Automation and analysis of high-dimensionality experiments in biocatalytic reaction screening

A thesis submitted to University College London for the degree of Doctor of Engineering

By

Shama Chilakwad

Department of Biochemical Engineering
University College London
Torrington Place
London
WC1E 7JE

2021

I, Shama Chilakwad, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed……

Date……

Abstract

Biological catalysts are increasingly used in industry in high-throughput screening for drug discovery or for the biocatalytic synthesis of active pharmaceutical intermediates (APIs). Their activity is dependent on high-dimensionality physiochemical processes which are affected by numerous potentially interacting factors such as temperature, pH, substrates, solvents, salinity, and so on. To generate accurate models that map the performance of such systems, it is critical to developing effective experimental and analytical frameworks. However, investigating numerous factors of interest can become unfeasible for conventional manual experimentation which can be time-consuming and prone to human error.

In this thesis, an effective framework for the execution and analysis of high-dimensionality experiments that implement a Design of Experiments (DoE) methodology was created. DoE applies a statistical framework to the simultaneous investigation of multiple factors of interest. To convert the DoE design into a physically executable experiment, the Synthace Life Sciences R&D cloud platform was used where experimental conditions were translated into liquid handling instructions and executed on multiple automated devices. The framework was exemplified by quantifying the activity of an industrially relevant biocatalyst, the CV2025 ω-transaminase enzyme from *Chromobacterium violaceum*, for the conversion of S-methylbenzylamine (MBA) and pyruvate into acetophenone and sodium alanine.

The automation and analysis of high-dimensionality experiments for screening of the CV2025 TAm biocatalytic reaction were carried out in three sequential stages. In the first stage, the basic process of Synthace-driven automated DoE execution was demonstrated by executing traditional DoE studies. This comprised of a screening study that investigated the impact of nine factors of interest, after which an optimisation study was conducted by taking forward five factors of interest using two automated devices to optimise assay conditions further. In total, 480 experimental conditions were executed and analysed to generate mathematical models that identified an optimum. Robust assay conditions were identified which increased enzyme activity >3-fold over the starting conditions. In the second stage, non-biological considerations that impact absorbance-based assay performance were systematically investigated. These considerations were critical to ensuring reliable and precise data generation from future high-dimensionality experiments and include confirming spectrophotometer settings, selecting microplate type and

reaction volume, testing device precision, and managing evaporation as a function of time.

The final stage of the work involved development of a framework for the implementation of a modern type of DoE design called a space-filling design (SFD). SFDs sample factors of interest at numerous settings and can provide a fine-grained characterisation of high-dimensional systems in a single experimental run. However, they are rarely used in biological research due to a large number of experiments required and their demanding, highly variable pipetting requirements. The established framework enabled the execution and analysis of an automated end-to-end SFD where 3,456 experimental conditions were prepared to investigate a 12-dimensional space characterising CV2025 TAm activity. Factors of interest included temperature, pH, buffering agent types, enzyme stability, co-factor, substrate, salt, and solvent concentrations. MATLAB scripts were developed to calculate important biocatalysis metrics of product yield and initial rate which were then used to build mathematical models that were physically validated to confirm successful model prediction. The implementation of the framework provided greater insight into numerous factors influencing CV2025 TAm activity in more dimensions than what was previously reported in the literature and to our knowledge is the first large-scale study that employs a SFD for assay characterisation.

The developed framework is generic in nature and represents a powerful tool for rapid one-step characterisation of high-dimensionality systems. Industrial implementation of the framework could help reduce the time and costs involved in the development of high throughput screens and biocatalytic reaction optimisation.

Impact Statement

In the pharmaceutical discovery and development of new medicines, significant interest has been shown towards the increasing use of automation and digitisation of experimental workflows. This thesis combines the capabilities of the Synthace Life Sciences R&D Cloud Platform and existing laboratory automation solutions to demonstrate the automated execution and analysis of high-dimensionality experiments, specifically using a statistical Design of Experiments (DoE) methodology. The model system used for this work was the bioconversion of *S-MBA* and Pyruvate to Acetophenone and Sodium Alanine catalysed by the industrially relevant CV2025 ω-transaminase enzyme from *Chromobacterium violaceum*.

The research described in the thesis shows that automated DoE execution can be carried out without manually generating complex device-specific protocols. This was exemplified by using a number of commercially available laboratory liquid handling platforms for iterative DoE studies. A traditional DoE campaign, comprising of scoping, screening, and optimisation studies was executed using automation to identify robust CV2025 ω-transaminase bioconversion conditions that can be used for high-throughput screening. This further expanded into the development of a generalizable framework for the automation and analysis of a space-filling DoE (SFD) design that can be used for rapid assay characterization. This was exemplified by characterizing a 12-dimensional assay space in one four-day experiment. The large quantity of data generated was processed using MATLAB scripts to generate important bioprocessing responses of yield and the initial rate which were then used to build meaningful mathematical models.

Overall, the research presented in this thesis demonstrates a new framework that can be used to reduce the time and effort spent characterizing and optimising enzymatic bioconversions in both academic and industrial laboratories. The approaches described are generic and can be applied to a range of end-user applications including assay development, cell-free protein synthesis, and drug discovery. In collaboration with the sponsoring company, this work has already been applied by end-users of the Synthace platform for developing biological and cellular assays.

Acknowledgments

I would like to thank my academic supervisor, Professor Gary Lye, and my industrial supervisors, Dr James Arpino and Dr Jordan Ang, for their constant guidance, support, patience, and encouragement throughout this project. You all have taught me so much and I really could not have asked for better supervisors for my EngD studies.

At Synthace, I am indebted to Dr Markus Gershater and Dr Tim Fell for creating this opportunity for me. Being able to do this research whilst seeing the company grow and evolve has been a real privilege, and for that you have my enduring gratitude. I thank Dr Michael Sadowski for introducing me to space-filling designs and helping me with DoE analysis. It has been fundamental to the success of this project. To Dr Chris Grant who helped troubleshoot any issues I ran into with my DoE workflows, to Dr Matt Couch for his insight into data analysis, and to Dr Xander Anderson for building the dragonfly® discovery driver, I also give my sincere thanks. I would also like to thank all the members of lab team for their help over the years, without whom this research would have been far more difficult. I give special thanks to Dr Vishal Sanchania for his advice and support, and to and Dr Jennifer Man for her incredible effort in managing the Synthace lab. I gratefully acknowledge Dr Charlie Bird and Dr Emilie Fritsch for their help and the generosity of their time. Lastly, I would like to thank the whole Synthace team for creating such an incredible work environment. I could not imagine doing this EngD anywhere else.

At SPT Labtech, I am also deeply thankful to Dr Anne Hammerstein and her team for lending me the dragonfly® discovery on numerous occasions. This partnership has been central to my project and has enabled me to do some truly exciting research.

At UCL, I would like to thank Professor John Ward, Professor Helen Hailes, and Dr Fiona Truscott for providing the initial CV2025 protocols and helping me get started in the lab.

I would like to thank all my friends for their love and support, especially my London friends for keeping me grounded during the times when my research was most trying. I thank Jules in particular for keeping me cultured.

I cannot forget to give my immeasurable thanks to Amma and Appa for everything they have done to get me this far and for always believing in me.

Finally, and most importantly, I thank my brother Ashveen, for the non-stop banter.

Table of Contents

List of Figures

List of Tables

Nomenclature

OFAT	One Factor at a Time
A	Absorbance
A.U.	Arbitrary Units
ANN	Artificial Neural Network
AP	Acetophenone
C	Celsius
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
CV	Coefficient of Variance
DMSO	Dimethyl Sulfoxide
DoE	Design of Experiments
E. coli	*Escherichia coli*
ELN	Electronic Lab Notebook
EMA	European Medicines Agency
FDA	Food and Drug Administration
FFF	Fast Flexible Filling
GPM	Gaussian Process Model
GUI	Graphical User Interface
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
ICH	International Conference on Harmonisation
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JSL	JMP Scripting Language
kD	kilo Dalton
LB	*Luaria Bertani*
LHP	Liquid Handling Policy
nm	nanometres
OD	Optical Density
PCR	Polymerase Chain Reaction
PLP	Pyridoxal 5'- Phosphate
PMP	Pyridoxamine 5-Phosphate
QbD	Quality by Design
R&D	Research and Development
RPM	Revolutions per minute
RSM	Response Surface Methodology
RT PCR	Reverse transcription Polymerase Chain Reaction
S-MBA	S-methylbenzylamine
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SFD	Space Filling Design
SRM	Stepwise Regression Model
TAm	Transaminase
TRIS-HCl	tris(hydroxymethyl)aminomethane

UV	Ultraviolet
v/v	volume per volume

1 Introduction

This thesis focuses on the study of the automated execution and analysis of high-dimensionality experiments that follow a Design of Experiments (DoE) methodology. This approach was used to investigate and optimise the activity of an industrially relevant enzyme, the CV2025 ω-transaminase from *Chromobacterium violaceum*, useful for the stereoselective amination of pharmaceutical intermediates. Automated experiments were designed and implemented using the Synthace R&D Cloud Platform, JMP (SAS Institute, USA) was used for DoE design generation, analysis, and visualisation of results, and the data processing pipeline involved MATLAB (MathWorks, USA) for automated processing of the raw absorbance data. This chapter provides background information on each of these aspects and reviews relevant literature before presenting the thesis aim and objectives.

1.1 Design of Experiments methodology for high-dimensionality experiments

1.1.1 The DoE approach

Biological systems and processes are inherently complex. They function by balancing the effect of a myriad potentially interacting variables within a dynamic physicochemical space. Understanding these systems in turn requires the application of sophisticated experimental methods and data analysis approaches. Despite this known complexity, most biological research is still typically executed using a One Factor at a Time (OFAT) approach (Anderson and Whitcomb, 2007; Fellermann *et al.*, 2019). Here, the impact of one variable of interest (also known as a factor) on a process is investigated first while all other factor settings are held constant. The first factor is then fixed at the best performing setting before moving onto evaluate the impact of a second factor (Figure 1.1 A & B). Consequently, the effect of factor interactions, that are characteristic of biological processes, are lost as factors are

explored in isolation to one another (Erbach *et al.*, 2004; Lendrem *et al.*, 2015). Despite the significant limitation of OFAT experimentation, its use remains ubiquitous due to its straight-forward and intuitive implementation and the fact that it is commonly taught as the standard scientific method (Lendrem *et al.*, 2016).

Limitations of the OFAT approach can be overcome by using DoE approaches. DoE represents a suite of methods that apply a statistical framework to experimental design, execution, and analysis. DoE methodology enables the investigation of multiple factors that impact a process in parallel (as explained in Figure 1.1). Experiments that use this approach are often called 'designed experiments'. Statistical DoE was first developed by Sir Ronald Fisher who used it to determine the effect of different types of fertilisers, volume of water and type of hybrid crop used to differentiate their effects on maximising crop production (Fisher, 1926). It has been further developed over the last century (Ronald A. Fisher, 1935; Box and Wilson, 1951; Elfving, 1952) and is now considered one of the most powerful tools for process characterisation and optimisation (Farooq *et al.*, 2016; Sadowski, Grant and Fell, 2016). It is also widely used across a variety of industries, including manufacturing, finance, food, and aviation (Antony, 2014a; Midilli and Parshutin, 2019; Singh and Rathi, 2019).

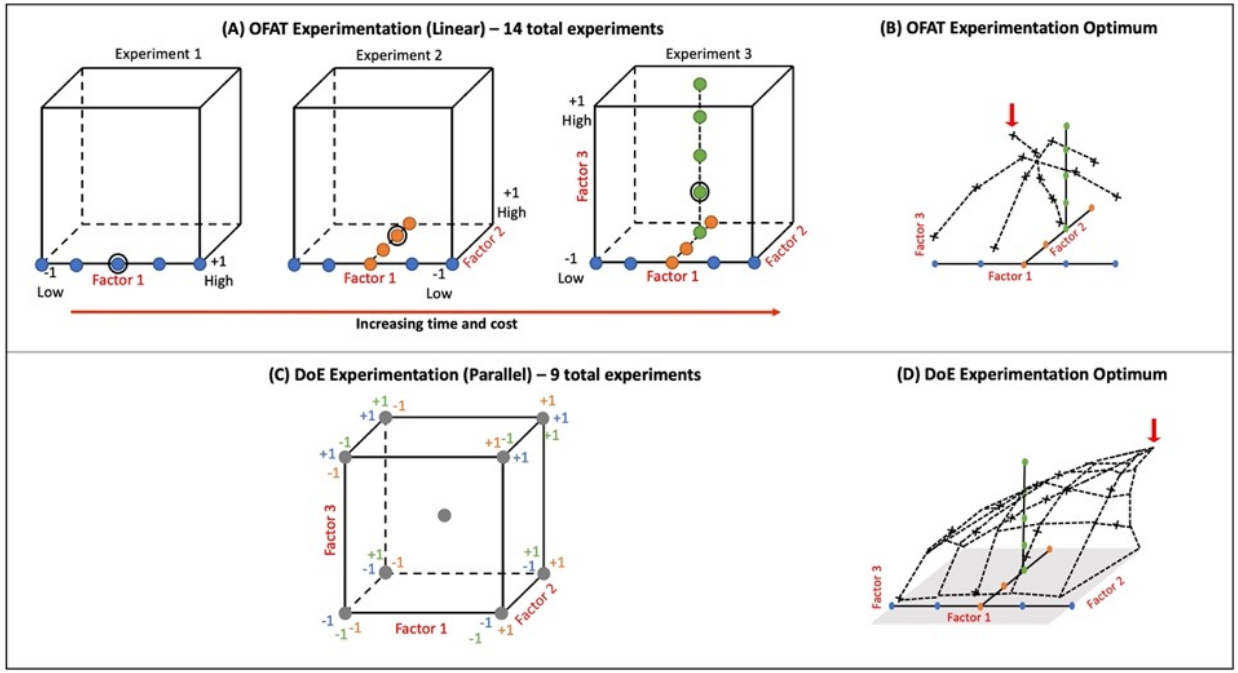


Figure 1.1. Comparison of OFAT and DoE approaches.

The figure shows the design space explored, and the location of optima found for both methods when 3 factors are investigated across 2 levels (-1 and +1 represent low and high levels, respectively). (1A) The OFAT approach investigates one factor in isolation with linear experimentation. The circled data point from Experiment 1 shows the optimal value that is concluded from experimental investigation of Factor 1. This value for Factor 1 is then carried forward to Experiment 2 where an optimal value is chosen for Factor 2 and so on. A total of 14 experiments are required to draw these conclusions. (1B) The research space explored for the three factors independently represented. Here, each factor is tested linearly with no parallel factor interactions. The red arrow indicates the researcher's perception of where the optimum might lie which is determined after the three experiments are performed. (1C) The multifactorial DoE approach varies Factors 1, 2 and 3 in parallel, covering a larger design space in only 9 experimental runs. Experimentation including the centre point, the midpoint of all the factors, allows curvature to be evaluated using data visualisation tools that are used in combination with the multifactorial approach to experimental design. (1D) The higher optimum that is found by performing a multifactorial DoE experiment is shown. Here, varying values for all three factors are incorporated into the experiment design in parallel. This method tests a wider design space allowing the researcher to understand how Factors 1, 2 and 3 interact with each other at both levels.

1.1.2 DoE terminology

In DoE terminology, a ‘factor’ is a variable of interest that is intentionally controlled to assess its impact on the performance of a process. Individual settings termed ‘levels’ are used to examine factors. Temperature is a factor that is sampled at three levels, for example, if temperature is investigated at 25, 35, and 45 °C. Factors are typically grouped into two types: categorical and continuous. Factors are categorical when there is no relationship between different levels, for example, buffer type, salt type, growth medium, bioreactor type etc. Factors are continuous when their levels are intrinsically linked, for example, temperature, pH, salt concentration etc. When a collection of factors at different levels is modified simultaneously in a designed experiment, the experimental space they cover is referred to as the 'design space'. Each factor accounts for one dimension of the design space. The DoE design uses factors and their levels as inputs to generate a set of algorithmically specified treatments, each of which is referred to as a ‘run’ or ‘run condition’. The collection of runs is represented in a DoE design table with each row containing run specific factor settings. These settings are used as instructions to execute the design which generates data that is used to calculate a ‘response’. Responses are metrics defined by the researcher to evaluate the process under investigation e.g., yield or rate. They are utilized to generate mathematical models to describe the relationship between the factors and response. The term ‘DoE’ is sometimes colloquially used when discussing an experiment that uses a DoE approach.

1.1.3 Statistical principles underpinning DoE

To ensure that robust mathematical models are generated, a designed experiment incorporates three underpinning statistical principles to account for biological and technical noise. These are: randomisation, replication and blocking (Anderson and Whitcomb, 2007; Montgomery, 2013; Antony, 2014b). Runs are executed in a random order. This ensures all the responses and errors are independent of each other. Replication of runs is performed either by including the entire experiment or

just selective runs e.g., the design centre point. This ensures that both an accurate estimation of the factor and factor interactions, and the experimental error, are captured. A designed experiment may often need to be 'blocked' according to a limiting factor. Blocks are created by separating the design into smaller groups of randomised runs to ensure that all factor effects can be appropriately estimated. For example, a DoE may need to be blocked by days, raw materials, or equipment availability. Overall, these fundamentals are used to account for technical error and biological noise and lead to more precise estimation of factor effects.

1.1.4 Scale of DoE designs

Due to the complexity of biological systems, it is easy to list multiple factors (k) over a number of levels (x) that might affect a process where x^k represents a 'full factorial' experiment as it investigates all possible combinations of factors and levels. For example, a full factorial experiment testing 5, 6 or 7 factors across 3 levels each would require 3^5 (243), 3^6 (729), and 3^7 (2,187) experimental runs, respectively. Depending on the resources and time available, execution of these large experiments could be feasible. However, a full factorial experiment with 10, 15 or 20 factors would require 3^{10} (59,049), 3^{15} (14,348,907), and 3^{20} (3,486,784,401) experimental runs respectively, to cover the entire design space. This number of runs is both unnecessary and unfeasible to implement in practice. Consequently, full factorial designs are not typically executed. Instead, the size of a full factorial design is significantly reduced by assuming that higher-order interactions, typically third order and above are unexpected and unimportant. These designs are called fractional factorial designs; they typically focus on main effects and two-factor interactions (Anderson and Whitcomb, 2007).

Selecting too small a run number may lead to 'confounding'. This is the term used when the influence of one factor cannot be differentiated from that of another. To avoid this, a summary statistic called 'design resolution' is used to understand if the total run number is sufficient to identify main effects and interactions; this can be

used to select an appropriate design (Montgomery, 2013; Onyeogaziri and Papaneophytou, 2019).

Using fractional factorial designs ensures that the design space is sampled in a sparse and unbiased manner. DoE studies are most effectively performed in an iterative manner (Coleman and Montgomery, 1993; Nabifar *et al.*, 2010), with each iteration narrowing in on the optimum (Figure 1.2). Typically, less experiments are required to optimise a system using DoE compared to OFAT methods which in turn reduces cost and time (Figure 1.1).

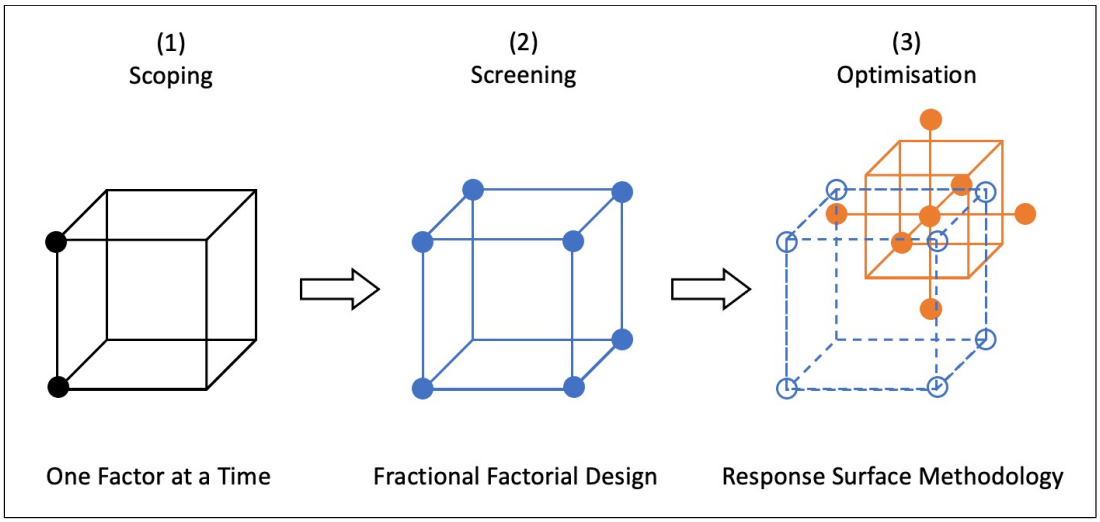


Figure 1.2. Illustration of iterative DoE experimentation.
Dots represent experimental run conditions. (1) Factors of interest are first tested to identify suitable factor levels for screening typically using an OFAT approach. (2) Selections of factors are screened typically using a fractional factorial design to efficiently explore the design space. (3) An optimization DoE design, such as a response surface design is applied to a smaller subset of significant factors to identify an optimum region. Permutations of this process may be required depending on the system being explored, e.g., screening can take place directly if sufficient prior knowledge of the system is available. Alternatively, a number of DoE screening designs may be required before optimisation.

1.1.5 Application of DoE in biological research and barriers to entry

The impact of DoE on biological research is most significant within industrial pharmaceutical research and development. Here DoE is an integral part of the statistical toolbox (reviewed in (Rathore, 2009; Mercier et al., 2014; Politis et al., 2017)(Wasalathanthri *et al.*, 2021)). Regulatory bodies (such as the US Food and Drug

Administration (FDA) and the European Medicines Agency (EMA)) increasingly recommend pharmaceuticals to be manufactured by processes that follow a Quality by Design (QbD) framework which encompasses a body of practices geared toward a systematic understanding of the products and processes to mitigate risk and minimise variability (Montgomery, 2012). It recognises that "quality cannot be tested into products, i.e., quality should be built in by design" (Rathore, 2009). To meet manufacturing standards consistently, DoE is used to systematically determine the bounds of the design space within which the manufacturing process must occur. These guidelines are outlined in the International Conference on Harmonisation (ICH) guidelines, ICH Q8 (R2) Pharmaceutical Development (ICH Q8 (R2), 2014).

In contrast, within basic biological research, DoE is used infrequently, and its acceptance has been slow. However, some examples of its use can be found in a range of biological fields including academic drug discovery (Tye, 2004), synthetic biology (Gilman *et al.*, 2020), mass spectrometry (Hecht, Oberg and Muddiman, 2016), liposome development (Jain, Hurkat and Jain, 2019), recombinant protein production (Uhoraningoga *et al.*, 2018), metabolomics (Jacyna, Kordalewska and Markuszewski, 2019), microfluidics (Sun, Lin and Barron, 2011), fermentation and bioreactor optimisation (Adinarayana and Ellaiah, 2002; Islam *et al.*, 2007; Fricke *et al.*, 2013), and cell therapy bioprocessing (Acosta-Martinez *et al.*, 2010).

The infrequent use of DoE is typically attributed to the lack of formal statistical training, where researchers are completely unaware of DoE methodology (Santos, Rato and Reis, 2019). Novice users are also faced with an initial learning curve which can be challenging and a barrier to adoption. To help overcome this, a number of commercial software packages have been designed to guide users through DoE design and analysis. These include JMP (SAS Institute, USA) Design Expert (Stat-Ease Inc., USA), Minitab (Minitab, LLC, USA), and Modde® (Sartorious, Germany) (Tye, 2004). The requirement of defining suitable factors and levels of interest *a priori* can also pose a challenge if sufficient expert knowledge is unavailable (Cambray, Guimaraes and Arkin, 2018). Additionally, manual pipetting that is required to

prepare experiments in a randomised order can be unintuitive and error-prone despite extensive upfront planning and organisation. These perceived barriers can thus make DoE an unattractive framework for researchers to follow (Erbach *et al.*, 2004; Farooq *et al.*, 2016).

Automation of DoE execution, using commercially available liquid handling technologies, represents a promising approach to making DoE more accessible. Currently available liquid handling solutions are described in Section 1.2.

1.2 Automated liquid handling solutions for biological experimentation

1.2.1 Automated liquid handling devices

Even though automated devices and liquid handling robots have been available for decades (Pauwels *et al.*, 1995), the majority of biological research is still performed manually using handheld pipettes (Baillargeon *et al.*, 2019). In this section, currently available hardware, and software solutions for automating biological experiments will be described. The emphasis will be on commercially available liquid-handling platforms ('liquid handlers'), the barriers to entry, and the software tools available to automate DoE execution.

Liquid handlers are devices that are designed to automate routine pipetting actions. The most basic class of liquid handlers use Cartesian robotics; where pipetting heads aspirate, dispense and mix solutions by moving in XYZ planes (Kong *et al.*, 2012). Device-specific control software is used to create an experimental protocol and monitor execution. Many liquid handlers of varying sizes and capabilities are commercially available. A subset of the most commonly used devices available today is listed in Table 1.1 together with an indication of their relative cost.

By using liquid handlers, manual pipetting errors and user-to-user variability when handling microliter volumes can be replaced with consistency, accuracy, and increased experimental throughput (Nealon *et al.*, 2005). Efficiency is increased as the researcher can focus on intellectual tasks due to the 'walk-away' time that is created when the liquid handler is in operation. Furthermore, experiments that would normally be difficult to execute by humans can be made accessible by the application of automation e.g., more frequent passaging of cultured cells (Hussain *et al.*, 2013).

Table 1.1. Range of commercially available liquid handling platforms and their relative capital cost.

Platform/ range name	Company	Price range*	Primary control software/ Graphical User Interface (GUI)	Secondary software and functions
OT-One S Robot	Open Trons (New York, NY, USA) (opentrons.com)	Very Low	Opentrons App	–
Andrew	Andrew Alliance (Geneve, Switzerland) (andrewalliance.com)	Low	Andrew Lab	–
Gilson Pipetmax	Gilson (Middleton, WI, USA) (gilson.com)	Low	Gilson Trilution® LH	• qPCR Assistant: qPCR method creation • Normalisation Assistant: DNA/RNA normalisation tool
CyBio® Felix	Analytik Jena (Jena, Germany) (analytik-jena.de)	Low	CyBio Composer®	• CyBio Scheduler®: predictive planning
Versette	Thermo Fisher Scientific (Waltham, MA, USA) (thermofisher.com)	Low–Medium	Versette™ ControlMate®	–
Dragonfly Discovery	TTP Labtech (Hertfordshire,UK) (ttplabtech.com)	Medium	Dragonfly	–
Mantis	Formulatrix (MA, US) (formulatrix.com)	Medium	Mantis	–
TECAN Freedom Evo® Series	Tecan (Mannedorf, Switzerland) (lifesciences.tecan.com)	Medium	Freedom EVOware®	–
Microlab STAR Series	Hamilton Robotics (Bonaduz, GR, Switzerland) (hamiltoncompany.com)	High	Venus	• Hamilton STAR Program Assistant (HSPA): Assistant software to Venus.
Biomek Series	Beckman Coulter (Brea, CA, USA) (beckmancoulter.com)	High	Biomek Software	Comprises of a collection of software for varying functions: • Biomek Method Launcher: launch screen • Biomek PowerPack: for easier programming • SAMI EX: flexible process control • SAMI Process Management: workflow management tool • DART: data gathering
Echo® Series	LabCyte Inc. (San Jose, CA, USA) (labcyte.com)	Very High	Echo® Software Application	Comprises of a collection of softwares for unit operations: • Echo® Array Maker: custom array design • Echo® Cherry Pick: transfers from well to well • Echo® Combination Screen: curve editor • Echo® Dose-Response: dose titration results • Echo® Plate Audit: validate and monitor sample libraries • Echo® Plate Reformat: plate transfer functions

* Very low cost: <£5K, Low cost: £20-30K, Medium cost: >£80K, High cost: £100-200K, Very high: >£230K. Categories based on 2021 UK prices.

Despite generally limited adoption in biological research, automation has become the default mode of operation within a few specific areas. The most predominant of these is in drug discovery, where small molecule libraries containing tens of thousands of compounds undergo high-throughput screening (HTS) against target assays to identify candidate therapeutics (Meyer *et al.*, 2017; Schneider, 2018). HTS methods were developed alongside the wide adoption of standardised multi-well microplate formats that were compliant according to the Society of Biomolecular Sciences (SBS). Starting in the mid-1990s, 96- and 384-well plate formats were most widely used (Armstrong, 1999), however, over the past decades, high-density plate formats containing 1536 and 3456 microwells have gained popularity and have been used to execute many biological assays at a total volume of 1-2 µL, thus, making it possible to conduct up to 100,000 assays per day (Szymański, Markowicz and Mikiciuk-Olasik, 2012; Beal, Lu and Weiss, 2021; Landeta and Mejia-Santana, 2021).

In the last two years, automation has played a significant role in the response to the SARS – CoV – 2 pandemic to meet the urgent need for hundreds of thousands of clinical diagnostic tests, which has, in turn, led to the execution of thousands of automated PCR/RT-PCR tests per day by reconfiguring commercially available liquid handlers (Crone *et al.*, 2020; Macaulay, 2020; Sanders, 2020). Examples of automated experimentation are also scattered within other areas of biological research, including automated DNA synthesis (Kaplan, 1985), rapid mutant library generation (Quaglia *et al.*, 2017), multi-part DNA construct assembly (Chao *et al.*, 2017), and end-to-end protein production (Bos *et al.*, 2015).

The above studies exemplify how automation can be used to perform tasks that would be challenging and time-consuming to attempt manually. Once the automation hardware is in place, there is then the need to consider how conventional lab processes are adapted for automation purposes and how the experimental protocol is programmed in the device control software. These aspects are considered in Section 1.2.2.

1.2.2 Generation of automation protocols

Each of the liquid handling devices listed in Table 1.1. come with their own control software. This is used by the operator to create protocol instructions; a set of instructions is often referred to as a 'script'. All the liquid-handling actions required to execute pipetting commands that make up a particular experiment are defined step-by-step within a script. Depending on the complexity of the automation device, the script may also contain instructions for other operations e.g., for the movement of plates around the deck of the robot or for the automated operation of integrated analytical devices such as plate readers or HPLCs. The complexity of each control software varies greatly. Some vendors have recognised the need to make intuitive and user-friendly Graphical User Interfaces (GUIs) e.g., the Freedom EVOware that controls the Tecan Evo Series liquid handling platforms (Männedorf, Switzerland) contains a drag-and-drop feature to easily add various actions such as aspirate, dispense, mix, etc. into a script.

In contrast, some liquid handlers can only be used when customised instructions are created by the vendor (Pandya *et al.*, 2010); this is the case for Hamilton Microlab STAR liquid handlers (Bonaduz, GR, Switzerland) that use a control software called Venus. Generating scripts using Venus has previously been called 'daunting' due to the programming skills required and often requires supporting secondary software to access the full functionality of the device (Fan, 2016; Meyer *et al.*, 2017).

Automated protocols require the researcher to become an expert user of the control software as they need to learn to specify not only the sequence of operations but also the underlying detail of how each operation is performed e.g., the specific location of each aspirate and dispense step, the volume and speed of each addition, how liquids are mixed, etc (Pandya *et al.*, 2010). Even for a straightforward experiment, this can become tedious as it requires continuous iterative adjustment of the script (Naugler and Church, 2019; Christensen *et al.*, 2021). It also limits the flexibility and potential of these devices and leaves automated protocols exposed to the same errors as manual protocols (Oritz *et al.*, 2017). Consequently, automation

is generally used to execute a limited number of tasks and is primarily used to replace repetitive pipetting tasks by reusing one script repeatedly.

When considering the automation of DoE designs, the complexity of the pipetting actions required to prepare multifactorial randomised runs can make current automated solutions difficult to use. Recent advances in the automation of DoE experimentation are discussed in Section 1.2.3.

1.2.3 Current solutions for automated DoE execution

To automate DoE execution, some researchers have taken on the task of creating their own support software to enable them to perform DoE studies. An early example of this was the use of a Tecan Freedom EVO 200 liquid handler to carry out a DoE study that investigated lysozyme solubility with an aim to identify optimal ion strength and pH in a precipitation process design. To implement this DoE protocol, an intelligent architecture programmed in C# that enabled liquid handling and analytical device integration, was used in combination with the Tecan control software (Wu and Zhou, 2014). Similarly, to study enzyme kinetics, the R programming language has been used to develop software that integrates automated protocol execution on a Tecan Genesis RSP 159 liquid handler and subsequent data analysis (Bonowski *et al.*, 2010). The Visual Basic programming language has also been used to import DoE design files into Tecan EVOware for optimising a ligand-binding assay (Tsoi, Patel and Shih, 2014). These examples illustrate that expert programming knowledge is generally required for automating DoE execution. Many labs may not have this skill set, time, or resources to develop such software tools for specific experiments.

Increasingly, the equipment vendors themselves have been exploring solutions for automated DoE execution. This has been driven by end-user requirements and the increasing use of DoE in pharmaceutical development as described in Section 1.1.5. Beckman Coulter was an early pioneer of this approach with a software package called SAGIAN™ Automated Assay Optimisation (AAO) (Erbach *et al.*, 2004) that could be used for assay optimisation DOEs on the Biomek FX Laboratory Automation Workstation or the Beckman Coulter BioRAPTR™ Microfluidic Dispenser. The DoE design file has to first be imported from Design Expert™ into AAO, the user must then specify parameters such as labware and device deck layouts. Once the DOE has been executed, the data can be collected and then exported using AAO (Taylor *et al.*, 2000). Another company, SPT Labtech (Melbourn, UK) has used their dragonfly® discovery liquid handling platform to perform DoEs by expanding their file inputs to allow input

and conversion of DoE files created using the JMP software (SAS Institute). This platform accepts all microplate formats up to 1,536 well plates to allow for low volume and high-throughput assay development (Scott *et al.*, 2018).

The limitation of these examples is that they are both device-specific and experiment-specific. They do, however, highlight the need for software solutions that can define biological protocols in a standardised high-level manner, translate them into instructions for automated devices and collect and analyse data as required (Bonowski *et al.*, 2010). Approaches to tackle this bottleneck are described in Section 1.2.4.

1.2.4 Digitisation and automation of biological research and development

The ease by which researchers can interface with robotic platforms, such as liquid handlers, is critical to automating biological research. This has led to the development of software tools that aim to digitise and automate different areas of biological research. A number of the most recent developments in this area are described below.

Aquarium is an open-source, web-based application that aims to digitize manual benchtop experiments. It can be used for the integration of experiment design, inventory management, protocol execution, and data capture. Once experiments are prepared digitally, a researcher follows step-by-step instructions to manually execute biological workflows by leveraging existing equipment (Vrana *et al.*, 2021). This has the advantage that existing labware can be leveraged to formalise the descriptions and execution of scientific workflows by researchers that may not have a high-level understanding of the protocol being performed, however, it is still limited to manual workflow execution and no experimental design and analysis done within the application.

Pr-Pr is an open-source biology-friendly high-level programming language that was developed with the aim to create standardised liquid-handling protocols. User-defined protocol steps are compiled into low-level instructions for Tecan liquid handlers, microfluidic devices, and manual execution (Linshiz *et al.*, 2013, 2014). Future development of Pr-Pr could potentially enable the adoption of standardised automated protocols across various laboratories despite any differences in automation equipment they have available. However, building Pr-Pr translators for devices other than Tecan liquid handlers is currently limited to individuals in the scientific community with the necessary expertise and time.

Many companies have also joined the effort to digitize biological research (Sadowski, Grant and Fell, 2016; Appleton *et al.*, 2017; Whitehead *et al.*, 2018). For example, Riffyn (riffyn.com) provides software for experiment planning and automated data collection and processing. Benchling (benchling.com) is a cloud platform that enables the centralisation of scientific data; its core function is an Electronic Lab Notebook (ELN) that can be used to track experiments, processes, and samples. It also consists of a molecular biology suite that has user-friendly tools for DNA construct assembly. Similarly, Teselagen (teselagen.com) focuses on DNA assembly design. Dotmatics (dotmatics.com) has a platform that has both ELN and data analysis capabilities. To bridge the gap between experiment design and automation, companies like Riffyn and Teselagen also have the functionality to create worklists; these are files that contain instructions for the Beckman Coulter Echo liquid handler control software, although they don't directly communicate with the device at present (Crone *et al.*, 2020).

Companies such as Emerald Cloud Labs (emeraldcoloudlab.com) and Strateos (known as Transcriptic prior to a merger in 2019, strateos.com) focus more heavily on automation as they provide a remote 'cloud lab' service that consists of integrated automated devices for end-to-end execution of common molecular biology protocols such as PCR and cloning. Clients outsource their experimentation to these companies and monitor progress online (Hayden, 2014). Strateos also contains a library of

predefined protocols that can be customised, validated, and queued up for execution directly by the clients (https://developers.strateos.com/docs).

Finally, the Synthace Life Sciences R&D Cloud Platform (referred to as Synthace, synthace.com) can be used to design, perform, and analyse experiments (Sadowski, Grant and Fell, 2016). Synthace and SynthaceHub are two pieces of software that comprise the Synthace platform. Experiments, including DoE designs, can be designed, and confirmed *in silico* in the Synthace user-friendly interface. Once experiments are ready to be executed, they are directly sent to Synthace Hub, which is installed on a lab computer connected to the automated device. This is used to start automated execution, operating through the device-specific control software. Furthermore, Synthace is a device-agnostic platform that allows the same experiments to be run on a variety of devices. Synthace was the software platform selected for use in this work hence it is described in more detail in Section 1.3.

1.3 Synthace Life Sciences R&D Cloud Platform

The Synthace Life Sciences R&D Cloud Platform is one of the latest software packages that aims to seamlessly automate biological experimentation and analysis. Synthace allows users to define high-level experimental workflows which are encoded in a programming language derived from Google Go. Workflows are interpreted and optimized by the Synthace 'planner' and are then compiled into automated device-specific liquid handling instructions using specific device drivers. Many automated devices are supported by Synthace, the majority of which are liquid handlers.

Experiments are designed and executed over a series of pages within the Synthace graphical user interface (Figure 1.3). The first step of this process is the creation of a Synthace workflow using a page called the 'Builder'. The Builder contains a canvas where a workflow is created, along with libraries of devices, plate types, tip types, and liquid handling policies.

The Builder is used to create a ‘workflow’, which is a user-defined digital protocol that contains high-level details of the automated experiment. A workflow is created by using pre-existing building blocks called ‘elements’. Each element is responsible for a specific function. Elements contain ‘input’ and ‘output’ parameters, inputs are user-defined descriptions of components required by the element to function, data generated from processing these inputs creates outputs. Typically, the outputs of one element can be joined to the input of another element. Thus, information can flow between elements which enable the creation of complex biological workflows which then get compiled to generate device-specific liquid-handling instructions.

Crucially, once a workflow is prepared, it can be simulated to enable the user to check for errors. Here, Synthace uses high-level details within the workflow and translates them into a device-specific automated protocol. Transfer volumes, plate layouts, and device preparation instructions are automatically generated. A workflow that is successfully simulated is called a ‘simulation’. If a simulation fails, an error message guides the user to the element that caused the error; often this process can be iterative as manual errors, missing information or incorrect inputs are identified.
The details of a simulation are accessed through a series of three pages collectively called ‘Simulation Details’; these include the ‘Overview’, ‘Setup’ and ‘Preview’ pages. The Overview page summarises the resources e.g., labware and reagents, required for the experiment. It also contains all the files created from that simulation. The Setup page shows the contents of each input and output plate required for the experiment. The reagents that need to be provided to the experiment are already organised in a pre-determined layout at required volumes, therefore, this information is used to prepare input plates prior to execution. The Preview page contains a visual representation of the automated experiment, it allows the user to step through every liquid-handling step one pipetting action at a time. As each step of the experiment can be visualised, this page is used to confirm that the automated experiment will operate as expected. It is also used to instruct the user on how to set up the deck of the automation platform being used i.e., where plates, reagent reservoirs, and tip boxes should be placed.

A simulation becomes an 'execution' when it is scheduled from the Simulation Details page. This process sends the execution to the pre-installed SynthaceHub software on the computer that controls the automated device (a manual instruction file upload is required for a few devices). Once the device is set up as previously shown in Synthace, the execution can be selected and run. A range of experiments, each of varying complexity can be automated as the same underlying process is followed each time.

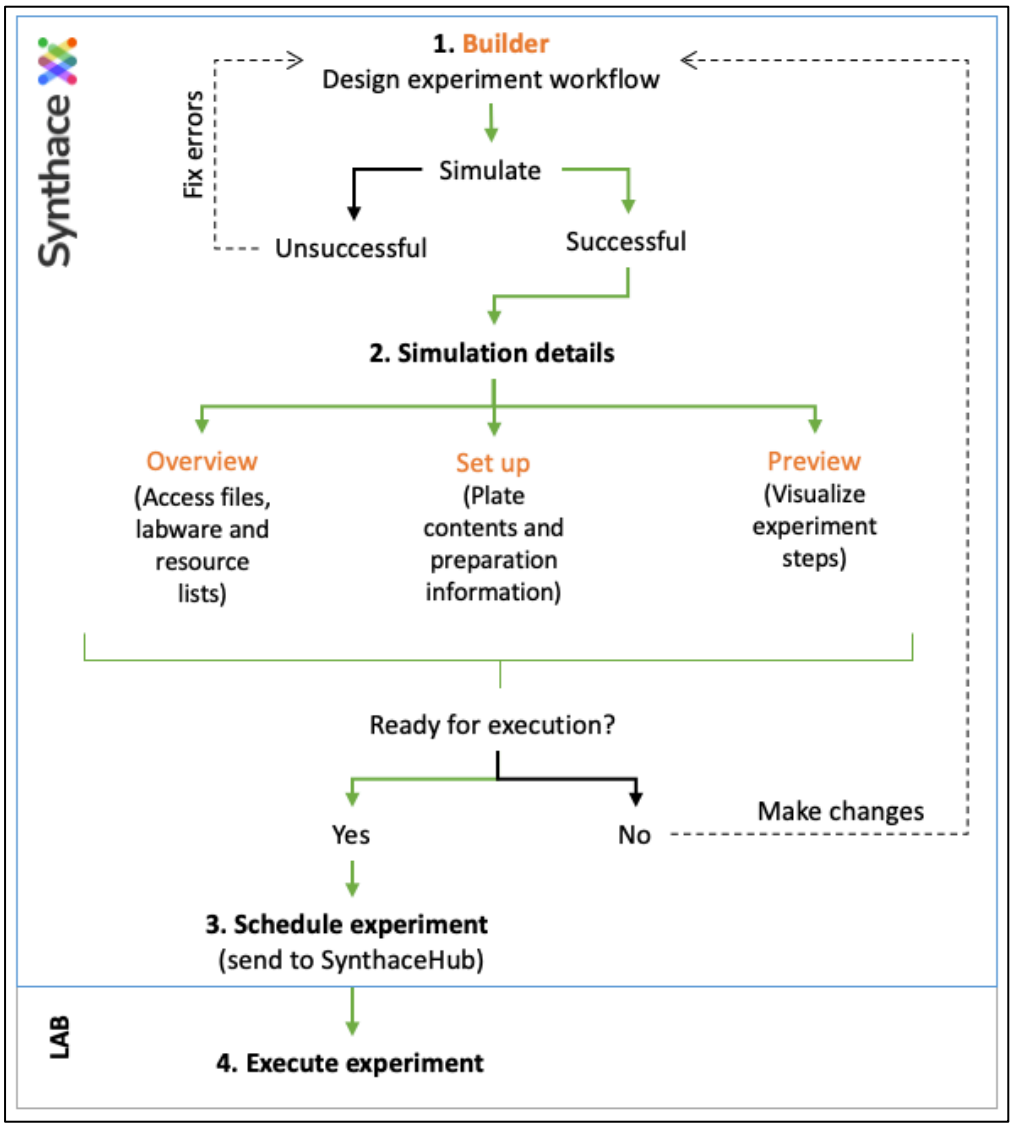


Figure 1.3. Flowchart outlining the process for automating an experiment using Synthace. Synthace specific pages are named in orange text. Green arrows show the process of creating an experimental workflow, generating a simulation and physical execution. Dashed black arrows show stages where reiteration of the workflow may be necessary.

1.4 Transaminases

Due to the generic nature of Synthace, as described in Section 1.3, it can be used to automate virtually any type of biological experiment on a wide range of automated devices. To date, most of the applications include molecular biology protocols such as DNA assembly, qPCR (quantitative polymerase chain reaction), and ELISAs (enzyme-linked immunosorbent assay) (https://www.synthace.com/#protocols), this thesis will examine the application of Synthace-driven DoE approaches for biocatalytic reaction screening and optimisation. Specifically, it will focus on industrially relevant transaminase catalysed bioconversions and activity assays that can be used in enzyme improvement programs or to guide optimisation of larger-scale bioreactor operations.

1.4.1 Introduction to transaminases

Transaminases (TAm) or aminotransferases are a class of enzymes that catalyse the transfer of an amino group from an amine donor to a ketone substrate (Brunhuber and Blanchard, 1994; Hyun and Davidson, 1995). They are dependent on the cofactor pyridoxal 5'- phosphate (PLP) to catalyse the reaction (Figure 1.4) (John, 1995). *In vivo*, transaminases catalyse the formation of amino acids by breaking down proteins (Braunstein. and Kritzman., 1937).

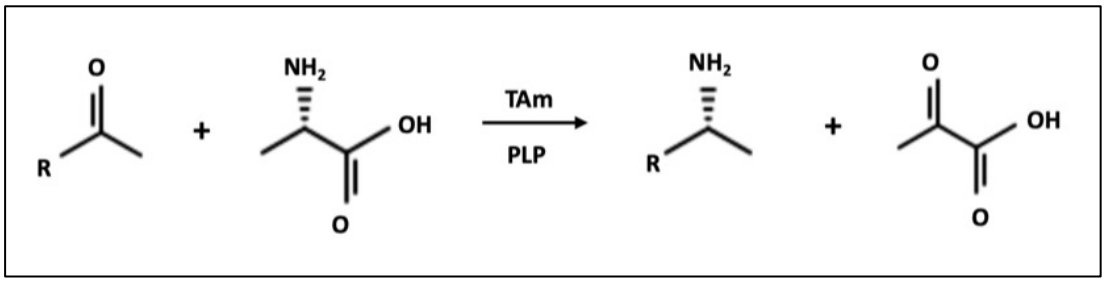


Figure 1.4. Schematic of a general mechanism of a TAm reaction showing the transfer of an amino group to a carbonyl group.

Transaminases are classified into four groups (I – IV) based on their primary (Mehta, Hale and Christen, 1993; Sayer, Isupov and Littlechild, 2007). All groups are co-factor dependent. Groups I, III and IV are comprised of α-TAms and Group II is comprised of

ω-TAMs (Mehta and Christen, 1994). The ω-TAMs have been studied more extensively and are considered the most important group of TAm as they can access a wider range of substrates that cannot be accessed by α-TAm groups and do not require the presence of an α-amino or α-keto acid (Shin and Kim, 1999; Koszelewski *et al.*, 2010). Their significance within industry has grown over the last few decades due to their ability to synthesise amino acids and chiral amines (Stewart, 2001).

1.4.2 Structure of transaminase

Consequently, many ω-TAMs have been studied over the last two decades, early examples include *Klebsiella pneumoniae, Bacillus thuringiensis JS64, and Vibrio fluvialis JS17* (Shin and Kim, 1997, 1998; Shin *et al.*, 2003). In 2007, the ω-TAm enzyme from *Chromobacterium violaceum* called CV2025 was identified from PSI-BLAST searches by querying the *Vibrio fluvialis JS17* TAm gene sequence and cloned into pET29a vector and expressed in *E. coli* BL21 cells (Kaulmann *et al.*, 2007). It exhibited strong kinetic resolution due to high selectivity towards aromatic (S)-amines such as S-MBA and asymmetric synthesis in the forward direction, making it a promising catalyst for the preparation of enantiopure amines and amino alcohols (Chen, Berglund and Humble, 2018; Voss *et al.*, 2018). CV2025 will be used as the model enzyme for the work done in this thesis. Figure 1.5 shows the ribbon structure of CV2025 which is a homodimer with a molecular mass of ~100 kDa (Humble *et al.*, 2012).

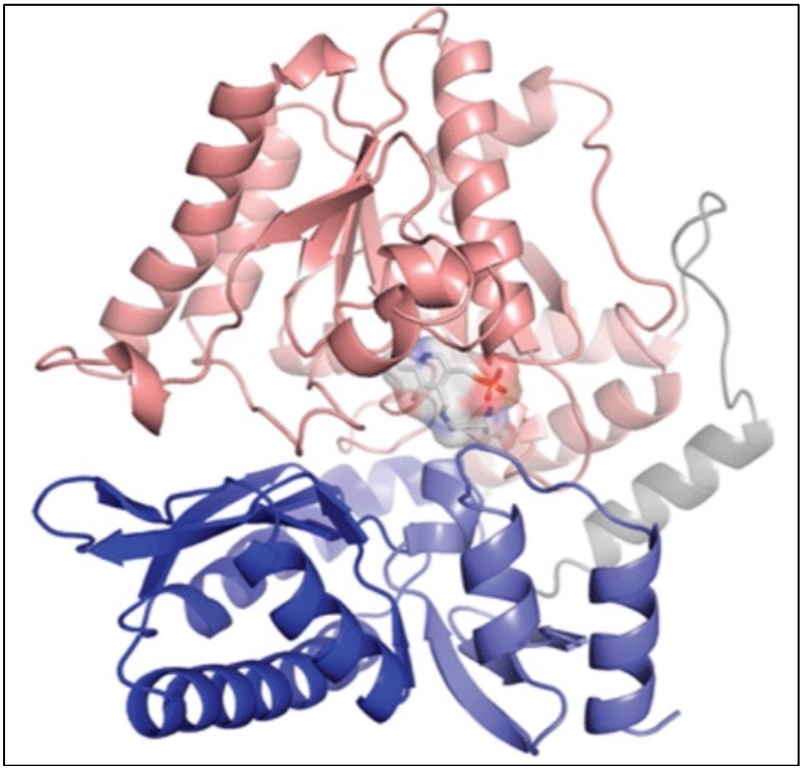


Figure 1.5. Ribbon structure of holo form TAm from *Chromobacterium violaceum* CV2025. A large domain (pink) and two smaller domains (grey and blue, respectively) are shown along with a linker (grey), and the active site is shown using a semi-transparent surface. Adapted from (Humble *et al.*, 2012)

1.4.3 Transaminase reaction mechanism

Transaminases (TAm) catalyse the transfer of an amino group from an amine donor to a ketone substrate via a ping-pong bi-bi mechanism resulting in the synthesis of optically pure chiral amines (Bulos and Handler, 1965a; Kuramitsu *et al.*, 1990a). In this mechanism, biocatalysis takes place in two half-reactions, in the first half, the amine donor binds to the TAm to transfer the amino group to PLP to form pyridoxamine 5-phosphate (PMP) and release the keto product. In the second half, the PMP-TAm complex transfers the amino group to the amine acceptor substrate which leads to PLP regeneration and the synthesis of the aminated compound.

1.4.4 Industrial applications of transaminases

Numerous TAms with broad substrate specificities have been identified *ex vivo* over the last few decades, which has increased their industrial popularity because they

can be used for the synthesis of amino acids and chiral amines, which are critical in the development of fine chemicals and pharmaceuticals, making TAms industrially appealing biocatalysts. (Satyawali *et al.*, 2017; Kelly *et al.*, 2020). TAms offer a sustainable green alternative to chemo catalytic methods and have been successfully incorporated into large-scale industrial synthesis (reviewed in (Kelly *et al.*, 2020), (Kohls, Steffen-Munsberg and Höhne, 2014)). The production of Sitagliptin, an anti-diabetic medication whose rhodium-catalysed synthesis was substituted with a TAm-catalysed method, is the most notable example (Saville *et al.*, 2010). This was a significant advancement over chemical synthesis, which requires toxic heavy metal catalysts and harsh conditions.

The search for industrially capable TAms with high activity, thermostability, solvent stability, enantioselectivity, and other desirable properties has led to researchers using metagenomics and protein engineering methods to identify and develop a variety of TAms (Gupta, 1992; Deszcz *et al.*, 2015; Leipold *et al.*, 2019). In turn, efficient and cost-effective biocatalytic reaction screening methods that can evaluate the characteristics of TAms are required before they can be considered industrially suitable biocatalysts.

1.4.5 High-throughput screening assays

As the interest in TAms has increased, various high-throughput assays have been developed for investigating TAms that have suitable characteristics of industrial catalysts. High-throughput assays are more efficient than techniques such as HPLC (High-performance liquid chromatography), mass spectroscopy, and NMR (nuclear magnetic resonance) which can often be low-throughput, manual, and time-consuming. Therefore, developing high-throughput assays that can be conducted in microtiter plates while maintaining sensitivity and robustness are needed for screening and characterising TAms.

In turn, many high throughput assays based on various principles have been developed with an aim to quickly evaluate important enzyme characteristics. These include enrichment methodology (Shin and Kim, 2001), pH sensor methods (Hopwood *et al.*, 2011), spectrophotometric detection methods (Hwang and Kim, 2004; Schätzle *et al.*, 2009; Hopwood *et al.*, 2011; Sehl *et al.*, 2012), colorimetric detection methods (Baud *et al.*, 2015) and conductometric methods (Ho, Robins and Bornscheuer, 2010). In this thesis, a spectrophotometric assay that quantifies the conversion of the amine donor, S-methylbenzylamine (*S-MBA*) and amine acceptor, Sodium Pyruvate into Sodium Alanine and Acetophenone will be used as the model system (Schätzle *et al.*, 2009).

While a range of high-throughput assays are now available, there is further scope for both improving biocatalysis screening conditions enzymes are subjected to while also improving the process of executing the assay by automating assay preparation. Additionally, optimising many non-biological considerations or 'assay logistics' that impact assay precision, such as ensuring appropriate labware, reaction volumes, automated devices, etc., can also be improved to create efficient and cost-effective frameworks for biocatalytic reaction screening.

1.5 Aim and objectives

DoE approaches provide a powerful and efficient way to systematically investigate complex biological systems and have been extensively adopted within industrial QbD frameworks (Section 1.1.5) Its adoption in other biological research has been slow due to many perceived limitations, one of which is its manual nature of preparing run conditions. Executing a DoE study requires the preparation of randomised run conditions, which in turn requires highly variable and unintuitive pipetting steps, making the process of executing a DoE study complicated and error-prone (Section 1.1.5). As a result, only a limited set of DoE designs are manually executed, whereas other modern design types that are suited to exploring high-dimensionality spaces are not currently used since they are beyond the scope of manual execution (Gilman *et al.*, 2020).

Automating DoE execution may seem like an ideal solution for executing DoE studies, however, this is largely unfeasible as current automation platforms are primarily used to conduct repetitive experiments where the same liquid-handling script can be prepared once and used repeatedly e.g., preparation of end-point assays. Often, specific programming skills and knowledge of device-specific control software are required to prepare liquid-handling scripts which limit the flexibility of automating various biological experiments, e.g., bioconversion reactions (Section 1.2.2). Currently, the lack of generic software tools that can interface with numerous liquid handlers and associated platforms and be applied to all types of biological experiments is still very limited. In this thesis, automated experimentation will be established using Synthace (Synthace Ltd) (Section 1.3) throughout. This has been chosen as it enables the automation of virtually any type of liquid-handling experiment across a range of automated devices.

In this thesis, the activity of an ω-transaminase enzyme from *Chromobacterium violaceum* called CV2025 (Kaulmann *et al.*, 2007) was screened using a spectrophotometric activity assay (Schätzle *et al.*, 2009) (Section 1.4.5). It was used

as the model system to demonstrate the application of high-dimensionality experiments which could provide insight into significant factors and factor interactions and lead to thorough characterisation of the assay's design space. It was selected over assays that require equipment such as HPLC, GC, and NMR as they can quickly generate large quantities of data and are capable of parallel measurements.

Using a spectrophotometric assay is beneficial as large quantities of absorbance data can be generated from automated DoE studies to calculate many metrics that are necessary when understanding biocatalytic reaction screening such as reaction rate, yield, stability, etc. However, handling large quantities of data and calculating these metrics require data processing frameworks that can use absorbance data to automatically calculate the metrics of interest. This functionality is missing from most current packages and is essential to the automation and analysis of high-dimensionality experiments.

The overall aim of this thesis is to establish a framework for the automated execution and analysis of high-dimensionality experiments created using DoE methodology. Successful implementation of these approaches will provide generic frameworks that can be implemented in industry by groups focusing on biocatalytic reaction screening and assay development.

The objectives necessary to achieve this aim are detailed below.

- The first objective is to establish an automated protocol for a manual high-throughput TAm activity assay. Experiments that follow DoE methodology will be conducted to identify fast and precise assay conditions by performing screening and optimisation DoE studies. All automated experiments will be defined and executed using Synthace. This work is described in Chapter 3.
- The second objective is to refine the non-biological considerations that could impact automated high-throughput spectrophotometric assays. As many things can impact assay performance, this process was carried out to ensure

any future high-dimensionality experiments can be executed This work is described in Chapter 4.

- The next objective is to integrate the methods established in Chapter 3 and 4 in order to establish a generalised framework for the one-step characterisation of a large unfamiliar design space with automated execution and data analysis. A modern type of DoE design, called a space-filling design, that was previously out of reach for physical experimentation will be used to explore an unfamiliar 12-dimensional design space to gain further insight into the design space, MATLAB scripts were used to process absorbance data and automatically generate responses of initial rate and yield for model generation. This work is described in Chapter 5.
- Lastly, the learnings from Chapter 5 will be critically evaluated in order to outline the key challenges and considerations that surround the commercial implementation of space-filling design preparation, execution, and analysis. This work is described in Chapter 6 and is a requirement for the award of a UCL Doctor of Engineering (EngD).

Furthermore, Chapter 2 describes the materials and methods used to prepare, execute, and analyse all experiments. Chapter 7 discusses the main experimental findings and conclusions of this project and suggests ideas for future work.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

Deionised water was used throughout this work. Unless stated otherwise, materials were purchased from Sigma Aldrich, Dorset, UK and were of analytical grade or better. They were stored according to the manufacturer's instructions.

Kanamycin was prepared at a stock concentration of 50 μg/ml in water, sterilised by filtration through a 0.22 μm cellulose acetate syringe filter (VWR, Leicestershire, UK) and stored at -20 °C. It was thawed and used at a working concentration of 25 μg/ml to supplement all bacterial growth media. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was prepared at a stock concentration of 1 M in water, sterilised by filtration through a 0.22 μm syringe filter and stored at -20 °C. It was used at a working concentration of 1 mM to induce protein expression.

Pyridoxal 5'-Phosphate (PLP) was prepared at a stock concentration of 10 mM in water. During preparation, the beaker was covered in aluminium foil as PLP is light sensitive and aliquots were stored at -20 °C in black micro centrifuge tubes. Tartrazine (Thermo Fisher Scientific, Leicestershire, UK) was prepared in water to a concentration of 1.6 mM and was stored at room temperature, away from bright light. Sodium chloride (NaCl) and Potassium chloride (KCL) (both Thermo Fisher Scientific, Leicestershire, UK) were prepared at a stock concentration of 2M, sterilized by autoclaving at 121 °C for 15 minutes, and stored at room temperature.

Phosphate (pH 7.4), HEPES (pH 7.4) and TRIS-HCl (pH 9) buffers at a stock concentration of 1 M, and CHES buffers (Alfa Aesar, Lancashire, UK) at pH 9, 9.5 and 10 at a stock concentration of 500 mM were purchased and stored according to the manufacturer's instructions. 1 M solutions of Hydrochloric Acid (HCl) and Sodium

hydroxide (NaOH) were used to adjust buffer pH if required. S-methylbenzylamine (S-MBA) and Acetophenone were prepared fresh daily at specified concentrations. Sodium Pyruvate ("Pyruvate") was purchased as a solution in water at a concentration of 100 mM.

2.1.2 Bacterial cell strains

Two plasmids with different versions of the CV2025 enzyme were used (Table 2.1). The first plasmid "pQR801" (Baud *et al.*, 2015) was obtained from the UCL Department of Biochemical Engineering. The second plasmid "pJexpress401_110499" was obtained from a previous project executed at Synthace Ltd.

Table 2.1 Characteristics of the two plasmids used during this work each containing a copy of the CV2025 transaminase enzyme.

Plasmid	Characteristics
pQR801 **(Baud *et al.*, 2015)**	pET-29a(+), T7 promoter, lacI, Km^r, encoding the CV2025 TAm gene from *Chromobacterium violaceum* DSM30191.
pJexpress401_110499	pJexpress401 Cloning cassette, T5 promoter, lacI, Km^R, codon optimised encoding the CV2025 TAm gene from *Chromobacterium violaceum* DSM30191.

Table 2.2 Characteristics of cell E. coli host strains used during this work.

Strain name	Characteristics
E. coli BL21 (DE3)	*F^-, ompT, hsdSB (r_B^- m_B^-), λ(DE3 [lacI lacUV5 T7 gene 1 Sam7* **Δ***nin5])*
E. coli BL21	*fhuA2 [lon] ompT gal [dcm]* **ΔhsdS**
E. coli DH5α	F^- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rk^-, mk^+) *pho*A *sup*E44 λ^- *thi^-* 1 *gyr*A96 *rel*A1

pQR801 was purified from 10 ml of *E. coli* DH5α cell culture broth (New England Biolabs Inc) overnight culture using the QIAprep Spin Miniprep Kit (Hilden, Germany) following the manufacturer's protocol with elution in deionised water. *E. coli* BL21 (DE3) competent cells (New England Biolabs Inc.) were transformed using 1 μL of purified pQR801 by heat shock as per the manufacturer's protocol.

A glycerol stock for the pJexpress401_110499 in *E. coli* BL21 cells was used to make fresh glycerol stocks by growing overnight cultures. Glycerol stocks of both strains were made by mixing overnight cultures to 50% v/v with pure glycerol. All glycerol stocks were stored at -80 °C.

2.1.3 Bacterial growth media

Luaria Bertani (LB) Broth (Merck) and LB Agar (Merck) were prepared according to the manufacturer's instructions. All media were sterilised by autoclaving at autoclave at 121 °C for 15 minutes and allowed to cool to room temperature before use.

2.1.4 Automated devices and labware

Two liquid handling devices, PIPETMAX® 268 (Gilson, Inc, Middleton, USA) and Hamilton Microlab STAR (Hamilton Company, Reno, NV) and one dispenser dragonfly®discovery (SPT Labtech, Melbourn, UK) were used for this project. A CLARIOstar Microplate Reader was used to generate all absorbance measurements (Section 2.4.1).

In terms of plasticware, all microplates adhered to standard microplate dimensions as specified by the Society of Biomolecular Screening (SBS). For any mixtures that required absorbance measurement, UV-clear flat bottom 96- and 384-well plates (UV Star, Greiner Bio-One) were used throughout. Reagents required for automated experiments were housed in the following types of 'input' plates, their suitability was

manually selected based on their volume capacity; Hard-Shell® 96-well PCR plates (Capacity: 5 – 200 µL, Part no: HSS9901, Bio-Rad Laboratories Inc.), Nunc™ 96-well Polypropylene Storage Microplates (Capacity: 10 – 450 µL, Part no: 249946, Thermo Fisher Scientific, Leicestershire, UK), 96-well deep well plate (Capacity: 420 µL – 2 ml (on the PipetMax) or 100 µL to 2 ml (on the Hamilton), Part no: 278743, Thermo Fisher Scientific, Leicestershire, UK), 12-Channel, Multi-well reservoirs (Capacity: 5 – 15 ml, Part no: E2999-8412, StarLab, Hamburg, Germany).

Microtiter plates in protocols executed using the PipetMax were placed on 18 mm risers (Model: SPL-2221A-HDW, Gilson, Inc, Middleton, USA). To control evaporation from the plates films was used. These included MicroAmp™ Optical Adhesive Film (Applied Biosystems®, Leicestershire, UK) and Aluminum Sealing Films (Starlab Ltd, Hamburg, Germany). Device specific pipette tips and syringes were used throughout.

2.2 Bacterial growth and cell lysis

2.2.1 CV2025 transaminase (TAm) expression

Overnight cultures were prepared by inoculating 5 ml of LB broth containing 50 µg/ml of Kanamycin in 50 ml centrifuge tubes (Sigma-Aldrich, Dorset, UK). They were grown overnight at 37 °C, with orbital shaking at 250 RPM in a MaxQ™ 6000 Incubated Stackable Shaker (Thermo Scientific). Overnight cultures were then used to subcultures fresh media at a ratio of 1:100 in sterile baffled Erlenmeyer flasks with a vented cap (Sigma-Aldrich, Dorset, UK) and incubated at the same settings as the overnight culture. Protein production was induced by adding IPTG at a final concentration of 1 mM when the optical density of the cell culture reached an OD_{600} of 0.4-0.6 A.U.. Growth was then allowed to continue for a further 20 hours. Cultures were then transferred to 50 ml centrifuge tubes, with 10-20 ml per tube. Cells were harvested by centrifugation at 3000 RPM, at 12 °C for 20 minutes using a Sorvall Evolution RC (Thermo Scientific, Rotor: SLA-600TC) centrifuge. The supernatant was discarded, and the cell pellets were stored at -80 °C or lysed immediately.

2.2.2 Lysis buffer preparation

Lysis buffer was prepared fresh from 1 M Phosphate buffer, pH 7.4, 50 mg/ml Lysosyme (Part no: 90082, Thermo Fisher Scientific, Leicestershire, UK), 10 mM PLP, 10X BugBuster™ (Merck Millipore) stock solutions to 100 mM Phosphate buffer, 0.2mg/ml Lysosyme, 1mM, 0.2 mM PLP and 1X BugBuster™ final concentrations. 1 µL of Benzonase Nuclease was added per ml of BugBuster reagent used, to reduce the viscosity of the lysate by digesting nucleic acids.

2.2.3 Clarified cell lysate preparation

A cell pellet harvested from 10 ml of cell culture, produced as described in Section 2.2.1, was resuspended in 3.125 ml of 100 mM Phosphate buffer, pH 7.4, and added to 23.4380 ml of lysis buffer. The cell lysis mixture was then statically incubated at room temperature for 20 minutes prior to centrifugation at 10,000 RPM using a Sorvall Evolution RC (Thermo Scientific, Rotor: SLA-600TC) to prepare clarified cell lysate, or 'lysate' is it is referred to from here on. Lysate was stored at -80 °C for up to six months and aliquots were thawed on ice when required.

2.2.4 SDS-PAGE analysis for quantification of protein expression

All reagents and equipment were purchased from Bio-Rad Laboratories Inc, unless stated otherwise. The quantity of protein in the lysate (Section 2.2.3) was measured on a Qbit using the Qbit Protein Assay Kit according to manufacturer's instructions (Life Technologies). Sample concentrations were normalised to 0.4 µg/µL using water followed by addition to Laemmli sample buffer (2% SDS, 10% glycerol, 5% β-meracaptoethanol, 63 mM Tris-HCl pH 6.8, bromophenol blue) and heated for 5 minutes at 95 °C to denature the samples. Each gel lane was loaded with 4 µg of protein. 15 µL of the Precision Plus Protein™ All Blue Prestained Protein Standard (10 – 250 kDa) was used as the protein ladder.

SDS-PAGE (Sodium dodecyl sulphate – polyacrylamide gel electrophoresis) analysis was performed using a vertical Mini-PROTEAN Tetra Cell gel electrophoresis system which uses pre-cast 4-20 % v/v Bis-Tris gels (Bio-Rad Laboratories Inc). The gel tank was filled with 1X Tris Glycine SDS Buffer that was diluted from a stock concentration of 10X using deionised water. Gel electrophoresis was carried out for 15 minutes at 240V. Gels were carefully removed from the casting according to manufacturer's instructions and stained using InstantBlue® Coomassie Protein Stain (Abcam, Cambridge, UK) by gently shaking (200 revolutions/minute) for 30 minutes on an orbital shaker (Model: S05, Stuart Scientific, Staffordshire, UK). Gels were then separated from the plastic cast according to the manufacturer's instructions, rinsed with water and placed on the rocking platform in water with two knotted tissues to absorb any remaining dye overnight prior to imaging.

2.3 CV2025 TAm activity assays

2.3.1 Colorimetric CV2025 TAm assay

CV2025 TAm activity assays were performed both manually and in an automated fashion. Transamination reactions were prepared at a final volume of 200 μL in clear, flat-bottom 96-well plates (UV Star, Greiner Bio-One). 55 μL of CV2025 TAm clarified cell lysate was added to a solution containing a final concentration of 100 mM phosphate buffer (pH 7), 0.2 mM PLP, 25 mM of 2-(4-nitrophenyl) ethan-1-amine hydrochloride (the amine donor), 10 mM of Sodium pyruvate (amine acceptor) to begin the bioconversion. Reactions were incubated for up to 24 hours at 30 °C in a thermomixer (Thermomixer C, Eppendorf, Hamburg, Germany) at 500 RPM for the formation of the red precipitate. The reaction between the amine donor (2-(4-nitrophenyl) ethan-1-amine) and its bioconverted aldehyde, followed by tautomerisation gives rise to a visible red precipitate (reaction mechanism can be found in Appendix A) Negative controls excluded the amine donor from the bioconversion mixture.

2.3.2 Acetophenone Spectrophotometric Assay (AP Assay)

Acetophenone formation from CV2025 TAm bioconversions (Figure 3.1) was determined spectrophotometrically at 245 nm, unless started otherwise. Assays was prepared in 96- or 384- well UV clear microtiter plates (UV Star, Greiner Bio-One). When the assay was prepared at various pH values, the pH of specified buffers was adjusted and monitored with a calibrated pH probe (MP220 pH meter, Mettler Toledo, Leicester, UK) beforehand using acids and bases specified in Section 2.1.1.

2.3.2.1 Original AP assay

The original protocol was adapted from a protocol developed by Dr. Fiona Truscott based on the paper where this assay was first described (Schätzle *et al.*, 2009). It was initially used to automate the AP assay and conduct scoping studies prior to the Screening DoE.

Transamination reactions were prepared at a final volume of 200 µL. 55 µL of CV2025 TAm lysate, prepared as described in Section 2.2.3, was used to initiate bioconversion by adding to a reaction mixture containing 2.5 mM S-MBA, 1 mM Sodium Pyruvate, 0.2 mM PLP in 100 mM Phosphate Buffer at pH 7.4 at 30 °C. Change in absorbance was measured immediately after lysate addition.

2.3.2.2 Improved AP assay

Assay conditions defined in Section 2.3.2.1 were improved after executing the Screening DoE study (Section 3.2.5). The resulting analysis was used to identify the following improved reaction conditions; 5 mM of S-MBA and 5 mM of Pyruvate, 10 % v/v DMSO in 100 mM TRIS-HCl buffer at pH 9. These reaction conditions were prepared at specified volumes, transamination at 25 °C was initiated by a adding an appropriate quantity of lysate, after which, endpoint or time-course absorbance

measurements were recorded after a specified amount of time. These reaction conditions were used as a reference reaction in Chapters 4 and 5.

2.4 Spectrophotometer settings and calibration

2.4.1 Measurement settings

A CLARIOstar Microplate Reader (BMG Labtech, Ortenberg, Germany) was used for all spectrophotometric measurements. Measurements were recorded after 22 flashes of the light beam per well per measurement cycle, the plate was shaken for 10 seconds at 500 RPM prior to measurement before each cycle. Here, a cycle refers to one full measurement of all specified wells a microtiter plate. A bidirectional, vertical left to right, top to bottom reading direction was used for all experiments. If the experiment required measurement at specific wavelengths, up to 8 discrete wavelengths were defined, otherwise, absorption spectra was measured in increments of 1 nm over a maximum range of 220 – 1000 nm.

2.4.2 Determination of the linear range

2.4.2.1 Calibration Curves

Calibration curves were prepared by measuring the absorbance of various reagents at 245 nm using the CLARIOstar plate reader. All standard curves were generated using automated liquid handlers, manually prepared reagent stock solutions were used to generate up to eleven 2-fold serial dilutions in triplicate with no tip-reuse throughout. Spectral scans ranging from 220 – 1000 nm were measured for specific dilutions of the calibration curves to compare the absorption spectra of various AP assay components; examples of calibration curves and spectra are shown in Appendix B.

2.4.2.2 Calculation of absorbance as a function of reaction volume and product concentration

The following equation was used to understand if the absorbance generated from acetophenone concentration curves prepared at various reaction volumes in both 96- and 384-well plates were comparable:

$$\text{Normalised VC } (\mu L.\ mM.\ mm^{-2}) = (V .C) / a \quad \ldots\ldots\ldots\ldots 2.1$$

where V = Volume (µL), C = Concentration (mM), a = area of the well base (mm^2)

The area was calculated from the manufacturer's plate maps. 32.07 mm^2 and 10.89 mm^2 were used as the area of the well base for 96- and 384- well plates, respectively.

2.4.2.3 Calculation of maximum detectable Acetophenone concentration

To identify the maximum concentration of Acetophenone that could be quantified within the linear range of the plate reader for various reaction volumes and plate types, the following equation was used:

Max. detectable Acetophenone concentration (mM) = (Normalised VC x a)/V……..2.2

where Normalised VC values are taken from Section 2.4.2.2, a = area of the well base (mm^2), V = Volume (µL).

2.4.3 Calculation of change in absorbance (ΔAbsorbance)

ΔAbsorbance (Figure 2.1) was the response used to analyse the screening and optimisation DoE studies described in Chapter 3 and was calculated using the raw absorbance files with Excel.

ΔAbsorbance calculated for the analysis of the screening DoE study used the first time point absorbance measurement as the blank value. As only end point

measurements were recorded for the optimisation DoE study, the blank value for each reaction was generated by preparing a separate set of reaction mixtures where lysate was excluded, these reactions were prepared using the same Synthace workflows as those that prepared actual runs.

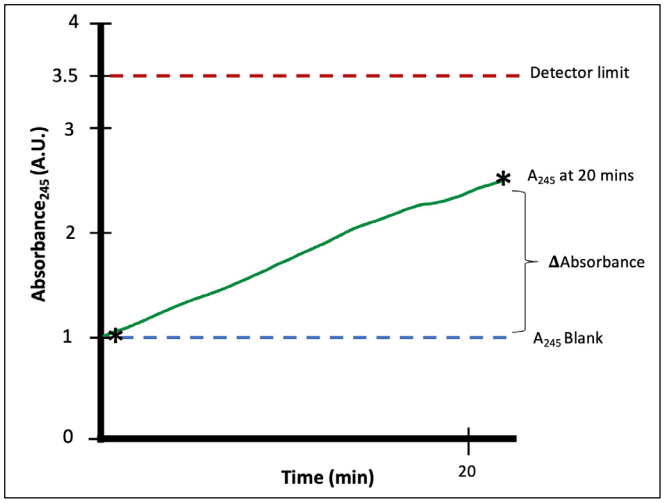


Figure 2.1. Schematic diagram showing an example of the measured change in absorbance over time and how ΔAbsorbance was calculated.

2.5 Liquid-handling precision testing

2.5.1 Artel MVS Verification System

A MVS Multichannel Verification System (Artel, Westbrook, ME) was used to evaluate liquid-handling precision by utilising a dual-dye, dual-wavelength volume measurement process. The ratio of the expected quantity of red dye (detected at 520 nm) to that of a fixed concentration of the blue dye (detected at 730 nm) is used to determine the precision of dispensing the quantity of red dye. Baseline solution, dyes, diluent, and verification plates were all ordered from Artel, and the precision tests were conducted as dictated by the Artel software, which automatically generated instructions and a test report containing the results (Bradshaw *et al.*, 2005).

2.5.2 Tartrazine precision testing

The liquid handling precision of the automation platforms was also quantified by preparing solutions at a fixed Tartrazine concentration of 0.1 mM. A tartrazine stock solution was prepared as previously described and was quantified by measuring absorbance at 425 nm.

2.6 DoE software, design, and analysis

2.6.1 Software

The majority of the work described in this thesis followed a statistical DoE approach (Section 1.1). JMP Pro Software (Versions: 14, 15, and 16, SAS Institute, North Carolina, USA) were used was used for DoE design and analysis in Chapter 3 and 5. All figures that contain Actual by Predicted plots, Prediction Profilers and Effect Summary tables were created within JMP. Excel files (Microsoft Office) in the .xlsx format were used to import and export data in and out of JMP e.g., into the Synthace software (Section 2.7.1) used for generating liquid-handling instructions to execute automated DoE studies.

2.6.2 Experimental designs

The 'custom design' function in JMP was used to create both the design used for the initial Screening DoE (Section 3.2.5.1) and the subsequent Optimisation DoE studies (Section 3.2.6.3). The design was a type of fractional factorial design, selected to enable the prediction of all main effects and two factor interactions, and consisted of 186 runs. The D-optimality criterion (Goos, 2002) is used by default. Six midpoints were added manually after the design was generated, totalling 192 runs (Appendix C and Appendix D).

The Optimisation DoE study used Response Surface Methodology (Anderson and Whitcomb, 2016) which automatically adds main effect, interactions, and quadratic terms up to second order for continuous factors, was selected. This adds the I-

optimality criterion (Goos and Jones, 2011) by default. The design consisted of 48 runs (Appendix E).

The space-filling DoE design was created using the 'Space Filling Design' function under the Special Purpose section of the JMP software. Only one type of space-filling design called 'Fast Flexible Filling' (FFF) could accommodate the categorical factors included in the design (Joseph, Gul and Ba, 2015). It was used to generate a 138-run design (Appendix J). To ensure that runs using TRIS buffer had a pH 7, 8 or 9 and CHES buffer have a pH of 9 or 10, a script written in the JMP Scripting Language (JSL) was added to the Disallowed Combinations box prior to design generation.

2.6.3 Stepwise regression model

Stepwise regression models (Miller, 2002) were used to analyse all DoE studies in this thesis and were created using JMP's Fit Model function. All single, two-factor and polynomial terms were provided to the model. P-value threshold with a lower and upper significance level of 0.1 and 0.25, respectively, in the mixed direction, was used to select the filtered model terms. Any effects with a p-value of > 0.05 were manually excluded from the model once the report was generated.

2.6.4 Gaussian Process model

Gaussian Process modelling (Jones and Johnson, 2009), housed under JMP's Specialized Modelling was used to analyse the space-filling design data. The model was created after incorporating the nugget parameter was incorporated which introduces a ridge parameter that smooths over noisy regions of the design space (Erickson, Ankenman and Sanchez, 2018). The Gaussian correlation type was selected.

The prediction formula and the jackknife predicted values were saved as columns on the data table. The former was used to calculate predicted values for validation experiments and the latter was used to create the actual by predicted plot which was used to check the goodness of model fit.

2.6.5 Artificial Neural Network model

Neural Network modelling (Gotwalt, 2012), housed under JMP's specialized modelling function was also used to analyse the space-filling design data. The random holdback validation method that divides 2/3 and 1/3 of the dataset into training and validation groups was used. The hidden layer structure consisted of a single layer with 3 TanH notes to activate the model. Goodness of Fit was determined by the R^2 values found in the model reports for both training and validation datasets.

2.7 Automated execution and data processing

2.7.1 Synthace software and workflow preparation

Synthace software (Synthace Ltd, London, UK), as described in Section 1.3 was used to generate liquid-handling workflows and control the various liquid handling platforms (Section 1.1.4). DoE designs generated in JMP were imported into Synthace via .xlsx files as described in Section 1.3 and shown diagrammatically in Figure 3.6. A summary of the various Synthace elements used to create all experimental workflows for automated execution are listed in Table 2.3.

Each element is further configured by instance parameters that need to be manually populated to generate device-specific liquid-handling instructions. Instance parameters are the 'inputs' needed by each element to perform its function, e.g., the 'Define Liquid' element has an instance parameter called 'Liquid Policy' that is populated to determine how the liquid will be physically handled by the automated device. The instance parameters of the 'Define Liquid' element are shown in Figure 2.2, another example can be found in Figure 3.11.

Table 2.3. Synthace elements used to create generate automated workflows.

Element Name	Function
Define Liquid	Defines a single custom-described liquid.
Define Liquid Set	Defines a set of custom-described liquids.
Append Liquid Sets	Joins single and/or sets of liquids sets together into one liquid set.
Split Workflow	Ensures that all liquid-handling actions on a set of liquids that were defined upstream of the element will be performed prior to actions defined downstream.
Aliquot Liquid Set	Instruct a liquid handler or dispenser to aliquot some liquids.
Upload Mix Set Plan	Uploads an XLSX-formatted file that specifies a set of liquid mixtures, detailing the individual components to be mixed and their order of mixture, for precisely-controlled mixing by a downstream Mix Set element.
Mix Set	This element makes a set of liquid mixtures, based on a "mix-set plan" defined by an upstream element.
Mix Set Onto Set	Transfers aliquots of a set of liquids, at user-specified volumes, onto a different set of liquids, in an ordered, pairwise fashion.
Upload Plate Layout File	Uploads a CSV plate file that contains liquids defined in a well-by-well layout of a single plate.
Prompt on Liquid Set	Pauses a workflow via a set of liquids and displays a user-defined prompt message via the liquid-handling robot software.
Serial Dilute by Factor	Takes a set of liquids and makes a serial dilution series from each, following user-specified dilution factors.
Parse DoE file	This element parses a DoE file exported from Design Expert (DX) or JMP in XLSX format.
Auto generate stock solutions*	Identifies liquids in your DoE design, then generates stock solutions for them.
Run DoE	Simplified DoE setup element, experimental stages are defined within this element.
Export DoE file	Exports DoE runs into an output file.

* Auto generate stock solutions element was used in combination with the Define Liquid Set element to define all input solutions needed for the space-filling DoE.

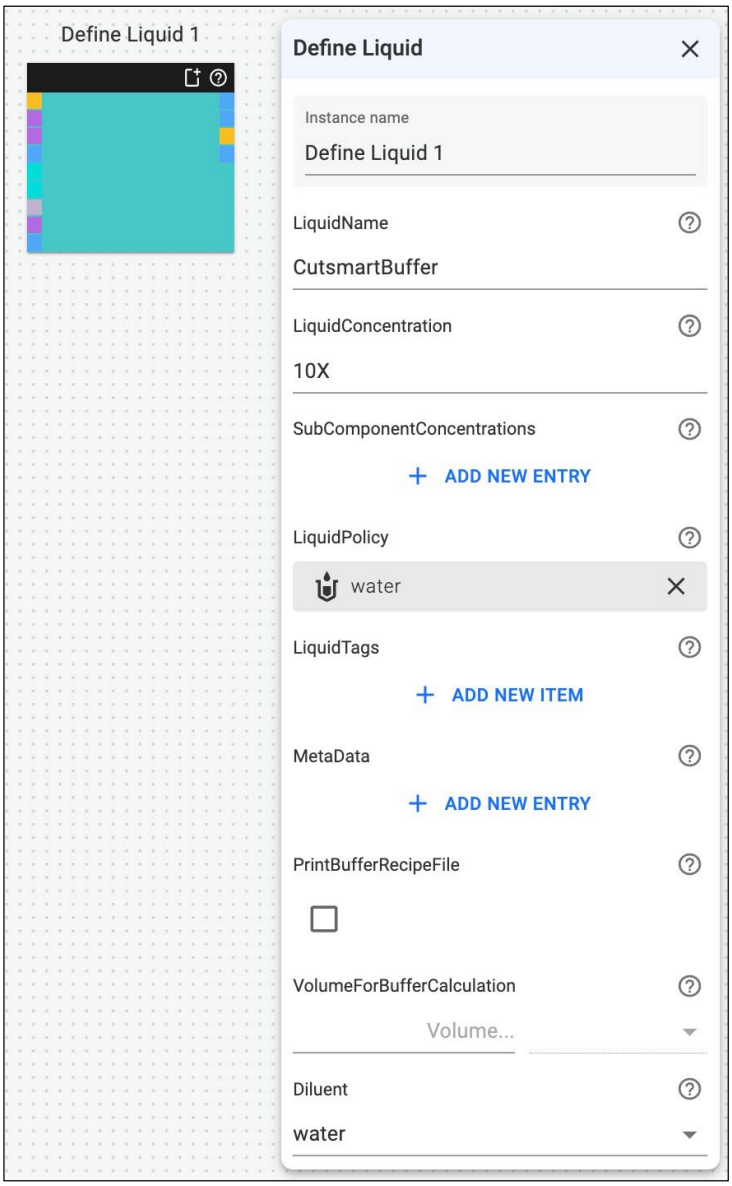


Figure 2.2. Example of instance parameters of a Synthace element.
The 'Define Liquid' element (blue box) consists of various instance parameters (white panel) that are manually populated by the user to describe the inputs of element within a Synthace workflow.

2.7.2 Data processing and calculation of responses

As part of this research, automated approaches for processing the large quantity of absorbance data generated during the space-fill DoE study (Section 2.6.2) were developed as for analysis described in Chapter 5.

2.7.2.1 MATLAB Data processing pipeline

During automated DoE execution, absorbance data indicating the change in absorbance at 245 nm with time in individual microwells was generated by the CLARIOstar plate reader and stored in .xlsx files. Figure 5.6 illustrates the type of data generated.

The steps taken to calculate the two responses used to analyse the space-filling DoE data; yield (% conversion of starting substrate) and initial rate (ΔA/s) responses from the raw time-course absorbance files are outlined in the flowchart Figure 2.3 (Further described in Section 2.7.2.2).

MATLAB (MathWorks, MA, USA) was used to develop scripts for automated data processing. All functions and scripts that were developed can be found in the following link: https://drive.google.com/drive/folders/1PCW5q23IPNJY0xob3KljWNxE29XRYDjj?usp=sharing.

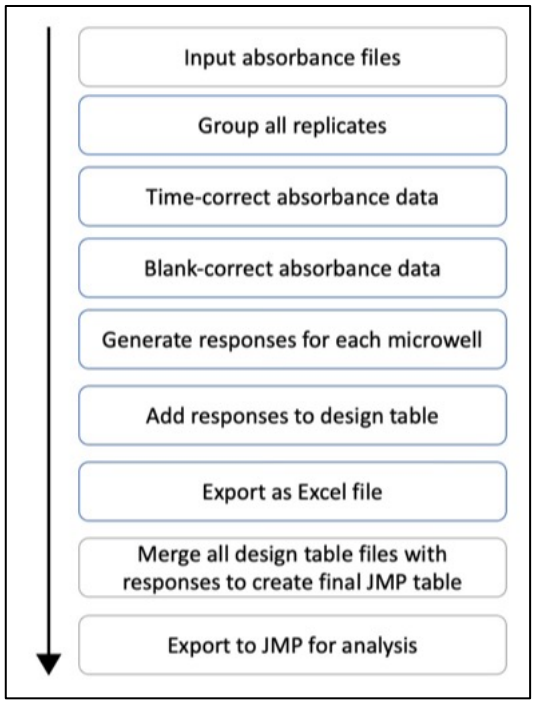


Figure 2.3. Schematic diagram of the data processing pipeline used to calculate responses. Manual (grey) and automated (blue, MATLAB) data processing steps followed for response generation and subsequent JMP analysis are listed in order of execution (arrow).

MATLAB scripts were used a set to calculate each response. As replicates are spread within and across plates, progress curves were first grouped to allow cumulative visualisation of each replicate run condition in the same subplot (Figure 5.6). A smoothing spline was fit to each reaction using the Shape Language Modelling (SLM) toolbox. The Y-intercept of the spline fit was used as the blank value to blank correct each progress curve. This is described further in Figure 2.5.

Reactions were also time corrected as each reaction in a 96-well was initiated and measured at slightly different time points than recorded in the raw absorbance file; this is a consequence of automation platform operation since only one reaction is kickstarted on a 96-well plate at a time and it took a period of time to transfer each plate to the plate reader for the first absorbance measurement. Figure 2.4 outlines the relationship between time and the order of lysate addition and one cycle of absorbance measurement within the CLARIOstar plate reader. The Dragonfly dispenser adds the final layer of reagents which is either lysate or substrates depending on the plate, which requires 40 or 90 seconds respectively and is executed in a snaking pattern going from right to left, top to bottom. All plates were then sealed with a MicroAmp^{TM} Optical Adhesive Film and spun down ('Short' setting, Centrifuge 5430 R, Eppendorf, Stevenage, UK) for a few seconds to remove any bubbles and transferred to the spectrophotometer for measurement, adding a further ~60 seconds. Once the spectrophotometer starts taking OD measurements, 2 minutes is required to measure and save absorbance data for all 96 wells. This was done bidirectionally, from top to bottom and left to right. Actual timepoints were calculated for each well based on these patterns to ensure that each measurement, and subsequent interpolation is associated with an accurate time point.

Once blocking and time correction were accounted for, the MATLAB script calculated the responses (Section 2.7.2.2) required for JMP analysis (Sections 2.6.3, 2.6.4, 2.6.5).

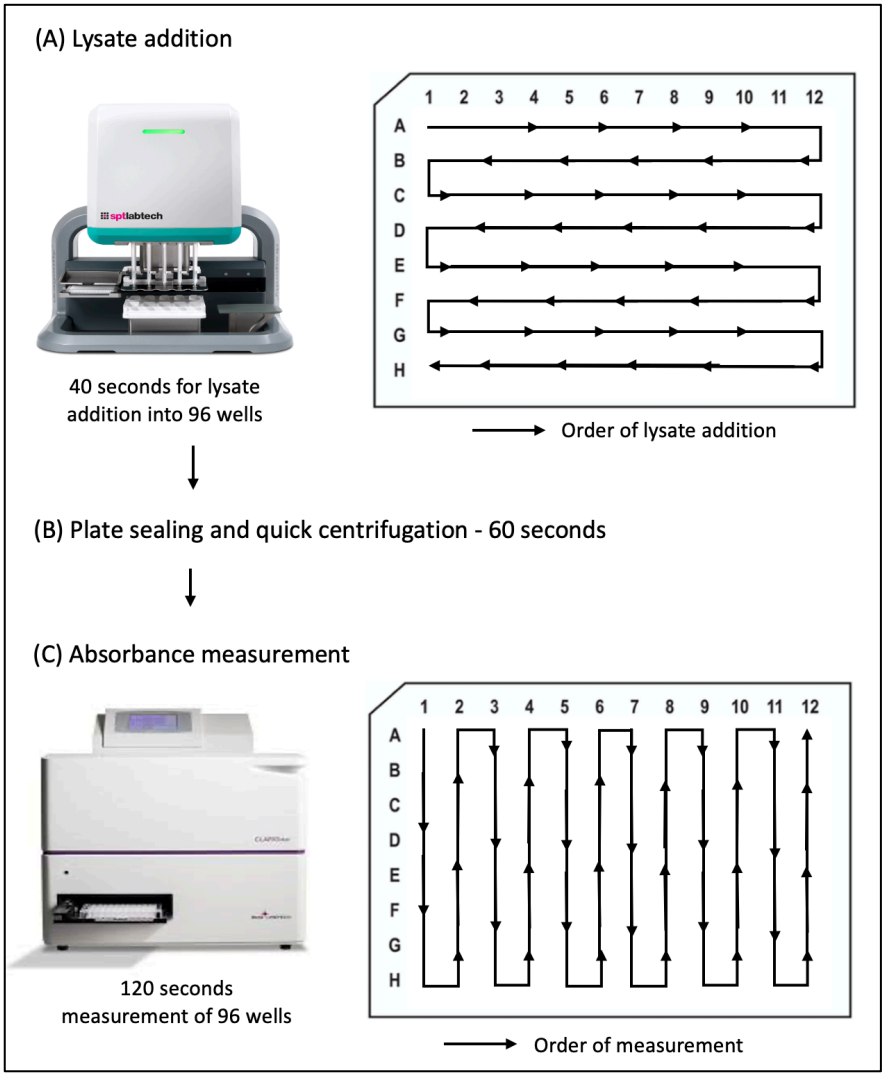


Figure 2.4. Illustration of the relationship between the start of bioconversion reactions (by lysate addition) and optical density measurement as performed by the laboratory automation devices.

The time and steps required from lysate addition to the first spectrophotometric measurement cycle is shown in (A), (B) and (C) with corresponding plate maps illustrating the order of dispensing and measurement using arrows.

2.7.2.2 Responses used to analyse space-filling DoE data

The space-filling DoE data was analysed using two responses that were calculated from the MATLAB script outlined in Section 2.7.2.1. The two responses were yield (% conversion of starting substrate) and initial rate (ΔA/s). Figure 2.5 summarise the approach taken to response calculation.

The yield response was based on the blank-corrected increase in absorbance after a fixed time of 60 minutes of reaction within the plate reader (Figure 2.5 A). Absorbance values were converted to acetophenone concentration using a standard curve (Figure 2.5 B). The calculated acetophenone concentration was then divided by the starting pyruvate concentration to calculate the percentage of substrate conversion, which was used as the yield response for JMP analysis.

The initial rate response was calculated by fitting a linear model to the first 4 minutes of absorbance (A) versus time (s) data which comprised of the first three recorded absorbance measurements; the slope of this linear fit is used as the initial rate response for JMP analysis and has the units of ΔA/s (Figure 2.5 C).

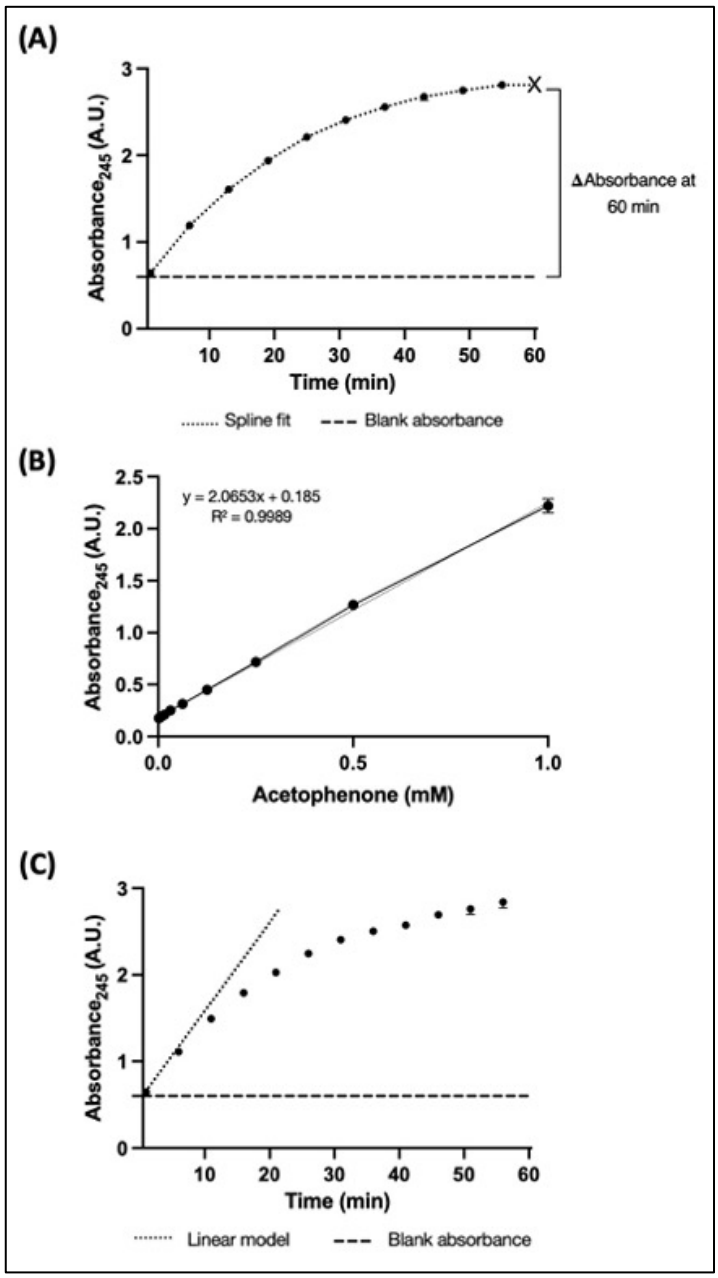


Figure 2.5. Visualisation of MATLAB response calculation from the OD versus time data from each microwell.

(A) Schematic illustration of calculation of the absorbance value used to generate the Yield (%) response. The black dots represent microplate absorbance readings obtained for one set of bioconversion conditions. The dotted line shows the fitted spline curve as described in Section 2.7.2.1. The change in absorbance was then calculated by interpolating the fitted absorbance values at 60 minutes (x) from the spline curve, and subtracting the value from the Y-intercept, providing a blank-corrected ΔAbsorbance value. (B) A typical calibration curve showing measured OD values (black circles) for known acetophenone concentrations and the fitted least squares linear regression model (solid line). The linear regression model was then used to convert absorbance measurements from (A) into acetophenone concentrations used to calculate yield (%). (C) Schematic illustration of fitting a linear model to the first 4 minutes of absorbance versus time data from each bioconversion to calculate the change in absorbance per second used as the initial rate (ΔA/s) response.

3 Automating DoE execution with Synthace

3.1 Introduction

The physical execution of experimentation is a major bottleneck for DoE implementation. Randomization of runs removes much of the structure (or pattern) from the initial conditions of an experiment, making manual setup slow, tedious, and error prone (Tye, 2004). This is exacerbated when a large number of factors need to be evaluated or when replication is required to quantify noisy experimental systems, both of which amplify the run number (Antony, 2014). In general, manual DoE execution becomes unrealistic for designs that require more than tens of runs (Gilman et al., 2021).

Automating DoE execution has the potential to overcome these challenges as described in Section 1.1.5. This is because automated devices can function without the need for structure or patterns typically required by scientists during manual run preparation. Current solutions for automating DoE execution, however, are limited either to specific devices or to specific applications that require significant programming expertise to generate scripts that can perform the liquid-handling actions required for DoE execution (Oritz et al., 2017) (Section 1.2.3). Software platforms like Synthace are currently being developed to enable scientists to easily interface with automated devices as described in Section 1.3.

3.1.1 Aim and objectives

The aim of this chapter is to demonstrate how Synthace can be used to automate the execution of standard, lower dimensionality, screening, and optimisation DoE studies. As described in Section 1.4.3, transaminase enzymes catalyse the synthesis of chiral amines from achiral substrates. The specific reaction studied will be the synthesis of sodium alanine from *S-MBA* (a widely used amine donor) and pyruvate

(a natural amine acceptor) by the CV2025 TAm (Figure 3.1). This reaction results in the production of acetophenone as a co-product which absorbs strongly at 245 nm (Schätzle *et al.*, 2009). The reaction can thus be continuously monitored spectrophotometrically providing quantitative data on the reaction rate and degree of conversion.

The specific objectives of this chapter are as follows:

1. Establish methods for the reliable production and recovery of the CV2025 TAm and replicate the published manual acetophenone assay.
2. Evaluate the assay design space, based on the published literature, and undertake initial scoping studies to identify factors to be used in an automated screening DoE study for refinement of selected factors and ranges.
3. Undertake an automated optimisation DoE study to characterise the improved design space and quantify the increases in the change in absorbance that can be achieved using standard I-optimal designs (Goos, Jones and Syafitri, 2000).

The optimisation DoE was executed in triplicate to map the noise in the design space arising from uncontrollable biological sources or technical sources such as pipetting discrepancies (Gilman *et al.*, 2021). It was also repeated using two automated devices (Section 2.1.4), a PipetMax liquid handler and a Dragonfly dispenser to compare the reproducibility of the response, and understand the impact of device type on the process of automated DoE execution, and to exemplify the device-agnostic capabilities of Synthace.

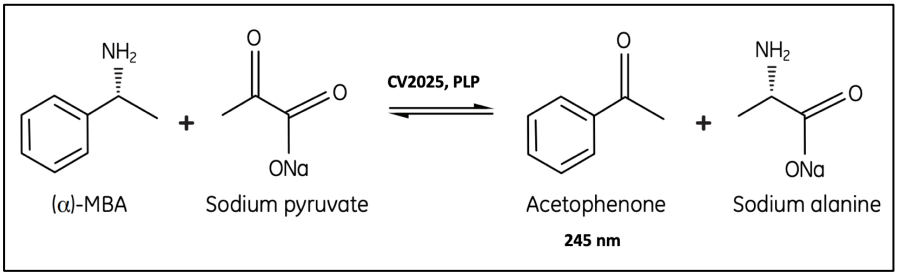


Figure 3.1. Principle of the acetophenone assay for monitoring w-TAm activity.
The amine donor and acceptor, *S-MBA* and sodium pyruvate are converted to acetophenone (AP) and sodium alanine by the pyridoxal phosphate (PLP) dependant CV2025 ω-transaminase (Schätzle et al., 2009). Acetophenone product formation at 245 nm is used to monitor ω-transaminase activity.

3.2 Results

3.2.1 Expression system selection

Two different plasmids containing the CV2025 TAm gene were initially evaluated with the intention of selecting the genetic construct that would yield the greatest quantity of soluble enzyme for use throughout this thesis. The plasmids, pJexpress401_110499 and pQR801 both contained the CV2025 TAm gene from *Chromobacterium violaceum* DSM3019 (Section 2.1.2). The former contained a codon optimised CV2025 gene cloned into a pJexpress401 cloning cassette containing a T5 promoter and was expressed in *E. coli* BL21 cells. The latter used the wild type CV2025 gene cloned into pET-29a(+) vector with a T7 promoter and was expressed in *E. coli* BL21 and *E. coli* BL21 (DE3) cells. Both plasmids contained the IPTG (Isopropyl β-D-1-thiogalactopyranoside)-inducible LacI repressor gene which regulated the transcription of the CV2025 TAm gene and conferred Kanamycin resistance (Section 2.1.2)

Each of the expression systems were cultivated overnight in shake flasks as described in Section 2.1.3. Soluble and insoluble protein fractions were then prepared and analysed by SDS-PAGE as shown in Figure 3.2. The pJexpress401_110499 plasmid was selected based on the results of SDS-PAGE analysis. While both enzymes expressed well, the results for the pJexpress401_110499 plasmid indicated a higher quantity of soluble protein in the soluble fraction. Clarified cell lysate from the pJexpress401_110499 plasmid, referred to as 'CV2025 TAm', from here on, was used to carry out all bioconversions for this project.

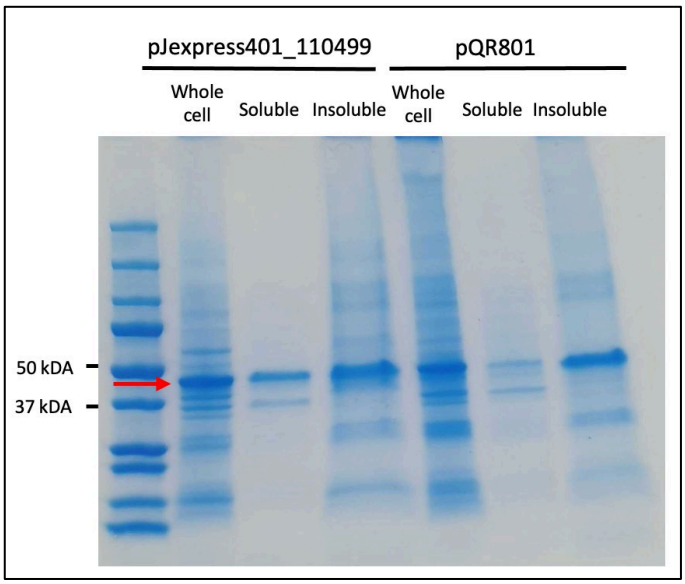


Figure 3.2. SDS-PAGE analysis of CV2025 ω-TAm expression in *E. coli* using two different plasmids (pJexpress401_110499 and pQR801).
4 µg of total protein containing whole cell, soluble and insoluble fractions isolated from CV2025 TAm expression from two different plasmids are compared following identical growth and lysis steps (Section 2.2.4). CV2025 TAm enzyme production was confirmed from both expression strains by the presence of a ~47 kDa band (red arrow). The presence of a more intense band in the soluble fraction of pJexpress401_110499 indicates that more soluble CV2025 is produced by this strain, whereas a more intense band in the insoluble fraction and fainter band in the soluble fraction for strain pQR801 indicate less soluble CV2025 production.

3.2.2 Assay selection

Initially, two transaminase activity assays were tested, with the aim of selecting one as the model system for the research carried out in this project. The first was a qualitative colorimetric assay (Baud *et al.*, 2015) where the reaction between the amine donor (2-(4-nitrophenyl) ethan-1-amine) and its bioconverted aldehyde, followed by tautomerisation gives rise to a visible red precipitate (Appendix A). The second was the AP assay described in Figure 3.1.

The advantage of the Baud assay was that it gave a clear, colorimetric signal and it was initially investigated to see if a qualitative relationship between the amine donor (2-(4-nitrophenyl)ethan-1-amine) and the amount of red precipitate formed could be sufficient for future DoE study analysis. Unfortunately, this was not selected as the assay of choice as the AP assay gave a rapid and quantitative measure of the CV2025

TAm bioconversions from spectrophotometric measurements and had the potential of generating a greater number of responses from time-course or end-point measurements. Furthermore, transamination could be measured within 1 hour using the AP assay while the colorimetric assay required an overnight incubation. For these reasons, the AP assay was selected as the model system for the work done in this thesis.

3.2.3 Confirming the accuracy of an automated AP assay

The first step toward automating a whole DoE experiment was to automate the assay manual handling steps. To do so, a PipetMax liquid handler was used (Figure 3.3).

Figure 3.3. A PipetMax liquid-handling device.
This is a benchtop liquid-handling device with nine deck positions that can house pipetting tip boxes, waste, and microtiter plates. Liquids are transferred and mixed from one position to another using a single- or multi-channel pipette that use up to eight pipette tips. (Image taken from https://ie.gilson.com/system-pipetmax.html).

To generate the liquid handling instructions for the PipetMax liquid handler, a Synthace workflow was created (Figure 3.4 A). The list of accessible elements used to create Synthace workflows can be found in Section 2.7.1. This workflow was physically executed using the PipetMax liquid handler to prepare the AP assay reaction mixtures and kickstart bioconversions to generate automated progress curves. These were compared to progress curves generated from normal manual preparation (Figure 3.4 B). An average increase in absorbance of 0.37 and 0.31 units

was observed at 245 nm over 60 minutes for both manual and automated AP assays, respectively. As both progress curves are largely comparable as they showed similar initial rates, however, the end-point absorbance of the manual assay was slightly greater than the automated version. It was concluded that automating the AP assay using the Synthace platform on the PipetMax liquid handler was as successful and reliable as the manual version as both versions gave similar reaction rates and overall conversion.

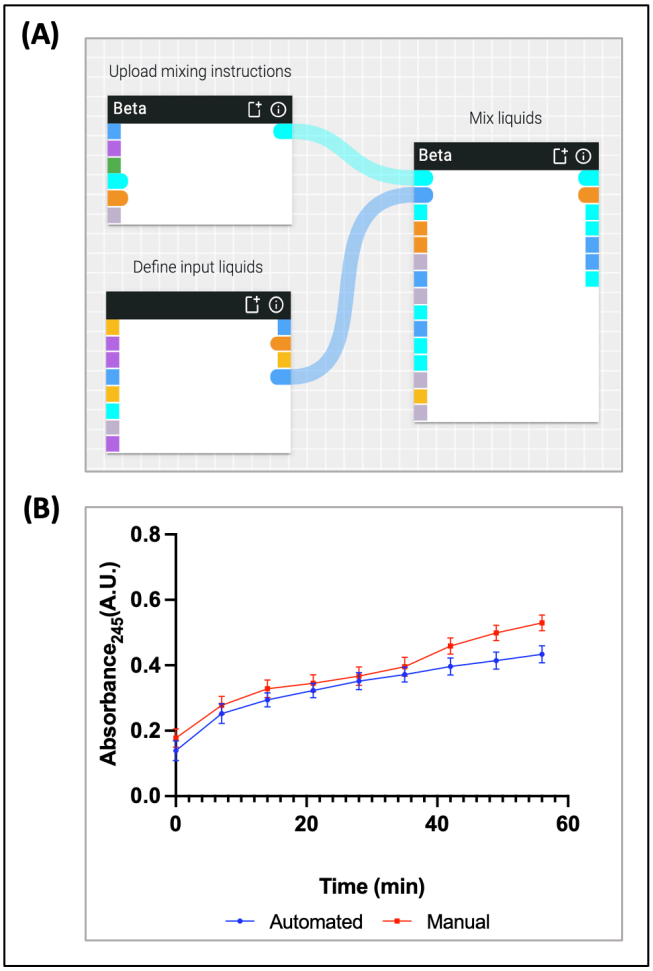


Figure 3.4. Synthace workflow and progress curves used for initial AP assay automation. (A) The Synthace workflow used to execute the automated AP assay. Synthace workflows are made by connecting Synthace elements (white boxes), each of which represents a core liquid-handling function (Section 2.7.1). In this workflow, the 'Upload Mix Set Plan' element was used to upload mixing instructions for the AP assay using an Excel spreadsheet that described the composition of a set of mixtures. The 'Define Liquid Set' element was used to define the input reagent stock solutions. This information was wired into the 'Mix Set' element which instructs the liquid handler to make the mixtures defined in the previous element using the reagents provided. In this case, mixing pyruvate, *S-MBA*, buffering agent, PLP and water together was followed by the addition of clarified cell lysate to trigger the conversion of pyruvate to acetophenone. (B) Comparison of automated and manual AP assay progress curves. The bioconversion of pyruvate to acetophenone by CV2025 was monitored by acetophenone absorbance at 245 nm every 2 minutes for 60 minutes (Section 2.4.1). Error bars represent one standard deviation from the mean (n=8 and n=3 for automated and manual assays, respectively).

3.2.4 Scoping studies for the screening DoE

A DoE campaign typically starts with one or more scoping studies as described in Section 1.1.5. A scoping study is typically an OFAT experiment conducted to determine if a factor is important enough to be included in a DoE design and to identify levels at which that factor must be sampled (Coleman and Montgomery, 1996; Politis *et al.*, 2017a).

For nine out of the 12 factors selected for screening, upper and lower levels were identified from literature mining. The finalised factor table can be found in Appendix C and Appendix D. However, it was not clear from the literature what the levels for PLP, DMSO, or buffer type should be. Therefore, scoping studies were performed to identify reasonable factor levels to include within the Screening DoE design.

Scoping experiments were performed to test PLP concentration, DMSO concentration and buffering agent (Figure 3.5). Transaminases are PLP-dependent enzymes, and the effect of supplemental PLP on long-term enzyme stability has been previously reported (Shin *et al.*, 2003; Kaulmann *et al.*, 2007). However, the benefit of supplementing additional PLP into an assay that is measured immediately after bioconversion is initiated (when the lysate already contains 0.2 mM PLP) had not been demonstrated. Therefore, investigation of whether additional supplementation of 0.2 mM PLP into the reaction immediately prior to substrate addition was performed (Figure 3.5 B). Although reactions supplemented with 0.2 mM PLP showed a marginal but insignificant increase in absorbance, it was therefore excluded as a factor for the screening DoE study. Additionally, PLP significantly contributes to the baseline absorbance (Appendix B) reducing the dynamic range of the assay, which is undesirable.

When the AP assay was first described, 0.25 % v/v DMSO was used in the bioconversion reaction (Schätzle et al., 2009) while other research has since tested CV2025 stability when stored in up to 50 % v/v DMSO (Chen *et al.*, 2016). To identify suitable factor levels, 0-25 % v/v DMSO was tested (Figure 3.5 A). Measurements

from reactions that used >15 % v/v DMSO saturated the plate reader's detector within 40 minutes, therefore, 0 and 10 % v/v were selected as factor levels as this ensured that the impact of DMSO could be investigated while staying within the physical limitations of the analytical equipment selected.

The most common buffering agents cited in the literature for CV2025 bioconversions are Phosphate, HEPES, and Tris HCl. However, these publications provide contradictory findings on which buffering agent provided the fastest reaction rates (Bea *et al.*, 2010; Chen, Berglund and Humble, 2018; Chen *et al.*, 2018). To identify the most suitable buffering agent, a comparative study on the impact of Phosphate, HEPES and TRIS HCl buffering agents at pH 7.4 was conducted to understand if the agent alone rather than the pH has an impact on reaction progression (Figure 3.5 C). An average increase in absorbance of 0.25 was observed with HEPES buffer while an increase of 0.47 was observed for both Phosphate and TRIS HCl buffers. CV2025 bioconversions are executed primarily between pH 7 – 9, therefore, TRIS HCl Buffer was selected for future work, as its pKa of ~8.1 makes it suitable for buffering within this range, unlike, HEPES and Phosphate Buffer, which have a pKa of 7.31 and 7.21, respectively. Additionally, inorganic phosphate was found to inhibit CV2025 activity by competing with PLP and blocking the phosphate group binding cup in the active site (Chen, Berglund and Humble, 2018). Therefore, it was decided that further studies using Phosphate based buffers would be avoided.

In summary, PLP was removed as a factor of interest, TRIS HCl was selected as a suitable buffering agent and the lower and upper levels of DMSO concentration for screening the AP assay were identified from these OFAT scoping studies.

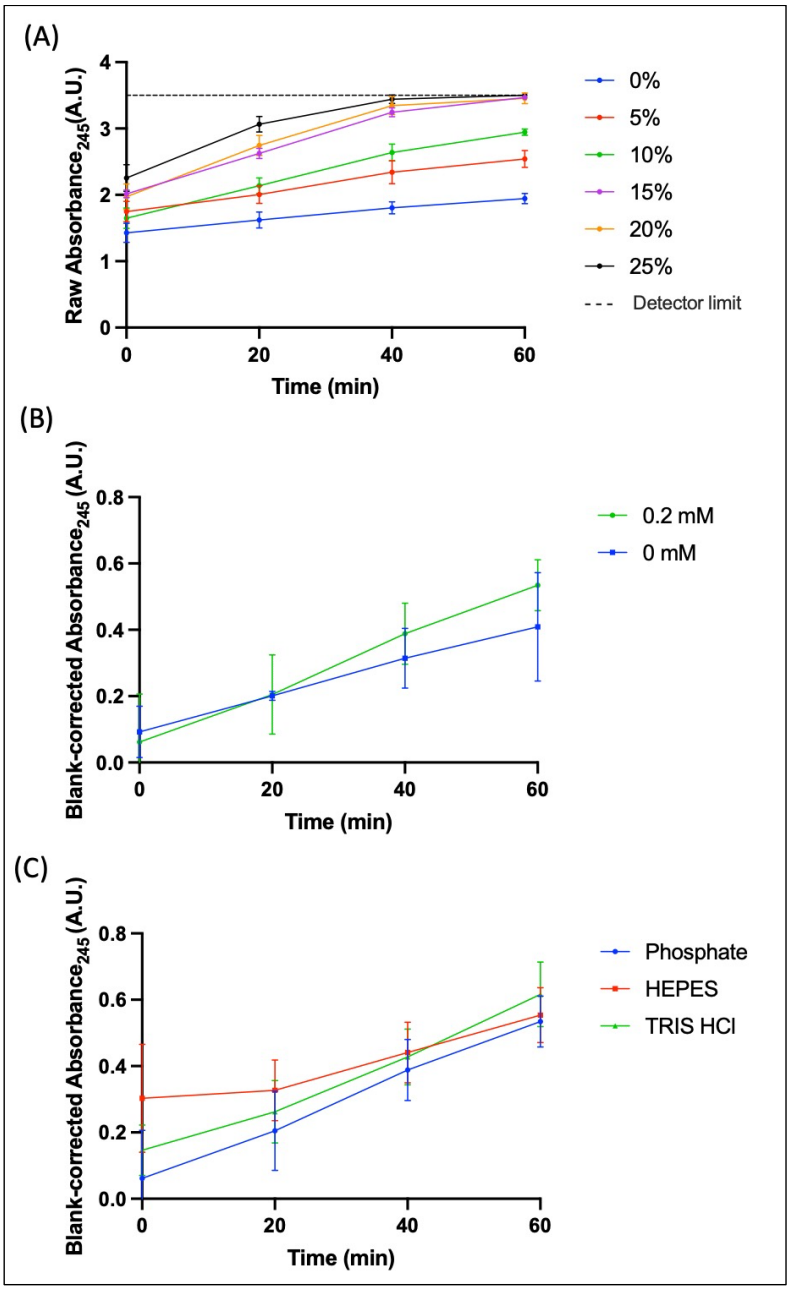


Figure 3.5. Results of AP assay scoping studies executed prior to the screening DoE.
(A) The effect of DMSO on the AP assay was tested by varying DMSO concentrations between 0-25% v/v to identify high and low DMSO factor settings. The figure plots raw data to highlight the increasing baseline absorbance from DMSO interference. (B) Effect of PLP supplementation on assay signal. The standard assay conditions were tested against reactions that contained an additional 0.2 mM PLP to identify if PLP should be included in the screening DoE. (C) Effect of buffer type on assay signal. Assays containing 100 mM of Phosphate, HEPES and TRIS HCl buffers at pH 7.4 were compared to select the best buffer type for future work. Assays were performed in triplicate as described in Section 2.3.2.1, and time-course absorbance data was generated every 2 minutes for 60 minutes. (A) and (C) show blank corrected data, where the blank reaction contained all reagents except for Sodium Pyruvate, the amine acceptor. (B) Error bars represent one standard deviation from the mean (n=3).

3.2.5 Screening DoE study

Following factor selection and initial OFAT experimentation (Section 3.2.4), the first step of DoE application involves the use of screening designs (Section 1.1.4). The objective at this stage is to screen a group of factors of interest to identify active factors. Automated DoE execution of the screening design consists of three steps as shown in Figure 3.6: (a) DoE design generation (Section 3.2.5.1); (b) manual design manipulation (optional) (Section 3.2.5.2); and (c) *in silico* simulation and physical execution through Synthace (Section 3.2.5.3). Each step is described in more detail in the following sections.

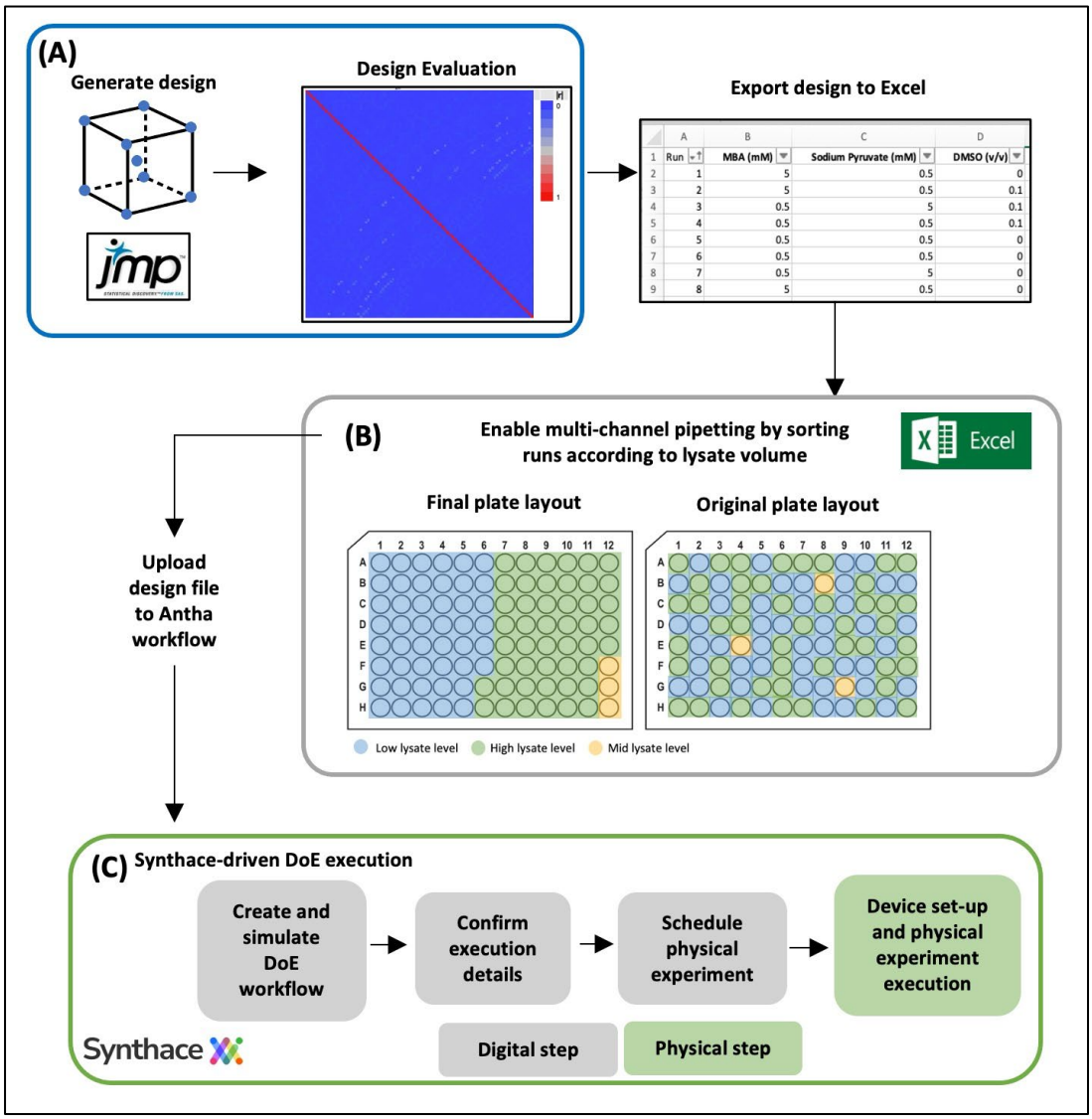


Figure 3.6. Schematic overview of the automated DoE experiment conducted using a PipetMax liquid handler.

(A) JMP is used to generate and evaluate the DoE design. (B) The design table was exported from JMP as an Excel file which was edited to ensure rapid lysate addition using multi-channel pipetting by grouping runs by column according to the lysate level. (C) The Synthace-driven DoE execution pipeline uses this modified design file as an input, which is parsed and used to parameterize the liquid-handling workflow. Upon successful *in silico* simulation, a physical experiment is scheduled where the instructions are sent directly to the selected device.

3.2.5.1 DoE design generation

A screening DoE study was performed to identify any key factors or factor interactions and ultimately identify fast and precise assay conditions. A two-level fractional factorial design with a D-optimality (determinant) criterion consisting of 186 run conditions (Appendix C and Appendix D) was chosen to test eight factors (Table 3.1). The D-optimal criteria is used to generate precise estimates of model effects, making it appropriate when identifying active factors, i.e., factors that significantly impact response values. The ranges for factors not tested in the scoping study (Section 3.2.4) were informed by literature and prior knowledge.

Selected factors and ranges (Table 3.1) were used to generate the DoE design using JMP, a third-party software (Figure 3.6 A). A minimum of 93 runs was required to test all main effects and two-factor interactions, however, as this experiment was executed using automation, the typical challenges faced with manual execution, such as time-consuming, error-prone, and unintuitive pipetting were removed (Tye, 2004). Therefore, the run number was doubled to 186 runs by 93 runs. As biological systems can be inherently noisy, generating more data could lead to potentially developing a more precise model.

Additionally, six replicates of a center point (also known as midpoint) run was also added to the design. A center point is a run where factor levels are fixed at the middle of all the low and high factor ranges. These runs are used to determine if curvature is present in the response and indicate if a relationship is linear or non-linear (Montgomery, 2013).

Evaluation of the 192-run design showed that it was more orthogonal than the 93-run design (Figure 3.7). This means that with the 192-run design, main effects and two-factor interactions can be estimated more independently compared to the 93-run design (Politis *et al.*, 2017b).

Table 3.1. Factors and levels chosen for the AP assay screening DoE study.

	Factors	Low level	High level	Reference
1	MBA (mM)	0.5	5	(Schätzle *et al.*, 2009a; Chen *et al.*, 2016; Voss *et al.*, 2020)
2	Sodium pyruvate (mM)	0.5	5	(Schätzle *et al.*, 2009a; Chen *et al.*, 2016; Voss *et al.*, 2020)
3	DMSO (% v/v)	0	10	Levels selected from scoping study (Section 3.2.4).
4	Salt (mM)	0	50	(Cerioli *et al.*, 2015)
5	Clarified cell lysate (% v/v)	10	25	(Schätzle *et al.*, 2009a)
6	Buffer (mM)	100	200	(Rios-Solis *et al.*, 2013)
7	Buffer pH	7	9	(Kaulmann *et al.*, 2007; Schätzle *et al.*, 2009a)
8	Temperature (°C)	25	37	(Kaulmann *et al.*, 2007; Schätzle *et al.*, 2009a; Du *et al.*, 2014)

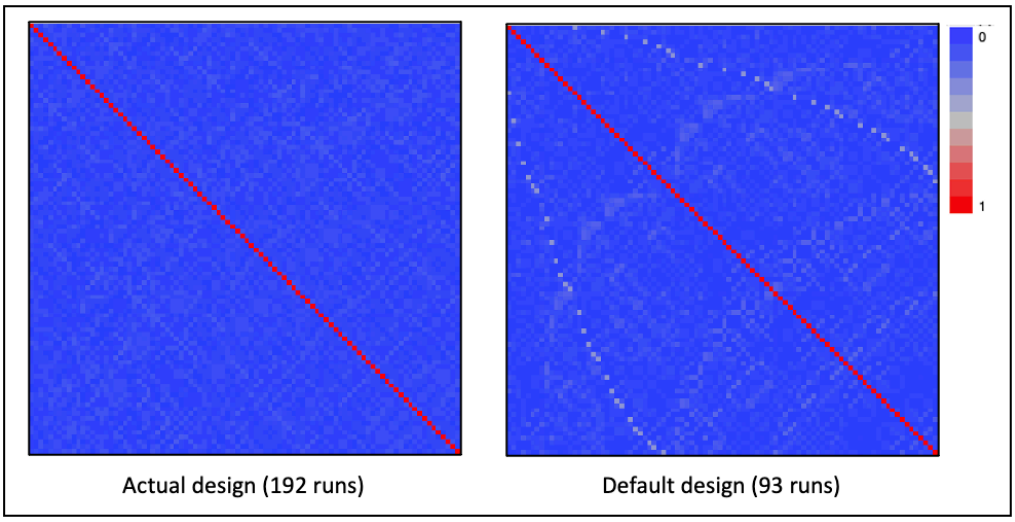


Figure 3.7. Design evaluation of two screening DoE design options.
Colour map on correlations plots of the actual design (192 runs) and the default design recommended by JMP (93 runs). Both X and Y axis contain variables (one pixel per variable) that are listed in the same order, they form a grid of correlations where each pixel represents the absolute correlation between any two variables which is shown using an intensity scale that ranges from 0-1 (blue to red). Values closer to 0 (blue) highlight the variables that are not highly correlated and can be estimated independently.

3.2.5.2 Manual design manipulation (optional step)

Once an appropriate DoE design is selected, it can be exported as an Excel worksheet and parsed directly into a Synthace workflow. However, in some cases, these files need to be manually modified to streamline downstream liquid handling before being parsed by a Synthace workflow (Figure 3.6 B). The PipetMax contains a fixed pipetting head that can aspirate and dispense one fixed volume using either a single channel pipette or a multi-channel (8 channel) pipette. This means that lysate addition, to initiate bioconversion reactions, is performed one well at a time as runs are listed in random order in the design file. Adding lysate in this way is time-consuming and may fail to capture meaningful measurement of the initial reaction rate in many cases since reactions may occur before the plates are passed to the spectrophotometer.

To speed up lysate addition and account for the operational characteristics of the PipetMax, the order of runs within the design file was adjusted to enable multi-channel pipetting. The runs for each lysate level were grouped together and by making this change, lysate is added in 14 steps, instead of 96 (Figure 3.6 B). These design files were then uploaded to the Synthace workflow to generate liquid-handling instructions.

This step is considered optional as other liquid handlers are capable of variable multichannel pipetting, i.e., where each channel can aspirate and dispense different volumes. In this case, manual manipulation of the design file is not considered necessary. Additionally, devices such as non-contact dispensers do not require this step as the same dispensing tool (e.g., a syringe) is used to transfer one reagent throughout the microtiter plate.

3.2.5.3 *In silico* simulation and physical execution with Synthace

The first step of Synthace-driven DoE execution is the preparation of a liquid-handling workflow. This is done in the Builder (Figure 3.8). Synthace elements are connected, and further parameterized by the scientist and the resulting workflow is compiled to generate device-specific liquid-handling instructions (Section 1.3).

To ensure that lysate was added to reaction mixtures last, the reagents required to prepare reaction mixtures were separated into two stages, first that contained all reagents except the lysate, and second with just the lysate (Figure 3.8 B). To generate stage-specific liquid-handling instructions, the workflow was simulated twice, once for each specified stage.

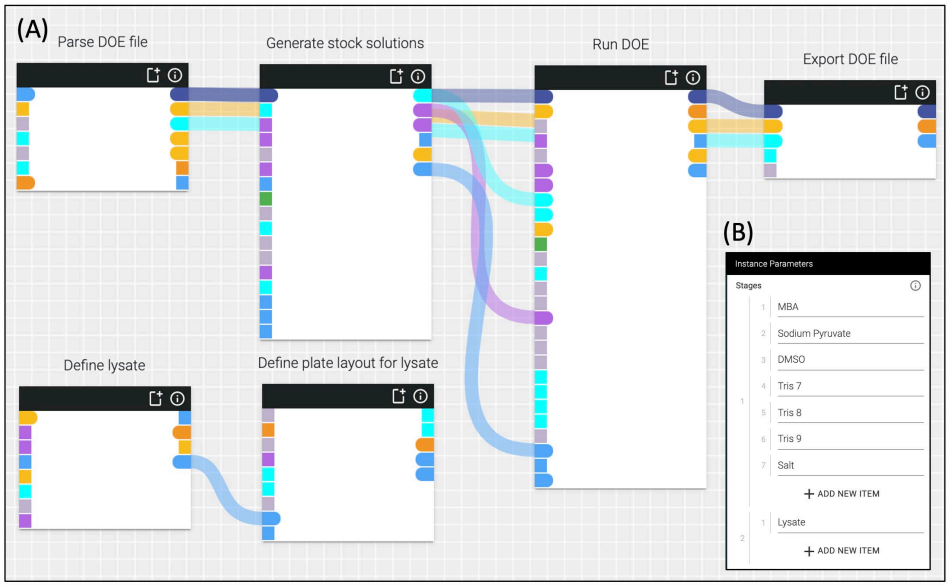


Figure 3.8. Synthace workflow for screening DoE Execution.
Elements (white boxes) are connected to facilitate information transfer; element titles indicate their functionality; see Section 2.7.1 for element descriptions (B) The Run DOE element parameters allow separation of reagent addition into different stages. Lysate is separated into a different stage. Specifying the 'StageToRun' parameter generates a simulation that contains liquid-handling instructions for the specified stage.

Upon successful workflow simulation, physical execution details can be accessed in three new pages; Overview, Setup, and Preview. The Overview page (Figure 3.9) summarizes all the files, labware, reagents, and devices associated with the workflow. The Setup page contains information on the contents of each plate needed or created by the protocol (Figure 3.10). This is primarily used to manually prepare the input plates when setting up the device.

The Preview (Figure 3.11, center) page is used to visualize each step of the experiment. The slider can be used to step through the experiment one pipetting step

at a time. Each well can be selected to gather isolated information at any stage in the experiment (Figure 3.11 right). Thus, the preview can be used as the 'debugging' tool to confirm that all liquid handling instructions are producing the desired outcome. Once the execution details are confirmed *in silico*, the device can be set up for physical execution by following the layout in the Preview tab, and the plate layouts in Setup tab.

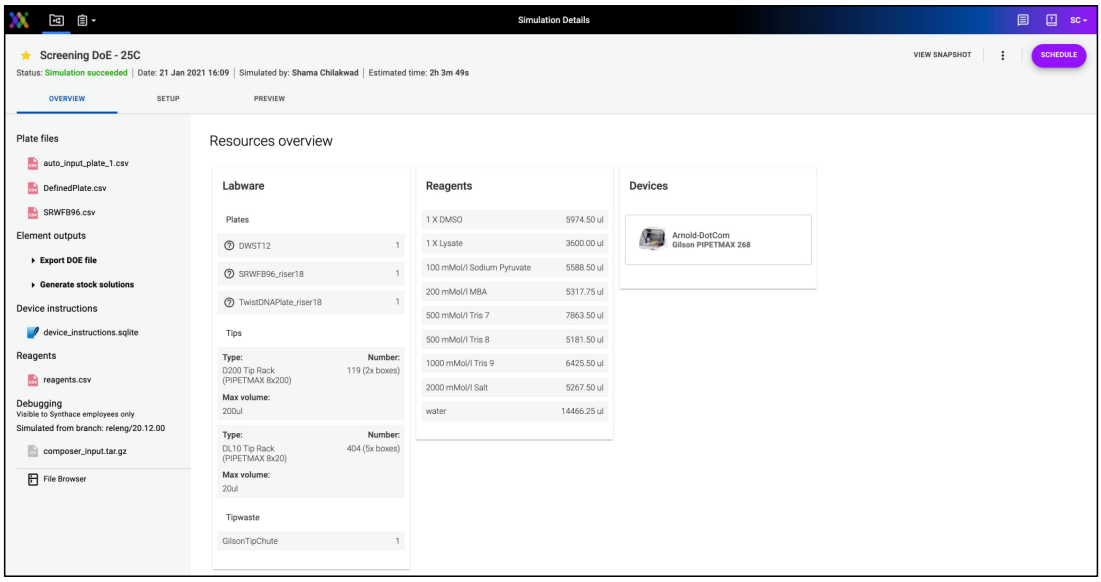


Figure 3.9. Overview of the execution details of a successfully simulated DoE workflow. Synthace software automatically collates and summarizes experimental details to facilitate setup. This includes liquid layout information (left) and hardware and consumables (right).

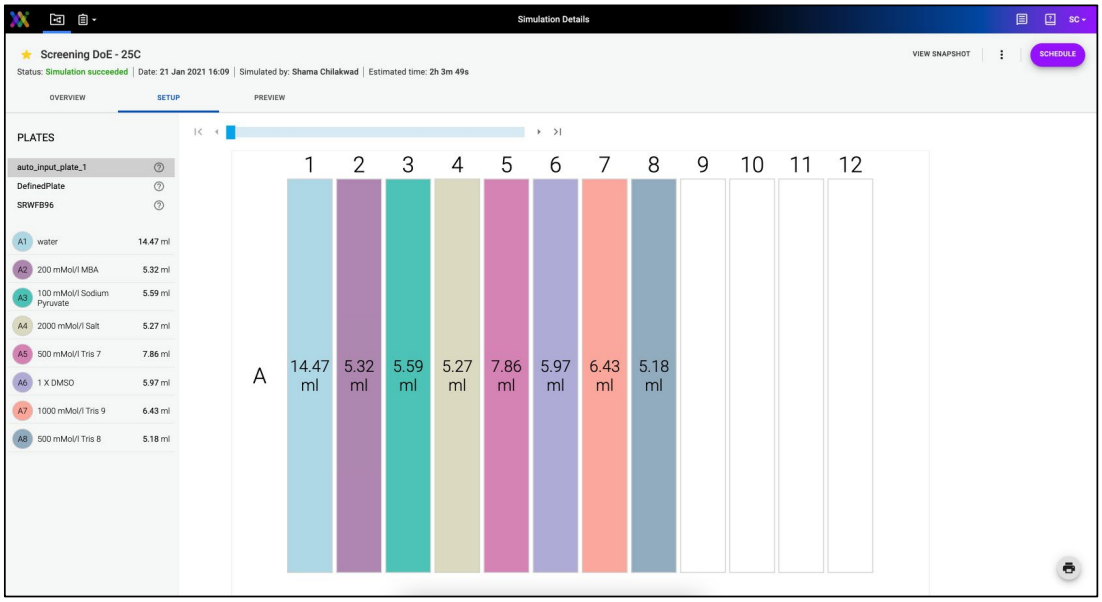


Figure 3.10. Setup page of a simulated DoE workflow for a PipetMax liquid handler. Automatically generated layouts and contents of all the microtiter plates required by a simulation are shown. This page is typically used to prepare input plates for physical execution.

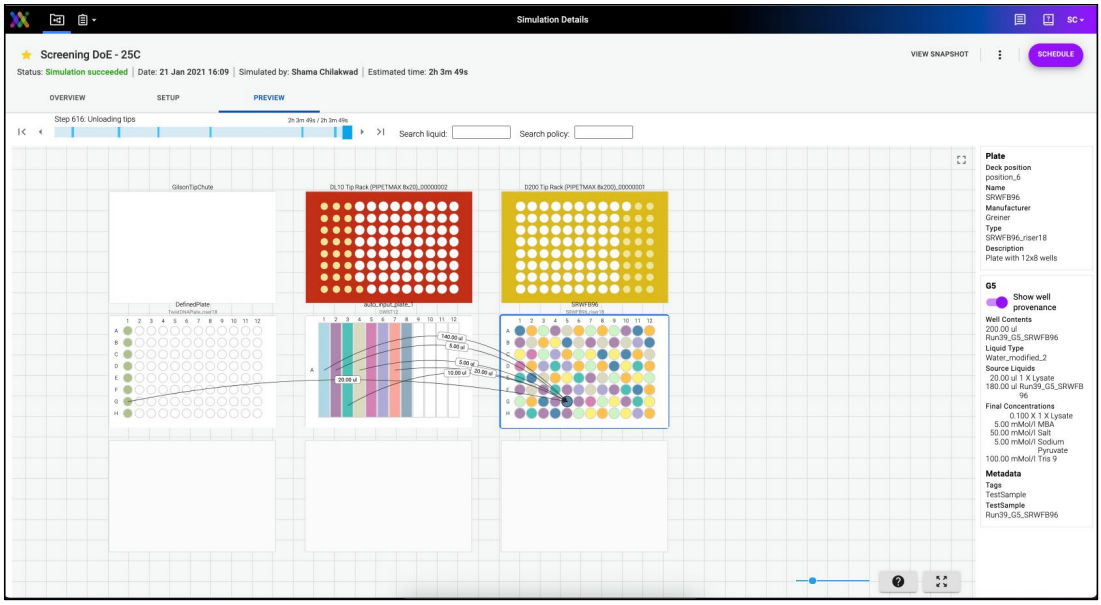


Figure 3.11. Preview page of a simulated DoE workflow for a PipetMax liquid handler. Experimental preview. Synthace's preview screen allows you to step through a simulated experiment one pipetting action at a time and allows inspection of volume, timing, and location information. Well G5 is selected to show the information that can be gathered from each well location (black lines with volumes and panel on the right).

Once the details of *in silico* simulation are confirmed, it is ready for physical execution. To initiate the process of physical execution, the simulation is scheduled, this action sends the liquid handling instructions to the SynthaceHub on a computer that controls the PipetMax liquid handler. The deck is prepared by placing input, and output microtiter plate, along with pipetting tips, as described within the Preview and Setup pages. Physical execution is started using SynthaceHub.

3.2.5.4 Analysis of the screening DoE study

When operated successfully, the screening DoE study could be executed in one day. Runs at 25 and 37 °C both took approximately 2 hours to prepare using the PipetMax liquid handler. This was followed by 2 hours of time-course spectrophotometric measurement. A wide range of progress curves were generated, a subset of which can be seen in Figure 3.12, highlighting the variety of information that can be generated from just one DoE screening experiment.

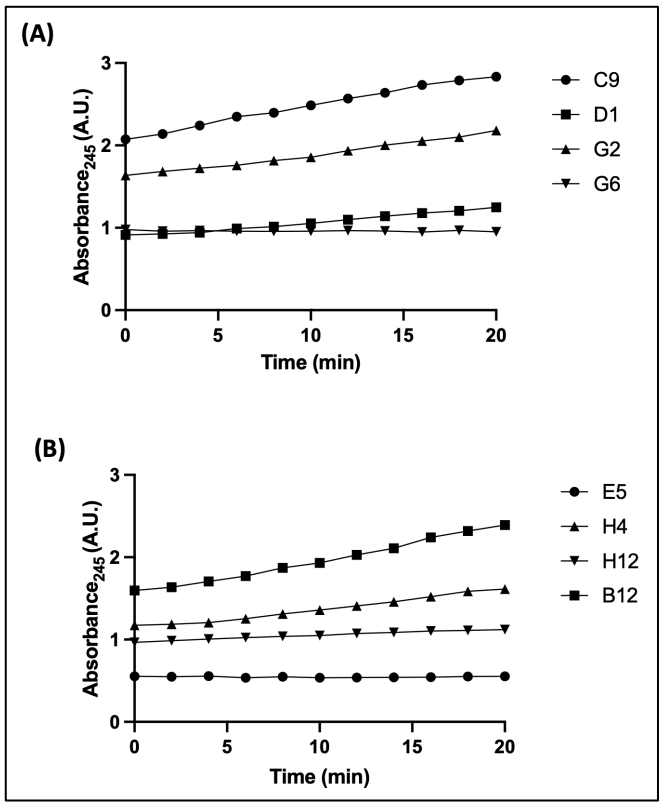


Figure 3.12. Example AP assay progress curves from the screening DoE.
The screening DoE consisted of 192 runs, with 96 runs at each of the two temperatures; (A) 25 °C and (B) 37 °C. A subset of the progress curves generated are shown to highlight the wide range of responses generated. The full data table with responses can be found in Appendix C and Appendix D.

ΔAbsorbance (Section 2.4.3) was the response used to analyse the screening DoE. To understand the relationship between the eight factors and the response, a statistical model is built. A stepwise regression model was selected as it condenses the set of potential explanatory variables down to a subset of the most significant variables before making a model, thus ensuring that a parsimonious model is created. Main effects, two-factor interaction effects, and polynomial terms were provided in the model (Section 2.6.3). Here, an effect refers to the impact of a factor or factor interaction on the response and is calculated by a statistical significance test e.g., T-test, F-test, Chi-square test etc.

Once the model is created, it was further refined in two ways. Firstly, any effects that had a p-value > 0.05 were removed to minimize the number of model terms. Secondly, outliers detected by the studentized residuals generated in the model report, i.e., the data points that fell outside the 95% individual *t* distribution limits (n

= 5, not shown) were excluded. A new model was generated using responses from 187 runs after accounting for these adjustments (Figure 3.13). The prediction formula for the model can be found in Appendix F.

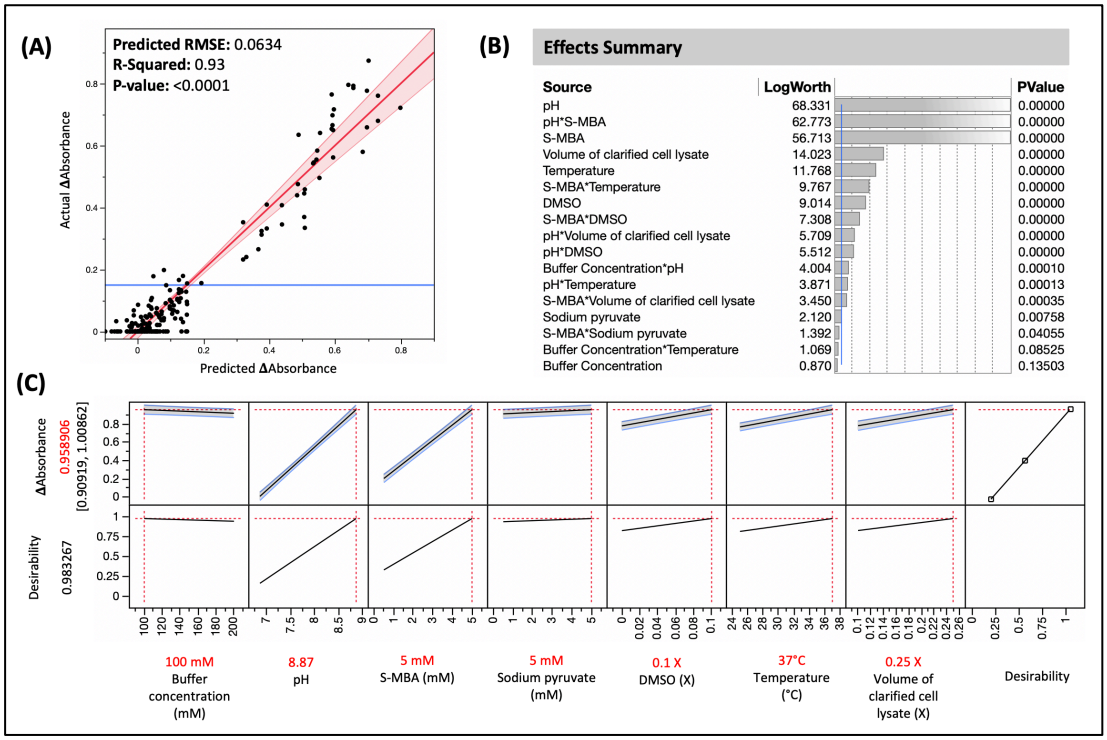


Figure 3.13. Stepwise regression modelling of the AP assay screening DoE response.
(A) Actual by Predicted plot. Correlation between the measured and predicted **Δ**Absorbance generated from a stepwise regression model indicates the goodness of the model fit. A linear fit (red) of data points (black, n=187, excluding 5 outliers) is shown with a 95% confidence interval (shaded red) along with the mean response value (blue). (B) Effects Summary Report. Significant effects ranked by p-value for effects with <0.01 (blue line). (C) Prediction profiler shows model predictions fixed at factor settings that predict maximum **Δ**Absorbance. Cross-sectional factor profiles (black) with 95% model confidence (shaded blue) and optimal factor conditions (red text and crosshairs) show maximum predicted **Δ**Absorbance with 95% confidence interval values (in brackets) when the desirability function (black, far right) is set to maximize **Δ**Absorbance. For DMSO and volume of clarified cell lysate, X represents the volume fraction (v/v).

The stepwise regression model was found to be statistically significant, meaning that the relationship between factors and the response was not due to chance. This can be seen in the actual by predicted plot (Figure 3.13 A) which has an R^2 of 0.93 showing that the predicted **Δ**Absorbance values correlated closely with the actual **Δ**Absorbance values. Furthermore, a p-value of <0.0001 was generated from the F-

test from the Analysis of Variance (ANOVA) report. Such a small p-value is evidence that at least one significant effect was detected.

The most significant effects i.e., the factors and factor interactions that were found to be significant were identified by conducting an F-test and are listed within the Effect Summary (Figure 3.13 B) in ascending order of p-value. Of the 186 runs, 70 runs executed at pH 7 had almost no impact on **Δ**Absorbance and led to the identification of an inactive region of the design space. While this may seem undesirable, knowing where there are regions of little or no enzyme activity is important in navigating towards desirable assay conditions. However, this is contradictory to literature data where the AP assay is routinely executed between pH 7-8 (Rios-Solis *et al.*, 2013; Gruber *et al.*, 2017).

Of the 8 factors tested, the effect of salt concentration had a p-value greater than 0.05 and was excluded from the model. All other factors have significant linear effects on **Δ**Absorbance, with pH being the most influential effect. This can be seen in the steep factor profiles, as increasing the pH to pH 9 steadily increases **Δ**Absorbance and agrees with what was found in the literature (Schätzle *et al.*, 2009a; Schell, Wohlgemuth and Ward, 2009). Higher levels of *S-MBA*, Sodium Pyruvate, the volume of clarified cell lysate, and temperature were also found to increase **Δ**Absorbance, however, this can be expected as increasing substrate and enzyme concentrations along with temperature typically increases enzyme activity (Schätzle *et al.*, 2009b; Stepankova *et al.*, 2013). Increased DMSO at 10% v/v was also found to increase **Δ**Absorbance. The CV2025 TAm enzyme's capacity to maintain activity at increasing DMSO concentrations has been previously reported (Chen et al., 2016; Leipold et al., 2019).

Furthermore, nine significant two-factor interactions with p-values under 0.05 were also observed. The top three are *S-MBA* and pH, *S-MBA* and Temperature, and *S-MBA* and DMSO (Figure 3.14). The interaction between *S-MBA* and pH and *S-MBA* and temperature can be explained using the reasoning that bioconversions at optimal

settings of pH and temperature accelerate the reaction rate, especially in the presence of excess *S-MBA* which strongly favours the forward reaction, resulting in maximising **Δ**Absorbance(Shin and Kim, 1998). Increasing **Δ**Absorbance due to increasing S-MBA and DMSO could be attributed to DMSO increasing S-MBA solubility, thus making it more available to the active site of the enzyme, in turn accelerating transamination (Meng *et al.*, 2020). Some of the two-factor interactions identified were speculated to be artifacts of the model as they were difficult to explain scientifically and may not be real effects, for example, the interaction between *S-MBA* and the volume of clarified cell lysate or *S-MBA* and Pyruvate.

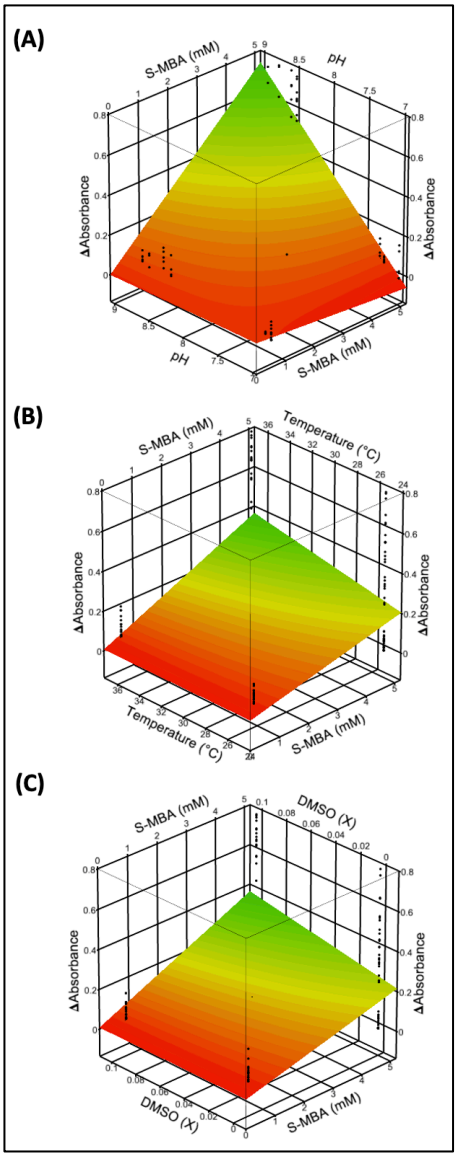


Figure 3.14. Predicted response surface plots for the AP assay showing significant two-factor interactions.
Significant two-factor interactions on the **Δ**Absorbance for: (A) pH and *S-MBA*, (B) *S-MBA* and Temperature, (C) *S-MBA* and DMSO. Black dots represent actual data points present above the response surface. Surfaces are calculated from the stepwise regression model of the screening DoE data (Section 2.6.3) and the prediction formula given in Appendix F. For DMSO, X represents the volume fraction (v/v). Factor settings and response data can be found in Appendix C and Appendix D.

Based on the DoE results and the prediction formula of the stepwise regression model presented in Appendix F, the following conditions were predicted as the optimal conditions to maximise **Δ**Absorbance: 100 mM TRIS Buffer at pH 8.87 with 5 mM of *S-MBA* and Pyruvate, 10 % v/v DMSO with 0.25X Lysate at 37 °C with a

predicted **Δ**Absorbance of 0.96 ± 0.05 units in 20 minutes. Some of the optima defined in these conditions e.g., for pH and temperature have also been previously reported (Shin and Kim, 2001; Schätzle *et al.*, 2009b; Schell, Wohlgemuth and Ward, 2009).

3.2.5.5 Implications of the screening DoE study

Improved reaction conditions that gave the greatest **Δ**Absorbance were predicted after executing and analysing the screening DoE study. In this section, some implications of these predictions on the next iteration of the DoE campaign are discussed.

The scoping study, executed prior to the screening DoE study, to identify suitable DMSO concentrations (Figure 3.5 A) showed that DMSO concentrations greater than 15 % v/v increase baseline absorbance and in turn saturated the plate reader detector. This result led to using 0 and 10% v/v DMSO as factor levels in the screening DoE design. However, analysis of the screening DoE study found the higher DMSO concentration of 10 % v/v gave rise to a greater **Δ**Absorbance indicating an improvement in TAm reaction rate (Figure 3.13). Furthermore, prior to the screening DoE, scoping studies were executed at pH 7.4, which was later identified as a non-optimal pH, this could potentially explain the need to use up to 50 µL of clarified cell lysate for bioconversions, which also contributed to increasing baseline absorbance (Appendix B). As improved reaction conditions have been identified, the quantity of clarified cell lysate required for bioconversions might be reduced, this, in turn, would reduce the lysate contribution to baseline absorbance and allow investigating the impact of even greater DMSO concentrations on CV2025 activity. These implications were considered when planning the next DoE iteration.

3.2.6 Optimisation DoE study

As the analysis of the screening DoE study identified improved reaction conditions for CV2025 bioconversions (Section 3.2.5.5), a second DoE iteration was conducted to map a narrower region of the design space. This iteration aimed to optimise

reaction conditions within the new design space and accurately map the experimental noise of that space. An overview of the optimisation DoE can be found in Figure 3.19. To better understand the noise of the design space, every run was executed in triplicate to allow calculations of metrics such as standard deviation. The first step of this process was to validate the reaction conditions identified by the screening DoE study (Section 3.2.5.4).

Initially, the optimisation DoE was to be executed using just the PipetMax liquid handler, however, another automated device, the Dragonfly dispenser became available at a later date. Therefore, the optimisation DoE was repeated using the Dragonfly dispenser. This was done to understand the effect of device type on the process of automated DoE execution and to test how reproducible results would be when the same experiment was executed using different devices three months apart. As the DoE design had not included device type in the original design, it was analysed separately to maintain the power of the design.

3.2.6.1 Validating the screening DoE model

Desirable reaction conditions identified from the screening DoE model (Section 3.2.5.4) were physically validated, this experiment was also used to identify an appropriate quantity of clarified cell lysate that can be used for the optimisation DoE study (Figure 3.15). The reaction conditions identified by the screening DoE led to a significant increase in **Δ**Absorbance while simultaneously reducing the quantity of clarified cell lysate was required by the AP assay. Therefore, the quantity of clarified cell lysate that resulted in a **Δ**Absorbance of 1 unit was selected for the optimisation DoE study as this signal would neither be high enough to saturate the plate reader detector nor be too low that no signal is measured. The precise volumes were calculated from the linear fit shown in Figure 3.15.

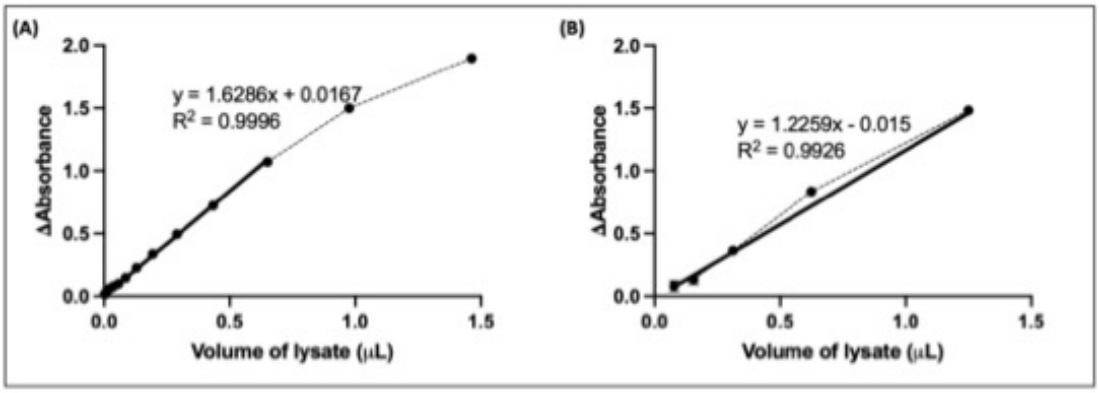


Figure 3.15. Experimental confirmation of the AP assay screening DoE model and identification of the appropriate quantity of clarified cell lysate for a subsequent optimisation DoE study.
Reaction conditions that predicted maximum **Δ**Absorbance from the screening DoE model (Section 2.3.2.2) were experimentally confirmed using both a PipetMax liquid handler (A) and the Dragonfly dispenser (B). **Δ**Absorbance after 20 minutes of measurement across a range of lysate volumes was plotted (dashed line) along with the line of best fit (solid line) that was used to determine appropriate clarified cell lysate volume. Error bars represent one standard deviation from the mean (n=3).

3.2.6.2 Scoping study for identifying suitable DMSO levels

Testing the improved reaction conditions generated from the screening DoE study identified that a lower quantity of clarified cell lysate (<1 µL compared to the 20 and 50 µL investigated in the screening DoE study) was required to generate a greater **Δ**Absorbance. Therefore, the contribution of clarified cell lysate on baseline absorbance was minimised allowing a further investigation of the impact of DMSO concentration on **Δ**Absorbance (previously explained in Section 3.2.5.5). Therefore, a scoping study was performed to identify the appropriate levels of testing DMSO in the optimisation DoE study.

Increasing DMSO concentrations reduced **Δ**Absorbance while increasing baseline absorbance, in turn decreasing the dynamic range of the assay (Figure 3.16). Low and high factor levels 0 and 30 % v/v DMSO were chosen for the optimisation DoE study as it allows for further investigation of the impact of DMSO on CV2025 bioconversions.

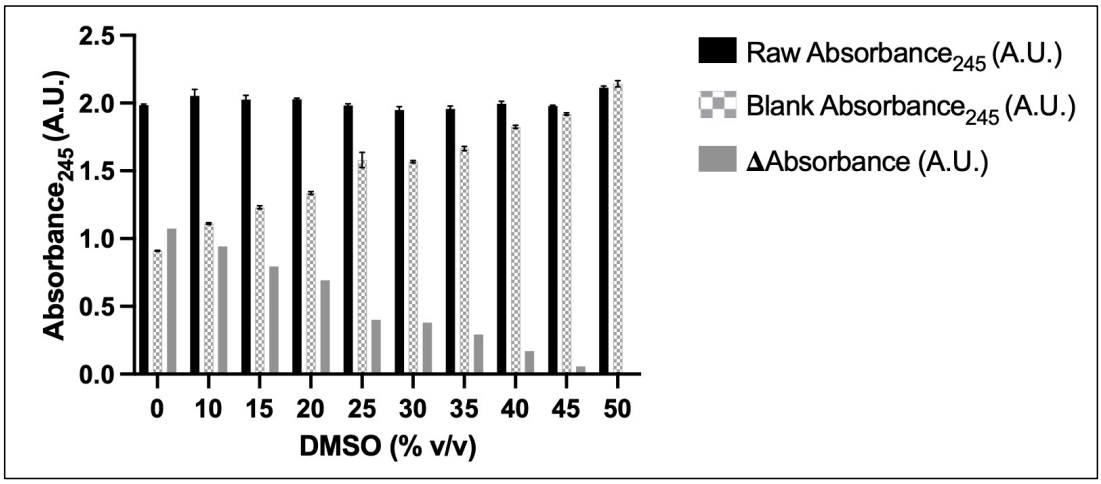


Figure 3.16. Scoping study to identify suitable DMSO concentration levels for the AP assay optimisation DoE.
The impact of DMSO concentration on CV2025 activity was evaluated using improved conditions identified from the screening DoE. Experiments were prepared, in triplicate, as described in Section 2.3.2.2 and executed using the PipetMax liquid handler. Error bars represent one standard deviation from the mean (n=3).

3.2.6.3 Optimisation DoE design

Five factors were taken forward for the optimisation DoE study (Table 3.2), the overlapping design space between the screening DoE study and the optimisation DoE study are shown in Figure 3.17. to help visualise how the design space was narrowed down in this iternation. The factor levels were adjusted to narrow in on the optimal region found in the screening DoE study.

Salt was removed as a factor of interest from this design, buffer concentration was also excluded and fixed at 100 mM as they both had a negligible impact on **Δ**Absorbance. The volume of clarified cell lysate was removed as a factor of interest, instead, the quantity of clarified cell lysate that resulted in a **Δ**Absorbance of 1 unit was used (Figure 3.16). The pH range was narrowed down to pH 8 and 9, pH 7 was excluded as it contributed significantly to inactive run conditions found in the Screening DoE. Temperature was tested at 18 and 25 °C to mimic a wide range of what could be considered room temperature. This was done despite the screening DoE analysis indicating that higher temperatures of 37 °C contributed to higher signal, as reducing bioconversion cost at scale up is often the industrial driver for

optimisation, further optimise conditions at and around room temperature was done to try and exceed the signal seen at higher temperatures in the previous DoE iteration. DMSO was tested at 0 and 30 % v/v (Figure 3.16) as solvent stability is a key feature of many industrially relevant bioconversions, as co-solvents are frequently used to aid the dissolution of various components such as substrates and products.

A design based on response surface methodology (RSM) with an I-optimality criterion was used to create a 48-run design using JMP (Appendix E). I-optimal criteria aims to reduce the average variance of prediction over the design space and places importance on the precise estimation of the response (Goos, Jones and Syafitri, 2000). This design could be executed in 21 runs; however, this number was increased to 48 to gather more data as the process of automated execution would be the same for either design. The colour map on correlations comparing both designs (Figure 3.18) shows that the 48-run design provides greater orthogonality for isolating main effects and two factor interactions. An overview of the optimisation DoE study can be found in Figure 3.19**.**

Table 3.2. Optimisation DoE study factors and levels.

	Factors	Low level	Mid level*	High level
1	MBA (mM)	4	5	6
2	Sodium Pyruvate (mM)	4	5	6
3	DMSO (% v/v)**	0	15	30
4	pH	8	8.5	9
5	Reaction temperature (°C)	18	21.5	25

* Automatically generated by JMP.

**DMSO is represented by the unit 'X' which represents the volume fraction (v/v) during Synthace workflow generation and JMP analysis.

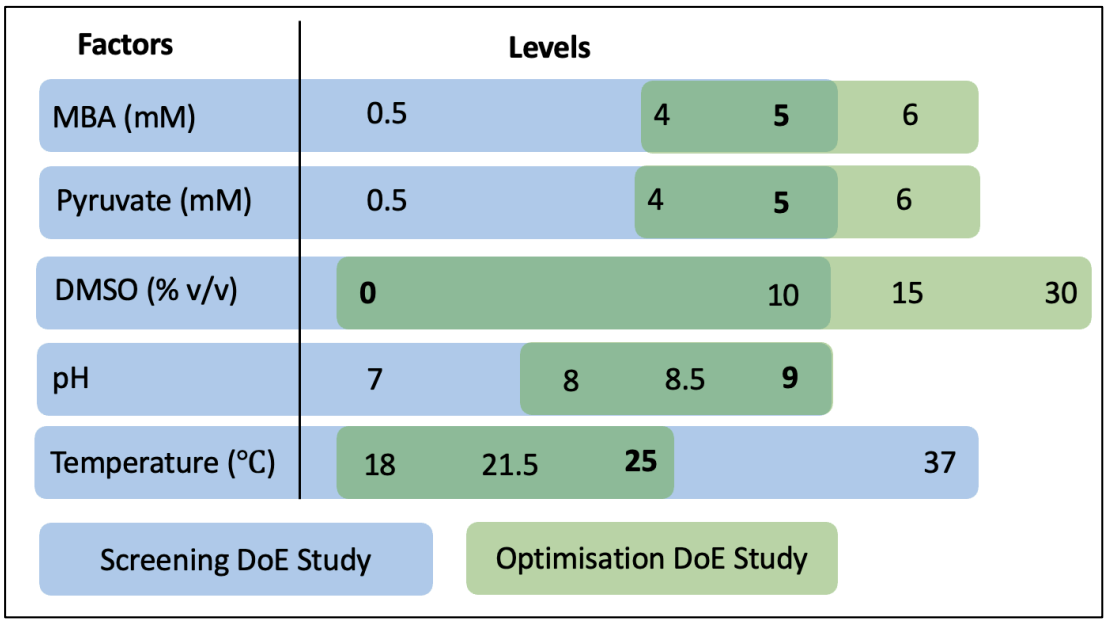


Figure 3.17: The overlap of design space covered between the screening and optimisation DoE studies for five factors.

Overlapping factor levels between both studies are indicated in bold.

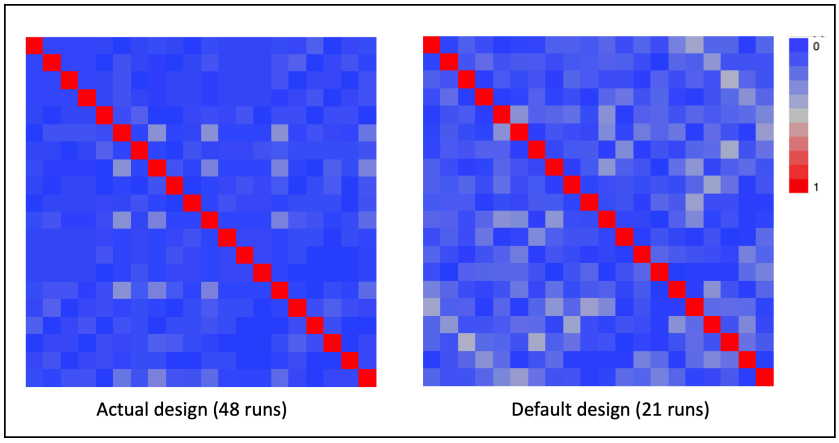


Figure 3.18. Evaluation of two Optimisation DoE design options.
Colour map on correlations plots of the actual design (192 runs) and the default design recommended by JMP (93 runs). Both X and Y axis contain variables (one pixel per variable) that are listed in the same order, they form a grid of correlations where each pixel represents the absolute correlation between any two variables which is shown using an intensity scale that ranges from 0-1 (blue to red). Values closer to 0 (blue) highlight the variables that are not highly correlated and can be estimated independently.

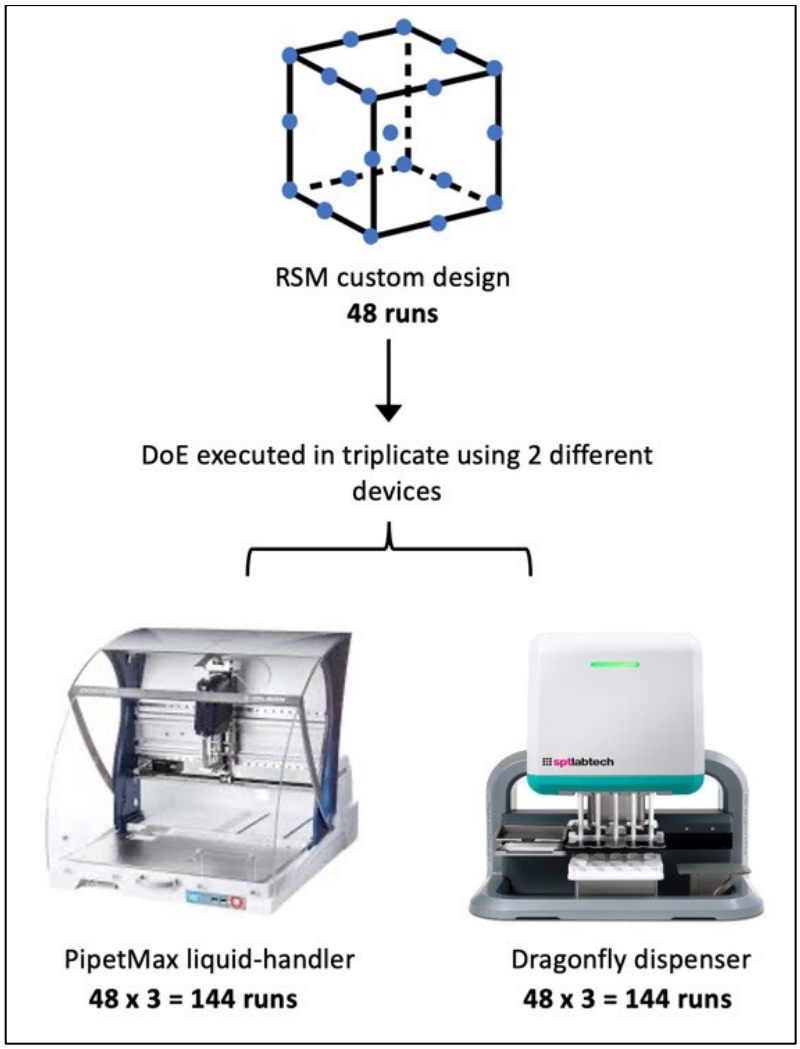


Figure 3.19. AP assay optimisation DoE study overview and implementation.
A 48-run RSM design (Section 2.6.1) was executed in triplicate. Reproducing the design provided data to assess the noise found within the design space while testing the design at two lysate volumes provide information on the dynamic range of the assay. This experiment was repeated on two devices: a PipetMax liquid handler and a Dragonfly dispenser to investigate how the device type impacts the process of automated DoE execution and analysis (Section 2.1.4).

3.2.6.4 Automated DoE execution using two devices

The process of generating liquid handling instructions for both devices is the same as previously outlined for the screening DoE study (Section 3.2.5.3). However, as the Dragonfly dispenser has 10 dispensing syringes and one output plate position the Synthace preview page changes to reflect this (Figure 3.20).

The summarised statistics of the time required to execute the designed experiment and the accompanying execution details for both devices are summarised in Table

3.3 and Table 3.4. The Dragonfly dispenser is a rapid, non-contact dispenser that uses 10 syringes to aspirate and dispense liquids using positive displacement. Using the PipetMax liquid handler was found to be ~37 times faster, additionally, to execute the design in triplicate once (144 runs) only 13 syringes and reservoirs, each, were needed to execute the whole design once compared to the 677 pipette tips (> 7 boxes) and six microtiter plates required by the PipetMax liquid handler.

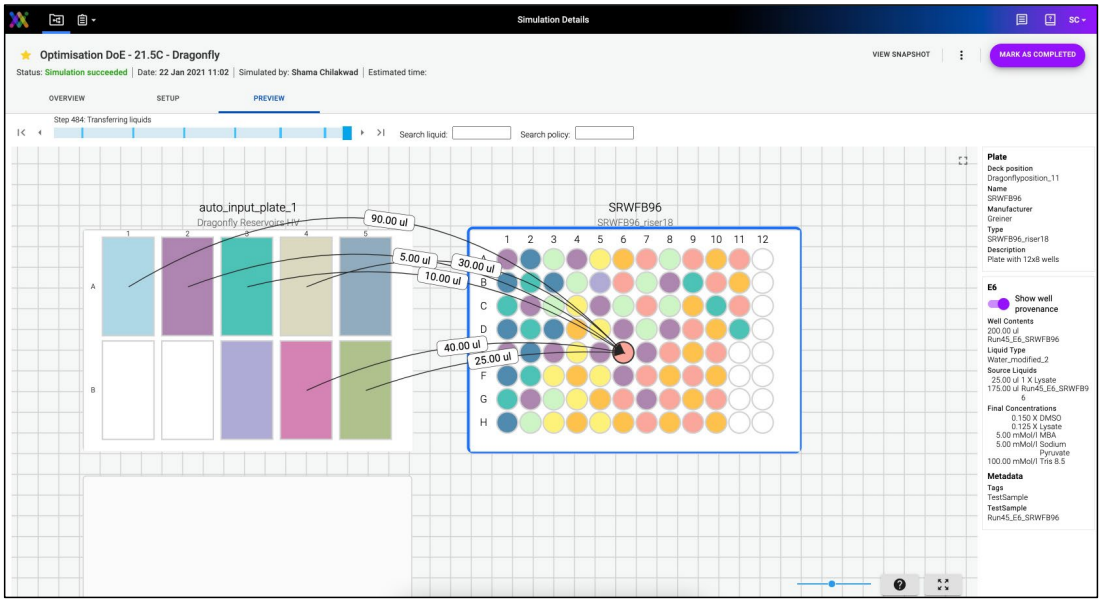


Figure 3.20. Preview page of a simulated DoE workflow for a Dragonfly dispenser. Reagents in the Dragonfly dispenser's reservoir tray (left) are transferred into the output microtiter plate (black lines with volumes). Well E6 is highlighted to show well contents on the right.

Table 3.3. Time required for automated execution of the AP assay optimization DoE.

		Gilson PipetMax		SPT Dragonfly	
	Workflow block	Plate preparation	Lysate addition	Plate preparation	Lysate addition
1	21.5 °C	2h 32m	6m 26s	3m 04s	0m 45s
2	18 °C	55m 38s	2m 28s	1m 34s	0m 16s
3	25 °C	57m 12s	2m 28s	1m 36s	0m 16s
	Total LH time*	**4h 25m**	**11m 22s**	**6m 14s**	**1m 17s**

*Accounts for the time required to run all three workflows once (144 runs).

Table 3.4. Summarized execution details for automated AP assay optimisation DoE execution.

	Gilson PipetMax			SPT Dragonfly		
Workflow	21.5 °C	18 °C	25 °C	21.5 °C	18 °C	25 °C
No. of liquid handling actions	612	224	217	485	170	139
No. of input plates	2	2	2	13 reservoirs*		
No. of tips	429	139	109	13		

* The Dragonfly dispenser's reagent tray holds 10 reservoirs where 10 standard reservoirs were used for plate preparation and 3 low-volume reservoirs were used for lysate addition to start the bioconversion.

3.2.6.5 Analysis of optimisation DoE study

The average **Δ**Absorbance calculated from three replicates (Appendix E) was used as the response to analyse the optimisation DoE study (Figure 3.21). The screening DoE study showed that the activity of the CV2025 enzyme could be measured within 20 minutes, therefore, end-point measurements after 20-minute incubations in benchtop thermomixers were measured to further simplified the execution.

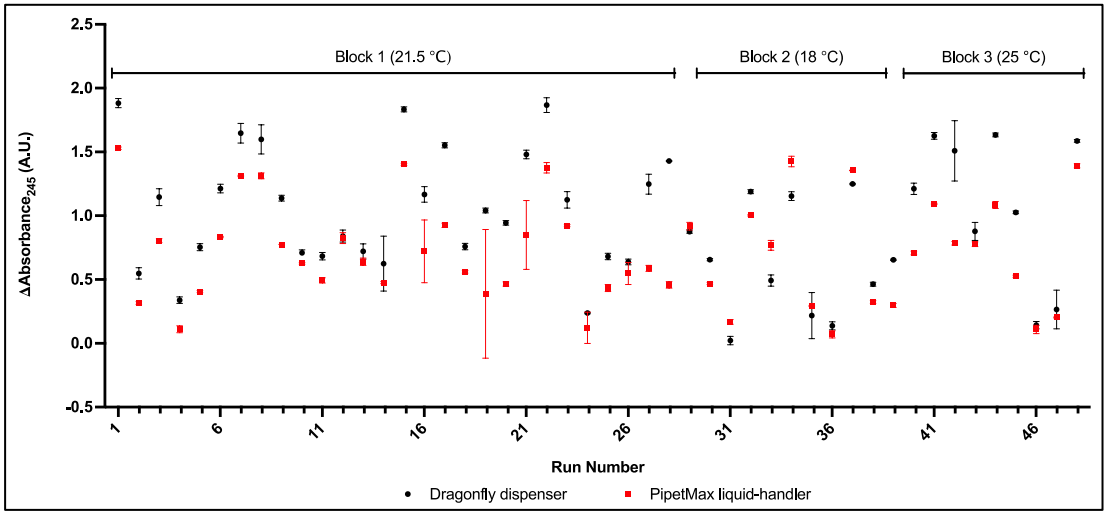


Figure 3.21. Comparison of ΔAbsorbance generated from executing the AP assay optimisation DoE study using two devices.
A 48-run design was executed in triplicate at two enzyme concentrations and on two devices, the PipetMax liquid handler and a Dragonfly dispenser (Appendix E). Runs were separated onto different plates using temperature as the blocking factor. **Δ**Absorbance was calculated as described in Section 2.4.3. Factor settings and response data can be found in Appendix E. Error bars represent one standard deviation from the mean (n=3).

For the optimisation studies executed on both devices, stepwise regression models were significant as the actual **Δ**Absorbance closely correlated with predicted **Δ**Absorbance values (Figure 3.22 A). Furthermore, a p-value of <0.0001 was generated from the F-test from the Analysis of Variance (ANOVA) report. High R^2 values and small p-value are evidence that at least one significant effect has been detected and that the response being analysed was not generated by chance. In general, R^2 values were marginally higher when the Dragonfly dispenser was used for execution. The maximum predicted **Δ**Absorbance was 2.11 ± 0.28 using the PipetMax liquid handler and 2.31 ± 0.18 units using the Dragonfly dispenser. Both prediction formulas can be found in Appendix G.

When the models were fixed to predict maximum **Δ**Absorbance, the underlying assay conditions were also similar between both models (Figure 3.22 B). Increasing pH and decreasing DMSO has the most significant influence on maximising the response; these were identified as the top two contributing factors in both models (Figure 3.23). Neither model found pyruvate concentration to have a significant impact and only the optimisation study, executed using the PipetMax liquid handler, found

temperature to also be insignificant. Significant quadratic effects were identified for DMSO, pH and temperature while no curvature was found with any of the factors.

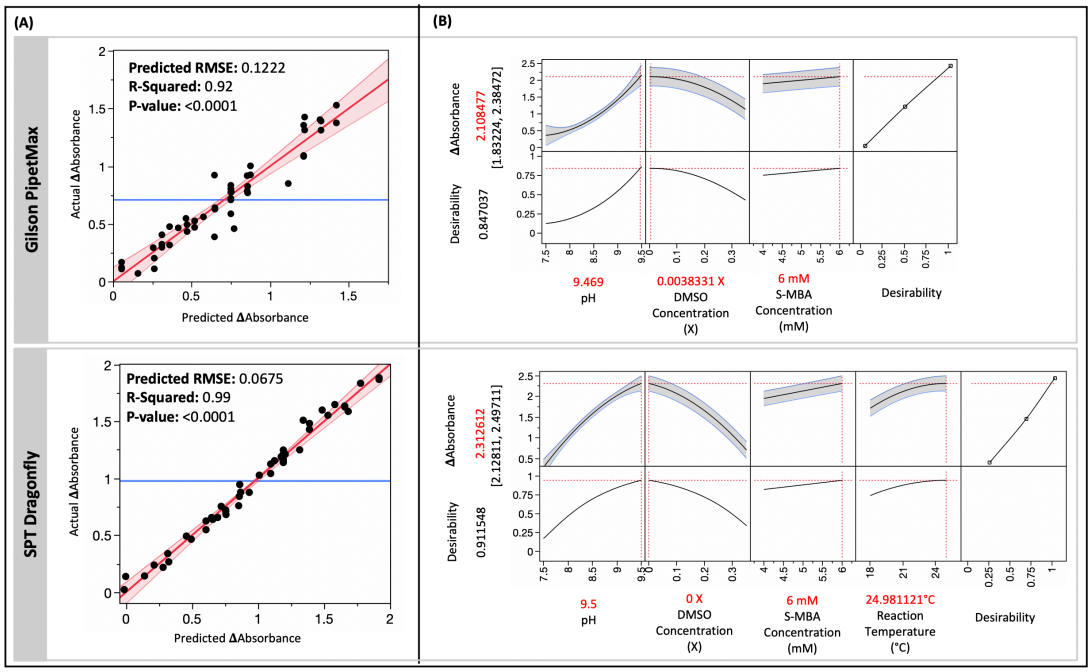


Figure 3.22. Stepwise regression modelling of the AP assay optimization DoE response generated using two devices at the high lysate level.
(A) Actual by Predicted plot. Correlation between the measured and predicted **Δ**Absorbance generated from a stepwise regression model indicate the goodness of the model fit. A linear fit (red) of data points (n = 48) is shown with 95% confidence (shaded red) along with the mean response value (blue). (B) Prediction profiler shows model predictions fixed at factor settings that predict maximum **Δ**Absorbance. Cross-sectional factor profiles (black) with 95% model confidence (shaded blue) and optimal factor conditions (red text and crosshairs) show maximum predicted **Δ**Absorbance with 95% confidence interval values (in brackets) when the desirability function (black, far right) is set to maximize **Δ**Absorbance. For DMSO, X represents the volume fraction (v/v).

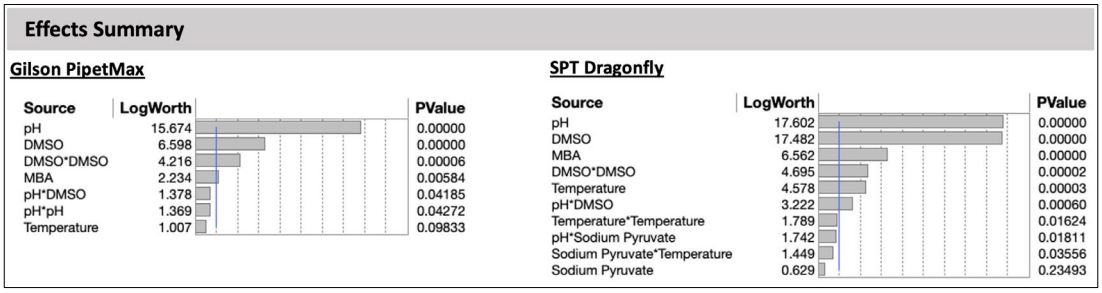


Figure 3.23. Comparison of significant model effects generated from modelling the AP assay optimisation DoE response.
Significant effects are grouped by device type and lysate level ranked by p-value for effects with p-values <0.01 (blue line).

The most significant two-factor interaction observed was between pH and DMSO, where increasing pH and decreasing DMSO concentrations maximised **Δ**Absorbance (Figure 3.24). Previously, the screening DoE study found that increasing DMSO concentration contributed to increasing **Δ**Absorbance required for bioconversion (Figure 3.13), however, increasing DMSO concentrations were found to decrease **Δ**Absorbance consistently across both models of the optimisation DoE study. This is likely to be because the screening DoE study investigated CV2025 Tam activity between a shorter range of 0-10 % v/v, where the upper level may have aided *S-MBA* solubility, in turn increasing **Δ**Absorbance. However, the optimisation DoE study investigated the impact of DMSO across a much wider factor range of 0-30 % v/v, the upper end of the range may have possibly led to denaturing the enzyme and in turn reducing **Δ**Absorbance when investigated beyond 10% v/v as seen in the response surface in Figure 3.24.

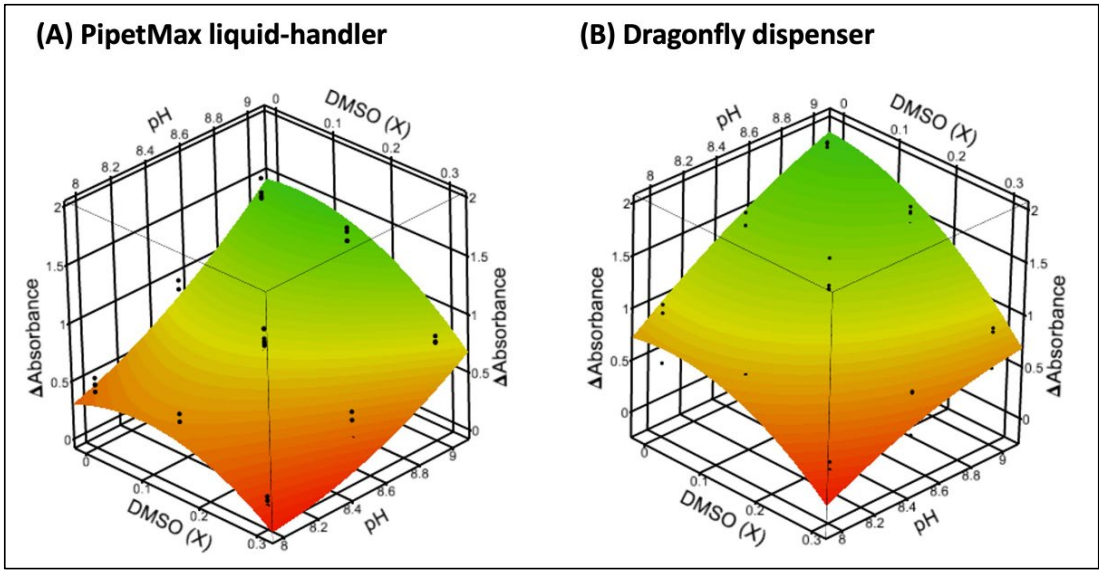


Figure 3.24. Response surface analysis of AP assay optimisation DoE results showing a significant two-factor interaction of pH with DMSO.
Dots represent actual data points present above the response surface. Surfaces are calculated from the stepwise regression models of the optimization DoE data (Section 2.6.3) and the prediction formula given in Appendix G. DMSO is analysed in units of relative volume fraction (X). Factor settings and response data can be found in Appendix E.

3.2.6.6 Analysis of noise generated from the optimisation DoE study

As one of the aims of the optimisation DoE was to characterise the noise of the AP assay each run was executed in triplicate. Standard deviation of **Δ**Absorbance was used as the response to map the design space using stepwise regression models (data not shown). However, the standard deviations calculated from all runs were negligible with the median standard deviation of **Δ**Absorbance of 0.0287 and 0.0288 for the PipetMax liquid handler and the Dragonfly dispenser datasets, respectively. Therefore, they were not modelled further.

However, one pattern contributing to increased standard deviations was observed in both datasets (Figure 3.25). For runs that needed 15 and 30 % v/v DMSO, a greater standard deviation was observed with the Dragonfly dispenser, with a mean standard deviation of 0.060, the same pattern was seen to a lesser extent using the PipetMax liquid handler, where the mean standard deviation was 0.053. One possible explanation for this might be that solvents are harder to handle due to low surface tension making aspiration and dispensing less precise than with aqueous reagents (Parshley, Bradshaw and Albert, 2014; Christensen *et al.*, 2021)

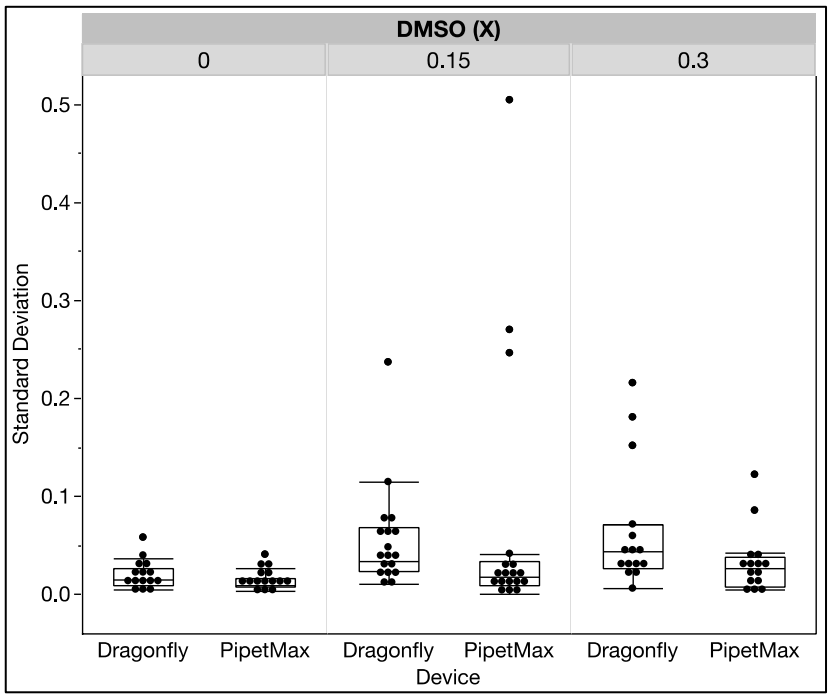


Figure 3.25. Effect of device type and DMSO concentration on standard deviation of ΔAbsorbance measurements in the AP assay.
Dots represent individual data points (n = 96, 48 per device). The median is indicated by the horizontal line within the box, whiskers indicate the lower and upper quartiles, any data points outside whiskers represent outliers. For DMSO, X represents the volume fraction (v/v).

3.3 Discussion

In this chapter, the AP assay (Figure 3.1) was used as a model to explain the method of DoE execution with Synthace. Automating manual assay methodology, scoping studies, screening, and optimisation DoE studies were all part of the DoE campaign. They were carried out with the goal of establishing rapid and robust reaction conditions capable of generating a precise response. All automation was executed using a PipetMax liquid handler, while the optimisation DoE study was also repeated using the Dragonfly dispenser.

The goal of identifying fast and robust reaction conditions were met as the analysis of the eight factors (Table 3.1) investigated in the screening DoE study identified reaction conditions that increased **Δ**Absorbance from ~0.1 units to ~ 0.9 units in 20 minutes. This was further improved by taking five factors forward (Table 3.2) to the optimisation DoE study which doubled the **Δ**Absorbance to > 2 units within 20 minutes of measurement.

The screening DoE study provided the most insight into active factors (Figure 3.13) and two factor interactions (Figure 3.14). Optimal values for pH and temperature, along with the impact of DMSO on **Δ**Absorbance were identified. Significant two factor interactions involving temperature, substrate concentrations, DMSO were also identified. Many of the findings were in agreement with data found in the literature, which in turn increased the reliability of the statistical model (Figure 3.13 Appendix F) (Schätzle *et al.*, 2009; Schell, Wohlgemuth and Ward, 2009; Rios-Solis *et al.*, 2013; Gruber *et al.*, 2017).

Validating the improved run conditions predicted from the screening DoE model (Section 3.2.5.5) resulted in identifying that only <1 µL of clarified cell lysate was required to generate the expected **Δ**Absorbance (Figure 3.15), which is almost 50 fold lower than the 25 – 50 µL volumes being used previously. As pH was identified as the factor with the most significant effect, increasing the pH from 7.4 to 9, as identified

by the screening DoE might explain why a far lower quantity of clarified cell lysate. A known drawback of the AP assay is the reduced dynamic range due to the contribution of protein to initial absorbance measurements (Schätzle *et al.*, 2009). Therefore, this finding consequently increased the dynamic range of the assay as it reduced the absorbance contribution of the lysate. This in turn enabled the investigation of increasing DMSO concentration up to 30 % v/v in the optimisation DoE study which would not have been possible previously as DMSO also increases the baseline absorbance (Section 3.2.6.2).

Five factors were then taken forward in the optimisation DoE study (Section 3.2.6.3) to further characterise the improved design space identified by the screening DoE study. The optimisation DoE study was executed on two devices, the PipetMax liquid handler, and the Dragonfly dispenser. Statistical models for both devices generated collected were practically identical and predicted similar optimal reaction conditions (Figure 3.22). A pH optimum of ~9 continued to generate maximum **Δ**Absorbance, however, retrospectively, a wider pH range should have been investigated to confirm if the peak really was at pH 9. DMSO concentrations up to 30% v/v were investigated and were found to reduce **Δ**Absorbance. This is contradictory to what was observed in the screening DoE study where a greater **Δ**Absorbance was attributed to reactions with 10% DMSO instead of ones that contained no DMSO, with the reasoning that DMSO may have led to increased *S-MBA* solubility and consequently it's availability (Figure 3.13). One possible reason for this effect could be that the optimisation DoE study was sampling high substrate concentrations while the screening DoE study sampled at 0.5 and 5 mM, making the effect of the lower level of *S-MBA* solubility by DMSO more apparent. Higher levels of both substrates led to increased **Δ**Absorbance which is typical of enzyme-catalysed reactions. These patterns remained consistent with those observed in the screening DoE study (Figure 3.13).

The optimisation DoE study was executed in triplicate to map the noise of the design space and identify reaction conditions that maximise **Δ**Absorbance while minimising noise (Figure 3.19). However, very low standard deviations were calculated across all

runs, indicating that none of the five factors were significantly influencing standard deviation. However, reactions that contained 15 and 30 % v/v DMSO, resulted in a greater mean standard deviation (Figure 3.25). This was observed on both devices and could be attributed to the challenges associated with aspirating and dispensing DMSO due to its low surface tension. In the future, a liquid-handling precision test could be conducted to investigate device precision prior to executing automated DoE studies as it may result in falsely making conclusions about certain factors.

Overall, the DoE campaign was successful as it identified fast and robust assay conditions. These settings are ideal for high-throughput screening as they can be executed at room temperature, requiring no incubator thus reducing operating costs, while generating end-point measurements after only 20 minutes. Furthermore, both statistical models could be used to predict reaction conditions for specific scenarios e.g., conditions that maximise **Δ**Absorbance at 37 °C, or at 30 % v/v DMSO. One potential use case for these conditions is the high-throughput screening of engineered libraries of CV2025 mutants.

Automating the manual assay protocol into an automated workflow using Synthace was straightforward with automated bioconversions closely matching manual experiments (Figure 3.4), this provided confidence in the liquid-handling instructions that were generated from Synthace and led to the execution of scoping studies using automated workflows.

DoE execution benefitted greatly from automation with Synthace. A total of 480 runs (192 from screening and 288 from optimisation DoE studies) were executed, each DoE iteration was executed in just one day. The same underlying Synthace DoE workflow was used throughout, despite device type, highlighting the device-agnostic nature of Synthace (Figure 3.8). Minimal upfront planning was required as input plate layouts and device deck layouts are determined by Synthace, in turn, the interaction with the device control software was limited to initiating liquid handling, this removed a major barrier of entry to automated DoE execution (Section 1.2.4). The

optimisation DoE study was executed on the PipetMax liquid handler and the Dragonfly dispenser three months apart using the same Synthace protocol and comparable models were generated from both datasets highlighting reproducibility from automating complex protocols. The impact of device type on the process of DoE execution was heightened as the Dragonfly dispenser required just 8.5 minutes to prepare microplates needed for the optimisation DoE study compared to the PipetMax liquid handler which required over 4.5 hours, which is 32 times greater (Table 3.3). Furthermore, less labware e.g., dispensing syringes and reservoirs were needed by the Dragonfly dispenser which reduces the quantity of plastic used, making certain device types a more environmentally friendly option.

In Chapter 3, the combination of automation and Synthace for the execution of larger DoE designs was demonstrated. However, now that larger automated DoE studies can be executed, more sophisticated designs that were previously completely out of reach should be explored, instead of simply expanding traditional DoE designs, one example of such a design, called a space-filling design is explored in Chapter 5 for characterising a 12-dimensional design space of the AP assay. However, to execute more challenging design types, numerous important characteristics that impact high-throughput spectrophotometric assay performance must be investigated and confirmed, this is further investigated in Chapter 4.

4 Refining assay logistics

4.1 Introduction

The conclusions from Chapter 3 showed that combining rapid automated liquid handling devices with Synthace-driven DoE implementation can push the boundaries of DoE execution away from using traditional DoE designs towards attempting the execution of large, high-dimensionality DoE designs that would have been previously out of reach. However, before executing these designs, many non-biological considerations that impact the performance, precision, execution, and cost of assay implementation must be investigated and confirmed. These considerations are termed 'assay logistics' in this thesis. These are important characteristics that are rarely emphasised in the academic literature but are critical to the successful implementation of high-throughput spectrometric assays.

Spectrophotometric measurements form the basis of numerous enzymatic and drug screening assays (Acker and Auld, 2014) including the AP assay which is used as the model system in this work (Figure 3.1). A plate reader measures the quantity of ultraviolet (UV) or visible radiation absorbed by a substance of interest in solution at a specified wavelength (nm). Measurements are generated by the fundamental law of spectrophotometry, the Beer – Lambert's Law (Bouguer, 1729; Lambert, 1760; Beer, 1852):

$$A = \varepsilon.c.l \qquad 4.1$$

which states that absorbance (A) is equal to the concentration (c) of the substance, the path length (l) travelled by the light, and the extinction coefficient (ε).

Some of the logistics that impact spectrophotometric assays include determining the linear range of the plate reader, selecting suitable microplate types, reaction

volumes, optimal measurement wavelengths, measuring the precision of liquid handling devices, managing evaporation, and so on (Acker and Auld, 2014; Hentz and Knaide, 2014; Sancenon *et al.*, 2015; Siguemoto and Gut, 2017).

4.1.1 Aim and objectives

The aim of this chapter was to systematically investigate a subset of assay logistics and decide on the most appropriate conditions that aid assay performance. The specific objectives of this chapter are as follows:

1. Develop a flow chart of the assay logistics which outlines the key decisions that need to be taken
2. Ensure that the flowchart is generalisable and can be used to investigate the impact of both 96- and 384-well microplates and, numerous automated devices to cover any future use cases.

4.2 Results

4.2.1 Assay logistics flow chart and implementation considerations

Figure 4.1 presents a flow chart describing the sequence of key decisions that need to be taken and the factors that need to be considered when executing high-throughput spectrophotometric assays (Roskoski, 2014). This was done to ensure that a solid foundation is present for the automated execution of a 3,456 run, a 12-dimensional space-filling-DoE study executed in Chapter 5, however, it can be used generically to define logistics that commonly need to be selected for the execution of high-throughput assays.

The first decision was to confirm the linear range of the spectrophotometer as the AP assay is an absorption-based assay, i.e., the product, acetophenone, is directly measured based on its absorbance properties (Schätzle *et al.*, 2009). Absorbance is directly proportional to pathlength (and therefore volume), and to concentration (Equation 4.1), therefore, a decision was made to determine the linear range of the spectrophotometer's detector for a range of acetophenone concentration curves prepared at varying volumes in both 96- and 384-well plates first. Doing so would identify the linear range of the spectrophotometer for varying reaction volumes and allow comparison between plate types. Additionally, a wide range of spectrophotometers are available, all with varying dynamic ranges and methods to calculate absorbance. However, a pathlength of 1 cm is often used to standardise absorbance measurements. Therefore, confirming the precision of the spectrophotometer's pathlength correction calculation and, if needed, developing methods for path length correction might be useful.

Once the linear regime of the spectrophotometer is determined, reaction volume and plate type can be selected. Typically, absorption-based assays are executed by preparing a reaction vessel e.g., a cuvette, microplate, etc. that contains all reagents except one. The missing component is normally a small volume of the enzyme or a

substrate that is used to initiate the reaction. It is critical to ensure that all the reagents can be accommodated into the reaction volume, this decision is especially important as a wide range of reaction mixtures will need to be prepared for DoE studies as each run condition is composed of different factor levels.

Following this, it is important to pre-determine an appropriate measurement format for the DoE study as this will directly impact the type of responses that can be calculated. Pre-determining responses are critical to collecting data in a format that is fit for purpose. Absorption-based assays can be measured continuously or after a specified incubation time, i.e., end-point assays. The former can be used to generate large quantities of data and capture responses to determine enzyme kinetic responses e.g., initial rate. End-point assays can be advantageous in terms of throughput; however, care should be taken to ensure that the time-point selected is within the window where the product formation is linear.

As the foundation of absorption-based assays is to measure the product based on its absorbance properties, it is important to ensure that the wavelength used to monitor the reaction is one with the biggest difference between the product and the rest of the reagents. This ensures that the clearest signal is recorded.

More generically, it is important to understand the advantages and the limitations of the automated devices being considered for DoE study execution with respect to precision, speed, and the range of dispensing volumes. Lastly, if reaction mixtures require incubation, it is important to seal the microtiter plates using plate films that minimise evaporation over the desired incubation period. Furthermore, for time-course measurements, microtiter plates may be sealed during measurement to manage evaporation over time. Therefore, finding a plate film that will manage evaporation while minimising interference over the measurement wavelengths is required.

The following sections provide further details on the decisions to be made and provide experimental data obtained in order to make the most appropriate decisions for the AP assay used here

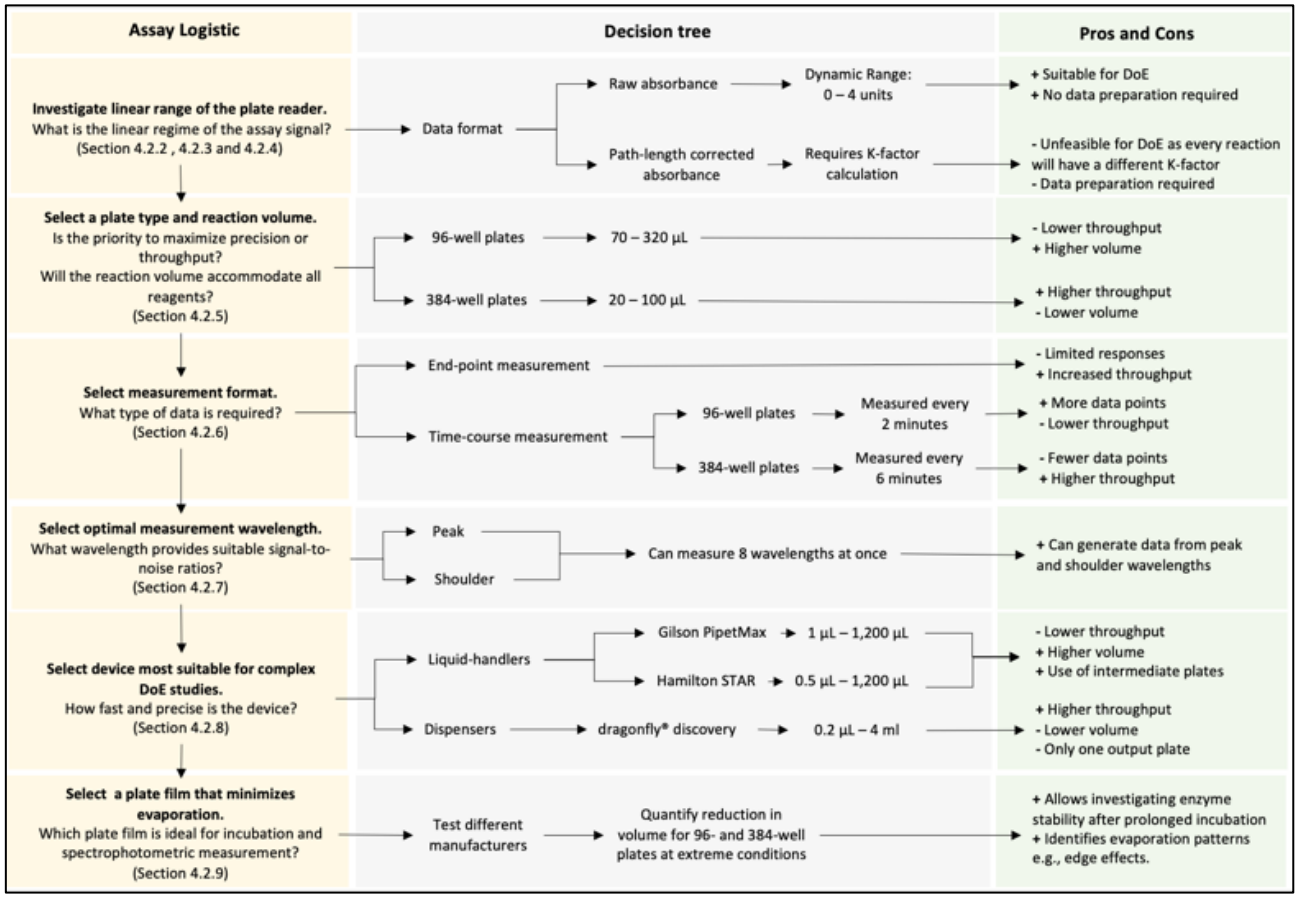


Figure 4.1 Flowchart generated to aid in making decisions on key assay logistics.
Individual logistic details and corresponding results are shown in yellow boxes, the subsequent decisions that need to be made are shown in grey boxes and the pros and cons of each option are shown in green boxes using “+” and”-” symbols, respectively.

4.2.2 Assay miniaturisation

Miniaturisation is a key aspect of assay development, with high-throughput screening groups (HTS) leading the way on automated, miniaturised screening reactions to maximise hit selection while simultaneously minimising reagent cost and increasing throughput (Kricka, 1998). Reducing reaction volume is the first step in miniaturisation. Various reaction volumes are tested, and one volume is selected, however, the impact of the minimised volume on assay signal and performance may or may not be tested if that volume is already fit-for-purpose (Sancenon *et al.*, 2015).

All AP assays in Chapter 3 were executed at a reaction volume of 200 µL, as described in the original method (Schätzle *et al.*, 2009a). This volume has been used as default in literature since it was first developed (Meng *et al.*, 2020; Heckmann, Dominguez and Paradisi, 2021). To investigate whether miniaturisation and increased throughput can be achieved simultaneously, assay performance was investigated by reducing the reaction volume to 100 µL in both 96- and 384-well plates (Figure 4.2).

In general, both progress (Figure 4.2 A and B) and standard curves (Figure 4.2 C) prepared at a 100 µL reaction volume in 384-well plates saturated the plate reader detector as the absorbance contribution of each reagent (Table 4.1) is greater in 384-well plates than 96-well plates for the same reaction. In contrast, measurements recorded from 96-well plates stayed within the plate reader's detection limit. Due to the well geometry of the narrow 384-well plates, the light beam from the plate reader must travel a greater vertical distance (path length) through a 100 µL solution in a 384-well plate compared to a 96-well plate. As path length is directly proportional to absorbance, the higher path length of 384-well plate generates higher absorbances thus saturating the signal.

This experiment showed that a systematic process is required when selecting a reaction volume. This process will be the topic of the next section (Section 4.2.3.)

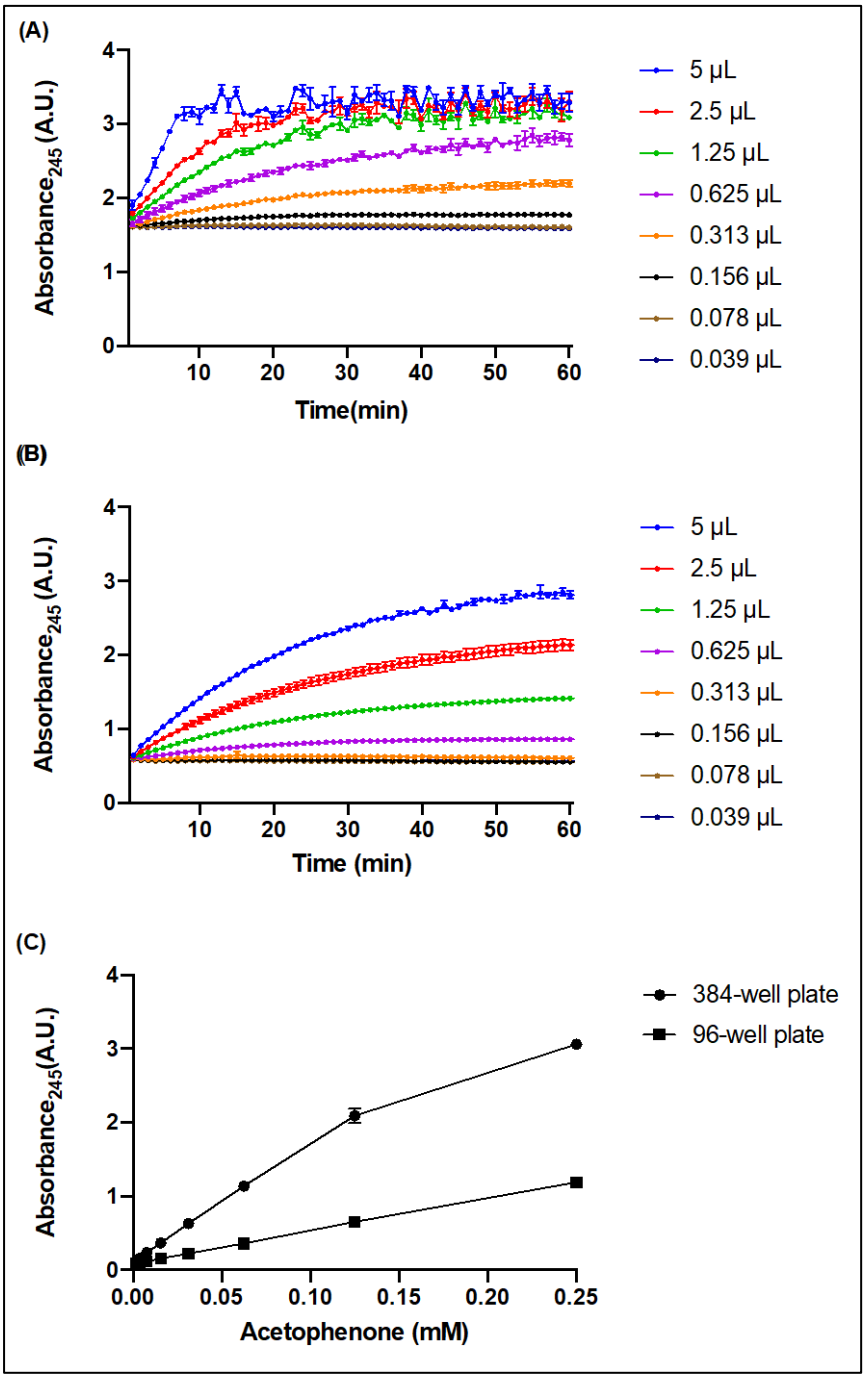


Figure 4.2. Comparison of AP assay progress and calibration curves prepared in 96- and 384-well plates.

(A), (B) The effect of varying lysate volume on the bioconversion of pyruvate to acetophenone in 100 µL total volume reactions (Section 2.3.2.2) was monitored by measuring acetophenone absorbance at 245 nm for 60 minutes in both 96- (A) and 384-well plates (B). (C) Acetophenone calibration curves. Identical serial dilutions (Section 2.4.2.1) were prepared at a final volume of 100 µL in 96- and 384-well plates. Error bars represent one standard deviation from the mean (n=3).

Table 4.1. Absorbance measurements of individual AP assay components in a 100 µL reaction volume.

	Absorbance$_{245}$(A.U.)	
Reagent	**96-well plate**	**384-well plate**
TRIS (100 mM, pH 9)	0.00	0.01
DMSO (10% v/v)	0.11	0.29
MBA (5 mM)	0.19	0.58
Pyruvate (5 mM)	0.25	0.73
Water	0.08	0.08
Expected assay solution absorbance	**0.62**	**1.68**
Measured assay solution absorbance	**0.62**	**1.69**

4.2.3 Determining the dynamic range of the spectrophotometer

The CLARIOstar Plus Microplate Reader (CLARIOstar plate reader) was used to monitor all reactions performed in this thesis and generate all absorbance measurements. According to the manufacturer, it has a dynamic range of 0 – 4 arbitrary units (A.U.) (https://www.bmglabtech.com/clariostar-plus/).

To start the process of miniaturization, the linear range of the plate reader was first confirmed. As the AP assay is based on the detection of acetophenone absorbance at 245 nm, serial dilutions of acetophenone were prepared at a range of volumes in both 96- and 384-well plates. The volumes ranged from 70 µL to 320 µL for 96-well plates and 20 µL to 100 µL in 384-well plates as these covered the widest range of reaction volumes for both plate types.

Comparing standard curves showed that increasing the volume decreases the concentration that can be detected within the linear range in both plate types which spans from 0 to ~2 units (A.U.) (Figure 4.3 A, B, and C). This can be observed in Figure 4.3 A and B, as the measured absorbance of acetophenone goes beyond the linear range after ~2 units (A.U.) indicating that any absorbance detected between 2 and 3.5 units is part of the non-linear range. No values were measured between 3.5 and

4 units, indicating that the actual dynamic range of the CLARIOstar was 0 – 3.5 units instead of up to 4 units as stated by the manufacturer.

The plate type was found to have no impact on the linear range. This was confirmed by normalising the volume and concentrations of each serial dilution by the well area (Figure 4.3 C). Furthermore, lower reaction volumes allow the measurement of higher Acetophenone concentrations for both plate types (Figure 4.3 D). Therefore, selecting a lower reaction volume would allow the capture of reaction conditions that produce higher quantities of Acetophenone, which is beneficial for future biocatalytic reaction screening experiments (Chapter 5).

In summary, this experiment confirmed the dynamic range of the plate reader and identified that a lower reaction volume must be selected to precisely detect greater acetophenone concentrations, with no preference for plate type.

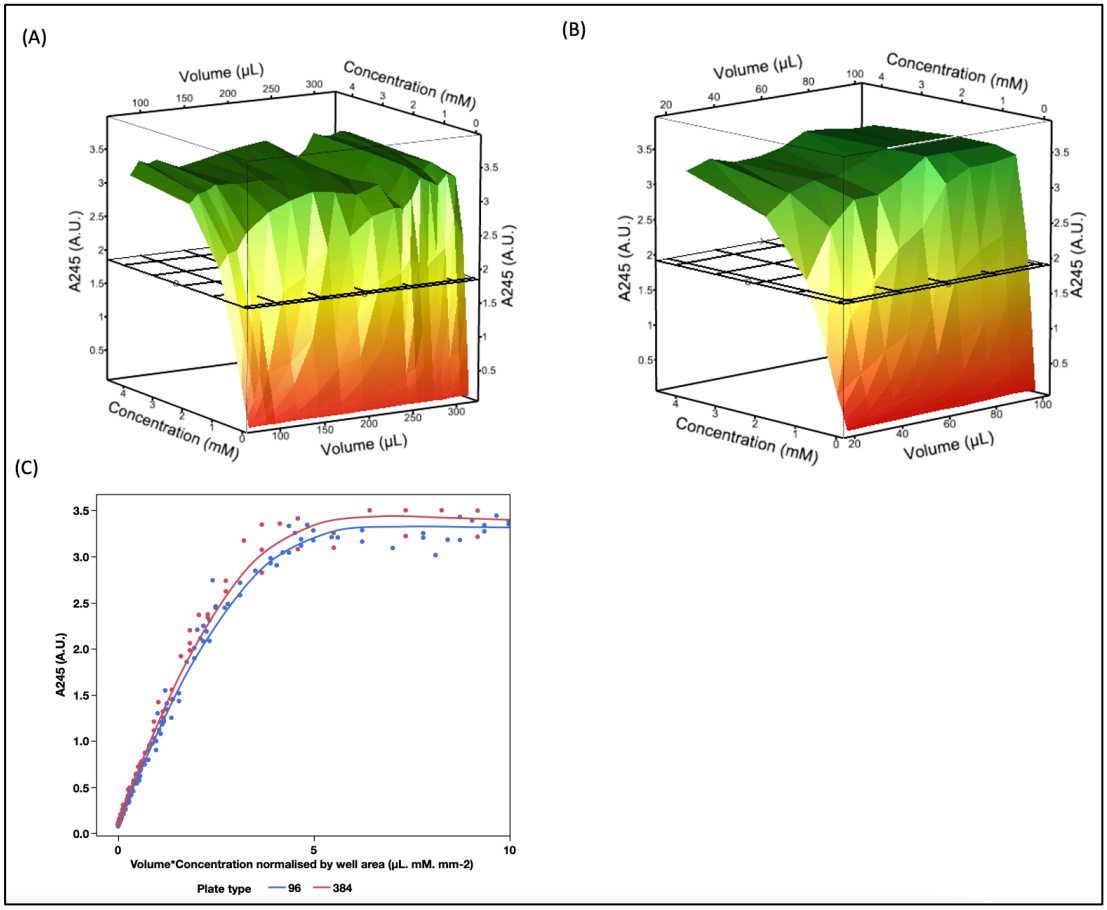


Figure 4.3. Determination of the dynamic range of the plate reader.
(A), (B) The effect of concentration and volume on absorbance in 96-well (A) and 384-well plates (B) is shown using a response surface plot. The surface was produced by preparing Acetophenone concentration curves at varying volumes as described in Section 2.4.2.2. Red to green shading represents increasing absorbance. The grids (black) indicate the maximum absorbance value of the linear regime. (C) Absorbance as a function of volume and concentration normalized by well area. Equations used to calculate the values on the X-axis are shown in Section 2.4.2.3.

4.2.4 Path-length correction

Path-length correction is a method of normalising absorbance values to the liquid path length of a standard square type cuvette of 1 cm. The overall dimensions of which are as follows: the external dimension of 1.25 x 1.25 cm, the height of 4.5 cm and, internal size of 1 x 1 cm. This is done to generate standardised absorbance measurements that can be used to compare data from different experiments and brands of plate readers (Roskoski, 2014).

Path-length correction calculations from raw absorbance measurements require a correction factor called the 'K-factor' (Figure 4.4 A) which is calculated by measuring the absorbance of the solution of interest at 977 nm and 900 nm in both a cuvette and a microplate. As water has an absorbance peak at 977 nm, the 900 nm measurement is used as the blank to eliminate the background absorbance of the plate and cuvette materials (Figure 4.4 D). The ratio of the blank corrected water peak of the microplate and the cuvette is the K-factor, which is then used to correct raw absorbance values (Lampinen *et al.*, 2012).

The CLARIOstar plate reader has an automatic setting to path-length correct absorbance values using the pre-set K-factor of water of 0.18. However, this feature was found to have many limitations. The K-factor of water can only be used for primarily aqueous solutions (Lampinen *et al.*, 2012; Siguemoto and Gut, 2017). If it is used for reactions with a high protein or solvent content, as used here, in AP reactions, then inaccurate data can be generated. Therefore, K-factors for each solution of interest must be calculated so that factors that affect path length, such as reagent concentrations and meniscus shape, can be accounted for. The discrepancies between default path length values used by the CLARIOstar Plus (https://www.bmglabtech.com/path-length-correction/) and the actual path length values generated for AP are shown in Figure 4.4 E.

(A)

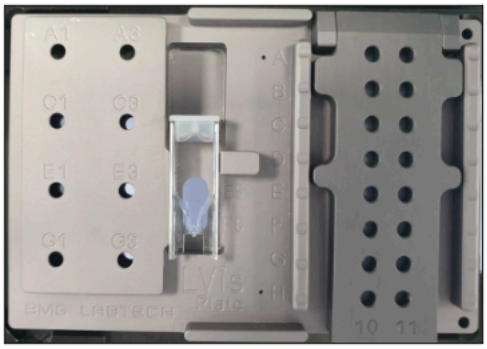
Path length (mm) = (A977 (well) – A900 (well) / A977 (cuvette) – A900 (cuvette)) x 10

K-factor

(B)

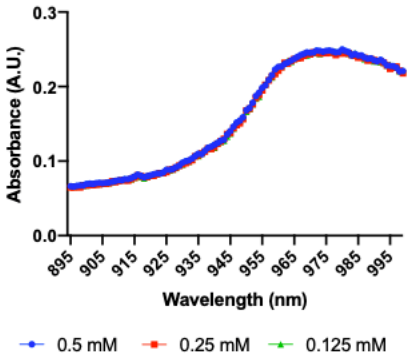
	0.5 mM		0.25 mM		0.125 mM	
Wavelength	**Absorbance**	**SD**	**Absorbance**	**SD**	**Absorbance**	**SD**
A900	0.0695	0.0007	0.0675	0.0007	0.0675	0.0007
A977	0.2475	0.0007	0.2455	0.0021	0.2460	0.0000
K-factor (A977-A900)	0.1780		0.1780		0.1785	

(C)



Measurement wells: U18 and U19 of a 384-well plate.

(D)



(E)

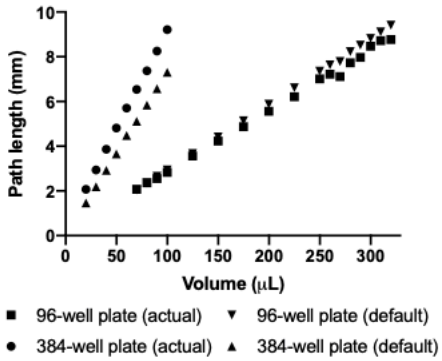


Figure 4.4. Path-length correction methodology.
(A) Equation to calculate the path length of a liquid in a microtiter plate. (B) Calculated K-factor for Acetophenone at three different concentrations. One standard deviation (SD) from the mean was calculated (n=3). (C) CLARIOstar's LVis plate showing the cuvette slot required to generate the A977 and A900 measurements for calculating (A). (D) Spectrum curves (895-1000 nm) for Acetophenone at varying concentrations highlighting the 'water peak' at 977 nm. (E) Comparison of default path-lengths generated by the CLARIOstar with path-length calculated using Acetophenone's K-factor from (B) in relation to volume and plate type.

4.2.5 Plate type, and reaction volume selection

Identifying the linear range of the plate reader and measuring the path-length for various reaction volumes, as presented in Sections 4.2.3 and 4.2.4, determined that using a 70 µL reaction volume in a 96-well plate and a 20 µL reaction volume in a 384-well plate would have the same path length of 2.07 mm. Under these assay conditions, the resulting absorbance measurements should be identical according to theory. To confirm this, and to gather information to select a suitable plate type and reaction volume, AP assays were performed at these reaction volumes and with these plate types.

Almost identical progress curves were generated (Figure 4.5) with an increase of 1.56 and 1.73 A.U. after 60 minutes of bioconversion for reactions prepared in 96- and 384-well plates respectively. This confirmed that the AP assay can be miniaturized from 200 µL to 70 µL in a 96-well plate successfully, and that experimental throughput can be further increased by 4-fold if a 20 µL reaction is prepared in a 384-well plate. However, the average standard deviation of the 70 µL progress curve in the 96-well plate was 0.0257 units while an average standard deviation of 0.0454 units was measured in the 20 µL progress curve in the 384-well plate indicating that 384-well plates generate nosier measurements. A tradeoff between throughput and precise measurements must therefore be made when evaluating AP assay logistics for subsequent DoE studies (Chapter 5).

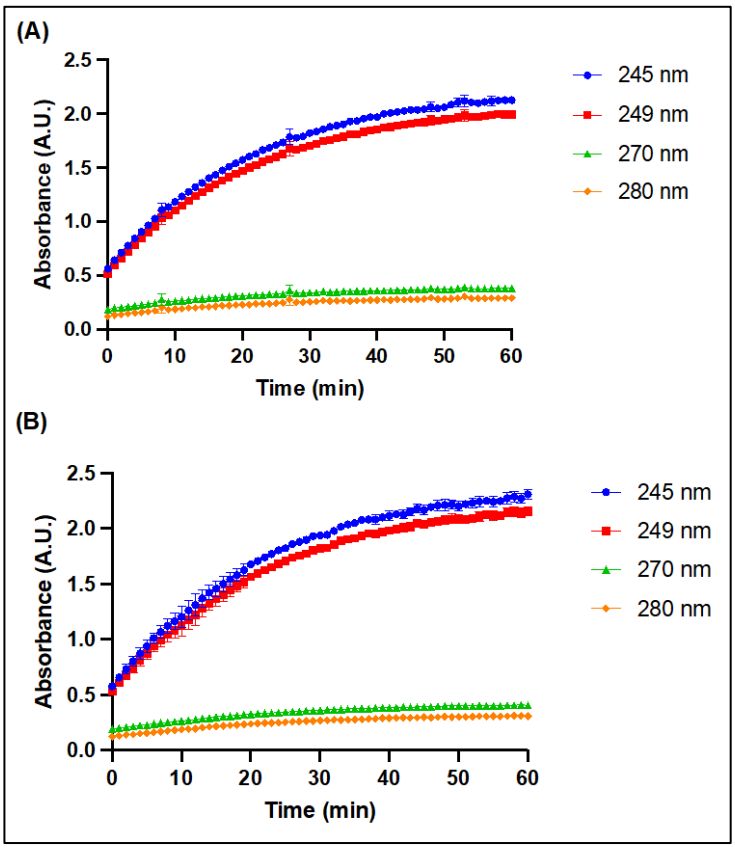


Figure 4.5. Comparison of AP assay progress curves in 96-well and 384-well microtiter plates.
Progress curves generated in a 96-well plate at a final volume of 70 µL (A) were compared to those in 384-well plates at a final volume of 20 µL (B) since both reaction solutions have the same theoretical path length (Figure 4.4). The bioconversion of pyruvate to acetophenone was monitored by measuring acetophenone absorbance at 4 different wavelengths for 60 minutes. Assays were prepared according to the reaction conditions described in Section 2.3.2.2. Error bars represent one standard deviation from the mean (n=3).

4.2.6 Selecting suitable measurement formats

Successful implementation of a DoE approach requires a thorough understanding of which responses are best to investigate in order to test a particular hypothesis (Hecht, Oberg and Muddiman, 2016). When responses are not well thought out prior to execution, there is a risk of collecting insufficient data or using an incorrect data format. With absorbance measurements, data gathered can be continuous i.e., time-course measurements or discontinuous i.e., end-point measurements (Acker and Auld, 2014).

Time-course data was selected as the preferred measurement format for DoE studies as regular measurement of assay signal provides more data points, and therefore more information about the assay. It is better suited to isolate responses important for bioprocessing responses such as initial rate and final reaction yield which would not be possible with an end-point measurement (Roskoski, 2014). Additionally, end-point values can be interpolated from the progress curves at set time points if needed.

The CLARIOstar plate reader scans a spectrum from 220 to 1000 nm for all selected wells and can record data for a maximum of 8 wavelengths for each well. Measurements showed that it requires 2 minutes to scan a full 96-well plate and 6 minutes to scan a 384-well plate. Based on this, the 96-well plate was favoured despite the 4-fold reduction in throughput to produce a more detailed dataset to enable a thorough understanding of the design space being investigated.

4.2.7 Identifying optimal measurement wavelength

Typically, the wavelength used to measure spectrophotometric assays is the one that gives the largest difference between the background absorption of the assay solution and the product being measured (Roskoski, 2014). In literature, the transaminase AP assay is most commonly measured at 245 nm, as originally developed (Schätzle *et al.*, 2009a), however, the signal can be measured up to 290 nm to avoid background interference (Jouda *et al.*, 2012; Chen *et al.*, 2016). As DoE studies can investigate the impact of numerous reagents on reaction kinetics, it is important to confirm the best measurement wavelength that maximizes signal while minimizing the absorption of the assay solution.

Consequently, the absorption spectra of all reagents that could be used to characterize the AP assay in Chapter 3 were measured (Figure 4.6). This included substrates, product, co-factor, enzyme, solvent, salts, and buffers (Section 3.2.4). As expected, acetophenone had a peak at 245 nm, however, at 249 the absorbance of other assay reagents such as DMSO, pyruvate, CHES buffer (pH 9.5, 125 mM), and

MBA was slightly lower. The baseline absorbance of a hypothetical assay solution was then calculated at both 245 nm and 249 nm. At 249 nm, the overall absorbance was 0.37 units lower than the measurement at 245 nm (Figure 4.7 A). Absorbance measurements beyond 270 nm were past the shoulder of the acetophenone detection peak and would be less beneficial in detecting differences in signal between varying reaction conditions.

However, if the AP assay is measured at a wavelength beyond 270 nm a higher concentration of acetophenone can be measured accurately (Figure 4.7 B); a maximum of 1.5 mM acetophenone can be detected at 245 and 249 nm while up to 12.5 mM acetophenone can be detected at 270 and 280 nm. The latter provides a better opportunity to monitor the time course of reactions with higher acetophenone yields.

As the CLARIOstar can record 8 wavelengths simultaneously, it was decided to measure reaction progress at wavelengths of 245, 249, 255, 270, 275, and 280 nm in case reaction conditions that saturate the detector at 245 and 249 nm are identified in future DoE studies (Chapter 5). This would ensure that no information about improved reactions would not be lost. Additionally, 900 and 977 nm wavelengths were also recorded for any future work that could be conducted on the topic of path-length correction as they would enable K-factor calculations described in Section 4.2.4.

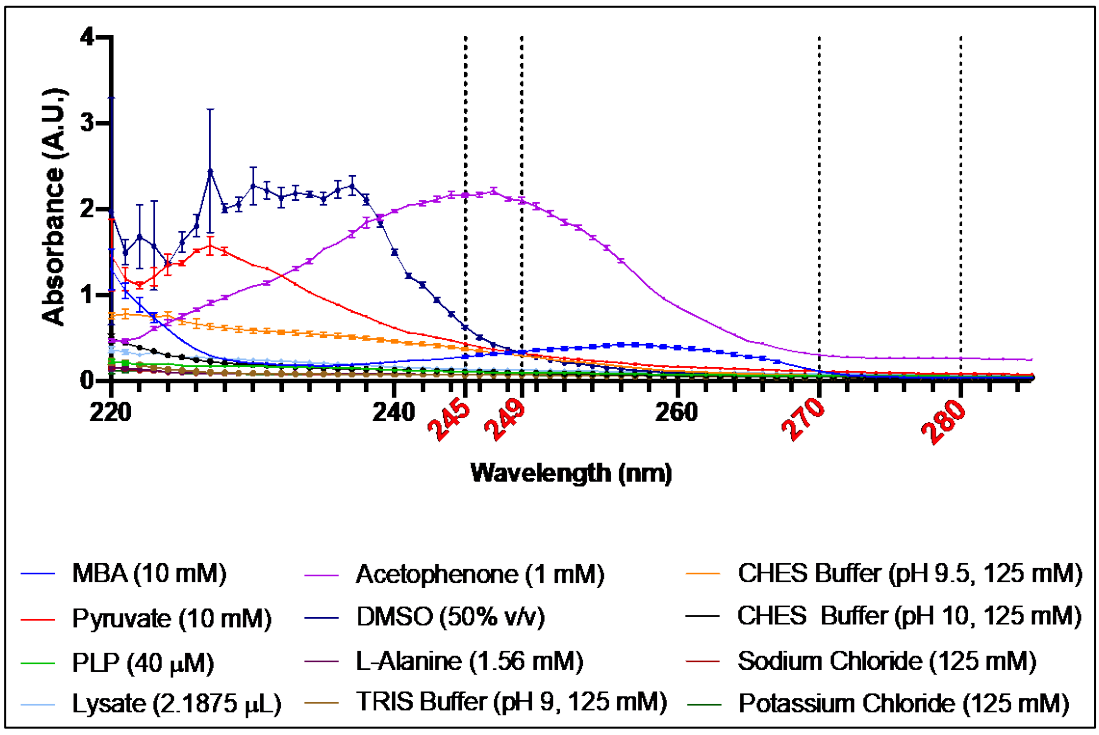


Figure 4.6. Absorbance spectra of individual AP assay reagents.
Spectrum profiles (220 – 285 nm) of reagents (70 µL, 96-well plate) that might be incorporated into DoE investigation of AP assay conditions are shown. The black vertical lines and corresponding red text indicate potential measurement wavelengths to capture acetophenone signal (purple). Reagents were prepared as described in Section 2.1.1. Error bars represent one standard deviation from the mean (n=3).

(A)

Reagent	Absorbance$_{245}$(A.U.)	Absorbance$_{249}$(A.U.)
MBA (10 mM)	0.2893	0.3480
Pyruvate (10 mM)	0.4367	0.3287
PLP (40 µM)	0.1113	0.1023
L-Alanine (1.6 mM)	0.0820	0.0793
DMSO (50%)	0.6360	0.3350
Buffer (125 mM, TRIS pH 9)	0.0780	0.0747
Salt (any)	0.0750	0.0750
Water	0.0750	0.0750
Total absorbance	**1.7833**	**1.4180**

(B)

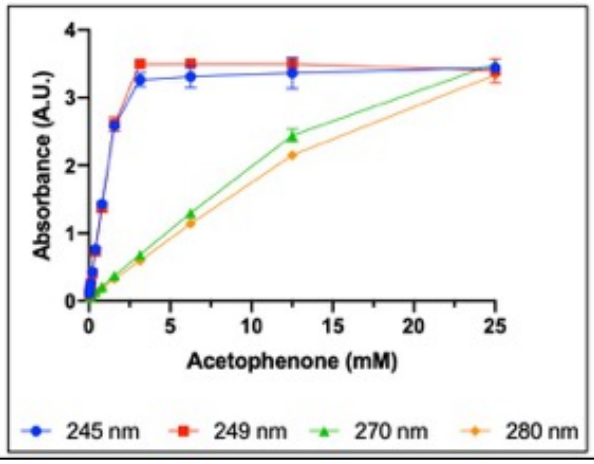


Figure 4.7. Total absorbance of AP assay reaction reagents and effect of measurement wavelength on Acetophenone concentration.
(A) Comparison of the absorbance contribution of AP assay reagents with respect to measurement wavelengths. (B) Acetophenone concentration curves at varying AP assay measurement wavelengths. Concentration curves were prepared as described in Section **Error! Reference source not found.**. Error bars represent one standard deviation from the mean (n=3).

4.2.8 Liquid-handling precision comparison of different devices

Liquid-handling accuracy and precision is a prerequisite of successful automated experimentation. Precision refers to how close measured replicate values are to each other, while accuracy refers to how close measured values are to the true value. When preparing automated experiments, it is crucial that automated devices are properly calibrated to ensure that the resulting liquid-handling actions are

transferring and mixing liquid volumes that are both precise and accurate thus avoiding the generation of erroneous data (Hentz and Knaide, 2014). Therefore, Routine Quality Control (RQC) must be performed regularly to maintain device calibration (Albert, 2007).

Two precision tests were conducted for the three automation devices used in this work comprising two liquid handling devices (Hamilton STAR and PipetMax), and one dispenser: (Dragonfly). These platforms were described in Section 2.1.4.

The first test used the Artel MVS® Verification System, a ratiometric dual dye-based test that is used to quantify specified dispense volumes (Bradshaw *et al.*, 2005). The precision of 1 µL dye addition to 199 µL of diluent was measured by using each channel or syringe to dispense 8 replicates (Figure 4.8). A 1 µL transfer was selected as it is in the lower end of the dispensing range for all three devices. The PipetMax and Hamilton STAR liquid handling devices have a dispense range of 1 – 1000 µL depending on tip type, while the Dragonfly dispenser can dispense between 0.2 µL – 4 ml. For a 1 µL dispense, Coefficient of Variances (CV) of 8, 6, and 5% can be expected from the PipetMax liquid handler, Hamilton STAR and Dragonfly® discovery, respectively according to the manufacturers (SPT Labtech, no date; Hamilton, 2015; Gilson Inc., 2018)

Based on the experimental results obtained here, the precision of the Hamilton STAR and the Dragonfly dispenser were closely matched. The Hamilton STAR was found to be the most precise, with an average CV of 2.90 %; 5 of the 8 channels had a CV of < 3% and an overall CV of < 5 %. The Dragonfly dispenser closely matched the precision of the Hamilton STAR with an average CV of 3.43 % and maintained a CV of < 5 % across all syringes. The PipetMax liquid handler was most imprecise with the highest CV of 107.79 % for Channel 1, due to missed dispenses for 4 out of 8 replicates. Under dispensing was also observed for one replicate in both Channels 2, and 8. This showed that the device would need to be serviced or recalibrated before it can be used for any other experiments.

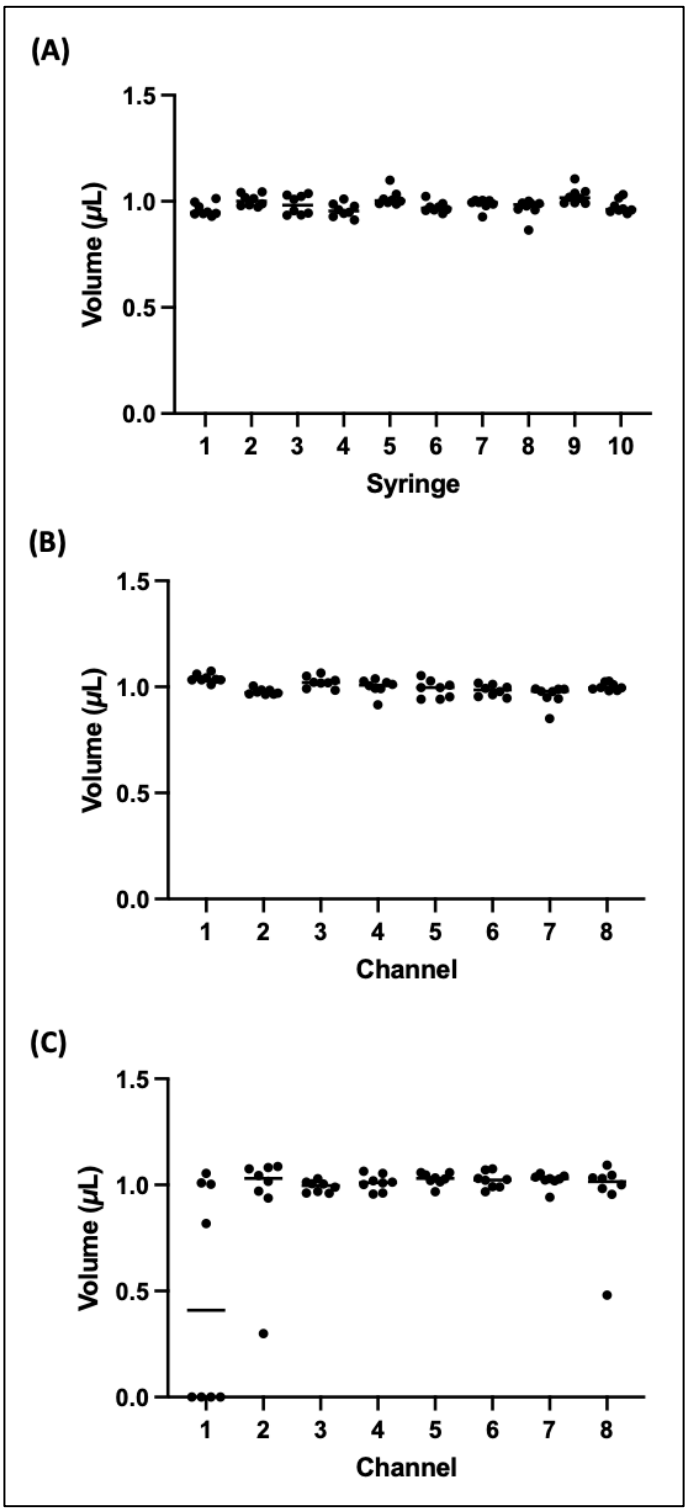


Figure 4.8. Liquid handling precision of the three automation platforms used in this study. All syringes and channels of the (A) Dragonfly dispenser, (B) Hamilton STAR, and (C) PipetMax liquid handlers were used to dispense 1 µL dye into 199 µL diluent to test dispensing precision using the Artel MVS System as described in Section 2.5.1. Dots represent detected volume; horizontal lines represent the median (n = 8).

As the Artel system is not commonly found in most laboratories due to its capital cost and specific function, a second precision test was designed using Tartrazine, an inexpensive yellow food dye, to test liquid handling precision (Jones, Clark and Clulow, 2003). The precision of preparing dilutions at a target concentration of 0.1 mM for spectrophotometric detection at 425 nm using various sampling volumes in both 96-well and 384-well plates at a final volume of 70 and 20 µL, respectively, for all three devices was investigated (Table 4.2). The stock solution was prepared by setting up a 2-fold serial dilution first and sampled to produce the final stock concentration of 0.1 mM. This ensured that the precision of the serial dilution is decoupled from that of the liquid transfer steps used to prepare the final solution.

Table 4.2. Transfer volumes used for tartrazine-based liquid-handling precision and accuracy test.

	Tartrazine concentration (mM)		96-well plate (Final volume = 70 µL)		384-well plate (Final volume = 20 µL)	
	Stock	Final	Tartrazine volume (µL)	Diluent volume (µL)	Tartrazine volume (µL)	Diluent volume (µL)
Dilution 1	1.6	0.1	4.38	65.63	1.25	18.75
Dilution 2	0.8	0.1	8.75	61.25	2.5	17.5
Dilution 3	0.4	0.1	17.5	55	5	15

Liquid handling policies (LHPs), i.e., instructions that control how a liquid is handled when transferred from one microwell (Appendix H) can be defined for both liquid handlers. Four such policies were selected based on variations in pre- and post-mixing steps for both liquid handlers to ensure that a range of LHPs can be investigated. However, no obvious differences were observed with any of the policies in any of the datasets. The Dragonfly dispenser contains only one hardcoded default liquid handling policy; therefore, the device was used as supplied by the manufacturer.

Figure 4.9 summarizes the results of the tartrazine-based liquid handling precision and accuracy test performed across all three devices and both 96- and 384-well

microplates. Due to the larger liquid volumes required, preparation of the 96-well plates for all devices yielded both precise and accurate results for all liquid handling policies and dilution volumes as all absorbance was within the expected range of 0.5 – 0.6 A.U. (Figure 4.9 A and C).

However, more inconsistent results were found when lower dilution volumes were sampled in 384-well plates for both devices. With the PipetMax liquid handler, the dilutions get more precise when larger dispensing volumes are used, indicating that higher transfer volumes are preferred when using 384-well plates despite the manufacturer stating a minimum dispensing volume of 1 µL.

More inconsistent results were found with the Hamilton STAR. The 1.25 µL dispenses were the most accurate and precise with recorded absorbances of ~ 0.5 A.U. or slightly lower, indicating slight under pipetting. The 5 µL transfers also largely clustered around the expected absorbance, however, each group had 2-3 outliers that indicated missed or under-dispensing. However, the majority of the 2.5 µL transfers had an absorbance of ~ 0.3 A.U. with only 1-3 data points generating expected absorbance. With the Hamilton STAR, two entities are required to execute a liquid handling action: the pipetting mode (such as a Liquid Handling Policy (LHP)) and a 'liquid class' which stores background parameter information e.g., aspiration and dispense rates. Standard liquid classes are provided by the manufacturer, however custom classes might need to be developed for certain liquids (Hamilton, 2015). The data generated for the 384-well plate indicates that the underlying liquid class needs to be recalibrated before repeating the investigation of liquid handling precision and accuracy.

Of the three devices evaluated, the Dragonfly dispenser was the most precise device at dispensing low volumes as all absorbance measurements were clustered around 0.5 A.U. The variable nature of the results generated across the three devices highlights the importance of performing precision tests prior to experimentation and of ensuring the devices are appropriately serviced and maintained. Overall, the

Dragonfly dispenser performed the best across both accuracy and precision tests and with both plate types. Among the devices tested, it was also the only device capable of fast multi-reagent mixing over wide volume ranges. The combination of speed and easy instruction generation using Synthace makes it best suited for future DoE studies as will be described later in Chapter 5.

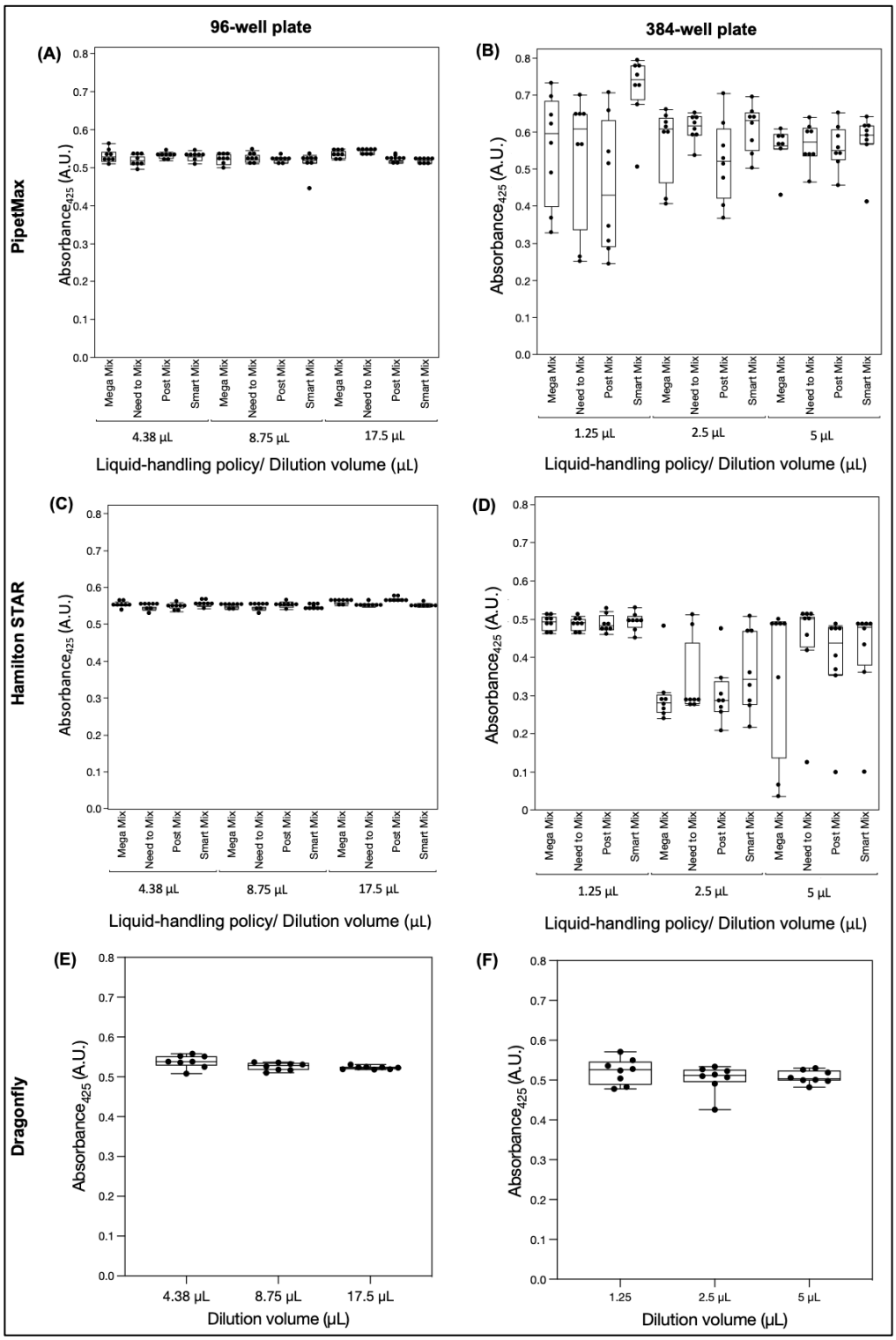


Figure 4.9. Box plots showing liquid-handling accuracy and precision of three liquid handling platforms with 96well- and 384-well plates.

(A), (B) Precision of preparing 0.1 mM tartrazine in 70 µL in a 96-well plate (A) and in 20 µL in a 384-well plate (B). Precision of liquid transfers was tested by sampling 3 different stock solutions (Dilutions 1, 2 and 3) 8 times each (black dots) using 4 different liquid handling policies (Appendix H) for the Hamilton and PipetMax liquid handling devices. The absorbance of tartrazine in the final solution was measured at 425 nm for each sample. Data points that fall outside the whiskers are considered outliers.

4.2.9 Effect of plate film types on evaporation and absorbance measurements

For any screening assay or enzyme reaction, it is important that evaporation is minimal over the duration of any measurements. Additionally, absorbance measurements need to be consistent across the whole plate without any "edge effects", i.e., heightened evaporation that is typically observed at the boundary of the microtiter plate due to the plate film not being attached firmly throughout. Incubating an enzyme in the assay solution, in the absence of substrates that initiate a bioconversion, over a prolonged period is a method of testing enzyme stability under various bioconversion conditions (Bos *et al.*, 2015). Consequently, two plate films; MicroAmp™ Optical Adhesive Film (Applied Biosystems®) and Aluminum Sealing Films (Starlab Ltd.) were investigated to identify one that is best fit for preventing evaporation during the AP reaction.

Tartrazine (0.1 mM) was added to each well in both 96-well and 384-well plate formats at a final volume of 70 and 20 µL respectively. Absorbance at 245 nm was then recorded before and after a 24 incubation at 45 °C (Appendix I) to investigate if the absorbance measurement remained constant which would indicate no evaporation. The temperature was set to 45 °C which is the maximum temperature that can be set on the spectrophotometer and was selected as it represents the most extreme assay conditions that would be investigated in the future (Chapter 5).

No edge effects were observed in all four plates between both plate films and microtiter plate types. However, the MicroAmp film was selected for future work as it performed ~3 fold better at minimizing evaporation than the Aluminum films for both plates (Table 4.3). The average reduction in absorbance was 0.0188 and 0.0072 units for 96- and 384-well plates, respectively which indicates almost no detectable evaporation as the precision of the plate reader is 0.02 units.

Table 4.3. Effect of plate film on evaporation (% v/v) in 96-well and 384-well plate formats.

Plate type	96-well		384-well	
Plate film	MicroAmp	Aluminium	MicroAmp	Aluminium
Evaporation (%)	3.45	9.26	1.39	4.27
SD	1.80	3.70	1.36	2.10

To prevent evaporation while time-course measurements are taking place inside the plate reader, a sealing film that minimises both evaporation and interference with the absorbance measurements is required. As the MicroAmp films were found to be suitable at preventing evaporation over prolonged periods of time, its impact on absorbance measurements was also investigated.

AP reaction progress curves were compared between unsealed plates and those sealed with MicroAmp films as shown in Figure 4.10. Largely identical progress curves were generated, with an increase of 1.044 and 1.151 A.U. calculated from reactions from unsealed and sealed plates respectively. Due to the addition of the MicroAmp plate film, the baseline absorbance increased slightly by 0.186 A.U. but this was not found to significantly reduce the dynamic range of the assay and was therefore deemed suitable for sealing microplates for absorbance measurements.

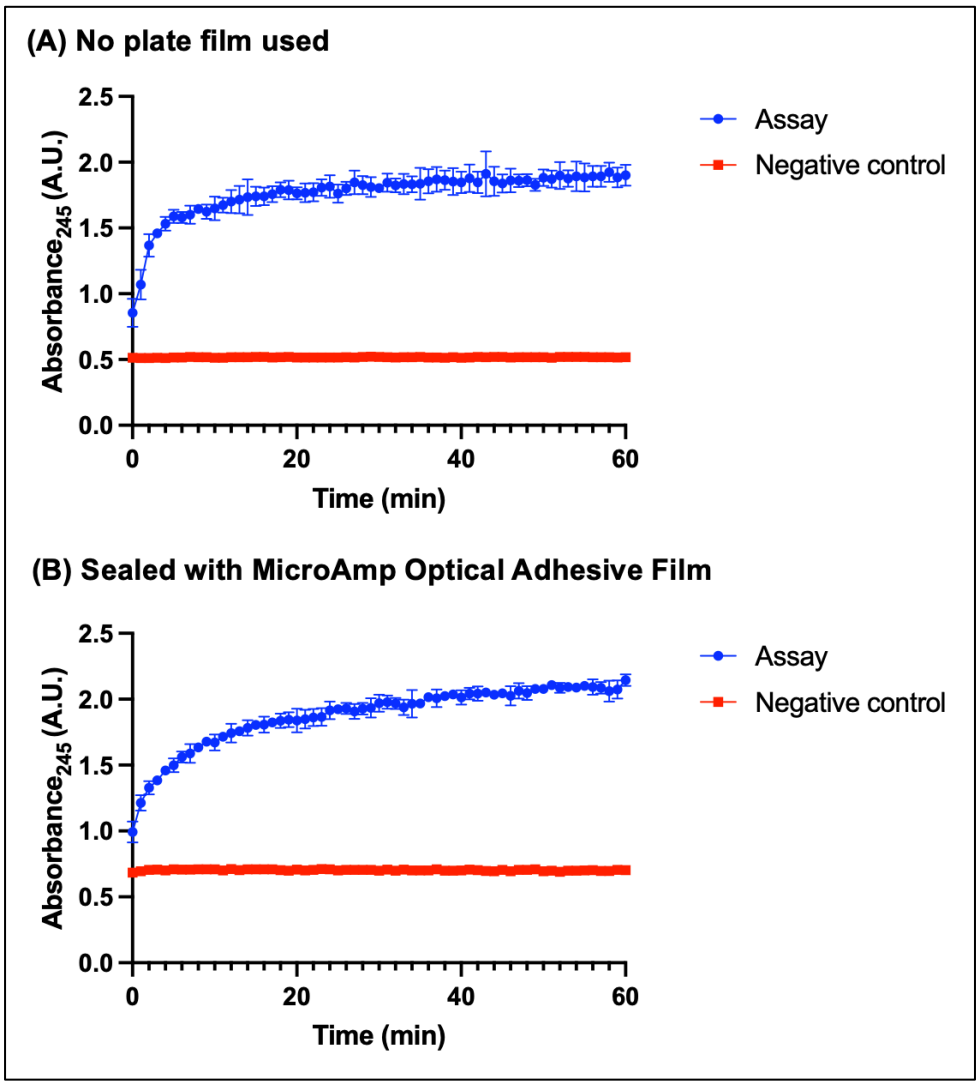


Figure 4.10. Comparison of AP reaction progress curves prepared in unsealed and sealed microtiter plates.

Progress curves generated with and without the MicroAmp plate film were compared to identify the impact of an additional plate film on the baseline absorbance. Reactions were prepared as described in Section 2.3.2.2 in 96-well plates at a reaction volume of 70 µL using a PipetMax liquid handler. Error bars represent one standard deviation from the mean (n=3).

4.3 Discussion

The aim of this chapter was to systematically investigate a subset of assay logistics and select optimised assay logistics for carrying out the AP assay for investigation of transaminase activity The intention was to establish a solid foundation for automated, high-dimensionality experiments to be conducted later in Chapter 5.

The approach to optimizing assay logistics was conceptualized in a flowchart as shown in Figure 4.1. This was established based on knowledge gained from experiments conducted in Chapter 3 and literature based on the AP assay and automated high-throughput assay execution. Six different assay logistics were investigated, and the optimal setting was selected for each based on a set of decisions (Figure 4.1).

The first step indicated in Figure 4.1 was to determine the linear range of the plate reader provided confidence in the quality of absorbance data being generated from the particular CLARIOstar plate reader used in this work. The linear range of the plate reader was found to be between 0-2 units A.U. (Figure 4.3) with a maximum limit of 3.5 units for both 96-well and 384-well plates. Therefore, a linear model would need to be used for absorbance-based calculations for up to 2 units, however, a non-linear model may be required to distinguish between data points greater than 2 units. Absorbance measurements generated from both 96-well and 384- well plates of comparable path length were largely comparable, indicating that plate type did not further impact with generating absorbance data. However, only one brand of microplates (UV Star, Grenier-Bio) was compared, therefore, any new plate types would need further confirmation.

Next, a method for path length correction using the CLARIOstar plate reader was outlined as path length corrected values are found routinely when discussing absorbance-based assays (Figure 4.4). The default path length setting of the CLARIOstar plate reader was found to be imprecise as the calculation used the path

length of water (0.18) to normalize absorbance data from every microwell, when, in reality, reaction mixture composition could vary vastly. However, path length correction using the method outlined in Figure 4.4 would be inefficient for DoE studies as every microwell contains a different reaction mixture, and in turn, has a different K-factor making K-factor calculations unfeasible. However, this method can be used in future use cases, for example, to normalise absorbance measurements for reaction conditions used in high-throughput screening.

In Section 4.2.5, the logistic of selecting microplate type and reaction volume was investigated. The AP assay reaction volume was miniaturised from 200 µL, which was consistently used in the literature (Schätzle *et al.*, 2009a; Meng *et al.*, 2020; Heckmann, Dominguez and Paradisi, 2021), to 70 µL in a 96-well plate. This had the advantage of reducing reagent costs by almost 3-fold. Additionally, using a 384-well plate can further miniaturise the AP assay to a 20 µL reaction, as 20 µL in a 384-well plate and a 70 µL reaction in a 96-well plate both have the same path-length of 2.07 mm (Figure 4.5). This would increase throughput by 4-fold and simultaneously reduce reagent cost by a further 3.5-fold.

AP assays executed at these reduced volumes generated almost identical progress curves as indicated in Figure 4.5. However, a number of considerations favor the selection of 96-well plates for future work; the average standard deviation values of absorbance measured from 70 µL progress curve in a 96-well plate (0.02570) were half as much as that calculated from 20 µL progress curves in a 384-well plate (0.0454) (Figure 4.5). As the standard deviation generated from the 384-well plate is still very small, it would be appropriate to select the 20 µL reaction in a 384-well plate for future DoE studies. However, as DoE studies investigate unknown and potentially noisy design spaces, logistics that produce the most precise data were favored over throughput. Additionally, a larger reaction volume has more capacity to accommodate a greater number of reagents, this is particularly important in DoE studies where many reagents at varying volumes are used to prepare each experimental run condition and minimum dispense volumes are limited by the

capabilities of the automated device available. Furthermore, more data points can be generated from a 96-well plate as one measurement cycle takes 2 minutes for 96-microwells compared to 6 minutes for 384-wells using the CLARIOstar plate reader (Section 4.2.6). Time-course absorbance data with more data points provide a richer dataset to generate responses for DoE analysis.

It is also important to consider the logistic of selecting an appropriate measurement wavelength for absorbance measurements that capture the best assay signal (Section 4.2.7). As shown in Figure 4.7, 245 nm is a suitable measurement wavelength to measure acetophenone signal but careful consideration of the absorbance contribution from other assay reagents such as substrates and DMSO helped identify 249 nm as the wavelength that further decreases background interference (Figure 4.6 and Figure 4.7.).

Once the plate format and assay conditions were established, the next step was to investigate the performance of the automated liquid handling devices that would set up the assays. For these automated devices precision is critical as it provides confidence that the expected volume transfers and mixing are taking place. The liquid-handling precision of 3 devices: Hamilton STAR and PipetMax liquid handlers, and the Dragonfly dispensers were tested using two methods; using the Artel MVS Verification System and preparing Tartrazine solution for spectrophotometric detection (Figure 4.8 and Figure 4.9). Overall, the Dragonfly dispenser was found to be capable of precise dispensing for both tests with CVs staying within the manufacturer's range. It was selected as the most suitable automated device for executing large DoE studies that require numerous variable pipetting actions. Additionally, the precision testing identified issues with both Hamilton and PipetMax liquid handlers in the Synthace lab and highlighted the importance of confirming liquid-handling precision, especially prior to executing any experiments that have a large quantity of liquid-handling instructions which could lead to the generation of noisy data.

Finally, the impact of plate film on evaporation over 24 hours at 45 °C was investigated as described in (Section 4.2.9). The MicroAmp plates were found to ensure consistent assay conditions could be maintained and absorbance readings could be recorded with minimal interference as they showed a negligible amount of evaporation with no edge effects for both 96-and 384-well plates (Table 4.3).

In Chapter 5, a complex DoE study of 3,456 experimental runs will be defined and carried out in order to characterize the AP assay, using a space-filling design that would be out of reach for manual execution. The work presented in this chapter has confirmed the assay logistics applied to the ambitious large-scale DoE investigation.

5 Space-filling designs for biocatalytic reaction screening

5.1 Introduction

Space-filling designs (SFD) are a modern class of DoE design geared toward exploring complex, non-linear factor-response landscapes (Montgomery, 2013; Wu, 2015; Joseph, 2016). Unlike traditional DoE designs, which sample the design space at the boundaries defined by the low, high, and sometimes mid-point level of a factor range (Figure 5.1 A), a SFD samples throughout the factor ranges across numerous levels (Figure 5.1 B).

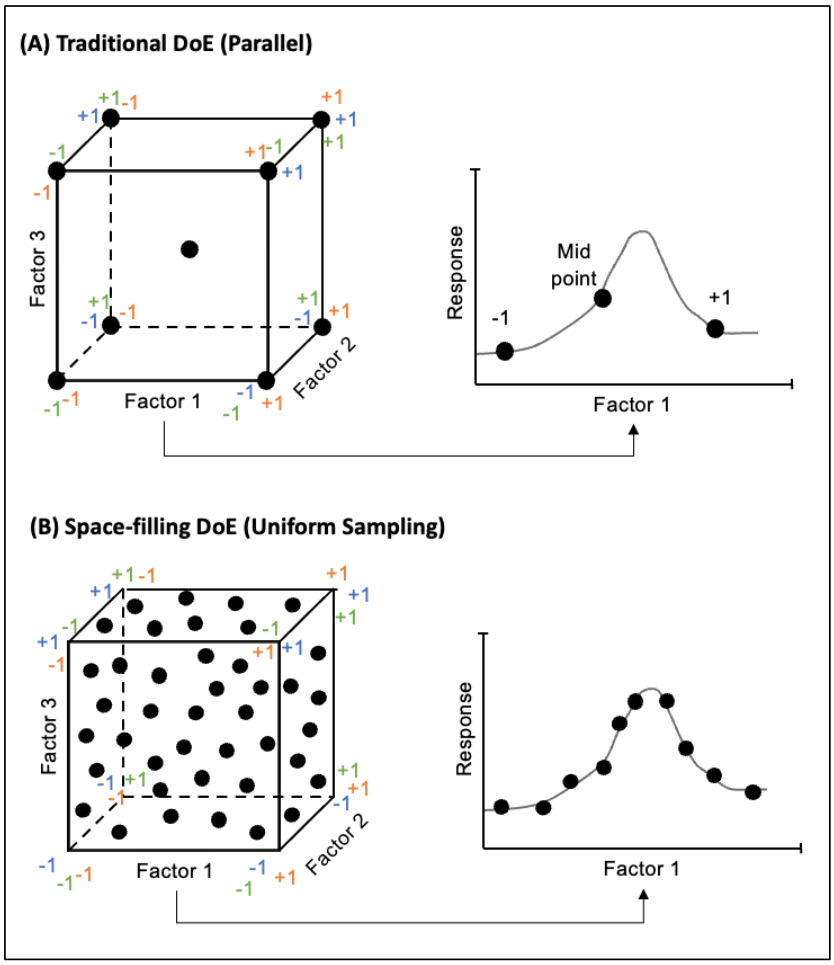


Figure 5.1. Comparison of a traditional DoE design to a space-filling DoE design. The distribution of run conditions (dots) and the impact of those conditions on the response in relation to a single factor are compared between the two design types.

SFDs can provide many advantages over traditional DoE designs. Their ability to investigate wide factor profiles reduces the pressure to conduct numerous scoping studies, and theoretically, they can eliminate the need to conduct scoping studies completely. Furthermore, they can map large high-dimensional spaces in potentially just one experimental step providing a precise view of the physicochemical spaces that enzymes function in while generating more data than that from traditional DoE designs.

However, traditional DoE designs are already minimally used in biology, due to various challenges surrounding laborious and error-prone physical execution along with the learning curve that comes with using statistical software (Section 1.1.5). Compared to traditional DoE designs, SFDs are significantly more complicated as they require high run numbers which need to be prepared using highly variable multi-reagent sampling. For example, a reagent for a continuous factor e.g., substrate concentration, would have to be prepared at potentially more than one stock concentration, each of which would then be sampled to prepare varying final concentrations which enables uniform sampling of that factor range. These challenges are highlighted in the literature as there is only one known example of a 100-run SFD study being employed to characterise enzyme activity where the researchers had to create bespoke software architecture to automate the execution of the experiment (Bonowski et al., 2010). Therefore, SFDs are near impossible to physically execute and are reserved for application with computer experiments that analyse a deterministic response (Kleijnen, 2005; Jones and Johnson, 2009).

However, the work carried out in this thesis aims to bridge the gap of executing and analysing the execution of SFDs by employing the combination of a rapid dispenser such as the Dragonfly dispenser (as used in Section 3.2.6.4) and Synthace's capability to generate liquid handling instruction files from a DoE design file (Section 3.2.5.1), potentially allows physical execution of a SFD.

5.1.1 Aim and objectives

The aim of this chapter was to demonstrate how the complex, high-dimensional physiochemical environments that biocatalysts function within can be finely and rapidly characterised by implementing just one automated SFD study to develop precise mathematical models that help understand what factors and factor settings lead to desired enzymatic activity. The AP assay (Figure 3.1) was used as the model system to demonstrate this framework as it allows the investigation of numerous generic factors of interest e.g., substrates, co-factors, temperature, solvents etc while generating quantitative time-course absorbance data which can be used to generate useful biocatalytic responses for model generation. Specific findings from Chapter 3 are ignored here so that all potential factors would be tested as if nothing was known about the model system.

The specific objectives of this chapter are as follows:

- Develop a generalised automated framework for rapid assay characterisation with automated SFD execution and data analysis (Figure 5.2).
- Execute a carefully controlled SFD experiment to characterise a large high-dimensional design space that can capture bioprocessing metrics of yield and initial rate where the design included all the key variables of interest to an enzymatic assay.
- To minimise or eliminate the time spent on scoping studies typically required prior to DoE execution to determine if a SFD can be executed as a one-step characterisation experiment.
- To generate and physically validate models for yield and initial rate.

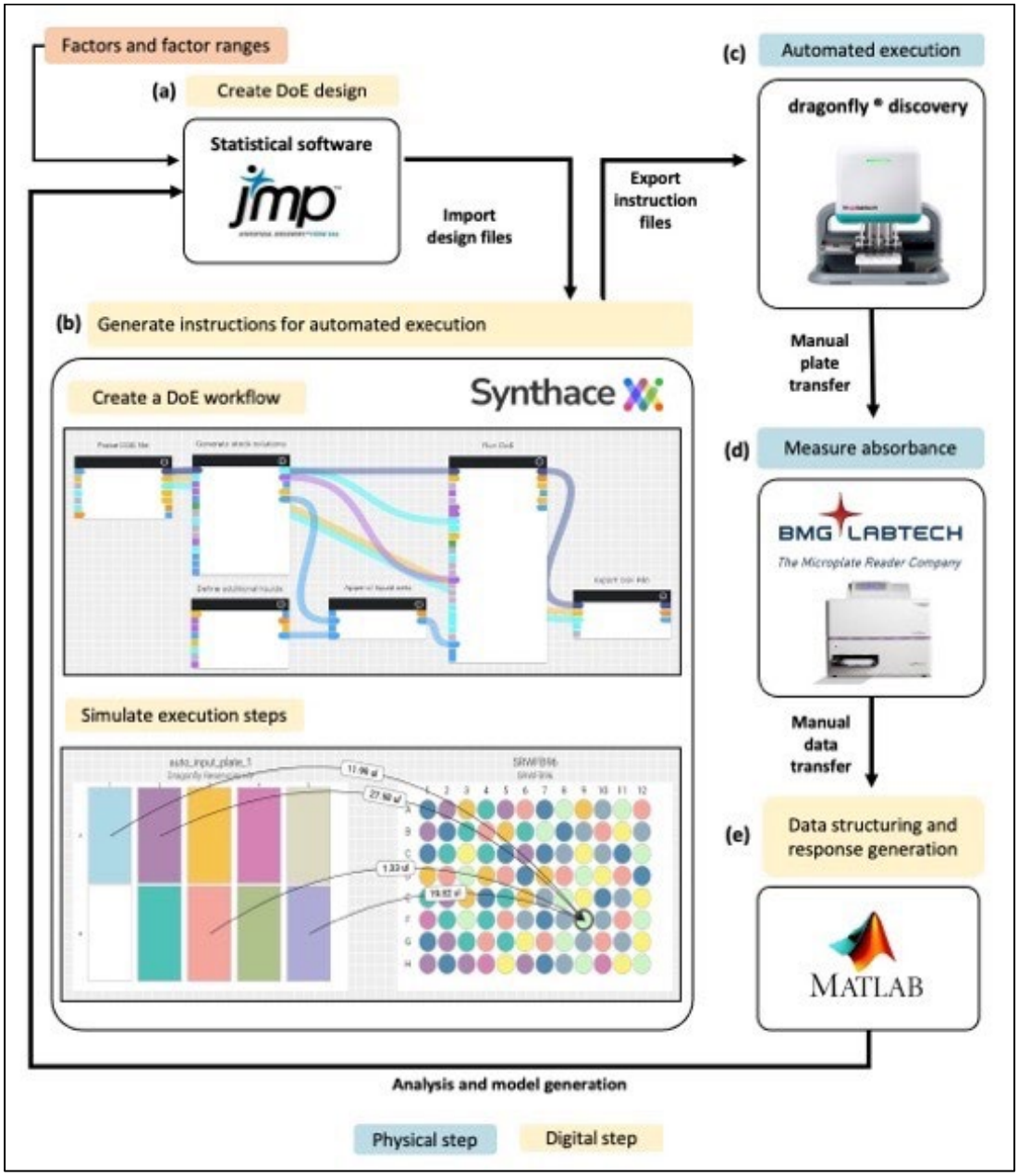


Figure 5.2. Framework for end-to-end automated execution and analysis of space-filling experiments.

A design file generated by JMP (SAS Institute Inc) is parsed into the Synthace platform, where a DoE workflow is used to convert details of individual run conditions into liquid handling instructions for physical execution on the Dragonfly dispenser. Once each plate is prepared, time-course spectrophotometric measurements are recorded (CLARIOstar, BMG Labtech). Once all plates are measured, the absorbance files are used as inputs to custom-made MATLAB (MathWorks) scripts (Section 2.7.2.1) that process the data by grouping replicates, time-correction, black-correction, and response generation. The responses are then joined to the original design file for analysis and model generation using JMP.

5.2 Results

5.2.1 Space-filling design details

In this section, the process of choosing factors, executing any scoping studies, and determining the number of runs needed for the SFD are first described before moving on to implementation.

5.2.1.1 Factor table and levels

The first step in generating a design file for a SFD is to define all factors of interest and their respective ranges. As the aim was to characterise a large design space of the AP assay in just one experiment, 12 key factors that commonly impact enzyme activity were selected and varied across wide ranges (Table 5.1). The structure of this SFD is considered to be 'generic' in that it could be applied to many different classes and types of enzyme-catalysed reaction

11 of the 12 factors were used to define the factor table that generated the design. These included factors such as substrate, enzyme, product, and co-factor (PLP) concentrations along with pH and temperature. Additionally, industrially important factors such as solvent (DMSO) concentration and salt concentration (both sodium chloride and potassium chloride) were also included as solvent and salt stability of an enzyme are characteristics that industrial biocatalysts are commonly optimised for (Stepankova *et al.*, 2013). While, in reality, pH and temperature are continuous factors, they were treated as categoric factors for design generation as they are considered hard-to-change factors. For example, if pH was treated as a continuous factor, a large number of buffers at varying pHs would be required to facilitate the design. Similarly, temperature is used as a blocking factor to divide experiments by microplate, each of which is controlled at a specified temperature as dictated by the incubator. Therefore, both pH and temperature were described as categorical factors with specific levels in the factor table. To cover a reasonably wide pH range, two buffer types (TRIS and CHES) were used, which allowed spanning a pH range of 7 to 10.

The 12th factor was 'enzyme pre-incubation' since understanding the stability of CV2025 TAm over time in different physiochemical environments was also of interest. This was investigated by incubating the CV2025 TAm lysate in the reaction mixture without substrates for 0, 8 and 24 hours, after which the substrates would be added to start the reaction. The enzyme pre-incubation factor was not included in the original design, instead, the design execution was replicated after three different enzyme pre-incubation time periods of 0, 8 and 24 hours so data could be captured for the same run conditions over time. However, to ensure that the effect of pre-enzyme incubation was captured, it was added as a continuous factor during the subsequent generation of statistical models as described in Section 5.2.6 and 5.2.7.

5.2.1.2 Scoping studies executed prior to confirming SFD design

Typically, multiple scoping studies are required prior to executing traditional DoE designs as factors are only sampled at the levels stated, at the boundary of the design space (Coleman and Montgomery, 1996; Politis *et al.*, 2017a). However, as a SFD samples the design space uniformly throughout, the amount of time and number of experiments required before executing the designed experiment was minimised.

Only two scoping studies were executed prior to generating the factor table (Table 5.1). Firstly, as every cycle of cell growth and lysis can vary, a scoping study that determined the range of lysate volumes used to vary enzyme concentration was executed (Figure 5.3). The reference reaction conditions generated from Chapter 3 (Section 3.2.5.5) were used to determine that 1-3 µL of lysate would be sufficient to quantify CV2025 TAm activity while not saturating the plate reader detector (Section 4.2.3). Secondly, to confirm that 50 mM was a suitable lower limit for buffer concentration to maintain the specified pH, a small number of reaction mixtures were prepared manually and tested with pH strips (data not shown) which confirmed that 50 mM was an appropriate lower limit.

Table 5.1. Factors and levels for a generic space-filling DoE to investigate enzyme reaction kinetics.
The specific application is exemplified in brackets for the TAM catalysed conversion of MBA and pyruvate to acetophenone and sodium alanine.

	Factor	Range	Factor type
1	Enzyme concentration (CV2025 lysate)	1-3 μL	Continuous
2	Substrate 1 (MBA)	0.25 - 5 mM	Continuous
3	Substrate 2 (Pyruvate)	0.25 - 5 mM	Continuous
4	Co-factor (PLP)	0 - 35 μM	Continuous
5	Product 1 (Sodium Alanine)	0.25 - 5 mM	Continuous
6	Temperature	25, 35, 45 °C	Continuous*
7	Solvent (DMSO)	0 - 40 % v/v	Continuous
8	Buffer type & pH	7 – 10 TRIS (7, 8, 9) & CHES (9, 10)	Categorical**
9	Buffer concentration	50 - 200 mM	Continuous
10	Salt type	NaCl and KCl	Categorical
11	Salt concentration	0 - 200 mM	Continuous
12	Enzyme pre-incubation	0, 8 and 24 hours	Continuous+

*Executed as a categorical factor but modelled as a continuous factor.
** pH modelled as a continuous factor; buffer type was excluded from analysis.
+ The whole design was created using Factors 1-12 and was replicated at these enzyme pre-incubation time points.

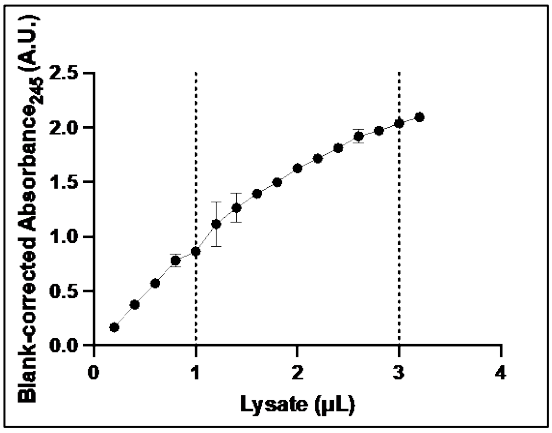


Figure 5.3. Scoping study to determine lysate volume.
Endpoint, blank-corrected absorbance measurements after 60 minutes (dots) were calculated for a range of lysate volumes; dotted lines indicate the volume range selected for the design. Experiments were conducted, in triplicate, using the Dragonfly dispenser as described in Section 2.3.2.2. Error bars represent one standard deviation from the mean (n=3).

5.2.1.3 Selecting a suitable run number

The run number refers to the total number of individual run conditions that are defined in a DoE design table. JMP automatically assigns 10 runs per factor, however, it is up to the user to determine if this provides sufficient coverage of their design space (Montgomery, 2013). For the SFD, a 138-run design was chosen (Appendix J) which expands to 414 runs during analysis to include the data from the three sets of enzyme pre-incubations. This run number was selected with a goal of sampling the 11-dimension design space outlined in Table 5.1 uniformly and, as minimally as possible. To determine an appropriate run number, two aspects were considered:

- The run number must provide sufficient coverage of the design space. JMP's Fast-Flexible Filling algorithm was used as it is the only option that accepts categorical inputs (Piepel *et al.*, 2019). This algorithm clusters run conditions into subgroups of categorical factors first, continuous factors are then distributed within these clusters. Therefore, it was critical to ensure that all subgroups were represented in the design, if the run number is too small, some subgroups would have been ignored, which would be unfavourable as

gaps would be introduced into the design space. With 138 runs, 3-4 runs are present in each of the 30 subgroups (2 salts x 5 buffer pH's x 3 temperatures).

- As the total number of runs is blocked by temperature, each block must be of an easily executable size for the plate type selected. As a 96-well plate was selected for execution (Section 4.2.5), 138 run conditions can be divided into 46 runs per temperature block. This is appropriate for the plate type selected, as a 96-well plate would accommodate 2 sets of replicates and 4 control reaction conditions.

5.2.1.4 Visualising the space-filling DoE

A scatterplot matrix Figure 5.4 is one of the easiest ways to view high-dimensional experiments as it shows the coverage of the design space with respect to the location of each run condition (Piepel *et al.*, 2019). It gives an idea of how densely the space is sampled and if the selected run number needs to be adjusted further. The figure shows how the design points from the ranges of eleven factors are distributed with respect to every other factor and provides a clear visualisation of how the chosen design space will be sampled.

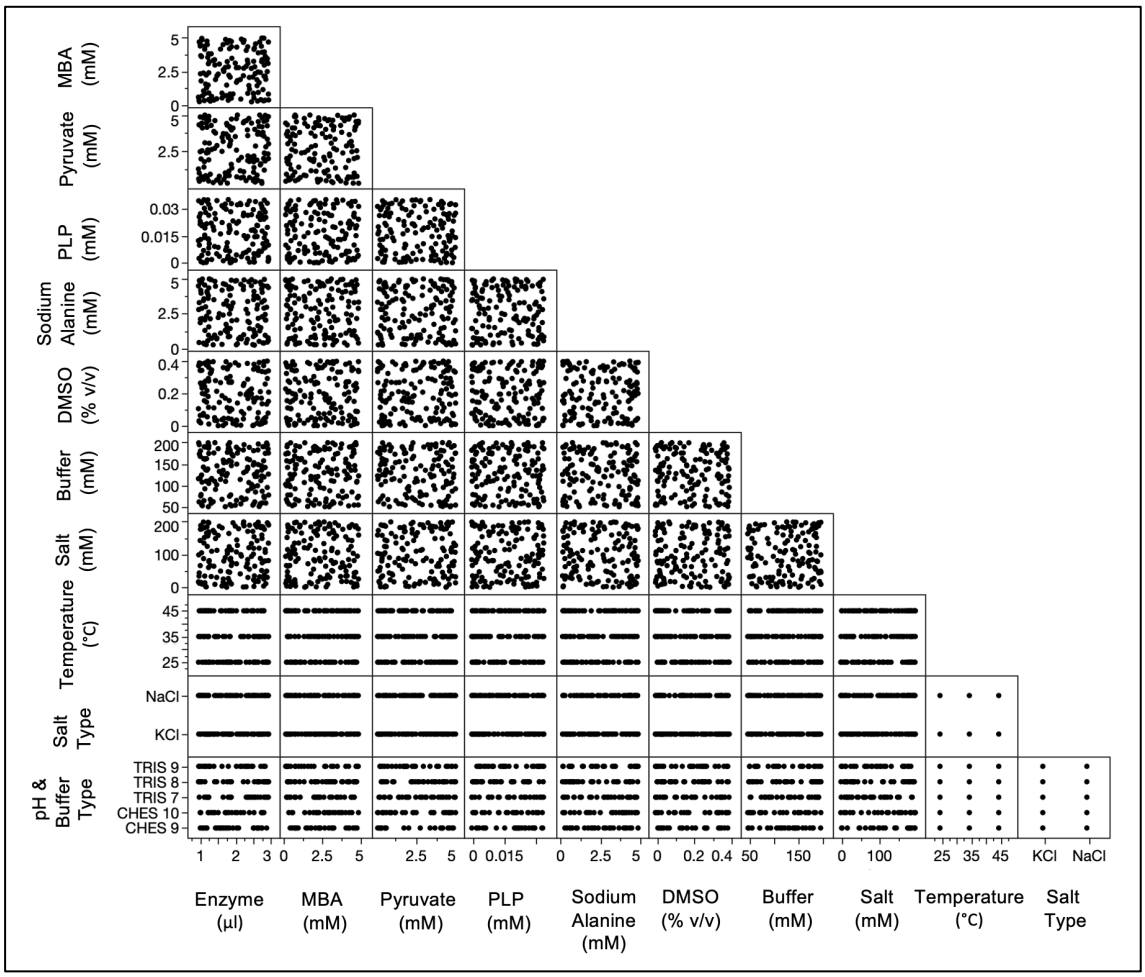


Figure 5.4. A scatterplot matrix of the 138-run space-filling DoE used to characterise a 11-dimensional space of the AP assay.

Each dot represents one run condition (n=138). Each square shows how one factor on the X-axis is related to another factor on the Y-axis for both continuous and categorical factors.

5.2.2 Experiment structure and execution details

To physically execute any DoE study successfully, the three fundamentals of DoE: randomisation, replication, and blocking, must be considered when structuring the experiment for execution (Anderson and Whitcomb, 2007; Montgomery, 2013; Antony, 2014b). Figure 5.5 provides an overview of the plate layout and experiment structure used while accounting for these fundamentals.

Randomisation was achieved by maintaining the order of run conditions as generated by the design file. This was easily done with the dragonfly dispenser as it contains 10 non-contact dispensing syringes, that can work in parallel while rapidly dispensing a range of volumes (1 µL – 4 mL) as required for each factor.

The SFD contained 432 individual runs (including controls). As this experiment aimed to characterise a large and unknown design space, each run was replicated 8 times to get an accurate understanding of noise in the event where noisier regions of the design space were encountered. This amounted to a total of 3,456 runs, which were prepared across thirty-six 96-well plates. The plate layout and experiment structure used are shown in Figure 5.5 A and C. An experiment of this scale was only possible as the dragonfly dispenser required just 6 minutes and 40 seconds to prepare each plate, totalling ~4 hours of liquid handling for >30,000 liquid handling actions.

Blocking was dictated by the 'hard-to-change' temperature factor. Three Synthace workflows, one for each temperature were created (Section 2.6.2). Each workflow was then used to generate 4 instruction files, one for each executable layer (Figure 5.5 B). Plates for each temperature were grouped for measurement.

To further streamline experimental execution, reagent dispensing was done in layers (Figure 5.5 B) as it minimised time spent changing syringes, reservoirs, and handling the dragonfly dispenser software. This, therefore, reduced the possibility of making any software errors. By doing so, only 68 syringes (34 syringes per 4 replicates) were required to execute the whole experiment.

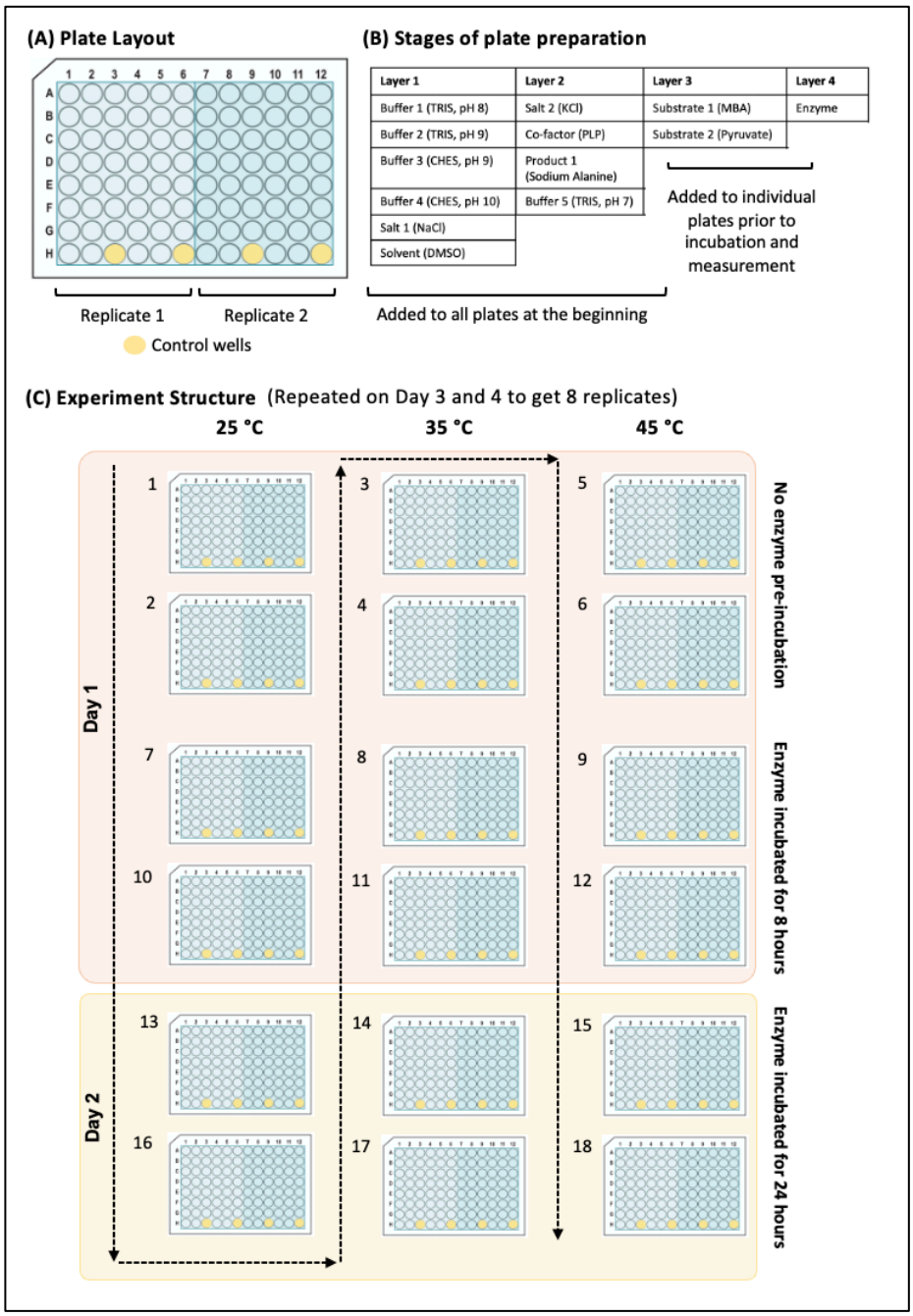


Figure 5.5. Overview of plate layout and SFD execution structure.
(A) Plate layout used for the space-filling DoE. (B) Division of reagents into layers for execution. Each plate was prepared in 4 layers each requiring a Synthace-generated instruction file. While the experiment structure is generic, reagents specific to TAm catalysed conversion of pyruvate to acetophenone are stated in brackets. Layers 1 and 2 were dispensed into all plates first. Layers 3 and 4 were added next if that plate was to be measured with no enzyme pre-incubation. For plates investigating enzyme stability, Layer 4 was added after Layer 1 and 2 and incubated for either 8 or 24 hours after which substrates from Layer 3 were added to kickstart the reaction. (C) Experiment structure followed for plate preparation and measurement. Two days were required to prepare and replicate the design four times. On Day 1, all 18 plates were prepared and plates 1-12 were measured. Plates 13-18 were measured after a 24-hour incubation on Day 2. The arrow indicates the order of plate preparation and the numbers indicate the order of measurement.

5.2.3 Data processing and visualisation of responses used for model generation

5.2.3.1 Progress curves

The 36 absorbance files generated from the SFD were processed using MATLAB scripts that group, time-correct, and blank-correct the data prior to generating responses used for model generation in JMP Pro (flowchart can be found in Section 2.7.2). Replicate time-course absorbance data were first grouped together, next, the data was time-corrected to ensure that every data point was linked into a precise time stamp. This was necessary because the absorbance files generated from the plate reader apply the same timestamp to all 96 microwells despite requiring two minutes to measure every well (detailed in Figure 2.4). This was followed by blank correction where the intercept of a spline fit of each progress curve was used as the blank value. This was necessary to calculate the absolute increase in absorbance over time which was then used to calculate the yield response.

These scripts also create graphs to visualise the raw data. Figure 5.6 is an example of this visualisation. Progress curves from 24 selected run conditions indicate the diversity of reactions captured from one SFD and the type of data that the MATLAB code can deal with.

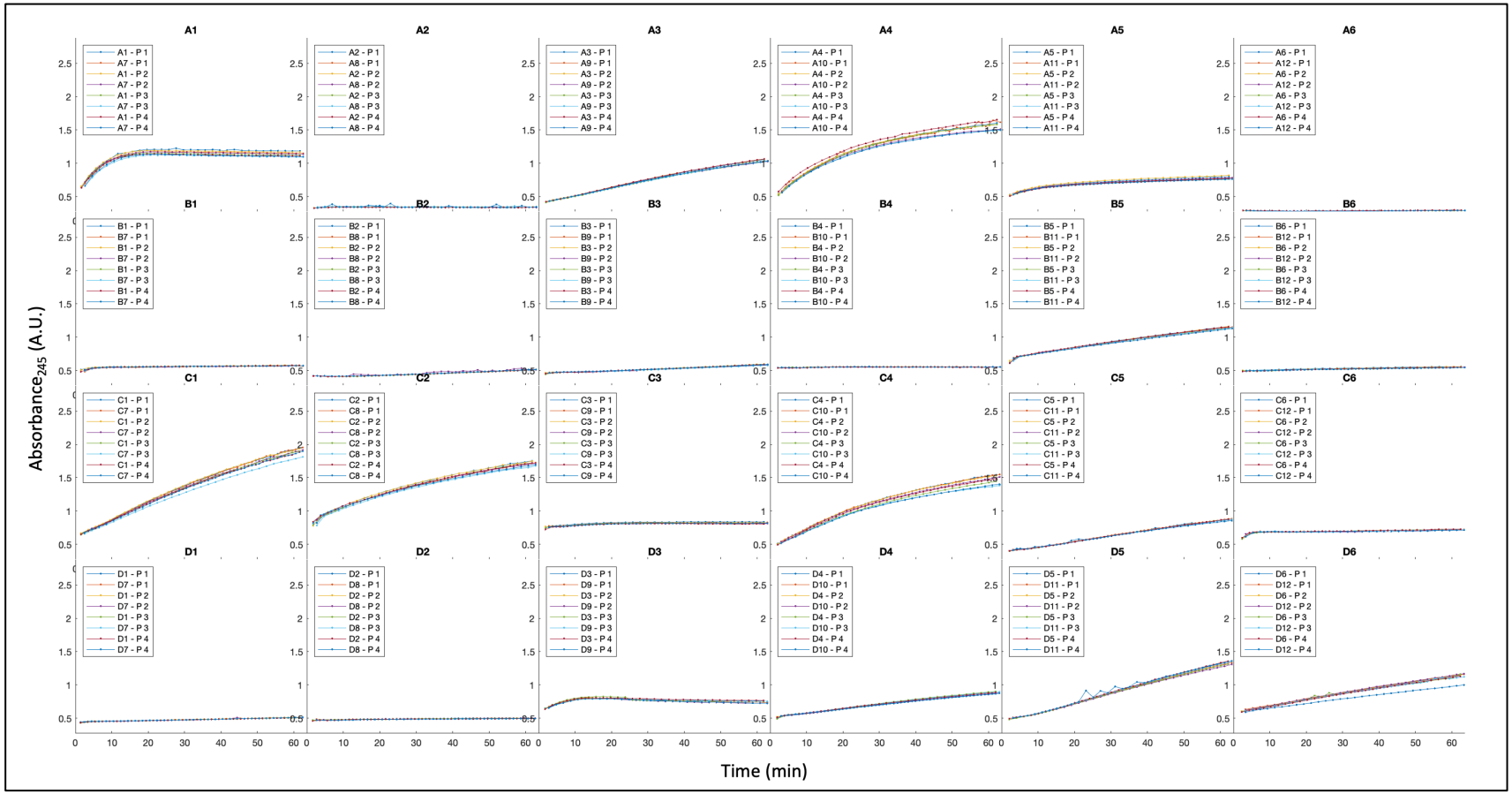


Figure 5.6. Examples of the range of progress curves generated from the space-filling DoE.
Progress curves from all eight replicates generated from 24 run conditions measured at 35 °C with no enzyme pre-incubation are shown. The well name of the first replicate is used to identify each sub-plot that contains a legend that shows the well name and plate number for each progress curve. The figure was generated using the MATLAB script described in Section 2.7.2.

5.2.3.2 Calculation of yield and initial rate responses

Yield and initial rate were the two responses calculated using the MATLAB script (detailed in Section 2.7.2), these responses were each used to generate statistical models in JMP Pro. Yield (%) was the percentage of starting substrate, pyruvate, that was converted to the product, acetophenone, after 60 minutes of bioconversion. It was calculated by converting the blank-corrected increase in absorbance into acetophenone concentration (Section 2.7.2.2). Initial rate was the immediate rate of the reaction, it was measured by calculating the slope of the progress curve from the first 5 minutes which quantifies the change in absorbance units per second (ΔA/s). By time-correcting the absorbance data prior to response generation, it was ensured that the same time period was used to generate the initial rate response.

The distribution of yield (Figure 5.7) and initial rate (Figure 5.8) responses are shown below. These values were used to generate various types of models in JMP Pro. The yield response ranged from -1.34 to 65.79 %, with a median value of 1.69%. The initial rate response ranged from -1.28×10^{-5} to 0.0006 ΔA/s, with a median value of 1.87×10^{-5}. This shows that while both responses are spread across a wide range, most of the data was centred around zero as the majority of reaction conditions were inactive or minimally active. This was not unexpected due to the wide sampling of the design space and provides useful information on optimal, sub-optimal and inactive areas of the design of design space.

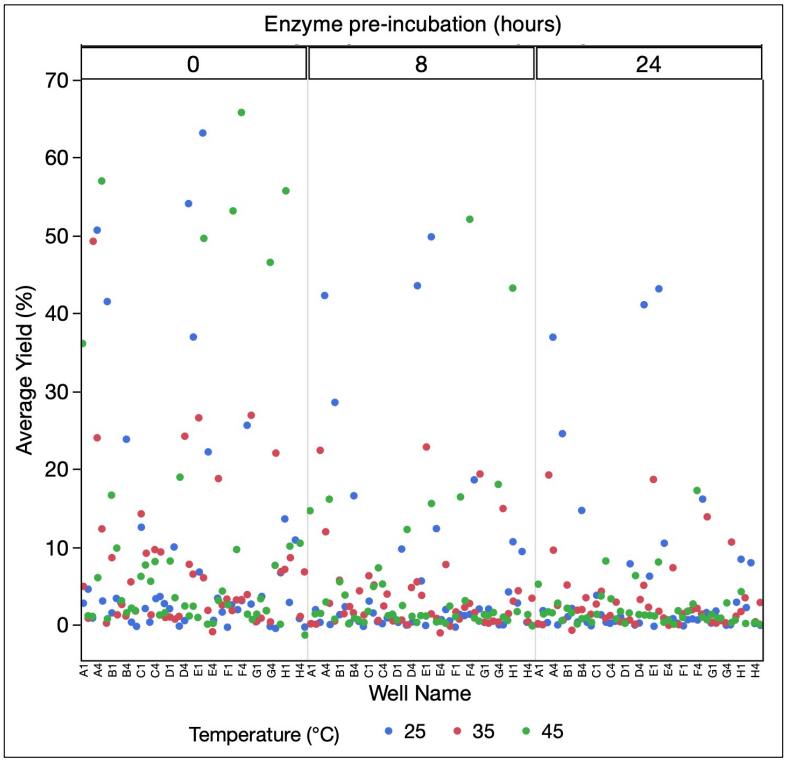


Figure 5.7. Overview of the yield response used for model generation in JMP Pro for the TAm catalysed conversion of pyruvate to acetophenone.
Each dot (n=414) represents the average yield value calculated from eight replicates.

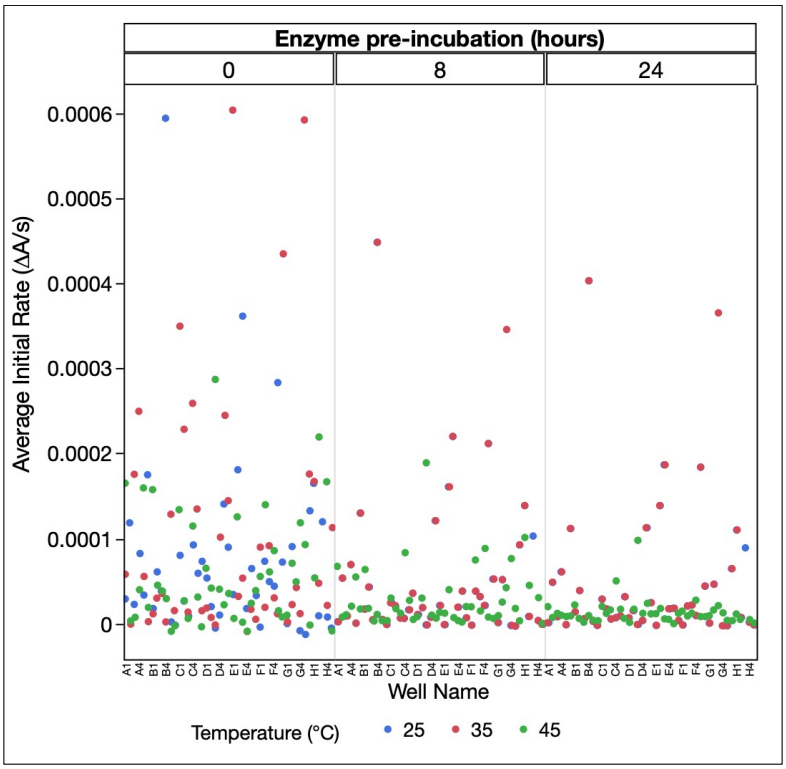


Figure 5.8. Overview of the initial rate response used for model generation in JMP Pro for the TAm catalysed conversion of pyruvate to acetophenone.
Each dot (n=414) represents the average initial rate value calculated from eight replicates.

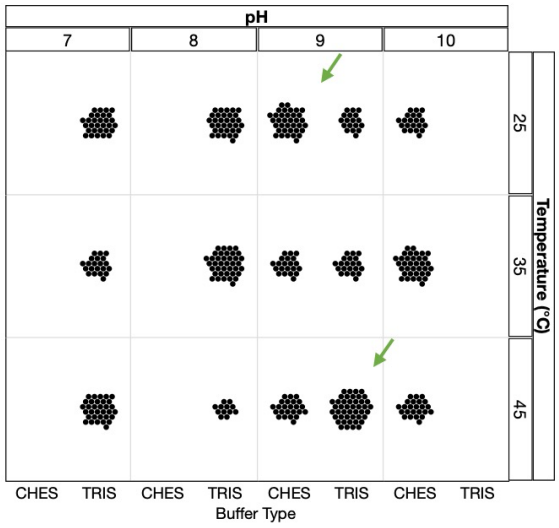
5.2.4 Excluding buffer type as a factor in model generation

In order to capture the distribution of enzyme activity over spanning a reasonably wide pH range the use of two buffer types was required. This was easily described in the format of SFD design used in this framework and would be suitable for the investigation of most classes of enzyme reactions.

Regarding analysis, buffer type can be viewed as a secondary factor as two buffer types, TRIS and CHES, were required to span the pH range of 7 to 10, where pH was the actual factor of interest. TRIS buffer has a pKa of ~8.1 while CHES Buffer has a pKa of ~9.3, this makes the former buffering agent suited to maintaining pH values ranging from 7.1 to 9.1 and the latter from 8.3 to 10.3. As buffer type was only present to control the pH, it was excluded from model generation. pH 9 is the only pH level that overlaps both buffer types, therefore, the responses generated from the subset of runs executed at pH 9 could be compared to identify if buffer type influences the responses by conducting a T-test.

However, it was not possible to investigate this as TRIS buffer is temperature-sensitive with its pH decreasing ~0.25 units with every 10°C increase (Figure 5.9 B). Therefore, the actual pH of reactions that use TRIS buffer across different temperatures is different to the originally intended pH value. Due to the temperature sensitivity of TRIS, only a small subgroup of runs executed at 25 °C and pH 9, could be compared. However, of the 54 runs present in this subgroup, 36 used CHES buffer while only 18 used TRIS buffer (Figure 5.9 A, top row). The reverse can be seen in the 45 °C subset (Figure 5.9 A, bottom row). As there are twice as many runs that use CHES buffer compared to TRIS buffer, a wider range of responses is seen with the former (Figure 5.9 C & D). Due to these reasons, no further analysis was conducted on buffer type. However, physical validation of models generated for both responses was conducted at optimal conditions at pH 9 using both buffer types.

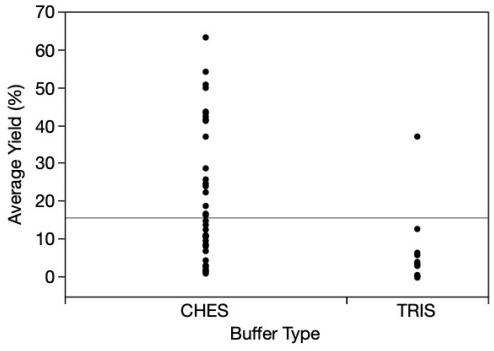
(A) Individual run condition distribution



(B) Effect of temperature on the pH of TRIS Buffer

Stock	25°C	35°C	45°C
pH 7	7.01	6.76	6.51
pH 8	7.98	7.77	7.52
pH 9	**9.01**	8.84	8.58

(C) Average Yield vs Buffer Type



(D) Average Initial Rate vs Buffer Type

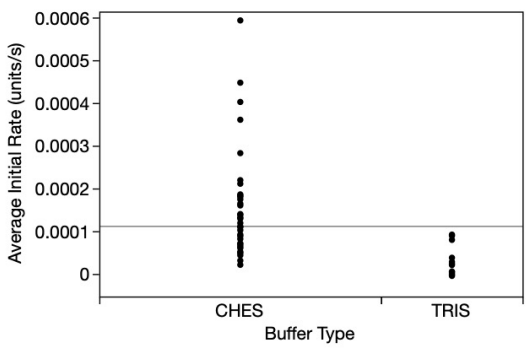


Figure 5.9. Impact of buffer type and pH on responses for the TAm catalysed conversion of pyruvate to acetophenone.

(A) The distribution of runs (N = 414) when grouped by pH, buffer type and temperature. Both TRIS and CHES buffers only overlap at pH 9, however, the distribution of runs is uneven (green arrows) when grouped by temperature. (B) Impact of temperature on TRIS buffer. (C), (D) The spread of responses for runs executed at 25°C, pH 9 with respect to buffer type. Runs that were executed using CHES buffer generated a wider range of responses, however, due to the uneven distribution seen in (A), 36 runs used CHES buffer while only 18 runs used TRIS buffer.

5.2.5 Model generation

Yield and Initial Rate responses were analysed separately. For both responses, three types of models were generated using JMP Pro. These included a Gaussian Process Model (GPM), an Artificial Neural Network (ANN) model and a Stepwise Regression Model (SRM).

GPMs, also known as 'kriging' models, are the default choice for analysing SFDs (Sacks *et al.*, 1989). They are non-parametric, while being parsimonious, meaning that they use the same number of parameters as the number of input factors, along with three other parameters: one for the mean, standard deviation, and a smoothing parameter that allows stochasticity of the response. Their flexibility via the process of spline interpolation can fit a wide variety of surfaces, making them a popular choice of model, especially when the underlying function is unknown (Rasmussen and Williams, 2006; Jones and Johnson, 2009; Erickson, Ankenman and Sanchez, 2018).

ANNs are biologically inspired models that mimic how signals are transferred in the nervous system (Kleene, 2016). Like GPMs, they are also flexible model types ideal for capturing non-linear landscapes. They require the user to select an underlying architecture that defines the layers of nodes used to process data, from a large set of options. This makes them prone to overfitting and they are typically seen as a "black box" approach as this underlying architecture can easily become overcomplicated, while making it challenging to assess the contribution of each factor on the response (Hastie, Tibshirani and Friedman, 2001; Singleton *et al.*, 2019).

SRMs are a type of linear regression model commonly used to fit responses generated from traditional DoE designs (previously demonstrated in Chapter 3), therefore, it would be interesting to evaluate their effectivity when fitting SFD data (Burke, 2018). Like ANNs, SRMs also require the user to select an underlying architecture, where every model term is specified upfront. This includes the selection of quadratic, cubic or higher order terms making the underlying models complex while remaining rigid compared to ANNs and GPMs (Rasmussen and Williams, 2006).

For all models, the average response values were used. There are two main reasons for analysing the average response values compared to treating responses generated from each replicate as an individual response. First, minimal noise was observed for both responses, making the average a suitable representative response for each run (Models analysing standard deviation for both responses can be found in Appendix K which show that greater standard deviations are unsurprisingly associated with regions of higher signal). Secondly, replicated points can cause the underlying correlation matrix generated in the GPM to become singular, which can lead to drawing less meaningful conclusions from the overall model (Jones and Johnson, 2009).

All 12 factors (Table 5.1) were used as inputs to all models generated in the following sections. A JMP Pro table was created for each response from the output files of the MATLAB data processing script (Section 2.7.2). No runs were excluded from the analysis, therefore, a total of 414 runs were analysed for each of the models described in Section 5.2.6 and 5.2.7.

5.2.6 Analysis of yield

5.2.6.1 Gaussian Process Model for yield

In this section, the procedure and outcome of fitting the yield response using a GPM is described and discussions comparing findings to literature are made in Section 5.3. First, a jackknife fit is used to get the predicted values of the model. Jackknifed values are generated by excluding each data point systematically one-by-one from the predicted model while retaining its correlation parameters (Kleijnen and Van Beers, 2004). These values were in good agreement with the actual data as most of the data points lie on the 45° diagonal line with an R^2 value of 0.875 (Figure 5.10 A). Second, the prediction profiler is used to show the effect of the modelled response on each factor. The GPM predicted a maximum yield of 79.6 ± 9.5 % (Figure 5.10 B). As the actual maximum yield generated from the SFD was 65.8 %, there is a possibility that the GPM is overfit.

According to the prediction profiler (Figure 5.10 B), and the report (Appendix L) generated from JMP Pro, pH causes the greatest variation in the response, accounting for 31% of response variation. It has a predicted optimum value of 9.1, with yields reducing to 30 % when the pH values are reduced to 8.5 or 10, this peak is seen in Figure 5.10 B.

Two factor interactions between pH and pyruvate, pH and enzyme pre-incubation, and pH and DMSO concentration, were the strongest two-factor interactions identified. However, these interactions are potentially unrealistic and are created as an artefact in the model due to the strong effect rendered by the pH factor. Model parameter estimates for main effects and two factor interactions can be found in Appendix I.

Of the other factors, pyruvate, MBA, enzyme pre-incubation, DMSO, and enzyme concentrations were ranked as having the strongest to weakest effect on yield respectively, causing a response variation of ~14 – 1.64%. Low pyruvate concentrations combined with high MBA concentrations led to a higher yield. The highest yield was obtained when the reactions were measured immediately after enzyme addition with no pre-incubation, which is expected as the enzyme has the least amount of time to degrade when biocatalysis is initiated immediately. Consequently, yield steadily decreased across 24 hours, where the enzyme was held in the reaction mixture; this can also be observed in the yield data overview in Figure 5.7.

High DMSO concentrations lead to lower yields, however, a maximum predicted yield of 56 % was predicted even with a DMSO concentration of 40 % v/v, suggesting that CV2025 is a solvent stable enzyme making it ideal for many industrial applications (Chen *et al.*, 2016). The factors that had a negligible impact on yield were salt (and salt type), buffer concentration, PLP concentration, sodium alanine concentration and temperature.

In literature, storage of CV2025 in PLP and the addition of supplemental PLP have been shown to improve enzyme stability, this was observed as according to the prediction profiler, higher PLP concentration leads to a higher yield, and without supplemental PLP, the yield reduces by ~15% (Shin *et al.*, 2003; Kaulmann *et al.*, 2007). It was expected that sodium alanine concentrations would minimally impact enzyme activity as the mechanistic model for CV2025 activity shows minimal inhibition from the product (Shin and Kim, 1998).

Generally, increasing temperature increases enzyme activity, however, in this case, 25 °C produced the maximum yields overall, with yields reducing by 20 % when the temperature was increased to 45 °C. This is contradictory to the results found in Chapter 3 which showed increased activity at 37 °C compared to 25 °C and to the paper that first described the AP assay which found an optimum of ~40 °C (Schätzle *et al.*, 2009b). One potential reason for this effect could be the distribution of underlying run conditions within each temperature block within the design.

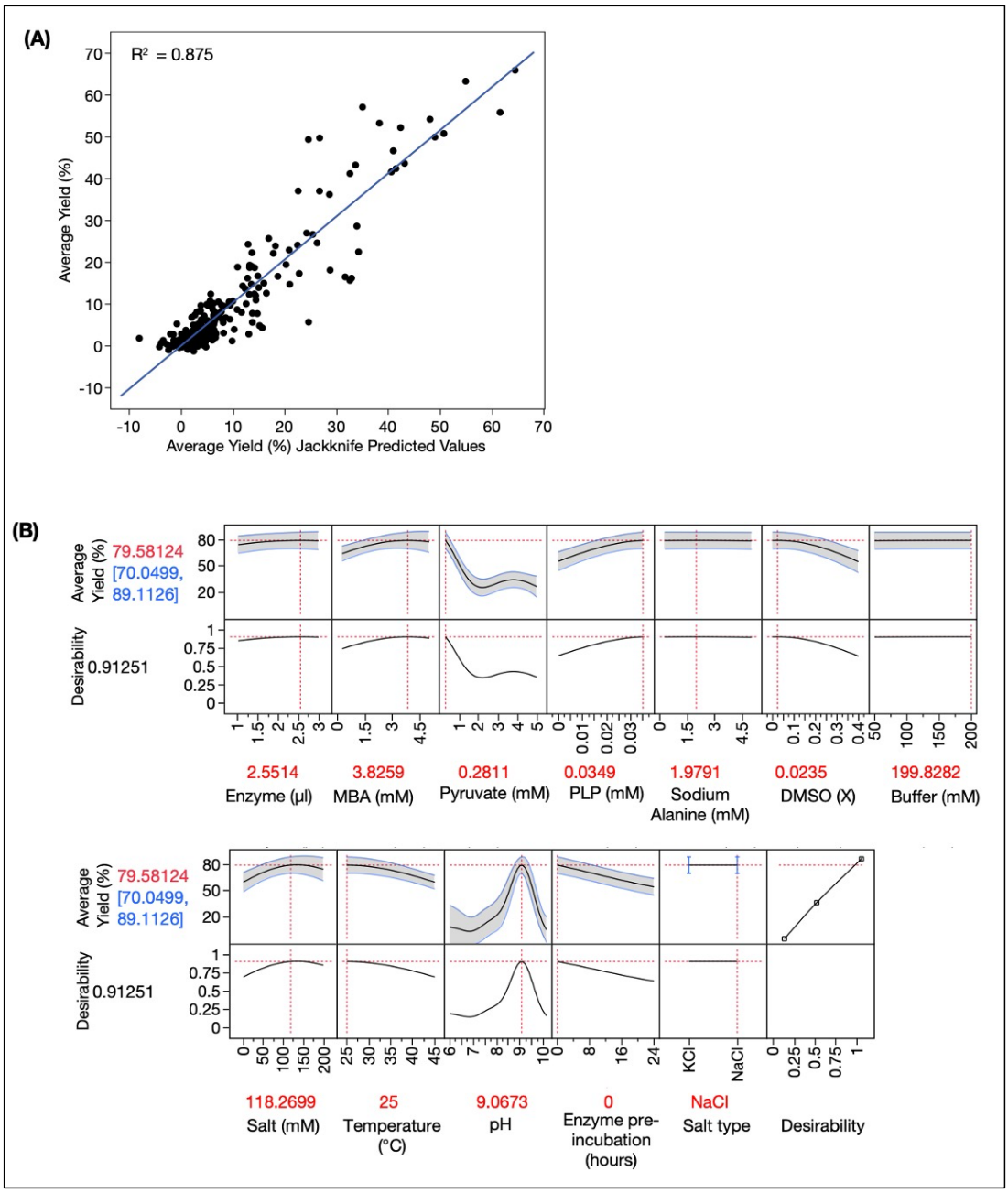


Figure 5.10. Gaussian process model predicting the yield response for the TAm catalysed conversion of pyruvate to acetophenone.

(A) Actual by Predicted Plot. Data points (black, n = 414) were fit with a GPM where predicted values are determined using a jackknife validation method. Blue line represents a line of fit. (B) Prediction Profiler is fixed at factor settings that predict maximum average yield. Factors and desirability profiles (black) with 95% confidence intervals (shaded grey) and optimal factor settings (red text and crosshairs) show the maximum average yield (red text on Y-axis) with a 95% confidence interval (blue text) predicted by the GPM. For DMSO, X represents the volume fraction (v/v). The GPM was fit as described in Section 2.6.4.

5.2.6.2 Artificial Neural Net for Yield

The ANN split the dataset into training and validation datasets, each containing 276 (66.66%) and 138 (33.33%) of the runs, respectively. The actual by predicted plots showed good agreement, with R^2 values of 0.86 and 0.61 for the training and validation datasets, respectively (Figure 5.11 A). However, the prediction profiler showed that the ANN had failed to recognise any of the patterns observed in the unanalysed data. For example, both the optimum values of pH and the effect of the no pre-enzyme incubation can be observed by simply plotting those factors against the response, however, these patterns were not recognised by the ANN, unlike the GPM. One potential reason for this could be that the splitting of the dataset into training and validation datasets could have led to gaps in the response landscape that was too challenging for the ANN architecture to capture indicating that a denser dataset would be required to use ANNs as a suitable model type. Furthermore, adding additional nodes and layers was not found to improve model fit (data not shown) and would have unnecessarily complicated the underlying model architecture.

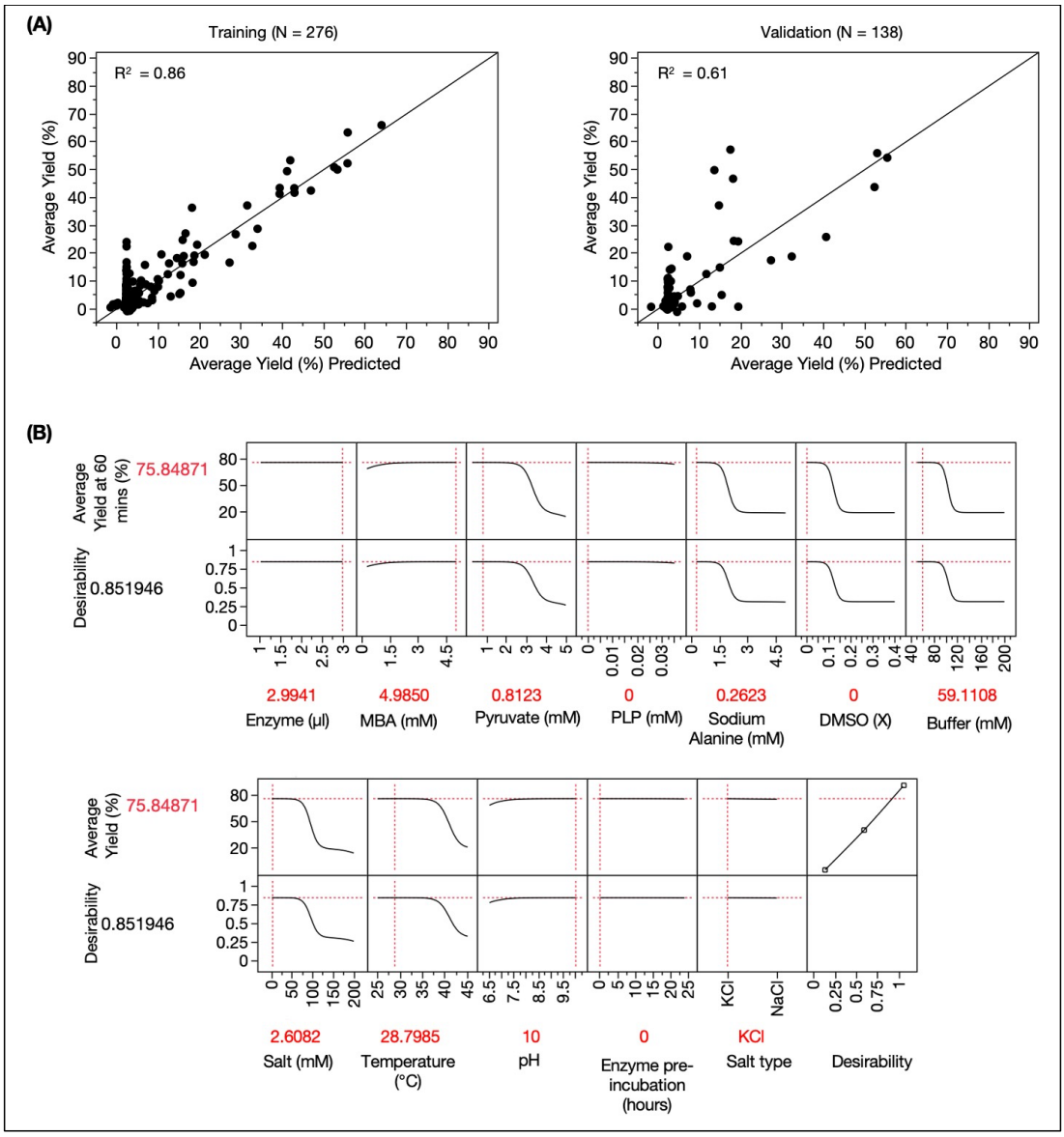


Figure 5.11. Artificial neural network model predicting the yield response for the TAm catalysed conversion of pyruvate to acetophenone.
(A) Actual by Predicted Plots. The dataset (n = 414) was divided into training (left, n = 276) and validation (right, n = 138) sets where data points (black) were fit with an ANN. Black line represents a line of fit. (B) Prediction Profiler fixed at factor settings that predict maximum average yield. Factors and desirability profiles (black) and optimal factor settings (red text and crosshairs) show the maximum average yield (red text on Y-axis) predicted by the GPM. For DMSO, X represents the volume fraction (v/v). The ANN was fit as described in Section 2.6.5.

5.2.6.3 Stepwise regression model for yield

To fit a stepwise regression model, the yield response had first to be transformed as this model type assume the response is normally distributed. Therefore, a logit transformation was applied prior to model generation (Section 2.6.3).

The predicted data showed good agreement with the actual data as the model had an R^2 value of 0.79 and a p-value of <0.0001 (Figure 5.12 A). However, an overinflated yield of ~100 ± $1.11x10^{10}$ % was predicted when factor conditions were set to maximum desirability. The underlying model was complex, with 63 model terms, which include quadratic and polynomial terms. It was also rigid as the non-linear landscape is only roughly captured. However, the factor settings and the shape of the factor profiles largely agree with those seen in the GPM in Section 5.2.6.1. For example, the overall patterns of the factor profiles from the SRM (Figure 5.12 B) for the enzyme, MBA, PLP, DMSO, salt, temperature, enzyme pre-incubation and salt type match closely with those in the GPM (Figure 5.10 B). With the SRM, increasing buffer concentration and sodium alanine concentration leads to higher yield, however, these factors have no impact on yield in the GPM. The pronounced peak of the pH factor is not seen with the SRM as that shape cannot be replicated unless higher-order terms are included in the model, however, an optimal value of 8.92 is identified, which closely matches the value from the GPM of 9.07.

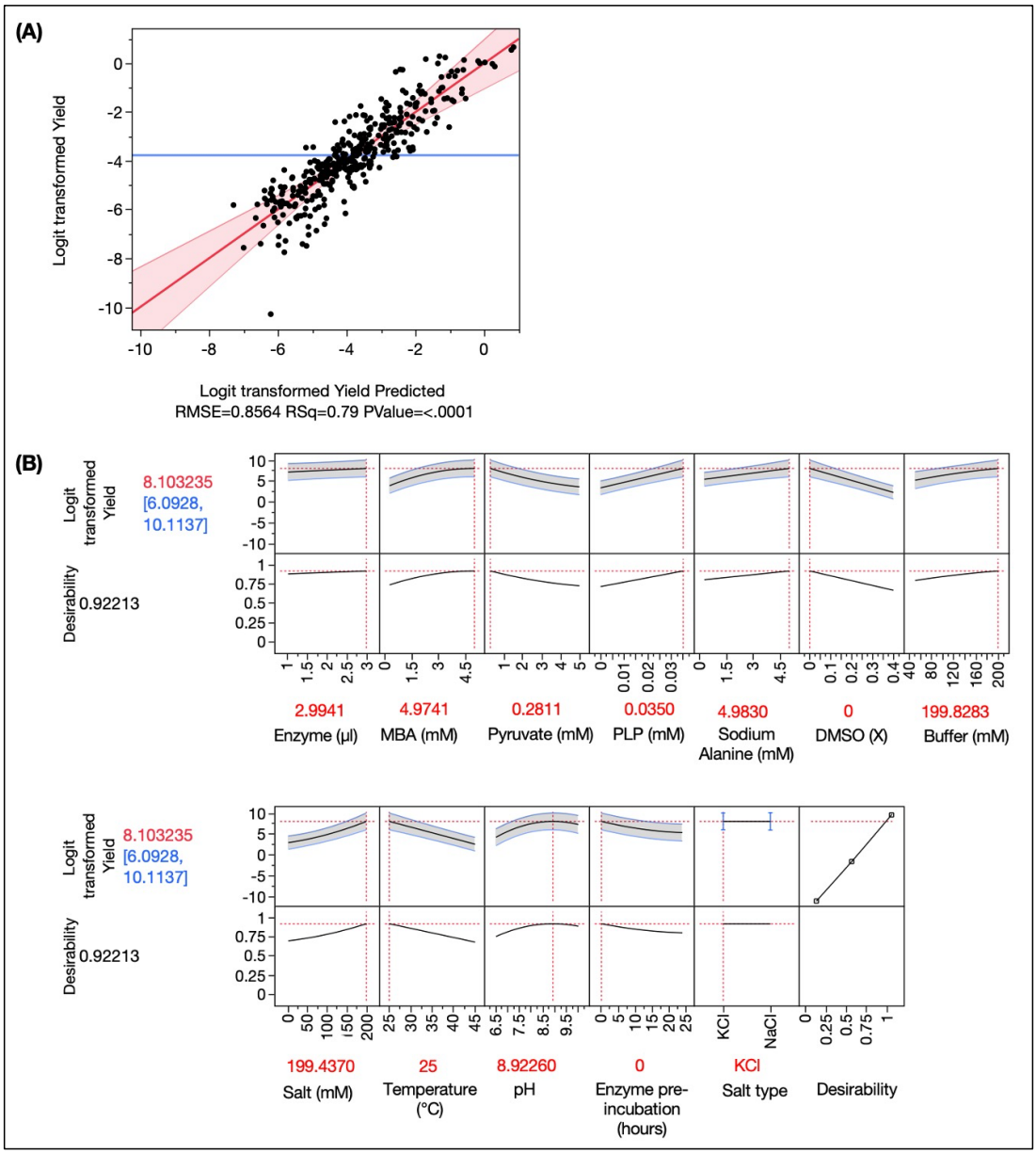


Figure 5.12. Stepwise regression model predicting the yield response TAm catalysed conversion of pyruvate to acetophenone.

(A) Actual by Predicted Plot. Logit transformed yield data points (black, n = 414) were fit with an SRM (red) with 95% confidence (shaded red) and a mean response value (blue line). (B) Prediction Profiler fixed at factor settings that predict maximum logit transformed yield. Factors and desirability profiles (black) with 95% confidence intervals (shaded grey/blue) and optimal factor settings (red text and crosshairs) show the maximum logit transformed yield (red text on Y-axis) with 95% confidence interval (blue text) predicted by the SRM. For DMSO, X represents the volume fraction (v/v). The SRM was fit as described in Section 2.6.3.

5.2.7 Analysis of Initial Rate

5.2.7.1 Gaussian Process Model for initial rate

Initial rate (ΔA/s) was the second response analysed from the SFD data. Like the yield response, models from a GPM, ANN and a SRM were compared for initial rate. Discussions comparing findings to literature are made in Section 5.3. The predicted values generated from the jackknife fit had an R^2 value of 0.567 (Figure 5.13 A). A maximum initial rate of $5.91 \times 10^4 \pm 9.1 \times 10^5$ ΔA/s was predicted, this is within the range of initial rate response generated from the SFD.

According to the prediction profiler (Figure 5.13 B) and the model report (Appendix L), the factor that lends the highest variability to the response is pH, which was also observed with the yield response, with the initial rate dropping by ~89% when the pH was reduced by 0.5 units, either side of the optimal value of 8.7. This optimum value of pH has also been identified in the literature previously in Chapter 3 (Section 3.2.6.5). The other factors that impact initial rate, from the strongest to the weakest are enzyme pre-incubation, DMSO, pyruvate, MBA, salt, and enzyme concentrations, and temperature. Initial Rate decreased by 30% when the enzyme was incubated in the reaction mixture for 8 hours and 50% when 40% v/v DMSO was used.

In general, high enzyme and substrate concentrations that are typical of enzyme catalysed reactions, when combined with no enzyme pre-incubation lead to the fastest initial rate. PLP, buffer and sodium alanine had negligible impact on the response. However, in literature PLP has been shown to further stabilise CV2025 activity over time. No significant information regarding the comparison of TRIS and CHES buffer has been noted previously and the kinetic model of TAm catalysed reactions converting pyruvate to acetophenone has shown no inhibition from sodium alanine.

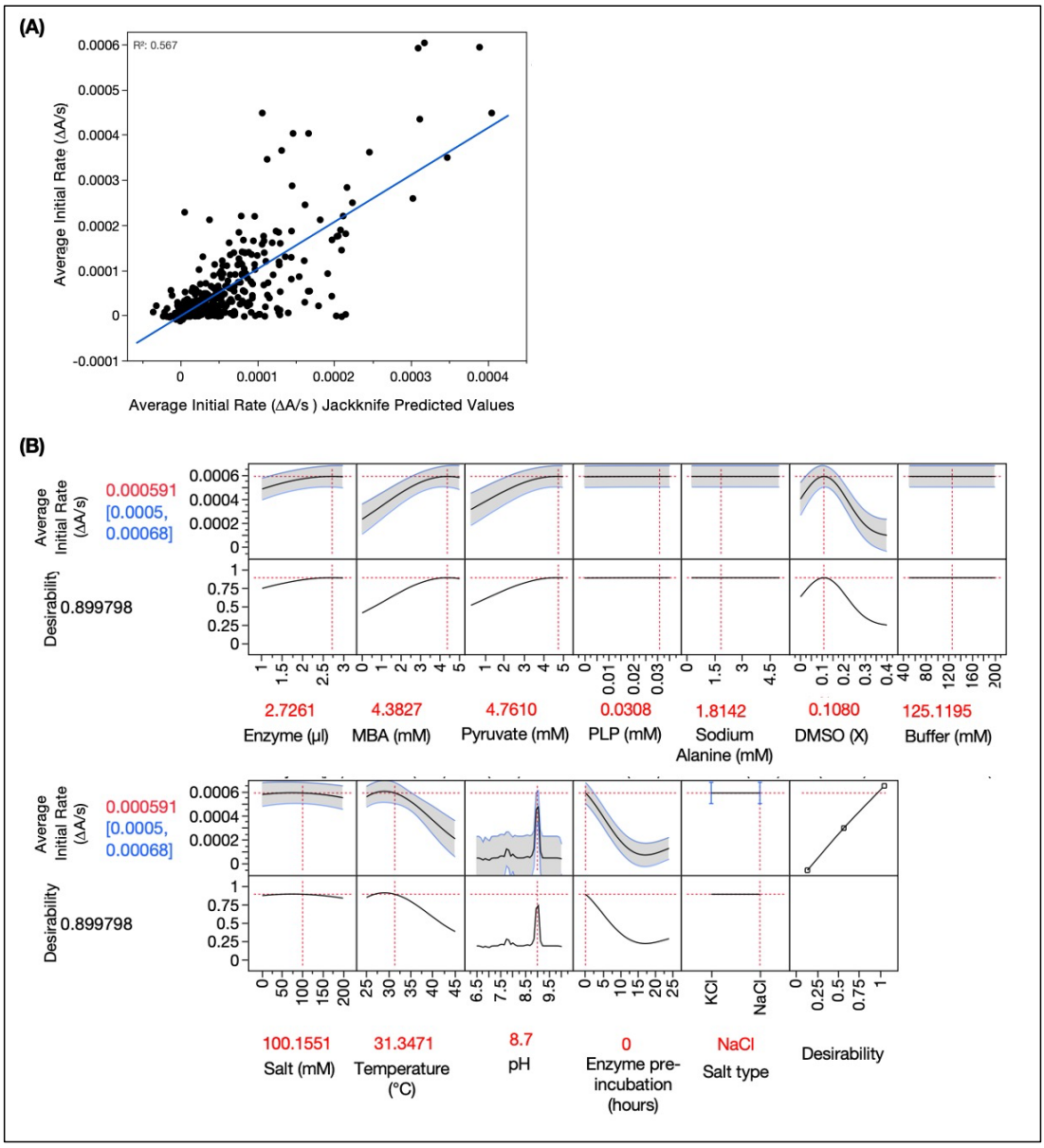


Figure 5.13. Gaussian process model predicting the initial rate response TAm catalysed conversion of pyruvate to acetophenone.

(A) Actual by Predicted Plot. Data points (black, n = 414) were fit with a GPM where predicted values are determined using a jackknife validation method. Blue line represents a line of fit. (B) Prediction Profiler fixed at factor settings that predict maximum average initial rate. Factors and desirability profiles (black) with 95% confidence intervals (shaded grey/blue) and optimal factor settings (red text and crosshairs) show the maximum average initial rate (red text on Y-axis) with a 95% confidence interval (blue text) predicted by the GPM. For DMSO, X represents the volume fraction (v/v). The GPM was fit as described in Section 2.6.4.

5.2.7.2 Artificial Neural Network model for initial rate

Data points were split into training and validation datasets, each containing 276 (66.66%) and 138 (33.33%) of the runs, respectively. The same split was used for the ANN analysing yield in Section 5.2.6.2. For both sets, the predicted values failed to predict beyond 0.0001 ΔA/s despite actual data points ranging up to 0.0006 ΔA/s. R^2 values of 0.10 and 0.12 were calculated for the training and validation datasets, respectively, indicating that the underlying model did not describe the data appropriately (Figure 5.14 A). Furthermore, the prediction profiler did not capture any patterns for individual factors, as it showed that none of the factors influenced the initial rate response (Figure 5.14 B). These issues were also encountered for the ANN describing yield in Section 5.2.6.2 and are likely caused by the same reasons detailed there.

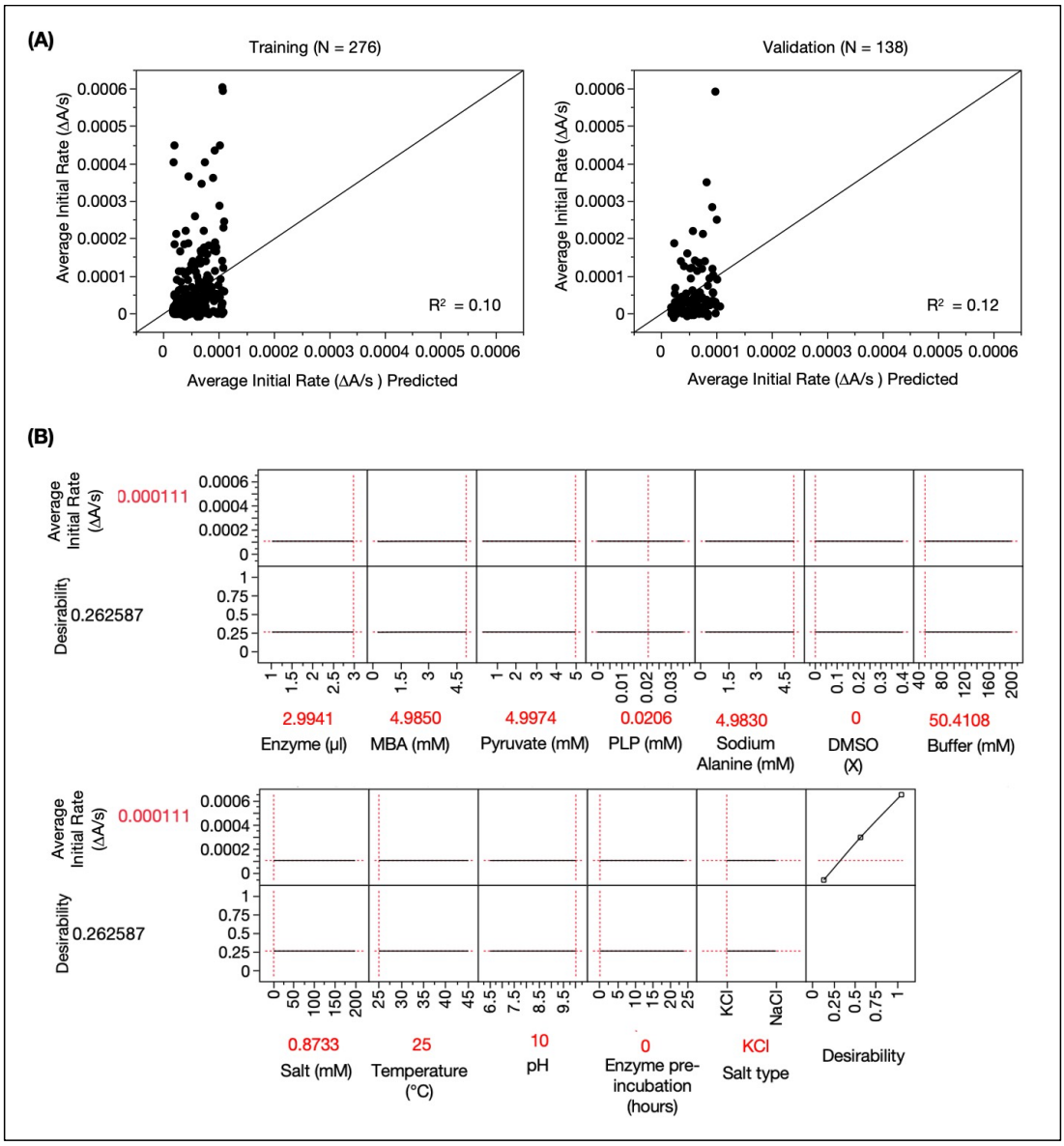


Figure 5.14. Artificial neural network model predicting the initial rate response TAm catalysed conversion of pyruvate to acetophenone.

(A) Actual by Predicted Plots. The dataset (n = 414) was divided into training (left, n = 276) and validation (right, n = 138) sets where data points (black) were fit with an ANN. Black line represents a line of fit. (B) Prediction Profiler fixed at factor settings that predict maximum average initial rate. Factors and desirability profiles (black) and optimal factor settings (red text and crosshairs) show the maximum average initial rate (red text on Y-axis) predicted by the GPM. For DMSO, X represents the volume fraction (v/v). The ANN was fit as described in Section 2.6.5.

5.2.7.3 Stepwise regression model for Initial Rate

Initial Rate values were transformed using a logit transformation to generate normally distributed values for stepwise regression as described previously (Section 2.6.3). The model includes quadratic and polynomial terms, along with all two factor interactions and required 53 model terms. The predicted values agree reasonably with the actual data as the model has an R^2 value of 0.60 and a p-value of <0.0001. However, the maximum predicted initial rates are extremely underestimated at $6.90 \times 10^{-6} \pm 6.75 \times 10^{-6}$ ΔA/s (Figure 5.15), indicating that the model is not representative of the underlying dataset.

However, this model can be broadly compared to the GPM in Figure 5.13 B, as the shape of the factor profiles, and the fixed factor settings show similarity for factors such as enzyme concentration, MBA, DMSO, temperature, pre-enzyme incubation, and pH. However, factor profiles and settings for factors such as pyruvate, PLP, sodium alanine, buffer concentration, salt concentration and type, did not show a high degree of similarity.

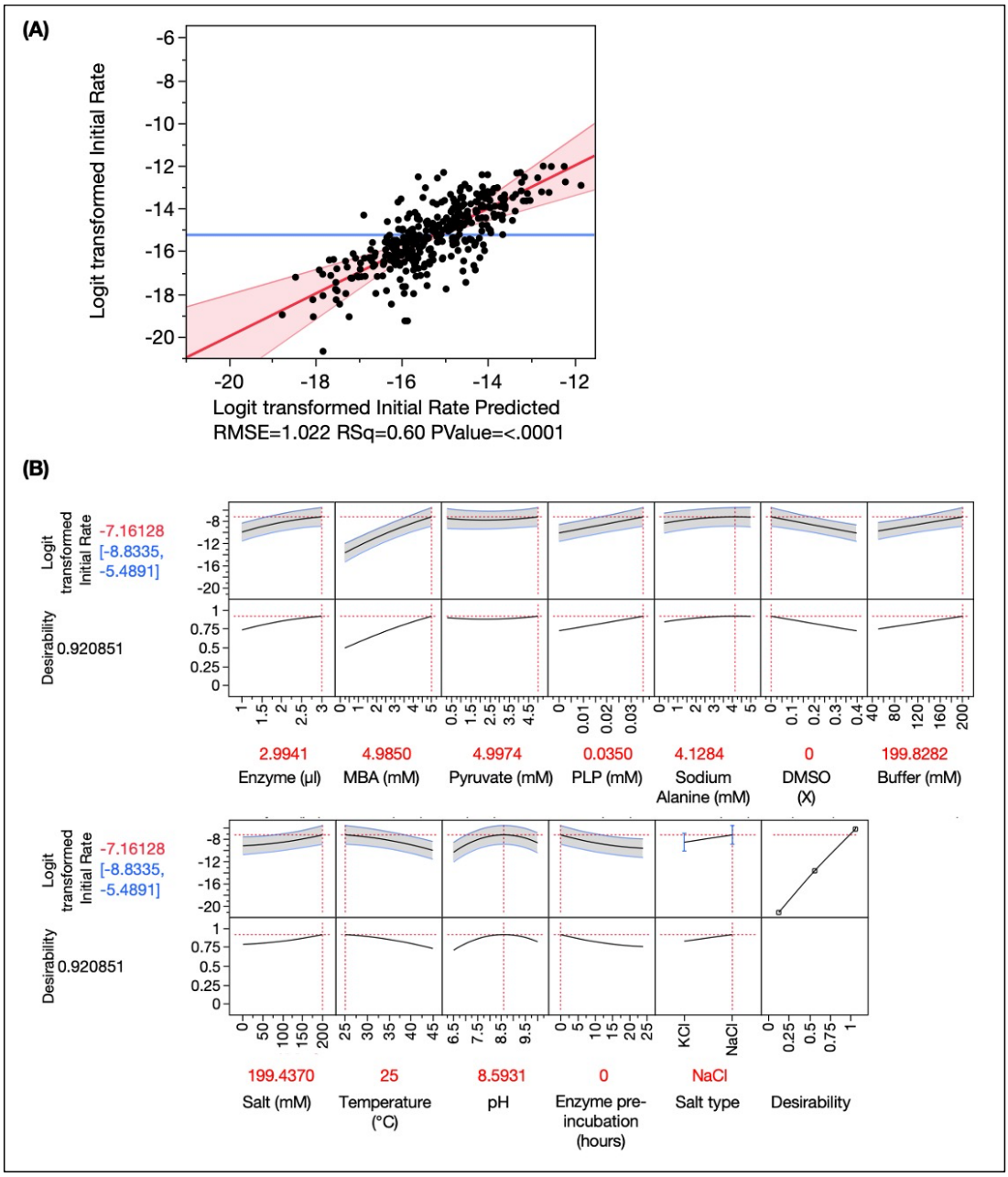


Figure 5.15. Stepwise regression model predicting the initial rate response TAm catalysed conversion of pyruvate to acetophenone.

(A) Actual by Predicted Plot. Logit transformed initial rate data points (black, n = 414) were fit with an SRM (red) with 95% confidence (shaded red) and a mean response value (blue line). (B) Prediction Profiler fixed at factor settings that predict maximum logit transformed initial rate. Factors and desirability profiles (black) with 95% confidence intervals (shaded grey/blue) and optimal factor settings (red text and crosshairs) show the maximum logit transformed initial rate (red text on Y-axis) with 95% confidence interval (blue text) predicted by the SRM. For DMSO, X represents the volume fraction (v/v). The SRM was fit as described in Section 2.6.3.

5.2.1 Comparison of main effects

The GPMs were used to compare factors that have the strongest to weakest impact on each response. Consequently, the main effect values from yield and initial rate models were compared to understand how different factors can influence each response (Figure 5.16). Here, the main effect is a parameter that indicates the total variation caused by each individual factor and is generated in the GPM model report, it can be used to understand which factors individually have the greatest impact on the response. Detailed model reports for each response can be found in Appendix L.

In Figure 5.16, the blue to red shading indicates factors that have the strongest to weakest effects on the respective response values. As described in Sections 5.2.6.1 and 5.2.7.1, pH was identified as having the strongest effect on both responses, however, this was followed by pyruvate, and MBA concentrations for yield, and in contrast, enzyme pre-incubation, and DMSO concentrations for initial rate. These differences highlight how one SFD can be used to screen and optimise for multiple responses thus enabling the identification of reaction conditions best suited to each response. For example, models generated in Sections 5.2.6.1 and 5.2.7.1 could be used to predict reaction conditions for various scenarios e.g., conditions that maximise yield at 30 % v/v DMSO.

	Main Effect	
Factor	**Yield**	**Initial Rate**
pH	0.3100	0.1104
Pyruvate (mM)	0.1384	0.0023
MBA (mM)	0.0605	0.0013
Enzyme pre-incubation (hours)	0.0585	0.0259
DMSO (% v/v)	0.0277	0.0031
Enzyme (µL)	0.0164	0.0003
Salt Concentration (mM)	0.0014	0.0003
PLP (mM)	0.0008	0.0000
Buffer Concentration (mM)	0.0004	0.0000
Sodium Alanine (mM)	0.0004	0.0000
Temperature (°C)	0.0002	0.0002

Figure 5.16. Comparing strongest to weakest main effects on yield and initial rate TAm catalysed conversion of pyruvate to acetophenone.
Blue to red colours show ranked model parameter estimates generated from respective yield and initial rate GPMs. Each column is coloured separately.

5.2.2 Physical validation of GPMs for yield and initial rate responses

One of the most important questions asked after generating models for yield and initial rate is if those models adequately represented reality (Jones and Johnson, 2009). To answer this question, specific scenarios generated from the GPMs for both yield and initial rate were isolated for physical validation.

The prediction profiler was fixed at 30 °C, a previously untested region of the design space and hence a good test of the predictive capability of the model. Different industrially important scenarios were then defined, generating corresponding factor settings. These settings were isolated from respective GPMs and converted to an input file that was parsed into a Synthace workflow. To keep the time for model validation minimal, only one 96-plate was prepared, however, it housed 3 replicates of 32 run conditions from across the whole design space and included 16 scenarios from the GPM predicting yield, and 15 scenarios from the GPM predicting initial rate and one control reaction. The scenarios selected for physical validation, along with their corresponding factor settings are found in Table 5.2.

Both responses were generated from these experimental runs of all scenarios and compared to the prediction formula generated by the models (Figure 5.17). Overall, the 31 actual responses were in good agreement with the predictions from both models. In general, yield was overpredicted by the GPM, as a cluster of 10 runs produced yield values ranging from 40 – 56 %, instead of the predicted range of 71 – 77 %. The GPM for initial rate underpredicted the actual response. A cluster of 7 runs produced initial rates over twice as much as the predicted values, ranging from 0.0013 – 0.0016 ΔA/s instead of the predicted 0.00052 – 0.00062 ΔA/s.

The actual values were compared to predicted values for each response in Table 5.3. Comparing model predictions with physically generated responses. For yield, six out of 16 scenarios had actual values within the predicted range, however, the remaining scenarios had actual values ~12 – 39 % lower than the predictions. Actual values from seven scenarios (Scenario 3,4,7,9,10,12, and 13) contained DMSO ranging from 0 - 40 % v/v were in good agreement with the predicted values, indicating that the GPM for yield is sufficiently reliable when predicting factor settings for individual factors.

Actual initial rate values were found to be greater than predicted values. Only two out of 16 scenarios had actual values within the predicted range. In contrast, nine of the scenarios had initial rates that were approximately twice as fast as predicted. These poorer predictions could be attributed to the original model being generated from a dataset that contained a mostly inactive design space and could have been improved if a validation study in a known design space was performed or if a model generated from a more focused experiment mapping the active region was conducted prior to validation.

Table 5.2. Scenarios selected for physical validation of yield and initial rate.

Each row contains the factor settings the model predicts will generate optimal response values for each scenario.

	Validation Scenarios for Yield	Enzyme (µL)	MBA (mM)	Pyruvate (mM)	PLP (mM)	Sodium Alanine (mM)	DMSO (% v/v)	Buffer Concentration (mM)	Salt Concentration (mM)	Predicted pH	Acutal pH and buffer type	Salt Type	Temperature (°C)	Enzyme pre-incubation (hours)
1	Optimal conditions	2.5253	4.2576	0.2811	0.0350	0.3348	0.0554	157.5417	100.2053	9.1158	CHES 9	NaCl	30	0
2	Optimal conditions with alternative salt	2.6677	4.1859	0.2811	0.0350	0.3057	0.0640	125.1210	117.2809	9.1140	CHES 9	KCl	30	0
3	pH 7	2.9941	4.9850	0.2811	0.0000	0.2623	0.1895	50.4108	0.8733	7.0000	TRIS 7	KCl	30	0
4	pH 8	2.9941	4.9850	0.2811	0.0000	0.2623	0.0287	125.1167	29.2071	8.0000	TRIS 8	KCl	30	0
5	pH 9, CHES Buffer	2.7693	4.1471	0.2811	0.0350	0.2623	0.0473	125.1205	119.3014	9.0000	TRIS 9	NaCl	30	0
6	pH 9, TRIS Buffer	2.7693	4.1471	0.2811	0.0350	0.2623	0.0473	125.1205	119.3014	9.0000	CHES 9	NaCl	30	0
7	pH 10	1.5488	4.9850	0.2811	0.0350	0.2623	0.2107	125.1185	42.6505	10.0000	CHES 10	KCl	30	0
8	Low DMSO concentration	2.4516	4.0056	0.2811	0.0350	0.4827	0.0000	125.5706	106.9655	9.1007	CHES 9	NaCl	30	0
9	High DMSO concentration	2.9941	4.9850	0.2811	0.0176	0.2623	0.4000	125.1195	132.7511	9.1771	CHES 9	NaCl	30	0
10	Low PLP concentration	2.7863	4.9850	0.2811	0.0000	0.2623	0.1821	199.8767	116.7782	9.1015	CHES 9	NaCl	30	0
11	High PLP concentration	2.5270	4.2759	0.2811	0.0350	0.2955	0.0580	125.1226	100.2520	9.1155	CHES 9	NaCl	30	0
12	Medium Pyruvate concentration	2.9941	4.9850	2.0000	0.0350	0.2623	0.1067	50.4108	199.4370	9.0947	CHES 9	NaCl	30	0
13	High Pyruvate concentration	2.9941	4.9850	5.0000	0.0350	0.2623	0.0000	50.4108	100.1236	9.1624	CHES 9	KCl	30	0
14	Medium MBA concentration	2.7771	2.5000	0.2811	0.0350	0.3053	0.0447	125.1203	121.4590	9.1102	CHES 9	NaCl	30	0
15	Medium Salt concentration	2.5234	4.2615	0.2811	0.0350	0.3355	0.0553	152.2113	100.0000	9.1158	CHES 9	NaCl	30	0
16	High Salt concentration	2.9941	3.7278	0.2811	0.0350	0.2623	0.0980	100.3879	199.0000	9.1070	CHES 9	NaCl	30	0

	Validation Scenarios for Initial Rate	Enzyme (µL)	MBA (mM)	Pyruvate (mM)	PLP (mM)	Sodium Alanine (mM)	DMSO (% v/v)	Buffer Concentration (mM)	Salt Concentration (mM)	Predicted pH	Acutal pH and buffer type	Salt Type	Temperature (°C)	Enzyme pre-incubation (hours)
1	Optimal conditions	2.9941	3.8414	4.9974	0.0228	1.9715	0.0639	174.0150	100.1551	9.0018	CHES 9	NaCl	30	0
2	Optimal conditions with alternative salt	2.9941	3.8408	4.9974	0.0228	1.9715	0.0639	174.0150	100.1550	9.0018	CHES 9	KCl	30	0
3	pH 7	2.9941	4.9850	0.2811	0.0000	4.9830	0.0000	50.4108	199.4370	7.0000	TRIS 7	KCl	30	0
4	pH 8	2.9941	4.7494	4.9956	0.0173	0.2623	0.0399	128.6670	38.7956	8.0000	TRIS 8	KCl	30	0
5	pH 9, CHES Buffer	2.9941	3.8393	4.9974	0.0230	1.9758	0.0634	152.3541	100.1551	9.0000	TRIS 9	NaCl	30	0
6	pH 9, TRIS Buffer	2.9941	3.8393	4.9974	0.0230	1.9758	0.0634	152.3541	100.1551	9.0000	CHES 9	NaCl	30	0
7	pH 10	2.9941	4.7490	4.9974	0.0053	0.2623	0.3992	50.4108	7.5723	10.0000	CHES 10	KCl	30	0
8	Low DMSO concentration	2.9924	3.8405	4.9974	0.0233	1.9715	0.0000	174.0150	100.1550	9.0016	CHES 9	NaCl	30	0
9	High DMSO concentration	2.9941	4.2750	4.4312	0.0162	1.4448	0.4000	199.8283	100.1551	8.9990	CHES 9	NaCl	30	0
10	Low enzyme concentration	1.0000	3.8895	4.9974	0.0285	1.2296	0.0071	125.1195	100.1551	9.0030	CHES 9	NaCl	30	0
11	Low Pyruvate concentration	2.9941	4.0822	0.3000	0.0245	1.8739	0.0690	199.8282	100.1552	9.0027	CHES 9	NaCl	30	0
12	Medium enzyme concentration	2.0000	3.8676	4.9974	0.0256	1.6226	0.0361	125.1196	100.1551	9.0025	CHES 9	NaCl	30	0
13	Low MBA concentration	2.9941	1.0000	4.9974	0.0225	3.0563	0.0714	159.6399	38.7956	9.0075	CHES 9	NaCl	30	0
14	High MBA concentration	2.9941	5.0000	4.7621	0.0221	1.8253	0.0665	125.1196	100.1551	9.0003	CHES 9	NaCl	30	0
15	No salt	2.9941	3.7496	4.9974	0.0192	2.1012	0.0605	199.8282	0.0000	9.0005	CHES 9	NaCl	30	0

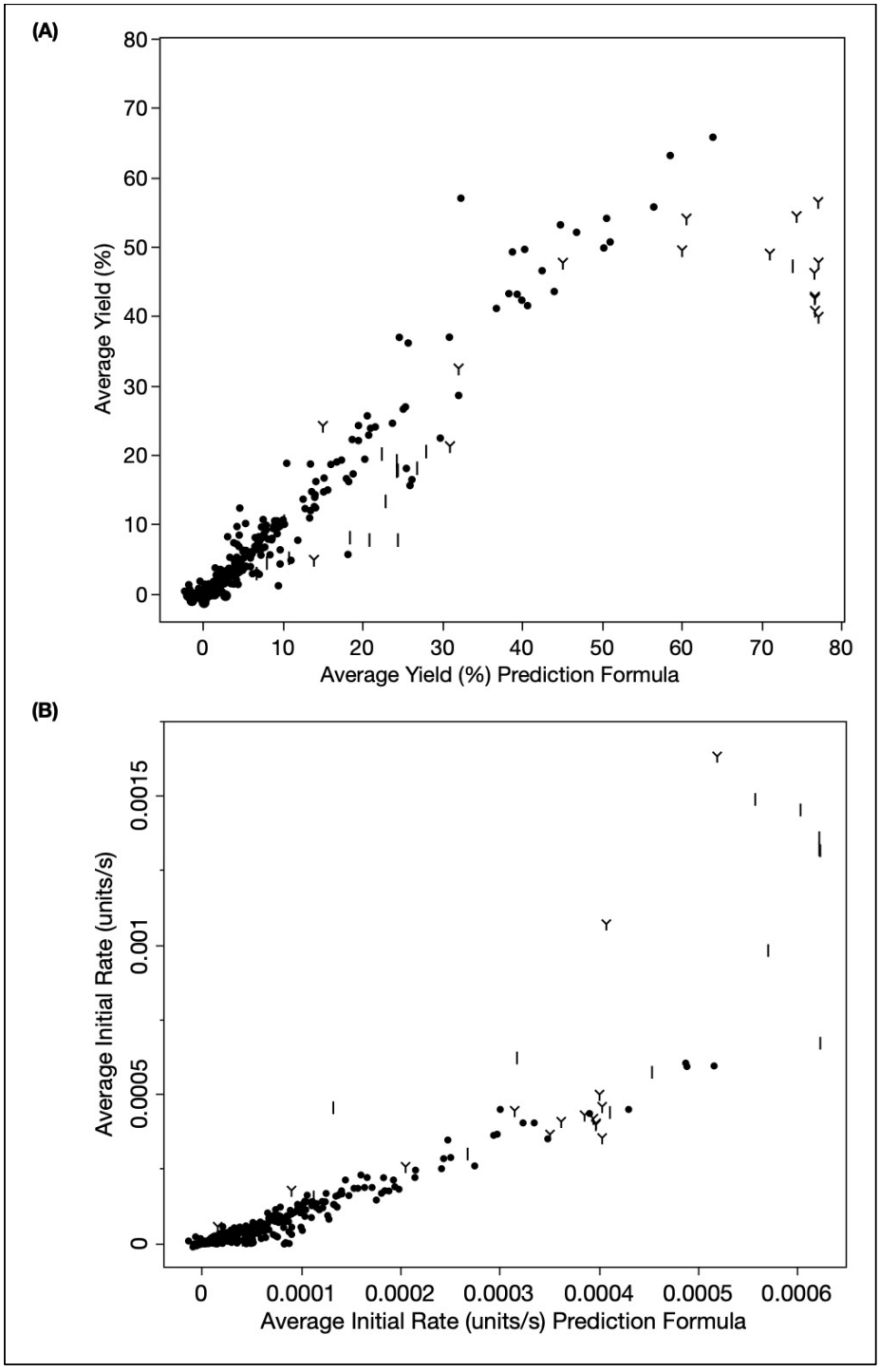


Figure 5.17. Correlation between SFD validation experiments and the original model prediction formula.

Data points from the initial space-filling DoE (black, n = 414) are consolidated with initial run validation experiments (I, n = 15) and yield validation experiments (Y, n = 16). The prediction formula generated from GMPs for average yield (A) and initial rate (B) was used to evaluate responses from the validation experiment to interpret correlation. Original DoE data was taken from GPMs generated in Section 5.2.6.1and 5.2.7.1 while validation data was obtained by applying respective prediction formulas generated from those GPMs.

Table 5.3. Comparing model predictions with physically generated responses. Response generated from the reaction conditions outlined in Table 5.2 and experimental data shown in Figure 5.17.

	Validation Scenarios for Yield	Average Yield (%) - Predicted	Prediction C.I.	Desirability	Average Yield (%) - Actual	Standard Deviation of Actual Average Yield (%)
1	Optimal conditions	80.0669	9.4799	0.88	42.7777	1.6590
2	Optimal conditions with alternative salt	80.4419	9.8375	0.88	56.4328	1.6546
3*	pH 7	15.2528	14.8707	0.25	24.1414	4.2559
4*	pH 8	47.5152	14.0810	0.56	47.6703	2.2000
5	pH 9, CHES Buffer	78.6054	9.0819	0.87	39.8680	0.7816
6	pH 9, TRIS Buffer	78.6054	9.0819	0.87	47.7044	0.9420
7*	pH 10	20.3583	14.3538	0.30	4.8226	2.2264
8	Low DMSO concentration	79.4916	9.3753	0.87	40.7408	2.4210
9*	High DMSO concentration	61.7213	11.6248	0.70	49.4695	4.4613
10	Low PLP concentration	66.6197	9.8782	0.75	54.0564	1.2191
11	High PLP concentration	79.9969	9.4317	0.88	46.2120	2.2315
12*	Medium Pyruvate concentration	36.6762	12.9550	0.45	32.4214	0.4811
13*	High Pyruvate concentration	35.0630	12.4849	0.44	21.2166	0.1793
14	Medium MBA concentration	78.1431	8.6073	0.86	54.4052	4.9145
15	Medium Salt concentration	80.0564	9.4844	0.88	42.5264	1.3612
16	High Salt concentration	70.9076	13.4931	0.79	48.9876	2.0763

	Validation Scenarios for Initial Rate	Average Initial Rate (units/s) - Predicted	Prediction C.I.	Desirability	Average Initial Rate (units/s) - Actual	Standard Deviation of Actual Average Initial Rate (units/s)
1	Optimal conditions	0.00062	1.16E-04	0.94	0.00132	7.12E-06
2	Optimal conditions with alternative salt	0.00062	1.16E-04	0.94	0.00136	1.25E-05
3	pH 7	0.00004	1.77E-04	0.18	0.00001	7.26E-06
4	pH 8	0.00011	1.53E-04	0.27	0.00015	4.88E-05
5	pH 9, CHES Buffer	0.00062	1.15E-04	0.94	0.00067	1.74E-05
6	pH 9, TRIS Buffer	0.00062	1.15E-04	0.94	0.00131	1.36E-05
7*	pH 10	0.00013	1.34E-04	0.29	0.00045	7.41E-05
8	Low DMSO concentration	0.00060	1.27E-04	0.92	0.00145	2.32E-05
9	High DMSO concentration	0.00028	1.38E-04	0.48	0.00030	1.22E-04
10*	Low enzyme concentration	0.00047	1.15E-04	0.73	0.00057	1.06E-05
11	Low Pyruvate concentration	0.00041	1.41E-04	0.65	0.00044	1.81E-05
12	Medium enzyme concentration	0.00057	1.10E-04	0.87	0.00098	2.61E-05
13	Low MBA concentration	0.00032	1.38E-04	0.52	0.00062	1.64E-05
14	High MBA concentration	0.00057	1.32E-04	0.87	0.00149	8.35E-06
15	No salt	0.00056	1.29E-04	0.86	0.00149	2.96E-05

*Scenarios where actual response values are within range of predicted values.
For each response, high to low response values are formatted using a blue to red colour scale.

5.3 Discussion

The main aim of this chapter was to develop a generalised end-to-end automated framework for rapid characterisation and analysis of a high-dimensional design space. To exemplify this framework, a proof-of-concept experiment that applied a space-filling DoE design was used to explore and optimise the AP assay (Figure 3.1).

Historically, SFDs have almost exclusively been used for deterministic computer experiments despite being well suited to biological experimentation due to their capability to explore complex, non-linear landscapes (Montgomery, 2013; Wu, 2015; Joseph, 2016). Much of the lack of SFD use can be attributed to the high execution effort and cost, which makes them near impossible to execute physically (Pronzato and Müller, 2012).

In biological research, there is just one known study that employed a 20-run SFD to iteratively (100 runs in total) characterise the design space of three factors impacting the activity, and stability of the enzyme, pyruvate kinase. To do so, researchers developed a "robot lab assistant" using the R programming language to automatically translate experiment instructions into commands for a Tecan Genesis RSP 159 liquid handler for reaction preparation. Additional software was developed to communicate between the liquid handler and the Tecan Ultra II plate reader, where time course absorbance measurements were collected and processed (Bonowski *et al.*, 2010). This study exemplifies the significant effort required to utilise specific automated devices for complex biological experimentation. In contrast, the work done in this chapter shows that straight-forward and efficient frameworks (Figure 5.2**)** can be developed using Synthace to execute SFD studies that necessitate complex liquid handling. No bespoke programming was required to generate device instructions, furthermore, due to Synthace's device-agnostic capabilities, liquid-handling instructions could be generated for a collection of automated devices with ease. Thus, the framework developed in this chapter makes SFD studies more accessible to biological research than ever before.

SFDs could be more advantageous than executing traditional DoE designs. For example, one major disadvantage to executing traditional DoE studies (such as the examples described in Chapter 3) is that significant time can be spent identifying suitable levels to investigate factors of interest which are only avoided if context-specific *a priori* process knowledge is available to researchers. Typically, OFAT experiments are executed to identify these levels to ensure that the appropriate design space is captured when sampled at those points (Coleman and Montgomery, 1993; Whitehead *et al.*, 2018). Often selecting factor levels is a major cause for concern as the risk of selecting an inappropriate factor range can lead to unusable data as the measured response can be either too small to detect or too large that it will "fall off a cliff" (Coleman and Montgomery, 1996; Onyeogaziri and Papaneophytou, 2019). Consequently, this trial-and-error process can take months and can lengthen the already challenging process of executing automated DoE (Section 1.2.4). However, in this chapter, OFAT experimentation was almost completely avoided except for one experiment that was required to confirm suitable lysate quantities (Figure 5.3). As the SFD sampled across wide factor ranges, no other scoping studies were required prior to executing the 3,456 run SFD which sampled a large 12-dimensional design space (Table 5.1).

When performed successfully, this SFD could be physically executed in just four days, thus demonstrating that the framework developed in this chapter could be used to characterise unknown design spaces rapidly. However, it should be noted that skipping the execution of scoping studies may not be possible if precious reagents are required, as they may be wasted. Furthermore, eliminating scoping studies could in turn lead to characterising inactive design spaces. This was observed in the SFD conducted in this chapter as most of the experiment (75%) explored in the 12-dimensional space was comprised of inactive regions (Figure 5.7 and Figure 5.8). However, this was not considered a disadvantage as it still provided knowledge about the AP assay and sufficient information was present to generate models for yield and initial rate. Furthermore, having knowledge of areas of the design space that are inactive is critical when selecting appropriate assay conditions that are suitable and

precise for many areas where biocatalytic reaction screening is used e.g., in therapeutic drug discovery.

The 414-run SFD study was replicated 8 times, yielding 3,456 run conditions. Here, the aim was to execute a design with minimal coverage of the design space, to avoid executing excess runs, while gathering precise information on the noise of the design space by replicating run conditions eight times. In this case, generating precise measurements of the noise of the design space (i.e., the standard deviation of eight replicate runs) was prioritised over increasing the total run number and reducing the number of replicates thus creating a trade-off between denser sampling of the design space and replication.

In reality, the standard deviations calculated were small (Appendix K). In retrospect, these results could have been generated with a 2-day experiment that included only 4 replicates, halving physical execution time, and reducing labware and reducing costs by half. Alternatively, the run number could have been increased to generate a more information-rich dataset which would have been beneficial when fitting models such as ANNs that failed to capture expected factor profiles when the data set needed to be split into subsets of training and validation sets which created pockets of missing data within the design space (Section 5.2.6.2 and Section 5.5.2.7.2). If the design was executed with a higher number of runs, the design space could have been mapped more clearly while enabling the use of alternative model types. Therefore, in the future, increasing the run number would be beneficial.

Generating precise response data for each of the 3,456 run conditions was done automatically using MATLAB scripts as described in Section 2.7.2. These used time-course absorbance files as inputs and generated yield and initial rate of acetophenone production as outputs. Developing these scripts allowed sophisticated data processing automatically and was critical to the framework developed in this chapter. For example, absorbance data from replicate run conditions spread over different microplates (36 in total) were grouped, time-corrected, and blank-corrected automatically prior to any response calculation (Section 2.7.2.1). Typically, executing

any of these data processing manually using tools such as Excel would be completely unfeasible due to the large quantity of data generated across 36 files. However, developing and editing these scripts required programming skills, therefore, it may cause a bottleneck for the generalised implementation of this workflow. However, with basic programming skills, the existing MATLAB scripts could be adapted for 384-well plates, calculate more responses and so on making the existing scripts more generalisable.

Three types of statistical models were generated to analyse the SFD data. Gaussian Process Models (GPMs) were most suited to capturing the non-linear response landscapes (Figure 5.10 and Figure 5.13). This was expected as they are the default model choice for analysing SFD (Rasmussen & Williams, 2006). ANNs, despite potentially being a flexible model type, failed to capture both yield and initial rate responses. This was possibly due to gaps of information created when the dataset is split into the training and validation subsets (Figure 5.11 and Figure 5.14). The run number selected for the SFD study ensured minimal coverage of the design space (Section 5.2.1.3), which would in turn require the dataset to be analysed as a whole, therefore splitting the dataset into validation and training subsets would introduce gaps in the design space making the ANN model challenging to capture the response surface appropriately. For future applications, a larger run number, in turn creating a denser and more data-rich design, may be better suited for ANNs. Finally, the stepwise Regression Models (SRMs) required many model terms (63 and 53 for yield and initial rate, respectively) but were incapable of capturing the non-linearity of some factor profiles as the underlying architecture only contained flexibility rendered from a two-parameter fit (Section 2.6.3). However, the SRM models (Figure 5.12 and Figure 5.15) matched most patterns found by the GPMs (Figure 5.10 and Figure 5.13). This further increased the trust in the conclusions drawn from the GPMs.

As SFDs allow fine-grain mapping of the design space, a wealth of information that characterised the AP assay and CV2025 TAm activity was generated with just one study. For example, pH 9 was found to be the optimal pH value for maximising both

yield and initial rate responses. This result agrees with the literature (Schätzle *et al.*, 2009; Schell, Wohlgemuth and Ward, 2009) and the findings of Chapter 3 (Figure 3.13) which also identified pH 9 as the optimum and further expands on those findings by confirming the peak at pH 9 due to sampling across the range of 7 – 10 units. However, these findings continue to be slightly contradictory to the literature as AP assays are routinely executed at pHs of 7 – 8 (Rios-Solis *et al.*, 2013; Gruber *et al.*, 2017). One potential reason for activity at low pH values could be the use of excess enzyme or substrates leading to some enzyme activity.

Buffer concentrations, ranging from 50-200 mM, had little to no impact on either response, this is a beneficial finding as buffer concentrations up to 200 mM are routinely cited in the literature which can be significantly reduced, thus minimising reagent costs (Yun, Cho and Kim, 2004; Rios-Solis *et al.*, 2015). Buffer type was excluded from analysis as described in Section 5.2.4, additionally, insufficient data was present to compare any differences between TRIS and CHES buffer at the overlapping pH setting. However, the impact of buffer type on TAm activity has been previously reported, especially between buffers such as HEPES and Phosphate buffer, therefore, in retrospect, a larger number of commonly used buffer types could have been included in the original design in an in-depth characterisation of buffer types across multiple pH values (Chen et al., 2018; Kaulmann et al., 2007).

As CV2025 TAm is PLP dependent, its impact on TAm activity and stability has been an important point of interest. The SFD study found that both yield and initial rate responses were minimally impacted by PLP concentration when bioconversions were started with no enzyme pre-incubation (Figure 5.10 and Figure 5.13). However, after 24 hours of enzyme pre-incubation, both yield and initial rate responses were maximised when incubated with a high PLP concentration, indicating that the presence of PLP helped maintain CV2025 TAm activity (data not shown, captured when pre-enzyme incubation was fixed at 24 hours using the prediction profilers shown in Figure 5.10 and Figure 5.13). This trend has been previously reported when CV2025 TAm has been incubated in HEPES and Phosphate buffer (Kaulmann *et al.*,

2007; Chen, Berglund and Humble, 2018). PLP has also been previously reported to have an inhibitory effect over time, this was not observed during the SFD study, potentially as low PLP concentrations (maximum of 0.035 mM) were used unlike in literature where inhibitory effects were observed when concentrations of over 1 mM PLP were used to enhance enzyme stability (Shin *et al.*, 2003; Yun *et al.*, 2004).

High concentrations of MBA were linked to maximising both yield and initial rate responses over 24 hours (Figure 5.10 and Figure 5.13) and also align with findings from Chapter 3 (Section 3.2.6.5). MBA concentrations of 5 mM have previously been reported to cause substrate inhibition in TAms from other organisms, however, this was not observed during the SFD study (Schell, Wohlgemuth and Ward, 2009), additionally, MBA toxicity has also been previously reported, however, this was also not observed in the SFD study but could potentially be observed at higher concentrations of MBA (Rios-Solis *et al.*, 2015).

The solvent stability of CV2025 TAm was investigated by preparing reactions that contained between 0 to 40 % v/v DMSO (Figure 5.10 and Figure 5.13). While both yield and initial rate responses generate the most desirable results in the absence of DMSO, yield and initial rate are only reduced by 30 and 50% respectively when subject to 40 % v/v DMSO. This was later confirmed by physical validation and agrees with the literature where CV2025 was found to be stable in 50 % v/v (Chen *et al.*, 2016). The results of the SFD also match the findings in Chapter 3 (Section 3.2.5.4) that show an increase in activity at 10 % v/v DMSO potentially due to the increased MBA solubility it renders and can be seen in both prediction profilers in Figure 5.10 and Figure 5.13**.**

A major advantage of executing one large space-filling DoE was the sampling of wide factor ranges which in turn led identification of optimal areas of the design space while enabling the visualisation of a high-dimensional design space, additionally, this knowledge can be used as a foundation for a wide-range of more focused iterative experiments e.g., iterative experiments focusing on optimising DMSO stability,

maximising initial rate etc. This can be advantageous over executing smaller iterative SFD studies or traditional DoE designs to arrive at the same findings as one large SFD generates a wealth of information immediately, while iteration campaigns consisting of smaller experiments can be expensive as a lot of resources can be required to design, execute, and analyse each iterative study. However, it is important to note some of the disadvantages of SFD, for example, there is a possibility that no active regions of the design space are identified, in which case, more focused scoping experiments may be required to identify alternative factor ranges. Additionally, while SFDs provide information on factor profiles and allow the visualisation of high-dimensional spaces, they struggle to identify two- or three- factor interactions as easily as traditional DoE designs, therefore, if such interactions are expected, a traditional iterative DoE campaign or a DoE campaign that combines small SFD studies with more traditional DoE studies may be better suited than investing experimental bandwidth on conducting a large SFD study.

Physical validation found that optimal yield was overestimated, while the optimal initial rate was underestimated. In general, both GPMs fail to capture the extrema of the response values accurately but provided reasonably good predictions overall (Figure 5.17 & Table 5.3). These experiments can be integrated into the original design to generate more accurate models. Retrospectively, conducting an iterative validation study that investigates a previously known design space before exploring an unknown design space would be preferred as it would confirm the accuracy of the models being validated first as it would be executed in the optimal known area of the design space.

In conclusion, one SFD study was found to be sufficient to characterise a 12-dimensional space on the AP assay and provided a wealth of information of the AP assay and CV2025 TAm activity as many trends that were found in existing literature were confirmed. The framework and data processing tools developed in this chapter can be adopted by biocatalysis and assay development groups to characterise and

even optimise the design space of enzymatic assays in a matter of weeks. In the following chapter, the industrial impact of using SFDs is considered.

6 Key considerations for the industrial implementation of space-filling designs[†]

6.1 Introduction

The work described in this thesis has led to the creation of a framework for executing and analysing space-filling designs (Figure 5.2). This culminated in a proof-of-concept experiment where it was shown that it was possible for an unfamiliar, high-dimensional design space could be explored rapidly to generate models for important bioprocessing responses, such as reaction yield and initial rate (Section 5.2.3.2). This could then be physically validated by a further round of experimentation where it was shown that the created GPMs could provide sufficiently accurate predictions of bioconversion performance for various experimental conditions not specifically studied in the initial SFD. This framework makes space-filling designs accessible for physical execution for the first time and can greatly benefit research groups that are focused on optimising biocatalytic reactions and developing assays by reducing overall time and cost while generating information-rich datasets.

6.1.1 Aim and objectives

The aim of this chapter is to highlight the key requirements and considerations for the successful implementation of this framework within an industrial context. The specific objectives are to outline these considerations in three broad categories:

1. DoE software considerations (Section 6.2).
2. Synthace and automation platform considerations (Section 6.3).
3. Data-processing considerations (Section 6.4).

[†] This chapter is included as part of the requirement for the award of a UCL EngD in Bioprocess Engineering and Bioprocess Leadership.

6.2 DoE software considerations

Throughout this thesis, JMP Pro, a commercial third-party software, for design and model building was used (Section 2.6.1). It was particularly useful for the SFD study as it accommodated many considerations that were made for SFD design creation, visualisation, and analysis, and provided many important benefits. These included the following:

- Design generation, visualisation, and analysis were all conducted within one software package. For example, the SFD tool can be used to generate a design, and once it is created, it can be visualised with functions such as the Scatterplot Matrix (Figure 5.4). Once the design is executed, a mathematical model can be generated using a variety of approaches (exemplified in Section 5.2.6).
- It has seven space-filling design options, including one option called Fast Flexible Filling (FFF) which accommodates categorical factors (Lekivetz and Jones, 2015). Categorical factors such as buffer type, salt type or device type are often investigated in biological research as they are common, therefore, SFD options that incorporate categorical inputs are critical to suitable design generation.
- The DoE platform is flexible as it enables the user to specify run numbers and specify disallowed combinations, and linear constraints. For example, a disallowed combination was defined to split pH values according to buffering agent types in Chapter 5 (Section 2.6.1), also a very common trait in biological experiments.
- It provides access to significant documentation, technical support, training, moderated user communities, and books, which are all invaluable resources for new users of SFDs (https://www.jmp.com/en_us/support/jmp-documentation.html).
- The DoE platform is well established and has been used in many industries for over 40 years.

The benefits and considerations listed above enable researchers that are new to space-filling designs to proceed to physical experimentation quickly. However, it may be undesirable for groups that are financially constrained as one license has an annual cost for commercial use in industry of £1,550 and £12,520 for JMP® and JMP Pro®, respectively (https://www.jmp.com/en_us/software/buy-jmp.html). Alternatively, another commercial software package, Stat-Ease® 360 (the Pro version of Design Expert®, Stat-Ease, Inc. costing $1,295/user) can also be used to create SFDs with categorical inputs as these capabilities were added in October 2021.

Free, open-source programming languages such as Python and R can also be used to develop programs that can design and analyse SFDs. Often these programs are bespoke, however, both Python and R contain pre-existing packages that can be modified or used for implementing DoE designs and analysis (Erickson, Ankenman and Sanchez, 2018).

It is important to note that the performance of GPMs varies between different packages. Previously, GPMs in R, JMP, Python, and MATLAB (note that MATLAB is also a commercial package) has been compared to highlight that significant differences can emerge from different packages, thus urging the user to be aware of their options when choosing modelling software (Schulz, Speekenbrink and Krause, 2018).

At present, SFDs are rarely used in biological research (Section 5.1), however, one research group has also developed a framework for computer-controlled experiments to study enzymatic assays. Bespoke R scripts were developed to control an automated device and plate reader, design five iterations of a 20-run space-filling DoE that contained 3 factors, and analysed the response by using a previously available Gaussian Random Process Regression function (Bonowski *et al.*, 2010). This work exemplifies that significant specialist knowledge is required to develop the programming architecture required for SFD design generation and analysis

frameworks, which may be unfeasible for many users likely due to the lack of expert knowledge and time. From an industrial standpoint, proceeding with commercial packages like JMP, which comprises of software that has been validated and improved across many years, provides the fastest route to implementation where the cost of the license fee can be rapidly recouped as part of customer projects.

6.3 Synthace and automation platform considerations

Key aspects of the framework for implementing SFDs are process miniaturisation and automation to carry out physical experimentation in a reproducible and timely fashion (Section 4.2.1). In this work, the Dragonfly dispenser was used as the automation platform of choice and liquid handling instructions were generated from experimental workflows created using the Synthace platform. Decisions made in Synthace workflows, and the Dragonfly software directly impact the efficiency of the physical execution process, therefore, the device specifications must be at the forefront of all *in silico* considerations (Figure 6.1).

For this work, the Dragonfly dispenser was chosen for SFD execution due to its ability to dispense variable volumes rapidly via non-contact dispensing, thus making the preparation of reaction mixtures incredibly fast (Section 5.2.2). The use of a non-contact dispenser is recommended compared to using liquid handling devices that require liquid transfers from one location to another as they need to change pipetting tips with every transfer and would require significantly longer to prepare reaction mixtures. This would not be well suited to SFD preparation due to prolonged reaction mixture preparation time which in turn would leave all reagents exposed to evaporation. However, it must be noted that confirming device precision is critical prior to physical experimentation and there is a need for regular equipment calibration and validation to ensure device performance is maintained.

Furthermore, for companies that wish to study new assay systems, numerous assay logistics that directly influence assay performance also need to be investigated prior to committing to a large-scale characterisation study such as a SFD to ensure that the best possible data is generated. These include non-biological considerations such as miniaturisation, selection of appropriate labware and reaction volumes, identifying suitable measurement wavelengths, confirming the linear regime of the spectrophotometer, confirming automated device precision, and accounting for methods to control evaporation etc., that can often be overlooked.

SFD studies can require numerous reagents, for example, the SFD study executed in Chapter 5 required 27 reagents in total (7 different reagents at different starting stock concentrations). However, the Dragonfly dispenser's reservoir tray can only house up to 10 reagents, therefore the reagents need to be grouped into various stages. This specification can be incorporated into the Synthace workflow. Within the 'Run DoE' element, the user can group the reagents required for the whole experiment into various experimental stages. Limiting each stage to 10 reagents ensures no manual grouping of reagents is required within the Dragonfly control software. As the same underlying workflow is used to generate all instruction files for all experimental blocks, reagents within each stage will be the same (Figure 6.1 A).

Reagents can be grouped into stages based on their function. For example, stable reagents such as water, buffers, salt solutions, solvents, can be grouped and dispensed at once. However, reagents that need to be added at specific time intervals, such as substrates or enzyme, can be grouped separately. This ensures that complex reaction mixtures can be prepared fully before adding time-sensitive reagents (Figure 6.1 B).

The Dragonfly dispenser control software allows the user to upload numerous instruction files at once, this ensures that all dispensing instructions can be accessed from one place. When instruction files are uploaded, reagents are allocated to random reservoir tray positions by the vendor's software. Also, It is time-consuming and inefficient to prepare one plate at a time as syringes need to be initiated and removed between each stage to avoid cross-contamination. To make the plate preparation process efficient, a pre-determined layout can be used to organise reagents within reservoir tray positions for each stage (Figure 6.1 B). As all stages contain the same set of reagents, ensuring that they are in the same location across different instruction files ensures that the same syringe can be used to dispense the reagent from different instruction files. Stages containing non-time sensitive reagents can be dispensed at the start of the experiment. This organisation of reagent layouts streamlines SFD execution significantly.

In the SFD study executed in Chapter 5, Stages 1 and 2 were added to all 36 microplates at the beginning of the experiment. Twenty syringes were used to dispense 20 reagents (out of 27 in total) in just 2 hours. Stage 3 and 4 (substrates and lysate, respectively) were added prior to a fixed incubation period or to kickstart biocatalysis periodically over the day (Figure 6.1 C). This method is efficient as it minimised interaction with the vendor's software, which may lead to additional errors while simultaneously minimising the quantity of syringes required.

By minimising human interaction with automated platforms, the reproducibility of experimentation increases as the devices being used are already calibrated appropriately. This will ensure reliable data generation, which is especially important due to the ongoing reproducibility crisis that is persistent in biological research (Begley and Ellis, 2012; Freedman, Cockburn and Simcoe, 2015)

In summary, from an industrial standpoint using a platform such as Synthace provides a generic solution to automate a wide range of biological experiments across a range of automated devices, thus eliminating the need to learn device-specific control software. This would significantly reduce the time scientists spend learning new software and attempting to automate biological experimentation.

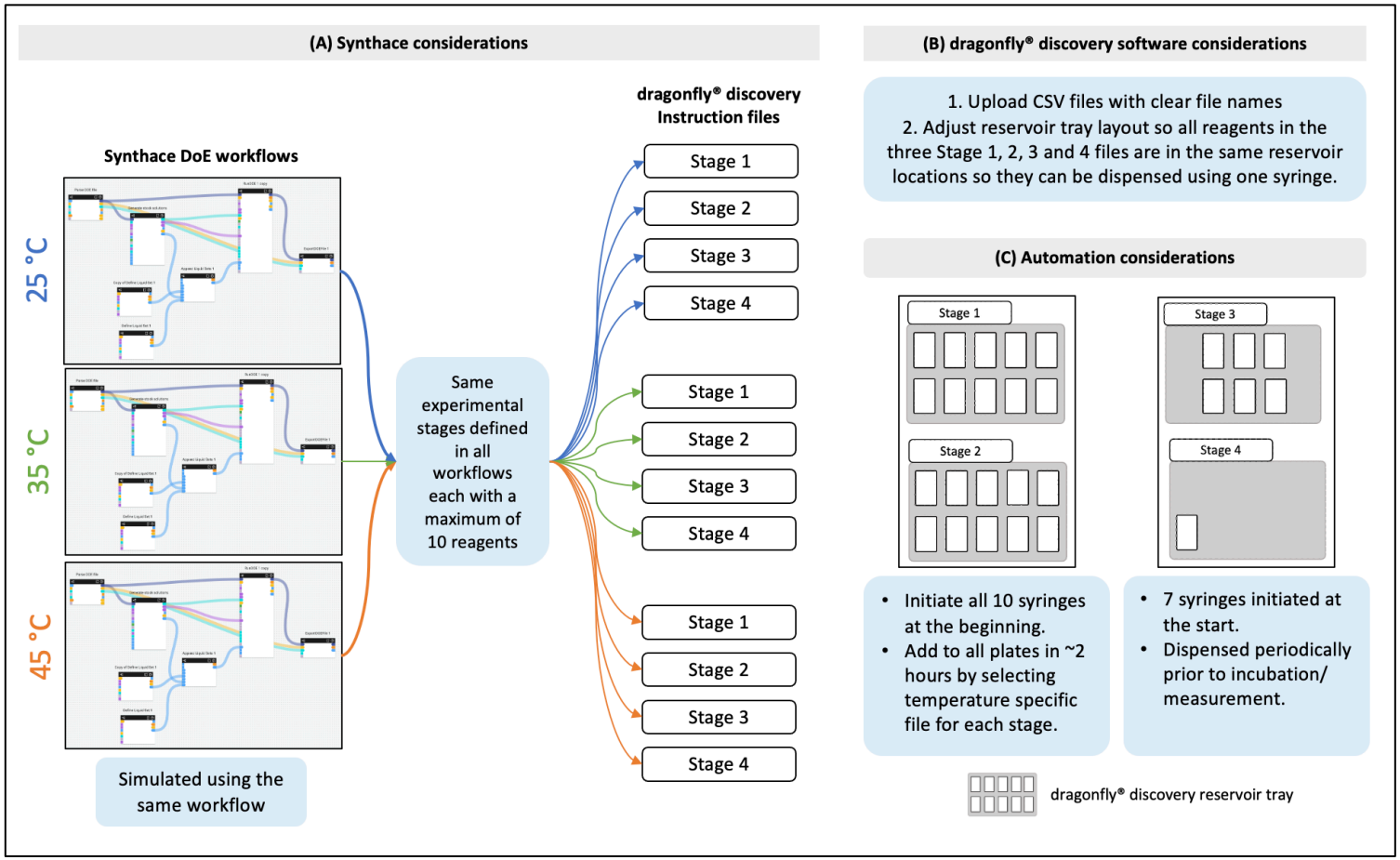


Figure 6.1.Software and automation considerations for streamlining SFD execution.
Blue boxes highlight the considerations made at each stage. Coloured arrows show the flow of information leading to the creation of stage-specific instruction files

6.4 Data-processing considerations

High—dimensionality experiments can generate a large quantity of data making manual analysis using tools like Microsoft Excel unfeasible for accurate response generation. The data processing pipeline developed here using MATLAB can group replicates that are spread within and across microplates, and blank- and time-correct progress curves before generating responses (Section 2.7.2). This pipeline was specifically developed to process time-course absorbance data generated from the SFD study in Chapter 5, however, it can be useful for future applications once a few considerations are made.

While significant time has been taken to generate the MATLAB code, debug it, and verify that the results being generated are correct, further validation would be required before it is considered for widespread use.

The MATLAB data-processing scripts group replicates based on well location. The SFD study in Chapter 5 houses two sets of replicates on one 96-well plate. Each replicate was six columns apart, for example, wells A1 and A7, wells B4 and B10 etc. This pattern is identified across all plates and used to group each individual replicate reaction together. Therefore, the existing scripts can only be used for experiments that follow a similar layout. Modifications that require changing plate type, e.g., to a 384-well plate, or increasing the number of replicates while following a similar layout require basic programming skills. However, identifying randomised replicate locations would require more in-depth specialist programming knowledge.

A novelty of the data-processing pipe that is not commonly found in literature is that each reaction within a 96-well plate is time-corrected (Section 2.7.2.1 and 5.2.3.1). This was designed to correct for the time discrepancy that is introduced when reactions in a 96-well plate are started one by one, in a specific pattern and then measured inside a plate reader at set intervals, again, in a pre-defined pattern (Figure 2.4). Time-correcting absorbance data ensures that a realistic time stamp is allocated

to each absorbance measurement which in turn ensures response precision and reduces confounding the impact of a time discrepancy with noise. However, the function developed to time-correct progress curves uses hard-coded time values and patterns for lysate addition and absorbance measurements; therefore, these time values and patterns may need to be altered for future use cases which would also require programming skills.

Progress curves were blank corrected by subtracting the intercept value generated from a spline fit. Alternatively, a function that blank corrects data using measured absorbance values was also developed. Originally, using a 'blank' value generated from measuring the end-point absorbance of reaction mixtures that contained all reaction components apart from the lysate was intended. However, it was later found that lysate addition noticeably reduced the path length of the final reaction solution. This in turn meant that the blank absorbance value was greater than the first absorbance measurement, which when subtracted yielded a negative starting absorbance. Therefore, it was deemed more accurate to use the intercept of a spline fit. For future use cases, the impact of the enzyme of interest on the meniscus of the reaction solutions should be considered prior to analysis.

Lastly, the yield response was calculated from a function that uses an acetophenone standard curve, alternative standards might need to be included in the scripts if different products are being measured. This is also the case if responses other than yield or initial rate are of interest.

Overall, the considerations highlighted in this section would make the existing MATLAB scripts more generalisable, however, they would require basic programming skills for intended modifications as the pipeline is already developed and can be used by inexperienced MATLAB users. To implement these scripts within industry, further debugging and generalisation of data processing is required. Additionally, the security of the software, its storage on different systems and networks, and the process of data storage would need to be addressed.

In general, data processing software is likely to be bespoke to the specific application of each company. This specificity in turn requires the generation of specific software which can be a time-consuming process that must be validated on a case-by-case basis. Therefore, employing programming experts capable of producing, validating and modifying code is recommended. Alternatively, specific software companies that focus on data analysis and rapid code generation can be contracted to develop bespoke scripts to ensure that the gap between generating large quantities of data through high-dimensional studies and processing that data is minimised.

6.5 Summary

From an industrial perspective, the three categories of considerations outlined in this chapter should be carefully considered and understood prior to attempting the execution of high-dimensionality experiments as they capture the thought process and decision-making required for the generic use of the developed framework. The choice of DoE software directly impacts the process of design generation, visualisation and model generation; therefore, the choice of software and its advantages and limitations should be clearly understood for successful DoE implementation. The choice of software used to generate liquid handling instructions and the selection and preparation of automated devices is critical in ensuring that the barriers of automated experimentation are overcome (previously outline in Section 1.2.4). Finally, having a verified data processing pipeline in place can significantly reduce the time between experiment execution and analysis, ensuring that the data collected is comprehended as soon as feasible.

Lastly, the final consideration is the need for the company manager to understand the benefits of automated experimentation and to act as a champion for change by promoting the implementation of automated high-dimensionality studies for rapid assay characterisation.

Having now considered the industrial implementation of the SFD experimental platform established in this thesis, Chapter 7 outlines the major conclusions arising from this research and presents suggestions for future work.

7 Conclusions and Future Work

The main aim of this thesis was to create an effective framework for the automated execution and analysis of high-dimensionality experiments based upon DoE methodology. Creation of the framework relied upon integration of a number of key components as illustrated sequentially in Figure 3.6, Figure 4.1 and Figure 5.2. These included: statistical DoE software (for initial design generation and final data analysis), microtiter plate experimentation (for assay miniaturisation and precision), robotic liquid handling platforms (for assay automation and parallelisation), Synthace software (for efficient conversion of DoE designs into automated liquid handling protocols), a microplate reader (for time-course data generation), and a MATLAB script (for automated data processing). The framework was exemplified by using a CV2025 TAm activity assay, the AP assay, as the model system which measured the rate and extent of Acetophenone production spectrophotometrically at 245 nm (Figure 3.1).

In Chapter 3, the use of JMP for DOE design generation and Synthace-driven DoE execution of traditional DoE screening and optimisation studies using a PipetMax liquid handler and a Dragonfly dispenser was exemplified (Figure 3.6 and Figure 3.19). Chapter 4 showed the systematic process of refining many non-biological assay logistics that are fundamental to miniaturised and automated assay performance (Figure 4.1). These considerations were incorporated into Chapter 5 where a framework for rapid assay characterisation was developed by automating the execution and analysis of a modern DoE design type, called a space-filling DoE design (Figure 5.2).

7.1 Conclusions

The main conclusions drawn from this thesis are divided into two main sections. The suitability of the AP assay as the model system and its impact on creating effective frameworks are outlined in Section 7.1.1, and the extensibility of the Synthace

platform when executing automated experiments is outlined in Section 7.1.2 The future work suggested for this project is detailed in Section 7.2.

7.1.1 AP assay as a model system

The AP assay was found to be a suitable model system for the experiments executed in this thesis. It quantified the activity of the PLP-dependent ω-transaminase, CV2025, which catalyses enzymatic amino group transfer by a ping-pong bi-bi mechanism (Bulos and Handler, 1965b; Kuramitsu *et al.*, 1990b), which is more complex than catalysis of the more commonly used systems that implement Michaelis-Menten kinetics (Vasić-Rački, Findrik and Vrsalović Presečki, 2011). This complexity enabled the investigation of the impact of two substrates and two products on bioconversion kinetics. Furthermore, the AP assay model system also allowed the investigation of important generic factors that are critical to fully characterising novel biocatalysts. These include the impact of co-factors such as PLP, solvent stability due to the high solvent tolerance of the CV2025 TAm, product inhibition, the impact of various buffering agents and salts, the impact of bioconversions over a wide pH range, reaction temperatures, and the stability of CV2025 as a function of time (Table 3.1, Table 3.2 and Table 5.1).

Additionally, the AP assay is an absorbance-based assay, which made it suitable for analysing DoE studies as the reaction progress could be easily monitored in real-time using a microplate reader. Generation of both endpoint and time-course absorbance data generated quantitative responses that were critical to generating all mathematical models in this thesis. Furthermore, important biocatalytic responses such as initial reaction rate and final yield (used in Chapter 5) were only possible to calculate due to the measurement of time-course absorbance generated from the AP assay and would not have been possible if a qualitative or endpoint, assay was used as a model system instead.

As the AP assay has been used previously in biological research, direct comparisons could be drawn between findings from DoE studies and those found in the literature. Analysis of DoE studies in both Chapters 3 and 5 generated findings that confirmed conclusions that were found in the literature based on more minimal data sets. The more extensive DoE studies reported here, for example, showed that the optimal pH value of ~ 9 units was identified in the Screening DoE (Section 3.2.5.4), however, the peak at pH 9 was confirmed in the SFD (Section 5.2.6.1 and 5.2.7.1) due to sampling over a wider range of 7 to 10. The majority of the experiments found in the literature were based on context-specific OFAT experiments (Shin and Kim, 1998; Shin *et al.*, 2003; Kaulmann *et al.*, 2007; Schätzle *et al.*, 2009; Bea *et al.*, 2010; Chen *et al.*, 2016; Chen, Berglund and Humble, 2018; Meng *et al.*, 2020). However, the work done in this thesis not only identified the optimum factor levels but also provided insight into how each factor is influenced by all other factors. Furthermore, the work performed in this thesis has greatly enhanced the characterisation of the AP assay as mathematical models were generated from multi-factorial experiments which provide insight not only into the impact of individual factors and settings but more importantly gives insight into the factor interactions that exist (Section 3.3 and 5.3).

Since replication of DoE studies comprising of high run numbers were executed, information-rich datasets that finely characterise the AP assay have been created. These datasets and corresponding mathematical models can be used to identify suitable reaction conditions for different industrially relevant use cases. For example, the findings from this thesis could be beneficial for transaminase catalysed reactions that are a part of cascade systems, where the products from transamination are reused as substrates for subsequent reactions as identifying suitable reaction solutions for such systems can often be difficult (Abaházi *et al.*, 2018; Ferrandi and Monti, 2018).

One significant limitation of this assay should be recognised. This was that Acetophenone, one of the products of the reaction, could not be used as a factor in any of the DoE designs as it is used to quantify CV2025 activity. This was a drawback

of the AP assay as acetophenone inhibition on the forward reaction has been previously described (Shin and Kim, 1998) and investigation of the substrate and product inhibition are key aspects of biocatalyst characterisation. Increasing initial concentrations of AP would have saturated the plate reader detector reducing the dynamic range of the assay (Figure 4.6). If necessary, this problem could be overcome by employing an alternative assay such as HPLC.

The main conclusion from studying the CV2025 TAm catalysed bioconversion, however, is that due to the numerous components that are required by the AP assay, a detailed systematic framework to guide researchers in confirming various assay logistics that are often ignored or overlooked was developed in Chapter 4, this framework is generic enough to be applicable to a wide range of bioconversion reactions over a wide range of conditions.

7.1.2 The extensibility of the Synthace platform and overcoming barriers to automated DoE execution

To use automation effectively for the execution of biological experiments, expert knowledge of device-specific control software is often a prerequisite. Depending on the complexity of the device and the control software interface, it is common to define every single liquid handling action required by an experiment (Section 1.2.4) which also makes the approach prone to error. This cumbersome process is often seen as one of the biggest barriers to accessing automation to its full potential, even for the most straightforward biological experiments (Oritz *et al.*, 2017). This barrier is further intensified when automating DoE studies due to their requirement for the preparation of numerous, highly-variable, randomised run conditions whose instructions are challenging to create using device-specific control software (Section 1.2.3).

In this thesis, the ways in which the Synthace software overcomes these barriers to enable the creation of effective protocols for automated experiments are demonstrated. It was the sole software used to flexibly generate liquid-handling

instructions for all automated experiments of varying complexity, from the preparation of standard curves to the execution of DoE studies comprising of thousands of run conditions. With Synthace, physical execution required limited interaction with the control software of the devices. As device-specific liquid handling instructions are automatically generated with successful workflow simulation, physical experiments executed with the PipetMax and Hamilton liquid handlers were started from SynthaceHub housed on the computer controlling the device (Figure 1.3). The experiments executed using the Dragonfly dispenser required a manual upload of Synthace-generated instruction files with the control software used to organise the input reagents and initiate syringes prior to dispensing (Figure 5.5). Minimising the interaction with control software greatly reduced the occurrence of manual errors and makes the whole process of automated experimentation more user-friendly.

The same underlying Synthace workflow was used to execute all DoE studies (Section 3.2.5.3). By using a standardised workflow, DoE designs were translated into liquid handling instructions following the same steps for both the Pipetmax liquid handler and the Dragonfly dispenser. This overcomes the barrier of creating both device-specific and experiment-specific protocols as the user simply needs to select the device needed for automated execution during workflow preparation. However, it is important to note that only a selection of automated devices is supported on the Synthace platform at the present time.

The simplification of liquid-handling instruction generation via Synthace combined with the automated physical implementation led to the creation of a framework for rapid DoE execution. Traditional DoE designs, such as the screening and optimisation DoE studies executed in Chapter 3 prepared using the PipetMax liquid handler took just 4 and 7 hours, respectively (Table 3.3. Time required for automated execution of the AP assay optimization DoE.). This time was drastically minimised to minutes when the Dragonfly dispenser was used to replicate the process of executing the optimisation DoE study while yielding almost identical results (Section 3.2.5.3 and

Table 3.4). This work exemplifies that standardised liquid-handling instructions can not only be generated quickly using Synthace but that they are also trustworthy across various devices.

Furthermore, as the process of generating liquid-handling instructions using Synthace is the same despite the underlying complexity of the DoE study, modern design types such as space-filling designs (Chapter 5) could be executed in a straight-forward manner, in turn making DoE design types that were previously beyond the scope of physical execution just as accessible as any other automated experiment, this has previously never been demonstrated in the literature. This was exemplified in Chapter 5 where a DoE study using a SFD design finely characterised a 12-dimensional design space of the AP assay by executing 3,456 run conditions in just 4 days (Figure 5.4 and Figure 5.5). Due to the large quantity of data generated, sophisticated data processing scripts using MATLAB were necessary to generate responses (Section 2.7.2 and Section 5.2.3). These MATLAB scripts provided routines for automated data collection and processing and were essential in developing an end-to-end automated framework. Without them, the bottleneck and risk of error would have been shifted from automated execution to data processing and analysis and would have drastically increased the time and effort required to create and physically validate mathematical models for yield and initial rate (Figure 5.10, Figure 5.13 and Figure 5.17). The future work that could be associated with this thesis is described in Section 7.2

7.2 Future Work

7.2.1 Reproduction of key experiments

Prior to launching into new areas of future work, some of the experiments conducted in each chapter could be improved. In Chapter 3, iterative DoE studies could be repeated without the inflation of the total run number, this would minimise the number of experiments that needed to be prepared while generating examples of automated studies that use minimal, software-recommended designs. This would

provide a more realistic example of the automated version of a typical DoE pipeline. Repetition of iterative DoE studies across a range of automated devices would further increase the understanding of the process of automated DoE execution regarding the benefits and limitations of each device. It would also further confirm the device-agnostic capabilities of the Synthace platform.

In Chapter 4, a systematic investigation of various assay logistics was generated for 96- and 384-well microplate formats, however, expanding to the use of 1,536 well microplates would be beneficial as the application of the work described in this thesis can be used in high-throughput screening and drug discovery where 1,536 well microplates are routinely used (Di and Kerns, 2006; Schneider, 2018). This would also highlight which devices are best suited to dispensing into 1,536 microplates.

In Chapter 5, the aim of creating a one-step characterisation framework was prioritised. However, in retrospect, a more detailed framework could be developed by executing another DoE study that focused on the active area of the design space, thus moving into optimisation rather than just characterisation of the assay alone. The models generated from this study could be used to conduct a validation study that evaluated the model in a known region of the design space before validating an unknown region of the design space. This process would be more comprehensive and could potentially improve model prediction.

7.2.2 Validation of the framework on additional assay systems

The framework developed in Chapter 4 for selecting suitable assay logistics and the framework developed in Chapter 5 for implementing, analysing, and validating space-filling designs were successfully exemplified by the AP assay. However, reusing those frameworks to characterise other assay systems, that are either less well-studied than CV2025 TAm or completely novel, would further highlight the applicability, advantages, and the generic nature of the framework described in this thesis.

If the developed framework is reused on a poorly characterised or unknown biocatalytic reaction, the methods described in this thesis can be further validated and improved. This would include the impact of a large one-step characterisation experiment, the suitability of the data processing pipeline, and the suitability of the Gaussian Process Model, along with the precision of model predictions. Overall, this would ensure that the framework is more generic and user-friendly to research groups focused on biocatalytic reaction screening, assay development, drug discovery, cell-free systems, etc.

7.2.3 Broadening the suite of automated devices and labware

As the DoE designs described in Chapter 3 were executed using liquid-handling devices, they could be repeated using other devices that are supported by Synthace, such as the Hamilton STAR or Tecan Freedom Evo (Tecan Group). Other dispensers such as the Mantis dispenser (Formulatrix), Echo acoustic liquid handler (Beckman Coulter), or Certus Flex (Gyger) could also be used. This would further confirm the device-agnostic capabilities of Synthace while identifying the advantages and disadvantages of using each device for automated DoE execution regarding the challenges surrounding device setup, interaction with control software, issues that lead to execution errors, points at which manual intervention is required, etc.

For more complex high-dimensionality experiments such as SFDs, using a device such as the Echo acoustic liquid handler could provide insight into preparing experiments at much lower reaction volumes than was used in this thesis as it can transfer volumes as low as 2.5 nanolitres. Using this device could lead to preparing reaction mixtures at lower final volumes while mixing many reagents into each reaction, as required by the specific run condition. This could greatly minimise reagent costs. By using lower reaction volumes, throughput can be increased by using 384- and 1536-well microplates for DoE studies. This would drastically reduce the time required to complete entire experiments. However, the assay logistics that surround response generation and signal precision would need to be confirmed prior to changing microplate formats. These experiments would make the execution of experiments

that use DoE methodology even more accessible than the frameworks created in this thesis.

7.2.4 Expansion of the data processing pipeline

The existing MATLAB data processing pipeline could be expanded to include additional features such as outlier detection. Currently, no data curation takes place within the existing data processing script. One example of data curation is the identification and removal of outliers. Outliers from noisy data are routinely present in biological research. Therefore, creating an addition to the scripts where potential outliers can be detected from a set of replicates would allow the researcher to curate the dataset and identify any patterns that may have led to the generation of noisy data. This would also impact the quality and predictive capability of the statistical model generated from data downstream.

Furthermore, the set of scripts could be merged into one main script that can generate a set of responses, currently, only one response is generated per script. Currently, grouping replicates is dependent on a specific layout within a 96-well plate. Expanding this feature to group completely randomised run conditions across various plate types of different formats would increase the flexibility of the pipeline. As MATLAB is a commercial third-party software, converting data processing scripts to function in free, open-source programming languages such as R or Python would make them more user-friendly.

By using such languages, various analysis packages for Gaussian Process Modelling can also be used to analyse the response generated. This could potentially reduce or eliminate the use of JMP, yet another commercial third-party software used to generate mathematical models. However, differing results significantly varying in precision have been previously reported when different GPM packages have been used to analyse the same dataset, therefore, any new packages selected would need to be physically validated to confirm model predictions (Erickson et al., 2018).

Furthermore, one of the most beneficial additions to the existing framework is the expansion of Synthace's current data visualisation and analysis capabilities. The ability to link absorbance data, visualise it, generate and analyses responses all within Synthace would be greatly beneficial as it would conserve provenance and provide a no-code solution to sophisticated data processing and analysis, thus reducing and potentially removing the need to be proficient at programming languages.

7.2.5 Understanding the power of space-filling designs and expanding on benefits of their physical execution

By following the framework developed in this thesis, SFDs can be easily executed and analysed within a matter of weeks and have been found to be a beneficial tool for mapping high-dimensional design spaces. However, due to their inherent lack of replication within factor levels, they are less statistically powerful than traditional DoE designs (Santner et al, 2018), therefore, understanding the power of SFDs and gathering more information on suitable design strategies for physical execution in biological research is an exciting area of future work. For example, the SFD executed in Chapter 5 aimed to characterise a 12-dimensional space using a 414-run design replicated 8 times as the aim was to execute a design that minimally covered the design space while capturing the noise of unexplored spaces. By analysing the power of space-filling designs, the following questions could be answered. Could the dataset generated from this experiment be thinned down further while still maintaining the same model predictions? Is the dataset dense enough to predict missing data points? Does the model fit increase when responses are generated from a high number of replicated run conditions, if so, is there a point of diminishing returns? How many minimum run conditions are required to capture a factor effect? How many replicates are required to capture noise? These questions can be answered by creating analysis scripts that automatically build GPMs from different subsets of the datasets while including a varying number of replicates. Such analysis would be beneficial when conducting both exploratory or exploitative work of new biological systems and would greatly be beneficial in scenarios where it is not possible to invest resources to execute large experiments.

Overall, the work described in this thesis has demonstrated the potential of performing automated, high dimensionality DoE experiments to enable the rapid and detailed characterisation of biological reactions.

8 References

1. Abaházi, E. *et al.* (2018) “Covalently immobilized Trp60Cys mutant of Ω-transaminase from *Chromobacterium violaceum* for kinetic resolution of racemic amines in batch and continuous-flow modes,” *Biochemical Engineering Journal*, 132, pp. 270–278. doi:10.1016/j.bej.2018.01.022.
2. Acker, M.G. and Auld, D.S. (2014) “Considerations for the design and reporting of enzyme assays in high-throughput screening applications,” *Perspectives in Science*, 1(1–6), pp. 56–73. doi:10.1016/j.pisc.2013.12.001.
3. Acosta-Martinez, J.P. *et al.* (2010) “Ultra scale-down stress analysis of the bioprocessing of whole human cells as a basis for cancer vaccines,” *Biotechnology and Bioengineering*, 107(6), pp. 953–963. doi:10.1002/bit.22888.
4. Adinarayana, K. and Ellaiah, P. (2002) “Response surface optimization of the critical medium components for the production of alkaline protease by a newly isolated Bacillus sp,” *Journal of Pharmacy and Pharmaceutical Sciences*, 5(3), pp. 272–278.
5. Albert, K.J. (2007) “Minimizing Liquid Delivery Risk: Automated Liquid Handlers as Sources of Error.,” *Am. Lab.*, June-July, pp. 8–10.
6. Anderson, M.J. and Whitcomb, P.J. (2007) *DOE simplified: practical tools for effective experimentation*.
7. Anderson, M.J. and Whitcomb, P.J. (2016) *RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments.* 2nd Edition. CRC Press.
8. Antony, J. (2014a) *Design of Experiments and its Applications in the Service Industry*. Second Edi, *Design of Experiments for Engineers and Scientists*. Second Edi. Elsevier Ltd. doi:10.1016/B978-0-08-099417-8.00010-9.
9. Antony, J. (2014b) “Fundamentals of Design of Experiments,” *Design of Experiments for Engineers and Scientists*, Second Edi, pp. 7–17. doi:10.1016/B978-0-08-099417-8.00002-X.
10. Appleton, E. *et al.* (2017) “Needs and opportunities in bio-design automation: four areas for focus,” *Current Opinion in Chemical Biology*, 40, pp. 111–118. doi:10.1016/j.cbpa.2017.08.005.
11. Armstrong, J.W. (1999) “A review of high-throughput screening approaches for drug discovery,” *American Biotechnology Laboratory*, 17(5), pp. 26–28.
12. Baillargeon, P. *et al.* (2019) “Design of Microplate-Compatible Illumination Panels for a Semiautomated Benchtop Pipetting System,” *SLAS Technology*, 24(4), pp. 399–407. doi:10.1177/2472630318822476.
13. Baud, D. *et al.* (2015) “A rapid, sensitive colorimetric assay for the high-throughput screening of transaminases in liquid or solid-phase,” *Chem. Commun.*, 51(97), pp. 17225–17228. doi:10.1039/C5CC06817G.

14. Bea, H.S. *et al.* (2010) “Kinetic resolution of α-methylbenzylamine by recombinant *Pichia pastoris* expressing ω-transaminase,” *Biotechnology and Bioprocess Engineering*, 15(3), pp. 429–434. doi:10.1007/s12257-009-3093-1.
15. Beal, J., Lu, T. and Weiss, R. (2021) “Automatic compilation from high-level biologically-oriented programming language to genetic regulatory networks,” in *PLoS ONE*. doi:10.1371/journal.pone.0022490.
16. Beer (1852) “Determination of the absorption of red light in coloured liquids.,” *Annalen der Physik und Chemie*, 86, pp. 78–88.
17. Begley, C.G. and Ellis, L.M. (2012) “Raise standards for preclinical cancer research,” *Nature*, 483, pp. 531–533.
18. Bonowski, F. *et al.* (2010) “Computer controlled automated assay for comprehensive studies of enzyme kinetic parameters,” *PLoS ONE*, 5(5). doi:10.1371/journal.pone.0010727.
19. Bos, A.B. *et al.* (2015) “Optimization and automation of an end-to-end high throughput microscale transient protein production process,” *Biotechnology and Bioengineering*, 112(9), pp. 1832–1842. doi:10.1002/bit.25601.
20. Bouguer, P. (1729) “Optics essay on the attenuation of light,” (Claude Jombert), pp. 16–22.
21. Box, G.E.P. and Wilson, B. (1951) “On the Experimental Attainment of Optimum Conditions Published by: Wiley for the Royal Statistical Society Stable URL : http://www.jstor.org/stable/2983966,” *Journal of the Royal Statistical Society -Series B*, 13(1), pp. 1–45.
22. Bradshaw, J.T. *et al.* (2005) “Multichannel Verification System (MVS): A Dual-Dye Ratiometric Photometry system for performance verification of multichannel liquid delivery devices,” *JALA - Journal of the Association for Laboratory Automation*, 10(1), pp. 35–42. doi:10.1016/j.jala.2004.08.012.
23. Braunstein., A.E. and Kritzman., M.G. (1937) “Formation and Breakdown of Amino-acids by Intermolecular Transfer of the Amino Group.,” *Nature*, pp. 503–504.
24. Brunhuber, N.M.W. and Blanchard, J.S. (1994) “The Biochemistry and Enzymology of Amino Acid Dehydrogenases,” *Critical Reviews in Biochemistry and Molecular Biology*, 29(6), pp. 415–467. doi:10.3109/10409239409083486.
25. Bulos, B. and Handler, P. (1965a) “Kinetics of Beef Heart Glutamic-Alanine Transaminase.,” *The Journal of biological chemistry*, 240(8), pp. 3283–3294. doi:10.1016/s0021-9258(18)97216-9.
26. Bulos, B. and Handler, P. (1965b) “Kinetics of Beef Heart Glutamic-Alanine Transaminase.,” *The Journal of biological chemistry*, 240(8), pp. 3283–3294. doi:10.1016/s0021-9258(18)97216-9.
27. Burke, S. (2018) “Computer Experiments: Space Filling Design and Gaussian Process Modelling,” (September).
28. Cambray, G., Guimaraes, J.C. and Arkin, A.P. (2018) “Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*.” *Nature Biotechnology*, 36(10), p. 1005. doi:10.1038/nbt.4238.

29. Chao, R. *et al.* (2017) “Fully Automated One-Step Synthesis of Single-Transcript TALEN Pairs Using a Biological Foundry,” *ACS Synthetic Biology*, 6(4), pp. 678–685. doi:10.1021/acssynbio.6b00293.
30. Chen, S. *et al.* (2016) “Stabilization of an amine transaminase for biocatalysis,” *Journal of Molecular Catalysis B: Enzymatic*, 124, pp. 20–28. doi:10.1016/j.molcatb.2015.11.022.
31. Chen, S. *et al.* (2018) “Characterization of the stability of *Vibrio fluvialis* JS17 amine transaminase,” *Journal of Biotechnology*, 282(November 2017), pp. 10–17. doi:10.1016/j.jbiotec.2018.06.309.
32. Chen, S., Berglund, P. and Humble, M.S. (2018) “The effect of phosphate group binding cup coordination on the stability of the amine transaminase from *Chromobacterium violaceum*,” *Molecular Catalysis*, 446, pp. 115–123. doi:10.1016/j.mcat.2017.12.033.
33. Christensen, M. *et al.* (2021) “Automation isn’t automatic,” *Chemical Science*, pp. 15473–15490. doi:10.1039/d1sc04588a.
34. Coleman, D.E. and Montgomery, D.C. (1993) “A Systematic Approach to Planning for a Designed Industrial Experiment,” *Technometrics*, 35(1), pp. 1–12. doi:10.1080/00401706.1993.10484984.
35. Coleman, D.E. and Montgomery, D.C. (1996) “A systematic approach to planning for a designed industrial experiment,” *IEEE Engineering Management Review*, 24(1), pp. 75–86. doi:10.2307/1269280.
36. Crone, M.A. *et al.* (2020) “A role for Biofoundries in rapid development clinical diagnostics,” *Nature Communications*, pp. 1–11. doi:10.1038/s41467-020-18130-3.
37. Deszcz, D. *et al.* (2015) “Dalby.1466729_Templated.FEBSJ._revised.pdf,” *FEBS Journal* [Preprint]. doi:10.1111/febs.13293.
38. Di, L. and Kerns, E.H. (2006) “Biological assay challenges from compound solubility: strategies for bioassay optimization,” *Drug Discovery Today*, 11(9–10), pp. 446–451. doi:10.1016/j.drudis.2006.03.004.
39. Elfving, G. (1952) “Optimum Allocation in Linear Regression Theory.,” 23(2), pp. 255–262.
40. Erbach, T. *et al.* (2004) “How Experimental Design Optimizes Assay Automation,” pp. 1–8.
41. Erickson, C.B., Ankenman, B.E. and Sanchez, S.M. (2018) “Comparison of Gaussian process modelling software,” *European Journal of Operational Research*, 266(1), pp. 179–192. doi:10.1016/j.ejor.2017.10.002.
42. Fan, L. (2016) “HSPA-A universal graphical user interface for the Hamilton Microlab STAR liquid handler,” *Journal of Applied Bioanalysis*, 2(1), pp. 38–44. doi:10.17145/jab.16.006.
43. Farooq, M.A. *et al.* (2016) “An innovative approach for planning and execution of pre-experimental runs for Design of Experiments,” *European Research on Management and Business Economics*, 22(3), pp. 155–161. doi:10.1016/j.iedee.2014.12.003.
44. Fellermann, H. *et al.* (2019) “Design of experiments and the virtual PCR simulator: An online game for pharmaceutical scientists and

biotechnologists," *Pharmaceutical Statistics*, 18(4), pp. 402–406. doi:10.1002/pst.1932.
45. Ferrandi, E.E. and Monti, D. (2018) "Amine transaminases in chiral amines synthesis: recent advances and challenges," *World Journal of Microbiology and Biotechnology*, 34(1), pp. 1–10. doi:10.1007/s11274-017-2395-2.
46. Fisher, R. (1926) "The arrangement of field experiments" *J Minist Agric (GB)*.
47. Freedman, L.P., Cockburn, I.M. and Simcoe, T.S. (2015) "The Economics of Reproducibility in Preclinical Research," *PLOS Biology*, 13(6), p. e1002165. Available at: https://doi.org/10.1371/journal.pbio.1002165.
48. Fricke, J. *et al.* (2013) "Designing a fully automated multi-bioreactor plant for fast DoE optimization of pharmaceutical protein production," *Biotechnology Journal*, 8(6), pp. 738–747. doi:10.1002/biot.201200190.
49. Gilman, J. *et al.* (2020) "Statistical Design of Experiments for Synthetic Biology." doi:10.1021/acssynbio.0c00385.
50. Gilman, J. *et al.* (2021) "Statistical Design of Experiments for Synthetic Biology," *ACS Synthetic Biology*, 10(1), pp. 1–18. doi:10.1021/acssynbio.0c00385.
51. Gilson Inc. (2018) *PIPETMAX ® 268 User's Guide*.
52. Goos, P. (2002) *The Optimal Design of Blocked and Split Plot Experiments.* New York: Springer.
53. Goos, P. and Jones, B. (2011) "Optimal Design of Experiments: A Case Study Approach.," in. New York: John Wiley and Sons.
54. Goos, P., Jones, B. and Syafitri, U. (2000) "I-optimal mixture designs," *Faculty of Applied Economics*, 32(0).
55. Gotwalt, C.M. (2012) "JMP Neural Network Methodology," pp. 1–11.
56. Gruber, P. *et al.* (2017) "Real-time pH monitoring of industrially relevant enzymatic reactions in a microfluidic side-entry reactor (μSER) shows potential for pH control," *Biotechnology Journal*, 12(6). doi:10.1002/biot.201600475.
57. Gupta, M.N. (1992) "Enzyme function in organic solvents.," *European journal of biochemistry / FEBS*, 203(1992), pp. 25–32. doi:10.1111/j.1432-1033.1992.tb19823.x.
58. Hamilton (2015) "Microlab ® STAR User Manual for User Software 3.0 including Options and Accessories."
59. Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*. Second Edi, *Revista Espanola de las Enfermedades del Aparato Digestivo*. Second Edi. Springer US.
60. Hayden, E.C. (2014) "The automated lab.," *Nature*, pp. 5–6.
61. Hecht, E.S., Oberg, A.L. and Muddiman, D.C. (2016) "Optimizing Mass Spectrometry Analyses: A Tailored Review on the Utility of Design of Experiments," *Journal of the American Society for Mass Spectrometry*, 27(5), pp. 767–785. doi:10.1007/s13361-016-1344-x.
62. Heckmann, C.M., Dominguez, B. and Paradisi, F. (2021) "Enantio-Complementary Continuous-Flow Synthesis of 2-Aminobutane Using Covalently Immobilized Transaminases," *ACS Sustainable Chemistry and Engineering*, 9(11), pp. 4122–4129. doi:10.1021/acssuschemeng.0c09075.

63. Hentz, N.G. and Knaide, T.R. (2014) “Effect of Liquid-Handling Accuracy on Assay Performance,” *Journal of Laboratory Automation*, 19(2), pp. 153–162. doi:10.1177/2211068213504095.
64. Ho, M., Robins, K. and Bornscheuer, U.T. (2010) “Conductometric Method for the Rapid Characterization of the Substrate Specificity of,” 82(5), pp. 2082–2086.
65. Hopwood, J. *et al.* (2011) “A fast and sensitive assay for measuring the activity and enantioselectivity of transaminases,” *Chemical Communications*, 47(2), pp. 773–775. doi:10.1039/c0cc02919j.
66. Humble, M.S. *et al.* (2012) “Crystal structures of the *Chromobacterium violaceum* ω-transaminase reveal major structural rearrangements upon binding of coenzyme PLP,” *FEBS Journal*, 279(5), pp. 779–792. doi:10.1111/j.1742-4658.2012.08468.x.
67. Hussain, W. *et al.* (2013) “Reproducible culture and differentiation of mouse embryonic stem cells using an automated microwell platform,” *Biochemical Engineering Journal*, 77, pp. 246–257. doi:10.1016/j.bej.2013.05.008.
68. Hwang, B.-Y. and Kim, B.-G. (2004) “High-throughput screening method for the identification of active and enantioselective ω-transaminases,” *Enzyme and Microbial Technology*, 34(5), pp. 429–436. doi:https://doi.org/10.1016/j.enzmictec.2003.11.019.
69. Hyun, Y.L. and Davidson, V.L. (1995) “Mechanistic Studies of Aromatic Amine Dehydrogenase, a Tryptophan Tryptophylquinone Enzyme,” *Biochemistry*, 34(3), pp. 816–823. doi:10.1021/bi00003a015.
70. ICH Q8 (R2) (2014) *ICH Topic Q 8 (R2) Pharmaceutical Development, Step 5: Note for Guidance on Pharmaceutical Development.*, *Regulatory ICH*. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002872.pdf.
71. Islam, R.S. *et al.* (2007) “Framework for the Rapid Optimization of Soluble Protein Expression in Escherichia coli Combining Microscale Experiments and Statistical Experimental Design,” *Biotechnology progress*, 23, pp. 785–793.
72. Jacyna, J., Kordalewska, M. and Markuszewski, M.J. (2019) “Design of Experiments in metabolomics-related studies: An overview,” *Journal of Pharmaceutical and Biomedical Analysis*, 164, pp. 598–606. doi:10.1016/j.jpba.2018.11.027.
73. Jain, A., Hurkat, P. and Jain, S.K. (2019) “Development of liposomes using formulation by design: Basics to recent advances,” *Chemistry and Physics of Lipids*, 224(February), p. 104764. doi:10.1016/j.chemphyslip.2019.03.017.
74. John, R.A. (1995) “Pyridoxal phosphate-dependent enzymes.,” *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular Enzymology*, 1248(2), pp. 81–96. doi:10.1016/0167-4838(95)00025-P.
75. Jones, B. and Johnson, R.T. (2009) “Design and analysis for the gaussian process model,” *Quality and Reliability Engineering International*, 25(5), pp. 515–524. doi:10.1002/qre.1044.
76. Jones, M., Clark, V. and Clulow, S. (2003) “The importance of the quality control of laboratory automation,” *JALA - Journal of the Association for*

Laboratory Automation, 8(2), pp. 55–57. doi:10.1016/S1535-5535(04)00253-9.
77. Joseph, V.R. (2016) "Space-filling designs for computer experiments: A review," *Quality Engineering*, 28(1), pp. 28–35. doi:10.1080/08982112.2015.1100447.
78. Joseph, V.R., Gul, E. and Ba, S. (2015) "Maximum projection designs for computer experiments," *Biometrika*, 102(2), pp. 371–380. doi:10.1093/biomet/asv002.
79. Jouda, S. *et al.* (2012) *Purification and immobilization of a transaminase for the preparation of an enzyme bioreactor*.
80. Kaplan, B.E. (1985) "The automated synthesis of oligodeoxyribonucleotides," *Trends in Biotechnology*, 3(10), pp. 253–256. doi:10.1016/0167-7799(85)90024-1.
81. Kaulmann, U. *et al.* (2007) "Substrate spectrum of ω -transaminase from *Chromobacterium violaceum* DSM30191 and its potential for biocatalysis," *Enzyme and Microbial Technology*, 41(5), pp. 628–637. doi:10.1016/j.enzmictec.2007.05.011.
82. Kelly, S.A. *et al.* (2020) "Transaminases for industrial biocatalysis: novel enzyme discovery," *Applied Microbiology and Biotechnology*, 104(11), pp. 4781–4794. doi:10.1007/s00253-020-10585-0.
83. Kleene, S.C. (2016) "Representation of Events in Nerve Nets and Finite Automata," in Shannon, C.E. and McCarthy, J. (eds) *Automata Studies. (AM-34), Volume 34*. Princeton University Press, pp. 3–42. doi:doi:10.1515/9781400882618-002.
84. Kleijnen, J.P.C. (2005) "An overview of the design and analysis of simulation experiments for sensitivity analysis," *European Journal of Operational Research*, 164(2), pp. 287–300. doi:10.1016/j.ejor.2004.02.005.
85. Kleijnen, J.P.C. and Van Beers, W.C.M. (2004) "Application-driven sequential designs for simulation experiments: Kriging metamodelling," *Journal of the Operational Research Society*, 55(8), pp. 876–883. doi:10.1057/palgrave.jors.2601747.
86. Kohls, H., Steffen-Munsberg, F. and Höhne, M. (2014) "Recent achievements in developing the biocatalytic toolbox for chiral amine synthesis," *Current Opinion in Chemical Biology*, 19(1), pp. 180–192. doi:10.1016/j.cbpa.2014.02.021.
87. Kong, F. *et al.* (2012) "Automatic liquid handling for life science: A critical review of the current state of the art," *Journal of Laboratory Automation*, 17(3), pp. 169–185. doi:10.1177/2211068211435302.
88. Koszelewski, D. *et al.* (2010) "Synthesis of optically active amines employing recombinant ω-transaminases in E. coli cells," *ChemCatChem*, 2(1), pp. 73–77. doi:10.1002/cctc.200900220.
89. Kricka, L.J. (1998) "Miniaturization of analytical systems," *Clinical Chemistry*, 44(9), pp. 2008–2014. doi:10.1093/clinchem/44.9.2008.
90. Kuramitsu, S. *et al.* (1990a) "Pre-Steady-State Kinetics of Escherichia coli Aspartate Aminotransferase Catalysed Reactions and Thermodynamic

Aspects of Its Substrate Specificity," *Biochemistry*, 29(23), pp. 5469–5476. doi:10.1021/bi00475a010.
91. Kuramitsu, S. *et al.* (1990b) "Pre-Steady-State Kinetics of Escherichia coli Aspartate Aminotransferase Catalysed Reactions and Thermodynamic Aspects of Its Substrate Specificity," *Biochemistry*, 29(23), pp. 5469–5476. doi:10.1021/bi00475a010.
92. Lambert, J.H. (1760) "Photometry, or, On the measure and gradations of light intensity, colors, and shade.," Augsburg (, p. p.391.
93. Lampinen, J. *et al.* (2012) "Correction Method for Photometric DNA Quantification Assay," *Thermo Fisher Scientific Application Notes*, pp. 1–7.
94. Landeta, C. and Mejia-Santana, A. (2021) "Union is strength: target-based and whole-cell high throughput screens in antibacterial discovery," *Journal of Bacteriology* [Preprint], (November). doi:10.1128/jb.00477-21.
95. Leipold, L. *et al.* (2019) "The identification and use of robust transaminases from a domestic drain metagenome," *Green Chemistry*, 21(1), pp. 75–86. doi:10.1039/c8gc02986e.
96. Lekivetz, R. and Jones, B. (2015) "Fast Flexible Space-Filling Designs for Nonrectangular Regions," *Quality and Reliability Engineering International*, 31(5), pp. 829–837. doi:10.1002/qre.1640.
97. Lendrem, D.W. *et al.* (2015) "Lost in space: Design of experiments and scientific exploration in a Hogarth Universe," *Drug Discovery Today*, 20(11), pp. 1365–1371. doi:10.1016/j.drudis.2015.09.015.
98. Lendrem, D.W. *et al.* (2016) "Teaching examples for the design of experiments: Geographical sensitivity and the self-fulfilling prophecy," *Pharmaceutical Statistics*, 15(1), pp. 90–92. doi:10.1002/pst.1723.
99. Linshiz, G. *et al.* (2013) "PaR-PaR laboratory automation platform," *ACS Synthetic Biology*, 2(5), pp. 216–222. doi:10.1021/sb300075t.
100. Linshiz, G. *et al.* (2014) "PR-PR: Cross-platform laboratory automation system," *ACS Synthetic Biology*, 3(8), pp. 515–524. doi:10.1021/sb4001728.
101. Macaulay, T. (2020) *Spain plans to use robots to test 80,000 people a day for the coronavirus.* Available at: https://thenextweb.com/news/spain-plans-to-use-robots-to-test-80000-people-a-day-for-the-coronavirus (Accessed: December 22, 2021).
102. Mehta, P.K. and Christen, P. (1994) "Homology of 1-aminocyclopropane-1-carboxylate synthase, 8-amino-7-oxononanoate synthase, 2-amino-6-caprolactam racemase, 2,2- dialkylglycine decarboxylase, glutamate-1-semialdehyde 2,1-aminomutase and isopenicillin-N-epimerase with aminotransferases.," *Biochemical and Biophysical Research Communications.*, 194.
103. Mehta, P.K., Hale, T.I. and Christen, P. (1993) "Aminotransferases: demonstration of homology and division into evolutionary subgroups," *European Journal of Biochemistry*, 214(2), pp. 549–561. doi:10.1111/j.1432-1033.1993.tb17953.x.
104. Meng, Q. *et al.* (2020) "Robust ω-Transaminases by Computational Stabilization of the Subunit Interface," *ACS Catalysis*, 10(5), pp. 2915–2928. doi:10.1021/acscatal.9b05223.

105. Mercier, S.M. *et al.* (2014) “Multivariate PAT solutions for biopharmaceutical cultivation: Current progress and limitations,” *Trends in Biotechnology*, 32(6), pp. 329–336. doi:10.1016/j.tibtech.2014.03.008.
106. Meyer, H.-J. *et al.* (2017) “High throughput screening identifies novel, cell cycle-arresting small molecule enhancers of transient protein expression,” *Biotechnology Progress*, pp. 1–10. doi:10.1002/btpr.2517.
107. Midilli, Y.E. and Parshutin, S. (2019) “Review for Optimisation of Neural Networks With Genetic Algorithms and Design of Experiments in Stock Market Prediction,” *Information Technology and Management Science*, 22(December), pp. 15–21. doi:10.7250/itms-2019-0003.
108. Miller, A. (2002) “Subset selection in regression.” Taylor Francis.
109. Montgomery, D.C. (2012) *Statistical Quality Control: A Modern Introduction, 7th Edition International Student Version.*
110. Montgomery, D.C. (2013) *Design and Analysis of Experiments, Eighth Edition.* Eighth Edi, *Copyright*. Eighth Edi. n.
111. Nabifar, A. *et al.* (2010) “A sequential iterative scheme for design of experiments in complex polymerizations,” *Chemical Engineering and Technology*, 33(11), pp. 1814–1824. doi:10.1002/ceat.201000237.
112. Naugler, C. and Church, D.L. (2019) “Automation and artificial intelligence in the clinical laboratory,” *Critical Reviews in Clinical Laboratory Sciences*, 56(2), pp. 98–110. doi:10.1080/10408363.2018.1561640.
113. Nealon, A.J. *et al.* (2005) “Use of operating windows in the assessment of integrated robotic systems for the measurement of bioprocess kinetics,” *Biotechnology Progress*, 21(1), pp. 283–291. doi:10.1021/bp049868+.
114. Onyeogaziri, F.C. and Papaneophytou, C. (2019) “A General Guide for the Optimization of Enzyme Assay Conditions Using the Design of Experiments Approach,” *SLAS Discovery*, 24(5), pp. 587–596. doi:10.1177/2472555219830084.
115. Oritz, L. *et al.* (2017) “Automated Robotic Liquid Handling Assembly of Modular DNA Devices.”
116. Pandya, K. *et al.* (2010) “Strategies to minimize variability and bias associated with manual pipetting in ligand binding assays to assure data quality of protein therapeutic quantification,” *Journal of Pharmaceutical and Biomedical Analysis*, 53(3), pp. 623–630. doi:10.1016/j.jpba.2010.04.025.
117. Parshley, R., Bradshaw, J.T. and Albert, K.J. (2014) “Pipetting Comparison of DMSO Versus Aqueous Solutions as Measured Using Dual- Dye Technology,” 6480, p. 6480.
118. Pauwels, R. *et al.* (1995) “Automated techniques in biotechnology,” *Current Opinion in Biotechnology*, 6(1), pp. 111–117. doi:10.1016/0958-1669(95)80017-4.
119. Piepel, G.F. *et al.* (2019) “Developing a space-filling mixture experiment design when the components are subject to linear and nonlinear constraints,” *Quality Engineering*, 31(3), pp. 463–472. doi:10.1080/08982112.2018.1517887.

120. Politis, S.N. *et al.* (2017a) “Design of experiments (DoE) in pharmaceutical development,” *Drug Development and Industrial Pharmacy*, 43(6), pp. 889–901. doi:10.1080/03639045.2017.1291672.
121. Pronzato, L. and Müller, W.G. (2012) “Design of computer experiments: Space filling and beyond,” *Statistics and Computing*, 22(3), pp. 681–701. doi:10.1007/s11222-011-9242-3.
122. Quaglia, D. *et al.* (2017) “Enzyme engineering: A synthetic biology approach for more effective library generation and automated high-throughput screening,” *PLoS ONE*, 12(2), pp. 1–14. doi:10.1371/journal.pone.0171741.
123. Rasmussen, C.E. and Williams, C.K.I. (2006) *Gaussian Processes for Machine Learning*, *MIT Press*.
124. Rathore, A.S. (2009) “Roadmap for implementation of quality by design (QbD) for biotechnology products,” *Trends in Biotechnology*, 27(9), pp. 546–553. doi:10.1016/j.tibtech.2009.06.006.
125. Rios-Solis, L. *et al.* (2013) “Non-linear kinetic modelling of reversible bioconversions: Application to the transaminase catalyzed synthesis of chiral amino-alcohols,” *Biochemical Engineering Journal*, 73, pp. 38–48. doi:10.1016/j.bej.2013.01.010.
126. Rios-Solis, L. *et al.* (2015) “Modelling and optimisation of the one-pot, multi-enzymatic synthesis of chiral amino-alcohols based on microscale kinetic parameter determination,” *Chemical Engineering Science*, 122(June 2014), pp. 360–372. doi:10.1016/j.ces.2014.09.046.
127. Ronald A. Fisher (1935) “The Design of Experiments.”
128. Roskoski, R. (2014) *Enzyme Assays ☆*, *Reference Module in Biomedical Sciences*. Elsevier Inc. doi:10.1016/b978-0-12-801238-3.05006-6.
129. Sacks, J. *et al.* (1989) “Design and analysis of computer experiments,” *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, 4(4), pp. 118–128. doi:10.2514/6.1998-4757.
130. Sadowski, M.I., Grant, C. and Fell, T.S. (2016) “Harnessing QbD, Programming Languages, and Automation for Reproducible Biology,” *Trends in Biotechnology*, pp. 214–227. doi:10.1016/j.tibtech.2015.11.006.
131. Sancenon, V. *et al.* (2015) “Development, validation and quantitative assessment of an enzymatic assay suitable for small molecule screening and profiling: A case-study,” *Biomolecular Detection and Quantification*, 4, pp. 1–9. doi:10.1016/j.bdq.2015.03.001.
132. Sanders, R. (2020) *UC Berkeley scientists spin up a robotic COVID-19 testing lab.*, *Berkley News*. Available at: https://news.berkeley.edu/2020/03/30/uc-berkeley-scientists-spin-up-a-robotic-covid-19-testing-lab/ (Accessed: December 22, 2021).
133. Santos, C.P., Rato, T.J. and Reis, M.S. (2019) “Design of Experiments: A comparison study from the non-expert user’s perspective,” *Journal of Chemometrics*, 33(1), pp. 1–18. doi:10.1002/cem.3087.
134. Satyawali, Y. *et al.* (2017) “Asymmetric synthesis of chiral amine in organic solvent and in-situ product recovery for process intensification: A case

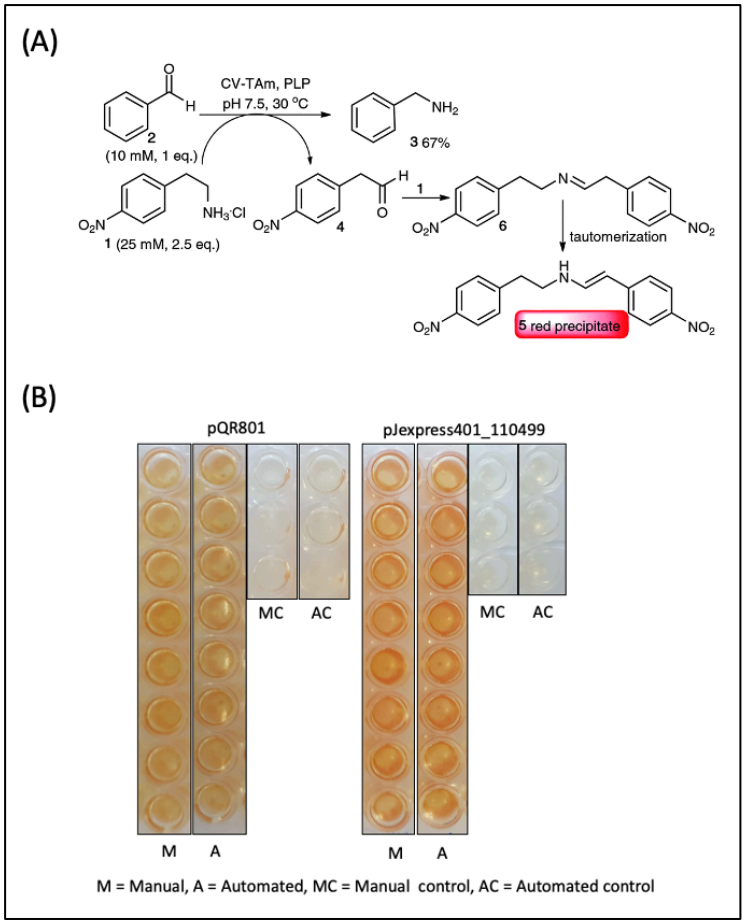
study," *Biochemical Engineering Journal*, 117, pp. 97–104. doi:10.1016/j.bej.2016.11.006.
135. Saville, C.K. *et al.* (2010) "Biocatalytic Asymmetric Synthesis of Sitagliptin Manufacture," *Science*, 329(July), pp. 305–310. doi:10.1126/science.1188934.
136. Sayer, C., Isupov, M.N. and Littlechild, J.A. (2007) "Crystallization and preliminary X-ray diffraction analysis of ω-amino acid:pyruvate transaminase from Chromobacterium violaceum," *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 63(2), pp. 117–119. doi:10.1107/S1744309107000863.
137. Schätzle, S. *et al.* (2009) "Rapid and sensitive kinetic assay for characterization of ω-transaminases," *Analytical Chemistry*, 81(19), pp. 8244–8248. doi:10.1021/ac901640q.
138. Schell, U., Wohlgemuth, R. and Ward, J.M. (2009) "Synthesis of pyridoxamine 5'-phosphate using an MBA:pyruvate transaminase as biocatalyst," *Journal of Molecular Catalysis B: Enzymatic*, 59(4), pp. 279–285. doi:10.1016/j.molcatb.2008.10.005.
139. Schneider, G. (2018) "Automating drug discovery.," *Nature Reviews Drug Discovery*, 17(February).
140. Schulz, E., Speekenbrink, M. and Krause, A. (2018) "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, 85, pp. 1–16. doi:10.1016/j.jmp.2018.03.001.
141. Scott, L. *et al.* (2018) "The application of Design of Experiment and the miniaturise , automate and accelerate assay optimization dragonfly discovery to," p. 300.
142. Sehl, T. *et al.* (2012) "TTC-based screening assay for ω-transaminases: A rapid method to detect reduction of 2-hydroxy ketones," *Journal of Biotechnology*, 159(3), pp. 188–194. doi:https://doi.org/10.1016/j.jbiotec.2011.12.023.
143. Shin, J.S. *et al.* (2003) "Purification, characterization, and molecular cloning of a novel amine:pyruvate transaminase from Vibrio fluvialis JS17," *Applied Microbiology and Biotechnology*, 61(5–6), pp. 463–471. doi:10.1007/s00253-003-1250-6.
144. Shin, J.S. and Kim, B.G. (1997) "Kinetic resolution of α-methylbenzylamine with ω-transaminase screened from soil microorganisms: Application of a biphasic system to overcome product inhibition," *Biotechnology and Bioengineering*, 55(2), pp. 348–358. doi:10.1002/(SICI)1097-0290(19970720)55:2<348::AID-BIT12>3.0.CO;2-D.
145. Shin, J.S. and Kim, B.G. (1998) "Kinetic modelling of ω-transamination for enzymatic kinetic resolution of α-methylbenzylamine," *Biotechnology and Bioengineering*, 60(5), pp. 534–540. doi:10.1002/(SICI)1097-0290(19981205)60:5<534::AID-BIT3>3.0.CO;2-L.
146. Shin, J.S. and Kim, B.G. (1999) "Asymmetric synthesis of chiral amines with ω-transaminase," *Biotechnology and Bioengineering*, 65(2), pp. 206–

211. doi:10.1002/(SICI)1097-0290(19991020)65:2<206::AID-BIT11>3.0.CO;2-9.
147. Shin, J.S. and Kim, B.G. (2001) “Comparison of the ω-transaminases from different microorganisms and application to production of chiral amines,” *Bioscience, Biotechnology and Biochemistry*, 65(8), pp. 1782–1788. doi:10.1271/bbb.65.1782.
148. Siguemoto, É.S. and Gut, J.A.W. (2017) “Validation of spectrophotometric microplate methods for polyphenol oxidase and peroxidase activities analysis in fruits and vegetables,” *Food Science and Technology*, 37, pp. 148–153. doi:10.1590/1678-457X.36216.
149. Singh, M. and Rathi, R. (2019) *A structured review of Lean Six Sigma in various industrial sectors*, *International Journal of Lean Six Sigma*. doi:10.1108/IJLSS-03-2018-0018.
150. Singleton, C. *et al.* (2019) “A design of experiments approach for the rapid formulation of a chemically defined medium for metabolic profiling of industrially important microbes,” *PLoS ONE*, 14(6), pp. 7–11. doi:10.1371/journal.pone.0218208.
151. SPT Labtech (no date) *https://www.sptlabtech.com/products/liquid-handling/dragonfly-discovery/*. Available at: https://www.sptlabtech.com/products/liquid-handling/dragonfly-discovery/.
152. Stepankova, V. *et al.* (2013) “Strategies for stabilization of enzymes in organic solvents,” *ACS Catalysis*, pp. 2823–2836. doi:10.1021/cs400684x.
153. Stewart, J.D. (2001) “Dehydrogenases and transaminases in asymmetric synthesis,” *Current Opinion in Chemical Biology*, 5(2), pp. 120–129. doi:10.1016/S1367-5931(00)00180-0.
154. Sun, M., Lin, J.S. and Barron, A.E. (2011) “Ultrafast, efficient separations of large-sized dsDNA in a blended polymer matrix by microfluidic chip electrophoresis: A design of experiments approach,” *Electrophoresis*, 32(22), pp. 3233–3240. doi:10.1002/elps.201100260.
155. Szymański, P., Markowicz, M. and Mikiciuk-Olasik, E. (2012) “Adaptation of high-throughput screening in drug discovery-toxicological screening tests,” *International Journal of Molecular Sciences*, 13(1), pp. 427–452. doi:10.3390/ijms13010427.
156. Taylor, P.B. *et al.* (2000) “Automated Assay Optimisation with Integrated Statistics and Smart Robotics,” Journal of.
157. Tsoi, J., Patel, V. and Shih, J. (2014) “A practical approach to automate randomized design of experiments for ligand-binding assays,” *Bioanalysis*, 6(5), pp. 705–713. doi:10.4155/bio.13.317.
158. Tye, H. (2004) “Application of statistical ‘design of experiments’ methods in drug discovery,” *Drug Discovery Today*, 9(11), pp. 485–491. doi:10.1016/S1359-6446(04)03086-7.
159. Uhoraningoga, A. *et al.* (2018) “The goldilocks approach: A review of employing design of experiments in prokaryotic recombinant protein production,” *Bioengineering*, 5(4). doi:10.3390/bioengineering5040089.

160. Vasić-Rački, D., Findrik, Z. and Vrsalović Presečki, A. (2011) “Modelling as a tool of enzyme reaction engineering for enzyme reactor development,” *Applied Microbiology and Biotechnology*, 91(4), pp. 845–856. doi:10.1007/s00253-011-3414-0.
161. Voss, M. *et al.* (2018) “*In Silico* Based Engineering Approach to Improve Transaminases for the Conversion of Bulky Substrates,” *ACS Catalysis*, 8(12), pp. 11524–11533. doi:10.1021/acscatal.8b03900.
162. Vrana, J. *et al.* (2021) “Aquarium: open-source laboratory software for design, execution and data management,” *Synthetic Biology*, 6(1), pp. 1–9. doi:10.1093/synbio/ysab006.
163. Wasalathanthri, D.P. *et al.* (2021) “Process analytics 4.0: A paradigm shift in rapid analytics for biologics development,” *Biotechnology Progress*, 37(4), pp. 1–13. doi:10.1002/btpr.3177.
164. Whitehead, E. *et al.* (2018) “Automated Planning Enables Complex Protocols on Liquid-Handling Robots,” *ACS Synthetic Biology*, 7(3), pp. 922–932. doi:10.1021/acssynbio.8b00021.
165. Wu, C.F.J. (2015) “Post-Fisherian Experimentation: From Physical to Virtual,” *Journal of the American Statistical Association*, 110(510), pp. 612–620. doi:10.1080/01621459.2014.914441.
166. Wu, T. and Zhou, Y. (2014) “An Intelligent Automation Platform for Rapid Bioprocess Design,” *Journal of Laboratory Automation*, 19(4), pp. 381–393. doi:10.1177/2211068213499756.
167. Yun, H. *et al.* (2004) “ω-Amino Acid:Pyruvate Transaminase from *Alcaligenes denitrificans* Y2k-2: A New Catalyst for Kinetic Resolution of β-Amino Acids and Amines,” *Applied and Environmental Microbiology*, 70(4), pp. 2529–2534. doi:10.1128/AEM.70.4.2529-2534.2004.
168. Yun, H., Cho, B.K. and Kim, B.G. (2004) “Kinetic resolution of (R,S)-sec-butylamine using omega-transaminase from *Vibrio fluvialis* JS17 under reduced pressure,” *Biotechnology and Bioengineering*, 87(6), pp. 772–778. doi:10.1002/bit.20186.

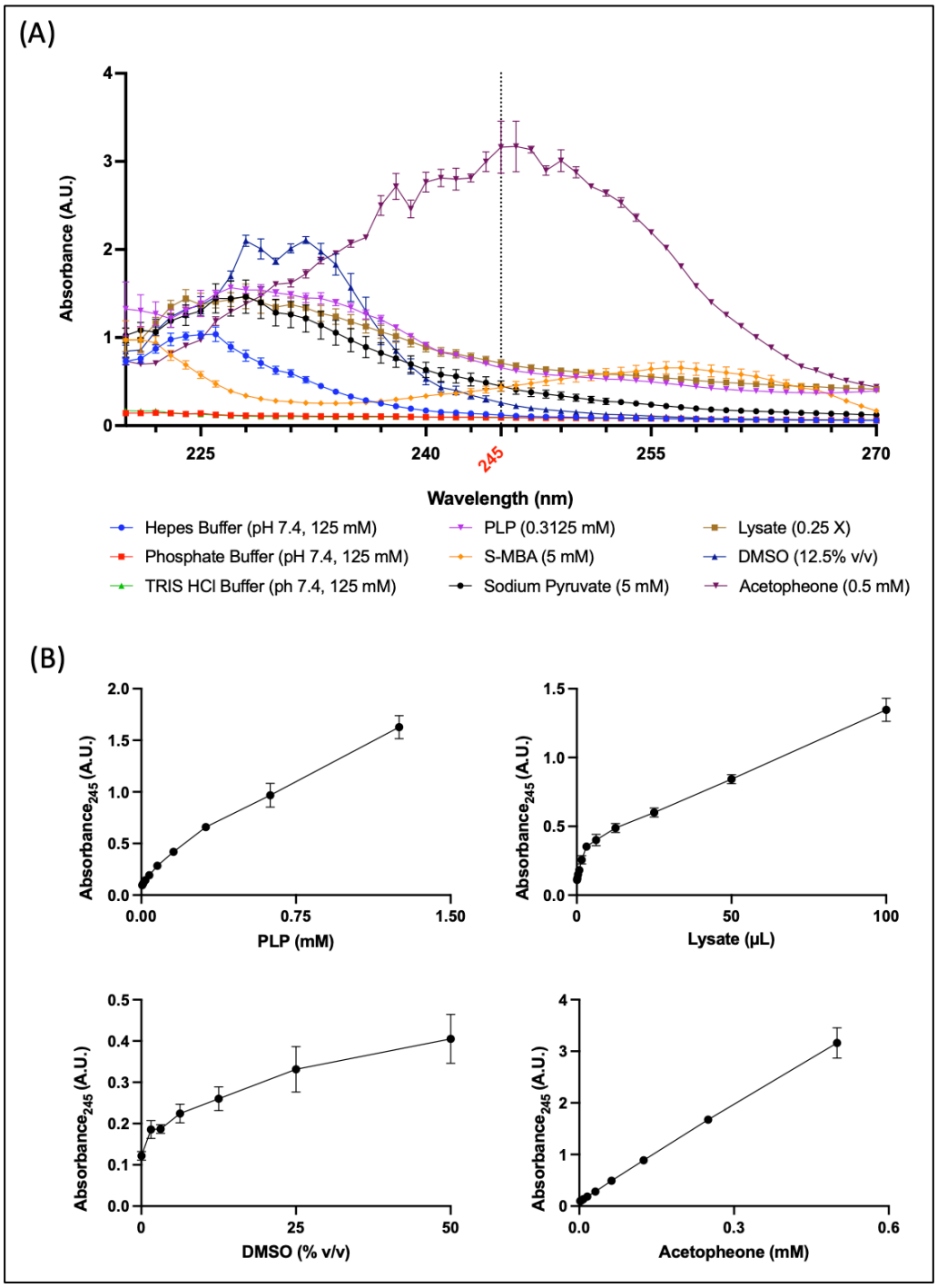
Appendices

Appendix A



Colorimetric assay for CV2025 ω-TAm lysate screening.
(A) Colorimetric assay mechanism. (B) Colorimetric assay comparison of two CV2025 TAm lysates. Assays were prepared manually and using automation and incubated prior to analysis as described in Section 2.3.1. Control reactions contained all reagents except for 2-(4-nitrophenyl) ethan-1-amine, the amine donor. 8 and 3 replicates were prepared for assay reactions and negative controls, respectively.

Appendix B



UV-visible absorption spectra and standard curves of AP assay reagents.
Absorption spectra, from 220 – 270 nm for each reagent considered for the screening DoE and the measured product, Acetophenone. (B) Standard curves for PLP (top left), Lysate (top right), DMSO (bottom left) and Acetophenone (bottom right). Data was generated in triplicate at a final volume of 200 µL as described in Section 2.4.2.1.

Appendix C

Screening DoE design and response (25 °C block).

Run	Buffer Concentration (mM)	Buffer pH	S-MBA (mM)	Sodium pyruvate (mM)	DMSO (X)	Salt Concentration (mM)	Temperature (C)	Volume of lysate (X)	Change in A245 (A.U.)
1	200	9	5	0.5	0	50	37	0.1	0.409
2	100	7	5	0.5	0.1	50	37	0.1	0
3	100	7	5	5	0.1	0	37	0.1	0.042
4	100	7	5	0.5	0	0	37	0.1	0
5	100	9	0.5	0.5	0.1	50	37	0.1	0
6	100	9	0.5	0.5	0.1	50	37	0.1	0
7	200	7	5	5	0.1	50	37	0.1	0
8	100	9	5	5	0.1	50	37	0.1	0.89
9	200	9	5	0.5	0.1	50	37	0.1	0.583
10	200	9	5	5	0.1	0	37	0.1	0.697
11	100	9	0.5	0.5	0	0	37	0.1	0.008
12	200	7	5	0.5	0.1	50	37	0.1	0.035
13	100	9	0.5	5	0.1	0	37	0.1	0
14	200	7	0.5	5	0	50	37	0.1	0
15	200	7	5	0.5	0.1	0	37	0.1	0
16	200	9	0.5	0.5	0.1	0	37	0.1	0.049
17	100	9	0.5	5	0	50	37	0.1	0.027
18	200	9	0.5	5	0	0	37	0.1	0.001
19	100	7	0.5	0.5	0.1	0	37	0.1	0
20	100	9	0.5	5	0.1	0	37	0.1	0.021
21	200	7	0.5	0.5	0.1	50	37	0.1	0
22	100	9	0.5	0.5	0	0	37	0.1	0.021
23	200	9	0.5	5	0.1	50	37	0.1	0
24	200	7	5	5	0	0	37	0.1	0.026
25	200	7	5	5	0	50	37	0.1	0
26	100	7	5	5	0	50	37	0.1	0
27	200	9	0.5	5	0.1	50	37	0.1	0.178
28	200	7	0.5	0.5	0	50	37	0.1	0
29	100	9	5	0.5	0.1	0	37	0.1	0.716
30	100	7	0.5	5	0.1	50	37	0.1	0
31	100	7	0.5	0.5	0	50	37	0.1	0
32	100	9	5	5	0	0	37	0.1	0.439
33	200	7	0.5	5	0.1	0	37	0.1	0
34	200	9	0.5	0.5	0.1	0	37	0.1	0
35	200	9	5	0.5	0	0	37	0.1	0.332
36	100	7	0.5	5	0	50	37	0.1	0
37	200	9	0.5	0.5	0	50	37	0.1	0
38	200	9	5	5	0	50	37	0.1	0.345
39	200	9	5	5	0.1	0	37	0.1	0.665
40	100	7	0.5	5	0	0	37	0.1	0
41	100	9	5	0.5	0	50	37	0.1	0.407
42	200	7	0.5	0.5	0	0	37	0.1	0
43	200	7	5	0.5	0	0	37	0.1	0
44	100	7	5	0.5	0	50	37	0.1	0
45	200	9	0.5	5	0	0	37	0.1	0.019
46	200	9	0.5	5	0.1	0	37	0.25	0.128
47	200	7	0.5	0.5	0.1	0	37	0.25	0.052
48	100	9	0.5	0.5	0	50	37	0.25	0
49	200	9	0.5	0.5	0	50	37	0.25	0.043
50	100	9	0.5	5	0	0	37	0.25	0.081
51	100	9	5	0.5	0	0	37	0.25	0.649
52	200	9	5	0.5	0	50	37	0.25	0.554
53	100	9	5	0.5	0	50	37	0.25	0.561
54	200	7	0.5	5	0	0	37	0.25	0.108
55	200	7	0.5	5	0	50	37	0.25	0
56	100	7	0.5	5	0.1	0	37	0.25	0
57	200	7	5	5	0.1	50	37	0.25	0.136
58	200	7	5	5	0.1	0	37	0.25	0.166
59	100	9	0.5	0.5	0.1	0	37	0.25	0.107
60	100	7	5	0.5	0.1	50	37	0.25	0
61	200	9	0.5	5	0	50	37	0.25	0.149
62	200	9	5	5	0.1	50	37	0.25	0
63	200	7	5	0.5	0	0	37	0.25	0
64	100	9	5	0.5	0.1	50	37	0.25	0
65	100	7	5	5	0.1	50	37	0.25	0.156
66	100	9	0.5	5	0.1	50	37	0.25	0.03
67	200	7	5	0.5	0.1	50	37	0.25	0.838
68	100	7	5	0.5	0	0	37	0.25	0.052
69	200	9	0.5	0.5	0	0	37	0.25	0.198
70	100	7	0.5	0.5	0	0	37	0.25	0
71	100	9	5	5	0.1	0	37	0.25	0.721
72	200	9	0.5	0.5	0.1	50	37	0.25	0.095
73	200	7	5	0.5	0	50	37	0.25	0
74	100	7	0.5	0.5	0	50	37	0.25	0
75	200	9	5	0.5	0.1	0	37	0.25	0.658
76	200	9	5	5	0	0	37	0.25	0.764
77	100	7	5	5	0	0	37	0.25	0.068
78	100	7	5	5	0	50	37	0.25	0.084
79	100	7	0.5	5	0.1	0	37	0.25	0
80	200	7	5	5	0	0	37	0.25	0
81	200	9	5	0.5	0.1	0	37	0.25	0.776
82	100	7	0.5	0.5	0.1	50	37	0.25	0.033
83	100	9	0.5	5	0	0	37	0.25	0.074
84	200	7	0.5	5	0.1	50	37	0.25	0
85	100	7	5	0.5	0.1	0	37	0.25	0.128
86	200	9	0.5	5	0.1	0	37	0.25	0
87	100	7	0.5	5	0	50	37	0.25	0
88	100	9	5	5	0.1	0	37	0.25	0
89	200	7	0.5	0.5	0.1	50	37	0.25	0.068
90	100	9	5	5	0	50	37	0.25	0.795
91	200	9	5	5	0	50	37	0.25	0.653
92	200	7	0.5	0.5	0	0	37	0.25	0.055
93	100	9	0.5	5	0.1	50	37	0.25	0.179
94	150	8	2.75	2.75	0.05	25	37	0.175	0.088
95	150	8	2.75	2.75	0.05	25	37	0.175	0
96	150	8	2.75	2.75	0.05	25	37	0.175	0.155

Appendix D

Screening DoE design and response (37 °C block).

Run	Buffer Concentration (mM)	Buffer pH	S-MBA (mM)	Sodium pyruvate (mM)	DMSO (X)	Salt Concentration (mM)	Temperature (C)	Volume of lysate (X)	Change in A245 (A.U.)
97	200	7	5	0.5	0	50	25	0.1	0
98	200	7	0.5	5	0	0	25	0.1	0
99	100	7	5	5	0	50	25	0.1	0
100	200	9	5	0.5	0.1	50	25	0.1	0.334
101	200	7	0.5	0.5	0.1	0	25	0.1	0
102	100	9	5	0.5	0	0	25	0.1	0.352
103	200	9	0.5	0.5	0.1	50	25	0.1	0
104	100	7	5	5	0	0	25	0.1	0.036
105	200	9	5	0.5	0.1	0	25	0.1	0.458
106	200	9	5	0.5	0	0	25	0.1	0.24
107	200	9	0.5	5	0.1	0	25	0.1	0
108	100	7	0.5	5	0.1	0	25	0.1	0
109	200	9	0.5	5	0	50	25	0.1	0
110	200	9	0.5	0.5	0	50	25	0.1	0
111	100	9	5	5	0.1	0	25	0.1	0.545
112	200	9	0.5	0.5	0	0	25	0.1	0
113	200	7	5	5	0.1	50	25	0.1	0.012
114	100	9	0.5	0.5	0	50	25	0.1	0
115	100	9	0.5	5	0.1	50	25	0.1	0.021
116	100	7	0.5	5	0.1	50	25	0.1	0.021
117	200	7	0.5	5	0	50	25	0.1	0.091
118	100	9	0.5	5	0	0	25	0.1	0.017
119	100	9	0.5	0.5	0.1	0	25	0.1	0
120	200	7	0.5	0.5	0	0	25	0.1	0
121	200	7	5	5	0	0	25	0.1	0
122	100	9	0.5	5	0	50	25	0.1	0
123	100	7	5	0.5	0.1	50	25	0.1	0
124	200	9	0.5	5	0.1	0	25	0.1	0.086
125	100	7	5	0.5	0.1	0	25	0.1	0
126	100	7	0.5	0.5	0	0	25	0.1	0.009
127	100	7	0.5	5	0	0	25	0.1	0
128	200	7	5	0.5	0.1	0	25	0.1	0.014
129	100	9	5	0.5	0	50	25	0.1	0.232
130	200	9	5	5	0	50	25	0.1	0.312
131	100	7	0.5	0.5	0	50	25	0.1	0
132	100	7	5	5	0.1	50	25	0.1	0.021
133	200	7	0.5	5	0.1	50	25	0.1	0.073
134	200	9	5	5	0	0	25	0.1	0.324
135	100	9	5	5	0	50	25	0.1	0.265
136	100	7	5	0.5	0	0	25	0.1	0
137	100	9	5	0.5	0.1	50	25	0.1	0.634
138	200	7	5	0.5	0	50	25	0.1	0
139	200	9	5	5	0.1	50	25	0.1	0.64
140	100	7	0.5	0.5	0.1	50	25	0.1	0
141	200	7	5	5	0.1	0	25	0.1	0.045
142	200	7	0.5	0.5	0.1	50	25	0.1	0
143	100	7	0.5	0.5	0	0	25	0.25	0
144	200	9	0.5	0.5	0.1	50	25	0.25	0.102

Run	Buffer Concentration (mM)	Buffer pH	S-MBA (mM)	Sodium pyruvate (mM)	DMSO (X)	Salt Concentration (mM)	Temperature (C)	Volume of lysate (X)	Change in A245 (A.U.)
145	200	7	5	0.5	0.1	50	25	0.25	0.075
146	200	7	5	0.5	0	0	25	0.25	0.015
147	100	7	5	0.5	0.1	0	25	0.25	0.095
148	200	7	0.5	5	0.1	50	25	0.25	0.027
149	100	9	5	0.5	0	0	25	0.25	0.475
150	200	9	0.5	0.5	0	0	25	0.25	0.083
151	100	9	0.5	5	0	50	25	0.25	0.064
152	200	9	5	5	0.1	50	25	0.25	0.679
153	100	7	0.5	0.5	0.1	0	25	0.25	0.016
154	100	7	0.5	5	0	0	25	0.25	0.007
155	200	9	5	0.5	0	50	25	0.25	0.445
156	100	9	0.5	5	0.1	0	25	0.25	0.041
157	100	9	5	5	0	0	25	0.25	0.542
158	200	7	5	5	0	50	25	0.25	0.049
159	100	9	5	5	0.1	50	25	0.25	0.873
160	200	7	5	5	0.1	0	25	0.25	0.07
161	100	7	5	5	0.1	50	25	0.25	0.054
162	200	7	0.5	5	0	0	25	0.25	0.036
163	200	9	5	5	0.1	0	25	0.25	0.76
164	200	7	0.5	0.5	0	50	25	0.25	0.056
165	200	9	0.5	0.5	0.1	0	25	0.25	0.133
166	200	9	0.5	5	0	0	25	0.25	0.078
167	200	9	0.5	5	0.1	50	25	0.25	0.057
168	100	7	0.5	5	0.1	50	25	0.25	0.027
169	200	9	5	0.5	0	0	25	0.25	0.369
170	100	9	0.5	0.5	0	0	25	0.25	0.045
171	200	7	0.5	0.5	0.1	0	25	0.25	0
172	100	9	0.5	5	0.1	0	25	0.25	0.133
173	100	7	5	0.5	0	50	25	0.25	0.047
174	100	9	5	0.5	0.1	0	25	0.25	0.792
175	100	7	5	5	0	0	25	0.25	0.06
176	100	7	0.5	5	0	50	25	0.25	0
177	200	9	0.5	5	0	50	25	0.25	0.101
178	200	7	0.5	5	0.1	0	25	0.25	0.011
179	200	9	5	5	0	0	25	0.25	0.495
180	200	7	5	0.5	0.1	0	25	0.25	0.058
181	100	9	0.5	0.5	0.1	50	25	0.25	0.124
182	200	9	5	0.5	0.1	50	25	0.25	0.579
183	100	7	5	5	0.1	0	25	0.25	0.096
184	100	7	0.5	0.5	0.1	50	25	0.25	0
185	100	9	5	5	0	50	25	0.25	0.542
186	200	7	0.5	0.5	0	50	25	0.25	0.036
187	100	9	0.5	0.5	0	50	25	0.25	0.021
188	200	7	5	5	0	50	25	0.25	0.025
189	100	9	5	0.5	0.1	50	25	0.25	0.785
190	150	8	2.75	2.75	0.05	25	25	0.175	0.063
191	150	8	2.75	2.75	0.05	25	25	0.175	0.105
192	150	8	2.75	2.75	0.05	25	25	0.175	0.102

Appendix E

Optimisation DoE design and response.

Design								Dragonfly dispenser		PipetMax liquid handler	
Run	Buffer Concentration (mM)	Buffer	DMSO (X)	MBA (mM)	Sodium Pyruvate (mM)	Temperature (C)	Lysate (X)	Average ΔAbsorbance	Standard deviation of Average ΔAbsorbance	Average ΔAbsorbance	Standard deviation of Average ΔAbsorbance
1	100mM	Tris 9	0	6	5	21.5	0.1250	1.8823	0.0361	1.5280	0.0095
2	100mM	Tris 8.5	0.3	5	6	21.5	0.1250	0.5477	0.0445	0.3163	0.0106
3	100mM	Tris 8.5	0.15	5	5	21.5	0.1250	1.1457	0.0661	0.8043	0.0152
4	100mM	Tris 8	0.3	6	5	21.5	0.1250	0.3383	0.0261	0.1100	0.0271
5	100mM	Tris 8	0	4	4	21.5	0.1250	0.7530	0.0277	0.4050	0.0078
6	100mM	Tris 8.5	0.15	5	5	21.5	0.1250	1.2127	0.0342	0.8333	0.0057
7	100mM	Tris 9	0.15	6	4	21.5	0.1250	1.6467	0.0774	1.3113	0.0061
8	100mM	Tris 9	0.15	5	6	21.5	0.1250	1.5980	0.1145	1.3130	0.0231
9	100mM	Tris 8.5	0.15	5	5	21.5	0.1250	1.1373	0.0223	0.7737	0.0006
10	100mM	Tris 9	0.3	4	4	21.5	0.1250	0.7100	0.0218	0.6267	0.0065
11	100mM	Tris 8	0.15	5	5	21.5	0.1250	0.6833	0.0287	0.4927	0.0205
12	100mM	Tris 9	0.3	6	4	21.5	0.1250	0.8403	0.0469	0.8243	0.0417
13	100mM	Tris 9	0.3	4	6	21.5	0.1250	0.7203	0.0595	0.6393	0.0291
14	100mM	Tris 8.5	0.3	5	5	21.5	0.1250	0.6243	0.2155	0.4753	0.0081
15	100mM	Tris 9	0	5	4	21.5	0.1250	1.8333	0.0190	1.4030	0.0079
16	100mM	Tris 8.5	0.15	5	5	21.5	0.1250	1.1667	0.0606	0.7210	0.2461
17	100mM	Tris 8.5	0	6	6	21.5	0.1250	1.5517	0.0196	0.9247	0.0165
18	100mM	Tris 8	0.15	6	4	21.5	0.1250	0.7583	0.0245	0.5600	0.0095
19	100mM	Tris 8.5	0.15	4	5	21.5	0.1250	1.0410	0.0192	0.3870	0.5041
20	100mM	Tris 8	0	5	4	21.5	0.1250	0.9443	0.0179	0.4647	0.0162
21	100mM	Tris 9	0.15	4	5	21.5	0.1250	1.4803	0.0341	0.8490	0.2697
22	100mM	Tris 9	0	6	5	21.5	0.1250	1.8670	0.0579	1.3750	0.0406
23	100mM	Tris 8.5	0.15	4	4	21.5	0.1250	1.1247	0.0648	0.9220	0.0078
24	100mM	Tris 8	0.3	4	6	21.5	0.1250	0.2373	0.0059	0.1213	0.1221
25	100mM	Tris 8	0.15	5	6	21.5	0.1250	0.6807	0.0250	0.4333	0.0278
26	100mM	Tris 8.5	0.3	6	4	21.5	0.1250	0.6380	0.0226	0.5470	0.0855
27	100mM	Tris 8.5	0.15	5	5	21.5	0.1250	1.2473	0.0779	0.5870	0.0233
28	100mM	Tris 8.5	0	5	6	21.5	0.1250	1.4287	0.0040	0.4580	0.0255
29	100mM	Tris 8.5	0.15	6	5	18	0.1250	0.8753	0.0133	0.9177	0.0301
30	100mM	Tris 8	0	6	6	18	0.1250	0.6553	0.0093	0.4690	0.0079
31	100mM	Tris 8	0.3	4	6	18	0.1250	0.0220	0.0327	0.1673	0.0182
32	100mM	Tris 8.5	0	6	4	18	0.1250	1.1883	0.0151	1.0020	0.0050
33	100mM	Tris 9	0.3	6	6	18	0.1250	0.4917	0.0439	0.7680	0.0386
34	100mM	Tris 9	0.15	5	4	18	0.1250	1.1537	0.0340	1.4253	0.0414
35	100mM	Tris 8.5	0.3	4	4	18	0.1250	0.2170	0.1805	0.2930	0.0040
36	100mM	Tris 8	0.3	5	4	18	0.1250	0.1370	0.0310	0.0700	0.0282
37	100mM	Tris 9	0	4	6	18	0.1250	1.2487	0.0057	1.3547	0.0029
38	100mM	Tris 8	0	4	5	18	0.1250	0.4637	0.0153	0.3230	0.0115
39	100mM	Tris 8	0	4	6	25	0.1250	0.6547	0.0050	0.2963	0.0146
40	100mM	Tris 8.5	0.15	5	5	25	0.1250	1.2110	0.0450	0.7073	0.0146
41	100mM	Tris 9	0	4	6	25	0.1250	1.6257	0.0258	1.0920	0.0053
42	100mM	Tris 8.5	0.15	6	4	25	0.1250	1.5087	0.2367	0.7837	0.0136
43	100mM	Tris 9	0.3	5	5	25	0.1250	0.8770	0.0714	0.7767	0.0180
44	100mM	Tris 9	0	4	4	25	0.1250	1.6333	0.0152	1.0840	0.0270
45	100mM	Tris 8	0	6	5	25	0.1250	1.0263	0.0111	0.5250	0.0082
46	100mM	Tris 8	0.3	4	4	25	0.1250	0.1417	0.0289	0.1083	0.0330
47	100mM	Tris 8	0.3	6	6	25	0.1250	0.2653	0.1515	0.2020	0.0040
48	100mM	Tris 9	0.15	6	6	25	0.1250	1.5860	0.0108	1.3913	0.0160

Appendix F

Stepwise regression model prediction formula for predicting ΔAbsorbance generated from the analysis of the screening DoE study.

$$
\begin{aligned}
&-1.783416561 \\
&+ 0.0001412466 \cdot \text{Buffer Concentration} \\
&+ 0.1851250698 \cdot \text{pH} \\
&+ 0.0512562055 \cdot \text{S-MBA} \\
&+ 0.0056004927 \cdot \text{Sodium pyruvate} \\
&+ 0.608267736 \cdot \text{DMSO} \\
&+ 0.0059128078 \cdot \text{Temperature} \\
&+ 0.5321849692 \cdot \text{Volume of clarified cell lysate} \\
&+ (\text{Buffer Concentration} - 151.87165775) \cdot ((\text{pH} - 7.8401069519) \cdot -0.000488745) \\
&+ (\text{Buffer Concentration} - 151.87165775) \cdot ((\text{Temperature} - 30.839572193) \cdot 0.0000277905) \\
&+ (\text{pH} - 7.8401069519) \cdot ((\text{S-MBA} - 2.6417112299) \cdot 0.0745274575) \\
&+ (\text{pH} - 7.8401069519) \cdot ((\text{DMSO} - 0.0481283422) \cdot 0.5763989954) \\
&+ (\text{pH} - 7.8401069519) \cdot ((\text{Temperature} - 30.839572193) \cdot 0.0039936383) \\
&+ (\text{pH} - 7.8401069519) \cdot ((\text{Volume of clarified cell lysate} - 0.1754010695) \cdot 0.3917550249) \\
&+ (\text{S-MBA} - 2.6417112299) \cdot ((\text{Sodium pyruvate} - 2.7620320856) \cdot 0.0019390271) \\
&+ (\text{S-MBA} - 2.6417112299) \cdot ((\text{DMSO} - 0.0481283422) \cdot 0.239911563) \\
&+ (\text{S-MBA} - 2.6417112299) \cdot ((\text{Temperature} - 30.839572193) \cdot 0.0024218655) \\
&+ (\text{S-MBA} - 2.6417112299) \cdot ((\text{Volume of clarified cell lysate} - 0.1754010695) \cdot 0.1015796932)
\end{aligned}
$$

Appendix G

Stepwise regression model prediction formula for predicting ΔAbsorbance generated from the analysis of the optimisation DoE study.

(A) Optimisation DoE study using the Dragonfly dispenser

$$
\begin{aligned}
&-6.061028277 \\
&+ 0.7232101065 \bullet \text{pH} \\
&+ -2.631419513 \bullet \text{DMSO} \\
&+ 0.096027523 \bullet \text{MBA} \\
&+ 0.0474229335 \bullet \text{Temperature} \\
&+ (\text{pH} - 8.5104166667) \bullet ((\text{DMSO} - 0.15) \bullet -1.244110462) \\
&+ (\text{pH} - 8.5104166667) \bullet ((\text{Temperature} - 21.5) \bullet 0.0267909479) \\
&+ (\text{DMSO} - 0.15) \bullet ((\text{MBA} - 5.0208333333) \bullet -0.29627049) \\
&+ (\text{MBA} - 5.0208333333) \bullet ((\text{Temperature} - 21.5) \bullet 0.0115024286) \\
&+ (\text{pH} - 8.5104166667) \bullet ((\text{pH} - 8.5104166667) \bullet -0.280539022) \\
&+ (\text{DMSO} - 0.15) \bullet ((\text{DMSO} - 0.15) \bullet -8.638486719) \\
&+ (\text{Temperature} - 21.5) \bullet ((\text{Temperature} - 21.5) \bullet -0.012237235)
\end{aligned}
$$

(B) Optimisation DoE study using the PipetMax liquid-handler

$$
\begin{aligned}
&-5.991996565 \\
&+ 0.7561908188 \bullet \text{pH} \\
&+ -1.37999104 \bullet \text{DMSO} \\
&+ 0.1041686745 \bullet \text{MBA} \\
&+ (\text{pH} - 8.5104166667) \bullet ((\text{DMSO} - 0.15) \bullet -1.027493056) \\
&+ (\text{pH} - 8.5104166667) \bullet ((\text{pH} - 8.5104166667) \bullet 0.3826168568) \\
&+ (\text{DMSO} - 0.15) \bullet ((\text{DMSO} - 0.15) \bullet -8.198784077)
\end{aligned}
$$

Appendix H

Liquid handling policy	**Definition**
Mega Mix	10 post-mixes of the sample being transferred. No tip reuse permitted.
Need to Mix	3 pre-mixes and 3 post-mixes of the sample being transferred. No tip reuse permitted.
Post Mix	3 post-mixes of the sample being transferred. No tip reuse permitted.
Smart Mix	3 post-mixes of the sample being transferred. Volume is adjusted based upon the volume of liquid in the destination well. No tip reuse permitted.

Appendix I

(A)

	1	2	3	4	5	6	7	8	9	10	11	12
A	3.48	2.15	2.17	1.62	2.19	2.77	2.03	3.06	2.91	3.96	5.90	4.73
B	2.69	2.27	2.33	2.87	1.63	4.18	1.49	3.97	4.86	4.21	6.27	5.05
C	3.07	2.09	1.62	1.99	1.46	1.45	0.74	2.33	3.70	0.37	5.94	5.22
D	4.35	2.50	3.24	2.17	1.82	2.53	13.13	2.89	3.34	4.20	4.18	4.23
E	3.78	2.65	3.09	2.37	2.01	1.62	1.88	1.81	4.10	4.35	4.95	4.46
F	4.14	0.54	1.99	2.00	1.99	1.97	1.50	2.35	2.03	3.06	6.02	4.53
G	5.83	4.46	3.95	3.44	2.73	1.96	2.64	2.90	4.43	4.85	5.27	4.86
H	5.38	4.86	4.50	4.24	5.27	4.13	5.39	4.74	4.79	5.20	0.93	8.12

(B)

	1	2	3	4	5	6	7	8	9	10	11	12
A	8.18	7.13	9.02	13.88	10.44	10.80	13.92	13.61	10.99	8.75	12.09	7.92
B	6.87	6.58	6.39	8.51	11.46	11.24	12.23	10.20	14.34	10.94	12.43	12.22
C	6.95	8.84	11.66	11.91	11.79	12.71	15.22	13.14	13.51	11.92	12.13	7.91
D	10.15	5.36	8.21	8.14	11.19	13.18	13.84	14.18	13.49	13.37	13.21	10.79
E	5.95	4.99	10.66	6.82	11.41	12.24	14.08	13.42	13.72	12.87	7.90	9.93
F	9.71	9.51	11.45	6.57	11.09	12.07	12.80	12.01	13.36	9.06	9.98	9.28
G	8.69	7.09	5.25	6.84	4.93	8.55	9.98	10.17	0.00	8.30	7.85	6.72
H	3.70	3.95	3.44	2.20	1.27	2.92	2.94	2.39	1.64	1.30	2.54	2.81

Impact of plate film type on evaporation observed in a 96-well plate. (A) Plate sealed with MicroAmp film. (B) Plate sealed with Aluminum film. Percentage of evaporation in each well after incubation of 70 µL of Tartrazine (0.1 mM) at 45 °C for 24 hours is shown in each well location. Green to red shading indicates increasing evaporation.

(A)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	1.49	1.12	-0.80	2.10	0.55	1.57	1.73	-0.20	3.21	-1.53	-1.51	4.17	-0.40	3.58	5.08	1.66	1.32	1.39	5.76	0.55	0.00	-0.97	2.89	-1.61
B	1.92	1.36	0.00	-0.80	-1.17	0.57	-1.39	1.11	-0.95	0.80	0.95	1.13	1.49	0.97	0.56	1.59	1.54	2.28	1.20	1.93	1.51	1.77	1.30	2.66
C	0.18	0.77	0.81	0.96	0.00	0.99	0.00	-1.87	-0.61	0.77	0.00	1.59	1.26	2.27	1.20	1.47	2.27	1.77	3.07	1.86	3.14	1.56	1.64	2.79
D	0.00	2.14	1.13	1.78	-0.39	0.94	-1.57	-0.93	-0.38	0.59	1.51	1.90	1.08	2.33	0.56	2.17	1.18	1.53	3.55	1.73	0.77	1.39	2.59	1.74
E	-2.61	0.00	0.00	0.38	0.37	-0.20	0.19	-0.41	0.40	0.19	1.17	2.77	1.95	2.10	-0.59	2.37	2.69	2.33	1.75	1.88	1.35	2.92	1.83	3.93
F	-1.13	0.57	1.36	0.20	-0.58	0.75	0.58	0.00	0.00	-0.20	1.34	1.15	2.58	2.49	1.88	2.54	2.10	2.69	1.77	2.73	3.07	1.96	2.01	2.14
G	1.10	1.59	1.93	0.38	0.95	0.19	0.78	0.63	-0.20	1.54	0.79	1.00	3.26	1.15	1.59	2.00	3.30	3.68	2.33	3.04	3.11	2.53	1.89	1.80
H	0.75	4.14	1.76	0.98	-0.79	0.00	1.17	0.93	0.57	0.00	2.34	1.92	0.60	1.93	0.76	0.99	1.73	1.58	2.36	2.50	2.50	4.21	3.02	-0.39
I	-2.94	3.83	1.92	0.39	0.78	1.77	0.00	0.44	0.82	0.20	2.00	2.19	2.17	1.95	3.20	1.10	1.97	1.37	2.36	3.52	2.16	2.95	2.50	1.38
J	0.74	1.35	0.58	0.78	0.77	1.33	1.36	0.74	0.56	1.14	2.12	1.51	0.39	0.58	0.78	3.09	1.34	0.98	2.54	3.53	2.49	2.37	2.25	1.36
K	-0.37	-0.40	1.52	-0.39	0.39	1.36	1.36	1.08	0.80	0.00	0.80	0.39	0.90	1.74	-2.38	1.29	1.58	1.34	1.95	1.95	1.75	2.95	2.06	3.24
L	-1.85	0.77	1.59	0.97	0.39	1.72	0.19	0.74	0.38	0.76	1.16	2.49	0.87	-1.17	0.97	1.72	2.72	2.16	2.33	1.57	1.54	1.57	1.64	3.60
M	5.64	1.80	0.75	0.77	0.59	2.14	1.35	0.43	0.96	0.76	1.01	2.15	1.38	1.18	0.40	1.81	1.80	1.88	1.95	1.96	0.97	2.94	2.64	3.06
N	8.38	1.73	-0.99	1.16	0.78	0.77	1.16	0.00	0.77	0.91	1.72	1.56	1.37	1.57	0.39	2.46	0.97	2.38	3.08	2.37	2.51	1.19	3.10	2.83
O	-0.54	1.60	0.93	1.16	0.00	0.77	1.74	0.67	1.92	0.38	1.81	1.15	1.57	1.93	2.15	3.08	3.98	1.12	2.14	2.36	2.70	1.95	1.31	2.26
P	0.36	1.13	1.19	0.57	5.80	1.73	1.54	3.28	1.75	0.92	3.61	0.97	4.41	1.35	1.15	2.59	4.54	1.18	5.29	-4.86	1.16	-0.39	1.43	2.40

(B)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	20.53	2.82	-0.39	4.41	4.31	3.74	3.82	4.73	4.18	3.11	0.56	1.19	4.62	0.72	3.25	4.26	-1.32	1.80	4.60	3.14	2.97	2.88	-0.77	-0.40
B	-1.20	5.21	7.89	2.78	4.29	3.41	4.24	7.41	6.06	6.02	6.50	5.64	6.55	4.75	4.37	5.72	3.47	4.20	3.77	2.86	2.37	2.88	2.92	2.86
C	-0.57	5.92	4.35	2.33	4.78	3.75	4.20	7.77	5.93	5.96	6.19	4.54	4.71	0.95	6.02	6.18	5.49	4.30	4.58	4.72	2.76	2.35	1.95	3.60
D	4.37	3.70	7.01	4.35	3.27	3.40	5.93	5.47	6.67	5.95	6.15	6.74	4.80	4.74	6.13	5.50	5.11	3.89	5.89	4.41	4.22	3.66	2.73	0.78
E	4.96	4.99	3.16	4.82	4.15	3.56	4.86	4.65	6.23	6.55	5.88	4.57	6.38	5.88	5.80	5.58	3.47	4.35	4.43	3.60	3.39	3.93	3.90	0.97
F	-0.78	4.21	5.65	4.11	4.04	4.78	7.09	6.51	5.57	5.59	6.95	4.40	5.93	5.98	4.31	6.38	5.61	5.74	3.73	2.86	3.34	4.55	3.49	4.54
G	-2.50	1.99	5.08	4.12	4.55	4.01	4.73	6.17	6.53	6.76	7.43	6.45	5.85	5.73	6.14	5.02	3.52	8.56	5.58	5.28	4.33	3.21	4.49	4.23
H	-1.93	0.79	5.31	3.54	4.66	3.49	5.43	5.59	4.90	5.52	6.29	5.58	4.39	5.60	4.56	4.45	3.95	5.15	4.97	5.61	3.56	4.31	-0.58	-2.95
I	-3.26	2.64	5.64	5.87	6.12	0.80	5.11	5.13	6.33	6.18	6.41	6.39	2.96	5.93	6.55	5.29	3.94	3.39	4.76	4.67	5.39	3.74	3.64	-0.58
J	3.05	5.27	3.55	4.14	5.63	3.69	4.82	5.38	4.36	5.34	8.22	5.25	4.22	4.70	5.72	5.53	4.11	5.34	3.94	2.93	4.93	3.28	7.51	-0.79
K	3.45	4.17	5.19	5.62	3.83	5.34	4.34	5.56	5.09	6.59	6.76	5.91	4.31	3.60	4.84	4.02	5.33	6.20	4.30	3.28	4.88	4.25	4.20	3.57
L	3.38	4.05	0.38	2.72	4.83	3.47	4.77	6.35	6.26	7.00	5.51	6.18	4.81	4.69	5.33	4.34	4.47	4.79	4.86	4.49	3.45	5.07	4.39	0.60
M	0.38	5.20	5.20	5.48	4.25	4.36	4.57	4.95	4.55	6.32	7.57	6.05	4.49	3.85	5.36	4.59	4.63	4.99	4.31	4.03	2.99	3.08	3.41	2.84
N	-0.93	4.28	4.84	6.47	4.29	4.31	4.92	4.77	6.17	6.53	7.62	6.95	5.60	5.20	4.22	5.21	3.73	4.81	4.65	3.11	3.29	2.87	4.58	3.98
O	4.91	3.82	3.63	4.37	5.24	4.95	3.61	7.55	5.11	4.96	5.47	4.72	3.69	3.80	4.53	5.28	5.03	5.12	2.96	4.02	3.93	3.76	3.64	7.43
P	-0.91	1.72	0.19	5.21	3.08	0.18	2.97	0.38	3.84	3.61	1.17	0.77	5.29	0.18	2.79	1.64	-0.38	0.39	0.19	0.19	2.98	0.73	0.95	0.99

Impact of plate film type on evaporation observed in a 384-well plate. (A) Plate sealed with MicroAmp film. (B) Plate sealed with Aluminum film. Percentage of evaporation in each well after incubation of 20 µL of Tartrazine (0.1 mM) at 45 °C for 24 hours is shown in each well location. Green to red shading indicates increasing evaporation.

Appendix J

Space-filling DoE design file

25 °C

Run	Enzyme (µl)	MBA (mM)	Pyruavte (mM)	PLP (mM)	Sodium Alanine (mM)	DMSO (X)	Buffer concentration (mM)	Salt concentration (mM)	Temperature (°C)	Salt type	Buffer pH
101	2.8969	0.3818	1.8244	0.0160	2.4483	0.0215	142.0881	115.3020	25	NaCl	TRIS 9
22	2.6223	4.5526	2.8570	0.0339	0.9913	0.3590	84.6205	191.1187	25	NaCl	TRIS 7
102	2.1881	1.9052	1.1484	0.0329	0.8207	0.1961	113.9426	41.4521	25	KCl	TRIS 9
129	1.8038	1.4962	4.4167	0.0088	3.4204	0.2941	154.3674	0.8733	25	NaCl	CHES 10
108	1.9507	3.4826	1.3348	0.0020	2.0019	0.1006	172.8807	76.0465	25	KCl	CHES 10
30	1.3396	0.3086	0.9069	0.0301	2.1372	0.3228	67.2071	66.8110	25	KCl	TRIS 9
112	2.4525	1.1981	0.6771	0.0004	1.4381	0.2171	159.9339	189.2124	25	NaCl	TRIS 7
132	2.5772	0.5534	2.1658	0.0025	3.7885	0.1462	97.3595	3.2805	25	NaCl	CHES 9
13	1.8882	4.7975	4.2058	0.0192	0.8819	0.0894	199.8283	46.1353	25	KCl	TRIS 8
11	2.9137	1.9259	2.8012	0.0251	2.0409	0.0509	58.8822	11.6594	25	NaCl	TRIS 8
43	2.9877	4.6924	1.9871	0.0099	4.3674	0.1504	175.6402	81.8344	25	NaCl	TRIS 7
35	1.4215	3.3187	0.3278	0.0199	0.2847	0.3957	62.0460	173.3272	25	KCl	CHES 10
38	1.7229	2.5101	0.5376	0.0306	3.1832	0.0457	153.0289	80.3886	25	KCl	CHES 9
24	1.9402	4.3424	4.7202	0.0327	3.0107	0.2718	52.4724	114.5129	25	NaCl	TRIS 8
85	2.6003	2.8546	4.0085	0.0144	1.1401	0.1587	61.7526	64.7566	25	NaCl	TRIS 8
88	1.2910	3.8212	0.4383	0.0129	0.9693	0.1137	186.9937	60.8315	25	KCl	TRIS 7
104	2.8328	1.1619	3.7978	0.0106	0.4171	0.0341	190.4492	184.2605	25	NaCl	TRIS 8
119	1.2343	1.2858	2.3784	0.0216	2.3274	0.1456	162.6597	165.4771	25	KCl	TRIS 9
23	1.7871	0.7152	3.6892	0.0296	1.7433	0.3954	122.3329	9.8122	25	NaCl	CHES 10
12	1.2981	0.9136	4.9083	0.0263	4.3433	0.0104	192.6714	79.2516	25	NaCl	TRIS 7
15	1.5289	2.1866	2.8905	0.0344	0.7578	0.0318	145.7299	113.0788	25	NaCl	CHES 9
26	1.1160	0.6938	4.1054	0.0211	4.9747	0.1992	52.7415	179.6351	25	NaCl	CHES 9
6	2.1426	2.7930	1.8771	0.0228	0.7169	0.3740	183.9854	178.6756	25	KCl	CHES 9
32	1.8236	4.6349	0.2885	0.0024	1.1783	0.3498	177.3001	103.9578	25	NaCl	CHES 9
79	2.9378	3.3090	4.6348	0.0120	2.9642	0.1333	199.0172	49.7544	25	NaCl	CHES 9
37	1.1828	4.1545	4.7816	0.0153	4.5447	0.2418	161.2296	34.8356	25	NaCl	TRIS 9
21	2.9873	0.8293	4.8899	0.0030	0.5975	0.3389	126.4874	134.7750	25	KCl	TRIS 8
73	1.9382	4.7326	3.9649	0.0017	4.3093	0.0402	146.1976	10.6932	25	NaCl	TRIS 7
109	2.8973	2.1440	2.3302	0.0280	0.9625	0.2572	191.1624	80.7107	25	KCl	TRIS 8
31	1.2255	2.2421	4.4450	0.0003	0.2949	0.3754	109.9163	2.0858	25	NaCl	TRIS 7
66	1.0043	0.6042	1.2713	0.0123	4.4830	0.1053	115.7937	17.1938	25	NaCl	TRIS 8
14	2.1082	1.8314	4.9974	0.0321	0.3551	0.1400	175.4733	127.9333	25	NaCl	CHES 10
25	2.9554	2.4984	4.2553	0.0330	4.4660	0.3992	95.3727	30.9473	25	KCl	TRIS 7
36	1.1061	4.8595	2.5313	0.0205	2.1585	0.2331	107.7050	10.4385	25	NaCl	TRIS 8
107	2.5613	0.8899	0.4602	0.0294	0.8739	0.1687	147.8762	101.5582	25	KCl	CHES 9
50	2.9580	3.4579	3.3337	0.0196	0.4782	0.2917	194.9325	196.6170	25	KCl	TRIS 8
110	2.6621	1.8942	1.8899	0.0208	2.8413	0.1726	121.2303	184.2389	25	NaCl	CHES 9
127	1.8471	0.3725	3.5489	0.0037	3.0641	0.3631	194.6162	47.8932	25	KCl	TRIS 7
111	1.9190	0.5701	3.2286	0.0283	2.8380	0.1061	78.0444	88.8578	25	KCl	TRIS 7
116	1.5957	3.0581	0.7663	0.0036	2.5050	0.1442	132.7821	186.1585	25	NaCl	CHES 9
120	1.8770	0.7176	0.8080	0.0207	4.4119	0.1554	158.3602	124.0623	25	NaCl	CHES 10
114	1.8188	3.2362	4.3005	0.0112	1.5119	0.1113	127.5987	112.2584	25	KCl	TRIS 8
39	2.3547	2.8825	0.4624	0.0059	2.8077	0.1909	68.5996	2.1568	25	KCl	TRIS 9
33	2.0475	3.1297	3.5118	0.0014	4.8356	0.3887	53.8128	186.5941	25	NaCl	CHES 10
98	2.6992	0.9530	4.5380	0.0256	2.1683	0.3414	148.7531	198.4554	25	NaCl	CHES 9
113	1.6748	0.5471	3.4596	0.0223	4.1790	0.0375	126.0493	181.1388	25	KCl	CHES 9

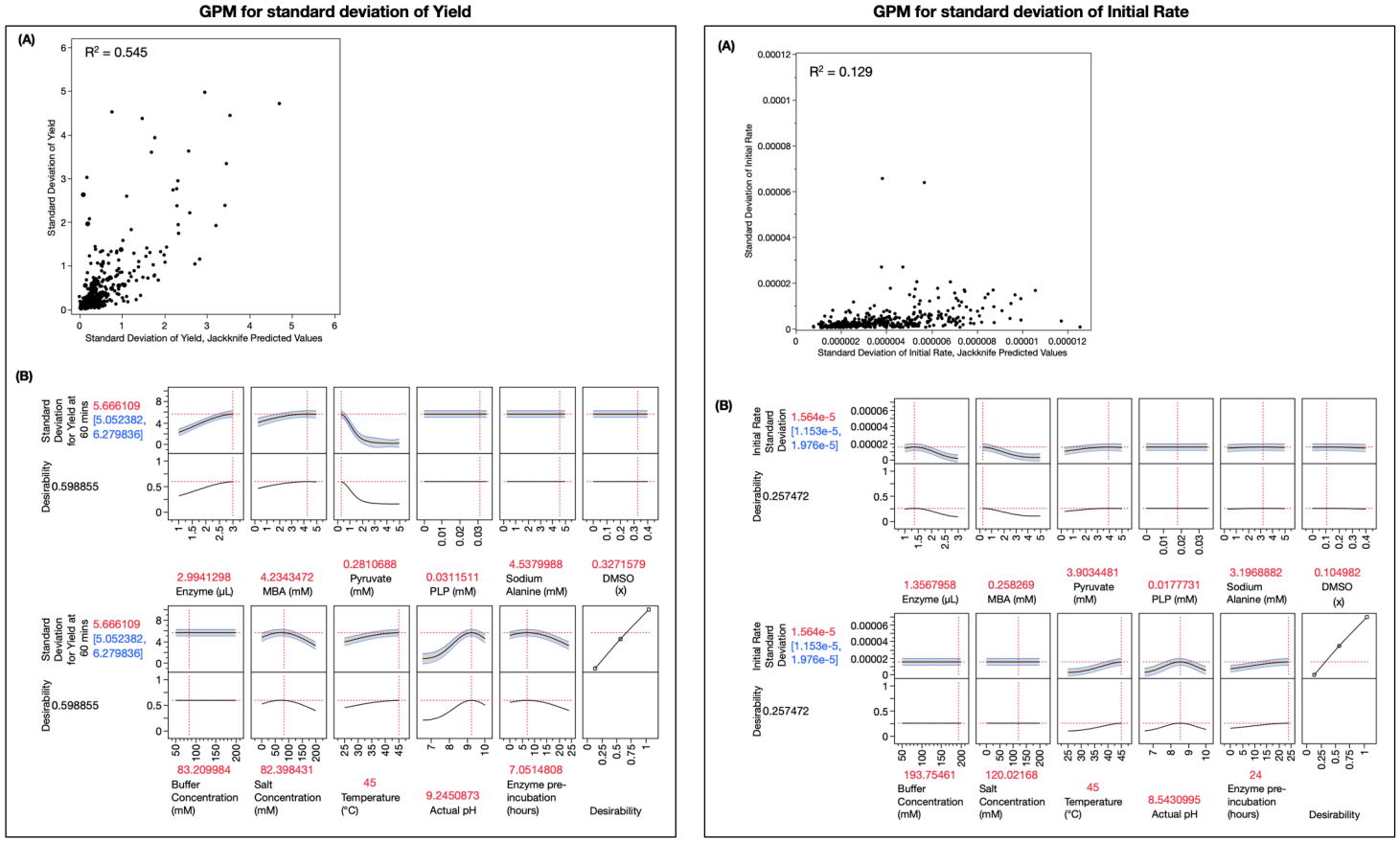
35 °C

Run	Enzyme (µl)	MBA (mM)	Pyruavte (mM)	PLP (mM)	Sodium Alanine (mM)	DMSO (X)	Buffer concentration (mM)	Salt concentration (mM)	Temperature (°C)	Salt type	Buffer pH
131	2.8120	3.9318	4.7937	0.0176	4.0234	0.0134	86.6339	75.0911	35	KCl	CHES 10
124	2.6118	2.3469	0.3646	0.0174	4.7815	0.3547	120.9806	17.9895	35	NaCl	TRIS 7
4	1.2456	4.7587	4.1526	0.0340	4.0826	0.2170	135.7076	184.3694	35	NaCl	CHES 9
94	2.9941	0.3942	3.0111	0.0019	4.4034	0.2073	77.7006	189.9035	35	KCl	TRIS 8
47	2.8078	3.3813	4.1867	0.0303	2.2837	0.1352	59.2488	149.1510	35	KCl	CHES 9
133	2.4334	1.4666	4.7646	0.0071	4.7890	0.1866	106.8542	121.0252	35	NaCl	TRIS 9
40	1.9029	2.0866	4.1201	0.0001	1.4464	0.3041	182.0097	84.6726	35	KCl	TRIS 7
75	2.7004	3.8889	3.8066	0.0042	3.6206	0.0809	196.6966	18.2616	35	KCl	TRIS 8
63	2.2047	0.4614	0.6543	0.0322	4.6001	0.0031	56.8977	100.1329	35	NaCl	CHES 10
18	1.7038	4.5027	3.1750	0.0314	2.4024	0.0012	135.4645	15.7404	35	KCl	TRIS 7
29	2.8595	4.9748	4.5660	0.0241	0.4087	0.3958	74.7838	25.7105	35	KCl	CHES 10
117	1.2965	0.4238	1.9041	0.0083	1.5623	0.2764	174.6615	169.1392	35	NaCl	TRIS 9
97	2.5790	4.1470	1.3688	0.0051	4.0922	0.3794	153.2836	154.3362	35	NaCl	CHES 10
128	1.0791	1.8085	1.9639	0.0006	3.9650	0.3833	78.7804	15.1886	35	KCl	CHES 10
10	1.0618	2.3370	4.2816	0.0272	2.1188	0.2636	180.5025	177.0029	35	NaCl	TRIS 8
134	2.6675	1.4527	0.8439	0.0146	3.7355	0.0152	131.6081	142.1326	35	NaCl	TRIS 8
34	2.5907	4.2420	0.5864	0.0005	0.2623	0.1536	125.5888	35.3956	35	KCl	TRIS 8
106	2.4065	3.1169	2.1740	0.0263	1.9555	0.1679	87.5871	30.5150	35	KCl	TRIS 7
80	1.0534	4.5496	2.4657	0.0214	4.7798	0.3415	167.9450	142.8169	35	NaCl	CHES 10
17	1.4710	2.9845	4.6110	0.0272	4.8105	0.0330	199.7458	9.6153	35	KCl	CHES 10
20	1.1866	0.8699	4.9934	0.0140	2.0579	0.1947	92.2079	14.3809	35	NaCl	TRIS 9
136	2.2784	2.1923	4.4622	0.0053	4.2234	0.0420	66.5986	193.2346	35	KCl	TRIS 8
54	1.2856	3.0788	4.6706	0.0133	3.7962	0.2916	50.4108	163.2507	35	NaCl	CHES 10
84	2.2220	4.9038	2.0710	0.0034	0.7101	0.0874	71.2404	95.1973	35	KCl	TRIS 9
7	1.0118	0.5762	0.5847	0.0343	2.7949	0.3703	199.4670	99.1796	35	KCl	TRIS 9
55	1.1447	2.3951	4.8040	0.0001	3.4419	0.0758	142.9746	199.0380	35	NaCl	CHES 9
68	2.2773	4.6657	0.6977	0.0261	4.4587	0.2485	182.6852	1.5069	35	KCl	TRIS 8
118	1.5643	2.3141	0.5565	0.0086	2.2265	0.0002	185.8174	158.4334	35	KCl	CHES 10
135	2.7583	2.4371	1.5513	0.0239	4.7476	0.2105	57.5127	88.5005	35	KCl	TRIS 7
51	1.0214	0.3696	4.3251	0.0025	0.4467	0.0300	61.6203	34.8291	35	NaCl	TRIS 8
121	1.7091	2.6338	3.7151	0.0012	4.6409	0.3590	159.2912	25.8092	35	KCl	TRIS 8
90	1.7638	2.8049	0.9978	0.0325	1.3109	0.1698	50.4833	90.9853	35	KCl	CHES 10
27	2.7250	1.6842	4.2167	0.0066	0.5300	0.3907	51.0416	119.6038	35	KCl	CHES 9
76	2.9782	3.1014	2.3123	0.0009	0.7816	0.0779	185.3158	193.5928	35	NaCl	TRIS 8
82	2.7408	4.6959	4.9962	0.0275	1.2771	0.0090	54.2012	32.6207	35	KCl	TRIS 8
41	2.8020	4.9850	0.2811	0.0312	3.3245	0.3348	122.4346	159.4335	35	KCl	CHES 10
60	1.5215	0.4825	0.5015	0.0050	0.7065	0.2127	101.7131	30.7097	35	KCl	TRIS 8
16	1.0553	4.3920	4.8650	0.0230	1.8580	0.0633	118.3706	66.5522	35	KCl	CHES 9
115	1.4627	3.3792	2.6915	0.0091	0.9750	0.0499	109.8436	166.6616	35	NaCl	TRIS 8
103	2.5680	0.8520	1.0453	0.0183	1.3512	0.0038	181.0095	23.4637	35	NaCl	TRIS 7
130	1.1598	1.1024	1.3970	0.0349	3.5647	0.2661	64.5079	135.2784	35	NaCl	TRIS 9
99	2.8991	3.6322	4.9744	0.0086	2.0618	0.3311	174.8504	107.0547	35	NaCl	TRIS 7
125	1.4876	1.9073	3.7515	0.0243	4.4614	0.2786	90.8137	148.8117	35	NaCl	CHES 9
105	2.4469	2.0645	1.7889	0.0342	1.8682	0.1073	188.8831	55.6160	35	NaCl	CHES 10
59	1.8847	4.1985	2.8627	0.0064	2.4914	0.2711	74.8047	18.4665	35	KCl	CHES 9
42	2.2966	4.8162	4.5440	0.0150	4.9151	0.2607	195.0913	187.3685	35	KCl	TRIS 9

45 °C

Run	Enzyme (µl)	MBA (mM)	Pyruavte (mM)	PLP (mM)	Sodium Alanine (mM)	DMSO (X)	Buffer concentration (mM)	Salt concentration (mM)	Temperature (°C)	Salt type	Buffer pH
122	1.2076	1.5751	0.8057	0.0279	2.9944	0.2757	136.6279	3.1337	45	KCl	CHES 9
137	2.4001	4.3368	1.8541	0.0026	3.3751	0.2067	142.6385	85.0640	45	KCl	TRIS 9
100	2.3912	3.3493	3.9713	0.0323	0.6647	0.3848	162.0455	73.1780	45	KCl	TRIS 9
86	2.5461	2.0932	1.9494	0.0103	0.3010	0.0007	54.0416	170.6795	45	NaCl	TRIS 9
58	1.9010	0.8335	2.9854	0.0183	3.8037	0.3755	55.3834	115.2421	45	NaCl	TRIS 7
53	1.0992	4.9625	4.4329	0.0041	2.8830	0.3420	54.4295	178.3251	45	NaCl	TRIS 9
123	2.8521	0.3283	1.9968	0.0294	4.4846	0.3713	157.5873	4.4082	45	KCl	TRIS 9
69	2.7963	3.2391	0.3962	0.0262	4.2747	0.0225	135.9763	51.7147	45	KCl	CHES 9
57	1.9349	0.5235	1.5918	0.0003	4.8915	0.2675	155.5402	66.9395	45	KCl	TRIS 7
74	1.2549	4.4977	0.7047	0.0016	3.9881	0.0550	190.4454	124.9569	45	KCl	TRIS 7
46	1.0931	3.6047	0.7311	0.0332	4.7254	0.2519	124.4004	105.5000	45	NaCl	TRIS 9
138	2.5052	2.7348	2.0835	0.0180	4.5005	0.1975	93.2209	133.5268	45	NaCl	TRIS 7
93	2.6890	2.4692	0.4899	0.0055	1.6360	0.3577	182.8698	190.2494	45	KCl	CHES 9
62	1.7740	4.8826	0.7577	0.0070	4.5793	0.0140	109.5294	4.5675	45	NaCl	CHES 9
70	1.3652	3.8176	4.2791	0.0219	4.9282	0.3317	58.5201	33.3813	45	KCl	CHES 10
44	1.9837	0.8397	4.0967	0.0232	3.2733	0.0137	168.6172	196.7454	45	NaCl	CHES 9
9	1.2694	3.5527	2.5936	0.0348	0.5745	0.3979	138.3439	156.5950	45	KCl	TRIS 7
78	2.0799	2.7876	4.7470	0.0044	0.2840	0.0044	189.2883	108.0628	45	NaCl	TRIS 9
1	1.0313	0.2679	0.4224	0.0098	4.8706	0.3880	173.5064	12.9046	45	NaCl	TRIS 9
45	2.8726	3.9601	2.3765	0.0347	4.9517	0.0450	71.3073	176.9216	45	NaCl	TRIS 8
67	2.6806	0.3137	3.6749	0.0206	3.1194	0.0467	99.3568	20.9187	45	KCl	TRIS 7
95	2.8711	2.9773	1.1923	0.0139	3.4519	0.3941	141.5374	37.9384	45	NaCl	TRIS 8
83	2.3590	4.0711	4.6037	0.0134	0.4254	0.1949	133.3257	63.3846	45	NaCl	TRIS 7
61	1.0275	4.7009	0.9864	0.0254	3.2912	0.0255	195.0490	44.3886	45	NaCl	TRIS 7
126	2.4014	0.8834	3.5741	0.0146	2.5180	0.2871	138.8794	170.7180	45	NaCl	TRIS 8
28	1.6717	0.8994	2.4874	0.0316	0.3158	0.1854	67.7718	4.6166	45	NaCl	TRIS 9
3	1.1881	4.3935	3.0803	0.0193	1.3079	0.3874	189.7079	182.2378	45	KCl	TRIS 8
71	1.2791	1.4994	2.4200	0.0052	3.5954	0.1808	199.5181	198.1223	45	KCl	CHES 10
87	1.6303	4.1676	0.3054	0.0160	1.1131	0.2660	114.1889	49.8873	45	NaCl	CHES 9
89	2.0899	4.3144	0.7272	0.0219	2.0506	0.0309	149.7488	144.7106	45	KCl	CHES 9
19	1.8508	2.8631	3.4504	0.0350	0.8537	0.0453	185.6774	27.5054	45	KCl	TRIS 9
72	1.1496	1.0707	0.5501	0.0068	0.4050	0.0646	162.5636	87.8336	45	KCl	CHES 9
96	2.2873	2.7493	1.0806	0.0275	1.0031	0.0732	129.8791	199.4370	45	KCl	CHES 10
49	2.8622	2.1619	3.5368	0.0081	2.8956	0.0265	119.7296	80.2392	45	KCl	TRIS 7
77	1.4419	0.3861	3.8651	0.0181	3.4996	0.1101	175.8558	64.7191	45	KCl	TRIS 9
91	2.7728	0.4151	1.6311	0.0325	1.5020	0.0113	111.3195	182.1729	45	NaCl	TRIS 9
56	1.3785	0.6588	2.6182	0.0019	0.7484	0.3450	116.1858	110.1622	45	KCl	CHES 10
65	2.0821	1.7041	2.5749	0.0166	4.8655	0.0349	65.1773	72.2262	45	NaCl	TRIS 9
81	1.0692	3.6774	0.5994	0.0130	4.1727	0.0172	75.6084	188.3798	45	NaCl	CHES 10
64	2.5089	0.2583	4.4398	0.0266	4.9263	0.2273	102.5854	94.6023	45	KCl	TRIS 9
48	2.7887	4.3786	3.6469	0.0063	4.5838	0.1699	145.9758	185.5930	45	NaCl	CHES 10
8	1.8669	2.2273	1.1164	0.0168	4.2648	0.2683	188.6918	164.2339	45	KCl	CHES 10
52	1.3108	3.2155	3.0512	0.0106	4.9830	0.3911	111.1807	81.7381	45	KCl	TRIS 9
92	2.5696	3.2612	1.7682	0.0005	3.5116	0.3488	56.9057	119.1239	45	KCl	TRIS 7
2	1.2969	2.4122	4.6384	0.0323	2.6163	0.3602	104.7745	196.4926	45	NaCl	CHES 10
5	1.1717	0.3484	1.7728	0.0302	0.3400	0.1549	190.3605	151.9195	45	NaCl	TRIS 7

Appendix K



Gaussian process models for standard deviation of Yield (left) and Initial Rate (right). (A) Actual by Jacknife Predicted plot. Shows the correlation between actual with predicted data points (dots) (B) Prediction Profiler set to conditions that generate maximum standard deviation. Impact of standard deviations on factors is shown in each sub plot using profiles (black line) with 95% confidence intervals (shaded grey). Values in red (X-axis) show the each of the factor settings. Maximum standard deviation ± 95% confidence interval predicted is shown in red and blue text, respectively, on the Y-axis.

Appendix L

(A) Gaussian Process Model Report for Yield (%)

Continuous factors	Theta	Total Sensitivity	Main Effect	Enzyme (µL) Interaction	MBA (mM) Interaction	Pyruvate (mM) Interaction	PLP (mM) Interaction	Sodium Alanine (mM) Interaction	DMSO (% v/v) Interaction	Buffer Concentration (mM) Interaction	Salt Concentration (mM) Interaction	Temperature (°C) Interaction	pH Interaction	Enzyme pre-incubation (hours) Interaction
Enzyme (µL)	0.0171	0.0226	0.0164	.	0.0000	0.0004	0.0000	0.0000	0.0001	0.0000	0.0001	0.0001	0.0052	0.0003
MBA (mM)	0.0113	0.0931	0.0605	0.0000	.	0.0177	0.0000	0.0000	0.0007	0.0000	0.0004	0.0003	0.0096	0.0039
Pyruvate (mM)	0.2238	0.2870	0.1384	0.0004	0.0177	.	0.0026	0.0000	0.0023	0.0000	0.0091	0.0017	0.0970	0.0178
PLP (mM)	199.9770	0.0182	0.0008	0.0000	0.0000	0.0026	.	0.0000	0.0020	0.0000	0.0030	0.0002	0.0094	0.0001
Sodium Alanine (mM)	0.0006	0.0006	0.0004	0.0000	0.0000	0.0000	0.0000	.	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000
DMSO (% v/v)	3.0835	0.0553	0.0277	0.0001	0.0007	0.0023	0.0020	0.0000	.	0.0000	0.0004	0.0019	0.0174	0.0026
Buffer Concentration (mM)	0.0000	0.0006	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	.	0.0000	0.0000	0.0001	0.0000
Salt Concentration (mM)	0.0000	0.0231	0.0014	0.0001	0.0004	0.0091	0.0030	0.0000	0.0004	0.0000	.	0.0028	0.0049	0.0010
Temperature (°C)	0.0008	0.0167	0.0002	0.0001	0.0003	0.0017	0.0002	0.0000	0.0019	0.0000	0.0028	.	0.0072	0.0023
pH	1.7157	0.5021	0.3100	0.0052	0.0096	0.0970	0.0094	0.0003	0.0174	0.0001	0.0049	0.0072	.	0.0411
Enzyme pre-incubation (hours)	0.0006	0.1275	0.0585	0.0003	0.0039	0.0178	0.0001	0.0000	0.0026	0.0000	0.0010	0.0023	0.0411	.
Categorical factor - Salt Type	KCl	NaCl												
KCl	1	1												
NaCl	1	1												

(B) Gaussian Process Model Report for Initial Rate (ΔA/s)

Continuous factors	Theta	Total Sensitivity	Main Effect	Enzyme (µL) Interaction	MBA (mM) Interaction	Pyruvate (mM) Interaction	PLP (mM) Interaction	Sodium Alanine (mM) Interaction	DMSO (% v/v) Interaction	Buffer Concentration (mM) Interaction	Salt Concentration (mM) Interaction	Temperature (°C) Interaction	pH Interaction	Enzyme pre-incubation (hours) Interaction
Enzyme (µL)	0.0755	0.0089	0.0003	.	6.20E-05	3.55E-06	3.15E-10	4.10E-12	2.37E-04	1.25E-11	3.92E-07	3.74E-04	7.51E-03	3.95E-04
MBA (mM)	0.0416	0.0284	0.0013	6.20E-05	.	3.28E-04	8.29E-09	4.38E-11	2.00E-03	6.68E-12	4.00E-05	4.93E-04	2.08E-02	3.38E-03
Pyruvate (mM)	0.0331	0.0279	0.0023	3.55E-06	3.28E-04	.	1.50E-09	4.03E-12	1.81E-03	3.25E-11	8.55E-06	3.88E-04	2.26E-02	3.96E-04
PLP (mM)	5.4960	0.0000	0.0000	3.15E-10	8.29E-09	1.50E-09	.	1.48E-15	1.08E-07	4.23E-15	4.60E-10	2.90E-09	2.52E-06	6.94E-08
Sodium Alanine (mM)	0.0000	0.0000	0.0000	4.10E-12	4.38E-11	4.03E-12	1.48E-15	.	1.56E-10	1.97E-16	1.09E-12	1.61E-12	4.00E-09	2.40E-11
DMSO (% v/v)	41.4233	0.2056	0.0031	2.37E-04	2.00E-03	1.81E-03	1.08E-07	1.56E-10	.	7.04E-11	2.67E-04	3.04E-03	1.85E-01	1.03E-02
Buffer Concentration (mM)	0.0000	0.0000	0.0000	1.25E-11	6.68E-12	3.25E-11	4.23E-15	1.97E-16	7.04E-11	.	1.16E-12	3.67E-10	5.00E-08	1.92E-10
Salt Concentration (mM)	0.0000	0.0061	0.0003	3.92E-07	4.00E-05	8.55E-06	4.60E-10	1.09E-12	2.67E-04	1.16E-12	.	9.67E-05	5.28E-03	1.61E-05
Temperature (°C)	0.0097	0.0921	0.0002	3.74E-04	4.93E-04	3.88E-04	2.90E-09	1.61E-12	3.04E-03	3.67E-10	9.67E-05	.	8.61E-02	1.43E-03
pH	141.3907	0.6519	0.1104	7.51E-03	2.08E-02	2.26E-02	2.52E-06	4.00E-09	1.85E-01	5.00E-08	5.28E-03	8.61E-02	.	2.14E-01
Enzyme pre-incubation (hours)	0.0036	0.2560	0.0259	3.95E-04	3.38E-03	3.96E-04	6.94E-08	2.40E-11	1.03E-02	1.92E-10	1.61E-05	1.43E-03	2.14E-01	.
Categorical factor - Salt Type	KCl	NaCl												
KCl	1	1												
NaCl	1	1												

Gaussian Process Model Reports for Yield and Initial Rate responses. Theta shows the model parameter estimates, Total Sensitivity is the sum of the main effect and all interactions. The Main Effect is the value for total variation caused by each factor. Main effects, and two-factor interactions are separately coloured from white to green to highlight smallest to biggest model parameter estimates.