Automated Kinetic Model Identification of Biocatalysts Under Process Conditions

Vom Promotionsausschuss
der Technischen Universität Hamburg
zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation (Monografie)

von
Leon Hennecke

aus
Braunschweig

2024

1. Gutachter: Prof. Dr. rer. nat. Andreas Liese

2. Gutachter: Prof. Dr.-Ing. Michael Schlüter

Prüfungsvorsitz: Prof. Dr. Alexander Penn

Tag der mündlichen Prüfung: 24.05.2024

Acknowledgement

First of all, I would like to thank Prof. Andreas Liese for giving me the opportunity to do my PhD at his institute. I am grateful for the trust, freedom and openness that I received during my time at the institute as a student, student assistant and research assistant. I would also like to thank Prof. Michael Schlüter for taking on the role of second examiner and Prof. Alexander Penn for chairing the exam. In addition, I would like to thank Prof. Mirko Skiborowski and Lucas Schaare of the Institute of Process Systems Engineering for being exceptional project partners with fruitful discussions and scientific support for the modelling procedures.

I would like to thank my students Jonathan Garlipp, Eric Nitschke, Grit Brauckmann, Fin Conje, Vincent Settler, Poom Rogge and Meike Kuhr for pushing the projects forward and teaching me a lot about being a better scientist and supervisor.

Special thanks to all past and present members of the Process Analytical Technologies (PAT) working group, including Dr. Stefan Wahlefeld, Dr. Joscha Kleber, Dr. Robert Hiessl, Victoria Bueschler, Luca Schmidt and Grit Brauckmann. Joscha always helped me to stay diplomatic and helped me remaining a positive mindset. Stefan supported me from the beginning to the end of my project, even beyond his time as group leader, and gave me honest and critical feedback, which I really appreciated. Robert always motivated me as a supervisor, colleague and friend and made sure I had a great time at the institute. Victoria, always had a good idea to throw across the monitors. I spent a lot of good times in and out of the lab with Luca and Grit, talking a lot of nonsense which made the corona pandemic much easier and make day-to-day lab work a lot funnier.

I would also like to thank the entire ITB team, with special thanks to the non-PAT people Jan-Ole Kundoch, Niklas Widderich, Daniela Eixenberger, Florian Kelsch and Catharina Kleist for the many conversations, coffees and beers at the pond.

Finally, I would like to thank my family and friends who have supported me throughout my academic career. Especially Saskia deserves my biggest thanks, because she always gave me motivating words, a sympathetic ear, love and a lot of patience during the final phase of my PhD.

I. Abstract

Kinetic model identification is important for the optimization and scale-up of enzymatic processes. The modelling process is strongly influenced by human error, experimental design and model selection, which can be overcome by automation. Therefore, in this work, an automated reactor platform has been developed to perform kinetic model identification for a model enzymatic reaction. This reaction is catalyzed by a formate dehydrogenase from the yeast *Candida boidinii*. The reduction of NAD^+ to NADH with the hydride donor sodium formate is monitored by inline UV/Vis spectroscopy at 340 nm in the design space of 0.35 to 2 mM NAD^+ and 50 to 300 mM formate. The enzyme is utilized in a packed bed reactor, where it is immobilized on functionalized porous resin particles. The steady-state concentration of NADH is measured to determine the corresponding reaction rate at various initial substrate concentrations, providing data for identifying a kinetic model from eight candidate models. These kinetic models serve as the basis for a model-based design of experiments using the Akaike weight design criterion. This method uses Akaike weights to measure the relative goodness of fit for candidate models and identify experiments with maximum information content. Once a successful model discrimination is achieved, parameter estimation is improved through additional experiments according to the e-optimal design criterion to reduce the confidence intervals of the model parameters of the discriminated kinetic model. The model identification process is based on six fully factorial designed experiments initially, which is then improved through nine additional experiments. This approach leads to the discovery of a kinetic model within 15 experiments, with standard deviations of less than 10% within the range of measurement error. The double substrate Michaelis-Menten model accurately represented the packed bed reactor data after implementing a multi-start heuristic and extending the design space. The kinetic model identification was optimized and applied to data from an enzyme membrane reactor process with formate dehydrogenase in solution. The resulting model showed non-competitive product inhibition, in contrast to the model for the immobilized enzyme. The reactor configuration and immobilization resulted in a distinct kinetic model and confidence intervals. This emphasizes the impact of experimental design on identifying kinetic models and the advantages of automation in this field.

II. Zusammenfassung

Die Identifizierung von kinetischen Modellen ist wichtig für die Optimierung und das Scale-up von enzymatischen Prozessen. Um den Einfluss des menschlichen Fehlers, der Versuchsplanung und der Modellauswahl zu minimieren, wurde in dieser Arbeit eine automatisierte Reaktorplattform entwickelt, um die kinetische Modellidentifizierung für eine enzymatische Modellreaktion durchzuführen. Die Reduktion von NAD^+ zu NADH mit dem Hydriddonator Natriumformiat wird durch eine Formiatdehydrogenase aus der Hefe *Candida boidinii* katalysiert. Die Reaktion wird im Konzentrationsbereich von 0,35 bis 2 mM NAD^+ und 50 bis 300 mM Formiat durch Inline-UV/Vis-Spektroskopie bei 340 nm verfolgt. Das Enzym ist auf funktionalisierten porösen Harzpartikeln immobilisiert und wird in einem Festbettreaktor verwendet. Die stationäre NADH-Konzentration wird gemessen, um die entsprechende Reaktionsgeschwindigkeit bei verschiedenen anfänglichen Substratkonzentrationen zu bestimmen. Die daraus resultierenden Daten dienen zur Identifizierung eines kinetischen Modells aus acht möglichen Modellen, die die Grundlage für eine modellbasierte Versuchsplanung unter Verwendung des Akaike-Gewichtskriteriums bilden. Bei dieser Methode werden Akaike-Gewichte verwendet, um die relative Anpassungsgüte von Modellkandidaten zu messen und Experimente mit maximalem Informationsgehalt zu ermitteln. Sobald eine erfolgreiche Modelldiskriminierung erreicht ist, wird die Parameterschätzung durch zusätzliche Experimente gemäß dem e-optimalen Designkriterium verbessert, um die Konfidenzintervalle der Modellparameter des diskriminierten kinetischen Modells zu reduzieren. Der Prozess zur Identifizierung des Modells basiert zunächst auf sechs Experimenten eines vollständigen Versuchsplans. Anschließend werden neun optimierte Experimente durchgeführt, um das Modell zu verbessern. Dieser Ansatz führt zur Entdeckung eines kinetischen Modells innerhalb von 15 Experimenten. Die Standardabweichungen liegen dabei innerhalb des Messfehlerbereichs und betragen weniger als 10%. Das Doppelsubstrat-Michaelis-Menten-Modell repräsentiert die Daten des Festbettreaktors mit der höchsten Genauigkeit, nachdem eine Mehrfachstart-Heuristik implementiert und der Designraum erweitert wurde. Die Identifizierung des kinetischen Modells wurde optimiert und auf die Daten eines Enzym-Membranreaktorprozesses mit Formiatdehydrogenase in Lösung angewendet. Das resultierende Modell zeigte eine nicht-kompetitive Produkthemmung, im Gegensatz zum

Modell für das immobilisierte Enzym. Die Reaktorkonfiguration und die Immobilisierung führten zu einem unterschiedlichen kinetischen Modell und Vertrauensintervall. Die produzierten Ergebnisse betonen die Bedeutung der Versuchsplanung für die Identifizierung kinetischer Modelle und die Vorteile der Automatisierung in diesem Bereich.

III. Table of Content

IV. List of Figures

V. List of Tables

VI. List of Symbols

Symbol	Unit	Meaning
μ	-	Mean of normal distribution
A	-	Absorbance
b	-	Best model
c	mM	Concentration
c_A	mM	NAD^+ concentration
c_B	mM	Formate concentration
c_{CSTR}	mM	CSTR concentration
c_E	mM	Enzyme concentration
$c_{E,0}$	mM	Initial enzyme concentration
c_{ES}	mM	Enzyme-substrate concentration
$CI_{95\%}$	%	95% confidence interval
c_{in}	mM	Inlet concentration
c_{out}	mM	Outlet concentration
c_P	mM	Product concentration
$c_{P,in}$	mM	Product inlet concentration
$c_{P,out}$	mM	Product outlet concentration
c_Q	mM	NADH concentration
c_S	mM	Substrate concentration
g	-	Number of parameters

i	-	Intercept
I_0	-	Baseline intensity
I_i	-	Measured intensity
j	-	Model index
k_2	s^{-1}	Enzyme substrate dissociation constant
k_{cat}	s^{-1}	Catalytic constant
K_i	mM	Inhibition constant
$K_{i,A}$	mM	Inhibition constant (NAD^+)
$K_{i,Q}$	mM	Inhibition constant (NADH)
K_m	mM	Michaelis Menten constant
$K_{m,A}$	mM	Michaelis Menten constant (NAD^+)
$K_{m,B}$	mM	Michaelis Menten constant (NADH)
L	-	Log-likelihood
n	-	Sample amount
n_p	mmol	Amount of substance
o	-	Observed model
p	-	Probability
P_1	mM^2	Combined parameter ($K_{i,A} \cdot K_{m,B}$)
s	-	Slope

$\mathbf{s}$	-	Sensitivity
$\mathbf{S}$	-	Sensitivity matrix
$\mathbf{t}$	s	Time
$\mathbf{u}$	-	Experimental conditions
$\mathbf{V}$	µL	Volume
$\mathbf{v}$	$mM \cdot min^{-1}$	Reaction rate
$\dot{\mathbf{V}}$	$\mu L \cdot min^{-1}$	Volume flow
$\mathbf{V_{CSTR}}$	µL	Volume of CSTR
$\dot{\mathbf{V}}_{\mathbf{in}}$	$\mu L \cdot min^{-1}$	Volume in-flow
$\mathbf{v_{max}}$	$mM \cdot min^{-1}$	Maximum reaction rate
$\dot{\mathbf{V}}_{\mathbf{out}}$	$\mu L \cdot min^{-1}$	Volume out-flow
$\mathbf{w_{AIC}}$	-	Akaike weight
$\mathbf{x}$	m	Axial coordinate
$\mathbf{y}$	-	Experimental data
$\bar{\mathbf{y}}$	-	Mean value
$\boldsymbol{\Theta}$	-	Parameter vector
$\boldsymbol{\lambda_{min}}$	-	Minimal eigenvalue
$\boldsymbol{\sigma}$	-	Standard deviation
$\boldsymbol{\tau}$	min	Residence time
$\boldsymbol{\Phi}$	-	Objective function

VII. List of Abbreviations

Abbreviation	Meaning
AI	Artificial intelligence
AIC	Akaike information criteria
AICc	Corrected Akaike information criteria
AWDC	Akaike weight design criterion
BSA	Bovine serum albumin
BIC	Bayesian information criteria
CDC	Classical design criterion
CPI	Competitive product inhibition model
CSTR	Continuous stirred tank reactor
DoE	Design of experiments
EMR	Enzyme membrane reactor
FDH	Formate dehydrogenase
FIM	Fisher information matrix
KP_i	Potassium phosphate buffer
M1	Model of original design space
M1*	Model of extended design space
MBDoE	Model-based design of experiments
MICH	Michalik *et al.* model
MM	Michaelis Menten model
MODI	Modified Michalik *et al.* model
TRR	Trust region reflective algorithm
LM	Levenberg-Marquardt algorithm
NAD^+	Nicotinamide adenine dinucleotide (ox.)
NADH	Nicotinamide adenine dinucleotide (red.)
NCPI	Non-competitive model
NMR	Nuclear magnetic resonance
PBR	Packed bed reactor
PFR	Plug flow reactor
PMMA	Polymethylmethacrylat
PTFE	Polytetrafluorethylene
SCHM	Schmidt *et al.* model

SCHMlump	Schmidt *et al.* lumped model
UPI	Uncompetitive product inhibition model

1. Introduction

1.1. Challenges for Biocatalysis

Biocatalysis is describing the synthesis of valuable chemicals (e.g. pharmaceuticals, fine chemicals, chiral pure chemicals) by isolated enzymes or whole cells. Enzymes are involved in many biochemical phenomena in nature. Since enzymes lower the activation energy of a reaction, they accelerate or enable biochemical reactions. Processes using biocatalysts are usually running at milder conditions by means of pressure, temperatures and hazardous chemicals in contrast to conventional chemical plants. In addition, biocatalysis often produces less waste, offer a high selectivity and efficiency. The disadvantages of biocatalysts are represented by comparable low space time yields, incompatibility with organic solvents due to low solubility and deactivation of the enzyme itself or its cofactor. Furthermore, the most common solvent applied in biocatalysis is still water, which interferes with the product solubility and therefore produces highly diluted product streams. Compared to chemical catalysts, the substrate scope can be limited by steric hindrance of the substrates functional groups in the active site of the enzyme. These challenges are tackled by genetic engineering of the enzyme (Arnold, 2018; Alonso *et al.*, 2020; Madhavan *et al.*, 2021; Sharma *et al.*, 2021), substrate or cofactor engineering (Miao *et al.*, 2015; Otun *et al.*, 2023) and process engineering of various reaction aspects including *in situ* product removal (Heintz *et al.*, 2017; Hülsewede *et al.*, 2019; Lindeque *et al.*, 2019), reactor engineering (Karande *et al.*, 2016; Baumann *et al.*, 2020; Burek *et al.*, 2022) and immobilization (Büscher *et al.*, 2019; Sheldon *et al.*, 2021). A biocatalyst can be a viable substitute for traditional chemical synthesis if an enzyme is discovered that can achieve the desired synthesis route with comparable productivity and turnover number. The turnover number or catalytic constant (k_{cat}) is defined as the conversion of substrate molecules per second for one active catalytic site and is a measure for the enzymes maximum productivity. The rise of biotechnology is also monitored by the revenue of products produced using recombinant DNA technology. Statistics show an increasing demand for these products from 2012 until now (cf. Figure 1), whose growing market is further extrapolated until 2027 (Radtke, 2023). Biotechnology processes themselves are becoming increasingly popular as demand for products grows.

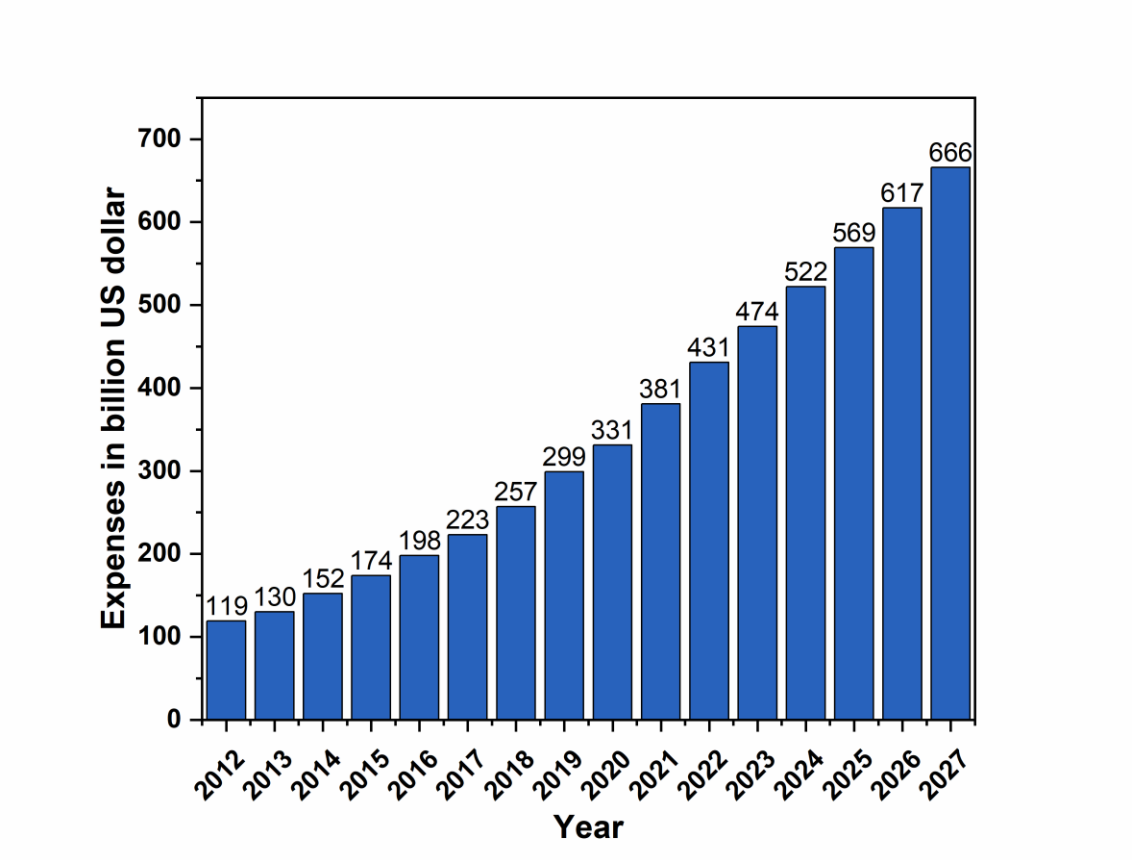

Figure 1: Global spending on biotechnology from 2012 to 2027 (in billions of US dollars). Spending after 2022 is extrapolated based analytics on previous years by the company IQVIA. Data of de.statista.com (Radtke, 2023)

Biocatalyis does not exclude classical chemistry in modern production plants, but substitute undesirable process parts like an excessive amount of reaction steps or safety and environmental critical reaction conditions (Sheldon *et al.*, 2018). However, to bring enzymes into the chemical industry, they must be able to compete with traditional chemical processes. This requires characterization and optimization of the processes enabled by enzyme variants produced via modern directed evolution and protein engineering. For a deep understanding of these processes, they have to be mathematically modeled. In addition to thermodynamic effects and mass transfer effects, enzyme kinetics play a major role in the modeling of these processes (Vasić-Rački *et al.*, 2011). A special interest lies in the kinetic modeling of continuous processes, since these show a high potential for achieving the required space-time yields necessary to compete with chemical processes (Martin *et al.*, 2018; Baumann *et al.*, 2020; De Santis *et al.*, 2020; Ötvös *et al.*,

2021). In these processes, the biocatalysts are often immobilized in order to be retained and stabilized (Binay *et al.*, 2016; Coloma *et al.*, 2021). However, no standardized screening procedures exist, especially for immobilized biocatalysts and continuous processes. This poses a major challenge for the optimization and implementation of novel processes in industry generated by advanced research due to the large heterogeneity of non-standardized data (Pleiss, 2021; Lauterbach *et al.*, 2023). In particular, the extent of libraries of enzyme variants requires rapid, accurate and comprehensive screening methods to increase the speed of implementation.

1.2. Continuous Biocatalysis

One commonly adopted way of competing with chemical reactions in industry in terms of sustainability is the implementation of continuous reactors to increase the productivity and cost-efficiency of biotechnological processes. Therefore, it is a rapidly growing field enabling green chemistry in conventional chemical production (Rogers *et al.*, 2019). In contrast to batch processes, continuous processes can run for long periods of time delivering a continuous product stream without human intervention and repetitive setup times (Croughan *et al.*, 2015; Lindeque *et al.*, 2019; Thompson *et al.*, 2019). Additionally, this type of process allows a more precise regulation of pH, temperature and substrate concentration which can lead to improved product quality and process stability. Furthermore, continuous processes are more ecological because of their reduced amount of waste and energy cost in average. (Tamborini *et al.*, 2018; Basso *et al.*, 2019; Bennett *et al.*, 2019; Thompson *et al.*, 2019) There are two main reactor types for continuous bioprocesses applying enzymes, which are described in detail below.

1.2.1. Continuous Stirred Tank Reactor

The continuous stirred tank reactor (CSTR) is a bioreactor, which consist of an agitated tank with continuous inlet of substrate and outlet of product stream. The incoming reactants are agitated to maintain a homogeneous environment of constantly high product concentrations in the whole reactor volume. Consequently, the substrate and product concentrations in the outlet are equal in the whole reactor volume (cf. Figure 2). Thus, product can be removed in a persistent quality and substrate inhibited enzymes can be utilized efficiently. When enzymes are applied in this type of reactor homogenously solubilized, they require a semipermeable membrane to separate the biocatalysts of the

product stream. However, enzymes can also be applied immobilized as a slurry retained mechanically with a mesh. (Jaeger *et al.*, 2018)

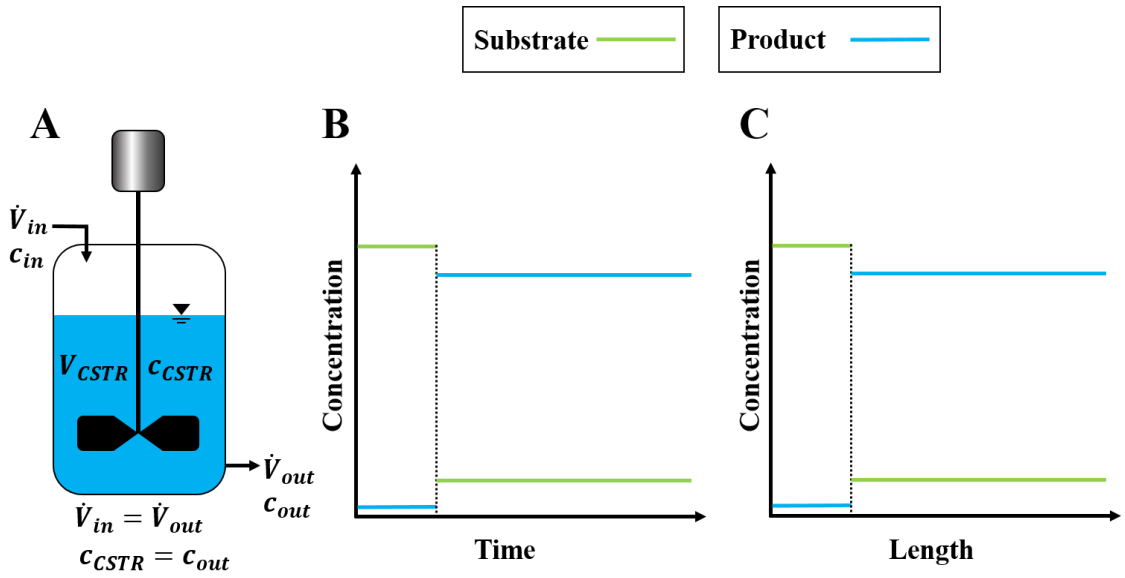

Figure 2: A: Schematic diagram for the CSTR. B: Concentration (c_i) over time (t_i) plot for the ideal CSTR process. C: Concentration (c_i) over reactor length (x_i) plot for the ideal CSTR process.

The mass balance of the CSTR is formulated as the difference in substrate concentrations ($c_{in} - c_{out}$) [mM] entering and leaving the reactor with the volume flow $\dot{V}$ adding the generation of product with the reaction rate v [mM·s^{-1}] in the CSTR with a volume of V_{CSTR} [L]. Under steady state conditions, the time dependent product generation $\frac{dn_P}{dt}$ [mmol·s^{-1}] equals zero as shown in equation 1.

$$\frac{dn_P}{dt} = \dot{V}_{in} \cdot c_{P,in} - \dot{V}_{out} \cdot c_{P,out} + v \cdot V_{CSTR} = \dot{V} \cdot \left(c_{P,in} - c_{P,out}\right) + v \cdot V_{CSTR} = 0 \qquad (1)$$

In equation 3 the residence time τ [min] defined as equation 2 is introduced as control variable.

$$\tau = \frac{V_{CSTR}}{\dot{V}} \qquad (2)$$

$$\frac{dc_P}{dt} = \frac{c_{P,in} - c_{P,out}}{\tau} + v = 0 \quad (3)$$

Thus, the reaction rate at steady state can be described by the set input concentrations, the measured output concentrations and the residence time. The reaction rate is defined by kinetics of the enzyme which is dependent on the steady state concentrations of the reactants, the reaction conditions of the process and the type of application (e.g. soluble enzyme and different types of immobilization). Various inhibition phenomena, require different reactor setups. Due to the characteristics of this reactor type, the product concentration is uniform over the reactor length and the reaction time. For this reason, product inhibition can be particularly noticeable in this reactor design and can therefore have a negative impact on productivity. This condition requires a different type of reactor to overcome the drawbacks of this inhibitions type. Therefore, the plug flow reactor (PFR) is introduced in the next chapter as a commonly adopted reactor concept that complements the use cases of CSTR. (Jaeger *et al.*, 2018; Lindeque *et al.*, 2019)

1.2.2. Plug Flow Reactors

The PFR generally comprises a lengthy tube or channel carrying reactants, with catalytic reaction occurring throughout the tube (cf. Figure 3). The resulting product concentration increases as the length of the reactor extends. The use of soluble or immobilized enzymes is also an option, similar to that of the CSTR. In chemical industry, immobilized enzymes are commonly utilized as heterogeneous catalysts in packed bed reactors (PBR), which are an important variation of the PFR. The PBR incorporates a reaction compartment in which immobilized enzymes are mechanically entrapped while the substrate feed flows through, with a controlled volume, via the so-called reactive bed. This arrangement allows the use of immobilized enzymes continuously without the need of agitation. It enhances the stability of the enzymes and simplifies the downstream processing of the product. Under optimal circumstances, the PBR demonstrates the characteristics of a PFR, however, the disparity between ideal and actual reactors grows as the reactor size increases. Similarly to the CSTR, the residence time determines the reaction rate and conversion in the reactor. However, PFR and PBR offer higher space-time yields than CSTR due to their lower reactor volume and also have lower enzyme consumption.

Furthermore, due to the concentration profile product inhibition becomes more significant close to the reactor outlet, making it possible to achieve higher conversions for enzymatic reactions inhibited by products. (Jaeger *et al.*, 2018; Lindeque *et al.*, 2019)

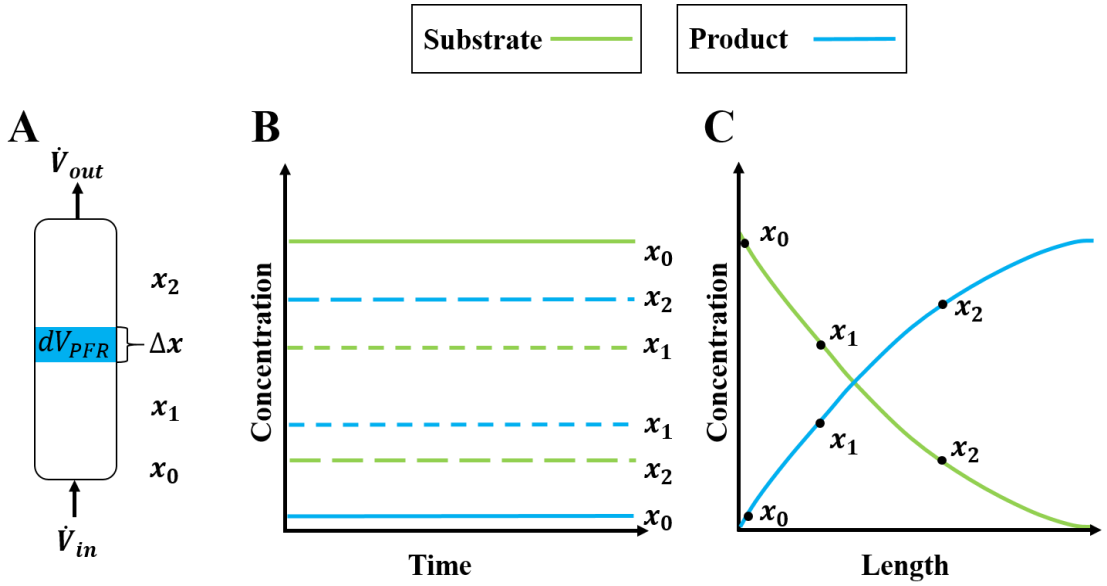

Figure 3: A: Schematic diagram for the PFR. B: Concentration (c_i) over time (t_i) plot for the ideal PFR process. C: Concentration (c_i) over reactor length (x_i) plot for the ideal PFR process.

According to the mass balance equation (1) of the CSTR, the balance term for the PFR can be introduced to describe the change in concentration. When assuming an ideal PFR, axial back mixing is negligible, so that the concentration is purely dependent on the axial coordinate, it results in equation 4. Thus, the product generation is not dependent on time in the reactor and the change in concentration is calculated in an infinitesimal small volume element based on the axial coordinate difference Δx [m].

$$\frac{dc_p}{dt} = \frac{c_{P,in} - c_{P,out}}{\tau}\bigg|_{x} - \frac{c_{P,in} - c_{P,out}}{\tau}\bigg|_{x+\Delta x} + v = 0 \qquad (4)$$

In this type of reactor, immobilization is commonly used to increase enzyme stability and reusability, thereby increasing profitability. There are two major types of immobilization, defined by their method of fixation. The first method involves retaining enzymes through physical means, such as membranes, adsorption, or encapsulation. Immobilization by

adsorption and membrane retention is a reversible process that minimally affects the enzyme structure. To maintain the solubilized enzyme in the process, downstream installation of additional membranes may be necessary. This measure is not necessary for encapsulation, where less to no leeching is observed. Encapsulation of enzymes, on the other hand, may lead to diffusion limitations, but it provides an opportunity to act as a second phase for enzymes that are incompatible with certain solvents. For instance, they can be encapsulated inside an aqueous gel and rinsed with a solvent to create a phase boundary. (Datta *et al.*, 2013; Maghraby *et al.*, 2023)

The second section of immobilization methods describing enzymes with an irreversible bond between an enzyme and a carrier matrix or by cross-linking between different enzyme units, which form catalytic active agglomerates. Both methods include covalent bond formation which can influence the quaternary structure of enzymes, changing the characteristics of the active site. This conformation alteration of the enzyme often results in activity loss, but can provide several benefits. When an enzyme is underlying conformational changes by a reactant in the form of non-competitive inhibition, a covalent single- or multipoint attachment can lead to a stabilization, which can decrease inhibitory influences. Furthermore, this structural change can affect the selectivity of a biocatalytic reaction, which facilitates downstream processing. (Datta *et al.*, 2013; Maghraby *et al.*, 2023)

Overall continuous biocatalysis offers several advantages over discontinuous processes, especially in the case of enzymatic conversions. Thus it can be beneficial for the implementation of biocatalysis in industry. Nevertheless, there are some disadvantages associated with this type of process. (De Santis *et al.*, 2020)

- The high cost of developing immobilized or stable enzymes may increase overall process costs.
- Scaling up a process can be a challenging task, demanding further optimization and validation to maintain reliable product quality and yield.
- Contamination in the reactor has the potential to cause production downtime and product loss.
- Enzyme inactivation or reactor fouling can potentially lead to reduced enzyme activity and lower productivity.

- The need for specialized equipment and infrastructure can prove expensive to incorporate.
- Continuous processes may prove more challenging to design, operate, and control in comparison to batch processes due to their inherent complexity.

In particular, the kinetic modelling of enzymes in continuous applications shows discrepancies with batch conversions due to the change in operation. This makes it difficult to move from laboratory to pilot plant or even industrial scale. This includes enzyme stability issues, which need to be incorporated into the model for long-term applications. Substrate and product inhibition also have different effects in continuous reactors. Furthermore, mass transfer limitations, if any, must be considered. Finally, structural changes during immobilization can lead to altered kinetic parameters. For this reason, continuous biocatalysis needs to be studied more intensively in research, in particular the modelling of reactions. In order to accelerate this research, new methodological and systematic approaches are needed. (Klimeš *et al.*, 2017; Häse *et al.*, 2019)

1.3. Enzyme Kinetics

Enzymes as sustainable and highly selective catalysts play a key role in continuous processes of a sustainable (bio-) chemical industry. Precise mathematical modelling of enzyme performance enables researchers and engineers to understand, optimize and scale-up enzymatic processes systematically. In their natural environment, enzyme activity is often regulated by inhibitory effects of substrate and product concentrations, as well as other regulatory substances (Sauro, 2017). These naturally occurring concentrations are usually lower than those found in industrial-scale processes. This makes it even more important for the success of biocatalytic processes to understand the characteristics of enzymes under industrial conditions. Consequently, the kinetic modelling of biocatalytic processes is crucial for the understanding of the underlying biochemical mechanisms of the enzyme and its performance under industrial relevant conditions. At the most fundamental level, a kinetic model describes how fast the reaction of a substrate S occurs in the presence of an enzyme E to a product P (cf. equation 5). In this concept, the binding of a substrate molecule to an enzyme is reversible, whereas the release of the product is not. (Cornish-Bowden, 2012; Bisswanger, 2017)

$$E + S \rightleftharpoons ES \longrightarrow E + P \qquad (5)$$

The principle and importance of enzyme kinetics are known for over 100 years since Michaelis and Menten (MM) achieved the mathematical characterization and experimental validation of enzyme kinetics in 1913 based on Henri's hypothesis in 1903 (Michaelis *et al.*, 1913). They showed on the example of invertase, that the reaction rate is directly proportional to the enzyme substrate complex concentration c_{ES} [mM]. This relation was formulated in the MM equation (6), which is widely applied in enzyme kinetics. The equation for the activity v [$mM \cdot s^{-1}$] incorporates the maximum activity v_{max} [$mM \cdot s^{-1}$] which, for saturation conditions at high substrate concentration c_S [mM] in relation to the enzyme concentration c_E [mM], is achieved by the enzyme when the enzyme substrate complex concentration c_{ES} is constant. The activity v is further dependent of the substrate concentration c_S [mM] and the MM constant K_m [mM]. The latter is a measure of affinity of the enzyme for the substrate and is the substrate concentration where 50% of the maximum reaction velocity is reached. (Cornish-Bowden, 2012; Bisswanger, 2017)

$$\frac{dc_S}{dt} = v = \frac{v_{max} \cdot c_S}{K_m + c_S} \qquad (6)$$

$$k_{cat} = v_{max} \cdot {c_{E,0}}^{-1} \qquad (7)$$

This formula is the basis for enzyme kinetic models defining the basic parameters describing enzymatic conversions. The formula is extended by additional K_m parameters and concentrations when more substrates are involved. To enhance comparability of experimental results, the catalytic constant k_{cat} [s^{-1}] is utilized as a measure of the catalytic activity that is independent of the applied enzyme concentration $c_{E,0}$ [mM] and is calculated using the equation (7). Further modifications are included when the enzyme is inhibited by a substrate, or a (co-) product, introducing inhibition constants K_i [mM] in dependence on the respective inhibitor concentration. In general, an enzyme can be

inhibited reversibly and irreversibly (Bisswanger, 2017). Thereby, the irreversible inhibition is less common in biocatalysis, since this type of inhibition reduces the turnover number and, thus, is contra productive to an efficient biotransformation. On the contrary, reversible enzyme inhibition is relatively common in biocatalysis. Therefore, several methods have been developed to circumvent different inhibition types (Lindeque *et al.*, 2019; Burek *et al.*, 2022).

For instance, to handle product inhibition, *in situ* product removal can be applied. This method maintains low product concentrations through follow-up reactions or product separation processes like adsorption or membrane separation, preventing negative effects on enzyme activity. Similar to this method, excess substrate inhibition can be adjusted by controlling the substrate addition at low concentration levels (fed batch). Cross-inhibition by coproducts or intermediates in multi-enzymatic systems can be minimized or even prevented by spatial compartmentalization of different enzymes (immobilization and membranes) or by completely separating the reaction steps using reaction sequences. (Heintz *et al.*, 2017; Burek *et al.*, 2022)

In general, optimizing the activity involves selecting and maintaining optimal reaction conditions, including the methods mentioned for dealing with inhibition. Optimizing activity beyond natural capacity can only be achieved through protein engineering, manipulating the attributes (e.g. activity, stability, specificity) of the enzyme to fit the requirements of the process (Arnold, 2018; Alonso *et al.*, 2020; Madhavan *et al.*, 2021; Sharma *et al.*, 2021). In order to choose the right process or protein engineering action, the kinetic models can be utilized to predict the performance of a biocatalyst in a process by knowing the sound kinetic expression and the operation mode specific material balance (Vasic-Racki *et al.*, 2003; Sudar *et al.*, 2021; Miličić *et al.*, 2022).

To extract the kinetic model of experiments, MM established the initial rate measurement, where initial reaction rates are measured with varying substrate concentrations to identify the kinetic parameters v_{max} and K_m experimentally (Tang *et al.*, 2010). By the application of initial rates, a quasi-steady state assumption can be made, where every enzyme molecule is saturated with a substrate molecule and the rate limiting step is the dissociation of the enzyme substrate complex [ES]. Thus, the MM equation is only viable with conditions fulfilling these assumptions. (Stroberg *et al.*, 2016; Choi *et al.*, 2017)

In the past, due to low computational power, scientists developed various approaches to evaluate non-linear reaction rate data from initial rate measurements. These approaches were based on various linearization methods by Lineweaver and Burk (Lineweaver *et al.*, 1934), Eadie and Hofstee (Hofstee, 1952) as well as Hanes (Hanes, 1932) allowing simple graphical evaluation of initial rate data. All linearization procedures are prone to deviations in experimental data and can lead to large variances of the derived kinetic parameters dependent on its linearization procedure and the quality of the measured data (Gygli, 2022). These deviations are affected by different substrate concentration extremes for the various linearization techniques. Consequentially, no linearization is suitable for the whole range of concentrations due to the nature of the data transformation and its associated error propagation. Nevertheless, this kind of experiments and evaluation methods supported the research on enzymes drastically and is still conducted nowadays (Tang *et al.*, 2010; Pinto *et al.*, 2019). They give clear first expressions of mechanistic behaviors but may lack in meaningfulness for process predictions (Kruger, 1995; Choi *et al.*, 2017). Although the Michaelis Menten model and the associated measurements are established for over a century, the resulting kinetic parameters may not be suitable to describe processes within production conditions which is desirable. This can happen when the assumptions for the MM equation are not met, when the equation does not take into account micro-kinetic considerations, or when diffusional and thermodynamic limitations affect the enzyme. Furthermore, immobilized enzymes are difficult to apply in standard assay procedures with diffusion limitations or significantly changed kinetic parameters due to structural changes of the enzyme (Cooney, 2017). Those unfavorable conditions can be handled efficiently in a reactor setup applying process conditions for either free or immobilized enzymes. If the course of the reaction is now measured (via substrate depletion or product formation), kinetic parameters can also be determined by means of non-linear curve fitting (Zimmerle *et al.*, 1989; Duggleby, 1995; Nikolova *et al.*, 2008). Since the complete conversion is considered here, it is named progress curve analysis.

In contrast to linearization of initial rates as a function of substrate concentration plots, especially with the increase of online and inline analytical technologies, the non-linear fit of progress curves becomes more popular in recent years (Nikolova *et al.*, 2008; Zavrel, Kochanowski, *et al.*, 2010; Choi *et al.*, 2017; Vang *et al.*, 2022). Although online and

inline analytical devices are highly practical for obtaining continuous high volumes of data, they can be expensive to acquire and require more complex data treatment and analysis to fully realize their potential. For the determination of kinetic parameters under process-relevant conditions, however, online analytics are indispensable, as the data density and real-time analysis is clearly superior to offline analytics. In contrast to initial rate measurements, which are conducted mostly photometrical or with chromatographic methods, progressive curves are more flexible in the choice of analytics. Here, the most efficient way to follow the whole reaction course is the application of non-invasive inline or online analytics like Fourier transformed infrared spectroscopy (Fagaschewski *et al.*, 2015), Raman spectroscopy (Geske *et al.*, 2013) and nuclear magnetic resonance spectroscopy (Friebel *et al.*, 2019; Claaßen *et al.*, 2020), since the amount of data points is only dependent of the measurement interval. Furthermore, multiple sensor locations enable a comprehensive measurement along the length of a tubular reactor or at different locations in a stirred tank reactor (Kara *et al.*, 2011). This allows to obtain a concentration profile of the reactor to increase the data of a single experiment and to analyze the axial reactor performance.

Contrary, offline analytics for progress curve analysis have the disadvantage that every data point needs a taken sample beforehand. In order to reach the same data density, a non-feasible quantity of samples have to be taken and analyzed manually. Every sample taken means a loss in volume of the reactor which can alter the behavior of the reaction mixture and, therefore, deteriorate the experiment. Thus, if offline analytics have to be applied, the sample size and sampling interval has to be chosen wisely to interfere the experimental conditions as less as possible. For that reason, the use of online and inline analysis is particularly useful because it eliminates the time delay caused by sampling and sample preparation, allows information to be generated directly from the measurements, and even allows iterative optimization approaches to be implemented. As a result, this type of analysis is useful for automating data analysis of experiments and enabling time-efficient approaches such as transient flow experiments. (Hess *et al.*, 2021; Taylor *et al.*, 2021)

When the data of a progress curve are available, the measured data has to be described by a differential equations for change of product over time to obtain kinetic parameters out of the progress curve. This type of data treatment was already established by MM in 1913,

but highly laborious since personal computers were not developed yet (Michaelis *et al.*, 1913). Thus, the integration rather than the differentiation of differential equations was manually done. Applying the quasi-steady-state assumption of MM, the differential equation results in (8) where k_2 corresponds to the dissociation constant of the enzyme substrate complex. (Cornish-Bowden, 2012)

$$t = \frac{1}{k_2 \cdot c_E} \cdot c_P + \frac{K_m}{k_2 \cdot c_E} ln \frac{c_{S,0}}{c_S} \tag{8}$$

Nowadays, a variety of computer programs exist, which can perform model identification on experimental data (e.g. Dynafit, FITSIM). These programs offer different methods for the numerical fit of experimental data based on different kinetic models based on reaction mechanisms (Zavrel, Kochanowski, *et al.*, 2010; Johnson, 2013). Today, even basic spreadsheet calculators such as Excel (Microsoft, Redmond, Washington, USA) and Origin (OriginLab, Northampton, Massachusetts, USA) have the capability to easily perform nonlinear fits. As a result, progress curve analysis has become simpler than ever before. The advantage of this type of kinetic modelling is the amount of data obtained of a single experiment which can lead to first results after one single experiment. In order to obtain highly precise kinetic parameters, a minimum of one experiment per parameter is recommended (Zimmerle *et al.*, 1989). Specialized kinetic model software has the benefit of performing Monte Carlo simulations to evaluate the measured data and the significance of the model parameters (Nikolova *et al.*, 2008). This is necessary since most experiments are still conducted manually, which can lead to poor experimental or model design.

Also the combination of progress curve analysis and initial rate measurement is possible. In the approach of Tang *et al.* (2010), the assumption of irreversibility and negligible product inhibition is achieved by a second enzyme, which further catalyzes the product of the monitored reaction to species on which the enzyme of interest is not active. Thus, full conversion is achieved and the product over time plot directly translates to many data points in the reaction rate over substrate concentration. Therefore, a double reciprocal linearization procedure can be applied to determine kinetic parameters. (Tang *et al.*, 2010) Ultimately, all these methods involve manual experimental design and execution, which come with inherent disadvantages for the precise determination of kinetic parameters. Manual approaches require careful planning and conduction of experiments, performing

repetitive tasks that can be time consuming and prone to errors. Especially, the human error is a significant concern, as inaccuracies in experimental planning, pipetting and timing can impact the accuracy and reliability of the results. Additionally, along the challenges of manual procedures, there are limitations in sample throughput. This hinders the execution of high throughput experiments and, therefore, limits the accuracy of the kinetic fits and valuable information about the dynamics of the reaction may be overlooked. Additionally, the selection of the underlying kinetic model can lead to subjective data interpretation, introducing biases and inconsistencies that hinder objectivity and reproducibility. Finally, lab-scale kinetic studies serve as valuable starting points for understanding enzyme characteristics and process design, but lack significance under process conditions, different operation modes, and enzyme preparations.

Recent advancements in automation and data processing have the potential to improve the described negative aspects of conventional kinetic studies. For instance, experiments can be designed closer to the actual process to ensure good scalability, automated execution of experiments can minimize human error and model identification methods can be used to make rational choices about kinetic models.

2. Motivation and Objectives

The transition from the classical chemical industry to a sustainable bio-based industry is a major goal in facing climate change. Biotechnological processes are considered influential drivers for this industrial transformation (Sheldon, 2020). Nevertheless, few biocatalytic processes have reached industrial maturity to replace existing processes, as they require extensive research and development (van Schie *et al.*, 2021). Understanding the kinetics of these processes and rationalizing them through mathematical models is necessary to develop competitive technologies (Miličić *et al.*, 2022). These models play a major role in reaction engineering. However, the experimental and mathematical methods used to identify them have not been standardized, which can lead to varying results based on the researcher's decisions and expertise (Pleiss, 2021; Gygli, 2022; Lauterbach *et al.*, 2023). This statement appears to contradict recent advancements in cost-efficient online analytics, straightforward automation procedures, and software accessibility (Häse *et al.*, 2019; Waldron *et al.*, 2020).

This thesis was conducted in collaboration with the Institute of Process System Engineering to create a continuous reactor platform that achieves automated and data-driven model identification for enzymatic reactions. The objective was to enhance the accuracy and reproducibility of enzyme kinetic modeling using modern statistical and software solutions for experimental automation. The aim of this work was to combine the statistical method of model-based design of experiments (MBDoE) with experimental automation. To qualitatively compare different model candidates, the Akaike information criteria (AIC) was chosen, which serves as a measure of the model's goodness of fit while taking into account the number of model parameters (Akaike, 1974; Michalik *et al.*, 2010). The parameter estimation should be further improved with optimally designed and automated executed experiments.

- Setup of a continuous automated reactor for incremental experiments
- Closed-loop identification of kinetic models of a model reaction
- Comparison of MBDoE and conventional statistical approach
- Investigation of influential factors on the model identification
- Evaluation of the AIC for model discrimination

3. Theoretical Background

This chapter presents the theoretical background needed for a better understanding of the subject. It presents the model system and model candidates to which the proposed model identification method is applied. Furthermore, it offers an explanation for classical methods of statistical experimental design and contrasts them with the data-driven and model-based approach.

3.1. Model Reaction

For the investigation of the proposed model identification approach, a model reaction is needed, which offers the following features:

- Commercially available enzyme
- Irreversible reaction without side products
- Available inline analytics in flow for reaction monitoring
- Variety of model candidates in literature
- No cofactor dependency

The enzyme formate dehydrogenase (FDH) from *Candida boidinii* (EC 1.17.1.9) is a well described biocatalyst in literature due to the importance as a cofactor regenerating enzyme (Hummel *et al.*, 1987; Kragl *et al.*, 1996; Liese *et al.*, 2006; Bolivar *et al.*, 2007; Schmidt *et al.*, 2009; Binay *et al.*, 2016). It catalyzes the reduction of nicotinamide adenine dinucleotide (NAD^+) to NADH with formate as hydride donor (referred to as $HCOO^-$ in figures), which serves as cofactor for enzymatic reactions in the class of oxidoreductases (cf. Figure 4). Thus, the kinetic modelling of this enzyme is of high interest for industrial applications. The NADH product can be easily detected through UV/Vis spectroscopy at 340 nm because of the isolated maximum at this wavelength. Furthermore, the enzyme activity is not affected significantly by the co-product CO_2 due to its low solubility and existence as carbonate at alkaline pH.

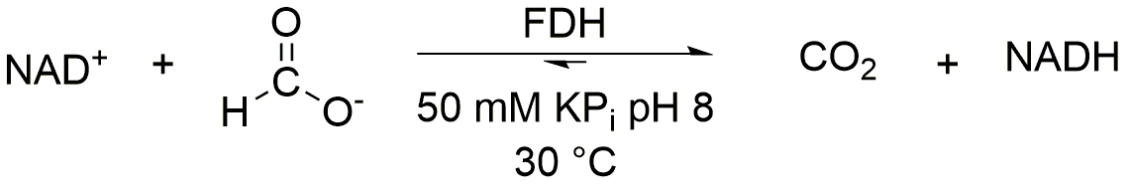

Figure 4: Reaction scheme for the reduction of NAD^+ to NADH catalyzed by FDH from *Candida boidinii*. The indexes of the kinetic parameters are labeled with the according to the figure with the letter A for NAD^+, B for formate and Q for NADH to increase readability in Table 1.

Furthermore, the large variance of kinetic models and parameters describing the catalyzed reaction serve as suitable candidates for the model identification approach in this work. The considered equations of the model candidates presented in Table 1 consist of macroscopic textbook models for enzymatic reactions with two substrates by MM without inhibition (MM), with competitive product inhibition (CPI), with uncompetitive product inhibition (UPI) and non-competitive product inhibition (NCPI). The product inhibition by NADH is represented with the kinetic parameter $K_{i,Q}$ [mM]. This list is extended by comprehensive mechanistic literature models of Schmidt *et al.*(2009) (SCHM) and Michalik *et al.*(2007) (MICH) and is complemented by their lumped variants (SCHMlump, MODI). These variants lump the parameters $K_{i,A}$ and $K_{m,B}$ into the single parameter P_1 to decrease model complexity and, therefore, improve the parameter confidence intervals (Michalik *et al.*, 2007). The non-lumped mechanistic models are less simplified than macroscopic models because they involve present micro-reaction steps and are validated by experimental data (Michalik *et al.*, 2007; Schmidt *et al.*, 2009). All models share the assumption that carbon dioxide does not affect the activity. In open systems with neutral or acidic pH, carbon dioxide has low solubility. At alkaline pH, carbon dioxide exists as carbonate, which is not usable by the FDH. Despite the large deviations in the number and value of the different model parameters derived at different assay conditions, all models are capable of predicting the reaction progress within the considered range. Consequently, this reaction system fulfills all features required for the demonstration of the proposed project.

Table 1: Kinetic model candidates for the FDH from *Candida boidinii*. The indices A, B and Q are accounted to NAD^+, formate and NADH respectively. P_1 is a lumped parameter combining $K_{i,A}$ and $K_{m,B}$ multiplicatively. The maximum reaction rate is $v_{max} = k_{cat} \cdot c_{E,0}$.

	Model
Michaelis Menten (MM)	$$v = \frac{v_{max} \cdot c_A \cdot c_B}{(K_{m,A} + c_A) \cdot (K_{m,B} + c_B)}$$
Competitive Product Inhibition (CPI)	$$v = \frac{v_{max} \cdot c_A \cdot c_B}{\left(\left(K_{m,A} \cdot \frac{1 + c_Q}{K_{i,Q}}\right) + c_A\right) \cdot (K_{m,B} + c_B)}$$
Uncompetitive Product Inhibition (UPI)	$$v = \frac{v_{max} \cdot c_A \cdot c_B}{\left(K_{m,A} + c_A \cdot \frac{1 + c_Q}{K_{i,Q}}\right) \cdot (K_{m,B} + c_B)}$$
Non-competitive Product Inhibition (NCPI)	$$v = \frac{v_{max} \cdot c_A \cdot c_B}{\left(\left(K_{m,A} \cdot \frac{1 + c_Q}{K_{i,Q}}\right) + \left(c_A \cdot \frac{1 + c_Q}{K_{i,Q}}\right)\right) \cdot (K_{m,B} + c_B)}$$
Schmidt *et al.*, 2009 (SCHM)	$$v = \frac{v_{max} \cdot c_A \cdot c_B}{K_{i,A} \cdot K_{m,B} + K_{m,A} \cdot c_B + K_{m,B} \cdot c_A + c_A \cdot c_B}$$
Michalik *et al.*, 2007 (MICH)	$$v = \frac{\frac{v_{max}}{K_{i,A} \cdot K_{m,B}} \cdot c_A \cdot c_B}{1 + \frac{c_A}{K_{i,A}} + \frac{K_{m,A} \cdot c_B}{K_{i,A} \cdot K_{m,B}} + \frac{c_A \cdot c_B}{K_{i,A} \cdot K_{m,B}} + \frac{c_Q}{K_{i,Q}} + \frac{K_{m,A} \cdot c_B \cdot c_Q}{K_{i,Q} \cdot K_{m,B} \cdot K_{i,A}}}$$
Michalik *et al.*, 2007 (MODI)	$$v = \frac{\frac{v_{max}}{P_1} \cdot c_A \cdot c_B}{1 + \frac{K_{m,A} \cdot c_B}{P_1} + \frac{c_A \cdot c_B}{P_1} + \frac{c_Q}{K_{i,Q}} + \frac{K_{m,A} \cdot c_B \cdot c_Q}{P_1 \cdot K_{m,B}}}$$
Schmidt *et al.*, 2009 (SCHMlump)	$$v = \frac{v_{max} \cdot c_A \cdot c_B}{P_1 + K_{m,A} \cdot c_B + c_A \cdot c_B}$$

Furthermore, the reaction can be applied in continuous reactor concepts easily, because the main product is a cofactor and does not need an additional cofactor addition, allowing an easy and cost-efficient detection of the product by flow-through UV/Vis spectroscopy. Consequentially, this model reaction is particularly suitable for the application in an automated continuous reactor system.

3.2. Design of Experiments

In order to enable the objective of an incremental working experimental platform with an MBDoE approach defined in chapter 2, an efficient way is needed to provide initial data to fit the model candidates to. This initial data set has to be established as the basis for the further incremental approach. In consequence, a statistical method is needed to extract the most information out of the least number of experiments. This method is called "Design of Experiments" (DoE) and includes different grids to cover a distinct parameter space. In general, a DoE approach is utilized to describe the response of a system when varying input factors (e.g. concentration, temperature, pressure, time) in a systematic manner. In order to design an experimental campaign, the upper and lower limits of the factors have to be specified. The decision is based on limitations of the system (e.g. detection limit of analytics, flow limits of pumps, solubility of substrates) or by preliminary results in literature or the experimenter itself. When the parameter limits are set, the combination of experiments is generated based on the design chosen. The most common designs are full factorial, fractured factorial and central composite design, which represent different distributions of factor combinations in the parameter space. (Draper *et al.*, 1996)

These designs are applied for different research targets, experiment workloads or responses. The **factorial design**, for example, serves as an initial exploration of the parameter space, by the combination of every possible factor level with each other. When the number of experiments exceeds feasibility, a **fractured factorial design** can lead to similar results with less experiments. Here, the resolution is a key feature, which dictates the number of main effects and interactions taken into account for the experimental screening. (Draper *et al.*, 1996) An example for a frequently applied design of experiments is the **central composite design** in Figure 5, which is often utilized for the so-called response surface methodology (Illanes *et al.*, 2002; Francis *et al.*, 2003; Bandara

et al., 2019; McMullen *et al.*, 2022). This design consists of a multi-level factorial design with a center point, describing the median of both factors. In addition, the design is completed with axial points, which are like center points, but for each axial point one factor is changed to a level outside of the parameter space. By this, the responses can be fitted with a quadratic model easily, whereas the curvature of the design space is specified with the axial points. When the best design for the desired application is selected, the experiments are conducted and the results are collected. Now the response can be fitted with a model, which can describe the main effects of the individual factor combinations. For response surface plots, the responses are often fitted with quadratic models to reach a certain curvature, for the determination of maxima and minima in the design space. This yields an intuitive visual form of the experimental response as a topographic graph with maxima and/or minima as peaks and valleys. The resulting model is then validated by predicting responses of experiments outside of the original experimental design.(Telford, 2007; Khuri *et al.*, 2010)

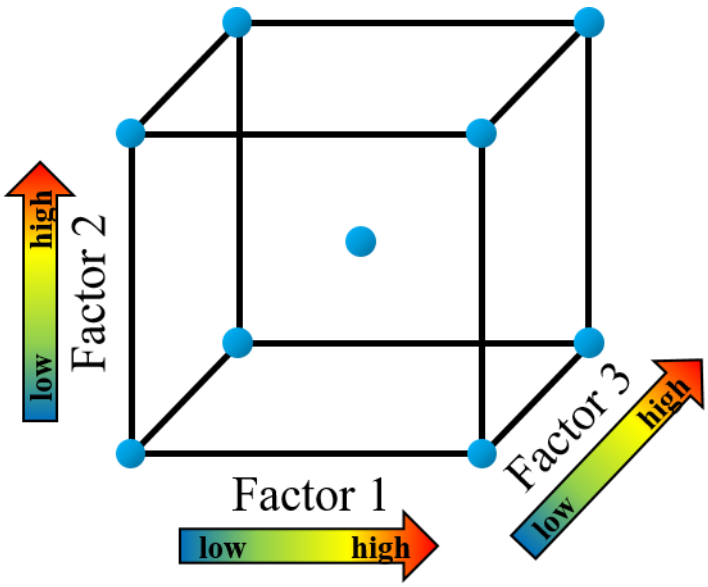

Figure 5: Exemplary design space of a 2^3 full factorial design with accordingly 2 levels (low and high), 3 input factors and a center point.

For this work, the responses are reaction rates dependent on the initial substrate concentrations and the fitted kinetic parameters on the established initial DoE according to the model candidates provided. Since the model candidates are known in this work, they can be applied to increase the relevance of the incrementally designed experiments.

Thus, the efficiency of the next experiments is increased in contrast to conventional DoE grid without comprehensive models in mind. This type of incremental experimental design is called MBDoE. (Franceschini *et al.*, 2008; Galvanin *et al.*, 2010; Michalik *et al.*, 2010; Zavrel, Michalik, *et al.*, 2010)

3.3. Model-Based Design of Experiments

The MBDoE procedure aims to optimize the sequential order of experiments, allowing for efficient exploration of the design space of complex systems. The objective of this approach is to gain a better understanding and comprehensive description of the underlying system by creating precise mathematical models with statistical relevance. Combining kinetic model candidates of enzymes and statistical techniques, experiments yield empirical data with the highest information content in relation to the objective functions. The objective functions are categorized into two separate targets within the kinetic modelling procedure: model discrimination and parameter estimation. These categories are automatically optimized within the experimental setup. Through initial discrimination of models and subsequent estimation of parameters, the aim of this work is to identify kinetic models efficiently.

3.3.1. Model Discrimination

The basis of the model discrimination performed in this project are several kinetic model candidates from literature and standard MM kinetics including different inhibition types (CPI, NCPI, UPI) are added in the pool of model candidates (Table 1). These models serve as regression models in the design space of the given parameters and their dependency to the experimental data. Since every model can fit the provided data, the suitability has to be evaluated. The most prominent way to evaluate model candidates is the goodness of fit. A measure for that is the log-likelihood which describes the location of the center of the normal distribution of experimental data around an estimation (Breitung *et al.*, 1994). The normal distribution is written as:

$$p(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} \tag{9}$$

Here, μ is the location of the mean for the normal distribution and σ is the standard deviation describing the width and the height of the distribution. In order to determine the likelihood L of a normal distribution given experimental data y, equation 10 is applied.

$$L(\mu, \sigma|y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(y-\mu)^2/2\sigma^2} \tag{10}$$

In order to simplify the derivation procedure of the likelihood to find its maximum, it is transformed with the natural logarithm, which has the same maximum as the original likelihood. For the computational procedure, the prior transformation results in shorter computation time regarding the optimization of the log-likelihood. Additionally, most optimization procedures are minimization functions and, therefore, the model discrimination is optimized by minimizing the negative log-likelihood according to equation (11). Here, with given data y_i, the parameter vector θ is minimized regarding the sum of the negative log-likelihood.

$$arg_\theta \max \sum_{i=1}^{n} \ln L(\theta|y_i) = arg_\theta \min - \sum_{i=1}^{n} \ln L(\theta|y_i) \tag{11}$$

Now, when comparing different regression models, a model with a larger number of parameters will naturally fit given data better due to the extra degrees of freedom. Thus, an information criterion can be applied to add penalties to the likelihood for a more reasonable model discrimination. One criterion which takes the quantity of parameters *k* into account is the AIC and defined in (12) (Akaike, 1974, Cavanaugh *et al.*, 2019). This information criterion was already successfully applied in different publications to compare kinetic models for enzymes. (Schmider *et al.*, 1996; Michalik *et al.*, 2010; Bezerra *et al.*, 2013, 2016; Mueller *et al.*, 2022)

$$AIC = 2g - \ln L(\theta|y_i) \tag{12}$$

With this criterion larger models are only considered, when the amount of parameters (g) add an increase in the goodness of fit overcoming the factor of 2g, which is considered as a balanced penalty between model complexity and precision (Cavanaugh *et al.*, 2019). Since this information criterion contains a fixed weighted penalty term, the needed computational power is the lowest among others. For this reason, this criterion is appropriate for automation without powerful computer hardware and optimized scripts. (Burnham *et al.*, 2002)

Finally, the differences of the obtained values of the AIC can be applied to estimate the relative likelihood of the models. By the introduction of Akaike weights w_b in (13), a relative measure for the goodness of fit, is implemented (Burnham *et al.*, 2002). Here the difference in the AIC of the best model b to the observed model o ($\Delta_{k,c}$) is calculated and put in contrast to the sum of differences of j model candidates.

$$w_b(\theta, u, y, \sigma^2) = \frac{\exp(-0.5\Delta_{b,o})}{\sum_{j=1}^{m} \exp(-0.5\Delta_{b,o})} \tag{13}$$

$$\text{With:} \qquad \Delta_{b,o} = AIC_b - AIC_o \tag{14}$$

The value between zero and one represents the probability that the observed model is the best among the candidates. The sum of all compared Akaike weights adds up to one. Models with an AIC weight close to one represent the experimental data better than models with an AIC weight close to zero. Therefore, Akaike weights can be used as a design criterion for the MBDoE approach, known as the Akaike weights design criterion (AWDC) (Michalik *et al.*, 2010):

$$\max_{u} \phi_{AWDC}(u) = \sum_{o=1}^{m} p_o w_{o,o}(\theta, u, x, \sigma^2) \tag{15}$$

The objective function (15) aims to maximize the Akaike weights such that among the model candidates, a single model reaches an AIC of 1 and the additional models strive to

decrease to an AIC of 0. In the AWDC, several scenarios are calculated in which one model is pretended to be the best model among the candidates. Each scenario includes a new experiment decreasing the AIC of the selected model. Meanwhile, the AWDC targets the maximization of the sum of Akaike weights calculated for all approximated scenarios. In the AWDC m scenarios are considered with an Akaike weight of the supposed best model $w_{o,o}$ and the prior probability or Akaike weight p_o of model o. Since the AWDC is based on the Akaike weights, a direct interpretation of the model discrimination based on different experimental scenarios can be deducted. (Michalik *et al.*, 2010)

For comparison, two additional information criterion are calculated, which are also common in model comparison. These are the Bayesian information criterion (BIC) defined in (16) and the corrected AIC (AICc) defined in (17). (Ward, 2008)

$$BIC = g \cdot \ln(n) - \ln L(\theta|y_i) \quad \textbf{(16)}$$

$$AICc = 2g - \ln L(\theta|y_i) + \frac{2g(g+1)}{(n-g-1)} \quad (17)$$

In this work, the AIC was utilized, which does not take into account the number of observations *n* which is equally low and limited in the applied reactor setup. While in this case BIC and AIC agree on the model discrimination, for larger data sets the BIC may identify the correct model in contrast AIC which would always designate the best performing model. (Chakrabarti *et al.*, 2011)

3.3.2. Parameter Estimation

After the initial model discrimination results in a final model structure with distinct model paramaters, the next target of the MBDoE is the improvement of the parameter accuracy. For this purpose, the Fisher information matrix (FIM) (cf. equation 19) is utilized to design optimal experiments to decrease the confidence intervals of the model parameters. The FIM measures the sensitivity of the likelihood function with respect to the model parameters. It provides information about the expected curvature and variability of the

likelihood function close to the true parameter values. Thus, it quantifies the information content of the data about the kinetic model parameters. The sensitivity matrix S is built with the partial derivates of the model predictions on the basis of the model parameters in (18). The derivates of the parameters are calculated numerically with the complex step approximation due to the absence of explicit solutions for the kinetic models in this work. (Franceschini *et al.*, 2008)

$$S = \left[\frac{\partial y}{\partial \theta_1} \cdots \frac{\partial y}{\partial \theta_n}\right]^T \tag{18}$$

$$FIM = S \cdot S^T \tag{19}$$

The FIM can be applied in different optimal design criterions to increase the accuracy of the parameter estimation. These criteria include the *d-optimal*, *a-optimal*, and *e-optimal* criteria. The *d-optimal* criterion maximizes the determinant of the FIM, which optimizes all parameters simultaneously. However, this criterion has limited applicability if a parameter is poorly determinable in the design space. The *a-optimal* criterion minimizes the trace of the inverse FIM to reduce the average variance of the model regression, or in other words, the uncertainty of the model. This criterion has the same application limitation as the *d-optimal* criterion. The *e-optimal* design criterion is considered the most applicable approach for advanced parameter estimation, as the parameters of kinetic models of enzymes naturally exhibit high correlations. This objective function aims to maximize the lowest eigenvalues λ_{min} of the FIM which represents the largest parameter error equation (20). Thus, the most erroneous parameter is optimized incrementally by minimizing the largest error of the covariance matrix. (Martins *et al.*, 2003; Ohs *et al.*, 2017)

$$\max_u \phi_E(u) = \lambda_{min}(FIM) \tag{20}$$

With this method, parameter estimation is improved once model discrimination has been performed and therefore is the efficient in estimating the desired parameters accurately while utilizing the available resources effectively. However, the error of the parameters is strongly dependent on the number of experiments performed and the error of the measurement method. This must be taken into account when evaluating the kinetic models and their parameters.

4. Setup and Results of the Automated Reactor Platform

Since enzymes can face challenges regarding their process stability, they are often immobilized on a solid support matrix. This not only offers an easy mechanical retention of the catalyst in the process but also provides higher stability. Moreover, the immobilization on spherical particles, monoliths and reactor walls makes enzymes accessible for different types of micro reactor technologies, which are of high interest in the production of fine chemicals and active pharmaceutical ingredients (Bennett *et al.*, 2019; Baumann *et al.*, 2020). In this context, high heat and mass transfer, high surface to volume ratios and constant product quality are driving forces for this technology. Porous spherical particles are used in this thesis due to their ability to achieve the highest surface to volume ratios. In addition, these features are extremely useful for the comprehensive analysis of enzymes in an automated manner as performed in this thesis. Highly stable catalysts can be applied for several experimental campaigns without a significant loss in activity, achieving high reproducibility in the resulting data. On the other hand this opens up the possibility to determine catalyst deactivation rates at real process conditions in a continuous operation mode. Furthermore, the miniaturization enabled the application of transient and fast changing conditions experiments. In this work, a PBR was employed to test several comparable iterations of automation on the same batch of immobilizate with a minimal application of enzyme, chemicals and solvent.[1]

4.1. Closed-Loop Setup

4.1.1. Design of Reactor System

A closed-loop setup describes a reactor setup with a sequential procedure of experiments designed to iteratively improve data for identifying relationships between input and output variables. In the case of kinetics of the FDH, the input variables are substrate concentrations (NAD^+ and formate) and residence times, which influence the reaction rate as an output variable. The relationship between the variables is described by the kinetic model, which is automatically identified in this work. Therefore, a closed-loop approach for kinetic modeling of enzymes in a PBR requires sufficient flexibility to create data-rich experiments without knowing the exact requirements beforehand. This leads to the use of a dosing system to apply varying substrate concentrations or a range of

[1]Parts of this chapter will be published in *"Closed-loop Identification of Enzyme Kinetics Applying Model-Based Design of Experiments"* by Hennecke *et al.*

residence times to control conversion in the reactor system. To screen the biocatalyst, a practical range of measurable concentrations, residence times, and dilution factors were selected, which were possible to achieve in the applied reactor setup (Figure 6). These framework conditions were set by considering the available hardware components and by searching the literature on kinetic parameters to specify concentration ranges.

With respect to achieve steady-state conditions in a continuous reactor system and to enable consistent measurement in the flow cell, reliable pumps were required. Furthermore, the pumps should be implementable in Python to realize an automation of different residence times and dilutions of substrate stock solutions. Python is a programming language that is widely used in research due to its clear syntax, versatility in areas such as data analysis and scientific computing, and extensive standard library (Short *et al.*, 2019; Gygli, 2022; Ryzhkov *et al.*, 2023). As an open-source programming language, Python promotes the open exchange of code and knowledge, providing researchers access to a wealth of freely available resources and facilitating collaboration within the scientific community. This programming language is well-suited for keeping the project free of commercial obstacles. First experiments indicate a high flow pulsation when applying a micro-gear pump (mzr-2905-1082, HNP Mikrosysteme GmbH, Schwerin, Germany), which disturbed the online UV/Vis measurement. Another alternative were syringe pumps (AL-1000, World Precision Instruments Germany GmbH, Saragosa, USA), which are cheap, reliable, programmable and easy to implement in a dosing system by the application of their connection in series (daisy chain). In contrast to micro-gear pumps, syringe pumps are resistant to backpressure and have low flow pulsation. The downside of these type of pumps is the limited amount of volume due to the limited volume of the syringes as reservoir. However, if experiments are optimized, the necessary component volume should not pose a challenge. In the end, the accurate dosing and uncomplicated implementation led to the choice for this type of pumps. Four of these pumps were implemented as dosing system to reach adjustable inlet concentrations. Desired inlet concentrations were set via the relative flow velocity of one substrate pump and the corresponding dilution pump. The concentration was set twice as high in the first mixing step of a T-mixer. In order to obtain the desired concentration with equally mixed substrate streams, a second T-mixer was introduced. The generated substrate stream was pumped through the PBR which consisted of an empty column filled with immobilized enzyme. The particles were retained by metal meshes at the in- and

outlet of the column. Ensuring a constant temperature of 30 °C in the reactor, the column was housed in an empty preparative chromatography column. Through this, the heating liquid of circulating thermostat was pumped with a volume flow of 40 mL min^{-1}. The outlet stream of the reactor is channeled through a flow cuvette with a volume of 62 µL. Here, the concentration of NADH is measured at 340 nm every 5 s. The volumetric flow is measured afterwards in a flowmeter and could be applied as an input argument for the flow control. The product stream (waste) is collected in a glass beaker.

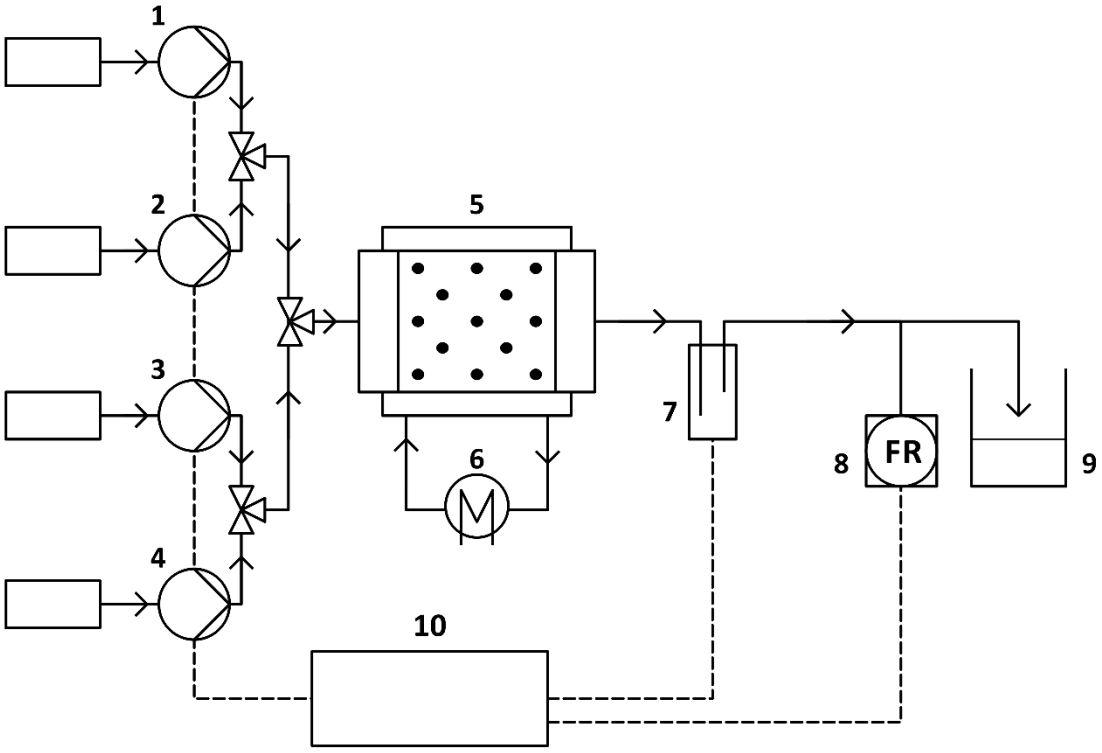

Figure 6: Schematic reactor setup of the automated platform for the MBDoE and model identification of enzymes. (1) Syringe pump with 4 mM NAD^+ in 50 mM KP_i at pH 8 (2) syringe pump with 50 mM KP_i at pH 8 (3) syringe pump with 600 mM formate in 50 mM KP_i at pH 8 (4) syringe pump with 50 mM KP_i at pH 8 (5) PBR (6) thermostate (7) UV/Vis spectrometer (8) flow meter (9) product container (10) measurement and control computer.

The applied UV/Vis spectrometry (HR4000, Ocean Insight, Orlando, USA) exhibits upper and lower detection limits dependent on the optical path length. Available diameters of flow cuvettes were 1 mm and 10 mm which translates to a theoretical measurable NADH concentration of 0.02 mM to 2 mM and 0.002 mM to 0.2 mM, respectively. However, applicability is only given in the linear range of the calibration

(cf. Figure 46). Preliminary assays and literature values indicate $K_{m,NAD}$ values of approx. 0.04 to 0.1 mM NAD^+ (Kato *et al.*, 1979; Kula *et al.*, 1980; Schmidt *et al.*, 2009; Guo *et al.*, 2016). When immobilizing enzymes, the K_m value tends to increase, due to structural changes, adsorption of product or even diffusion limitation. To estimate v_{max} and k_{cat}, it is common practice to experimentally apply a substrate concentration that is 10 times the K_m value. This ensures that the maximum activity is reached at a level of at least 90%. (Bisswanger, 2017). For the determination of K_m, the heuristic approach is to measure initial rates at a tenth of a prior estimated K_m. Thus, the 1 mm cuvette was selected to be applied in the automatic system to cover the region of v_{max} and most likely the apparent K_m.

The PBR consisted of a chromatography column (Xela, Isera GmbH, Düren, Germany) with an inner diameter of 3 mm and a length of 5 cm packed with FDH immobilized on a methyl acrylate epoxy resin matrix (Lifetech™ ECR8204M, Purolite Ltd, King of Prussia, USA). For the exact setting of the flow rate, the void volume of the packed PBR had to be determined. Therefore, the void volume of the PBR with ECR8204M resin was calculated based on the particle mass $m_{Particle}$ of 392.7 mg and water mass m_{Water} of 631.8 mg which was applied to reach a fixed total volume V_{Total} of 1 mL inside a measuring cylinder. According to equation (21) the water mass m_{Water} at 24 °C and the according density ρ_{Water} (24 °C) was used to calculate the water volume V_{Water} to 633.5 µL. With the resulting particle density of 1.07 g·cm^{-3} and the empty reactor volume $V_{Reactor}$ of 353.4 µL, the void volume was calculated with equation (23) and resulted in 145.4 µL.

$$\rho_{Particle} = \frac{m_{Particle}}{V_{Total} - \frac{m_{Water}}{\rho_{Water}}} \tag{21}$$

$$V_{Reactor} = V_{Particle} + V_{Void} = \frac{m_{Particle}}{\rho_{Particle}} + \frac{m_{Fluid}}{\rho_{Fluid}} \tag{22}$$

$$V_{Void} = V_{Reactor} - \frac{m_{Particle}}{\rho_{Particle}} = \frac{m_{Fluid}}{\rho_{Fluid}} \tag{23}$$

After selecting the pumps, the void volume was calculated to determine the appropriate stock concentrations and maximum dilution factors for the design space. Since the determination of v_{max} requires measurements at high substrate concentrations ($c_S >> K_{m,S}$) the maximum measurable concentration of NADH was selected as maximum inlet concentration for NAD^+ (2 mM). In order to utilize the linear force of all pumps and to avoid reaching the non-linearity of the UV/Vis measurement at maximum conversion, the full concentration was chosen close to the maximum (1.9 mM NAD^+) with 4 mM NAD^+ stock concentration. Preliminary experiments showed that a maximum flowrate of a single syringe pump of approx. 30 µL·min^{-1} was feasible with the backpressure by the applied periphery and reactor with 145.4 µL void volume. Therefore, the resulting minimum dilution ratio in form of relative volume flow resulted in 1:6. Accordingly, with a maximum inlet concentration of 1.9 mM, the minimum inlet concentration was adopted to be 0.35 mM. Although the limit of quantification is lower at 0.143 mM, the higher concentration ensures that scenarios with lower conversion rates do not fall below the limit of quantification. The maximum concentration of formate for the kinetic determination was determined according to literature values for $K_{m,formate}$ of 29.3 mM (Schmidt *et al.*, 1992) and preliminary results of kinetic measurements of soluble FDH in that range. Thus, applying the heuristics again, a maximum concentration of 290 mM formate was appointed. Due to the hardware limitations, the maximum dilution resulted in a minimal formate concentration of 50 mM.

For the automated process feasible residence times were specified to run the reactor platform without stalling of pumps. Already during the testing of the dosing system, the resistance of the reactor system and the resulting backpressure were limiting for choice of flow rates and dilution factors. The reason for the high backpressure is that polytetrafluorethylene (PTFE) tubing with an inner diameter of 0.25 mm were applied to reduce the system volume. For that reason, the backpressure of the system was already high without the catalytic bed. First experiments with methyl acrylate epoxy resin (Lifetech^{TM} ECR8204F, Purolite Ltd, King of Prussia, USA) as catalyst matrix, which have a particle diameter of 150 µm to 300 µm and a pore diameter of 30-60 nm, led to stalling pumps at very high (1.9 mM NAD^+ and 290 mM formate) and very low inlet concentrations (0.35 mM NAD^+ and 50 mM formate). In these experiments, the cause of the problem was that one pump of each dilution pair was running at significantly higher speeds than at medium concentrations. Therefore, particles with the same functional

groups and pore diameter but larger particle diameter of 300-710 nm (Lifetech™ ECR8204M, Purolite Ltd, King of Prussia, USA) were chosen. With the application of a minimum residence time of 2 min, inlet concentrations in the selected design space were freely adjustable with four 12 mL syringes mounted on syringe pumps, which made automated experiments possible.

One critical phenomenon, that could occur in heterogeneous catalysis and had to be evaluated, are mass transfer limitations. With the application of porous carriers, diffusion processes occur for the substrate and the product. The substrate requires convective transport through the liquid phase, film diffusion through the laminar boundary layer between the liquid phase and the solid particle, and potentially pore diffusion to the enzyme inside the particle. After the substrate is converted to product, it must undergo mass transfer from the pore to the bulk phase once more. Since diffusion limitations may lead to apparently low reaction rates, an evaluation of their presence is necessary. In order to analyze the presence of film diffusion, two PBRs with 5 and 3 cm length and an inner diameter of 3 mm were prepared with the same batch of immobilizate to compare their performance. The 5 cm reactor has a void volume of 149.4 µL and requires a higher volume flow to achieve the same hydrodynamic residence time as the 3 cm reactor with a void volume of 94.0 µL. When diffusion limitations are existent, the product stream of the 5 cm reactor would result in increased conversion due to the reduced film thickness at the interface between the liquid phase and solid particle phase with increased fluid velocity. Therefore, the diffusion of the substrate is not as limiting as with a higher film thickness at lower volume flow. By the application of the same residence time, the reaction rate of the enzyme should be theoretically equal and only be modified by diffusion limitation. For that reason, the same hydrodynamic residence time was applied experimentally at four different flow rates and two different initial substrate concentrations to test for diffusion limitation. The NADH concentration was measured at steady state conditions to compare the conversion of the 3 cm reactor and the 5 cm reactor (cf. Figure 7).

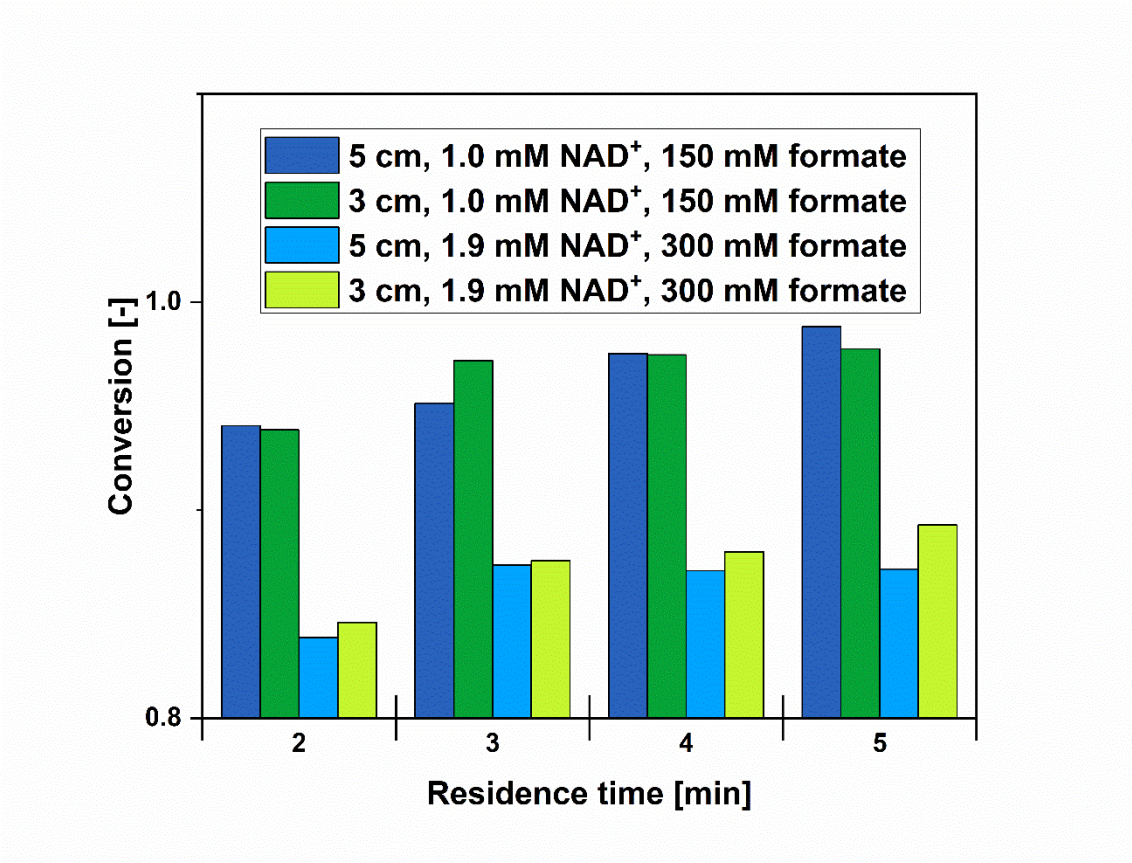

Figure 7: Investigation of film diffusion limitation at hydraulic residence times of 2, 3, 4 and 5 min and two bed length of 3 cm and 5 cm with the inner diameter of 3 mm. Applied substrate concentrations were 1 mM NAD^+ and 150 mM formate and 1.9 mM NAD^+ and 300 mM formate in 50 mM potassium phosphate buffer (KP_i) at pH8.

Figure 7 displays a paired column diagram of individual residence times ranging from 2 to 5 min. The reactor lengths are indicated by color (blue: 5 cm and green: 3 cm), and the two substrate concentrations are distinguished by color intensity (dark: 1.0 mM NAD^+ and 150 mM formate, light: 1.9 mM NAD^+ and 300 mM formate). Both concentration levels exhibit similar conversion trends for different residence times. For all experiments, conversion increased with longer residence times. However, the difference in conversion between residence times of 3 to 5 minutes was small, indicating that the near maximum conversion is converged at approximately 97% for 1 mM NAD^+ and 87% for 1.9 mM NAD^+ at steady state conditions. For the experiments with 1.9 mM NAD^+ the 3 cm reactor displayed a rather increased conversion which accounts for 1.1% in the mean or 0.02 mM. This deviation is insignificant compared to the limit of quantification of 0.14 mM (cf. Figure 46) and the smallest error of 0.06 mM observed in the repeatability

study (cf. Figure 13). Thus, a systematic increase in conversion was not observed for the 5 cm reactor with a higher volume flow. The small deviation was accounted to small differences in packing, particle size distribution or noise in the analytics.

Due to the low change in conversion at higher residence times, it is possible that the reaction has already completed before a volume element passes through the entire catalytic bed. To avoid distorting the data for the calculated reaction rate with an unnecessarily high contact time between biocatalyst without a significant increase in conversion, a minimum applicable residence time of 2 min was chosen as the fixed residence time. At this residence time, it was possible to dilute the substrate feed of NAD^+ between 0.35 and 1.9 mM (4 mM stock concentration) and of formate between 50 and 290 mM (600 mM stock concentration) without causing pump stalling in the runs. From a hardware perspective, this enabled the implementation of an automated process for identifying the kinetic model of FDH with manipulated concentrations of NAD^+ and formate. Due to the hardware limitations, however, the residence time had to be omitted as an input variable.

4.1.2. Experimental Automatization and Settings

The described setup (cf. chapter 4.1.1) in the previous chapter was operated without human intervention to realize a full automated closed-loop reactor system. To achieve this, the experimental execution was automated using Python (Version 3.9.12, Python Software Foundation) and linked with MATLAB (Version R2021b, Mathworks) to handle the experiment evaluations. An overview of the software structure is given in Figure 8 and displays the interconnection between Python, MATLAB and the different input formats as Json and Excel files. To transmit and receive data of the , the Sensirion SHDLC Python driver was applied (Sensirion AG, 2021, Stäfa, Switzerland). The implementation of the UV/Vis spectrometer was achieved by applying *python-seabreeze* by Poehlmann, 2019.

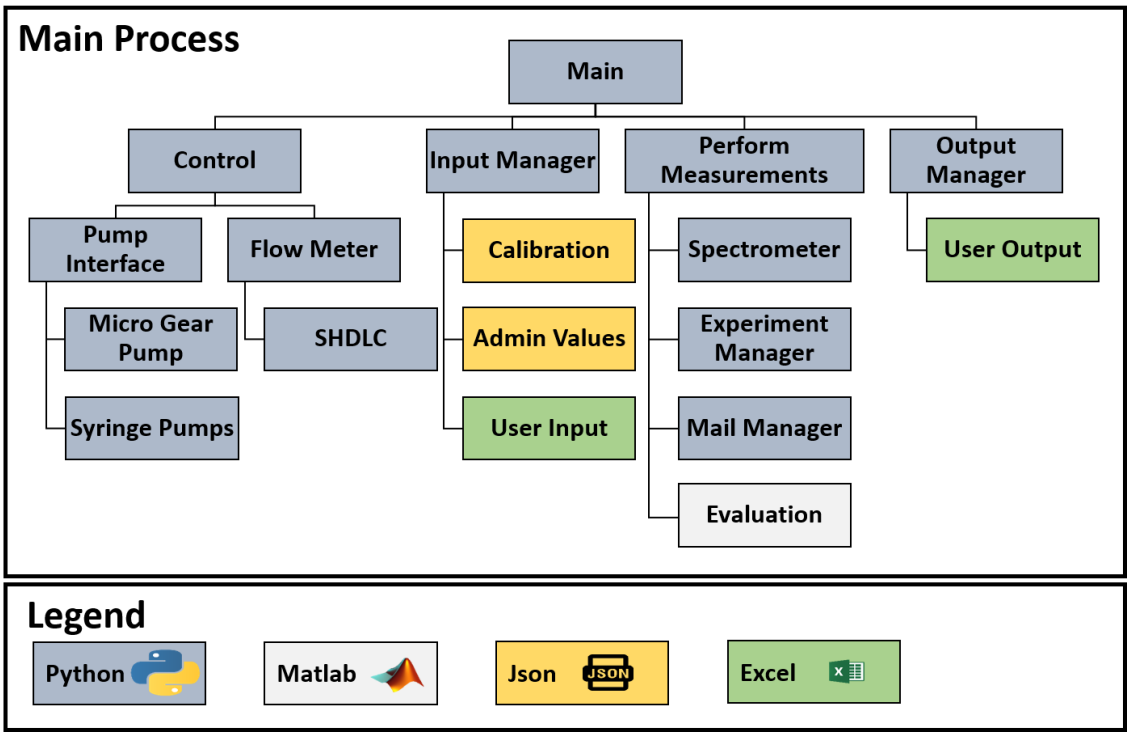

Figure 8: Software structure of the automated reactor platform written in Python (Version 3.9.12, Python Software Foundation) with Json and Excel files as Input. The script communicates with the MATLAB (Version R2021b, Mathworks) script for the model identification and MBDoE developed within the cooperation with Lucas Schaare (Institute of Process System Engineering, Hamburg University of Technology).

The inlet concentrations and important process parameters were set via the *UserInput.xlsx* (cf. Figure 9). The Excel format was employed to simplify the readability and accessibility of the settings for users without programming language experience.

1		**User Input**	**Unit**
2	Pump Mode	3	1, 2 or 3
3	Control Mode	1	0 or 1
4	Start Concentration	1.5	[mM]
5	Measure Intervals	5	[s]
6	Reactor Volume	0.14539	[mL]
7	OutputPath	C:/Entwicklung/seabreeze-kinetic-enzyme-pj/output	[-]
8	Name	Automated Campaign MM Starting Values	[-]
9	CuvetteDiameter	1	[mm]
10	Product	nadh	str or int
11	Educt	sodium formate	str or int
12	Solvent	water	str or int
13	Diameter S1	15.96	[mm]
14	Diameter S2	15.96	[mm]
15	Diameter S3	15.96	[mm]
16	Diameter S4	15.96	[mm]
17	Concentration S1	4	[mmol/L]
18	Concentration S3	600	[mmol/L]
19	Volume S1	10	[mL]
20	Volume S2	10	[mL]
21	Volume S3	10	[mL]
22	Volume S4	10	[mL]
23	Intermediate Purge	1	[boolean]

Figure 9: Userinput.xlsx screenshot for the settings of the built reactor platform. This list is applied for the setup of an automated experimental campaign.

Switching between *pump mode* 1, 2 and 3 allows to control one, two or four pumps. For pump mode 1 and 2 the user could insert a stock concentration, which is useful for experiments in the enzyme membrane reactor (cf. chapter 6) or reactions with one single substrate to vary. The *control mode* enabled and disabled the control of the flow using the current measured volume flow as input. The *measure intervals* defined the frequency of the UV/Vis spectrophotometer measurements. A crucial part for the calculation of the total volume flow $\dot{V}_{total}$ was the void volume V_{Void} of the attached reactor, which has to be changed if a reactor with different dimensions is applied or for repeating experiments with a different packing of the reactor. All obtained data was saved under the given output path with the given campaign name. Since the volume flow of the syringe pumps is controlled by a stepper motor, the diameter of the syringes was required as setting. With

a known concentration $c_{NAD,stock}$ and $c_{formate,stock}$ in the two substrate syringes, every combination of inlet concentrations was adjusted with (24) and (25).

$$V_1 \cdot c_{NAD,1} = \frac{\dot{V}_{NAD,stock}}{\dot{V}_{Dilution}} \cdot c_{NAD,stock} \cdot V_1 \tag{24}$$

$$0.5 \cdot c_{NAD,1} \cdot V_1 + 0.5 \cdot c_{formate,1} \cdot V_2 = (c_{NAD,2} + c_{formate,2}) \cdot V_{total} \tag{25}$$

The intermediate purge with buffer was enabled and disabled with a Boolean (0 = False, 1 = True) as switching variable. This purge procedure was conducted after each experiment to ensure the washout of residual NADH in the cuvette or other system parts. The duration of this purge and other rarely changed settings are saved in *AdminValues.json* (cf. Figure 10).


```
{
    "SensorCOM":        "COM3",
    "PumpCOM":          "COM6",
    "FlowInterval":     100,
    "ControlInterval":  1000,
    "MeasureInterval":  5,
    "TimeSecurity":     10,
    "MainFolder":       "C:\\Entwicklung\\seabreeze-kinetic-enzyme-pj\\OverhaulEN",
    "ControlAccuracy":  10,
    "ControlDamp":      0.25,
    "Wavelength":       "Auto",
    "WvLngthSpread":    "Auto",
    "StopAccuracy":     5,
    "MaxExperiments":   15,
    "PurgeTime":        4,
    "PurgeFlow":        0.035,
    "LastPurgeFactor":  4,
    "DecayRate":        0.000275,
    "ReceiverAddress":  [redacted]
}
```


Figure 10: Screenshot of *AdminValues.json* settings for the settings of the built reactor platform. This list is taken for the setup of an automated experimental campaign.

In this file, the emulated COM-ports of the USB-adapters for the flow sensor and the pumps are specified. For the flow sensor, the interval of flow measurements is set on

100 ms. This flow can be controlled via the implemented p-controller, which is the simplest controller possible applied with equation (26). When gas bubbles occurred in the cuvette or the outlet tubing was slightly moved, the controller led to errors in the controlling procedure, shutting down the whole measurement. Gas was only produced in soluble concentrations in the reactor system, but due to the heating of the liquid, the gas solubility decreased in the system. Thus, small amounts of air and CO_2 could accumulate in the periphery, causing irregular release and resulting in an error with active flow control. Without flow control, the data would only show a brief spike in flow and absorption. As result, the p-controller was not applied, since the accuracy of the applied syringe pumps of ±1% was high enough to ensure accurate flow rates. If the control mode is nevertheless enabled, the flowrate change $c(t)$ is calculated with the current flowrate $e(t)$ and set flowrate $s(t)$ in the set control interval with the dampening factor (*ControlDamp*) $K(p)$, when the applied control accuracy (*ControlAccuracy*) is not fulfilled.

$$c(t) = K(p) \cdot \frac{\big(s(t) - e(t)\big)}{s(t)} \cdot p(t) \tag{26}$$

The *MeasureInterval* is the time between the individual measurements of the spectrophotometer, here set to 5 s. *TimeSecurity* describes the number of residence times τ run before the steady state measurement of the current NADH concentration is performed. Thus, it is a multiplicator of the set residence time before the switch to the next set experiment. The second way of stopping the experiment was the *StopAccuracy*, which is given in percent and triggers the measurement of the steady state when the last measurements display an absorbance difference below the given percentage.

The measurement was taken at a wavelength of 340 nm, as specified in the calibration file, with the “Auto” setting. The wavelength spread (*WvLngthSpread*) was also set to 10 nm, defining the range in which the maximum absorbance is measured based on the set wavelength (e.g. 340 ± 10 nm). This compensation was necessary because non-screwed probes in the flow-through cuvette showed small deviations in maximum absorbance due to inaccurate leveling of optical probes in earlier setup iterations. To

address this issue, a wavelength spread was introduced. This feature became irrelevant with the implementation of screwed optical fibers.

Between the individual experiments, the reactor was flushed with pure buffer at the set flow rate and time (*PurgeFlow* and *PurgeTime*). To maintain the pressure in the system, the total flow was set similar to the residence time of the experiments (2 min), using only the buffer reservoirs. The intermediate purges were set to two residence times (4 min) to reduce the amount of applied buffer. Additionally, the experiments had a high number of set residence times, which eliminated the need for a lengthy washing procedure. However, to ensure that there were no residual substrates or products in the reactor for storage, the purge time was increased by a multiplier (*LastPurgeFactor*) after the final experiment. The final experiment was conducted with the maximum number of experiments (*MaxExperiments*), which included the initial six experiments. Fifteen experiments were found to be optimal when using 12 mL syringes. This corresponds to nine additional MBDoE experiments, with NAD^+ being the limiting factor. The *DecayRate* of NAD^+ was also adjusted to correct the inlet concentration of NAD^+ based on elapsed experimental time. This adjustment was necessary to improve the accuracy of the active NAD^+ concentration, which was shown experimentally proven in chapter 9.1. The last feature was the insertion of an e-mail address (*RecieverAddress*), which was applied to write an automated mail to the receiver when the experimental campaign was fully executed.

Further settings regarding the model identification of the reaction system were directly inserted in the MATLAB code. Important characteristics of the reactor setup and the experimental design were set here. For the correct fitting of v_{max}, the enzyme concentration in mM was needed. This concentration was calculated based on the molar immobilized enzyme concentration per void volume of the reactor. Furthermore, the variables for the kinetic parameters are created and set on a value according to literature data. The concentration ranges were fixated in the MATLAB script to ensure experiments in the set design space.

Three different types could be set for the optimal experimental design (*opt_types*), namely *a-optimal*, *e-optimal*, and *d-optimal* experimental. Due to the high correlation between the kinetic parameters, the *e-optimal* design offered the best performance (cf. Chapter

3.3.2). The optimization mode (*opt_mode*) could be set to local optimum with multiple initialization (*opt_mode = 1*) points which were defined (*opt_init = 2*). With *opt_mode = 2* the grid search is enabled originating of different grid points *np*. In this work, the local optimum method was applied.

4.1.3. Model Identification in MATLAB

The identification of the model with the AWDC was automated using MATLAB. The workflow for the reactor platform automation is shown in Figure 11.

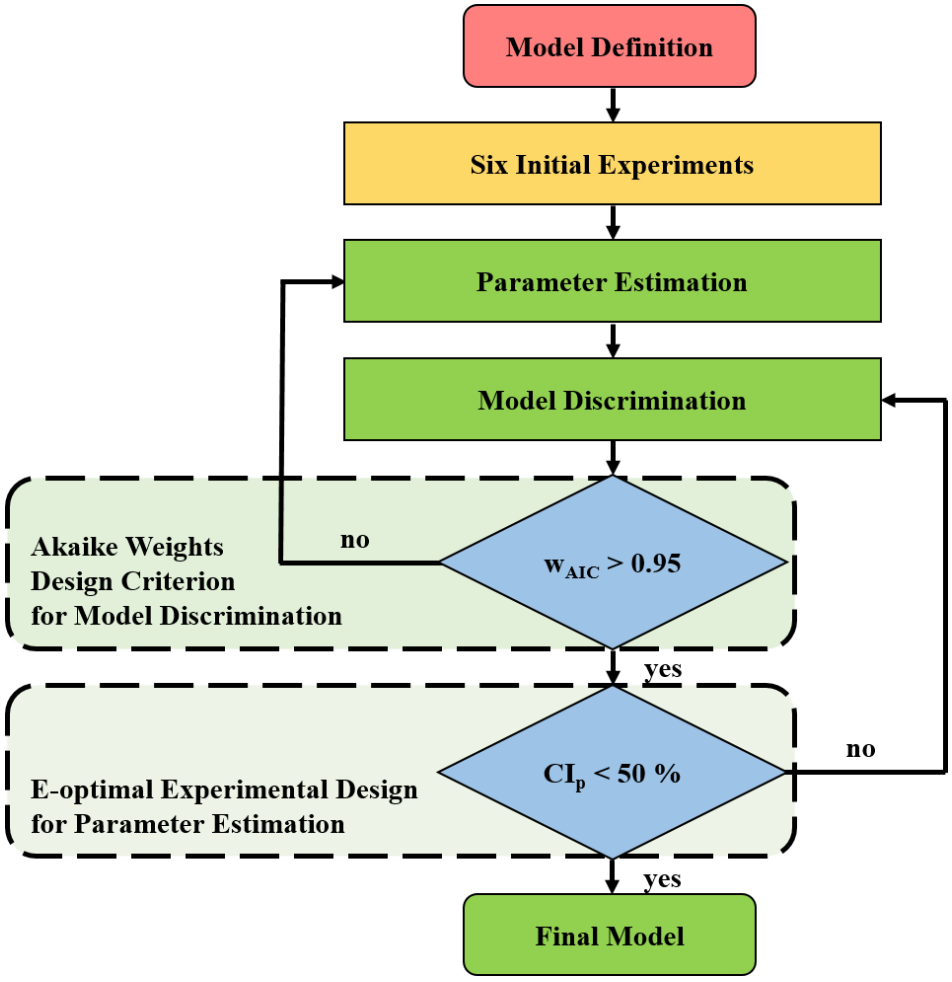

Figure 11: Systematic workflow for the closed-loop reactor setup for the model identification of immobilized FDH of *Candida boidinii*. In MATLAB the MBDoE is based on the AWDC for the model discrimination and based on the *e-optimal* design for the parameter estimation.

The reactor model was implemented as a plug-flow model (cf. Chapter 1.2.2), and the resulting system of ordinary differential equations was solved using the *ode45* solver in

MATLAB. The model candidates given in Table 1 were used to define the reaction rate laws. Initial values were specified as a starting point for solving the initial value problem (cf. Table 3). Applying MATLAB, initial experiments were conducted to determine the product concentration within the given design space (cf. Table 2). The initial parameter estimation was performed using the non-linear least-square curve-fitting (*lsqcurvefit*) regression method based on the input data (NAD^+ and formate concentration) and the response (NADH concentration) of the initial DoE. This procedure was repeated for each model candidate, and the 95% confidence intervals were calculated using *coefCI*. To discriminate between models, the AIC for all models was calculated based on the log-likelihood and the number of kinetic parameters, according to Equation 12. The comparison of models was conducted by calculating the Akaike weights w_{AIC} applying Equation 13. The stop criterion for the model discrimination was fulfilled when a model exceeded a w_{AIC} value of 0.95. When none of the models met this criterion, clear discrimination between models was not possible based on the initial experiments. Consequently, optimal experimental design with the AWDC for maximizing the deviation between model responses based on the previous parameters was performed until the stop criterion was satisfied. To evaluate the quality of the model in terms of parameter estimation, the 95% confidence intervals of the kinetic parameters were assessed. If the 95% confidence intervals of the selected model exceeded 50%, the parameter estimation was deemed unsatisfactory. Therefore, the objective of the subsequent optimal experimental design was to maximize the precision of the parameter estimates. The procedure utilized the FIM to approximate the variance of the parameter estimates. The *e-optimal* criterion aimed to maximize the minimal eigenvalue of the FIM to decrease the confidence intervals of the kinetic parameters of the best performing model. Model identification was completed when either the threshold of 50% or the maximum number of experiments was reached. The final results for model identification are displayed in the Python console. For each individual run, a comprehensive evaluation of the experiment was saved in an Excel file, including steady-state data for NADH concentration, input values for substrate concentrations, and residence time for the experimental campaign. Plots were generated to evaluate the model's goodness of fit, discrimination progress, parameter estimation progress, parity plot for all models, and sensitivity analysis for the final model.

4.1.4. Experimental Workflow

In order to prepare an automated experiment, all four syringes were filled with KP_i (50 mM, pH 8) to flush the system without the reactor first for 15 min at 300 µL/min with *purgeSystem.py*. During the rinsing of the system, the substrate stock solutions are prepared and kept on ice until usage. After stopping the rinsing procedure with *StopPumps.py*, the PBR was taken of the refrigerator and connected to the system. Restarting the purge procedure, the tubing of the dosing system to the reactor was filled with buffer previously to ensure no enclosure of air in the system before reconnecting the reactor. Immediately, the rinsing procedure was continued to check for leakages at the reactor. As a result, the reactor was not heated to observe potentially occurring droplets at the reactor screw fittings inside of the heating jacket. When the leak check was successful, the substrate syringes were connected and the buffer syringes were refilled.

A typical initial experimental campaign for the closed loop model identification was set in *UserInput.xlsx* as displayed in Table 2. The first set of experiments consisted of a total of six experiments with four experiments combining near-maximum and minimum concentrations of formate (min. 50 mM; max. 290 mM) and NAD^+ (min. 0.35 mM; max. 1.9 mM). The full maximum concentration was not applied to avoid the full halt of the dilution pumps, which could lead to a stalling of the substrate pumps and errors in the early iterations of the dosing program. Therefore, these concentrations were applied in later experiments to compare all initial experiments with each other to determine the overall standard deviation σ of the system. The first and last experiments of the initial campaign were chosen to be replicating center points (1 mM NAD^+ and 150 mM formate) of the design space to evaluate the internal error of the initial DoE. Furthermore, these experiments allow the evaluation of curvature, which is important for the fitting of kinetic models to the initial data, especially regarding the K_m values.

Table 2: Table for the setup of experimental procedures in the automated reactor platform with a typical initial design of experiment campaign in the *UserInput.xlsx* file.

	Mode 1+2+3	**Mode 2+3**	**Mode 3**
Experiment	**Residence times [min]**	**NAD^+ [mM]**	**Formate [mM]**
A	2	1	150
B	2	0.35	50
C	2	0.35	290
D	2	1.9	50
E	2	1.9	290
A* (repetition)	2	1	150

For all experiment modes, the settings for an experimental campaign can be implemented in this table. The residence time can be set for all modes. Different dilutions for mode 2 (two pumps) and 3 (four pumps) can be set. Mode 3 with two different substrate solutions in a specified concentrations range is only applied in the PBR setup with full dilution control.

When the initial experiments are inserted in *UserInput.xlsx* the *main_EnzymeVision.py* is started and the automated experimental campaign was initialized. Instantly, the mail address "automatedExperiment@gmail.com" can be typed in the console with the password to enable the reminder mail. After that, the initial experiments are initiated and the live view for the NADH concentration and flowrate appears in a separate window (cf. Figure 12).

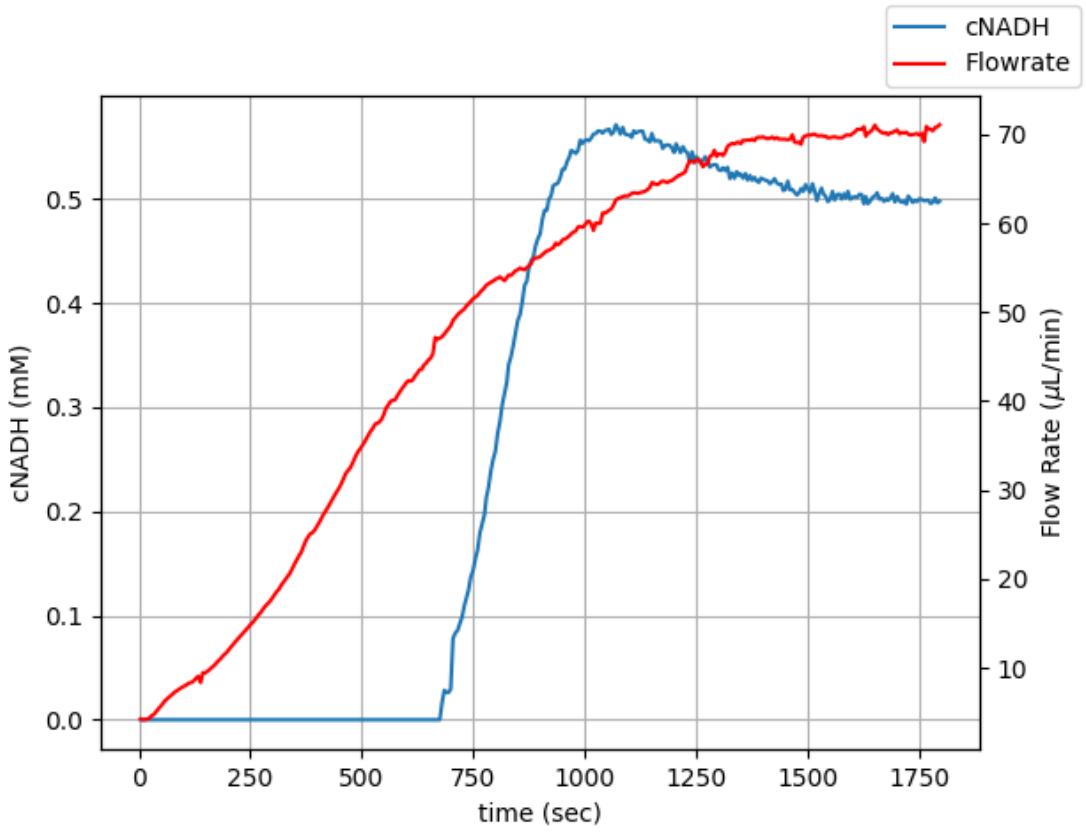

Figure 12: Exemplary report of the reaction progress in the automated reactor system in live view. The measured NADH concentration in mM and the flowrate in µL/min is shown in respect to time.

This real-time monitoring of concentration and flow allows the user to continuously monitor the reactor process. This provides immediate feedback and enables them to identify potential problems such as set point deviations, leaks, bubbles, pump failures or blockages. As a result, malfunctions can be detected in time and measures can be taken to regulate the process. Especially the continuous operation of the reactor system requires control mechanisms, if the automated procedures is successfully executed during the development of the reactor platform. Based on the presented data, automated control mechanisms could be implemented as well.

4.1.5. Measurement Error

In order to check the activity regularly, displaying a base amount of data for the model identification and check the reproducibility of the system, the same six experiments (cf. Table 2) were performed in the beginning of a closed-loop campaign (cf. Figure 13). These experiments are the result of a 2^2 full factorial design with a center point at the

beginning and at the end. Since the center point experiments are repetitions starting (A) and ending (A*) the initial experimental campaign, they are combined as center point in Figure 13, resulting in 14 averaged experiments for 7 campaigns (cf. Table 15). Consequently, seven repetitions are averaged for the other experiments in Figure 13 to obtain the standard deviation σ with the equation:

$$\sigma = \sqrt{\frac{\sum(y - \bar{y})^2}{n}} \tag{27}$$

The kinetic parameters estimated in this work are evaluated on the basis of 95% confidence intervals. These are calculated with the equation:

$$CI_{95\%} = \bar{y} \pm 1.96 \frac{\sigma}{\sqrt{n}} \tag{28}$$

The standard deviation for low NADH concentrations in experiments B and C accounts for 0.030 and 0.023 mM. Therefore, the relative standard deviations are 9.6% and 7.1% and appear high. However, due to the detection limit in this order of magnitude (0.047 mM) and the limit of quantification at 0.143 mM, this high deviation was to be expected (cf. Figure 46). Near the detection limit, the signal-to-noise ratio is lowest, resulting in higher deviations in this region. The relative error is approximately halved at higher concentrations with 0.042 mM at the center point (Experiment A), 0.069 mM at Experiment D and 0.054 mM at Experiment E. This corresponds to a relative error of 4.5, 5.1 and 3.4%, respectively. These deviations have to be taken into account to assess the quality of the identified kinetic models and its standard deviations or 95% confidence intervals.

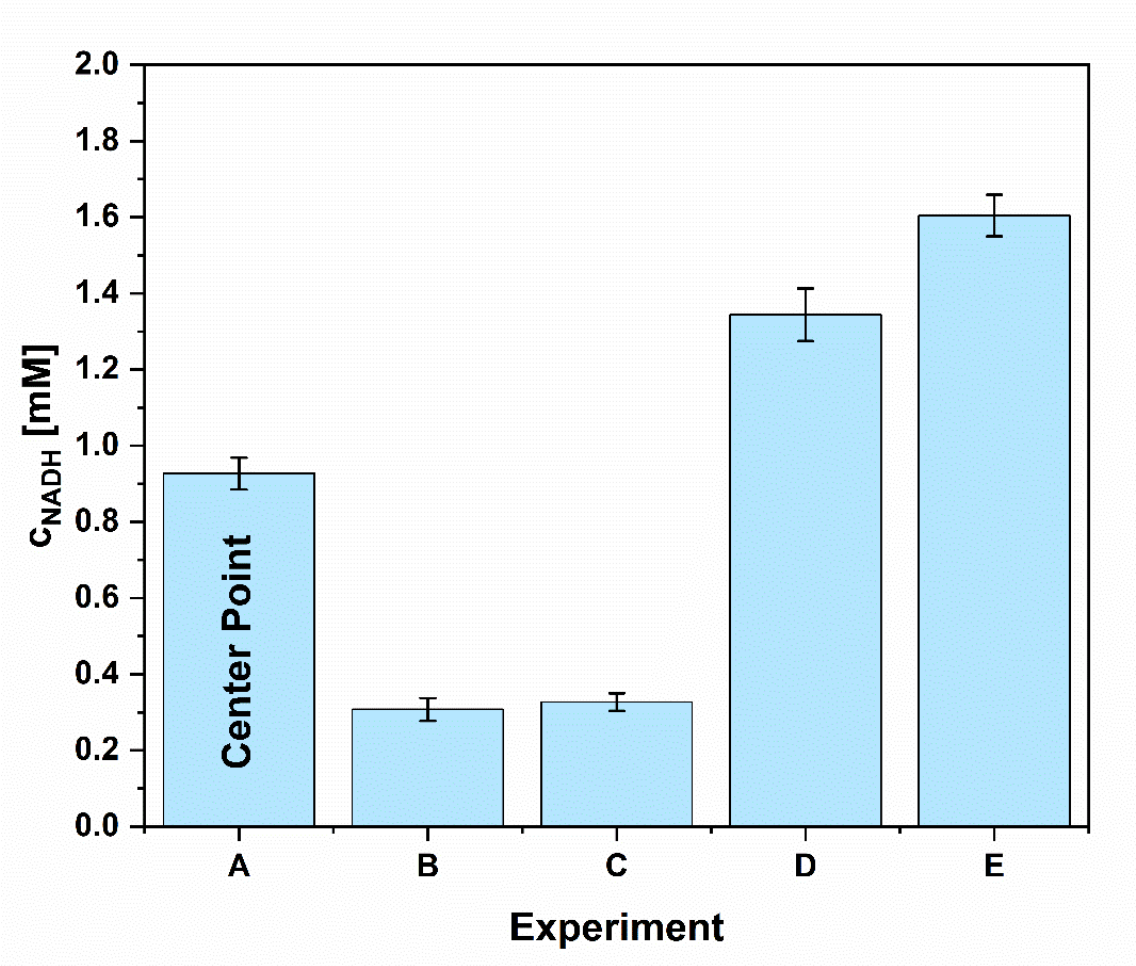

Figure 13: Repeatability shown by initial DoE Experiments; A: 1 mM NAD^+ 150 mM formate; B: 0.35 NAD^+ and 50 mM formate; C: 0.35 mM NAD^+ and 290 mM formate; D: 1.9 mM NAD^+ and 50 mM formate; E: 1.9 mM NAD^+ and 290 mM formate. All experiments were conducted in 50 mM KP_i buffer at pH 8 and 30 °C.

4.2. Parameter Estimation

4.2.1. Automated Model Identification

The constructed automated reactor setup was run initially with the assumption, that the model of Schmidt et al. is the correct model (Schmidt *et al.*, 2009). Therefore, the initial starting parameters were set on the values found in literature data (cf. Table 3). The closed-loop campaign was initiated with a 2^2 factorial design of the design space with two center points, which corresponds to the first 6 experiments (cf. Figure 13). Regarding a focus on a distinct model discrimination in the model-based experimental design, an Akaike weight of 0.95 was specified as the experimental campaign target. After reaching

the target, the system automatically switches to the *e-optimal* design of experiments to improve the accuracy of the parameters. The model discrimination was performed based on the AWDC (cf. chapter 3.3.1). For the least-squared curve fit of the model, lower and upper bounds were employed according to the entries in the Braunschweig Enzyme Database BRENDA for the enzyme commission number 1.17.1.9. Since the entries for *Candida boidinii* also include mutants and long-stored enzyme preparations but do not originate from immobilized enzymes, a further value of 50% was subtracted from the lower limit of the parameters and a value of 50% was added to the upper limit. This resulted in Table 3, summing up the applied limitations and starting values for the kinetic parameters. After the initial experiments, the model discrimination directly reached the target threshold of 0.95 Akaike weight for the SCHMlump model with an AIC weight of 0.99 with the kinetic parameters shown in Table 3. It should be noted that the experimental data was generated with an older script that was subsequently improved. Therefore, the optimal experimental design in this campaign is aimed at improving the confidence intervals. After changing the script, the experiments should actually improve the model discrimination, since now only an AIC weight of 0.69 of the SCHMlump model was reached (cf. Figure 14).

The quality of the model candidates CPI, SCHM, MODI, and SCHMlump is indicated by a negative log-likelihood. The model discrimination is further defined by the AIC, which penalizes the number of parameters in the model. This increases the difference of CPI, SCHM and MODI to SCHMlump by a value of two since all models incorporate one more parameter (cf. equation 12). With only six data points, the qualitative difference between the AIC and the log-likelihood is marginal. At this point, AICc provides a clearer model discrimination by the incorporation of dataset size (cf. equation 17). However, the goal is to achieve clear and distinct model discrimination based on optimally designed experiments. This is achieved through automated experiments, which increase data density if necessary. Therefore, AIC is used as the design criterion, since more optimal experimental data is added equally. For this reason, the AICc will be equal to the AIC after 10 experiments for four parameters and after 12 experiments for five parameters. Therefore, the AIC provides a constant penalty for more complex models.

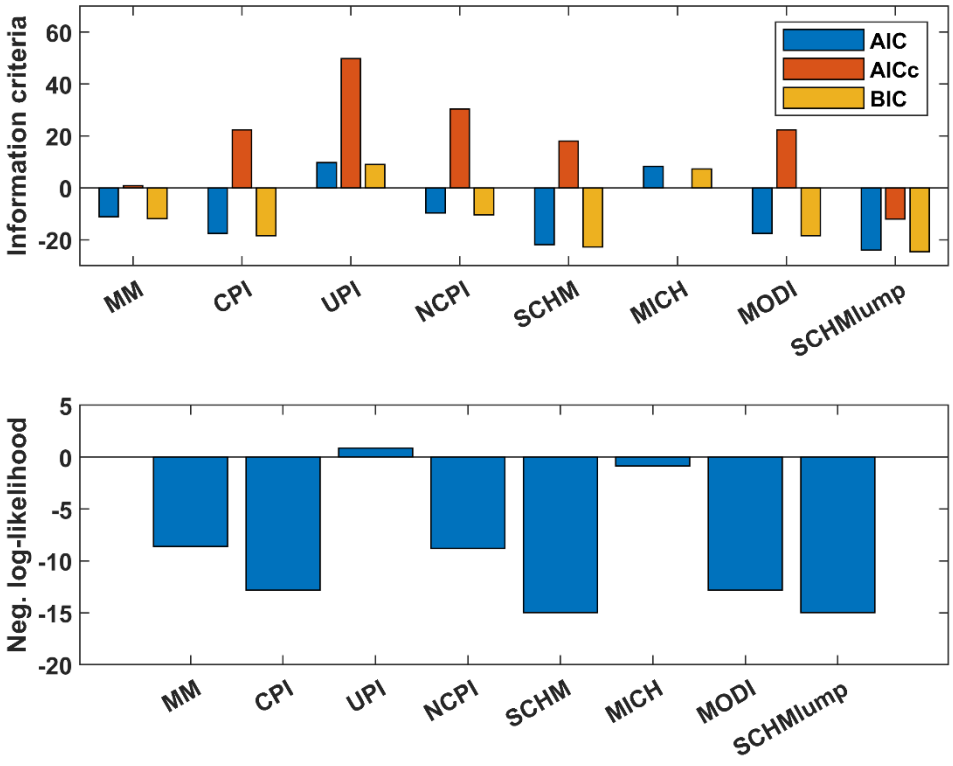

Figure 14: Information criteria and negative log-likelihood as an evaluation of the goodness of fit for candidate models of FDH kinetics based on six initial experiments. Higher absolute values indicate better fits.

Despite the results of the initial model discrimination, all models could fit the data well within the first 1 mM of NADH (cf. Figure 15). The highest difference is generated in the higher concentrations of the design space around 1.5 mM. Here, UPI and MICH fail to represent the experimental data and have therefore positive AIC values of 9.8 and 8.3 respectively. In contrast, the best model SCHMlump has an AIC value of -24.0, which is close to MODI and CPI with -17.6 each. Nevertheless, since the Akaike weight is calculated based on the summarized AIC values, the weight is low due to the high average AIC.

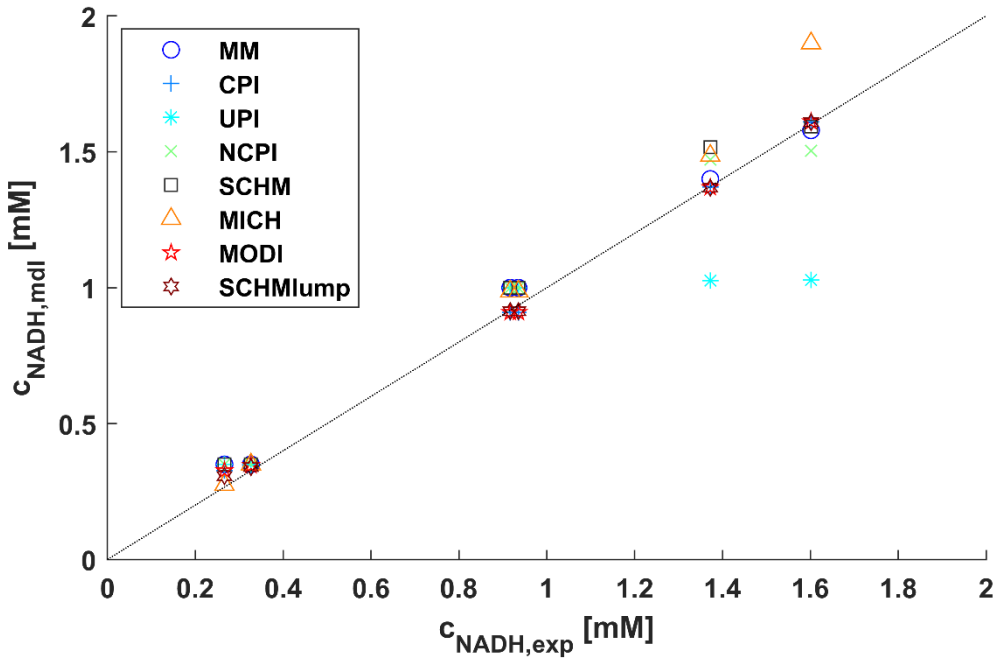

Figure 15: Parity plot for the modelled and experimental product concentration of the closed-loop results based on starting values according to Schmidt *et al.* (2009). All experiments were conducted in 50 mM KP_i buffer at pH 8 at 30 °C.

Based on the design criterion, the SCHMlump model was selected as the best candidate. The corresponding kinetic parameters based on the initial experiments are listed in Table 3. However, it is important to note that all kinetic parameters of the immobilized FDH should be considered as **apparent**, as the effects of immobilization have not been fully investigated. Additionally, the selected model candidate is chosen solely based on their goodness of fit without any physiochemical meaning beyond what is provided in literature. Therefore, it cannot be fully excluded that there is a structural influence or a change in kinetics caused by mass transport phenomena in the pores.

Table 3: Initial values, upper and lower bounds and estimates for the non-linear fit of the first closed-loop experimental campaign design of experiment approach.

Parameter	Lower Bound	Upper Bound	Start	Estimate
k_{cat} [s^{-1}]	$1.25 \cdot 10^{-4}$	4.8	0.18	0.31 ± 56.2%
$K_{m,NAD}$ [mM]	$7.5 \cdot 10^{-3}$	135	0.03	0.65 ± 118.5%
$K_{m,formate}$ [mM]	$1.85 \cdot 10^{-2}$	52.5	0.3	-
$K_{i,NADH}$ [mM]	$1 \cdot 10^{-2}$	0.3	0.1	-
$K_{i,NAD}$ [mM]	5	117.2	70	-
P_1 [mM^2]	$9.25 \cdot 10^{-2}$	6154	21	32.8 ± 98.1%

Due to the limited size of the data set, the confidence intervals for the kinetic parameters, ranging from 56.2 to 118.5%, still exhibit considerable width. Nonetheless, they provide an order of magnitude for the kinetic parameters. A comprehensive evaluation of the parameter estimation is, however, not currently reasonable, as the confidence intervals do not allow for a clear estimation. To improve the data set for parameter estimation, a series of experiments were designed based on the *e-optimal* MBDoE criterion. This was accomplished by maximizing the eigenvalues of the FIM , resulting in nine iteratively planned experiments that were automatically executed after the initial DoE.

To visualize the iterative improvement in parameter estimation, the 95%-confidence intervals are plotted against the number of added experiments (cf. Figure 16). In the MATLAB script, the parameters are fitted and represented with v_{max} in $mM \cdot s^{-1}$ which is dependent on the enzyme concentration. Thus, for better comparison with literature data, the turnover k_{cat} in s^{-1} is employed in discussions and for the presentation of results.

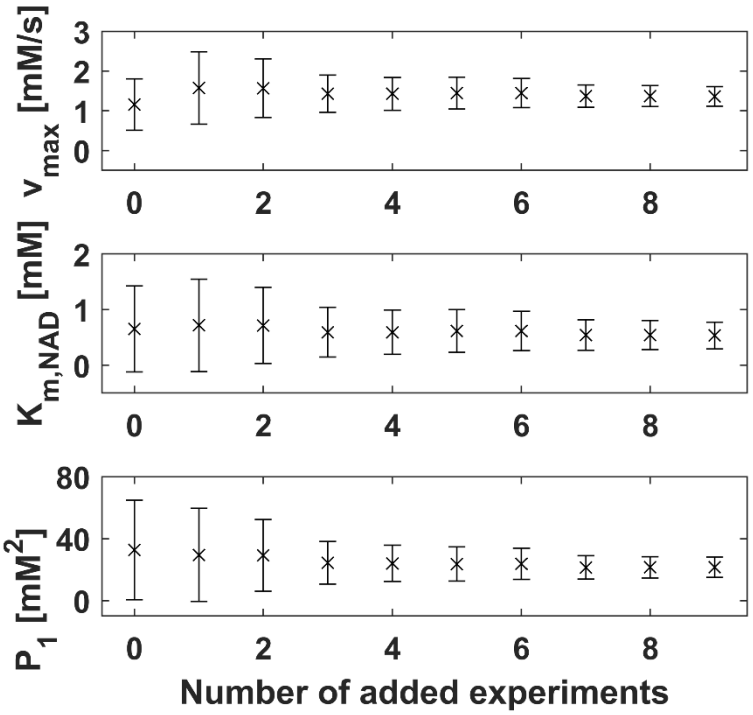

Figure 16: Confidence intervals of the model parameters for the model according to Schmidt *et al.* (2009) plotted over the number of added experiments during the automated MBDoE 1.1 campaign.

The most significant parameter improvements resulted between the added experiments 2 and 3 and the experiment 6 and 7. This is congruent with the plotted minimized eigenvalues (cf. Figure 17) of the FIM. For both experiments the minimal formate concentration of 50 mM was applied with NAD^+ concentrations of 1.13 mM and 1.11 mM which resulted in 0.97 mM and 0.96 mM NADH, respectively (cf. Figure 18). These experiments are the extrapolated center points of the initial DoE producing almost identical amounts of NADH (0.94 mM and 0.92 mM for the center point experiments) at 1 mM NAD^+ and 150 mM formate. Therefore, the curvature of the model in dependence of the formate concentration was refined and confirmed, increasing the accuracy of kinetic parameter $K_{m,formate}$ which is included in the combined parameter P_1. Due to the high correlation of the kinetic parameters (cf. Chapter 3.3.2), all confidence intervals decreased in the same magnitude, but P_1 was improved the most. For the experiment 3 the confidence intervals for v_{max}, $K_{m,NAD}$ and P_1 decreased by 37%, 34% and 41%, and for experiment 7 by 24%, 20% and 26%, respectively (cf. Figure 16).

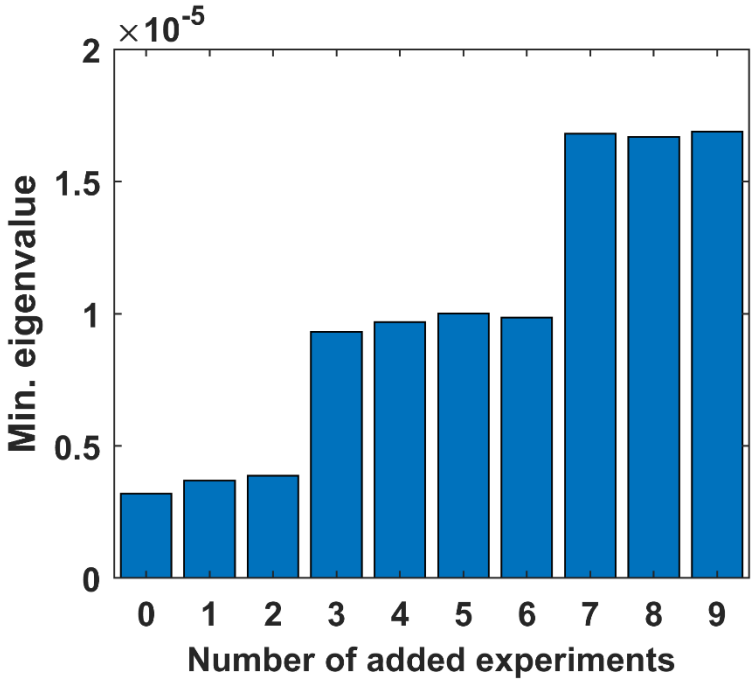

Figure 17: Observed maximization of minimized eigenvalues of the objective function plotted over the number of added experiments.

The first MBDoE campaign (cf. MBDoE-1.1 in Table 15) started with an added experiment at maximum NAD^+ concentration of 1.88 mM at that time point accounting for the deactivation over time and a minimal formate concentration of 50 mM. This reproduces the composition of experiment D (cf. Figure 13), which was repeated twice in the automated campaign (added experiment 1, 5 and 9). In each case, the maximum available NAD^+ concentration was applied with minimum formate concentration. Alternating to experiments with 50 mM formate, experiments with 95% of the maximum concentrations were conducted at 1.8 mM NAD^+ and 275.5 mM formate (added experiments 2, 4, 6, 8). In the added experiment 1 and 7, 50 mM formate were repeatedly applied, but the NAD^+ concentration of 1.1 mM was set to result in the same NADH concentration as the center points of the initial DoE of 0.94 and 0.92 mM. Thus, the curvature of the model was confirmed in that region with resulting NADH concentrations of 0.97 and 0.96 mM, respectively (cf. Figure 18).

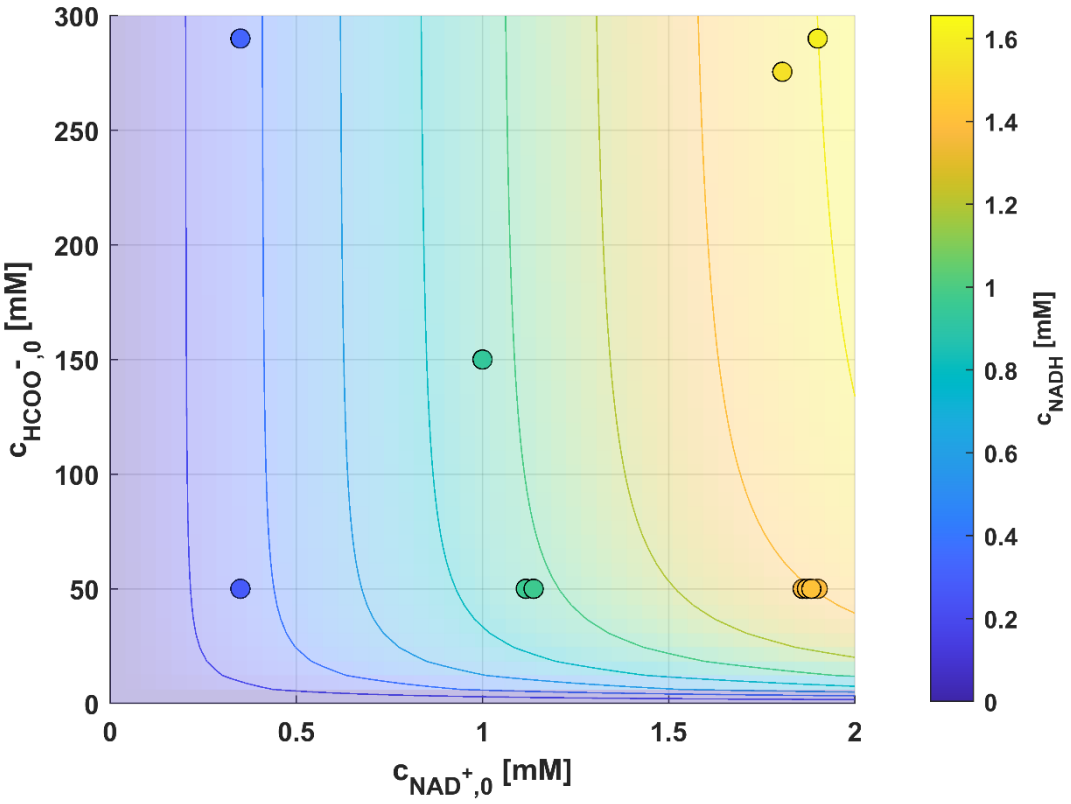

Figure 18: Two-dimensional surface plot of the experimental product concentration of the closed-loop results based on starting values according to Schmidt *et al.* (2009) dependent on the initial concentrations of formate and NAD^+. The surface is represented with colors from blue (0 mM NADH) to yellow (1.6 mM NADH).

Overall, the added experiments resulted in improved parameter accuracy of the resulting kinetics according to the SCHMlump model, which was chosen as starting values. Nevertheless, the majority of the experiments were conducted at minimum formate concentration and maximum NAD^+ concentration. The final results of the non-linear fit are summarized in Table 4.

Table 4: Initial values, upper and lower boundaries and estimates for the non-linear fit of the first closed loop experimental campaign based on starting values according to Schmidt *et al.* (2009) and with the application of the MBDoE design with the kinetic parameter estimates of the SCHMlump model.

Parameter	Lower Bound	Upper Bound	Start	Estimate
k_{cat} [s^{-1}]	$1.25 \cdot 10^{-4}$	4.8	0.18	0.28 ± 18.3%
$K_{m,NAD}$ [mM]	$7.5 \cdot 10^{-3}$	135	0.03	0.53 ± 44.9%
$K_{m,formate}$ [mM]	$1.85 \cdot 10^{-2}$	52.5	0.3	-
$K_{i,NADH}$ [mM]	$1 \cdot 10^{-2}$	0.3	0.1	-
$K_{i,NAD}$ [mM]	5	117.2	70	-
P_1 [mM^2]	$9.25 \cdot 10^{-2}$	6154	21	21.5 ± 30.5%

Regarding the starting solution for the non-linear regression, k_{cat} is in the same order of magnitude as the starting value, however, exceeding it by 80%. P_1, on the other hand, is nearly consistent with the starting value. Solely, $K_{m,NAD}$ exceeds the starting value by a factor of 17.6 indicating influences of the immobilization or the model choice. The immobilization could have an influence by changing the structure on the enzyme, decreasing the affinity of the enzyme to the substrate as well as introducing diffusion limitations. On the other hand, the model discrimination could have led to the wrong model identification, ignoring inhibition effects that increase the alleged apparent $K_{m,NAD}$. This is contradicted by the fact that no product-dependent inhibition model was selected from the candidates. However, both models involving substrate inhibition in the form of P_1 were selected. This supports the substrate inhibition, which could not be quantified unambiguously due to a linear dependence of the $K_{m,\,NAD}$ and $K_{i,\,NAD}$ values of the substrates and therefore had to be summarized by P_1.

The conclusion of the first experimental campaign is that a non-linear fit of the model candidates was successfully performed. The attempted model discrimination is possible, however, with limitations of the input parameters (concentration range and fixed residence time). The optimal experimental design is able to decrease the confidence intervals of the parameter estimates. Since the model discrimination was achieved after the initial DoE with a very limited number of experiments, the conducted campaign was repeated. Accordingly, the repeatability of the experimental campaign and subsequent model identification was tested.

4.2.2. Repeatability

Automated screening platforms as presented in this work have to be tested on repeatability to show the validity of their results. The initial DoE (Experiments A-A*) results are fitted with starting solutions for non-linear regression based on literature data. Measurement deviations or low repeatability may result in a different model discrimination. This would allow the MBDoE approach to focus on experiments, improving the accuracy of a model that initially fits erroneous data. Thus, three repetitions of the same experimental campaign were conducted to confirm repeatability (cf. Figure 19). Both the reaction conditions and the non-linear regression parameters were kept constant (according to Table 3). MBDoE Run 2 and MBDoE Run 3 are both repetitions of the experimental campaign discussed in chapter 4.2.1. Thus, the initial DoE was designed according to Figure 13 and the following experiments (Experiments 1-9) were purely designed based on the MBDoE applied on the results of the first experiments.

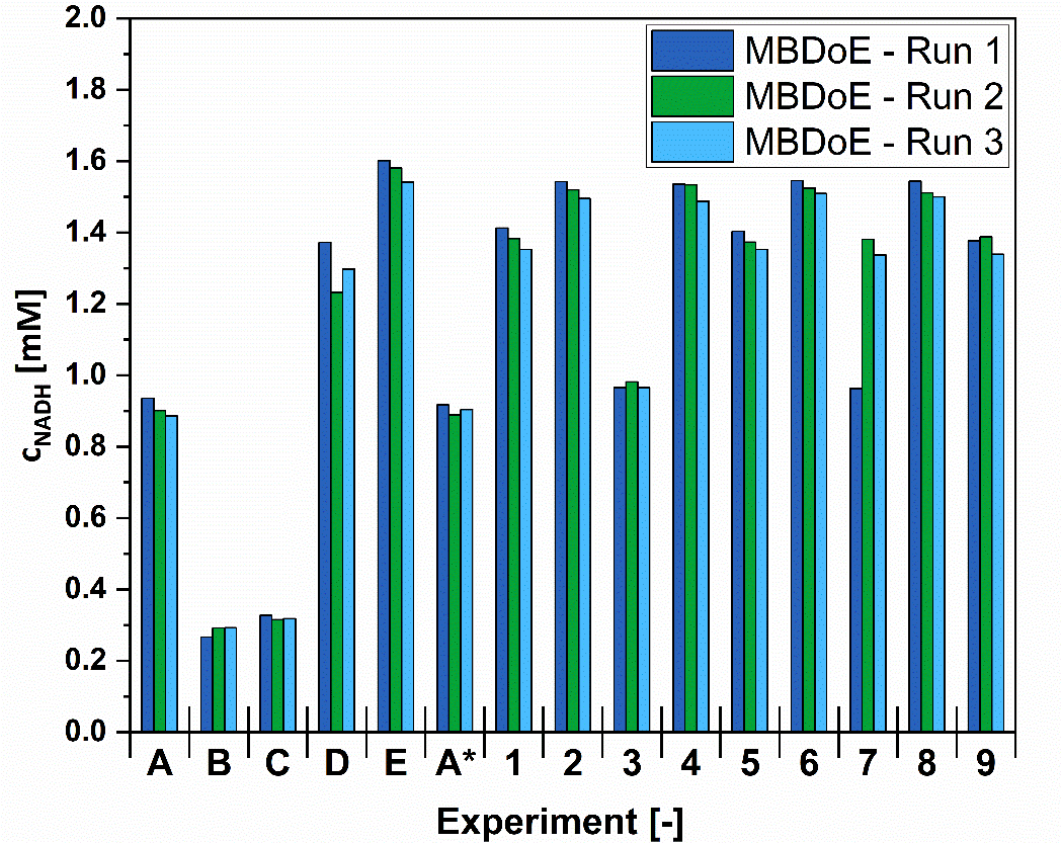

Figure 19: Repeatability analysis of the MBDoE campaign based on Schmidt *et al.* 2009 initial parameter estimates. Bars show the product concentration of the 15 experiments long campaigns. Experiments A, B, C, D, E and A* refer to the initial DoE. A* is the repetition of the center point A. Experiments 1-9 are automatically planned and executed experiments based on the MBDoE approach.

All experimental campaigns followed the same course of experiments after the initial DoE (cf. Figure 19). The only exception is experiment 7 of the first MBDoE run. Here, a repetition of the experiment using 50 mM formate and 1.12 mM as inlet concentration of NAD^+ is conducted. In the first MBDoE run, this experiment showed the greatest information gain in relation to the eigenvalues of the FIM. As stated in the previous chapter, this experiment is applied to confirm the NADH concentration predicted with the model parameters at minimal formate concentration according to the initial center points. The resulting NADH concentration of the experimental campaigns are in good agreement with each other. All runs finalized with a kinetic model according to the lumped model of Schmidt *et al.* (2009) after 15 experiments. The resulting apparent kinetic parameters and the corresponding 95% confidence intervals of the individual runs are summarized in Table 5.

Table 5: Comparison of three different automated MBDoE runs on starting values according to Schmidt *et al.* (2009) and with the application of the AWDC with the same settings.

Run	**k_{cat} [s^{-1}]**	**$CI_{95\%}$ [%]**	**K_{mNAD} [mM]**	**$CI_{95\%}$ [%]**	**P_1 [mM^2]**	**$CI_{95\%}$ [%]**
1	0.28	18.3	0.53	44.9	21.5	30.5
2	0.26	42.1	0.50	108.2	22.5	72.1
3	0.26	20.9	0.54	52.2	22.0	35.9
Mean	0.27	27.1	0.52	68.4	22.08	46.2
STD	0.01	13.1	0.02	34.6	0.50	22.6

The parameter comparison shows high consistency throughout the runs. The mean of the turnover number k_{cat} resulted in $0.27 \pm 0.01\ s^{-1}$, $K_{m,NAD}$ in 0.52 ± 0.02 mM and P_1 in $22.08 \pm 0.5\ mM^2$. Consequently, the method carried out in the automated reactor platform displays high repeatability with a standard deviation of the mean of <4%. The confidence intervals of the kinetic parameters varied significantly but exhibit a matching pattern with k_{cat} exhibiting the lowest error followed by P_1. $K_{m,NAD}$ shows the largest confidence interval in all runs with a ~2.5-fold value in contrast to k_{cat}. This is especially evident in the second run, where the confidence interval reaches a value of over 100%. The estimation of the initial parameters applied in the presented runs was based on literature data, which exhibited parameters for soluble enzymes at pH 7.5 and 25 °C (Schmidt *et al.*, 2009). These values were selected because the underlying model has been developed to a high level of detail to match the mechanistic kinetics of FDH under process

conditions, which is the aim of this work. However, a slight difference in conditions might influence the first estimation of the parameters strong enough to influence the result of the model identification. The kinetics of the reaction could also be influenced by the immobilization in addition to the reaction conditions. Nevertheless, the SCHMlump model has the highest agreement with the experimental data, but the narrow model discrimination suggests that other models with different starting parameters could also fit. It is important to note that the kinetic parameters of the SCHMlump model were used as starting values. To test the influence of the starting parameters, another campaign was carried out within the design space with different starting values.

4.2.3. Comparison to Full Factorial Design

After the introduction of the automated MBDoE approach and the confirmation of repeatability, it is compared to a standard factorial design approach. This statistical experimental design is distinctive in its design by equally distributed combinations of the input factors. This type of design is partially applicable on modelling procedures, due to the dependence on the input variables. In this case, only the inlet concentrations were changed, that a full factorial design with four levels result in 16 experiments, which was feasible for the reactor platform. The resulting product concentration are shown in Table 6.

Table 6: Full factorial design of experiments. All experiments were performed in 50 mM potassium phosphate buffer at pH 8 and 30 °C.

50 mM formate		**133.33 mM formate**	
c_{NAD} [mM]	c_{NADH} [mM]	c_{NAD} [mM]	c_{NADH} [mM]
0.35	0.26	0.35	0.28
0.90	0.74	0.90	0.79
1.45	1.13	1.45	1.22
1.90	1.36	1.90	1.50

216.67 mM formate		**300 mM fomate**	
c_{NAD} [mM]	c_{NADH} [mM]	c_{NAD} [mM]	c_{NADH} [mM]
0.35	0.27	0.35	0.30
0.90	0.81	0.90	0.82
1.45	1.25	1.45	1.27
1.90	1.56	1.90	1.57

In the applied concentration ranges, the results show a minimal influence of the substrate concentration of formate on the product formation. Here, higher concentrations of over 50 mM formate were favored, since industrial conditions for FDH apply formate in excess amounts to reach multiple cofactor regeneration cycles. The investigation of lower concentrations would have resulted in the given reactor setup in high dilution factors of over 10, which was avoided on purpose to minimize the concentration and pump errors. The NAD^+ concentration, on the other hand, directly influences the activity and the resulting product concentration.

For the model fitting of the full factorial data, the starting values for the model identification were taken of Schmidt *et al*. (2009), to compare the modelling results to the MBDoE campaign. Furthermore, Schmidt *et al.* provided profound research regarding the kinetic modeling of the FDH under industrially relevant conditions. This leads to the assumption, that this model represents the data with the highest accuracy. Nevertheless, the model discrimination based on the full factorial design resulted in MM kinetics as the best fitting model. Comparing the information criteria and the negative log-likelihood of the models, except for CPI, UPI and MICH all models can fit the data similarly well (cf. Figure 20).

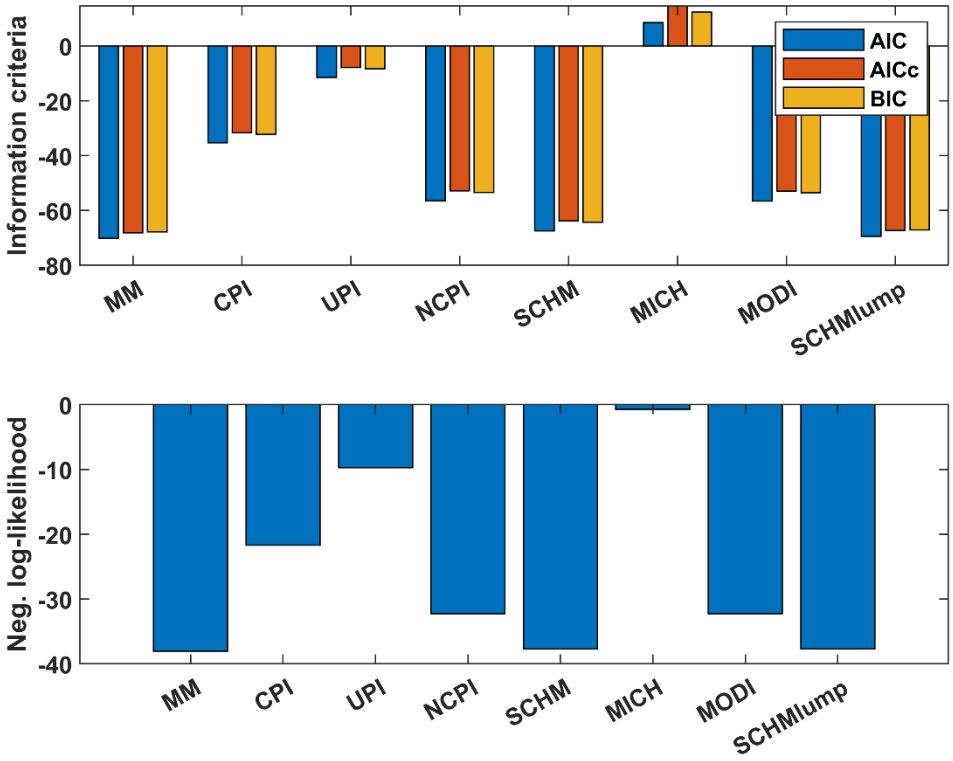

Figure 20: Information criteria and negative log-likelihood as an evaluation of the goodness of fit for candidate models of FDH kinetics based on a 2^4 full factorial design of experiments. Higher absolute values indicate better fits.

It is characteristic of the results that, although the initial values of the SCHM model were employed, the MM model is the closest representation. However, the SCHM (AIC = -67.42), SCHMlump (AIC = -69.37) and MM (AIC = -70.10) models had extremely close AIC values. Thus, model discrimination did not yield a clear result. NCPI stands out because this model is considered to be mechanistically incorrect at NAD^+ concentrations above 1 mM (Kato *et al.*, 1979). In literature 1 mM NAD^+ was considered as the saturation concentration in 50 mM KP_i buffer at pH 7.5 and a temperature of 30 °C. Below this concentration, NADH acts as a non-competitive inhibitor with respect to formate. These conditions can occur at the end of the reactor, when the conversion is advanced and the NAD^+ concentration is below saturation. Above the saturation, NADH is a competitive inhibitor. The MM model exhibited the largest Akaike weight of 0.51, which was close to the SCHMlump model with an Akaike weight of 0.35. However, the Akaike weights represent only relative likelihood of the models, so complete

discrimination cannot be observed. This is underlined by the fact that models with fewer parameters (MM, SCHMlump) are favored by a penalty on the number of parameters according to equation (12) due to the small difference of the log-likelihood. The resulting parameters are summarized in Table 7 and display similar confidence intervals comparing the results of the MBDoE approach. The kinetic parameter k_{cat} and $K_{m,NAD}$ are differing the most with k_{cat} being 64% higher with an accordingly increased $K_{m,NAD}$ of 1.5 mM which is 172% higher than the previously determined $K_{m,NAD}$ for the SCHMlump model (cf. Table 5).

Table 7: Initial values, upper and lower bounds and estimates for the non-linear fit of the full factorial design of experiment approach with the MM model.

Parameter	Lower Bound	Upper Bound	Start	Estimate
k_{cat} [s^{-1}]	$1.25 \cdot 10^{-4}$	4.8	0.15	0.46 ± 30.9%
$K_{m,NAD}$ [mM]	$7.5 \cdot 10^{-3}$	135	0.03	1.50 ± 45.3%
$K_{m,formate}$ [mM]	$1.85 \cdot 10^{-2}$	52.5	0.3	18.2 ± 31.9%
$K_{i,NADH}$ [mM]	$1 \cdot 10^{-2}$	0.3	0.1	-
$K_{i,NAD}$ [mM]	5	117.2	70	-
P_1 [mM^2]	$9.25 \cdot 10^{-2}$	6154	21	-

The apparent kinetic parameters and the 95% confidence intervals for the MM model were applied to model the data in Figure 21. Here, the uniform distribution of the experiments in the experimental space is further shown. The model is able to express the resulting product concentrations in the surface plot, but the regions of interest were not covered well. These regions are the higher concentrations of NAD^+, since the effect of inhibitions by NAD^+ and NADH was not observed and, therefore, not fitted with the CPI, MICH or SCHM model. In addition, an asymptotic approximation to the maximum reaction rate is assumed. However, this is not achieved within the experimental design space. Additionally, this can be observed by the high $K_{m,NAD}$ value compared to the initial value, which describes per definition the position of the substrate concentration at half maximum reaction rate. However, this high value may be an apparent value artificially increased by the inclusion of product inhibition.

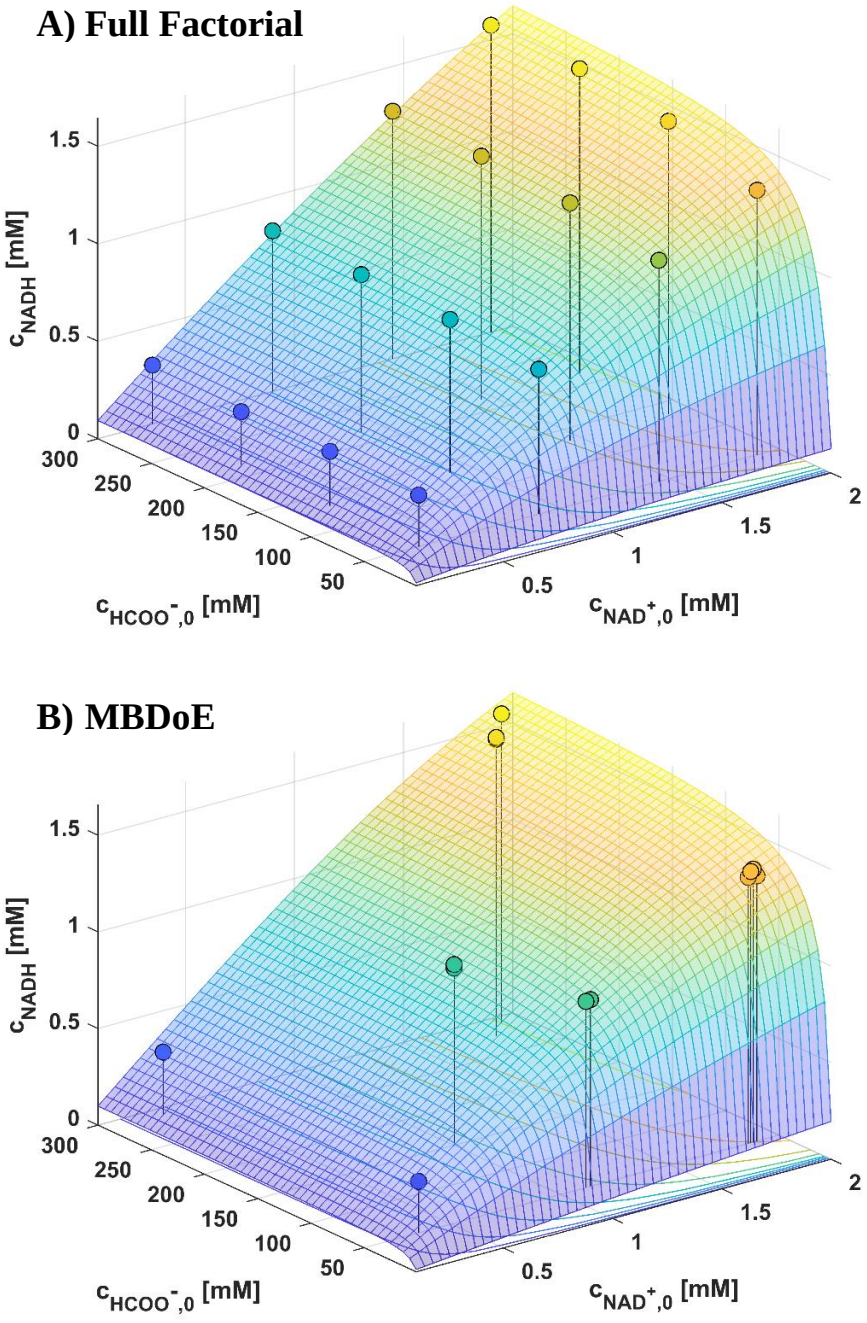

Figure 21: Surface plot for the experimental data of the full factorial experimental campaign with the MM model as surface (A) and the surface plot for the experimental data of the full factorial experimental campaign with the SCHMlump model as surface (B).

Eventually, the model identification with the MBDoE approach and the full factorial design identified two different models with similar confidence intervals for the designated model candidates. Both methods further indicate a high proximity of the model candidates in the design space, but the models MM, SCHM, SCHMlump share the highest AIC and

log-likelihood for both approaches. In contrast to the MBDoE approach, a lot of experiments are executed in a region of low interest for the parameter estimation (cf. Figure 21). Hence, the experimental data show evidence for the efficiency of the approach, but lack in information about the increased accuracy of the results due to the limited design space. However, the parameter estimation is satisfactory for both methods, but the model discrimination achieved no definiteness within the factorial design. Accordingly, it has to be shown whether the automated model discrimination was only successful because of the initial parameter solution by Schmidt *et al.* (2009). Therefore, the influence of the initial parameter estimates had to be investigated in detail.

4.2.4. Exchange of Starting Solution

The initial parameters of the first campaigns were exchanged to create a different scenario for the MBDoE model identification. Due to the limited amount of literature models, the model parameters of Kula *et al.* (1980) were chosen, which were determined by initial rate measurements with the same temperature (30 °C) and pH (8). Furthermore, the model is designed according to most literature models as a double substrate kinetic with a competitive product inhibition of NADH. The parameter limits, fit functions and reactor platform settings were kept constant to isolate the influence of the starting values. Only the starting value for $K_{i,NAD}$ and accordingly P_1 were set to 0 mM since the MATLAB code needs the assignment of a value to involve it in the model discrimination procedure. The turnover number k_{cat} was set to 1.35 s^{-1} which was determined by Schmidt *et al.* (2009), calculated on the basis of progress curves with the CPI model. The original results of Kula *et al.* (1980) suggest a k_{cat} of 4.8 s^{-1}, which was calculated with the molecular weight (41331 Da) of the FDH applied in this work (Megazyme, 2023). The molecular weight should be equal or close to other FDH of *Candida boidinii* wild types resulting in similar values for a v_{max} of 0.007 U/mg. The k_{cat} value of 4.8 s^{-1} value appeared to be overestimated compared to the results reported by Schmidt *et al.* (2009), who also considered their own result of 1.35 s^{-1} applying the same model as overestimated. Following the initial DoE, the model discrimination was unsatisfactory with a weighted AIC of 0.55, prompting further experiments to improve the model discrimination. The results of the initial estimate of the model parameters are shown in Table 8 and show satisfactory confidence intervals. However, changing the starting values significantly influenced the experimental design and resulting kinetic parameters. Although the kinetic

parameters used a double substrate kinetic with a single product inhibition (CPI), the model discrimination resulted in a simple double substrate kinetic (MM) without any inhibition constant. This was already accomplished in the full factorial design, where no experiments were conducted to enhance the SCHMlump model parameter estimation.

Table 8: Initial values, upper and lower bounds and estimates for the non-linear fit of the closed loop experimental campaign DoE approach with starting values according to a simple double substrate kinetic with CPI by Kula *et al.* (1980).

Parameter	Lower Bound	Upper Bound	Start	Estimate
k_{cat} [s^{-1}]	$1.25 \cdot 10^{-4}$	4.8	1.35	0.38 ± 24.7%
$K_{m,NAD}$ [mM]	$7.5 \cdot 10^{-3}$	135	0.07	1.04 ± 40.2%
$K_{m,formate}$ [mM]	$1.85 \cdot 10^{-2}$	52.5	29.3	19.2 ± 28.7%
$K_{i,NADH}$ [mM]	$1 \cdot 10^{-2}$	0.3	0.09	-
$K_{i,NAD}$ [mM]	5	117.2	0	-
P_1 [mM^2]	$9.25 \cdot 10^{-2}$	6154	0	-

The turnover number of 0.38 s^{-1} is in the same magnitude as the previously determined value for the SCHMlump model. $K_{m,formate}$ of 19.2 mM is in the order of magnitude of the starting solution. $K_{m,NAD}$ increased by a factor of 2 in contrast to the first MBDoE campaign and is strongly overestimated in contrast to the starting value of 0.07 mM. The apparent increase in $K_{m,NAD}$ was expected in a model design without a dedicated parameter accounting for product inhibition. This is because models with product inhibition affect the K_m value by the inhibitor concentration dependent constant K_i (cf. CPI, NCPI in Table 1).

After the first initial model discrimination did not reach the threshold, nine additional experiments were performed on the basis of the six first experiments. This time, the experiments were repetitions of the same experiment with a concentration of NAD^+ of 0.445 mM and a formate concentration of 64.5 mM. The progress of the model discrimination is shown in Figure 22. It is noticeable that the three repetitions of the experiments actually cause an increase in model discrimination. However, this is only the change from 0.55 to 0.61 for the MM model. Simultaneously, the weighted AIC of the SCHMlump model increased to 0.22 while the SCHM model decreased from 0.37 to 0.22 in the first experiment. Here, the weighted AIC of both models fluctuated around the value of 0.2 and did not show any trend during the MBDoE campaign. The fluctuation indicates a stagnation in a local minimum of the non-linear regression problem, which

could not be solved based on the initial parameter estimates and the corresponding MBDoE.

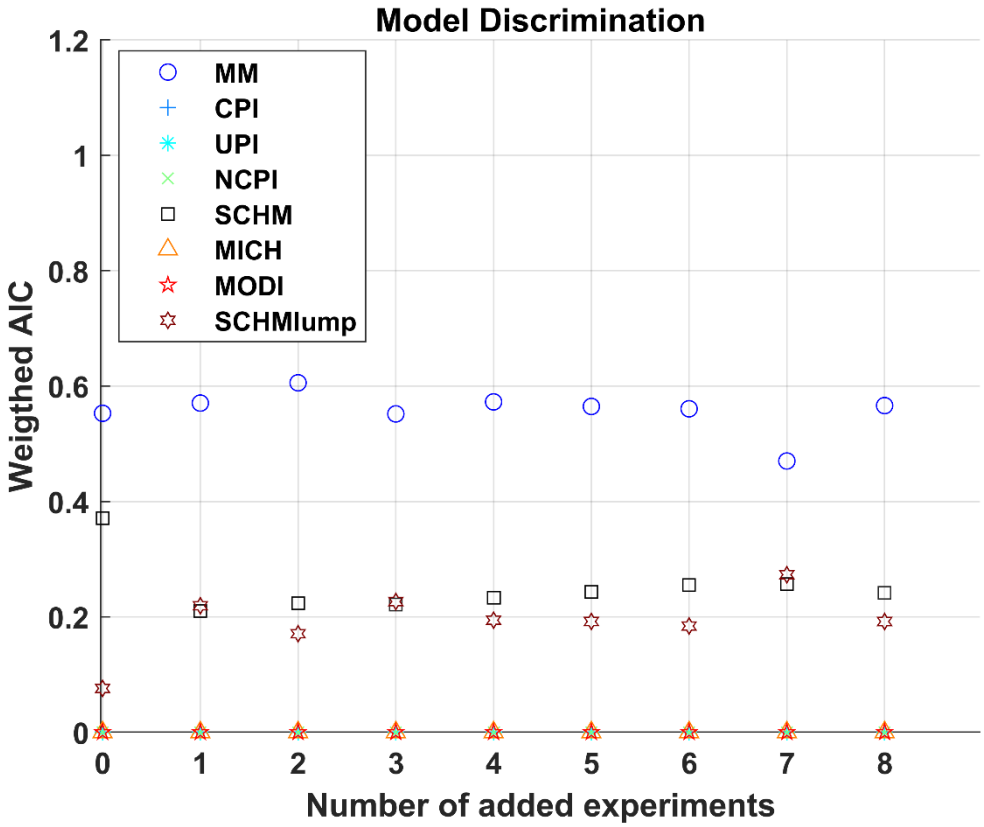

Figure 22: Change of the weighted AIC over the number of added experiments. Experimental conditions: KP_i buffer 50 mM pH 8, temperature: 30 °C. Substrate concentrations: 0.445 mM NAD^+ and 64.5 mM formate.

Nevertheless, the resulting confidence intervals of the MM parameters were drastically reduced by the addition of extra experiments. Although the MBDoE campaign did not aim for the increased parameter estimation, the amount experiments reduced the confidence intervals naturally. This is the case due to the fact that the confidence intervals are directly dependent on the amount of data points n and the standard deviation σ (cf. equation 24). The repetition of the experiment had a mean NADH concentration of 0.397 mM with a standard deviation of 0.008 mM, which accounts for a percentage standard deviation of 2%. Due to this small deviation and the increase of the experimental count, the confidence intervals could be reduced to 9.8%, 14.7% and 14.8% for k_{cat}, $K_{m,NAD}$ and $K_{m,formate}$ respectively (cf. Table 9). These confidence intervals are relatively

low when compared with an analytical standard deviation of 3.4% to 9.6%.The parameters itself did not change significantly after the nine added experiments. All parameters slightly increased with $K_{m,NAD}$ increasing the most from 1.04 mM to 1.15 mM.

Table 9: Initial values, upper and lower bounds and estimates for the non-linear fit of the closed loop experimental campaign DoE approach with starting values according to a simple double substrate kinetic with CPI by Kula *et al.* (1980).

Parameter	Lower Bound	Upper Bound	Start	Estimate
k_{cat} [s^{-1}]	$1.25 \cdot 10^{-4}$	4.8	1.35	0.40 ± 9.80%
$K_{m,NAD}$ [mM]	$7.5 \cdot 10^{-3}$	135	0.07	1.15 ± 14.7%
$K_{m,formate}$ [mM]	$1.85 \cdot 10^{-2}$	52.5	29.3	20.0 ± 14.8%
$K_{i,NADH}$ [mM]	$1 \cdot 10^{-2}$	0.3	0.09	-
$K_{i,NAD}$ [mM]	5	117.2	0	-
P_1 [mM^2]	$9.25 \cdot 10^{-2}$	6154	0	-

The difficulties in model discrimination in the available design space are observable in Figure 22 and Figure 23. Here, both the information criteria and the negative log-likelihood indicate the goodness of fit for three fitting models, namely MM, SCHM and SCHMlump are almost equal. The AIC of the models accounted for -96.18, -94.32 and -93.30 showing the close proximity of the fitting (cf. Figure 23). Only the number of parameters is displaying the difference between the SCHM model and the other two models, which is four in contrast to three for MM and SCHMlump, respectively. This is shown in the values of the negative log-likelihood of -51.16 for SCHM, whereas MM exhibits a value of -51.09 and SCHMlump of 49.65 (cf. Figure 23).

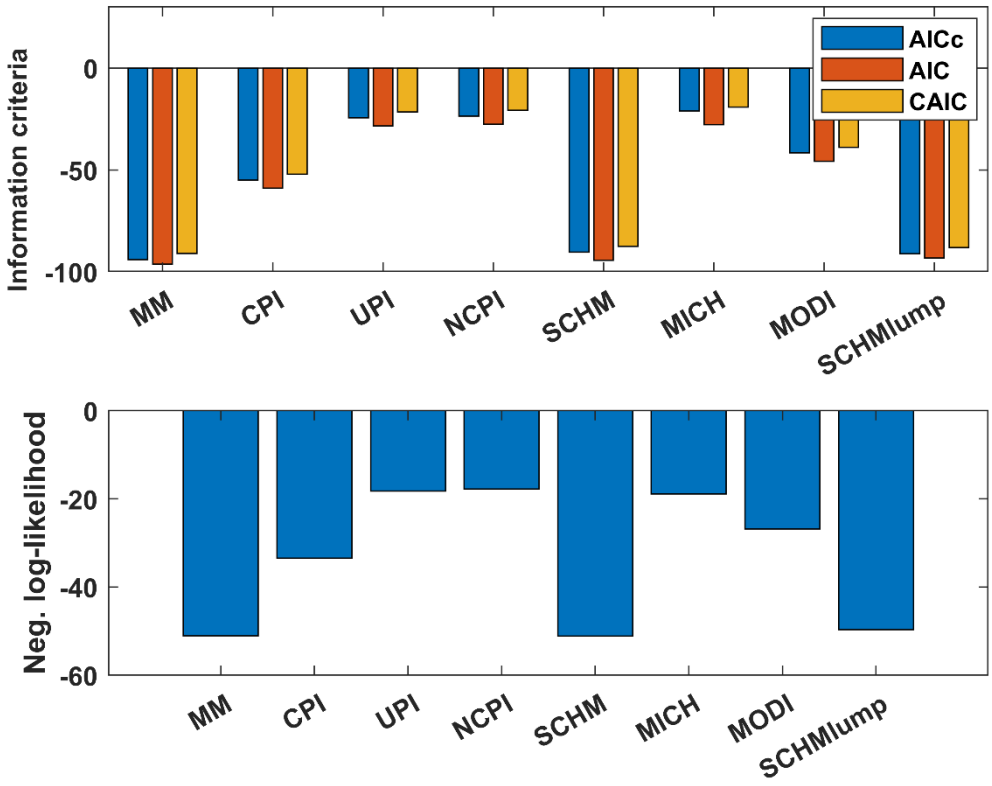

Figure 23: Comparison of the information criteria for the MBDoE campaign with starting values adapted by Kula *et al.* (1980). Higher absolute values indicate better fits.

After comparing the results of both experimental design campaigns, no clear trend for advantages of the MBDoE method over the full factorial method could be established with certainty. The increase of the parameter accuracy is mainly due to the number of experiments and the associated refinement of the standard deviation. This does not imply that the MBDoE offers any advantages over conventional statistical experimental design procedures, nevertheless, it is equally appropriate in the limited design space. According to the first closed-loop results, the method is repeatable and displays the best non-linear fit in the given design space even when the initial estimates after six experiments are showing discrepancies in experimental results. However, the applied procedure is prone to stagnation in a local minimum after the initial DoE, because the weighted AIC describes the relative performance of the model. It is calculated on the basis of all AIC present throughout the model candidates so that small differences in the log-likelihood in combination with the parameter penalty term result in apparently high levels of discrimination. Thus, the best model is discriminated on a small sample size of the given

design space. As a result, this could lead to an erroneous model discrimination in the first place, so that the parameter estimation threshold is triggered and the model discrimination is not further improved. This challenge could be overcome by the extension of the design space, to be able to observe larger differences in the model discrimination within the initial DoE experiments.

A change in the experimental approach is necessary to achieve improved model discrimination. One challenge is the non-convexity of the non-linear regression problem which forced the modelling procedure into a local minimum once a satisfactory solution was found. This can be explained by the one-dimensionality of the initial parameter solutions, which follow the gradient of the FIM in local solutions. The iteration of different starting parameters could lead to global solutions. Unfortunately, the execution of several iteration steps requires higher calculation times, which are not feasible in the continuous operation without optimized software structure and with limited processor capacity. The second change in the experimental design should be an extension of the concentration range. This leads to clearer distinction between competing models. Particularly the maximum reactions rate v_{max} was not reached with a maximum NAD^+ concentration of 1.9 mM. Furthermore, the minimal formate concentration of 50 mM was higher than the estimated $K_{m,formate}$ values which could lead to erroneous results. Unfortunately, the hardware was limited regarding pump linear force, detection limits of the UV/Vis analysis and computational power. These limitations dictated a certain design space regarding dilution steps, residence times and maximum measurable concentrations of NADH. These limitations could not be overcome inside the closed-loop setup but the limitations were circumvented by measuring steady state product concentrations offline. Thus, the concentration range could be extended within the optimal experimental design. For the reasons mentioned above, the automated closed-loop system reached its limits, so that a further investigation of the method was conducted offline. This evolved new possibilities in terms of calculation times and detection limits (cf. chapter 5). Thus, the MBDoE method could be investigated in detail with less limitations and its applicability was reevaluated.

4.3. Interim Summary

This chapter details the creation of the experimental platform, resulting in:

- A reactor system for automated experiments with limitations in product detection and system backpressure was designed.
- A software structure using Python, with comprehensive settings for experimental execution (syringe pump dosing system, liquid flow monitoring, and inline UV/Vis analytics) and a model fitting procedure in MATLAB was developed.
- An optimal experimental design was implemented in the automated reactor infrastructure to carry out efficient experiments based on a data set of six full factorial experiments.

The key findings utilizing the MBDoE method in the closed-loop system highlighted following aspects:

- The comparison of the conducted initial experiments on the automated platform resulted in a standard deviation range between 3.4% and 9.6% which confirm repeatability.
- Individual experimental runs result in repeatable MBDoE campaigns with diverse confidence intervals but consistent model parameters.
- Closed-loop experiments lead to a greater degree of model discrimination. However, they yield different outcomes based on the initial solution applied. The model discrimination was constrained by the design space limitations and the limited number of experiments.

Based on these findings, the following research questions and considerations arise:

- What is the impact of a non-linear regression approach with multiple starting values on results and calculation time?
- Can the fitting procedure influence the resulting parameters?
- Is the use of MBDoE significantly superior to standard methods for enhancing model discrimination

5. Post-processing

By reaching the limitations of the automated reactor platform, the evaluation of the model discrimination via the AIC had to be progressed *in silico*. In this chapter, the influence of the starting parameters was made negligible by a multi-start heuristic. This function runs repeating solvers to find multiple local minima and returns the lowest local minimum among the solver solutions in the given design space. This does not mean, that the solutions are the definite global minimum, however, it is the lowest minimum discovered from multiple starting points. In addition, the continuously measured results were extended and both the influence of the design space and the application of the multi-start heuristic to these data was analyzed. Finally, important parameters for the model identification were identified.[1]

5.1. Multi-start Model Identification

The results of chapter 4.2.4 indicated a high influence of the starting parameter on the non-linear regression. Therefore, the model discrimination and parameter estimation are highly dependent on the chosen initial parameter estimates, resulting in different final solutions. In order to minimize this influence, 200 iterations of the optimization problem were started in parallel applied on the data of the first MBDoE run. Therefore, multiple initial starting values for the non-linear regression were generated based on different starting points to avoid stagnation in local minima. This should result on the one hand in the most precise model representing the training data set and on the other hand avoid the mentioned optimization stagnation. This improvement is necessary for the MBDoE approach to optimize the experiment towards the correct model and not towards a local minimum. Furthermore, it exhibits insights, if the weakness in model discrimination is based on wrong initial solutions due to the non-convexity of the objective function or the constraints of the underlying design space.

5.1.1. Starting Parameters

The closed-loop results of all campaigns indicated a high error for the models MICH and UPI initially. This suboptimal result may be due to an initial solution that is misleading, causing a non-optimal minimum when starting with a high distance from the global minimum. For this reason, the multi-start heuristic was first applied on the data set of the

[1]Parts of this chapter will be published in *"Closed-loop Identification of Enzyme Kinetics Applying Model-Based Design of Experiments"* by Hennecke *et al.*

first closed-loop campaign to investigate the influence of that approach on the resulting model discrimination and parameter estimation. Therefore, each model fit is optimized based on the 200 iterations applied. The resulting parity plot (cf. Figure 24) shows a high agreement of all models in the design space. Only the MICH model is not able to represent the measured data with high accuracy. This is the first improvement of the multi-start heuristic. In the MBDoE approach, the use of a single set of initial parameters resulted in preferential model discrimination with respect to the model from which the parameters originated. Additionally, some models had poor fits due to these initial values.

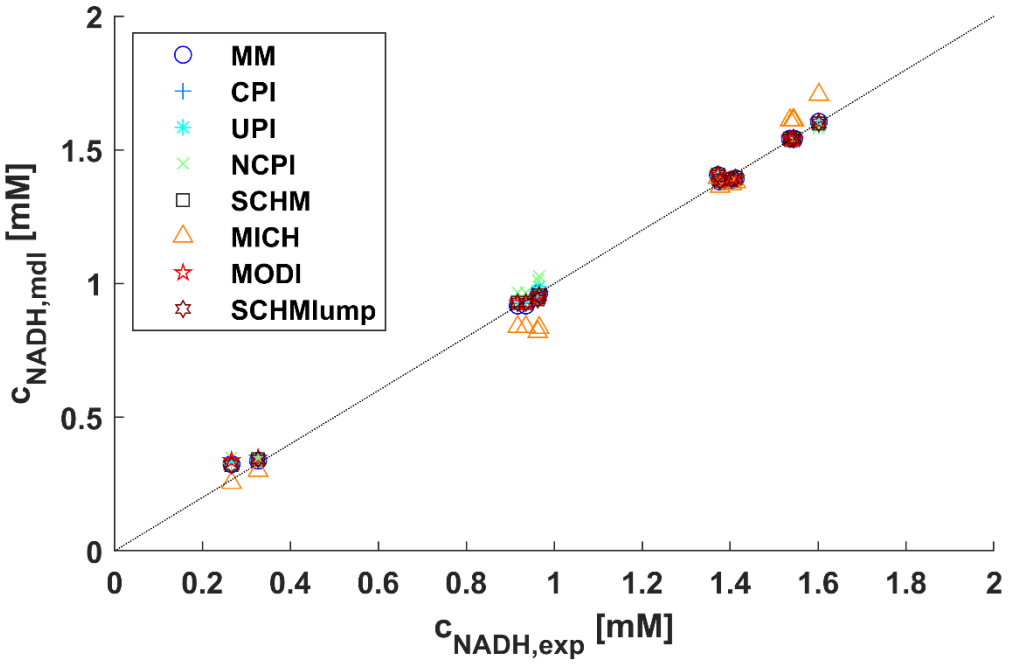

Figure 24: Parity plot for the modelled and experimental product concentration after the application of a multi-start heuristic with 200 iterations on the data of the closed-loop results based on starting values according to Schmidt *et al.* (2009).

The high accordance of all models is additionally represented in the AIC values, which show high conformity (Figure 25). The MM model is the best fitting model with the lowest AIC value of -69.2 among the candidates. The kinetic parameters of the MM model resulted in a k_{cat} of 0.34 s^{-1}± 20.0%, $K_{m,NAD}$ of 0.85 mM ± 37.1% and $K_{m,formate}$ of 15.6 mM± 21.9%. A small difference in the kinetic parameters is observed compared to the previously determined parameters with the same model of k_{cat} of 0.4 s^{-1} ± 9.8%, $K_{m,NAD}$ of 1.15 mM ± 14.7% and $K_{m,formate}$ of 20.0 mM ± 14.8%. The turnover number is

adjusted downwards whereas the Michaelis Menten constants are adjusted upwards. Moreover, the confidence intervals doubled in contrast to the MBDoE run, which aimed for the parameter accuracy of the MM model. Hence, the purpose of the MBDoE approach was achieved and is confirmed with the post-processed experimental data of the first closed-loop run, although the model discrimination was misled by the starting parameters.

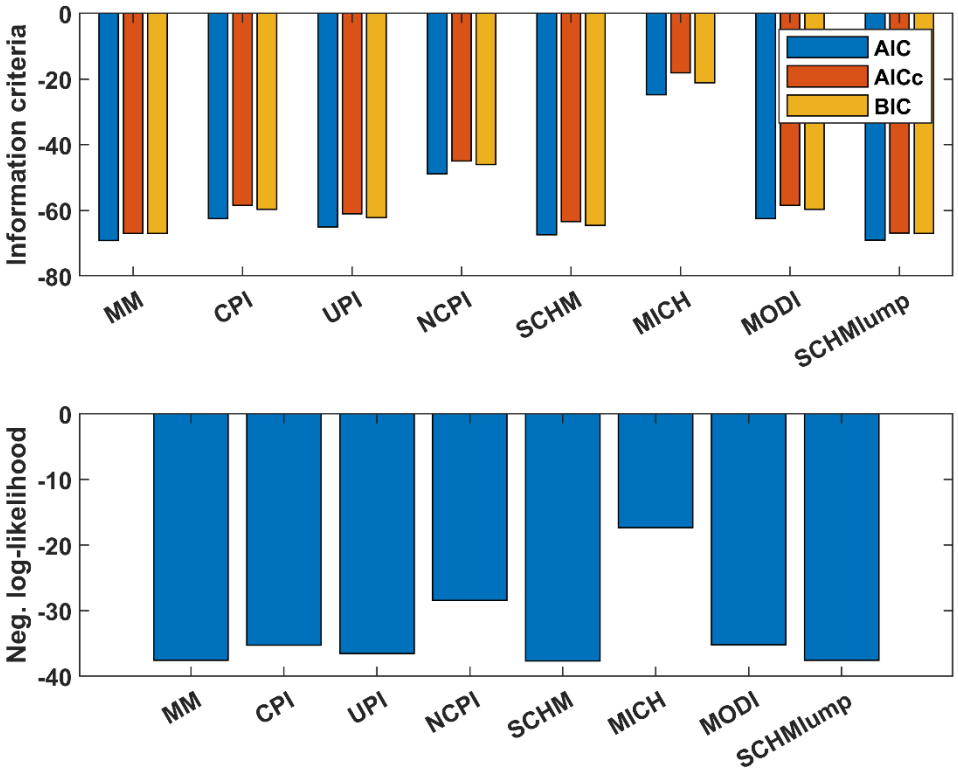

Figure 25: Information criteria and negative log-likelihood as an evaluation of the goodness of fit for candidate models of FDH kinetics based on the first MBDoE campaign fitted with the multi-start heuristic with 200 iterations. Higher absolute values indicate better fits.

Although the differences are small, the weighted AIC value for the MM model accounts for 0.39 (AIC = -69.2). The second highest weighted AIC is the SCHMlump model which represents the data equally well with a weighted AIC of 0.38 (AIC = -69.1). The SCHM model fits the data also well with an AIC of -67.4. Every other model is minimally worse fitting except for MICH which accounts for the highest AIC of -24.7.

The MM model was shown to represent the experimental data from the first MBDoE campaign with the highest AIC weight after the application of the multi-start heuristic. Consequentially, the impact on the model identification of the dataset from chapter 4.2.4 was investigated, which already showed high conformity with the MM model. Therefore, the multi-start approach was applied on the second MBDoE campaign to investigate any difference in the result. The parity plot of the modelled NADH concentration versus the experimental concentration shows a high accuracy of the model predictions except for the MICH model again (cf Figure 26).

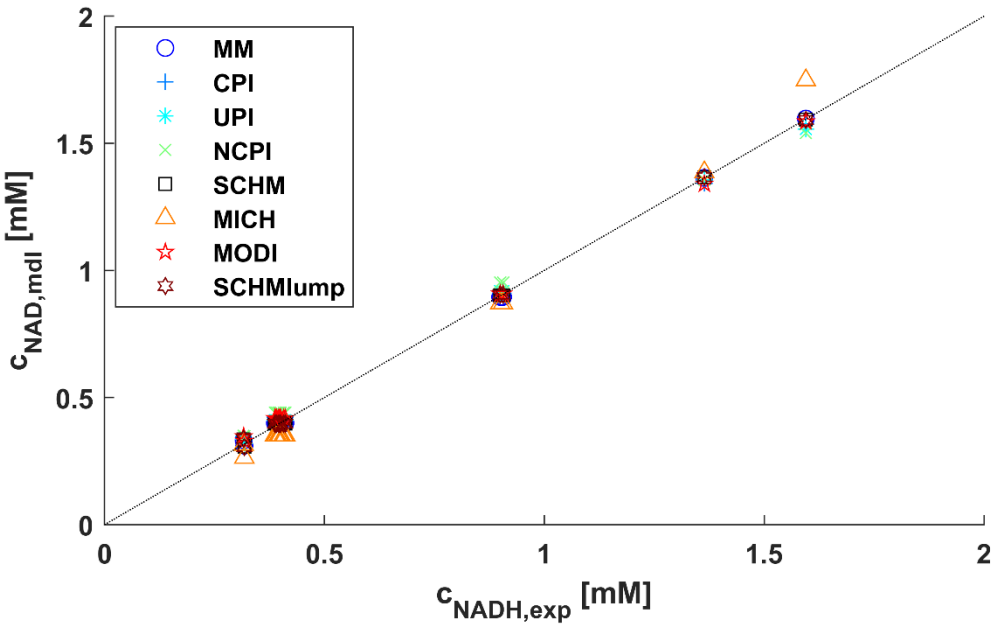

Figure 26: Parity plot for the modelled and experimental product concentration after the application of a multi-start heuristic with 200 iterations on the MBDoE-2 closed-loop campaign data based on starting values according to Kula *et al.* (1980).

This model predicts the correct NADH concentrations worse than other models at low (<0.5 mM) and high (>1.5 mM) NADH concentrations. All other models show a well agreement with the experimental data by applying the multi-start heuristic. Thus, it was shown once again that the use of multiple starting solutions enables to the fit all the models equally well to the data and that the initial values do not strongly favor one or more of the models. As displayed in Figure 27, the three best fitting models are represented in the reevaluation of the data, however, the negative log-likelihood of the

worst models decreased drastically (UPI, NCPI, and MODI). This further underlines the fact that every model is able to fit the data with proper starting estimates.

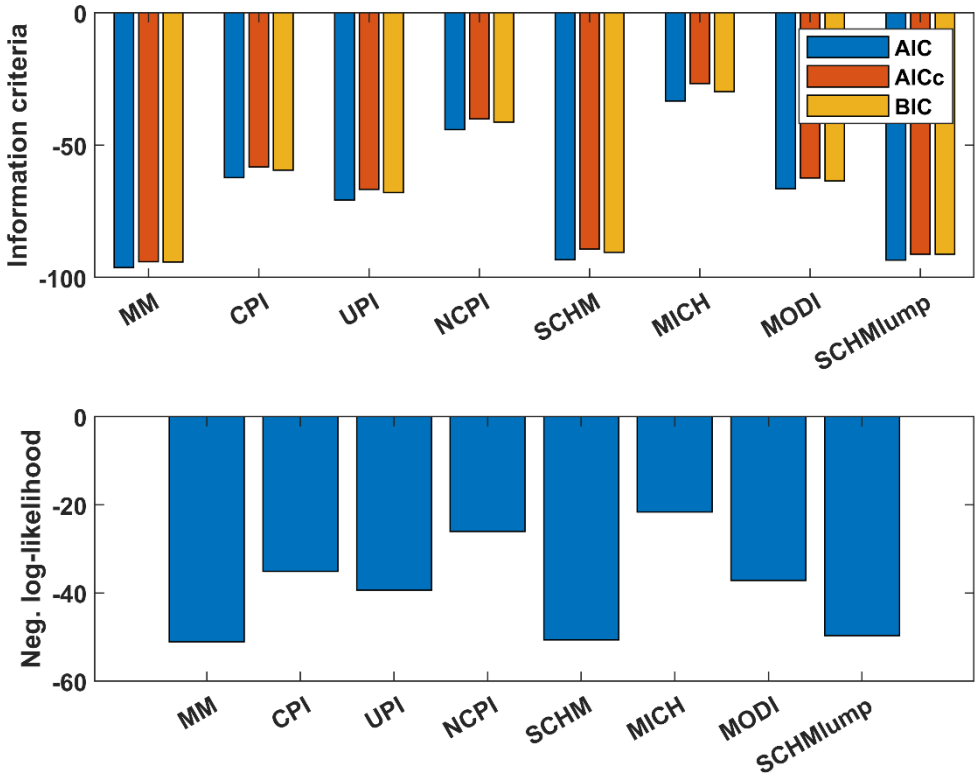

Figure 27: Information criteria and negative log-likelihood as an evaluation of the goodness of fit for candidate models of FDH kinetics based on the MBDoE campaign based on MM start parameters fitted with the multi-start heuristic. Higher absolute values indicate better fits.

With the application of the multi-start heuristic on the data set of the first MBDoE run, the model identification changed from the SCHMlump model to the MM model. This study confirmed the significant impact of the initial parameter regardless of the data set applied, as the data were generated by maximizing the parameter accuracy of the SCHMlump model through the *e-optimal* experimental design. Based on these results, the influences of other aspects of parameter estimation were examined, such as the applied optimization algorithm. In early iterations of the automated experimental platform, the Levenberg-Marquardt (LM) algorithm was utilized to fit the data. However, after several script updates, the numerical optimization algorithm was replaced with the trust region reflective (TRR) method and the finite difference method was applied to solve

the optimization problem. The regression results should not be affected by this change. However, the time required for the fitting procedure may be impacted. This may not be an issue for a small amount of data, but as the automated reactor platform is further developed, more data may be generated. Therefore, effective numerical optimization is necessary for MBDoE campaigns.

5.1.2. Optimization Algorithm

To investigate the influence of the optimization algorithm, the LM algorithm was used to fit the MM dataset in addition to the original fit with the TRR algorithm. The identical MM model and according parameters as those obtained with the TRR algorithm in the MBDoE were achieved using the trust region reflective algorithm with finite difference method. Once again, the kinetic parameters of the model achieved the values with 95% confidence intervals of k_{cat} of 0.4 s^{-1} ± 9.8%, $K_{m,NAD}$ of 1.15 mM± 14.7% and $K_{M,formate}$ of 20.0 mM± 14.8%. This indicates that the same local or even global solution has been found with the TRR algorithm. Thus, it was investigated, which algorithm is more suitable for automated model identification.

The LM algorithm is a numerical optimization algorithm for non-linear regression problems applying the method of least squares. It solves a quadratic minimization problem without constraints and is based on a second Taylor approximation to the objective function. The TRR method is a further development of the LM algorithm, however, exhibits differences in the solution of the quadratic sub-problem. In contrast to its predecessor, this method has deterministic step-size constraints and can follow the negative curvature of the objective function. By this, the trust-region reflective method is faster than the LM algorithm. (Nocedal *et al.*, 2006)

When extrapolating the purpose of the automatic reactor platform, it should provide rapid and accurate results. Thus, the TRR is preferred over the LM algorithm, because the calculation is more rapid and results in the same kinetic parameters within the small data set provided in the MBDoE campaign (cf. Table 10). Nonetheless, the TRR algorithm was applied with the same stiff starting parameters of Kula *et al.* (1980) on the second MBDoE data set and this resulted in the SCHM model with the kinetic parameters k_{cat} of 0.3 s^{-1}± 9.4%, $K_{m,NAD}$ of 0.73 mM ± 19.7% and P_1 of 31.9 mM^2 ± 16.8%. This shows that the optimization algorithm can have an influence on the resulting parameters if the data

set is limited and the starting parameters are not optimally chosen. Consequentially, the TRR is the superior choice for the automated approach but with the requirement of the multi-start heuristic to get comprehensive results.

Table 10: Influence of the optimization algorithm for the least squared curve fit of different data sets applying TRR and LM with different options on the calculation time. Time is based on the application of an AMD Ryzen 5 3600 6-Core Processor with 3.60 GHz using the Parallel Computing Toolbox™ of MATLAB.

Data set (number of experiments)	**Algorithm**	**Finite Difference**	**Time [s]**
MBDoE2 (15)	LM	-	170
MBDoE2 (15)	TRR	-	90
MBDoE2 (15)	LM	central	164
MBDoE2 (15)	LM	forward	165
MBDoE2 (15)	TRR	central	162
MBDoE2 (15)	TRR	forward	160
Meta (112)	LM	central	1588
Meta (112)	TRR	central	1143

For that reason, different options for TRR and LM are applied to fit the data of the second MBDoE and the Meta data set which contains the MBDoE campaigns and the full factorial campaign of chapter 4.2.1, 4.2.2, 4.2.3 and 4.2.4. All data was treated with an AMD Ryzen 5 3600 6-Core Processor with 3.60 GHz base clock applying the Parallel Computing Toolbox™ of MATLAB. This utilizes all cores in parallel for the calculation of the kinetic parameters and is therefore more suitable than the processor employed in the laptop of the automated reactor platform (AMD Ryzen 5 2500U 4-Core Processor

with 2 GHz base clock). For small data sets (15 data points), the only large difference is noticed between the TRR and all other optimization algorithm which takes 90 s in contrast to 160 s to 170 s. Especially the claimed superiority is not seen here when comparing TRR and LM with the finite difference method which take 162 s and 164 s respectively. However, this small discrepancy is drastically increased when calculating the optimum model, a larger data set (112 data points). TRR required 1143 s to fit the data whereas LM took 1588 s to fit the data. The difference of 7.4 min corresponds to almost four residence times in respect to the experimental platform, and is further increased when applying less or weaker processor cores or shorter residence times. Thus, for future work on the automatization of enzymatic modelling, especially on more complex reaction systems, the optimization algorithm displays an influence in results and time. Additionally, the optimization and parallelization of the model fitting is recommended to shorten the overall experimental time. This factor might be crucial for screenings of less stable enzymes.

Regardless of the time saving, the evaluation with both LM and TRR resulted in the same model for the 112 given data points. With a weighted AIC of 0.67 the MM resulted as the best model to fit all the data. The values for k_{cat} of $0.32\ s^{-1} \pm 10.1\%$, $K_{m,NAD}$ of $0.78\ mM \pm 19.2\%$ and $K_{M,formate}$ of $16.6\ mM \pm 12.6\%$ are the result including the comprehensive data set of 112 experiments. This shows on the one hand, that the confidence intervals are not drastically decreased by adding more data points and on the other hand, that the model discrimination stopped improving over an AIC of 0.67. In addition, the results are in good agreement with the results of the automated MBDoE approach resulting in k_{cat} of $0.34\ s^{-1} \pm 20.0\%$, $K_{m,NAD}$ of $0.85\ mM \pm 37.1\%$ and $K_{M,formate}$ of $15.6\ mM \pm 21.9\%$ which validates the result in the observed design space.

Ultimately, all models except for MICH can fit the data well inside the design space and the solution is neither dependent on the starting parameters nor on the optimization algorithm with the multi-start heuristic. Hence, the limitations of the evaluation scope were reached and the results can only be improved by increasing the design space.

5.1.3. Experimental Design

Since the starting values and the optimization algorithm displayed a high influence on the experimental and closed-loop results, the influence of the experimental design is

additionally evaluated by the multi-start approach. The research focuses on whether the superiority of the MBDoE approach in chapter 4.2.3 is only due to the choice of initial values, or whether the same tendency can be observed when multiple initial solutions are applied.

When applying the multi-start heuristic with 200 iterations on the experimental data of the full factorial design, similar Akaike weights for the model discrimination resulted in comparison to the fixed starting value evaluation. With the Akaike weight of 0.50 the MM model (AIC = -70.1) was the best model among the candidates followed by the SCHMlump model with the Akaike weight of 0.34 (AIC = -69.4) and the SCHM model with an Akaike weight of 0.16 (AIC = -67.8). Both the AIC values and the kinetic parameters did not change after application of the multi-start heuristic. Even the confidence intervals show high accordance to prior results. The parameters were estimated for k_{cat} of 0.46 s^{-1} ± 30.9%, $K_{m,NAD}$ of 1.50 mM ± 45.3% and $K_{m,formate}$ of 18.2 mM ± 31.9% again. Focusing on the model discrimination in Figure 28, it is observable, in combination with the previous results from chapter 5.1.1, that all models now fit more accurately. Nevertheless, the three best models remain MM, SCHMlump and SCHM.

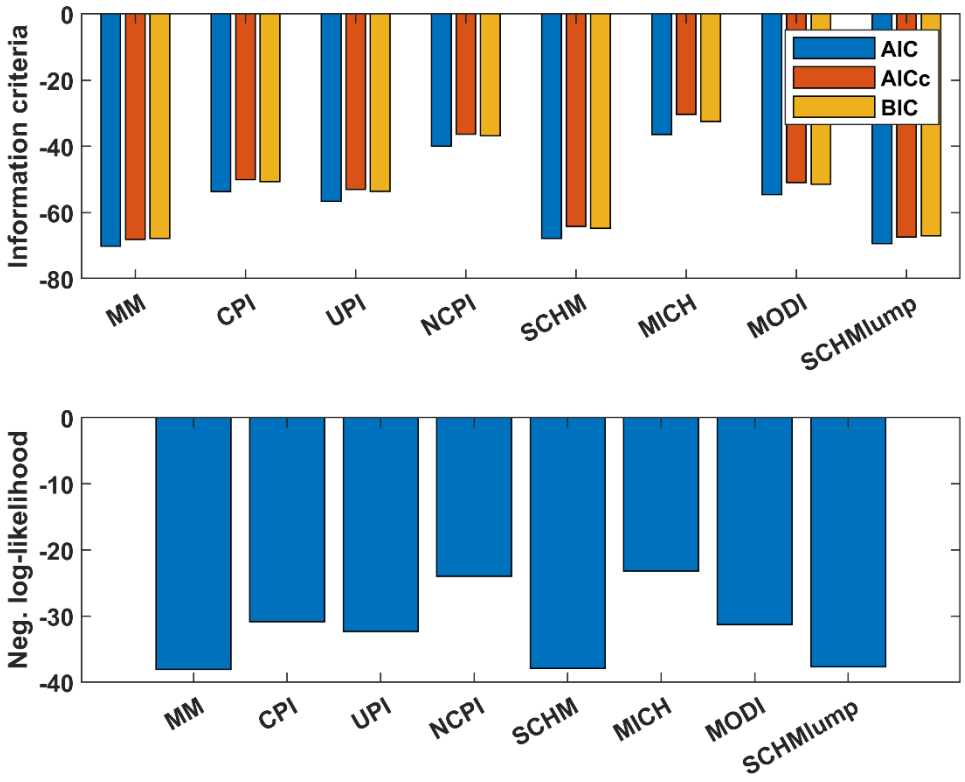

Figure 28: Information criteria and negative log-likelihood as an evaluation of the goodness of fit for candidate models of FDH kinetics based on a 2^4 full factorial DoE fitted with the multi-start heuristic with 200 iterations. Higher absolute values indicate better fits.

By applying the multi-start heuristic on the full factorial data set, no new insights could be generated by evaluating the information criteria or the resulting parameters. Therefore, a sensitivity analysis of these data is evaluated to compare the MBDoE approach and the full factorial approach. This analysis is part of the automated design experimental campaign and show the first derivatives of the best model function describing the product concentration of the experiments. The plotted data points show the experiments conducted and the color code provides the value of the derivative. The higher the value of the derivative, the higher is the respective sensitivity of the parameter to that experiment. A high sensitivity indicates a high deviation of the probability distribution at this point in the FIM, which changes the negative log-likelihood of the model and therefore the model accuracy. That means, that the higher absolute value of the derivative the higher the impact of the experiment on the model identification results, as shown in Figure 29. The algebraic sign indicates whether the change has a positive or negative

effect on the output of the model response. For instance, an increase in NAD^+ concentration exerts a strong impact on v_{max} and results in a higher overall reaction rate.

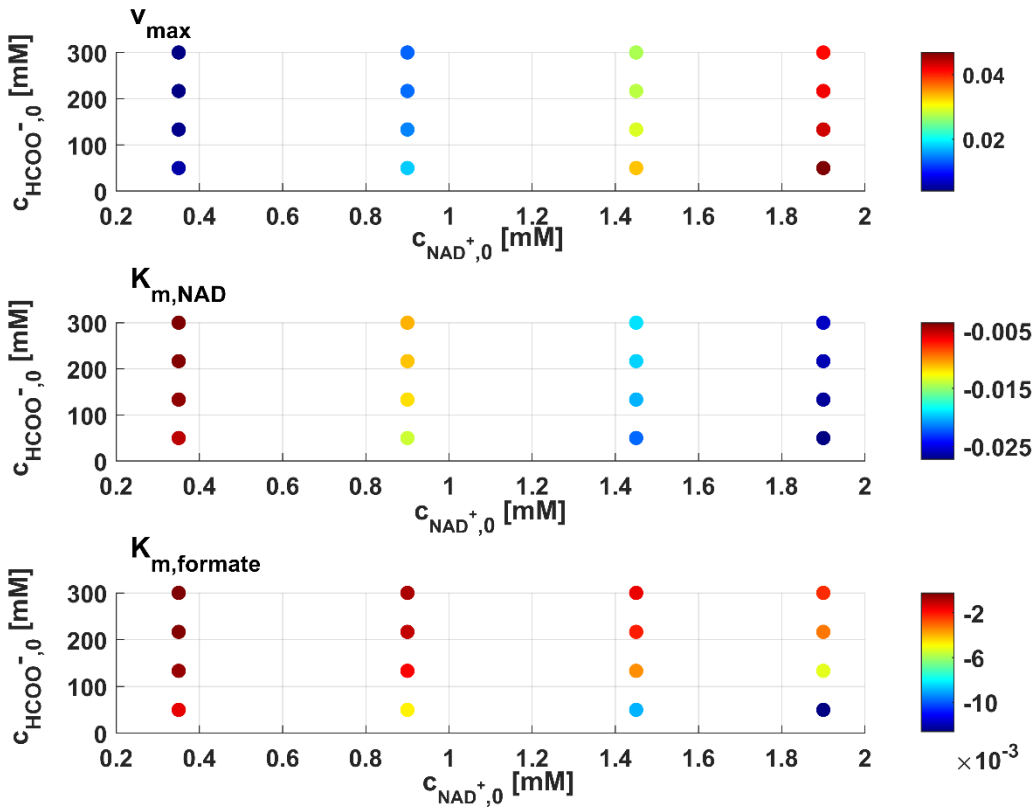

Figure 29: Sensitivity analysis of the full factorial data set for the resulted MM model. Larger absolute values indicate larger changes in the model response.

As observed in Figure 29, for v_{max} the sensitivity is the highest at maximum NAD^+ concentration and minimum formate concentration. The positive entries indicate an increasing model response at these experimental points. In the first MBDoE campaign, the experiment clusters at maximum NAD^+ concentration due to the highest absolute value among the parameter sensitivities (cf. Figure 18). This is the result of the *e-optimal* design, which focuses on the highest overall information content of an experiment for parameter estimation. The sensitivity of $K_{m,NAD}$ is similarly distributed as well as the information content. Accordingly, impactful data for v_{max} and $K_{m,NAD}$ is obtained at elevated NAD^+ concentrations. Typically, v_{max} is determined at high substrate concentrations, while K_m is determined in the region between zero and first order kinetics. In this region, the change in reaction rate decreases significantly. This suggests that v_{max} may not be achieved within the NAD^+ design space. Although the $K_{m,formate}$ value is comparably high at 18.2 mM ± 31.9%, the highest sensitivity is observed at the

experimental data point with 1.9 mM NAD^+ and 50 mM formate which indicates, that the concentration range was not sufficiently low to estimate the true value for $K_{m,formate}$. Nevertheless, the full factorial design resulted in a similar model quality with one extra experiment. It is observed that some experimental data points offer low sensitivity for every parameter and are therefore obsolete for an efficient model identification. Especially the experiments at low NAD^+ concentrations offered the lowest sensitivity for all parameters. This does not imply, that these experiments are valueless, but display the lowest impact on the parameter estimation. The results clearly show the advantages of the MBDoE approach over the full factorial DoE approach. In this work, the level of complexity of the input variables was rather low. As the level of complexity increases, the discrepancy between efficient experimental design by MBDoE and unguided experimental design increases accordingly. In addition, data-driven optimal design increases experimental efficiency regardless of the optimization objective.

In Figure 30, the experiments carried out in the first MBDoE campaign showed high sensitivities for all parameters at maximum NAD^+ concentration and minimum formate concentration (50 mM). The NAD^+ concentration decreased with time, so that the maximum NAD^+ concentration was adjusted by the degradation rate. Interestingly, the sensitivity of the experiments at 50 mM formate and 1.12 mM NAD^+ is medium for all parameters, but compared to the parameter estimation results in chapter 4.2.1, these experiments resulted in the highest increase in parameter estimation. This directly translates into the highest increase in the negative log-likelihood for the selected model candidate (cf. Figure 16).

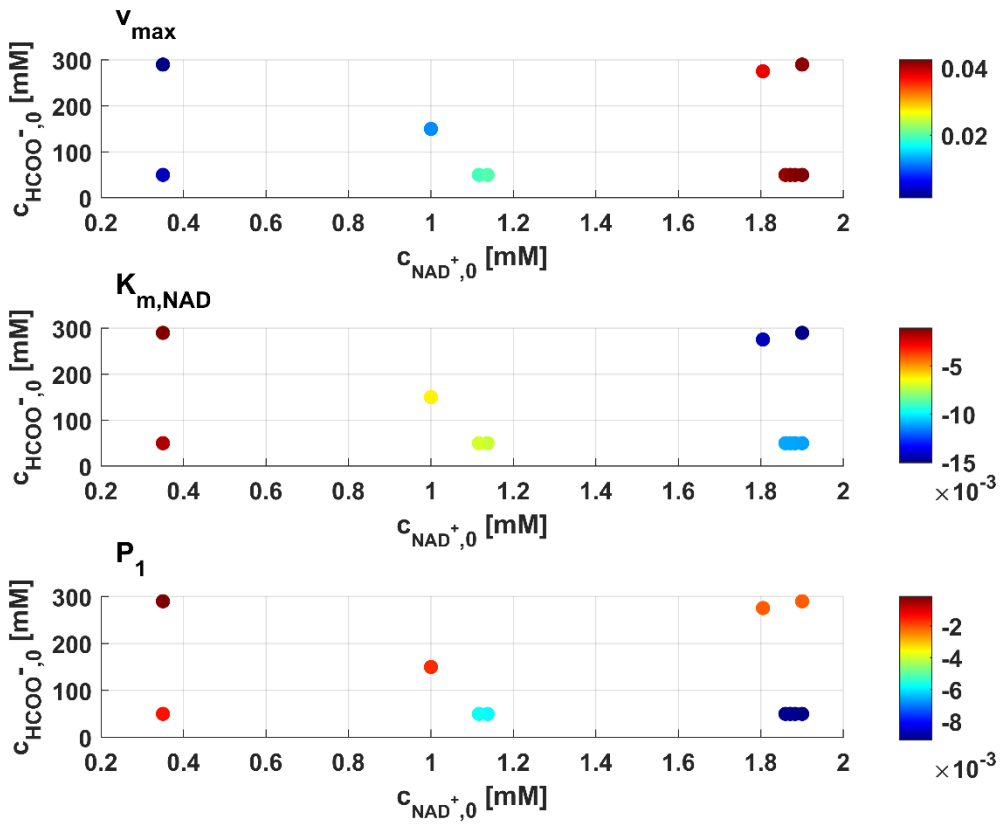

Figure 30: Sensitivity analysis of the first MBDoE approach for the SCHMlump model. Large absolute values indicate large changes in the model response.

The sensitivity analysis of the MBDoE approach with the initial values according to Kula *et al.* (1980) shows a comparable distribution of experiments. These experiments did not display high sensitivity for the parameters but for the model accuracy. Figure 31 clearly illustrates the difference between parameter accuracy and model discrimination in the MBDoE approach. The method achieves this by including two goals in the modeling procedure: model discrimination and parameter estimation. These objectives require different experiments.

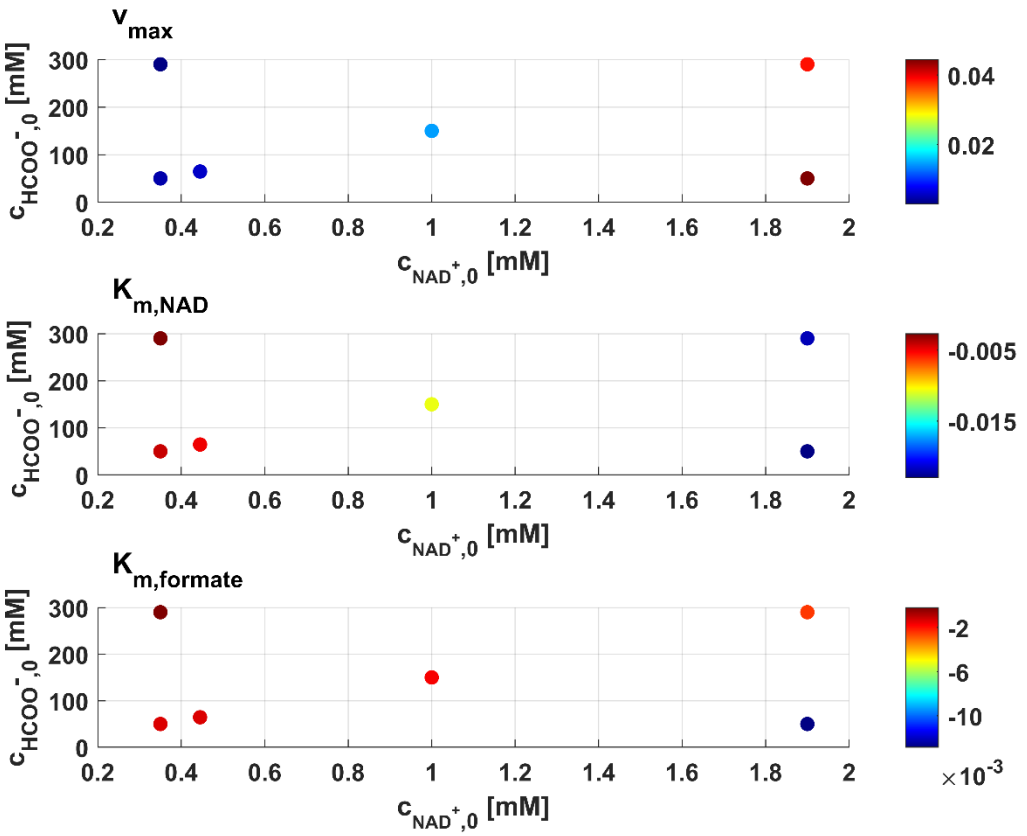

Figure 31: Sensitivity analysis of the second MBDoE approach for the MM model. Large absolute values indicate large changes in the model response.

The first MBDoE campaign resulted in many experiments with high parameter sensitivity because model discrimination reached the threshold of a weighted AIC of 0.95. However, the second MBDoE campaign did not achieve this level of accuracy and included less sensitive experiments in terms of parameters. Although some experiments were highly sensitive in terms of model discrimination, the sensitivity analysis indicated that the concentration limits of the design space should be considered. Model discrimination was not possible in the narrow design space, so the data set had to be extended through manual experiments.

5.2. Extension of Design Space

The kinetic parameters of the MM model and the SCHMlump model showed a maximum reaction rate v_{max} outside of the screened design space. Additionally, the model discrimination with the multi-start approach was not able to fully discriminate one model as the superior candidate. Correspondingly, the design space in the automated reactor

platform is too limited to fully reach the intended model identification. Therefore, the theoretical design space was extended based on the previously determined kinetic parameters. Due to the *in silico* application, the MBDoE design can be performed in different approaches separately to design experiments via the AWDC criterion (Michalik *et al.*, 2010) and the design criteria according to CDC (Hunter *et al.*, 1965). Thus, both objective criteria approaches were conducted based on of the first MBDoE run results to improve the model discrimination with the multi-start heuristic. The results of the extended MBDoE are shown in Figure 32 for the CDC and in Figure 33 for the AWDC.

A $K_{m,NAD}$ of approximately 1 mM for the MM model in the second closed-loop run suggested a maximum NAD^+ concentration of at least 10 mM (Bisswanger, 2017). Formate did not exhibit a high influence on the reaction rate, but is applied in industry in high concentration, because it is usually applied until complete consumption and does not display a significant effect on the FDH performance. Thus, according to Schmidt *et al.* (2009) a maximum concentration was set to 1500 mM. The minimum concentration of formate was initially reduced to 10 mM for the CDC. Preliminary experimental results indicated a positive effect with a further reduction in formate concentration. Therefore, the minimum concentration of formate was set at 1 mM to expand the design space in the lower direction for the AWDC. For the determination of the product concentration at elevated inlet concentration of NAD^+ (>2 mM), the product stream was not measured online in the UV/Vis flow through cuvette, but offline in a 1 mm glass cuvette. Therefore, the product stream was collected in reaction tubes to be diluted with 50 mM KP_i buffer (pH 8) and transferred with a pipette into the cuvette. The off-line measurements allowed dilution of the samples, which in turn increased the design space considered.

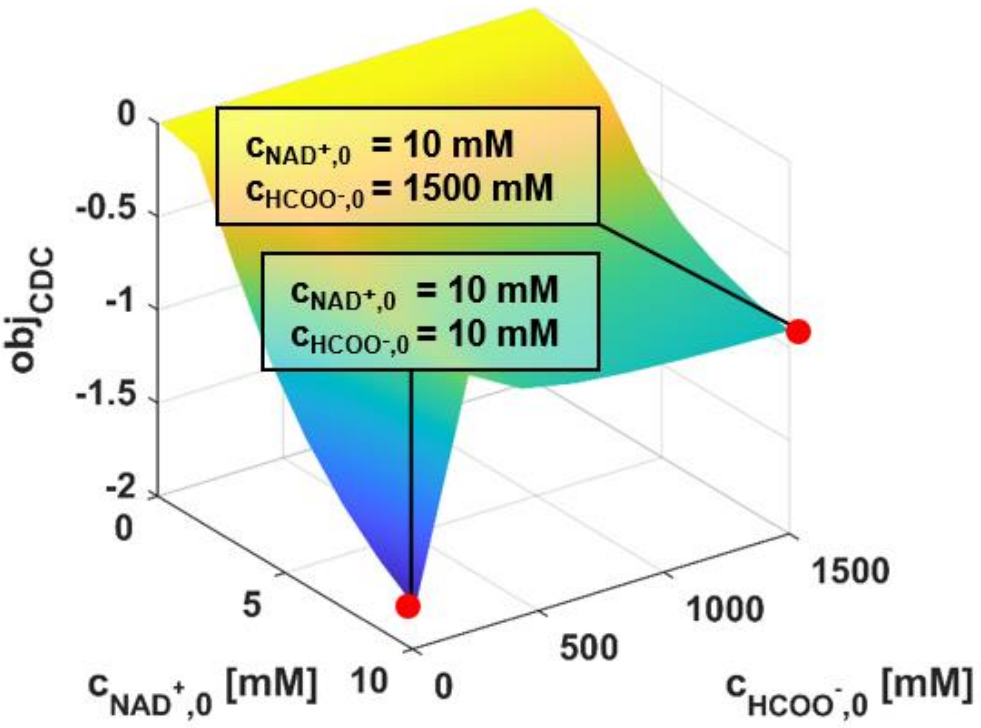

Figure 32: Surface plot for the objective function of the classical design criterion over the extended experimental design space within the concentration of 1 mM to 10 mM NAD^+ and 10 mM to 1500 mM formate.

For the CDC, a NAD^+ concentration of 10 mM has the highest influence on the model discrimination. They show the lowest results for the objective function (-1 to -2) which exhibit the highest influence on the model discrimination. This statement is logical, because v_{max} was not reached in the original design space. Additionally, inhibition effects are more severe in these concentration regions. When the formate concentration is included, low concentrations near the $K_{m,formate}$ value in combination with maximum NAD^+ concentration deliver the highest progress for the model discrimination with a value of -2. Noticeable is a peak at 300 mM formate which was the original maximum concentration of the design space which exhibits the lowest content for the model discrimination. The shape of the design criterion for the AWDC approach follows a similar trend towards NAD^+ concentrations of 10 mM and high formate concentrations in Figure 33. In contrast to the CDC approach, the highest function value is not obtained at a single spot with the highest NAD^+ concentration at low formate concentration, but at

an elongated front from 2.5 mM NAD^+ to 10 mM NAD^+. At this front of the AWDC, all experiments deliver the same information content for the model discrimination.

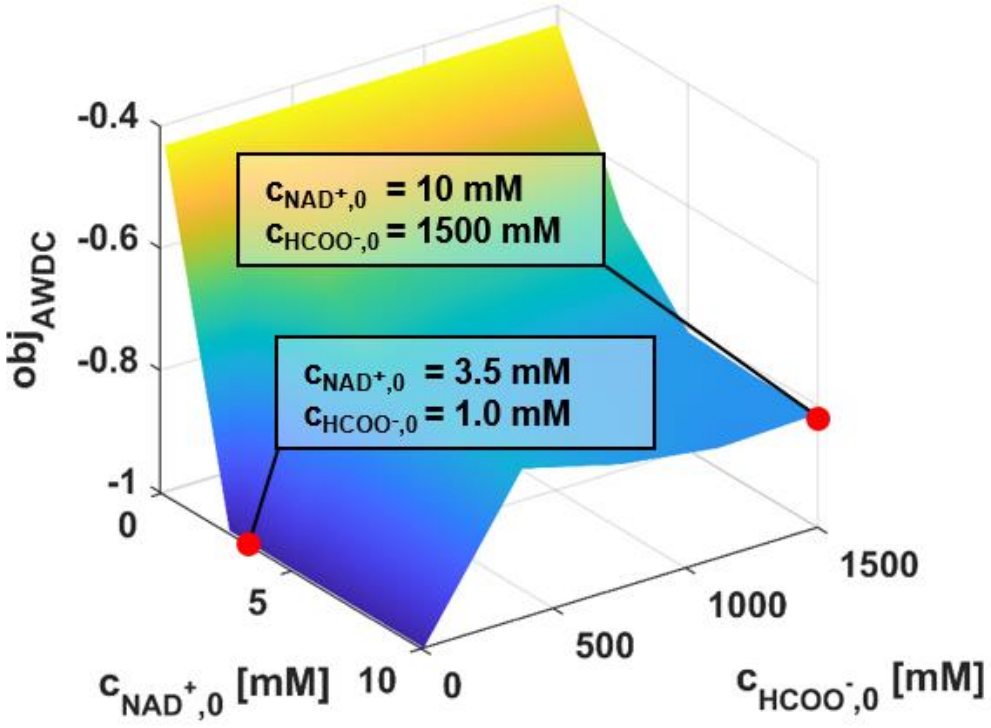

Figure 33: Surface plot for the objective function of the Akaike weights design criterion over the extended experimental design space within the concentration of 1 mM to 10 mM NAD^+ and 10 mM to 1500 mM formate.

According to the objective criteria results, three optimal experiments were chosen and carried out to increase the model discrimination. With the multi-start criterion applied on the data set of the first MBDoE and two extra experiments, the model identification was performed (Figure 34). The experiment at lower NAD^+ concentration (3.5 mM and 1 mM) was not included because the sensitivity analysis clearly showed higher trends with higher NAD^+ concentration due to the direct correlation of k_{cat} and $K_{m,NAD}$ to the initial substrate concentration. Furthermore, the inclusion of an experiment with a low formate and a high NAD^+ concentration was shown in the sensitivity analysis of the full factorial design to give the highest information content for the $K_{m,NAD}$. This fact was highlighted by the CDC, but the AWDC also attributed a high information content to lower NAD^+ concentrations.

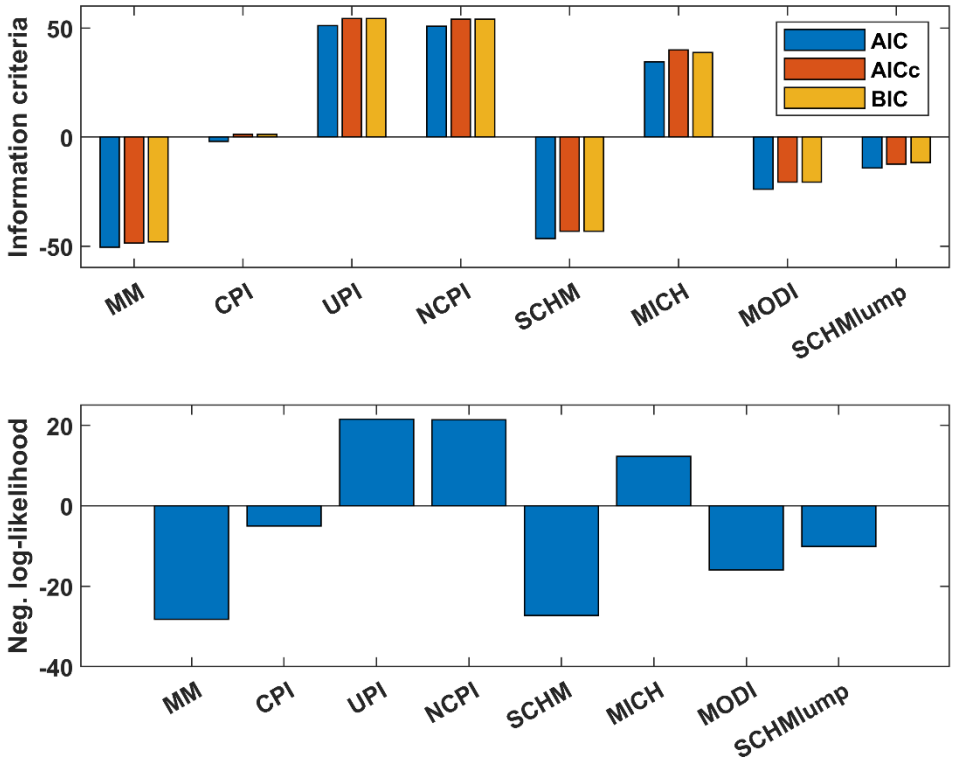

Figure 34: Model discrimination after the addition of two experiments according to the CDC. The data contained the steady state concentrations of the first MBDoE experiment.

When adding the experiments sharing the same information content according to the CDC and AWDC (10 mM NAD^+ with 10 and 1500 mM formate), the model identification results in the MM model (M1*) with a weighted AIC of 0.87 with the parameters (± 95% confidence intervals) of k_{cat} = 1.21 s^{-1} ± 6.6%, $K_{m,NAD}$ = 4.39 mM ± 13.5% and $K_{m,formate}$ = 31.04 mM ± 8.7%. The model discrimination clearly increases with the additional experiments (cf. Figure 34) and improves the prediction accuracy of the models which were already performing well in the original design space. Namely the MM, SCHM, MODI and SCHMlump model still better perform than the other candidates. In the extended design space, the SCHM model is currently superior to the SCHMlump model. The MODI model also fits better compared to the results before the design space extension. In the end, however, the simple double-substrate model MM still produces the best results for the available data. This shows that the influence of product inhibition in the PBR cannot be unambiguously represented by one of the chosen models.

Nevertheless, the good performance of the SCHM suggests that substrate inhibition occurs, which in the MM model can only be explained by apparently high parameters.

When the maximum activity was recalculated and employed for the calculation of the activity yield, it resulted in 66.3%, in contrast to 7.3% in the standard assay (1 mM NAD^+ and 200 mM formate). This corresponds to an active enzyme concentration increase from 0.082 mM to 0.742 mM active FDH, which is calculated based on the immobilization yield of 52.2% and a resulting enzyme loading of 0.03 $mg_{Enzyme} \cdot mg_{Carrier}^{-1}$. Comparing kinetic parameters of the free enzyme with literature values for $K_{m,NAD}$ of 0.04 mM and $K_{m,formate}$ of 0.47 mM (Schmidt *et al.*, 2009), the applied substrate concentrations are adequately chosen for the activity assay. However, with changing kinetics by immobilization, particularly in the K_m values, the maximum reaction rate was not reached within the conventional assay conditions. Accordingly, the $K_{m,NAD}$ is drastically overestimated by the application of a high active enzyme concentration (0.742 mM active FDH). Another possibility is that the assay was diffusion limited, although the particles were vortexed for a minute to ensure sufficient mixing inside the assay media before measuring the absorbance of NADH at 340 nm. A clear mechanistic model identification was not possible with this underestimated enzyme concentration and limited design space. Nevertheless, to show the model prediction performance of M1* model, the kinetic model of the extended design space model M1* was validated with the data set of the four level full factorial design experiments (cf. Table 6). Therefore, the concentration of NADH is simulated based on the applied inlet concentrations of NAD^+ and formate (Figure 35). The simulation demonstrates a strong correlation with the validation data set, achieving an R^2 value of 0.9944. It is evident that only the data points at an NAD^+ concentration of 1.9 mM were slightly overestimated. These points, however, have the same slope as the unity line, indicating a small systematic error in the experiments.

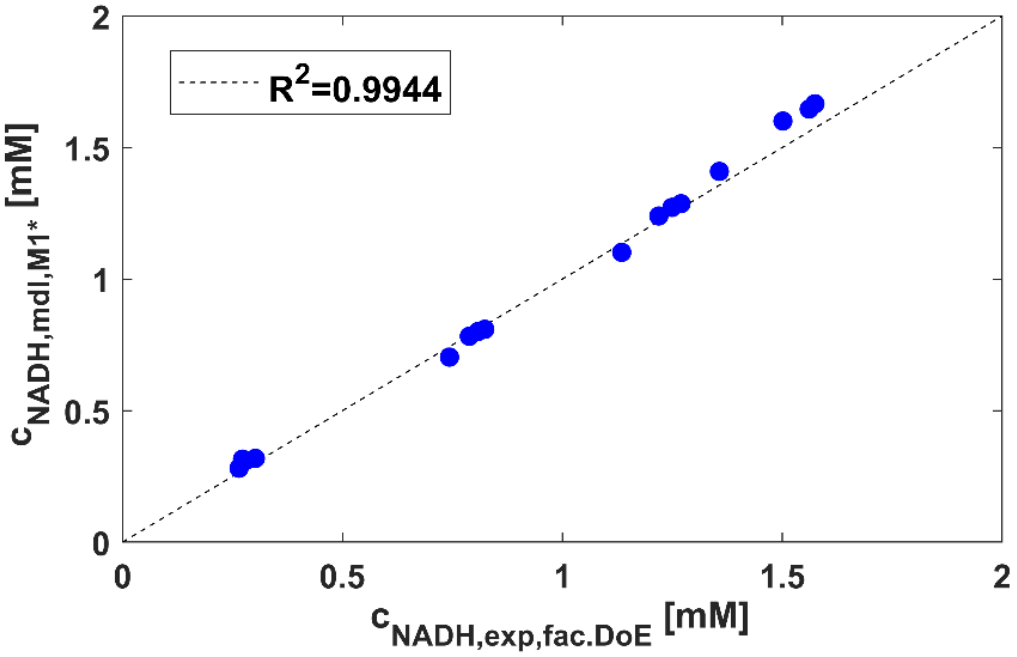

Figure 35: Model validation of the extended design space model M1* with the data set of the four level full factorial design. The substrate concentration of NAD^+ ranges from 0.35 mM to 1.9 mM and the formate concentration ranges from 50 mM to 290 mM.

Although the validation shows high model prediction accuracy, it is clear that the maximum reaction rate was significantly underestimated because the design space was too limited. Even with the extended design space, the maximum reaction rate was not reached, but it was represented much more accurately than before (cf. Figure 36). The MBDoE serves as a statistical experimental planning tool and is designed to optimize the experimental procedure for maximum information content to discriminate between models. However, its functionality is limited as it cannot assess mechanistic correctness. Similarly, when utilizing the AIC, which prioritizes parameter reduction if there is no significant improvement in accuracy, mechanistic correctness is not evaluated. However, the MBDoE technique has revealed that accurate kinetic models without a mechanistic basis can arise despite errors in experimental design and a limited range of concentrations utilized. This phenomenon is not observed when analyzing the enzyme's kinetics with less time or effort, especially when the enzyme is applied under different conditions than those of the assay (e.g. utilizing kinetic data from soluble enzyme data to model a PBR with immobilized enzyme). Especially beyond the assay concentrations, the predictive accuracy is somewhat limited, as evident when the design space is artificially expanded.

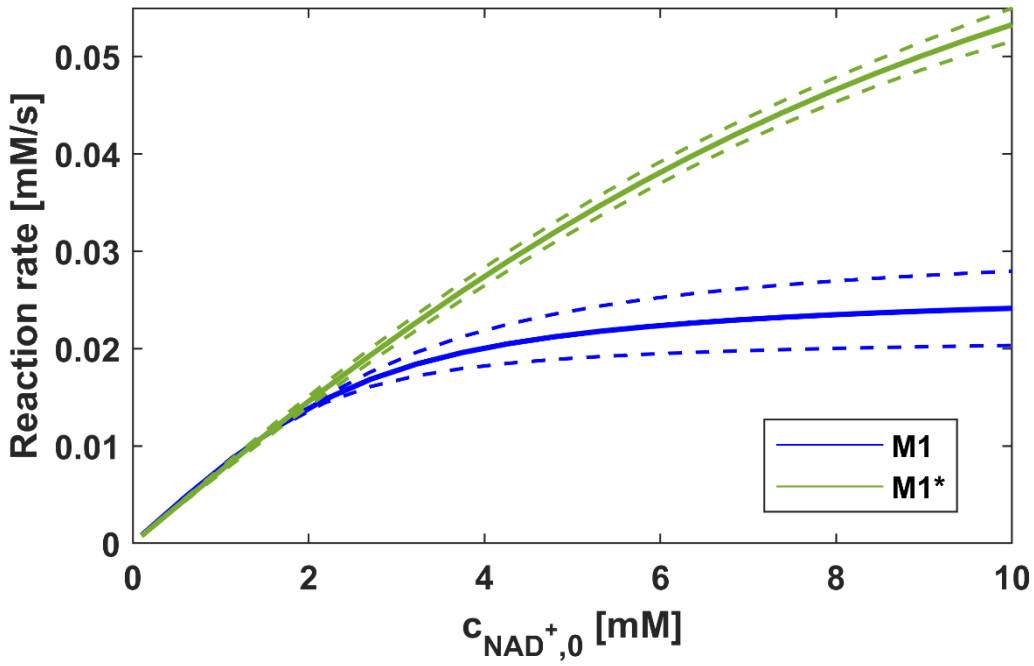

Figure 36: Simulation of reaction rates over the initial NAD^+ concentration for a constant formate concentration of 300 mM. M1 represents the model after closed-loop experiments and M1* represents the corrected model after the extension of the design space. The dashed lines represent the 95%-confidence intervals for the simulations.

In Figure 36 both, the original closed-loop derived kinetic model (M1) and the advanced model derived by the extended design space (M1*) are compared with the most influential input parameter, the NAD^+ concentration. It can be observed, that both models are able to represent the reaction rate in the original design space up to 2 mM NAD^+ but M1 displays a much lower extrapolated maximum reaction rate. With the estimated low $K_{m,NAD}$ and k_{cat}, M1 seems to reach a plateau in the higher range of NAD^+ concentration (8-10 mM) but with an increasingly high confidence interval. M1* does not reach the maximum reaction rate but has a much lower confidence interval in that region. In order to determine the correct k_{cat}, experiments have to be carried out at even higher NAD^+ concentrations or at shorter residence times. However, the curvature of the M1* model is supported by experimental data in contrast to the curvature of M1, which is solely extrapolated out of the original design space, which explains the high confidence intervals until 10 mM NAD^+. Therefore, the curvature of M1* is data based and can be extrapolated outside the considered design space until 40 mM NAD^+ plotted in Figure 37.

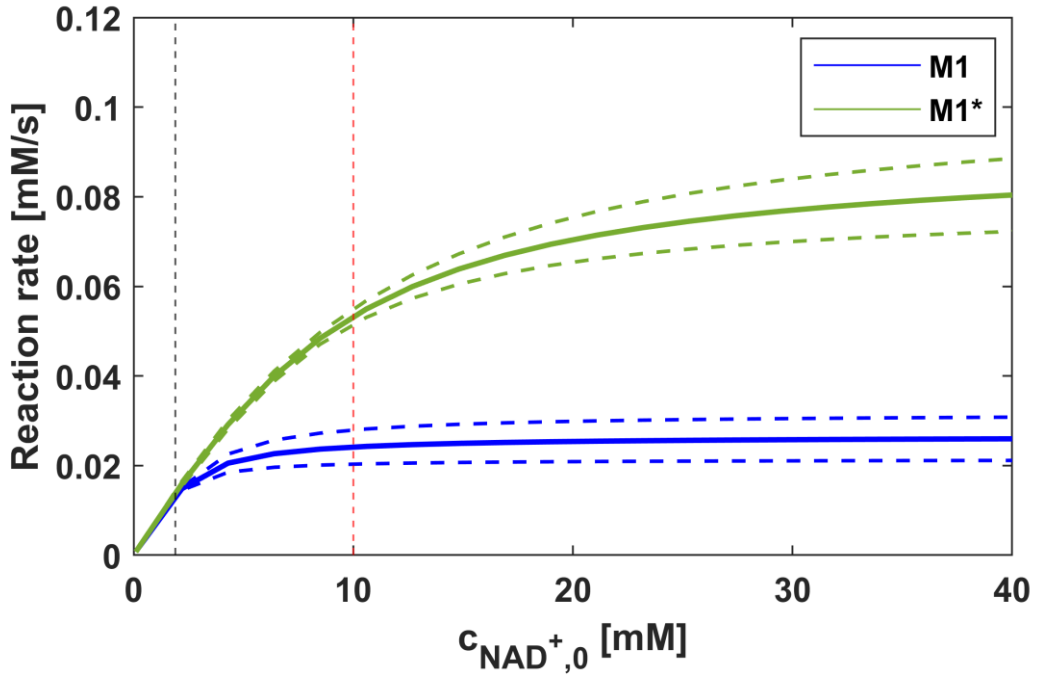

Figure 37: Simulation of reaction rates over the initial NAD^+ concentration for a constant formate concentration of 300 mM. M1 represents the model after closed-loop experiments and M1* represents the corrected model after the extension of the design space. The prediction was prolonged to 40 mM NAD^+. The original design space is indicated with a black dashed line and the extended design space is indicated with a red dashed line.

At 40 mM NAD^+ the increase of M1*in reaction rate is considerably low, indicating the approximation of the maximum reaction rate. At this substrate concentration the relative confidence intervals logically result in high absolute error (± 0.008 $mM \cdot s^{-1}$) in contrast to the observed design space (± 0.002 $mM \cdot s^{-1}$). This error can be minimized by adding more data points in the range of 10 to 40 mM NAD^+, but this concentration would be outside of a feasible range for the regeneration of NADH, which is a high cost factor in industry. This is already true for 10 mM NAD^+, but this concentration was necessary to compensate the high enzyme concentration in the PBR. At higher substrate concentrations, the immobilizate could reach high reaction rates, allowing for an accurate estimation of the maximum reaction rate. Another option to achieve the same result is to decrease the retention time, reducing the contact time between the enzyme and substrate. Therefore, higher reaction rates can also be achieved with the applied immobilizate within the design space. Unfortunately, this was not possible in the reactor setup due to the backpressure in the system. With the knowledge that the results produced, the use of a shorter reactor

would have been useful. This would have allowed for the use of the same enzyme preparation while achieving a lower enzyme concentration. Additionally, a smaller fixed bed volume would have reduced back pressure, enabling higher flow rates. However, a larger reactor volume was deliberately chosen initially to ensure a larger reactor volume in relation to the periphery. To prevent errors in the pumps from being minimized, it was important to consider this factor, especially for transient experiments with a constantly increasing and accurate volume flow. The optimization of the reactor system was ultimately limited by the analytics used. It is recommended to use a concentration range of 0.1 to 10 times K_m (Bisswanger, 2017), resulting in a factor of 100 between the lowest and highest concentration. However, the applied UV/Vis analysis has a measuring range with a limit of quantification of 0.143 mM up to a maximum concentration of 2 mM, covering only an approximately 10-fold increase in concentration. Thus, the analytical device limits the design space. To expand the measuring range, two spectrometers with different ranges or a completely different type analytical method, such as NMR, would be necessary. However, dilutions of 1:100 would be subject to error, so setting the reaction rate via residence time is the preferred control variable.

In conclusion, the study demonstrates that MBDoE is a valuable tool for conducting optimal experiments and estimating kinetic parameters. This is a common scenario in research, and guided experimental campaigns can aid in determining the actual kinetics, providing approximations of true values. As a result, this helps designing systematic experiments that produce relevant kinetic models for the studied biocatalyst and process. Particularly when kinetics may alter at various substrate levels, as explained by Kato *et al.* (1979) for the FDH, it is recommended that the experimental assessment be carried out under suitable process conditions. Thus, the kinetics can be identified directly within the chosen carrier matrix in the process design space and reactor setup of choice. This process is time-efficient due to automation and can potentially be extended to perform versatile screening of stability, diffusion, or residence time distributions. The approach presented here is entirely driven by experimental results and is not influenced by experimental procedures or human errors. However the gained efficiency by automated experiments requires investments in hardware, software, and knowledge of the possible model candidates.

5.3. Interim Summary

The post-processing of the experimental data in Chapter 4 revealed some interesting insights into the capabilities and characteristics of the MBDoE method. Additionally, this approach highlighted the limitations of the system.

- The employment of a multi-start heuristic yielded the MM model as the best fitting model for the first MBDoE campaign's dataset, which originally produced the SCHMlump model.
- The multi-start heuristic illustrates the challenge in distinguishing among the model candidates based on the NAD^+ and formate concentrations within the given design space, as nearly every model fits the data accurately.
- By broadening the design space and supplementing with manual experiments, the model discrimination increased to the AIC weight of 0.87, taking into account the new optimal experiments. These two experiments adhered to both the AWDC and CDC design criteria.
- The expanded design range resulted in the identification of an activity yield that had been previously underestimated and was subsequently adjusted to a ten-fold value of 66.3%.
- The utilization of fixed starting values during model identification can be impacted by the optimization algorithm. However, this is not the case when utilizing the multi-start heuristic approach.
- For more complex reactions, it is necessary to examine the optimization algorithm in terms of computation time. Parallelizing calculations and experiments is an advisable method to consider in future endeavors.

6. Evaluation Transfer to Continuous Stirred Tank Reactor

To evaluate the model discrimination and parameter estimation method in a different reactor type and with soluble enzymes, the evaluation method is applied to steady-state data from a CSTR in the form of an enzyme membrane reactor (EMR). It is a common reactor type in biocatalysis, making it relevant for the determination of enzyme kinetics under process conditions. Furthermore, this reactor concept offers complementary process characteristics compared to a PBR. An EMR, for example, has minimal diffusion limitations and does not require immobilization on solid surfaces. Therefore, an EMR may be used as an alternative option for continuous biocatalysis when the process requirements do not fit a PBR. In case of a heavily substrate inhibited enzyme, an EMR is preferred over a PBR because, at steady-state conditions, there are constantly low substrate concentrations in an EMR. If no inhibition occurs, both reactor types may be evaluated, and other factors will influence the selection of the biocatalytic reactor configuration. Requirements that have a significant impact on the reactor choice may include selectivity, the need for a secondary phase, or precise regulation of temperature or pH.

For the purpose of this study, the EMR provides certain benefits in terms of the conceptual demands of the analyzed biocatalyst. Specifically, employing a soluble enzyme retained by a membrane is advantageous for exploring new enzyme variants. Since they are utilized in solution, there is no need to generate an active enzyme immobilizate to evaluate new biocatalysts for biocatalytic reactions. In addition, potential diffusion limitations or structural changes have minimal impact on enzyme kinetics, resulting in authentic enzymatic kinetic data. Additionally, cell-free extract can be applied for preliminary tests of new enzymes to identify performance markers and estimate kinetic parameters. The continuous operation of the EMR and the PBR makes them ideal for automation, which is essential for implementing the MBDoE approach presented in this study. In this work, however, a semi-automated approach was implemented due to the unavailability of sufficiently large dosing equipment for the higher volume of the enzyme membrane reactor. Nonetheless, the system's lower back pressure and micro gear pump allowed for additional control over the retention time, aiding in the identification of the kinetic model. As a result, the experimental design did not utilize the MBDoE approach. The experiments were designed based on understanding of the system and adjusted manually

through visual examination of the resulting model parameters, parameter sensitivities and errors.

6.1. Semi-automated Setup

The semi-automated setup consists of an enzyme membrane reactor (EMR) (Forschungszentrum Jülich GmbH, Jülich, Germany) with a total empty reactor volume of 10 mL (Wichmann *et al.*, 1981). The reactor was assembled according to Figure 38 and consisted of a bottom part with an inlet. This inlet directs the substrate solution directly into the center opening of the reactor at the bottom. Holes for the thermostating water flow are recessed in the lower part. Thus, the incoming substrate solution is instantly dispersed by the magnetic stirrer on top of the opening. The magnetic stirrer is placed on a PTFE socket to circumvent high friction forces along the whole stirrer area. Above that, the membrane is placed between two rubber seal rings supported by a PTFE disc. The ultrafiltration membrane consists of regenerated cellulose with a cut-off of 10 kDa (Merck KGaA, Darmstadt, Germany) and with a diameter of 63.5 mm. The supporter plate is perforated to enable an equally spread flow through of the output stream through the whole membrane and a stable position of the membrane without bending, which could damage the membrane and result in enzyme leakage. The two seal rings which fixate the supporter plate and the membrane are housed between the bottom and the lid plate made of stainless steel. The whole construct is assembled by four screws.

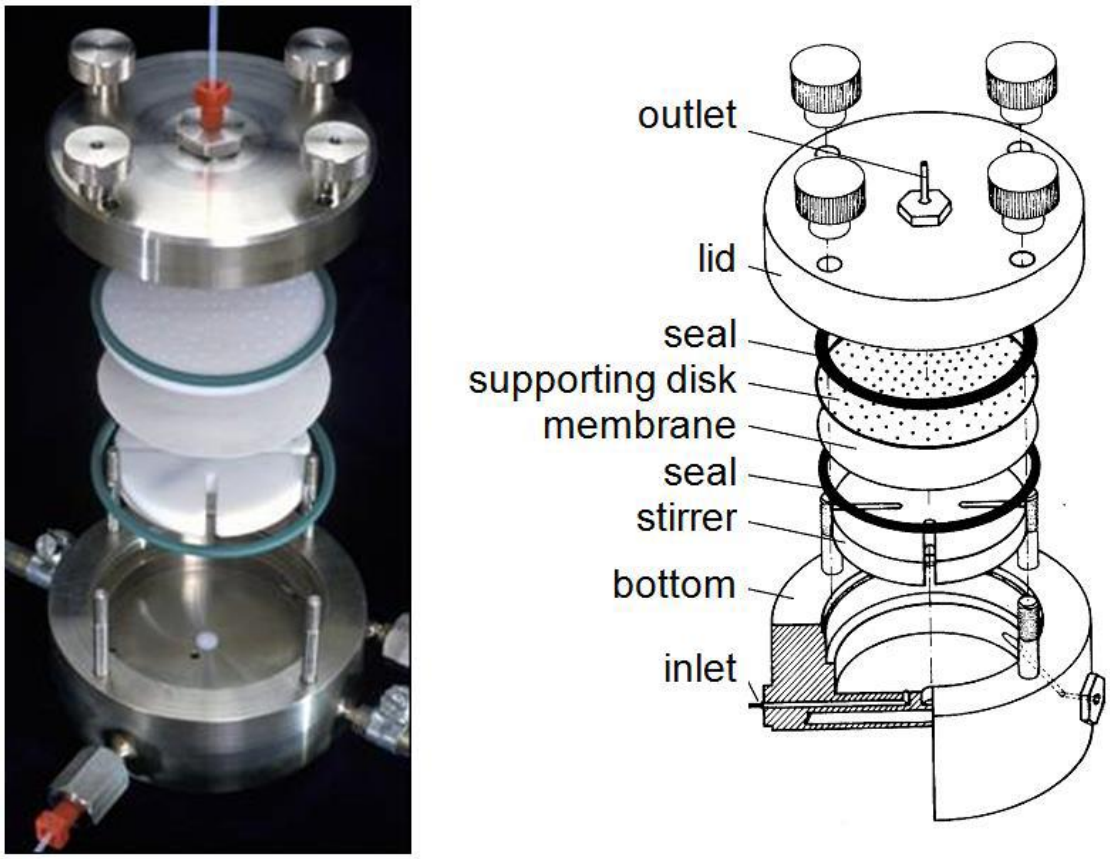

Figure 38: Setup of the enzyme membrane reactor. Left: photographic picture of the components, right: exploded-view drawing (EDS Maschinenbau GmbH, Germany, Linnich) (Wichmann *et al.*, 1981).

The complete setup, which was applied for the steady state experiments is shown in Figure 39. The substrate solution is pre-mixed with a fixed concentration of NAD^+ and formate and cooled for the duration of the experiment. For the supply of the substrate a micro-gear pump (mzr-2905-1082, HNP-Mikrosysteme GmbH, Schwerin, Germany) was utilized with 5-10 µm PTFE suction filter (Techlab, Braunschweig, Germany) to prevent dust or other particles from entering the pump. Between the pump and the reactor, a bubble trap made of polymethylmethacrylat (PMMA) was applied to prevent entrapment of air bubbles in the reactor. This would lead to a decrease in reaction volume and in consequence lowers the hydrodynamic residence time, which leads to errors in the determination of reaction rates. The outlet of the reactor was connected to a flow-through UV/Vis spectrophotometer (HR4000, Ocean Insight, Orlando, USA) and a cuvette (170.700 QS, Hellma GmbH &Co. KG, Müllheim im Marktgräferland, Germany) with a light path of 1 mm coupled with a UV light source (DH-2000-BAL, Ocean Insight,

Orlando, USA). The NADH concentration was determined via the absorbance at a wavelength of 340 nm. Afterwards, the flow was measured with a liquid flow meter (SLF3S-0600F, Sensirion AG, Stäfa, Switzerland) to observe failures in the experimental procedure, for example dropouts of the stirrer or changed pump efficiency.

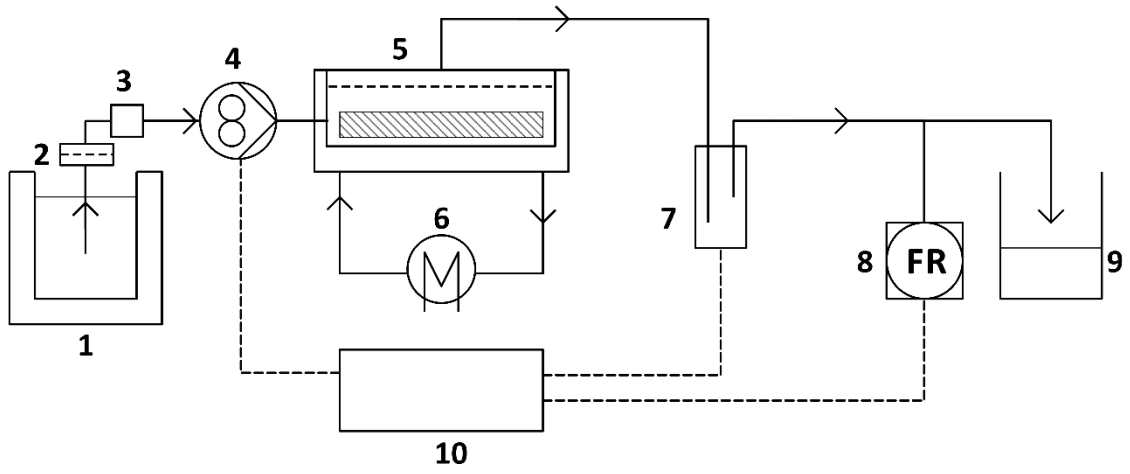

Figure 39: EMR Setup for the determination of steady state NADH concentrations with FDH as biocatalyst and formate and NAD^+ as substrates in 50 mM KP_i buffer at pH 8. (1) Cooled substrate reservoir at 4°C (2) micro gear pump (3) bubble trap (4) enzyme membrane reactor with magnetic stirring bar (5) thermostat (6) flow through UV/Vis spectrometer (7) flow meter (8) product container (9) measurement and control computer.

6.2. Experimental Workflow

The EMR was applied according to the shown specifications (Figure 38). Prior to conducting any experiments, the reactor setup was assembled by connecting the PTFE suction filter with the first PTFE tube between the substrate reservoir and the micro-gear pump. This pump is connected to a personal computer and controlled by Faulhaber Motion Manager 6 (FAULHABER Drive Systems, Schönaich, Germany), which is the manufacturer's software. Between the pump and the reactor, the bubble trap was connected to a tightly screwed silicon septum that served as the enzyme solution inlet.

The enzyme membrane reactor setup was applied according to the flow scheme in Figure 39, but without the membrane and support plate. The UV/Vis spectrophotometer and flow meter were connected to the reactor's outflow through PTFE tubes. The entire system was flushed with 50 mM KP_i buffer pH 8 to remove any air. When the system was flushed with at least five reactor volumes (50 mL), the pump was switched off. Subsequently, the

reactor was reopened and filled with additional buffer to obtain the maximum liquid volume. The membrane was then inserted with its active asymmetric side facing towards the reactor. The membrane was stored the day before in 50 mM KP_i buffer pH 8 to pre-swell overnight. Through the use of excess buffer, the membrane was submerged to fit onto the first seal ring without any bending. The membrane was fixated by placing the supporting plate on top of it. The support plate was rinsed once more with buffer to remove any air from the perforation holes. The reactor was then sealed with the lid and reconnected to the periphery. After setting up the system, it was tested for leaks by flushing it was 50 mM KP_i buffer pH 8.

To prepare for the experiment, 20 mg of bovine serum albumin (BSA) (Sigma)solubilized in 1 mL 50 mM KP_i buffer pH 8 was added with a syringe through the septum of the bubble trap at a low flow rate of 0.2 $mL \cdot min^{-1}$ to prevent adsorption of FDH to the membrane. The flow rate was increased to 5 mL/min, transferring the BSA into the reactor to adsorb for 10 min on the membrane. Afterwards, the flow rate was decreased to 0.2 $mL \cdot min^{-1}$ to inject the FDH solution, which had a concentration of 1 $mg \cdot mL^{-1}$, achieving a final concentration of 0.1 $mg \cdot mL^{-1}$ (0.012 mM) of FDH in the 10 mL reactor. Consequently, the flow rate was once again increased to match the first applied residence time. Another waiting period of 10 min was implemented to ensure uniform distribution of the FDH throughout the reactor volume.

The experimental campaign was pre-set in an Excel spreadsheet (cf. Table 2), which outlined the residence times for each run and the number of residence times conducted prior to measuring the steady state concentration of NADH. Upon evaluation of the initial experiments, it was determined that the system required seven residence times to achieve the set flow rate, steady state conditions and to remove the product from the previous run. The residence times were assigned based on individual pump velocities, which were automatically adjusted according to the previously calibrated flow rate to pump velocity relationship. Due to the setup, the system only allowed the introduction of one set of substrate concentrations at a time, referred to as a semi-automated setup. Although the MBDoE concept was not fully established at that time, the experimental data was utilized to generate optimal experiments. The fitting procedure was not optimized and lacked the multi-start heuristic and literature models in the beginning. Nevertheless, the results of

the final model identification were executed with the updated script to ensure full functionality and comparability.

6.3. Model Identification

Unlike the PBR system, no dosing system was implemented, and instead, a pre-mixed substrate solution was utilized for each experiment. Thus, the reaction rate was predominantly regulated by varying the residence time. The multi-start heuristic was employed to identify the model in a data set of 36 experimental points, with NAD^+ concentrations ranging from 1.08 mM to 2.16 mM and formate concentrations ranging from 10.01 mM to 275.5 mM. The residence time was also varied between 3.5 and 30 min. Based on the previous results, the optimization algorithm for the least-squares curve fit was configured using TRR with the central finite difference method. The starting values underwent 200 iterations within the ranges of chapter 4.2.1. The model differentiation led to the acquisition of Akaike values for eight possible kinetic model candidates, as displayed in Figure 40. The best models, MM, NCPI, SCHM and SCHMlump, have AIC values of -65.6, -90.3, -69.3 and -71.8, respectively.

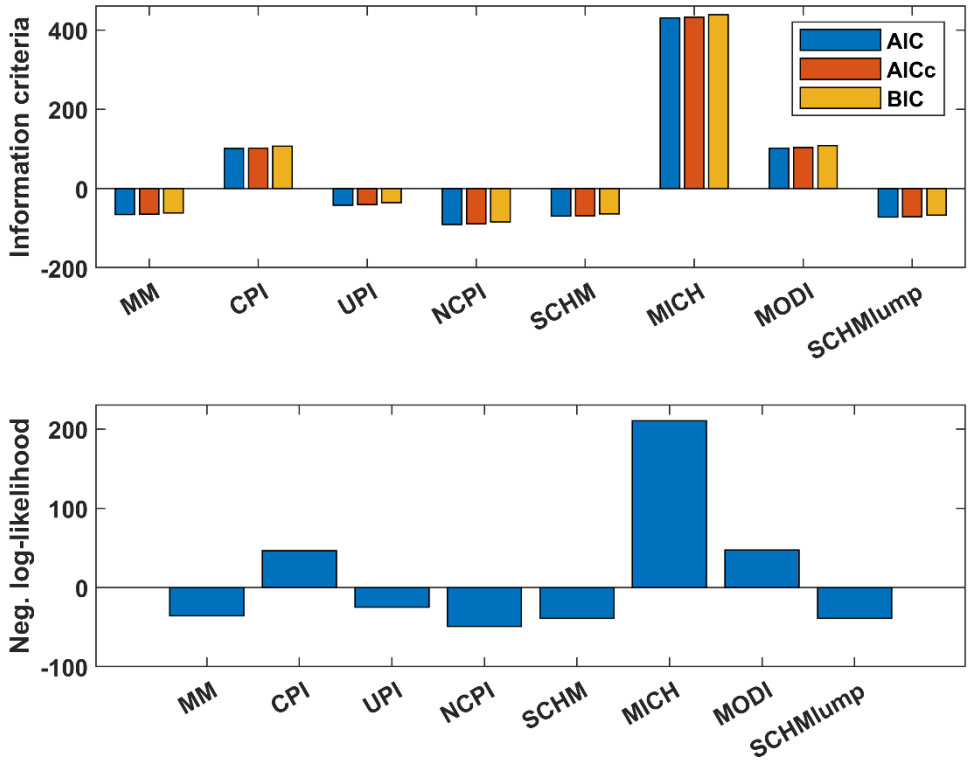

Figure 40: Information criteria and negative log-likelihood as an evaluation of the goodness of fit for candidate models of FDH kinetics based on 36 experiments in the EMR setup with NAD^+ (1.08-2.16 mM), formate (10-275.5 mM) and residence time (5-30 min) varied. The data is fitted with the multi-start heuristic including 200 iterations. Higher absolute values indicate better fits.

The three superior models of chapter 5 are still present in these results (MM, SCHM, SCHMlump), but the NCPI offers the highest AIC weight of 0.99 within the design space with kinetic parameters of k_{cat} = 2.1 s^{-1} ± 40.4%, $K_{m,NAD}$ = 0.44 mM± 27.7%, $K_{m,formate}$ = 52.5 mM ± 9.1% and $K_{i,NADH}$ = 0.3 mM ± 49.7%. This accounts for the highest degree of model discrimination among all post-processed data in this thesis. All comparable parameters have reasonable confidence intervals and are in the range of literature values, but there are no comparable results for a non-competitive inhibition constant in literature. Interestingly, literature generally only refers to the competitive product inhibition (Cleland, 1963; Kato *et al.*, 1979; Kula *et al.*, 1980) or refer to data-driven mechanistic models with substrate inhibition towards NAD^+ (Michalik *et al.*, 2007; Schmidt *et al.*, 2009). The presented model shows high accuracy for the prediction of the underlying data set obtain in the steady state mode of the CSTR setup in Figure 41.

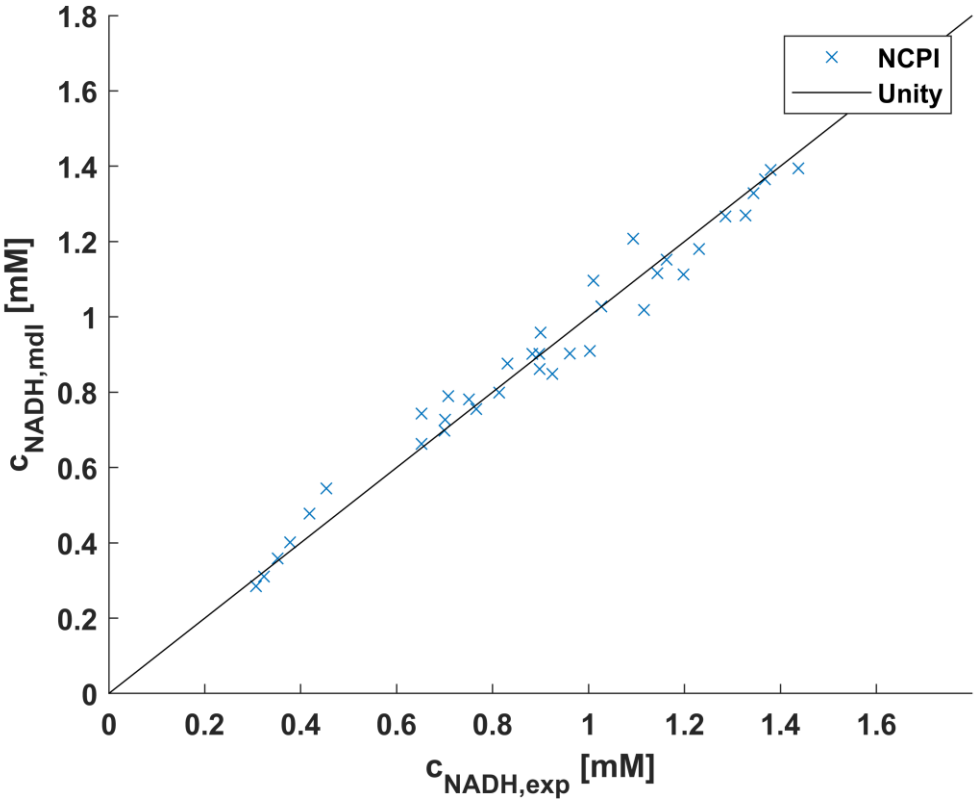

Figure 41: Parity plot for the best fitting model in the CSTR setup: NCPI. Modelled NADH concentration is plotted over experimental NADH concentration.

Nonetheless, a non-competitive inhibition by NADH in respect to NAD^+ is rarely reported in literature. The non-competitive inhibition is only published at NAD^+ concentrations below 1 mM, which accounts for the saturation concentration of NAD^+ at assay conditions of 1.5 mM NAD^+ and 150 mM sodium formate in 50 mM KP_i buffer at pH 7.5 (Kato *et al.*, 1979). Detailed analysis of a metal-dependent FDH from *Rhodobacter capsulatus* also revealed the coexistence of competitive and non-competitive inhibition (Laun *et al.*, 2022). In consequence, both inhibition types are possible dependent on the saturation concentration of NAD^+ under the applied assay conditions. The results of the model identification indicate unsaturated conditions in the EMR with an FDH concentration of 0.1 $mg \cdot mL^{-1}$, so that the NCPI is the most valid model. An indication for a wrong model identification in respect to the NCPI was the approximation of the set parameter limits regarding $K_{m,formate}$ and $K_{i,NADH}$ values which were set to 52.5 mM and 0.3 mM respectively (cf. Table 3). But when these limits were increased the AIC weight for the NCPI model reached 1 with the following kinetic parameters k_{cat} = 5.5 s^{-1}± 19.7%,

$K_{m,NAD}$ = 0.59 mM ± 18.4%, $K_{m,formate}$ = 98.6 mM ± 14.8% and $K_{i,NADH}$ = 0.13 mM ± 25.1%. Especially k_{cat} and $K_{m,formate}$ drastically increase in this case and overextend the literature values. However, the confidence intervals of all parameters decrease. Only the confidence intervals for $K_{m,formate}$ increase with the new limit for $K_{m,formate}$ and $K_{i,NADH}$. The information criterion increases from -90.3 to -103.37, enlarging the difference to the SCHMlump model (AIC = -71.8). The SCHMlump model is also capable of representing the experimental data (cf. Figure 42).

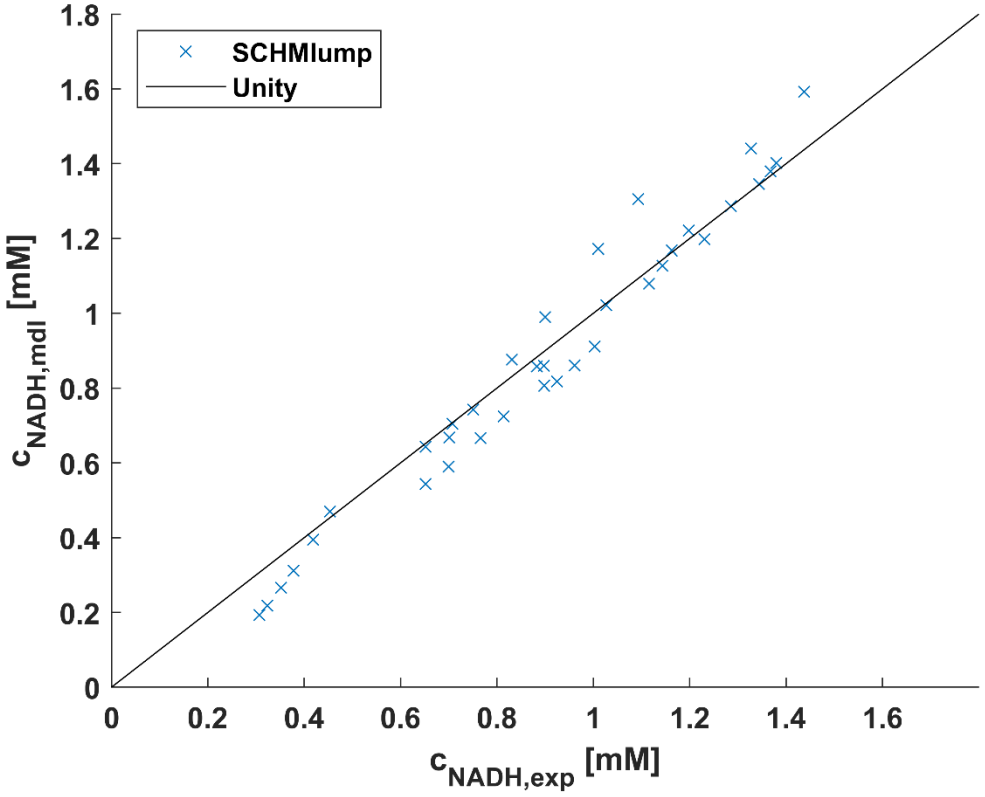

Figure 42: Parity plot for the second best fitting model in the CSTR setup: SCHMlump. Modelled NADH concentration is plotted over experimental NADH concentration.

The SCHMlump model consists of the kinetic parameters k_{cat} = 0.42 s^{-1} ± 31.2%, $K_{m,NAD}$ = 0.47 mM ± 78.4%, P_1 = 62.2 mM^2 ± 49.0% and show higher confidence intervals for the model fit. In contrast to the NCPI model, the SCHMlump model shows an underestimation in the lower NADH concentration range (<0.5 mM NADH) and an overestimation in the higher NAD^+ concentration range. Nevertheless, as shown by the information criteria, both models can represent the experimental data within the design

space at reasonable error levels. However, the empirical results clearly demonstrate that the NCPI model outperforms all other models in the EMR, indicating the substantial impact of inhibition in this reactor type and the accuracy of the model in this context.

As a coenzyme for cofactor regeneration, FDH is rarely exposed to high concentrations of NAD^+ due to the high cost of the cofactor, which is used in very low concentrations. Therefore, the use of NCPI models should be necessary more frequently. However, the exception of the NCPI model has been considered so rarely due to the use of assays performed absent from process conditions. Under these conditions, a low amount of enzyme and a wide range of substrate concentrations are typically used. It is important to note that the reaction rates obtained at assay conditions are evaluated using a single model. In this case, the CPI may outperform the NCPI under saturation conditions. Based on this, simple kinetic modelling with a limited number of model candidates for the enzymatic reaction cannot represent enzyme kinetics accurately. Hence, crystallography and spectroscopic methodologies are necessary to attain this level of specificity regarding an enzyme's inhibition type and catalytic mechanism. It is worth noting that in industrial production processes, modelling serves the sole purpose of productivity verification and optimization. If this objective is attained through simplified or potentially mechanistically incorrect models in the course of the procedure, it would be considered successful.

7. Overall Discussion and Outlook

The objective of this thesis was to develop an automated reactor platform that allows repeatable and efficient kinetic modelling of enzymes. This chapter presents and dissects the findings acquired. The overall research methodology and the developed reactor system are assessed, and their abilities and limitations are identified. The motivation and aims are analyzed from various perspectives, along with the corresponding results. Lastly, the gained knowledge is utilized to construct the future prospects and possibilities of the subject matter.

7.1. Automatization of Kinetic Modelling

This thesis demonstrates the possibility of automating kinetic modelling using the applied methods. To ensure experiment repeatability, a reactor platform was constructed initially. Syringe pumps and flow-through UV/Vis measurements were utilized, along with a measurement and control computer operated by a Python code. The precise dosing of the substrate solution resulted in different operation points in the steady state, which facilitated non-linear regression analysis of the experimental data. This method is considerably more time-consuming than acquiring a single data point in conventional kinetic assays, but the conditions are more representative of process conditions. It has been suggested that modeling results are more useful for process development than outcomes obtained from simplified laboratory settings. This assumption is supported by the fact that even manual decisions related to the experimental setup and design can impact the assay results. (Nikolova *et al.*, 2008; Yang, 2009; Stroberg *et al.*, 2016; Grosch *et al.*, 2017; van Schie *et al.*, 2021; Gygli, 2022). These factors also apply to the automated reactor setup, as stated in this thesis. However, it is important to note that the proximity to the process conditions implies a closer proximity to the true model results. This was evidenced by the distinct model identification outcomes for both operation modes (NCPI for EMR and MM for PBR). The identification of the NCPI model in the EMR revealed that the modeling can vary depending on the process conditions. Although this model has rarely been used in the literature, it provides the best results for the applied conditions in the EMR.

While this thesis yielded positive outcomes, it is important to acknowledge that all models fail to account for real factors associated with the desired close proximity to process conditions. Diffusion limitations, partitioning effects, temperature fluctuations, enzyme

deactivation, and contamination are factors that could improve the model's accuracy. However, incorporating these factors would increase the model's complexity, the experimental expense and computational effort. This is already a challenge for more complex reactions that require modeling of intermediate steps or by-products. Nonetheless, through the application and optimization of this approach, the model error is mainly impacted by the hardware employed (analytics, pump, and reactor). One major advantage is that experimental results are obtained without the use of reporter molecules, model substrates, high dilution factors, or extraction steps. The relative standard deviation of the measured NADH concentration for the presented setup ranges from 3.4% to 9.6% in the detection range. The minimal parameter errors are in the same range in the MM model of the post-processed extended design space data set with the parameters (± 95% confidence intervals) of k_{cat} = 1.21 s^{-1} ± 6.6%, $K_{m,NAD}$ = 4.39 mM ± 13.5% and $K_{m,formate}$ = 31.04 mM ± 8.7%. This results in a relative standard deviation from 3.3% to 6.8%. The finding is particularly emphasized when the entire data set is utilized to perform the model identification excluding the manual measurements of the extended design space. Here the parameters of the MM resulted in k_{cat} of 0.32 s^{-1} ± 10.1%, $K_{m,NAD}$ of 0.78 mM ± 19.2% and $K_{m,formate}$ of 16.6 mM ± 12.6%. The standard deviations of the parameters were 5.1%, 9.6% and 6.3%, respectively, which are within the analytical standard deviation. Thus, the study demonstrates that the automation of the experimental and analytical procedures is restricted by the systematic error of the reactor setup. This results in an analytical error of 3.6% at NADH concentrations of approximately 2 mM and 9.6% at NADH concentrations of approximately 0.35 mM. Additionally, the study identifies flow deviation as another source of error, which is consistently at 1%. These errors determine the accuracy of the model parameter estimation but are not directly related to the model's performance. In this thesis, the model's accuracy is assessed using the negative log-likelihood, which is refined using the AIC. This criterion penalizes the number of model parameters in addition to the model's accuracy (negative log-likelihood). While this method provides useful support for model discrimination, it does not provide an absolute measure of model prediction accuracy. Instead, it offers a relative measure of the set of candidate models available in the literature. During the development of the evaluation method, models were often predicted using parameter confidence intervals with three or more digits. This occurred because lower parameter accuracy can increase model precision, leading to overfitting of the data and high negative log-likelihoods. This

phenomenon is known as the 'Bias-Variance Tradeoff', which was limited by the introduction of parameter limits. These limitations and the corresponding model candidates were based on literature which should be plentiful for an industrial relevant enzyme like the FDH, but turned out to be rather limited (Kato *et al.*, 1979; Kula *et al.*, 1980; Michalik *et al.*, 2007; Schmidt *et al.*, 2009).

The presented approach has a weakness in that it requires a certain amount of prior knowledge. When applying this method to novel enzyme variants, diverse kinetic parameters must be available to adjust the limits in contrast to the wild type. As mutations are typically conducted to achieve specific process development goals, this information should be readily available. The reactor platform still depends on initial experiments with new enzyme variants or comprehensive data on wild-type variants to specify initial parameter estimates and design spaces. The same principle applies to the kinetic models for the prediction of experimental data of the reactor platform. To describe these kinetics, four textbook models (MM, CPI, NCPI, and UPI) were applied in this thesis. Additionally, four literature models (SCHM, SCHMlump, MODI, and MICH) were employed to broaden the spectrum of kinetic models with more specialized mechanistic models for the FDH from *Candida boidinii*.

The devised reactor platform and implemented evaluation method cannot generate new models independently, but rather select the most suitable model from a given set. To ensure correct model identification, the AIC was employed to account for model complexity. By applying the information criteria, the most appropriate model is determined. However, the outcomes suggest that the penalty on the number of parameters has no effect on the results as opposed to using the standard negative log-likelihood. The reason for this is that the analyzed models have between three and five parameters. The AIC penalizes a single parameter with a low AIC value of two, which is significantly lower than the absolute difference of the resulting AIC values in the magnitude of -20 to -80. In chapter 6.3 for example, the closest AIC values were calculated for the SCHM and SCHMlump model which account for -69.3 and -71.8, respectively. In this case, the SCHM model has four parameters and the SCHMlump model has three parameters. Comparing the log-likelihoods without the parameter penalty, the value for SCHM results in -77.3 and -77.8 for the SCHMlump model, respectively. This additionally applies to

the results of the extended design space in chapter 5.2, in which the narrow difference of two competing models is increased by the application of the AIC. Only in scenarios where model discrimination is ambiguous, as in chapter 5.1.1, the penalty changes the result. However, the conclusion of a model discrimination is not relevant in a parameter space where no unambiguous discrimination is feasible.

However, all applied models contain comparable parameters reliant on only two input variables, namely the NAD^+ concentration and formate concentration. If the model complexity is increased by introducing additional dependencies, such as temperature, pH value or diffusion limitation, the AIC provides a way to reduce parameters rationally and, therefore, to reduce the model complexity simultaneously. Still, this criterion fails to differentiate between thermodynamically or mechanistically correct and incorrect. Based on the AIC the Akaike weights are calculated which are the basis for the model discrimination approach. It could be shown, that the applied MBDoE is able to increase the model discrimination when the limited parameter design is increased to an Akaike weight of 0.87 (cf. chapter 5.2). However, this increase was only necessary because the active enzyme concentration was applied instead of the total enzyme concentration, which was therefore underestimated (7.3% activity yield in contrast to 66.3%).The extension of the design space was conducted *in silico* with higher inlet concentrations than possible in the automated setup leading to satisfactory model discrimination and parameter estimation with the *e-optimal* design of experiments. The results are supported by the practical suggestions by Ohs *et al.* (2017), who recommends *e-optimal* criteria, a large as possible design space and few optimal experiments of large non-optimal data sets.

In chapter 5.2 it was shown, that the experimental design is targeting experiments with the highest sensitivity for the model parameters for both, the CDC (Hunter *et al.*, 1965; Atkinson *et al.*, 1975; Buzzi-Ferraris *et al.*, 1990) and the AWDC (Michalik *et al.*, 2010), resulting in a better model discrimination. In contrast to Michalik *et al* (2010), the superiority of the AWDC was not observed in this work. In fact, the opposite is the case, so that the CDC is directly pointing at the most optimal experiment at 10 mM NAD^+ and 10 mM formate, whereas the AWDC shows a front of optimal experiments in the range of 3.5 to 10 mM NAD^+ (cf. Figure 32). Buzzi-Ferraris comments for this work (Buzzi-Ferraris, 2010) demanded dismissing of illogical models beforehand and referred to his

updated kinetic model analysis (Buzzi-Ferraris *et al.*, 2009). The improvements included some additional functionality like outlier detection which can be considered to be included in the next iteration of this automated reactor platform for kinetic model identification. Nevertheless, with the application of the *e-optimal* design and the AWDC criterion the objective of the reactor platform was fulfilled. However, there is still room for additional functionalities, such as correlation studies via eigenvalues, which are already calculated for the MBDoE. Additionally, model candidate reduction, outlier detection, and other features could be included to find the true model. It is important to note that each functionality requires higher computational power and a smart sequence of events in the experimental execution and parallel data evaluation.

The scope of kinetic modeling is vast and can encompass numerous influencing factors that cannot be tested within the presented platform. Mechanistic analyses can be better performed with molecular docking simulations based on crystallographic data and spectroscopic analysis. Furthermore, incremental mechanistic modeling combined with an MBDoE approach could lead to mechanistically correct models as a basis for experimental model identification. For FDH, the NCPI model was successfully identified under the EMR conditions. However, for PFRs, only the very simplified MM model was identified, despite the availability of more complex literature models that were validated by experiments. This highlights a critical point in kinetic modelling under process conditions and raises the question of how complex and accurate the kinetic model needs to be in order to predict the process satisfactorily. The presented method achieves precise results even when using simple models like the MM model for analysis, by employing the entire process window as the design space. Additionally, using the largest feasible design space for process conditions enables interpolation within appropriate conditions, rather than extrapolation of assay results, increasing the likelihood of identifying a reliable model for the application objective. However, it is currently uncertain whether scalability can be achieved for more complex systems. The number of required experiments increases as the number of undetermined parameters grows. Furthermore, establishing a complete model basis for optimizations requires more control over additional parameters such as temperature, pressure, or buffer type and capacity. As process complexity increases, the necessary sensors, analytics, and calculations also increase. However, a higher magnitude of complexity provides the opportunity to

effectively utilize information criteria such as the AIC. This is because the penalty term can have a greater influence on model discrimination due to the significant differences in the number of parameters. But this is only true if all parameters can be examined in the applied reactor platform. Therefore, the accuracy of automated modeling relies on the available resources. The next chapter analyzes these resources, emphasizing their potential limitations and capabilities.

7.2. Assessment of System Limitations and Potential Capabilities

The developed reactor platform can determine the most appropriate kinetic model for the model reaction of FDH catalyzed reduction of NAD^+ to NADH with formate as a proton donor. This is shown and analyzed in this thesis in contrast to traditional experimental design. The thesis discusses the benefits of independent modelling and outlines the restrictions of the model identification procedure. The experimental platform has limitations that could greatly impact its application to other systems if components are replaced.

Syringe pumps were utilized in the PFR experiments to ensure a low volume flow pulsation and enable a flexible substrate dosing setup. The pump setup was connected to a laptop via an USB to RS-232 adapter, which acted as a serial interface. The pumps applied in the experiment were capable of daisy-chaining with minimal Python programming. However, their use mandated low volume syringes (10 mL) to attain adequate flow rates, with required linear force to overcome the back pressure inherent in the reactor system. In addition, the backpressure limitation restricted the residence time to 2 min, which had previously served as a useful input factor for the DoE in the EMR experiments. Thus, the underestimated enzyme load could have been compensated for by shorter residence times. In addition, the limited capacity of each syringe limited the number of experiments that could be performed before emptying. Because of the incremental nature of the MBDoE approach, it was not possible to predict the volumes required in advance. Planning volumes could be made easier by conducting *in silico* simulations of experiments in advance.

A straightforward enhancement to a high-pressure liquid chromatography system, or similar setup with high-pressure piston pumps, would greatly increase flexibility in residence times and also allow for temperature control through the regulated column oven. Alternatively, the system can be miniaturized, which would reduce the necessary volume for a single residence time and decrease backpressure along the reactor's length. Additionally, it was previously anticipated that the necessary enzyme or immobilizate amount would be lower. However, the utilization of small-scale reactors requires the implementation of specific periphery, enzyme loading and analytical devices. Particularly, obtaining sufficient online monitoring results using analytical devices in microscale is challenging due to their limited availability and requires special

considerations (Yue *et al.*, 2012; Sans *et al.*, 2016; Trojanowicz, 2016; Sagmeister *et al.*, 2019). Parameters such as relaxation time, detection limits, and temperature are essential for Benchtop-NMR (Sans *et al.*, 2015; Gomez *et al.*, 2017; Giraudeau *et al.*, 2018; Friebel *et al.*, 2019; Claaßen *et al.*, 2020). For infrared spectroscopy and Raman spectroscopy, specialized flow cells and reactors are required (Geske *et al.*, 2013; Fagaschewski *et al.*, 2015). There are limited instrumentational solutions readily available, therefore, offline analytics and online UV/Vis continue to be the preferred analytics in this field for most researchers due to their accessibility, cost, handling and easy data interpretation (Calleri *et al.*, 2021). However, due to the increasing popularity of micro reactors and continuous flow technology, more methods are being developed and utilized for this purpose.

The application of UV/Vis analysis in the reactor system has limitations in two ways. Firstly, there are lower and upper detection thresholds due to physical principles and device constraints. Secondly, the measurement interval can only be adjusted to a certain extent by selecting different optical path lengths. This limitation is not ideal for maximizing the parameter space and measurement window. The reactor platform presented here has a limitation in terms of univariate data analysis obtained from UV/Vis spectroscopy. Although several products may be detected at different wavelengths with a diode array detector, overlapping spectra must be distinguished, which may further limit the detection window due to additive absorption at certain wavelengths and different magnitudes of adsorbance. Therefore, including multivariate quantitative data analysis would improve the platform's flexibility, particularly for complex reactions. However, this would require modifying the code, which is currently hardcoded for the model reaction and the applied periphery. It would be easier to set up and customize a combination of LabVIEW with MATLAB scripts for different biotransformations and peripherals. In addition, chemometric modeling is required for multivariate data analysis to obtain concentration profiles from recorded spectra. This effort is worthwhile as it significantly enhances the capabilities of the reactor platform.

The assessment of kinetic parameters could be improved by adding several measurement points, which would increase the information gathered in a single run and lessen the experimental workload. This would also provide a better understanding of the axial behavior of the PFR. However, it is not necessary for the EMR, as the concentration profile varies in time and not location. One way to improve the efficiency of continuous

reactors in determining kinetic parameters is by conducting transient flow experiments. These experiments involve conducting several steady states in a single experiment (Waldron *et al.*, 2019, 2020; Taylor *et al.*, 2021; McMullen *et al.*, 2022). Another option is to examine the impact of inhibitory product concentrations through a recycling stream. As previously mentioned in chapter 5.1.2, each additional data collection, degree of freedom, candidate model, or model complexity increases the necessary computing power. This could potentially bottleneck local systems. To reduce the downtime between experiments by long calculation times with local hardware, a possible solution would be to utilize server-based data evaluation with the aid of cluster processors for rapid data analysis. Ultimately, this thesis establishes a foundation for identifying kinetic models of continuous enzymatic processes. This can be expanded as required with more powerful hardware, modified operation, and more comprehensive evaluation through more flexible programming. The approach enables efficient experimentation with various immobilizates, reactor types, and reactions. To achieve this, the platform must be correctly adapted with the suggestions and knowledge acquired in this thesis.

Additionally, this approach benefits from the increasing accessibility of artificial intelligence (AI) and machine learning approaches, which makes software development more accessible to a wider range of users in biotechnology (Oliveira, 2019; Holzinger *et al.*, 2023). This leads to more efficient process development through the use of AI and machine learning tools that can handle the large amount of data provided by automated screening procedures. The generated knowledge can be used to optimize experimental procedures, such as this near-process MBDoE approach, process predictions (Adebar *et al.*, 2024) or enzyme engineering (Yang *et al.*, 2023). Processed data must be curated and made accessible in databases to ensure quality, conformity, and completeness for both AI and human use (Garg *et al.*, 2019; Plehiers *et al.*, 2020; Bell *et al.*, 2021; Finnigan *et al.*, 2021; Pleiss, 2021; Lauterbach *et al.*, 2023). This will accelerate process development in biocatalysis, enhancing sustainability and effectiveness in the modern (bio-) chemical industry.

8. Summary

This thesis presented the development of an automated experimental platform to showcase the strengths and challenges of computer-assisted model identification for enzyme kinetics. The platform utilized a combination of MBDoE according to the *e-optimal* design and AWDC. The reactor platform was set up to run without human intervention and showed fully automated experimental campaigns to demonstrate the feasibility, repeatability, and accuracy of the model identification approach. The post-processing of the generated experimental data addressed the challenges and bottlenecks of the automated approach by identifying and resolving the influence of different aspects. The transferability of the method was demonstrated by applying the kinetic model identification to experimental results of another reactor type. The main research findings are as follows:

- An automated reactor platform was built to perform the kinetic model identification on a model reaction. The reduction of NAD^+ to NADH with FDH from *Candida boidinii* was applied as the model reaction. The experimental results demonstrated low standard deviation between 3.4% and 9.6%, repeatability of the model identification and no need for human intervention.

- An experimental workflow for automated kinetic measurements was developed, starting with 6 initial experiments based on a standard full factorial design, followed by 9 MBDoE experiments. Directed model identification was achieved by separating the two approaches, with thresholds for Akaike weights of over 0.95 indicating successful model discrimination and subsequent optimized parameter estimation with a threshold of $CI_{95\%}$ of under 50%.

- Software was developed to enable simultaneous execution and analysis of steady-state experiments using Python and MATLAB. The combination of AWDC for model discrimination and *e-optimal* designed experiments was utilized. However, the AIC did not provide an advantage in model discrimination over the standard negative log-likelihood at the level of model complexity applied. The CDC showed the same trend and was more refined for increased model discrimination than the AWDC. The potential benefits of using AIC or AWDC for model

reduction are promising, particularly when dealing with complex models that have significant differences in the number of parameters.

- Two automated experimental campaigns produced accurate parameter results, but differed in the resulting model, demonstrating the high relevance of the initial solutions to the model fit and the resulting experimental design. However, compared to a full factorial design, MBDoE achieved a more efficient way to identify better models. This advantage is not fully exploited with this simple model reaction, but the full potential is anticipated for more complex reactions.

- To avoid stagnation in local optima during model identification, a multi-start heuristic is required due to the high relevance of the initial solution and the strong correlation of the kinetic parameters. The number of starting point iterations must be balanced between computational power and result accuracy. The optimization algorithm is also crucial to computation time as the number of experiments increases, especially in more complex reaction systems.

- To enhance model discrimination, it is important to select a design space that is as large as possible. In the closed-loop experiments with a limited design space, most models could fit the experimental data, making model discrimination difficult unless the design space is extended.

- The method is transferable to different reactors and reactions, contingent upon the provision of mass balances and the accessibility of relevant concentration data within the screened design space. It is important to note that continuous reactors are easier to automate, but introducing a second enzyme makes cofactor regeneration more challenging. Alternatively, the cofactor can be supplied within the feed and introduced as an additional control variable.

- The model identification for the immobilized FDH packed in the PBR resulted in the MM model with the apparent parameters: k_{cat} = 1.21 s^{-1}± 6.6%, $K_{m,NAD}$ = 4.39 mM ± 13.5% and $K_{m,formate}$ = 31.04 mM± 8.7%. In contrast to that, the model

identification for the soluble FDH in the EMR resulted in the NCPI model with the kinetic parameters: k_{cat} = 2.1 s^{-1} ± 40.4%, $K_{m,NAD}$ = 0.44 mM ± 27.7%, $K_{m,formate}$ = 52.5 mM ± 9.1% and $K_{i,NADH}$ = 0.3 mM ± 49.7%. In the example of the EMR, the results indicate that enzyme kinetics may vary depending on the steady-state conditions of the process. Further investigation is necessary to thoroughly examine the impact of immobilization and distinguish non-apparent kinetic and mass transfer parameters.

- Automated experiments conducted under process conditions yield data that are more representative of the industrial environment than any assay could provide. The automated continuous reactor approach is particularly advantageous for screening immobilized enzymes, as they tend to be more stable and therefore useful for numerous parameter screenings. However, soluble enzymes must be stable enough to maintain a consistent concentration throughout an experimental campaign to obtain reliable data. This task requires stability or degradation analysis as a crucial characteristic. The reactor platform has the potential for enhancement to achieve accurate kinetic models for process optimization.

9. Material and Methods

9.1.Activity Assay

The activity of the FDH from *Candida boidinii* was carried out in a UV/Vis spectrophotometer (UvikonXL, Bio-Tek Instruments, Bad Friedrichshall, Germany) at 340 nm wavelength for 3 min at 30 °C to follow the production of NADH. The linear increase in the NADH absorption was measured as initial rate to compare wash fraction after immobilization and the immobilizate itself. All components were solved in 50 mM KP_i at pH 8. Typically, 1 mL of assay solution consists of 600 µL NAD^+ solution (1.67 mM), 200 µL formate solution (1 M), 150 µL buffer and 50 µL enzyme solution. The substrates formate and NAD^+ had a final concentration of 200 mM and 1 mM respectively. 150 µL buffer was employed to have volume of adjustment for the enzyme concentration if the concentration was too low to detect activity in the given time frame of 3 min.

For the immobilized FDH, the same assay composition was employed but it was scaled to an assay volume of 2 mL. The assay solution was measured as blank. The solution was preheated to 30 °C. After that, a triplicate of prepared particle samples of the immobilized enzyme was incubated with 2 mL of the assay solution and vortexed for 2 min to reduce diffusion limitation. The resulting assay mixture was separated from the particles and measured at 340 nm to detect the produced NADH.

The active NAD^+ concentration was measured according to the activity test of the FDH with 1 mL of assay solution consisting of 600 µL NAD^+ solution (1.67 mM), 200 µL formate solution (1 M), 50 µL buffer and 150 µL enzyme solution (1 $mg \cdot mL^{-1}$) in a cuvette. A blank was prepared in a cuvette with 600 µL NAD^+ solution (1.67 mM), 200 µL formate solution (1 M) and 200 µL buffer. Both cuvette were left for 2 hours at room temperature and measured at 340 nm after that time. The measurement resulted in 95% and confirmed the manufacturer specifications of ≥ 95% NADH (Carl Roth, Karlsruhe, Germany).

9.2. Pierce Assay

The Pierce assay was carried out in UV/Vis microtiter plate reader (Infinite 200 pro, Tecan, Männedorf, Switzerland). Here, 150 µL Pierce reagent was pipetted in triplicates separate wells on the plate. After that, 10 µL of sample was added in each well with a following incubation time of 5 min at room temperature. The microtiter plates were placed in the plate reader after incubation to determine the enzyme concentration at 660 nm. The calibration for the assay was performed with BSA in the concentration range of 0 to 2000 $\mu g \cdot mL^{-1}$.

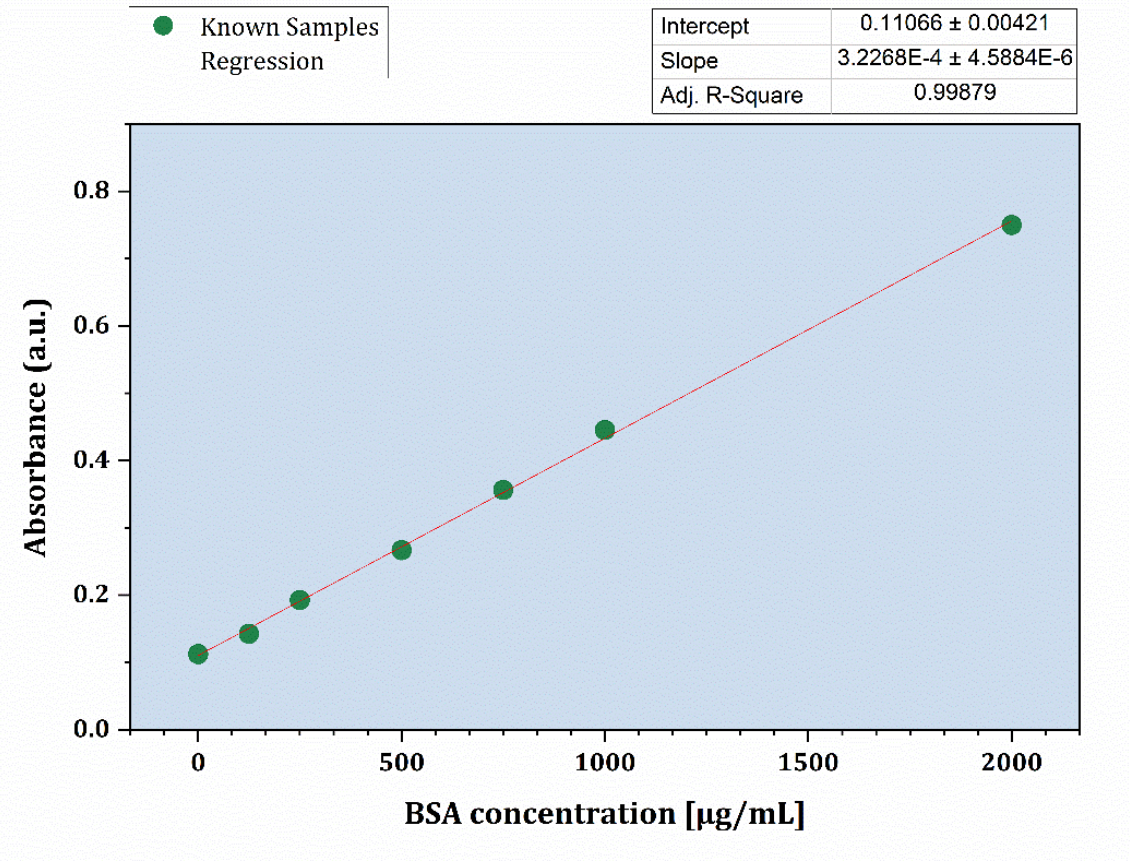

Figure 43: Pierce assay at 660 nm calibrated with BSA from 0 to 2000 $\mu g \cdot mL^{-1}$ in 50 mM KP_i buffer at pH 8.

9.3. Immobilization

For the retention and the stabilization of the enzyme FDH in a PBR, it was immobilized on methylmethacrylate epoxy functionalized resins (ECR Lifetech 8204M, Purolite, King of Prussia, USA). The employed particles had a diameter of 300 – 710 µm. Since the commercial enzyme formulation (Megazyme, Dublin, Ireland) was delivered as suspension in 3.2 M ammonium sulfate, it was centrifuged to separate the solid enzyme. Therefore, 1 mL of enzyme suspension was centrifuged 10 min at 15000 rpm and 4°C. The resulting solid pellet was resuspended in 500 mM KP_i buffer at pH 8 to a total volume of 10 mL. The carrier was first equilibrated 500 mM KP_i buffer at pH 8 before usage. Hereby, the carrier was rinsed with 1 mL buffer four times. The prepared enzyme solution was put in a 15 mL centrifuge tube with 328.8 mg enzyme carrier. The mixture was gently stirred in an axial stirrer at 4 °C for 23 h. After incubation, the immobilizate was washed eight times with 1 mL 50 mM KP_i pH 8 with gentle vortexing. The washing fractions were collected to determine the immobilization yield. The pierce assay of the supernatant and all washing fraction resulted in an immobilization yield of 52.2% (cf. Figure 44). After washing, the immobilizate was vacuum filtrated and stored at 4 °C until packing of the reactor. The calculated specific activities are listed in Table 11.

Table 11: Particle masses and according specific activity for the immobilized FDH measured in the UV/Vis spectrophotometer at 340 nm under assay conditions.

Particle mass [mg]	**Specific activity [mU·mg$_{Carrier}^{-1}$]**
4.5	3.15
4.1	3.75
2.5	4.23
	Mean = 3.71 ± 0.5

One unit is defined as the amount of enzyme required to convert one µmol NAD^+ per minute at assay conditions (50 mM KP_i buffer at pH 8 and a temperature of 30°C) to NADH and CO_2. The activity of the soluble enzyme with a concentration of 1.9 mg·mL^{-1} was 1.7 U·mg$_{Enzyme}^{-1}$. Applying the enzyme loading of the carrier of 0.03 mg$_{Enzyme}$·mg$_{Carrier}^{-1}$ and the mean measured activity of 3.71 mU·mg$_{Carrier}^{-1}$, an enzyme

specific activity of 0.12 $U \cdot mg_{Enzyme}^{-1}$ results. The activity yield accounts for 7.3% at assay conditions and is corrected with experimental results in the PBR to 66.3% (cf. chapter 5.2.).

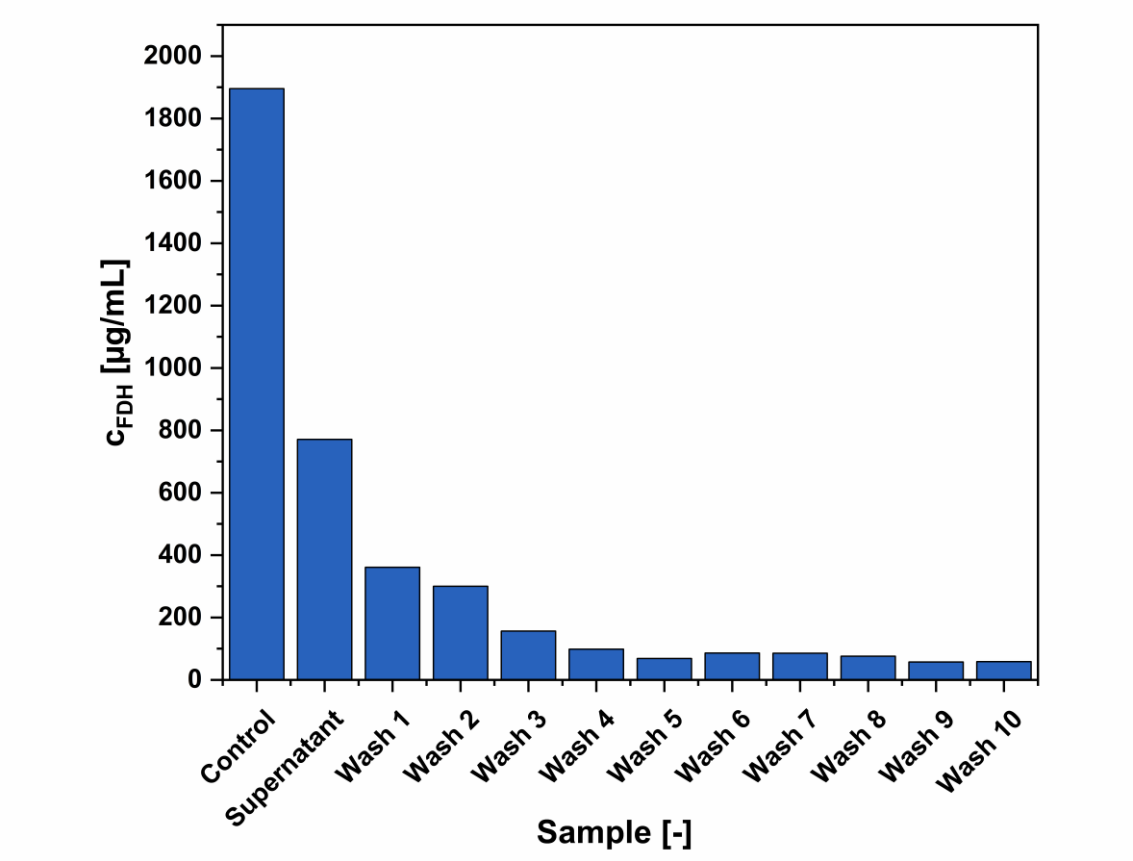

Figure 44: FDH concentration of the supernatant after the immobilization procedure, the control of the stock solution and the washing steps.

9.4. UV/Vis Analytics

For the determination of the NADH concentration in the outlet flow of the reactor platform, the spectrophotometer was calibrated in the range of 0 to 2 mM NADH at 340 nm (Figure 46). The intensity data was obtained via the *python-Seabreeze* package (Poehlmann, 2019) and converted to absorbance with equation 30. The NADH concentration was provided in 50 mM KP_i at pH 8 for calibration with a corrected concentration based on the purity according to the certificate of analysis of 85.6%. A typical spectrum is shown in Figure 45. The dotted lines indicate the wavelength spread of 10 nm, which was introduced to compensate a sensor offset with optical UV/Vis probes in early iterations of the reactor setup.

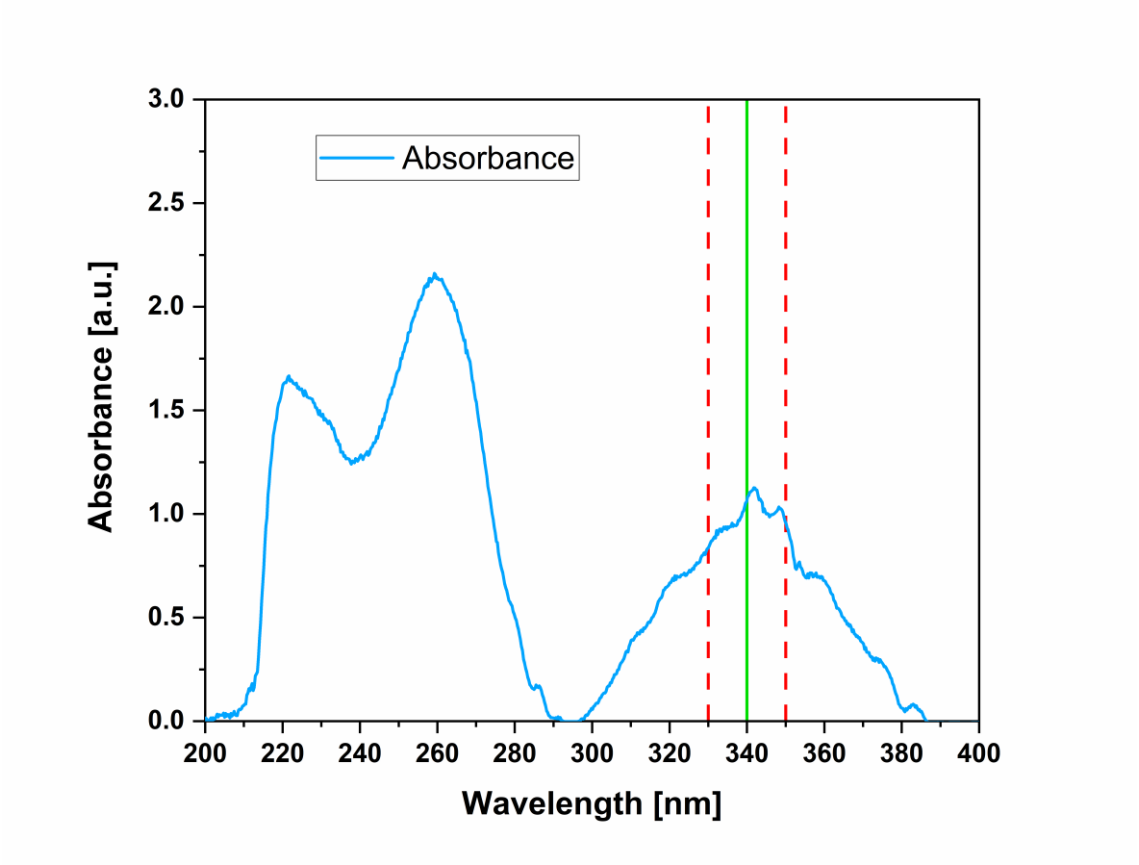

Figure 45: UV/Vis spectrum of NADH with reaction conditions of 1 mM NAD^+ and 150 mM formate in KP_i buffer at pH 8. The shown NADH concentration is 0.89 mM in a 1 mm flow-through cuvette. The green line indicate the measurement wavelength of 340 nm and the dotted lines represent the wavelength spread of 10 nm.

Linear Calibration

$$A = sc + i \tag{29}$$

Absorbance

$$A = \log_{10} \frac{I_0}{I_i} \tag{30}$$

Limit of Detection (LoD)

$$LoD(c) = \frac{3.3\sigma}{s} \tag{31}$$

Limit of Quantification (LoQ)

$$LoQ(c) = \frac{10\sigma}{s} \tag{32}$$

A = Absorbance

s = slope

c = concentration

i = Intercept

σ = standard deviation

I_0 = Baseline intensity

I_i = Measured intensity

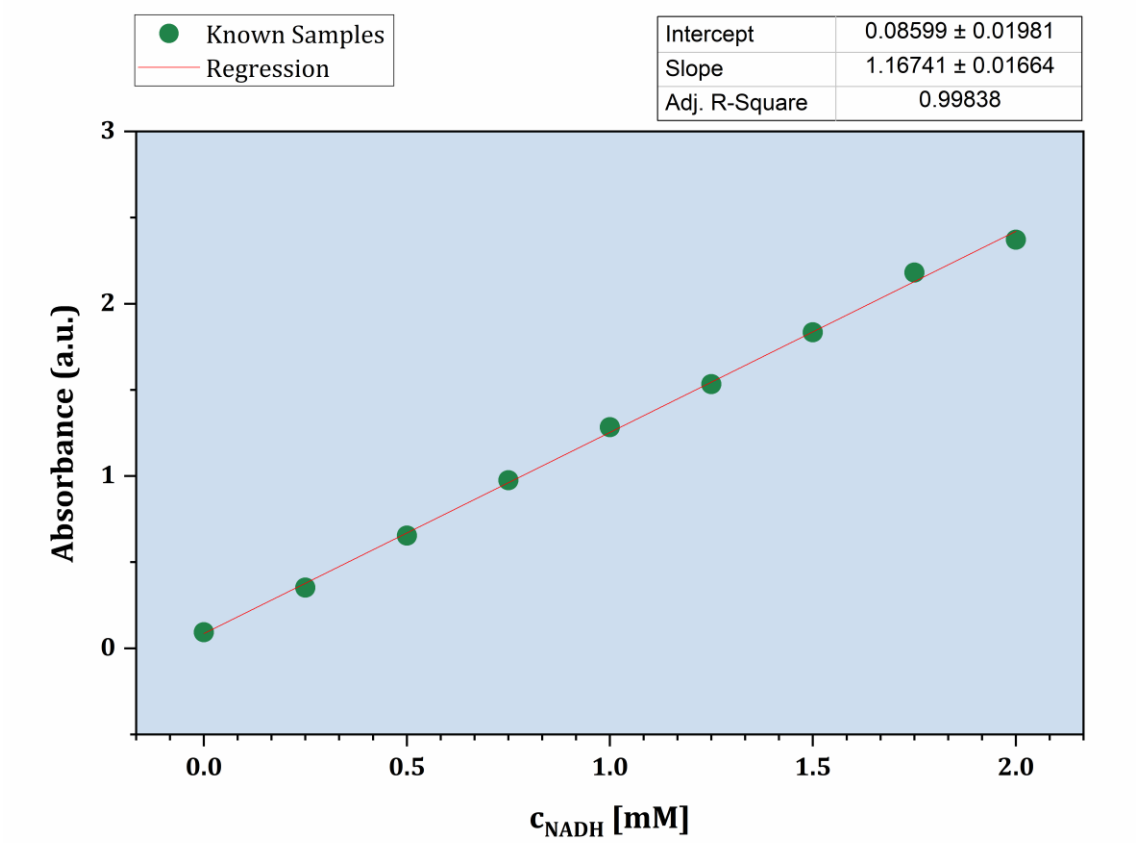

Figure 46: Calibration curve for NADH at 340 nm for the closed-loop determination of the product concentration in a 1 mm flow through cuvette. The concentration range is 0 mM to 2 mM NADH in 50 mM KP_i buffer pH 8. The same UV/Vis analytics were applied for the EMR setup. LoD of 0.047 mM and a LoQ of 0.143 mM.

In addition to the calibration of the automated experimental platform, manual measurements were carried out with a UV/Vis spectrophotometer (Shimadzu UV-1280, Duisburg, Germany) to extend the measurable product concentration range up to 9 mM NADH. This was achieved by shifting the wavelength to 380 nm and using a 1 mm glass cuvette. Concentrations over 9 mM were outside of the linear range (Figure 47).

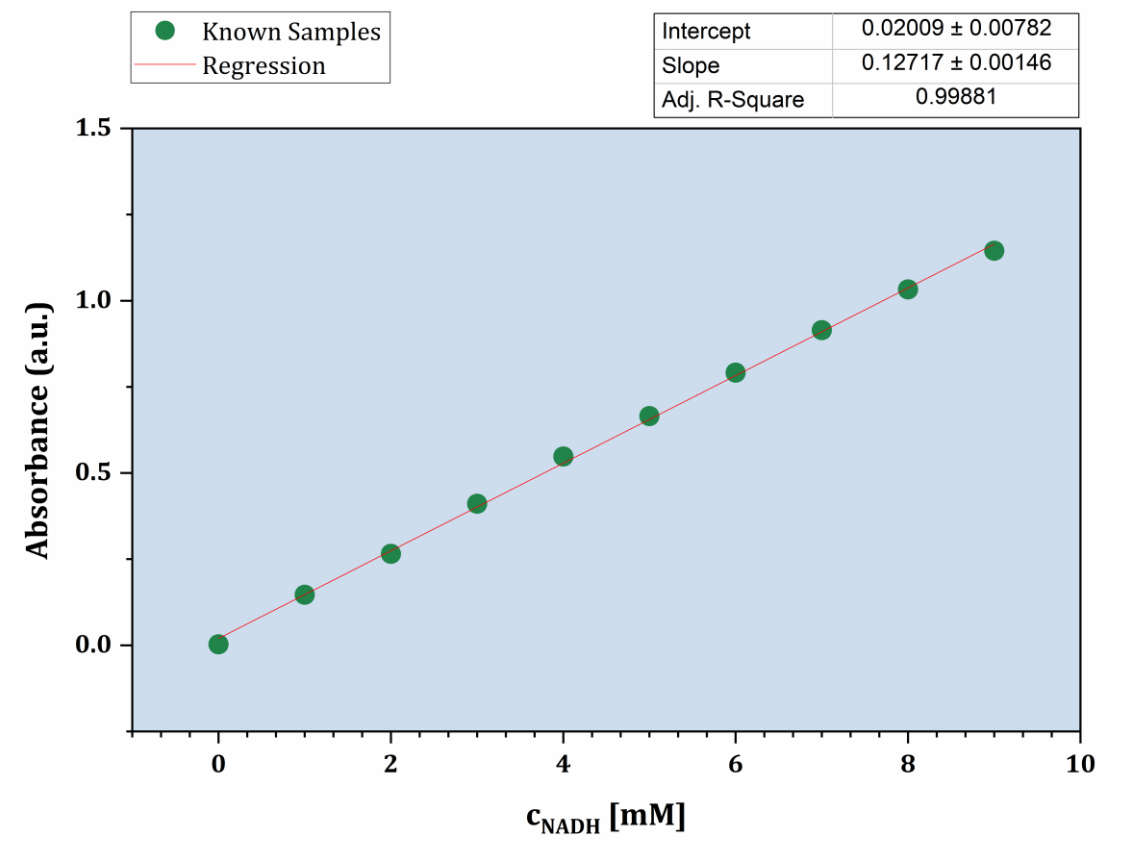

Figure 47: Calibration curve for NADH at 380 nm for the determination of the product concentration of the extended design space up to 10 mM NAD^+. The concentration range is 0 mM to 9 mM NADH in 50 mM KP_i buffer. LoD = 0.038 mM and LoQ = 0.115 mM.

Due to the instability of NAD^+ at room temperature (23°C), a calibration curve for NAD^+ was carried out at 300 nm to measure the concentration change when stored at room temperature in 50 mM buffer pH 8. This also had to be calibrated in the UV/Vis spectrophotometer (UvikonXL, Bio-Tek Instruments, Bad Friedrichshall, Germany).

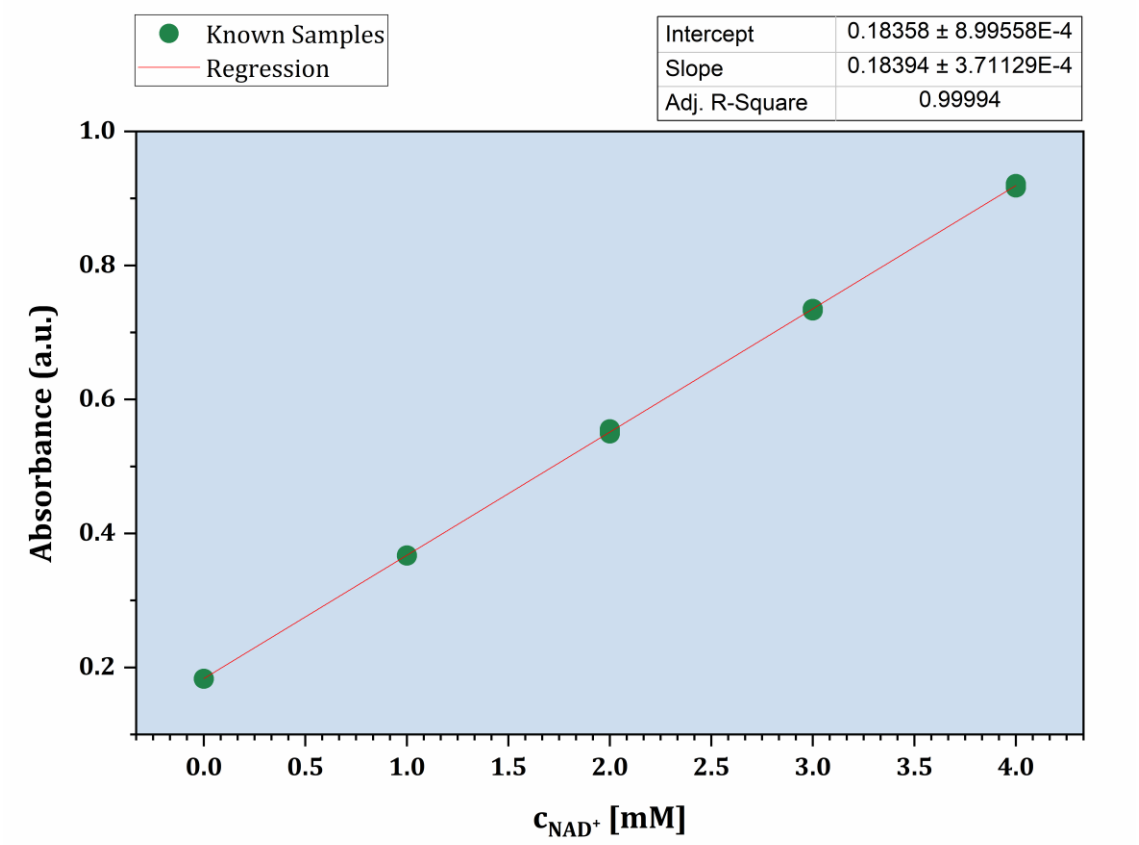

Figure 48: Calibration curve for NAD^+ at 300 nm for the determination of NAD^+ degradation rate. The concentration range is 0 mM to 9 mM NADH in 50 mM KP_i buffer at pH 8. LoD = 0.007 mM and LoQ = 0.02 mM

In the time scale of the experimental platform of up to 6 hours, no significant difference between an exponential (first order) and linear (zero order) decay of NAD^+ was found (cf. Figure 49). Therefore, the linear degradation rate of 0.275 $\mu M \cdot min^{-1}$ was taken into account for the dosing system.

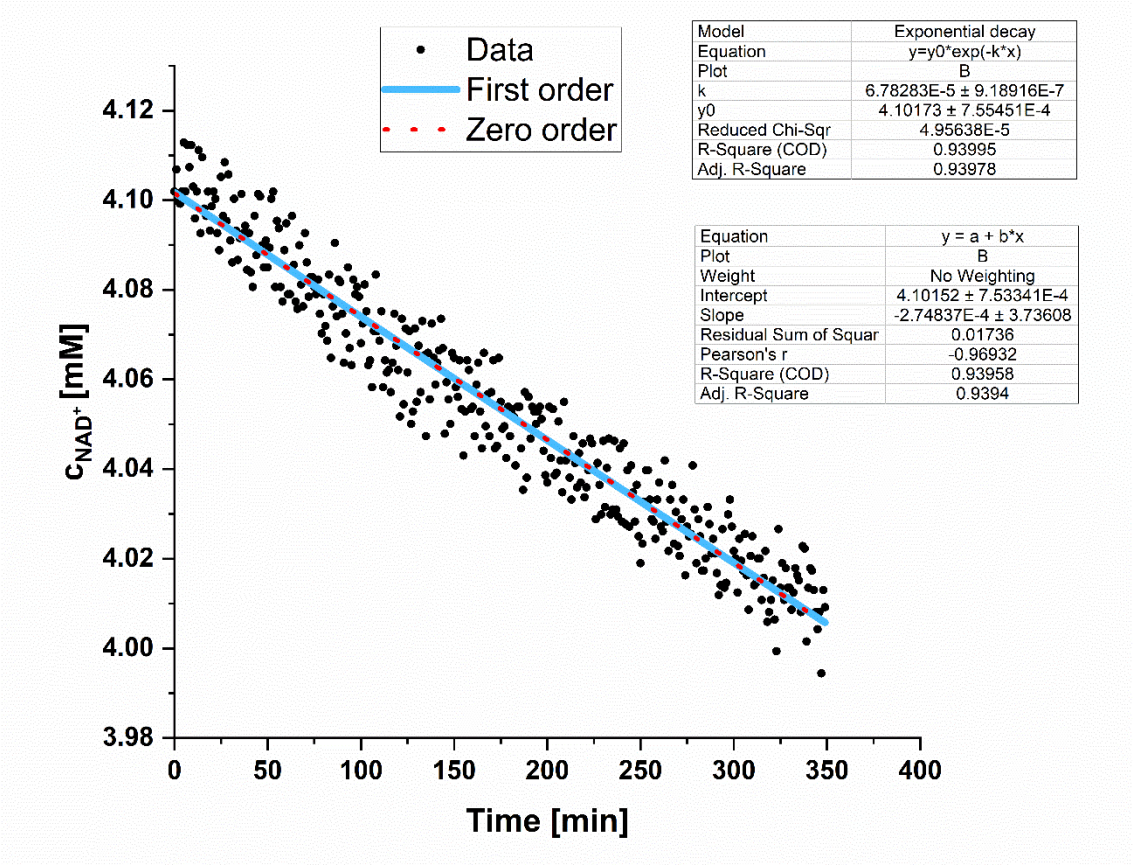

Figure 49: Fitting of first and zero order kinetics to the decaying concentration of NAD^+ in 50 mM KP_i at pH 8 and 24 °C for 6 hours. The measurement was performed at 300 nm.

9.5. Python and MATLAB Scripts

As part of this thesis, Python and MATLAB scripts were developed, which are available in the repository for Open Access publications and research data of the Hamburg University of Technology (TORE).

Name: EnzymeVision - Automated Kinetic Model Identification of Biocatalysts Under Process Conditions

DOI: https://doi.org/10.15480/882.9427

10. References

Adebar, N., **Keupp, J.**, **Emenike, V. N.**, **Kühlborn, J.**, **Vom Dahl, L.**, **Möckel, R.** and **Smiatek, J.** (**2024**) 'Scientific Deep Machine Learning Concepts for the Prediction of Concentration Profiles and Chemical Reaction Kinetics: Consideration of Reaction Conditions', *The journal of physical chemistry. A*, 128(5), pp. 929–944. doi: 10.1021/acs.jpca.3c06265.

Akaike, H. (**1974**) 'A New Look at the Statistical Model Identification', *IEEE Transactions on Automatic Control*, 19(6), pp. 716–723. doi: 10.1109/TAC.1974.1100705.

Alonso, S., **Santiago, G.**, **Cea-Rama, I.**, **Fernandez-Lopez, L.**, **Coscolín, C.**, **Modregger, J.**, **Ressmann, A. K.**, **Martínez-Martínez, M.**, **Marrero, H.**, **Bargiela, R.**, **Pita, M.**, **Gonzalez-Alfonso, J. L.**, **Briand, M. L.**, **Rojo, D.**, **Barbas, C.**, **Plou, F. J.**, **Golyshin, P. N.**, **Shahgaldian, P.**, **Sanz-Aparicio, J.**, **Guallar, V.** and **Ferrer, M.** (**2020**) 'Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis', *Nature Catalysis*, 3(3), pp. 319–328. doi: 10.1038/s41929-019-0394-4.

Arnold, F. H. (**2018**) 'Directed Evolution: Bringing New Chemistry to Life', *Angewandte Chemie - International Edition*, 57(16), pp. 4143–4148. doi: 10.1002/anie.201708408.

Atkinson, A. C. and **Fedorov, V. V.** (**1975**) 'The design of experiments for discriminating between two rival models', *Biometrika*, 62(1), pp. 57–70. doi: 10.1093/biomet/62.1.57.

Bandara, P. C., **Nadres, E. T.** and **Rodrigues, D. F.** (**2019**) 'Use of Response Surface Methodology to Develop and Optimize the Composition of a Chitosan-Polyethyleneimine-Graphene Oxide Nanocomposite Membrane Coating to More Effectively Remove Cr(VI) and Cu(II) from Water', *ACS Applied Materials and Interfaces*, 11(19), pp. 17784–17795. doi: 10.1021/acsami.9b03601.

Basso, A. and **Serban, S.** (**2019**) 'Industrial applications of immobilized enzymes—A review', *Molecular Catalysis*. Elsevier, 479(March), p. 110607. doi: 10.1016/j.mcat.2019.110607.

Baumann, M., **Moody, T. S.**, **Smyth, M.** and **Wharry, S.** (**2020**) 'A Perspective on Continuous Flow Chemistry in the Pharmaceutical Industry', *Organic Process Research and Development*, 24(10), pp. 1802–1813. doi: 10.1021/acs.oprd.9b00524.

Bell, E. L., **Finnigan, W.**, **France, S. P.**, **Green, A. P.**, **Hayes, M. A.**, **Hepworth, L. J.**, **Lovelock, S. L.**, **Niikura, H.**, **Osuna, S.**, **Romero, E.**, **Ryan, K. S.**, **Turner, N. J.** and **Flitsch, S. L.** (**2021**) 'Biocatalysis', *Nature Reviews Methods Primers*, 1(1), pp. 1–21. doi: 10.1038/s43586-021-00044-z.

Bennett, J. A., **Campbell, Z. S.** and **Abolhasani, M.** (**2019**) 'Role of continuous flow processes in green manufacturing of pharmaceuticals and specialty chemicals', *Current Opinion in Chemical Engineering*. Elsevier Ltd, 26, pp. 9–19. doi: 10.1016/j.coche.2019.07.007.

Bezerra, R. M. F., **Fraga, I.** and **Dias, A. A.** (**2013**) 'Utilization of integrated Michaelis-Menten equations for enzyme inhibition diagnosis and determination of kinetic constants using Solver supplement of Microsoft Office Excel', *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd, 109(1), pp. 26–31. doi: 10.1016/j.cmpb.2012.08.017.

Bezerra, R. M. F., **Pinto, P. A.**, **Fraga, I.** and **Dias, A. A.** (**2016**) 'Enzyme inhibition studies by integrated Michaelis-Menten equation considering simultaneous presence of

two inhibitors when one of them is a reaction product', *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd, 125, pp. 2–7. doi: 10.1016/j.cmpb.2015.12.013.
Binay, B., **Alagöz, D.**, **Yildirim, D.**, **Çelik, A.** and **Tükel, S. S.** (**2016**) 'Highly stable and reusable immobilized formate dehydrogenases: Promising biocatalysts for in situ regeneration of NADH', *Beilstein Journal of Organic Chemistry*, 12(February), pp. 271–277. doi: 10.3762/bjoc.12.29.
Bisswanger, H. (**2017**) *Enzyme Kinetics*. Wiley. doi: 10.1002/9783527806461.
Bolivar, J. M., **Wilson, L.**, **Ferrarotti, S. A.**, **Fernandez-Lafuente, R.**, **Guisan, J. M.** and **Mateo, C.** (**2007**) 'Evaluation of different immobilization strategies to prepare an industrial biocatalyst of formate dehydrogenase from Candida boidinii', *Enzyme and Microbial Technology*, 40(4), pp. 540–546. doi: 10.1016/j.enzmictec.2006.05.009.
Breitung, K. and **Faravelli, L.** (**1994**) 'Log-likelihood maximization and response surface in reliability assessment', *Nonlinear Dynamics*, 5(3), pp. 273–285. doi: 10.1007/BF00045337.
Burek, B. O., **Dawood, A. W. H.**, **Hollmann, F.**, **Liese, A.** and **Holtmann, D.** (**2022**) 'Process Intensification as Game Changer in Enzyme Catalysis', *Frontiers in Catalysis*, 2(March), pp. 1–18. doi: 10.3389/fctls.2022.858706.
Burnham, K. P. and **Anderson, D. R.** (**2002**) *Model Selection and Multimodel Inference*, *Model Selection and Multimodel Inference*. Springer New York. doi: 10.1007/B97636.
Büscher, N., **Sayoga, G. V.**, **Rübsam, K.**, **Jakob, F.**, **Schwaneberg, U.**, **Kara, S.** and **Liese, A.** (**2019**) 'Biocatalyst Immobilization by Anchor Peptides on an Additively Manufacturable Material', *Organic Process Research and Development*, 23(9), pp. 1852–1859. doi: 10.1021/acs.oprd.9b00152.
Buzzi-Ferraris, G. (**2010**) 'Some Observations on the paper "optimal experimental design for discriminating numerous model candidates: The AWDC criterion"', *Industrial and Engineering Chemistry Research*, 49(19), pp. 9561–9562. doi: 10.1021/ie100373t.
Buzzi-Ferraris, G., **Forzatti, P.** and **Paolo, C.** (**1990**) 'An improved version of a sequential design criterion for discriminating among rival multiresponse models', *Chemical Engineering Science*, 45(2), pp. 477–481. doi: 10.1016/0009-2509(90)87034-P.
Buzzi-Ferraris, G. and **Manenti, F.** (**2009**) 'Kinetic models analysis', *Chemical Engineering Science*, 64(5), pp. 1061–1074. doi: 10.1016/j.ces.2008.10.062.
Calleri, E., **Temporini, C.**, **Colombo, R.**, **Tengattini, S.**, **Rinaldi, F.**, **Brusotti, G.**, **Furlanetto, S.** and **Massolini, G.** (**2021**) 'Analytical settings for in-flow biocatalytic reaction monitoring', *TrAC - Trends in Analytical Chemistry*. Elsevier Ltd, 143, p. 116348. doi: 10.1016/j.trac.2021.116348.
Cavanaugh, J. E. and **Neath, A. A.** (**2019**) 'The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements', *Wiley Interdisciplinary Reviews: Computational Statistics*, 11(3), pp. 1–11. doi: 10.1002/wics.1460.
Chakrabarti, A. and **Ghosh, J. K.** (**2011**) *AIC, BIC and Recent Advances in Model Selection*, *Philosophy of Statistics*. Elsevier B.V. doi: 10.1016/B978-0-444-51862-0.50018-6.
Choi, B., **Rempala, G. A.** and **Kim, J. K.** (**2017**) 'Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters', *Scientific Reports*. Springer US, 7(1), pp. 1–11. doi: 10.1038/s41598-017-17072-z.
Claaßen, C., **Mack, K.** and **Rother, D.** (**2020**) 'Benchtop NMR for Online Reaction Monitoring of the Biocatalytic Synthesis of Aromatic Amino Alcohols', *ChemCatChem*, 12(4), pp. 1190–1199. doi: 10.1002/cctc.201901910.

Cleland, W. . (**1963**) 'The kinetics of enzyme-catalyzed reactions with two or more substrates or products', *Biochimica et Biophysica Acta (BBA) - Specialized Section on Enzymological Subjects*, 67(2), pp. 104–137. doi: 10.1016/0926-6569(63)90211-6.
Coloma, J., **Guiavarc', Y.**, **Hagedoorn, P.-L.** and **Hanefeld, U.** (**2021**) 'Immobilisation and flow chemistry: tools for implementing biocatalysis', *11416 | Chem. Commun*, 57, p. 11416. doi: 10.1039/d1cc04315c.
Cooney, M. J. (**2017**) 'Kinetic measurements for enzyme immobilization', *Methods in Molecular Biology*, 1504, pp. 215–232. doi: 10.1007/978-1-4939-6499-4_17.
Cornish-Bowden, A. (**2012**) *Fundamentals of enzyme kinetics 4.* 4., comple. Weinheim: Wiley-Blackwell.
Croughan, M. S., **Konstantinov, K. B.** and **Cooney, C.** (**2015**) 'The future of industrial bioprocessing: Batch or continuous?', *Biotechnology and Bioengineering*, 112(4), pp. 648–651. doi: 10.1002/bit.25529.
Datta, S., **Christena, L. R.** and **Rajaram, Y. R. S.** (**2013**) 'Enzyme immobilization: an overview on techniques and support materials', *3 Biotech*, 3(1), pp. 1–9. doi: 10.1007/s13205-012-0071-7.
Draper, N. R. and **Pukelsheim, F.** (**1996**) 'An overview of design of experiments', *Statistical Papers*, 37, pp. 1–32. doi: 10.1007/BF02926157.
Duggleby, R. G. (**1995**) 'Analysis of enzyme progress curves by nonlinear regression', *Methods in Enzymology*, 249(C), pp. 61–90. doi: 10.1016/0076-6879(95)49031-0.
Fagaschewski, J., **Sellin, D.**, **Wiedenhöfer, C.**, **Bohne, S.**, **Trieu, H. K.** and **Hilterhaus, L.** (**2015**) 'Spatially resolved in situ determination of reaction progress using microfluidic systems and FT-IR spectroscopy as a tool for biocatalytic process development', *Bioprocess and Biosystems Engineering*, 38(7), pp. 1399–1405. doi: 10.1007/s00449-015-1381-z.
Finnigan, W., **Hepworth, L. J.**, **Flitsch, S. L.** and **Turner, N. J.** (**2021**) 'RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades', *Nature Catalysis*. Springer US, 4(2), pp. 98–104. doi: 10.1038/s41929-020-00556-z.
Franceschini, G. and **Macchietto, S.** (**2008**) 'Model-based design of experiments for parameter precision: State of the art', *Chemical Engineering Science*. Pergamon, 63(19), pp. 4846–4872. doi: 10.1016/j.ces.2007.11.034.
Francis, F., **Sabu, A.**, **Nampoothiri, K. M.**, **Ramachandran, S.**, **Ghosh, S.**, **Szakacs, G.** and **Pandey, A.** (**2003**) 'Use of response surface methodology for optimizing process parameters for the production of α-amylase by Aspergillus oryzae', *Biochemical Engineering Journal*, 15(2), pp. 107–115. doi: 10.1016/S1369-703X(02)00192-4.
Friebel, A., **Von Harbou, E.**, **Münnemann, K.** and **Hasse, H.** (**2019**) 'Reaction Monitoring by Benchtop NMR Spectroscopy Using a Novel Stationary Flow Reactor Setup', *Industrial and Engineering Chemistry Research*, 58(39), pp. 18125–18133. doi: 10.1021/acs.iecr.9b03048.
Galvanin, F., **Barolo, M.** and **Bezzo, F.** (**2010**) 'A framework for model-based design of experiments in the presence of continuous measurement systems', *IFAC Proceedings Volumes*. IFAC, 43(5), pp. 571–576. doi: 10.3182/20100705-3-be-2011.00095.
Garg, N., **Woodley, J. M.**, **Gani, R.** and **Kontogeorgis, G. M.** (**2019**) 'Sustainable solutions by integrating process synthesis-intensification', *Computers and Chemical Engineering*. Elsevier Ltd, 126, pp. 499–519. doi: 10.1016/j.compchemeng.2019.04.030.
Geske, M., **Korup, O.** and **Horn, R.** (**2013**) 'Resolving kinetics and dynamics of a catalytic reaction inside a fixed bed reactor by combined kinetic and spectroscopic profiling', *Catalysis Science and Technology*, 3(1), pp. 169–175. doi: 10.1039/c2cy20489d.

Giraudeau, P. and **Felpin, F. X.** (**2018**) 'Flow reactors integrated with in-line monitoring using benchtop NMR spectroscopy', *Reaction Chemistry and Engineering*, 3(4), pp. 399–413. doi: 10.1039/c8re00083b.

Gomez, M. V. and **De La Hoz, A.** (**2017**) 'NMR reaction monitoring in flow synthesis', *Beilstein Journal of Organic Chemistry*, 13, pp. 285–300. doi: 10.3762/bjoc.13.31.

Grosch, J. H., **Wagner, D.**, **Knaup, N.**, **Keil, T.** and **Spieß, A. C.** (**2017**) 'Influence of the experimental setup on the determination of enzyme kinetic parameters', *Biotechnology Progress*. John Wiley and Sons Inc., 33(1), pp. 87–95. doi: 10.1002/btpr.2390.

Guo, Q., **Gakhar, L.**, **Wickersham, K.**, **Francis, K.**, **Vardi-Kilshtain, A.**, **Major, D. T.**, **Cheatum, C. M.** and **Kohen, A.** (**2016**) 'Structural and Kinetic Studies of Formate Dehydrogenase from Candida boidinii', *Biochemistry*, 55(19), pp. 2760–2771. doi: 10.1021/acs.biochem.6b00181.

Gygli, G. (**2022**) 'On the reproducibility of enzyme reactions and kinetic modelling', *Biological Chemistry*, (Ibg 1), pp. 1–29. doi: 10.1515/hsz-2021-0393.

Hanes, C. S. (**1932**) 'Studies on plant amylases: The effect of starch concentration upon the velocity of hydrolysis by the amylase of germinated barley.', *The Biochemical journal*, 26(5), pp. 1406–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16744959%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1261052.

Häse, F., **Roch, L. M.** and **Aspuru-Guzik, A.** (**2019**) 'Next-Generation Experimentation with Self-Driving Laboratories', *Trends in Chemistry*, 1(3), pp. 282–291. doi: 10.1016/j.trechm.2019.02.007.

Heintz, S., **Börner, T.**, **Ringborg, R. H.**, **Rehn, G.**, **Grey, C.**, **Nordblad, M.**, **Krühne, U.**, **Gernaey, K. V.**, **Adlercreutz, P.** and **Woodley, J. M.** (**2017**) 'Development of in situ product removal strategies in biocatalysis applying scaled-down unit operations', *Biotechnology and Bioengineering*, 114(3), pp. 600–609. doi: 10.1002/bit.26191.

Hess, D., **Dockalova, V.**, **Kokkonen, P.**, **Bednar, D.**, **Damborsky, J.**, **deMello, A.**, **Prokop, Z.** and **Stavrakis, S.** (**2021**) 'Exploring mechanism of enzyme catalysis by on-chip transient kinetics coupled with global data analysis and molecular modeling', *Chem*. Elsevier Inc., 7(4), pp. 1066–1079. doi: 10.1016/j.chempr.2021.02.011.

Hofstee, B. H. J. (**1952**) 'On the evaluation of the constants Vm and KM in enzyme reactions', *Science*, 116(3013), pp. 329–331. doi: 10.1126/science.116.3013.329.

Holzinger, A., **Keiblinger, K.**, **Holub, P.**, **Zatloukal, K.** and **Müller, H.** (**2023**) 'AI for life: Trends in artificial intelligence for biotechnology', *New Biotechnology*. Elsevier B.V., 74(February), pp. 16–24. doi: 10.1016/j.nbt.2023.02.001.

Hülsewede, D., **Meyer, L. E.** and **von Langermann, J.** (**2019**) 'Application of In Situ Product Crystallization and Related Techniques in Biocatalytic Processes', *Chemistry - A European Journal*, 25(19), pp. 4871–4884. doi: 10.1002/chem.201804970.

Hummel, W., **Schütte, H.**, **Schmidt, E.**, **Wandrey, C.** and **Kula, M. R.** (**1987**) 'Isolation of l-phenylalanine dehydrogenase from Rhodococcus sp. M4 and its application for the production of l-phenylalanine', *Applied Microbiology and Biotechnology*, 26(5), pp. 409–416. doi: 10.1007/BF00253523.

Hunter, W. G. and **Reiner, A. M.** (**1965**) 'Designs for Discriminating Between Two Rival Models', *Technometrics*, 7(3), pp. 307–323. doi: 10.1080/00401706.1965.10490265.

Illanes, A., **Anjarí, S.**, **Arrieta, R.** and **Aguirre, C.** (**2002**) 'Optimization of yield in kinetically controlled synthesis of ampicillin with immobilized penicillin acylase in organic media', *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering*

and Biotechnology, 97(3), pp. 165–179. doi: 10.1385/ABAB:97:3:165.
Jaeger, K.-E., **Liese, A.** and **Syldatk, C.** (**2018**) *Einführung in die Enzymtechnologie*, *Einführung in die Enzymtechnologie*. Edited by K.-E. Jaeger, A. Liese, and C. Syldatk. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-662-57619-9.
Johnson, K. A. (**2013**) 'A century of enzyme kinetic analysis, 1913 to 2013', *FEBS Letters*, 587(17), pp. 2753–2766. doi: 10.1016/j.febslet.2013.07.012.
Kara, S., **Müller, J. J.** and **Liese, A.** (**2011**) 'Online analysis methods for monitoring of bioprocesses', *Chemistry Today*, 29(2), pp. 38–41.
Karande, R., **Schmid, A.** and **Buehler, K.** (**2016**) 'Applications of Multiphasic Microreactors for Biocatalytic Reactions', *Organic Process Research and Development*, 20(2), pp. 361–370. doi: 10.1021/acs.oprd.5b00352.
Kato, N., **Sahm, H.** and **Wagner, F.** (**1979**) 'Steady-state kinetics of formaldehyde dehydrogenase and formate dehydrogenase from a methanol-utilizing yeast, Candida boidinii', *BBA - Enzymology*, 566(1), pp. 12–20. doi: 10.1016/0005-2744(79)90243-2.
Khuri, A. I. and **Mukhopadhyay, S.** (**2010**) 'Response surface methodology', *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2), pp. 128–149. doi: 10.1002/wics.73.
Klimeš, P., **Mazura, P.**, **Turek, D.** and **Brzobohatý, B.** (**2017**) 'An automated method to evaluate the enzyme kinetics of β-glucosidases', *Protein Science*, 26(2), pp. 382–388. doi: 10.1002/pro.3078.
Kragl, U., **Kruse, W.**, **Hummel, W.** and **Wandrey, C.** (**1996**) 'Enzyme engineering aspects of biocatalysis: Cofactor regeneration as example', *Biotechnology and Bioengineering*, 52(2), pp. 309–319. doi: 10.1002/(SICI)1097-0290(19961020)52:2<309::AID-BIT11>3.0.CO;2-E.
Kruger, N. J. (**1995**) 'Errors and artifacts in coupled spectrophotometric assays of enzyme activity', *Phytochemistry*, 38(5), pp. 1065–1071. doi: 10.1016/0031-9422(94)00787-T.
Kula, M., **Wichmann, R.**, **En, U. O. I.** and **Wandrey, C.** (**1980**) 'Influence of substrate or product inhibition on the performance of reactors', *Biochimie*, 62(8–9), pp. 523–536. doi: https://doi.org/10.1016/S0300-9084(80)80097-6.
Laun, K., **Duffus, B. R.**, **Wahlefeld, S.**, **Katz, S.**, **Belger, D.**, **Hildebrandt, P.**, **Mroginski, M. A.**, **Leimkühler, S.** and **Zebger, I.** (**2022**) 'Infrared Spectroscopy Elucidates the Inhibitor Binding Sites in a Metal-Dependent Formate Dehydrogenase', *Chemistry - A European Journal*, 28(54). doi: 10.1002/chem.202201091.
Lauterbach, S., **Dienhart, H.**, **Range, J.**, **Malzacher, S.**, **Spöring, J. D.**, **Rother, D.**, **Pinto, M. F.**, **Martins, P.**, **Lagerman, C. E.**, **Bommarius, A. S.**, **Høst, A. V.**, **Woodley, J. M.**, **Ngubane, S.**, **Kudanga, T.**, **Bergmann, F. T.**, **Rohwer, J. M.**, **Iglezakis, D.**, **Weidemann, A.**, **Wittig, U.**, **Kettner, C.**, **Swainston, N.**, **Schnell, S.** and **Pleiss, J.** (**2023**) 'EnzymeML: seamless data flow and modeling of enzymatic data', *Nature Methods*. Springer US, 20(3), pp. 400–402. doi: 10.1038/s41592-022-01763-1.
Liese, A., **Seelbach, K.** and **Wandrey, C.** (**2006**) *Industrial Biotransformations*. Second Edi. Edited by A. Liese, K. Seelbach, and C. Wandrey.
Lindeque, R. M. and **Woodley, J. M.** (**2019**) 'Reactor selection for effective continuous biocatalytic production of pharmaceuticals', *Catalysts*, 9(3). doi: 10.3390/catal9030262.
Lineweaver, H. and **Burk, D.** (**1934**) 'The Determination of Enzyme Dissociation Constants', *Journal of the American Chemical Society*, 56(3), pp. 658–666. doi: 10.1021/ja01318a036.
Madhavan, A., **Arun, K. B.**, **Binod, P.**, **Sirohi, R.**, **Tarafdar, A.**, **Reshmy, R.**, **Kumar Awasthi, M.** and **Sindhu, R.** (**2021**) 'Design of novel enzyme biocatalysts for industrial

bioprocess: Harnessing the power of protein engineering, high throughput screening and synthetic biology', *Bioresource Technology*. Elsevier Ltd, 325(December 2020), p. 124617. doi: 10.1016/j.biortech.2020.124617.

Maghraby, Y. R., **El-Shabasy, R. M.**, **Ibrahim, A. H.** and **Azzazy, H. M. E. S.** (**2023**) 'Enzyme Immobilization Technologies and Industrial Applications', *ACS Omega*, 8(6), pp. 5184–5196. doi: 10.1021/acsomega.2c07560.

Martin, B., **Lehmann, H.**, **Yang, H.**, **Chen, L.**, **Tian, X.**, **Polenk, J.** and **Schenkel, B.** (**2018**) 'Continuous manufacturing as an enabling tool with green credentials in early-phase pharmaceutical chemistry', *Current Opinion in Green and Sustainable Chemistry*. Elsevier, 11, pp. 27–33. doi: 10.1016/J.COGSC.2018.03.005.

Martins, J. R. R. A., **Sturdza, P.** and **Alonso, J. J.** (**2003**) 'The complex-step derivative approximation', *ACM Transactions on Mathematical Software (TOMS)*. ACM PUB27 New York, NY, USA , 29(3), pp. 245–262. doi: 10.1145/838250.838251.

McMullen, J. P. and **Wyvratt, B. M.** (**2022**) 'Automated optimization under dynamic flow conditions', *Reaction Chemistry and Engineering*. Royal Society of Chemistry, 8(1), pp. 137–151. doi: 10.1039/d2re00256f.

Megazyme (**2023**) *Formate dehydrogenase (Candida boidinii)*. Available at: https://www.megazyme.com/formate-dehydrogenase-candida-boidinii.

Miao, Y., **Rahimi, M.**, **Geertsema, E. M.** and **Poelarends, G. J.** (**2015**) 'Recent developments in enzyme promiscuity for carbon-carbon bond-forming reactions', *Current Opinion in Chemical Biology*. Elsevier Ltd, 25, pp. 115–123. doi: 10.1016/j.cbpa.2014.12.020.

Michaelis, L. and **Menten, M. L.** (**1913**) 'The kinetics of invertase action', *Biochem. Zeitung*, 49, pp. 79–90. doi: 10.1007/BF02477924.

Michalik, C., **Schmidt, T.**, **Zavrel, M.**, **Ansorge-Schumacher, M.**, **Spiess, A.** and **Marquardt, W.** (**2007**) 'Application of the incremental identification method to the formate oxidation using formate dehydrogenase', *Chemical Engineering Science*, 62(18–20), pp. 5592–5597. doi: 10.1016/j.ces.2006.12.072.

Michalik, C., **Stuckert, M.** and **Marquardt, W.** (**2010**) 'Optimal Experimental Design for Discriminating Numerous Model Candidates: The AWDC Criterion', *Industrial & Engineering Chemistry Research*. American Chemical Society, 49(2), pp. 913–919. doi: 10.1021/ie900903u.

Miličić, N., **Ćevid, I.**, **Çakar, M. M.**, **Sudar, M.** and **Blažević, Z. F.** (**2022**) 'Enzyme Reaction Engineering as a Tool to Investigate the Potential Application of Enzyme Reaction Systems', *Hungarian Journal of Industry and Chemistry*, 50(1), pp. 45–55. doi: 10.33927/hjic-2022-08.

Mueller, I., **Runne, E.** and **Hamel, C.** (**2022**) 'Comparative Study on Mechanistic Kinetic Modeling of the Enzymatic Synthesis of Galacto-Oligosaccharides', *Chemie-Ingenieur-Technik*, 94(5), pp. 774–779. doi: 10.1002/cite.202100190.

Nikolova, N., **Tenekedjiev, K.** and **Kolev, K.** (**2008**) 'Uses and misuses of progress curve analysis in enzyme kinetics', *Central European Journal of Biology*, 3(4), pp. 345–350. doi: 10.2478/S11535-008-0035-4.

Nocedal, J. and **Wright, S. J.** (**2006**) *Numerical optimization*, *Springer Series in Operations Research and Financial Engineering*. doi: 10.1201/b19115-11.

Ohs, R., **Wendlandt, J.** and **Spiess, A. C.** (**2017**) 'How graphical analysis helps interpreting optimal experimental designs for nonlinear enzyme kinetic models', *AIChE Journal*, 63(11), pp. 4870–4880. doi: 10.1002/aic.15814.

Oliveira, A. L. (**2019**) 'Biotechnology, Big Data and Artificial Intelligence',

Biotechnology Journal. John Wiley & Sons, Ltd, 14(8), pp. 1–6. doi: 10.1002/biot.201800613.
Otun, S. O., **Lerma-Escalera, J. A.**, **Ntushelo, K.** and **Achilonu, I.** (**2023**) 'Protein engineering for natural product biosynthesis: expanding diversity for therapeutic applications', *Journal of Bio-X Research*, 6(2), pp. 49–60. doi: 10.1097/JBR.0000000000000141.
Ötvös, S. B. and **Kappe, C. O.** (**2021**) 'Continuous flow asymmetric synthesis of chiral active pharmaceutical ingredients and their advanced intermediates', *Green Chemistry*. Royal Society of Chemistry, 23(17), pp. 6117–6138. doi: 10.1039/d1gc01615f.
Pinto, M. F., **Ripoll-Rozada, J.**, **Ramos, H.**, **Watson, E. E.**, **Franck, C.**, **Payne, R. J.**, **Saraiva, L.**, **Pereira, P. J. B.**, **Pastore, A.**, **Rocha, F.** and **Martins, P. M.** (**2019**) 'A simple linearization method unveils hidden enzymatic assay interferences', *Biophysical Chemistry*, 252, pp. 1–25. doi: 10.1016/j.bpc.2019.106193.
Plehiers, P. P., **Coley, C. W.**, **Gao, H.**, **Vermeire, F. H.**, **Dobbelaere, M. R.**, **Stevens, C. V.**, **Van Geem, K. M.** and **Green, W. H.** (**2020**) 'Artificial Intelligence for Computer-Aided Synthesis In Flow: Analysis and Selection of Reaction Components', *Frontiers in Chemical Engineering*. Frontiers, 2, p. 5. doi: 10.3389/FCENG.2020.00005.
Pleiss, J. (**2021**) 'Standardized Data, Scalable Documentation, Sustainable Storage – EnzymeML As A Basis For FAIR Data Management In Biocatalysis', *ChemCatChem*, 13(18), pp. 3909–3913. doi: 10.1002/cctc.202100822.
Poehlmann, A. (**2019**) *Python Seabreeze*. Available at: https://python-seabreeze.readthedocs.io/en/latest/.
Radtke, R. (**2023**) *Weltweite Ausgaben für Biotechnologie in den Jahren 2012 bis 2027*. Available at: https://de.statista.com/statistik/daten/studie/1368790/umfrage/weltweite-ausgaben-fuer-biotech/.
Rogers, L. and **Jensen, K. F.** (**2019**) 'Continuous manufacturing-the Green Chemistry promise?', *Green Chemistry*. Royal Society of Chemistry, 21(13), pp. 3481–3498. doi: 10.1039/c9gc00773c.
Ryzhkov, F. V., **Ryzhkova, Y. E.** and **Elinson, M. N.** (**2023**) 'Python in Chemistry: Physicochemical Tools', *Processes*, 11(10). doi: 10.3390/pr11102897.
Sagmeister, P., **Williams, J. D.**, **Hone, C. A.** and **Kappe, C. O.** (**2019**) 'Laboratory of the future: A modular flow platform with multiple integrated PAT tools for multistep reactions', *Reaction Chemistry and Engineering*. Royal Society of Chemistry, 4(9), pp. 1571–1578. doi: 10.1039/c9re00087a.
Sans, V. and **Cronin, L.** (**2016**) 'Towards dial-a-molecule by integrating continuous flow, analytics and self-optimisation', *Chemical Society Reviews*. Royal Society of Chemistry, 45(8), pp. 2032–2043. doi: 10.1039/c5cs00793c.
Sans, V., **Porwol, L.**, **Dragone, V.** and **Cronin, L.** (**2015**) 'A self optimizing synthetic organic reactor system using real-time in-line NMR spectroscopy', *Chemical Science*. Royal Society of Chemistry, 6(2), pp. 1258–1264. doi: 10.1039/c4sc03075c.
De Santis, P., **Meyer, L. E.** and **Kara, S.** (**2020**) 'The rise of continuous flow biocatalysis-fundamentals, very recent developments and future perspectives', *Reaction Chemistry and Engineering*. Royal Society of Chemistry, 5(12), pp. 2155–2184. doi: 10.1039/d0re00335b.
Sauro, H. M. (**2017**) 'Control and regulation of pathways via negative feedback', *Journal of the Royal Society Interface*, 14(127). doi: 10.1098/rsif.2016.0848.
van Schie, M. M. C. H., **Spöring, J. D.**, **Bocola, M.**, **Domínguez de María, P.** and **Rother, D.** (**2021**) 'Applied biocatalysis beyond just buffers - From aqueous to unconventional media. Options and guidelines', *Green Chemistry*. Royal Society of

Chemistry, 23(9), pp. 3191–3206. doi: 10.1039/d1gc00561h.
Schmider, J., **Greenblatt, D. J.**, **Harmatz, J. S.** and **Shader, R. I.** (**1996**) 'Enzyme kinetic modelling as a tool to analyse the behaviour of cytochrome P450 catalysed reactions: Application to amitriptyline N-demethylation', *British Journal of Clinical Pharmacology*, 41(6), pp. 593–604. doi: 10.1046/j.1365-2125.1996.35717.x.
Schmidt, E., **Ghisalba, O.**, **Gygax, D.** and **Sedelmeier, G.** (**1992**) 'Optimization of a process for the production of (R)-2-hydroxy-4-phenylbutyric acid - an intermediate for inhibitors of angiotensin converting enzyme', *Journal of Biotechnology*, 24(3), pp. 315–327. doi: 10.1016/0168-1656(92)90040-G.
Schmidt, T., **Michalik, C.**, **Zavrel, M.**, **Spieß, A.**, **Marquardt, W.** and **Ansorge-Schumacher, M. B.** (**2009**) 'Mechanistic Model for Prediction of Formate Dehydrogenase Kinetics Under Industrially Relevant Conditions', *American Institute of Chemical Engineers Biotechnol. Prog*, 26, pp. 73–78. doi: 10.1002/btpr.282.
Sensirion AG (**2021**) *Sensirion SHDLC Python Driver*. Available at: https://sensirion.github.io/python-shdlc-driver/index.html.
Sharma, A., **Gupta, G.**, **Ahmad, T.**, **Mansoor, S.** and **Kaur, B.** (**2021**) 'Enzyme Engineering: Current Trends and Future Perspectives', *Food Reviews International*. Taylor & Francis, 37(2), pp. 121–154. doi: 10.1080/87559129.2019.1695835.
Sheldon, R. A. (**2020**) 'Biocatalysis and biomass conversion: enabling a circular economy', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2176), p. 20190274. doi: 10.1098/rsta.2019.0274.
Sheldon, R. A., **Basso, A.** and **Brady, D.** (**2021**) 'New frontiers in enzyme immobilisation: Robust biocatalysts for a circular bio-based economy', *Chemical Society Reviews*. Royal Society of Chemistry, 50(10), pp. 5850–5862. doi: 10.1039/d1cs00015b.
Sheldon, R. A. and **Woodley, J. M.** (**2018**) 'Role of Biocatalysis in Sustainable Chemistry', *Chemical Reviews*, 118(2), pp. 801–838. doi: 10.1021/acs.chemrev.7b00203.
Short, M., **Schenk, C.**, **Thierry, D.**, **Rodriguez, J. S.**, **Biegler, L. T.** and **Garcia-Muñoz, S.** (**2019**) *KIPET – An Open-Source Kinetic Parameter Estimation Toolkit*, *Computer Aided Chemical Engineering*. Elsevier Masson SAS. doi: 10.1016/B978-0-12-818597-1.50047-3.
Stroberg, W. and **Schnell, S.** (**2016**) 'On the estimation errors of KM and V from time-course experiments using the Michaelis–Menten equation', *Biophysical Chemistry*. Elsevier B.V., 219, pp. 17–27. doi: 10.1016/j.bpc.2016.09.004.
Sudar, M. and **Blažević, Z. F.** (**2021**) *Enzyme cascade kinetic modelling*, *Enzyme Cascade Design and Modelling*. doi: 10.1007/978-3-030-65718-5_6.
Tamborini, L., **Fernandes, P.**, **Paradisi, F.** and **Molinari, F.** (**2018**) 'Flow Bioreactors as Complementary Tools for Biocatalytic Process Intensification', *Trends in Biotechnology*. Elsevier Ltd, 36(1), pp. 73–88. doi: 10.1016/j.tibtech.2017.09.005.
Tang, Q. and **Leyh, T. S.** (**2010**) 'Precise, facile initial rate measurements', *Journal of Physical Chemistry B*, 114(49), pp. 16131–16136. doi: 10.1021/jp1055528.
Taylor, C. J., **Booth, M.**, **Manson, J. A.**, **Willis, M. J.**, **Clemens, G.**, **Taylor, B. A.**, **Chamberlain, T. W.** and **Bourne, R. A.** (**2021**) 'Rapid, automated determination of reaction models and kinetic parameters', *Chemical Engineering Journal*. Elsevier B.V., 413, p. 127017. doi: 10.1016/j.cej.2020.127017.
Telford, J. K. (**2007**) 'A brief introduction to design of experiments', *Johns Hopkins APL Technical Digest (Applied Physics Laboratory)*, 27(3), pp. 224–232.
Thompson, M. P., **Peñafiel, I.**, **Cosgrove, S. C.** and **Turner, N. J.** (**2019**) 'Biocatalysis Using Immobilized Enzymes in Continuous Flow for the Synthesis of Fine Chemicals', *Organic Process Research and Development*, 23(1), pp. 9–18. doi:

10.1021/acs.oprd.8b00305.
Trojanowicz, M. (**2016**) 'Flow chemistry vs. flow analysis', *Talanta*. Elsevier, 146, pp. 621–640. doi: 10.1016/j.talanta.2015.07.043.
Vang, J. Y., **Breceda, C.**, **Her, C.** and **Krishnan, V. V.** (**2022**) 'Enzyme kinetics by real-time quantitative NMR (qNMR) spectroscopy with progress curve analysis', *Analytical Biochemistry*. Elsevier Inc., 658(May), p. 114919. doi: 10.1016/j.ab.2022.114919.
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.
Vasic-Racki, D., **Kragl, U.** and **Liese, A.** (**2003**) 'Benefits of enzyme kinetics modelling', *Chemical and Biochemical Engineering Quarterly*, 17(1), pp. 7–18.
Waldron, C., **Pankajakshan, A.**, **Quaglio, M.**, **Cao, E.**, **Galvanin, F.** and **Gavriilidis, A.** (**2019**) 'An autonomous microreactor platform for the rapid identification of kinetic models', *Reaction Chemistry and Engineering*. Royal Society of Chemistry, 4(9), pp. 1623–1636. doi: 10.1039/c8re00345a.
Waldron, C., **Pankajakshan, A.**, **Quaglio, M.**, **Cao, E.**, **Galvanin, F.** and **Gavriilidis, A.** (**2020**) 'Model-based design of transient flow experiments for the identification of kinetic parameters †'. doi: 10.1039/c9re00342h.
Ward, E. J. (**2008**) 'A review and comparison of four commonly used Bayesian and maximum likelihood model selection tools', *Ecological Modelling*, 211(1–2), pp. 1–10. doi: 10.1016/j.ecolmodel.2007.10.030.
Wichmann, R., **Wandrey, C.**, **Bückmann, A. F.** and **Kula, M.-R.** (**1981**) 'Continuous enzymatic transformation in an enzyme membrane reactor with simultaneous NAD(H) regeneration', *Biotechnology and Bioengineering*, 13(12), pp. 2789–2802. doi: 10.1002/(SICI)1097-0290(20000320)67:6<791::AID-BIT15>3.0.CO;2-I.
Yang, J., **Li, F. Z.** and **Arnold, F. H.** (**2023**) 'Opportunities and Challenges for Machine Learning-Assisted Enzyme Engineering', *ACS Central Science*. doi: 10.1021/acscentsci.3c01275.
Yang, Z. (**2009**) 'Hofmeister effects: an explanation for the impact of ionic liquids on biocatalysis', *Journal of Biotechnology*, 144(1), pp. 12–22. doi: 10.1016/j.jbiotec.2009.04.011.
Yue, J., **Schouten, J. C.** and **Alexander Nijhuis, T.** (**2012**) 'Integration of microreactors with spectroscopic detection for online reaction monitoring and catalyst characterization', *Industrial and Engineering Chemistry Research*, 51(45), pp. 14583–14609. doi: 10.1021/ie301258j.
Zavrel, M., **Kochanowski, K.** and **Spiess, A. C.** (**2010**) 'Comparison of different approaches and computer programs for progress curve analysis of enzyme kinetics', *Engineering in Life Sciences*, 10(3), pp. 191–200. doi: 10.1002/elsc.200900083.
Zavrel, M., **Michalik, C.**, **Schwendt, T.**, **Schmidt, T.**, **Ansorge-Schumacher, M.**, **Janzen, C.**, **Marquardt, W.**, **Büchs, J.** and **Spiess, A. C.** (**2010**) 'Systematic determination of intrinsic reaction parameters in enzyme immobilizates', *Chemical Engineering Science*, 65(8), pp. 2491–2499. doi: 10.1016/j.ces.2009.12.026.
Zimmerle, C. T. and **Frieden, C.** (**1989**) 'Analysis of progress curves by simulations generated by numerical integration', *Biochemical Journal*, 258(2), pp. 381–387. doi: 10.1042/bj2580381.

11. Appendix

11.1. Material List

Table 12: Applied chemicals

Substance	CAS	Purity	Manufacturer	City	Country
NAD^+	53-84-9	≥ 95%	Carl Roth	Karlsruhe	Germany
NADH disodium salt	606-68-8	≥ 84%	Carl Roth	Karlsruhe	Germany
Sodium formate			Carl Roth	Karlsruhe	Germany
Potassium dihydrogen phosphate	7778-77-0	≥ 99%	Carl Roth	Karlsruhe	Germany
Di-potassium hydrogen phosphate	7758-11-4	≥ 99%	Carl Roth	Karlsruhe	Germany
Bovine Serum Albumin	9048-46-8	>98%	Sigma-Aldich Chemie GmbH	Steinheim	Germany
FDH (*Candida boidinii*) solid	9028-85-7	0.92 U/mg solid	Merck	Darmstadt	Germany
FDH (*Candida boidinii*) liquid	9028-85-7	75 U/mL in 3.2 M ammonium sulphate	Megazyme	Bray	Ireland
Lifetech ECR 8204M epoxy resin	-	-	Purolite	King of Prussia	USA
Pierce 660 nm protein assay reagent	-	-	Thermo Scientific	Waltham	USA

Table 13: Software

Software	Version	Publisher	City	Country
Chemdraw Prime	8.2	PerkinElmer Informatics	Köln	Germany
Excel	2013	Microsoft	Redmond	USA
Faulhaber Motion Manager	6	FAULHABER Drive Systems	Schönaich	Germany
MATLAB	R2021b	Mathworks	Aachen	Germany
Origin 2020	9.7.0.188	OriginLab Corporation	Northhampton	USA
Powerpoint	2013	Microsoft	Redmond	USA

Python	3.9.12	Python Software Foundation	-	-
Word	2013	Microsoft	Redmond	USA

Table 14: Applied apparatuses

Apparatus	**Specification**	**Manufacturer**	**City**	**Country**
Chromatography column	Xela ID: 3 mm, 5 cm length	Isera GmbH	Düren	Germany
Circulation thermostat	Lauda L100	LAUDA DR. R. WOBSER GMBH & CO. KG	Lauda-Königshofen	Germany
Enzyme Membrane Reactor	10 mL	Forschungszentrum Jülich GmbH,	Jülich	Germany
Flowmeter	Sensirion SLF3S-0600F	Sensirion AG	Staefa	Switzerland
Flow-through cuvette	170.700 QS	Hellma GmbH & Co. KG	Müllheim	Germany
Magnetic Stirrer	IKAMAG RCT	IKA	Breisgau	Germany
Micro gear pump	mzr-2905-1082	HNP Mikrosysteme GmbH	Schwerin	Germany
PTFE Suction Filter	5-10 µm pores	Techlab	Braunschweig	Germany
Regenerated Cellulose Membrane Filter	Ultracel®, 63.5 mm, 10 kDa	Merck KGaA	Darmstadt	Germany
Syringe pump	AL-1000	World Precision Instruments	Sarasota	USA
Syringes	Omnifix	Braun	Kronberg	Germany
UV/Vis Inline Spectrophotometer	HR4000	Ocean Insight	Dunedin	USA
UV/Vis light source	DH-2000-BAL	Ocean Insight	Dunedin	USA
UV/Vis Photometer	Shimadzu UV-1280	Shimadzu Deutschland GmbH	Duisburg	Germany
UV/Vis Photometer (heated)	Uvikon XL	Secomam	Oullins	France

11.2. Data sets

Table 15: Meta data set of steady state NADH concentrations with the corresponding NAD^+ concentration, formate concentration and residence time. All the data was produced in the automated experimental reactor platform in a packed bed reactor.

No	NAD^+ [mM]	Formate [mM]	τ [min]	NADH [mM]	Note
1	1	150	2	0.94	MBDoE-1.1
2	0.35	50	2	0.27	MBDoE-1.1
3	0.35	290	2	0.33	MBDoE-1.1
4	1.9	50	2	1.37	MBDoE-1.1
5	1.9	290	2	1.60	MBDoE-1.1
6	1	150	2	0.92	MBDoE-1.1
7	1.88	50	2	1.41	MBDoE-1.1
8	1.81	275.5	2	1.54	MBDoE-1.1
9	1.14	50	2	0.97	MBDoE-1.1
10	1.81	275.5	2	1.54	MBDoE-1.1
11	1.87	50	2	1.40	MBDoE-1.1
12	1.81	275.5	2	1.55	MBDoE-1.1
13	1.12	50	2	0.96	MBDoE-1.1
14	1.81	275.5	2	1.54	MBDoE-1.1
15	1.86	50	2	1.38	MBDoE-1.1
16	1	150	2	0.90	MBDoE-1.2
17	0.35	50	2	0.29	MBDoE-1.2
18	0.35	290	2	0.32	MBDoE-1.2
19	1.9	50	2	1.23	MBDoE-1.2
20	1.9	290	2	1.58	MBDoE-1.2
21	1	150	2	0.89	MBDoE-1.2
22	1.88	50	2	1.38	MBDoE-1.2
23	1.81	275.5	2	1.52	MBDoE-1.2
24	1.11	50	2	0.98	MBDoE-1.2
25	1.81	275.5	2	1.53	MBDoE-1.2
26	1.87	50	2	1.37	MBDoE-1.2
27	1.81	275.5	2	1.52	MBDoE-1.2
28	1.87	50	2	1.38	MBDoE-1.2
29	1.81	275.5	2	1.51	MBDoE-1.2
30	1.86	50	2	1.39	MBDoE-1.2
31	1	150	2	0.89	MBDoE-1.3
32	0.35	50	2	0.29	MBDoE-1.3
33	0.35	290	2	0.32	MBDoE-1.3
34	1.9	50	2	1.30	MBDoE-1.3
35	1.9	290	2	1.54	MBDoE-1.3
36	1	150	2	0.90	MBDoE-1.3
37	1.88	50	2	1.35	MBDoE-1.3
38	1.81	275.5	2	1.50	MBDoE-1.3
39	1.13	50	2	0.97	MBDoE-1.3
40	1.81	275.5	2	1.49	MBDoE-1.3

41	1.87	50	2	1.35	MBDoE-1.3
42	1.81	275.5	2	1.51	MBDoE-1.3
43	1.87	50	2	1.34	MBDoE-1.3
44	1.81	275.5	2	1.50	MBDoE-1.3
45	1.86	50	2	1.34	MBDoE-1.3
46	0.35	50	2	0.27	Full Factorial
47	0.35	133.3	2	0.28	Full Factorial
48	0.35	216.7	2	0.27	Full Factorial
49	0.35	300	2	0.30	Full Factorial
50	0.9	50	2	0.74	Full Factorial
51	0.9	133.3	2	0.79	Full Factorial
52	0.9	216.7	2	0.81	Full Factorial
53	0.9	300	2	0.82	Full Factorial
54	1.45	50	2	1.14	Full Factorial
55	1.45	133.3	2	1.22	Full Factorial
56	1.45	216.7	2	1.25	Full Factorial
57	1.45	300	2	1.27	Full Factorial
58	1.9	50	2	1.36	Full Factorial
59	1.9	133.3	2	1.50	Full Factorial
60	1.9	216.7	2	1.56	Full Factorial
61	1.9	300	2	1.57	Full Factorial
62	1	150	2	0.90	MBDoE-2
63	0.35	50	2	0.32	MBDoE-2
64	0.35	290	2	0.32	MBDoE-2
65	1.9	50	2	1.36	MBDoE-2
66	1.9	290	2	1.60	MBDoE-2
67	1	150	2	0.91	MBDoE-2
68	0.45	64.5	2	0.39	MBDoE-2
69	0.45	64.5	2	0.40	MBDoE-2
70	0.45	64.5	2	0.39	MBDoE-2
71	0.45	64.5	2	0.40	MBDoE-2
72	0.45	64.5	2	0.40	MBDoE-2
73	0.45	64.5	2	0.40	MBDoE-2
74	0.45	64.5	2	0.39	MBDoE-2
75	0.45	64.5	2	0.41	MBDoE-2
76	0.45	64.5	2	0.41	MBDoE-2
77	1	150	2	1.01	MBDoE-2 Repetition
78	0.35	50	2	0.35	MBDoE-2 Repetition
79	0.35	290	2	0.37	MBDoE-2 Repetition
80	1.9	50	2	1.45	MBDoE-2 Repetition
81	1.9	290	2	1.71	MBDoE-2 Repetition
82	1	150	2	1.01	MBDoE-2 Repetition
83	0.45	64.5	2	0.44	MBDoE-2 Repetition
84	0.45	64.5	2	0.46	MBDoE-2 Repetition
85	0.45	64.5	2	0.43	MBDoE-2 Repetition

86	0.45	64.5	2	0.45	MBDoE-2 Repetition
87	0.45	64.5	2	0.44	MBDoE-2 Repetition
88	0.45	64.5	2	0.45	MBDoE-2 Repetition
89	0.45	64.5	2	0.44	MBDoE-2 Repetition
90	0.45	64.5	2	0.44	MBDoE-2 Repetition
91	0.45	64.5	2	0.45	MBDoE-2 Repetition
92	1	150	2	0.93	Test Initial DoE
93	0.35	50	2	0.33	Test Initial DoE
94	0.35	290	2	0.33	Test Initial DoE
95	1.9	50	2	1.38	Test Initial DoE
96	1.9	290	2	1.60	Test Initial DoE
97	1	150	2	0.93	Test Initial DoE
98	1	150	2	0.89	MBDoE-1 Repetition
99	0.35	50	2	0.32	MBDoE-1 Repetition
100	0.35	290	2	0.30	MBDoE-1 Repetition
101	1.9	50	2	1.33	MBDoE-1 Repetition
102	1.9	290	2	1.60	MBDoE-1 Repetition
103	1	150	2	0.92	MBDoE-1 Repetition
104	1.88	50	2	1.35	MBDoE-1 Repetition
105	1.81	275.5	2	1.53	MBDoE-1 Repetition
106	1.12	50	2	1.00	MBDoE-1 Repetition
107	1.81	275.5	2	1.53	MBDoE-1 Repetition
108	1.87	50	2	1.33	MBDoE-1 Repetition
109	1.81	275.5	2	1.53	MBDoE-1 Repetition
110	1.87	50	2	1.34	MBDoE-1 Repetition
111	1.81	275.5	2	1.52	MBDoE-1 Repetition
112	1.86	50	2	1.33	MBDoE-1 Repetition
113	10	10	2	1.81	Extended Design Space
114	3.5	1	2	0.50	Extended Design Space
115	0.4	7	2	0.28	Extended Design Space
116	10	1500	2	6.79	Extended Design Space

Table 16 Meta data set of steady state NADH concentrations with the corresponding NAD^+ concentration, formate concentration and residence time. All the data was produced in the non-automated experimental reactor setup in an enzyme membrane reactor platform

No	NAD^+ [mM]	Formate [mM]	τ [s]	NADH [mM]
1	1.50	275.5	300	0.90
2	1.50	275.5	600	1.16
3	1.50	275.5	900	1.29
4	1.50	275.5	1200	1.34
5	1.50	275.5	1500	1.37
6	1.50	275.5	1800	1.38
7	1.50	278.0	300	0.88
8	1.50	278.0	420	1.03
9	1.50	278.0	540	1.14
10	1.50	278.0	660	1.23
11	1.39	266.0	240	0.71
12	1.39	266.0	210	0.65
13	1.08	270.4	210	0.65
14	1.08	270.4	240	0.70
15	1.08	270.4	300	0.77
16	1.08	270.4	360	0.81
17	1.08	270.4	480	0.90
18	1.08	270.4	600	0.96
19	2.16	266.3	210	0.93
20	2.16	266.4	240	1.00
21	2.16	266.4	300	1.12
22	2.16	266.4	360	1.20
23	2.16	266.4	480	1.33
24	2.16	266.4	600	1.44
25	1.61	10.02	210	0.31
26	1.61	10.02	240	0.32
27	1.61	10.02	300	0.35
28	1.61	10.02	360	0.38
29	1.61	10.02	480	0.42
30	1.61	10.02	600	0.45
31	1.94	99.98	210	0.70
32	1.94	99.98	240	0.75
33	1.94	99.98	300	0.83
34	1.94	99.98	360	0.90
35	1.94	99.98	480	1.01
36	1.94	99.98	600	1.09