CHALMERS
UNIVERSITY OF TECHNOLOGY

Optimizing Compound Library Synthesis in Drug Discovery using Empirical Modelling Techniques

Application to greener solvent selection for amide coupling reactions

Master's thesis in Innovative and Sustainable Chemical Engineering at AstraZeneca

LINUS KARLING

Department of Chemistry and Chemical Engineering

Chalmers University of Technology
Gothenburg, Sweden 2022
www.chalmers.se

Master thesis KBTX12

Optimizing Compound Library Synthesis in Drug Discovery using Empirical Modelling Techniques

Application to greener solvent selection for amide coupling reactions

LINUS KARLING

CHALMERS
UNIVERSITY OF TECHNOLOGY

Institution of Chemical Engineering
Department of Chemistry and Chemical Engineering
Chalmers University
Gothenburg, Sweden 2022

Optimizing Compound Library Synthesis in Drug Discovery using Empirical Modelling Techniques
Application to greener solvent selection for amide coupling reactions
Master's thesis in Chemistry and Chemical Engineering
Linus Karling

AstraZeneca supervisors:
Michael Kossenjans
Lisa A. Holmes
Nicholas Holmes
Examiner:
Gunnar Westman, Associate Professor in Chemistry and Biochemistry

Master thesis at the department of Chemistry and Chemical Engineering.
Chalmers University
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2022

Abstract

In this master thesis project a methodological workflow for solvent selection was established. The method was applied to find a green solvent for amide coupling reactions with the aim of substituting DMF which will be restricted from December 2023 due to health hazards. The approach was to use High-Throughput experiments (HTE) in combination with Design of Experiments (DoE) and Principal Component Analysis (PCA) to effectively explore the chemical space while maintaining a methodological workflow. The results show that the combination of these techniques is a powerful method to cover a wide chemical space with few experiments and propylene carbonate was selected as alternative to DMF for amide coupling. The performance of both solvents were comparable when tested on a small amide coupling library.

Acknowledgements

I would like to express my deepest gratitude to Lisa A. Holmes and Nicholas Holmes for their invaluable patience, feedback, for all the time put in to answer questions and for forming the work. I would also like to say a special thanks to Micheal Kossenjans for providing me with the opportunity to do this master thesis project. A big thanks to Gunnar Westman for being my examiner. I am also grateful for Tina Seifert, Maria Johansson, Tove Slagbrand and Sara Phalén which have all helped me in the lab. Thanks should also go to Tierry Kogej for his attempts in deriving predictive models for my project. I would be remiss in not mentioning David Hose for his help in personally explaining his research more in-depth. Lastly I would like to thank all my colleagues at the AstraZeneca iLAB.

Contents

1 Background

In this section an introduction to the scope of this thesis is presented, to demonstrate the significance of the work and how it aligns to current challenges in green chemistry and drug discovery.

1.1 Green Chemistry

Green chemistry is a response to the effects some chemicals and the chemical industry has had on public health, the climate and biodiversity. In an effort to tackle these issues the American Chemical Society set up the Green Chemistry Institute (ACS GCI) and promoted The 12 Principles of Green Chemistry [1]. The 12 principles are presented in Figure 1.

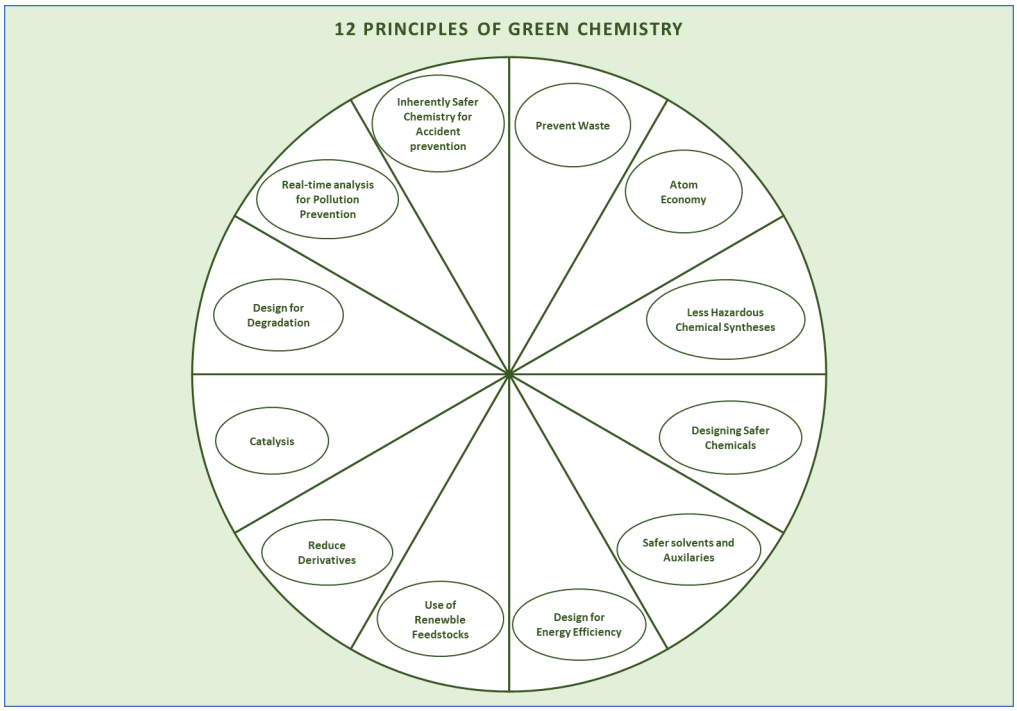


Figure 1: The 12 principles of Green Chemistry. Authors own picture. CC-BY 2.0

These principles were created to highlight areas where significant improvements should be made to develop greener chemistry, and reduce the negative impacts on health and the environment. Today most chemical industries and corporations have some sustainability plan to reduce the climate impact of their businesses and incorporate the green chemistry principles. AstraZeneca has set targets to reduce their Green House Gas (GHG) emission from their own operations (site and fleet) by 98% until end of year 2026. AstraZeneca has also set targets to contribute to the Sustainable Development Goals (SDG) which overlaps with many of the green

principles [2] [3]. Among the principles are *Less Hazardous Chemicals Syntheses* and *Design for Energy Efficiency* which are heavily correlated with the topic of this thesis.

1.1.1 Safer solvents and Auxiliaries

Besides the hazards to people and the environment there are more reasons to choose some solvents over others. To evaluate the "greenness" of a solvent the most common criteria for which solvents are ranked for include: *Waste management*, *Environmental impact*, *Health*, *Flammability and Explosion risk*, *Reactivity and stability* [4]. Commonly these parameters are scored and if one solvent scores better than another it is said to be "greener".

At many companies it is encouraged to use these greener solvents over traditional ones for the sake of health and environment but also since legislation may take away the accessibility to use a certain solvent, it is necessary to transition into the use of another to avoid an abrupt change which might affect performance of production and synthesis. However, for well established chemistry it can be difficult to justify the time and resources required to investigate new, greener solvents and methodologies. Moreover, switching to a green solvent which may perform less well in terms of yield can lead to sub-optimization due to less efficient use of substrates and a more energy and material intensive work up process [5].

1.1.2 Global Warming Potential of the Pharmaceutical industry

According to the American Chemical Society Green Chemistry Institute Pharmaceutical Roundtable (ACS GCIPR) the process mass intensity (PMI) for materials used to manufacture an active pharmaceutical ingredient (API) consisted of 56% solvent and 32% water in 2008. The distribution of the PMI is shown in Figure 2

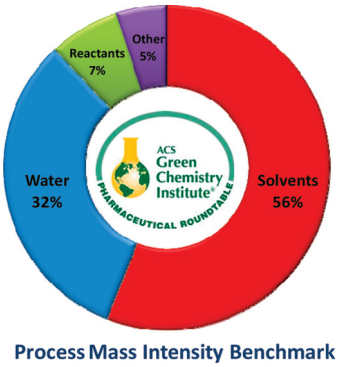


Figure 2: PMI for production of an API. From *Using the Right Green Yardstick: Why Process Mass Intensity Is Used in the Pharmaceutical Industry To Drive More Sustainable Processes* Reprinted with permission from Org. Process Res. Dev. 2011, 15, 4, 912–917. Copyright (2011) American Chemical Society.

PMI is calculated according to Equation 1.

$$process\ mass\ intensity\ (PMI) = \frac{total\ mass\ in\ process\ or\ process\ step\ (kg)}{mass\ of\ product\ (kg)} \quad (1)$$

PMI is a good measurement for the GWP of the pharmaceutical industry because the major contributing category to GWP is raw materials used in the process [5]. According to a study from 2004 where a cradle-to-gate life cycle assessment (LCA) was made on the GlaxoSmithKline's pharmaceutical products, solvents contributed the most towards the environmental footprint in all the API's that were examined [6].

1.1.3 Classification of Hazardous Substances

The European Chemicals Agency (ECHA) has been adding substances to their *Substances of Very High Concern Candidate list* (SVHC Candidate List) of hazardous substances in order to protect workers, consumers and the public of risks associated with chemicals. These risk includes toxicity to reproduction, endocrine disrupting properties and carcinogenic properties, among others. The purpose of the list is to inform industry that the compounds may become restricted in the near future. Among these compounds included on the list are solvents which are used in the chemical industry for production and synthesis of compounds and products. An included solvent is *N,N-Dimethylformamide* (DMF) which is used extensively as a solvent in the pharmaceutical industry but will be restricted from December 2023 [7]. The reason DMF is being restricted is due to its toxicity to the unborn child and potential damage to fertility [8].

1.2 Drug Discovery Process

The pharmaceutical industry is a heavily controlled industry with legislation and rules implemented to ensure safety of pharmaceutical products for the public [9]. This has led to a methodological way of conducting research in drug discovery science which is shared among companies across the globe. Typically research is done to asses the nature of the disease, this is called target validation and its aim is to find a target for the drug to interact with [10]. This could for example be a receptor or an enzyme. After the target validation, a cycle of research follows called the DMTA cycle which stands for Design, Make, Test, Analyze [11]. A schematic workflow of the DMTA cycle is presented in Figure 3.

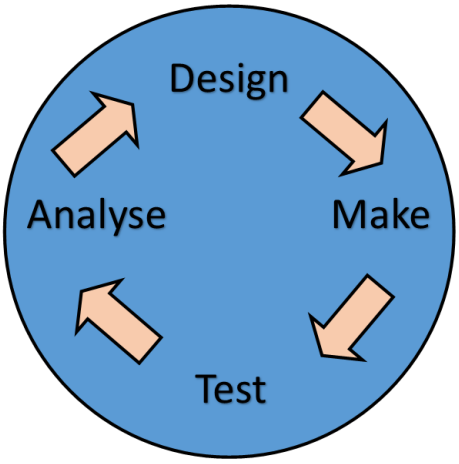


Figure 3: A schematic figure of the DMTA workflow in drug discovery research. Authors own picture. CC-BY 2.0

In the first design iteration an assay of compounds are designed under the hypothesis that they could with interact with drug target. After the assay is designed the next step in the cycle is make. Since the compounds are designed to interact with the same target they are often alike in terms of chemical structure and an efficient way of making the compounds is therefore library synthesis. Library synthesis can be done via high-throughput experiments which is described in section 1.3.1. The next step in the cycle is to test. Depending on how far into the DMTA cycle the research has progressed different types of testing are in question. Early in the discovery process *in vitro* testing is the predominant type. *In vitro* means testing in glass and refers to testing on for example cell cultures, enzymes or tissues [10]. The last step of the cycle is to analyse. This refers to analysing which compounds had a desired effect of bonding to the target, a compound bearing this property is called a *hit*. The results from this analysis is used in the next iteration of DMTA cycle to design a new, hopefully more target accurate, assay of compounds.

The cycle of research boils down to a compound which shows great potential as a clinical candidate called a *lead*. However the requirements of a *lead* compound is not only its chemical activity but also related to bio-chemical stability, optimization potential and patentability [12]. The next step is to optimize the *lead*, which is also part of the DMTA cycle, in terms of previously mentioned prerequisites. Since the following steps are far more costly, measures are taken at this stage to perfect the clinical candidate, such as reducing toxicity, making the compound selective to its dedicated target and achieving high potency [13]. Once a lead compound is optimized it can be suitable for *in vivo* testing, which means animal testing. This stage is still part of the DMTA cycle as tweaks to the compound could be needed if the compound for example shows no effect. If the drug candidate shows good results in the *in vivo* testing it could move into clinical trials where drug discovery

stops and process development and scale up begins.

In the scale-up considerations can be taken to how the substance was synthesised in lab scale but the synthetic route generally differs from the route conducted in lab since more considerations to GWP of substrates and toxicity substances in trace amounts needs to be taken. [14]. In the scale up process it is also easier, compared to library synthesis, to optimize the reaction conditions and yield since by this stage, the process chemists can focus solely on a single compound [15].

1.3 Optimization methods

In this section different tools for optimization are presented, with a description of when they are commonly used in the drug discovery process.

1.3.1 High-throughput experiments

One simple way to reduce the GWP in the research area of the pharmaceutical industry is to reduce the scale of which the chemical reactions are done and this can be achieved by utilizing the technique of High-throughput experiments (HTE). HTE can be described as running several reactions in parallel. With special equipment small scale reactions can be run in parallel on a "block" which holds several reaction vessels. Combined with automation, such as automated dispensing of substrates and solvents, this can give a time and resource efficient research technique where a broad chemical space can be examined within a limited amount of time. This research method has increased in use greatly in the past decade and is used widely in the pharmaceutical industry today [16].

1.3.2 Design of Experiments

Design of Experiments (DoE) is an experimental approach for achieving the most available data with as few experiments as possible. The method utilizes multiple linear regression (MLR), or Partial Least Squares (PLS) regression, to fit a model to the experimental data along with analysis of variance (ANOVA) to analyze the results of the experiments and determine which variables had an significant effect on the outcome.

1.3.2.1 Multiple linear regression

When analysing experimental data sets MLR is an effective way to model the response in terms of the input variables, MLR can be described by Equation 2

$$y(i) = \beta_0 + \beta_1 x_1(i) + \beta_2 x_2(i) \dots \beta_n x_n(i) + \epsilon(i) \tag{2}$$

where $y(i)$ is the response, $x_p(i)$ are the independent variables, β_p are the parameters and $\epsilon(i)$ the error term. The equation can be written on matrix form according to Equation 3

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon \tag{3}$$

when expanded is described by Equation 4

$$\begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(n) \end{bmatrix} = \begin{bmatrix} 1 & x_1(1) & x_2(1) & \dots & x_p(1) \\ 1 & x_1(2) & x_2(2) & \dots & x_p(2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_1(n) & x_2(n) & \dots & x_p(n) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \tag{4}$$

Where X is the design matrix containing, on each row, the individual factor settings, interaction factor setting and higher order factor settings for the n^{th} experiment, Y the vector containing the response on each row, β the vector containing the parameters for the factors on each row and ϵ containing the error on each row [17]. To calculate the parameter vector, β, the matrix operations in presented in Equation 5 can be used.

$$\beta = inv(X^T X)X^T Y \tag{5}$$

Although a value for each $\beta(i)$ is obtained from Equation 5 it is not certain that the variable is of statistical significance, to determine this ANOVA is needed.

1.3.2.2 ANOVA

A fundamental of measure of how well a model fits a data set is sum of square errors (SSE) which is defined as

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{6}$$

where y_i is the response and $\hat{y}_i$ is the predicted response. To calculate the estimated variance, s^2, in the data the SSE is divided by the degrees of freedom according to Equation 7

$$s^2 = \frac{SSE}{n-p} \tag{7}$$

where n is the number of data points and p the number of parameters. The standard error of the i_{th} parameter,$se(b_i)$ is calculated according to Equation 8

$$se(b_i) = s\sqrt{\{(X'X)\}_{ii}^{-1}} \tag{8}$$

where $\{(X'X)\}_{ii}^{-1}$ is the i_{th} diagonal term of the matrix $(X'X)^{-1}$ and s the standard error. Lastly to determine the significance of the parameter the standard error can be compared to confidence interval for the parameter according to Equation 9

$$b_i \pm se(b_i)t(n-p); \frac{\alpha}{2} \tag{9}$$

where $t(n-p); \frac{\alpha}{2}$ is the student t-distribution for $n-p$ degrees of freedom at significance level α. If the standard error lies within the confidence interval of the parameter the parameter is statistically significant. If the standard error lies outside the confidence interval the parameter should be removed and the model refitted and ANOVA applied for the new model again.

A common measure of how well a model fits the data is the coefficient of determination R^2 [18]. To calculate R^2 Equation 10 is used

$$R^2 = 1 - \frac{SSE}{SS_{tot}} \tag{10}$$

where SS_{tot} is the total sum of squares defined as:

$$SS_{tot} = \sum_{i=1}^{n} (y(i) - \bar{y}) \tag{11}$$

where $\bar{y}$ is the mean response. An $R^2 = 1$ is achieved when $SSE = 0$ and corresponds to a model without any error [17].

1.3.2.3 Advantages of DoE

To use DoE one must plan the experiments in detail before execution. Common experimental plans include full factorial designs, fractional factorial designs and central composite designs. Factorial designs are suitable for screening experiments with few independent variables. This is due to the fact that no quadratic effects can be estimated and that the number of experiments needed are equal to 2^p where p is the number of independent variables. For optimization designs a central composite design may be used which includes more than 3 levels for each independent variable to estimate quadratic effects [18]. Compared to the traditional and straight forward *one-variable-at-a-time* (OVAT) experimental approach, the method of DoE is superior in terms of resource efficiency and retrieving information of the underlying process [19]. In Figure 4 an example of an optimization problem with two variables comparing the two approaches are presented.

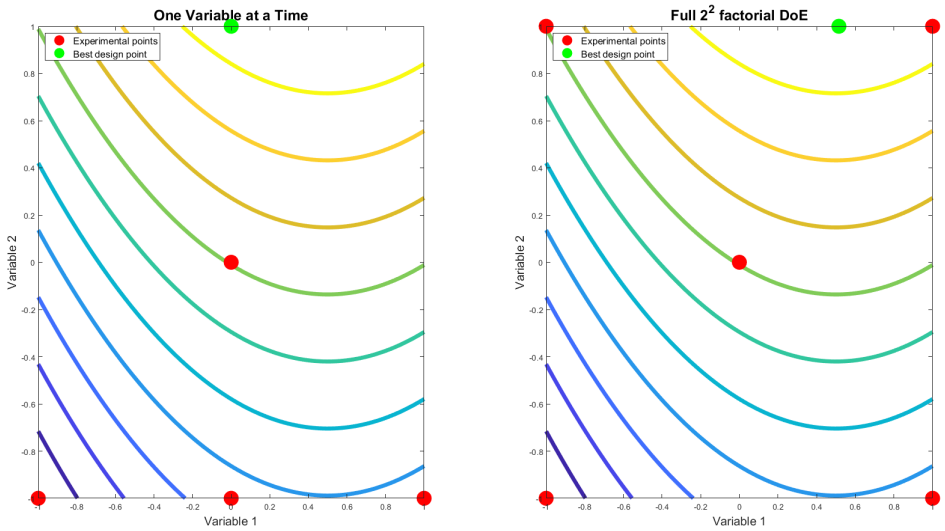


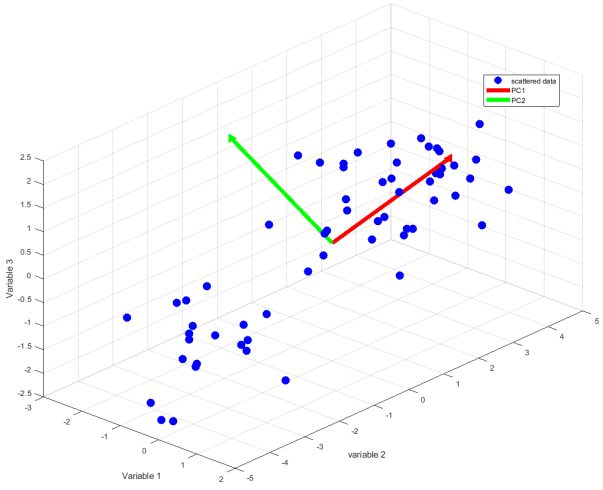
Figure 4: OVAT (left) v.s. DoE full factorial approach (right) for a two variable optimization problem, visualizing the response through a contour plot. Authors own picture. CC-BY 2.0

As seen in Figure 4 (left) using OVAT approach there is a possibility to end up at a local maximum design point, represented as the top green experimental point, interpreted as the global maximum by the experimenter. When employing the DoE approach, Figure 4 (right), multiple linear regression (MLR) together with ANOVA will reveal the real character of the surface and the experimenter can pinpoint the best design point (green).

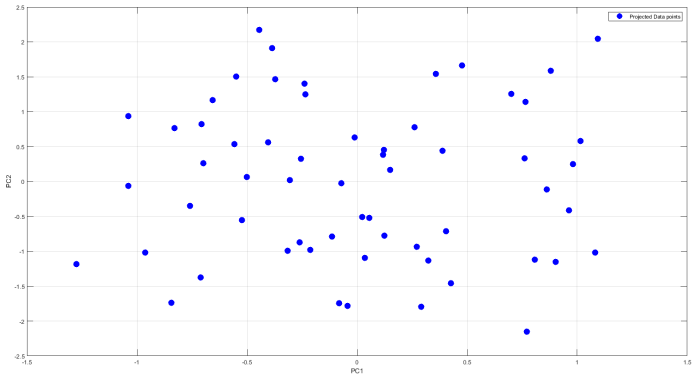
The Food and Drug Administration (FDA) suggested in the beginning of the millennium that pharmaceutical products should be a quality by design product rather than a quality by testing product. As of today DoE is used frequently in the production chain and scale up process of pharmaceuticals but this can be considered a rather slow adoption acknowledging that DoE has been around since the 19th century [19] [20]. However, it is less frequently used in early drug discovery, despite HTE equipment being set-up able to handle rapid screening of experimental conditions.

1.3.3 Principal component analysis

PCA is a mathematical technique that reduces the dimension of a data set by performing linear regressions, orthogonal projections and linear transformations. The outcome is a new continuous coordinate system which describes the variation in the data with some lack of fit. Another advantage of the method is that it is possible to do a PCA model on a non continuous data set and the outcome will still be a continuous coordinate system spanned by the principal components [21]. PCA is used in the pharmaceutical industry for example in analysis of spectroscopy data [22]. In Figure 5 an example of a dimension reduction from $R^3 \rightarrow R^2$ is shown.



(a) Data scattered in R^3 along with the vectors of PC1 and PC2.



(b) Data projected onto PC1 and PC2 and transformed to R^2.

Figure 5: Example of an R^3 data set along with PC1 and PC2 (a), and the reduced and transformed data set in R^2(b). Authors own picture. CC-BY 2.0

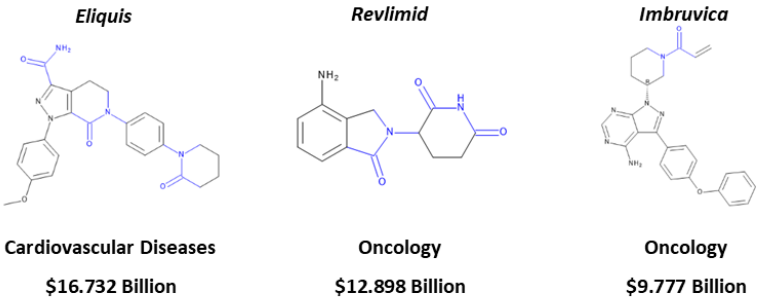
In Figure 5a a R^3 data set is scattered along with the Principal component 1, which is the vector which spans across the most variation in the data-set. The vector, $w_{(1)}$, is found by solving Equation 12

$$w_{(1)} = argmax(\frac{w^T X^T X w}{w^T w}) \tag{12}$$

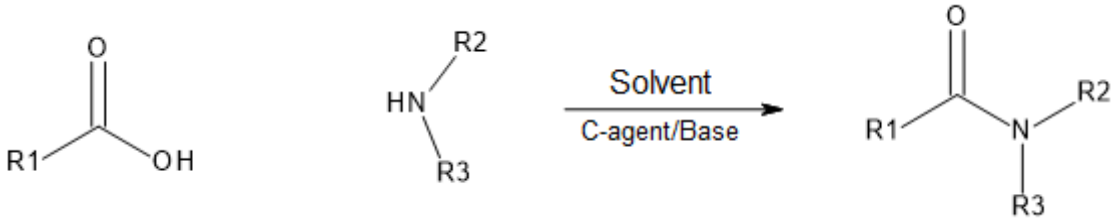
where X is the data set. The w that full fills this criteria is the first eigenvector of the matrix $X^T X$. The second principle component, PC2, is the second eigenvector of the same matrix. The transformed coordinate of the data point onto PC1, $t_1(i)$, are then given by $t_1(i) = x_{(i)} \cdot w_{(1)}$ and similarly for PC2 but with vector $w_{(2)}$ [23]. The transformed system is then effectively reduced from R^3 to R^2 and the transformed system is shown in Figure 5b.

1.4 Amide coupling reactions

Amide couplings are a fundamental chemical reaction that occurs in nature in the form of peptide bonds which links the amino acids to form proteins.It is the chemical reaction that links the amino acids that form the DNA of all living organisms. Moreover it is present several top selling pharmaceuticals as of 2021. The generic amide coupling reaction scheme as well as three out of the ten most sold pharmaceuticals of 2021, *Eliquis* and *Revlimid* and *Imbruvica*, with the amide bond highlighted in blue, are presented in Figure 6 [24].



(a) Three of the ten most sold brand name pharmaceuticals in 2021, with amide bonds highlighted in blue, including their name, therapeutic use and total retail sales.



(b) A carboxylic acid (left) reacting with an amine (middle), in solvent, coupling agent (c-agent) and base, to form a amide (right) [24]

Figure 6: Three top selling brand name pharmaceuticals as of 2021 (a) as well as the generic amide coupling reaction (b) [24].

As seen in Figure 6 the generic amide coupling reaction consists of a condensation reaction of a carboxylic acid with an amine, which could be primary, secondary or of anilinic character, to form an amide. When the reaction is carried out in a lab environment a base is usually added to activate the acid and a coupling agent to drive the reaction.

The gold standard solvent and coupling agent for small molecule amide couplings at AstraZeneca is *DMF* and *HATU*. As previously mentioned *DMF* will be restricted

from December 2023 due to its toxic properties and *HATU* can form explosive byproducts during reaction. They both therefore can be said to directly contradict the green chemistry principle of *Less Hazardous Chemicals Syntheses*.

2 Aim

In this master thesis is a methodological workflow will be established for modelling compound library synthesis in order to optimize reaction conditions for a high average yield. The method will be applied to amide coupling reactions with the aim of finding a green solvent substitute for *DMF*, while avoiding *HATU* as coupling agent in the essence of green chemistry. This will be done by first doing a screening of viable amide coupling agent options. Next a solvent screening will be done using the empirical modelling techniques PCA and DoE in synergy with HTE. Then optimization design will be carried out using DoE in order to find the optimal reaction conditions for the green solvent. Lastly the optimal reaction conditions for the green solvent will be tested on a amide coupling library, and compared against *DMF* and *HATU*.

3 Method

In this section the general experimental procedure, analysis method and design of experiments will be presented along with the workflow. The workflow can be split into four stages. First viable coupling agent options were evaluated for a single reaction, with the reactions carried out in *DMF* as a baseline. Next solvent options were examined for a small library with the help of PCA and DoE as modelling techniques. Then the reaction conditions were optimized using another round of DoE. Lastly the optimal reaction conditions were tested for a small novel amide coupling library.

3.1 General Experimental Procedure

The general experimental procedure is described in a schematic figure with the different steps and automation systems presented in Figure 7.

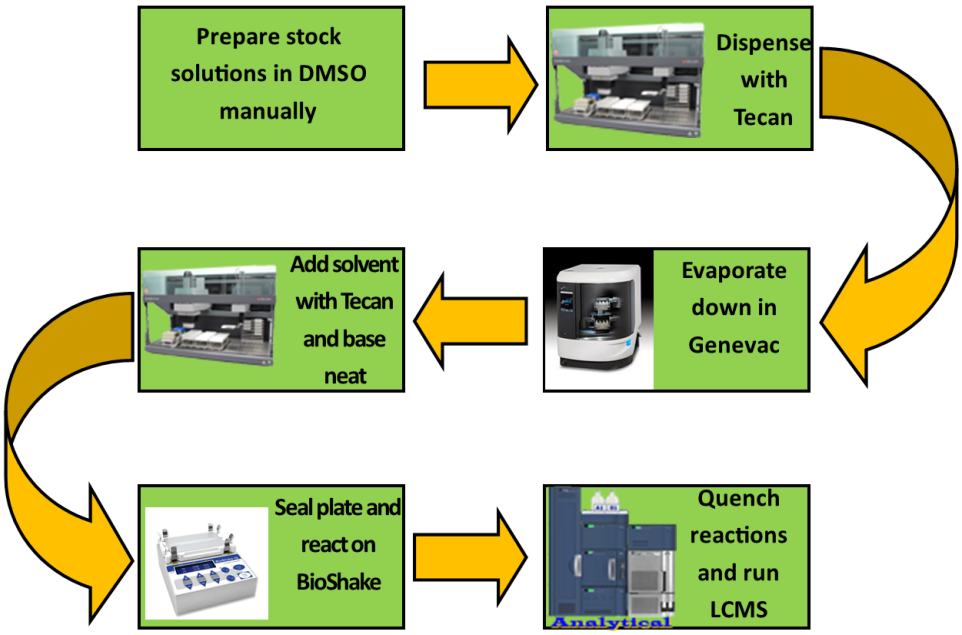


Figure 7: A schematic of the step-by-step workflow throughout the project.

Acid, amine and coupling agent were weighed out into separate 4ml High-recovery vials. Then Dimethylsulfoxide (DMSO) was added manually with *Eppendorf* pipettes to a reach a specified stock solution concentration and the vials were shaken by hand to dissolve the compounds. In the case of solubility issues the vials were submerged into the *Bandelin SONOREX* ultrasonic bath until the compounds had dissolved. Next acid, amine and coupling agent were dispensed, in accordance with the experimental plan to reach desired concentrations, with the automated liquid handler *Tecan Freedom EVO* into the wells of a *Greiner-bio-one Microplate 96 well v-bottom (335µL)*. The plate(s) were then evaporated down with the vaccum centrifuge *Genevac SP Scientific*. Next the solvent desired for the reaction was added with the automated liquid handler *Tecan Freedom EVO*. The base was added manually with a multichannel *Eppendorf* pipette due to lower liquid volume limitations of the *Tecan Fredom EVO*. The plates were then sealed with plastic film with the *Agilent PlateLoc Seal*. Lastly, the plates were placed on the agitator and hotplate *BioShake IQ*, at agitation of 700 RMP, for the desired reaction time and temperature.

When the desired reaction time had passed the plates were un-sealed by hand and the reactions were quenched simultaneously as analysis plates were prepared by adding 20µL of reaction mix, 90µL of DMSO and 2µL of formic acid with the

Tecan Freedom EVO into a fresh *Greiner-bio-one Microplate 96 well v-bottom (335µL)*. The samples could then be analysed through HPLC-MS and HPLC-UV.

3.2 Analysis method

All samples were analysed by reversed phase chromatography using an *ACQUITY-Ultra Performance LC* instrument, equipped with an *HSS C18 column (1.8 µm particle size, 1 µL injection volume)*, a *Waters PDA detector (254 nm)* and a *Waters 3100 Mass detector*. All samples were analysed using the following method: H_2O (pH3)/MeCN (pH3) 90:10 to 10:90 over 2 min, 1 mL min^{-1}.

In all reactions carried out through this project the acid was the limiting substrate. Due to the acid being the limiting substrate the conversion of the reaction was calculated according to Equation 13

$$Conversion = \frac{A_{Amide}}{A_{Acid} + A_{Amide}} \tag{13}$$

where A_{Amide} is the UV-peak area for the amide and A_{Acid} is the UV-peak area for the acid. Equation 13is valid under two assumptions, the first one being that $\epsilon_{Amide} = \epsilon_{Acid}$ where ϵ is the molar absorption coefficient for the respective compound. The second assumption is that all UV-peaks in the chromatogram are separated indicating that no co-elution of compounds occurs.

3.3 Design of Experiments and analysis of experimental data

All design plans, fitting of data, and ANOVA were done with *MODDE 12*. This includes the generation of design plans, MLR of experimental data as well as ANOVA. The D-optimal design plan was also generated by an algorithm within the software. The ANOVA carried out includes, t-test of fitted parameters, calculation of R^2, Q^2, Lof-tests and creation of contour plots. For finding optimal conditions for derived models the add-in Solver in *Excel* was used utilizing the *GRG Non Linear* method.

3.4 Coupling Agent Screening

Four different coupling agents were tested for a single reaction with the general experimental procedure described in section 3.1. The coupling agents as well as their structures are presented in Figure 8.

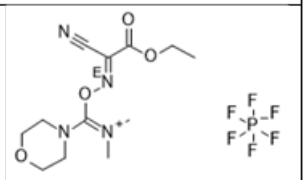
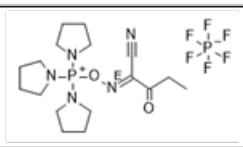
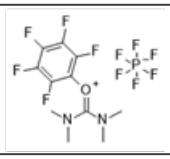
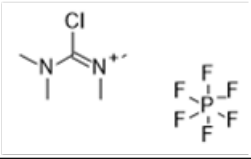
Coupling Agent	Structure
COMU	
PyOxim	
PFTU	
TCFH	

Figure 8: The structure of the coupling agents and their acronyms screened in this study.

The base used in all the reactions was *DIPEA* with the exception of the reaction with *TCFH* as coupling agent where *1-Methylimidazole* was used instead. The reaction conditions are presented in Table 1

Reaction condition	**Value**
Reaction volume [µL]	200
Acid concentration [mM]	10
Coupling agent Equivalents	2
Base Equivalents	4
Amine Equivalents	1.5
Temperature[°C]	25
Reaction time [h]	18

Table 1: Reaction conditions used for the coupling agent screening.

The reactions were then analyzed according to the analysis method described in Section 3.2.

3.5 Solvent screening

For the solvent screening a open-access PCA model for solvents provided by ACS GCIPR was used [25]. The model was created based on 18 physical properties and contains 272 solvents. Three principal components, PC1, PC2 and PC3, were used

as independent variables in a full factorial design. The intervals for the independent variables were derived by a custom *MATLAB (R2020b)* script with the aim of finding the most orthogonal set of design points available, while maintaining a useful span of the independent variables. To achieve this an algorithm was created that for a fixed center point, which was the solvent closest to origin, a set of loop calculations was done. In each loop a span for each PC was set where for each span an fully orthogonal box was created within the PCA space. Next the solvents closest to each corner point of the fully orthogonal cube was chosen as a candidate. Then the sum of distances to each corner point, named *skewness*, was calculated and compiled in a list. When all loop calculations were done a list containing the *skewness* for each calculated combination of spans was obtained and the results could then be plotted for evaluation of *skewness* vs span.

The solvents were tested for 8 different reactions with the general experimental procedure and the reaction conditions presented in Table 2.

Condition	**Value**
Reaction Volume [µL]	200
Acid Concentration [mM]	10
Acid Equivalent	1
Amine Equivalents	1.5
Coupling Agent Equivalents	2
Base Equivalents	5.7
Reaction time [h]	18
Reaction temperature [°C]	25

Table 2: Reaction conditions for solvent screen.

3.6 Reaction conditions optimization

To optimize the reaction conditions a D-optimal design was used. The D-optimal design was generated in *MODDE 12* and the variables examined were: reaction temperature, acid concentration, coupling agent equivalents and base equivalents. The reaction intervals for the parameters are presented in Table 3.

Condition	**Low level**	**High level**	**Center level**
Temperature [°C]	25	50	37.5
Acid concentration [mM]	0.0075	0.0125	0.01
Coupling agent Equivalents	1	3	2
Base Equivalents	3	9	6

Table 3: Variable intervals used for the D-optimal design.

The reaction conditions were tested for 7 different acid substrates with the general experimental procedure and the reaction conditions that were kept constant presented

in Table 4.

Condition	Value
Acid Equivalent	1
Amine Equivalent	1.5
Reaction Time [h]	18

Table 4: Constant reaction conditions for the D-optimal design

The derived optimal reaction conditions were then tested for a new amide coupling library. The library consisted of 1 acid substrate and 12 amine substrates, due to the substrates being proprietary compounds their structures can not be presented in this report.

4 Results and discussion

In this section the results of the coupling agent screening, solvent screening and reaction condition optimization will be presented. A test of the optimal reaction conditions for a novel small library will also be presented and discussed.

4.1 Coupling Agent Screening

As mentioned in the Background, for amide coupling reaction a coupling agent is added to drive the reaction. There are numerous options but the gold standard for AstraZeneca lab workers is HATU with DMF as solvent. HATU does however form explosive byproducts and is a respiratory sensitiser, therefore other coupling agents were screened. In the spirit of green chemistry non-hazardous options that were available in-house, were considered. The coupling agents were tested for the reaction presented in Figure 9.

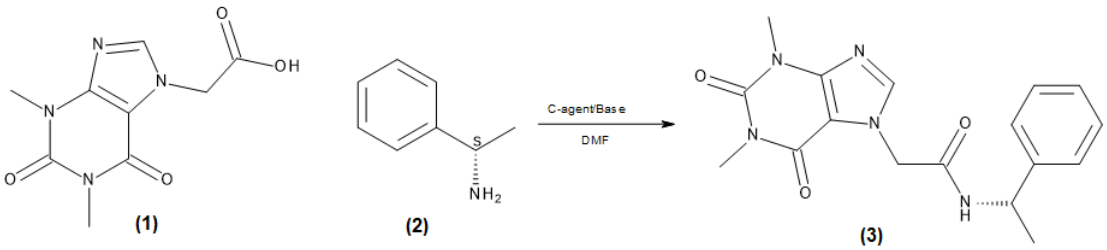


Figure 9: Initial test reaction used for the coupling agent screening. With the acid (1) reacting with the amine (2) to form the amide (3). (C-agent = PFTU, TCFH, COMU or PyOxim. Base = DIPEA or 1-methylimidazole).

The reactions were carried out in *DMF* and with *DIPEA* as base for all coupling agents except *TCFH* for which *1-Methylimidazole* was used since it was suggested in literature to work only with this base [26]. None of the starting material acid

was detected in any of the chromatograms, so it was assumed that full conversion was achieved with each coupling agent. Due to co-elution of by products of the coupling agents TCFH and PyOxim and the product, Equation 13 would have failed to give an accurate result in the case of partial conversion. To investigate this further standards of the coupling agents were prepared in DMSO and analysed using LCMS. The UV chromatograms of these standards are presented in Figure 10.

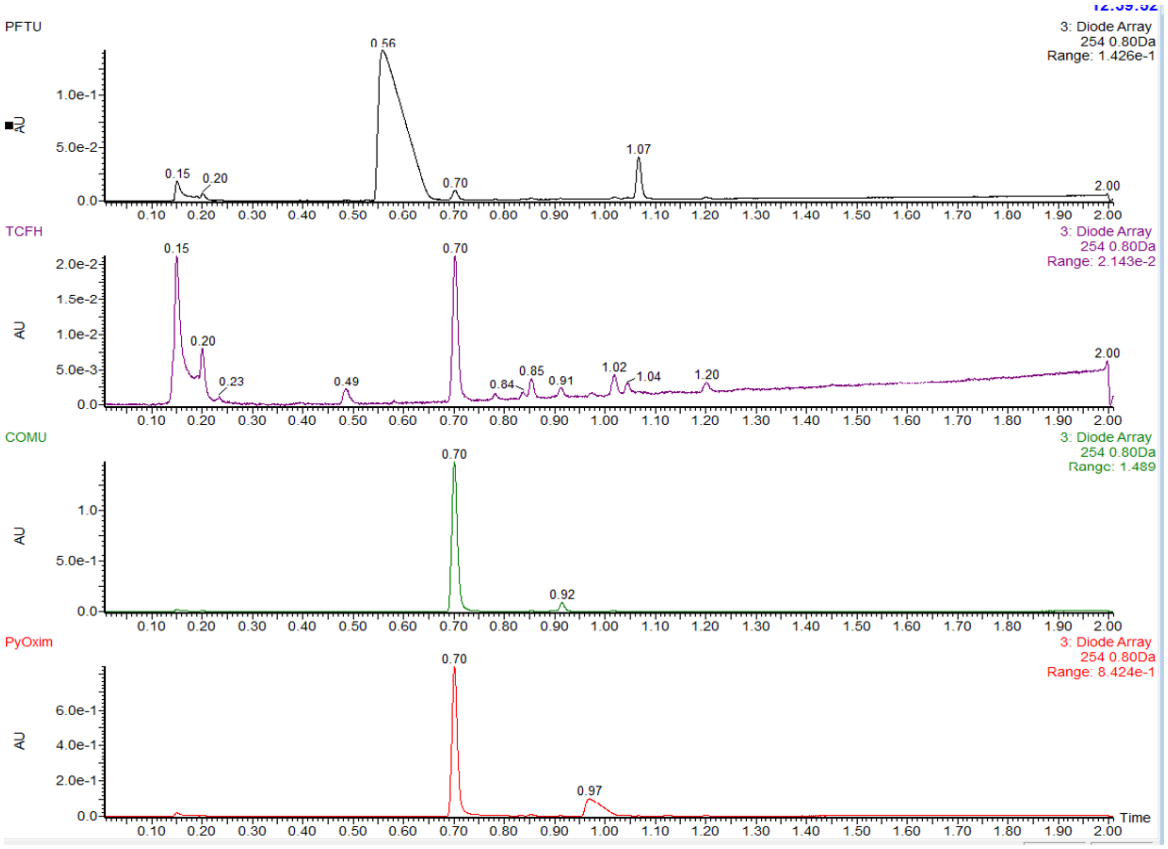


Figure 10: UV-Chromatograms for the four different coupling agents.(Shown from top to bottom: PFTU, TCFH, COMU and PyOxim.)

As seen in Figure 10 COMU and PyOxim showed similar UV-chromatograms. The TCFH had the most peaks and the peak at 0.56 for PFTU was not seen in reaction mixture. The peak causing co-elution with the product was the one at retention time (Rt)=0.7. Based on these results a decision to move on with COMU and PFTU for further testing was made.

To avoid the problem of co-elution in further testing an attempt with another acid substrate was made and the reaction scheme is presented in Figure 11. The reactions were carried out in five different solvents, Dimethyl carbonate (DMC), Ethyl acetate (EtOAc), 2-Methyltetrahydrofuran(2-Me-THF), Isopropyl alcohol (IPA) and DMSO. This was done to establish if there were interaction effects between the type of solvent and type of coupling agent. The five solvents tried were suggested

by literature to work well for amide coupling reactions [27]. The acid new acid was analysed through LCMS beforehand to ensure that it did not co-elute with the coupling agent fragments. The reaction scheme is presented in Figure 11.

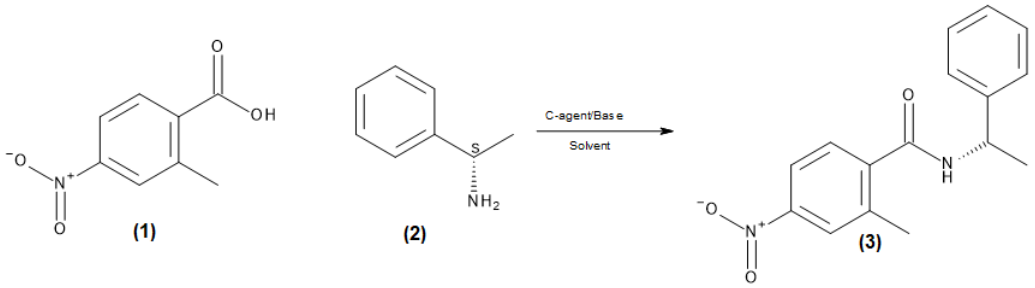


Figure 11: Reaction used for the coupling agent and solvent screening. With the alternative acid substrate (1) reacting with the amine (2) to form amide (3) (C-agent = PyOxim or PFTU, Base = DIPEA, Solvent = DMC, EtOAc, 2-Me-THF, IPA or DMSO.)

PyOxim and PFTU were tested for the reaction presented in Figure 11 with the five different solvents, the results are presented in Table 5

Coupling agent	Solvent	Conversion
PyOxim	DMC	1.00
	EtOAC	1.00
	2-Me-THF	0.97
	IPA	0.85
	DMSO	1.00
PFTU	DMC	1.00
	EtOAC	1.00
	2-Me-THF	1.00
	IPA	0.43
	DMSO	1.00

Table 5: Conversion for the tested solvents with *PFTU* and *PyOXim* as coupling agent respectively.

As seen in Table 5 both PyOXim and PFTU performed well with the combination of PFTU and IPA being the only exception. However, PFTU mediated reactions generated chromatograms with fewer peaks and therefore would have a lower risk of co-eluting with other substrates than PyOxim. Therefore PFTU was selected as coupling agent for the solvent screening.

4.2 Solvent screening

The examine if there were more viable options for choosing a green solvent for amide coupling reactions the use of a PCA model that included 272 different solvents

was used in a full factorial design with the 3 PCs as independent variables. The intention was to discover solvents that was previously not widely discussed in literature. As previously mentioned, in drug discovery science an approach in the early drug discovery research is to run parallel reactions to generate a compound library. The objective in library synthesis is to generate the highest number of products with a yield $>$ 20% since the products can be purified by HPLC. It is of less interest to maximise the yield of a single reaction rather than to get a high average yield. With this reasoning the approach of the solvent screening was therefore to use the average yield of a set of reactions as response in the modelling. The screenings of reaction conditions were done on 3 primary amines, 2 secondary amines, 2 primary anilines and 1 secondary aniline reacting with the acid used in the coupling agent screening. This was done to mimic the distribution among type of amines used in library amide coupling reactions. The amines used in the experiments are presented in Figure 12

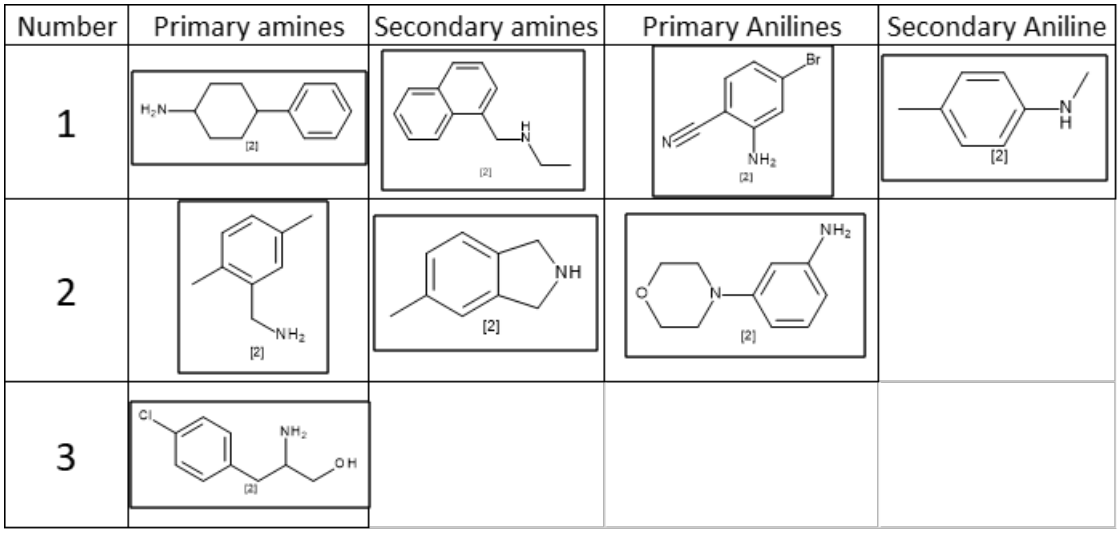


Figure 12: Amine substrates used in the solvent screening and optimization experiments.

Using the MATLAB script the *skeweness* as a function of the span of the PC1 and PC2 could be obtained. However, before deploying the MATLAB script the list of solvents was filtered to remove acids and amines since these would alter the reaction. The *skewness* as a function of the span, in number of standard deviations, for PC1 and PC2 was plotted to help guide the choice. The result can be seen in Figure 13.

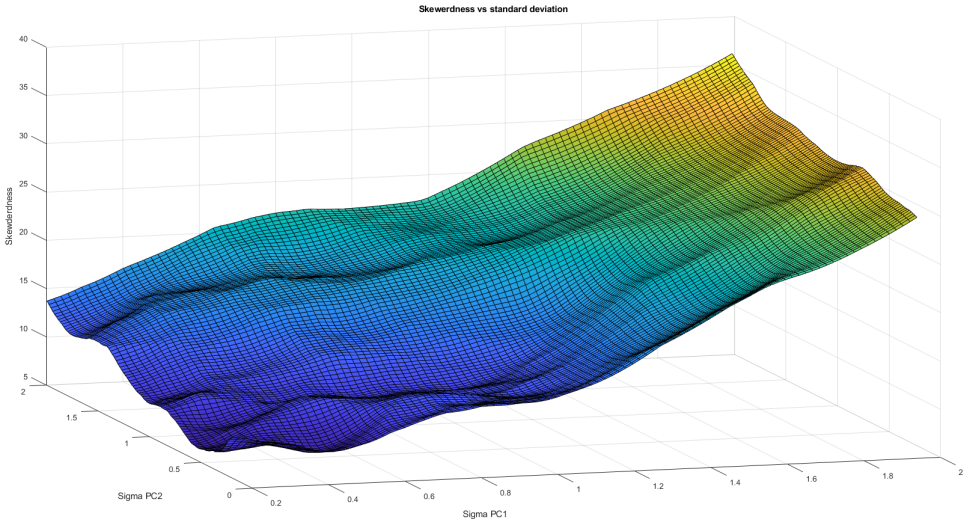


Figure 13: *Skewerdness* as a function of the span (in standard deviations) of PC1 and PC2.

As seen in Figure 13 the *skewness* increased as the span for PC1 and PC2 increased meaning that the choice of span was a balance between covering a large chemical space and having a more orthogonal design. This was because the PCA model was more sparsely occupied by solvents when moving away from the origin. Based on the results seen in Figure 13 and the fact that span of PC3 had little effect on the *skewness* of the cube a choice of intervals for the PC's was made and the intervals for the parameters are presented in Table 6

Parameter	**Low level**	**High level**
PC1	-6.96	7.00
PC2	-4.43	4.77
PC3	-3.07	2.99

Table 6: Intervals for the PC's used to generate a full factorial design plan.

The intervals presented in 6 corresponded to nine solvents, eight corner points and one center point. The solvents chosen as this stage was not restricted to being green as the aim was to model the chemical space within the box rather then to evaluate each solvent. The solvents and their coordinates and ideal normalized coordinates (INC) are presented in Table 7.

Solvent	PC1	PC2	PC3	INC [PC1, PC2, PC3]
Acetonitrile	-4.65	-4.75	-3.28	-1, -1, -1
Cyclohexane	5.48	-5.68	-0.566	1, -1, -1
DMSO	-7.16	4.00	-4.00	-1, 1, -1
Ethanol	-6.84	-3.68	1.12	-1, -1, 1
5-Nonanone	3.79	2.87	0.485	1, 1, -1
n-Heptane	7.46	-4.17	1.01	1, -1, 1
Diethylene Glycol	-7.96	2.83	3.45	-1, 1, 1
Benzyl Benzoate	7.93	5.18	3.83	1, 1, 1
Hexanenitrile	0.642	-0.828	-0.863	0, 0, 0

Table 7: Real coordinates and ideal normalized coordinates (INC) of the most orthogonal solvent selection for the chosen intervals of the PCs.

Due to the previously mentioned fact that there was only 272 solvents in the PCA model the design found was not fully orthogonal. To visualize deviations from the INC the solvents, plotted in the R^3 PCA space, are presented in Figure 14.

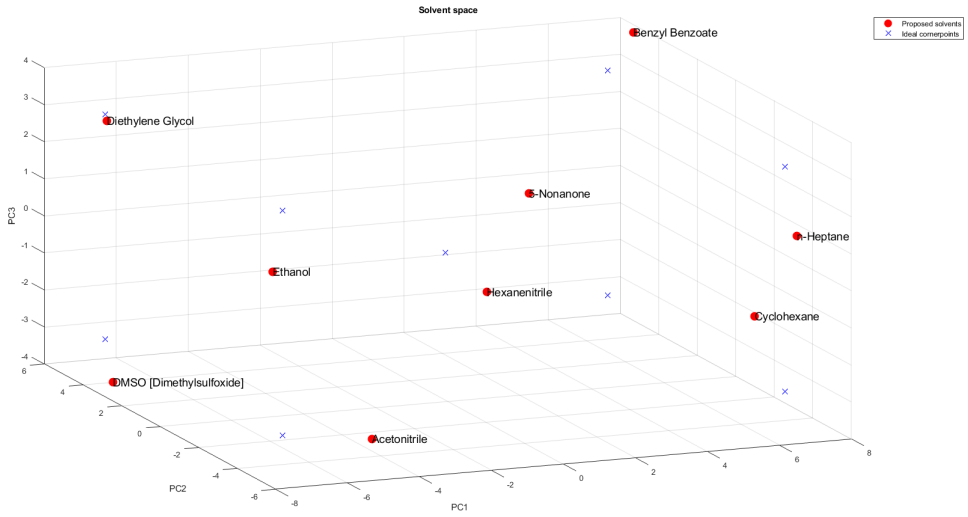


Figure 14: Solvents (red) plotted in the 3-D solvent space along with the ideal orthogonal corner points (blue).

With the nine solvents a full factorial design experiment was made with three center replicates. The experiments were done with the general experimental procedure, with PFTU as coupling agent and DIPEA as base, and the reaction conditions are available in Table 2. The average conversion, as well as the INC, for each solvent are presented in Table 8.

Solvent	**Coordinate [PC1,PC2,PC3]**	X_{avg}
Diethylene glycol	”-1,1,1”	0.25
Ethanol	”-1,-1,1”	0.55
5-nonaone	”1,1-1”	0.22
Acetontirile	”-1,-1,-1”	0.58
Benzyl Benzoate	”1,1,1”	0.19
DMSO	”-1,1,-1”	0.77
Cyclohexane	”1,-1,-1”	0.46
n-heptane	”1,-1,1”	0.09
Hexanenitrile	”0,0,0”	0.55
Hexanenitrile	”0,0,0”	0.45
Hexanenitrile	”0,0,0”	0.55

Table 8: Average conversion (X_{avg}) obtained from the test library solvent screening including the ideal normalized coordinate.

As seen in Table 8 the results from the full factorial design experiment showed that DMSO, Acetonitrile and Ethanol gave the highest average conversion while n-heptane and benzyl benzoate performed poorly. From the results MLR and ANOVA was carried out and revealed that only PC1 and PC3 had a significant effect on the response. The fitted prediction model is presented in Equation 14.

$$\hat{X}_{avg}(PC1,PC3) = 0.4137 - 0.0157PC_1 - 0.0610PC_3 \tag{14}$$

Where $\hat{X}_{avg}$ is the predicted average conversion. The model had an $R^2 = 0.81$. Based on Equation 14 a list of all solvents in the PCA model fulfilling the criteria PC_1,$PC_2 < 0$ was compiled which contained 84 solvents in total. One of the solvents predicted to perform the best was *DMF*. As *DMF* is currently widely used for amide coupling reactions, and performs well, this suggest that the model predictability is valid. The 84 solvents that fulfilled the criteria were then filtered down to include only those solvents that were categorized as green, to provide *Dihyrdolevoglucosenone (Cyrene)*, *Butanone (MEK)* and *Propylene carbonate*. The predicted average conversion for the green solvents and *DMF*, along with their coordinates in the solvent space are presented in Table 9

Solvent	**PC1**	**PC3**	$\hat{X}_{avg}$
Dihyrdolevoglucosenone (Cyrene)	-14.147	-3.010	0.819
Butanone (MEK)	-1.479	-3.009	0.62
Propylne carbonate	-3.564	-2.168	0.602
DMF	-5.08	-3.72	0.73

Table 9: Coordinates and predicted average conversion ($\hat{X}_{avg}$) for the green solvents predicted to perform well in amide couplings, including *DMF* as a comparison to demonstrate model validity.

As seen in Table 9 *Cyrene* has a value of PC1 that lies far outside the modelled

region. However, due to the limited number of green solvents included on the compiled list, it was still included for further testing. The three green solvent candidates were tested with same reaction conditions and test library substrates as for the solvent screening, each with 4 replicates. The results from the experiments are visualized in the form of a predicted vs observed plot presented in Figure 15

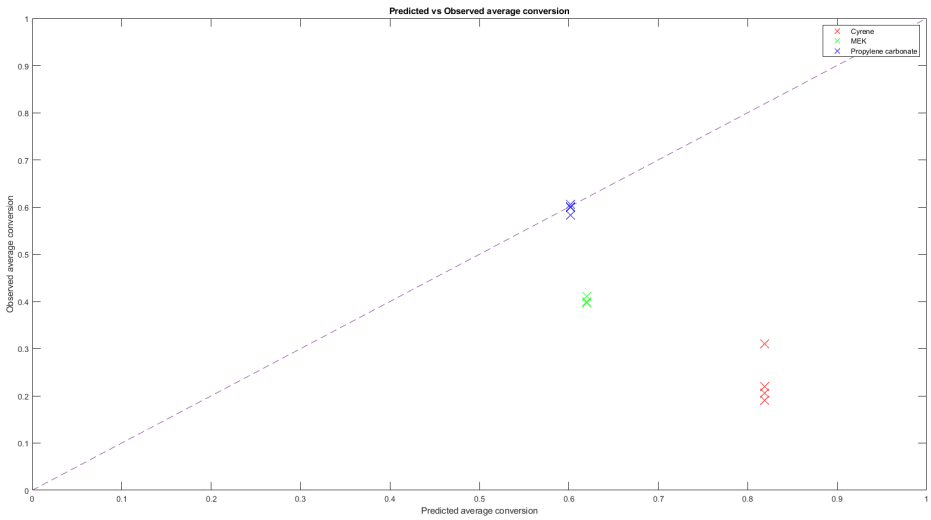


Figure 15: Predicted vs observed average conversion for the test library of amide couplings using *cyrene* (red), *MEK* (green) and *Propylene carbonate* (blue).

As seen in Figure 15 *Cyrene* stray far from the predicted value and could be due to the fact that it was heavily extrapolated. *MEK* did not perform as well as the prediction while the *Propylene carbonate* results correlated well with the predicted value. The approach of using the PCA model in combination with DoE resulted in finding a solvent, *Propylene carbonate*, that has not been widely described in literature as a potential replacement for *DMF*. This was achieved by methodologically selecting 9 out of 272 solvents in the PCA model and conducting a total of 11 experiments with the test library, excluding the experiments to confirm the predicted result. To properly establish *Propylene carbonate* as a replacement for *DMF* it was taken for further testing and optimization of the reaction conditions.

4.3 Reaction Conditions and Optimization Design

After *Propylene carbonate* was selected as the optimal green solvent, an optimization design was carried out to determine the optimal reaction conditions. The variables examined were *reaction temperature*, *acid concentration*, *coupling agent equivalents* and *base equivalents*. Due to the HTE methodology of combining reaction to run in plates the temperature could not be varied across the plate which gave a constraint to the experimental design of having a multilevel parameter in the design. This constraint, in combination with a limitation to three temperature levels, ruled out the option of using a central composite design and a D-optimal design was

deployed instead. As mentioned previously a D-optimal design is a design plan calculated by an algorithm with considerations taken to the constraints. Using the previous reaction conditions as center point a D-optimal design plan was generated with the help of the software *MODDE 12* is presented in Table 10

Exp Name	Temperature [°C]	Acid concentration [mM]	Coupling agent Equivalents	Base Equivalents
N1	25	8	1.00	3
N2	25	13	3.00	3
N3	25	8	3.00	9
N4	25	8	3.00	5
N5	25	13	1.00	5
N6	25	13	2.33	9
N7	25	9	1.00	9
N8	25	11	3.00	3
N9	37.5	13	1.00	3
N10	37.5	8	1.00	9
N11	37.5	13	3.00	9
N12	37.5	8	1.67	3
N13	37.5	10	2.00	6
N14	50	8	3.00	3
N15	50	13	1.00	9
N16	50	8	1.00	7
N17	50	8	2.33	9
N18	50	13	3.00	5
N19	50	13	1.67	3
N20	50	9	1.00	3
N21	50	9	3.00	9
N22	37.5	10	2.00	6
N23	37.5	10	2.00	6
N24	37.5	10	2.00	6

Table 10: D-Optimal design plan for optimization of amide coupling reaction conditions, using *propylene carbonate* as the solvent.

As seen in Table 10 the total number of design point were 24 distributed equally across the three temperature levels $T = 25°C$, $T = 37.5°C$ and $T = 50°C$. The wells at the edges of the V-96 plate were avoided to reduce edge-effects due to heating which limited the number of reactions to 60 reactions per plate. Secondary amine number 2 in Figure 12 was excluded from the design since this had reached full conversion for all reactions previously carried out and would not provide further information. The exclusion was done to be able to fit all reactions into three plates. The experiments were done with the general experimental procedure with reaction conditions from the D-optimal design along with the reaction conditions kept constant available in Table 4. The results of the D-optimal design are presented in Table 11.

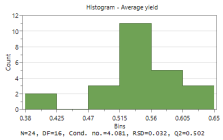
Experiment	X_{Avg}
N1	0.38
N2	0.56
N3	0.54
N4	0.54
N5	0.49
N6	0.58
N7	0.56
N8	0.53
N9	0.42
N10	0.52
N11	0.56
N12	0.56
N13	0.61
N14	0.54
N15	0.48
N16	0.47
N17	0.59
N18	0.55
N19	0.59
N20	0.56
N21	0.53
N22	0.60
N23	0.62
N24	0.62

Table 11: Average conversion (X_{Avg}) achieved from the D-optimal design experiments for the optimization of amide coupling reaction conditions.

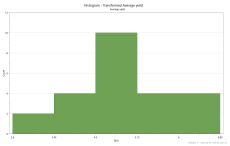
As seen in Table 11 experiment $N13$, $N23$, $N23$ and $N24$ that were the center points showed the best results. The conditions were the same as previous experiments except elevated reaction temperature of $T = 37.5°C$. For the fitting of the data a transform of the conversion was applied, the transform is presented in Equation 15

$$X_{avg}^{Transform} = \frac{1}{(X_{avg} - 1)^2} \tag{15}$$

where X_{avg} is the average conversion of the 7 amines and $X_{avg}^{Transform}$ the transform of the average conversion. The transform was done to achieve a normal distribution in the data in order to perform *ANOVA*. The distribution of the responses before and after the transformation is presented in Figure 16



(a) The distribution of the response in the optimization design (X_{avg}) without any transform applied to it.



(b) The distribution of the response (X_{avg}) in the optimization design with the transform presented in 15 applied to it.

Figure 16: The distribution of responses in the optimization design with no transform applied to it (a) and the transform presented in 15 applied to it.

After the transform was applied to the average conversion MLR and ANOVA was done and the statically significant transformed parameters for the effects are presented in Table 12

Effect	**Transformed effect parameter**
Constant	6.54
Temperature (T)	0.074
Acid concentration (C_A)	0.080
Coupling agent equivalents (Eq_C)	0.428
Base equivalents (Eq_B)	0.114
$C_A \cdot C_A$	-0.844
$Eq_C \cdot Eq_C$	-1.52
$T \cdot Eq_b$	-0.320

Table 12: The effects and their transformed parameters.

As seen in Table 12 the quadratic effect of the equivalents of coupling agent followed by the quadratic effect of the concentration of acid had the greatest impact on the average conversion, both with negative parameters. The full transformed average conversion equation is presented in Equation 16

$$X_{avg}^{Transform} = 6.54 + 0.074T + 0.080C_A + 0.428Eq_C + 0.114Eq_B - 0.844C_A^2 - 1.52{Eq_C}^2 - 0.32T \cdot Eq_B \quad (16)$$

With Equation 16 the optimal conditions within the modeled region were calculated with the add-in Solver in Excel. The best conditions was found for $T = 25°C$ and the second best was found for $T = 50°C$.The conditions are presented in Table 13

Condition	Optimal conditions 1	Optimal conditions 2
Temperature [°C]	25	50
Acid concentration [mM]	10	10
Coupling agent equivalens (PFTU)	2.14	2.14
Base Eq	9	3
$\hat{X}_{avg}$	0.620	0.618

Table 13: The two most optimal set of reaction conditions and their predicted average conversions.

Both the conditions presented in Table 13 was tested with 3 replicates each and the results is visualized in the form of a predicted v.s. observed plot and are presented in Figure 17.

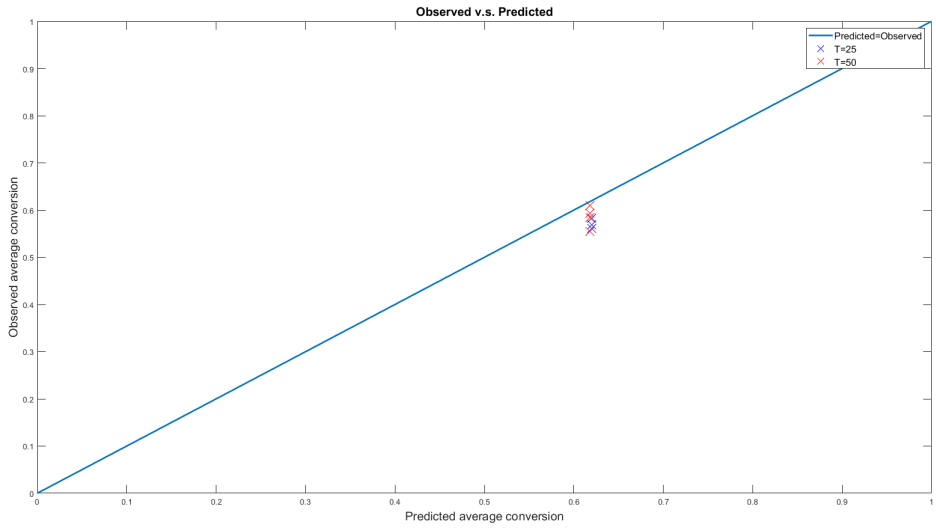


Figure 17: Observed v.s. predicted conversion for the optimal reaction conditions in $T = 25°C$ (blue) and $T = 50°C$ (red).

As seen in Figure 17 the runs with $T = 50°C$ performed better but was also under the predicted average conversion. To evaluate the performance of the optimal reaction conditions for general amide coupling reactions it was tested for a novel small library consisting of 12 different acid substrates and 1 amine. The 12 reactions were also carried out with DMF as solvent and HATU as coupling agent to get comparison between the gold standard reaction conditions at AstraZeneca and the greener alternative of propylene carbonate and PFTU. Due to substrates being proprietary compounds their structures can not be presented here. The reactions were carried out with the general experimental procedure with three replicates each, and the reaction conditions for both solvents are presented in Table 14

Condition	PFTU & Propylene carbonate	DMF & HATU
Reaction volume [µL]	200	200
Acid concentration [mM]	10	10
Acid Equivalents	1	1
Coupling Agent Equivalents	2.14	1.2
Amine Equivalents	1.5	1.5
Base Equivalents	3	3
Reaction Temperature [°C]	50	25
Reaction Time [h]	24	24

Table 14: Reaction conditions for the AstraZeneca amide coupling library performed with both *Propylene carbonate* and *DMF*.

The reaction conditions for the reactions carried out with DMF and HATU presented in Table 14 were based on general experimental procedures at AstraZeneca. Analysing the data problems with co-elution were evident in four out of twelve reactions. These four reactions were therefore excluded from analysis leaving eight reactions for evaluation. The average conversion for the eight reactions are presented in Table 15.

Experiment	PFTU & Propylene carbonate	DMF & HATU
Replicate 1 X_{avg}	0.98	0.98
Replicate 2 X_{avg}	0.98	0.99
Replicate 3 X_{avg}	0.98	0.99

Table 15: Average conversion (X_{avg}) for each AstraZeneca amide coupling library replicate experiment, using *DMF* as solvent with HATU as coupling agent or *Propylene carbonate* as solvent and PFTU as coupling agent.

As seen in Table 15 both methods provided almost full conversion for all three replicates. These results shows that using propylene carbonate as solvent with PFTU as coupling agent is not only a viable green option but performs equally well as DMF and HATU for amide coupling reactions.

5 Conclusion

In the current challenges of green chemistry there are many areas which needs improvement in order achieve sustainability and put less stress on the climate. One of the twelve principles of green chemistry is safer solvents and auxiliaries which aims to reduce the use of toxic solvents in chemical synthesis and production processes. Besides toxicity of solvents, studies have shown that in the pharmaceutical industry the major contributor to the environmental footprint comes from solvents used in the production of API's. AstraZeneca has set targets to reduce their green house gas emission from their own operation (site and fleet) by 98% until 2026. To help industry in using safer solvents and auxiliaries ECHA has created the SVHC candidate list which includes *DMF* that will be restricted from 2023. DMF is a widely used solvent and is, along with the sensitising coupling agent *HATU*, the gold standard for amide couplings at AstraZeneca.

On these premises, in this master thesis project a methodological workflow for optimizing compound library synthesis has been established. It was applied to amide coupling library synthesis with aim of finding a green solvent substitute for *DMF* while avoiding *HATU* as coupling agent in the essence of green chemistry. It was done by using a HTE experimental procedure throughout the project which significantly increased the experiment efficiency.

This was done by first conducting a coupling agent screening which resulted in PFTU as the best option for this project since it showed high conversion while having a clean UV-chromatogram which was an important factor for an accurate calculation of reaction yield. With PFTU as coupling agent a solvent screening was done using DoE in combination with a PCA model provided by ACS GCIPR. The PC's was used as independent variables in a 2^3 full factorial design with the average yield for a small amide coupling library as response which resulted in a prediction model. The prediction model could then be used to predict the average conversion for green solvents within the PCA model. This resulted in propylene carbonate being discovered as a potential green solvent to replace *DMF*. The reaction conditions, with propylene carbonate as solvent and PFTU as coupling agent, were then optimized by conducting a D-optimal design with temperature, acid concentration, coupling agent equivalents and base equivalents as independent variables. Optimal reaction conditions were established and tried for a new amide coupling library along with *DMF* and *HATU* for comparison. The results showed that both combinations of solvent and coupling agent provided almost full conversion which indicates that propylene carbonate is equally good as the gold standard for amide coupling reactions.

Further the results show how effective the modelling techniques DoE and PCA are in combination, while the HTE workflow allowed for sufficient number of experiments to produce accurate general prediction models. In future work, to further establish how effective the method of using DoE and PCA in combination is, the method could be applied to other types of chemical transformations to help AstraZeneca in their transition towards green solvents.

References

[1] Paul Anastas and John Charles Warner. *Green chemistry: theory and practive*. OXFORD University Press, 1998.

[2] AstraZeneca. Ambition Zero Carbon. page 2030, 2020.

[3] AstraZeneca. Sustainability Data. pages 3–8, 2016.

[4] Concepcion Jimenez-Gonzalez, Celia S Ponder, Quirinus B Broxterman, and Julie B Manley. Using the Right Green Yardstick: Why Process Mass Intensity Is Used in the Pharmaceutical Industry To Drive More Sustainable Processes. *Organic Process Research & Development*, 15(4):912–917, 2011.

[5] Catherine M Alder, John D Hayler, Richard K Henderson, Anikó M Redman, Lena Shukla, Leanna E Shuster, and Helen F Sneddon. Updating and further expanding GSK{'}s solvent sustainability guide. *Green Chem.*, 18(13):3879–3890, 2016.

[6] Concepción Jiménez-González, Alan D. Curzons, David J.C. Constable, and Virginia L. Cunningham. Cradle-to-Gate Life Cycle Inventory and Assessment of Pharmaceutical Compounds. *International Journal of Life Cycle Assessment*, 9(2):114–121, 2004.

[7] European Chemicals Agency. Substance Inforcard: N,N-dimethylfromamide, 2021.

[8] ECHA. Support Document for Identification of N,N-Dimethylformamide As a Substance of Very High Concern Because of Its Cmr Properties. 57(November):1–6, 2012.

[9] Food and Drug Administration. Guidance for Industry: Postmarketing Studies and Clinical Trials—Implementation of Section 505 (o)(3) of the Federal Food, Drug, and Cosmetic Act. AAdministration, F. and D., 2012. Guidance for Industry: Postmarketing Studies and Clinical Trials—Implement. 505(April), 2012.

[10] Graham L. Patrick. *An Introduction to Medicinal Chemistry*. OXFORD University Press, second edi edition, 2001.

[11] Alleyn T Plowright, Craig Johnstone, Jan Kihlberg, Jonas Pettersson, Graeme Robb, and Richard A Thompson. Hypothesis driven drug design: improving quality and effectiveness of the design-make-test-analyse cycle. *Drug Discovery Today*, 17(1):56–62, 2012.

[12] György M Keserű and Gergely M Makara. Hit discovery and hit-to-lead

approaches. *Drug Discovery Today*, 11(15):741–748, 2006.

[13] Konrad H. Bleicher, Hans Joachim Böhm, Klaus Müller, and Alexander I. Alanine. Hit and lead generation: Beyond high-throughput screening. *Nature Reviews Drug Discovery*, 2(5):369–378, 2003.

[14] Michael Levin. *Pharmaceutical Process Scale-up*. MARCEL DEKKER, INC., 2001.

[15] J. P. Hughes, S. S. Rees, S. B. Kalindjian, and K. L. Philpott. Principles of early drug discovery. *British Journal of Pharmacology*, 162(6):1239–1249, 2011.

[16] Steven M. Mennen, Carolina Alhambra, C. Liana Allen, Mario Barberis, Simon Berritt, Thomas A. Brandt, Andrew D. Campbell, Jesús Castañón, Alan H. Cherney, Melodie Christensen, David B. Damon, J. Eugenio De Diego, Susana García-Cerrada, Pablo García-Losada, Rubén Haro, Jacob Janey, David C. Leitch, Ling Li, Fangfang Liu, Paul C. Lobben, David W.C. Macmillan, Javier Magano, Emma McInturff, Sebastien Monfette, Ronald J. Post, Danielle Schultz, Barbara J. Sitter, Jason M. Stevens, Iulia I. Strambeanu, Jack Twilton, Ke Wang, and Matthew A. Zajac. The Evolution of High-Throughput Experimentation in Pharmaceutical Development and Perspectives on the Future. *Organic Process Research and Development*, 23(6):1213–1242, 2019.

[17] Anders Rasmusson, Bengt Andersson, Louise Olsson, and Ronnie Andersson. *Mathematical Modelling in Chemical Engineering*. Cambridge University Press, 2014.

[18] Douglas C. Montgomery. *Design and Analysis of Experiments*. John Wiley & sons, internatio edition, 2013.

[19] Stavros N Politis, Paolo Colombo, Gaia Colombo, and Dimitrios M Rekkas. Design of experiments (DoE) in pharmaceutical development. *Drug Development and Industrial Pharmacy*, 43(6):889–901, 2017.

[20] Amit S. Patil and Anil M. Pethe. Quality by design (QbD): A new concept for development of quality pharmaceuticals. *International Journal of Pharmaceutical Quality Assurance*, 4(2):13–19, 2013.

[21] Hervé Abdi and Lynne J. Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010.

[22] Hyunku Shin, Hyesun Jeong, Jaena Park, Sunghoi Hong, and Yeonho Choi. Correlation between Cancerous Exosomes and Protein Markers Based on Surface-Enhanced Raman Spectroscopy (SERS) and Principal Component

Analysis (PCA). *ACS Sensors*, 3(12):2637–2643, 2018.

[23] I. T. Joliffe. *Prinipal Component Analysis*. Springer New York, 2002.

[24] M Haziq Qureshi and The Njar■arson. Top 200 Pharmaceuticals by Retail Sales in 2020. page 2020, 2020.

[25] Louis J Diorazio, David R J Hose, and Neil K Adlington. Toward a More Holistic Framework for Solvent Selection. *Organic Process Research & Development*, 20(4):760–773, 2016.

[26] Gregory L Beutner, Ian S Young, Merrill L Davies, Matthew R Hickey, Hyunsoo Park, Jason M Stevens, and Qingmei Ye. TCFH–NMI: Direct Access to N-Acyl Imidazoliums for Challenging Amide Bond Formations. *Organic Letters*, 20(14):4218–4222, 2018.

[27] Donna S. MacMillan, Jane Murray, Helen F. Sneddon, Craig Jamieson, and Allan J.B. Watson. Evaluation of alternative solvents in common amide coupling reactions: Replacement of dichloromethane and N,N-dimethylformamide. *Green Chemistry*, 15(3):596–600, 2013.

Department of Cemistry and Chemical Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY