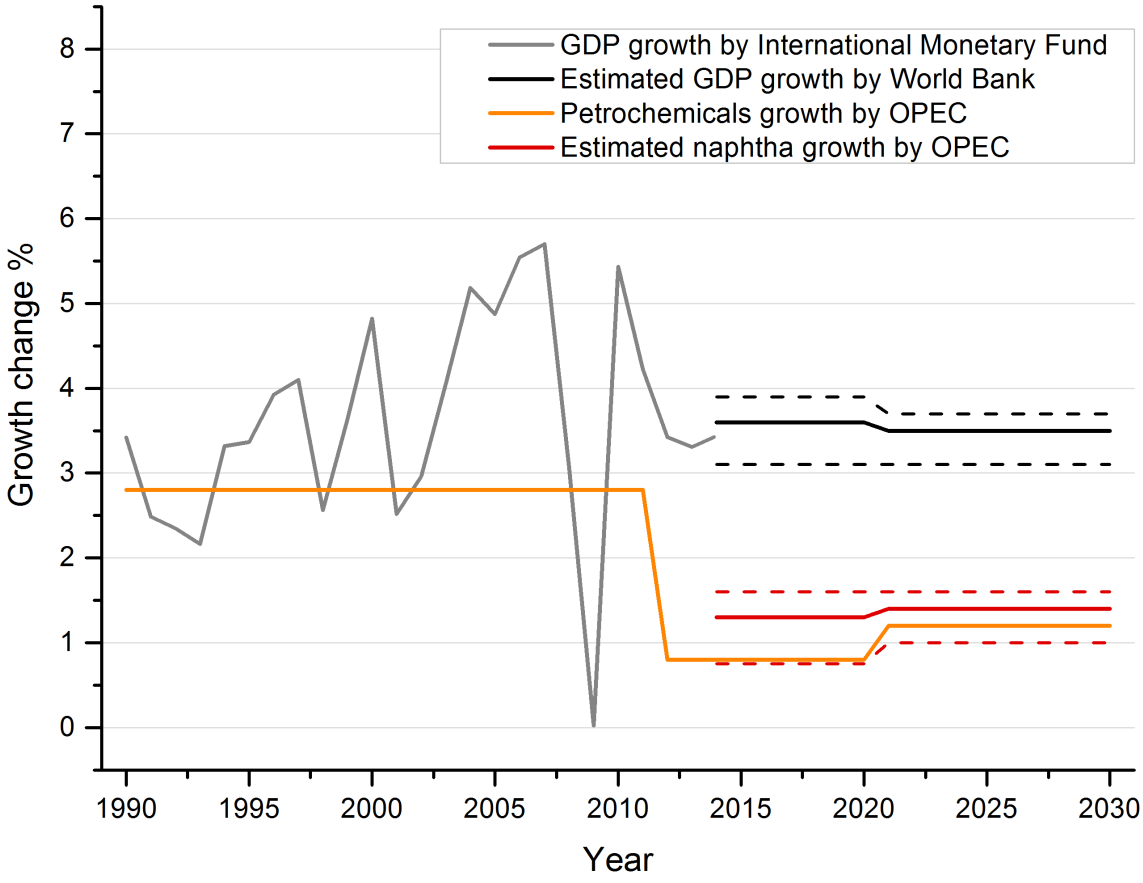


Figure 6.2: Parallel of GDP trends and petrochemical growth. Actual average values up to 2013 and reported forecasts up to 2030 with corresponding error bands.

In Figure 6.2 petrochemical growth rates reported by OPEC and trends in global Gross Domestic Product (GDP) reported by International Monetary Fund (IMF) are overlapped.[192,193] It can be seen that historically these two parameters were in strong correlation, and GDP growth could indeed be used as a reliable estimate for petrochemical growth as a whole.[192]

However, since 2009, the effects of the global financial crisis have deeply affected the growth of the petrochemical sector, establishing a large differential between GDP which recovered quickly (3.6% average growth from now up to 2020) and the current estimated petrochemical growth which did not (0.8% average growth up to 2020).[192,193] OPEC's report also supplies an individual estimate for the growth of naphtha as a petrochemical product, currently standing at 1.3% growth and expected to raise by 1.4% from 2020 onwards. However, no individual account of toluene growth was provided.

6.2.2 Extrapolation of trends

For the purpose of this study, and in absence of a direct account of toluene trends, the growth of naphtha was assumed to be reasonably applicable to toluene. Although naphtha growth reported by OPEC is expressed in error-free values, a confidence range may be indirectly derived by adopting intervals defined by IMF for their GDP forecasts. These intervals are represented in Figure 6.2 as dashed lined defining the lower or upper economic scenario around the average lines for the two trends of interest, *i.e.* GDP and naphtha growth. Table 6.2 also lists the average, upper and lower growth rates, which were later applied to estimate growth of naphtha and toluene. By applying the growth percentages of Table 6.2 to the latest available data on global toluene consumption allowed a realistic forecast to be obtained. Knowing the total consumption was 14.79 million MT in 2014,[22] it was estimated that toluene's consumption on average will be 15.98 million MT in 2020 and 18.36 million MT in 2030.

Table 6.2: Expected growth values for GDP and toluene (= naphtha) as reported respectively by IMF and OPEC.[192,193] Column 2014* shows real values, unavailable for toluene.

	GDP growth[193]			Toluene growth (= Naphtha)[192]		
Scenario	2014*	by 2020	by 2030	2014*	by 2020	by 2030
Lower	3.35%	3.1%	3.1%	n.a.	0.8%	1.0%
Average	3.39%	3.6%	3.5%	n.a.	1.3%	1.4%
Upper	3.43%	3.9%	3.7%	n.a.	1.6%	1.6%

Finally, these volumes were scaled according to the ratio of toluene employed in the application, *i.e.* 5.5% of total use.[22] Abiding to the assumption that this percentage shall not change significantly in the time-frame of study, it was obtained an estimated average consumption of toluene as a cleaning agent of 996,200 MT in 2030. Details are provided in Table 6.3. In relation to the UN's SDG of 30% bio-based substitution, approximately 300,000 MT/y of toluene should be supplied from renewable feedstocks in 2030.

Table 6.3: Calculated values for toluene global consumption as cleaning agent.

	Toluene use in cleaning (MT/y)			UN's SDG (MT/y)
Scenario	2014	2020	2030	2030
Lower	-	846,500	933,200	280,000
Average	**813,400**	**867,700**	**996,200**	**298,800**
Upper	-	880,600	1,032,000	309,600

6.3 Limonene from citrus-waste

Limonene is an oil extract obtained from citrus skins typically by steam distillation.[57] Although no indication of global limonene production volume was available, an estimate of its potential supply capacity was obtained by considering data on its main resource of origin: citrus fruits. This type of information is publicly accessible from databases such as of Food and Agriculture Organisation's Statistics Department (FAOSTAT), amongst other sources referenced in the text.[194,195]

The global geographical distribution of citrus production is presented in Figure 6.3 in a map developed using 2014 data available from FAOSTAT.[194,196] This shows that around 130 million MT of citrus fruits are being produced globally year on year, with[194] the combined harvests of China and Brazil making up nearly half of the world's entire production. Then, the ranking of biggest producers follows with United States and India as critical players in the citrus market.

In addition, FAOSTAT provides access to historical archives of the yearly production

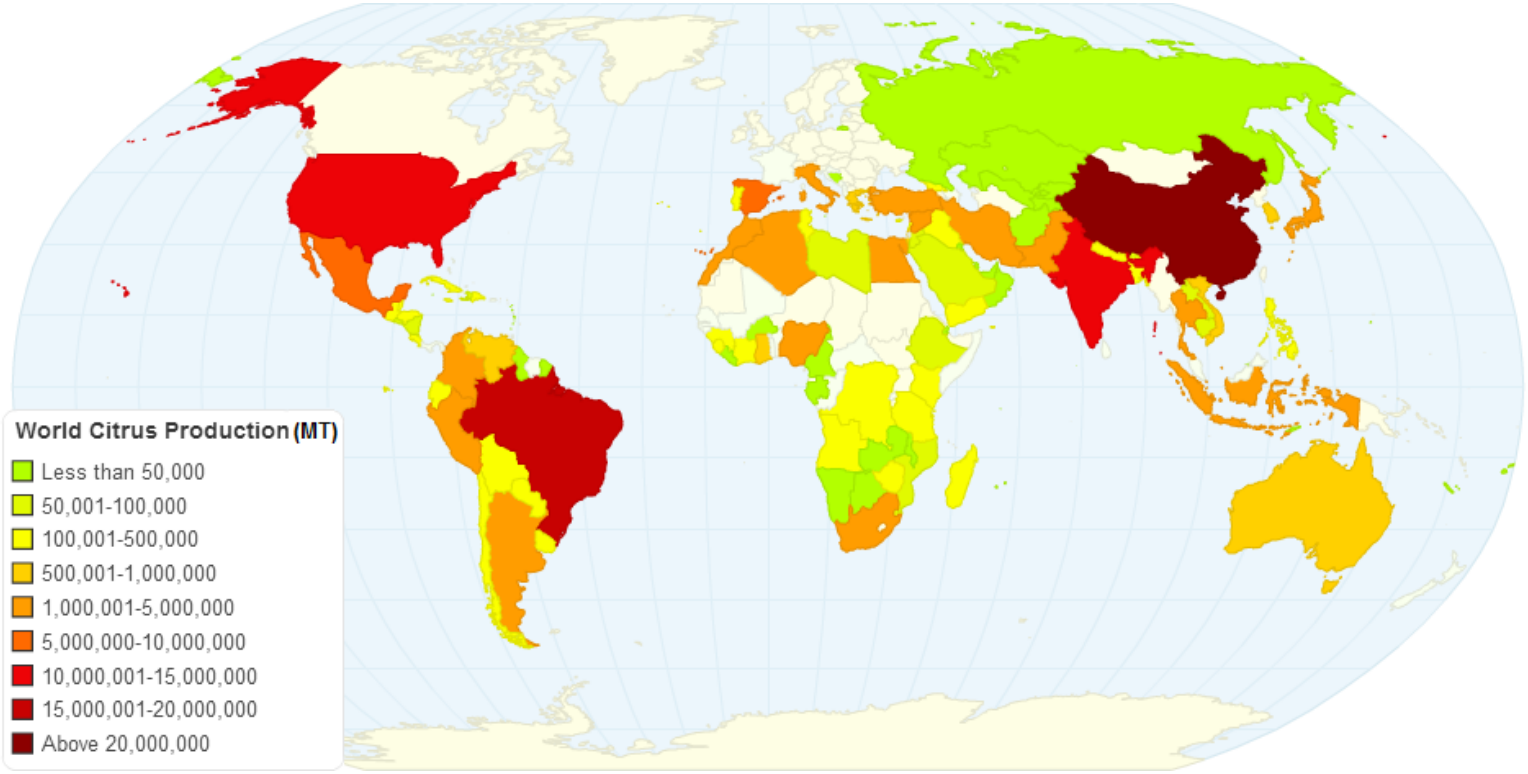


Figure 6.3: Global map of citrus production, adapted from FAOSTAT data.[194,196] Values in MT, not including trade statistics.

of all most common citrus fruit varieties. As shown in Figure 6.4, in 2014, oranges were the dominant citrus fruit type in the world, accounting for 53% of harvests, followed by tangerines & mandarins, and limes & lemons. Altogether, grapefruits & pomelos and all other less common citrus varieties represented approximately 15% of the world's production in that year. A very similar breakdown is reflected in the juice production figures.[194] However, while these represent a global breakdown, it is known that such distribution can vary greatly from country to country. For instance, the Brazilian citrus production is predominantly based on oranges, while the Chinese market is mainly dedicated to growing tangerines.[194]

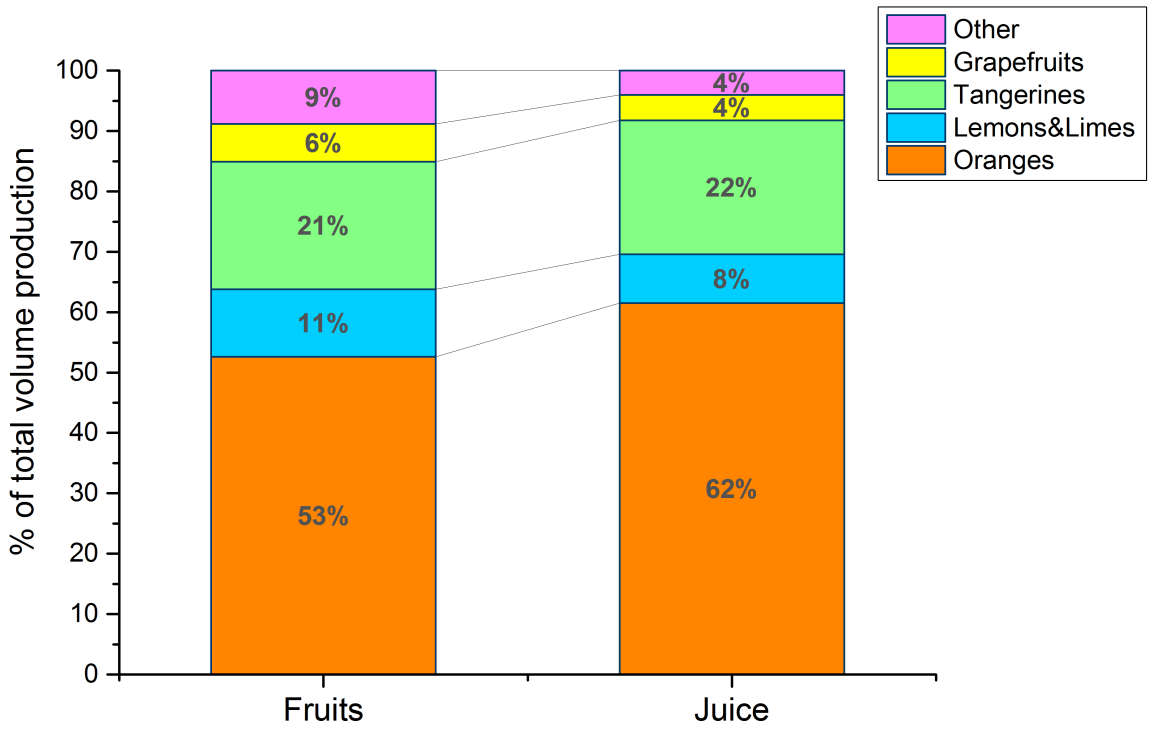


Figure 6.4: Breakdown of main citrus varieties produced worldwide as percentage of total of a) fruits harvests, b) juice production. Data adapted from FAOSTAT, based on year 2014.[194]

6.3.1 Juice production data

The volumes of juice production used in this study were obtained by combination of two fields of data from FAOSTAT:

1. 'single strength', meaning volume of juice produced via mechanical pressing,

2. 'concentrate', representing volumes of juice placed in the market after being subjected to a drying process that reduced its volume and concentrated its content.

In order to combine these volumes, it appeared necessary to convert the concentrated juice volumes into their original single strength equivalents. This was achieved by utilising tables of Brix standard concentration. Brix values define juices' concentration based on the content of soluble solids in solution (measured as v/v at 20°C). They are used as a international standards for juices to be classed either as single strength or as concentrate,[197] and are specific of each fruit variety. Table 6.4 presents the standard Brix values of concentrated and single strength juices from the most common citrus fruit varieties.[197,198] Brix values allowed to calculate a useful conversion factor for determining the total juice production for each fruit type according to equation 6.3.1. Ultimately, the total volume for all citrus was obtained by summing all results from individual citrus types.

Table 6.4: Standard Brix values for single strength and concentrate juices of the most common citrus varieties. In this study, values for 'Other citrus' were assumed equivalent to oranges.

Fruit type	Brix_{conc} [197]	Brix_{single} [198]	$\text{Brix}_{conc}/\text{Brix}_{single}$
Oranges	62.5	11.8	5.3
Lemons	62.5	4.5	13.9
Limes	62.5	4.5	13.9
Mandarins	62.5	11.8	5.3
Tangerines	62.5	11.8	5.3
Grapefruits	62.5	10.0	6.3
Other citrus*	62.5*	11.8*	5.3*

$$TotalJuice = Volume_{fresh} + \left(Volume_{conc} \cdot \frac{Brix_{conc}}{Brix_{single}} \right) \tag{6.3.1}$$

6.3.2 Global citrus growth trends and extrapolation

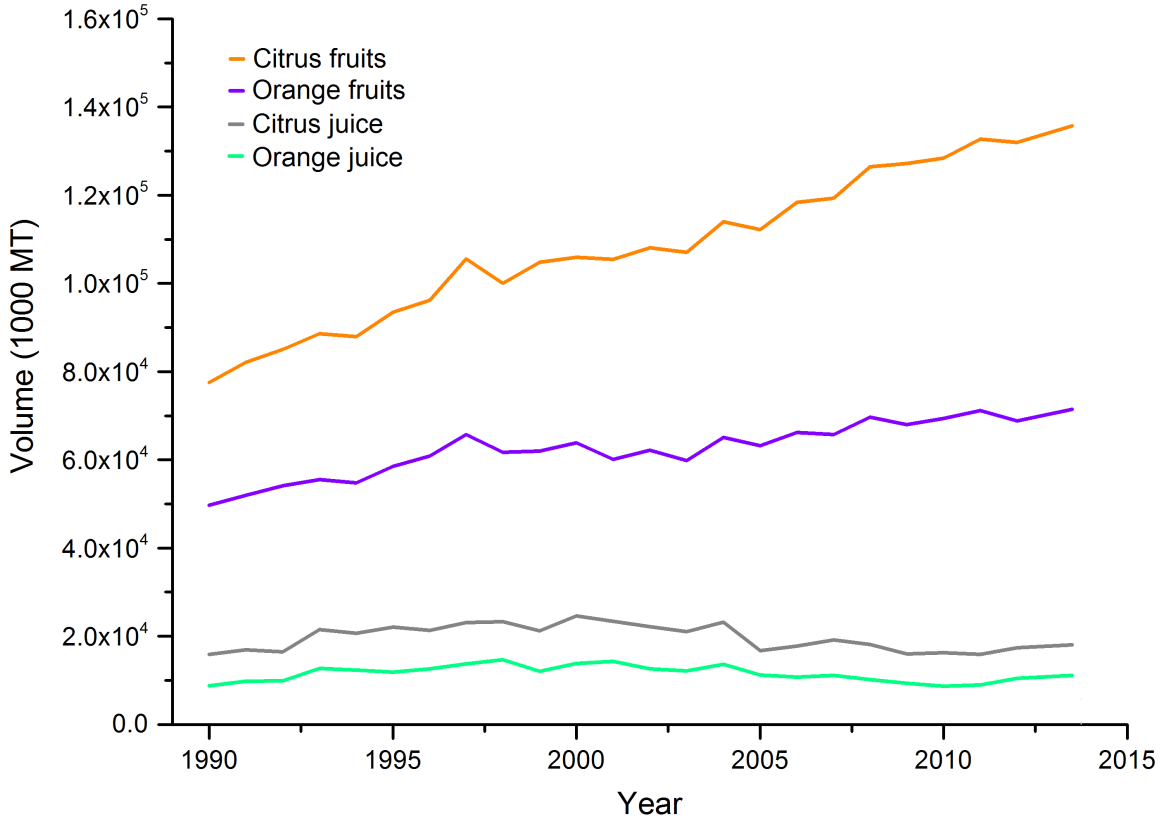


Figure 6.5: FAOSTAT historical data related to production of citrus fruits and juices as a whole, as well as orange fruits and juice.[194]

Figure 6.5 illustrates the historic trends gathered from FAOSTAT for the period 1990 to 2014.[194] The graph compares production volumes of citrus fruits as well as citrus juices, while also indicating the fraction represented by oranges. Orange trends appear to dictate the oscillations observed in the larger citrus trends. As a whole, fruits' harvests appear to be in a steady rise, reportedly due to a general increase in agricultural land.[51] In contrast, juice production for both oranges and overall citrus appears to have peaked around 10,000 MT/y and 20,000 MT/y respectively for many years. Overall, juice trends have witnessed a slow decline over the last 15 years, in line with what reported by other sources.[195] Reportedly, the juice market of United States has been recently affected by an invasive disease known as HuangLongBin (HLB) or 'citrus greening', which causes lower productivity including lower juice content in fruits, and ultimately leading to premature death of the trees.[199]

In order to extrapolate the trends up to year 2030, a linear regression was applied. This

allowed citrus forecasts to be developed for the same timeframe previously determined for toluene, which will later be used for estimating limonene potential production. Figure 6.6 demonstrates a representative example of the regression method employed. In this procedure, a confidence test was used to identify outliers in the set, which indicated six data points falling outside of the band. These points were excluded from the determination of the linear fitting equation and CAGR%. Figure 6.7 shows the resulting extrapolation lines and statistical details for the four scenarios considered.

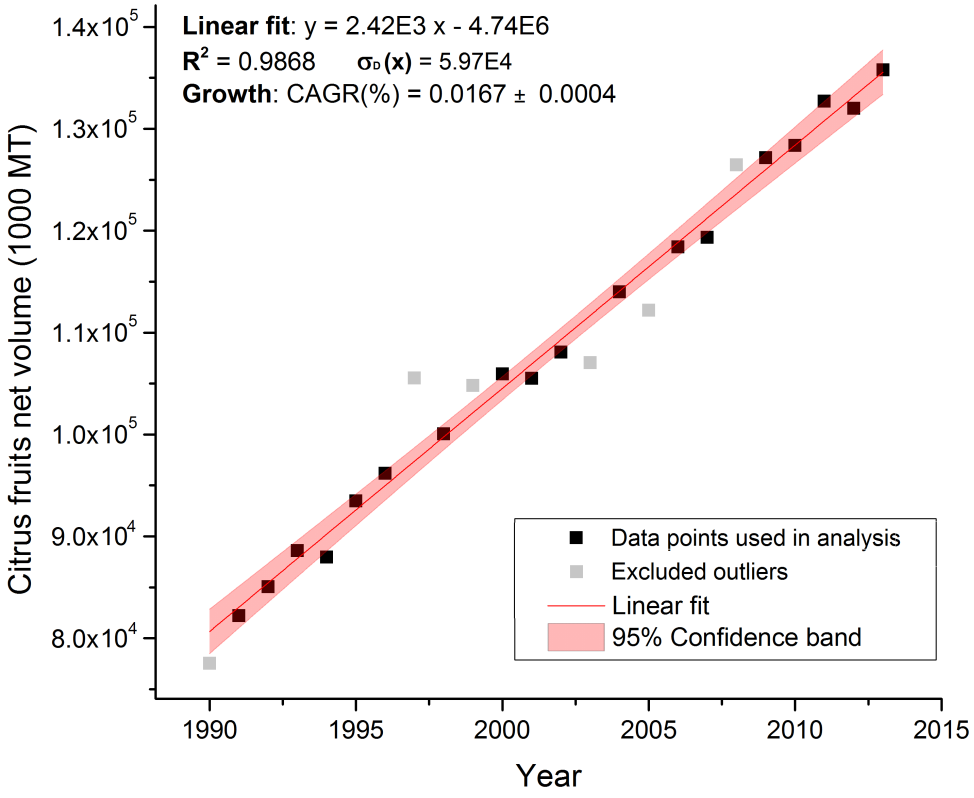


Figure 6.6: Example of linear regression carried out on citrus fruits production trend applying a 95% confidence band to exclude outliers from the data set.

6.3.3 Limonene potential supply capacity

Based on the extrapolated profiles it became possible to estimate the potential limonene availability from each given source. Values of typical limonene content present in citrus peel were drawn from the literature as summarised herein.

One large scale study by Pourbafrani reported 7.5 kg of limonene extracted via steam

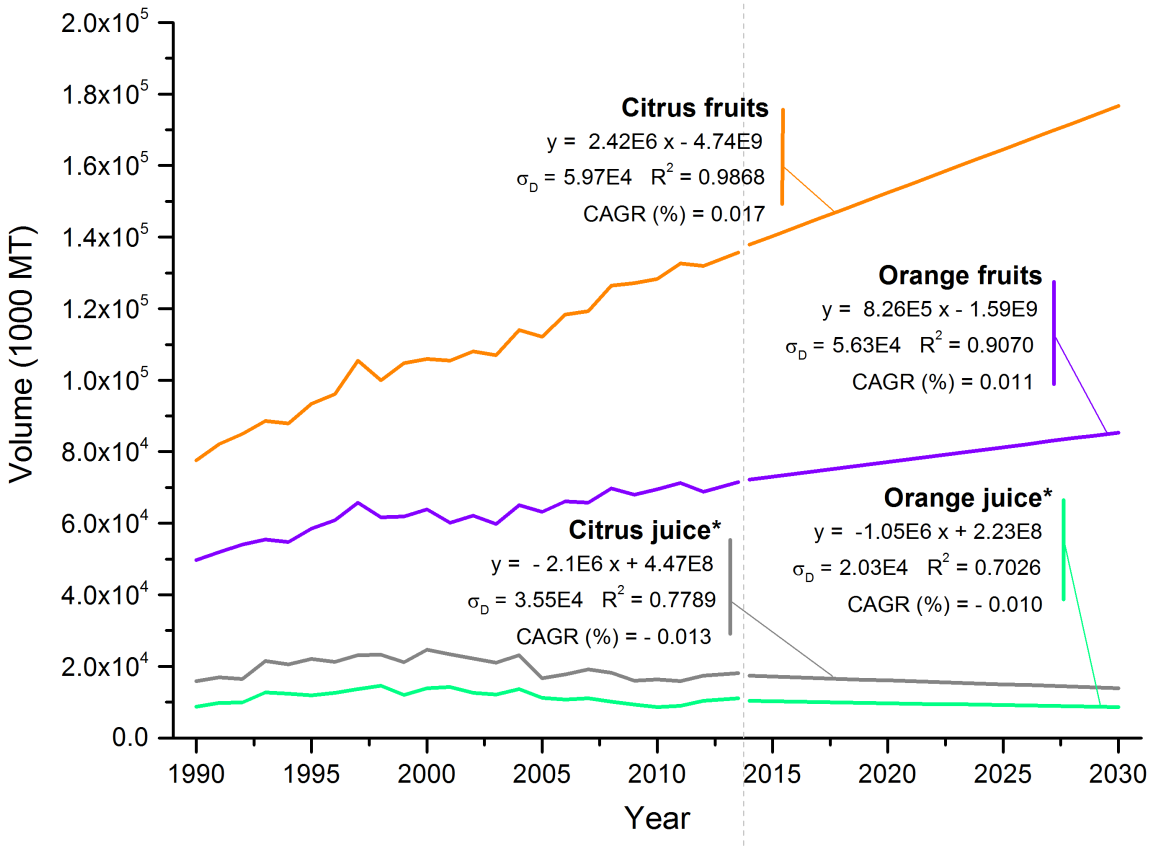


Figure 6.7: Results of extrapolated historic trends of citrus fruits and juice production up to 2030. Statistical outcomes of linear regressions displayed on the graph.

distillation from 1 ton of wet peel, *i.e.* 0.75 wt%.[56] In a similar work by Ceron-Salazar *et al.*, 0.62 wt% yield was achieved, equivalent to 3.08 wt% on a dry basis.[200] In some examples the use of modern technologies brought a significant advantage in the extraction yields achieved. For instance, Pfalzgraff reported 1.03 wt% yield of oil extracted on a wet basis by solvent-free microwave assisted extraction;[52] and Davila *et al.* observed 1.5 wt% yield of oil attained in a continuous flow extractor.[46] The evident fluctuation in the reported yields may be due to the different level of purity of the isolated fractions. Also, limonene content is typically dependent on the fruit variety considered and its origin.[52]

In the present work, a conservative value of 0.75 wt% from wet basis was adopted for the evaluation of potential limonene supply, in line with figures quoted majorly in reviews.[57] This percentage was applied to the estimated fruit and juice production according to equations 6.3.2 and 6.3.3. As mentioned before, the weight of citrus waste is typically equivalent to the weight of juice obtained from a fruit, each accounting for 50% of the over-

all mass of the fruit. The values were thus calculated and results compared against toluene consumption, as discussed in the evaluation of global substitution potential presented in the next Section.

$$Limonene(Juicing) = JuiceVolume \cdot 0.75\% \tag{6.3.2}$$

$$Limonene(Fruits, incl.Juicing) = \frac{FruitVolume}{2} \cdot 0.75\% \tag{6.3.3}$$

6.4 Global substitution potential

The combined forecasts obtained from the above analyses for toluene consumption and limonene potential supply are illustrated in Figure 6.8. A black solid line on the top of the graph represents the average toluene consumption in the cleaning sector; either side of this, dashed lines represent the lower and upper economic growth scenarios. The calculated trends of maximum limonene capacity are presented according to the four feedstocks: orange fruits grown and juiced, and overall citrus fruits grown and juiced. All the citrus trends are accompanied by error bars determined by the standard deviation of the original data.

As shown in Figure 6.8, extracting limonene from residues of total citrus being juiced would presently yield approximately 130,000 MT/y. However, under the steady downward trend affecting the juice industry, limonene capacity from such feedstock may decrease to 100,000 MT/y by 2030. On the other hand, the overall limonene capacity calculated from the use of any residue of grown fruits regardless of the type of consumption is estimated to supply by 2030 approximately 660,000 MT/y. By comparison, global consumption of toluene as cleaning agent in 2030 is expected to be approximately 996,000 MT/y.

It conclusion, it appears that even the maximum limonene capacity from all citrus residues would fall 30% short of the demand for a complete substitution. Considering that the cleaning sector constitutes only a 5.5% fraction of the total toluene use, this highlights the colossal scale of modern days consumption and our extreme dependency on

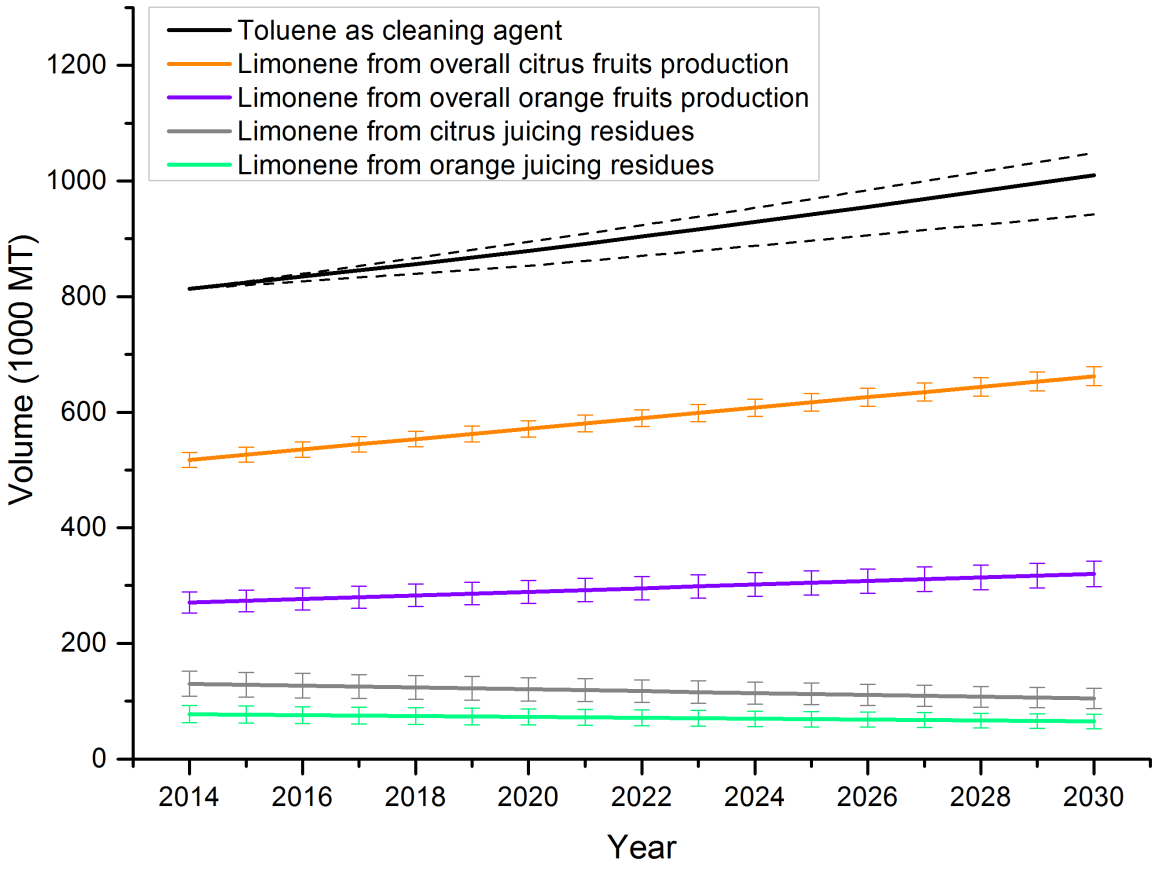


Figure 6.8: Comparison of extrapolated trends of toluene consumption and limonene potential supply from four feedstocks.

fossil-derived products. Yet, a less ambitious but still challenging goal such as the 2030 UN's SDG target would require about 300,000 MT of the demand to be satisfied by renewable alternatives such as limonene. According to the estimates achieved in this study, this volume could be largely met under two possible scenarios: 1) exploiting all waste coming from oranges whether juiced or used for whole consumption; 2) exploiting at least half of all citrus fruits waste.

6.5 Regional substitution potential

As part of the principles of a bioeconomy, the regional dimension has a critical role in supporting social, environmental and economic development.[9] In fact, one of the most debated aspects of globalisation is the ripped link between local resources and demand prioritising international/overseas trade.[201] However, global trade faces higher difficulties and

added costs in dealing with bio-based feedstock and waste due to: 1) the fast perishability of the materials requiring costly refrigeration on long-distance hauling; 2) the complex and diverse matrix of natural feedstocks from different sources which hinders global scale processing.[201,202] These factors may be more easily addressed within a regional economy. Herein are presented the results of the assessment of toluene-to-limonene substitution at regional level, which breaks-down the global substitution potential presented earlier into its regional and sub-regional components.

6.5.1 Regional trends of toluene consumption

Once again data availability proved to be challenging. In fact, while toluene consumption as a whole at national/regional level was obtained, no report of its proportion of use as cleaning agent was found for the regions of interest. Therefore an approximation was applied. From the global analysis, a 5.5% factor had been applied as a global average to account for the fraction of toluene use in cleaning applications. Herein, the same factor was used to scale down the overall consumption to estimate the use as cleaning agent in individual regions. While this procedure does not take into account each country's actual use of toluene as cleaner and carries inaccuracies and limitation, it provided for an effective estimate and consistent methodology for comparing consumption of toluene across countries.

Data on consumption volumes were available for year 2013 for most regions of interest, and, therefore, this year was selected as a benchmark for this analysis. Two exceptions had to be made, for China and South Africa, where the most recent available data on toluene consumption was found respectively for year 2012 and 2011. All original data and references are reported Table 6.5 and further information is supplied in Appendix B.

6.5.2 Regional trends of citrus production

Original citrus-related data is reported in Table 6.6 in terms of total juice production and net overall fruit production. A principal source for this data was an official report by the United States' Department of Agriculture (USDA) published in 2016 on global citrus

Table 6.5: Regional toluene data based on production and consumption volumes in year 2013 - unless indicated (*). References stated in table.

Region	**Total toluene consumption** MT (2013)	**Toluene as cleaner** MT (2013)
USA[48]	2,177,000	120,000
Florida,USA[48]	94,000	5,000
EU[203]	1,250,000	69,000
Germany[203]	357,000	20,000
Spain[203]	56,000	3,000
Brazil[204]	406,000	22,000
India[205]	108,000	6,000
China[206]	5,050,000* *(2012)*	278,000
South Africa[207]	86,000* *(2011)*	5,000

Table 6.6: Regional citrus data based on production and consumption volumes in year 2013 - unless indicated (*).

Region	**Total citrus juice production** MT (2013)	**Total citrus fruits production** MT (2013)
USA[195]	2,551,000	8,577,000
Florida,USA[195,208]	2,325,000	5,091,000
EU[195,209]	2,134,000	12,383,000
Germany[195,209]	459,500	1,140,000
Spain[195,209]	1,451,000	827,000
Brazil[195]	6,732,000	23,811,000
India[195,210]	83,500	7,408,000* *(2009)*
China[195]	662,000	28,264,000
South Africa[195]	373,000	953,000

trade.[195] These values were cross-checked for correlation with those present in FAOSTAT database related to 2013.[194] In the case of India, gaps in information had to be filled with data from other sources, which were only found available for year 2009.[210]

These figures were used as an account of citrus-waste availability. It becomes apparent that the volume of such waste is not only dependent on the volume of fruits produced in a region, but is also affected by the amount of citrus traded into and out of each region. In other words, the import of citrus fruits would increase the availability of citrus waste in a given region, whilst exports would decrease it by moving it to another destination market. A full account of these statistics and corresponding references is provided in Appendix B, while in Table 6.6 are reported the final net values calculated after import volumes were added and export volumes were subtracted. From this, the overall waste generated and the overall limonene extraction potential were then estimated by applying equations 6.3.2 and 6.3.3 used earlier.

6.5.3 Results of regional substitution case-studies

Figure 6.9 illustrates the substitution potential provided by the methodology described above. Herein, the two sources of limonene are represented with red and orange bars, respectively from juiced fruits and from total net volume of fruits. The estimated toluene consumption as cleaning agent is stacked for comparison as a black bar. It is evident that the potential for solvent substitution is highly region-specific, and the range of regions proposed in this case-study emphasises the characteristic diversity of these markets, discussed in detail below.

United States of America

United States show a trend not dissimilar from the global substitution potential previously discussed. Being the third largest citrus-producing country in the world, it is estimated that it could potentially produce 32,000 MT of limonene per year, although toluene consumption in cleaning is approximately 4-times that volume, *ca.* 120,000 MT/y. In fact, United States are also the second largest consumers of petrochemicals.

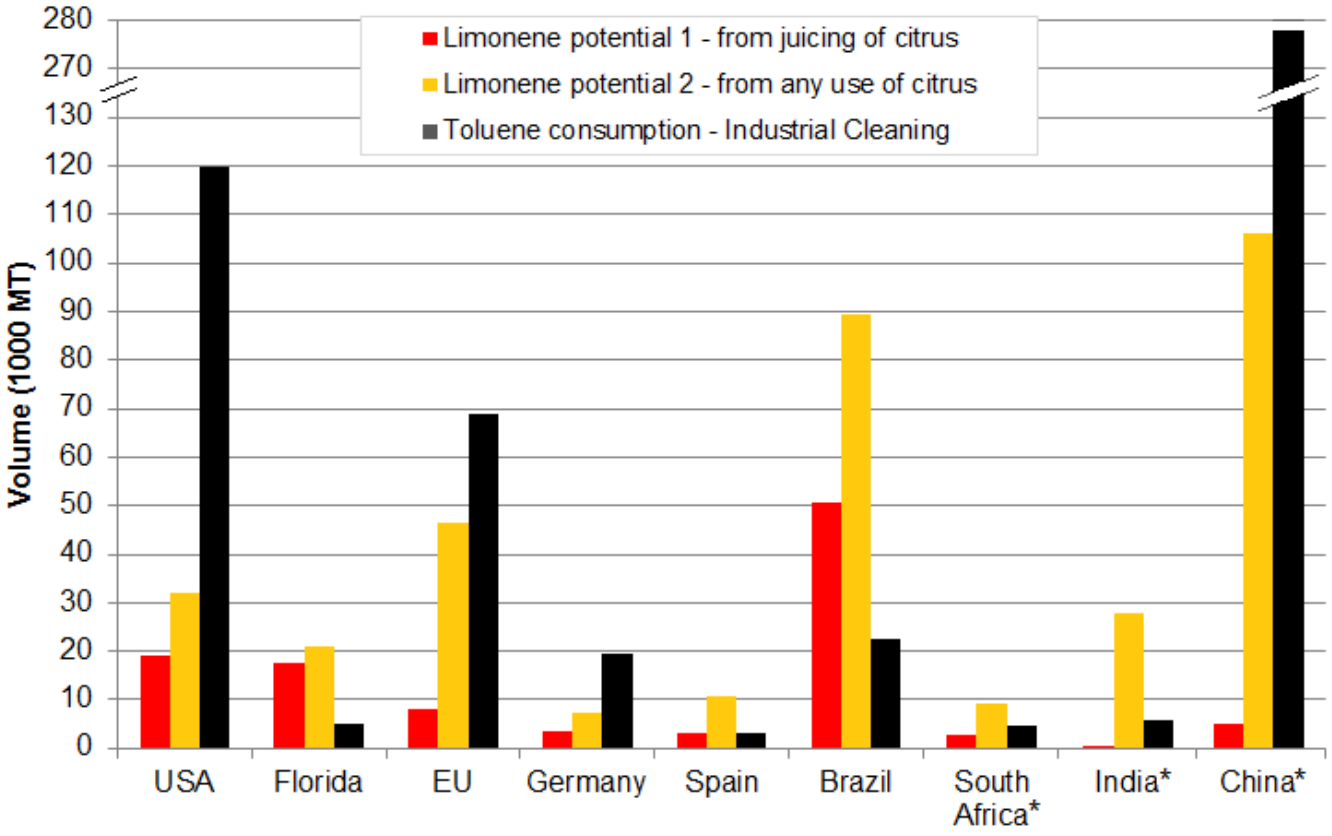


Figure 6.9: Results of regional substitution case-studies based on estimates of 2013.

A different perspective is found when considering the case-study of Florida state, as a sub-region of the above case-study. It is reported that Florida produces 70% of the country's citrus fruits, and provides almost for the entirety of the national citrus juice demand.[48] The state has a developed chemical industry mainly focused on fine and speciality chemicals, which accounts for 4% of the petrochemical consumption of the federal country.[48] By applying these percentages, it is estimated that Florida would consume approximately 5,000 MT of toluene as cleaner per year. Given the outstanding citrus production, it is expected that the region could largely supply for its own toluene demand by means of locally-sourced bio-derived limonene, potentially producing more than 20,000 MT/y of limonene. Therefore, Florida shows a high potential for successfully implementing this substitution, and for extending it to other sectors or exporting its surplus to neighbouring states or countries.

European Union

The European Union as a whole (27 member states) also reflects a similar profile to the one observed for United States (Figure 6.9). In this case-study, both limonene scenarios were strongly influenced by import/export volumes, with high overseas imports contributing positively to increasing the volume of peel available for limonene extraction in the region (above 45,000 MT/y). If compared with toluene consumed in Europe as cleaning agent estimated as 68,750 MT/y, it can be concluded that approximately 60% of it may be potentially replaced by limonene produced in the region.

Further insight into the European case was sought by examining the substitution potential within two member states: 1) Germany, a non-citrus growing country yet major importer of citrus products, and 2) Spain, a citrus producer with high international exports. Despite their opposite status and role in relation to the citrus market, the two countries showed a similar potential for internal limonene production. Although Germany produces virtually no citrus fruits, it imports more than 1 million MT/y thus creating a considerable potential for producing limonene at a scale of 7,000 MT/y. Interestingly, this result suggest that even a non-citrus growing country could potentially generate significant quantities of

limonene from its citrus waste. By these means, Germany may replace more than 30% of its toluene consumption by means of a renewable feedstock.

An analysis of the case of Spain shows a potential limonene supply of 3,103 MT/y from juiced fruits, and 10,886 MT/y from all fruits. The estimated toluene consumption as cleaner in Spain in 2013 is approximately 3,000 MT/y, thus showing promising potential for complete substitution even simply by means of residues from juicing processes. It is worth mentioning that approximately than 60% of citrus harvests from Spain are typically exported to other countries, mostly within Europe. Therefore, changes to this ratio would highly influence the potential availability of waste biomass for limonene extraction in the country. In 2013, Spain grew more citrus than Florida, and contributed nearly half of Europe's total production.[209]

BRICS countries

Outside of the major western economies described above, some of the biggest world producers of citrus fruits are most of the so-called BRICS countries. These emerging fast-growing economies play an increasingly determinant role in global arena, as they are characteristically rich in natural and fossil resources.[211] With reference to Figure 6.9, these countries display a staggering potential for citrus-waste exploitation and to become leading players in the limonene market.

For instance, Brazil has a long history and expertise in utilisation of biomass, best known for the production of bioethanol from sugar-cane.[212] Overall, given its abundant and diverse biomass, Brazil is believed to stand a high chance of becoming an exemplary model of bio-based economy.[213] In terms of citrus, Brazil is the world's largest producer of oranges (*ca.* 20,000,000 MT/y) and of citrus juices (overall 13,434,000 MT/y), the latter accounting for 75% of the world's production. It is estimated that nearly 90,000 MT/y of limonene could be potentially produced in the country, whilst toluene consumption is estimated to be approximately 22,000 MT/y in cleaning applications. Thus, Brazil could supply more than four times its toluene demand in industrial cleaning by means of limonene. Half of this may be supplied from the juice industry alone.

A very different scenario is presented in the case of China. In the recent years China overtook Brazil as the world leader in citrus fruits production, growing mainly tangerines and mandarins. Despite the highest citrus feedstock availability, its potential for toluene substitution with bio-derived limonene appears less encouraging, due to the current out-of-scale petrochemical consumption and fast industrial development that the country is experiencing. Moreover, as mentioned earlier, China is not traditionally involved in juice-making, meaning that the relevant waste may be highly distributed across the vast country and, therefore, more difficult to collect.

Similarly, in the case-studies on India and South Africa a very small percentage of production undergoes juicing every year. However, in these examples the maximum limonene capacity is estimated to largely outstrip toluene demand in the application of interest.

6.6 Chapter summary

In conclusion, this study quantified the transition potential from toluene to limonene within an exemplary sector. The case-studies showing the most promise are Brazil, Florida (USA), Spain, India and South Africa; in these regions, limonene extraction from currently underused citrus waste resources has the capacity to exceed toluene's demand as cleaning agent and to supply for other sectors and applications. Overall in this analysis, both citrus-growing and citrus-importing countries showed good outcomes in the substitution evaluation.

At a global level, it was estimated that the maximum limonene supply capacity could currently provide approximately 535,000 MT/y, expected to reach 662,000 MT/y in 2030. Unfortunately, this volume would account for only 3.6% of the world's total consumption of toluene and not allow for full substitution even within the relatively small cleaning sector (5.5% of total). This result highlights the scale of the challenges faced by the transition to a bio-based economy and the strong need to reduce consumption in general as a key effort to achieve broader sustainability.

Since limonene alone could not supply total toluene demand in this or other sectors,

it becomes evident that a substitution strategy should also be accompanied by efforts to reduce, reuse and recycle solvents. Effectively, in the same way in which the current rate of consumption of petrochemicals is not sustainable, a direct substitution with bio-chemicals without any reduction in usage would also inevitably stay unsustainable. Therefore, reducing consumption may be the only effective way of reducing the gap between current demand and a sustainable supply. This said, substitution of petrochemical solvents with those derived from waste biomass remains a cardinal to tackle chemical sourcing in the future. Industrial cleaning is one of such application where good performance from bio-based alternatives is consolidated and regulation is pressing for substitution of more volatile solvents.[214] Overall, the conclusions reached in this study may be plausibly extended and applied to the consideration of substitution potential between other chemical products used in different sectors.

Chapter 7

Experimental methods

7.1 Analytical equipment and methods

7.1.1 UV/Vis spectroscopy

Spectroscopy in the ultraviolet/visible range was employed for the measurement of Kamlet-Taft solvent parameters. Table 7.1 reports the specifications and settings adopted.

Table 7.1: UV/Vis set-up and conditions.

Instrument	Perkin Elmer Lambda 2 version 3.6 UV/Vis
Cuvette	Quartz
Scan region	800 nm - 200 nm
Cycle	1, interval 0.2
Scan speed	480 nm/min

7.1.2 Gas Chromatography - Flame Ionisation Detector (GC-FID)

Most of the analyses presented in this thesis were performed by gas chromatography with flame ionization detector (GC-FID). The device used was an Agilent 6890N equipped with FID detector with a ZB-5HT column (15m x 0.25mm x 0.25mm). The method used a temperature gradient to control the oven according to the conditions reported in Table 7.2 and illustrated in Figure 7.1. For reactions in certain solvents, the peak of the solvent from the reaction mixture had similar retention time as the reactant peak, in which case an adapted method was used where a milder gradient was adopted near the problematic retention time.

7.1.3 Gas Chromatography - Mass Spectroscopy (GC-MS)

Gas chromatography coupled with mass spectrometry (GCMS) allowed identification of product mass and ionisation pattern. The apparatus used was a Perkin Elmer Clarus 500 GCMS. The method used employed equivalent settings and column to those described above for GC-FID instrument.

Table 7.2: GC-FID set-up and conditions

Instrument	Agilent Technologies 6890N Network GC System
Column	ZB-5HT column (15m x 0.25mm x 0.25mm)
Flow	Constant, 2.0 ml/min
Pressure	20.2 psi
Gas carrier	He at 1 MPa (20 ml/min total flow)
Inlet	Split mode
Split ratio	20:1
Heater	250°C
Pressure	20.2 psi
Total flow	44.1 ml/min
Oven	from 50°C to 400°C
Equilibration time	1 min
Detector	Flame Ionisation Detector (FID)
Heater	340°C
Air flow	350.0 ml/min
H_2 flow	35.0 ml/min

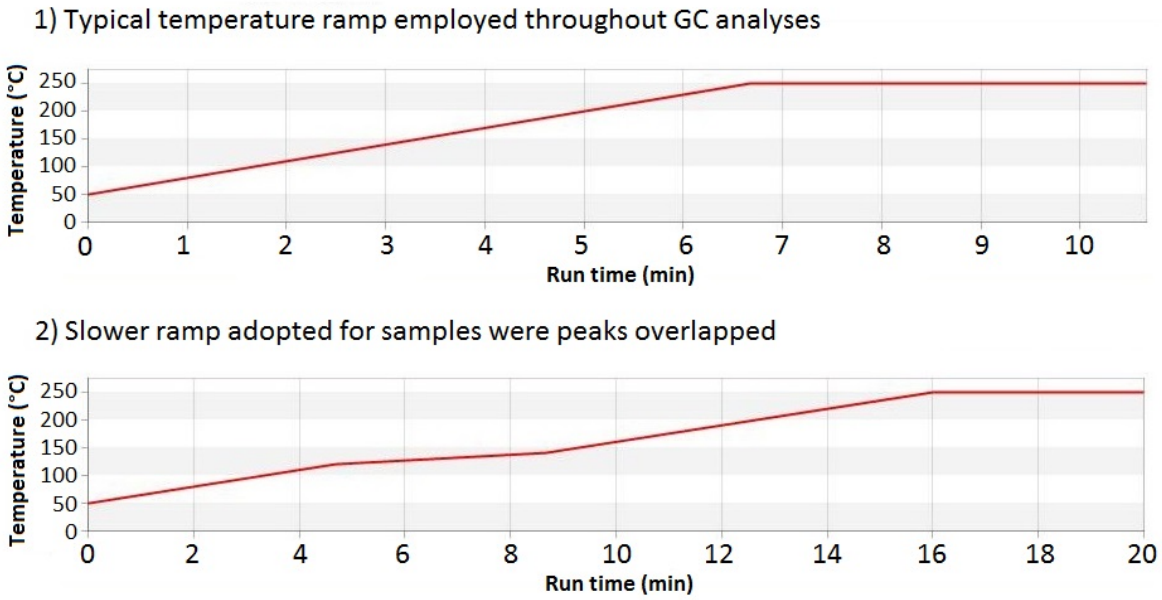


Figure 7.1: GC temperature ramps used in this work.

7.1.4 High-Performance Liquid Chromatography (HPLC)

HPLC analysis was also performed in alignment with procedures and materials provided by project partners at Pfizer. Details of methods are provided in Table 7.3.

Table 7.3: HPLC set-up and conditions.

Instrument	Waters 2695 HPLC Separator System
Detector	Photodiode array detector
Column	Fortis H_2O C_{18}
Dimensions	150mm x 4.6mm x 5μm
Temperature	20°C
Flow	1.5 ml/min
Injection	1 μl
Method	Reverse phase, isocratic
Buffer solution	10 mM NH_4OAc
Eluent A	90% Buffer : 10% MeOH
Eluent B	10% Buffer : 90% MeOH
Eluent ratio	65% eluent A : 55% eluent B
Run time	36 min
Scan wavelength	276 nm

7.1.5 Flash Chromatography

Biotage Isolera Four Flash Chromatography System fitted with a UV-Vis detector was in some cases employed to separate or purify desired products. The program used was automatically determined by the apparatus by applying the retention factor (rf) obtained from a previously run thin layer chromatography (TLC) in an optimised solvent system.

7.1.6 Proton Nuclear Magnetic Resonance Spectroscopy (^{1}H-NMR)

Products were characterized via proton nuclear magnetic resonance spectroscopy (^{1}H-NMR). The instrument details used for these analyses are reported in Table 7.4.

Table 7.4: ^{1}H-NMR set-up and conditions.

Instrument	Jeol ECS400 400MHz
Magnet	9.4 T Oxford AS400
Probe	TH5 (Jeol) 5mm NMR probe
Gradient	Pulsed-field along Z

7.2 Typical chromatograms and calibration methods

7.2.1 GC

Retention times of reaction components and corresponding calibration factors are reported in figures below as from prepared reaction samples and standard solutions.

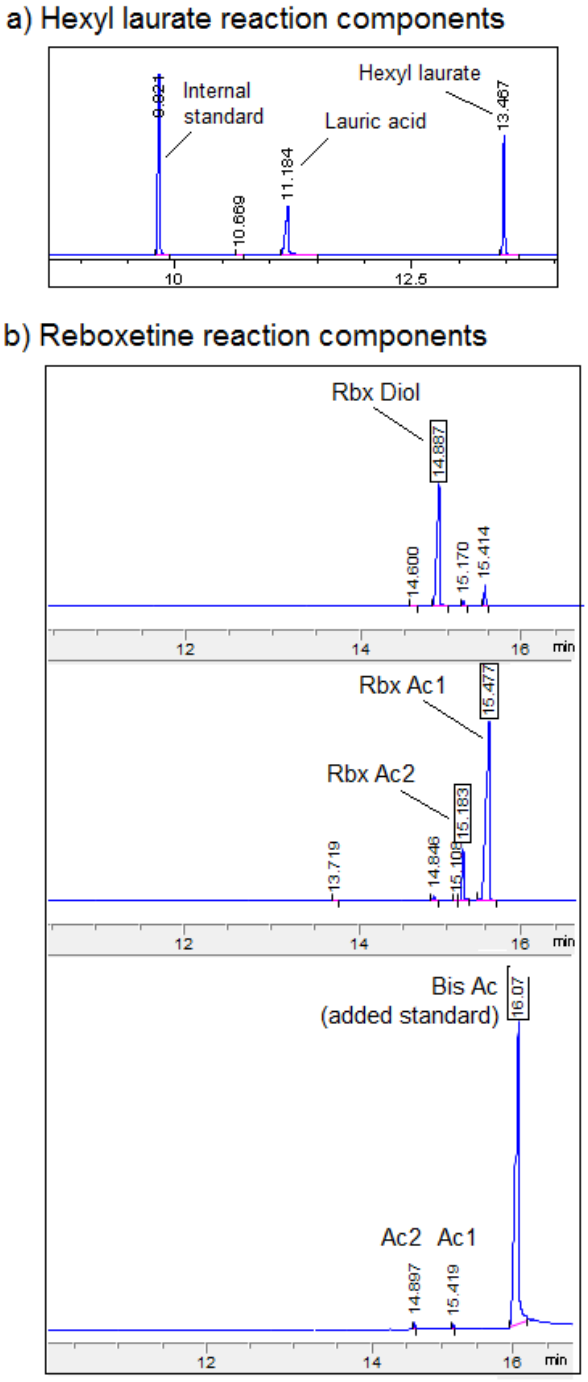


Figure 7.2: Typical GC chromatograms highlighting retention times of reaction components in the formation of a) hexyl laurate, b) Reboxetine acetate.

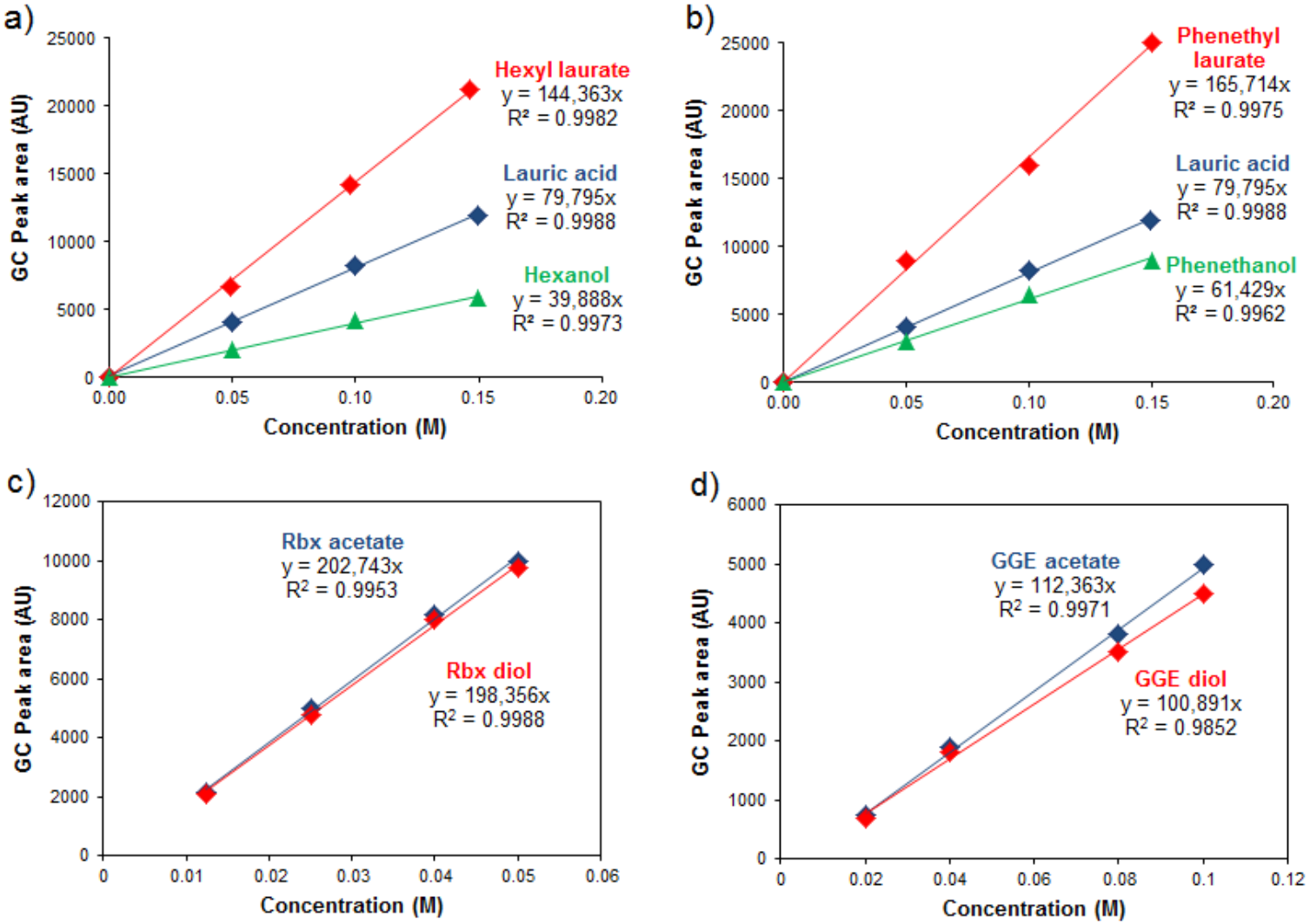
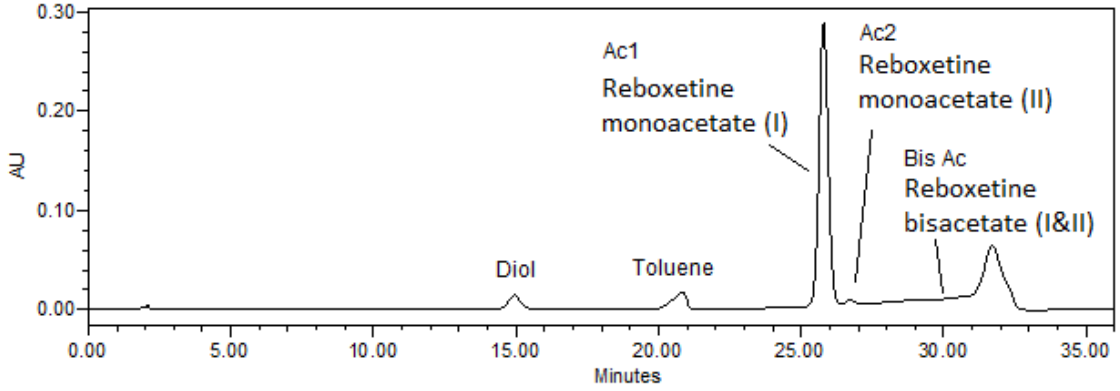


Figure 7.3: GC calibration of reaction components in the synthesis of a) hexyl laurate, b) phenethyl laurate, c) Reboxetine acetate, d) GGE acetate.

7.2.2 HPLC

Retention times of reaction components and corresponding calibration factors are reported in figures below as from prepared standard solutions.

1) Typical HPLC chromatogram of a reaction mixture



2) Same HPLC chromatogram after addition of Reboxetine bisacetate standard (ZOOMED)

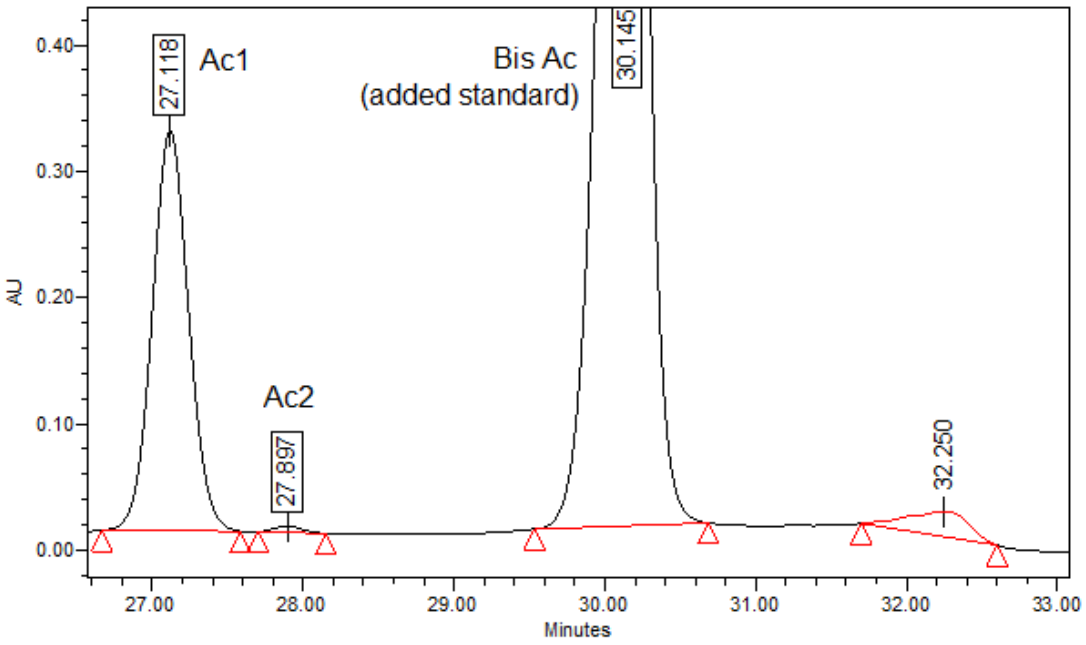


Figure 7.4: Typical HPLC chromatograms and peaks identification.

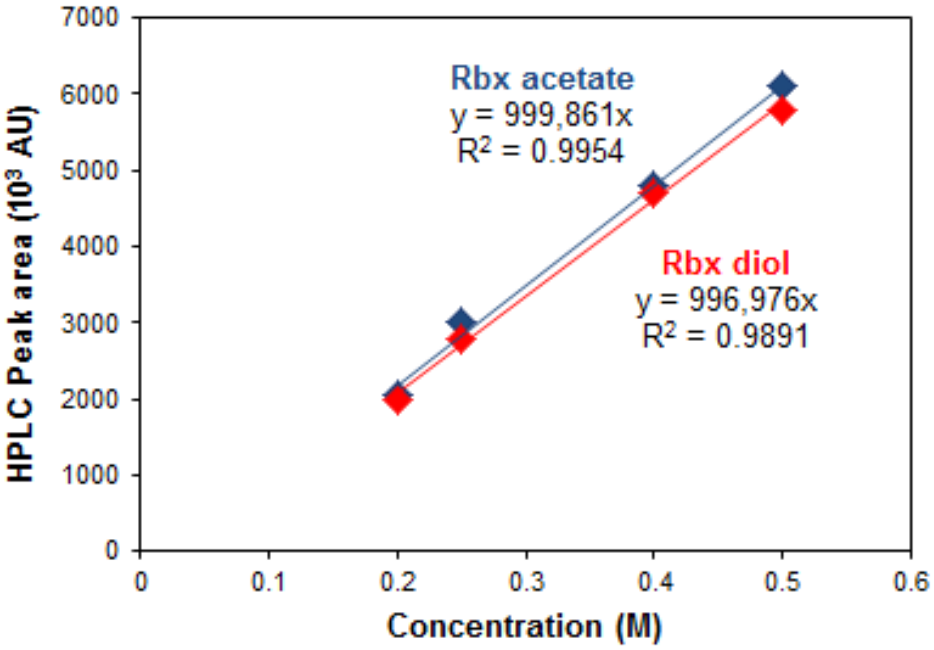


Figure 7.5: HPLC calibration of Reboxetine diol and acetate.

7.3 Determination of solvent properties

7.3.1 Kamlet-Taft parameters

Kamlet-Taft solvatochromic parameters were employed in the study of solvent properties amongst others. Individual values were obtained from published literature.[17,215,216] In the case of 2-methylfuran, no published data was found. Attempts were made to measure its α, β and π^* via spectroscopic method,[216,217] yet the solvent's natural absorbance was found to interfere with the wavelength range of interest for the analysis. Figure 7.6 presents the overlap of 2-methylfuran containing a trace of the three solvatochromic dyes (NR = nile red, NA = 4-nitroaniline, NN = *N*,*N*-diethyl-4-nitroaniline). A neat absorption peak of each dye would usually be found within the range of 200-400 nm. The results show high level of noise in this region due to the subtraction of the reference absorbance of the clean solvent.

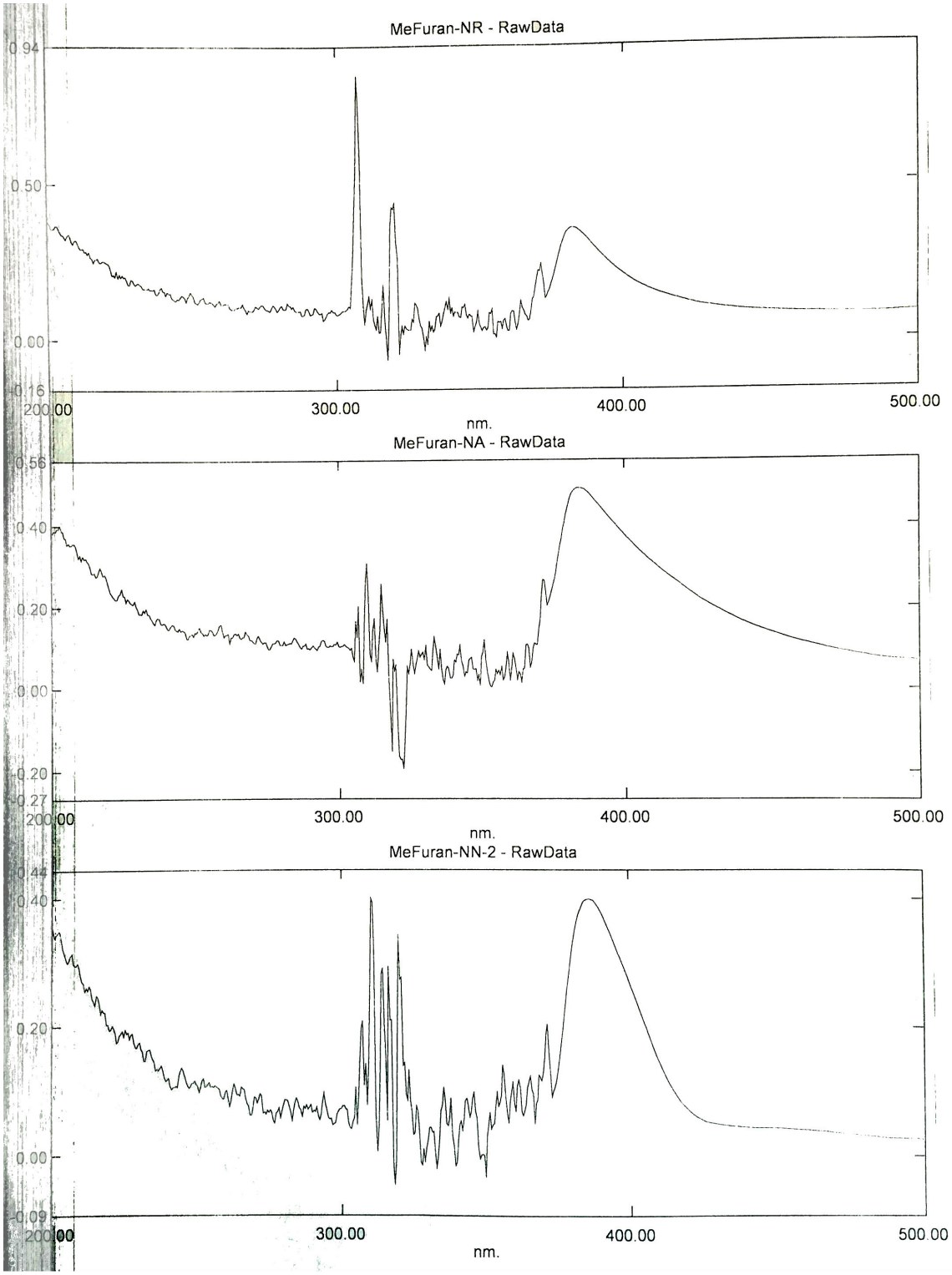


Figure 7.6: UV spectra of 2-methylfuran solutions of the three solvatochromic dyes - NR, NA, NN.

7.3.2 Hansen's parameters

Hansen's solubility parameters were obtained with HSPiP software (Figure 7.7).[59] Data for most solvents was already available from the software's database. For missing solvents, it was possible to predict Hansen's data through a dedicated function of the software by simply feeding the Simplified Molecular Input Line Entry System string (SMILES) of the solvent - *e.g.* cyrene: O=C1C(OC2)OC2CC1. This string can be generated for any molecule via most chemical editors, such as ChemDraw®.

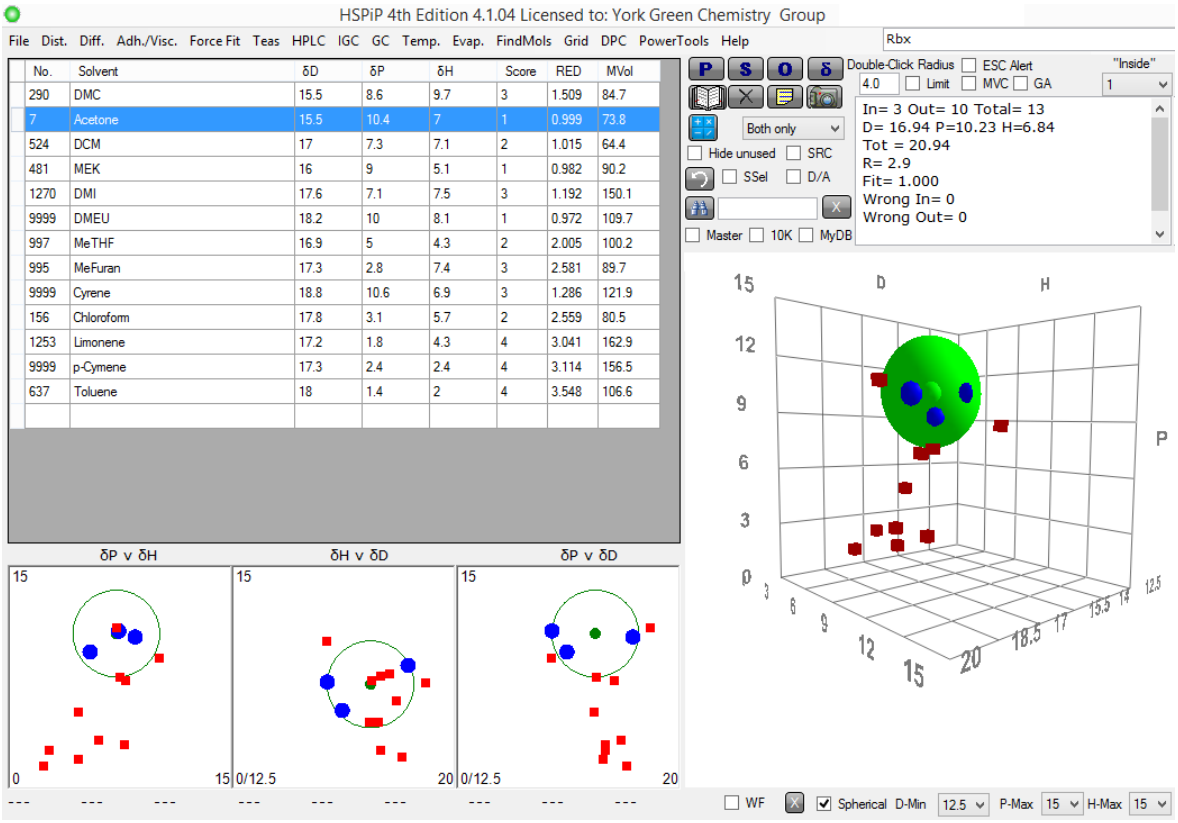


Figure 7.7: Use of HSPiP software for plotting Hansen parameters from the software's database.[59]

7.3.3 Water content

Water content was determined for each solvent via Karl-Fischer titration method without applying any previous drying.

The coloumetric procedure was performed with a Mettler-Toledo kit, where the active oxidising agent, I_2, was electrochemically generated *in situ*, while a methanolic solution of each solvent of was added volumetrically. Solvents containing carbonyl or acid/base functionalities are likely to alter the chemistry and/or stoichiometry of the Karl-Fisher reaction and should be avoided in standard conditions.[218]

The method is based on the redox chemistry reported in equations 7.3.1 and 7.3.2, where 1 equivalent of iodine is consumed per every equivalent of water present in the sample (B stands for base).

$$CH_3OH + SO_2 + B^- \rightarrow [CH_3SO_3H^+B^-] \tag{7.3.1}$$

$$H_2O + I_2 + [CH_3SO_3H^+B^-] + 2B^- \rightarrow [CH_3SO_4H^+B^-] + 2B^-[HI]^+ \tag{7.3.2}$$

7.3.4 Solubility of Reboxetine diol

For each of the solvents reported in Table 7.5, a 0.5 ml aliquot was placed into a small sample vial with a screw cap. Reboxetine diol was made into a fine powder into a mortar, and was added to each vial up to the point were no more solute was taken in solution. Vials were closed and agitated manually, then left for 24 hours on a rolling platform for continuous mixing. In cases where no precipitate remained, additional Reboxetine diol was added to the samples, which were left to mix for a further 24 hours. Once saturated solutions were obtained, their concentration was quantified by means of a GC calibration curve (Figure 7.3) employing an external standard.

Table 7.5: Data used for chromatographic measurement of Reboxetine diol content in saturated solutions.

Solvents	diol area (AU)	[standard] (M)	standard area (AU)	[diol] (M)	ln[diol]
DMEU	19376	0.0113	401.3	0.570	0.562
Acetone	12817	0.0113	399.3	0.379	0.970
MEK	12573	0.0113	399.6	0.371	0.990
2-MeTHF	10354	0.0113	398.8	0.307	1.182
Prop. Carb.	6524	0.0113	388.9	0.198	1.619
Cyrene	7230	0.0113	431.1	0.198	1.619
DiMe-isosorbide	6748	0.0113	403.7	0.197	1.623
2-MeFuran	5174	0.0113	394.4	0.155	1.865
Toluene	2351	0.0113	398.4	0.070	2.664
Limonene	507	0.0113	411.3	0.015	4.230
p-Cymene	340	0.0113	384.5	0.010	4.563
Water	13	0.0113	399.0	0.001	7.874
DCM	9017	0.0113	388.6	0.274	1.295
Chloroform	8724	0.0113	387.5	0.266	1.325

7.4 Experimental procedures

7.4.1 Reactions presented in Chapter 2

Synthesis of hexyl laurate phenethyl laurate with CALB - large scale

CALB Novozyme435 (0.100 g) and dodecanoic acid (10.000 g, 50 mmol, 1 eq) were weighed into a round-bottomed flask. Subsequently, either 1-hexanol (9.400 ml, 7.650 g, 75 mmol, 1.5 eq) or phenethyl alcohol (8.970 ml, 9.150 g, 75 mmol, 1.5 eq) were added to the flask in solvent-free conditions. The mixture was stirred at 40°C and followed by TLC until completion (3:1 ethyl acetate/cyclohexane, $rf_{hexyllaurate}$ = 0.70, $rf_{phenethyllaurate}$ = 0.81). The reaction mixture was transferred to a separating funnel and washed with water (3x2 ml). The organic phase was then submitted to flash chromatography (3:1 ethyl acetate/cyclohexane) to isolate the product.

Initial scoping reactions

CALB Novozyme435 (20 mg - amounts varied in loading screening) and dodecanoic acid (400 mg, 2 mmol, 1 eq) were weighed into a V-shaped Wheaton® vial. Subsequently, 1-hexanol (250 μl, 204 mg, 2 mmol, 1 eq - in specified reactions added as 10 eq excess) in 4 ml of solvent (heptane or other solvent screened) were added to the flask and stirred at 40°C. Reactions were followed by GC-FID.

Enzyme denaturation

CALB Novozyme435 (20 mg) was refluxed in boiling water for 30 minutes. The catalyst was filtered and dried in the oven overnight. The denaturated enzyme was then added to a solution of lauric acid and hexanol and residual catalytic activity was monitored over 8 hours.

Hydrolysis screening

CALB Novozyme435 (20 mg) and hexyl laurate (400 mg, 2 mmol, 1 eq) were weighed into a V-shaped Wheaton® vial. Subsequently, water (36 μl, 36 mg, 2 mmol, 2 eq) was added in 4 ml of solvent that had been pre-dried overnight using molecular sieves. The mixture was stirred at 40°C. Reactions were followed by GC-FID.

Screening of solvent effects

In a 5 ml Wheaton® sample vial was added 10 mg of CALB Novozyme435. The solvent was added by mass (dependent on density) to ensure accurate volume of 10 ml. Then, lauric acid (200 mg, 1 mmol, 1 eq) was added, and the reaction mixture and left stirring for about 15 minutes on a heating plate to reach a temperature of 40°C. An internal standard, tetradecane, was used and it was added with a calibrated pipette (27.1 μl, 19.8 mg, 0.1 mmol). A first sample for GC analysis was collected before the addition of the second reagent. Then 1-hexanol (102 mg, 1 mmol, 1 eq) was added to the reaction mixture using an automated pipette. Samples were collected at frequent intervals until the

reaction was stopped after 75 minutes. Each sample was collected using a pipette, filtered through cotton-wool and washed with DCM for the preparation of each GC vial. Dichloromethane was employed for analytical purposes at the beginning of the study and maintained throughout the screening for consistency. However, the use of alternative solvents would be preferred (*e.g.* acetone). In the case of the reaction carried out in water, the collected sample was washed with acetone through cotton-wool and transferred into a GC vial for the analysis.

Alternative H_2SO_4 catalysed hexyl laurate formation

Dodecanoic acid (100 mg, 0.50 mmol, 1 eq) and 1-hexanol (76 mg, 0.75 mmol, 1.5 eq) were dissolved in 4 ml of toluene. Subsequently, conc H_2SO_4 (18.4 M, 54.3 μl, 1 mmol) were added to the solution and stirred at 40°C for 8 hours. The reaction was followed by GC-FID although only traces of product were observed under these conditions.

Alternative *p*-toluenesulfonic acid catalysed hexyl laurate formation

Dodecanoic acid (100 mg, 0.50 mmol, 1 eq) and 1-hexanol (76 mg, 0.75 mmol, 1.5 eq) were dissolved in 4 ml of toluene. Subsequently, *p*-toluenesulfonic acid (172 mg, 1 mmol) was added to the solution and stirred at 40°C for 8 hours. The reaction was followed by GC-FID although no product was formed under these conditions.

Alternative zeolite catalysed hexyl laurate formation

Dodecanoic acid (100 mg, 0.50 mmol, 1 eq) and 1-hexanol (76 mg, 0.75 mmol, 1.5 eq) were dissolved in 4 ml of toluene. Subsequently, 5 mg of H-β Zeolite (5 wt%) were added to the solution and stirred at 40°C for 8 hours. The reaction was followed by GC-FID although no product was formed under these conditions.

7.4.2 Reactions presented in Chapter 3

Screening for regioselectivity

Reboxetine diol (100 mg, 0.347 mmol, 1.0 eq) and CALB Novozyme435 (2 mg, 3 wt%) were weighed in a 3 ml sample vial. The solvent (1.0 ml) and isopropenyl acetate (0.0764 ml, 0.694 mmol, 2 eq) were added with a micropipette and the mixture was taken to 50°C in a orbital thermo-shaker for 6 hours. In preparation for the HPLC analysis, the samples were centrifuged and 20 μl of supernatant were diluted to 1 ml in MeCN.

Synthesis of Reboxetine acetate with CALB

CALB Novozyme435 (3 mg) was weighted into a V-shaped 4 ml Wheaton® vial. Reboxetine diol (100 mg, 0.35 mmol, 1 eq) and isopropenyl acetate (70 mg, 0.69 mmol, 2 eq) in *p*-cymene (5 ml) were added to the vial. The mixture was stirred at 40°C and followed by TLC until completion (3:1 ethyl acetate/cyclohexane, rf_{diol} = 0.42, $rf_{acetate}$ = 0.70). The reaction mixture was transferred to a separating funnel and washed with water (3x2 ml). The organic phase was then submitted to flash chromatography (3:1 ethyl acetate/cyclohexane) to isolate the acetate product.

Synthesis of Reboxetine bisacetate with acetyl chloride

Reboxetine diol (100 mg, 0.35 mmol, 1 eq) and triethylamine (70 mg, 0.69 mmol, 2 eq) were added to *p*-cymene (3 ml). The mixture was stirred at room temperature and acetyl chloride (54.5 mg, 0.69 mmol, 2 eq) was added dropwise over 10 min. The reaction mixture was transferred to a separating funnel and washed with water (3x2 ml). The organic phase was then submitted to flash chromatography (3:1 ethyl acetate/cyclohexane, rf_{diol} = 0.41, $rf_{monoacetate}$ = 0.69, $rf_{bisacetate}$ = 0.77) to isolate the bisacetate product.

Screening for kinetic analysis

Reboxetine diol (100 mg, 0.347 mmol, 1.0 eq), CALB Novozyme435 (2 mg, 3 wt%) and the internal standard 4-methylbiphenyl (5.5 mg, 0.033 mmol, 0.09 eq) were weighed in a V-

shaped 4 ml Wheaton® vial. The solvent (3.0 ml) and isopropenyl acetate (0.0764 ml, 0.694 mmol, 2 eq) were heated to 40°C in a separate vial. The heated liquid was then added to the V-shaped vial. The mixture was magnetically stirred at 40°C for 3 hours while samples were taken at multiple time-points. The samples were collected with a glass pipette and filtered through cotton-wool and washed with DCM for preparation of GC samples. This procedure was reproduced for the acetylation of guaiacol glyceryl ether (GGE).

Telescoped series: diol to zwitterion

Step 1: Acetylation To a 25 ml round-bottomed flask were added *p*-Cymene (2.5 m) and isopropenyl acetate (400 μl, 347 mg, 3.47 mmol, 2 eq) followed by Reboxetine diol (500 mg, 1.736 mol, 1 eq). Novozyme435 (10 mg, 0.2 wt%) was added, and the reaction mixture was stirred and heated to 40°C until the complete conversion of the starting material was observed by TLC (3:2 ethyl acetate/cyclohexane, rf_{diol} = 0.42, $\mathrm{rf}_{acetate}$ = 0.71). The catalyst was removed by filtration, and the crude filtrate solution was progressed to step 2.

Step 2: Mesylation The solution from step 1 was stirred at 20°C and triethylamine (350 μl, 252 mg, 2.50 mol, 1.7 eq) was added in one portion. A solution of methanesulfonyl chloride (160 μl, 235.6 mg, 2.06 mol, 1.4 eq) in *p*-cymene (2 ml) was then added dropwise over 3 h. Reaction was followed by TLC until completion (3:2 ethyl acetate/cyclohexane, $\mathrm{rf}_{acetate}$ = 0.56, $\mathrm{rf}_{acetate-mesylate}$ = 0.85, $\mathrm{rf}_{mesylate}$ = 0.69) then progressed to step 3 .

Step 3: Epoxide Formation To the solution from step 2, aqueous HCl (1 M, 2.5 ml) was added in a single portion to cleave off the acetate group. The biphasic mixture was agitated for 30 min, then transferred to a separating funnel. The phases were separated, and the upper organic phase was retained. To the extracted organic stream, a solution of NaOH (464 mg, 11.60 mmol, 10 eq) in water (2 ml) and methyltributylammonium chloride (27.2 mg, 115 mmol, 0.1 eq) was added to favour the epoxide ring closure with elimination of the mesylate sodium salt. The resulting mixture was agitated for 3.5 h. The phases were

allowed to separate and the lower aqueous phase was discarded. The upper organic phase was washed with water (2 ml) and progressed to step 4.

Step 4: Zwitterion Formation A separate solution of 2-aminoethyl hydrogen sulfate (409 mg, 2.90 mmol, 2.5 equiv), 1,8-diazabicyclo-[5.4.0]undec-7-ene (DBU, 429 μl, 2.90 mmol, 2.5 eq) in *p*-cymene (0.5 ml) and ethanol (0.5 ml) was prepared. The solution was heated to 65°C and stirred at this temperature for 1 h. The organic solution from step 3 was diluted with *p*-cymene (1 ml) and added to the activated amine mixture over 1 h. The temperature was risen to 70°C and the mixture was further agitated for 2 h. After cooling to room temperature, the mixture was washed with an aqueous solution of sodium hydroxide (103 mg in 3 ml of water, 3.0 eq NaOH). The upper organic phase was discarded, and the lower aqueous phase was treated with concentrated HCl (to pH = 5) resulting in crystallization of the product. The latter was isolated by filtration, washed with water (5 ml), and dried to yield the desired product as an off white solid.

7.4.3 Reactions presented in Chapter 4

Screening for thermodynamic analysis

Reboxetine diol (50 mg, 0.174 mmol, 1.0 eq), CALB Novozyme435 (2 mg, 3 wt%) and the internal standard 4-methylbiphenyl (5.5 mg, 0.033 mmol, 0.18 eq) were weighed in a V-shaped 4 ml vial. In a separate vial, the solvent (3.0 ml) and isopropenyl acetate (0.0764 ml, 0.694 mmol, 2 eq) were taken to desired temperature (25°C, 32.5°C, 40°C, 55°C) on a heating plate. The liquid phase was then added to the solid mixture in the V-shaped vial. The mixture was maintained at the set temperature and magnetically stirred for 8 hours with samples taken every 20 min. The samples were collected with a glass pipette and filtered through cotton-wool and washed with DCM for GC analysis. Procedure was reproduced for the acetylation of guaiacol glyceryl ether (GGE).

7.5 Product characterisation

_n_-Hexyl laurate

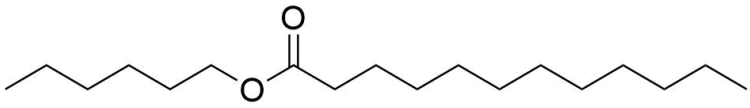


Figure 7.8: Chemical structure of hexyl laurate.

^{1}H-NMR (400 MHz, $CDCl_3$) δ 4.08 (t, J = 6 Hz, 2H), 2.30 (t, J = 7.6 Hz, 2H), 1.66-1.59 (m, 4H), 1.40-1.25 (m, 26H), 0.92-0.88 (m, 6H). **MS (+EI)** *m/z* 285 (M^+).[219]

Phenethyl laurate

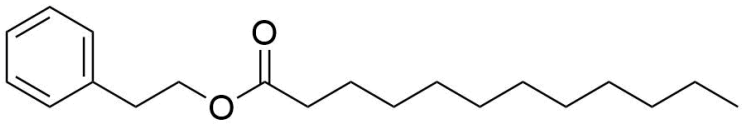


Figure 7.9: Chemical structure of phenethyl laurate.

^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.35-7.18 (m, 5H), 4.32 (t, J = 7, 2H), 2.91 (t, J = 7.1, 2H), 1.62 (m, 2H), 1.34-1.25 (m, 16H), 0.90 (t, J = 8, 3H). **MS (+EI)** *m/z* 305 (M^+).[219]

Reboxetine acetate

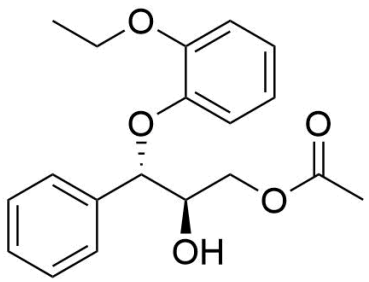


Figure 7.10: Chemical structure of mono-acetylated reboxetine diol.

^{1}H-NMR (400 MHz, DMSO-D_6) δ 7.54–7.32 (m, 2H), 7.32–7.24 (m, 2H), 7.24–7.12 (m,

1H), 6.89 (dd, J = 8.1, 1.4 Hz, 1H), 6.78 (td, J = 7.6, 1.9 Hz, 1H), 6.73–6.55 (m, 2H), 5.25 (d, J = 5.9 Hz, 1H), 5.07 (d, J = 6.5 Hz, 1H), 4.26 (dd, J = 11.2, 3.1 Hz, 1H), 4.09 (dd, J = 11.2, 7.1 Hz, 1H), 3.99 (q, J = 7.0 Hz, 3H), 3.32 (s, 2H), 2.46 (dt, J = 3.6, 1.8 Hz, 1H), 1.94 (d, J = 4.5 Hz, 4H), 1.32 (t, J = 7.0 Hz, 3H), 1.13 (t, J = 7.1 Hz, 1H). **MS (+EI)** *m/z* 331 (M^+).[141]

GGE acetate

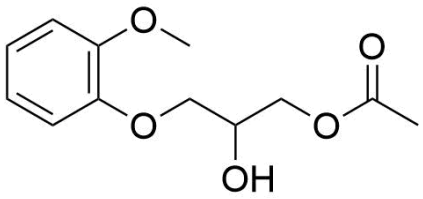


Figure 7.11: Chemical structure of mono-acetylated guaiacol glyceryl ether.

^{1}**H-NMR (400 MHz, DMSO-D$_6$)** δ 6.92-6.88 (m, 4H), 4.19 (q, 2H), 4.10-3.79 (m, 5H), 3.13 (s, 1H), 2.20 (s, 3H), 1.33 (t, J = 8.0 Hz, 3H). **MS (+EI)** *m/z* 241 (M^+).[220]

Reboxetine bisacetate

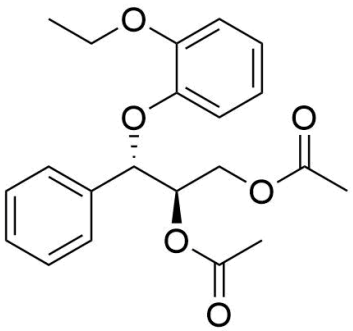


Figure 7.12: Chemical structure of bis-acetylated reboxetine diol.

^{1}**H-NMR (400 MHz, DMSO-D$_6$)** δ 7.39-7.35 (m, 5H), 6.91-6.87 (m, 4H), 5.60 (d, J = 7.1 Hz, 1H), 5.45 (q, J = 7.0 Hz, 1H), 4.25-4.39 (m, 2H), 4.09 (q, J = 8.0 Hz, 2H), 2.20 (s, 6H), 1.31 (t, J = 8.0 Hz, 3H). **MS (+EI)** *m/z* 373 (M^+).[141]

Reboxetine mesylate

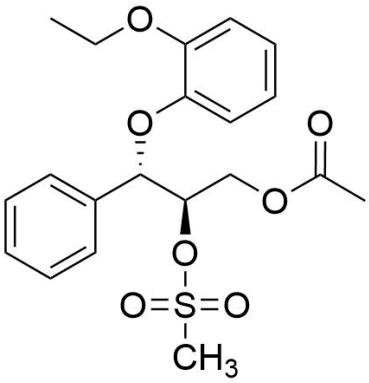


Figure 7.13: Chemical structure of Reboxetine mesylate.

^{1}H-NMR (400 MHz, DMSO-D$_6$) δ 7.54–7.19 (m, 5H), 6.92–6.63 (m, 4H), 5.61 (d, J = 4.0 Hz, 1H), 5.19 (dt, J = 7.2, 3.3 Hz, 1H), 4.40 (dd, J = 12.2, 3.1 Hz, 1H), 4.04 (dd, J = 11.6, 7.7 Hz, 1H), 3.99 (dd, J = 14.3, 7.3 Hz, 2H), 3.34 (s, 3H), 3.18 (s, 3H), 1.31 (t, J = 7.0 Hz, 3H). **MS (+EI)** *m/z* 409 (M^+).[141]

Reboxetine zwitterion

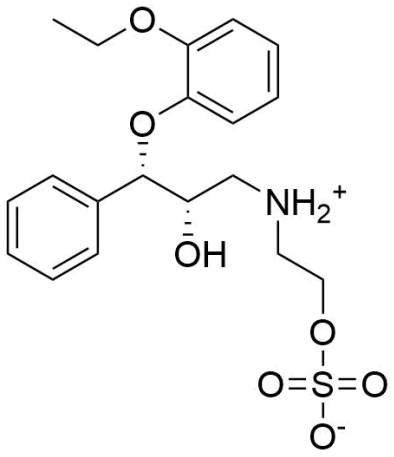


Figure 7.14: Chemical structure of Reboxetine zwitterion.

^{1}H-NMR (400 MHz, CD$_3$OD) δ 7.40-7.27 (m, 5H), 6.93-6.84 (m, 2H), 6.74-6.68 (m, 2H), 5.14 (d, J = 5.3 Hz, 1H), 4.26-4.22 (m, 3H), 4.12-4.10 (m, 2H), 3.35 (m, 2H), 3.15 (m, 2H), 1.43 (t, J = 7.1 Hz, 3H). **MS (+EI)** *m/z* 412 (M^+).[141]

Chapter 8

Conclusion and future work

The aim of this final Chapter is to provide:

1. an overview of the key outcomes of this work,
2. an evaluation of the adopted approach identifying potential for improvement,
3. a vision for future research in light of the findings presented.

Detailed summary and conclusions based on the results from each study are discussed at the end of individual chapters.

In Chapters 2, 3 and 4 were illustrated various solvent screenings carried out on esterifications catalysed by *Candida Antarctica* lipase type B (CALB), a commercial biocatalyst supported on acrylic resin. These assays had the dual purpose of 1) identifying the best classes of bio-based solvents for such reactions as alternatives to conventional petrochemical solvents, and 2) providing data for statistical multivariate analyses aimed at clarifying the mechanisms of solvent influence. Past reported attempts had encountered difficulties due to the inherent complexity of non-aqueous enzyme-catalysed systems. Typically, these were limited by the use of single-variable correlation approaches which had predominantly fallen short in providing a comprehensive explanation to the observed effects. This difficulty is often attributed to the dichotomy surrounding the very nature of solvents: continuum media with bulk properties? Or chemical species with molecular interactions?

Herein, in-depth statistical analyses of solvent effects were performed, where initial rate of reaction (also under the influence of temperature) was used as a dependent variable against a large set of solvent parameters as independent variables. A preliminary investigation was carried out by screening a large set of green and conventional solvents in the formation of hexyl laurate, a fatty ester of commercial relevance. A linear regression analysis was performed on the data, where a gradient-based reaction rate measurement was used as dependent variable (y) and six solvent properties as independent variables (x_i), including Kamlet-Taft solvatochromic parameters, molar volume, Hildebrand's parameter, and water content. A Partial Least Squares method was employed to assess the correlation using various combinations of independent variables, which where manually refined and narrowed down based on statistical parameters such as F-test, P-value, and least squares

(R^2). This procedure identified the model equation 8.0.1, which describes the reaction system of study with an error confidence of 95%.

$$ln(rate) = -3.70 - 4.53\beta + 0.044V_m \tag{8.0.1}$$

Pleasingly, this model equation, which is based on Kamlet's hydrogen bond accepting ability (*i.e.* 'basicity') and molar volume, provided significant improvement over the traditional and long established description based solely on partition coefficient logP, bringing over 10% increase in accuracy (R^2 increasing to 0.9017 from 0.7863). Overall, the publication of this work in 2014 constituted a first example of 1) citrus-derived solvents being used as media in a biocatalysed reaction, and 2) a multivariate regression applied to the study of solvent effects in a biocatalysed system.

In Chapter 3, a similar yet more advanced approach was explored as part of a second case study surrounding the lipase-catalysed acetylation of Reboxetine diol, a commercial pharmaceutical intermediate. In summary, the analysis was improved in that: 1) screening was focused on bio-based or potentially bio-based solvents, 2) the solvent property set was expanded to include substrate solubility and Hansen's parameters, 3) a more rigorous kinetic analysis and determination of reaction rate was carried out, 4) a more advanced statistical method was employed which encompassed an assessment for multicollinearity and variable significance. As a result, this approach yielded a more robust and statistically relevant interpretative model, based on equation 8.0.2. The model equation developed through this analysis highlighted the dominant impact of Hansen's hydrogen bond parameter in determining changes in rate, with smaller contribution from other variables such as molar volume and Hansen's polar forces.

$$ln(r) = -7.80 + 0.00366V_m + 0.0430\delta_p - 0.261\delta_h \tag{8.0.2}$$

Overall, the two individual model equations indicate that the ability of a solvent to engage (or rather not to engage) in hydrogen bonding is critical in driving observed enzymatic activity and reaction performance. To a lesser extent, the solvents' molar volume can

also affect the initial rate of reaction.

In order to be able to interpret these findings in a mechanistic context, the study was progressed by introducing another variable in the reaction system: temperature. This work, presented in Chapter 4, revealed the presence of a solvent-induced isokinetic effect and an energy compensation effect resulting from a thermal screenings in multiple solvents. Firstly, the isokinetic effect indicated that with increasing temperature, the influence of the by solvent on individual reaction rates decreased, ultimately leading to convergence to a common isokinetic temperature, T_{iso}. Secondly, the evidence for compensation effect demonstrated that enthalpy and entropy exert opposite energetic drivers on the process, with the former dominating in driving the course of reaction. Interestingly, the enthalpic excess for the reaction in each solvent is such that the benefit in overall free activation energy is constant, which defines the linear compensation effect according to equation 8.0.3:

$$\Delta\Delta G^{\ddagger}_{i,j} = \Delta\Delta H^{\ddagger}_{i,j} - T\Delta\Delta S^{\ddagger}_{i,j} = 0 \tag{8.0.3}$$

This phenomenon also implies the presence of a common compensation temperature, T_c, the temperature at which the effect of solvent selection upon thermodynamics of the process will be equal. While it is not unusual for isokinetic temperature and compensation temperature to be equivalent or comparable - such as in this case -, the existence of one effect does not necessarily imply the existence of the other. By changing the substrate of study, the isokinetic and compensation effects were confirmed. The equivalence of T_{iso} and T_c was confirmed with the second substrate, GGE, although showing to be substrate-dependent in value:

Reboxetine acylation: T_{iso} = 66°C; T_c = (71.1 $\pm$ 5.8)°C

GGE acylation: T_{iso} = 84°C; T_c = (86.4 $\pm$ 4.3)°C

Ultimately, the outcomes of the kinetic and thermodynamic studies allowed a preliminary model explaining the role of the solvent in biocatalysed esterifications to be developed. The best solvents for the lipase-catalysed esterifications appeared to be those that

1) are least capable of coordinating, solvating or stabilising the reactive -OH functionalities of the substrates, due to a low Hansen's δ_h, and 2) bring the most favoured enthalpic contribution, by lowering $\Delta H^{\ddagger}$ at the expense of the entropic contribution, $\Delta S^{\ddagger}$. Thanks to the double nature of Hansen's parameters, they can be used to describe solvent behaviour either in terms of molecular electronic effects or in terms of bulk cohesive forces. Therefore, it can be concluded that δ_h acts both at molecular level and at bulk level to impart energy benefits/penalties on the reaction course, and thus influences both solute coordination and solute partition across the solvent-aqueous interface.

From a molecular perspective, this translates in low-δ_h solvents driving the reaction force by:

1. creating an energy incentive through product stabilisation;
2. leaving substrates and transition state uncoordinated;

At bulk level, the results suggest that low-δ_h solvents can:

1. solubilise the product in the organic phase and, therefore, prevent it from accumulating in the active-site's surroundings where it may be reversed through the hydrolytic process;
2. form a loosely structured medium, which can favour diffusion and partition of solutes to and from the active site.

While this interpretation is made on the bases of experimental evidence in this work, it may still be an over-simplification. The origin and significance of the energy compensation effect are still highly disputed in the literature, and the limitations of this theory and interpretation are acknowledged. Further work could help elucidate the origin of the enthalpy effect by investigating the solvent effects on equilibrium constants (k_{eq}) and energy of formation (ΔH_f), and thus complementing the analysis pursued here on rate constants (k_{rate}) and energy of activation ($\Delta H^{\ddagger}$).

In general, it would be of particular interest to compare and validate these outcomes against other transformations and other enzymes, especially amongst those relevant to

pharmaceutical synthesis. Some additional work was carried out investigating the solvent effects on the activity of *Pseudomonas Stutzeri* lipase (PSL, also known as Amano-TL) in catalysing the formation of amides on drug-like model structures. This ongoing work in collaboration with CatSci, a CHEM21 partner, showed similar trends to those described in this thesis in relation to *Candida Antarctica* catalysed esterifications (results summarised in Appendix C). Other preliminary screenings involved ω-transaminase and galactose oxidase, able to catalyse ketone-to-amine transformations - and *viceversa* -, and the selective oxidation of alcohols to aldehydes (summarised in Appendix C). Consideration of these enzymes was only partially undertaken due to the higher complexity of the systems (*e.g.* requiring cofactor, co-enzyme partner). The potential for applying these enzymes in organic solvents remains vastly unexplored, as well as the understanding of solvents effects on their activity. These enzymes are naturally capable of catalysing transformations that are difficult to achieve chemically and can do so in mild conditions and with high selectivity. Therefore, there would be great scope and value for the scientific and industrial community in investigating their solvent-tolerance, especially in sustainable solvents.

Beyond the study of solvent effects, another key aim of this work was to identify the most suitable bio-based solvent candidates for use in lipase-catalysed reactions, thus simplifying future screening efforts and suggesting a direction for research in medium optimisation. Across the screening work here presented, *p*-cymene and limonene gave consistently outstanding performance. The greener credentials of these solvents over more traditional petrochemical counterparts (*e.g.* hexane and toluene) were also discussed. From experience gained in this work, *p*-cymene and limonene should be considered as top media choices in future studies of lipase-catalysed processes. A recommended list for future bio-based solvent screenings would also include 2-MeTHF, propylene carbonate, and acetone. These solvents represent different classes and areas of the solvent space and have shown good performance in the reactions of study - yet subject to more substrate variability (Figure 8.1). In addition, there is scope for consideration of additional solvents not screened in this study, including new commercial green solvents, as well as miscible or biphasic

co-solvent systems.

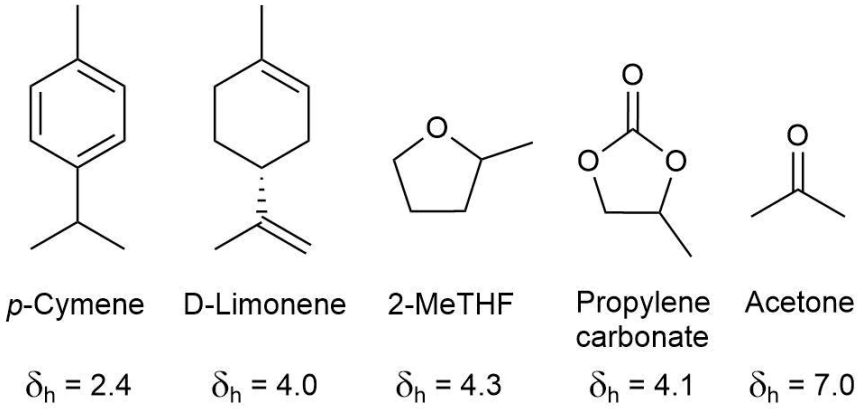


Figure 8.1: Most recommended bio-based solvents for future enzymatic screenings.

In the Reboxetine case study, these were proven to be the best alternatives to toluene in terms of performance as well as greenness. Moreover, the substitution with *p*-cymene was successfully taken forward through various downstream synthetic steps leading to isolation of the final product with comparable yield to toluene. In this case the desired product was a zwitterion salt which partitioned out of the organic and into the aqueous phase from which it was later crystallised. In general, pharmaceutical intermediates tend to increase in hydrophilicity as they progress towards the final API structure, which makes a very appropriate scenario for the use of citrus-derived solvents in late stage synthetic steps where the product is most likely to phase out of the organic medium. Since distillation of these solvents is not recommended due to the high boiling points, emerging separation techniques may be considered as effective and greener alternatives, *e.g.* membrane separation, supercritical fluid chromatography, vapour permeation, among others.

In the second part of this thesis, the potential for applicability of green and bio-based solvents was explored from an alternative angle, which benefited from the use of socio-economic methodologies. In Chapter 5, a stakeholder perception survey was carried out in order to identify barriers that individuals working in the solvent sector feel are hindering the adoption of green solvents. These were grouped into categories and then quantified and prioritised. This study was envisaged as a proof-of-concept to the development of a more comprehensive future survey targeting a wider pool of experts and stakeholders, most

preferably with the support of a relevant industry group or policy-advising organisation. Future work in this direction would allow more robust conclusions to be drawn and add substantial depth to the analysis.

From the survey, it appears that technological issues related to the removal and recycling of high boiling solvents mentioned earlier are, in fact, not perceived as significant barriers to the adoption of green solvents. Technological factors consistently ranked bottom of the eight categories of concern. The voting activity carried out during the participatory workshop highlighted the top perceived barriers to be addressed as a priority: cost, lack of data, and availability & supply.

The dominant issue, cost, received 46% of votes, and pointed to the fact that green solvents are often more expensive or costly to develop than common-use petroleum-derived solvents. This fact is typical of any novel technology in comparison to an established one, and is usually due to economy of scale. In the comparison of limonene and toluene for bulk use as cleaning solvent, there is currently 3.5-fold difference in price (1.4 USD/ton *versus* 0.4 USD/ton).[188] That said, it is important to bear in mind that the difference in total market volume between the two solvents is of approximately three orders of magnitude (10^4 MT *versus* 10^7 MT). Further scale-up of this technology and more wide-spread citrus-waste exploitation will help to narrow these differences.

Secondly, lack of data attracted the 42% of votes, stressing the issue that many green solvents still have an incomplete EHS profile, especially with regards to long term effects on humans and the environment. Many of the less-common solvents contemplated in this study have never yet been employed such in large scale chemical processing and therefore limited know-how can be gathered concerning their use. Of special relevance to the pharmaceutical case-study here presented, the lack of guidance on the classification of solvent impurities for emerging green solvents such as limonene and *p*-cymene can constitute a significant obstacle to the uptake of these solvents in the industry (*e.g.* ICH limits).[221]

Thirdly, the survey highlights concerns over the availability of bio-based feedstocks and the overall issues of supply capacity for bio-based solvents. This aspect received 26% of total votes. Although its importance emerged clearly from this study, very limited research

has investigated this issue so far.

The work presented in Chapter 6 intended to fill this gap and answer the question: can the supply of bio-based solvents meet demand? This work was carried out in the form of a case-study with a focus on various scenarios of potential limonene supply evaluated against the demand for toluene in a specific application, *i.e.* as cleaning solvent.

The results showed that complete substitution at global level will remain out of reach unless significant reduction in consumption rates takes place. In 2014, global toluene consumption was quoted as 14,790,000 MT, of which approximately 813,400 MT (5.5%) were employed in cleaning applications. The calculations carried out by extrapolation of historic trends suggest that toluene consumption in cleaning will approach 1,000,000 MT/y by 2030, while the maximum potential supply of bio-derived limonene from citrus-waste is estimated to reach *ca.* 660,000 MT/y. In the study, a breakdown of this this maximum capacity was given which highlighted the contribution of juicing processes to the global volume of citrus-waste available for limonene extraction.

Substitution appears more feasible and encouraging from a regional perspective. Data on toluene consumption and citrus production was gathered for nine regions/countries in order to bring a different geographic dimension into the analysis. Overall, citrus-producing and citrus-juicing countries appear very well placed to achieve full substitution of toluene with citrus-based limonene in the cleaning sector. Indeed, regions such as Brazil and Florida show the highest potential for replacing their toluene consumption as cleaning agent by means of regionally sourced renewable limonene. For these top citrus-producing regions it would suffice to extract limonene from the citrus residues generated from local juicing facilities, which would largely facilitate enabling the logistics of this substitution. Spain, India, and South Africa also have promising potential to substitute toluene with limonene derived from their citrus produce, however a more challenging complex system for collection exploitation of peels from all stream of consumption would be required.

Although leading citrus-growing regions such as China, USA and Europe have the capacity to produce the largest volume for limonene in the world (combined they could

currently supply 180,000 MT/y), they also are the world's biggest consumers of toluene. In 2013, it is estimated that China alone consumed approximately 278,000 MT of toluene in cleaning applications, followed by USA with 120,000 MT, and Europe with 69,000 MT. For these vast regions with such strong chemical industries, limonene could partially contribute to replacing petroleum-derived toluene as a cleaner, although a full substitution scenario appears unlikely.

Ultimately, at global as much as at regional level, efforts towards effective reduction, reuse and recycling of solvents will remain critical in achieving sustainability. Nevertheless, great promise for a bio-based transition in the cleaning sector is demonstrated across many citrus-producing countries. Interestingly, even a non-citrus-growing country such as Germany showed significant capability for limonene production, based on the large volumes of citrus fruits that are imported in the country every year. By exploiting the waste from these fruits, Germany has the potential to achieve 30% substitution of it toluene consumption as cleaning agent by means of renewable limonene.

Up to now, the substitution potential for a bio-derived chemical had never been quantified against demand of its common petrochemical counterpart. This very underexplored field of research has a vast potential for expansion in terms of design, scope, and impact. Key to achieving this development, is the availability of data in publicly accessible databases, which in the future will possibly provide increasingly accurate information on many aspects of production and trade that are relevant for the growth of green products.

Overall, the work presented in this thesis demonstrates that there is extensive scope for the use of bio-based solvents in the chemical industry, by means of experimental proof-of-principle backed by mechanistic interpretation, as well as by socio-economic and geo-technical assessment of the potential and the challenges faced. The results also indicate that one bio-based chemical alone will not be able to sustainably replace the global consumption of petrochemicals (such as discussed in relation to toluene). Significant market diversification would be required to allow multiple bio-based options to target specific applications, taking advantage of the incredible level of functionalisation and diversity that

bio-based chemicals are able to provide. Ultimately, reducing consumption and shifting to a circular utilisation of materials and products will likely be the most determinant factors in driving sustainability, with waste utilisation and bio-based technologies as fundamental tools in the future design of materials and solvents.

Appendix A

A.1 Kinetic study

A.1.1 Kinetic profiles

Reactions in different solvents were monitored via GC over 25 hours. Data herein shows the relevant interval used for deriving each reaction's rate constant. In all figures, clockwise from top-left corner: concentration of starting material; linearised conversion of starting material yielding reaction rate constant from gradient; concentration of regioisomer byproduct, concentration of desired product.

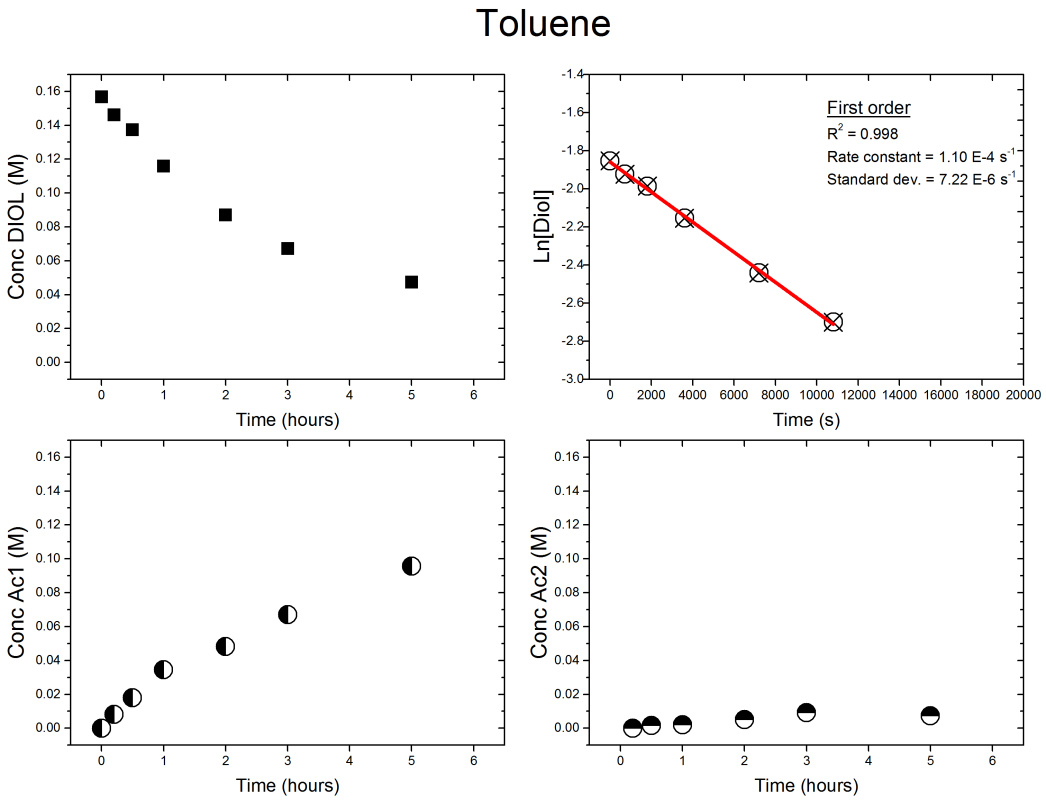


Figure A.2: Reaction profiles in toluene.

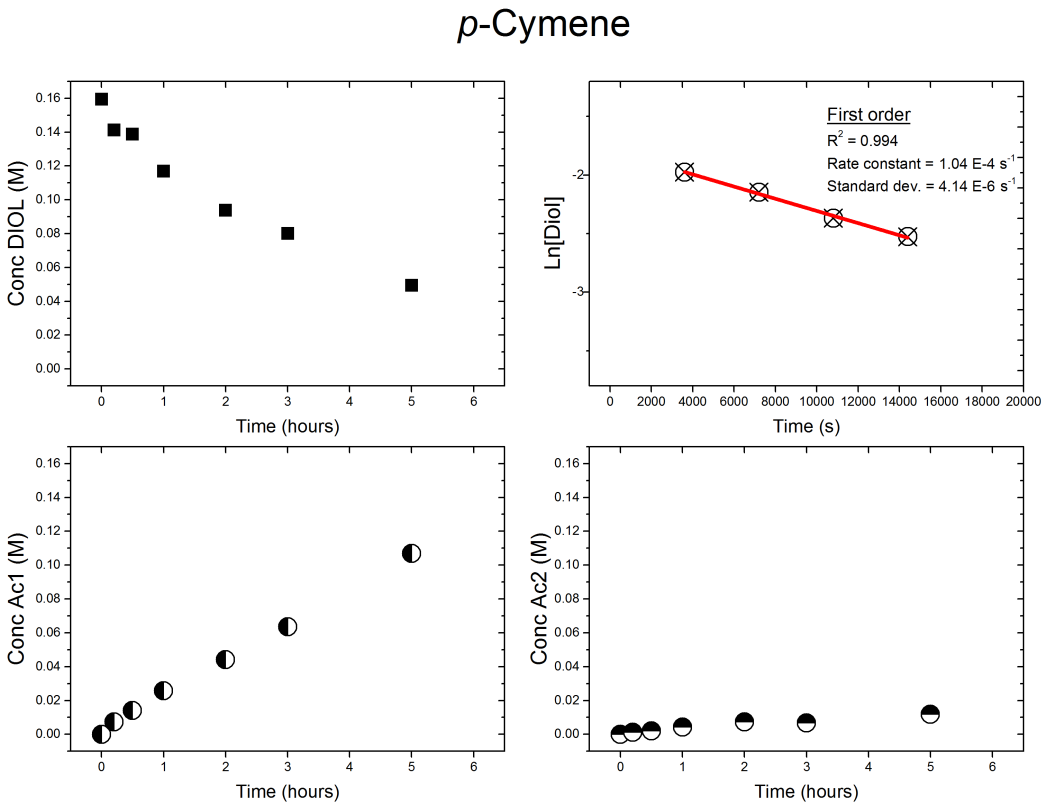


Figure A.3: Reaction profiles in *p*-cymene.

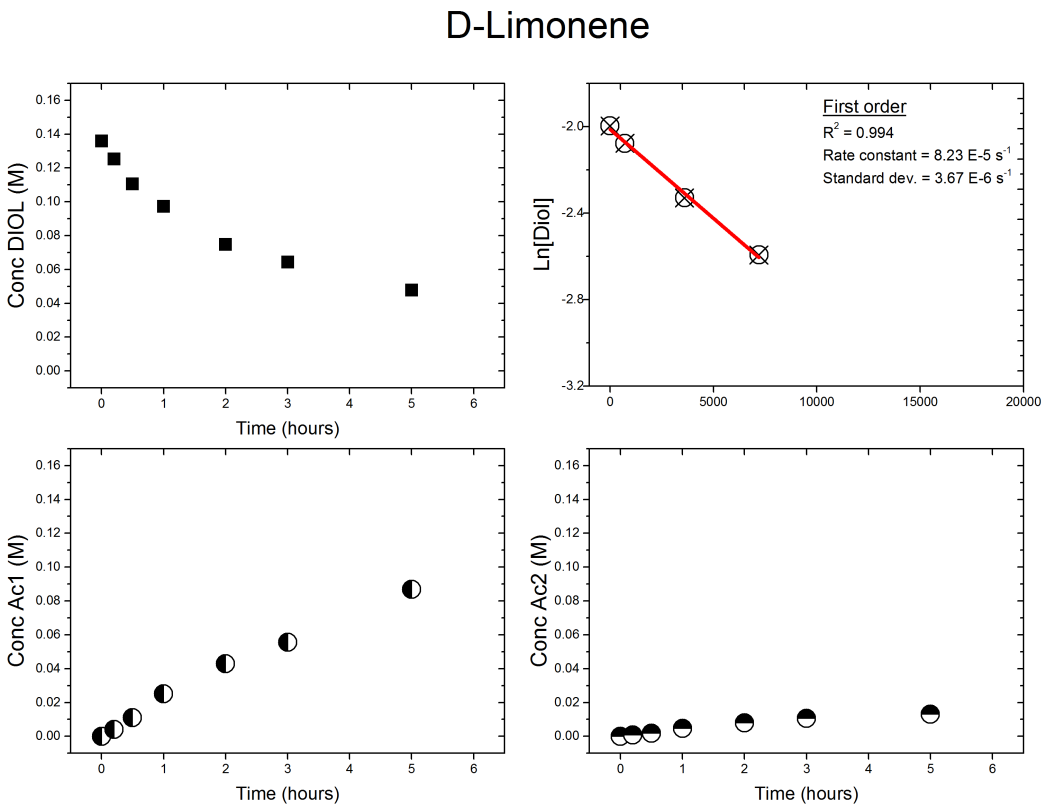


Figure A.4: Reaction profiles in limonene.

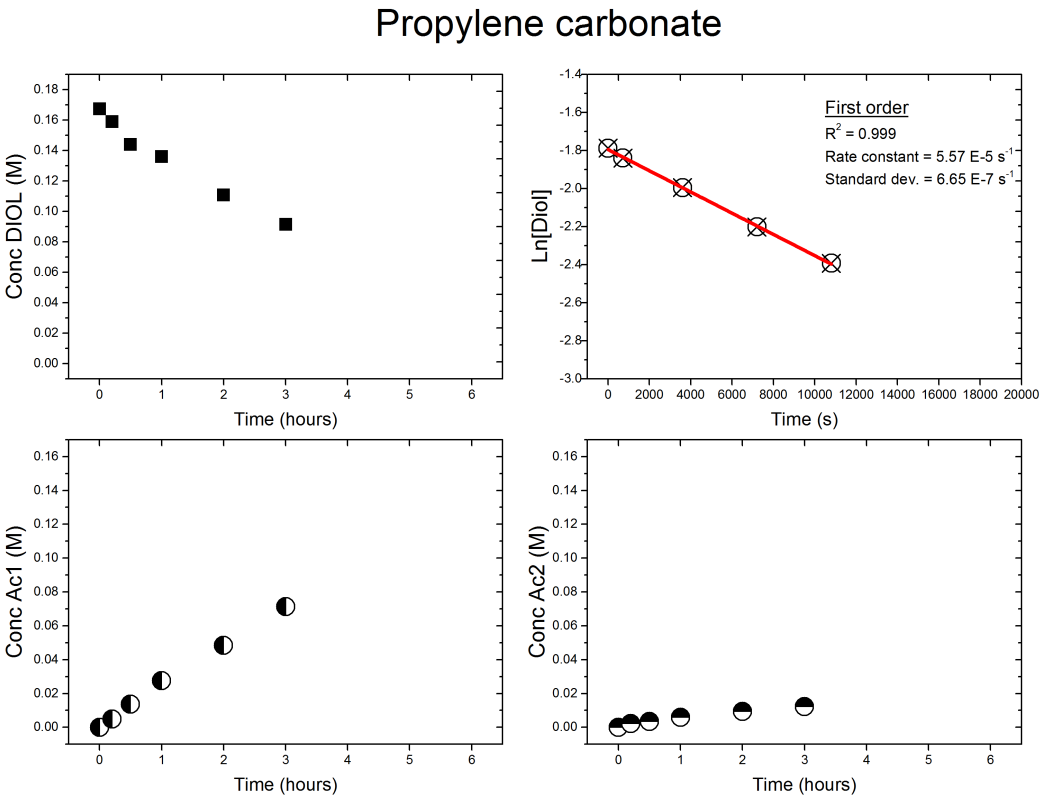


Figure A.5: Reaction profiles in propylene carbonate.

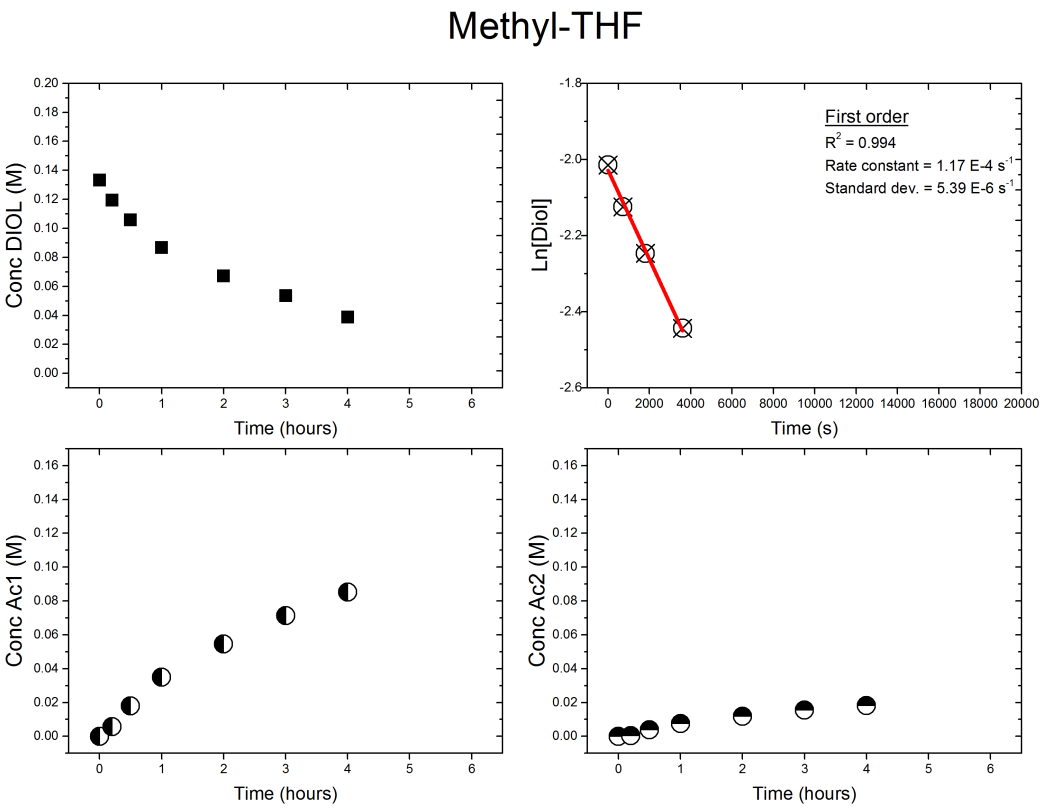


Figure A.6: Reaction profiles in 2-methyltetrahydrofuran.

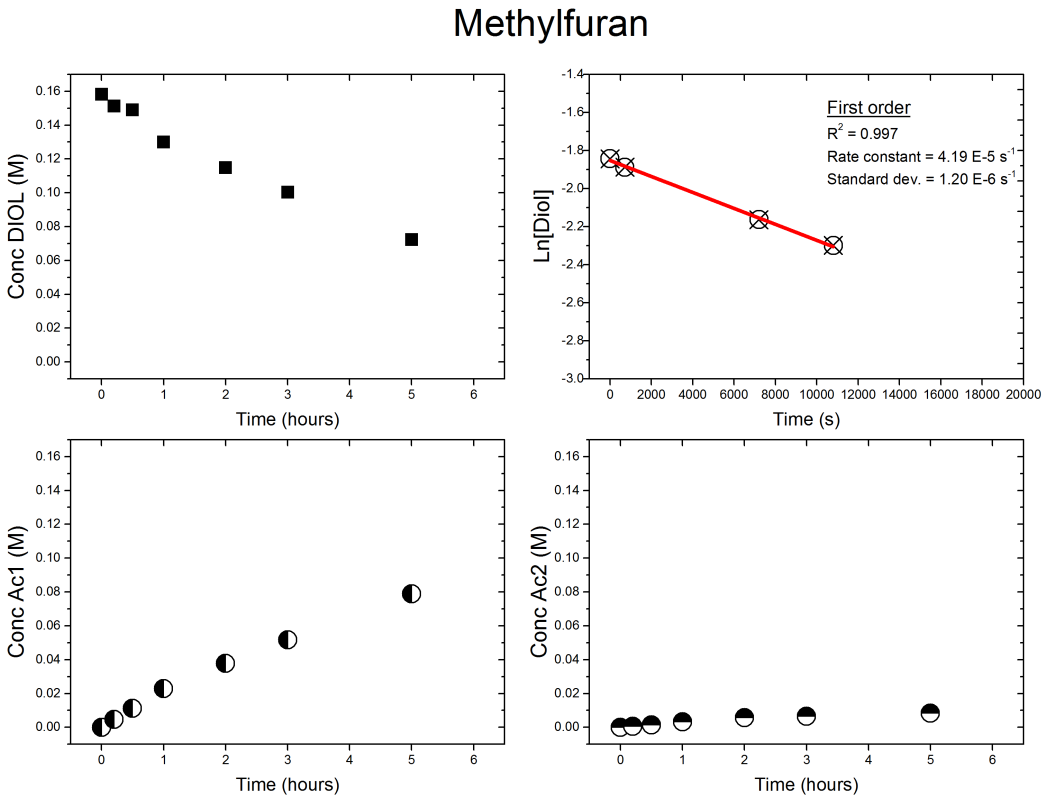


Figure A.7: Reaction profiles in 2-methylfuran.

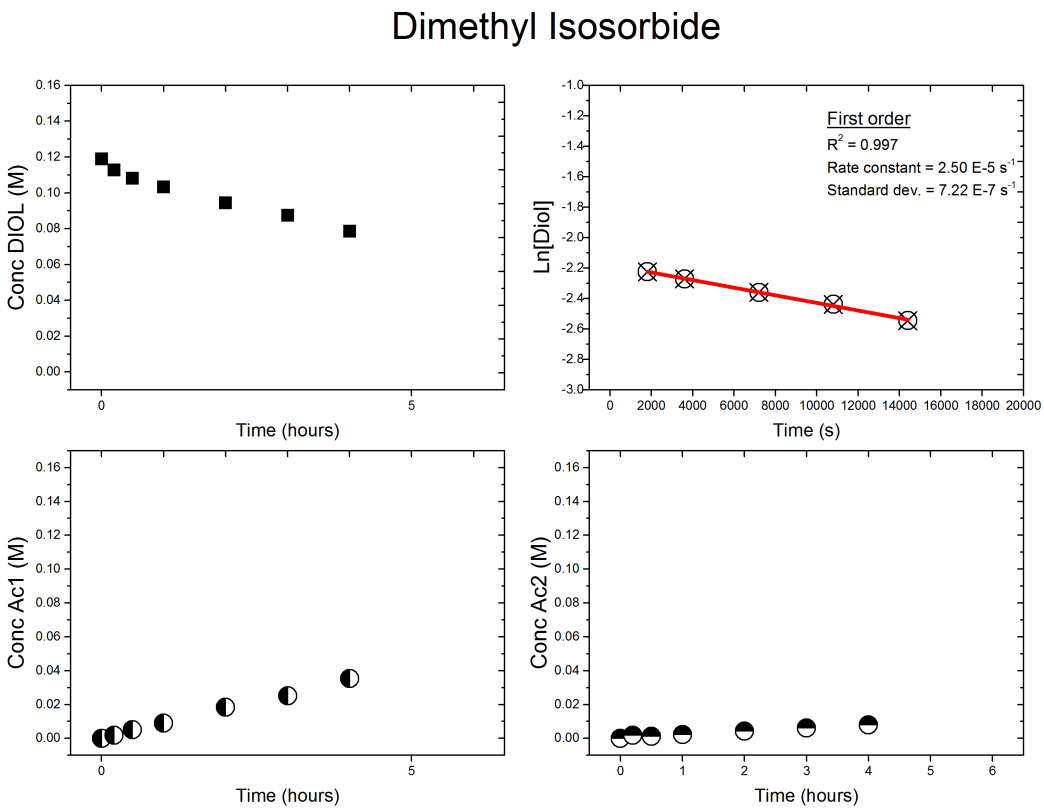


Figure A.8: Reaction profiles in dimethylisosorbide.

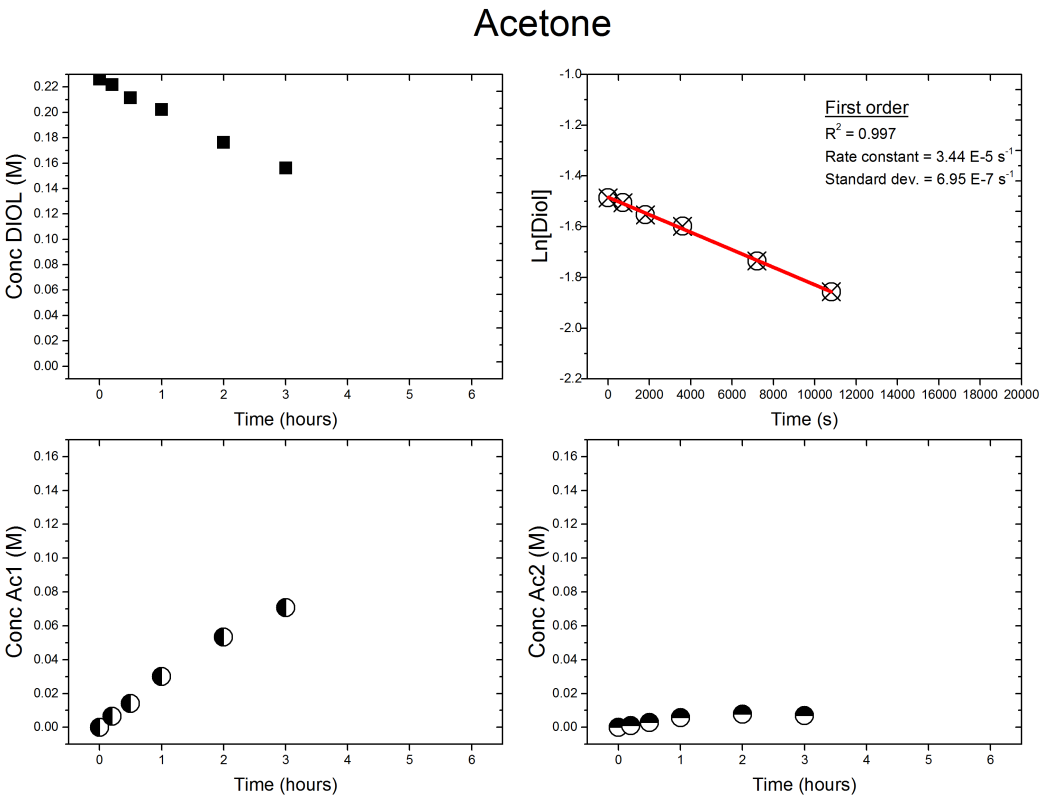


Figure A.9: Reaction profiles in acetone.

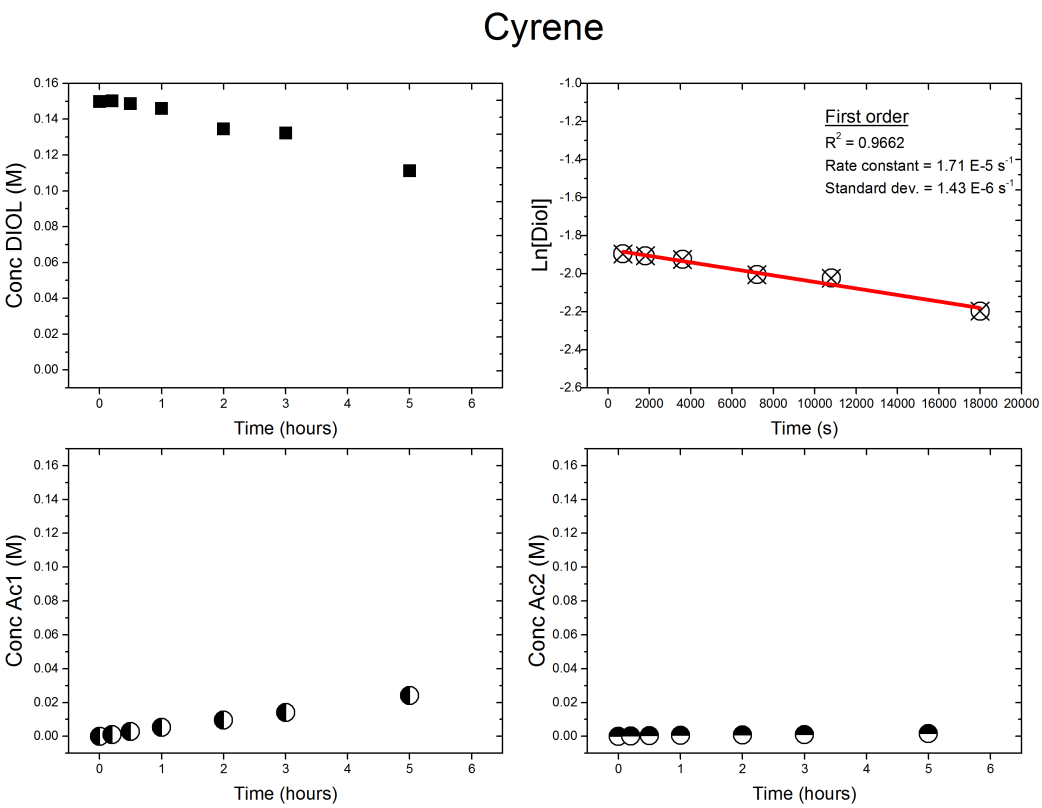


Figure A.10: Reaction profiles in cyrene.

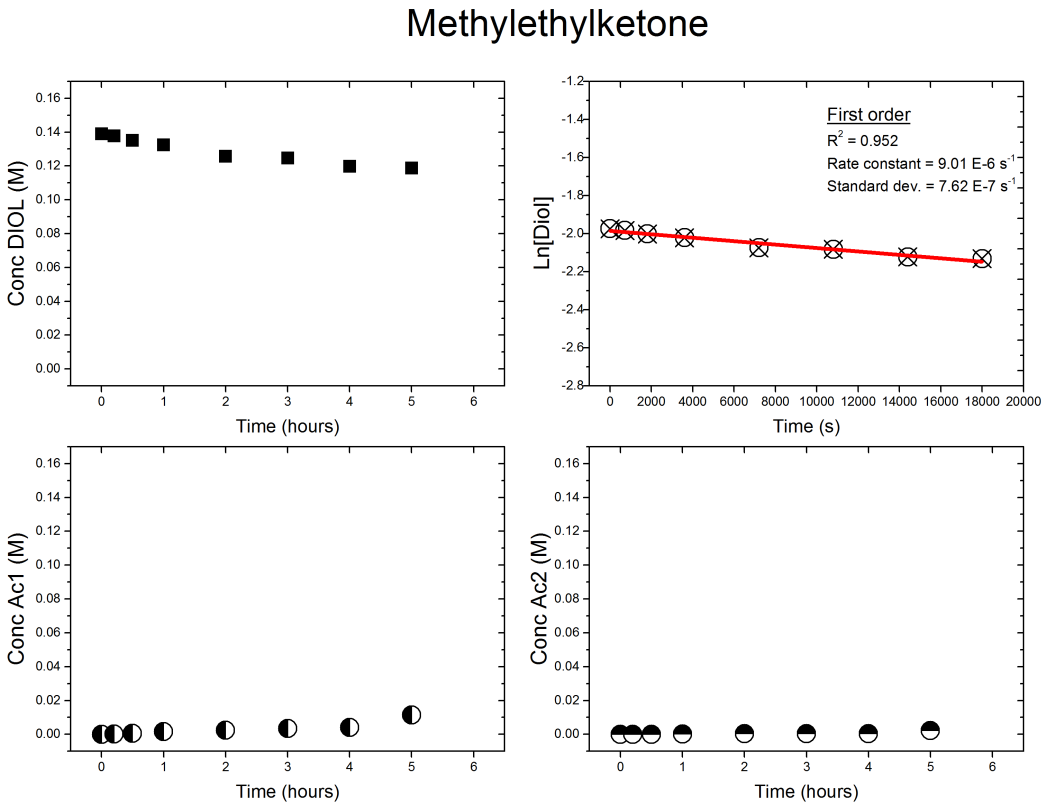


Figure A.11: Reaction profiles in MEK.

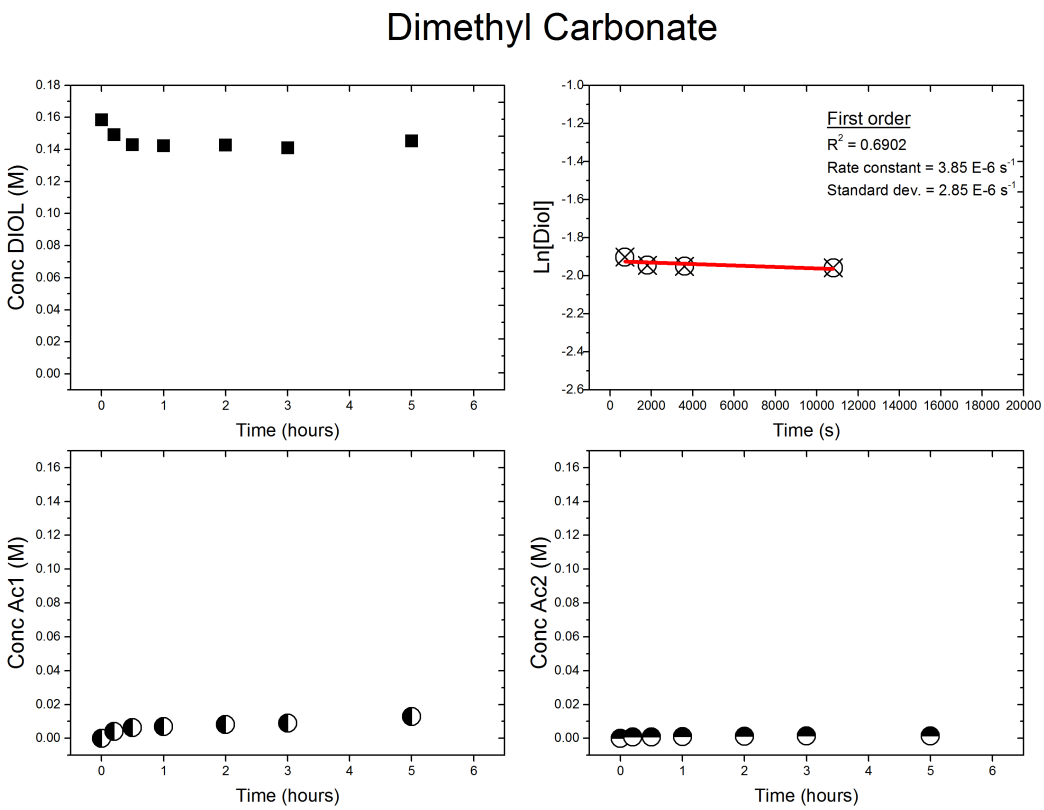


Figure A.12: Reaction profiles in dimethylcarbonate.

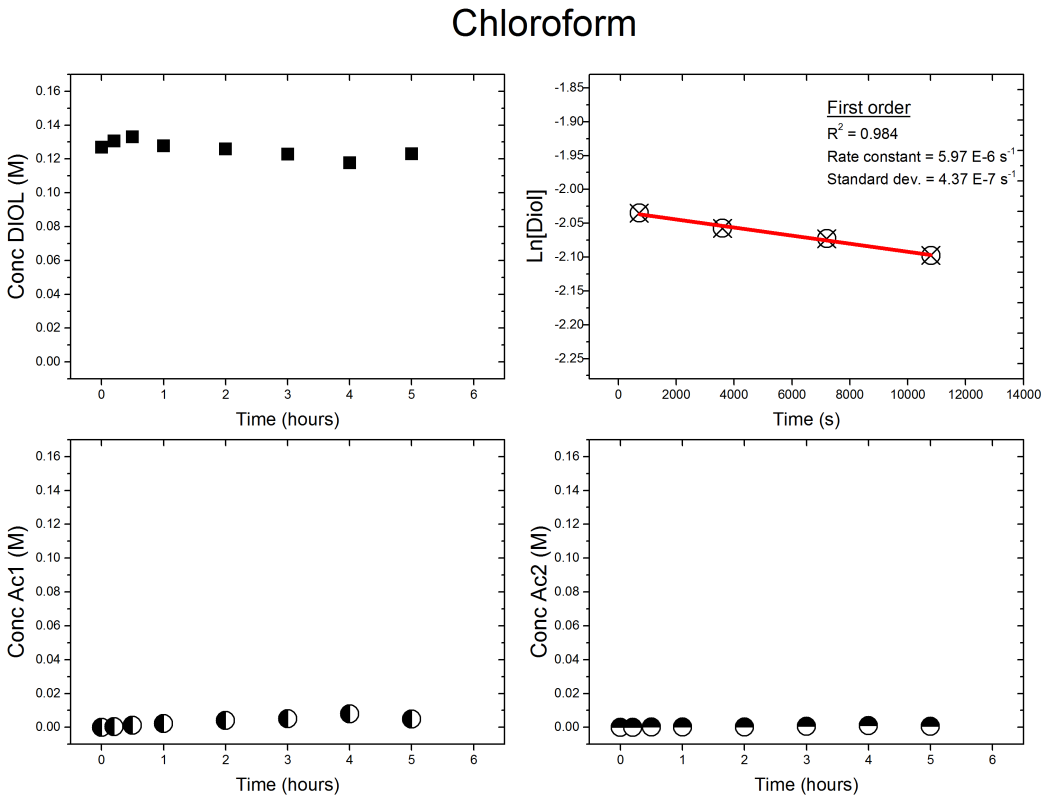


Figure A.13: Reaction profiles in chloroform.

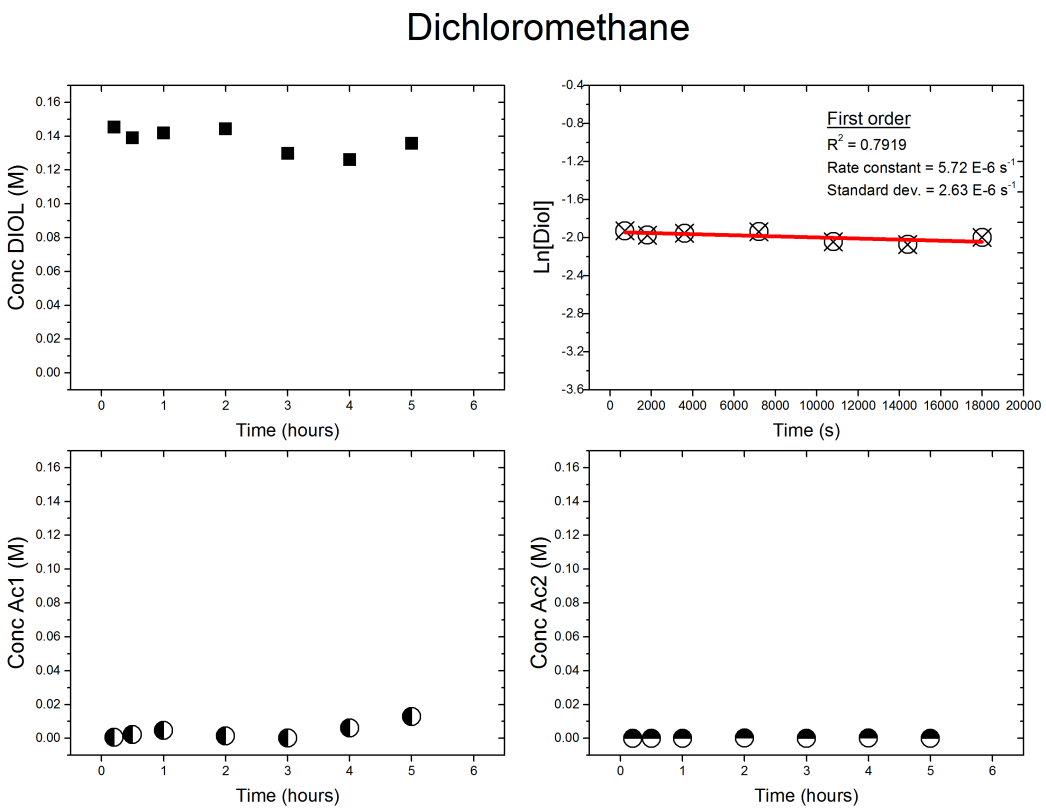


Figure A.14: Reaction profiles in dichloromethane.

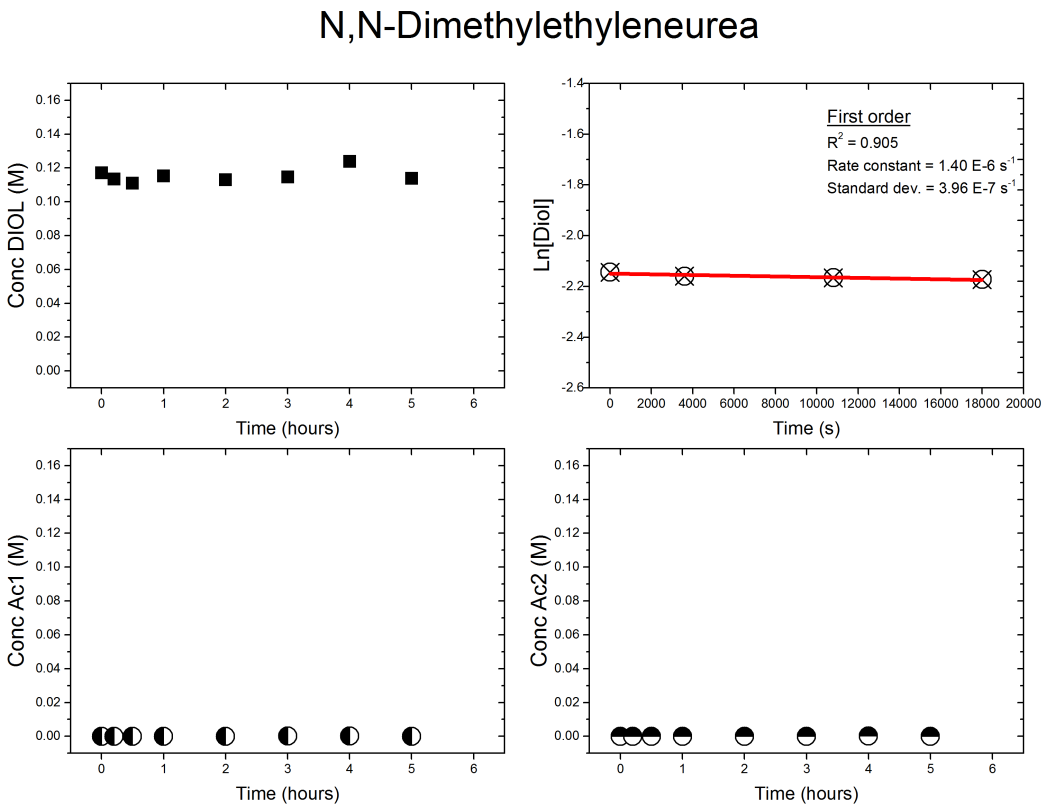


Figure A.15: Reaction profiles in DMEU.

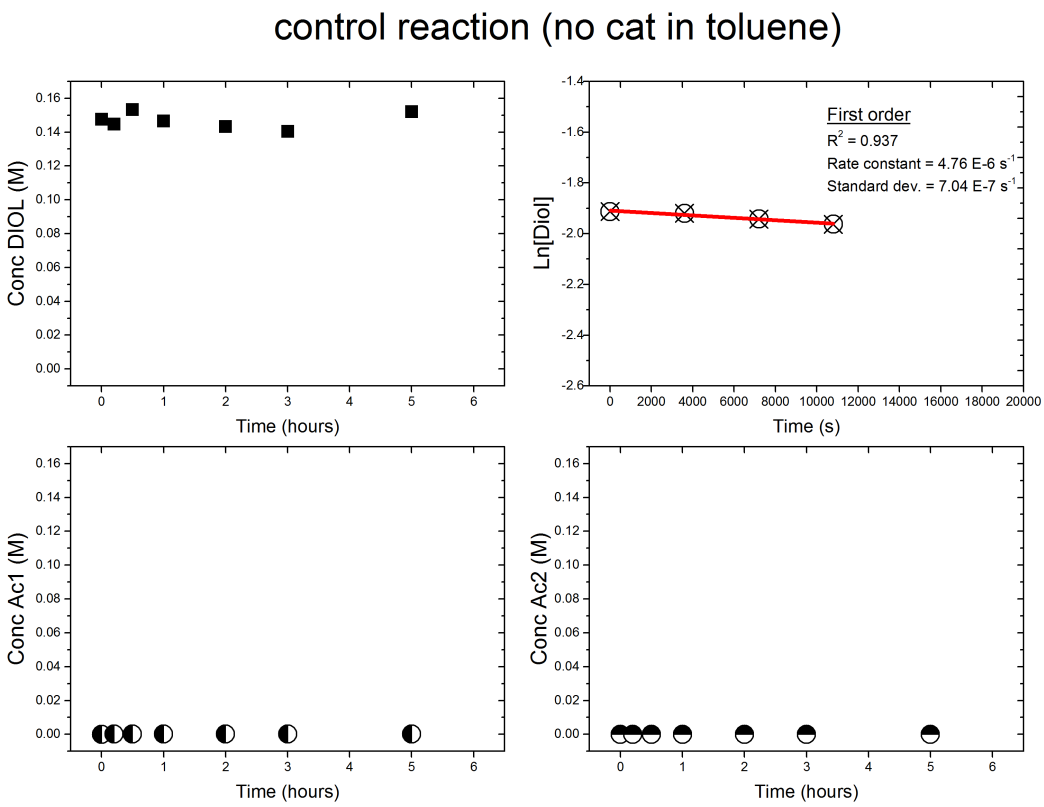


Figure A.16: Control reaction with no catalyst in toluene.

A.1.2 Normalised curves

The normalised curve of the reaction in chloroform and DMEU appear highly affected by the relative error associated with the low conversions.

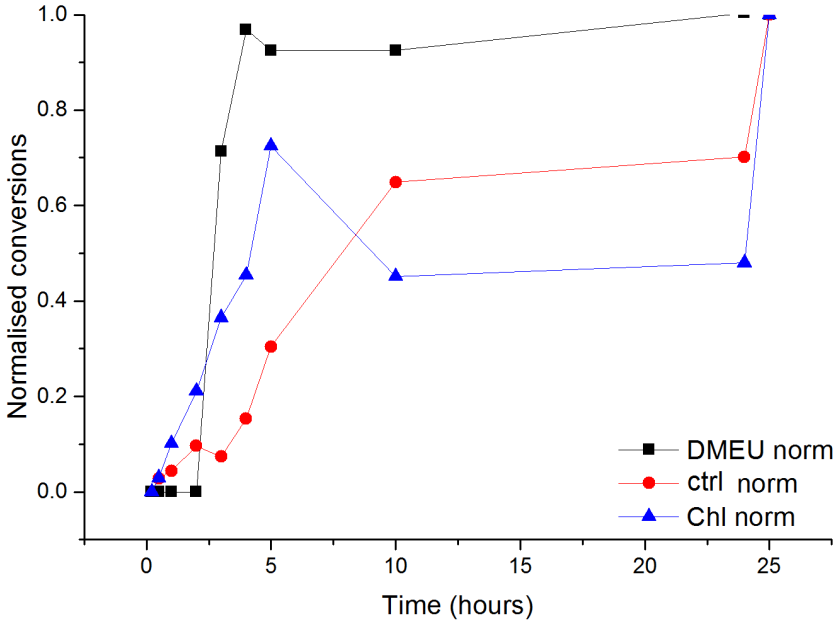


Figure A.17: Normalised curves of remaining low performing reactions.

A.1.3 Reproducibility tests

Reactions in three solvents were repeated three times in order to provide proof-of-concept reproducibility as well as an estimate of typical error bars associated with the data.

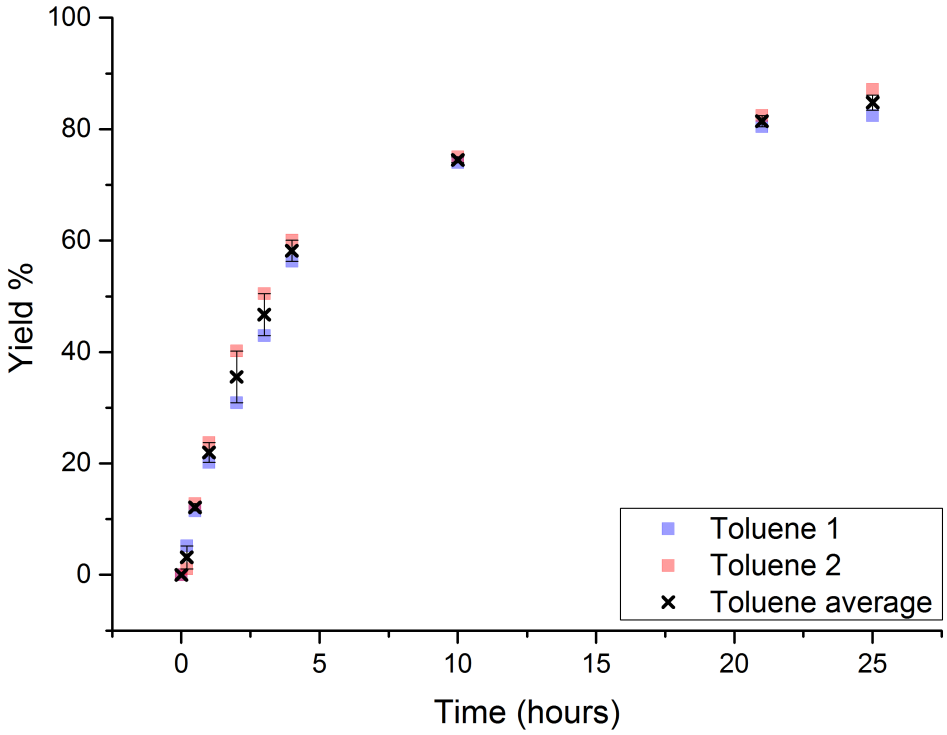


Figure A.18: Profiles of repeated reactions in toluene.

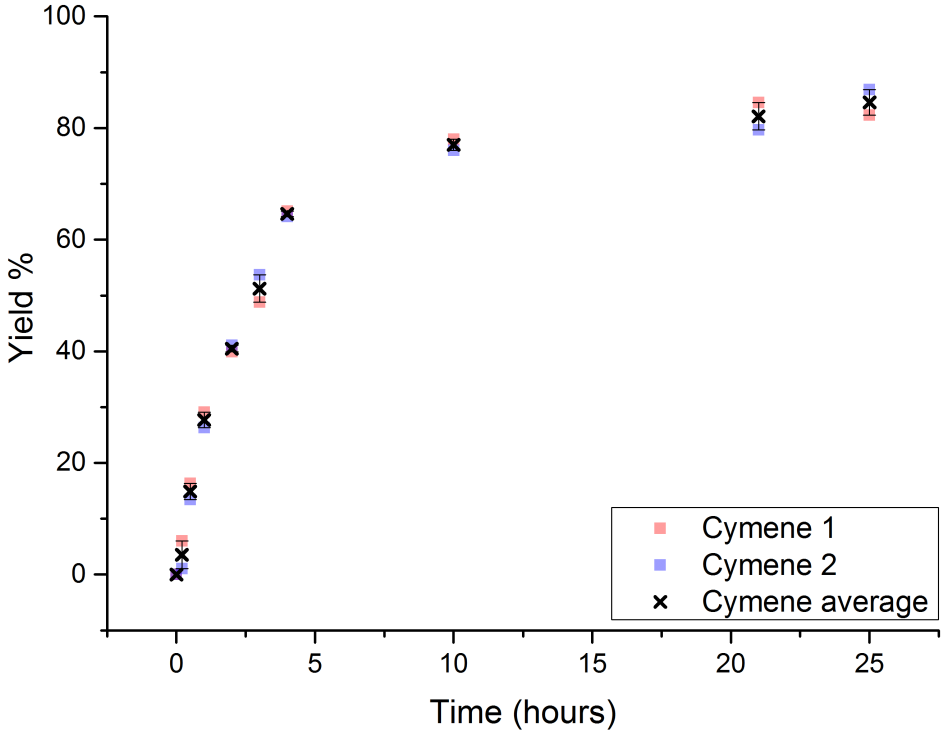


Figure A.19: Profiles of repeated reactions in *p*-cymene.

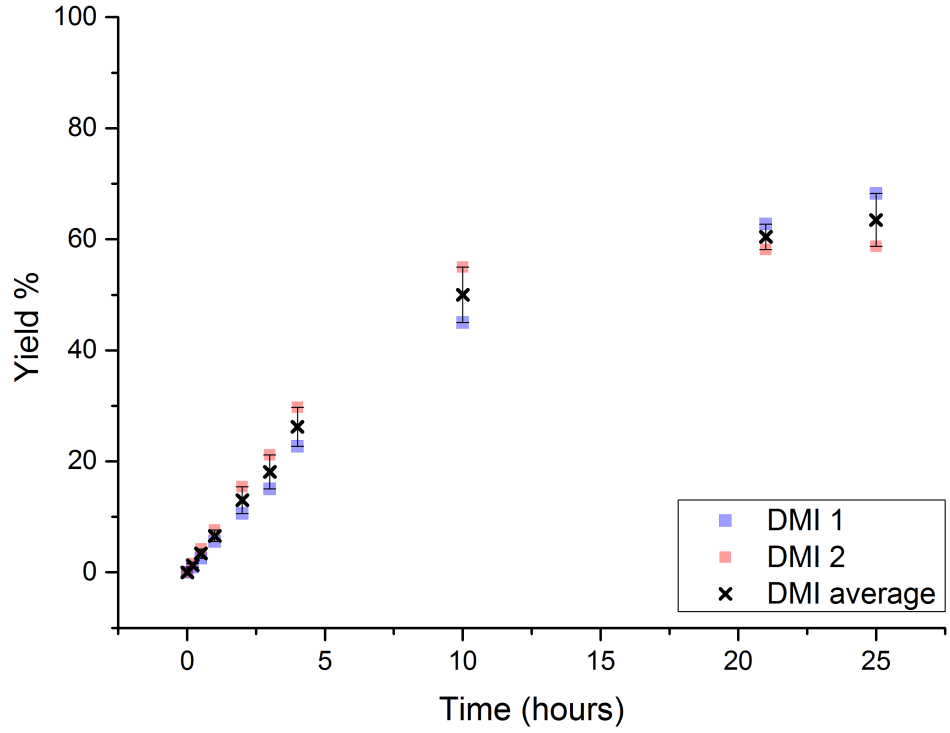


Figure A.20: Profiles of repeated reactions in dimethylisosorbide

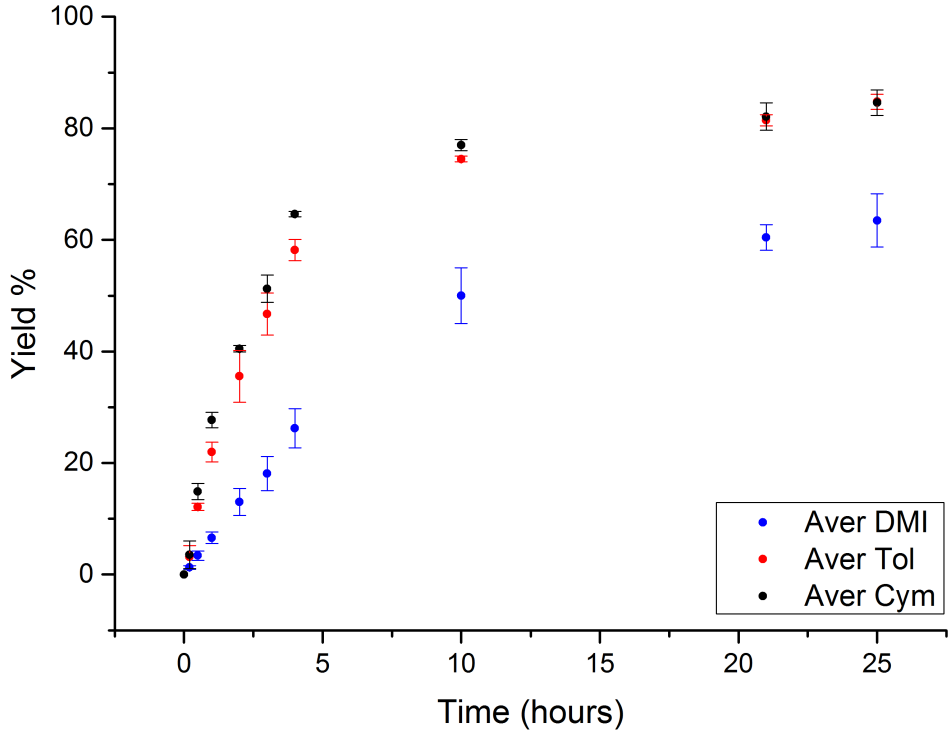


Figure A.21: Overlap of average profiles in the three solvents and corresponding error bars derived from repeated experiments.

A.1.4 Regression statitics

Multivariate regression analysis was performed using OriginLab software. The processing of the data yielded the following reports.

a)

Parameters

		Value	Standard Error
M	Intercept	-2.97968	1.5116
	Slope	0.71751	0.14237

Some input data points are missing.

Statistics

	M
Number of Points	12
Degrees of Freedom	10
Residual Sum of Squares	3.57865
Pearson's r	0.84706
Adj. R-Square	0.78926

Summary

	Intercept		Slope		Statistics
	Value	Standard Error	Value	Standard Error	Adj. R-Square
M	-2.97968	1.5116	0.71751	0.14237	0.78926

ANOVA

		DF	Sum of Squares	Mean Square	F Value	Prob>F
M	Model	1	9.08965	9.08965	25.39966	5.06771E-4
	Error	10	3.57865	0.35787		
	Total	11	12.6683			

b)

Parameters

		Value	Standard Error
O	Intercept	-1.10682	0.88284
	Slope	0.89193	0.08961

Statistics

	O
Number of Points	9
Degrees of Freedom	7
Residual Sum of Squares	0.21894
Pearson's r	0.96644
Adj. R-Square	0.92458

Summary

	Intercept		Slope		Statistics
	Value	Standard Error	Value	Standard Error	Adj. R-Square
O	-1.10682	0.88284	0.89193	0.08961	0.92458

ANOVA

		DF	Sum of Squares	Mean Square	F Value	Prob>F
O	Model	1	3.09865	3.09865	99.07246	2.20614E-5
	Error	7	0.21894	0.03128		
	Total	8	3.31759			

O: At the 0.05 level, the slope is significantly different from zero.

Figure A.22: Statistical outcome of linear fit and ANOVA analysis on a) 12 solvents (all), and b) 9 solvents (best).

A.2 Thermodynamic study

A.2.1 Temperature profiles

Temperature screenings in different solvents were monitored via GC. The data presented below shows the linearised form of the datasets from which the individual rate constants were derived. These show all reactions following a first order kinetics, since a straight line is observed when plotting the natural logarithm of the concentration of starting material against time.

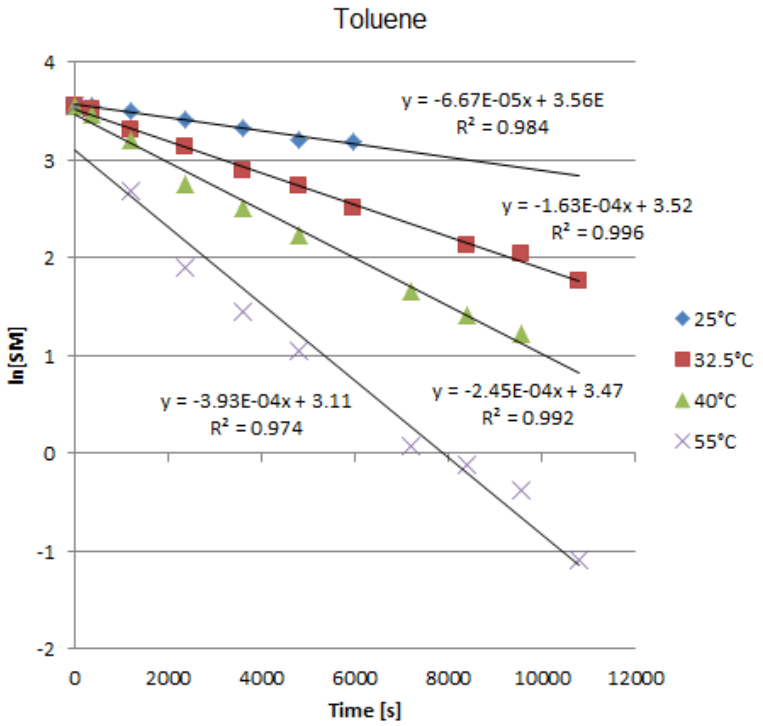


Figure A.23: Linearised data from conversion of Reboxetine starting material in toluene at different temperatures.

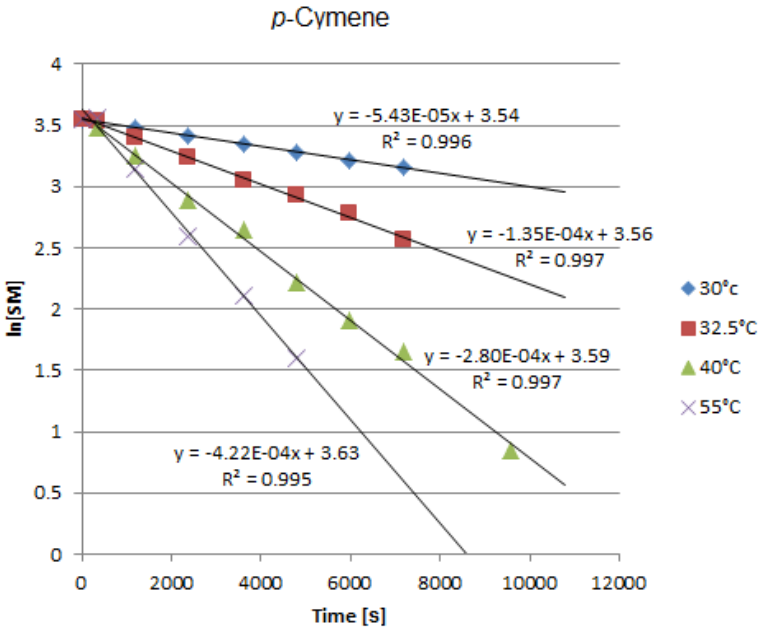


Figure A.24: Linearised data from conversion of Reboxetine starting material in *p*-cymene at different temperatures.

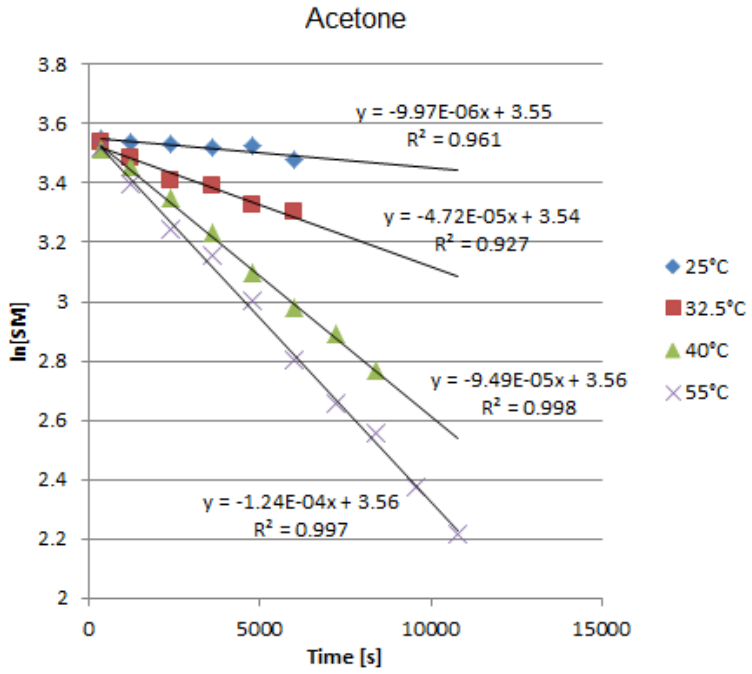


Figure A.25: Linearised data from conversion of Reboxetine starting material in acetone at different temperatures.

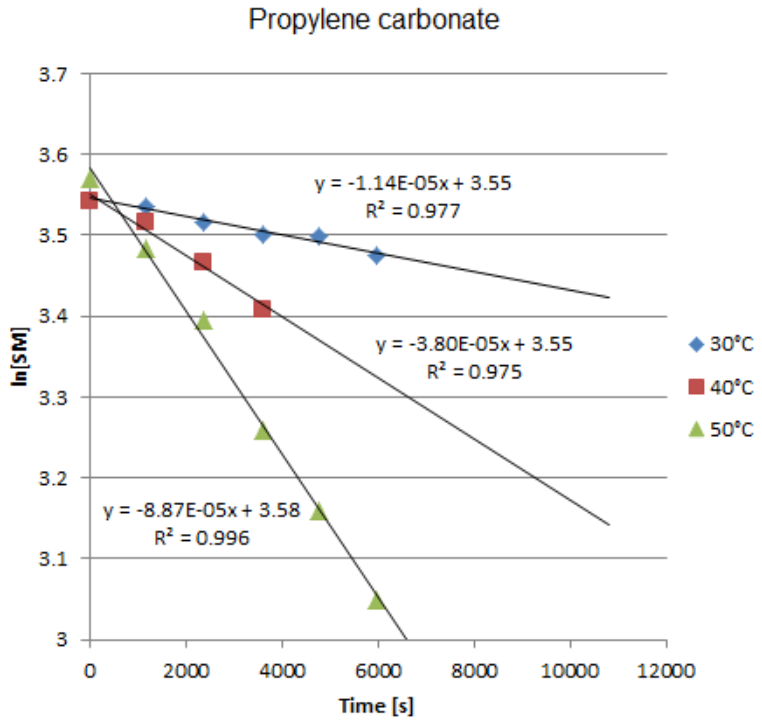


Figure A.26: Linearised data from conversion of Reboxetine starting material in propylene carbonate at different temperatures.

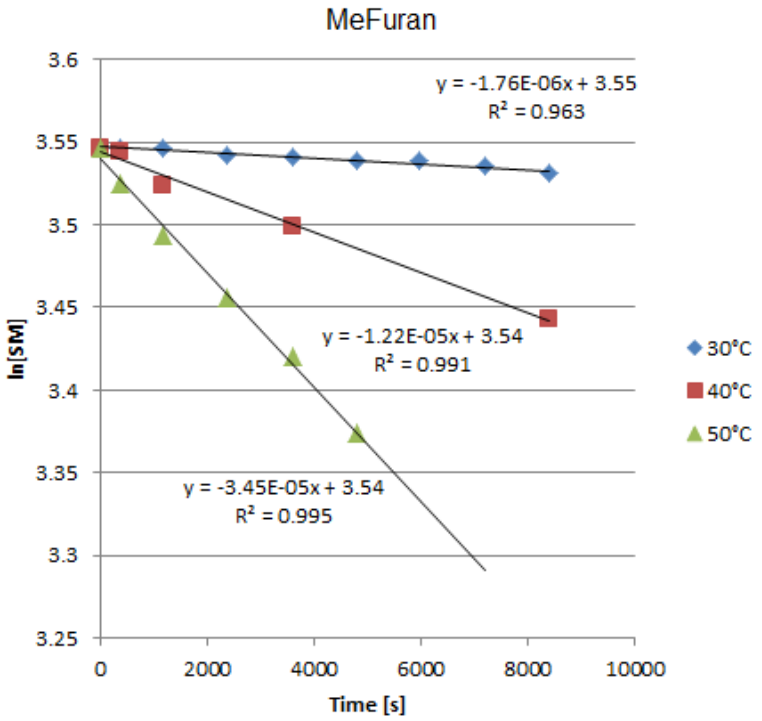


Figure A.27: Linearised data from conversion of Reboxetine starting material in 2-methylfuran at different temperatures.

A.3 Validation with model substrate GGE

The thermodynamic study was repeated on a model substrate (guaiacol glyceryl ether - GGE) in order to provide validation of the results. All reactions were monitored via GC.

A.3.1 GGE Temperature profiles

Acetylation of GGE was screened in four solvents at three temperatures, as presented in figures below.

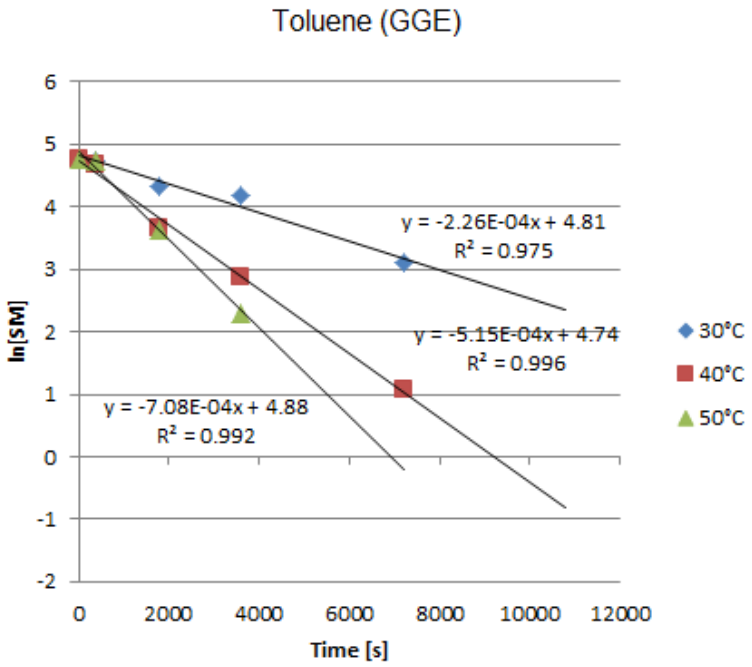


Figure A.28: Linearised data from conversion of GGE starting material in toluene at different temperatures.

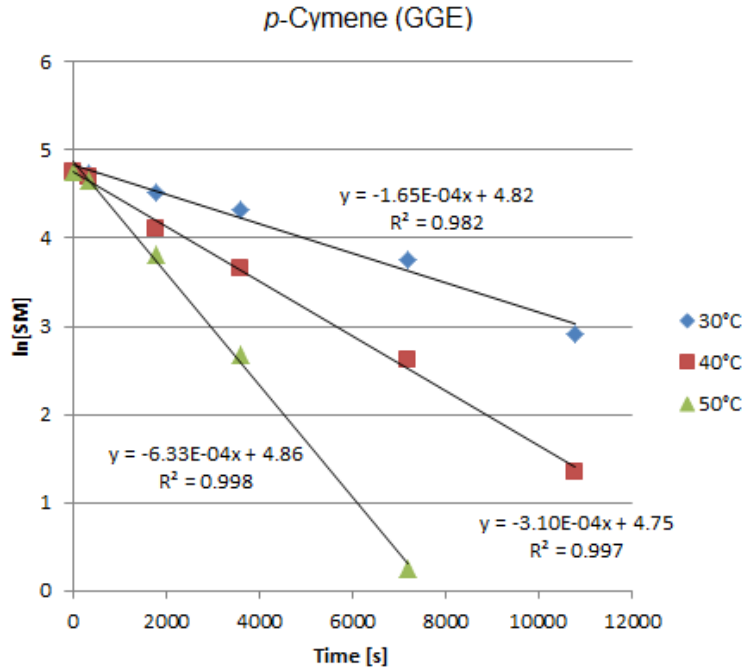


Figure A.29: Linearised data from conversion of GGE starting material in *p*-cymene at different temperatures.

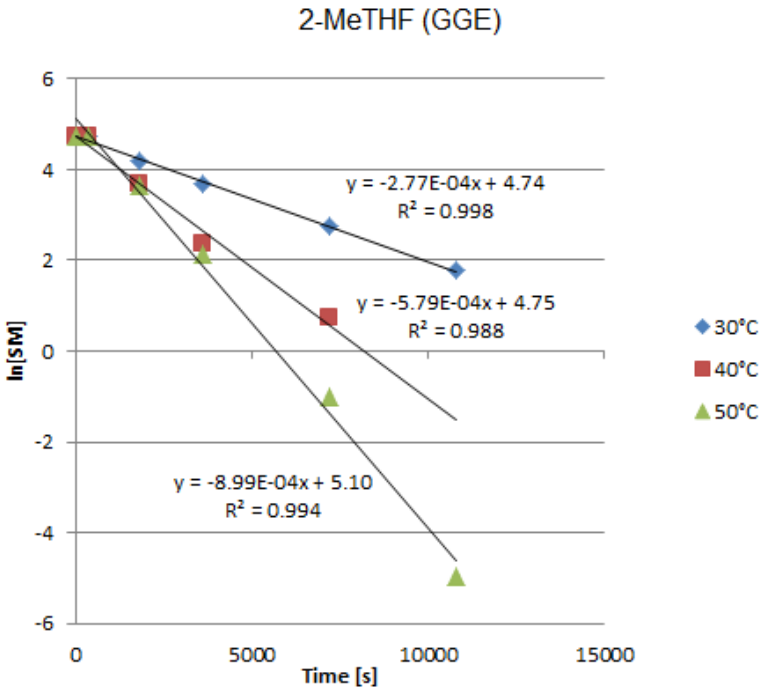


Figure A.30: Linearised data from conversion of GGE starting material in 2-methylTHF at different temperatures.

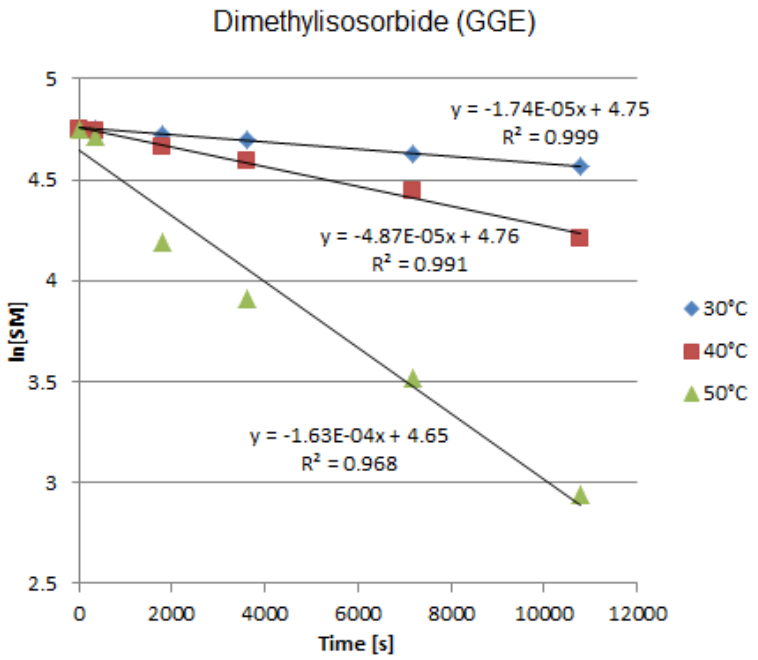


Figure A.31: Linearised data from conversion of GGE starting material in dimethylisosorbide at different temperatures.

Appendix B

Table B.1: Total citrus juice global production from FAOSTAT archives.

TOTAL JUICE PRODUCTION – as single strength equivalent						
	Oranges	Lemons & Limes	Mandarins & Tangerines	Grapefruits	Other citrus	TOTAL JUICE
Year	Volume (MT)	Volume (MT)	Volume (MT)	Volume (MT)	Volume (MT)	Volume (MT)
1990	8761897	454289	6786	1007283	297449	15866306
1991	9763643	482430	5632	998823	378140	16922429
1992	9890055	222847	4530	946384	363510	16443160
1993	12741211	216460	5387	1298011	410790	21551315
1994	12349946	250041	6272	1195457	493027	20630663
1995	11899941	549350	3456	1436563	538657	22041751
1996	12576260	662627	2842	1184702	584283	21289632
1997	13711904	675697	2530	1282357	625583	23094559
1998	14632861	559226	1730	1182480	614276	23257713
1999	12066298	628972	1781	1232294	740066	21200566
2000	13864539	705018	1781	1502860	588417	24627776
2001	14264817	540821	1781	1268935	627505	23429215
2002	12607071	1225670	1781	1243681	523861	22193573
2003	12192268	1138979	1781	1135264	536795	21021982
2004	13633843	1193597	1781	1240663	556762	23202161
2005	11202715	1331426	1781	562931	617324	16699710
2006	10734227	1304756	1781	802076	661572	17755417
2007	11143398	1273465	3887	951993	720656	19138964
2008	10142785	1259665	2381	954981	761289	18182496
2009	9298230	1203339	2381	756539	757948	16028093
2010	8643460	1413032	2381	862987	809680	16305374
2011	8933353	1443832	2381	755495	797206	15936391
2012	10429766	1456219	2381	760647	685292	17365737
2013	11127735	1462248	2381	756552	732011	18090653

Table B.2: Total citrus fruits global production per fruit type.

TOTAL FRUITS PRODUCTION						
	Oranges	Lemons & Limes	Mandarins & Tangerines	Grapefruits	Other citrus	TOTAL FRUITS
Year	Volume (MT)	Volume (MT)	Volume (MT)	Volume (MT)	Volume (MT)	Volume (MT)
1990	49705740	7251262	12542391	4148893	3882655	77530941
1991	51973010	7846120	13655742	4375410	4350198	82200480
1992	54076791	7979127	14030067	4360623	4615021	85061629
1993	55515310	8459921	14670505	4973971	4972802	88592509
1994	54759216	8478629	15108172	4845491	4782923	87974431
1995	58475619	8669010	16153733	5309865	4855444	93463671
1996	60817105	9166550	15813034	5332677	5055308	96184674
1997	65706138	9785854	19101121	5541945	5402128	105537186
1998	61702122	9825572	17656822	5231460	5627210	100043186
1999	61948743	1.06E+07	20310760	5275154	6704717	104822623
2000	63833109	1.14E+07	18366494	5774457	6577453	105942285
2001	60127384	1.21E+07	20902286	5456776	6883681	105517851
2002	62126596	1.23E+07	21282465	5364097	7029496	108099877
2003	59808627	1.26E+07	21982370	5301015	7324556	107048620
2004	65029785	1.23E+07	23413821	5447487	7815080	114005972
2005	63200633	1.23E+07	23985646	4543809	8138216	112206114
2006	66148896	1.37E+07	26130914	5036076	7420340	118402061
2007	65702187	1.63E+07	20327176	7149605	9905522	119336436
2008	69724718	1.72E+07	21521445	7480796	10560020	126446293
2009	67995342	1.73E+07	22497667	7483671	11942276	127170587
2010	69461798	1.50E+07	24073926	7533359	12300883	128373876
2011	71241218	1.52E+07	27473504	7906138	10908826	132696520
2012	68817475	1.51E+07	27983469	8083622	12059413	131998376
2013	71445353	1.52E+07	28678214	8453446	11992686	135761181

Table B.3: Predicted limonene volumes up to 2030 based on extrapolated citrus trends.

	ESTIMATED LIMONENE FROM JUICING			
	FROM JUICING		FROM ANY USE	
	Orange	All citrus	Orange	All citrus
Year	Volume (MT)	Volume (MT)	Volume (MT)	Volume (MT)
2014	77590	130184	270631	517103
2015	76799	128583	273729	526181
2016	76008	126982	276828	535260
2017	75216	125382	279927	544339
2018	74425	123780	283025	553418
2019	73634	122180	286124	562496
2020	72843	120579	289223	571575
2021	72052	118978	292321	580654
2022	71261	117377	295420	589733
2023	70470	115776	298518	598811
2024	69679	114175	301617	607890
2025	68888	112575	304716	616969
2026	68097	110974	307814	626048
2027	67306	109373	310913	635126
2028	66514	107772	314012	644205
2029	65723	106171	317110	653288
2030	64932	104571	320209	662363

Table B.4: Breakdown of data used for the case studies on USA and State of Florida, USA. Some values were converted from 'boxes' to volume metric tonnes equivalents according to conversion factors suggested in the reference document. All volumes expressed as of fruits used in juice production (JUICE) or total independent of use (FRUITS), expressed in MT.

USA	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	4,312,000	6,132,000	506,000	143,000	5,769,000	16,170	21,633	
Grapefruit	451,000	950,000	147,000	18,000	821,000	1,691	3,0788	
Lemon&Lime	169,000	748,000	127,000	481,000	1,102,000	634	4,133	
Tangerine	170,000	702,000	na	183,000	885,000	637	3,319	
TOTAL	5,102,000	9,400,000	780,000	825,000	8,577,000	19,133	32,164	119,730

Florida, USA	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	4,209,004	4,707,000	na	na	4,707,000	15,784	17,651	
Grapefruit	380,686	665,000	na	na	284,000	1,428	1,066	
Lemon&Lime	20,651	37,000	na	na	17,000	77	63	
Tangerine	40,580	123,000	na	na	83,000	152	310	
TOTAL	4,650,921	5,550,000	na	na	5,091,000	17,441	19,090	5,150

Table B.5: Breakdown of data used for the case study on the European Union (EU27) and Germany. All volumes expressed as of fruits used in juice production (JUICE) or total independent of use (FRUITS), expressed in MT.

EU	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	1,487,000	6,607,000	346,000	818,000	7,079,000	5,576	24,776	
Grapefruit	26,000	112,000	19,000	364,000	457,000	98	420	
Lemon&Lime	312,000	1,316,000	101,000	370,000	1,585,000	1,170	4,935	
Tangerine	309,000	3,245,000	350,000	367,000	3,262,000	1,159	12,169	
TOTAL	2,134,000	11,280,000	816,000	1,919,000	12,383,000	8,000	42,300	68,750

Germany	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	851,240	0	0	558,200	1,409,440	3,192	5,285	
Grapefruit	50,030	0	13,000	68,000	55,000	188	206	
Lemon&Lime	17,918	0	22,300	151,800	129,500	67	486	
Tangerine	0	0	19,400	361,600	342,200	0	1,283	
TOTAL	919,188	0	54,700	1,139,600	1,936,140	3,447	7,261	19,638

Table B.6: Breakdown of data used for the case studies on Spain and South Africa. All volumes expressed as of fruits used in juice production (JUICE) or total independent of use (FRUITS), expressed in MT.

Spain	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	719,320	3,548,000	1,811,000	126,000	1,863,000	2,697	6,986	
Grapefruit	14,554	58,800	na	7,000	65,800	55	247	
Lemon&Lime	93,480	818,000	537,900	56,000	336,100	350	1,260	
Tangerine	na	2,199,000	1,567,000	6,000	638,000	0	2,393	
TOTAL	827,354	6,623,800	3,915,900	195,000	2,902,900	3,103	10,886	3,095

South Africa	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	463,000	1,715,000	1,144,000	12,000	583,000	1,736	6,431	
Grapefruit	203,000	413,000	217,000	12,000	208,000	761	1,549	
Lemon&Lime	80,000	312,000	150,000	na	162,000	300	1,170	
Tangerine	na	153,000	153,000	na	0	na	0	
TOTAL	746,000	2,593,000	1,664,000	24,000	953,000	2,798	9,724	4,709

Table B.7: Breakdown of data used for the case studies on Brazil and India. All volumes expressed as of fruits used in juice production (JUICE) or total independent of use (FRUITS), expressed in MT.

Brazil	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	11,832,000	16,034,400	21,256	15,000	19,053,000	na	71,430	
Grapefruit	na	78,000	na	na	78,000	na	293	
Lemon&Lime	na	2,014,855	na	na	2,014,855	na	7,556	
Tangerine	na	2,619,787	na	na	2,619,787	na	9,824	
TOTAL	13,567,990	23,811,250	na	na	23,811,250	50,879	89,292	22,349

India	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	41,659	1,851,941	na	na	1,851,941	156	7,101	
Grapefruit	11,664	530,208	na	na	530,208	44	1,988	
Lemon&Lime	41,659	1,893,600	na	na	1,893,600	156	7,101	
Tangerine	71,654	3,256,992	na	na	3,256,992	269	12,214	
TOTAL	166,637	7,574,400	na	na	7,574,400	625	28,404	5,940

Table B.8: Breakdown of data used for the case study on China. All volumes expressed as of fruits used in juice production (JUICE) or total independent of use (FRUITS), expressed in MT.

China	FOR JUICE (MT)	TOT FRUITS (MT)				LIMONENE (MT)		TOLUENE (MT)
Type Citrus	Production	Production	Exports	Imports	Net vol.	Juicing	Any use	Cleaning
Orange	715,000	7,600,000	108,000	108,000	7,580,000	2,681	28,425	
Grapefruit	na	3,717,000	165,000	165,000	3,578,000	na	13,418	
Lemon&Lime	8,921,000	1,937,880	na	na	na	33	na	
Tangerine	600,000	17,850,000	16,524,000	744,000	17,106,000	2,250	64,148	
TOTAL	1,323,921	31,110,880	1,017,000	114,000	28,264,000	4,964	105,991	22,349

Appendix C

C.4 Amidation with *Pseudomonas Stutzeri* lipase (PSL)

Biocatalysed *N*-acylations have the potential to bring significant improvements in the current API synthetic practice. At present, most amidations are carried out using acid chlorides as starting materials or coupling reagents such as 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDCI.HCl), 1-propylphosphonic acid cyclic anhydride, and *N,N*-carbonyldiimidazole (CDI). A greenness assessment on these practices found them inefficient from an atom economy perspective and dangerous for human health and the environment.[84] A review published by Montalbetti and Falque (2005) provides an comprehensive overview of the conventional non-catalytic methods for amide bond formation, yet several catalysed methods have also been developed.[222] Some examples are the use of boron-based catalysts, heterogeneous sulfated tungstate catalyst and activated heterogeneous silica.[223] In 2009, Comerford *et al.* compared the ability of different catalysts to perform *N*-acylations, providing figures for E factor, atom economy, mass intensity and yield associated with the use of different catalysts or coupling agents: thionile chloride, boron-based catalyst, DCC and silica.[224] Biocatalysed amidation was not included in the study.

Van Pelt *et al.* compared two commercially available highly active enzymes for amide formation catalysis: *Candida Antarctica* lipase type B (CALB) and *Pseudomonas Stutzeri* lipase (PSL). They screened them against a range of benzyl and benzylic esters and with primary and secondary amines as summarised in Figure C.32. Strikingly, they observed amides were formed even with very electron-poor amines such as aniline, a difficult substrate to react by conventional chemical means. In particular, PSL showed the ability to well accept secondary amines such as *N,N*-methylbenzylamine and piperidine, while CALB outperformed PSL when more substituted α-methyl esters were employed (Figure C.32).

Moreover, another recent study by Maraite *et al.* has approached the complementarity of the two lipases with computer modelling.[225] They focused on benzoin as substrates for acylation reactions (Figure C.33), and revealed remarkable differences in the two catalytic sites and interesting considerations as to the binding mode. PSL, renowned for its

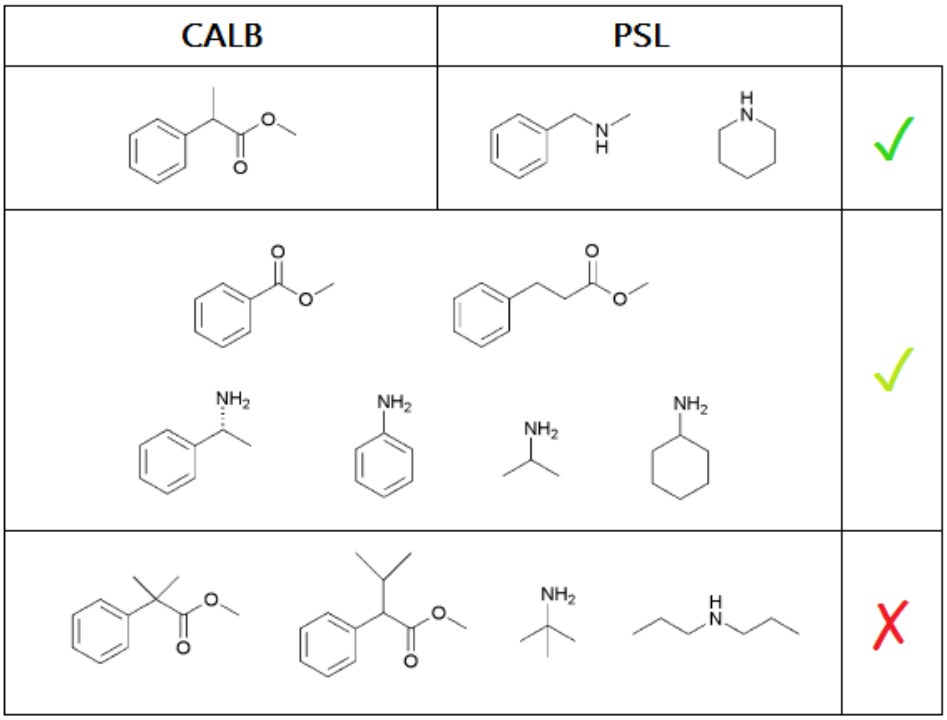


Figure C.32: Substrate scope of CALB and PSL with different esters and amines.[223]

ability to accept particularly bulky substrates, confirmed activity with benzoin substrates whereas CALB was proven inactive. This fact was investigated further using computer modelling, and the conformation of the two active sites was directly compared, as shown in Figure C.33. The active site of PSL, in grey, was found to be larger and therefore able to accommodate bulkier ligands. Instead, the slightly different conformation of CALB causes the residues highlighted in blue to become an obstacle to bulky ligands.

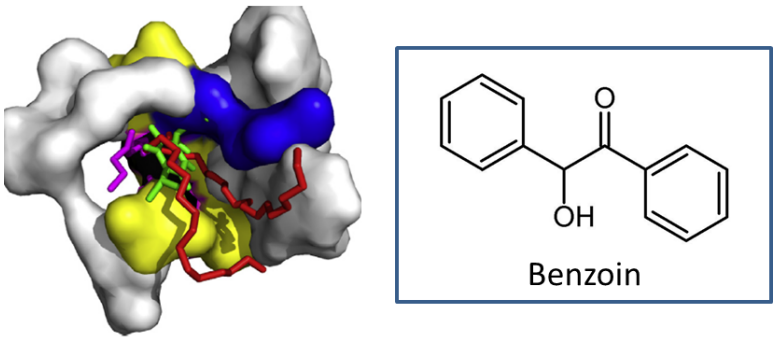


Figure C.33: Comparison between PSL (grey) and CALB (yellow and blue) active site conformation via computer modelling.[225]

Pseudomonas Stutzeri Lipase has so far tested in a few organic solvents and reported to be active in MTBE and in 2-MeTHF.[128,223] According to the procedure published by Van

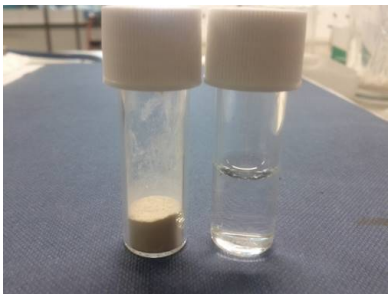


Figure C.34: Catalyst load required for amidation according to the procedure published by Van Pelt *et al.*.[223]

Pelt *et al.* the catalyst load required to obtain 43% yield with PSL is fairly high, 60 mg per ml of reaction mixture (Figure C.34). In this work, a screening was carried out to test *Pseudomonas Stutzeri* Lipase in a larger set of green solvents. The enzyme was obtained from Meito-Sangyo group as Lipase TL (PSL). The tests were initially carried out with eight solvents, adjusting a protocol published by Van Pelt *et al.*. By initially employing 1/3 of this quantity, the following results were obtained for the reaction in Figure C.35. The screening was run in parallel with CALB for comparison.

It is clear from Figure C.35 that the catalyst load employed was not enough for providing any yield above 10%. Thus, no real trend could be identified from the data. Later, further screenings were performed at CatSci utilising the catalyst load stated originally in the paper. In Figure C.36 are reported the reaction profiles in 5 solvents. The reaction were followed by LCMS monitoring the formation of the desired amide as well as of the undesired hydrolytic byproduct, as reported in Figure C.37 for a larger set of solvents.

In conclusion, solvent influence shows to vary significantly across the plots, yet some general consideration can be made: 1) the more non-polar solvents (MTBE, terpinolene, *p*-cymene, limonene) facilitate the aminolysis giving the highest amide to acid ratios; 2) 2-MeTHF and γ-valerolactone show the highest variability in behaviour, but seem to naturally favour the reaction with water. Additional work was carried out and is currently in writing for publication. Overall, the results of this project provide guidance for conscious selection of substrate combinations and solvents, and highlights the need of further investigation on the enzyme's activity towards *para*- or 4-substituted substrates. Tackling the competing hydrolysis reaction will continue to be a priority in PSL chemistry. Morever,

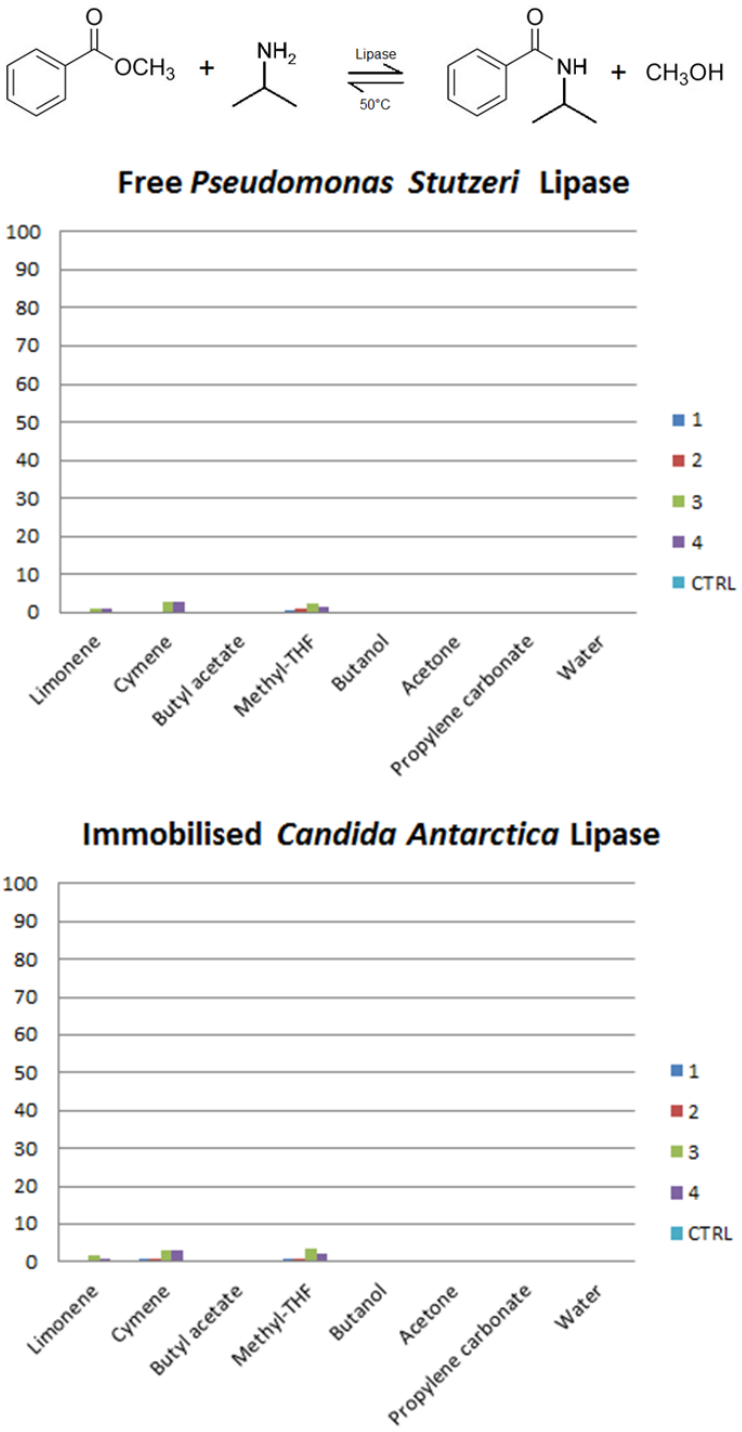


Figure C.35: Results of screening PSL and CALB in amidation reaction.

further protein optimisation, purification and immobilisation developments[226] combined with better understanding of the chemical potential will help in driving the application of this enzyme in fine chemical manufacturing.

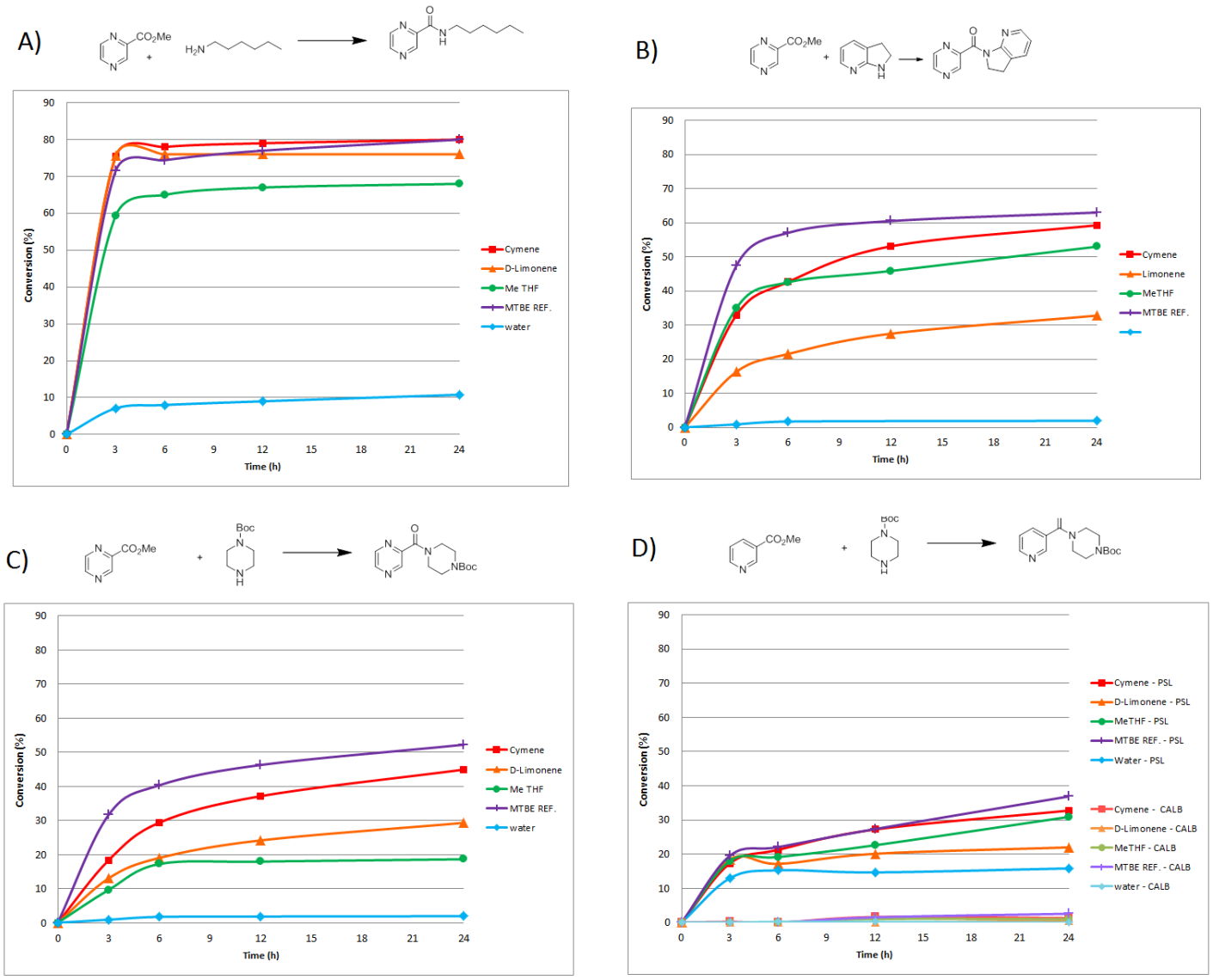


Figure C.36: Yield plots of four reactions in different solvents. In A, B, C the same ester is compared against reaction with different amines. Plot D compares PSL and CALB activity against amine **17**.

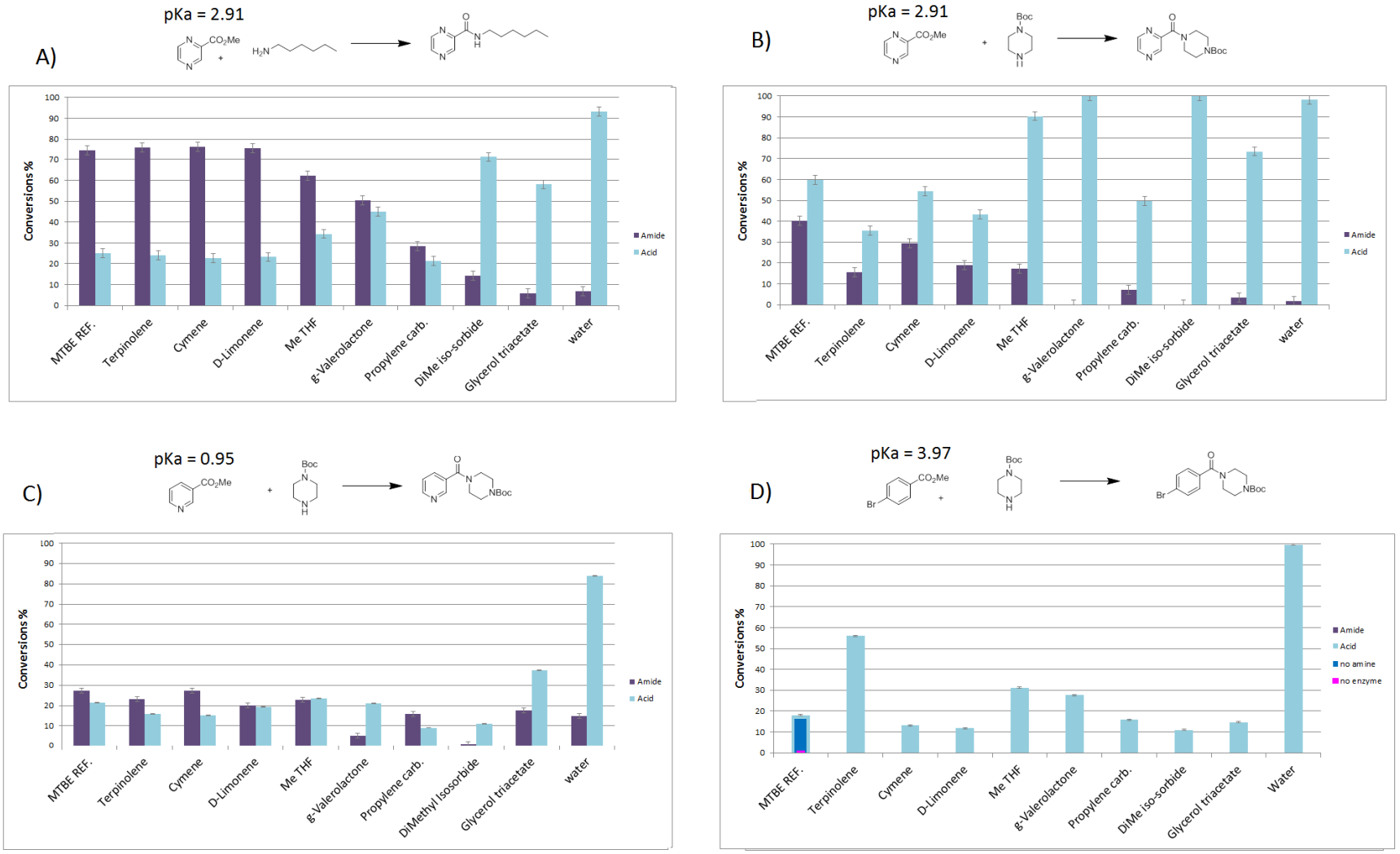


Figure C.37: Yield of aminolysis (purple) *versus* hydrolysis (blue) after 6 hours of reaction. Error bars have been obtained by repetition of the MTBE reference reaction and extended across the dataset.

C.5 De-amination with ω-transaminase ATA-113

Enzymes belonging to the transaminase class have the ability to exchange carbonyl and amino-functionalities. This can be useful for the synthesis of asymmetric amines starting from a carbonyl-containing substrates and also to deracemise an amine racemate.[227] Transaminases operate with the aid of pyridoxal phoshate (PLP) as a cofactor as illustrated in Figure C.38, and in conjunction with an amino-acceptor or amino-donor depending on the synthetic direction preferred.

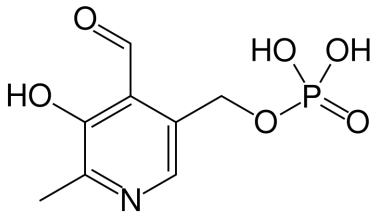


Figure C.38: Pyridoxal phosphate cofactor.

Both directions of the equilibrium reaction represent a greatly useful synthetic tool, especially for the pharmaceutical industry where the need to produce chiral amines, or their derivatives, is pressing.[227] The natural activity of transaminases is towards the formation of carbonyl groups starting from an amine and an amine-acceptor. The opposite route, operating with a ketone and an amine-donor, generally requires a large excess of the amine-donor - *e.g.* alanine - in order to push the equilibrium towards the amine synthesis. Koszelewski *et al.* have explored the effect of a few relatively water-soluble co-solvents in transaminase catalysed reactions, obtaining discouraging outcomes in comparison to the performance in water.[228] We are not aware of any other work investigating the effect of neat organic solvents on the activity of transaminases.

For this study, a transaminase supplied by University of Manchester was tested for the de-amination of (S)-methylbenzylamine to acetophenone in *green* solvents (Figure C.39). The enzyme used is the (S)-selective variant, although no investigation into the chirality of the product was here pursued.

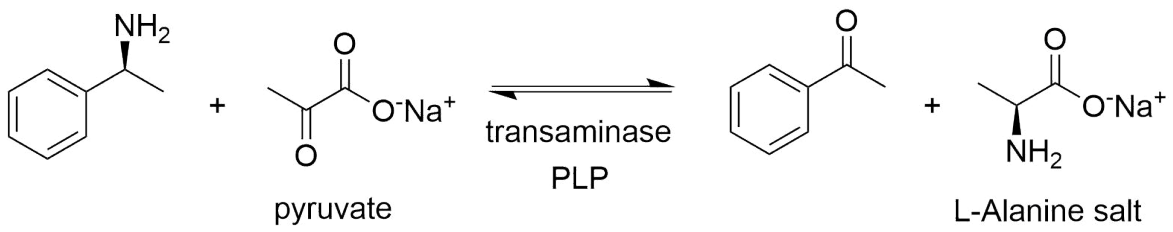


Figure C.39: ω-Transaminase ATA-113 (10 mg) catalysed deamination of (S)-methylbenzylamine (25 mM) to acetophenone, in 1 ml stock solution, 2 mg PLP cofactor.

Sodium pyruvate was used as an amine-acceptor, and added in excess in order to push the equilibrium of the reaction towards the formation of the ketone. The enzyme requires the aid of a prosthetic group to perform the catalysis. The latter needs to be added to the reaction mixture in catalytic quantity, yet proportionate to the mass of enzyme employed. The results of the assay in organic solvents are presented in Figure C.40.

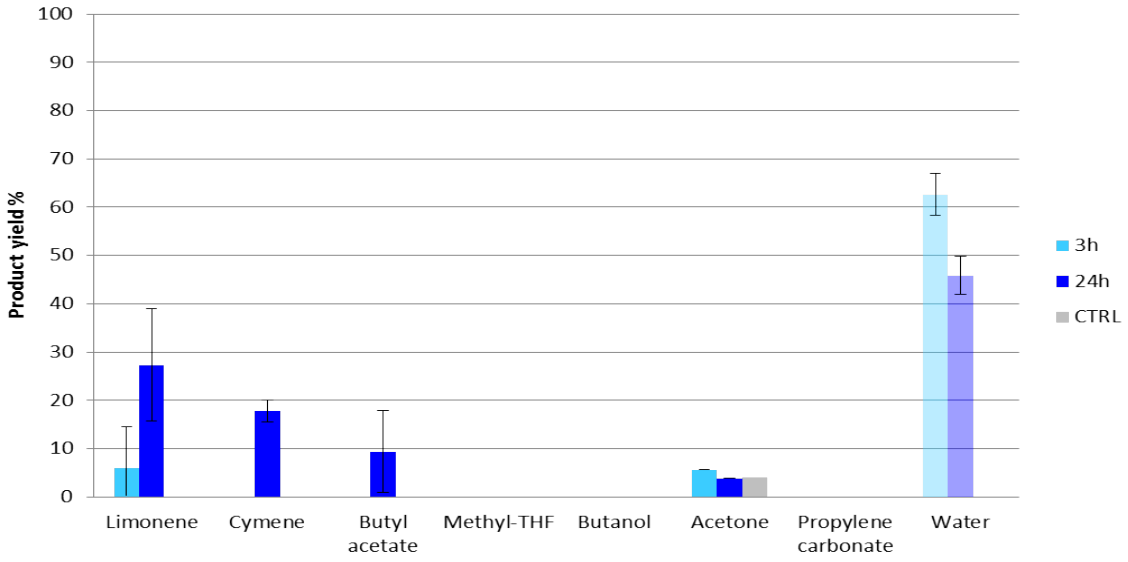


Figure C.40: Screening of ω-transaminase ATA-113 catalysed deamination of (S)-methylbenzylamine to acetophenone

Some issues have been encountered in performing this reaction in organic solvents. At this stage, challenges were faced in terms of:

1. Solubility of pyruvate salt and cofactor in the organic medium, making the preparation of the stock solutions and their equal split into the reaction vessels potentially inaccurate;

2. Volatility and polarity of substrate and product, being very similar to those of the

some of the solvents employed. This fact caused complications in the analytical stage.

Nevertheless, the trend is clear as well as surprising. The reaction was performed efficiently in water, with high yields after 3 hours. It appears that after 24 hours the product concentration decreased, possibly due to equilibrium shift taking place, with re-conversion of the ketone to the amine.

Peculiarly, no conversion was observed in the most hydrophilic solvents except from acetone, namely propylene carbonate, 1-butanol, 2-MeTHF. In the case of acetone, the solvent may have acted as an amine-acceptor itself, therefore providing a strong driving force towards unfavoured side product. Indeed, the reaction in acetone seems to have worked also in absence of the enzyme. This fact may be explained by: 1) the driving force exercised by a vast excess of acetone in the system, or 2) presence of the cofactor in the blank reaction tube, which might have played a catalytic role itself.

In summary, the transaminase of study showed potential in the most hydrophobic solvents and hardly any catalytic activity in the more hydrophilic organic solvents tested. This fact is surprising, since cofactor and pyruvate solubility in these solvents is extremely low.

C.6 Alcohol oxidation with galactose oxidase (GOase)

Oxidases are a class of oxidative enzymes which work by engaging directly with molecular oxygen, as opposed to oxygenases which require the addition of cofactors.[229] They attracted particular interest in the present years for reasons including:[229]

- Ability to use O_2 or air to directly oxidise organic compounds;
- Only hydrogen peroxide is produced along with the oxidised substrate, potentially allowing for additional profit;
- Optimum atom economy;
- Minimised E-factor and energy intensity;
- High selectivity (*e.g.* galactose oxidase oxidises primary alcohols selectively to aldehydes);
- One-step clean process;
- Mild and safe working conditions.

So far, the use of oxidases in organic solvents has been only sporadic.[230 231 232 233] More specifically, in regards to galactose oxidases only one study has addressed the challenge.[234] Morgan and Clark (2004) compared the activity of different oxidative enzymes including oxidases, peroxidase and alcohol dehydrogenase, in both water and butyl acetate. All the enzymes tested were previously freeze-dried and shown to retain activity in the organic solvent. Galactose oxidase, however, retained only 6% of the aqueous activity, improved to 14% when KCl was added to the enzyme preparation before the freeze-drying step. It is not clear whether this apparent salt-activation had an actual role in enhancing the enzyme's activity in organic solvents or it only protected it from denaturation during the freeze-drying step.[234]

In most cases, it is observed that the formation and accumulation of hydrogen peroxyde as a side product of the reaction causes deactivation of the enzyme. Hence, the use of oxidases is usually coupled with a second enzyme, a catalase or peroxidase, which converts H_2O_2 into harmless H_2O. Perhaps, a wide organic solvent screening may reveal that this coupled reaction may be avoidable by H_2O_2 phasing out or, on the contrary, by being more easily dispersed in the medium.

With this intent, galactose oxidase variant M_{3-5} was screened in organic solvents, as kindly supplied by University of Manchester from Nicholas Turner's research group. The reactions were carried out in eight different solvents, including water. In contrast with the standard procedure, no additional catalase or peroxidase was used to demolish the hydrogen peroxide side-product. The active oxidative agent in the transformation is molecular oxygen, made available simply by leaving a reasonable volume of air in the head-space of the sealed reaction tube.

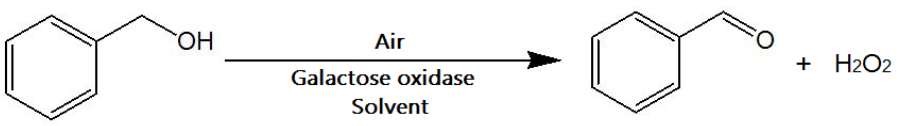


Figure C.41: Galactose oxidase (0.34 mg) catalysed oxidation of benzyl alcohol (50 mM) to benzaldehyde, in 1 ml stock solution in a 10 ml glass vial.

Figure C.42 shows the results obtained in the oxidation of benzyl alcohol to benzaldehyde with galactose oxidase. The different solvents employed have been ordered by decreasing $logP$ value, from limonene to propylene carbonate, and compared to the benchmark process in water.

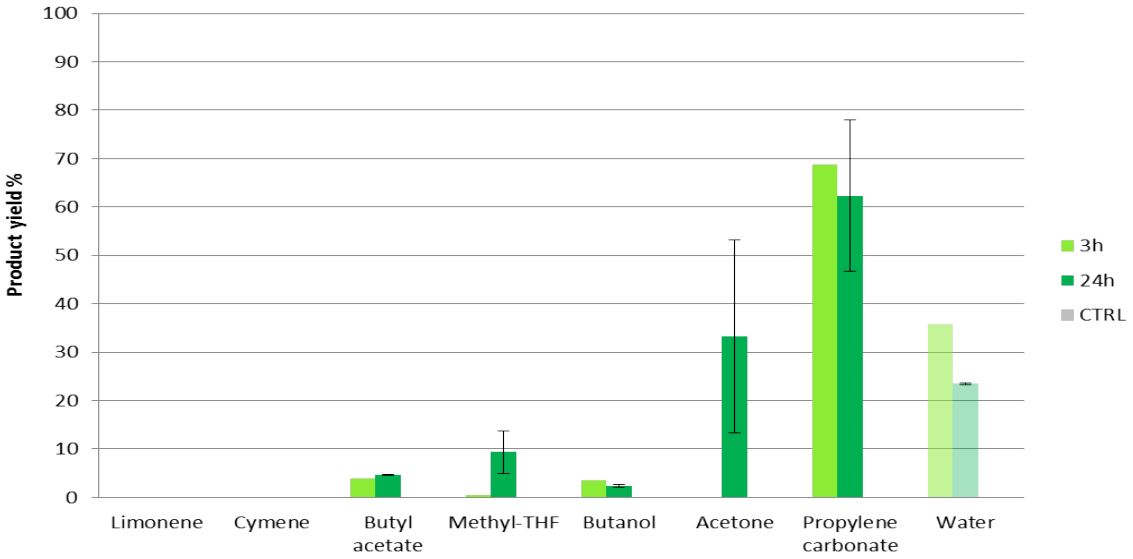


Figure C.42: Results of screening of galactose oxidase catalysed oxidation of benzyl alcohol to benzaldehyde in different organic solvents ordered by increasing polarity.

Hydrophilicity seems to play a key role in this process. No product was obtained in either limonene or *p*-cymene, the most hydrophobic solvents employed. Little conversion occurred in the moderately polar solvents, butyl acetate, 2-MeTHF and 1-butanol. Instead, striking yields were observed in the most polar solvents, with outstanding results for the use of propylene carbonate as a medium for galactose oxidase catalysed processes.

This behaviour may be explained as follows. The enzyme preparation was an aqueous phosphate buffer solution in which the enzyme was dissolved/suspended. It is likely the most hydrophobic media created a biphasic aqueous-organic system, where the organic solution containing the substrate was unable to reach the enzyme though the aqueous layer in its surroundings. The extremely favoured solubility of the substrate - benzyl alcohol - in the organic phase would keep it from diffusing into the aqueous buffer and, thus, from undergoing the transformation. Similarly, the highly hydrophilic hydrogen peroxide is

less likely to diffuse away from the enzyme's surroundings into the organic phase. These may be especially true for strongly hydrophobic solvents, by which H_2O_2 is likely to be constrained into the aqueous 'bubble' where its accumulation will lead to the enzyme's inactivation. This theory is in accordance with the statements made by Laane *et al.*, in which the ability of the substrate to partition between the organic continuum and the aqueous enzymatic environment was expected to be a crucial parameter for process optimisation.[104]

According to this premise, neat propylene carbonate may be a better medium than water for this reaction. By stripping some of the water located around the enzyme, propylene carbonate is likely to also disperse the peroxide dissolved within, hence preventing the latter from affecting the enzyme.

In conclusion, the organic solvent coating the enzyme's aqueous layer has appeared to have a significant effect on the catalytic activity of galactose oxidase. The use of highly polar solvents has shown to be promising, and to potentially outperform the conventional aqueous reaction media.

Notably, these reactions were performed without the aid of a peroxide-consuming enzyme. The promising results in polar solvents, especially in propylene carbonate, show that through medium optimisation it may be possible to completely overcome the need of a second enzyme in the process. Moreover, continuous-phase extraction may provide a way to extract and purify this hydrogen peroxide as a valuable co-product of the process.

C.7 Methods

C.7.1 Amidation

60 mg of *Pseudomonas Stutzeri* lipase (dried free enzyme preparation) were weighed into tubes using Flexi-weight. Then 100 mg molecular sieves (powder) were added to all tubes before adding the substrate 50 mg and the amine. Solvent was then added by Eppendorf and the reactions stirred at 50°C for 42 h. Reactions were run on the Mettler-Toledo Mini-Mapper in a standard double-sealed reaction block at 50°C with stirring with a simple magnetic flea. The reactions were automatically sampled on the Mini-Mapper at 3, 6, 12,

24 and 42 hours.

UV absorbance was converted to mmol/ml using the formula from the linear regression of the calibration curve. A calibration was done with the acid, SM and product at the following concentration. The acid was more difficult to calibrate due to its poor solubility.

The conversion was calculated as Product/(Product+Acid+SM) and the hydrolysis calculated as Acid/(Acid+Product+SM). The solvents peaks did not interfere with the samples. Samples were taken on mini-mapper at, 3, 6, 12 and 24 hours, by taking 100 μl into 600 μl of diluent. Vials were filled with 4:1 MeCN/water with 0.2 mg/ml of internal standard (4,4'-*tert*-butyl-biphenyl) pre-dissolved in it. Needle wash was 3.0 ml DMAc. Samples were spun down on the centrifuge before being run on the LCMS using BEH Shied RP18 column 2.1x100 mm with CatSci Generic 2.1 method (Acquity).

C.7.2 De-amination

In a reaction tube, weighed 2 mg of dry enzyme ω-transaminase (ATA-113), and 0.5 mg of PLP cofactor. Prepared stock solutions (values for 10 ml of solvent) of methylbenzylamine (0.24 mmol, 29 mg, 30 μl), methyl pyruvate (51 mg, 45 μl). Transfered 2 ml of each stock solution in the correspondent reaction tube. Ran reactions at 40°C, medium stirring using small light flees.

C.7.3 Oxidation

Prepared stock solutions (values for 10 ml of solvent) of benzyl alcohol (1 mmol, 108 mg, 104 μl). Transfered 2 ml of each stock solution in the correspondent reaction tube. With a micropipette, added 200 μl of GOase preparation (1.7 mg/ml concentration in phosphate buffer) into each reaction tube. Run at 30°C, medium stirring using small light flees.

List of Abbreviations

2-MeFuran = 2-Methylfuran

2-MeTHF = 2-Methyltetrahydrofuran

ABE = Acetone-Butanol-Ethanol process

AIT = Autoignition point

API = Active Pharmaceutical Ingredient

Asp = Aspartic acid aminoacid

AU = Absorption Unit

B2B = Business-to-business

Bp = Boiling point

BRICS = Brazil, Russia, India, China and South Africa

BTX = Benzene Toluene Xylenes

CAGR = Compound Annual Growth Rate

CALA = *Candida Antarctica* Lipase type A

CALB = *Candida Antarctica* Lipase type B

CDI = Carbodiimidazole

CEN = European Committee for Standardization

CMR = Carcinogenic Mutagenic Reprotoxic

CPME = Cyclopentylmethylether

DCC = *N,N'*-Dicyclohexylcarbodiimide

DCE = 1,2-Dichloroethane

DCM = Dichloromethane

DIAD = Diisopropyl azodicarboxylate

DKR = Dynamic Kinetic Resolution

DMAP = 4-Dimethylaminopyridine

DMC = Dimethylcarbonate

DMF = Dimethylformamide

DMI = Dimethyl isosorbide

DMSO = Dimethyl sulfoxide

EDCI = 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide

EHS = Environment Health and Safety

EtOH = Ethanol

EU = European Union

FAOSTAT = Food and Agriculture Organization's Statistics Department

Fp = Flash point

GC = Gas Chromatography

GC-FID = Gas Chromatography with Flame Ionisation Detector

GC-MS = Gas Chromatography with Mass Spectroscopy

GDP = Gross Domestic Product

GGE = Guaiacol glycerol ether

GHS = Globally Harmonized System of Classification and Labelling of Chemicals

Glu = Glutamic acid aminoacid

GOase = Galactose oxidase

His = Histidine aminoacid

HIV = Human immunodeficiency virus

HLB = Huanglongbing

^{1}H-NMR = Proton Nuclear Magnetic Resonance

HPLC = High Performance Liquid Chromatography

HSP = Hansen's Solubility Parameter

HSPiP = Hansen's Solubiliry Parameters in Practice

IMF = International Monetary Fund

IR = Infrared

LFER = Linear Free Energy Relationship

LSER = Linear Solvation Energy Relationship

MEK = Methylethylketone

MsCl = Methanesulfonyl chloride

MT = Metric tonne

MTBE = Methyl *tert*-butyl ether

NA = 4-Nitroaniline

nIR = Near-Infrared

NMP = *N*-Methyl-2-pyrrolidone

NN = *N*,*N*-diethyl-4-nitroaniline

NR = Nile Red

NRI = Norepinephrine reuptake inhibitor

OPEC = Organization of the Petroleum Exporting Countries

PRER = Product-Related Environmental Regulation

PSL = Pseudomonas Stutzeri Lipase

Rbx = Reboxetine

REACH = Registration, Evaluation, Authorisation and restriction of Chemicals

Rf = Retention factor

Rt = Retention time

$ScCO_2$ = Supercritical carbon dioxide

SDG = Sustainable Development Goal

Ser = Serine aminoacid

SMILES = Simplified Molecular-Input Line-Entry System

THBP = Tetrahydroxybenzophenone

THF = Tetrahydrofuran

TMS = Tetramethylsylane

TMSCl = Trimethylsilyl chloride

UNEP = United Nations Environment Programme

USA = United States of America

USD = United States Dollar

USDA = United States' Department of Agriculture

UV/Vis = Ultraviolet-visible light

VOC = Volatile Organic Compound

Bibliography

[1] G. Brundtland, M. Khalid, S. Agnelli, S. Al-Athel, B. Chidzero, L. Fadika, V. Hauff, I. Lang, M. Shijun and M. M. de Botero, *Our common future*, Oxford University Press, USA, 1987.

[2] J. Elkington, *Enter the triple bottom line*, Oxford University Press, 2004.

[3] J. Morelli, *Journal of Environmental Sustainability*, 2013, **1**, 2.

[4] R. Heijungs, G. Huppes and J. B. Guinée, *Polymer Degradation & Stability*, 2010, **95**, 422–428.

[5] N. Dempsey, G. Bramley, S. Power and C. Brown, *Sustainable Development*, 2011, **19**, 289–300.

[6] M. Missimer, *Ph.D. thesis*, Blekinge Institute of Technology, 2015.

[7] S. Anand and A. Sen, *World development*, 2000, **28**, 2029–2049.

[8] S. Baumgärtner and M. Quaas, *Ecological Economics*, 2010, **69**, 445–450.

[9] UNEP, *Towards a green economy: pathways to sustainable development and poverty eradication*, United Nations Environment Programme, 2011.

[10] United Nations' Sustainable Development, *http://www.un.org/*, last accessed 2016.

[11] P. T. Anastas and J. C. Warner, *Green chemistry: Theory and practice*, 1998, 29–56.

[12] P. Anastas and N. Eghbali, *Chemical Society Reviews*, 2010, **39**, 301–312.

[13] R. A. Sheldon, *Green Chem.*, 2007, **9**, 1273–1283.

[14] F. M. Kerton and R. Marriott, *Alternative solvents for green chemistry*, Royal Society of chemistry, 2013, vol. 20.

[15] The Council of the European Union, *Solvent Emission Directive: Council Directive 1999/13/EC, article 2*, 1999.

[16] C. Reichardt and T. Welton, *Solvents and solvent effects in organic chemistry*, John Wiley & Sons, 2011.

[17] P. G. Jessop, D. A. Jessop, D. Fu and L. Phan, *Green Chemistry*, 2012, **14**, 1245–1259.

[18] M. J. Kamlet, M. H. Abraham, R. M. Doherty and R. Taft, *Journal of the American Chemical Society*, 1984, **106**, 464–466.

[19] C. M. Alder, J. D. Hayler, R. K. Henderson, A. M. Redman, L. Shukla, L. E. Shuster and H. F. Sneddon, *Green Chemistry*, 2016, 471–477.

[20] Transparency market research. Global solvents market, *http://www.transparencymarketresearch.com/solvents-market.html/*, last accessed 2016.

[21] Markets and Markets. Solvent Market by type, application and source, *http://www.marketsandmarkets.com/PressReleases/solvent.asp/*, last accessed 2016.

[22] Merchant Research & Consulting, *Toluene: 2015 World Market Outlook and Forecast up to 2019*, 2015.

[23] G. Wypych, *Handbook of Solvents*, 2001, 847–922.

[24] D. J. Constable, C. Jimenez-Gonzalez and R. K. Henderson, *Organic Process Research & Development*, 2007, **11**, 133–137.

[25] REACH. Registration, Evaluation, Authorisation and Restriction of Chemicals, *http://ec.europa.eu/enterprise/sectors/chemicals/reach/*, last accessed 2016.

[26] D. Prat, A. Wells, J. Hayler, H. Sneddon, C. R. McElroy, S. Abou-Shehada and P. J. Dunn, *Green Chemistry*, 2016, **18**, 288–296.

[27] K. Grodowska and A. Parczewski, *Acta Poloniae Pharmaceutica*, 2010, **67**, 3–12.

[28] M. J. Raymond, C. S. Slater and M. J. Savelski, *Green Chemistry*, 2010, **12**, 1826–1834.

[29] J. H. Clark, T. J. Farmer, A. J. Hunt and J. Sherwood, *International journal of molecular sciences*, 2015, **16**, 17101–17159.

[30] C. Capello, U. Fischer and K. Hungerbühler, *Green Chemistry*, 2007, **9**, 927–934.

[31] J. M. DeSimone, *Science*, 2002, **297**, 799–803.

[32] W. M. Nelson, *Green Solvents for Chemistry: Perspectives and Practice: Perspectives and Practice*, Oxford University Press, 2003.

[33] Energy Global. Hydrocarbon Engineering, *http://www.energyglobal.com/*, last accessed 2016.

[34] A. W. Sheppard, I. Gillespie, M. Hirsch and C. Begley, *Current Opinion in Environmental Sustainability*, 2011, **3**, 4–10.

[35] L. A. Pfaltzgraff, E. C. Cooper, V. Budarin and J. H. Clark, *Green Chemistry*, 2013, **15**, 307–314.

[36] CEN Technical Committee WG 2, *EN 16575:2014 Bio-based products – Vocabulary*, 2014.

[37] A. Mohammad, *Green Solvents I: Properties and Applications in Chemistry*, Springer Verlag, 2012, vol. 1.

[38] P. G. Jessop, *Green Chemistry*, 2011, **13**, 1391–1398.

[39] Y. Gu and F. Jérôme, *Chemical Society Reviews*, 2013, **42**, 9550–9570.

[40] R. A. Sheldon, *Green Chemistry*, 2005, **7**, 267–278.

[41] J. H. Clark and S. J. Tavener, *Organic process research & development*, 2007, **11**, 149–155.

[42] J. Fabri, U. Graeser and T. A. Simo, *Ullmann's Encyclopedia of Industrial Chemistry - Toluene*, Wiley-VCH Verlag GmbH, 2005.

[43] E. D. Ozokwelu, *Kirk-Othmer Encyclopedia of Chemical Technology - Toluene*, John Wiley & Sons, 2000.

[44] M. Martin-Luengo, M. Yates, M. M. Domingo, B. Casal, M. Iglesias, M. Esteban and E. Ruiz-Hitzky, *Applied Catalysis B: Environmental*, 2008, **81**, 218–224.

[45] M. Martin-Luengo, M. Yates, E. S. Rojo, D. Huerta Arribas, D. Aguilar and E. Ruiz Hitzky, *Applied Catalysis A: General*, 2010, **387**, 141–146.

[46] J. A. Dávila, M. Rosenberg and C. A. Cardona, *Waste & Biomass Valorization*, 2015, **6**, 253–261.

[47] E. Jongedijk, K. Cankar, M. Buchhaupt, J. Schrader, H. Bouwmeester and J. Beekwilder, *Applied Microbiology & Biotechnology*, 2016, **100**, 2927–2938.

[48] M. Bender, *Global Aromatics Supply Today and Tomorrow*, BASF, 2013.

[49] M. Röper, E. Gehrer, T. Narbeshuber and W. Siegel, *Ullmann's Encyclopedia of Industrial Chemistry - Acylation and alkylation*, Wiley Online Library, 2000.

[50] H. Fiege, H.-W. Voges, T. Hamamoto, S. Umemura, T. Iwata, H. Miki, Y. Fujita, H.-J. Buysch, D. Garbe and W. Paulus, *Ullmann's Encyclopedia of Industrial Chemistry - Phenol derivatives*, Wiley Online Library, 2000.

[51] M. F. Neves, V. G. Trombin, F. F. Lopes, R. Kalaki and P. Milan, *The orange juice business: A Brazilian perspective*, Springer Science & Business Media, 2012.

[52] L. Pfaltzgraff, *Ph.D. thesis*, University of York, 2014.

[53] M. Pourbafrani, *Bioresource Technology*, 2010, **101**, 4246 – 4250.

[54] N. I. Tracy, D. Chen, D. W. Crunkleton and G. L. Price, *Fuel*, 2009, **88**, 2238–2240.

[55] M. Firdaus and M. A. Meier, *Green Chemistry*, 2013, **15**, 370–380.

[56] M. Pourbafrani, *Ph.D. thesis*, Chalmers University of Technology, 2010.

[57] R. Ciriminna, M. Lomeli-Rodriguez, P. D. Carà, J. A. Lopez-Sanchez and M. Pagliaro, *Chemical Communications*, 2014, **50**, 15288–15296.

[58] SigmaAldrich Safety Data Sheets of individual chemicals, *http://www.sigmaaldrich.com/*, last accessed May 2016.

[59] Hansen Solubility Parameters in Practice software, *http://www.hansen-solubility.com/HSPiP.html/*, downloaded in Aug 2014 under UoY licence.

[60] WHO, *Air quality guidelines for Europe*, European Series, Regional Publications, No. 91, 2000.

[61] United States OSHA. Permissible Exposure Limits, *https://www.osha.gov/dsg/annotated-pels/*, last accessed 2016.

[62] P. Campo, K. Maguin, S. Gabriel, A. Moller, E. Nies, M. D. Sole Gomez and E. Toppila, *Combined exposure to noise and ototoxic substances*, European Agency for Safety and Health at Work, 2009.

[63] European Chemical Agency. Safety Data Sheet (SDS), *http://echa.europa.eu/regulations/reach/safety-data-sheets*, last accessed 2016.

[64] G. Roberts, *Life Cycle Assessment - Renewable and Sustainable Citrus Oils*, Environment Resources Management, 2012.

[65] A. J. Straathof, S. Panke and A. Schmid, *Current Opinion in Biotechnology*, 2002, **13**, 548–556.

[66] D. J. Constable, P. J. Dunn, J. D. Hayler, G. R. Humphrey, J. L. Leazer Jr, R. J. Linderman, K. Lorenz, J. Manley, B. A. Pearlman and A. Wells, *Green Chemistry*, 2007, **9**, 411–420.

[67] C. Jiménez-González, P. Poechlauer, Q. B. Broxterman, B.-S. Yang, D. am Ende, J. Baird, C. Bertsch, R. E. Hannah, P. Dell'Orco and H. Noorman, *Organic Process Research & Development*, 2011, **15**, 900–911.

[68] N. J. Turner and E. O'Reilly, *Nature Chemical Biology*, 2013, **9**, 285–288.

[69] N. Ran, L. Zhao, Z. Chen and J. Tao, *Green Chemistry*, 2008, **10**, 361–372.

[70] C. M. Clouthier and J. N. Pelletier, *Chemical Society Reviews*, 2012, **41**, 1585–1605.

[71] G. Grogan, *Annual Reports Section B (Organic Chemistry)*, 2012, **108**, 202–227.

[72] J. Tao and J.-H. Xu, *Current Opinion in Chemical Biology*, 2009, **13**, 43–50.

[73] R. N. Patel, *Coordination Chemistry Reviews*, 2008, **252**, 659–701.

[74] S. Wenda, S. Illner, A. Mell and U. Kragl, *Green Chemistry*, 2011, **13**, 3007–3047.

[75] E. A. Sym, *Biochemical Journal*, 1935, **30**, 609.

[76] A. Zaks and A. M. Klibanov, *Proceedings of the National Academy of Sciences*, 1985, **82**, 3192–3196.

[77] G. Carrea and S. Riva, *Organic synthesis with enzymes in non-aqueous media*, Wiley-VCH, 2008.

[78] A. G Correa, M. W Paixao and R. S Schwab, *Current Organic Synthesis*, 2015, **12**, 675–695.

[79] M. Perez-Sanchez, M. Sandoval, M. J. Hernaiz and P. D. Maria, *Current Organic Chemistry*, 2013, **17**, 1188–1199.

[80] M. J. Hernáiz, A. R. Alcántara, J. I. García and J. V. Sinisterra, *Chemistry - A European Journal*, 2010, **16**, 9422–9437.

[81] R. Sharma, Y. Chisti and U. C. Banerjee, *Biotechnology Advances*, 2001, **19**, 627–662.

[82] D. G. Brown and J. Boström, *Journal of Medicinal Chemistry*, 2015, 511–525.

[83] J. S. Carey, D. Laffan, C. Thomson and M. T. Williams, *Organic & Biomolecular Chemistry*, 2006, **4**, 2337–2347.

[84] J. P. Adams, C. M. Alder, I. Andrews, A. M. Bullion, M. Campbell-Crawford, M. G. Darcy, J. D. Hayler, R. K. Henderson, C. A. Oare, I. Pendrak, R. Aniko, L. E. Shuster, H. F. Sneddon and M. D. Walker, *Green Chemistry*, 2013, **15**, 1542–1549.

[85] E. Fischer and A. Speier, *Chemistry*, 1895, **28**, 3252–3258.

[86] M. Tsakos, E. S. Schaffert, L. L. Clement, N. L. Villadsen and T. B. Poulsen, *Natural Product Reports*, 2015, **32**, 605–632.

[87] E. Haslam, *Tetrahedron*, 1980, **36**, 2409–2433.

[88] J. Otera and J. Nishikido, *Esterification: methods, reactions, and applications*, John Wiley & Sons, 2009.

[89] C. A. Martinez, S. Hu, Y. Dumond, J. Tao, P. Kelleher and L. Tully, *Organic Process Research & Development*, 2008, **12**, 392–398.

[90] O. Kirk and M. W. Christensen, *Organic Process Research & Development*, 2002, **6**, 446–451.

[91] S. Cao, Y. Feng, Z. Liu, Z. Ding and Y. Cheng, *Applied Biochemistry & Biotechnology*, 1992, **32**, 7–13.

[92] A. L. Paiva, V. M. Balcão and F. X. Malcata, *Enzyme & Microbial Technology*, 2000, **27**, 187–204.

[93] A. Cornish-Bowden, *Fundamentals of enzyme kinetics*, John Wiley & Sons, 2013.

[94] A. M. Koskinen, *Enzymatic reactions in organic media*, Springer, 1996.

[95] C.-S. Chen and C. J. Sih, *Angewandte Chemie*, 1989, **28**, 695–707.

[96] J. S. Dordick, *Enzyme & Microbial Technology*, 1989, **11**, 194–211.

[97] P. J. Dyson and P. G. Jessop, *Catalysis Science & Technology*, 2016, **6**, 3302–3316.

[98] J. L. Schmitke, C. R. Wescott and A. M. Klibanov, *Journal of the American Chemical Society*, 1996, **118**, 3360–3365.

[99] E. Wehtje, I. Svensson, P. Adlercreutz and B. Mattiasson, *Biotechnology Techniques*, 1993, **7**, 873–878.

[100] R. H. Valivety, P. J. Halling, A. D. Peilow and A. R. Macrae, *Biochimica et Biophysica Acta - Protein Structure & Molecular Enzymology*, 1992, **1122**, 143–146.

[101] R. H. Valivety, P. J. Halling and A. R. Macrae, *FEBS Letters*, 1992, **301**, 258–260.

[102] N. M. Micaelo and C. M. Soares, *FEBS Journal*, 2007, **274**, 2424–2436.

[103] D. Lousa, A. M. Baptista and C. M. Soares, *Physical Chemistry Chemical Physics*, 2013, **15**, 13723–13736.

[104] C. Laane, S. Boeren, K. Vos and C. Veeger, *Biotechnology & Bioengineering*, 1987, **30**, 81–87.

[105] P. J. Halling, *Enzyme & Microbial Technology*, 1994, **16**, 178 – 206.

[106] S. H. van Erp, E. O. Kamenskaya and Y. L. Khmelnitsky, *European Journal of Biochemistry*, 1991, **202**, 379–384.

[107] N. Doukyu and H. Ogino, *Biochemical Engineering Journal*, 2010, **48**, 270 – 282.

[108] D. Herbst, S. Peper and B. Niemeyer, *Journal of Biotechnology*, 2012, **162**, 398–403.

[109] L. V. Schneider, *Biotechnology & Bioengineering*, 1991, **37**, 627–638.

[110] M. Reslow, P. Adlercreutz and B. Mattiasson, *Applied Microbiology & Biotechnology*, 1987, **26**, 1–8.

[111] M. D. Gagnon and P. T. Vasudevan, *Energy & Fuels*, 2011, **25**, 4669–4674.

[112] J. F. Martins, T. C. de Sampaio, I. B. de Carvalho and S. Barreiros, *Biotechnology & Bioengineering*, 1994, **44**, 119–124.

[113] Y. B. Tewari, M. M. Schantz and D. J. Vanderah, *Journal of Chemical & Engineering Data*, 1999, **44**, 641–647.

[114] K. Ryu and J. S. Dordick, *Biochemistry*, 1992, **31**, 2588–2598.

[115] L. E. S. Brink and J. Tramper, *Biotechnology & Bioengineering*, 1985, **27**, 1258–1269.

[116] R. H. Valivety, G. A. Johnston, C. J. Suckling and P. J. Halling, *Biotechnology & Bioengineering*, 1991, **38**, 1137–1143.

[117] A. A. Clifford, *NATO ASI Series E Applied Sciences*, 1994, **273**, 449–480.

[118] P. Lozano, G. Villora, D. Gomez, A. Gayo, J. Sanchez-Conesa, M. Rubio and J. Iborra, *The Journal of Supercritical Fluids*, 2004, **29**, 121–128.

[119] K. Rezaei, F. Temelli and E. Jenab, *Biotechnology Advances*, 2007, **25**, 272–280.

[120] A. M. Klibanov, *Nature*, 2001, **409**, 241–246.

[121] B. Stauch, S. J. Fisher and M. Cianci, *Journal of Lipid Research*, 2015, **56**, 2348–2358.

[122] K. Drauz, *Enzyme catalysis in organic synthesis*, Wiley-VCH, 2012, vol. 1.

[123] L. Casas-Godoy, S. Duquesne, F. Bordes, G. Sandoval and A. Marty, *Lipases and Phospholipases: Methods and Protocols*, 2012, 3–30.

[124] P. Lozano, *Green Chemistry*, 2010, **12**, 555–569.

[125] Z.-Q. Duan and F. Hu, *Green Chemistry*, 2012, **14**, 1581–1583.

[126] W.-L. Gao, H. Liu, N. Li and M.-H. Zong, *Bioresource Technology*, 2012, **118**, 82–88.

[127] S. Shanmuganathan, D. Natalia, A. van den Wittenboer, C. Kohlmann, L. Greiner and P. D. de María, *Green Chemistry*, 2010, **12**, 2240–2245.

[128] P. Hoyos, M. A. Quezada, J. V. Sinisterra and A. R. Alcántara, *Journal of Molecular Catalysis B: Enzymatic*, 2011, **72**, 20–24.

[129] X.-M. Wu, W. Sun, J.-Y. Xin and C.-G. Xia, *World Journal of Microbiology and Biotechnology*, 2008, **24**, 2421–2424.

[130] Z.-G. Chen, D.-N. Zhang, L. Cao and Y.-B. Han, *Bioresource Technology*, 2013, **133**, 82–86.

[131] T. Bavaro, P. Torres-Salas, D. Ubiali and M. Terreni, *RSC Advances*, 2013, 471–477.

[132] G. W. Huisman and S. J. Collier, *Current Opinion in Chemical Biology*, 2013, **17**, 284–292.

[133] D. M. Solano, P. Hoyos, M. Hernáiz, A. Alcántara and J. Sánchez-Montero, *Bioresource Technology*, 2012, **115**, 196–207.

[134] R. N. Patel, *ACS Catalysis*, 2011, **1**, 1056–1074.

[135] A. L. Margolin, D. L. Delinck and M. R. Whalon, *Journal of the American Chemical Society*, 1990, **112**, 2849–2854.

[136] A. Ghanem, M. N. Aboul-Enein, A. El-Azzouny and M. F. El-Behairy, *Journal of Chromatography A*, 2010, **1217**, 1063–1074.

[137] A. Mustranta, *Applied Microbiology and Biotechnology*, 1992, **38**, 61–66.

[138] W. H. Lee, K.-J. Kim, M. G. Kim and S. B. Lee, *Journal of Fermentation & Bioengineering*, 1995, **80**, 613–615.

[139] M. Goto, S. Noda, N. Kamiya and F. Nakashio, *Biotechnology Letters*, 1996, **18**, 839–844.

[140] N. D'Antona, P. Lombardi, G. Nicolosi and G. Salvo, *Process Biochemistry*, 2002, **38**, 373–377.

[141] S. T. Hayes, G. Assaf, G. Checksfield, C. Cheung, D. Critcher, L. Harris, R. Howard, S. Mathew, C. Regius, G. Scotney and A. Scott, *Organic Process Research & Development*, 2011, **15**, 1305–1314.

[142] A. Goswami, J. M. Howell, E. Y. Hua, K. D. Mirfakhrae, M. C. Soumeillant, S. Swaminathan, X. Qian, F. A. Quiroz, T. C. Vu, X. Wang, B. Zheng, D. R. Kronenthal and R. N. Patel, *Organic Process Research & Development*, 2001, **5**, 415–420.

[143] M. J. Homann, R. Vail, B. Morgan, V. Sabesan, C. Levy, D. R. Dodds and A. Zaks, *Advanced Synthesis & Catalysis*, 2001, **343**, 744–749.

[144] L. Murtagh, C. Dunne, G. Gabellone, N. J. Panesar, S. Field, L. M. Reeder, J. Saenz, G. P. Smith, K. Kissick, C. Martinez, J. G. van Alsten, M. C. Evans, L. C. Franklin and T. Nanninga, *Organic Process Research & Development*, 2011, **15**, 1315–1327.

[145] E. Weringer, *Reflections on twentieth-century psychopharmacology - The history of psychotropic drugs at Pfizer*, Animula Publishing House, 2004, pp. 58–64.

[146] G. Assaf, G. Checksfield, D. Critcher, P. J. Dunn, S. Field, L. J. Harris, R. M. Howard, G. Scotney, A. Scott, S. Mathew, G. M. H. Walker and A. Wilder, *Green Chemistry*, 2012, **14**, 123–129.

[147] K. E. Henegar and M. Cebula, *Organic Process Research & Development*, 2007, **11**, 354–358.

[148] K. E. Henegar, C. T. Ball, C. M. Horvath, K. D. Maisto and S. E. Mancini, *Organic Process Research & Development*, 2007, **11**, 346–353.

[149] E. M. Anderson, K. M. Larsson and O. Kirk, *Biocatalysis & Biotransformation*, 1998, **16**, 181–204.

[150] G. Grogan, *Practical biotransformations: a beginner's guide*, John Wiley & Sons, 2009.

[151] M. Mascal, *Biofuels, Bioproducts & Biorefining*, 2012, **6**, 483–493.

[152] M. Anitha, S. K. Kamarudin, N. Shamsul and N. T. Kofli, *International Journal of Hydrogen Energy*, 2015, **40**, 11791–11799.

[153] M. Sauer, *FEMS Microbiology Letters*, 2016, **363**, 120–134.

[154] S.-W. Chang, J.-F. Shaw, K.-H. Yang, L. Shih, C.-H. Hsieh and C.-J. Shieh, *Green Chemistry*, 2005, **7**, 547–551.

[155] N. R. Khan and V. K. Rathod, *Process Biochemistry*, 2015, **50**, 1793–1806.

[156] C. R. McElroy, A. Constantinou, L. C. Jones, L. Summerton and J. H. Clark, *Green Chemistry*, 2015, **17**, 3111–3121.

[157] C. Reichardt, *Organic Process Research & Development*, 2007, **11**, 105–113.

[158] C. M. Hansen, *Hansen solubility parameters: a user's handbook*, CRC press, 2007.

[159] R. Howard and S. Spencer, *Project Report Esreboxetine - Acylation (Confidential report)*, Pfizer Global Research & Development, Biocatalysis Centre of Emphasis, 2009.

[160] S. T. Hayes, *(S,S)-Reboxetine Succinate: Chemical adventures in chasing the cheapest route (Confidential presentation)*, Pfizer Global Research & Development, 2010.

[161] M. J. Pilling and P. W. Seakins, *Reaction kinetics*, Oxford University Press, 1996.

[162] D. A. Kraut, K. S. Carroll and D. Herschlag, *Annual Review of Biochemistry*, 2003, **72**, 517–571.

[163] J. J. Rooney, *Journal of Molecular Catalysis A: Chemical*, 1998, **133**, 303–305.

[164] M.-P. Bousquet-Dubouch, M. Graber, N. Sousa, S. Lamare and M.-D. Legoy, *Biochimica et Biophysica Acta*, 2001, **1550**, 90–99.

[165] L. Liu and Q.-X. Guo, *Chemical Reviews*, 2001, **101**, 673–696.

[166] D. H. Leung, R. G. Bergman and K. N. Raymond, *Journal of the American Chemical Society*, 2008, **130**, 2798–2805.

[167] D. M. Ford, *Journal of the American Chemical Society*, 2005, **127**, 16167–16170.

[168] R. C. Petersen, J. H. Markgraf and S. D. Ross, *Journal of the American Chemical Society*, 1961, **83**, 3819–3823.

[169] D. Taylor, P. Mongelard, A. Mulkern, P. Piatkiewicz, M. Pitts, J. Reid and D. Santillo, *A Guide to Substitution*, UK Chemicals Stakeholder Forum, 2010.

[170] W. J. Sutherland, E. Fleishman, M. B. Mascia, J. Pretty and M. A. Rudd, *Methods in Ecology & Evolution*, 2011, **2**, 238–247.

[171] A. B. Boxall, M. A. Rudd, B. W. Brooks, D. J. Caldwell, K. Choi, S. Hickmann, E. Innes, K. Ostapyk, J. P. Staveley, T. Verslycke, G. T. Ankley, K. F. Beazley, S. E. Belanger, J. P. Berninger, P. Carriquiriborde, A. Coors, P. C. DeLeo, S. D. Dyer, J. F. Ericson, F. Gagne, J. P. Giesy, T. Gouin, L. Hallstrom, M. V. Karlsson, D. G. J. Larsson, J. M. Lazorchak, F. Mastrocco, A. McLaughlin, M. E. McMaster, R. D. Meyerhoff, R. Moore, J. L. Parrott, J. R. Snape, R. Murray-Smith, M. R. Servos, P. K. Sibley, J. O. Straub, N. D. Szabo, E. Topp, G. R. Tetreault, V. L. Trudeau and G. Van Der Kraak, *Environmental Health Perspectives*, 2012, **120**, 1221.

[172] E. Fleishman, D. E. Blockstein, J. A. Hall, M. B. Mascia, M. A. Rudd, J. M. Scott, W. J. Sutherland, A. M. Bartuska, A. G. Brown, C. A. Christen, J. P. Clement, D. Dellasala, C. S. Duke, M. Eaton, S. J. Fiske, H. Gosnell, J. C. Haney, M. Hutchins, M. L. Klein, J. Marqusee, B. R. Noon, J. R. Nordgren, P. M. Orbuch, J. Powell, S. P. Quarles, K. A. Saterson, C. C. Savitt, B. A. Stein and W. M. S., *BioScience*, 2011, **61**, 290–300.

[173] M. A. Rudd, *Frontiers in Marine Science*, 2014, **1**, 36.

[174] W. J. Sutherland and H. J. Woodroof, *Trends in Ecology & Evolution*, 2009, **24**, 523–527.

[175] M. A. Rudd, *Conservation Biology*, 2011, **25**, 860–866.

[176] C. W. Craighead, J. Blackhurst, M. J. Rungtusanatham and R. B. Handfield, *Decision Sciences*, 2007, **38**, 131–156.

[177] J. E. van Dam, B. de Klerk-Engels, P. C. Struik and R. Rabbinge, *Industrial Crops & Products*, 2005, **21**, 129–144.

[178] H. Langeveld, J. Sanders and M. Meeusen, *The biobased economy: biofuels, materials, and chemicals in the post-oil era*, Earthscan, 2012, pp. 83–100.

[179] S. Piotrowski, M. Carus and R. Essel, *Industrial Biotechnology*, 2015, **11**, 308–315.

[180] M. Kircher, *Biofuels, Bioproducts & Biorefining*, 2012, **6**, 369–375.

[181] M. Narodoslawsky, A. Niederl-Schmidinger and L. Halasz, *Journal of Cleaner Production*, 2008, **16**, 164–170.

[182] M. K. Benkovich, *Solvent Substitution for Electronic Products*, Kansas Allied-Signal Aerospace Division, 1992.

[183] K. R. Stone and J. Springer, *Environmental Progress*, 1995, **14**, 261–265.

[184] T. Toplisek and R. Gustafson, *Precision Cleaning*, 1995, **3**, 17–20.

[185] K. Fisher, *Terpenes Replacing BTEX In Oil Field*, American Oil & Gas Reporter, 2013.

[186] E. Besson, *BP cleanup extends beyond area waters*, Houma Times, 2012.

[187] C. S. Pereira, V. M. Silva and A. E. Rodrigues, *Green Chemistry*, 2011, **13**, 2658–2671.

[188] J. Durkee, *Cleaning with solvents: Science and technology*, William Andrew, 2013.

[189] J. Sherman, B. Chin, P. Huibers, R. Garcia-Valls and T. A. Hatton, *Environmental Health Perspectives*, 1998, **106**, 253.

[190] A. M. Niziolek, O. Onel, J. A. Elia, R. C. Baliban and C. A. Floudas, *AIChE Journal*, 2015, **61**, 831–856.

[191] S. K. Green, R. E. Patet, N. Nikbin, C. L. Williams, C.-C. Chang, J. Yu, R. J. Gorte, S. Caratzoulas, W. Fan, D. G. Vlachos and P. J. Dauenhauer, *Applied Catalysis B: Environmental*, 2016, **180**, 487–496.

[192] Organisation of Petroleum Exporting Countries (OPEC) - 2014 World Oil Outlook, *http://www.opec.org*, last accessed 2016.

[193] International Monetary Fund (IMF) - World Economic Outlook 2015, *http://www.imf.org*, last accessed 2016.

[194] FAOSTAT. Food and Agriculture Organisation Statistics Department, *http://faostat3.fao.org/browse/Q/QC/E/*, last accessed 2016.

[195] USDA, *Citrus: World Markets and Trade*, USDA Foreign Agricultural Service, 2016.

[196] Map developed with free online tool on ChartsBin.com under creative commons licence, *http://chartsbin.com/view/37266*, last accessed 2016.

[197] R. P. Bates, J. R. Morris and P. G. Crandall, *Principles and practices of small- and medium-scale fruit juice processing*, FAO, 2001.

[198] Codex Alimentarius International Food Standards, *General Standard for Fruit Juices and Nectars (CODEX STAN 247-2005)*, 2005.

[199] J. M. Bové, *Journal of Plant Pathology*, 2006, 7–37.

[200] I. Cerón-Salazar and C. Cardona-Alzate, *Ingeniería y Ciencia*, 2011, **7**, 65–86.

[201] S. F. Pfau, J. E. Hagens, B. Dankbaar and A. J. Smits, *Sustainability*, 2014, **6**, 1222–1249.

[202] E. Barbier, *Natural Resources Forum*, 2011, **35**, 233–245.

[203] EUROSTAT - PRODCOM list, *http://ec.europa.eu/eurostat/web/prodcom/data/database*, last accessed 2016.

[204] ANP. Homologocao dos volumes comercializados, *http://www.anp.gov.br*, last accessed 2015.

[205] P. Rajesh, *Benzene and toluene: The Indian scenario*, Platt's - Global Aromatics Conference, 2013.

[206] C. C. Reporter, *Production and consumption of toluene grow in Chinahttps://www.highbeam.com/doc/1G1-277676317.html*, HighBeam Research, 2012.

[207] Indexmundi. South Africa Yearly Imports in US Dollars - Toluene, *http://www.indexmundi.com*, last accessed 2012.

[208] A. H. Putnam, *Florida Citrus Statistics 2013-2014*, Florida Department of Agriculture and Consumer Services, 2015.

[209] EUROSTAT - Agriculture, forestry and fisheries - Crop statistics, *http://ec.europa.eu/eurostat/data/database*, last accessed 2016.

[210] M. Saxena and C. P. Gandhi, *Indian Horticulture Database*, National Horticulture Information Service, 2014.

[211] G. Resch, A. Held, T. Faber, C. Panzer, F. Toro and R. Haas, *Energy Policy*, 2008, **36**, 4048–4056.

[212] A. E. Wheals, L. C. Basso, D. M. Alves and H. V. Amorim, *Trends in Biotechnology*, 1999, **17**, 482–487.

[213] M. Kircher, *Biofuels, Bioproducts & Biorefining*, 2012, **6**, 240–245.

[214] Indiana Administrative Code (Article 8) - Volatile Organic Compound Rules, *http://www.in.gov/legislative/iac/*, 2015.

[215] Y. Marcus, *Chemical Society Reviews*, 1993, **22**, 409–416.

[216] M. J. Kamlet, J. L. M. Abboud, M. H. Abraham and R. Taft, *Journal of Organic Chemistry*, 1983, **48**, 2877–2887.

[217] C. Reichardt, *Chemical Reviews*, 1994, **94**, 2319–2358.

[218] J. Verhoef and E. Barendrecht, *Journal of Electroanalytical Chemistry & Interfacial Electrochemistry*, 1976, **71**, 305–315.

[219] S. Zhan, X. Tao, L. Cai, X. Liu and T. Liu, *Green Chemistry*, 2014, **16**, 4649–4653.

[220] H. B. Mereyala, S. K. Mamidyala, K. P. Chigurupati and S. R. Srinivasa, *Synthesis*, 2003, **2003**, 2378–2384.

[221] ICH. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use, *http://www.ich.org/*, last accessed 2016.

[222] C. A. Montalbetti and V. Falque, *Tetrahedron*, 2005, **61**, 10827–10852.

[223] S. Van Pelt, R. Teeuwen, M. Janssen, R. Sheldon, P. Dunn, R. Howard, R. Kumar, I. Martínez and J. Wong, *Green Chemistry*, 2011, **13**, 1791–1798.

[224] J. W. Comerford, T. J. Farmer, D. J. Macquarrie, S. W. Breeden and J. H. Clark, *ARKIVOC*, 2009, **7**, 282–293.

[225] A. Maraite, P. Hoyos, J. D. Carballeira, Á. C. Cabrera, M. B. Ansorge-Schumacher and A. R. Alcántara, *Journal of Molecular Catalysis B: Enzymatic*, 2013, 210–217.

[226] Y. Cao, Y. Zhuang, C. Yao, B. Wu and B. He, *Biochemical Engineering Journal*, 2012, **64**, 55–60.

[227] M. D. Truppo and N. J. Turner, *Organic & Biomolecular Chemistry*, 2010, **8**, 1280–1283.

[228] D. Koszelewski, I. Lavandera, D. Clay, D. Rozzell and W. Kroutil, *Advanced Synthesis & Catalysis*, 2008, **350**, 2761–2766.

[229] B. Yuan, A. Page, C. P. Worrall, F. Escalettes, S. C. Willies, J. J. McDouall, N. J. Turner and J. Clayden, *Angewandte Chemie*, 2010, **122**, 7164–7167.

[230] N. Doukyu and R. Aono, *Applied & Environmental Microbiology*, 1998, **64**, 1929–1932.

[231] P. Estrada, W. Baroto, M. P. Castillon, C. Acebal and L. R. Arche, *Journal of Chemical Technology & Biotechnology*, 1993, **56**, 59–65.

[232] M. Bona, M. Fabian and M. Sedlák, *Biochimica et Biophysica Acta*, 1990, **1020**, 94 – 100.

[233] E. Skrika-Alexopoulos, J. Muir and R. B. Freedman, *Biotechnology & Bioengineering*, 1993, **41**, 894–899.

[234] J. A. Morgan and D. S. Clark, *Biotechnology & Bioengineering*, 2004, **85**, 456–459.