Model-assisted design and optimization of biotechnological processes

Vom Promotionsausschuss der
Technischen Universität Hamburg
zur Erlangung des akademischen Grades

Doktor-Ingenieur (Dr.-Ing.)

genehmigte Dissertation (Monografie)

von
Lukas Arndt, M.Sc.
aus
Dannenberg (Elbe, Deutschland)

2024

Gutachter:

1. Gutachter: Prof. Dr.-Ing. Ralf Pörtner

2. Gutachter: Prof. Dr. Andreas Liese

Vorsitzender: Prof. Dr.-Ing. Mirko Skiborowski

Tag der mündlichen Prüfung:

08. Oktober 2024

Identifier

DOI: https://doi.org/10.15480/882.13739

Handle: https://tore.tuhh.de/handle/11420/52130

Abstract

Today, new developments in the field of biocatalysis enable the replacement of a wide range of chemical synthesis processes with biotransformations. The biocatalysts used in these processes are non-toxic and have a low environmental impact. To fully exploit the biocatalytic potential already in the early stages of process development, knowledge-based optimization of reaction conditions is crucial, but time-consuming. Design of Experiments, process modeling and regression analysis are effective tools for systematically planning experiments, describing parameter interactions, and determining appropriate process conditions. In the model-assisted Design of Experiments approach, these tools are combined. Initially, a process model is set up, model parameters are estimated and suggested experimental combinations are simulated. After evaluation of the predicted responses only a reduced number of factor combinations consistent with a pre-defined optimization objective are performed experimentally.

In this work, the model-assisted Design of Experiments methodology is discussed for the design and optimization of a continuously operated cytidine-5'-triphosphate regeneration reaction by polyphosphate kinase under high hydrostatic pressure. The process development was particularly complex since the packed-bed bioreactor system used and the enzymatic reaction were only partially described in the literature. An adequate process model was formulated to describe the time- and location-dependent changes of substrates and product as well as the influence of process conditions on the enzyme kinetics. The formulation of the process model proved to be extensive due to the use of immobilized enzyme and the consideration of a disperse flow within the reactor. In addition, a suitable modeling approach was elaborated to describe the influence of the process parameter pressure on the enzymatic reaction. To reduce the number of time-consuming individual optimization studies, the process parameters and their interactions were to be optimized and described simultaneously using the model-assisted Design of Experiments methodology.

The proposed methodology, integrated into a user-friendly iterative workflow, accompanied the interdisciplinary development process throughout and the process understanding of the reaction system was enhanced. Furthermore, the experimental effort was significantly reduced and process conditions were optimized in terms of maximal cytidine-5'-triphosphate concentration after three iteration rounds. Overall, the applicability of the knowledge-based model-assisted Design of Experiments methodology was shown and proven for a complex biocatalytic reaction system with a limited amount of prior knowledge. The transferability to other reaction systems and a further application of the methodology are discussed at the end of the work.

Kurzfassung

Neue Entwicklungen auf dem Gebiet der Biokatalyse ermöglichen es, eine Vielzahl chemischer Syntheseverfahren durch Biotransformationen zu ersetzen. Die eingesetzten Biokatalysatoren sind ungiftig und umweltverträglich. Um das volle biokatalytische Potenzial schon zu Beginn der Prozessentwicklung auszuschöpfen, ist eine wissensbasierte Optimierung der Reaktionsbedingungen entscheidend, aber zeitintensiv. Versuchsplanung, Prozessmodellierung und Regressionsanalyse sind effiziente Werkzeuge zur systematischen Planung und Durchführung von Experimenten, zur Beschreibung von Parameterinteraktionen und zur Bestimmung geeigneter Prozessbedingungen. In der modellgestützten Versuchsplanung werden diese Methoden kombiniert. Zunächst wird ein Prozessmodell erstellt, die Modellparameter werden bestimmt und der vorgeschlagene Versuchsplan wird simuliert. Nach der Auswertung der vorhergesagten Antworten wird nur eine reduzierte Anzahl von Faktorkombinationen, übereinstimmend mit einem vordefinierten Ziel, experimentell durchgeführt.

In dieser Arbeit wird die modellgestützte Versuchsplanung für die Auslegung und Optimierung einer kontinuierlichen Cytidin-5'-Triphosphat-Regenerationsreaktion durch Polyphosphatkinase unter hohem hydrostatischem Druck diskutiert. Die Prozessentwicklung war besonders komplex, da der verwendete Festbettbioreaktor und die enzymatische Reaktion nur teilweise in der Literatur beschrieben waren. Ein Prozessmodell wurde formuliert, um die zeit- und ortsabhängigen Änderungen von Substraten und Produkt sowie den Einfluss der Prozessbedingungen auf die Enzymkinetik zu beschreiben. Die Formulierung des Prozessmodells erwies sich aufgrund der Verwendung von Enzymimmobilisat und der Berücksichtigung disperser Strömung im Reaktor als umfangreich. Außerdem wurde ein geeigneter Modellierungsansatz formuliert, um den Druckeinfluss auf die enzymatische Reaktion zu berücksichtigen. Zur Reduzierung der Anzahl zeitintensiver Einzeloptimierungsstudien, sollten die Prozessparameter und ihre Wechselwirkungen mit Hilfe der modellgestützten Versuchsplanung optimiert und beschrieben werden.

Durch die Integration der vorgeschlagenen Methodik in einen benutzerfreundlichen iterativen Workflow wurde der Entwicklungsprozess durchgängig begleitet und das Prozessverständnis erhöht. Darüber hinaus konnte der experimentelle Aufwand reduziert werden und die Prozessbedingungen wurden nach drei Iterationsrunden im Hinblick auf die maximale Cytidin-5'-Triphosphat-Konzentration optimiert. Insgesamt wurde die Anwendbarkeit des wissensbasierten mDoE für ein komplexes biokatalytisches Reaktionssystem mit wenig Vorwissen gezeigt und nachgewiesen. Die Übertragbarkeit und weitere Verwendung der Methodik werden am Ende der Arbeit diskutiert.

Contents

List of Figures

List of Tables

Abbreviations, constants und symbols

Abbreviations

ANN	*Artificial neural network* [-]
ANOVA	*Analysis of variance* [-]
BBD	*Box-Behnken design* [-]
CCD	*Central composite design* [-]
CDP	*Cytidine-5'-diphosphate* [-]
CFD	*Computational fluid dynamics* [-]
CI	*Confidence interval* [-]
CQA	*Critical quality attribute* [-]
CSTR	*Continuously stirred tank reactor* [-]
CTP	*Cytidine-5'-triphosphate* [-]
DoE	*Design of Experiments* [-]
DT	*Digital twin* [-]
FDM	*Finite difference method* [-]
GSA	*Global sensitivity analysis* [-]
LHS	*Latin hypercube sampling* [-]
LHSD	*Latin hypercube sampling design* [-]
LSA	*Local sensitivity analysis* [-]
mDoE	*Model-assisted Design of Experiments* [-]
MOL	*Method of lines approach* [-]
MPC	*Model-predictive controller* [-]
MSE	*Mean square error* [-]
NMP	*Nucleoside monophosphate* [-]
NRMSD	*Normalized root mean square deviation* [-]
NTP	*Nucleoside triphosphate* [-]
PAT	*Process analytical technology* [-]
PB	*Plackett-Burman design* [-]
PDE	*Partial differential equation* [-]
PPDE	*Parabolic partial differential equation* [-]

PPK	*Polyphosphate kinase* [-]
polyP	*Polyphosphate* [-]
RMSD	*Root mean square deviation* [-]
RSM	*Response surface methodology* [-]
SRC	*Standardized regression coefficient* [-]
STR	*Stirred tank reactor* [-]
STY	*Space-time yield* [-]
TC	*Termination criterion* [-]
UHPLC	*Ultra high performance liquid chromatography* [-]

Constants and symbols

β_0	Intercept regression coefficient [-]
β_i	Linear regression coefficient [-]
β_ii	Quadratic regression coefficient [-]
β_ij	Interaction regression coefficient [-]
$\hat{\beta}_\mathrm{i}$	Estimated linear regression coefficient [-]
$\hat{\beta}_\mathrm{ii}$	Estimated quadratic regression coefficient [-]
$\hat{\beta}_\mathrm{ij}$	Estimated interaction regression coefficient [-]
β'_i	Linear standardized regression coefficient [-]
$\beta'_\mathrm{i,i}$	Quadratic standardized regression coefficient [-]
$\beta'_\mathrm{i,j}$	Interaction standardized regression coefficient [-]
ϵ	Error term regression model [-]
σ	Standard deviation [mmol L^{-1}]
$\hat{\sigma}$	Standard deviation model parameter sample [-]
σ^2	Variance [-]
σ^2_θ	Dimensionless variance [-]
τ	Hydrodynamic residence time [min]
Δt	Time difference [min]
θ	Dimensionless time [-]
Bo	Bodenstein number [-]
c_S	Substrate concentration [mmol L^{-1}]
d_i	Desirability [-]

D	Overall desirability [-]
D_{ax}	Axial dispersion coefficient [dm^2 min^{-1}]
e	Reactor bulk density [-]
$E\,(t)$	Exit age distribution [-]
f_{Dsig}	Normalized double sigmoidal function [-]
H_0	Test hypothesis [-]
i	Workflow round [-]
k	Number of factor reduction [-]
$K_{\mathrm{I,polyP}}$	Inhibition constant polyphosphate [mmol L^{-1}]
$K_{\mathrm{M,S}}$	Michaelis-Menten constant [mmol L^{-1}]
K_{sl}	Slope of double sigmoidal function [-]
L	Reactor length [dm]
L_{i}	Lower acceptable response [mmol L^{-1}]
L_{step}	Step length MOL [dm]
n	Number of grid nodes and factors [-]
p	Significance factor [-]
Q^2	Goodness of prediction [-]
R^2	Coefficient of determination [-]
R^2_{adj}	Adjusted coefficient of determination [-]
t	Workflow test round, observation [-]
$\bar{t}$	Mean residence time [min]
u	Flow velocity [dm min^{-1}]
U_{i}	Upper acceptable response [mmol L^{-1}]
u_{xx}	Partial differential equation [-]
v	Reaction rate [min^{-1}]
$v_{\mathrm{i,d}}$	Expected process variability [-]
V	Reactor volume [dm^3]
$\dot{V}$	Flow rate [dm^3 min^{-1}]
$\overline{v}_{\mathrm{i}}$	Process variability [mmol L^{-1}]
v_{max}	Maximum reaction rate [min^{-1}]]
$v_{\mathrm{p,Dsig}}$	Normalized reaction rate depending on hydr. pressure [-]
v_{red}	Normalized reduction of reaction rate [-]
w	Weighting constant [-]
x	Location index [dm]

x'	Standardized value [-]
x_{centered}	Centered model parameter value [-]
x_{Dsig}	Normalized parameter double sigmoidal function [-]
x_{i}	Factor level factor one [mmol L^{-1}]
$x_{\text{i,sample}}$	Model parameter sample value [-]
$y_{\text{i,t}}$	Independent variable value [-]
x_{j}	Factor level factor two [mmol L^{-1}]
$\overline{x}_{\text{i}}$	Factor level mean value [mmol L^{-1}]
$X_{50,\text{l}}$	Location parameter double sigmoidal function (low side) [-]
$X_{50,\text{h}}$	Location parameter double sigmoidal function (high side) [-]
$\overline{x}_{\text{sample}}$	Mean model parameter sample value [-]
y'	Standardized value [-]
y_{i}	Response [mmol L^{-1}]
$y_{\text{i,m}}$	Experimental response value [mmol L^{-1}]
$\hat{y}_{\text{i}}$	Estimated response [mmol L^{-1}]
$\overline{y}_{\text{i}}$	Expected, mean response [mmol L^{-1}]
Y_{h}	Function value at high x_{Dsig} [-]
Y_{l}	Function value at low x_{Dsig} [-]
Y_{mid}	Mid value $X_{50,\text{l}}$ and $X_{50,\text{h}}$ [-]
$y_{\text{s,m}}$	Simulated response value [mmol L^{-1}]
y_{t}	Dependent variable value [-]
$Y_{\text{x,t}}$	Standard regression model [-]

1 Introduction and motivation

Efforts have been made in the fine chemical industry to integrate [1] or replace chemical synthesis by biotransformations using enzymes as biocatalysts [2–4]. The use of biocatalysts has increased continuously since the 1960's, and in 2017, the global market for industrial enzymes was valued at $7 billion dollars and expected to rise to $10.5 billion dollars by 2024 [5]. Beside the low environmental impact [6], biocatalysts offer high specific activity, including substrate-, stereo-, regio- and chemoselectivity [7–13]. However, compared to chemical catalysts, biocatalysts in their native state possess relatively low stability and activity under harsh process conditions [6, 8, 14, 15]. In particular, a better understanding of enzyme stability is crucial for the design of competitive biotechnological processes [16, 17], but the selection of ideal reaction conditions is challenging due to a variety of influencing factors [18]. The scope of the process development is further increased if the enzymatic activity is not only dependent on pH [5, 16], temperature [5, 16, 19] or reaction medium [16] but also on the presence of cofactors [20]. Cofactors, often small organic molecules, metal ions and metal organic complexes, are needed to accomplish enzymatic activity and new cofactor-dependent reactions are still continuously discovered [20]. Among others, nicotinamide cofactors and nucleoside triphosphates, essential in biological systems, are used in a vast number of enzymatic reactions [20, 21]. As these cofactors can provide functional groups transferred to the substrate, stoichiometric amounts are required during catalysis [21]. Therefore, due to the costs and in the interest of an economical process, *in situ* regeneration of cofactors is proposed [21]. Together with the practical challenges associated with the use of enzymes, such as recovery or reusability [22], and cofactor-dependency, there is a need to develop reliable cofactor regeneration systems [21].

New developments in biocatalysis include enzyme design and protein engineering to tailor enzymes to specific process requirements and to improve their properties, e.g., through directed evolution [23–26], and multi-step catalysis to decrease the number of isolation steps and improve the process sustainability [25, 26]. Furthermore, flow biocatalysis is increasingly used to reduce the reactor footprint and achieve a better reactor control [26]. In flow biocatalysis, enzymes are usually entrapped via immobilization in or on a support material and the flow is adjusted to achieve a defined residence time [26]. The need for enzyme immobilization has also led to new developments in immobilizate materials, paving the way for enhanced enzyme and process performance as well as reduced process costs [27, 28].

However, these methods are only used to improve or modify the enzyme function before the enzyme is applied in a reaction and cannot be adapted during the actual reaction process. In contrast, the application of high hydrostatic pressure to enzymatic reactions represents a more process-related improvement method, derived from applications in the food industry, that has gained increasing attention in recent years [29]. On the one hand, high hydrostatic pressure can be used to stabilize enzymes against heat-induced structural changes. On the other hand, the thermodynamic equilibrium of the reaction, internal interactions or the reaction kinetics are affected. [29, 30] However, a comprehensive description of the influence of high pressures and the use of high pressure as a possible complementary process parameter has been sparsely reported in the literature. Additional methods are needed to capture the resulting process dynamics and make sound decisions in process design and optimization.

Today, mathematical process models are widely used to develop, optimize and validate bioprocesses in the fields of fermentation [31–33] and cell culture [34–36], but only to a limited extent for biocatalytic processes [1, 37]. They include state-of-the-art process knowledge [31, 38] and provide insight into the most relevant variables and process parameters as well as their impact on model outputs, also called responses [31, 39, 40]. In addition, analogies are used in modeling to understand complex systems and support decision-making [41]. Therefore, with respect to process design, process models can be seen as "an initial starting point to obtain a deeper process understanding during the bioprocess life cycle" [42]. In process development, mathematical models can also be used to predict various process scenarios [43] and suggest non-intuitive operating conditions [44], which may be potentially hidden by the complexity of the system and address the empirical and challenging procedure of traditional process development and optimization approaches [44]. Nevertheless, the use of process models is still heavily dependent on the quality and reliability of collected process data and predictions [40, 45].

Design of Experiments (DoE) is a suitable tool for systematically collecting experimental data and assessing the interaction of process variables and responses [38, 45–48]. Particularly if prior knowledge about the investigated system is not available, DoE methods are useful to explore the defined design space [44]. Unlike one-factor-at-a-time (OFAT) approaches, multiple variables, also called factors, can be examined simultaneously, and process knowledge is maximized with a minimum use of resources [45, 49–51]. Therefore, interactions between input factors and output responses are revealed [52]. The number of factors and factor boundary conditions must be defined beforehand and carefully selected [42, 53]. In addition, DoE methods also enable efficient planning, exploration and delimitation of the design space

[42, 48–50, 54]. The gained experimental results can be then used to define an optimal experimental set-point based on the previously defined factors. Despite the applicability of high-throughput technologies, the main disadvantage of common DoE methods, which are often used in data-driven optimization approaches [41, 55, 56], is the high workload particularly in the early stages of process development due to the complexity and sensitivity of factors [33, 45, 49, 57]. Furthermore, the user-defined boundary conditions might lead to heuristic restrictions and reduced gain in process knowledge [42, 45, 57].

Innovative model-based approaches such as the model-assisted DoE (mDoE) methodology, which combines traditional DoE methods with mathematical process models, can be used to overcome the specific drawbacks of data quality dependency and high workload of purely experimental DoEs. In addition to process modeling and DoE, the mDoE methodology includes the consideration of input uncertainties, e.g., experimental variations. Uncertainty propagation and additional process knowledge is quantified and gained. [35] As a result, the mDoE methodology meets the demand for more model-based tools to reduce the experimental effort and improve the process understanding as stated in literature [35, 44, 45, 58, 59]. In contrast to *a priori* modeling, non-linear model predictive controller and the sole application of response surface methodology used for process design and optimization, the mDoE methodology is based and relies on the systematic and efficient nature of DoE methods [35]. Today, the application of the mDoE methodology is mainly limited to the cultivation of mammalian cells, such as Chinese hamster ovary [45] and human mesenchymal stem cells [60], and microbial fermentation, e.g., with *Saccharomyces cerevisiae* [57]. Since biocatalytic processes are also highly complex and model-based or model-assisted tools are rarely used, the mDoE methodology could be the starting point for further process improvement and understanding alongside traditional optimization methods.

2 Aim of the thesis

To establish sustainable and efficient biocatalytic processes that can compete with chemical synthesis processes, more systematic and comprehensive design and optimization approaches need to be developed. These approaches should ensure transferability to other biotechnological processes to avoid case-by-case decisions and to promote their development from early stage to the final process. [40] Traditional DoE methods used in early stage process development still rely on planning and executing a large number of experiments to explore a predefined design space [44]. Additional development tools such as response surface methodology (RSM) [44], process and kinetic modeling as well as sensitivity and uncertainty analysis are available, but not used in combination [40]. Also, the parallel implementation of design and optimization methods is not state of the art in biotechnology.

Recently, several authors suggested the use of the mDoE methodology, which combines different development tools with varying degrees of complexity [45, 57, 60]. The methodology was mostly related to cell culture [45, 60] and fermentation processes [57], but the application to biocatalytic processes, which differ from cell culture or fermentation in the description of their kinetics, is still open. Furthermore, the use of this methodology right at the start of the development process and integrated into an iterative and uncertainty-based workflow has not yet been proven. Therefore, this work focuses on the evaluation of the mDoE methodology as a comprehensive process design and optimization tool for the biocatalytic regeneration of cytidine-5'-triphosphate (CTP) using polyphosphate kinase and polyphosphate as phosphate donor. In addition, it should be assessed whether an accelerated and reduced design and optimization process can be achieved.

The regeneration reaction is particularly important because the CTP is consumed in a multi-enzyme cascade for the production of 3'-sialyllactose including three main reactions and four different enzymes [61]. By regenerating CTP with a polyphosphate kinase, the costly supply of reasonable starting concentrations or continuous addition of CTP is replaced by the addition of comparatively inexpensive polyphosphate salts and cytidine-5'-diphosphate (CDP). The design, implementation and optimization of the individual parts of the multi-enzyme cascade and the 3'-sialyllactose production process were part of an industrial cooperation with GALAB Laboratories GmbH (Hamburg, Germany) and the Institute of Technical Biocatalysis (Hamburg University of Technology).

The starting point of this work is to define the framework of an iterative workflow based on the requirements of the biocatalytic process. The iterative workflow and its individual modules should allow a model-based description of process variables and their interactions, experimental design and evaluation of the results using various statistical tools, and subsequent optimization with respect to a predefined objective. The iterative structure of the workflow should also allow the incorporation of gained experimental data and process understanding into the database and process model. All modules of the workflow will be evaluated and critically reviewed.

A fundamental part of the workflow is the formulation of an appropriate process model to describe the time- and location-dependent changes in substrate and product concentrations in a packed-bed reactor, as well as the influence of the process conditions on the enzyme kinetics. Initially, this process model should be able to reflect the process dynamics observed in preliminary experiments. The formulation of the process model is complicated by the novelty of the reactor design and the continuous reactor operation. In addition, the use of immobilized enzyme retained in the reactor and the need to account for possible disperse flow within the reactor packed-bed increases the level of model complexity. Furthermore, only limited information on the enzymatic reaction is available in the literature, and an adequate modeling approach is needed to describe the nonlinear influence of the hydrostatic pressure on the enzymatic reaction at the process level. The integration of this approach into the process model should enable the assessment of its suitability as an unconventional but complementary process parameter. However, a trade-off between the process model complexity and preservation of the computational performance has to be found as the process model parameters are repeatedly adapted to consider experimental uncertainty.

After formulation of the process model, the process parameters and their interactions are to be described and optimized simultaneously using the actual mDoE methodology. Process modeling, DoE, RSM and statistical tools are combined to reduce the number of time-consuming individual reactor characterization and optimization studies, and accelerate the development process. The final evaluation of the mDoE methodology is defined as the maximization of the final CTP concentration reached at the end of the reactor bed. Nevertheless, the mDoE methodology should also allow the utilization of different optimization objectives such as the space-time yield (STY).

In the following paragraphs, current challenges in the development of biocatalytic processes and the fundamentals of biocatalytic reaction systems as well as process control strategies

will be discussed, representing the basis for the process model formulation. Furthermore, the essential parts of the mDoE methodology such as DoE and response surface methodology will be addressed to complete the iterative and uncertainty-based workflow for process design and optimization. Subsequently, the procedure, advantages and challenges of the workflow used for biocatalytic reaction systems will be elaborated in detail. Overall, the establishment of the mDoE methodology as a fundamental design and optimization tool alongside other optimization methods should enable the development of a variety of promising biocatalytic reactions that have not yet been fully developed or are still under investigation [40]. As a result, more economically competitive processes with an improved process understanding can be realized.

3 State of the art

In the following, the current challenges in the development of biocatalytical processes are elaborated and the basic understanding of the influence of various process conditions on the activity and stability of enzymes is summarized. Furthermore, the principles of different operation modes and the use of enzymes in continuous operated reactors are discussed. In addition, the fundamentals of process modeling and experimental design, as well as their combined application in the model-assisted DoE methodology, which is essential for process optimization and understanding, are described.

3.1 Current challenges in the development of biocatalytical processes

Industrial biotechnology is regarded as an alternative and sustainable approach that offers the possibility of replacing petroleum-based materials with renewable raw materials in the manufacture of chemicals [62]. Despite the increasing interest and practical relevance, the development of industrial bioprocesses faces a variety of challenges, e.g., high development costs and times [58], comparably sensitive processes in terms of acceptable reaction conditions and product concentrations [63] and the lack of systematic process design approaches[40], which leads to slower development as expected [63–66]. In addition, due to the complexity and variability inherent in bioprocesses the digitalization and automatization process is more complicated than in conventional manufacturing processes [67].

Biocatalysts used in biotechnological processes offer advantageous features including high substrate selectivity, chemoselectivity and the ability to carry out reactions at ambient pressure and temperature with a wide range of substrates [8, 12, 65, 66, 68–71]. The biocatalyst itself can be produced from biodegradable or renewable resources and is considered non-toxic [70]. However, many biocatalysts are comparatively weak and not fully functional or stable under harsh process conditions that can be encountered when they are used as replacement for conventional catalysts in chemical reactions. Consequently, the biotechnological products produced are less competitive than those from chemical processing, which makes them more difficult to market. [63, 72–74] This fact is further reinforced by low oil prices and expanding shale gas capacities, which stimulated new investments in chemical industry [62]. As a result, only low volume and high price products, such as chiral monomers, drugs and other products for biomedical and pharmaceutical applications, are generally more competitive because the

high prices compensate for low efficiency and production costs [63, 66, 75].

To overcome the current challenges, among others, genetic [76–78] and protein engineering [79], application of new reactor concepts [62, 80], enzyme immobilization [73] as well as statistical and process modeling [81, 82] have been implemented in the last years to create more efficient and sustainable systems [63]. In addition, high-throughput technologies [83–85] and Digital Twins (DT) are used in enzyme screening as well as process design and optimization [86]. Using high-throughput devices combined with sample handling stations, high load of experimental data can be produced with reduced material and personnel costs [84]. DTs in particular represent *in silico* alternatives to purely experimental-driven approaches by using mathematical process models [42]. It is to be expected that DTs will be become a more important and valuable optimization method in many areas of application [86], but the widespread use of DTs in biocatalysis is still to come today.

The judicious selection of available methods could facilitate efficient design and optimization, gain process knowledge and understanding throughout the process, and improve the process performance of existing or new processes [62, 73]. As a result, refined and efficient enzymatic processes will be more widely applied in industry, reducing the gap between biocatalysts and chemical catalysts [75].

One enzyme class for which improving process performance and efficiency is particularly important in large-scale industrial applications are transferases (enzyme class two) [87], including, among others, phosphorylases, transaminases and kinases [88]. Especially kinases, have been subject of various optimization studies [89–92] and are of interest for many biotransformations, e.g., for the production of amino acids [93], terpenoids [94] and 5'-nucleotides [95, 96] in food, flavor and pharmaceutical industry [93, 96]. However, the yield of transferases can be moderate, e.g., in the production of optically pure stereoisomers [93]. Furthermore, kinases, that can use inexpensive substrates as phosphate donors in 5'-nucleotide regenerations, potentially face substrate inhibition by polyphosphates and are often dependent on the presence of certain magnesia ion concentrations [96, 97]. Therefore, despite the promising applicability of these biocatalysts in industry, the detailed influence of the process conditions on the biocatalyst has to be investigated in more detail and summarized in a design and optimization framework.

3.2 Influence of reaction and process conditions on enzymes

To enable the use of enzymes in industrial scale, their catalytic properties have to be reviewed and improved [15]. A variety of different improvement methods are available today to modify and improve enzymes and process conditions.

3.2.1 Activation and stabilization of enzymes

In biocatalytic processes, high enzymatic activity, stability and stereoselectivity are sought [98, 99]. However, enzyme activation and deactivation as well as the enzyme stability are strongly influenced by the applied reaction conditions and physical environment [14, 99]. The influence of pH, temperature, salts and reaction medium has been comprehensively discussed in the literature [16, 100, 101]. In the last decade, the effect of hydrostatic pressure on proteins in microbial, biochemical and biotechnological processes [102, 103] and in particular on enzyme stability and activity has been discovered and investigated [8]. On the one hand, hydrostatic pressure showed a proven impact on enzyme activity due to changes in, i.e., the enzyme conformation, solvation, chemical equilibrium, and on the immobilization support and immobilization bonding [8]. On the other hand, it is presumed that the pressure-induced stabilization effect appears due to, i.e., intramolecular interactions, hydration of charged groups and stabilization of hydrogen bonds [8]. At relatively low hydrostatic pressures of up to 2000 bar and at room temperature, hydrostatic pressure only impacts relatively weak chemical bonds, i.e., hydrogen, hydrophobic and ionic bonds [103]. The enzyme stability, activity and structure itself, the physical solvent properties (e.g., pH, density, viscosity) that impacts the enzyme or the rate-limiting step are influenced by hydrostatic pressure [3, 8]. Based on the LeChatelier's principle, enzymatic reactions with a change in the reaction volume are affected. More specifically, the enzymatic reaction shows a dependence on the magnitude and sign of the change in reaction volume. [8, 18, 103–107]. Hydrostatic pressure have been also used in biocatalysis to alter the reaction conditions. As shown by Bolivar et al. (2019), gas-liquid transfer limitation were overcome by applying hydrostatic pressures of up to 34 bar to increase the solubility of oxygen in the reaction medium. [108] Due to the decreasing price of high hydrostatic pressure processing and the beneficial influence on biocatalysts [8], it is to be expected that the application of high hydrostatic pressure will find its way into many areas of biocatalysis as an independent or complementary process parameter. However, enzyme-specific hydrostatic pressure-dependencies and interactions with other influencing process conditions have to be investigated.

3.2.2 Effect of hydrostatic pressure on transferases

In the past years, transferases (enzyme class 2) have been used for the generation of different drug candidates and metabolites. They are capable of transferring a variety of functional groups such as formyl or phosphate groups in the case of polyphosphate kinases (PPKs) [109–112]. The transfer of phosphate groups is of particular interest as different nucleoside triphosphates (NTPs), e.g., adensoine-5'-triphosphate (ATP) and guanosine-5'-triphosphate (GTP), can be regenerated, which are essential for almost all biosynthetic activities of cellular life and common enzyme cofactors [96, 111, 113]. PPKs are able to use inexpensive and stable polyphosphate salts as the phosphate donors and are therefore considered promising candidates to meet the demand for industrial enzyme cofactor regeneration [96]. As described by Tavanti et al. (2021), established NTP regeneration systems based on other kinases are hampered by cost, availability of the phosphate donor and the need for enzyme cascades to regenerate NTPs from nucleoside monophosphates (NMPs) [96]. The regeneration of additional NTPs such as CTP is also possible due to the broad substrate specificity of PPKs [114]. Compared to hydrolases (enzyme class 3) [73, 104, 115, 116], for which the influence of hydrostatic pressure has already been studied, the investigation to describe the influence on transferases, in particular on PPKs, has not yet been described in the literature.
Recently, Schmalle and Arndt et al. (2024) published first findings on the influence of hydrostatic pressure on the stability and activity of a PPK from *Ruegeria pomeroyi* (RpPPK2-3). In a hydrostatic pressure range of 0 to 800 bar, the specific activity initially increased with increasing pressure and then decreased at 800 bar. The results and knowledge gained were then used to increase the final product concentration in a cofactor regeneration process. [117] This study demonstrated that a systematic and efficient process development is useful when investigating novel enzyme activity and stability improvement techniques.

3.3 Bioreactor operation modes

The choice of reaction system and mode of operation is crucial in the design and optimization an effective and economical bioprocess. Based on the selection, considerably different biological performance, control effort and costs can be expected. [118, 119] Generally, process strategies in the field of biotechnology can be distinguished into discontinuous, i.e., batch and fed-batch, and continuous operation mode [80, 120, 121].

3.3.1 Batch operation mode

In batch operation mode, the substrate concentration decreases over time and the product is recovered after the reaction is complete. Generally, the composition and conversion of the reactor content is defined as a function of time and full conversion is often achieved after sufficient time if a favorable equilibrium exists. [40, 74, 122] Despite the ease of operation and robustness, disadvantages of batch operation are the low volumetric productivities and labor intensity [123]. Today, a variety of enzymatic reactions are still performed in batch operation mode [124–126].

Stirred tank reactors (STR) are often operated in batch operation mode. Ideal hydrodynamics in terms of mixing with uniform temperature and composition profiles are generally assumed. [40, 74, 118] STRs in batch mode are mainly used for liquid phase reactions and regarded as simple, reliable and flexible. They are extensively used in biotechnological and pharmaceutical industry at a small scale for the production of several products with the same reactor equipment [118]. In addition, batch operation in STRs is favored to avoid contamination and fouling due to simple cleaning and sanitation procedures. Furthermore, the capital investment for construction, control and instrumentation is relatively low compared to reactors operated in continuous operation mode. However, batch operation in STRs can be associated with high material handling costs and should not be used if the reaction is strongly inhibited by the initial substrate concentration or if rather unstable products are formed. In these cases, fed-batch (semi-batch) operation in which additional reactants are gradually added to the batch operated reactor over time, or continuous operation can be considered, respectively. [40, 80, 118, 120, 122, 127].

3.3.2 Continuous operation mode

Continuously operated reactors are considered to be “more efficient and sustainable” compared to reactors operated in batch mode [80]. In the continuous operation, the reactor is continuously supplied with material, e.g., substrates. At the same time, to maintain a constant volume level and working volume, an equal quantity of reactor content is discarded [118, 122, 128]. In contrast to the batch operation mode, the costs per product unit are lower and the reaction conditions can be better controlled, resulting in high process stability, but a thorough understanding of chemical and engineering aspects is crucial, particularly when transferring a process from batch to continuous operation mode [128]. With regard to the reaction performed, continuous operation is favored if the reaction is inhibited by the substrate or if large production

capacities are involved [40, 118]. Therefore, also rather unstable products can be produced due to their low residence time and low accumulation of toxic or inhibitory byproducts in the reactor [119].

The continuously stirred tank reactor (CSTR) is a common type of continuously operated reactor. Similar to the STR, ideal mixing and homogeneous distribution of the reactor contents can be assumed. Since effluent is continuously discharged at the reactor outlet, the operating conditions must be well controlled to minimize substrate loss due to incomplete reaction, and (bio-)catalysts must be retained within the reactor. [122] In addition to the CSTR, tubular plug flow reactors can be used in continuous operation [129]. Under ideal and steady-state conditions, a flat velocity profile with no axial diffusion or back-mixing is assumed [74, 118, 128]. Therefore, the concentration profile over the length of the reactor is the same as for a batch operation over time [122]. In biocatalysis, this reactor type can be used as an alternative to a CSTR, especially to determine specific characteristics and parameters such as diffusion limitation, turnover number, retention time, leaching and enzyme deactivation [40, 74].

3.4 Continuously operated packed-bed reactor

The packed-bed reactor is a common type of continuously operated reactor, often used with heterogeneous catalysts in chemical transformations and biotechnology, paving the way for process intensification [74, 128, 130, 131]. The catalyst material, among others glass, polymers or stainless steel in the form of columns or cartridges, is embedded and retained within the reactor [72, 128, 132]. As a result, subsequent separation steps can be omitted and maintenance is reduced leading to lower overall production costs [128, 130]. Biocatalysis carried out in continuous packed-bed reactors can be more productive and sustainable compared to batch processes [133]. Furthermore, enzymes can be reused if they are used in immobilized form [80, 134]. Among others, this was shown for immobilized lipase by Chen et al. (2011) [135] and for invertase by Albertini et al. (2012) [136]. However, leaching of the biocatalyst can occur and the conversion strongly depends on the catalyst particle size. Larger particles possess a relatively low volume-to-surface ratio which leads to reduced reaction rates at the catalyst surface. Small particles potentially lead to high back pressures and clogging of the filter units at the end of the reactor column. [128] The usage of porous particles can compensate for the smaller volume-to-surface ratio of larger particles [137].

An advantage over STRs is the reduction of shear forces in packed-bed reactors and the higher loading of biocatalysts, i.e., biocatalyst surface area per unit of reaction volume [72],

resulting in an increased productivity [122, 130, 131]. Furthermore, it has been shown that hydrodynamic limitations due to high viscosity can be reduced and comparably high substrate conversion rates as well as operational stability are achieved [72]. However, packed-bed compaction and flow channelling potentially limiting the mass and heat transfer are the major disadvantages of packed-bed reactors [72]. To compare the conversion of the catalyzed reaction in continuous and batch reactors, it is useful to express the conversion in terms of productivity or space-time yield, as the conversion is determined either under stable process conditions at the steady state or for the entire process time including the start-up phase [128]. When working with a packed-bed reactor, a number of operational variables are of interest, including residence time, flow behavior and diameter-to-height ratio [80]. The time between start and termination of a continuous process and the position where constant substrate or product concentrations are reached can be defined as the residence time [128]. By manipulating the residence time, the conversion is increased, generally at the cost of lower productivity [80]. In addition to the residence time, the residence time distribution is considered to be important especially if the reaction volume is smaller than the actual reactor volume. It is defined as "the probability distribution of time that solid or fluid materials stay inside one or more unit operations in a continuous flow" [138]. In packed-bed reactors, a laminar or parabolic flow profile rather than an ideal plug flow can be observed, which results from axial convection and radial diffusion as a consequence of the use of mechanical pumps and a strong pressure drop along the column wall. [128, 139] This velocity profile leads to sample dispersion along the path of the fluid, i.e., not all fluid elements entering the reactor leave the reactor outlet at the same time, and to the formation of a residence time distribution [128]. Furthermore, the diameter-to-height ratio can affect the overall conversion and productivity as smaller reactor diameter increases the mass and heat transfer. In order to compensate for the small diameter and to be able to use larger quantities of biocatalysts as required in industrial applications, long reactor columns are desirable. [80] Finally, due to the use of different biocatalysts, immobilizates, reaction media and flow conditions, compromises must be made on a case-by-case basis.

3.5 Process modeling

Mathematical process modeling is an essential building block to overcome current challenges in biotechnology and to enable knowledge-based process design and optimization. Today, new modeling concepts can be used to simulate and predict process dynamics in different

bioreactor configurations and modes of operation and are considered key to the digitalization of biotechnology. [40, 44, 67, 140]

3.5.1 Fundamentals of process modeling

Mathematical process models, including kinetic models, are increasingly used to efficiently predict a bioprocess under different reaction conditions and to evaluate the process operability, limitations and robustness [32, 39, 40, 58, 65, 141, 142]. The increased interest in process models and simulation also originates from growing manufacturing costs, shorter development times required and the need to eliminate process inefficiencies [58]. They can not only be seen as a quantitative description of natural phenomena [143] using analogies [41], but also as tool for decision-making[58] and reaction engineering [82]. Generally, process models are considered as "an initial starting point to obtain a deeper process understanding during the bioprocess life cycle" [42]. In biotechnology, process modeling has been used to describe the dynamics of cell culture [35, 144–146], microalgae [142, 147, 148], fermentation [32, 57] and biocatalytic reaction processes [149, 150].

A process model should not only describe different types of bioprocesses, but ideally also be applicable to different bioreactor configurations and scales [42]. Furthermore, appropriate mathematical bioprocess models not only form the basis for the description and efficiency of bioprocesses, but also enable the application of model-based control strategies and the improvement of process knowledge [38, 58, 147, 151–153]. Among others, the characteristics of complex interactions occurring in bioprocesses can be studied and biological information can be stored as equations [34]. Despite the wide range of applications, mathematical process models fundamentally require high quality information and knowledge about the input, output and their relationships as well as internal variables of the bioprocess [39, 58, 65, 154]. Furthermore, the dynamics, the high complexity resulting from the non-linear biological behavior, bioprocess variability and limited amount of high-quality data, can complicate the modeling process and the application of model-based methods [147, 148, 152, 153, 155, 156]. As a result, an adequate compromise must be found between the complexity of the process model, the adaptability, predictive capability and the experimental effort, e.g., for estimating the model parameters, particularly in the beginning of model and process development [42].

A special application of mathematical process models is their use in DTs. DTs are considered a promising *in silico* alternative to purely experimental-driven approaches. [42] According to Barricelli et al. (2019), a DT is defined as a virtual machine "simulating, emulating, mirroring, or "twinning" the life of a physical entity, which may be an object, a process, a human, or a

human-related feature" [157]. In other words, DTs are virtual copies of processes and their inherent mathematical process models can vary in their requirements for expert knowledge, computational needs and applicability [67]. Today, DT-based concepts are already used in the manufacturing of biopharmaceuticals and other biotechnological fields [42, 157] to describe the mechanical, physical and (bio-)chemical structures of a bioprocess [42]. The predictions made with process modeling and knowledge-based methods which are part of DT also aimed at reducing resources in terms of experimental effort [67, 158]. Fundamental to the successful use and implementation of a DT is the clear definition of the objective to overcome the challenges of biotransformation development, data collection and processing [67].

3.5.2 Classification of mathematical process models

Before creating a process model, a number of aspects must be considered, including the biophysical and biochemical mechanisms, the components involved and their interactions, the mode of operation, the reactor design and the action of process control, but also the fundamental problem the model intends to address [34, 50, 65]. On the one hand, a process model can be classified as deterministic or stochastic. Compared to deterministic models which are based on experimental observations, knowledge about the structure of the reaction network and predict a single outcome for a given state, stochastic models take uncertainty and fluctuations inherent in the investigated systems into account and predict a range of possible outcomes. [34, 143, 159] This is done by implementation of probabilistic based variation of input variables. An advantage of determinsitic process models is their capability to be calibrated to a specific data set by adjusting biological and physical process parameters [160]. Disadvantages are the complex structure and high number of kinetic parameters and their interactions [161].

On the other hand, further classification can be made for mechanistic, based on deterministic principles, and empirical process models [31]. Mechanistic models rely on physical and chemical laws, considerable understanding of the modeled system and knowledge as well as relations are summarized in material and energy balances. Only relatively low amount of data are needed, but the set of equations can be complex. They are used to identify critical process variables and to plan experiments in microbial fermentation, biocatalysis and mammalian cell culture. [41, 42, 162, 163] Empirical models are derived from the analysis of data with less system knowledge and non-mechanistic parametric equations [36, 41]. They are also named black-box model [164]. The fundamental classification of black-, gray- and white-box models is comprehensively described in literature [34, 36, 165]. Using empirical models, input-output

relationships can be characterized even in complex systems, but many experimental data may be required to investigate relationships between model inputs and outputs. Today, empirical models are predominantly applied in the field of manufacturing and artificial neural networks and machine learning. [41, 50] Advantages and disadvantages of mechanistic and empirical process models are further discussed and summarized in literature [36, 42, 50].
The formulation of the existing equations as dynamic differential equations or time-independent static equations is decisive for the subsequent usability of the process model. Dynamic process models determine the progression of the represented process over a predetermined time horizon and not only for a certain time point, e.g, steady-state. Disadvantages of dynamic models are the higher complexity, i.e., high nonlinear terms and large number of model parameters, and the resulting computational demand. [34, 166] Nevertheless, their are widely used in many areas of application today.
Recently, new process modeling concepts have been proposed. In the hybrid modeling approach, the advantages of different process model classes, e.g., mechanistic and empirical models, are combined. The basis of the hybrid model is often a mechanistic model describing the phenomena observed in the bioprocess. The proportion that cannot be expressed by the mechanistic model using known kinetic equations or balance equation, is then described by the hybrid model counterpart, e.g., data-driven machine learning or artificial neural networks. These hybrid models can be used for accurate online monitoring, bioprocess prediction and optimization. [36, 41, 42, 148, 167] Despite the variety of process models that are available today, the selection of an appropriate process model can be difficult. In addition, the combination of different modeling techniques and corresponding evaluation methods in a systematic framework do not exist.

3.5.3 Kinetic models

Sin et al. (2009) proposed the distinction of models used in the field of biocatalysis into four levels depending on their scope and purpose. In the first level, catalytic models are used to describe the catalysis at the molecular level and to predict reaction selectivity. The second level is covered by reaction models that incorporate kinetic models to describe mechanisms and intrinsic rates as a function of concentration, pH and temperature. The third level is represented by reactor models, which use of the reaction models and account for mass balances as well as hydrodynamic conditions. Process models are the top level and incorporate all other levels. They are frequently used as a complement to experimental studies and as an *in silico* tool to evaluate the overall process behavior and process performance. Due to this structure, the

complexity of the final process model increases with increasing detail of the individual levels. [37] As a consequence, simplified kinetic models, e.g., the Michaelis-Menten type kinetic reaction model, are widely used in the second level. [37, 168] These kinetic models can be often expressed by single rate equations and include a kinetic dependence from pH and temperature as well as substrate or product inhibition. [37]

Reaction and kinetic models are still predominantly used in biocatalysis, among others, to find optimal operation points and increase the process understanding [82]. Jurado et al. (2004) used different kinetic models to describe the enzymatic activity in dependence on pH and temperature as well as the resulting deactivation [16]. Vasic-Racki et al. (2003) comprehensively showed the benefits of kinetic models in process modeling for several enzymatic reactions performed in batch and continuously operated reactors also considering different types of inhibition, e.g., for the continuous (R)-mandelic acid production. [82] Reaction condition optimization was presented by Zhong et al. (2017) for the synthesis of cellobiose from sucrose using three thermophilic enzymes and a Michaelis-Menten-based kinetic model [169]. In addition, a xylan and cellulose hydrolysis process was modeled using a combined transport and kinetic model with further validation in biorefinery experiments by Prunescu and Sin (2013) [170]. Since the kinetic models are to be used not only for kinetic investigations, but also in combination with reactor models for process design and optimization, the resulting process variables and cause-effect relationships that can be observed outside the enzyme's microenvironment must be investigated.

3.6 Design of Experiments

In bioprocess development the performance of the process can be affected by numerous process variables and raw material properties [49]. As a result, these processes require effective statistical methods to generate and collect large amounts of information-rich data to draw sound conclusions in research. One solution is the use of statistically designed experiments. [50]

3.6.1 Design of Experiments: Fundamentals

The basics of statistical design of experiments have been known since the 1900s [52] and used in engineering, agriculture, medical and logical science [50, 51, 171]. In the past, however, Design of Experiments (DoE) has not been widely used in biological research and bioprocess engineering [50]. Due to the increasing complexity of processes and process variables [42], the

availability of mathematical process models, the emergence of process analytical technology (PAT) and process intensification as well as the need to improve the bioprocess efficiency, DoE methods are now being increasingly utilized in biotechnology applications [42, 50, 52]. Among others, DoE has been used for biotechnological unit operations[38, 49], media formulation [46, 52, 172], cell culture processes[45, 50], microbial fermentation [33, 173] and microalgae processes [47]. In biocatalysis, DoE has only been used to a comparatively small extent, for example, in the optimization of biocatalytic biodiesel production [174].

In general, DoE methods are based on sound and logical statistical principals [175] and represent systematic tools in bioprocess development to assess the interdependence of multiple process variables on outcomes, also called responses and factors, which may be critical quality attributes (CQA) of a process, e.g., the final product quality. In addition to conventional bioprocess design, DoE can be used in process validation with an optimized number of experiments [38, 42, 45–48, 53]. Unlike OFAT approaches, which only change one variable at a time, multiple factors are examined simultaneously by DoE [45, 46, 49] by a small set of carefully planned experiments [50]. As a result, the process knowledge and information are increased and maximized with a minimum use of resources, i.e., time, costs and material [50, 51, 175]. DoE methods also allow for an efficient planning, exploration and narrowing of the design space which is defined by the selected minimum and maximum factor values [42, 48–50, 54]. The design space is mainly based on extensive expert knowledge and findings from preliminary experiments and is referred to as the area in which product quality can be assured [42, 53, 176]. The number of factors and factor minimum and maximum values as well as the design plan and objective of the DoE itself are critical and have to be selected carefully beforehand [42, 53]. An inappropriate choice of the factor limits, leading to a design space that is too narrow or too wide, could reduce the usefulness of the DoE outcome [52]. In addition, in reality, DoE methods can lead to a huge process development effort with multiple rounds of time-consuming experimentation [42].

After exploring the design space and identification of the interdependence of factors and responses, the outcome of the experimental design can be used to create an empirical model. The empirical model correlates the responses to the systematically varied critical factors [33, 49–53]. Considering the challenges emerging using DoE methods, systematic approaches for the evaluation and selection of an optimal DoE design and the understanding how factors affect responses and process dynamics are essential today [33, 49].

3.6.2 Common DoE approaches

To overcome the limitations of OFAT approaches, the DoE method is used as an efficient tool for the evaluation of experiments and main effects on process outcomes [45, 49, 50, 177, 178]. To use DoE in process development or optimization, range-finding studies and process characterization have to be conducted first [38]. The factors to be investigated are often selected based on the experience of the researcher [175]. After the factors are defined, process relevant factors can be identified using a full factorial or fractional factorial design, and the experimental space is evaluated, also called initial screening [52]. More precisely, the impact of factors on responses is determined, [38, 45, 52, 179] which precedes the optimization of the bioprocess [180]. As described in literature, the Plackett-Burman (PB) design can be also used to screen and efficiently estimate linear effects between factors and the approximation of the local response surface generated by the resulting responses [179–181]. However, these experimental designs are not used for the subsequent optimization of bioprocesses [182].
For the full factorial design, 2^n trials are needed for n factors representing a complete design. The incomplete fractional factorial (2^{n-k}, reduced by k steps) and PB design require fewer trials. [52, 180, 181] It is assumed that main effects and interactions of one or more factors on the predefined response can be evaluated using the full factorial design [181]. For the PB design, it can be assumed that main effects are dominant and interaction effects (higher order effects) are neglectable [172]. Kalil et al. (2000) propose the use of the PB design for the identification of relevant factors in the beginning of optimization with subsequent use of factorial design and surface response analysis [180]. The special feature of the PB design is that the number of trials is a multiple of 4 and not a power of 2, as for fractional factorial design, to obtain an orthogonal main effect plan. In the PB design for n factors and two levels, the levels are either -1 or 1 (low and high factor value). [172, 175, 179]. However, the reduction of factor experiments in the fractional factorial and PB design can reduce the statistical quality. Once the factor combinations have been determined, it is suggested that the entire test series be carried out randomly. [52] To avoid areas of free design space and avoid bias, so called space-filling designs as Latin hypercube sampling design (LHSD) can be used. Compared to other DoE methods, the LHSD results in an evenly filled room with an even but random distribution of factor combinations[181], which, however, leads to a high number of trials.
As the bioprocess optimization process progresses, the maximization of the predefined response via an appropriate experimental design and RSM is suggested [48, 49, 179]. Compared to

screening approaches, more complex designs are often used in RSM to mathematically describe the influence of factors on responses. Various combinations of screening and optimization designs used in bioprocess optimization are summarized in the work of Kasemiire et al. (2021). [182] On the one hand, the selection of the optimization design depends on the nature of the study, i.e., optimization or robustness study, and the number of factors to be investigated and the interactions itself, e.g., linear interactions [48, 182]. On the other hand, the selection depends on available resources, i.e., cost, materials and time [48]. For first-order designs applied in RSM, e.g., a full factorial design, the first-order regression models represent a plane. These designs can only be used until the true local response surface exhibit greater curvature closer to the area of the peak. The curvature can only be represented by second- and higher-order designs and models. [179] Commonly used second-order designs for optimization are the I-optimal, D-optimal, Box-Behnken (BBD), Central-Composite (CCD) and space filling designs [49, 179]. The D-Optimal, BBD and CCD are widely used for RSM [49, 183]. The BBD, CCD and D-optimal designs for 3 factors and 3 levels are depicted in Figure 3.1 [149, 179, 184]:

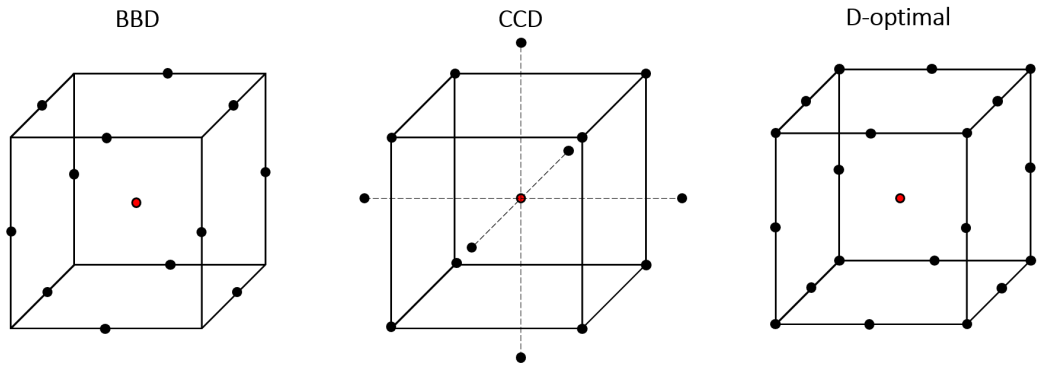

Figure 3.1: Left: Box-Behnken-Design. Mid: Central-Composite-Design with trial points outside the experimental space. Right: D-optimal design. Black points represent the experimental trial points and the red point represents the central point. Modified from Davim (2016) [185].

Mid-points of edges in the experimental space are considered in the BBD, whereas corner points and out-of-boundary points are omitted. This allows experimentation around extreme factor combinations. Compared to the CCD, fewer factor combinations (trial points) are received [42, 185, 186]. According to Annadurai et al. (1998), the BBD is preferable because a relatively low number of experimental combinations are needed to sufficiently estimate complex response functions and second-order response models [177, 187]. However, the BBD should be only used if the experimental boundaries are known [185]. CCD has already been

used in the investigation of biocatalytic reactions together with RSM [178, 188]. Depending on the purpose, the CCD can be favorable since the design covers a larger volume with five factor levels (built from 2^n design, center points with corner and star points) and can better capture strong curvature and cubic responses [52, 176]. Furthermore, compared to the BBD, the accuracy of the estimates over the design space is higher for CCD due to the location of the experimental factors. In addition, extreme regions at the corners are poorly estimated for the BBD [186]. Compared to the BBD and CCD, quantitative and qualitative factors can be included into the D-optimal design, and variance of the regression coefficients is minimized [183]. Using the D-optimal design, the D-criterion is used which is further described in literature [189, 190].

The experimental results obtained can then be used to predict the parameters of an empirical RSM, to describe the factor interactions on the responses and to define an optimal experimental factor combination. The main disadvantages of common DoE methods are still the high number of time-consuming experiments and the high costs, particularly in the early stages of process design due to the high complexity and sensitivity of factors, even when using high-throughput technologies. [33, 45, 49, 57] In addition, user or expert knowledge is critical and a prerequisite for determining an appropriate design space which can lead to multiple rounds of iteration and wrongly chosen limits in the design and optimization of a bioprocess, which results in a reduced increase in process knowledge [42, 45, 57].

3.6.3 DoE modeling

The data obtained from the DoE can be further analyzed by regression analysis and is represented by a linear or multiple-linear regression model for two or more independent and quantitative factors. Using the regression analysis, the relationship between independent factors and responses is described and illustrated as response surface [48, 49, 179].

Standard first-order designs for first-order regression models are the 2^n factorial experiment and the 2^{n-k} fractional factorial design [176, 179]. The standard first-order polynomial regression model for one factor and experimental combination, also called treatment combination x, can be formulated as a linear function [179]:

$$Y_{x,t} = \beta_0 + \beta_1 x_1 + \epsilon \tag{3.1}$$

with $Y_{x,t}$ as the t-th observation of treatment combination $x_{1,\ldots,n}$, β_0 as the intercept, β_1 as the regression coefficients describing the linear effect of x_1 . The ϵ represents all minor sources

of variation (random-error variable with $\epsilon \sim N\left(0, \sigma^2\right)$). For more factors n and treatment combinations $x_{1,...,n}$, the regression model can be written in a more general form:

$$Y_{x,t} = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \epsilon \tag{3.2}$$

The linear effects $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimated using least square estimates to fit the first-order model to collected data at different treatment combinations $x_{1,...,n}$. By this, the estimated mean response $\hat{y}$ can be formulated as a function of x_1 [179]:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 \tag{3.3}$$

$\hat{y}$, also called fitted model, enables the prediction of new response values for any treatment combination [179]. Similar to Equation (3.1), $\hat{y}$ for n factors can be formulated as:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_n x_n \tag{3.4}$$

First-order polynomial regression models can be further extended to higher-order models with two or more quantitative factors. This can be necessary in the case of a lack of fit of the first-order model and an inadequately approximated local response surface. Additional factor combinations should be collected. The lack of fit can be attributed to, e.g., two-factor interactions or quadratic effects which can be described by *p*-th-degree polynomial regression models. [176, 179] As a result, the resulting mean response is a function of additional variables and defines a response surface instead of a regression line in three dimensions [179].

3.7 Response surface methodology

The RSM is considered an extension of the standard regression analysis and effective tool commonly used in conjuncton with DoE to reduce the experimental effort and time in optimization approaches [149, 176, 178]. This methodology relies on the generated regression models and provides improved process understanding and information about the relationships between user-defined experimental factors and observed responses y_i in the experimental space [45, 60, 176, 185, 187]. The responses are affected by different levels of the defined factors, and an optimum factor combinations can be located using statistical tools [176, 178, 185]. A recognizable drawback of RSM is the “only one point in time” consideration in the development of the regression model. In most cases, the endpoints, also called endpoint description, are used as responses and the course of the process with possible deviations over

time is not assessed [167]. Therefore, the careful selection of sampling points or process termination is of great importance in RSM.

Regression models for RSM

To generate the response surface, the order of the regression is defined and the coefficients of an empirical response surface model, which describes the response surface in 1+p dimensions, are predicted first using the least-square method [179, 187]. The least-square method is comprehensively explained in Dean et al. (2017) [179]. Then, the mean responses $Y_{x,t}$ are plotted as a function of the factor combinations suggested by the design plan[179]. In addition, an optimal factor combination is predicted to minimize or maximize the response [45, 177, 187].

The standard second-order model $Y_{x,t}$ considering two-factor interactions and higher-order effects can be formulated as [179]:

$$Y_{x,t} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \beta_{ii} x_i^2 + \epsilon \tag{3.5}$$

$Y_{x,t}$ denotes the t-th response observed for all factor levels x_i. β_i, β_{ij} and β_{ii} are the linear, interaction (cross product coefficient) and the quadratic effects, respectively. In a multiple-linear regression model, interactions higher than the first order are neglected resulting in fewer experimental combinations needed to estimate the complex response function [177, 179].

For higher-order designs and models (order p), the polynomial standard form of the fitted regression model with the mean response $\hat{y}_{x,t}$ (general form see Equ. (3.4)) with treatment combinations $x_{1,\ldots,n}$ can be written as a p-th-degree polynomial regression model:

$$\hat{y}_{x,t} = \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_i x_i + \sum_{i<j} \hat{\beta}_{ij} x_i x_j + \sum_{i=1}^{n} \hat{\beta}_{ii} x_i^p + \epsilon \tag{3.6}$$

Usually, the exact form of the polynomial function is unknown but it is assumed that the true response curve is approximated by the regression model. According to Dean et al. (2017), it is recommended to use the simplest model that provides an adequate fit to prevent overfitting of the linear or polynomial model [179]. Overall, the three main purposes of regression models are [191]:

1. Definition of a relationship between $Y_{x,t}$ and the treatment combinations $x_{1,...,n}$ to predict process responses for a given set of factor values.

2. Determination of significant factors.

3. Determination of an optimum set point of factors to maximize or minimize the response for a predefined experimental space.

3.8 Model-assisted Design of Experiments

The systematically evaluation of different design plans and DoE modeling is needed to predict process responses and correlate them to experimental factors [49]. To overcome the still high workload associated with first- and second-order designs, the mDoE is proposed [45, 57, 60]. In the mDoE methodology, a mathematical process model is used to simulate factor combinations suggested by an experimental design plan in a predefined experimental space [45]. The experimental space reflects user-defined constraints and represents, among others, technical limitations [40]. The resulting responses are assessed and optimal process conditions according to the optimization objective are determined. By this, only favorable factor combinations are experimentally performed and undesirable factor combinations are ruled out in advance. [38, 45, 60]. In addition, the drawback of RSM, the “only one point in time” consideration and endpoint description, is overcome as various time- and endpoints of the DoE factor combinations can be simulated. Therefore, the mDoE methodology has not only descriptive but also predictive capabilities. Similar to the previously described DoE methods, prior knowledge is needed to define factors, set factor limits and select an appropriate regression model.
The general mDoE methodology for the simulation and evaluation of experimental designs can be divided into six parts and is shown in Figure 3.2.

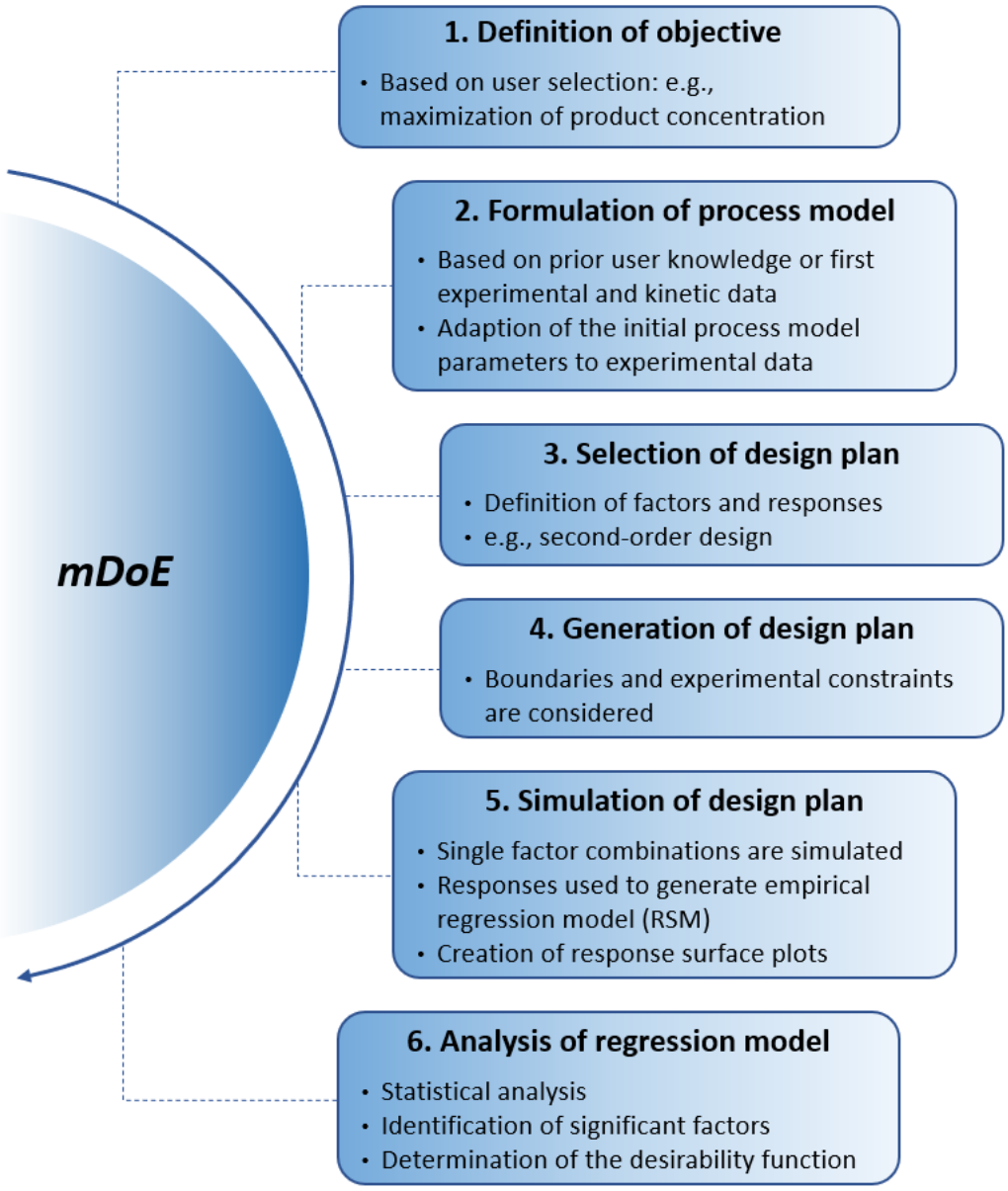

Figure 3.2: General structure of the model-assisted Design of Experiments methodology [45, 57, 60].

First, the objective of the study, i.e., the optimization target, is defined [57]. Second, the mathematical process model is formulated by prior knowledge or first-stage experimentation. According to Arndt et al. (2020) [144] and Moser et al. (2020) [57], model parameters that are part of the process model are repeatedly adapted to a low number of experimental data. Third, a first- or second-order design plan is selected depending on the expected structure and curvature of the response surface (RSM), and the factors as well as responses which are to be investigated are defined. Fourth, the experimental design plan is generated under consideration of the experimental space. [45] Fifth, the single factor combinations are simulated and responses are used to generate an empirical regression model and response surface plots. Therefore, the

simulated responses y_i are treated in the same way as experimental data. At the end of the mDoE methodology, the regression model and results are analyzed and the desirability function D can be calculated for y_i with user-defined constraints. The main advantage of the desirability function is the standardization of the multidimensional optimization problem to one function. [42, 45] The mDoE methodology enables the selection of an appropriate experimental design, the establishment of optimization criteria, the definition of a well-defined design space, and the evaluation of the importance of model parameters and factors before costly experiments are fully conducted [45, 57, 60]. However, the mDoE methodology represents only one part of the overall bioprocess design and optimization approach. Together with the preliminary analysis of the process, the building of the process model and the final evaluation of the results, a complementary approach is created. A continuous improvement of the predictive capabilities of the process model and the process understanding, would be achieved by integrating the mDoE methodology into an iterative workflow.

3.9 Alternative model-based approaches for the design of bioprocesses

In addition to the mDoE methodology, alternative approaches have been used in bioprocess design and optimization. In the following, *a priori* models and the use of model-predictive controller are discussed.

3.9.1 A priori modeling

A priori modeling is mainly based on knowledge gained from former experiments or literature to describe underlying phenomena or cause-effect relationships influencing the bioprocess dynamics [192]. In general, the collection of *a priori* knowledge represents the first step in building a model, however, in the field of bioprocess modeling, *a priori* biochemical knowledge is often incomplete and only the most relevant reactions are considered. This can lead the reduction of the overall complexity of the *a priori* modeling approach. [193] After the *a priori* model is formulated and the process is modeled, experiments are performed and evaluated. Crucial to the continuous improvement of bioprocess models and understanding is the extension of existing *a priori* knowledge through new insights and information gained by analyzing experimental or simulated data. The *a priori* knowledge is also used in hybrid models, e.g., to take intracellular reaction kinetics into account [194].
Another *a priori* approach is to use of the developed process model to generate a response surface without setting up an empirical regression model. The factor combinations suggested

by a DoE and needed for RSM application in mDoE could be extended by an arbitrary number of combinations in the experimental space. As a result, the entire experimental space is simulated and investigated, process dynamics are considered and rather extreme peaks or minima in the response surface can be detected. Furthermore, compared to conventional RSM, not only an endpoint analysis is conducted, but the regression of the response surfaces over time can also be displayed. Latin hypercube design would be a possible sampling method for this approach. Subsequently, common RSM analysis is performed. However, depending on the number of factor combinations and the process model complexity, the computational effort could be high.

3.9.2 Model predictive control

The optimization and control of bioprocesses is a challenge due to the non-linearity and non-stationarity, which causes conventional methods, e.g., model-free controller, to not succeed [195]. As part of the process analytical technology in bioprocessing, a model predictive controller (MPC) enables an enhanced bioprocess understanding with a minimized process variability, i.e., minimal deviation of process states, by the control of real-time CQAs [196]. Until now, MPCs have been used in many industrial fields and today, are recognized as advanced control technique [195]. With a MPC, the courses of predefined CQAs are predicted on-line and not before the start of the experiments as with *a priori* models, while the bioprocess inputs are manipulated and adjusted accordingly to a target function [196, 197]. The need for an efficient process control arises from the variability of batch-to-batch and model parameters which leads to different process outputs for same process inputs [195]. According to Sommeregger et al. (2017), a prerequisite for MPC is the measurement of important variables and a mathematical process model that describes the relationship and interaction between process parameters and CQAs. The mathematical process model is used to repeatedly predict a so called finite prediction horizon, for example from time point t to $t + 1$ h, and estimate optimal process inputs of a predefined design space [196]. In contrast to the mDoE method in which model parameters are only adapted to existing data and bioprocess inputs are potentially changed after the process is finally evaluated, dynamic real-time and continuous adaption of process model parameters and operational variables is achieved using MPCs [198].
Today, MPCs and non-linear MPCs, which use nonlinear system equations, are frequently used in optimization of biotechnological processes. Ashoori et al. (2009) applied a MPC to a penicillin fed-batch fermentation process. Instead of nonlinear models, neuro-fuzzy piecewise linear models were used to avoid high computational costs. In addition to the bioprocess

control, optimization of the product concentration was achieved. [195] Another MPC was used by Craven et al. (2014) for the control of the glucose concentration in a fed-batch mammalian cell bioprocess. Despite process variability, measurement noise and time-intensive measurements, the nonlinear MPC was successfully applied. However, they pointed out that an appropriate process model and representative experimental data are needed to optimize feeding strategies or bioprocess variables. [199] In the biocatalytic field, Horsch et al. (2022) discussed the application of a linear MPC for the transesterification of butanol and hexyl acetate. The process control variables were quickly adjusted after disturbances were observed in the process. According to the authors, the main challenge in the applicability of MPCs is the design of the controller which requires experience and know-how in control engineering principles and process dynamics. [200]

Despite the advantages of the *a priori* modeling and the application of MPCs, the mDoE methodology is particularly suitable for the application of a rather simple model from the beginning and a knowledge-based process development, optimization and adaption of variables. Depending on the process to be optimized, the time and the computational effort required to estimate new model parameters and adjust the operational variables trajectory between the short prediction horizons could be too high with a MPC. Furthermore, the implementation of experimental design to systematically assess the impact of process variables [196] has not been shown yet.
The additional implementation of the mDoE methodology into a workflow would enable and facilitate decision-making at various points in the process development and a step-wise adaption as it is done in a modified form for MPCs between the time-horizons. Thus, due to the specific advantages, the mDoE methodology represents the most promising method.

4 Experimental material and methods

In the following paragraph, the biocatalytic reaction system used in this work is described and the kinetic model is formulated. Furthermore, the determination of the influence of high hydrostatic pressure on the enzyme and the related analytics are summarized. As the reaction system is part of an ongoing project with the cooperation partners Institute of Technical Biocatalysis (Hamburg University of Technology) and GALAB Laboratories GmbH (Hamburg), the reaction system is also subject of the doctoral thesis of Marlene Schmalle (Institute of Technical Biocatalysis, Hamburg University of Technology). The conceptualization of the reactor setup, the formulation of the kinetic model and the pressure investigation experiments were performed by Marlene Schmalle. The analytics were provided by GALAB Laboratories.

4.1 Biocatalytical reaction system and reactor concept

In this work, the biocatalytic regeneration of CTP from CDP using polyphosphate (polyP) by immobilized polyphosphate kinase [201] from *Ruegeria pomeroyi* (RpPPK2-3) is investigated in a continuously operated high hydrostatic pressure flow reactor. The enzyme RpPPK2-3 has a wide range of acceptable substrates and uses polyphospahte salts as phosphate donor [117, 201]. One phosphate group is transferred per molecule of CDP regenerated. The reaction formula is displayed in Figure 4.1.

$$\mathrm{CDP} + \mathrm{polyP_n} \xrightarrow{\mathrm{RpPPK2\text{-}3}} \mathrm{CTP} + \mathrm{polyP_{n-1}}$$

Figure 4.1: Reaction formula CTP regeneration.

The reactor concept and the applicability for enzymatic reactions was recently shown by Schmalle and Arndt et al. (2024) [117]. The overview of the reactor setup is shown in Figure 4.2.

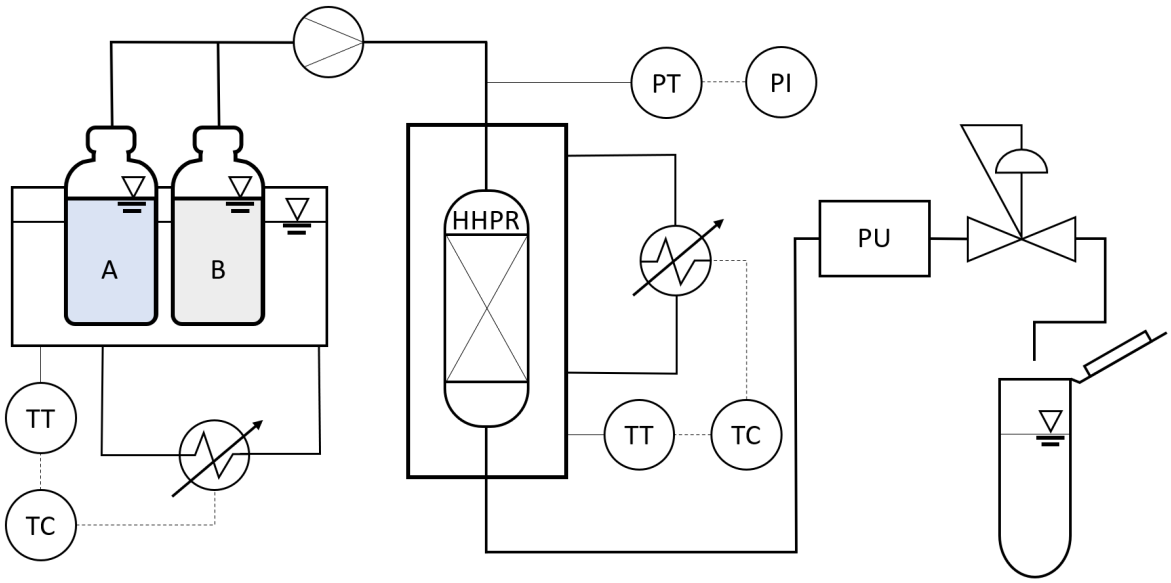

Figure 4.2: Overview of high hydrostatic pressure reactor, modified from Schmalle et al. (2024) [117].

The main parts of the reactor concept are an ultra-high performance liquid chromatography pump (Waters Corporation, USA), a high pressure reactor chamber (Isera GmbH, Germany, see Fig, 4.2, denoted as HHPR) and a pressure generation unit (VICI Jour, Switzerland, see Fig, 4.2, PU). Together with the use of small diameter stainless steel capillaries (0.005", VICI Jour, Switzerland), hydrostatic pressures up to 100 MPa can be realized using this reactor setup. The HHPR and a water bath (Julabo GmbH, Germany) where two substrate bottles A and B can be stored, are temperature-controlled. Depending on the reaction or investigation performed, various compositions of substrate A and B can be set using a mixing chamber. At the end of the reactor, samples can be taken.

The immobilized RpPPK2-3 retained inside the HHPR was covalently bound on a glutaraldehyde-pre-activated epoxy methacrylate carrier (ECR8209M, Purolite, USA). The detailed immobilisation procedure and related stability tests are comprehensively described in Hölting et al. (2024). [201]

4.2 Analytics

Samples of the substrates and effluent taken at the outlet of the reactor were analyzed by a HPLC (Agilent, USA) using an Exsil 300 C18 column (Exmere Ltd, United Kingdom). The mobile phase contained 80% 20 mmol L^{-1} tetrabutylammonium bromide in 20 mmol L^{-1} potassium phosphate buffer with pH 7 and 20% acetonitrile. The temperature was set to 40°C and the flow rate to 0.8 mL min^{-1}. The detector wavelength was 272 nm. 100μL of a

collected sample was mixed for 5 seconds with 50μL isopropanol on a vortex mixer (Scientific Industries Inc., USA). Subsequently, the sample was centrifuged for 1 minute and 25μL of the supernatant was diluted with 975μL water. The resulting sample was then transferred to a HPLC vial and measured.

From the experimental data obtained in first experimentation, a systematic deviation between measured CDP concentration and weighted and prepared CDP substrate solution was observed (see Suppl. Tab. A.1). Since the deviations were systematic and consistent, i.e., measured 36 mmol L^{-1} instead of 30 mmol L^{-1}, and the origin could not be clarified during the time of study, the proposed workflow and mDoE were performed with the measured values. It was assumed that the substrate and product concentrations were influenced to the same extent and the factor boundary conditions were not affected. In addition, an improved process understanding and knowledge was considered as more important compared than an exact prediction of the actual final concentration.

5 Process model development

In the following paragraph, the development of the mathematical process model describing the continuously operated biocatalytic cofactor regeneration reaction is discussed in detail. First, the modeling approach is selected (see Fig. 5.1).

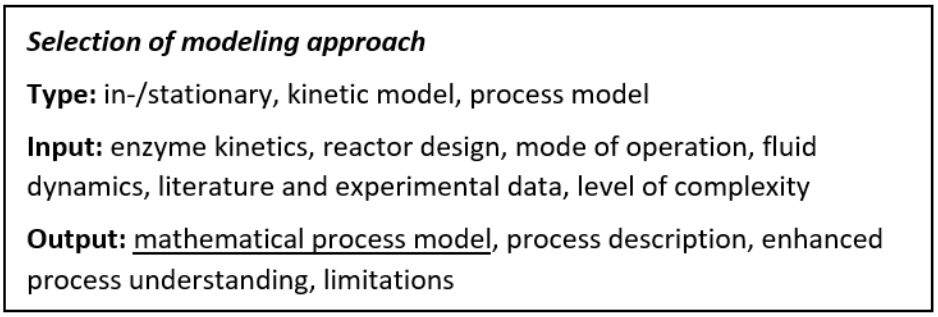

Figure 5.1: Considerations for the selection of a modeling approach.

Therefore, the type of the model must be defined in a way that the model reflects the complexity and scope of the system to be modeled. A combination of different models, such as kinetic and process models, is possible. After selecting the model type, the model inputs are defined. Among others, enzyme kinetics and the corresponding kinetic equations are incorporated into the model, the flow through the packed-bed reactor is characterized and well-considered assumptions are made, to reduce the overall model complexity. As a result, a mathematical process model is formulated that allows a description of the process and an enhanced process understanding through a set of kinetic and differential equations.

5.1 Process model approach: Method-of-Lines

The transport of different species through the stationary phase of the continuously operated packed-bed is modeled using a dispersion-convection model and the method-of-lines approach (MOL). To ensure fast adaption of model parameters and to preserve computational performance, only one-dimensional axial dispersion and convection is considered in this thesis. In the reaction-dispersion-convection model, all non-idealities which arise due to a non-ideal flow, e.g., Eddy-diffusion and molecular diffusion, and superimpose the plug flow behavior are summed up in the apparent axial dispersion coefficient D_{ax}. The resulting

one-dimensional partial differential equations (PDE) with time- and location-independent D_{ax} neglecting velocity gradients can be written as [118, 202–204]:

$$\frac{\partial c_i}{\partial t} = D_{ax} \frac{\partial^2 c_i}{\partial x^2} - u \frac{\partial c_i}{\partial x} \tag{5.1}$$

In addition to the one-dimensionality of flow, constant D_{ax} is also assumed for varying flow rates. The PDE system possesses the following Danckwert's initial and boundary conditions for a closed-closed system [202]:

$$c_{0,x} = 0, 0 \geq x \geq L, t \geq 0 \tag{5.2}$$

$$\frac{\partial c_i}{\partial t}(x = L) = 0 \tag{5.3}$$

Equation 5.1 is called a parabolic partial differential equation (PPDE) which describes changing phenomena at various points from moment to moment. Compared to elliptic partial differential equations ($u_{xx} = 0$), PPDE are rather used to describe instationary, dynamic processes ($u_{xx} = u_t$), e.g., diffusion, heat conduction and oscillatory phenomena [205–207].
The formulated PDE system is then approximated by the MOL approach which leads to a time- and location-dependent description along the axial reactor bed (x = 0 to x = L). [208–211] For this, the reactor length is uniformly segregated into a predefined number of grid nodes n (spatial segregation, see Fig. 5.2).

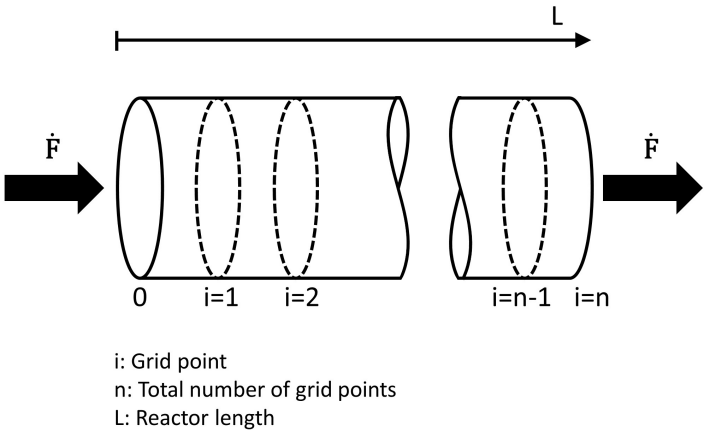

Figure 5.2: Segregation of the reactor length L into n grid-nodes according to the MOL approach [208].

The change of concentration is then calculated consecutively for every node i [208, 212]. L_{step} is considered as the resulting step length.

In the MOL, the first and second derivative of the convection and diffusion terms are approximated point-wise using the finite difference method (FDM) in one dimension. A second order, central approximation with adequate initial and boundary conditions is used (Taylor series expansion) [208, 212]. The resulting system of algebraic first order differential equations can be then solved by an ordinary differential equation solver, e.g., the *ode15s* in MATLAB [186]:

$$\frac{\partial c_i}{\partial t} = D_{ax} \frac{c_{i-1} - 2 \cdot c_i + c_{i+1}}{L_{step}^2} - \frac{u}{e} \frac{c_{i+1} - c_{i-1}}{2 \cdot L_{step}} \tag{5.4}$$

To estimate the axial dispersion coefficient D_{ax}, the flow through and type of the mass transport within the packed-bed reactor has to be characterized.

5.2 Characterization of the flow behavior

The flow of the mobile phase through the packed-bed is characterized by the flow velocity u, the bulk density e and the dimensionless Bodenstein number Bo. The Bo describes the rate of the convective to the dispersive mass transport within the packed-bed [213]. In reactor systems with Bo > 100 only slight deviations from an ideal plug flow can be observed and ideal plug flow behavior is achieved for Bo > 1000 [214]. Non-homogeneities and the degree of axial mixing can be determined by analyzing the residence time distribution using the dispersion model. In brief, a suitable tracer is injected into the reaction system and the response at the reactor outlet is measured [202, 214–217]. The mean residence time $\bar{t}$ and space time τ are defined as [202, 218, 219]:

$$\bar{t} = \sum_{t=0}^{\infty} t \cdot E(t) \, \Delta t \tag{5.5}$$

$$\tau = \frac{V}{\dot{V}} \tag{5.6}$$

$E(t)$ denotes the exit age distribution and indicates the fraction of moleculese leaving the reactor with a residence time between a time interval $t + dt$ [220]. The Bo can then be approximated from the variance σ^2 of the residence time distribution with respect to the mean residence time $\bar{t}$ [202, 218, 219]:

$$\theta = \frac{t}{\bar{t}}, \; \bar{t} : \textit{mean residence time} \tag{5.7}$$

$$\sigma^2 = \int_0^{\infty} \left(t - \bar{t}\right)^2 E(t) \, dt \tag{5.8}$$

$$\sigma_\theta^2 = \frac{\sigma^2}{\bar{t}^2} \tag{5.9}$$

For closed-closed system conditions Equation 5.10 is used to approximate Bo with the dimensionless variance σ_θ^2 [219, 221, 222]:

$$\sigma_\theta^2 = \frac{2}{Bo} - \frac{8}{Bo^2} \cdot (1 - exp(-Bo)) \tag{5.10}$$

The D_{ax} is calculated based on Bo, the flow velocity u and reactor length L [223]:

$$D_{ax} = \frac{u \cdot L}{Bo} \tag{5.11}$$

The D_{ax} is then used in the dispersion-convection model to describe the disperse flow.

Analysis of the flow through the packed-bed reactor

To characterize the flow through the packed-bed reactor, the residence time distributions for five flow rates $\dot{V}$ varying from 0.5 mL min^{-1} to 1.5 mL min^{-1} and constant inlet concentration (step tracer inlet) were analyzed. Prior to experimentation, the reactor was filled with immobilizate and the packed-bed was flushed with substrate-free sodium phosphate buffer. The feed was then initiated (t=0 min) and samples were taken at the reactor outlet. The feed composition and the time Δt between the sampling points are summarized in Table 5.1. A total of five samples were taken for each flow rate with six replicates for each.

Table 5.1: Sampling points and experimental conditions for flow characterization experiments, LxID=30x3 mm, $m_{Carrier}$ = 0.166 g, 50 mmol L^{-1} sodium phosphate buffer, $c_{Mg^{2+}}$ = 30 mmol L^{-1}, T=40°C, pH=7.4, Purolite ECR8209M carrier.

Δ t Sampling [s]	**Flow rate** [mL min^{-1}]	**CDP_0** [mmol L^{-1}]	**$polyP_0$** [mmol L^{-1}]	**n [−]**
30	0.5	44	0	6
20	0.5	44	0	6
15	0.5	44	0	6
12	0.5	44	0	6
10	0.5	44	0	6

After each measurement, the packed-bed was flushed again with substrate-free sodium phosphate buffer and the measurements were continued according to Table 5.1. The dimensionless cumulative F(t) distribution, defined as $c_{Tracer,t}/c_{Tracer,0}$, and dimensionless exit age E(θ) distribution with respect to the space time τ are exemplarily shown in Figure 5.3.

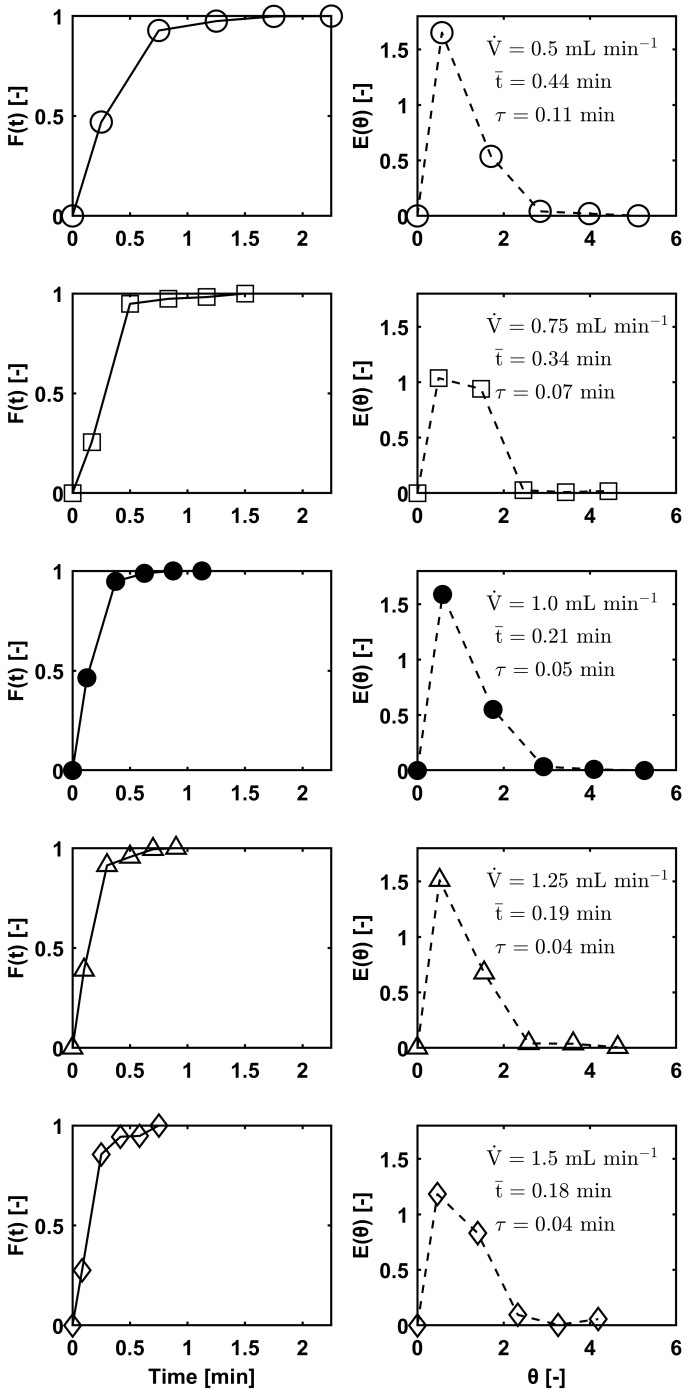

Figure 5.3: Cumulative F(t) and external age E(θ) distributions for different flow rates and constant inlet concentration. Top to bottom: 0.5, 0.75, 1.0, 1.25 and 1.5 mL min^{-1} with $c_{CDP,inlet}$ = 44 mmol L^{-1}. n=1.

As can be seen from Figure 5.3, the time needed to reach F(t)=1 decreased with increasing volumetric flow rate $\dot{V}$ as sapce times τ decreased (from top to bottom). Furthermore, similar E(θ) distribution widths and decreasing mean residence times $\bar{t}$ from 0.44 to 0.18 min were observed. As described in Subsection 5.2 the Bo was then approximated by calculating the dimensionless variance σ_θ^2 with closed-closed boundary conditions (Equ. 5.10). For the E(θ)

distributions shown, σ_θ^2 of 0.60, 0.35, 0.51, 0.34 and 0.51 as well as Bo of 1.10, 0.84, 1.58, 1.31 and 1.77 were determined. Therefore, deviations from the ideal plug flow reactor (Bo > 1000) can be assumed [214]. In the formulated process model and definition of process parameters, an average Bo of 1.5 was used as the volumetric flow rate had a relatively small effect on the Bo in the selected flow range and reactor dimensions presented.

5.3 Kinetic model

To model the overall biocatalytic cofactor regeneration reaction in the continuously operated packed-bed reactor, the dispersion-convection model has to be extended by a kinetic model. The kinetic model should account for the dependence of the reaction on the initial substrate concentrations and the hydrostatic pressure. The choice of the kinetic equations and assumptions is critical for further process design and optimization.

5.3.1 Influence of initial substrate concentration

The Michaelis-Menten kinetics is a simple and frequently used approach to describe the volumetric reaction rate of an enzymatic reaction [224, 225]. It can be formulated as [224]:

$$v = \frac{v_{\max} \cdot c_s}{K_m + c_s} \tag{5.12}$$

$v_{\max}$ denotes the maximum rate of reaction at infinite reactant concentration, K_m the Michaelis-Menten constant and c_s the reactant concentration. K_m is equal to the reactant concentration at $v_{\max}/2$ and expresses the affinity of the enzyme for the substrate. As the reactant concentration increases, the catalyst's active center becomes saturated, which leads to a decrease in the slope of the reaction rate curve. If the reactant concentration $c_s \gg K_m$, the v approaches $v_{\max}$ in the zero-order kinetics region [224].

In this work, the volumetric reaction rate is to be described as a double substrate reaction with the substrates CDP and polyP. Therefore, an additional term is added to Equation 5.12 [226]:

$$v = v_{\max} \cdot \frac{c_{CDP}}{c_{CDP} + K_{m,CDP}} \cdot \frac{c_{polyP}}{c_{polyP} + K_{m,polyP}} \tag{5.13}$$

In kinetic experiments with varying concentrations of the substrates CDP and polyP, Schmalle et al. (2024) observed an inhibitory effect by polyP. As a result, different inhibition models were tested to investigate which inhibition type, i.e., competitive, non-competitive and uncompetitive inhibition, accurately describes the experimental data. Double-substrate Michaelis-Menten

kinetics and uncompetitive enzyme inhibition relative to the substrate polyP ($K_{I,polyP}$) best described the data and were therefore assumed for RpPPK2-3. [117] The resulting kinetic equation considering uncompetitive inhibition can be formulated as [227]:

$$v = v_{max} \cdot \frac{c_{CDP}}{K_{m,CDP} + c_{CDP}} \cdot \frac{c_{polyP}}{K_{m,polyP} + c_{polyP} \cdot \left(1 + \frac{c_{polyP}}{K_{I,polyP}}\right)} \tag{5.14}$$

In Figure 5.4 the related experimental data from the kinetic investigation are shown.

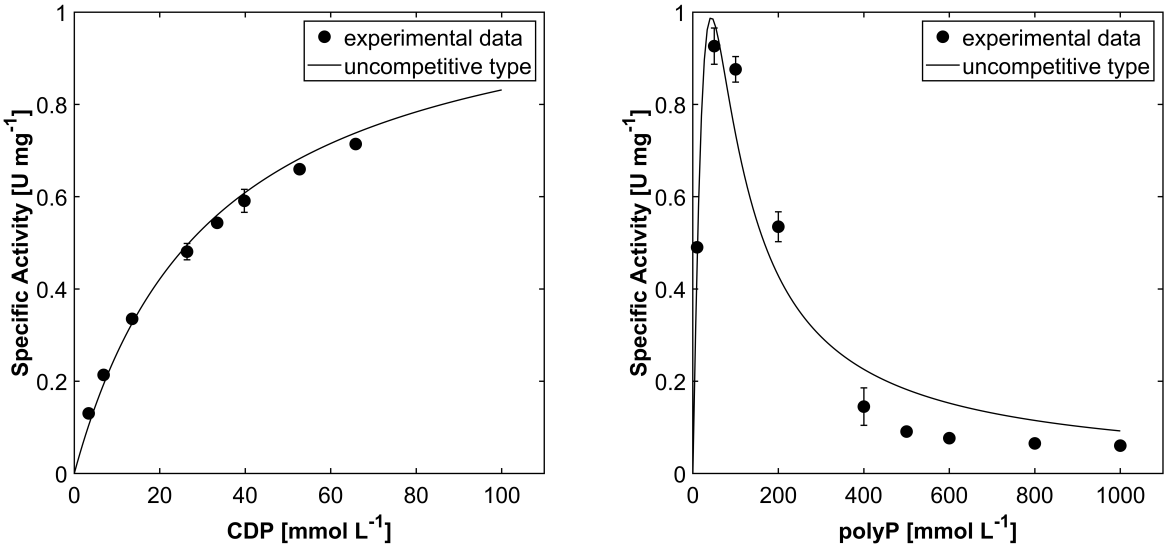

Figure 5.4: Investigation of substrate inhibition for varying initial CDP and polyP concentrations in the continuously operated high hydrostatic pressure reactor, for experimental conditions see Schmalle et al. (2024) [117].

As can be seen, no enzyme inhibition was observed for CDP in the concentration range shown. However, for increasing polyP concentrations, the reaction rate initially increased and then decreased for polyP concentrations higher 50 mmol L^{-1}. The experimental procedure is described in detail in Schmalle et al. (2024) [117].

To account for time-dependent enzyme deactivation, the deactivation rate k_d is calculated according to the Arrhenius model [228, 229]. The equation is defined as:

$$k_d = -t \cdot ln\left(\frac{v}{v_0}\right) \tag{5.15}$$

v and v_0 denote the reaction rate at time t and the initial reaction rate observed. The determination of the deactivation rate k_d is presented in Schmalle et al. (2024) [117]. The kinetic equations and the related estimated parameters represent the starting point for the model-assisted design and optimization which is integrated into the iterative workflow (see

Subs. 6). With every iteration round, the kinetic equations and parameters can be adapted. The resulting kinetic equations are summarized in the Supplementary A.2.

5.3.2 Formulation of the influence of hydrostatic pressure

As described in Subsection 3.2.2, high hydrostatic pressure can have an influence on the enzymatic activity. Therefore, the specific activity of the enzyme was determined for varying pressures in the high hydrostatic pressure reactor. In brief, the reactor was packed with immobilizate ($m_{RpPPK2-3} = 8.1$mg) and the pressure was adjusted from 0 to 800 bar. Subsequently, the reactor was re-packed with new immobilizate and the specific activity was determined again ($m_{RpPPK2-3} = 7.2$mg). A starting concentration of $c_{CDP} = 30$ mmol L^{-1} and $c_{polyP} = 44$ mmol L^{-1} was used (for further information see Schmalle et al. (2024) [117]). The specific activity values were normalized to the maximal value measured. The normalized specific activity values are shown in Figure 5.5.

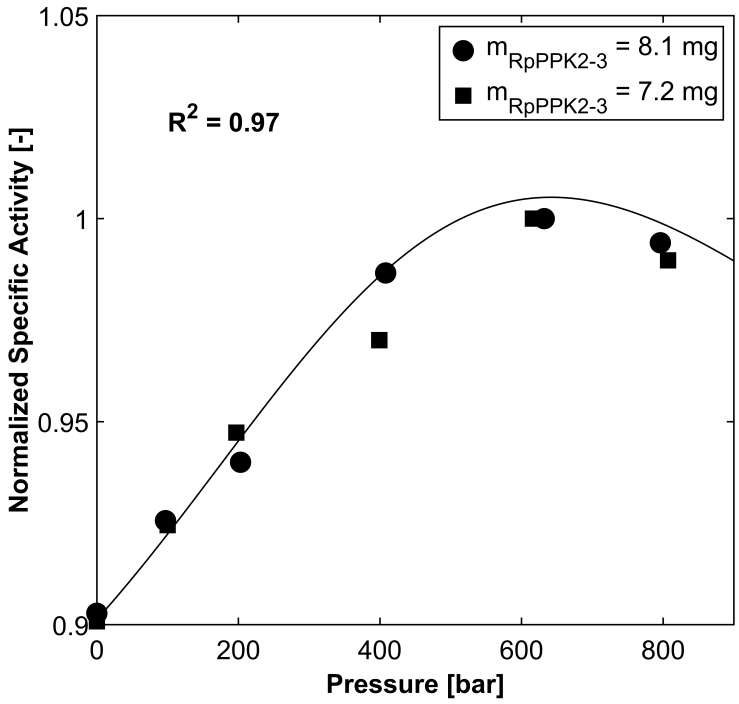

Figure 5.5: Investigation of the effect of hydrostatic pressure on specific activity for varying pressures in the continuously operated high hydrostatic pressure reactor. The solid line is based on the adapted parameters of the double sigmoidal function f_{Dsig}. For experimental conditions see Schmalle et al. (2024) [117].

In both experiments the specific activity of 0.56 U mg^{-1} and 0.70 U mg^{-1} at 0 bar increased for increasing hydrostatic pressures (normalized values of 0.903 and 0.901, respectively). The

highest specific activity of 0.62 U mg^{-1} and 0.78 U mg^{-1} was measured at 632 and 616 bar, respectively. At the maximum pressure of 796 and 807 bar, a decrease in the specific activity was observed.

The mathematical formulation of the dependence of the enzyme activity on the reaction conditions can be expressed using different types of equations. Jurado et al. (2004), proposed a deprotonation and a first-order deactivation model to describe the enzyme activation and deactivation by pH and temperature [16]. Another approach was presented by Prunescu and Sin (2013), who expressed the pH and temperature dependence of an enzymatic hydrolysis process, resulting in an efficiency factor, using individual arbitrary Gaussian curve functions [170]. However, these type of equations are limited in the shape and skewness of the resulting curve describing the influence of a particular reaction condition. Alternatively, the influence of pH or temperature on biochemical reactions or cell growth can be expressed by sigmoidal [230] and double sigmoidal functions [57, 86].

The concept of double sigmoidal functions is used and adopted in this thesis to express the dependence of the biocatalytic activity on the hydrostatic pressure [231]. As shown in the work of Appl et al. (2021), the advantage of the double sigmoidal function lies in the possibility to define a rather unusual shape of the curve, which may better represent the influence of the reaction conditions [86]. The resulting function is then integrated into the kinetic equations and the process model. The general double sigmoidal function f_{Dsig} can be written as [57]:

$$f_{Dsig}\left(x_{Dsig}\right) = \left(Y_l + \frac{Y_{mid} - Y_l}{1 + exp\left(-K_{sl} \cdot \left(x_{Dsig} - X_{50,l}\right)\right)}\right) \cdot \left(1 + \frac{Y_h/Y_{mid} - 1}{1 + exp\left(-K_{sh} \cdot \left(x_{Dsig} - X_{50,h}\right)\right)}\right) \tag{5.16}$$

f_{Dsig} denotes the normalized biocatalytic activity and x_{Dsig} the hydrostatic pressure values. As a result, f_{Dsig} ranges from zero to one. A non-linear least-squares solver is used to adapt the initial parameter values. The initial and adapted parameter values are summarized in Table 5.2.

Table 5.2: Initial and adapted parameters of the double sigmoidal function describing the influence of hydrostatic pressure on the specific activity of the enzyme.

Parameter	Initial value [-]	Adapted value [-]
Y_l	0.9	0.9
Y_h	0.9	0.9
Y_{mid}	0.6	3.71
$X_{50,l}$	200	415.2
$X_{50,h}$	500	0.44
K_{sl}	0.2	0.003
K_{sh}	0	0.003

Despite the parameters Y_l and Y_h all parameters changed from their initial values to minimize the sum of squares between the experimental and function data. As can be seen in Figure 5.5 the normalized specific activity data were reflected well by the double sigmoidal function with a calculated R^2 of 0.97. Therefore, the double sigmodial function and the adapted parameter values were incorporated into the process model.

5.4 Reaction-dispersion-convection model

The resulting partial differential equation including the dispersion-convection model based on the MOL approach (hydrodynamic model part, see Equ. 5.4) and the kinetic model (see Equ. 5.14) reflecting the biocatalytic reaction and dependence on the reaction conditions can be formulated as:

$$\frac{\partial c_i}{\partial t} = \underbrace{D_{ax}\,\frac{\partial^2 c_i}{\partial x^2} - u\,\frac{\partial c_i}{\partial x}}_{\text{Part hydrodynamic model}} + \underbrace{f\,(i,t)}_{\text{Part kinetic model}} \tag{5.17}$$

The kinetic model part $f\,(i,t)$ of the differential equation includes the kinetic equation v as well as the description of the hydrostatic pressure dependence of the immobilized enzyme which is represented by the double sigmoidal function f_{Dsig} (see Equ. 16):

$$f\,(i,t) = v \cdot f_{Dsig} \tag{5.18}$$

The overall model is called a reaction-dispersion-convection model. Similar to Equation (5.4), a second order, central approximation with adequate initial and boundary conditions (see

Equ. 5.2 and 5.3) is used for the final differential equation [208, 212]:

$$\frac{\partial c_i}{\partial t} = D_{ax} \frac{c_{i-1} - 2 \cdot c_i + c_{i+1}}{L_{step}^2} - \frac{u}{e} \frac{c_{i+1} - c_{i-1}}{2 \cdot L_{step}} + f\,(i,t) \tag{5.19}$$

It should be noted that the final differential equation and the process model presented serve as a basis and can be modified in the course of process development. Model assumptions and boundary conditions can also be adjusted. The complete process model used in this thesis is summarized in Supplementary Table A.2.

6 Iterative workflow

To efficiently design and optimize the biocatalytic cofactor regeneration process, an iterative mDoE workflow is created. The underlying idea of the workflow is based on the suggestions of Möller et al. (2019) and Kuchemüller et al. (2020), who developed a workflow for cell culture processes, mainly operated in batch and fed-batch mode. In contrast to these workflows, the workflow presented here is based on a different modeling approach and used for a continuously operated process.

6.1 Development of the iterative workflow

The aim of this workflow is to improve the process development procedure and to enhance the process understanding when only limited amount of information is available, e.g., on enzyme kinetics, suitable experimental space and bioreactor design characteristics. As reported in literature, by implementing the mDoE methodology in a workflow, a more knowledge-based optimization with consideration of experimental variations and reduced experimental effort can be achieved [35, 144, 232]. In Figure 6.1, the suggested workflow consisting of different modules with varying tasks is shown.

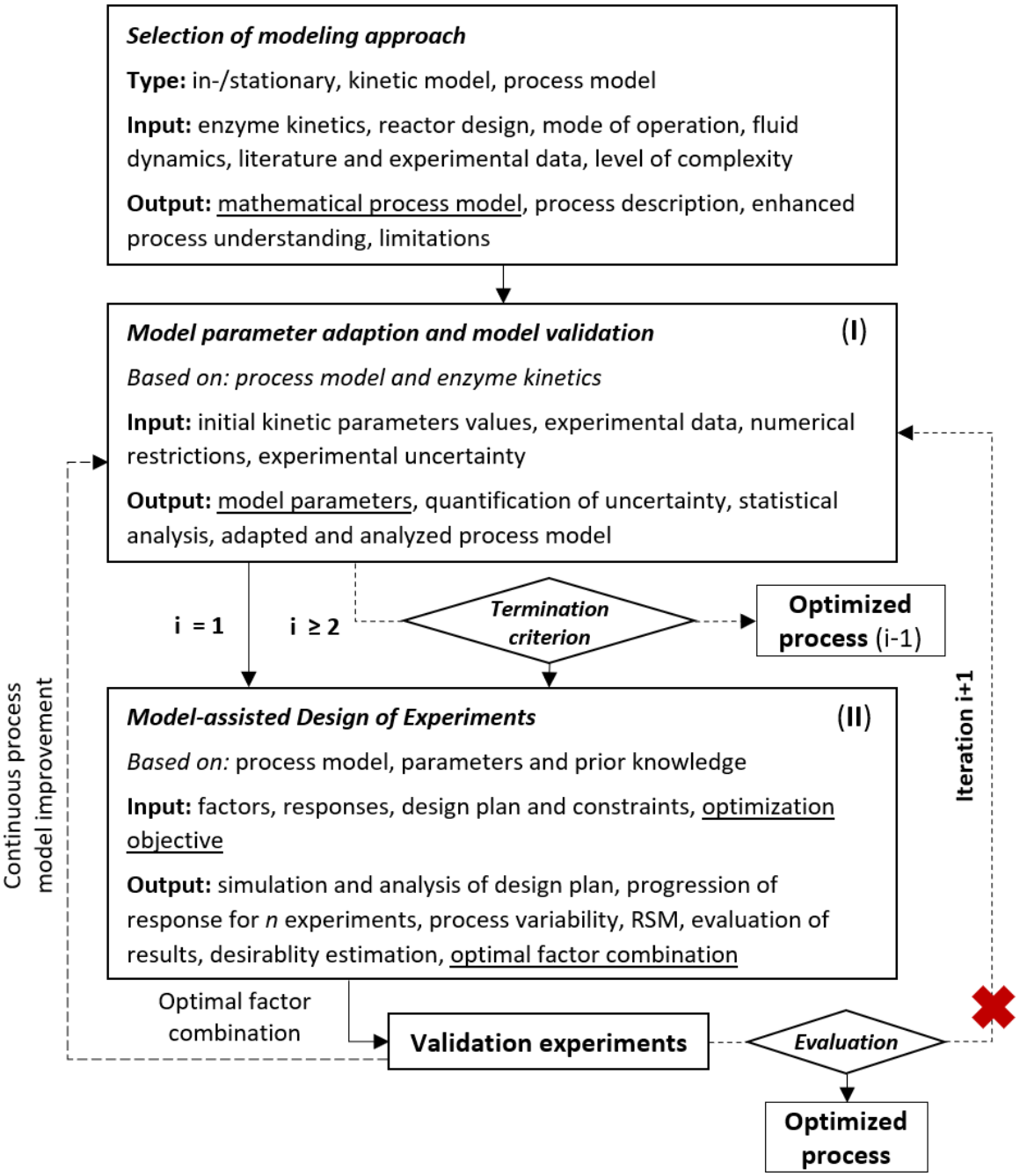

Figure 6.1: Proposed iterative model-assisted Design of Experiments workflow for process design and optimization [35, 57, 144].

Initially, a mathematical process model (see Sub. 5) is formulated using first experimental data, literature or general information about enzyme kinetics and reactor specific flow characteristics (Fig. 6.1, top). The process model also comprises information on the reactor design, operation mode and user-defined limitations (see Subs. 5). Subsequently, based on the process model and available data under varying process conditions, e.g., substrate starting concentration, the model-parametric uncertainties are determined (Monte-Carlo method see

Subs. 6.2, Fig. 6.1). The process is simulated and model parameters are adapted multiple times based on an assumed normally distributed measurement error of the experimental data [35, 144]. In this way, experimental variability is taken into account both in the biocatalytic reactions and measurements. In case of deviations between the experimental and simulated data, additional selective experiments have to be performed or the process model has to be extended, while overfitting of the process model should be avoided. This section can also be used to validate the process model.

To reduce the experimental effort, mDoE methods are applied (see Subs. 6.3, Fig. 6.1, II). The optimization objective is defined first and experimental factors as well as responses, e.g., product concentration, are specified. Second, a suitable DoE design is selected depending on boundary conditions delimited by prior experimental findings or expert knowledge. Third, the suggested experimental combinations are simulated using the process model and adapted model parameters. All responses are treated in the same way as experimental data.

Subsequently, Monte-Carlo-based simulations are performed. Therefore, sets of model parameter values are randomly drawn by Latin-Hypercube sampling (LHS) n-times and used to simulate the responses. This procedure enables the prediction of the propagated process variability by the pre-quantified process variability of the model parameter distributions (Fig. 6.1, I). Then, the obtained responses are evaluated by RSM. An appropriate regression model is selected, regression coefficients $\hat{\beta}_i$ are determined and the response surface is plotted using $\hat{y}_{\mathrm{x,t}}$. Furthermore, the weighted desirability, which is a measure for the quality of the expected response within acceptable factor ranges, is calculated and an optimal factor combination, which represents the new favorable process condition is obtained by further analysis.

In case of deviation between simulated and experimental data from validation experiments, model parameters and the process model itself are re-adapted considering the new experimental data and the workflow is repeated (i+1) until the termination criteria is reached (valid for second and subsequent iteration rounds). In brief, re-adapted model parameters from round i are used to predict a new response surface $(\mathrm{i}-1)_{\mathrm{t}}$ using the previous input data (i-1) and the optimal responses of both rounds are compared. If the increase in the optimal response was less than 10%, the workflow was stopped.

By using this proposed iterative workflow, more focused initial and optimization experiments are performed, the number of validation experiments is reduced and the bioprocess dynamics are described comprehensively. The experimental data gained from validation experiments should be used in the subsequent workflow rounds to further improve the process model.

6.2 Adaption of model parameters and uncertainty analysis

6.2.1 General optimization procedure

Initial approximations or expert knowledge from previous studies and literature can be used to set the model parameter starting values [144]. The adaption of these process model parameters is based on the minimization of a weighted cost function J considering experimental and simulated process data. The number of data sets and the weighting factor are defined as j and w_k.

$$J = \sum_{k=1}^{j} w_k \cdot \sqrt{\sum_{m=1}^{n} \frac{\left(y_{\mathrm{s,m}} - y_{\mathrm{i,m}}\right)^2}{n}} \tag{6.1}$$

The model parameter starting values are adapted using a minimization algorithm which is based on the Nelder-Mead algorithm implemented in MATLAB as *fminsearch* function [233]. Similar to DoE methods and analogous to factors and responses, only a selection of model parameters is to be adapted and to be defined beforehand. To reduce the computational effort and time needed for the quantification of the model-parametric uncertainty, model parameter starting values are adapted to a reduced experimental data set (training set). In this thesis, 4 data sets out of the available data sets are randomly chosen as training set for the individual adaption runs. Nevertheless, the process dynamics resulting from varying process conditions are captured due to the repeated model parameter adaptions. The model parameter adaption is terminated if a certain termination criterion is reached or if the maximum number of iterations is exceeded [144].

6.2.2 Quantification of model-parametric uncertainty

To evaluate the process model parameters and possible model outcomes with their associated probability of occurrence, the model-parametric uncertainties are determined based on the experimental uncertainty and the process dynamics are simulated by the process model. Thus, when measuring an unchanging quantity, a distribution of possible model outputs is generated. [35, 58, 144, 224, 234]

In this work, the variation of the model inputs is defined, model parameters are adapted repeatedly (1000 times) to a randomly chosen training set and the model outputs are sampled for each run. The samples are drawn by Monte-Carlo methods, with an experimental normally distributed relative error of 2% standard deviation. [35, 144, 235] The relative error of

2% was assumed due to the use of high precision flow and hydrostatic pressure equipment which is also used in analytics. In addition, the substrate solution were prepared accurately. Subsequently, the mean, median and uncertainty bands representing the individual 10% and 90% quantiles of the model parameter distributions are determined. The model parameter adaption objective is defined as minimization of the Root Mean Square Deviation (RMSD) between the experimental and simulated data [236].
Another uncertainty analysis approach is proposed by Prunescu and Sin (2013). Compared to the procedure used in this work, the authors defined the input uncertainty with their range first and then sampled kinetic and process parameters such as feed parameters using Latin hypercube sampling. In contrast to the model parameter adaption procedure shown in this work, model parameters are only adapted (also referred as calibrated) once on a reduced set of recorded experimental data. Subsequently, simulations are performed and the results are evaluated. [170] However, the sound selection of the data sets to which the model parameters are to be adjusted is critical. In addition, this approach runs the risk that decisive effects remain unrecognized.

6.2.3 Monte-Carlo-based uncertainty bands

The uncertainty of a process can be considered and evaluated by assigning probability distributions to the process input factors, e.g., random substrate concentration, and repeatedly simulating the process outputs resulting in output probability ranges [58, 234]. This method differs from a model calibration situation, where for a "specific set of model inputs, the model outputs differ from the observed values" [234]. After calibration, for "those model inputs, the observed values are always the same" with no randomness [234]. In this work, the process variability is simulated based on the former determined model parametric uncertainty (see Subs. 6.2.2). Therefore, uncertainty bands, representing the uncertainty propagation, are determined by calculation of the mean and the 10% and 90% quantiles of each individual model parameter distribution by the *prctile* function in MATLAB. The process is then simulated using the three model parameter values (10% quantile, mean, 90% quantile). For normally distributed model parameter values, the median value can be used as an alternative to the mean value.
The confidence interval (CI) serves as an additional way to evaluate and compare the results of a hypothesis test about model parameters and their distributions for different bioprocess conditions [35, 237]. In general, the estimated mean value of a model parameter distribution is different from the actual mean value. However, with sufficient sample size this mean can be

approximated. According to Scott (2020), the CI is regarded as an informative summary and gives a range of the most likely values of the actual model parameter. The 95% CI is defined as the corresponding quantity of a test with the 5% level. The acceptance region for testing the null hypothesis H_0 with known variance σ^2 is defined as:

$$z_{0.025} \leq \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \leq z_{0.975} \tag{6.2}$$

with the random sample mean $\overline{X}$ and the known mean value μ_0 based on the number of samples n at the 5% level test [237]. This formulation can be re-formulated to:

$$\overline{X} - z_{0.025}\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \overline{X} + z_{0.975}\frac{\sigma}{\sqrt{n}} \tag{6.3}$$

For a 95% CI, $z_{0.025}$ and $z_{0.975}$ are 1.96. The expression 95% CI does not mean that the fixed parameter μ_0 is with a probability of 95% within the random interval. It rather means that if 100 experiments are performed with a true H_0 and 95% confidence, 95 experiments and their CIs will include the μ_0 and 5 experiments will not [237, 238]. The CIs can be also calculated if the σ^2 is unknown [237].

Similar to the simulation of the 10% and 90% quantile uncertainty band, the process can be simulated using the calculated values from the 95% CI and the mean model parameter value. However, as this work focuses on the determined interval of possible process outcomes and not on the interval resulting from the distribution that includes the true model parameter with a defined certainty, the uncertainty bands are simulated. Nevertheless, the model parameter confidence interval can be determined, as it is a measure for the level of certainty in the parameter estimation.

6.3 RSM and Monte-Carlo-based simulations

6.3.1 Applied RSM procedure

As proposed in the mDoE workflow, the objective is defined and an appropriate regression model as well as experimental design is selected first to perform RSM. In addition, factor combinations defined by the DoE are simulated and model coefficients of an empirical RSM, which describes the response surface in 1+n dimensions, are predicted using the least square method. [179] As described in literature, an empirical multiple-linear regression models of second-order is used due to the higher-dimensional structure. In the end, an optimal factor

combination is predicted to maximize the response and the adequacy of the model is evaluated using statistical tools [45, 177, 187].

In this work, four influencing factors are considered: the hydrostatic pressure, the substrate concentration CDP and polyP as well as the flow rate. The corresponding fitted second-order polynomial regression model for 4 factors with intercept, linear, interaction and quadratic terms can be written as:

$$\hat{y}_{\mathrm{x,t}} = \hat{\beta}_0 + \sum_{i=1}^{4} \hat{\beta} x_i + \sum_{i<j} \hat{\beta}_{ij} x_i x_j + \sum_{i=1}^{4} \hat{\beta}_{ii} x_i^2 + \epsilon \tag{6.4}$$

The quadratic terms are necessary to account for curvature in the response surface [179].

6.3.2 RSM evaluation and analysis of variance

The estimation of the regression coefficients and the test of significance for the regression model are performed by the polynomial regression model and ANOVA (for further information see Subs. 6.5.6). For this, the *regstats* algorithm and *anova* function were used (MATLAB, R2018b). The calculation of single model coefficient significance values (p-values, t-statistics) for first- and second order models is described in the literature [179]. The p-value together with a defined significance level indicates whether the single model coefficients are significant or not [177]. Regression coefficients which do not significantly contribute to the model can be excluded from the final regression model resulting in the reduced form [239]. Based on the F-statistics provided by the *anova* function, the proposed regression model was classified as significant or non-significant (ANOVA, see Subs. 6.5.6) [49]. In addition, the R^2 indicates what proportion of the response variation is explained by the regression model [176]. Further evaluation was performed by two dimensional factor plots to assess the individual effects of single factors within the factor range on the response [177].

6.3.3 Monte-Carlo-based simulation of experiments

To fully exploit the mDoE capability (see Subs. 3.8), Monte-Carlo simulations for individual factor combinations suggested by the DoE are performed. According to Arndt et al. (2020) [144] and Moser et al. (2020) [57], based on the adapted model parameters and the resulting model parameter distribution, 30 simulations were executed to quantify the propagated uncertainty for every factor combination. The corresponding model parameter values were drawn by LHS and the median and the 10% and 90% quantiles were determined. The results, which represent the propagated experimental uncertainty were used to generate a new response

surface and determine optimal factor combinations (see Subs. 6.3.1) [57]. Thus, this approach enables a risk assessment by determining the range of possible responses [58].
The applied LHS is a non-collapsing and mostly space-filling design, which implies, that there are no unwanted replicated and wasteful responses in the experimental design. Therefore, *n* distinct values are obtained from *n* distinct points [179]. In Figure 6.2, a LHS design generated with the *lhsdesign* algorithm compared to a random design using the *rand* function (MATLAB, R2018b) is exemplarily shown for 100 combinations of two factors or parameters x_1 and x_2:

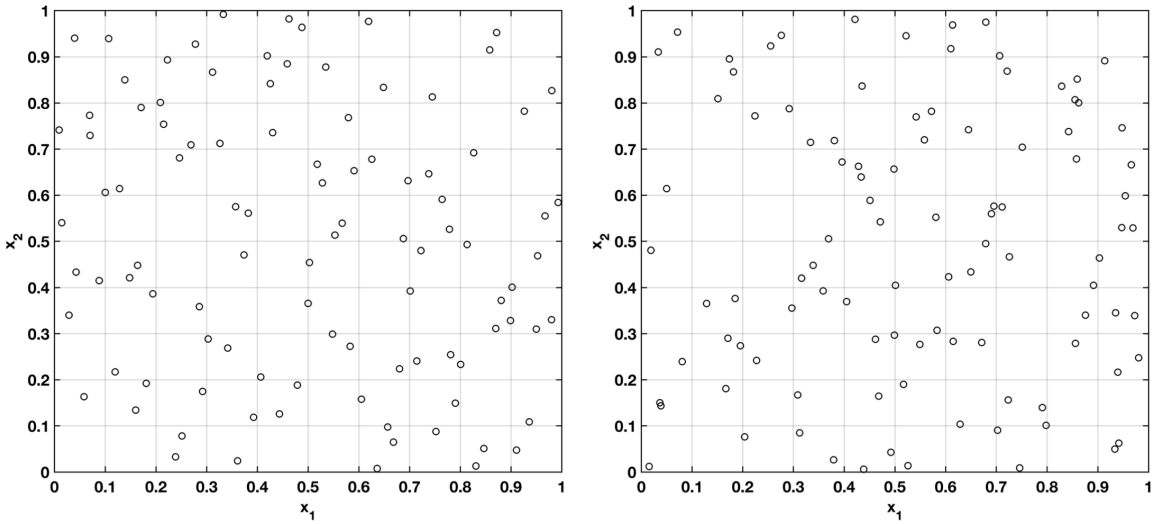

Figure 6.2: Latin-Hypercube design (left) and random experimental design (right) for 100 combinations of two factors x_1 and x_2.

Compared to the random design, where the experimental design is not fully covered, the Latin-Hypercube covers the experimental space better due to the evenly distributed random values. Out of 100 combinations, 79 are covered by LHS and only 71 are covered by the random design. As a result, the model parameters are drawn more evenly from the model parameter distributions, resulting in an improved representation of variability.

6.4 Desirability evaluation

As briefly described in Subsection 3.8, further evaluation of the mDoE and the resulting response surface is conducted by using the results from the Monte-Carlo-based simulations and the calculation of the desirability function $d_i(\overline{y}_i)$. The desirability ranges between values from 0 and 1. A desirability function is 1 if the optimization and the expected response $\overline{y}_i$ is highly desirable. In this work,$\overline{y}_i$ was defined as the maximized product concentration at the end of the reactor. $d_i(\overline{y}_i)$ is calculated based in the lower and upper acceptable response L_i

and U_i as shown in Equation 6.5 [45, 57, 185]:

$$d_i\left(\overline{y}_i\right) = \begin{bmatrix} 0 & \text{if } \overline{y}_i < L_i, \\ \left(\dfrac{\overline{y}_i - L_i}{U_i - L_i}\right) & \text{if } L_i < \overline{y}_i < U_i, \\ 1 & \text{if } \overline{y}_i > U_i. \end{bmatrix}. \tag{6.5}$$

The overall desirability D is calculated as the product of the desirability function values [45]:

$$D = \left(\prod_{n=1}^{n} d_i\left(\overline{y}_i\right)\right) = d_1\left(\overline{y}_1\right) \times d_2\left(\overline{y}_2\right) \times ... \times d_n\left(\overline{y}_n\right) \tag{6.6}$$

According to Moser et al. (2020) [57] and Arndt et al. [231], the desirability can be reformulated as the combination of the maximization of the expected response $\overline{y}_i$ as well as the minimization of the expected process variability $\overline{v}_{i,d}$ including the weighting factors w_i ($\sum w_i = 1$). $\overline{v}_i$ is provided by the difference of the 10% and 90% quantiles. To maximize the expected response, the desirability $d\left(\overline{y}_i\right)$ is calculated as follows:

$$d\left(\overline{y}_i\right) = \left(\frac{\overline{r}_i - L\left(\overline{r}\right)}{U\left(\overline{r}\right) - L\left(\overline{r}\right)}\right) \tag{6.7}$$

The minimization of the expected process variability $d\left(\overline{v}_{i,d}\right)$ and the resulting overall desirability D with the weighting factors w_1 and w_2 of 0.8 and 0.2 [57] can be formulated as:

$$d\left(\overline{v}_i\right) = \left(\frac{\overline{v}_{i,d} - U\left(\overline{v}\right)}{L\left(\overline{v}\right) - U\left(\overline{v}\right)}\right) \tag{6.8}$$

$$D = w_1 \times d\left(\overline{y}_i\right) + w_2 \times d\left(\overline{v}_{i,d}\right) \tag{6.9}$$

Based on Equation 6.9, the calculated D serves as an additional tool to assess the trade-off between the expected response and variabilty of the results from the RSM and Monte-Carlo-based simulations.

6.5 Evaluation of experimental and simulated data

6.5.1 Root mean square deviation RMSD

As described in Subsection 6.2.1, the initial values of the model parameters are adapted by minimizing an objective function J (Equ. 6.1) defined as the weighted sum of squared residuals, also referred as root mean square deviation (RMSD), between the simulated $y_{\mathrm{s,m}}$ and experimental data $y_{\mathrm{i,m}}$ for all time points $t_1, t_2, ..., t_{\mathrm{end}}$. The objective J is minimized using the Nelder-Mead algorithm as commonly used for model parameter adaption and model evaluation [45, 59, 144, 240–242]. Alternative methods for model parameter identification and adaption are the adaptive experimental redesign [243–245] and the Bayesian interference method [57, 156]. The RSMD is calculated from the squared difference between the experimental $y_{\mathrm{i,m}}$ and simulated values $y_{\mathrm{s,m}}$ divided by the number of data points n in the data set. k_{w} is used as weighting factor for individual data points.

$$RMSD = k_{\mathrm{w}} \cdot \sqrt{\sum_{m=1}^{n} \frac{\left(y_{\mathrm{s,m}} - y_{\mathrm{i,m}}\right)^2}{n}} \tag{6.10}$$

6.5.2 Normalized root mean square deviation NRMSD

For the normalized root mean square deviation (NRMSD), the RMSD is normalized to the range of the experimental data from the minimum ($min\left(y_{\mathrm{i,m}}\right)$) to the maximum ($max\left(y_{\mathrm{i,m}}\right)$) values [32, 246]:

$$NRMSD = \frac{k_{\mathrm{w}} \cdot \sqrt{\sum_{m=1}^{n} \frac{\left(y_{\mathrm{s,m}} - y_{\mathrm{i,m}}\right)^2}{n}}}{max(y_{\mathrm{i,m}}) - min(y_{\mathrm{i,m}})} \tag{6.11}$$

According to literature, the NRMSD value is used rather than the RSMD value when comparing experimental and simulated data sets of different dimensions [247].

6.5.3 Coefficient of determination R^2

To evaluate the adapted model parameters and the resulting progression of the simulated data, the coefficient of determination R^2 is calculated. The R^2 can range between 1 and minus infinity and is defined as the residual of one minus the ratio of the squared differences between the experimental-simulated data ($y_{\mathrm{i,m}}$ - $y_{\mathrm{s,m}}$) as well as the differences between the

experimental data and their mean ($y_{i,m}$ - $\overline{y}_i$). R^2 is equal to 1, if the simulated data corresponds to the experimental data measured [35, 248].

$$R^2 = 1 - \frac{\sum_{m=1}^{n} \left(y_{i,m} - y_{s,m}\right)}{\sum_{i=1}^{n} \left(y_{i,m} - \overline{y}_i\right)} \tag{6.12}$$

The R^2 can be also used in the evaluation of DoE approaches. In this case, R^2 is a measure for the goodness of the model [249] and is defined as the fraction of the responses variation (variability in the response) that can be explained by the model and attributed to the independent variables (linear or multiple-linear regression model, see Subs. 3.7) [52, 179, 250].
In addition to the R^2, the goodness of fit Q^2 is used to describe the goodness of prediction and is defined as the fraction of the responses variation predicted by the model according to cross validation. Furthermore, the Q^2 can be used to predicted how well new factors will be simulated by the model and model parameters. In literature, typical values of Q^2 and R^2 for a good model prediction and description are >0.6 and >0.75, respectively. [52]

6.5.4 Adjusted coefficient of determination R^2_{adj}

Similar to R^2, the adjusted coefficient of determination R^2_{adj} is used in statistics. As R^2 increases when additional terms are added to the regression model, the use of R^2_{adj} is preferred. R^2_{adj} even decreases when unnecessary terms are added. It can be formulated as [251]:

$$R^2_{adj} = 1 - \frac{n-1}{n-p}\left(1 - R^2\right) \tag{6.13}$$

with n number of observations and p number of explanatory parameters (regression coefficients) included in the regression model (see Subs. 6.3.1).

6.5.5 Mean squared error MSE

The mean squared error (MSE) is used in the modeling of the response surface as a measure for the quality of the prediction [252] and to assess the accuracy of a model [253]. Furthermore, its minimization can be used as a design criterion. In addition to the variance of the prediction, a measure of bias caused by model mis-specification is included in the MSE of prediction. [252] The MSE is defined as [253]:

$$MSE = \sum_{m=1}^{n} \frac{\left(y_{s,m} - y_{i,m}\right)^2}{n} \tag{6.14}$$

6.5.6 Analysis of variance

Significant differences in the model parameter distributions can be identified using analysis of variance (ANOVA) [35, 179]. Furthermore, ANOVA can be used to compare and assess experimental and simulated observations as well as regression models. If one categorical independent variable is used, it is referred to as one-way ANOVA. If two categorical independent variables are used, it is referred to as two-way ANOVA. First, the test value F is calculated and it is determined if one pair of all pairs is significantly different (p-value). According to Mishra et al. (2019), the F value is defined as "the ratio of the variability between groups with the variability of the observations within the groups" [254]. Second, the significantly different pair has to be identified by multiple comparison tests, e.g., least significant difference, Bonferroni test, Tukey's test [254]. A prerequisite to perform ANOVA is variance homogeneity [255] and a continuous scale [254]. The Bartlett's or Levene's test can be used to test for variance homogeneity for normally and non-normally distributed samples, respectively [238]. The H_0 should be tested if the treatment effects τ_i (e.g., change in the process conditions) are all the same [179, 238, 255]:

$$H_0 : \{\tau_1 = \tau_2 = ... = \tau_3\} \tag{6.15}$$

against the alternative hypothesis H_A that treatment effects differ on the response:

$$H_A : \{\text{at least two } \tau \text{ differ}\} \tag{6.16}$$

α is defined as the significance level of the test and reflects the probability of rejecting H_0. The determined p-value indicates the smallest value of α that would allow a rejection of H_0 (difference is significant: reject H_0 if $p < \alpha$) [35, 179].
In this thesis, three different significance levels (p-values) were chosen ($p < 0.001$= highly significant -**, $p < 0.05$= significant -*, $p \geq 0.05$= not significant -n.sig). Subsequently, the means and sample variance were determined.

6.5.7 Sensitivity analysis

According to Lencastre Fernandes et al. (2013), a "mechanistic model should be combined with proper model analysis tools, such as uncertainty and sensitivity analysis" to identify which parameters are responsible for the output uncertainty [163]. Sensitivity analysis can be used to describe and understand relative changes in the bioprocess performance and the

influence of key attributes and interactions, resulting in a more focused optimization [58, 234]. Sensitivity analysis might be particularly required to identify the most critical inputs particularly when highly non-linear or differential equations are used. This could reduce the amount of experimental work.

In sensitivity analysis, a distinction can be made between the local (LSA) and global sensitivity analysis (GSA) [58]. The LSA determines the influence of changes $\Delta\theta$ in a given individual input parameter value θ on the relative change at a given point (location θ_0). The input parameter is varied on the basis of its derived probability distribution. The responses, e.g., final product concentration, are simulated and the relative width of the probability distribution represented by the 10% and 90% quantiles is compared to the input probability distribution [58, 234]:

$$S = \frac{\Delta_{target}}{\Delta_\theta} \tag{6.17}$$

An input variable is classified as significant, if S is above 5% [234]. However, LSA is limited to the analysis of one input parameter at a time. In contrast to this, GSA is a multivariable analysis assessing all input variables simultaneously for all points. The input variables are ranked to determine the average contribution of every parameter over a "stipulated set of input ranges". Therefore, the number of design variables and laboratory effort may be reduced [58].

6.5.8 Standardized regression coefficients

In order to understand the results from multiple linear regression and to identify the most influential factors represented by the estimated response surface regression coefficients $\hat{\beta}_\mathrm{i}$, $\hat{\beta}_\mathrm{i,j}$ and $\hat{\beta}_\mathrm{ii}$, standardized regression coefficients (SRCs) are calculated. The advantages of the SRC are the absence of variable and response units and simpler interpretation of the regression results. The SRCs β'_i, $\beta'_\mathrm{i,j}$ and β'_ii are determined by subtracting the mean $\overline{\mathrm{y}}$ and $\overline{\mathrm{x}}_\mathrm{i}$ of the dependent and independent variables from $\mathrm{y_t}$ and $\mathrm{x_{i,t}}$, dividing by the calculated standard deviation σ_y and $\sigma_\mathrm{x,i}$ and re-run the regression analysis (see Subs. 6.3.1) [256, 257]:

$$y' = \frac{y_t - \overline{y}}{\sigma_y}\ ,\ x' = \frac{x_{i,t} - \overline{x}_i}{\sigma_{x,i}} \tag{6.18}$$

According to Allam et al. (2020), SRCs can be also considered as a global sensitivity analysis method to reveal the influence of different process variables[256] which can be important for process model development and application [258]. The methods of global sensitivity analysis used for process models are described in the literature [234].

6.5.9 Discrete Kolmogorov-Smirnov-Test

The two-sided Kolmogorov-Smirnov test (ks-test) tests a random sample of independent observations to determine whether it originates from a population with normal distribution or other theoretical probability distribution functions [255, 259, 260]. The ks-test is of particular interest since the normal distribution is a main assumption of, e.g., the paired and unpaired t-test, linear regression analysis or ANOVA. In case of a non-normal distribution, an interpretation may not be valid [260]. The ks-test is based on the largest vertical difference between two cumulative distribution functions: the hypothesized (e.g., normal) and empirical distribution (e.g., random sample or distribution of model parameter values). The H_0 is rejected if the test value exceeds the 1-α quantile which can be found in the table of quantiles for the ks-test statistics at the given level of significance α [259–261]. Before performing the ks-test, the sample distributions have to be centered and scaled. Therefore, the mean value of the individual model parameter distribution $\overline{x}$ is subtracted from the model parameter value x_i and then divided by the standard deviation $\hat{\sigma}$ of the sample [186]:

$$x_{centered} = \frac{x_{\mathrm{i,sample}} - \overline{x}_{\mathrm{sample}}}{\hat{\sigma}} \tag{6.19}$$

6.6 Programming structure

All modules of the iterative and uncertainty-based workflow were programmed in MATLAB (2018b). Similar to the structure of the iterative workflow, the program is divided into individual sections which are consecutively executed. An overview is shown in Figure 6.3. The dotted lines indicate a transfer or exchange of data.

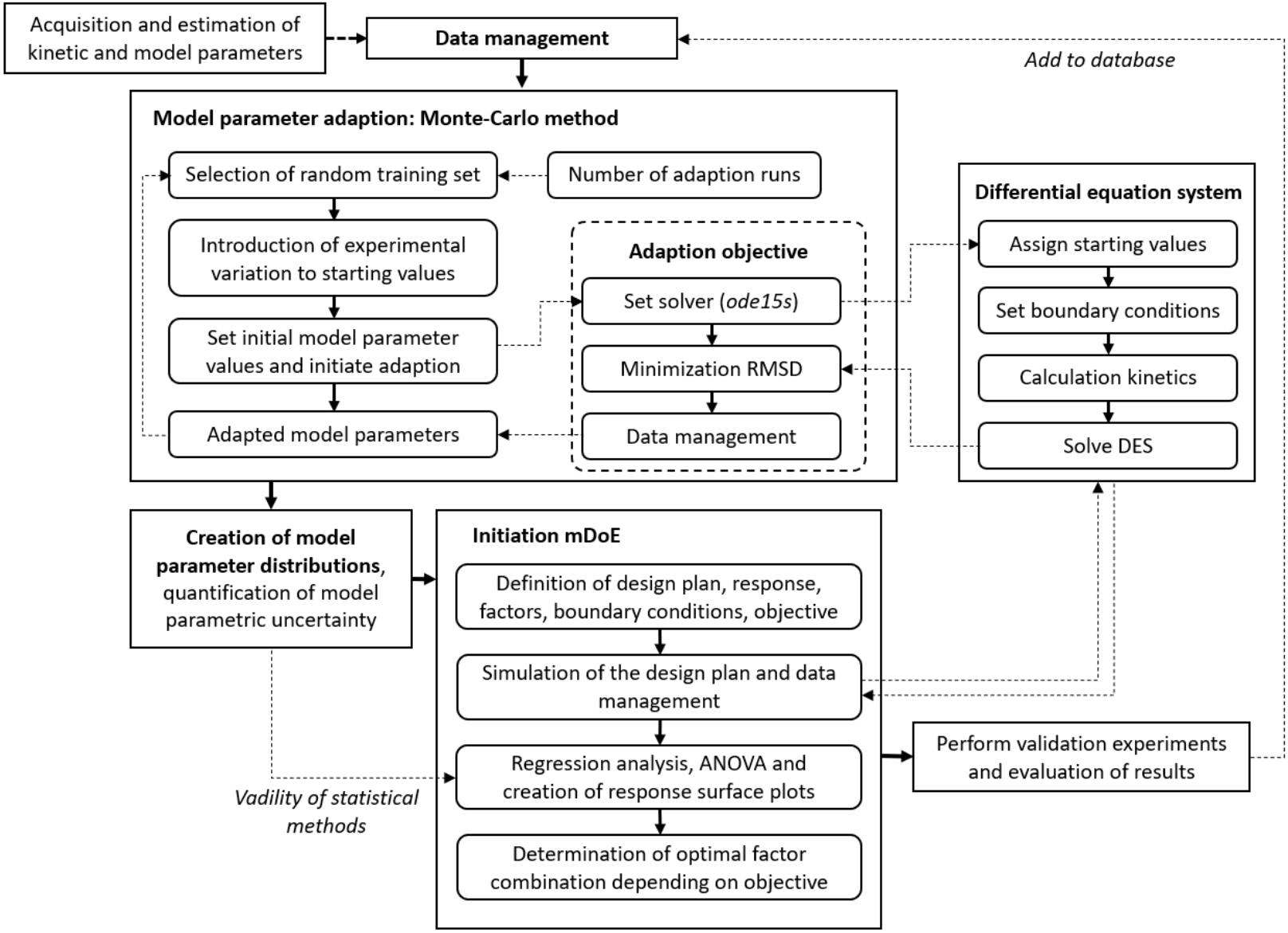

Figure 6.3: Overview programming structure of the model parameter adaption, process model and mDoE, programming code written in MATLAB (2018b).

First, data from various sources, such as Excel workbooks, is acquired and embedded into the program (stored in matrices). In addition, among others, kinetic parameters are determined using parameter estimation methods (*nlinfit* algorithm) and data from preliminary experiments. The kinetic parameters are stored as initial model parameters and can be recalled during the course of the program.

Second, the model parameter adaption module, in which the model parameters are repeatedly adapted under consideration of experimental uncertainty, is initiated by selection of a random training set (*randperm* algorithm). The number of training sets out of the total set of experimental data is defined beforehand. The resulting experimental data are then transferred to another function where the experimental variation is introduced to the starting values of the individual training sets. This is done by the *randn* algorithm. In addition, the parameters of the reactor, e.g., reactor length and diameter, are specified and saved as parameters. Parameters which are to be adapted to experimental data are stored separately from parameters which are constant during the model parameter adaption process, e.g., number of grid nodes and reactor bed bulk density. Subsequently, the *fminsearch* algorithm is used to start the model parameter adaption. The algorithm options are defined, all training set data and parameters are passed to the adaption objective function, and the type and options of an *ode15s* solver are set. In addition, the time interval to be simulated and initial values for all simulated species are defined. In this work, the same time intervals as they are found in the experimental data are used. From this, the differential equation system function is called and after assigning the initial values and setting boundary conditions for the used modeling approach, the kinetics are calculated and the system of differential equations is solved step-wise. Due to the use of the method of lines approach, on the one hand, the differential equations are calculated step-wise in terms of number of grid nodes defined beforehand as well as time-wise. The time step width of the simulated time is specified by the selected solver. After the respective time interval is simulated, the data matrix is passed back to the adaption objective and the RMSD is calculated. Therefore, the simulated data corresponding to the time points of the experimental data are located and stored for RMSD determination. This is done using the *ismembertol* and *find* algorithm. As described in Subsection 6.2.2 the model parameter adaption is repeated until a predefined criterion is reached. The adapted set of model parameters is then saved and the procedure is repeated automatically for a new training data set (*for* loop).

Third, the sets of adapted model parameter values are merged and model parameter distributions are created. On the one hand, this is done to determine the 10% and 90% quantiles (*perc* function), mean (*mean* function) and median (*median* function) values of each individual

distribution. On the other hand, the distributions are tested for normality (*kstest* algorithm) which is prerequisite for statistical test in the subsequent programming module.
Fourth, the mDoE methodology part is initiated. The user is ask which DoE plan is to be used. Accordingly, the corresponding algorithm generating the experimental design plan is selected and integrated into the programming. The resulting factor combination matrix is stored. Furthermore, boundary conditions for each factor are defined. The simulation of the design plan is conducted by transferring the individual factor combinations as initial starting values to the differential equation system. The median values and the 10% and 90% quantile values are defined as parameters. Individual simulations are conducted consecutively using a for loop and the results are stored into a new data matrix. With this data, regression analysis is performed as described in Subsection 6.3.2. For the identification of an optimal factor combination, a minimization algorithm (*fmincon* algorithm) is incorporated into the programming structure and initiated. In addition, the standardized regression coefficients and the desirability are determined as described in Subsection 6.5.8 and 6.4. The calculation is performed automatically and without further user input. As described in the workflow, validation experiments are performed based on the optimal factor combination. After analyzing the experimental data, the new data is stored in the database. The described program can be restarted or used to evaluate, e.g., different experimental designs or factor boundary conditions.

7 Experimental validation of the iterative workflow

The proposed iterative model-assisted workflow was introduced to design and optimize a biocatalytic cofactor regeneration process considering CDP and polyP starting concentrations, flow rate and hydrostatic pressure as experimental factors. Due to the novelty of this reactor concept and reduced information about the reaction system and influence of high hydrostatic pressure on the biocatalyst RpPPK2-3, the application of the mDoE methodology for the design and optimization was justified. The process model was first formulated, the model parameters were estimated from kinetic investigations, and both were iteratively adapted. Process dynamics were predicted well and as a result of the mDoE, a reduced number of experiments had to be performed to evaluate optimal factor combinations for a user-defined experimental space and optimization objective. Due to the iterative structure of the workflow, new experimental data sets, changes in the process model and re-adapted model parameters were added and used after the first iteration round.

7.1 Model-assisted experimental design

As presented in the model-assisted workflow (see Fig. 6.1, I), the process model and parameters of the process were initially formulated and adapted to available experimental data from kinetic experiments by modeling the time- and location-dependent progression of the substrate and CTP concentration and minimizing the RMSD between simulated and experimental data (see Subs. 6.2). The model parameters starting values were estimated from kinetic measurements (see Subs. 5.3.1). Before each model parameter adaption run, a training set with four randomly selected data sets was drawn from all data sets. 1000 runs were performed in total with a standard error of 2% (see Subs. 6.2.2). The individual workflow rounds performed are denoted as mDoE-1, mDoE-2 and following. The rounds in which the termination criterion is checked are denoted as mDoE-1 TC, mDoE-2 TC and following.

7.1.1 Quantification of model parametric uncertainty, mDoE-1

In mDoE-1, the model parameters were adapted to 23 data sets, whereby a training set of 4 data sets was randomly drawn for each adaption run beforehand (see Suppl. Tab. A1 and A.3). The model parameter starting values are summarized in Table 7.1. From the model parameter distributions resulting from the individual model parameter adaption runs, the median and the

10% and 90% quantile model parameter as well as the 95% CI values were determined (see Tab. 7.1).

Table 7.1: Adapted process model parameters: Median, 10% and 90% quantile and 95% CI values (mDoE-1).

	v_{max} [mmol min^{-1} mg^{-1}]	$K_{M,CDP}$ [mmol L^{-1}]	$K_{M,polyP}$ [mmol L^{-1}]	$K_{I,polyP}$ [mmol L^{-1}]	k_d [min^{-1}]
Start value	0.0166	32.16	219.8	8.40	3.0 10^{-4}
Mean	0.0249	20.59	54.22	13.69	2.7 10^{-4}
10% quantile	0.0209	8.135	0.131	10.72	1.2 10^{-4}
Median	0.0244	20.61	35.53	13.13	3.0 10^{-4}
90% quantile	0.0292	30.32	132.9	17.35	3.7 10^{-4}
$CI_{0.025}$	0.0245	19.88	50.46	13.49	2.7 10^{-4}
$CI_{0.975}$	0.0253	21.31	57.98	13.87	2.8 10^{-4}

Normality of the model parameter distributions is one of the main assumption of various statistical tools [260]. Using the ks-test, the individual model parameter distributions of v_{max}, $K_{M,CDP}$, $K_{M,polyP}$, $K_{I,polyP}$ and k_d were tested for normality (see Subs. 6.5.9). For all model parameter distributions, the H_0 was rejected at the 10% significance level. Therefore, the individual centered and scaled model parameter distributions did not follow a normal distribution. However, according to Eckey et al. (2005), the deviations of the model parameter distributions from the ideal normal distribution curve can be neglected in case of a sufficiently large scope of observations (central limit theorem). The obtained distributions converge and are approximated by a normal distribution. [262] Since the 10% and 90% quantile values can be used to determine the described uncertainty propagation and the corresponding uncertainty band, these model parameter values were used for further simulation. The 95% CI only defines the probability that the true mean value is contained in the interval.

Based on this, the progression of the CTP concentration over time and the corresponding uncertainty band were simulated which is exemplarily shown together with the experimental data for data set 23 (see Fig. 7.1).

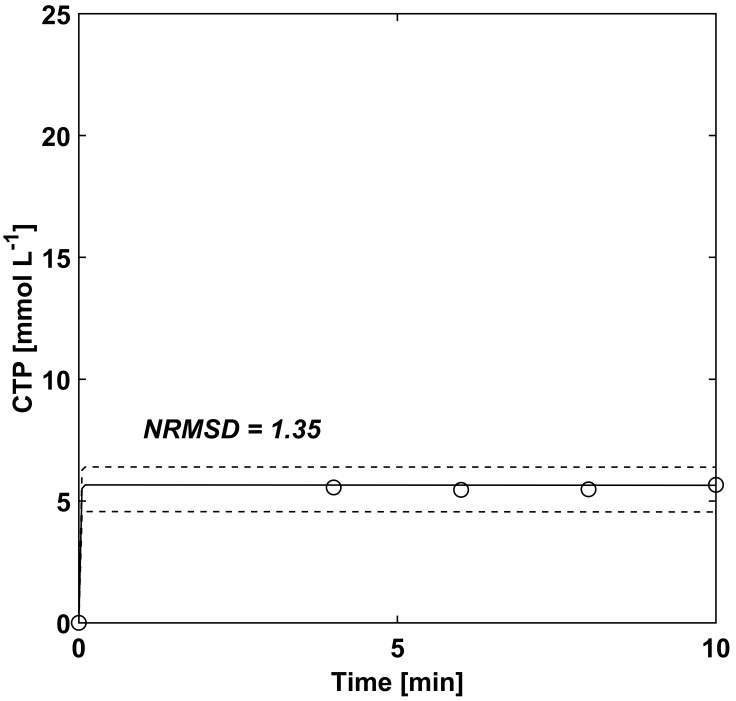

Figure 7.1: Progression of the experimental CTP data (circle marker) and predicted CTP concentration data over time for median (solid line) and 10% and 90% quantile model parameter values (dashed lines), experimental data set 23 (mDoE-1). For experimental conditions see Suppl. Tab. A.1

In Figure 7.1, the simulated CTP concentration at the reactor outlet over time based on the median and the 10% and 90% quantile values (uncertainty band) is represented by the solid and the dashed lines, respectively. A NRMSD between the simulated and experimental data of 1.35 was determined. The experimental data points of data set 23 (circle marker) were predicted well.

Furthermore, the experimental data of all data sets were adequately simulated, but over-predicted at the beginning of the regeneration process with a calculated NRMSD value of 1.81 (deviation from the dashed line, see Fig. 7.2).

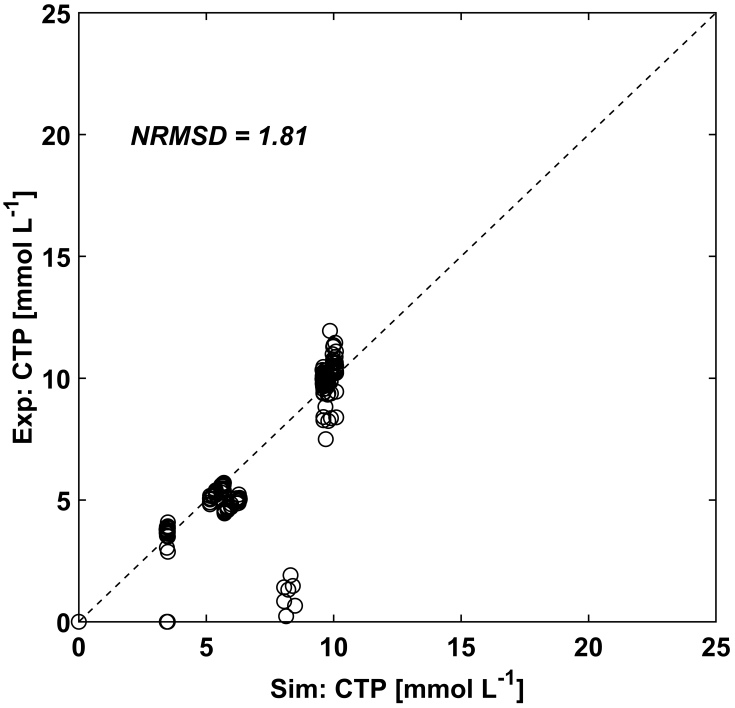

Figure 7.2: Comparison of experimental and simulated CTP concentration for median model parameter values, all data sets mDoE-1. For experimental conditions see Suppl. Tab. A.1.

In particular, the experimental data of the data sets one to nine measured between the start of the process and a process time of one minute were overpredicted by the process model. This could be attributed to start-up effects or rearrangement of the biocatalytic packed reactor bed [117]. Nevertheless, the formulated process model provided a sufficient prediction based on the estimated model parameter distributions.

7.1.2 Selection of the experimental design

As proposed in the workflow (see Fig. 6.1, II), a suitable experimental design was selected to further increase the process understanding and enable RSM. First, to limit the experimental and computational effort, four relevant continuous process variables that have an impact on process performance were selected: initial substrate concentration (polyP and CDP), flow rate and hydrostatic pressure. Maximizing the CTP concentration at the reactor outlet after a process time of 15 minutes was selected as the optimization objective and taken into account in the subsequent desirability evaluation. Second, the CCD, BBD and D-optimal designs were compared with regard to R^2, R^2_{adj} and MSE values (see Suppl. Tab. A.4 for coded experimental design matrices). These experimental designs were selected as they are commonly used for RSM [178, 188]. Furthermore, the adequacy of first-order linear (see Equ. 3.4) and interaction as well as second-order regression models with quadratic terms (see Equ. 3.6) were assessed.

In Table 7.2 the determined values for R^2, R^2_{adj} and MSE are summarized.

Table 7.2: Selection of the experimental design for process design and optimization.

Design	**Model structure**	R^2	R^2_{adj}	**MSE**
CCD	linear	0.78	0.74	8.94
	interaction	0.85	0.78	7.62
	quadratic	0.96	0.92	2.78
BBD	linear	0.63	0.57	37.44
	interaction	0.75	0.59	35.85
	quadratic	0.94	0.86	11.83
D-optimal	linear	0.70	0.65	39.63
	interaction	0.91	0.86	15.73
	quadratic	0.96	0.93	8.47

For all three experimental designs tested, the R^2 and R^2_{adj} values increased with increasing complexity of the RSM regression model (from linear to quadratic). The R^2_{adj} for CCD increased from 0.75 to 0.96 which means that 21% more variance was explained by the regression model. In contrast, the MSE value decreased from 8.94 to 2.78. Therefore, it can be assumed that the additional quadratic terms and the corresponding regression coefficients are needed to adequately reflect the curvature and the simulated responses. This is supported by the fact that the use of a second-order regression model is often recommended in RSM. The results of the three individual design plans showed that the obtained coefficients for the second-order regression model of the CCD and D-optimal design plans are similar, but higher than for the BBD. Due to the lower MSE values, 2.78 compared to 8.47 for CCD and D-optimal design, the CCD design plan and second-order regression model were selected for further use in the mDoE methodology and iterative workflow rounds.

In Table 7.3 the boundary conditions for every factor with five factor levels based on the CCD are summarized.

Table 7.3: Factor levels of flow rate, CDP and polyP concentration, hydrostatic pressure for Central Composite Design plan with 5 factor levels.

Factor	**Level -2**	**Level -1**	**Level 0**	**Level 1**	**Level 2**
Flow rate [mL min^{-1}]	0.15	0.3	0.75	1.05	1.35
CDP [mmol L^{-1}]	5	15	35	55	65
polyP [mmol L^{-1}]	5	15	35	55	65
Pressure [bar]	0	200	400	600	800

As described in Subsection 3.6.2, the design of the CCD consisted of 18 corner points, 7 replicated center points and 6 star points. It should be noted that the concentrations of CDP and polyP were constraint to be in a range in which the reaction was not significantly affected by inhibition, as shown by kinetic data or solubility problems (data not shown). The experimental combinations for the CCD resulting from the coded combinations and defined boundary conditions for every factor are summarized in Table 7.4.

Table 7.4: Factor combinations for four factor CCD and all workflow rounds.

Run	**Flow** [L min^{-1}]	$\mathbf{c_{CDP}}$ [mmol L^{-1}]	$\mathbf{c_{polyP}}$ [mmol L^{-1}]	**Pressure** [bar]
1	0.0003	15	15	200
2	0.0003	15	15	600
3	0.0003	15	55	200
4	0.0003	15	55	600
5	0.0003	55	15	200
6	0.0003	55	15	600
7	0.0003	55	55	200
8	0.0003	55	55	600
9	0.00105	15	15	200
10	0.00105	15	15	600
11	0.00105	15	55	200
12	0.00105	15	55	600
13	0.00105	55	15	200
14	0.00105	55	15	600
15	0.00105	55	55	200
16	0.00105	55	55	600
17	0.00015	35	35	400
18	0.00135	35	35	400
19	0.00075	5	35	400
20	0.00075	65	35	400
21	0.00075	35	5	0
22	0.00075	35	65	800
23	0.00075	35	35	400
24	0.00075	35	35	400
25	0.00075	35	35	400
26	0.00075	35	35	400
27	0.00075	35	35	400
28	0.00075	35	35	400
29	0.00075	35	35	400
30	0.00075	35	35	400
31	0.00075	35	35	400

7.1.3 Experimental analysis of the response surface, mDoE-1

Prior to creating the regression model and analyzing the response surface of mDoE-1 by regression analysis, the content of the packed-bed reactor used for kinetic and preliminary experiments was exchanged for new and unused immobilizate. Subsequently, all factor combinations suggested by the CCD for 4 factors (see Tab. 7.4) were simulated for the specifications of the newly packed-bed, as even slight deviations in the immobilizate weight and packed-bed bulk density have a considerable effect on the results (see Suppl. Tab. A.1). The regression coefficients of a second-order polynomial regression model were estimated by multi-linear regression analysis and the regression model was tested for significance using ANOVA (see Subs. 6.5.6). All regression coefficients, the corresponding p-values and the results of the model evaluation are summarized in Table 7.5 (for all values of all workflow rounds see Suppl. Tab. A.5 and Tab. A.6).

Table 7.5: Estimated coefficients of second-order linear regression model considering linear, interaction and quadratic effects and corresponding p-values for mDoE-1. *=significant, n.sig.=not significant (0.05 significance level).

Coefficient	Coefficient value	p-value	Significance
$\hat{\beta}_0$	8.346	0.063	n.sig.
$\hat{\beta}_{\text{Flow}}$	$-3.04 \cdot 10^{4}$	$5.70 \cdot 10^{-5}$	*
$\hat{\beta}_{\text{CDP}}$	0.541	$1.51 \cdot 10^{-4}$	*
$\hat{\beta}_{\text{polyP}}$	0.173	$1.34 \cdot 10^{-1}$	n.sig.
$\hat{\beta}_{\text{Pressure}}$	$2.91 \cdot 10^{-3}$	0.752	n.sig.
$\hat{\beta}_{\text{Flow,CDP}}$	-282.0	$1.09 \cdot 10^{-4}$	*
$\hat{\beta}_{\text{Flow,polyP}}$	65.02	0.257	n.sig.
$\hat{\beta}_{\text{Flow,Pressure}}$	-1.332	0.813	n.sig.
$\hat{\beta}_{\text{CDP,polyP}}$	$-1.21 \cdot 10^{-3}$	0.263	n.sig.
$\hat{\beta}_{\text{CDP,Pressure}}$	$2.76 \cdot 10^{-5}$	0.795	n.sig.
$\hat{\beta}_{\text{polyP,Pressure}}$	$-4.55 \cdot 10^{-6}$	0.966	n.sig.
$\hat{\beta}_{\text{Flow}^2}$	$1.79 \cdot 10^{7}$	$2.94 \cdot 10^{-5}$	*
$\hat{\beta}_{\text{CDP}^2}$	$-2.12 \cdot 10^{-3}$	$9.90 \cdot 10^{-2}$	n.sig.
$\hat{\beta}_{\text{polyP}^2}$	$-2.97 \cdot 10^{-3}$	$2.59 \cdot 10^{-2}$	*
$\hat{\beta}_{\text{Pressure}^2}$	$-2.19 \cdot 10^{-6}$	0.779	n.sig.

The estimated regression coefficients differ in the coefficient estimate and their corresponding p-values. Only the coefficients for the variable flow and CDP, the interaction effect of flow and CDP as well as the quadratic effect of flow and polyP contributed significantly to the regression

model output. All other coefficients are classified as insignificant. However, since new data sets are to be added and model parameters are to be adapted as the workflow progresses, all regression coefficients were still included in the regression model.

To create the three-dimensional response surface, the design space was calculated completely based on the estimated regression coefficients. Therefore, two factors were randomly varied 1000 times between the upper and lower factor limits and the responses were calculated (see Subs. 6.3.1, Equ. 6.4), while the other two factors were held constant at mid-level values. The resulting response surface for the observed responses depending on the CDP and polyP starting concentrations and constant mid-level flow rate of 0.75 mL min^{-1} and hydrostatic pressure of 400 bar is shown in Figure 7.3 (part A).

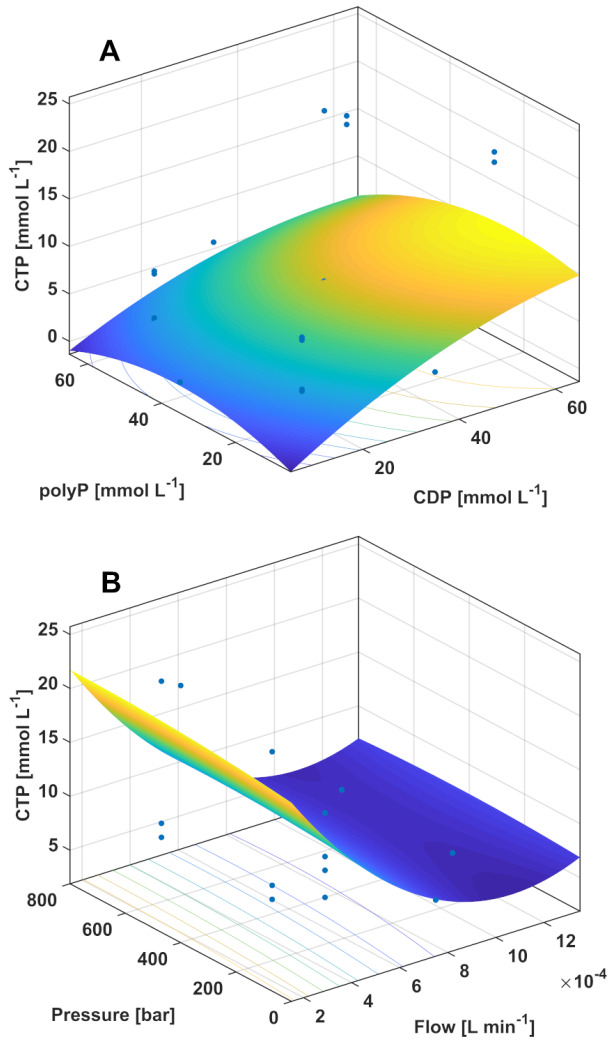

Figure 7.3: **A**: Response surface for predicted CTP concentration at reactor outlet based on CCD and shown for varying CDP and polyP starting values. Other factors set to constant mid-level values. **B**: Response surface for predicted CTP concentration at reactor outlet based on CCD and shown for varying flow rate and hydrostatic pressure starting values. Other factors set to constant mid-level values. For experimental conditions see Suppl. Tab. A.1 run 23.

An adjusted R^2 of 0.92 was obtained for the response surface regression model considering the four factors with a good fit based on F-statistics (p-value of $2.64 \cdot 10^{-8}$ with significance level $p<0.05$, see Subs. 6.5.6) indicating an adequate approximation of the simulated responses by the regression predictions. It can be seen that the CTP concentration increased with increasing starting concentrations of CDP and polyP. The highest CTP concentration was obtained at polyP concentrations above 20 mmol L^{-1}, but decreased with further increases of polyP concentration.
In Figure 7.3 (part B), the response surface depending on the flow rate and hydrostatic pressure with constant mid-level CDP and polyP starting concentration is depicted. With increasing flow rate, the CTP concentration decreased and increased again for flow rates higher 1.05 mL min^{-1}. Compared to the flow rate, the increasing hydrostatic pressure had only a comparatively minor influence on the CTP concentration. It was assumed that the proven effect of hydrostatic pressure (see Subs. 5.3.2) may have been masked by the greater influence of the flow rate on the response. Nevertheless, highest CTP concentrations were obtained for minimum flow rate and maximum hydrostatic pressure. In contrast to the response surface reflecting the influence of CDP and polyP, negligible interaction effects between the flow rate and hydrostatic pressure were observed.

To identify the most important regression coefficient, the SRCs were determined and summarized in Table 7.6 (for all values of all workflow rounds see Suppl. Tab. A.7).

Table 7.6: Standardized regression coefficients of multiple linear regression model considering linear, interaction and quadratic effects (mDoE-1).

Coefficient	**SRC value**
β'_0	-0.055
β'_{Flow}	-0.705
β'_{CDP}	0.451
β'_{polyP}	-0.092
β'_{Pressure}	0.031
$\beta'_{\text{Flow,CDP}}$	-0.251
$\beta'_{\text{Flow,polyP}}$	0.058
$\beta'_{\text{Flow,Pressure}}$	-0.013
$\beta'_{\text{CDP,polyP}}$	-0.056
$\beta'_{\text{CDP,Pressure}}$	0.014
$\beta'_{\text{polyP,Pressure}}$	-0.002
β'_{Flow^2}	0.304
β'_{CDP^2}	-0.098
β'_{polyP^2}	-0.138
$\beta'_{\text{Pressure}^2}$	-0.012

With an estimated absolute value of 0.705 and 0.451 the flow rate and CDP starting concentration had the greatest influence on the response. This was confirmed by the strong curvature in the response surface shown in Figure 7.3 (part A and B). The SRC $\beta'_{\text{flow,CDP}}$ with a value of 0.251 describing the interaction effect of the flow rate and CDP starting concentration was also classified as important. The SRCs for polyP and hydrostatic pressure as well as the flow-rate-hydrostatic-pressure interaction, with 0.091, 0.031 and 0.013, were comparatively low. Subsequently, the output of the response function including the regression coefficients and 1000 random factor combinations was maximized using a minimization function *fmincon* and predefined objective (MATLAB, R2018b).

The highest CTP concentration of 29.98 mmol L^{-1} was predicted for a flow rate of 0.15 mL min^{-1}, 65 mmol L^{-1} CDP, 17.2 mmol L^{-1} polyP and 400.5 bar hydrostatic pressure. The resulting factor combination should then be validated and carried out experimentally. The progression of the measured CTP concentration over time is shown in Figure 7.4.

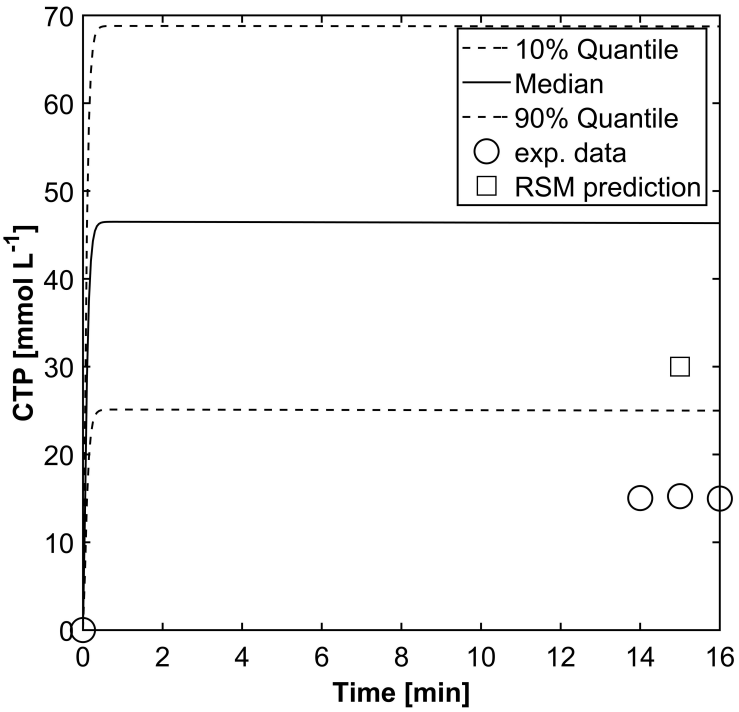

Figure 7.4: Workflow round mDoE-1. Predicted CTP concentration by RSM (square), experimental CTP data for the suggested optimal factor combination (circle) and simulated CTP concentration by the process model (solid and dashed lines, median and 10% and 90% quantile model parameter values) over time. For experimental conditions see Suppl. Tab. A.1.

As depicted in Figure 7.4, a CTP concentration of 14.97 mmol L^{-1} was measured (shown as open circle marker) at the reactor outlet after a process time of 15 min. As a consequence, the predicted CTP concentration of 29.98 mmol L^{-1} (denoted as square marker, based on RSM) was not reached with the suggested factor combination. Furthermore, it can be seen that the process model over-predicted the CTP concentration for the median and 10% and 90% quantile model parameter values. Therefore, the process model was not capable to reflect the process dynamics for the given experimental point. This could be related to a lack of experimental data in the lower flow range area below 0.5 mL min^{-1} for the adaption of model parameters (see Subs. 6.2). The lack in prediction capabilities for extreme areas of the design space can therefore be seen as detrimental to the quality of description.

As a result, the estimation of optimal process conditions based on the Monte-Carlo simulations (Subs. 6.3.3) was not performed. Instead, obtained data sets were used to re-adapt the model parameters and to improve the predictive capabilities of the process model over the entire design space, as proposed in the workflow (see Fig. 6.1). Furthermore, changes in the process model were made and an equation describing the reduction of the reaction rate with decreasing flow rate was implemented. The data from the first round of mDoE experiments was then

added to the database in order to re-adapt the model parameters of the process model (30 data sets in total).

7.1.4 Experimental analysis of the response surface, mDoE-2

In Figure 7.5 the reduction of the specific activity with decreasing flow rate (circle marker) is shown [117].

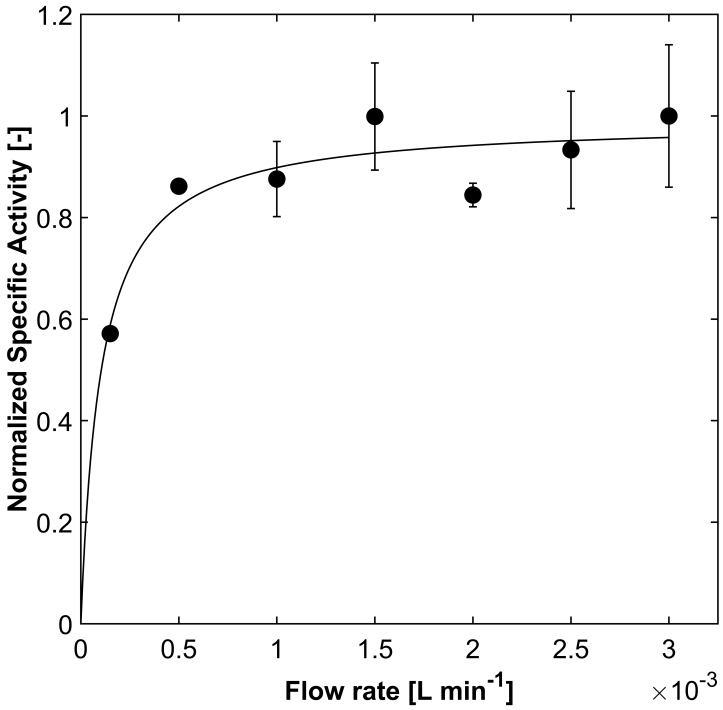

Figure 7.5: Determination of mass transfer limitation as a function of the flow rate. Flow rate = 0.15 mL min^{-1}-3.0 mL min^{-1}, $c_{CDP,0}$ = 30 mmol L^{-1}, $c_{polyP,0}$ = 44 mmol L^{-1}, $m_{RpPPK2-3}$ = 0.166 g, n = 3.

The kinetic equation considering the reduction in the reaction rate v_{red} with decreasing flow rate $\dot{V}$ was formulated as:

$$v_{red} = \frac{0.99 \cdot \dot{V}}{1.02 \cdot 10^{-4} + \dot{V}} \tag{7.1}$$

The predicted normalized specific activity v_{red} (solid line) is depicted in Figure 7.5 as a function of the flow rate from $0.15 - 3.0$ mL min^{-1}. Based on this, the adapted kinetic equation was then changed to:

$$v\,(\mathrm{i}) = v_{\mathrm{red}} \cdot v_{\mathrm{p,Dsig}} \cdot c_{\mathrm{Enz}}\,(\mathrm{i}) \cdot v_{\mathrm{max}} \cdot \frac{c_{\mathrm{CDP}}\,(\mathrm{i})}{K_{\mathrm{m,CDP}} + c_{\mathrm{CDP}}\,(\mathrm{i})} \cdot \frac{c_{\mathrm{polyP}}\,(\mathrm{i})}{K_{\mathrm{m,polyP}} + c_{\mathrm{polyP}}\,(\mathrm{i}) \cdot \left(1 + \frac{c_{\mathrm{polyP}}\,(\mathrm{i})}{K_{\mathrm{I,polyP}}}\right)} \quad (7.2)$$

After implementing the kinetic equation into the process model, the second workflow round (mDoE-2) was started and model parameters were re-adapted (see Suppl. Tab. A.3) using the adapted process model and data sets (see Fig. 6.1, I). The median, 10% and 90% quantile model parameter values were calculated for all model parameter distributions (see Tab. 7.7).

Table 7.7: Adapted process model parameters: Median, 10%/90% quantile and 95% CI values (mDoE-2).

	v_{max} [mmol min^{-1} mg^{-1}]	$K_{\mathrm{M,CDP}}$ [mmol L^{-1}]	$K_{\mathrm{M,polyP}}$ [mmol L^{-1}]	$K_{\mathrm{I,polyP}}$ [mmol L^{-1}]	k_{d} [min^{-1}]
Start value	0.0166	32.16	219.8	8.40	3.0 10^{-4}
Mean	0.0319	18.88	74.75	15.69	2.7 10^{-4}
10% quantile	0.026	7.89	0.19	11.88	$1.33 \cdot 10^{-4}$
median	0.030	17.07	77.69	15.29	$2.83 \cdot 10^{-4}$
90% quantile	0.037	27.52	125.84	19.73	$3.86 \cdot 10^{-4}$
$\mathrm{CI}_{0.025}$	0.0308	17.63	71.22	15.47	$2.6 \cdot 10^{-4}$
$\mathrm{CI}_{0.975}$	0.0330	20.14	78.29	15.90	$2.8 \cdot 10^{-4}$

Compared to the adapted model parameters of mDoE-1, all model parameter values changed after re-adaption to new data sets. In particular, the maximum specific activity v_{max} and $\mathrm{K_{M,polyP}}$ increased and decreased, respectively. The deactivation rate $\mathrm{k_d}$ remained nearly unchanged. Using the re-adapted model parameter, all experimental data sets were simulated. The results of mDoE-2 TC (check of termination criterion) are shown in Figure 7.6.

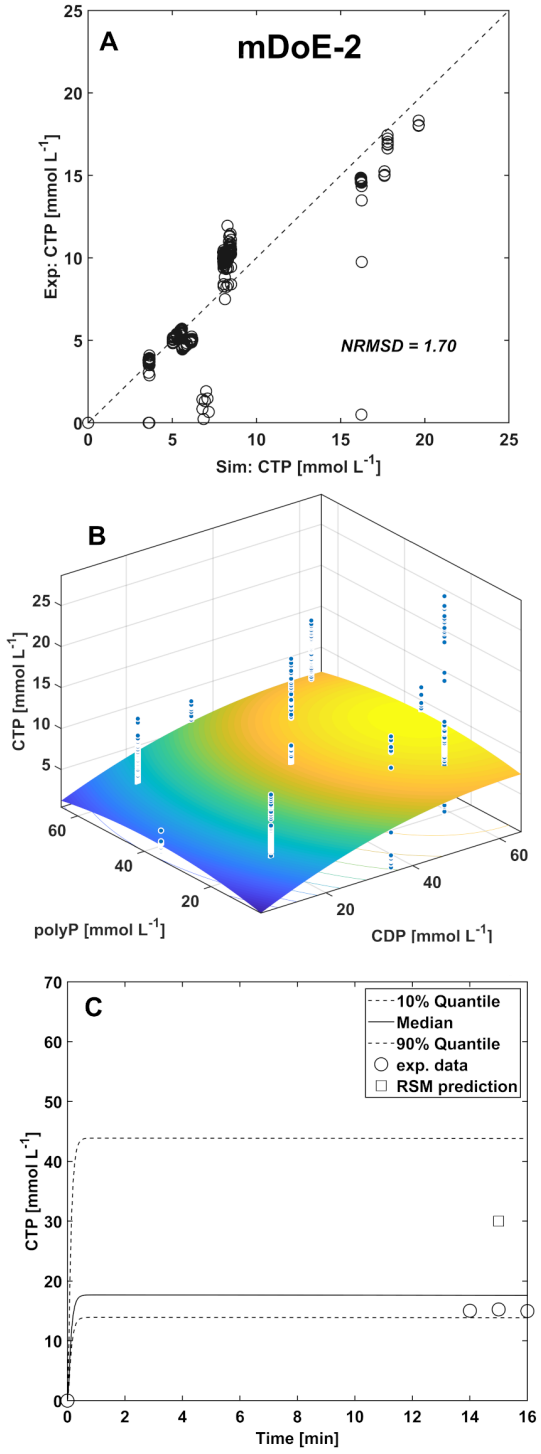

Figure 7.6: Workflow round mDoE-2 TC. **A**: Comparison of experimental and simulated CTP concentration for median model parameter values, all data sets (mDoE-2 TC). **B**: Monte-Carlo-based response surface (n=30) for predicted CTP concentration at reactor outlet and varying CDP-polyP starting concentrations. Flow rate and hydrostatic pressure set to constant mid-level values (mDoE-2 TC). **C**: Predicted CTP concentration by RSM (square, mDoE-1), experimental CTP data for the suggested optimal factor combination (circle, mDoE-1) and simulated CTP concentration by the process model (solid and dashed lines, median and 10%/90% quantile model parameter values of mDoE-2 TC, respectively) over time. For experimental conditions see Suppl. Tab. A.1.

The comparison of all experimental and simulated data points showed that experimental data points were particularly over-predicted at the beginning of the process for data sets with comparable high CDP and polyP starting concentrations (deviation from the dashed line, see Fig. 7.6, part A). Nevertheless, the NRMSD value of 1.70 indicates that the final CTP concentration can be predicted well over a wide experimental space from 5 to 65 mmol L^{-1} CDP and polyP. As shown in Figure 7.6 (part C), the prediction of the CTP concentration at the reactor outlet for the favorable factor combination determined for mDoE-1 was improved. The experimental data were located within the uncertainty band.

Similar to the first round of the workflow (mDoE-1) the applied enzyme concentration as well as the bulk density of the packed-bed were first calculated before the workflow was continued. The calculated values are summarized in Supplementary Table A.1. Subsequently, the termination criteria was checked (mDoE-2 TC, see Subs. 6.1). Therefore, the design plan was simulated for the reactor specifications of mDoE-1 and the maximal CTP concentration at the reactor outlet based on RSM, Monte-Carlo simulations, response surface analysis and the resulting optimal factor combination was determined. A CTP concentration of 18.2 mmol L^{-1} at a flow rate of 0.15 mL min^{-1}, 65 mmol L^{-1} CDP, 33.3 mmol L^{-1} polyP and 400.8 bar hydrostatic pressure was predicted. The corresponding CDP-polyP response surface with constant mid-level flow rate and hydrostatic pressure values is shown in Figure 7.6 (part B). A theoretical CTP concentration increase of 21.3% compared to the mDoE-1 experiment (CTP concentration of 14.97 mmol L^{-1}) was achieved, thus the workflow and the simulation of the design plan were continued as proposed.

On this basis, all factor combinations of the design plan were simulated (see Suppl. Tab. A.8), regression coefficients of the response surface were estimated (see Suppl. Tab. A.5) and the surface was plotted for the CDP-polyP factor combinations (flow rate and hydrostatic pressure at constant mid-level, see Fig. 7.7).

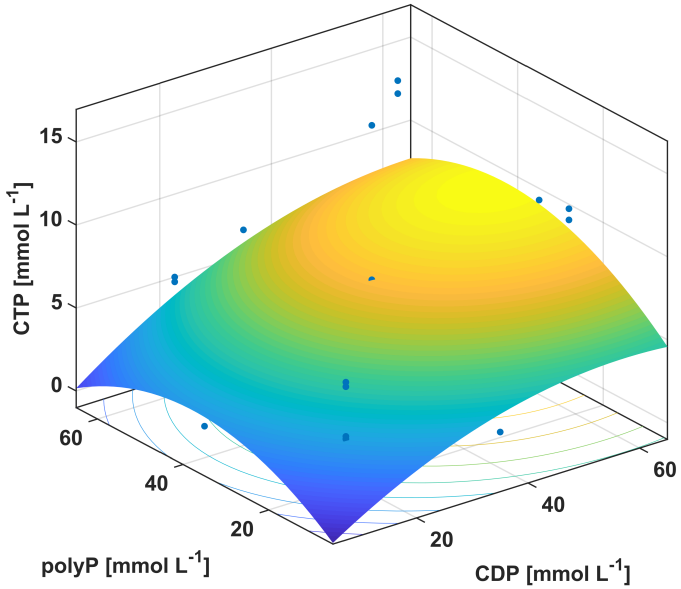

Figure 7.7: Response surface for predicted CTP concentration at reactor outlet based on CCD and shown for varying CDP and polyP starting concentrations. Other factors set to constant mid-level values. For experimental conditions see Suppl. Tab. A.1 run 26 (mDoE-2).

With increasing CDP and polyP concentration, CTP concentration increased whereas the highest concentration was reached for maximum CDP and 40 mmol L^{-1} polyP. This was confirmed by the determined absolute SRCs for CDP and polyP of 0.527 and 0.1663, respectively (see Suppl. Tab. A.7). As shown for mDoE-1, the flow rate and hydrostatic pressure increased the CTP concentration with decreasing and increasing factor values, respectively (see Fig. 7.8).

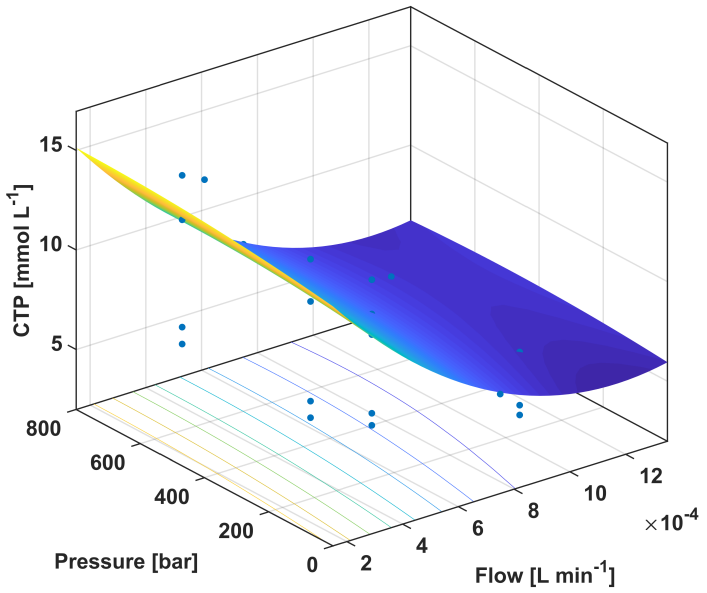

Figure 7.8: Response surface for predicted CTP concentration at reactor outlet based on CCD and shown for varying flow rate and hydrostatic pressure starting values. Other factors set to constant mid-level values. For experimental conditions see Suppl. Tab. A.1 run 26 (mDoE-2).

Despite the comparatively low influence of the hydrostatic pressure reflected by the regression model and regression coefficients, it contributed to the maximization of the defined response, as shown in preliminary experiments. In these experiments, an activity increase of up to 12.5% was observed at a hydrostatic pressure of 600 bar [117]. SRCs of 0.681 and 0.045 were determined for the flow rate and hydrostatic pressure (see all SRCs Suppl. Tab. A.7). Consequently, the flow rate can be regarded as the most important factor.

To quantify the propagated uncertainty for every factor combination and the resulting possible range of all observed responses, Monte-Carlo simulations were performed (n=30, see Subs. 6.3.3). The model parameter values were drawn by LHS from the model parameter distributions and the 10% and 90% quantile values were calculated. Based on this, the response surface representing the propagated experimental uncertainty was estimated and an optimal factor combination was determined. The response surface for the CDP-polyP factor combination and constant flow rate and hydrostatic pressure factor values is shown in Figure 7.9 (part A).

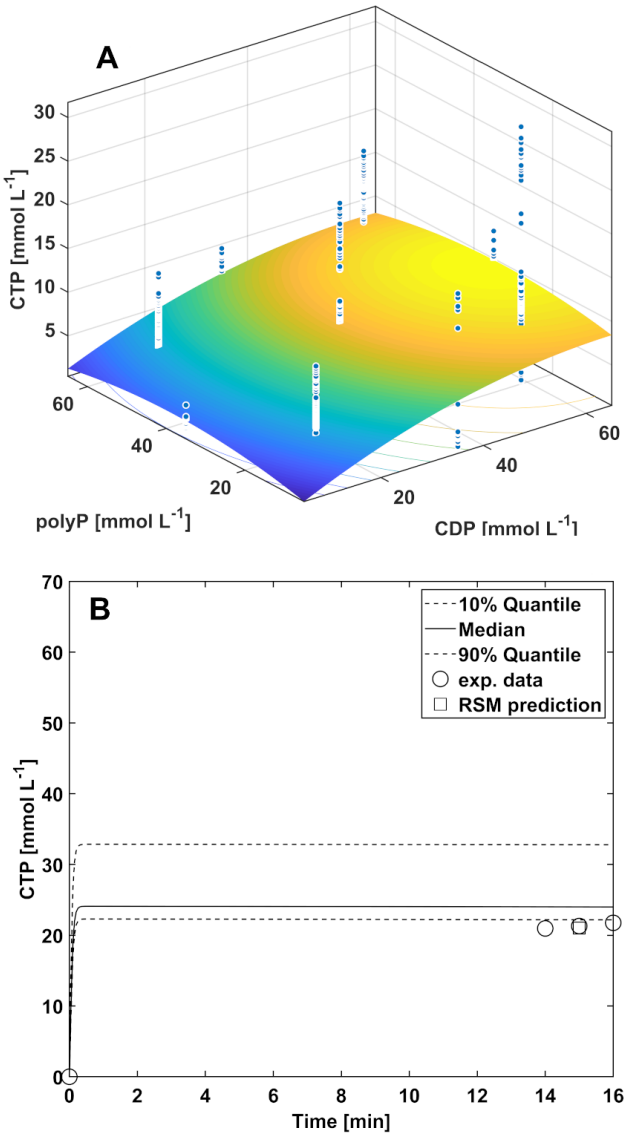

Figure 7.9: Workflow round mDoE-2. **A**: Monte-Carlo-based response surface (n=30) for predicted CTP concentration at reactor outlet and varying CDP-polyP starting concentrations. Flow rate and hydrostatic pressure set to constant mid-level values (mDoE-2). **B**: Predicted CTP concentration by RSM (square), experimental CTP data for the suggested optimal factor combination (circle) and simulated CTP concentration by the process model (solid and dashed lines, median and 10%/90% quantile model parameter values, respectively) over time (mDoE-2). For experimental conditions see Suppl. Tab. A.1.

A total of 930 experiments were simulated, which are represented by the blue circle markers (see Fig. 7.9, part B)). The mean, 10% and 90% quantiles and interquartile distances of the responses are summarized in Supplementary Table A.9. The regression coefficients of the estimated response surface can be found in Supplementary Table A.5. It can be seen that CDP starting concentration had an impact on the final CTP concentration whereas the influence of polyP was comparatively low, as indicated by the slight response surface curvature. A greater

influence would have been expected for higher polyP concentrations as shown by the kinetic data of Schmalle et al. (2024) [117].

As previously described and proposed in the workflow, the weighted desirability of the Monte-Carlo simulations should be determined to evaluate the trade-off between maximizing the predicted response and minimizing the process variability (see Fig. 6.1, II). 0.8 and 0.2 were used as weighting factors for mean response and interquartile distance, reflecting the variability of the Monte-Carlo-based simulations. The resulting desirability surfaces for the CDP-polyP and flow-rate-hydrostatic-pressure factor combinations are shown in Figure 7.10.

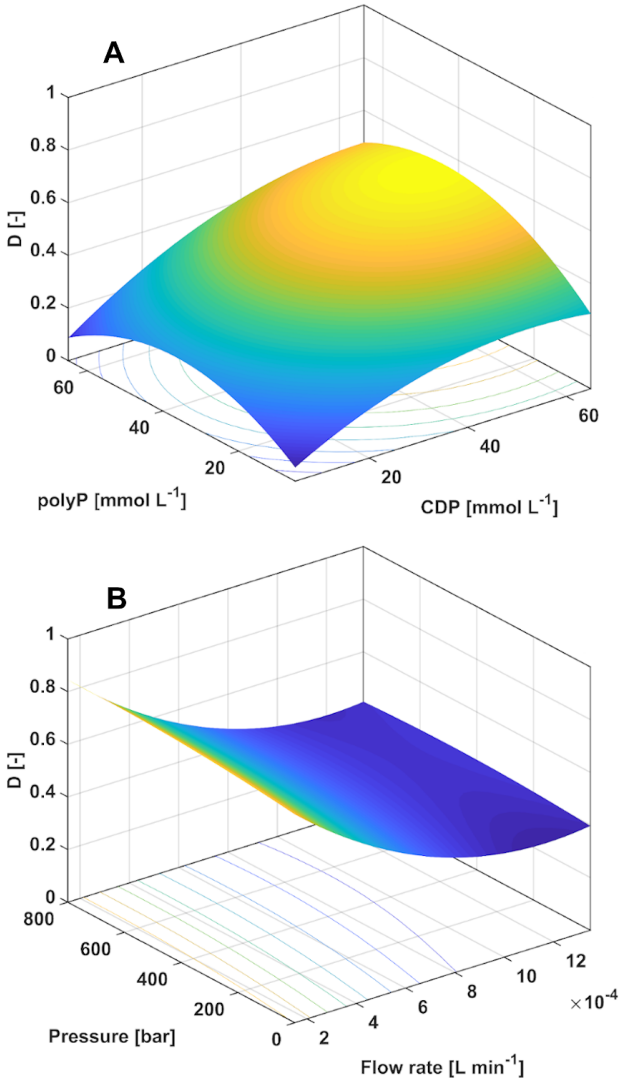

Figure 7.10: Weighted desirability surface for predicted CTP concentration at reactor outlet based on CCD and Monte-Carlo simulations (n=30). **A**: Shown for varying CDP-polyP starting concentrations. **B**: Shown for varying flow-rate-hydrostatic-pressure starting values. Other factors set to constant mid-level values (mDoE-2).

For the CDP-polyP factor combination, D values of 0.5 and higher were reached from 45 to 65 mmol L^{-1} CDP and 30 to 55 mmol L^{-1} polyP (see Fig. 7.10, part A). D higher 0.8 were obtained for flow rates lower 0.2 ml min^{-1} and hydrostatic pressures of 300 bar and higher. Similar to mDoE-1, the maximum response was determined for the response surface and given regression model. A final CTP concentration of 21.01 mmol L^{-1} was predicted for a flow rate of 0.15 mL min^{-1}, 65 mmol L^{-1} CDP, 34.5 mmol L^{-1} polyP and a hydrostatic pressure of 799.9 bar. Thus, the suggested factor combination was located in the area of the highest weighted D (see Fig. 7.10). Compared to mDoE-1, where the maximum response was reached for 400.5 bar, the suggested hydrostatic pressure to be applied was doubled. To validate the workflow and factor combination, the desirable experiment was performed. In Figure 7.9 (part B), the simulated progression of the CTP concentration, the predicted maximum response and experimental data over time are shown.

After a process time of 15 min, a CTP concentration of 21.3 mmol L^{-1} was measured at the reactor outlet (circle marker) for a flow rate of 0.15 mL min^{-1}, 799.9 bar hydrostatic pressure and starting concentrations of 65 mmol L^{-1} CDP and 34.5 mmol L^{-1} polyP. As can be seen in Figure 7.9 (part B), the predicted CTP concentration of 21.01 mmol L^{-1} based on the RSM (square marker) was over-predicted by the process model and model parameters (solid and dashed lines). Furthermore, both the experimental and the predicted data were located outside the uncertainty band. Therefore, new experimental data were added to the data base (33 data sets in total) and the workflow was repeated.

7.1.5 Experimental analysis of the response surface, mDoE-3

In the beginning of the workflow round mDoE-3, the initial (Equ. 7.3) and boundary conditions (Equ. 7.4) for the differential equations (see Suppl. Tab. A1), except for the differential equation describing the deactivation of the enzyme over time, were changed to:

$$c\,(1) = c_{\text{inlet}} - \frac{D_{\text{ax}}}{u}\left(\frac{\partial c}{\partial x}\right)\Big|_{\text{x=0}} \tag{7.3}$$

$$\frac{\partial \text{c}}{dt}(\text{end}) = 0 \tag{7.4}$$

As a result, the prediction of the initial concentration change at the inlet of the packed-bed and the numerical stability of the differential equation solver was improved. Subsequently,

the model parameters were re-adapted to all experimental data sets. The estimated model parameter values based on these changes are summarized in Table 7.8.

Table 7.8: Adapted process model parameters: Median, 10%/90% quantile and 95% CI values (mDoE-3 TC).

	v_{max} [mmol min^{-1} mg^{-1}]	$K_{M,CDP}$ [mmol L^{-1}]	$K_{M,polyP}$ [mmol L^{-1}]	$K_{I,polyP}$ [mmol L^{-1}]	k_d [min^{-1}]
Start value	0.0166	32.16	219.8	8.40	$3.0\ 10^{-4}$
Mean	0.0253	25.85	125.3	12.46	$2.8\ 10^{-4}$
10% quantile	0.019	6.51	29.97	7.80	$2.01 \cdot 10^{-5}$
median	0.024	23.60	117.76	11.64	$2.97 \cdot 10^{-4}$
90% quantile	0.033	46.62	214.24	17.05	$4.49 \cdot 10^{-4}$
$CI_{0.025}$	0.0248	24.60	120.5	12.13	$2.7 \cdot 10^{-4}$
$CI_{0.975}$	0.0257	27.10	130.1	12.80	$2.9 \cdot 10^{-4}$

Similar to the workflow round mDoE-2, all model parameters changed in their mean, median, 10% and 90% quantile and CI values. As a result of the revised boundary conditions and the adjusted ordinary differential equation solver options, process model and model parameter re-adaption the accuracy of prediction was improved. A NRMSD value of 1.35 for all experimental data sets compared to all simulated data (see Fig. 7.11, part A) was determined and the absolute width of the uncertainty band was reduced (10%/90% quantile, see Fig. 7.11, part C).

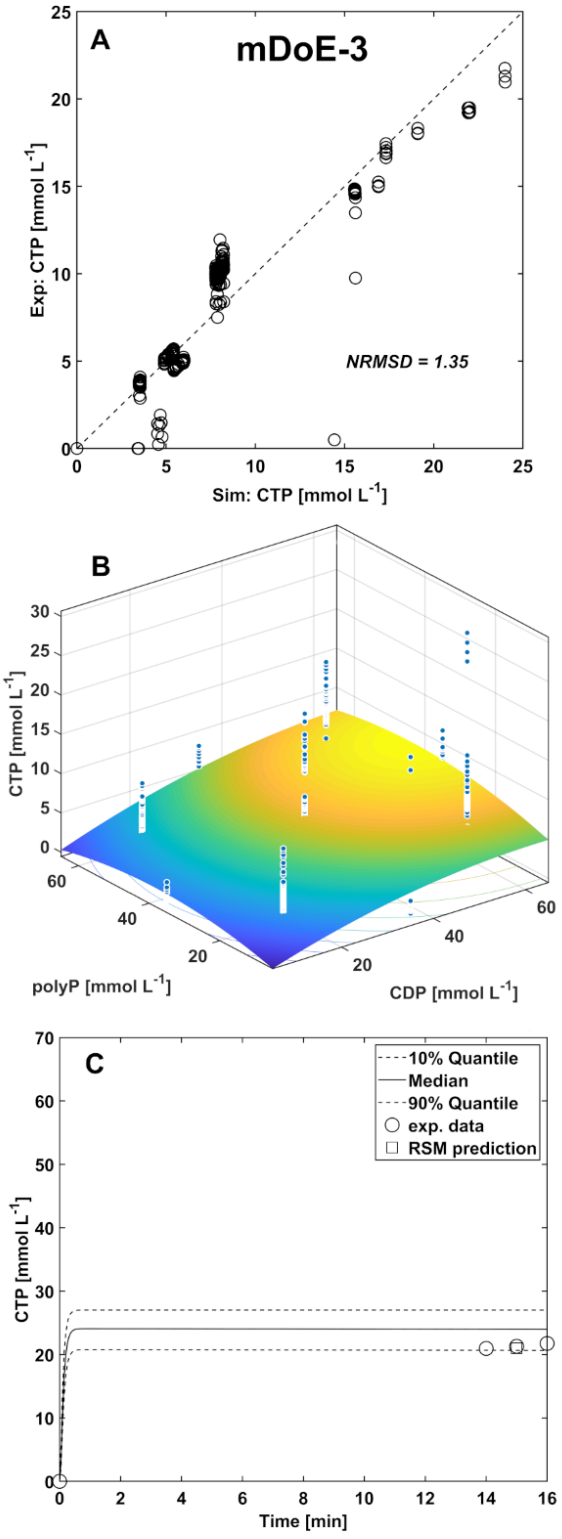

Figure 7.11: Workflow round mDoE-3 TC. **A**: Comparison of experimental and predicted CTP concentration for median model parameter values, all data sets. **B**: Monte-Carlo-based response surface (n=30) for predicted CTP concentration at reactor outlet based on CCD and varying CDP-polyP starting concentrations, flow rate and hydrostatic pressure set to constant mid-level values. **C**: Predicted CTP concentration by RSM (square, mDoE-2), experimental CTP data for the suggested optimal factor combination (circle, mDoE-2) and simulated CTP concentration by the process model (solid and dashed lines, median and 10%/90% quantile model parameter values) over time. For experimental conditions see Suppl. Tab. A.1

Compared to mDoE-2 (see Fig. 7.9, part B), the experimental data points were located within the uncertainty band. As described in the workflow, the termination criterion was reviewed (mDoE-3 TC). Based on the reactor specifications of mDoE-2 ($m_{RpPPK2-3}$ and bulk density), a new optimal factor combination was determined. A CTP concentration of 19.3 mmol L^{-1} at a flow rate of 0.15 mL min^{-1}, 65 mmol L^{-1} CDP, 34.5 mmol L^{-1} polyP and 799.9 bar hydrostatic pressure was predicted. The predefined termination criterion has been reached (difference $c_i/c_{i-1} < 10\%$) and the workflow ended. As indicated by the p-values of the response surface regression coefficients (see Suppl. Tab. A.6), the coefficients considering the flow, CDP and polyP starting concentration had a significant linear and quadratic effect on the response with p-values lower 0.05 (correlation between the dependent and independent variables). The coefficient describing the interaction effect of the flow rate and CDP starting concentration also contributed significantly to the regression model. p-values higher than 0.05 were determined for all other linear, interaction and quadratic effects including the hydrostatic pressure effect. However, the regression coefficients of the multiple linear regression model are only estimated in such a way that the responses of individual factor combinations are well described and the true response surface is approximated. Since hydrostatic pressure has already been shown to have an effect on the final CTP concentration, it might be possible that this effect was superimposed by the other factors and the regression model complexity.

7.2 Optimization objective: Space-time yield STY

As previously described, the iterative workflow could be also used for optimization objectives other than maximizing the final product concentration at the reactor outlet. For example, the space-time yield objective which is defined as the product formed per unit volume and time and frequently used in biotechnology to assess the continuous process performance, could be defined [225, 263]. Based on the process model and process model parameters adapted in the third workflow round, the factor boundary values (see Tab. 7.3) and the results from the Monte-Carlo-based simulations, the maximum STY was determined equivalently to the maximum CTP concentration. A maximum STY of 238.1 mmol L^{-1} min^{-1} was reached for maximum flow rate of 1.35 mL min^{-1}, 65 mmol L^{-1} CDP, 36 mmol L^{-1} polyP and 799.9 bar hydrostatic pressure. The corresponding response surface plots are shown in Figure 7.12.

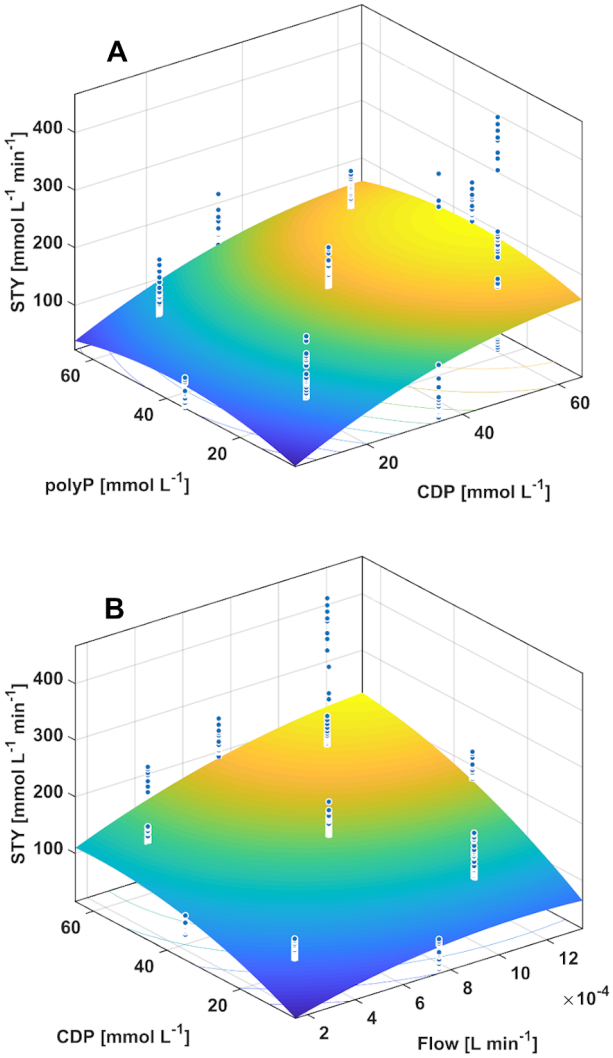

Figure 7.12: **A**: Response surface for predicted maximum STY based on user-constraint CCD and Monte-Carlo-based simulations of mDoE-3 shown for varying CDP and polyP starting concentrations. Flow rate and hydrostatic pressure set to constant mid-level values. **B**: Response surface for predicted maximum STY based on user-constraint CCD and Monte-Carlo-based simulations of mDoE-3 shown for varying flow rate and CDP starting concentrations. PolyP and hydrostatic pressure set to constant mid-level values.

Compared to the response surface plots for maximized CTP concentration (see Fig. 7.11, part B), the maximum STY was also reached for high CDP and mid-range polyP starting concentration (see Fig. 7.12, part A). However, as shown in Figure 7.12 (part B), maximum flow rates were necessary to reach the highest possible STY. This behavior can be explained by the process dynamics of the continuously operated reactor and the change in the optimization objective from achieving maximum product concentration to producing as much product as possible per unit of time. Considering the costs of the substrate and the product as well as the effort of product downstreaming, an extended optimization objective should be formulated

to design an economical and competitive production process. In addition, since cofactor regeneration processes are often only part of more complex multi-step enzymatic processes, the actual CTP requirement of one or more reactions should be assessed.

8 Discussion

In this thesis, the applicability of the mDoE methodology was demonstrated for the model-assisted design and optimization of a cofactor regeneration reaction in a continuously operated packed-bed reactor under high hydrostatic pressure conditions and using an immobilized enzyme. In order to improve the process understanding during the development process and enable a knowledge-based optimization, the mDoE methodology was to be integrated into an iterative workflow. No one-factor-at-a-time or traditional DoE approach, common in process development, was used at the beginning to avoid time-consuming and cost-intensive experimentation. Instead, the aim was to demonstrate the usability of the proposed workflow and the mDoE methodology from the early stages of process design, even when only limited information about the reaction system is available. In addition, insight into the process dynamics of the reaction system should be gained with a reduced number of experiments.

8.1 Applicability of the model-assisted DoE and iterative workflow

First, by selecting the modeling approach and formulating the process model for the presented CTP regeneration reaction in the high hydrostatic pressure reactor, selective experiments were initially performed to validate the process model, improve the model accuracy and describe the process dynamics. These selective experiments were partly necessary due to the novelty of the reactor concept and the lack of information in the literature on the enzymatic reaction and the influencing variables and process conditions. Typically, such experiments are not performed in design and optimization studies. In order to accurately predict the time- and location-dependent as well as the final product concentration with the selected reaction-dispersion-convection model, experiments on the flow behavior and the resulting parameters such as the residence time distribution, Bodenstein number and the resulting axial dispersion coefficient were needed. Relatively low Bodenstein numbers confirmed the assumption of a dispersive axial flow in the packed-bed for flow rates between 0.5 and 1.5 mL min^{-1}. The results obtained also indicated that, among others, the bulk density of the packed-bed may have an influence on the fluid dynamics and on mass transfer effects. However, the accuracy of the process model was sufficient for the specific application shown. Investigation of additional effects caused by the bulk density or structure of the packed-bed would require more extensive

studies beyond the scope of this thesis.

Second, based on the four factors flow rate, initial CDP and polyP concentration and hydrostatic pressure, and their factor limits to be investigated, different experimental designs, such as the central composite, Box-Behnken, and D-optimal design plans, could be evaluated in the beginning of the process design and optimization. Therefore, the design plans were generated and the responses were simulated using the formulated process model. The designs were evaluated with regard to the R^2, R^2_{adj} and MSE values for a first- and second-order regression model determined by regression analysis. The simulated responses were best reflected by a combination of a CCD and second-order regression model, which was then used for the further development process. As a result, an *in silico* and rational rather than a visual selection of an adequate design plan was made. This eliminated the need to perform time-consuming experiments with different experimental designs and enables the accelerated assessment different DoE plans varying in the number of suggested factor combinations and their distribution in the experimental space.

Third, the process dynamics and individual as well as interaction effects of four factors, such as the compensating effect of increasing initial CDP concentration and flow rate, were described and predicted by the process model and model parameter values which may not be explored in initial screening or solely kinetic experiments. In addition, it was shown that individual factors have a varying impact on the process outcome and the optimization objective. Overall, the flow rate followed by the initial CDP concentration were considered as most important. The mean residence time and the mass transfer limitations encountered in the selected factor ranges were directly determined by the flow rate, but further reduction of the flow rate should not be considered as the process productivity would decrease. The impact of the polyP concentration and hydrostatic pressure was comparatively low in the predefined experimental space. Among others, the latter could be attributed to an unfavorable combination of the hydrostatic pressure and reaction temperature, since only one temperature was set for the regeneration reaction. The importance of an ideal hydrostatic pressure and temperature combination has been reported in the literature [13, 115]. Nevertheless, with the use of a suitable pump and resistant reactor system components, the hydrostatic pressure and reaction temperature could be further increased and varied, to investigate the full range of the influence of hydrostatic pressure and temperature on the enzyme activity and stability, as partially shown by Zhao et al. (2015) for immobilized *C. rugosa* lipase up to 2000 bar [116] and Mozhaev et al. (1996)

for α-chymotrypsin up to 4700 bar [115]. The varying impact of the individual factors was emphasized by the regression analysis and determined p-values as well as the determined SRCs of the empirical regression model. The use of SRCs provided an additional method for assessing the importance of each factor on the response. The importance of factor interactions and quadratic effects are assessed quantitatively and not just qualitatively as in standard regression analysis. Compared to other global sensitivity analysis methods, SRCs can be determined rather simple based on the empirical regression model and the corresponding responses. A global sensitivity analysis method, as presented by Möller et al. (2020), could be integrated into the workflow for further studies to also determine important process model parameters and their impact on model outcomes [35].

Fourth, the predefined optimization objective of maximizing the product concentration at the reactor outlet was evaluated not only qualitatively, but also quantitatively for factor combinations different from those suggested by the selected design plan using RSM. In addition, variation of the response can be described on the basis of the uncertainty band represented by the 10% and 90% quantiles. In traditional optimization approaches, optimal factor combinations would have been explored only if the costly and time-consuming experiments suggested by the DoE plan followed by RSM had been conducted fully experimentally. For example, Fabbri et al. (2021) used a full-factorial design with three factor levels and four factors resulting in 30 experiments to analyze the thermostability and selectivity of different hydrolases [264]. A full-factorial design with three factor levels and four factors in combination with RSM was also used by Gupta et al. (2016) to optimize a biodiesel production process [265].

Fifth, at this point, using the regression and process model, further DoE plans can be simulated and evaluated in order to re-evaluate the design plan selected at the really beginning and to facilitate and accelerate the introduction of new optimization approaches.

8.2 Comprehensive discussion

Based on the results presented, the application of the mDoE methodology integrated into the proposed iterative workflow can be justified. After each iteration round, the knowledge and experimental data gained were incorporated into the process model and added to the data base to improve the predictive capabilities and to re-adapt the model parameters from the

previous workflow round. The results clearly show that the accuracy of the process model improved after the second (mDoE-2) and third mDoE (mDoE-3) workflow round, as the width of the uncertainty band is a measure for process model prediction quality [266] and reliability of model parameter estimation [57]. The relative width, relative to the simulated CTP concentration based on the median model parameter values, decreased from 170.2% to 26.6% for mDoE-2 and mDoE-3, respectively. Furthermore, the improvement of the process model accuracy was also confirmed by the decreased NRMSD value for all experimental and simulated data. The NRMSD decreased from 1.81 for mDoE-1 to 1.35 for mDoE-3. Overpredicted CTP concentrations at the beginning of the simulation could be related to start-up effects or packed-bed rearrangement. These types of effects have not been described for small-scale biocatalytic processes in the literature and need to be investigated as a function of the immobilizate material and particle size, flow rate and hydrostatic pressure. In addition to the uncertainty-based methods, a risk-based approach was enabled by using the weighted desirability function as described previously in Subsection 6.4. Potentially, the optimization based solely on the final concentration could be replaced by the estimated desirability. For further optimization studies, the criterion for terminating the workflow when the expected change in the response is less than 10% could be revised to better meet the needs and expectations of the user. However, there will be a trade-off between the number of iteration rounds and the resulting costs and the time required for the optimization.

The versatile usability of the results and the iterative workflow has been demonstrated as the maximum STY was additionally determined for the given factor boundaries despite the described potential for improvements and further development. Generally, in contrast to maximizing the final product concentration at the reactor outlet, optimizing the STY can be regarded as more relevant in continuous biocatalysis [267], but it is important to consider whether the increased STY compensates for the increased product and waste stream [268]. In order to design an economical and competitive biocatalytic process based on the STY, the optimization objective has be to extended to also account for the cost of substrates, energy, equipment and subsequent downstreaming and purification of the product. Furthermore, ecological criteria are becoming more important and have to be accounted as well. [43] For example, a multi-objective dynamic optimization based on STY, enzyme consumption and cofactor consumption for an enzymatic cascade producing α -ketoglutarate was presented by Paschalidis et al. (2022) [269].

8.3 Scale-up

In order to reduce the cost of the final product, scale-up of the regeneration process is needed. Similar to the scale-up of STR or CSTR reactors, scale-up design criteria exist for flow and packed-bed reactors. According to Afandizadeh and Foumeny (2001), the most significant criterion in the scale-up of a packed-bed reactor is the reactor to particle diameter ratio [270]. However, it is to be expected that the overall performance is affected by scaling-up [270]. Due to the complexity of the CTP regeneration reaction it is assumed that beside this criterion a more comprehensive description of the influence of the bed structure and hydrodynamics inside and along the packed-bed would be also necessary.

Compared to the small-scale reactor presented in this thesis, which was mainly derived from a UHPLC concept, different immobilizate particles in terms of size and size distribution are needed, resulting in changed axial and radial dispersion, internal and external mass transfer and possible zones of inadequate flow and conversion rates (often referred as dead volume). Particularly, diffusion and heat transfer processes are not linearly scalable. [43, 270] The size and its variation also directly influences the specific area available, void volume and back pressure [270]. In addition, back pressure results from the relatively high flow rates used in flow reactors due to the high loading of enzymes [271]. The back pressure can be overcome by changing the shape of the particles [270] or application of silica monoliths [133]. To transfer the concept of reactions performed under high hydrostatic pressure to large scale, the design of the pressure generation unit, reactor wall thickness, tube connectors and suitable back pressure regulator have to be adapted as well.

As a result, the prior evaluation of different scaling criteria, immobilizate sizes and shapes as well as the resulting conversions and STYs, even if the flow rate is similar to the small scale, is recommended. Therefore, the presented process model could be adapted or extended and used as a knowledge-based scale-up tool. The basis for the model-based scale-up could be provided by the mDoE together with a model-predictive controller concept for the adaption of model parameters and the control of process variables, such as the NMPC. In Figure 8.1 an adapted workflow for the scale-up of biocatalytic production processes is proposed.

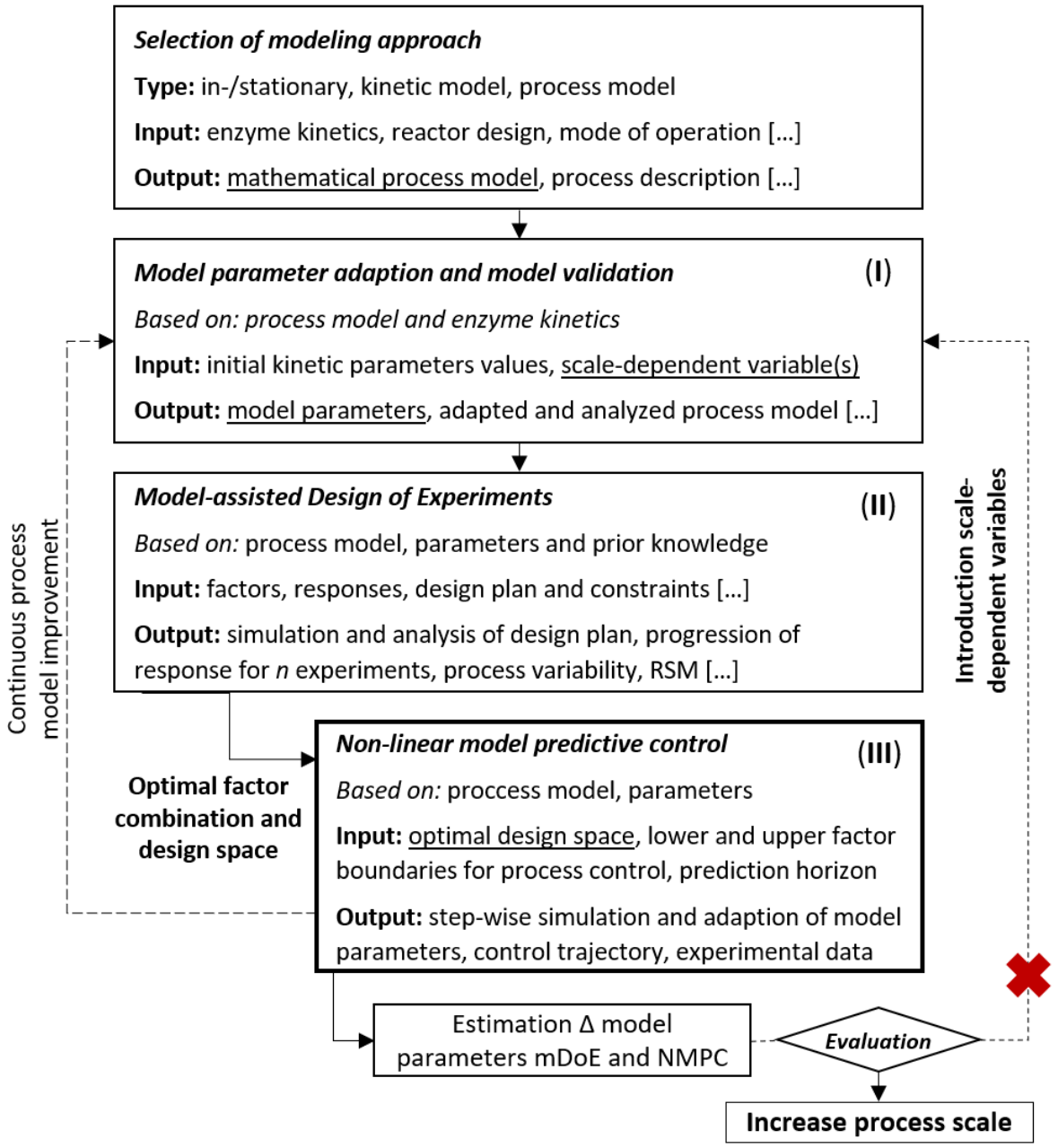

Figure 8.1: Combined iterative model-assisted Design of Experiments workflow and non-linear model predictive control for process scale-up.

In the beginning of process development and scale-up, the mDoE methodology is performed as described in Subsection 6. Adapted model parameters from the smallest scale are initially used and it is assumed that the kinetics are unchanged. The mDoE is now used to predict an optimal factor combination which is then used in the NMPC part (see Fig. 8.1, part III) as starting point for experiments. In addition, a reduced design space is determined with the mDoE, which represents the lower and upper factor boundaries for the NMPC. In the NMPC part, model parameters are adapted and process variables are controlled according

to a predefined optimization objective. For example, the flow rate can be adjusted over the process time to compensate for enzyme deactivation and to achieve a constant product concentration. After the process is simulated, the differences in the model parameter values (Δ) and distributions between the mDoE and NMPC from the higher scale are to be compared. Significant differences would indicate that process model parameters or process variables are scale-dependent. In this case, the process model would be adapted, scale-dependent variables would be introduced or declared, e.g., the D_{ax}, and the resulting process model would be validated. With this concept, the progression of the process model structure and model parameter values can be summarized over the different process scales.

Furthermore, a model-based scale-up strategy can be developed and the reactor or packed-bed design can be adapted, e.g., the reactor diameter and length, and optimized. Also, the maximum time of operation due to enzyme deactivation or leaching could be determined. A model-based scale-up is also suggested by Afandizadeh and Foumeny (2001) who proposed the use of a catalyst-reactor model to study physical and operating variables and their main as well as interaction effects if catalyst particles are changed [270]. Therefore, the NMPC can be seen as an addition to the mDoE methodology and as a further step towards process intensification, digitalization and replacement of traditional scale-up concepts. Common *a priori* models are assumed to fail in scale-up as process dynamics and process variables change between small and larger scale [272].

9 Conclusion and outlook

The mDoE methodology was successfully used as part of an iterative workflow to support the design and optimization of the complex enzymatic CTP cofactor regeneration process. Compared to other design and optimization approaches, the iterative workflow represents a user-friendly method as the user can rely on a predefined decision path and criteria. Furthermore, the workflow and programming structure also allows for efficient interaction between the individual workflow modules. Despite the low amount of data about, e.g., enzyme kinetics, flow characteristics and the experimental space, it was possible to formulate a suitable process model and improve the understanding of the process by taking experimental variations into account. In addition to the application of different flow rates and initial substrate concentrations, high hydrostatic pressure was used as a complementary process parameter and its effect was described and incorporated into the process model.
Based on the process model, model parameters and pre-evaluated experimental design, insights into the process dynamics and factor interactions as well as importance of hydrodynamics inside the packed-bed were gained. Using the proposed workflow, focused experiments were carried out in the early stage of process development, and the reaction conditions were efficiently optimized with regard to a maximized CTP concentration at the reactor outlet. 21.3 mmol L^{-1} CTP was obtained for a flow rate of 0.15 mL min^{-1}, 800 bar hydrostatic pressure and starting concentrations of 65 mmol L^{-1} CDP and 34.5 mmol L^{-1} polyP after the third workflow round. In addition, it has been shown, that other optimization objectives such as the STY, can be used and evaluated using the workflow.
Overall, the experimental effort was significantly reduced compared to traditional optimization approaches where several rounds of DoE can be necessary to optimize a process. In further works, the workflow could be repeated not only to improve the predictive capability of the process model, but also to step-wise reduce the design space or vary the optimization objective.

Despite the described deviations between the measured and set substrate concentrations, the process understanding of the reaction system has been improved. However, these findings have implications for the original mDoE methodology, so that in further applications, accurate analysis of the experimental data should be a fundamental part of the iterative workflow to avoid or account for differences between the simulated and measured values. As can be seen from the figures showing the progression of the CTP concentration over time, a more detailed analysis

of the flow characteristics and kinetics is required to adequately reflect the experimental data in the initial phase of the process. These deviations could be due to time-dependent changes in the process dynamics and kinetics that are different from those at a hypothetical steady-state condition [273]. Changes in the packed-bed could also be a potential explanation. Therefore, additional terms considering this behavior in the initial phase of the process and reaching of steady-state conditions could be implemented. A promising time scale analysis approach was presented by Sinner et al. (2020) to describe the time- and operation condition-dependent dynamics of process parameters for a fermentation process. This approach could be adapted and used for the continuously operated packed-bed reactor [273]. Besides, more precise simulations and uncertainty analysis can be also achieved by individual starting value standard deviations. This was shown by Bayer et al. (2020), who replaced the generally assumed standard deviation with individual standard deviations estimated in repeated experiments at the beginning of the modeling procedure [167].

With regard to the considered convective and dispersive flow, the stability and predictive capability of process models could be improved by increasing the level of detail, e.g., by consideration of spatial changes in axial and radial direction [274]. Furthermore, a comprehensive study of the influence of the packed-bed bulk density on the flow behavior should be taken into account. Describing the Bodenstein number as a function of the bulk density and flow rate would also be conceivable. However, this would potentially lead to longer simulation times, especially if used in the proposed iterative workflow with adaption of the process model parameters. Nevertheless, these simulations can decoupled from the workflow procedure and only used at the start. Computational fluid dynamics (CFD) and other simulation platforms are promising and well developed programs to further analyze the flow through the catalytic bed. For example, this was shown by Rezvani et al. (2015) for phenol removal by immobilized peroxidase in a packed-bed reactor. It was found that the CFD approach provides useful information on the hydrodynamics and possible bottlenecks in the process design, which is needed for further industrial applications. [275]
A design and optimization approach that focuses more on the use of hybrid or black box ANN models could further increase the predictive capabilities as these approaches are able to describe phenomena that can only be described by time-consuming and cost-intensive experiments. It would even be possible to integrate them into an workflow to increase the number of experimental data sets and adapt the process model parameters and the design space. However, it should be kept in mind that the chosen modeling approach is consistent with the expectations

and further use of a purely descriptive, e.g., ANN, or rather explanatory, e.g., mechanistic or hybrid, process model. Compared to the workflow proposed, the adaption of process model parameters and quantification of model parametric uncertainty would be omitted in a novel iterative workflow incorporating an ANN. Depending on the type of ANN, the following workflow parts including RSM and evaluation of results are still usable. The ANN could also replace the RSM as described by Huang et al. (2017). Experiments were performed according to a CCD and the ANN was then used to predict and optimize the process variables (RSM and ANN modeling-based optimization). [276] Similar to the workflow presented in this work, using this type of optimization procedure, the experimental data generated in validation experiments would be added to the data base. Nevertheless, a rational and systematic decision for type of workflow should be made to maximize the information and insights gained.

With the parts and tools presented and used in the model-assisted workflow, the basis for further research and the application of the process model as part of a digital twin (DT) has been established. The model requirement to quantitatively and qualitatively represent the characteristics of the regeneration reaction and process variable interactions were met [232]. However, additional DT building blocks, such as continuous storage of operation data and process visualization, are to be implemented [277] to successfully perform *in silico* predictions and optimize real world processes [140]. A DT approach in the field of enzymatic hydrolysis was presented and applied by Appl et al. (2021) [86]. Based on these promising results and together with appropriate on- or at-line process control, soft sensors and non-linear model predictive controllers, the DT concept could be used to accompany current or future biotechnological production processes.

Bibliography

[1] Schmid, A, Hollmann, F, Park, JB, and Bühler, B. "The use of enzymes in the chemical industry in Europe". *Current Opinion in Biotechnology* 13.4 (2002), 359–366.

[2] Maria, G and Crisan, M. "Evaluation of optimal operation alternatives of reactors used for d-glucose oxidation in a bi-enzymatic system with a complex deactivation kinetics". *Asia-Pacific Journal of Chemical Engineering* 10.1 (2015), 22–44.

[3] Osório, N, Ribeiro, M, Fonseca, M da, and Ferreira-Dias, S. "Interesterification of fat blends rich in ω-3 polyunsaturated fatty acids catalysed by immobilized Thermomyces lanuginosa lipase under high pressure". *Journal of Molecular Catalysis B: Enzymatic* 52 (2008), 58–66.

[4] Sanchez, S and Demain, AL. "Enzymes and bioconversions of industrial, pharmaceutical, and biotechnological significance". *Organic Process Research & Development* 15.1 (2011), 224–230.

[5] Atalah, J, Cáceres-Moreno, P, Espina, G, and Blamey, JM. "Thermophiles and the applications of their enzymes as new biocatalysts". *Bioresource technology* 280 (2019), 478–488.

[6] Liszka, MJ, Clark, ME, Schneider, E, and Clark, DS. "Nature versus nurture: developing enzymes that function under extreme conditions". *Annual review of chemical and biomolecular engineering* 3 (2012), 77–102.

[7] Berglund, P. "Controlling lipase enantioselectivity for organic synthesis". *Biomolecular engineering* 18.1 (2001), 13–22.

[8] Eisenmenger, MJ and Reyes-De-Corcuera, JI. "High pressure enhancement of enzymes: a review". *Enzyme and microbial technology* 45.5 (2009), 331–347.

[9] Klibanov, AM. "Improving enzymes by using them in organic solvents". *nature* 409.6817 (2001), 241–246.

[10] Krishna, SH, Divakar, S, Prapulla, S, and Karanth, N. "Enzymatic synthesis of isoamyl acetate using immobilized lipase from Rhizomucor miehei". *Journal of biotechnology* 87.3 (2001), 193–201.

[11] Lanza, M, Priamo, WL, Oliveira, JV, Dariva, C, and De Oliveira, D. "The effect of temperature, pressure, exposure time, and depressurization rate on lipase activity in SCCO 2". *Applied biochemistry and biotechnology* 113.1-3 (2004), 181–187.

[12] Liese, A, Seelbach, K, and Wandrey, C. *Industrial biotransformations*. John Wiley & Sons, 2006.

[13] Marques, J, Vila-Real, HJ, Alfaia, AJ, and Ribeiro, MH. "Modelling of the high pressure-temperature effects on naringin hydrolysis based on response surface methodology". *Food Chemistry* 105.2 (2007), 504–510.

[14] Chitnis, A and Sadana, A. "pH-Dependent enzyme deactivation models". *Biotechnology and bioengineering* 34.6 (1989), 804–818.

[15] Mateo, C, Palomo, JM, Fernandez-Lorente, G, Guisan, JM, and Fernandez-Lafuente, R. "Improvement of enzyme activity, stability and selectivity via immobilization techniques". *Enzyme and microbial technology* 40.6 (2007), 1451–1463.

[16] Jurado, E, Camacho, F, Luzón, G, and Vicaria, J. "Kinetic models of activity for β-galactosidases: influence of pH, ionic concentration and temperature". *Enzyme and Microbial Technology* 34.1 (2004), 33–40.

[17] Liese, A and Villela Filho, M. "Production of fine chemicals using biocatalysis". *Current Opinion in Biotechnology* 10.6 (1999), 595–603.

[18] Jaenicke, R and Závodszky, P. "Proteins under extreme physical conditions". *FEBS letters* 268.2 (1990), 344–349.

[19] Polaina, J and MacCabe, AP. *Industrial enzymes*. Springer, 2007.

[20] Richter, M. "Functional diversity of organic molecule enzyme cofactors". *Natural product reports* 30.10 (2013), 1324–1345.

[21] Kara, S, Schrittwieser, JH, Hollmann, F, and Ansorge-Schumacher, MB. "Recent trends and novel concepts in cofactor-dependent biotransformations". *Applied microbiology and biotechnology* 98 (2014), 1517–1529.

[22] Hartmann, M and Kostrov, X. "Immobilization of enzymes on porous silicas-benefits and challenges". *Chemical Society Reviews* 42.15 (2013), 6277–6289.

[23] Derat, E and Kamerlin, SCL. *Computational advances in protein engineering and enzyme design*. 2022.

[24] Miller, DC, Athavale, SV, and Arnold, FH. "Combining chemistry and protein engineering for new-to-nature biocatalysis". *Nature synthesis* 1.1 (2022), 18–23.

[25] Rocha, RA, Speight, RE, and Scott, C. "Engineering enzyme properties for improved biocatalytic processes in batch and continuous flow". *Organic Process Research & Development* 26.7 (2022), 1914–1924.

[26] Woodley, JM. "New frontiers in biocatalysis for sustainable synthesis". *Current Opinion in Green and Sustainable Chemistry* 21 (2020), 22–26.

[27] Federsel, HJ, Moody, TS, and Taylor, SJ. "Recent trends in enzyme immobilization - Concepts for expanding the biocatalysis toolbox". *Molecules* 26.9 (2021), 2822.

[28] Tripathi, P and Sinha, S. "Industrial biocatalysis: An insight into trends and future directions". *Current Sustainable/Renewable Energy Reports* 7 (2020), 66–72.

[29] Scepankova, H, Galante, D, Espinoza-Suaréz, E, Pinto, CA, Estevinho, LM, and Saraiva, J. "High Hydrostatic Pressure in the Modulation of Enzymatic and Organocatalysis and Life under Pressure: A Review". *Molecules* 28.10 (2023), 4172.

[30] Kirsch, C, Dahms, J, Kostko, AF, McHugh, MA, and Smirnova, I. "Pressure assisted stabilization of biocatalysts at elevated temperatures: characterization by dynamic light scattering". *Biotechnology and bioengineering* 110.6 (2013), 1674–1680.

[31] Gernaey, KV, Lantz, AE, Tufvesson, P, Woodley, JM, and Sin, G. "Application of mechanistic models to fermentation and biocatalysis for next-generation processes". *Trends in biotechnology* 28.7 (2010), 346–354.

[32] Sinner, P, Kager, J, Daume, S, and Herwig, C. "Model-based analysis and optimisation of a continuous Corynebacterium glutamicum bioprocess utilizing lignocellulosic waste". *IFAC-PapersOnLine* 52.26 (2019), 181–186.

[33] Stosch, M von, Hamelink, JM, and Oliveira, R. "Toward intensifying design of experiments in upstream bioprocess development: An industrial Escherichia coli feasibility study". *Biotechnology progress* 32.5 (2016), 1343–1352.

[34] Kiparissides, A, Koutinas, M, Pistikopoulos, EN, and Mantalaris, A. "Model development and analysis of mammalian cell culture systems". *Process Systems Engineering* (2010), 403–439.

[35] "Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes". 134 ().

[36] Tsopanoglou, A and Val, IJ del. “Moving towards an era of hybrid modelling: advantages and challenges of coupling mechanistic and data-driven models for upstream pharmaceutical bioprocesses”. *Current Opinion in Chemical Engineering* 32 (2021), 100691.

[37] Sin, G, Woodley, JM, and Gernaey, KV. “Application of modeling and simulation tools for the evaluation of biocatalytic processes: a future perspective”. *Biotechnology progress* 25.6 (2009), 1529–1538.

[38] Baur, D, Angelo, J, Chollangi, S, Müller-Späth, T, Xu, X, Ghose, S, Li, ZJ, and Morbidelli, M. “Model-assisted process characterization and validation for a continuous two-column protein A capture process”. *Biotechnology and bioengineering* 116.1 (2019), 87–98.

[39] Caramihai, M and Severin, I. “Bioprocess modeling and control”. *Biomass Now: Sustainable Growth and Use* (2013), 147.

[40] Lima-Ramos, J, Neto, W, and Woodley, JM. “Engineering of biocatalysts and biocatalytic processes”. *Topics in Catalysis* 57.5 (2014), 301–320.

[41] Fisher, OJ, Watson, NJ, Escrig, JE, Witt, R, Porcu, L, Bacon, D, Rigley, M, and Gomes, RL. “Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems”. *Computers & Chemical Engineering* 140 (2020), 106881.

[42] Möller, J and Pörtner, R. “Digital Twins for Tissue Culture Techniques: Concepts, Expectations, and State of the Art”. *Processes* 9.3 (2021), 447.

[43] Siedentop, R, Claaßen, C, Rother, D, Lütz, S, and Rosenthal, K. “Getting the most out of enzyme cascades: strategies to optimize in vitro multi-enzymatic reactions”. *Catalysts* 11.10 (2021), 1183.

[44] Narayanan, H, Luna, MF, Stosch, M von, Cruz Bournazou, MN, Polotti, G, Morbidelli, M, Butté, A, and Sokolov, M. “Bioprocessing in the digital age: the role of process models”. *Biotechnology journal* 15.1 (2020), 1900172.

[45] Möller, J, Kuchemüller, KB, Steinmetz, T, Koopmann, KS, and Pörtner, R. “Model-assisted design of experiments as a concept for knowledge-based bioprocess development”. *Bioprocess and biosystems engineering* 42.5 (2019), 867–882.

[46] Dong, J, Mandenius, CF, Lübberstedt, M, Urbaniak, T, Nüssler, AK, Knobeloch, D, Gerlach, JC, and Zeilinger, K. "Evaluation and optimization of hepatocyte culture media factors by design of experiments (DoE) methodology". *Cytotechnology* 57.3 (2008), 251–261.

[47] Hallenbeck, PC, Grogger, M, Mraz, M, and Veverka, D. "The use of Design of Experiments and Response Surface Methodology to optimize biomass and lipid production by the oleaginous marine green alga, Nannochloropsis gaditana in response to light intensity, inoculum size and CO2". *Bioresource technology* 184 (2015), 161–168.

[48] Shivhare, M and McCreath, G. "Practical considerations for DoE implementation in quality by design". *BioProcess Int* 8.6 (2010), 22–30.

[49] Kumar, V, Bhalla, A, and Rathore, AS. "Design of experiments applications in bioprocessing: Concepts and approach". *Biotechnology Progress* 30.1 (2014), 86–99.

[50] Lee, KM and Gilmore, DF. "Statistical experimental design for bioprocess modeling and optimization analysis". *Applied biochemistry and biotechnology* 135.2 (2006), 101–115.

[51] N. Politis, S, Colombo, P, Colombo, G, and M. Rekkas, D. "Design of experiments (DoE) in pharmaceutical development". *Drug development and industrial pharmacy* 43.6 (2017), 889–901.

[52] Mandenius, CF and Brundin, A. "Bioprocess optimization using design-of-experiments methodology". *Biotechnology progress* 24.6 (2008), 1191–1203.

[53] Borchert, D, Suarez-Zuluaga, DA, Thomassen, YE, and Herwig, C. "Risk assessment and integrated process modeling-an improved QbD approach for the development of the bioprocess control strategy [J]". *AIMS Bioengineering* 7.4 (2020), 254–271.

[54] Chhatre, S, Farid, SS, Coffman, J, Bird, P, Newcombe, AR, and Titchener-Hooker, NJ. "How implementation of quality by design and advances in biochemical engineering are enabling efficient bioprocess development and manufacture". *Journal of Chemical Technology & Biotechnology* 86.9 (2011), 1125–1129.

[55] Georgakis, C. "Design of dynamic experiments: A data-driven methodology for the optimization of time-varying processes". *Industrial & Engineering Chemistry Research* 52.35 (2013), 12369–12382.

[56] Wang, K, Han, L, Mustakis, J, Li, B, Magano, J, Damon, DB, Dion, A, Maloney, MT, Post, R, and Li, R. "Kinetic and data-driven reaction analysis for pharmaceutical process development". *Industrial & Engineering Chemistry Research* 59.6 (2019), 2409–2421.

[57] Moser, A, Kuchemüller, KB, Deppe, S, Rodríguez, TH, Frahm, B, Pörtner, R, Hass, VC, and Möller, J. "Model-assisted DoE software: optimization of growth and biocatalysis in Saccharomyces cerevisiae bioprocesses". *Bioprocess and Biosystems Engineering* (2021), 1–18.

[58] Chhatre, S. "Modelling approaches for bio-manufacturing operations". *Measurement, monitoring, modelling and control of bioprocesses* (2012), 85–107.

[59] Kroll, P, Hofer, A, Ulonska, S, Kager, J, and Herwig, C. "Model-based methods in the biopharmaceutical process lifecycle". *Pharmaceutical Research* 34.12 (2017), 2596–2613.

[60] Kuchemüller, KB, Pörtner, R, and Möller, J. "Efficient Optimization of Process Strategies with Model-Assisted Design of Experiments". *Animal Cell Biotechnology: Methods and Protocols*. Ed. by Pörtner, R. New York, NY: Springer US, 2020, 235–249.

[61] Schelch, S, Eibinger, M, Gross Belduma, S, Petschacher, B, Kuballa, J, and Nidetzky, B. "Engineering analysis of multienzyme cascade reactions for 3'-sialyllactose synthesis". *Biotechnology and Bioengineering* 118.11 (2021), 4290–4304.

[62] Burk, MJ and Van Dien, S. "Biotechnology for chemical production: challenges and opportunities". *Trends in biotechnology* 34.3 (2016), 187–190.

[63] Chen, GQ. "New challenges and opportunities for industrial biotechnology". *Microbial cell factories* 11.1 (2012), 1–3.

[64] Mahadevan, R, Burgard, AP, Famili, I, Van Dien, S, and Schilling, CH. "Applications of metabolic modeling to drive bioprocess development for the production of value-added chemicals". *Biotechnology and bioprocess engineering* 10.5 (2005), 408–417.

[65] Santacoloma, PA, Sin, G, Gernaey, KV, and Woodley, JM. "Multienzyme-catalyzed processes: next-generation biocatalysis". *Organic Process Research & Development* 15.1 (2011), 203–212.

[66] Thomas, SM, DiCosimo, R, and Nagarajan, V. "Biocatalysis: applications and potentials for the chemical industry". *TRENDS in Biotechnology* 20.6 (2002), 238–242.

[67] Gargalo, CL, Las Heras, SC de, Jones, MN, Udugama, I, Mansouri, SS, Krühne, U, and Gernaey, KV. "Towards the development of digital twins for the bio-manufacturing industry". *Digital twins: Tools and concepts for smart biomanufacturing* (2021), 1–34.

[68] Badgujar, KC and Bhanage, BM. "Enhanced biocatalytic activity of lipase immobilized on biodegradable copolymer of chitosan and polyvinyl alcohol support for synthesis of propionate ester: kinetic approach". *Industrial & Engineering Chemistry Research* 53.49 (2014), 18806–18815.

[69] Clayton, AD, Labes, R, and Blacker, AJ. "Combination of chemocatalysis and biocatalysis in flow". *Current Opinion in Green and Sustainable Chemistry* 26 (2020), 100378.

[70] Rodrigues, CJ and Carvalho, CC de. "Process development for benzyl alcohol production by whole-cell biocatalysis in stirred and packed bed reactors". *Microorganisms* 10.5 (2022), 966.

[71] Peschke, T, Bitterwolf, P, Rabe, KS, and Niemeyer, CM. "Self-Immobilizing Oxidoreductases for Flow Biocatalysis in Miniaturized Packed-Bed Reactors". *Chemical Engineering & Technology* 42.10 (2019), 2009–2017.

[72] Dal Magro, L, Pessoa, JPS, Klein, MP, Fernandez-Lafuente, R, and Rodrigues, RC. "Enzymatic clarification of orange juice in continuous bed reactors: Fluidized-bed versus packed-bed reactor". *Catalysis Today* 362 (2021), 184–191.

[73] Eisenmenger, MJ and Reyes-De-Corcuera, JI. "High hydrostatic pressure increased stability and activity of immobilized lipase in hexane". *Enzyme and Microbial Technology* 45.2 (2009), 118–125.

[74] Liese, A and Hilterhaus, L. "Evaluation of immobilized enzymes for industrial applications". *Chemical Society Reviews* 42.15 (2013), 6236–6249.

[75] Wandrey, C, Liese, A, and Kihumbu, D. "Industrial biocatalysis: past, present, and future". *Organic Process Research & Development* 4.4 (2000), 286–290.

[76] Ferreira, R, Skrekas, C, Hedin, A, Sánchez, BJ, Siewers, V, Nielsen, J, and David, F. "Model-assisted fine-tuning of central carbon metabolism in yeast through dCas9-based regulation". *ACS Synthetic Biology* 8.11 (2019), 2457–2463.

[77] Panke, S, Held, M, Wubbolts, MG, Witholt, B, and Schmid, A. "Pilot-scale production of (S)-styrene oxide from styrene by recombinant Escherichia coli synthesizing styrene monooxygenase". *Biotechnology and bioengineering* 80.1 (2002), 33–41.

[78] Peters-Wendisch, P, Schiel, B, Wendisch, VF, Katsoulidis, E, Mockel, B, Sahm, H, and Eikmanns, BJ. "Pyruvate carboxylase is a major bottleneck for glutamate and lysine production by Corynebacterium glutamicum". *Journal of molecular microbiology and biotechnology* 3.2 (2001).

[79] May, O, Nguyen, PT, and Arnold, FH. "Inverting enantioselectivity by directed evolution of hydantoinase for improved production of L-methionine". *Nature biotechnology* 18.3 (2000), 317–320.

[80] Guajardo, N and María, P Domínguez de. "Continuous biocatalysis in environmentally-friendly media: a triple synergy for future sustainable processes". *ChemCatChem* 11.14 (2019), 3128–3137.

[81] Barak, Y, Nov, Y, Ackerley, DF, and Matin, A. "Enzyme improvement in the absence of structural knowledge: a novel statistical approach". *The ISME Journal* 2.2 (2008), 171–179.

[82] Vasic-Racki, D, Kragl, U, and Liese, A. "Benefits of enzyme kinetics modelling". *Chemical and biochemical engineering quarterly* 17.1 (2003), 7–18.

[83] Longwell, CK, Labanieh, L, and Cochran, JR. "High-throughput screening technologies for enzyme engineering". *Current opinion in biotechnology* 48 (2017), 196–202.

[84] Neubauer, P, Cruz, N, Glauche, F, Junne, S, Knepper, A, and Raven, M. "Consistent development of bioprocesses from microliter cultures to the industrial scale". en. *Engineering in Life Sciences* 13.3 (May 2013), 224–238. (Visited on 02/10/2020).

[85] Nickel, DB, Cruz-Bournazou, MN, Wilms, T, Neubauer, P, and Knepper, A. *Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale*. Tech. rep. Wiley Online Library, 2017.

[86] Appl, C, Baganz, F, and Hass, VC. "Development of a digital twin for enzymatic hydrolysis processes". *Processes* 9.10 (2021), 1734.

[87] Ferrandi, EE and Monti, D. "Amine transaminases in chiral amines synthesis: recent advances and challenges". *World Journal of Microbiology and Biotechnology* 34.1 (2018), 13.

[88] Kaushik, N, Biswas, S, and Singh, J. "Biocatalysis and biotransformation processes-an insight". *The Science Technology Journal* 1.08 (2014), 15–17.

[89] Gauss, D, Schönenberger, B, Molla, GS, Kinfu, BM, Chow, J, Liese, A, Streit, WR, and Wohlgemuth, R. "Biocatalytic phosphorylation of metabolites". *Applied Biocatalysis: From Fundamental Science to Industrial Applications: From Fundamental Science to Industrial Applications* (2016), 147–177.

[90] Guo, F and Berglund, P. "Transaminase biocatalysis: optimization and application". *Green Chemistry* 19.2 (2017), 333–360.

[91] Mori, H, Iida, A, Fujio, T, and Teshiba, S. "A novel process of inosine 5'-monophosphate production using overexpressed guanosine/inosine kinase". *Applied microbiology and biotechnology* 48 (1997), 693–698.

[92] Siedentop, R, Siska, M, Möller, N, Lanzrath, H, Lieres, E von, Lütz, S, and Rosenthal, K. "Bayesian optimization for an ATP-regenerating in vitro enzyme cascade". *Catalysts* 13.3 (2023), 468.

[93] Panke, S, Held, M, and Wubbolts, M. "Trends and innovations in industrial biocatalysis for the production of fine chemicals". *Current Opinion in Biotechnology* 15.4 (2004), 272–279.

[94] Zhang, X, Wang, X, Zhang, Y, Wang, F, Zhang, C, and Li, X. "Development of isopentenyl phosphate kinases and their application in terpenoid biosynthesis". *Biotechnology Advances* 64 (2023), 108124.

[95] Suzuki, Ei, Ishikawa, K, Mihara, Y, Shimba, N, and Asano, Y. "Structural-based engineering for transferases to improve the industrial production of 5'-nucleotides". *Bulletin of the Chemical Society of Japan* 80.2 (2007), 276–286.

[96] Tavanti, M, Hosford, J, Lloyd, RC, and Brown, MJ. "Recent developments and challenges for the industrial implementation of polyphosphate kinases". *ChemCatChem* 13.16 (2021), 3565–3580.

[97] Suzuki, S, Hara, R, and Kino, K. "Production of aminoacyl prolines using the adenylation domain of nonribosomal peptide synthetase with class III polyphosphate kinase 2-mediated ATP regeneration". *Journal of bioscience and bioengineering* 125.6 (2018), 644–648.

[98] Ali, M, Ishqi, HM, and Husain, Q. "Enzyme engineering: Reshaping the biocatalytic functions". *Biotechnology and bioengineering* 117.6 (2020), 1877–1894.

[99] Chaloupkova, R, Prokop, Z, Sato, Y, Nagata, Y, and Damborsky, J. "Stereoselectivity and conformational stability of haloalkane dehalogenase DbjA from Bradyrhizobium japonicum USDA110: the effect of pH and temperature". *The FEBS Journal* 278.15 (2011), 2728–2738.

[100] Gouda, MD, Singh, SA, Rao, AA, Thakur, MS, and Karanth, NG. "Thermal inactivation of glucose oxidase: Mechanism and stabilization using additives". *Journal of Biological Chemistry* 278.27 (2003), 24324–24333.

[101] Magario, I, Neumann, A, Oliveros, E, and Syldatk, C. "Deactivation Kinetics and Response Surface Analysis of the Stability of α-l-Rhamnosidase from Penicillium decumbens". *Applied biochemistry and biotechnology* 152.1 (2009), 29–41.

[102] Gross, M and Jaenicke, R. "Proteins under pressure: the influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes". *European Journal of Biochemistry* 221.2 (1994), 617–630.

[103] Hendrickx, M, Ludikhuyze, L, Broeck, I Van den, and Weemaes, C. "Effects of high pressure on enzymes related to food quality". *Trends in Food Science & Technology* 9.5 (1998), 197–203.

[104] Luong, TQ, Erwin, N, Neumann, M, Schmidt, A, Loos, C, Schmidt, V, Fändrich, M, and Winter, R. "Hydrostatic pressure increases the catalytic activity of amyloid fibril enzymes". *Angewandte Chemie International Edition* 55.40 (2016), 12412–12416.

[105] Morild, E. "The theory of pressure effects on enzymes". *Advances in protein chemistry*. Vol. 34. Elsevier, 1981, 93–166.

[106] Seyderhelm, I, Boguslawski, S, Michaelis, G, and Knorr, D. "Pressure induced inactivation of selected food enzymes". *Journal of Food Science* 61.2 (1996), 308–310.

[107] Yancey, P, Rhea, M, Kemp, K, and Bailey, DM. "Trimethylamine oxide, betaine and other osmolytes in deep-sea animals: depth trends and effects on enzymes under hydrostatic pressure". *Cellular and molecular biology* 50.4 (2004), 371–376.

[108] Bolivar, JM, Mannsberger, A, Thomsen, MS, Tekautz, G, and Nidetzky, B. "Process intensification for O2-dependent enzymatic transformations in continuous single-phase pressurized flow". *Biotechnology and Bioengineering* 116.3 (2019), 503–514.

[109] Bonting, CF, Kortstee, G, and Zehnder, A. “Properties of polyphosphate: AMP phosphotransferase of Acinetobacter strain 210A”. *Journal of bacteriology* 173.20 (1991), 6484–6488.

[110] Ishige, K and Noguchi, T. “Polyphosphate: AMP phosphotransferase and polyphosphate: ADP phosphotransferase activities of Pseudomonas aeruginosa”. *Biochemical and biophysical research communications* 281.3 (2001), 821–826.

[111] Ishige, K, Zhang, H, and Kornberg, A. “Polyphosphate kinase (PPK2), a potent, polyphosphate-driven generator of GTP”. *Proceedings of the National Academy of Sciences* 99.26 (2002), 16684–16688.

[112] Meyer, HP, Eichhorn, E, Hanlon, S, Lütz, S, Schürmann, M, Wohlgemuth, R, and Coppolecchia, R. “The use of enzymes in organic synthesis and the life sciences: perspectives from the Swiss Industrial Biocatalysis Consortium (SIBC)”. *Catalysis Science & Technology* 3.1 (2013), 29–40.

[113] Kuroda, A and Kornberg, A. “Polyphosphate kinase as a nucleoside diphosphate kinase in Escherichia coli and Pseudomonas aeruginosa”. *Proceedings of the National Academy of Sciences* 94.2 (1997), 439–442.

[114] Zhao, H and Van Der Donk, WA. “Regeneration of cofactors for use in biocatalysis”. *Current opinion in biotechnology* 14.6 (2003), 583–589.

[115] Mozhaev, VV, Lange, R, Kudryashova, EV, and Balny, C. “Application of high hydrostatic pressure for increasing activity and stability of enzymes”. *Biotechnology and Bioengineering* 52.2 (1996), 320–331.

[116] Zhao, Z, Herbst, D, Niemeyer, B, and He, L. “High pressure enhances activity and selectivity of Candida rugosa lipase immobilized onto silica nanoparticles in organic solvent”. *Food and Bioproducts Processing* 96 (2015), 240–244.

[117] Schmalle, M, Arndt, L, Lopez Haro, F, Aßmann, M, Höltng, K, Götz, S, Kuballa, J, Bubenheim, P, Pörtner, R, and Liese, A. “High hydrostatic pressure effects and kinetics of immobilized Ruegeria pomeroyi polyphosphate kinases in a continuously operated packed bed reactor”.

[118] Hill, CG and Root, TW. *An introduction to chemical engineering kinetics & reactor design*. Wiley Online Library, 1977.

[119] Zhong, JJ. “Recent advances in bioreactor engineering”. *Korean Journal of Chemical Engineering* 27 (2010), 1035–1041.

[120] Eibl, R, ed. *Cell and tissue reaction engineering*. en. Principles and practice series. OCLC: ocn280347494. Berlin: Springer, 2009. ISBN: 978-3-540-68175-5 978-3-642-01874-9 978-3-540-68182-3.

[121] Pörtner, R, Nagel-Heyer, S, Goepfert, C, Adamietz, P, and Meenen, NM. "Bioreactor design for tissue engineering". *Journal of bioscience and bioengineering* 100.3 (2005), 235–245.

[122] Lindeque, RM and Woodley, JM. "Reactor selection for effective continuous biocatalytic production of pharmaceuticals". *Catalysts* 9.3 (2019), 262.

[123] Tufvesson, P, Lima-Ramos, J, Haque, NA, Gernaey, KV, and Woodley, JM. "Advances in the process development of biocatalytic processes". *Organic Process Research & Development* 17.10 (2013), 1233–1238.

[124] Khamseh, AAG and Miccio, M. "Comparison of batch, fed-batch and continuous well-mixed reactors for enzymatic hydrolysis of orange peel wastes". *Process biochemistry* 47.11 (2012), 1588–1594.

[125] Maria, G. "Model-based optimisation of a batch reactor with a coupled bi-enzymatic process for mannitol production". *Computers & Chemical Engineering* 133 (2020), 106628.

[126] Willeman, W, Straathof, A, and Heijnen, J. "Comparison of a batch, fed-batch and continuously operated stirred-tank reactor for the enzymatic synthesis of (R)-mandelonitrile by using a process model". *Bioprocess and biosystems engineering* 24.5 (2002), 281–287.

[127] Garcia, V, Cabassud, M, Le Lann, M, Pibouleau, L, and Casamatta, G. "Constrained optimization for fine chemical productions in batch reactors". *The Chemical Engineering Journal and the Biochemical Engineering Journal* 59.3 (1995), 229–241.

[128] Plutschack, MB, Pieber, B, Gilmore, K, and Seeberger, PH. "The hitchhiker's guide to flow chemistry". *Chemical reviews* 117.18 (2017), 11796–11893.

[129] Paschalidis, L, Arana-Peña, S, Sieber, V, and Burger, J. "Modeling Enzymatic Cascade Reactions Immobilized in Plug-Flow Reactors for Flow Biocatalysis". *Chemie Ingenieur Technik* 96.6 (2024), 741–748.

[130] Salvi, HM, Kamble, MP, and Yadav, GD. "Synthesis of geraniol esters in a continuous-flow packed-bed reactor of immobilized lipase: Optimization of process parameters and kinetic modeling". *Applied biochemistry and biotechnology* 184 (2018), 630–643.

[131] Yoshida, A, Hama, S, Tamadani, N, Fukuda, H, and Kondo, A. "Improved performance of a packed-bed reactor for biodiesel production through whole-cell biocatalysis employing a high-lipase-expression system". *Biochemical engineering journal* 63 (2012), 76–80.

[132] Brink, L, Tramper, J, Luyben, KCA, and Van't Riet, K. "Biocatalysis in organic media". *Enzyme and Microbial Technology* 10.12 (1988), 736–743.

[133] Tamborini, L, Fernandes, P, Paradisi, F, and Molinari, F. "Flow bioreactors as complementary tools for biocatalytic process intensification". *Trends in biotechnology* 36.1 (2018), 73–88.

[134] Aragon, CC, Santos, AF, Ruiz-Matute, AI, Corzo, N, Guisan, JM, Monti, R, and Mateo, C. "Continuous production of xylooligosaccharides in a packed bed reactor with immobilized-stabilized biocatalysts of xylanase from Aspergillus versicolor". *Journal of Molecular Catalysis B: Enzymatic* 98 (2013), 8–14.

[135] Chen, HC, Ju, HY, Wu, TT, Liu, YC, Lee, CC, Chang, C, Chung, YL, Shieh, CJ, et al. "Continuous production of lipase-catalyzed biodiesel in a packed-bed reactor: optimization and enzyme reuse study". *BioMed Research International* 2011 (2011).

[136] Albertini, A, Reis, A, Teles, F, Souza, J, Rolim Filho, J, Freire, V, Santos, R, Martins, J, Cavada, B, Martins, D, et al. "The new flow system approach in packed bed reactor applicable for immobilized enzyme". *Journal of Molecular Catalysis B: Enzymatic* 79 (2012), 1–7.

[137] Mokhtar, NF, Abd. Rahman, RNZR, Muhd Noor, ND, Mohd Shariff, F, and Mohamad Ali, MS. "The immobilization of lipases on porous support by adsorption and hydrophobic interaction method". *Catalysts* 10.7 (2020), 744.

[138] Gao, Y, Muzzio, FJ, and Ierapetritou, MG. "A review of the Residence Time Distribution (RTD) applications in solid unit operations". *Powder technology* 228 (2012), 416–423.

[139] Cazes, J et al. "Encyclopedia of chromatography" (2001).

[140] Scheper, T, Beutel, S, McGuinness, N, Heiden, S, Oldiges, M, Lammers, F, and Reardon, KF. "Digitalization and bioprocessing: Promises and challenges". *Digital Twins: Tools and Concepts for Smart Biomanufacturing* (2021), 57–69.

[141] Smiatek, J, Jung, A, and Bluhmki, E. “Towards a Digital Bioprocess Replica: Computational Approaches in Biopharmaceutical Development and Manufacturing”. *Trends in Biotechnology* (2020).

[142] Surendhiran, D, Vijay, M, Sivaprakash, B, and Sirajunnisa, A. “Kinetic modeling of microalgal growth and lipid synthesis for biodiesel production”. *3 Biotech* 5.5 (2015), 663–669.

[143] Pinsky, M and Karlin, S. *An introduction to stochastic modeling*. Academic press, 2010.

[144] Arndt, L, Wiegmann, V, Kuchemüller, KB, Baganz, F, Pörtner, R, and Möller, J. “Model-based workflow for scale-up of process strategies developed in miniaturized bioreactor systems”. *Biotechnology Progress* (2021), e3122.

[145] Consolo, F, Bariani, C, Mantalaris, A, Montevecchi, F, Redaelli, A, and Morbiducci, U. “Computational modeling for the optimization of a cardiogenic 3D bioprocess of encapsulated embryonic stem cells”. *Biomechanics and modeling in mechanobiology* 11.1 (2012), 261–277.

[146] Kyriakopoulos, S, Ang, KS, Lakshmanan, M, Huang, Z, Yoon, S, Gunawan, R, and Lee, DY. “Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing”. *Biotechnology journal* 13.3 (2018), 1700229.

[147] Del Rio-Chanona, EA, Cong, X, Bradford, E, Zhang, D, and Jing, K. “Review of advanced physical and data-driven models for dynamic bioprocess simulation: Case study of algae-bacteria consortium wastewater treatment”. *Biotechnology and bioengineering* 116.2 (2019), 342–353.

[148] Zhang, D, Del Rio-Chanona, EA, Petsagkourakis, P, and Wagner, J. “Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization”. *Biotechnology and bioengineering* 116.11 (2019), 2919–2930.

[149] Sontakke, JB and Yadav, GD. “Kinetic modeling and statistical optimization of lipase catalyzed enantioselective resolution of (R, S)-2-pentanol”. *Industrial & engineering chemistry research* 50.23 (2011), 12975–12983.

[150] Zimmermann, V, Hennemann, HG, Daußmann, T, and Kragl, U. “Modelling the reaction course of N-acetylneuraminic acid synthesis from N-acetyl-D-glucosamine - new strategies for the optimisation of neuraminic acid synthesis”. *Applied microbiology and biotechnology* 76.3 (2007), 597–605.

[151] Gosling, I. "Process simulation and modeling for industrial bioprocessing: tools and techniques". *Industrial Biotechnology* 1.2 (2005), 106–109.

[152] Van Impe, JF. "Power and limitations of model based bioprocess optimization". *Mathematics and Computers in Simulation* 42.2-3 (1996), 159–169.

[153] Liu, Y and Gunawan, R. "Bioprocess optimization under uncertainty using ensemble modeling". *Journal of biotechnology* 244 (2017), 34–44.

[154] Gao, J, Gorenflo, VM, Scharer, JM, and Budman, HM. "Dynamic metabolic modeling for a MAB bioprocess". *Biotechnology progress* 23.1 (2007), 168–181.

[155] Krichen, E, Rapaport, A, and Fouilland, E. "About frame estimation of growth functions and robust prediction in bioprocess modeling". *Journal of Process Control* 85 (2020), 121–135.

[156] Hernández Rodríguez, T, Posch, C, Schmutzhard, J, Stettner, J, Weihs, C, Pörtner, R, and Frahm, B. "Predicting industrial scale cell culture seed trains - a Bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using a MCMC method". *Biotechnology and Bioengineering* 116.11 (2019), 2944–2959. DOI: 10.1002/bit.27125.

[157] Barricelli, BR, Casiraghi, E, and Fogli, D. "A survey on digital twin: Definitions, characteristics, applications, and design implications". *IEEE Access* 7 (2019), 167653–167671.

[158] Trelea, IC, Titica, M, Landaud, S, Latrille, E, Corrieu, G, and Cheruy, A. "Predictive modelling of brewing fermentation: from knowledge-based to black-box models". *Mathematics and Computers in Simulation* 56.4-5 (2001), 405–424.

[159] Érdi, P and Tóth, J. *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press, 1989.

[160] Dalmau, M, Atanasova, N, Gabarrón, S, Rodriguez-Roda, I, and Comas, J. "Comparison of a deterministic and a data driven model to describe MBR fouling". *Chemical Engineering Journal* 260 (2015), 300–308.

[161] Rivera, EC, Da Costa, AC, Lunelli, BH, Maciel, MRW, and Filho, RM. "Kinetic modeling and parameter estimation in a tower bioreactor for bioethanol production". *Biotechnology for Fuels and Chemicals: Proceedings of the Twenty-Ninth Symposium*

on Biotechnology for Fuels and Chemicals Held April 29-May 2, 2007, in Denver, Colorado. Springer. 2008, 681–691.

[162] Chhatre, S. "Extreme Scale-Down Approaches for Rapid Chromatography Column Design and Scale-Up During Bioprocess Development". *Measurement, Monitoring, Modelling and Control of Bioprocesses* (2013), 109–135.

[163] Lencastre Fernandes, R, Bodla, VK, Carlquist, M, Heins, AL, Eliasson Lantz, A, Sin, G, and Gernaey, KV. "Applying mechanistic models in bioprocess development". *Measurement, monitoring, modelling and control of bioprocesses* (2013), 137–166.

[164] Velayudhan, A. "Overview of integrated models for bioprocess engineering". *Current opinion in chemical engineering* 6 (2014), 83–89.

[165] Fredrickson, A, Megee Iii, R, and Tsuchiya, H. "Mathematical models for fermentation processes". *Advances in applied microbiology*. Vol. 13. Elsevier, 1970, 419–465.

[166] Szabo, R and Lente, G. "A Comparison of the Stochastic and Deterministic Approaches in a Nucleation-Growth Type Model of Nanoparticle Formation". *Chemistry of Materials* 33.13 (2021), 5430–5436.

[167] Bayer, B, Stosch, M von, Striedner, G, and Duerkop, M. "Comparison of modeling methods for DoE-based holistic upstream process characterization". *Biotechnology journal* 15.5 (2020), 1900551.

[168] Tzafriri, AR. "Michaelis-Menten kinetics at high enzyme concentrations". *Bulletin of mathematical biology* 65.6 (2003), 1111–1129.

[169] Zhong, C, Wei, P, and Zhang, YHP. "A kinetic model of one-pot rapid biotransformation of cellobiose from sucrose catalyzed by three thermophilic enzymes". *Chemical Engineering Science* 161 (2017), 159–166.

[170] Prunescu, RM and Sin, G. "Dynamic modeling and validation of a lignocellulosic enzymatic hydrolysis process-A demonstration scale study". *Bioresource technology* 150 (2013), 393–403.

[171] Patil, AS and Pethe, AM. "Quality by Design (QbD) : A new concept for development of quality pharmaceuticals". en (2013), 7.

[172] Gurunathan, B and Sahadevan, R. "Design of experiments and artificial neural network linked genetic algorithm for modeling and optimization of L-asparaginase production by Aspergillus terreus MTCC 1782". *Biotechnology and bioprocess Engineering* 16.1 (2011), 50–58.

[173] Tholudur, A, Sorensen, T, Zhu, X, and Shepard, S. "Using design of experiments to assess Escherichia coli fermentation robustness". *BioProcess Int* 3.10 (2005).

[174] Carteret, C, Jacoby, J, and Blin, J. "Using factorial experimental design to optimize biocatalytic biodiesel production from Mucor Miehei Lipase immobilized onto ordered mesoporous materials". *Microporous and Mesoporous Materials* 268 (2018), 39–45.

[175] Vanaja, K and Shobha Rani, R. "Design of experiments: concept and applications of Plackett Burman design". *Clinical research and regulatory affairs* 24.1 (2007), 1–23.

[176] Nwabueze, TU. "Basic steps in adapting response surface methodology as mathematical modelling for bioprocess optimisation in the food systems". *International Journal of Food Science & Technology* 45.9 (2010), 1768–1776.

[177] Annadurai, G, Balan, SM, and Murugesan, T. "Design of experiments in the biodegradation of phenol using immobilized Pseudomonas pictorum (NICM-2077) on activated carbon". *Bioprocess Engineering* 22.2 (2000), 101–107.

[178] Bakkiyaraj, S, Syed, MB, Devanesan, M, and Thangavelu, V. "Production and optimization of biodiesel using mixed immobilized biocatalysts in packed bed reactor". *Environmental Science and Pollution Research* 23 (2016), 9276–9283.

[179] Dean, A. *Design and Analysis of Experiments*. Ed. by Voss, D. Second edition. Springer Texts in Statistics. Cham: Springer, 2017.

[180] Kalil, S, Maugeri, F, and Rodrigues, M. "Response surface analysis and simulation as a tool for bioprocess design and optimization". *Process biochemistry* 35.6 (2000), 539–550.

[181] Kuchemüller, KB, Pörtner, R, and Möller, J. "Digital twins and their role in model-assisted design of experiments". *Digital Twins: Applications to the Design and Optimization of Bioprocesses* (2021), 29–61.

[182] Kasemiire, A, Avohou, HT, De Bleye, C, Sacre, PY, Dumont, E, Hubert, P, and Ziemons, E. "Design of experiments and design space approaches in the pharmaceutical bioprocess optimization". *European Journal of Pharmaceutics and Biopharmaceutics* 166 (2021), 144–154.

[183] Oladipo, B and Betiku, E. "Process optimization of solvent extraction of seed oil from Moringa oleifera: An appraisal of quantitative and qualitative process variables on oil quality using D-optimal design". *Biocatalysis and Agricultural Biotechnology* 20 (2019), 101187.

[184] Abnisa, F, Daud, WW, and Sahu, J. "Optimization and characterization studies on bio-oil production from palm shell by pyrolysis using response surface methodology". *Biomass and bioenergy* 35.8 (2011), 3604–3616.

[185] Davim, JP and Aveiro, P. *Design of experiments in production engineering*. Springer, 2016.

[186] MATLAB. *9.7.0.1190202 (R2019b)*. Natick, Massachusetts: The MathWorks Inc., 2020.

[187] Annadurai, G and Sheeja, R. "Use of Box-Behnken design of experiments for the adsorption of verofix red using biopolymer". *Bioprocess engineering* 18.6 (1998), 463–466.

[188] Bilal, M and Iqbal, HM. "Lignin peroxidase immobilization on Ca-alginate beads and its dye degradation performance in a packed bed reactor system". *Biocatalysis and Agricultural Biotechnology* 20 (2019), 101205.

[189] López, F, Quintana, MD, and Fernández, AG. "The use of a D-optimal design to model the effects of temperature, NaCl, type and acid concentration on Lactobacillus pentosus IGLAC01". *Journal of applied microbiology* 101.4 (2006), 913–926.

[190] Telen, D, Logist, F, Van Derlinden, E, and Van Impe, J. "Approximate robust optimal experiment design in dynamic bioprocess models". *2012 20th Mediterranean Conference on Control & Automation (MED)*. IEEE. 2012, 157–162.

[191] Khuri, AI and Mukhopadhyay, S. "Response surface methodology". *Wiley interdisciplinary reviews: Computational statistics* 2.2 (2010), 128–149.

[192] Richelle, A and Bogaerts, P. "Systematic methodology for bioprocess model identification based on generalized kinetic functions". *Biochemical engineering journal* 100 (2015), 41–49.

[193] Bernaerts, K and Van Impe, JF. "Data-driven approaches to the modelling of bioprocesses". *Transactions of the Institute of Measurement and Control* 26.5 (2004), 349–372.

[194] Teixeira, AP, Alves, C, Alves, PM, Carrondo, MJ, and Oliveira, R. "Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control". *BMC bioinformatics* 8 (2007), 1–15.

[195] Ashoori, A, Moshiri, B, Khaki-Sedigh, A, and Bakhtiari, MR. "Optimal control of a nonlinear fed-batch fermentation process using model predictive approach". *Journal of Process Control* 19.7 (2009), 1162–1173.

[196] Sommeregger, W, Sissolak, B, Kandra, K, Stosch, M von, Mayer, M, and Striedner, G. "Quality by control: Towards model predictive control of mammalian cell culture bioprocesses". *Biotechnology Journal* 12.7 (2017), 1600546.

[197] Levermann, P, Freiberger, F, Katha, U, Zaun, H, Möller, J, Hass, VC, Schoop, KM, Kuballa, J, and Pörtner, R. "NMPC-Based Workflow for Simultaneous Process and Model Development Applied to a Fed-Batch Process for Recombinant C. glutamicum". *Processes* 8.10 (2020), 1313.

[198] Freitas, HFSd, Olivo, JE, and Andrade, CMG. "Optimization of bioethanol in silico production process in a fed-batch bioreactor using non-linear model predictive control and evolutionary computation techniques". *Energies* 10.11 (2017), 1763.

[199] Craven, S, Whelan, J, and Glennon, B. "Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller". *Journal of Process Control* 24.4 (2014), 344–357.

[200] Horsch, AS, Hamann, D, Egger, LS, Fieg, G, and Skiborowski, M. "Demonstration of applied linear model predictive control for an enzymatic reactive dividing wall column". *Chemical engineering research and design* 178 (2022), 251–266.

[201] Hölting, K, Götz, S, Aßmann, M, Bubenheim, P, Liese, A, and Kuballa, J. "Resilient Enzymes through Immobilisation: Stable NDP Polyphosphate Phosphotransferase from Ruegeria pomeroyi for Nucleotide Regeneration". *Catalysts* 14.3 (2024), 165.

[202] Bérard, A, Blais, B, and Patience, GS. "Experimental methods in chemical engineering: Residence time distribution - RTD". *The Canadian Journal of Chemical Engineering* 98.4 (2020), 848–867.

[203] Degaleesan, S and Duduković, MP. "Liquid backmixing in bubble columns and the axial dispersion coefficient". *AIChE journal* 44.11 (1998), 2369–2378.

[204] Schmidt, I, Lottes, F, Minceva, M, Arlt, W, and Stenby, EH. "Estimation of chromatographic columns performances using computer tomography and cfd simulations". *Chemie Ingenieur Technik* 83.1-2 (2011), 130–142.

[205] Urbas, J et al. “Lecture on second order linear partial differential equations”. *Instructional Workshop on Analysis and Geometry, Part 1*. Centre for Mathematics and its Applications, Mathematical Sciences Institute. 1996, 39–75.

[206] Wu, Z, Yin, J, and Wang, C. *Elliptic & parabolic equations*. World Scientific, 2006.

[207] Zauderer, E. *Partial differential equations of applied mathematics*. Vol. 71. John Wiley & Sons, 2011.

[208] Hamdi, S, Schiesser, WE, and Griffiths, GW. “Method of lines”. *Scholarpedia* 2.7 (2007). revision #124335, 2859. DOI: 10.4249/scholarpedia.2859.

[209] Mendoza, NAG, Dobrosz-Gómez, I, and García, MÁG. “Modeling and simulation of an industrial falling film reactor using the method of lines with adaptive mesh. Study case: Industrial sulfonation of tridecylbenzene”. *Computers & chemical engineering* 68 (2014), 233–241.

[210] Migliorini, C, Meissner, JP, Mazzotti, M, and Carta, G. “Regioselective Enzymatic Diol Esterification in Batch and Fixed-Bed Adsorptive Reactors: Experiments and Modeling”. *Biotechnology progress* 16.4 (2000), 600–609.

[211] Saucez, P, Schiesser, W, et al. *Adaptive method of lines*. CRC Press, 2001.

[212] Kuzmin, D. *A guide to numerical methods for transport equations*. 2010.

[213] Bachmann, P, Bück, A, and Tsotsas, E. “Experimental investigation and correlation of the Bodenstein number in horizontal fluidized beds with internal baffles”. *Powder Technology* 308 (2017), 378–387.

[214] Bogatykh, I and Osterland, T. “Characterization of residence time distribution in a plug flow reactor”. *Chemie Ingenieur Technik* 91.5 (2019), 668–672.

[215] Aubin, J, Prat, L, Xuereb, C, and Gourdon, C. “Effect of microchannel aspect ratio on residence time distributions and the axial dispersion coefficient”. *Chemical Engineering and Processing: Process Intensification* 48.1 (2009), 554–559.

[216] Taylor, GI. “The dispersion of matter in turbulent flow through a pipe”. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* 223.1155 (1954), 446–468.

[217] Thömmes, J, Weiher, M, Karau, A, and Kula, MR. “Hydrodynamics and performance in fluidized bed adsorption”. *Biotechnology and bioengineering* 48.4 (1995), 367–374.

[218] Coker, AK. *Modeling of Chemical Kinetics and Reactor Design*. Elsevier, 2001.

[219] Pietsch, S, Schönherr, M, Kleine Jäger, F, and Heinrich, S. “Measurement of residence time distributions in a continuously operated spouted bed”. *Chemical Engineering & Technology* 43.5 (2020), 804–812.

[220] Rodrigues, AE. “Residence time distribution (RTD) revisited”. *Chemical engineering science* 230 (2021), 116188.

[221] Baerns, M, Hofmann, H, and Renken, A. *Lehrbuch der Technischen Chemie Band 1: Chemische Reaktionstechnik*. 1992.

[222] Meller, T. “Design und Charakterisierung eines Gas-Feststoff-Reaktionskalorimeters”. PhD thesis. Staats-und Universitätsbibliothek Hamburg Carl von Ossietzky, 2019.

[223] Shadpoor, S, Pirouzi, A, Hamze, H, and Mazaheri, D. “Determination of Bodenstein number and axial dispersion of a triangular external loop airlift reactor”. *Chemical Engineering Research and Design* 165 (2021), 61–68.

[224] Doran, PM. *Bioprocess engineering principles*. Elsevier, 1995.

[225] Jaeger, KE, Liese, A, and Syldatk, C. *Einführung in die Enzymtechnologie*. Springer, 2018.

[226] Heils, R, Niesbach, A, Wierschem, M, Claus, D, Soboll, S, Lutze, P, and Smirnova, I. “Integration of enzymatic catalysts in a continuous reactive distillation column: reaction kinetics and process simulation”. *Industrial & engineering chemistry research* 53.50 (2014), 19612–19619.

[227] Sharma, R. “Enzyme inhibition: mechanisms and scope”. *Enzyme inhibition and bioapplications* (2012), 3–36.

[228] Aymard, C and Belarbi, A. “Kinetics of thermal deactivation of enzymes: a simple three parameters phenomenological model can describe the decay of enzyme activity, irrespectively of the mechanism.” *Enzyme and microbial Technology* 27.8 (2000), 612–618.

[229] Ricca, E, Calabrò, V, Curcio, S, and Iorio, G. “Optimization of inulin hydrolysis by inulinase accounting for enzyme time-and temperature-dependent deactivation”. *Biochemical engineering journal* 48.1 (2009), 81–86.

[230] Salom, D, Hill, BR, Lear, JD, and DeGrado, WF. “pH-dependent tetramerization and amantadine binding of the transmembrane helix of M2 from the influenza A virus”. *Biochemistry* 39.46 (2000), 14160–14170.

[231] Arndt, L, Schmalle, M, Aßmann, M, Höltng, K, Kuballa, J, Bubenheim, P, Liese, A, and Pörtner, R. "Model-assisted design and optimization of a continuous biocatalytical cofactor regeneration process under high hydrostatic pressure".

[232] Moser, A, Appl, C, Brüning, S, and Hass, VC. "Mechanistic mathematical models as a basis for digital twins". *Digital Twins: Tools and Concepts for Smart Biomanufacturing* (2021), 133–180.

[233] De, S, Wojtkiewicz, SF, and Johnson, EA. "Efficient optimal design and design-under-uncertainty of passive control devices with application to a cable-stayed bridge". *Structural Control and Health Monitoring* 24.2 (2017), e1846.

[234] Loucks, DP and Van Beek, E. *Water resource systems planning and management: An introduction to methods, models, and applications*. Springer, 2017.

[235] Wechselberger, P, Sagmeister, P, and Herwig, C. "Model-based analysis on the extractability of information from data in dynamic fed-batch experiments". en. *Biotechnology Progress* 29.1 (Jan. 2013), 285–296.

[236] Aptula, AO, Jeliazkova, NG, Schultz, TW, and Cronin, MT. "The better predictive model: high q2 for the training set or low root mean square error of prediction for the test set?" *Qsar & Combinatorial Science* 24.3 (2005), 385–396.

[237] Scott, DW. *Statistics: A Concise Mathematical Introduction for Students, Scientists, and Engineers*. John Wiley & Sons, 2020.

[238] Hogg, RV, McKean, JW, and Craig, AT. *Introduction to Mathematical Statistics 8th ed*. Pearson, 2019.

[239] Johnson, RT and Montgomery, DC. "Choice of second-order response surface designs for logistic and Poisson regression models". *International Journal of Experimental Design and Process Optimisation* 1.1 (2009), 2–23.

[240] Chai, T and Draxler, RR. "Root mean square error (RMSE) or mean absolute error (MAE)". *Geoscientific Model Development Discussions* 7.1 (2014), 1525–1534.

[241] Möller, J, Korte, K, Pörtner, R, Zeng, AP, and Jandt, U. "Model-based identification of cell-cycle-dependent metabolism and putative autocrine effects in antibody producing CHO cell culture". *Biotechnology and Bioengineering* 115.12 (2018), 2996–3008.

[242] Nelder, JA and Mead, R. "A simplex method for function minimization". *The computer journal* 7.4 (1965), 308–313.

[243] Barz, T, Körkel, S, Wozny, G, and Lopez Cárdenas, D. "Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design". *Computers & Chemical Engineering* 77 (2015), 24–42.

[244] Barz, T, Sommer, A, Wilms, T, Neubauer, P, and Bournazou, M. "Adaptive optimal operation of a parallel robotic liquid handling station". *IFAC-PapersOnLine* 51.2 (2018), 765–770.

[245] Cruz Bournazou, M, Barz, T, Nickel, D, Lopez Cárdenas, D, Glauche, F, Knepper, A, and Neubauer, P. "Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities". *Biotechnology and bioengineering* 114.3 (2017), 610–619.

[246] Pais, DA, Portela, RM, Carrondo, MJ, Isidro, IA, and Alves, PM. "Enabling PAT in insect cell bioprocesses: In situ monitoring of recombinant adeno-associated virus production by fluorescence spectroscopy". *Biotechnology and bioengineering* 116.11 (2019), 2803–2814.

[247] Carugo, O and Pongor, S. "A normalized root-mean-spuare distance for comparing protein three-dimensional structures". *Protein science* 10.7 (2001), 1470–1473.

[248] Cameron, AC and Windmeijer, FA. "An R-squared measure of goodness of fit for some common nonlinear regression models". *Journal of econometrics* 77.2 (1997), 329–342.

[249] Maran, JP and Priya, B. "Comparison of response surface methodology and artificial neural network approach towards efficient ultrasound-assisted biodiesel production from muskmelon oil". *Ultrasonics Sonochemistry* 23 (2015), 192–200.

[250] Ratnam, B, Rao, SS, Mendu, DR, Rao, MN, and Ayyanna, C. "Optimization of medium constituents and fermentation conditions for the production of ethanol from palmyra jaggery using response surface methodology". *World Journal of microbiology and Biotechnology* 21.4 (2005), 399–404.

[251] Myers, RH, Montgomery, DC, and Anderson-Cook, CM. *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, 2016.

[252] Mukhopadhyay, S and Khuri, AI. "A new graphical approach for comparing response surface designs on the basis of the mean squared error of prediction criterion". *Statistics and Applications* 6 (2008), 293–324.

[253] Simpson, T, Mistree, F, Korte, J, and Mauery, T. "Comparison of response surface and kriging models for multidisciplinary design optimization". *7th AIAA/USAF/-NASA/ISSMO symposium on multidisciplinary analysis and optimization*. 1998, 4755.

[254] Mishra, P, Singh, U, Pandey, CM, Mishra, P, and Pandey, G. "Application of student's t-test, analysis of variance, and covariance". *Annals of cardiac anaesthesia* 22.4 (2019), 407.

[255] Kuckartz, U, Rädiker, S, Ebert, T, and Schehl, J. *Statistik: eine verständliche Einführung*. Springer-Verlag, 2013.

[256] Allam, AS, Bassioni, HA, Kamel, W, and Ayoub, M. "Estimating the standardized regression coefficients of design variables in daylighting and energy performance of buildings in the face of multicollinearity". *Solar Energy* 211 (2020), 1184–1193.

[257] Menard, S. "Standards for standardized logistic regression coefficients". *Social Forces* 89.4 (2011), 1409–1428.

[258] Razavi, S and Gupta, HV. "What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models". *Water Resources Research* 51.5 (2015), 3070–3092.

[259] Lopes, RH, Reid, I, and Hobson, PR. "The two-dimensional Kolmogorov-Smirnov test" (2007).

[260] Razali, NM, Wah, YB, et al. "Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests". *Journal of statistical modeling and analytics* 2.1 (2011), 21–33.

[261] Conover, WJ. *Practical nonparametric statistics*. Vol. 350. John Wiley & Sons, 1998.

[262] Eckey, HF, Kosfeld, R, Türck, M, Eckey, HF, Kosfeld, R, and Türck, M. "Grenzwertsätze und Approximation von Verteilungen". *Wahrscheinlichkeitsrechnung und Induktive Statistik: Grundlagen-Methoden-Beispiele* (2005), 156–179.

[263] Polyzoidis, A, Altenburg, T, Schwarzer, M, Löbbecke, S, and Kaskel, S. "Continuous microreactor synthesis of ZIF-8 with high space-time-yield and tunable particle size". *Chemical Engineering Journal* 283 (2016), 971–977.

[264] Fabbri, F, Bertolini, FA, Guebitz, GM, and Pellis, A. "Biocatalyzed synthesis of flavor esters and polyesters: A Design of Experiments (DoE) approach". *International Journal of Molecular Sciences* 22.16 (2021), 8493.

[265] Gupta, J, Agarwal, M, and Dalai, A. "Optimization of biodiesel production from mixture of edible and nonedible vegetable oils". *Biocatalysis and Agricultural Biotechnology* 8 (2016), 112–120.

[266] Sin, G, Meyer, AS, and Gernaey, KV. "Assessing reliability of cellulose hydrolysis models to support biofuel process design: identifiability and uncertainty analysis". *Computers & Chemical Engineering* 34.9 (2010), 1385–1392.

[267] Romero-Fernández, M and Paradisi, F. "Protein immobilization technology for flow biocatalysis". *Current opinion in chemical biology* 55 (2020), 1–8.

[268] Roode, BM de, Franssen, MC, Padt, Avd, and Boom, RM. "Perspectives for the industrial enzymatic production of glycosides". *Biotechnology progress* 19.5 (2003), 1391–1402.

[269] Paschalidis, L, Beer, B, Sutiono, S, Sieber, V, and Burger, J. "Design of enzymatic cascade reactors through multi-objective dynamic optimization". *Biochemical Engineering Journal* 181 (2022), 108384.

[270] Afandizadeh, S and Foumeny, E. "Design of packed bed reactors: guides to catalyst shape, size, and loading selection". *Applied thermal engineering* 21.6 (2001), 669–682.

[271] Bolivar, JM and López-Gallego, F. "Characterization and evaluation of immobilized enzymes for applications in flow reactors". *Current Opinion in Green and Sustainable Chemistry* 25 (2020), 100349.

[272] Farhan, M, Larjo, A, Yli-Harja, O, and Aho, T. "Modeling bioprocess scale-up utilizing regularized linear and logistic regression". *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE. 2013, 1–6.

[273] Sinner, P, Herwig, C, and Kager, J. "Time scale analysis and optimization of a continuous microbial bioprocess". *Computer Aided Chemical Engineering*. Vol. 48. Elsevier, 2020, 1603–1608.

[274] Manenti, F, Cieri, S, Restelli, M, and Bozzano, G. "Dynamic modeling of the methanol synthesis fixed-bed reactor". *Computers & Chemical Engineering* 48 (2013), 325–334.

[275] Rezvani, F, Azargoshasb, H, Jamialahmadi, O, Hashemi-Najafabadi, S, Mousavi, SM, and Shojaosadati, SA. "Experimental study and CFD simulation of phenol removal by immobilization of soybean seed coat in a packed-bed bioreactor". *Biochemical Engineering Journal* 101 (2015), 32–43.

[276] Huang, SM, Li, HJ, Liu, YC, Kuo, CH, and Shieh, CJ. "An efficient approach for lipase-catalyzed synthesis of retinyl laurate nutraceutical by combining ultrasound assistance and artificial neural network optimization". *Molecules* 22.11 (2017), 1972.

[277] Chen, Y, Yang, O, Sampat, C, Bhalode, P, Ramachandran, R, and Ierapetritou, M. "Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review". *Processes* 8.9 (2020), 1088.

A Supplementary

Table A.1: Experimental conditions for all conducted experiments in high hydrostatic pressure reactor, LxID=30x3 mm, 50 mmol L^{-1} sodium phosphate buffer, $c_{Mg^{2+}} = 30$ mmol L^{-1}, T=40°C, pH=7.4, enzyme loading = 86.4 mg $g^{-1}_{Carrier}$, Purolite ECR8209M carrier.

Run	**Carrier** [mg]	**Load** [%]	**E** [mg L^{-1}]	**e** [–]	**Pressure** [bar]	$\mathbf{c_{CDP}}$ [mmol L^{-1}]	$\mathbf{c_{polyP}}$ [mmol L^{-1}]	**V** [L]	$\dot{V}$ [L min^{-1}]
1	163.6	50	33337.4	0.21	0	35.98	44	$2.12 \cdot 10^{-4}$	$0.5 \cdot 10^{-3}$
2	163.6	50	33337.4	0.21	50	35.98	44	$2.12 \cdot 10^{-4}$	$0.5 \cdot 10^{-3}$
3	163.6	50	33337.4	0.21	100	35.98	44	$2.12 \cdot 10^{-4}$	$0.5 \cdot 10^{-3}$
4	163.6	50	33337.4	0.21	150	35.98	44	$2.12 \cdot 10^{-4}$	$0.5 \cdot 10^{-3}$
5	163.6	50	33337.4	0.21	200	35.98	44	$2.12 \cdot 10^{-4}$	$0.5 \cdot 10^{-3}$
6	163.6	50	33337.4	0.21	250	35.98	44	$2.12 \cdot 10^{-4}$	$0.5 \cdot 10^{-3}$
7	163.6	50	33337.4	0.21	0	35.98	44	$2.12 \cdot 10^{-4}$	$0.5 \cdot 10^{-3}$
8	163.6	50	33337.4	0.21	50	35.98	44	$2.12 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$
9	163.6	50	33337.4	0.21	100	35.98	44	$2.12 \cdot 10^{-4}$	$1.5 \cdot 10^{-3}$
10	166.1	50	38062.9	0.25	0	36.0	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
11	166.1	50	38062.9	0.25	97	36.0	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
12	166.1	50	38062.9	0.25	203	36.0	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
13	166.1	50	38062.9	0.25	408	36.0	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
14	166.1	50	38062.9	0.25	0	36.0	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
15	166.1	50	38062.9	0.25	632	36.0	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
16	166.1	50	38062.9	0.25	796	36.0	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
17	166.1	50	38062.9	0.25	0	36.57	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
18	166.1	50	38062.9	0.25	100	36.57	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
19	166.1	50	38062.9	0.25	197	36.57	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
20	166.1	50	38062.9	0.25	399	36.57	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
21	166.1	50	38062.9	0.25	0	36.57	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
22	166.1	50	38062.9	0.25	616	36.57	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
23	166.1	50	38062.9	0.25	807	36.57	44	$2.12 \cdot 10^{-4}$	$1.0 \cdot 10^{-3}$
24	163.3	50	33276.2	0.30	0	78.0	17.2	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
25	163.3	50	33276.2	0.30	0	78.0	17.2	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
26	163.3	50	33276.2	0.30	400	78.0	17.2	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
27	163.3	50	33276.2	0.30	0	78.0	17.2	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
28	163.3	50	33276.2	0.30	0	78.0	21.5	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
29	163.3	50	33276.2	0.30	700	78.0	21.5	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
30	163.3	50	33276.2	0.30	0	78.0	21.5	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
31	186.9	50	38066.0	0.17	0	77.3	34.5	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
32	186.9	50	38066.0	0.17	799	77.3	34.5	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$
33	186.9	50	38066.0	0.17	0	79.0	34.5	$2.12 \cdot 10^{-4}$	$0.15 \cdot 10^{-3}$

Table A.2: Mathematical process model used in workflow rounds mDoE-1, mDoE-2 and mDoE-3.

Kinetics mDoE-1

$$v\,(\mathrm{i}) = v_{\mathrm{p,Dsig}} \cdot c_{\mathrm{Enz}}\,(\mathrm{i}) \cdot v_{\max} \cdot \frac{c_{\mathrm{CDP}}\,(\mathrm{i})}{K_{\mathrm{m,CDP}} + c_{\mathrm{CDP}}\,(\mathrm{i})} \cdot \frac{c_{\mathrm{polyP}}\,(\mathrm{i})}{K_{\mathrm{m,polyP}} + c_{\mathrm{polyP}}\,(\mathrm{i}) \cdot \left(1 + \frac{c_{\mathrm{polyP}}\,(\mathrm{i})}{K_{\mathrm{I,polyP}}}\right)}$$

Kinetics mDoE-2 and mDoE-3

$$v\,(\mathrm{i}) = v_{\mathrm{red}} \cdot v_{\mathrm{p,Dsig}} \cdot c_{\mathrm{Enz}}\,(\mathrm{i}) \cdot v_{\max} \cdot \frac{c_{\mathrm{CDP}}\,(\mathrm{i})}{K_{\mathrm{m,CDP}} + c_{\mathrm{CDP}}\,(\mathrm{i})} \cdot \frac{c_{\mathrm{polyP}}\,(\mathrm{i})}{K_{\mathrm{m,polyP}} + c_{\mathrm{polyP}}\,(\mathrm{i}) \cdot \left(1 + \frac{c_{\mathrm{polyP}}\,(\mathrm{i})}{K_{\mathrm{I,polyP}}}\right)}$$

Boundary conditions mDoE-1 and mDoE-2

$$c\,(1) = c_{\mathrm{inlet}}$$
$$\frac{\partial \mathrm{c}}{dt}(\mathrm{end}) = 0$$

Boundary conditions mDoE-3

$$c\,(1) = c_{\mathrm{inlet}} - \frac{D_{\mathrm{ax}}}{u}\left(\frac{\partial c}{\partial x}\right)\Big|_{\mathrm{x=0}}$$
$$\frac{\partial \mathrm{c}}{dt}(\mathrm{end}) = 0$$

Differential equations

$$\frac{\partial c_{\mathrm{CDP}}}{\partial \mathrm{t}}(\mathrm{i}) = D_{\mathrm{ax}} \frac{c_{\mathrm{CDP}}\,(\mathrm{i}-1) - 2 \cdot c_{\mathrm{CDP}}\,(\mathrm{i}) + c_{\mathrm{CDP}}\,(\mathrm{i}+1)}{L_{\mathrm{step}}^2} - u \frac{c_{\mathrm{CDP}}\,(\mathrm{i}+1) - c_{\mathrm{CDP}}\,(\mathrm{i}-1)}{2 \cdot L_{\mathrm{step}}} - f\,(i,t)$$
$$\frac{\partial c_{\mathrm{polyP}}}{\partial \mathrm{t}}(\mathrm{i}) = D_{\mathrm{ax}} \frac{c_{\mathrm{polyP}}\,(\mathrm{i}-1) - 2 \cdot c_{\mathrm{polyP}}\,(\mathrm{i}) + c_{\mathrm{polyP}}\,(\mathrm{i}+1)}{L_{\mathrm{step}}^2} - u \frac{c_{\mathrm{polyP}}\,(\mathrm{i}+1) - c_{\mathrm{polyP}}\,(\mathrm{i}-1)}{2 \cdot L_{\mathrm{step}}} - f\,(i,t)$$
$$\frac{\partial c_{\mathrm{CTP}}}{\partial \mathrm{t}}(\mathrm{i}) = D_{\mathrm{ax}} \frac{c_{\mathrm{CTP}}\,(\mathrm{i}-1) - 2 \cdot c_{\mathrm{CTP}}\,(\mathrm{i}) + c_{\mathrm{CTP}}\,(\mathrm{i}+1)}{L_{\mathrm{step}}^2} - u \frac{c_{\mathrm{CTP}}\,(\mathrm{i}+1) - c_{\mathrm{CTP}}\,(\mathrm{i}-1)}{2 \cdot L_{\mathrm{step}}} + f\,(i,t)$$
$$\frac{\partial c_{\mathrm{Enz}}}{\partial \mathrm{t}}(\mathrm{i}) = -k_{\mathrm{deact}} \cdot c_{\mathrm{Enz}}\,(\mathrm{i})$$

Table A.3: Adapted model parameter values mDoE-1, mDoE-2 and mDoE-3.

	Parameter	v_{max} [mmol min^{-1} mg^{-1}]	$K_{M,CDP}$ [mmol L^{-1}]	$K_{M,polyP}$ [mmol L^{-1}]	$K_{I,polyP}$ [mmol L^{-1}]	k_{deact} [min^{-1}]
	Start value	0.017	32.16	219.79	8.40	$3.00 \cdot 10^{-4}$
mDoE-1	**10% quantile**	0.021	8.14	0.13	10.73	$1.24 \cdot 10^{-4}$
	Median	0.024	20.61	35.53	13.13	$2.98 \cdot 10^{-4}$
	90% quantile	0.029	30.32	132.87	17.35	$3.73 \cdot 10^{-4}$
mDoE-2	**10% quantile**	0.026	7.89	0.19	11.88	$1.33 \cdot 10^{-4}$
	Median	0.030	17.07	77.69	15.29	$2.83 \cdot 10^{-4}$
	90% quantile	0.036	27.52	125.84	19.73	$3.86 \cdot 10^{-4}$
mDoE-3	**10% quantile**	0.019	6.51	29.97	7.80	$2.01 \cdot 10^{-5}$
	Median	0.024	23.60	117.76	11.64	$2.97 \cdot 10^{-4}$
	90% quantile	0.033	46.62	214.24	17.05	$4.49 \cdot 10^{-4}$

Table A.4: Coded factor combinations for four factor CCD, BBD and D-optimal Design.

	CCD					BBD					D-Optimal				
Run	**Flow** [L min^{-1}]	$\mathbf{c_{CDP}}$ [mmol L^{-1}]	$\mathbf{c_{polyP}}$ [mmol L^{-1}]	**Pressure** [bar]		**Flow** [L min^{-1}]	$\mathbf{c_{CDP}}$ [mmol L^{-1}]	$\mathbf{c_{polyP}}$ [mmol L^{-1}]	**Pressure** [bar]		**Flow** [L min^{-1}]	$\mathbf{c_{CDP}}$ [mmol L^{-1}]	$\mathbf{c_{polyP}}$ [mmol L^{-1}]	**Pressure** [bar]	
1	-1	-1	-1	-1		-1	-1	0	0		-1	-1	-1	-1	
2	-1	-1	-1	1		-1	1	0	0		1	-1	0	-1	
3	-1	-1	1	-1		1	-1	0	0		-1	-1	1	1	
4	-1	-1	1	1		1	1	0	0		-1	1	-1	1	
5	-1	1	-1	-1		0	0	-1	-1		-1	1	1	-1	
6	-1	1	-1	1		0	0	-1	1		1	1	-1	1	
7	-1	1	1	-1		0	0	1	-1		0	1	-1	-1	
8	-1	1	1	1		0	0	1	1		-1	-1	1	0	
9	1	-1	-1	-1		-1	0	0	-1		1	-1	1	-1	
10	1	-1	-1	1		-1	0	0	1		1	-1	1	1	
11	1	-1	1	-1		1	0	0	-1		0	-1	-1	1	
12	1	-1	1	1		1	0	0	1		1	1	1	-1	
13	1	1	-1	-1		0	-1	-1	0		-1	-1	-1	1	
14	1	1	-1	1		0	-1	1	0		0	1	1	1	
15	1	1	1	-1		0	1	-1	0		0	-1	1	-1	
16	1	1	1	1		0	1	1	0		-1	1	1	1	
17	-2	0	0	0		-1	0	-1	0		-1	1	0	-1	
18	2	0	0	0		-1	0	1	0		1	-1	-1	0	
19	0	-2	0	0		1	0	-1	0		1	0	-1	-1	
20	0	2	0	0		1	0	1	0		-1	0	-1	1	
21	0	0	-2	0		0	-1	0	-1		-1	1	-1	0	
22	0	0	2	0		0	-1	0	1		1	1	-1	-1	
23	0	0	0	-2		0	1	0	-1		1	1	0	1	
24	0	0	0	2		0	1	0	1		1	1	1	0	
25	0	0	0	0		0	0	0	0		1	-1	-1	1	
26	0	0	0	0		0	0	0	0		0	0	0	0	
27	0	0	0	0		0	0	0	0		-1	-1	0	1	
28	0	0	0	0		-	-	-	-		1	0	1	1	
29	0	0	0	0		-	-	-	-		0	0	0	0	
30	0	0	0	0		-	-	-	-		-1	0	1	-1	
31	0	0	0	0		-	-	-	-		-	-	-		

Table A.5: Estimated coefficients of second-order linear regression model considering linear, interaction and quadratic effects.

Coefficient	**mDoE-1**	**mDoE-2 TC**	**mDoE-2 TC MBS**	**mDoE-2**	**mDoE-2 MBS**	**mDoE-3 TC**	**mDoE-3 TC MBS**
$\hat{\beta}_0$	8.346	2.012	3.573	1.982	3.883	0.830	3.151
$\hat{\beta}_{\text{Flow}}$	$-3.04 \cdot 10^4$	$-1.24 \cdot 10^4$	$-1.31 \cdot 10^4$	$-1.34 \cdot 10^4$	$-1.42 \cdot 10^4$	$-1.17 \cdot 10^4$	$-1.31 \cdot 10^4$
$\hat{\beta}_{\text{CDP}}$	0.541	0.287	0.338	0.328	0.404	0.316	0.358
$\hat{\beta}_{\text{polyP}}$	0.173	0.231	0.137	0.257	0.145	0.262	0.154
$\hat{\beta}_{\text{Pressure}}$	$2.91 \cdot 10^{-3}$	$1.87 \cdot 10^{-3}$	$2.01 \cdot 10^{-3}$	$1.96 \cdot 10^{-3}$	$2.35 \cdot 10^{-3}$	$1.99 \cdot 10^{-3}$	$2.91 \cdot 10^{-3}$
$\hat{\beta}_{\text{Flow,CDP}}$	-282.030	-117.383	-135.934	-137.585	-161.629	-139.324	-149.050
$\hat{\beta}_{\text{Flow,polyP}}$	65.021	-19.622	2.050	-20.926	-5.612	-31.651	-6.195
$\hat{\beta}_{\text{Flow,Pressure}}$	-1.332	-0.798	-0.813	-0.860	-0.874	-0.826	-0.834
$\hat{\beta}_{\text{CDP,polyP}}$	$-1.21 \cdot 10^{-3}$	$4.86 \cdot 10^{-4}$	$-2.34 \cdot 10^{-4}$	$5.68 \cdot 10^{-4}$	$-3.35 \cdot 10^{-4}$	$8.20 \cdot 10^{-4}$	$-1.41 \cdot 10^{-4}$
$\hat{\beta}_{\text{CDP,Pressure}}$	$2.76 \cdot 10^{-5}$	$1.63 \cdot 10^{-5}$	$1.85 \cdot 10^{-5}$	$1.90 \cdot 10^{-5}$	$2.18 \cdot 10^{-5}$	$1.90 \cdot 10^{-5}$	$2.01 \cdot 10^{-5}$
$\hat{\beta}_{\text{polyP,Pressure}}$	$-4.55 \cdot 10^{-6}$	$2.95 \cdot 10^{-6}$	$2.39 \cdot 10^{-7}$	$3.16 \cdot 10^{-6}$	$1.20 \cdot 10^{-6}$	$4.41 \cdot 10^{-6}$	$1.23 \cdot 10^{-6}$
$\hat{\beta}_{\text{Flow}^2}$	$1.79 \cdot 10^7$	$7.45 \cdot 10^6$	$7.59 \cdot 10^6$	$8.12 \cdot 10^6$	$8.45 \cdot 10^6$	$7.55 \cdot 10^6$	$7.82 \cdot 10^6$
$\hat{\beta}_{\text{CDP}^2}$	$-2.12 \cdot 10^{-3}$	$-1.84 \cdot 10^{-3}$	$-1.87 \cdot 10^{-3}$	$-2.06 \cdot 10^{-3}$	$-2.23 \cdot 10^{-3}$	$-1.91 \cdot 10^{-3}$	$-1.84 \cdot 10^{-3}$
$\hat{\beta}_{\text{polyP}^2}$	$-2.97 \cdot 10^{-3}$	$-2.91 \cdot 10^{-3}$	$-1.83 \cdot 10^{-3}$	$-3.25 \cdot 10^{-3}$	$-1.79 \cdot 10^{-3}$	$-3.21 \cdot 10^{-3}$	$-1.95 \cdot 10^{-3}$
$\hat{\beta}_{\text{Pressure}^2}$	$-2.19 \cdot 10^{-6}$	$-1.47 \cdot 10^{-6}$	$-1.55 \cdot 10^{-6}$	$-1.54 \cdot 10^{-6}$	$-1.95 \cdot 10^{-6}$	$-1.72 \cdot 10^{-6}$	$-1.81 \cdot 10^{-6}$

Table A.6: p-values of second-order linear regression model coefficients considering linear, interaction and quadratic effects, based on t-statistics.

Coefficient	mDoE-1	mDoE-2 TC	mDoE-2 TC MBS	mDoE-2	mDoE-2 MBS	mDoE-3 TC	mDoE-3 TC MBS
$\hat{\beta}_0$	0.063	0.329	0.001	0.375	0.003	0.686	0.005
$\hat{\beta}_{\text{Flow}}$	$5.70 \cdot 10^{-5}$	$2.83 \cdot 10^{-4}$	$1.30 \cdot 10^{-19}$	$3.05 \cdot 10^{-4}$	$4.77 \cdot 10^{-16}$	$5.15 \cdot 10^{-4}$	$3.10 \cdot 10^{-17}$
$\hat{\beta}_{\text{CDP}}$	$1.51 \cdot 10^{-4}$	$5.19 \cdot 10^{-5}$	$4.26 \cdot 10^{-32}$	$2.89 \cdot 10^{-5}$	$6.52 \cdot 10^{-31}$	$1.99 \cdot 10^{-5}$	$3.40 \cdot 10^{-31}$
$\hat{\beta}_{\text{polyP}}$	$1.34 \cdot 10^{-1}$	$4.40 \cdot 10^{-4}$	$7.76 \cdot 10^{-7}$	$3.49 \cdot 10^{-4}$	$1.75 \cdot 10^{-5}$	$1.48 \cdot 10^{-4}$	$2.59 \cdot 10^{-7}$
$\hat{\beta}_{\text{Pressure}}$	0.752	0.672	0.380	0.683	0.399	0.655	0.373
$\hat{\beta}_{\text{Flow,CDP}}$	$1.09 \cdot 10^{-4}$	$4.18 \cdot 10^{-4}$	$1.84 \cdot 10^{-21}$	$2.01 \cdot 10^{-4}$	$1.56 \cdot 10^{-20}$	$8.57 \cdot 10^{-5}$	$3.35 \cdot 10^{-22}$
$\hat{\beta}_{\text{Flow,polyP}}$	0.257	0.469	0.883	0.477	0.741	0.254	0.679
$\hat{\beta}_{\text{Flow,Pressure}}$	0.813	0.767	0.559	0.768	0.606	0.761	0.577
$\hat{\beta}_{\text{CDP,polyP}}$	0.263	0.344	0.372	0.310	0.295	0.123	0.618
$\hat{\beta}_{\text{CDP,Pressure}}$	0.795	0.748	0.482	0.730	0.495	0.711	0.477
$\hat{\beta}_{\text{polyP,Pressure}}$	0.966	0.953	0.993	0.954	0.970	0.931	0.965
$\hat{\beta}_{\text{Flow}^2}$	$2.94 \cdot 10^{-5}$	$1.26 \cdot 10^{-4}$	$2.58 \cdot 10^{-21}$	$1.20 \cdot 10^{-4}$	$3.82 \cdot 10^{-18}$	$1.21 \cdot 10^{-4}$	$9.20 \cdot 10^{-20}$
$\hat{\beta}_{\text{CDP}^2}$	$9.90 \cdot 10^{-2}$	$5.88 \cdot 10^{-3}$	$1.28 \cdot 10^{-9}$	$4.73 \cdot 10^{-3}$	$2.79 \cdot 10^{-9}$	$4.83 \cdot 10^{-3}$	$2.30 \cdot 10^{-8}$
$\hat{\beta}_{\text{polyP}^2}$	$2.59 \cdot 10^{-2}$	$1.25 \cdot 10^{-4}$	$2.41 \cdot 10^{-9}$	$9.08 \cdot 10^{-5}$	$1.58 \cdot 10^{-6}$	$4.84 \cdot 10^{-5}$	$3.38 \cdot 10^{-9}$
$\hat{\beta}_{\text{Pressure}^2}$	0.779	0.694	0.424	0.705	0.408	0.650	0.384

Table A.7: Standardized regression coefficients of multiple linear regression model considering linear, interaction and quadratic effects.

Coefficient	**mDoE-1**	**mDoE-2**	**mDoE-2 MBS**	**mDoE-3 TC**	**mDoE-3 TC MBS**
β'_0	-0.055	0.185	0.065	0.189	0.070
β'_{Flow}	-0.705	-0.681	-0.565	-0.652	-0.565
β'_{CDP}	0.451	0.526	0.455	0.555	0.481
β'_{polyP}	-0.092	0.166	0.016	0.209	0.033
β'_{Pressure}	0.031	0.045	0.036	0.043	0.037
$\beta'_{\text{Flow,CDP}}$	-0.251	-0.201	-0.176	-0.205	-0.177
$\beta'_{\text{Flow,polyP}}$	0.058	-0.031	-0.006	-0.047	-0.007
$\beta'_{\text{Flow,Pressure}}$	-0.013	-0.014	-0.010	-0.013	-0.011
$\beta'_{\text{CDP,polyP}}$	-0.056	0.043	-0.019	0.063	-0.009
$\beta'_{\text{CDP,Pressure}}$	0.014	0.016	0.013	0.016	0.013
$\beta'_{\text{polyP,Pressure}}$	-0.002	0.003	0.001	0.004	0.001
β'_{Flow^2}	0.304	0.228	0.177	0.213	0.178
β'_{CDP^2}	-0.098	-0.157	-0.127	-0.147	-0.114
β'_{polyP^2}	-0.138	-0.248	-0.102	-0.247	-0.121
$\beta'_{\text{Pressure}^2}$	-0.012	-0.014	-0.013	-0.015	-0.013

Table A.8: Factor combinations and corresponding responses for four factor CCD mDoE-1, mDoE-2 and mDoE-3.

Run	**Flow** [L min^{-1}]	$\mathbf{c_{CDP}}$ [mmol L^{-1}]	$\mathbf{c_{polyP}}$ [mmol L^{-1}]	**Pressure** [bar]	$\mathbf{c_{CTP}}$ [mmol L^{-1}] mDoE-1	$\mathbf{c_{CTP}}$ [mmol L^{-1}] mDoE-2	$\mathbf{c_{CTP}}$ [mmol L^{-1}] mDoE-3
1	0.0003	15	15	200	8.9	6.1	5.3
2	0.0003	15	15	600	9.2	6.3	5.5
3	0.0003	15	55	200	7.6	6.9	6.4
4	0.0003	15	55	600	7.9	7.2	6.7
5	0.0003	55	15	200	21.3	11.9	11.1
6	0.0003	55	15	600	22.4	12.6	11.7
7	0.0003	55	55	200	16.9	14.0	14.1
8	0.0003	55	55	600	17.9	14.8	14.8
9	0.00105	15	15	200	3.5	2.9	2.5
10	0.00105	15	15	600	3.7	3.1	2.6
11	0.00105	15	55	200	2.8	3.4	3.1
12	0.00105	15	55	600	2.9	3.6	3.3
13	0.00105	55	15	200	6.7	5.1	4.8
14	0.00105	55	15	600	7.0	5.4	5.1
15	0.00105	55	55	200	5.2	6.1	6.1
16	0.00105	55	55	600	5.5	6.4	6.5
17	0.00015	35	35	400	25.7	16.9	16.1
18	0.00135	35	35	400	4.4	4.8	4.6
19	0.00075	5	35	400	1.9	1.9	1.7
20	0.00075	65	35	400	9.7	9.3	9.3
21	0.00075	35	5	0	4.4	2.6	2.3
22	0.00075	35	65	800	5.7	6.5	6.4
23	0.00075	35	35	400	7.1	6.9	6.7
24	0.00075	35	35	400	7.8	7.7	7.3
25	0.00075	35	35	400	7.7	7.6	7.2
26	0.00075	35	35	400	7.7	7.6	7.2
27	0.00075	35	35	400	7.7	7.6	7.2
28	0.00075	35	35	400	7.7	7.6	7.2
29	0.00075	35	35	400	7.7	7.6	7.2
30	0.00075	35	35	400	7.7	7.6	7.2
31	0.00075	35	35	400	7.7	7.6	7.2

Table A.9: Mean, 10% and 90% quantiles and interquartile distances of Monte-Carlo-based simulation based on Central Composite Design with four factors (mDoE-2).

Run	$\overline{y}$ [mmol L^{-1}]	**10% quantile** [mmol L^{-1}]	**90% quantile** [mmol L^{-1}]	**Interquantile** [mmol L^{-1}]
1	7.06	4.76	10.14	5.38
2	7.34	4.98	10.41	5.43
3	7.50	5.46	9.68	4.22
4	7.80	5.71	10.03	4.32
5	14.96	9.83	27.98	18.14
6	15.73	10.38	29.20	18.82
7	14.77	11.95	17.56	5.61
8	15.59	12.62	18.54	5.93
9	3.62	2.23	5.94	3.70
10	3.80	2.36	6.20	3.84
11	3.75	2.62	4.98	2.37
12	3.94	2.76	5.24	2.48
13	6.72	4.26	13.33	9.07
14	7.10	4.51	14.07	9.56
15	6.40	5.21	7.57	2.37
16	6.77	5.51	8.02	2.51
17	17.92	15.09	20.84	5.75
18	5.15	4.23	6.14	1.91
19	2.24	1.52	2.85	1.33
20	9.86	8.22	12.09	3.87
21	5.85	1.90	17.70	15.80
22	7.02	5.21	8.87	3.66
23	7.43	6.07	8.77	2.70
24	8.13	6.65	9.57	2.92
25	8.05	6.59	9.49	2.90
26	8.05	6.59	9.49	2.90
27	8.05	6.59	9.49	2.90
28	8.05	6.59	9.49	2.90
29	8.05	6.59	9.49	2.90
30	8.05	6.59	9.49	2.90
31	8.05	6.59	9.49	2.90

Acknowledgement

This thesis was written during my time as a research assistant at the Institute of Bioprocess and Biosystems Engineering at the Hamburg University of Technology. Many people were involved in various ways in its preparation. I would therefore like to express my special thanks to them.

First and foremost, I would like to thank my doctoral supervisor, Prof. Dr.-Ing. Ralf Pörtner, who provided me with suggestions and critical comments throughout the entire period. His confidence in me, his patience and the freedom he gave me contributed to the success of this thesis.

I would like to thank Prof. Dr. Andreas Liese for his interest in my work and for being the second examiner. I also thank Prof. Dr.-Ing. Mirko Skiborowski for being the chair of the board of examiners.

I would also like to thank my current and former colleagues at the institute for the time we have spent together and for their mutual support. In particular, I would like to mention my colleagues Christiane Schaffer, Dr.-Ing. Nikolai Jürgensen, Dr.-Ing. Johannes Möller, Dr.-Ing. Kim Beatrice Kuchemüller, Dr.-Ing. Fabian Freiberger, Dr.-Int. Grit Blume and Dr.-Ing. Ludwig Selder. I would also like to thank Giovanni Sayoga, Leon Klose and Luca Schmidt from the Institute of Technical Biocatalysis.

A special thanks goes to my cooperation partners Prof. Dr.-Ing. Volker Hass, André Moser and Dr. Christian Appl (Hochschule Furtwangen) as well as Dr. Jürgen Kuballa, Dr. Miriam Aßmann and Kristin Hölting (GALAB Laboratories GmbH) for the successful cooperation in the prot P.S.I. project.

The biggest thank goes to my family and friends, who have shaped and supported me throughout my studies and time as a research assistant, and without whom this work would not have been possible. I am deeply grateful to my wife Marlene for her tireless support, understanding, scientific discussions and constant motivation during the preparation of this thesis. I would also like to thank her family for their loving help and support.