SONG, JIN A. Mechanistic Model Development for Multi-Enzymatic Reactions in Lignin Biosynthesis. (Under the direction of Cranos Williams.)

Modeling an enzymatic reaction, one unit of biomass regulation, enables a more comprehensive analysis of biosynthesis. Such modeling enables one to control each reaction rate for the product and, by extension, to use transgenic perturbation strategies. Modeling approaches have traditionally been confined to reactions catalyzed by a single enzyme. Reactions controlled by multiple enzymes, however, comprise a large proportion of various biosynthesis pathways such as lignin. Modeling a multi-enzymatic reaction presents a challenge due to unknown enzyme-enzyme or enzyme-substrate interactions. New modeling approaches are needed to predict unknown interactions among components in multi-enzymatic reactions and to provide appropriate mathematical expressions.

The research in this dissertation presents two modeling approaches which develop mechanistic models for multi-enzymatic reactions. The first is for mechanistic models characterized by detailed inner components and interactions. This approach optimizes the model through analysis and assumptions based on mass action. It also develops the model structures and mathematical equations for multi-enzymatic reactions and provides reproduction and prediction of experimental reaction rates. The second provides simplified mechanistic models representing key relationships among the enzymes associated with the reaction rate. This approach is carried out through computational methodologies, a rule-based algorithm and an evolutionary computation algorithm, which contribute to a general insight about the mechanism of multi-enzyme reactions and also save model development time. This approach can be used as a tool which facilitates control of the early stages of a detailed mechanistic modeling.

These approaches were applied to multi-enzymatic reactions with different characteristics. These are important reactions to shape the final products in lignin biosynthesis. Biological relevance of the modeling results of these applications was assured by comparing them with previous literature and with the results of diverse experiment technologies. These tools will provide helpful insights in many research areas based on the functions of enzyme-regulated biochemical reactions.
Mechanistic Model Development for Multi-Enzymatic Reactions in Lignin Biosynthesis

by

Jin A Song

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Electrical Engineering

Raleigh, North Carolina

2014

APPROVED BY:

Winser Alexander

Huaiyu Dai

Joel Ducoste

Vincent Chiang

Cranos Williams
Chair of Advisory Committee
DEDICATION

To my Heavenly Father
To my parents, Jae-Myeong Song, Pal-Bok Heo,
To my parents-in-laws, Jeong-Ryeol Kim, Deok-Kyun Park,
   To my husband, Jongbeom Park,
   To my lovely daughter, Songee Park.
BIOGRAPHY

Jin A Song was born September, 24, 1978 in Seoul, Korea. In March 1997, she was admitted to Ewha Womans University in Korea to earn her B.S. She received her B.S. and M.S. degrees in Electronics Engineering from Ewha Womans University in 2001 and 2003, respectively. After earning her M.S. degree, she worked as a senior engineer at Magnachip Semiconductor Ltd. from January 2003 to January 2006. In August of 2007 she then moved to North Carolina State University to pursue her Ph.D. degree, where she has conducted her research in Systems Biology.
ACKNOWLEDGEMENTS

Most of all, I am deeply grateful to my God, my Father, who always accompany me and fills all my needs with His goodness. I praise the Load for His grace to me by which I could complete this dissertation.

I would like to express my heartfelt thanks to my advisor, Dr. Cranos Williams, for his guidance and encouragement. His advice and support could not be more for me to finish this study. I wish to express my gratitude to the other dissertation committee members, Dr. Winser Alexander, Dr. Huaiyu Dai, Dr. Vincent Chiang, and Dr. Joel Ducoste, whose knowledge and expertise contributed to the completion of this dissertation. I would also like to thank Dr. Ron Sederoff, Dr. Hsi-Chuan Chen, Dr. Jack Wang, and Chien Yuan, Lin (Kevin) in the National Science Foundation project. I could broaden by sight through the precious time with them. I would specially like to thanks to EnbiSys Lab. including Dr. Maria Angels De Luis Balaguer, Dr. Youngsoo Kim, Megan Matthews, Zohaib Qazi, and Alexlandr Koryachko for all your kind support.

I do not know how to thank my parents, Jae-Myeong Song and Pal-Bok Heo, for their pray and encouragement. I also extend my appreciation to my parents in law, Deok-Kyun Park and Jeong-Ryeol Kim, enough for love and support that allow me to study in the United States. My genuine thanks and deep appreciation is deserved to my husband, Jongbeom Park, for his valuable help and being patient all the time. His sunny personality makes me smile and being comfort whenever I felt exhausted. To my beloved daughter Songee Park, who joined us in my Ph.D period, I would like to express my thanks for unlimited happiness and pleasure.

Lastly, I would like to thank my friends and Duraleigh Presbyterian Church family for their love, kindness, and support.
# TABLE OF CONTENTS

**LIST OF TABLES** .......................................................... vii

**LIST OF FIGURES** .......................................................... viii

**Chapter 1 Introduction** ...................................................... 1
  1.1 Overview .............................................................. 1
  1.2 Motivation ........................................................... 2
  1.3 Contributions of This Work .......................................... 3
    1.3.1 Intellectual Merit .............................................. 3
    1.3.2 Broader Impacts ............................................. 4
  1.4 Organization ......................................................... 5

**Chapter 2 Background** ...................................................... 6
  2.1 Multi-Enzymatic Reactions ........................................... 6
  2.2 Classic Modeling Methods for Enzyme Reactions ..................... 7
    2.2.1 Mass Action Laws in Mechanistic Modeling .................. 7
    2.2.2 Artificial Neural Networks in Empirical Modeling .......... 8
  2.3 Recent Modeling Methods for Biological Systems .................... 8
    2.3.1 Rule-Based Modeling ......................................... 8
    2.3.2 Evolutionary Computation .................................. 9
  2.4 Lingnin Biosynthesis ............................................... 10

**Chapter 3 Mechanistic Modeling Approach based on Model Optimization Process** ...................................................... 12
  3.1 Motivation ........................................................... 12
  3.2 Proposed Modeling Approach ........................................ 14
    3.2.1 Modeling Process Flowchart ................................ 14
    3.2.2 Initial Process ............................................... 15
    3.2.3 Model Optimization Process .................................. 16
    3.2.4 Evaluation Process ......................................... 17
  3.3 Model Development .................................................. 18
    3.3.1 Application 1 : CAld5Hs Enzymatic Reaction with Functional Redundancy 18
    3.3.2 Application 2 : 4CLs Enzymatic Reaction with Enzyme Complex ....... 23
  3.4 Results ....................................................................... 33
    3.4.1 Case 1 : CAld5Hs Enzymatic Reaction with Functional Redundancy ... 33
    3.4.2 Case 2 : 4CLs Enzymatic Reaction with Enzyme Complex ................... 36
    3.4.3 Case 3 : 4CLs Enzymatic Reaction with Enzyme Complex considering Inhibitions by Multiple Substrates .................. 42
  3.5 Discussion ............................................................ 49
    3.5.1 Analysis ......................................................... 49
    3.5.2 Prediction ...................................................... 49
  3.6 Chapter Summary ...................................................... 53
LIST OF TABLES

Table 3.1 Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using 4-coumaric Acid as Substrate .................................................. 39
Table 3.2 Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using Caffeic Acid as Substrate .................................................. 41
Table 3.3 Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using 4-coumaric Acid as a Main Substrate with Inhibition ............. 45
Table 3.4 Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using Caffeic Acid as a Main Substrate with Inhibition ............. 47

Table 4.1 Multi-Enzymatic Reaction Models with the Linear Relationship between the Reaction Rate and the Enzymes .................................................. 63
Table 4.2 Multi-Enzymatic Reaction Models with the Nonlinear Relationship between the Reaction Rate and the Enzymes ............................................. 67
Table 4.3 Operands and Operators in GP for multi-enzymatic reaction modeling with two enzymes ................................................................. 69
Table 4.4 The optimized models of the reaction rates for CAld5H1 and CAld5H2 ........ 75
Table 4.5 The optimized models of the reaction rates for 4CL3 and 4CL5 ................ 79
Table 4.6 Michaelis-Menten Kinetics of 4CL3 and 4CL5 using 4-coumaric Acid as a Substrate ................................................................. 82
Table 4.7 The optimized models of the reaction rates for 4CLs and HCTs ................ 84
LIST OF FIGURES

Figure 2.1  Enzyme-Enzyme Interaction Examples .......................... 7
Figure 2.2  Lignin Biosynthesis .................................................. 10

Figure 3.1  General Flowchart of System Identification ....................... 13
Figure 3.2  Proposed Modeling Process Flowchart for Multi-Enzymatic Reaction Modeling 14
Figure 3.3  Initial Model for CAld5Hs Enzymatic Reactions .................. 19
Figure 3.4  Initial Model Analysis for CAld5Hs Enzymatic Reactions ........... 20
Figure 3.5  Model Structure Development with for CAld5Hs Enzymatic Reactions .... 21
Figure 3.6  Model Optimization Process for CAld5Hs Enzymatic Reactions ........ 22
Figure 3.7  Initial Enzyme Concentrations for 4CLs Enzymatic Reactions ........ 24
Figure 3.8  Initial Model for 4CLs Enzymatic Reactions ........................ 24
Figure 3.9  Initial Model Analysis for 4CLs Enzymatic Reactions (The concentration of 4-coumaric acid as substrate : 28 uM) .......................... 25
Figure 3.10 Model Structure Development for 4CLs Enzymatic Reactions ........ 26
Figure 3.11 Initial Model Structure for 4CLs Enzymatic Reactions .............. 27
Figure 3.12 Model Optimization Process for 4CLs Enzymatic Reactions .......... 32
Figure 3.13 Simulation Results along with the Experimental Rates using Coniferaldehyde as Substrate) .................................................. 34
Figure 3.14 Simulation Results along with the Experimental Rates using Coniferyl alcohol as Substrate) .............................................. 34
Figure 3.15 Model Structure Development for CAld5Hs Enzymatic Reaction with Inhibition 35
Figure 3.16 Simulation Results along with the Experimental Rates using Multiple Substrates with coniferyl alcohol as a Main Substrate .................. 36
Figure 3.17 Mechanistic Description for 4CLs Enzymatic Reaction with Single Substrate 37
Figure 3.18 Model Structure of 4CLs Enzymatic Reaction using 4-coumaric Acid as Substrate ......................................................... 37
Figure 3.19 Simulation Results along with the Experimental Rates using 4-coumaric Acid as Substrate ....................................................... 39
Figure 3.20 Model Structure of 4CLs Enzymatic Reaction using Caffeic Acid as Substrate 40
Figure 3.21 Simulation Results along with the Experimental Rates using Caffeic Acid as Substrate) ....................................................... 42
Figure 3.22 Mechanistic Description for 4CLs Enzymatic Reaction with Multiple Substrates 43
Figure 3.23 Model Structure of 4CLs Enzymatic Reaction using 4-coumaric Acid as a Main Substrate with Inhibition ......................... 44
Figure 3.24 Simulation Results along with the Experimental Rates using 4-coumaric Acid as a Main Substrate with Inhibition ..................... 46
Figure 3.25 Model Structure of 4CLs Enzymatic Reaction using Caffeic Acid as a Main Substrate with Inhibition ................................. 47
Figure 3.26 Simulation Results along with the Experimental Rates using Caffeic Acid as a Main Substrate with Inhibition ....................... 48
Figure 3.27 Model Description for CoA Ligation by 4CL enzymes using Caffeic Acid as a Main Substrate ......................................... 50
Figure 3.28 The Predicted Fraction of the Total Rate Attributed to Each of the Enzyme Entities (Multiple Enzymes and Single Substrate Model: 4-coumaric Acid as Substrate) .................................................. 51
Figure 3.29 The Predicted Fraction of the Total Rate Attributed to Each of the Enzyme Entities (Multiple Enzymes and Multiple Substrate Model: Caffeic Acid as Substrate) .................................................. 52

Figure 4.1 Model Structures for Multi-Enzymatic Reactions - A. Mechanistic Model B. Empirical Model C. Proposed Model (E1, E2: Single Enzyme, Ec: Enzyme Complex, S: Substrate, P: Product) .................................. 56
Figure 4.2 Proposed Modeling Process Flowchart with Rule-based Modeling and Evolutionary Computation ........................................... 58
Figure 4.3 Proposed Rule-based Modeling and Evolutionary Computation for Multi-Enzymatic Reaction Modeling Framework .............. 59
Figure 4.4 The effects of rule-based modeling ..................................... 60
Figure 4.5 Agent-based Evolutionary Computation for Multi-Enzymatic Reactions ...................................................... 61
Figure 4.6 The Optimization Example using Tree Structures of GP ............................................ 68
Figure 4.7 The basic tree structures of the functional operands (i: enzyme index, p: unknown parameters such as $z_1$, $z_2$, m, and n) .......................... 70
Figure 4.8 Numerical analysis (MSE and BIC) in optimization process ........................................ 72
Figure 4.9 Simulation Results of CAld5Hs enzymatic reaction models ........................................ 74
Figure 4.10 Numerical analysis (RMSE and BIC) in optimization process .................................. 76
Figure 4.11 Simulation Results of the reaction with 4-coumaric acid (A) ................................ 80
Figure 4.12 Simulation Results of the reaction with caffeic acid (B) ........................................ 80
Figure 4.13 Simulation Results of the reaction with 4-coumaric acid inhibited by caffeic acid and feruic acid (C) ........................................ 81
Figure 4.14 Simulation Results of the reaction with caffeic acid inhibited by 4-coumaric-acid and feruic acid(D) .................................. 81
Figure 4.15 Numerical analysis (RMSE and BIC) in optimization process .................................. 83
Figure 4.16 Simulation Results of the reaction for 4CL3 and HCT1 (A) .................................. 87
Figure 4.17 Simulation Results of the reaction with 4CL3 and HCT6 (B) ................................ 88
Figure 4.18 Simulation Results of the reaction with 4CL5 and HCT1 (C) ................................ 88
Figure 4.19 Simulation Results of the reaction with 4CL5 and HCT6 (D) ................................ 89
Chapter 1

Introduction

1.1 Overview

A cell is composed of biochemical pathways connecting components, such as DNA, RNA, proteins, and other biomass compounds. To understand the mechanism of a cell, it is beneficial to consider the cell as an interaction-based system rather than as a set of individual components [2]. Metabolic networks, one of biochemical pathways in a cell, is in charge of producing material for cell growth and maintenance and also is composed of the biochemical reactions of metabolism, as well as the regulatory - mostly enzyme - interactions that control these reactions [76]. A system-level mechanistic description of metabolic networks allows not only for the representation of current knowledge but also for in-depth insights and predictions into molecular mechanisms in a cell. For modeling, we need to know the dynamic behavior of the components involved in the reactions, and interactions among these components [46]. Since the completion of the sequencing of the human genome [21], advances in experimental techniques have satisfied the required data more completely than in the past. Nevertheless, many structures and parameters still remain unidentified.

Modeling for a multi-enzymatic reaction, a biochemical reaction influenced by two or more enzymes in metabolic networks, has the same modeling issues as mentioned above. The multi-enzymatic reaction is a key process in the biosynthesis of relevant biochemical compounds involved in plant cells. The interaction among multiple enzymes in similar enzyme families can seriously complicate the modeling of these metabolic pathways. Mathematical models for single enzyme systems, such as Michaelis-Menten kinetics, have been developed and used successfully for more than a century [26]. Michaelis-Menten kinetics forms the rate equation of an enzymatic reaction based on mass action law [53]. Michaelis-Menten kinetics rate equations are not typically used to model multi-enzymatic reactions due to unknown mass action relationships associated with multiple enzymes. Techniques including the empirical modeling approaches
have been suggested to overcome this problem [81] [8]. These algorithms build mathematical models from experimental data. These approaches use learning algorithms like artificial neural network algorithms [81] [8]. The algorithms are able to reproduce and predict the change of reaction rates with respect to changing enzyme amounts, but can lose biological significance easily because of empirical mathematical representation. Thus, there is a need for new modeling strategies, capable of identifying the interaction and plausible mathematical descriptions that can describe multi-enzymatic reactions.

The overall objective of this research is to develop a mechanistic modeling approach that will assess how the interactions of multiple enzymes influence changes in reaction fluxes associated with metabolic pathways. This research proposes two modeling approaches for mechanistic modeling of multi-enzymatic reactions. The first modeling algorithm suggests a mechanistic modeling approach with the model structure optimization process. This method is based on mass action law, one of classical models representing chemical reactions, which yield the mechanistic model comprised of detailed interactions among the involved components. The second modeling algorithm uses computational methods, such as a rule-based algorithm and evolutionary computation, for the identification of the key interactions and components requiring less time than the former process. These proposed modeling frameworks and their developed models provide a new strategy for the understanding of complex biological systems with a lack of information available associated with a detailed mechanism.

1.2 Motivation

The modeling of enzymatic systems for biomass regulation leads to a more comprehensive analysis of biosynthesis. This modeling will make it possible to use transgenic perturbation strategies to find the main pathways and to control reaction rates for the key products. Lignin biosynthesis is a good example of a way that we can ascertain the importance of enzymatic modeling. Lignin biosynthesis is composed of a series of biochemical reactions that must be modeled effectively if a predictive model of lignin is to be created. Most of the reactions in the main pathway in lignin biosynthesis are controlled by multiple enzymes. Good models for reactions catalyzed by a single enzyme are readily available [53]. Suitable models for multiple enzymes remain intensely researched topics [85] [77] [47] [64]. This is due to the unknown interactions among the involved enzymes and their unpredictable influence on the rate of production formation. Regulation by multiple enzymes comprise a large proportion of biomass regulation. However, model complexity, which increases exponentially with the number of involved components in the reaction, and lack of quantitative data, make it difficult to represent these reactions with conventional models. Empirical modeling such as artificial neural networks has been suggested recently as an alternative [81] [8]. Artificial neural network algorithms are regarded as attrac-
tive algorithms for building mathematical models from experimental data for complex biological systems [39]. There is, however, has one drawback: the researcher can easily lose the biological significance of the multi-enzymatic reactions (e.g. the existence and effects of potential enzyme interactions.). We have the need to generate new approaches that can overcome some issues due to the unknown and complicated interactions in multi-enzymatic reactions while retaining the core biological significance. This research suggests a new approach based on the pool of diverse algorithms in engineering to satisfy this need.

1.3 Contributions of This Work

1.3.1 Intellectual Merit

The research presented in this dissertation addresses issues associated with the modeling of multi-enzymatic reactions in metabolic pathways, yielding the following advantages:

1. *Improvement of the mechanistic modeling process for network-level complex biochemical reactions*

   We propose new system identification approaches for mechanistic modeling of multi-enzymatic reactions with undiscovered information about components and interactions. We insert the required processes for resolving the issues by unknown model structures in a general system identification flowchart. Computational algorithms substituting a time consuming optimization process result in the acceleration of the model development flow.

2. *Identify possible components and their interactions in multi-enzymatic reactions*

   The proposed algorithms provide methodologies for assessing and predicting the interactions among components involved in multi-enzymatic reactions, which identify and suggest new potential relationships, including potential enzyme-enzyme or enzyme-substrate interactions, and new components, such as enzyme complexes with a specific ratio. Modeling interactions between 4CL enzymes contributes to a better understanding of the lignin biosynthesis process. These reactions occur at a key branching point in the pathways.

3. *Identify the important influence of the identified components on the overall reaction rate*

   These modeling approaches can describes the change of the overall biochemical reaction rate along with the perturbation of the involved enzyme components, making it possible to assess and predict how control mechanisms of multiple enzymes leads to the linearity or nonlinearity of the overall reaction rate. This modeling reveals the contribution of interactions between 4CL enzymes to CoA ligation, including inhibition and activation. This is difficult or impossible in the case of most enzymes.
4. **Develop the model structures and mathematical equations representing the mechanism of multi-enzymatic reactions**

The developed model structures, representing the identified components and interactions, and the developed mathematical equations, representing their functions, are the outputs of the proposed modeling approaches. These can be useful expressions to explain the functions of general enzyme-regulated biochemical reactions.

### 1.3.2 Broader Impacts

**Application**

It is not hard to find research in plant biology that involve multi-enzyme interactions. These include examples such as the lignin pathway and the Calvin cycle in fatty acid metabolism [33]. The proposed model will lead to a better understanding of how multi-enzyme interactions can provide another level of control in modulating the reaction flux in plant metabolic pathways. Other applications include drug-drug interactions, which are very important in pharmacology research. The research for drug-drug interaction involving the complex characteristics of multi-enzymatic reactions is challenging and there is a need for suitable model-based approaches [78]. In bio-fuel research, the enzymatic fuel cell is a very attractive recent technology. The catalyzed power of single or multiple enzyme systems is used for converting bio-fuels to electrical currents [3]. Food science is also interested in multi-enzyme reactions which are an important part of food processing. To wit, papers related to multi-enzymatic reaction modeling have been published in food science [81][8]. The developed model in our research will be an attractive proposition for the requirements of diverse research issues in medical and biological areas.

**Integrative Research**

Remarkable research in the biological sciences enables us to collect data in high-throughput experiments at the genomic, proteomic, and metabolic scales [22]. These data demand the development of computational approaches to identify the systematic relationships embedded therein. The current proposed research leads to collaborations among biologists, mathematicians, and engineers. Diverse algorithms such as the one proposed in this manuscript support mathematical and computational analysis for biological systems and enhance the systematic understanding of these processes.

Multi-enzymatic reaction modeling is an example of true interdisciplinary research. The research is also based on diverse algorithms related to mathematical modeling, associated to engineering areas. Research such as this provide opportunities for biologists and engineers to work closely together to solve complex biological problems. This allows biologist to get information to design experiments effectively and to predict experimental results, which might
accelerate the advancement of biological systems research. In addition, this gives engineers the opportunity to develop new algorithms and to improve developed algorithms for stochastic approaches, parameter estimation and optimal control, which would have potential impacts on modeling complex biological systems at multiple scales and multiple levels.

1.4 Organization

This dissertation describes the proposed modeling approaches and their results using diverse applications. Chapter 2 presents the background of this research, which include the characteristics of multi-enzymatic reactions and classical modeling methods for biochemical reaction descriptions. In addition, state-of-the-art computational algorithms from the literature related to the proposed modeling algorithms are reviewed. Finally, diverse multi-enzymatic reactions in lignin biosynthesis are introduced as applications of my modeling approaches. Chapter 3 presents the mechanistic modeling approach based on model optimization process for multi-enzymatic reactions. The detailed processes of the modeling approach are explained. The realistic modeling process and the results of the several sub-cases in two main applications, show the representation and prediction ability of the proposed modeling approach. Chapter 4 presents the mechanistic modeling approach based on rule-based modeling and evolutionary computation for multi-enzymatic reactions. The processes related to rule-based modeling and it’s ability to decipher topological interactions in multi-enzymatic reactions, are explained. The optimization processes for evolutionary computations are described. The developed model structures, mathematical equations, and their simulation results, using several sub-cases in three main applications, demonstrate the accuracy and the convenience of the proposed modeling approach. Chapter 5 summarizes the work presented in this dissertation and describes future work.
Chapter 2

Background

2.1 Multi-Enzymatic Reactions

Multi-enzymatic reactions are metabolic reactions that involve two or more enzymes [1]. The relationships between enzymes in multi-enzymatic reactions decide the core characteristics of diverse biosynthetic pathways such as branched chain amino acid biosynthetic pathways in *E. coli* [85] and pyruvate dehydrogenase and alpha-ketoglutarate dehydrogenase in *E. coli* [77]. Michaelis-Menten kinetics was developed and used successfully over time for single enzymatic reactions. However, it is hard to decipher mass action relationships associated with multi-enzymatic processes. This is because it is very difficult to analyze and model the unknown and complicated interactions such as enzyme-enzyme interactions or enzyme-substrate interactions in multiple enzymatic reactions based on physical or chemical fundamentals.

Modeling for multi-enzymatic reactions requires a new point of view to resolve the main issues. Enzyme relationships may be roughly classified for this research, as shown in Fig. 2.1. When both enzymes are coexpressed in the reaction, the enzyme activities could be additive, suggesting functional redundancy and independence of the involved enzymes (Fig. 2.1A) [80], or the enzyme activity could be under the control of only one of the involved enzymes, suggesting functional specificity and independence of these enzymes (Fig. 2.1B)[58]. Changes in enzyme concentrations in the cases in Fig. 2.1A and Fig. 2.1B will have linear impacts on the total reaction rate regardless of any type of substrate. Alternatively, the involved enzymes could make an enzyme complex. The enzyme complex could be composed of the simple ratio 1:1 (Fig. 2.1C) [56] or the specific ratio $m : n$ (Fig. 2.1D) [15]. In these cases, changes in enzyme concentrations would have a influence on the total reaction rate. More complicated multiple enzymatic reactions can be represented as a compound of the relationships shown in Fig. 2.1.
### 2.2 Classic Modeling Methods for Enzyme Reactions

Models of biological systems can be generally divided into mechanistic models and empirical models. Enzymatic reaction models have, for the most part, been based on mass action, which is a representative mechanistic model for biosynthesis. In the single enzyme case, models based on mass action have been a suitable tool for an extended period of time. In the multiple enzyme case, it is hard to apply mass action analysis due to complicated and unknown interactions between enzymes. In this section, we will consider issues associated with multi-enzymatic reaction modeling based on mass action laws and we will present empirical modeling techniques, such as Artificial Neural Networks (ANN), from the literature that endeavor to address these issues.

#### 2.2.1 Mass Action Laws in Mechanistic Modeling

Elementary reactions such as M-M, are derived from mass action laws. It is not easy to define a “two or more enzymatic reaction” based on mass action law because it is difficult to get accurate enzyme interaction information from experimental data. The modeling process based on mass action laws for multi-enzymatic reactions must assume relationships among enzymes and derive the rate equation based on mass action laws supporting the relationship. This method can discern the biological meaning of the regulation and reaction, however it requires significant knowledge of the underlying unknown mechanism in multi-enzymatic reactions and significant development time. It is difficult to apply the method to highly complex multi-enzymatic reactions due to exponential increase in complexity when the number of components increase. For
these reasons, we are interested in other approaches that are more attractive.

2.2.2 Artificial Neural Networks in Empirical Modeling

An Artificial Neural Network (ANN) is a mathematical model or computational model inspired by biological neural networks [39]. An ANN is an adaptive system that learns its structure based on training data sets. This tool can identify arbitrary nonlinear statistical models from experimental data, even without first principles knowledge. The characteristics of ANNs are attractive for modeling complex biological systems, which are difficult to model mechanistically and have a number of unknown parameters. They have been used for representing complex relationships of multi-enzymatic reactions between input and output [81] [8]. This approach, however, has a significant downside because it can easily generate an empirical model that contains arbitrary connections that are not biologically significant. For this reason, we propose to develop new approaches for modeling multi-enzymatic reactions.

2.3 Recent Modeling Methods for Biological Systems

2.3.1 Rule-Based Modeling

The rule-based algorithm is an attractive approach for representing complex interactions in biochemical networks. The algorithm is a modeling approach, which builds the model from the individual components and biologically relevant rules or interactions. The rules for generating the patterns are based on the discovered characteristics of the biological systems, e.g. the kinetics of biochemical reactions [19]. One advantage of rule-based modeling is that it is more concise, more comprehensive and more easily extended than modeling with many chemical species and reactions [6]. In addition, this type of approaches can be automated through an optimization process. The generalized reaction, called rules, cannot only classify and simplify the components and the interactions but also suggests potentially new components and interactions required to model the system. It is also very useful for representing protein-protein interactions such as the enzyme binding and enzymatic activities [40], because the rules in this algorithm are able to express possible interactions and transformations among different domains of molecules in biochemical networks [28]. Rule-based modeling has been successfully used for modeling diverse interactions between ligands and receptors in signaling pathways, such as signaling by the epidermal growth factor receptor (EGFR) [11] [86]. The advantages have yielded diverse tools for rule-based modeling such as BioNetGen [10] and Kappa [20], as well as visualization tools [19] [72]. The results of modeling using the rule-based algorithm and related tools show that this new approach is a good alternative for resolving issues arising from the complexity of modeling based on mass action law [23].
2.3.2 Evolutionary Computation

Evolutionary Computation (EC) is a computational optimization approach that is inspired by the biological mechanisms of evolution and adaptation [88]. EC targets the development of more robust and efficient systems, rather than overly accurate and over-trained systems, for solving complex nonlinear problems [25]. A variety of complex nonlinear systems have used EC for diverse issues, such as optimization, modeling and simulation [32] [61] [43] [87]. The basic concept idea of EC is survival of the fittest. Search operators such as selection, algorithmic recombination and algorithmic mutation are performed iteratively until an appropriate solution is identified. Algorithmic recombination exchanges information from candidate solutions in the hope that combining two good solutions will lead to a better solution. Algorithmic mutation perturbs solutions to escape local minima in the optimization. Both operators are in charge of generating new possible solutions in the algorithm. Selection is used for testing which solutions are fitter than others. The most active and productive area in EC is that of numerical and combinatorial optimization [25]. The concept of EC has been applied to explain complex biological systems, such as the estimation of unknown parameters in nonlinear biochemical dynamic model and analysis of the metabolome, and has been proved as a suitable algorithm for solving biological modeling issues [7] [55] [34]. The strengths of this algorithm will help to predict unknown relationships among enzymes.

Significantly, agent-based evolutionary computation approaches, one type of the EC, are emerging as a new paradigm to efficiently solve complex problems [65] [49]. The approach combines an agent-based modeling with evolutionary algorithms. Agent-Based Modeling (ABM) is a new approach for deriving models based on identified agents and interactions among these agents [12]. The agent-based model is a computational model for simulating the actions and interactions of autonomous agents from the point of view of assessing their effects on the system [60]. The agents and the rules governing how the agents interact are first fed into the approach, with consideration of the limitations due to environment. This approach then learns the most suitable structure and parameters of the model through implementation processes and optimization methods. The developed models based on these rules and conditions have the ability to represent non-linear characteristics well, while preserving the relevance associate with identified interactions [12]. These advantages can have an effect producing effective on biochemical pathways models, which need to represent the behavior of complex nonlinear systems while maintaining the core biological meaning [60]. This agent-based modeling approach has successfully led to modeling and predicting of regulatory events such as activation and inhibition of NF-kB pathways [60].
2.4 Lignin Biosynthesis

Lignin, a phenolic structural polymer of plants, is essential for water transport, mechanical support and protection against biotic and abiotic stresses [27]. Lignin is also a major barrier to the process of the production of paper or for the biomass conversation to biofuel [17]. Further understanding of the lignin biosynthesis can improve wood processing and result in an increase in the quality of biomass. Lignin biosynthesis is composed of a series of multiple enzyme interactions that must be effectively modeled if a predictive model of lignin is to be created [80] [69] [16].

![Figure 2.2: Lignin Biosynthesis](image)

Fig. 2.2 represents the lignin biosynthesis pathways. Highlighted are three examples of multi-enzymatic reactions, coniferaldehyde 5-hydroxylase (CAld5H) enzymatic reactions, 4-Coumaric acid:coenzyme A ligase (4CL) enzymatic reactions, and the enzymatic reactions of one of 4CL enzymes and one of hydroxycinnamoyl CoA:shikimate/quinate hydroxycinnamoyl transferase (HCT) enzymes. The reactions catalyzed by these three enzyme families are key to influenc-
ing the concentration of the main units, S and G, that compose lignin. Therefore, it is very important to understand and analyze the mechanism of these reactions for lignin biosynthesis regulation. The multi-enzymatic reactions, mentioned above as the examples, have different characteristics respectively. First, coniferaldehyde 5-hydroxylase (CAld5H) initiates the syringyl subunits (S) in lignin biosynthesis [30]. CAld5H enzymes (CAld5H1 and CAld5H2) are associated with biosynthesis regulation in *Populus trichocarpa*. Conifealdehyde and coniferyl alcohol are important substrates involved in a main pathway in monolignol biosynthesis [57]. In addition, the activity of coniferyl alcohol is greatly inhibited when the two substrates are present in the reaction where inhibition mode of CAld5H1 is uncompetitive and inhibition of CAld5H2 is a combination of competitive and uncompetitive [80].

Second, 4CL regulates CoA ligation fluxes, which affect the lignin main units in secondary cell walls [42]. The role of 4CL enzymes in controlling lignin biosynthesis has been an active research topic [70]. Recently, it was revealed that two monolignol 4CL enzymes (Ptr4CL3 and Ptr4CL5) were mainly associated with the biosynthesis regulation in *Populus trichocarpa* [16]. Ptr4CL3 and Ptr4CL5 mainly catalyze the formation of CoA thioesters of 4-coumaric acid and caffein acid of several hydroxycinnamic acids. In addition, both enzymes are influenced by competitive and uncompetitive inhibitions as well as self-inhibition [16]. Third, HCT enzymes do not have any impact on the formation of CoA thioesters of 4-coumaric acid by themselves, but, when HCT enzymes are coexpressed with 4CL enzymes, HCT enzymes make changes in the original 4CL enzymatic reactions (under experimental observation).

CAld5H enzymatic reactions and 4CL enzymatic reactions are used as the applications of the proposed modeling algorithms in chapter 3 and chapter 4. 4CLs-HCTs enzymatic reactions are used as an application to identify interactions between enzymes that have yet to be experimentally verified (Chap 4 only).
Chapter 3

Mechanistic Modeling Approach based on Model Optimization Process

3.1 Motivation

The relationships among components, such as enzymes and substrates, involved in multi-enzymatic reactions affect the main characteristics of various biosynthetic pathways [85] [77]. Mechanistic models for multi-enzymatic reactions make it possible to quantify the relationships between enzymes and reactions and regulate productive reactions. It is hard to apply physical and chemical principles such as mass action law directly to modeling because of unknown interactions in multi-enzymatic reactions, as mentioned in chapter 2. We propose to use modeling methods for overcoming these barriers to develop a mechanistic model of the reaction.

The field of system identification is the construction of mathematical models of dynamic systems from experimental data. Fig. 3.1 describes the general flowchart of system identification [73]. In this process, we need to collect the observed input-output data for model development through a well-designed experiment. The appropriate model structure to represent the system is determined and unknown parameters in the structure are estimated. Based on the measured data, the developed model is validated. The procedure is iterated until the developed model reaches a satisfactory result. Although the process for system identification is attractive for modeling complex systems, we have some concerns about applying this flowchart to mechanistic modeling for multi-enzymatic reactions directly. First, model structure may not be fixed because we have no information about interactions among components. Second, many unknown parameters in model structures may not be properly dealt with in these mathematical equations. Third, the developed model may not satisfy the requirement for biological relevance. We
A new modeling method which takes into account these issues for multi-enzymatic reaction modeling.

This chapter is outlined as follows. Section 3.2 proposes the modeling process for multi-enzymatic reactions. The proposed process considers the modeling issues explained and suggests the appropriate problem-solving methods such as customized experimental design for data characteristics, model optimization process for unknown system structures and parameters, and model validation for establishing biological relevance. Section 3.3 presents how to develop the model, applying the proposed modeling process to three multi-enzymatic reactions. These applications have different interaction types between the enzymes involved with each i.e., functional redundancy, enzyme complex and enzyme complex with inhibition by multiple substrates. Section 3.4 shows the plausible model structures and mathematical equations of the developed models respectively. In addition, simulation results represent a good-fit of the developed models to experimental data. Section 3.5 discusses the ability of the developed models to recreate the experimental data and calculates fluxes associated with the individual drivers of product production. Section 3.6 summarizes the results obtained from the approach developed in this chapter.
3.2 Proposed Modeling Approach

3.2.1 Modeling Process Flowchart

This section shows the proposed modeling process for building a mechanistic model of multi-enzymatic reactions, considering the issues mentioned in the previous section. Our key focus is the unknown model structure resulting from lack of information about interactions among components. For solving this problem, we need a specific experimental design for capturing the influence that different concentration of enzymes have on product rate formation. Based on the collected data, model optimization process is performed to look for a potential model structure and to estimate unknown parameters. The process is comprised of analyzing evaluation results of the previous developed model, suggesting the modified model structure with newly assumed components and interactions and simplifying parameters. The model developed by the optimizing process passes tests for validating biological relevance. Fig. 3.2 shows the proposed
modeling process flowchart. From next section, we discuss each process in the flowchart in detail.

### 3.2.2 Initial Process

The proposed modeling process begins by collecting the data supporting the modeling and suggesting the initial model.

**Design Experiment** Multi-enzymatic reaction modeling requires the design of specific experimental scenarios for predicting unknown enzyme-enzyme and enzyme-substrate interactions. Single-enzymatic reaction modeling focuses only on the relationship between the amount of substrate and the reaction rate based on the assumption that the amount of the involved enzyme is constant. On the other hand, multi-enzymatic reaction modeling also reflects the reaction rate change due to change in enzyme concentrations and proportion. According to the characteristics of each multi-enzymatic reaction applications, the specific experimental designs would be used.

**Collect Experimental Data** The designed experiments are performed. The input data, the concentrations of substrates and the amounts of enzymes, and the related output data, the reaction rate, are collected in every experiment. The collected data may then be rearranged for the effective analysis to reveal interactions among the involved components (enzymes, substrates).

**First Principle Knowledge** The first principle knowledge can be described as the available prior knowledge used in the modeling. In multi-enzymatic reaction modeling, the main information of the first principle knowledge is related to single-enzymatic reaction. Initially, the information includes first principles explaining enzymatic reactions, such as mass action law [36] [37] [38], enzyme conservation law [50], chemical equilibrium [4] and quasi-steady-state approximation [13], and well-known models developed from the first principles such as Michaelis-Menten kinetics. Next, first principle knowledge also contains experimental data and kinetic parameter values acquired from each enzymatic reaction experiment under the same conditions. Diverse information from previous literature can be included in first principle knowledge to improve the accuracy of modeling.

**Develop Initial Model** Multi-enzymatic reaction modeling is started with the initial model suggested by conceivable assumptions. The first assumption for building the initial model is that each involved enzyme works independently, without any interaction among the enzymes. The second assumption is that each enzymatic reaction can be expressed with Michaelis-Menten kinetics. According to the second assumption, the equation representing each enzymatic reaction would be expressed by Michaelis-Menten kinetics where the kinetic parameters in the equation
would be obtained by *in-vitro* kinetics. The developed initial model will provide clues about interactions among components through the comparative analysis of the model simulation and experimental data.

### 3.2.3 Model Optimization Process

The main process of modeling multi-enzymatic reactions is to optimize the model structure and parameters. The process would run iteratively until reaching the goal to develop a plausible mechanistic model that fits the data. The process is comprised of these steps as shown below.

**For Structure Optimization**

**Analysis**  We analyze the potential characteristics of multi-enzymatic reactions based on evaluation results by comparing the initial model or the developing model simulation and experimental data. The comparison would highlight discrepancies due to unknown interactions between components. We would be able to predict the potential interactions, such as functional redundancy or some complex nonlinear effects, through the analysis.

**Assumption**  Some hypotheses for supporting the potential characteristics derived from the analysis are needed. In this step, new components or interactions could be suggested or existing components could be modified to represent the effects of the assumed interactions. The proposed factors are built based on first principle knowledge and are reflected in model modification steps.

**Model Structure Modification**  The model structure reflecting the proposed components and interactions would be built. In this step, it is important that the effects of the proposed factors on the previous model are well represented in the new model structure.

**For Parameter Optimization**

**Mathematical Equation**  The mathematical equation would be derived based on the proposed structure. Reasonable physical and chemical principles are given for the equation development from first principle knowledge. Specifically, the equation of multi-enzymatic reaction models would be derived with the principles used for Michaelis-Menten kinetics. The main components involved in the experiments, such as enzymes and substrates, would be the input variables and the rate of product formation would be the output variable in the equation.
Diverse parameters are used for representing the relationships among the variables. In particular, unknown parameters associated with the newly involved components and their interaction should be considered.

**Parameter Simplification** Parameter simplification is the process reducing the number of unknown parameters. As new components and interactions are involved in the model, many unknown parameters could be introduced into the mathematical equation. In addition, some unknown parameters could work independently, but the other unknown parameters could have an interrelationship. These issues from unknown parameters could increase the complexity of the equation and decrease the accuracy of the parameter estimation. We resolve these issues using two approaches. We first fixs the values of some unknown parameters with reasonable assumptions and diverse information in first principle knowledge. We then integrate unknown parameters associated with each other into one unknown parameter in the equation derivation. Parameter simplification would prevent the over-fitting of the developed model and simplify the optimization of the model [35].

**Parameter Estimation** The unknown parameters in the developed equation are inferred from the experimental data. Parameter estimation considers the values of unknown parameters in minimizing the objective function. The objective function in the optimization process indicates how we define the suitability of the model. The objective function is targeted toward representing the experiment data as closely as generally possible, for which the distance between the expression data and the simulation results is minimized. One common objective that functions to assess the goodness-fit is the root mean square error (RMSE). Parameter estimation uses the appropriate optimization methods of diverse global or local optimization algorithms. The mean values of the results would be used in the final developed equation.

### 3.2.4 Evaluation Process

It is necessary that biological system modeling should be evaluated not only in terms of goodness-of-fit but also for biological meaning, unlike artificial system modeling. First, the goodness-of-fit of the developed model for multi-enzymatic reactions could be assessed visually and numerically between model simulation results and experimental data [66]. Second, biological relevance of the developed model would be validated using diverse advanced technology, such as chemical crosslinking [59] and mass spectrometry [74].

**Model Evaluation** The goodness-of-fit of the developed model could be evaluated by two methods [66]. One is the visual comparison of similarities and differences between model simulation results and experimental data. Visual display helps to find out the characteristics of
multi-enzymatic reaction instinctively. In addition, visual presentation is useful to predict the rough degree of fit and to prevent the over-fitting. The other is numerical analysis measuring how well experimental data are replicated by the developed model. The quantitative measurement makes it possible to assess small differences between the model prediction and the measured data. The results of numerical analysis are used as an indicator representing the accuracy of the fits of model to data. Based on both evaluation methods, the nonlinearity and the accuracy of the multi-enzymatic reaction model would be evaluated.

**Biological Relevance Validation** Although we have the goodness-of-fit of the model mathematically, the model could be in conflict with biological importance. Therefore, we need to validate whether the results from the developed model are compatible with first principles, such as physical and chemical theories. We can check whether the diverse hypotheses suggested in the modeling process actually occur using experimental technologies for verifying interactions between enzymes such as chemical crosslinking [59], Biomolecular Fluorescence Complementation (BiFC) [44], and Mass spectrometry (MS) [74].

### 3.3 Model Development

We applied the proposed modeling approach to three multi-enzymatic reactions, with different interaction types among the involved components with each other, in lignin biosynthesis (Fig. 2.2). The first application is the metabolic reaction catalyzed by CAld5H enzymes with functional redundancy. The second application is the metabolic reactions catalyzed by 4CL enzymes with only enzyme complex. In the final model development, we consider not only enzyme complex but also inhibitions by multiple substrates except for the main productive reaction.

#### 3.3.1 Application 1 : CAld5Hs Enzymatic Reaction with Functional Redundancy

We analyze the function of both CAld5H1 and CAld5H2 enzymes to evaluate the extent of specificity or function redundancy. CAld5H1 and CAld5H2 are expressed specifically and abundantly in xylem [70], but it had not been determined whether these two enzymes have specific or redundant biochemical functions. We suggest the potential interaction, between CAld5H1 and CAld5H2, and the mathematical expressions using the proposed modeling flowchart and carry out a metabolic flux simulation.

**Experimental Design and Data Collection** The proposed multiple enzymatic reaction modeling for CAld5H1 and CAld5H2 requires the total reaction rate, when both enzymes are
coexpressed, and kinetic information, when two enzymes work independently, under same experimental conditions. If the total reaction rate depends on one enzyme activity of both enzymes, we will be able to consider the function of both CAld5H1 and CAld5H2 as functional specificity. Otherwise, if the total reaction rate approaches the sum of both enzyme activities, we will be able to regard an operation as functional redundancy. First, reaction rates are determined using the six concentrations of the substrates, such as coniferaldehyde and coniferyl alcohol, and the initial concentration amounts of CAld5H1 and CAld5H2 enzymes. Second, Michaelis-Menten kinetics and inhibition kinetics of CAld5H1 and CAld5H2 using coniferaldehyde and coniferyl alcohol as a substrate are collected from the literature [80]. The measured reaction rates would be used as the goal which our developed model should represent and be displayed in simulation results graphs presented in Section 3.4.

**Initial Model Analysis**  The initial model for recognizing the extent of specificity or functional redundancy in the enzyme activity of CAld5H1 and CAld5H2 would be suggested. The model assumes that one specific enzyme of two enzymes only works for the reaction in Fig. 3.3. Equation (3.1) represents the total rate of product formation.

\[
v_{\text{initial}} = v_{\text{CAld5H1}} \quad \text{or} \quad v_{\text{CAld5H2}},
\]

\[v_{\text{CAld5H1}} = \frac{k_{\text{cat}1}[E1][S]}{k_{M1} + [S]}, \quad (3.2)\]

**Figure 3.3: Initial Model for CAld5Hs Enzymatic Reactions**

\(v_{\text{CAld5H1}}\) and \(v_{\text{CAld5H2}}\) are derived from basic Michaelis-Menten kinetics [53] and use kinetic information from the experiment in the literature [80]. We compare this baseline equation to experimental measurements to evaluate specific situations where the actual product rate deviates from one specific enzyme. Fig. 3.4A measured the rate of product formation using coniferaldehyde as substrate, \(S\), when the amounts of both enzymes are \(1nM\). The equation shown in (3.1) predicts a total rate \((v_{\text{tot}})\) through Michaelis-Menten kinetics with kinetic constants limited to CAld5H1\((E1)\):

\[v_{\text{CAld5H1}} = \frac{k_{\text{cat}1}[E1][S]}{k_{M1} + [S]}, \quad (3.2)\]
where $k_{cat1}$ and $k_{M1}$, which are the known kinetic constants for CAld5H1. The difference between experimental data (the dots in Fig. 3.4A) and the initial predicted values (the dashed line in Fig. 3.4A) is displayed numerically: Root Mean Squared Error (RMSE) of 0.7304 and R-squared of 31.5089. Otherwise, the graph in Fig. 3.4B predicts a total rate ($v_{tot}$) with kinetic constants limited to CAld5H2($k_{cat2}$):

$$v_{CAld5H2} = \frac{k_{cat2}[E1][S]}{k_{M2} + [S]},$$

(3.3)

where $k_{cat2}$ and $k_{M2}$, the known kinetic constants, for CAld5H2. Associated with the difference between actual experimental data (the dots in Fig. 3.4B) and the initial predicted values (the dashed line in Fig. 3.4B), the numerical analysis results for the initial model are RMSE of 0.8507 and R-squared of 22.3591. In both graphs in Figure 3.4, we can recognize that the experimental rates are significantly higher than the predicted values by the initial assumption about the enzyme activity function. RMSE and R-squared values are also insufficiently satisfied. Through the initial model analysis, other assumptions, supporting the greater reaction rate values than expected, were considered in the next process.

Figure 3.4: Initial Model Analysis for CAld5Hs Enzymatic Reactions

Model Structure Development with Assumption  We recognized that experimental data have values greater than the simulation results of the initial model, which assumes that the re-
action is catalyzed by one specific enzyme of CAld5H1 and CAld5H2, in Fig. 3.4. It is inferred that the functional analysis of both enzymes would be functional redundancy. Based on this inference, we assume that both enzymes would have in independent impact on product formation. This assumption is presented as a block diagram (Fig. 3.5) to describe the functional redundancy on the total reaction rate associated with the related substrate. The total reaction rate would be a function of the activity of both CAld5H1 and CAld5H2 enzymes.

**Mathematical Equation Development**  We propose mathematical descriptions of the total reaction rate using coniferaldehyde or coniferyl alcohol as substrate. The rate of product formation considers the effects of CAld5H1($E_1$) and CAld5H2($E_2$) as:

$$ v = v_{CAld5H1} + v_{CAld5H2} = \frac{k_{cat1}[E_1][S]}{k_{M1} + [S]} + \frac{k_{cat2}[E_2][S]}{k_{M2} + [S]}, \quad (3.4) $$

where kinetic constants are same as (3.2) and (3.3). In consideration of functional redundancy, the model equation is modified depending on the type and the number of substrates in the experimental conditions respectively.

**Model Optimization Process**  In the model optimization process, structure development and mathematical equation derivation would be executed iteratively, until the results of the modeling are optimized. Fig. 3.6 briefly shows the model optimization process for CAld5Hs multi-enzymatic reaction modeling. The left side of Fig. 3.6 represents the assumptions applied to the unknown relationship between the involved enzymes and the right side represents the simulation results of the developed model based on the assumptions. The first row shows the initial model analysis considering one specific enzyme activity (CAld5H1). As we expected, the simulation results cannot reach the actual experimental data. The model in the second column

---

Figure 3.5: Model Structure Development with for CAld5Hs Enzymatic Reactions
considering other specific enzyme activity (CAld5H2) cannot express the experimental data, either. The next modified model considers functional redundancy of CAld5H1 and CAld5H2 based on the above analysis. Finally, the third modified model is a better-fit model. The models considering enzyme complex formation between the involved enzymes did not bring significant improvement to the numerical analysis, although the complexity of the model increases more than the model considering functional redundancy.

**Evaluation** The final optimized model for CAld5Hs enzymatic reaction is evaluated against goodness-of-fit and biological relevance. For goodness-of-fit, visual comparison and numerical analysis are used. Visual display evaluation considers how well the developed model can present the characteristics of the multi-enzymatic system, which resembles the graph curve of Michaelis-
Menten kinetics. RMSE is used as a numerical indicator of the accuracy of the fits of model to data, which is represented as

\[ RMSE = \sqrt{E((v - \hat{v})^2)} = \sqrt{\frac{\sum_{i=1}^{n} v_i - \hat{v}_i}{n}} \]  

(3.5)

where \( v \) represents a set of the measured reaction rates in the experiment, \( \hat{v} \) is a set of the associated model simulation values, and \( n \) is the number of data points. Biological relevance of this developed models is verified by analysis in Wang et al. (2013) [80]

3.3.2 Application 2: 4CLs Enzymatic Reaction with Enzyme Complex

We develop the kinetic model of two 4CL isoforms 4CL3 and 4CL5 involved in lignin biosynthesis during wood formation. 4CL (EC 6.2.1.12) catalyzes the formation of CoA thioesters of several cinnamic acids [16]. In our modeling, two cinnamic acids, 4-coumaric acid and caffeic acid highlighted in gray in Fig. 2.2, are considered as the main substrates. We discovered that the mixture of two isoforms did not lead to the sum of each enzymatic reaction in our previous research [16], which informs us of the potential for the interaction between two enzymes. The models for 4CLs enzymatic reactions, considering enzyme-enzyme interactions, was developed using the proposed modeling flow.

**Experimental Design** Specific experimental scenarios are needed for inferring the characteristics of unknown enzyme-enzyme interactions in 4CLs enzymatic reactions. By focusing on enzyme-enzyme interaction, we want to observe the reaction rate change due to the change in the proportion of the involved enzymes when the substrate concentration is held constant. This modeling suggests the two experimental scenarios shown in Fig. 3.7. The first scenario measures the rate of product formation when the 4CL3 molar concentration is held constant at 40nM and the 4CL5 molar concentration is gradually increased from 0nM to 40nM, in increments of 10nM. The reciprocal scenario measures the rate of product formation when 4CL5 is held constant at 40nM and 4CL3 is increased from 0nM to 40nM. The experimental data in these two scenarios would be used to identify the influence of enzyme-enzyme interactions by comparing and analyzing with the simulated model results.

**Data Collection** The data needed in the modeling framework can be acquired from not only the designed experiment but also first principle knowledge related to single enzymatic reactions. First, reaction rates would be determined using the six concentrations of the substrates, such as 4-coumaric and caffeic acids, and the initial amounts of enzymes changed following the experimental scenarios. Secondly, kinetic data and metabolic concentrations of each enzymatic
reaction catalyzed by 4CL3 and 4CL5 would be collected from the previous literature [16]. The Michaelis-Menten kinetics was determined using the same substrates and optimal conditions as those in the experiment. The measured reaction rates were used as the goal which our developed model should represent and be displayed in simulation results graphs in Section 3.4.

Initial Model Analysis  The initial model for recognizing the presence and absence of enzyme-enzyme interaction was suggested based on the assumption that each enzyme works independently as shown in Fig. 3.8. Equation (3.6) represents the total rate of product formation.

\[ v_{\text{initial}} = v_{4\text{CL3}} + v_{4\text{CL5}}, \]  

(3.6)

\(v_{4\text{CL3}}\) and \(v_{4\text{CL5}}\) are derived from basic Michaelis-Menten kinetics [53] and use kinetic information from experiments in previous literature [16]. We compare this baseline equation to experimental measurements to evaluate specific situations where the actual product rate deviates from enzyme independence under two experimental scenarios.
Fig. 3.9A measured the rate of product formation when the 4CL3 molar concentration was held constant at 40nM while the 4CL5 molar concentration was gradually increased from 0nM to 40nM, in increments of 10nM. Fig. 3.9B measured the rate of product formation when 4CL5 was held constant at 40nM and 4CL3 was increased from 0nM to 40nM. The substrate concentration was held constant. Equation (3.6) predicts a total rate ($v_{\text{tot}}$) that increases linearly with increasing total enzyme concentration. This linearity is expected because the individual rates for both 4CL3 and 4CL5 was linear and proportional to the total enzyme concentration when the substrate concentration was in excess and constant.

The expected baseline and experimental rates are shown for 4-coumaric acid as a substrate at 28 M in Fig. 3.9. The same experiments were conducted at concentrations of 21, 38, 67, and 90 M with 4-coumaric acid and a similar set was done with caffeic acid (in Section 3.4). All results confirm a nonlinear deviation of the experimental rates from the baseline rates. In Fig. 3.9A, for a 4CL3 concentration of 40nM and a 4CL5 concentration of 10nM (ratio 4:1), the rate drops significantly below baseline rates followed by a rise for both the 40nM: 20nM and 40nM: 30nM (4CL3: 4CL5 total concentration) conditions. Despite the subtle rise, the measured rates for the 40nM: 10nM, 40nM: 20nM, and 40nM: 30nM experiments were primarily below the baseline rates. The experimental rates for the 40nM: 40nM case (4CL3: 4CL5) were significantly above the baseline rate and displayed a distinct change from lower 4CL3: 4CL5 ratios. Similar results were observed in Fig. 3.9B, but a weaker reduction was observed.

Figure 3.9: Initial Model Analysis for 4CLs Enzymatic Reactions (The concentration of 4-coumaric acid as substrate : 28uM)
The difference of rate reductions between the two difference graphs in Fig. 3.9 indicate that low concentrations of 4CL5 mixed with high concentrations of 4CL3 result in inhibition of product formation. This inhibitory effect, representing as the red arrow in Fig. 3.9A, may be attributed to the interaction between 4CL3 and 4CL5, where 4CL3 is recruited by low concentrations of 4CL5, reducing the amount of 4CL3 that is available to generate product. Therefore, the recruitment of 4CL3 by low concentrations of 4CL5 results in a net decrease in the overall rate. Otherwise, the relatively small inhibitory effect, represented by the red arrow in Fig. 3.9B show that 4CL5 is not greatly influenced by a 'low' concentration of 4CL3. High concentrations of both 4CL3 and 4CL5 display an activation effect, represented by the green arrow of in both graphs in Fig. 3.9, where the rate of product formation is higher than the sum of the independent rates of 4CL3 and 4CL5. This paradoxical inhibition and activation may be explained by differential regulation, which is a function of different ratios of 4CL3 and 4CL5. Based on the analysis, model development is started in order to regenerate the experimental data and to represent possible interactions between enzymes.

Model Structure Development with Assumptions  We recognized that actual experimental data are not only inhibited but also activated according to the different ratio of 4CL3 and 4CL5, when compared with the simulation results of the initial model, which assumed no interactions between enzymes. It is inferred that the interactions among the involved components could lead to the inhibition and activation effects seen in Fig. 3.9. Considering the experimentally verified existence of a complex, the analysis had a reason to believe that the complex influenced product formation.
The assumption is presented as an interaction block diagram (Fig. 3.10) to describe the effects of the enzyme complex formation on the total CoA-ligation rate associated with the related substrate. The total CoA-ligation rate is a function of the activity of free 4CL3 (E1), free 4CL5 (E2), and their protein complex. The complex formed by 4CL3 and 4CL5 may have a different ratios based on the initial model analysis. The inhibition effects resulting from the formation of the complex and the activation path of the complex, affect the rate of product formation. First, 4CL3 binds with available 4CL5, reducing the rate associated with 4CL3 (Inhibition Path A in Fig. 3.10) and then 4CL5 binds with 4CL3, leading to a reduction in rate associated with 4CL5 (Inhibition Path B in Fig. 3.10). High amounts of 4CL3 and 4CL5 result in the formation of high concentrations of the 4CL3/4CL5 complex, which can then bind to an available substrate. The complex produces an alternative path towards product formation, resulting in a net increase in the product rate (Activation Path in Fig. 3.10).

![Diagram of enzyme interactions](image)

**Figure 3.11: Initial Model Structure for 4CLs Enzymatic Reactions**

A plausible mechanistic description of the inhibition and activation concept in Fig. 3.10 is presented in Fig. 3.11. All interactions between enzymes are assumed to occur at rate $k$, which is defined as $k = k_d/k_a$, where $k_d$ is the dissociation rate between 4CL3 and 4CL5 and $k_a$ is the association rate between both enzymes.

**Mathematical Equation Development** We propose mathematical description of the total CoA-ligation rate using 4-coumaric acid or caffeic acid as substrate, which includes the inhibition and activation effects attributed to the formation of the 4CL3/4CL5 complex. The rate of
product formation considering the effects of 4CL3\((E1)\), 4CL5\((E2)\) and 4CL3/4CL5 complex can be written as:

\[
v = v_{4CL3} + v_{4CL5} + v_{4CL3/4CL5} = k_{cat1}[E1S] + k_{cat2}[E2S] + k_{cat3}[E1E2S],
\]

(3.7)

where \(k_{cat1}\) and \(k_{cat2}\) are the known rate constants for 4CL3 and 4CL5, respectively, and \(k_{cat3}\) is the unknown rate constant for the enzyme complex. The equation development of the reaction rate follows previous assumptions, such as the quasi-equilibrium and the conservation of enzymes. Quasi-equilibrium assumes that the association and disassociation among the components related to enzymes and substrates are in binding equilibrium. Enzyme conservation assumes that the total enzyme concentration is equal to the sum of species with the enzyme. The key issue in deriving the rate equation for 4CL enzymatic reaction is how we will deal with rate parameters associated with the complex and the intermediates. In this section, we assume that the ratio between 4CL3 and 4CL5, constituting the enzyme complex, is 1:1. As the optimization progresses, diverse ratios of enzyme complex can be considered. The detailed process of the equation derivation considering the final optimized ratio between enzymes is presented in Appendix A.

First, quasi-equilibrium could be applied to the bindings between enzyme-related components and substrates, as shown in (3.8), (3.9) and (3.10).

\[
[E1S] = \frac{1}{k_{M1}}[E1][S]
\]

(3.8)

\[
[E2S] = \frac{1}{k_{M2}}[E2][S]
\]

(3.9)

\[
[E1E2S] = \frac{1}{k_{M3}}[E1E2][S]
\]

(3.10)

Equation (3.8) and (3.9) have \(k_{M1}\) and \(k_{M2}\), which are Michaelis constants for 4CL3 and 4CL5, respectively. These parameters are obtained based on \textit{in-vitro} kinetics. \(k_{M3}\) in (3.10) is the unknown Michaelis constants for 4CL5/4CL5 enzyme complex and is defined in (3.17). The enzyme complex in (3.10) would also follow the quasi-equilibrium assumption as:

\[
[E1E2] = \frac{1}{k}[E1][E2],
\]

(3.11)

where \(k\) is the unknown interaction constant between 4CL3 and 4CL5. Second, enzyme conservation assumption can be applied to represent 4CL3 and 4CL5 enzyme concentrations as:

\[
[E1] = [E1] + [E1S] + [E1E2] + [E1E2S],
\]

(3.12)
and

$$[E2_t] = [E2] + [E2S] + [E1E2] + [E1E2S]. \tag{3.13}$$

Third, based on the previous two assumptions, quasi-equilibrium and the total enzyme conservation, we conclude that amounts of $[E1]$ and $[E2]$ involved in enzyme complexes such as $[E1E2]$ and $[E1E2S]$ can be presented by a proportion of the total enzyme concentration amounts respectively. The condition allows the amounts of $[E1]$ and $[E2]$ involved in enzyme complexes to be represented as:

$$[E1] = \alpha [E1t], \quad 0 < \alpha < 1 \tag{3.14}$$

and

$$[E2] = \beta [E2t], \quad 0 < \beta < 1, \tag{3.15}$$

where $\alpha$ and $\beta$ represent the proportion of 4CL3 and 4CL5 in the enzyme complex.

The process developing the mathematical expression of the rate of 4CLs enzymatic reactions requires parameter simplification, which reduces the number of unknown parameters from 5 to 3. $k_{cat3}$ can represented by multiples of $k_{cat1}$ or $k_{cat2}$, giving us reasonable ranges and representing activation effects. It is described as:

$$k_{cat3} = \gamma \cdot (k_{cat1} \text{ or } k_{cat2}) \tag{3.16}$$

where $\gamma$ is an unknown constant. $k_{M3}$ can be fixed to have the value close to $k_{M1}$ or $k_{M2}$, because the enzyme complex is comprised of 4CL3 and 4CL5, and represented as

$$k_{M3} = k_{M1} \text{ or } k_{M2}. \tag{3.17}$$

Equation (3.18) is derived from the configuration of Fig. 3.11 using the above assumptions and is represented as:

$$v = \frac{k_{cat1} \cdot [E1t] \cdot [S]}{k_{M1} + [S] + \frac{k_{M1} \cdot \beta \cdot [E2t]}{k} \cdot (1 + \frac{[S]}{k_{M2}})} + \frac{k_{cat2} \cdot [E2t] \cdot [S] \cdot (1 + \frac{[E1t]}{k})}{(k_{M2} + [S]) \cdot (1 + \frac{\alpha \cdot [E1t]}{k})} \tag{3.18}$$

where, $k$ and $\beta$, and, $k$ and $\alpha$, are not working independently. It requires that two pairs of unknown parameters converge to one unknown parameter each as:

$$k_1 = k \frac{k}{\beta}, \tag{3.19}$$

29
and

\[ k_2 = \frac{k}{\alpha}. \tag{3.20} \]

Finally, through parameter simplification, the number of unknown parameters could be reduced in the equation, which is expressed as

\[
v = \frac{k_{\text{cat}1} \cdot [E_1t] \cdot [S]}{k_{M1} + [S] + \frac{k_{\text{cat}1} \cdot [E_2t]}{k_1} \cdot (1 + \frac{[S]}{k_{M2}})} + \frac{k_{\text{cat}2} \cdot [E_2t] \cdot [S] \cdot (1 + \gamma \cdot \frac{[E_1]}{k_2})}{(k_{M2} + [S]) \cdot (1 + \frac{[E_1]}{k_2})} \tag{3.21} \]

More detailed processes developing the equation refer to derivation for the final model equation in Appendix A. The equation would be completed by parameter optimization. The completed equation will be rebuilt, according to the altered assumptions and conditions, on model optimization process iteratively.

**Parameter Optimization** The rate equations we developed have unknown parameters related to the enzyme complex, such as \( k_1, k_2 \) and \( \gamma \). It is possible to infer these parameters from the experimental data. Our parameter optimization, based on experimental data, considers the values of unknown parameters in minimizing the objective function as

\[
\min_{\bar{p}} f(\bar{p}). \tag{3.22}
\]

where \( f(\bar{p}) \) is the objective function and \( \bar{p} \) is a vector of the parameters, \( k_1, k_2 \) and \( \gamma \), that will be estimated. One common objective function to assess the goodness-fit modeling is the mean square error (MSE) [79], which is defined as

\[
f(\bar{p}) = \frac{\sum_{i=1}^{N} (v_{e_i}(\bar{p}) - v_{m_i})^2}{N}. \tag{3.23}
\]

In (3.23), \( v_{e_i} \) is the \( i \)th estimated reaction rate by the developed model, \( v_{m_i} \) is the \( i \)th measured reaction rate, and \( N \) is the number of experimental data samples. Mean square error values that approach zero represent improved model fit to the experimental data.

We implemented a hybrid optimization algorithm [84] to optimize the unknown equations in the model. This process includes global and local optimization for effective searches in complex nonlinear systems. The Genetic Algorithm, used as the global optimization method, is a search method that mimics natural evolution. The random nature of its various starting points is the reason that it is not contained by local optima. The Genetic Algorithm uses the probabilistic selection rule with parallel searching, which is attractive to evaluating large search space and solving non-linear problems with a large number of parameters (global optimization problems).

Fmincon in the Optimization Toolbox of MATLAB [75] performs local optimization. This
function is a gradient-based implementation of the Interior Point algorithm [83]. This algorithm is suitable for solving problems with nonlinear constraints and multiple variables such as 4CL enzyme kinetic modeling. The optimization process was simulated by the optimization toolbox in MATLAB 7.11. 100 runs were performed. The mean values of the results were used in the final rate equations.

Model Optimization Process In the model optimization process, structure development, mathematical equation derivation, and parameter estimation would be executed iteratively, when the results of the modeling are satisfied with the evaluation conditions. Fig. 3.12 briefly shows the model optimization process for 4CLs multi-enzymatic reaction modeling. Each row represents the assumptions applied to the unknown relationship between the involved enzymes (the left side in Fig. 3.12) and the results, simulated following the experiment scenarios, of the developed model based on the assumptions (the right side in Fig. 3.12) sequentially. The first row shows the initial model analysis with no relationship between 4CL3 and 4CL5. As we expected, the simulation results have a linearity on the basis of the summation of the reaction rates catalyzed by 4CL3 and 4CL5, and cannot express the nonlinear characteristics, inhibition and activation, of the experimental data. The model in the second column suggests that the enzyme complex shows that the modified model can support nonlinearity to some extent. Even so, the gaps between the measured data and the predicted data are not satisfied. The next modified model considers the change of the ratio of 4CL3 and 4CL5 in the enzyme complex based on the analysis that 4CL5 may recruit more 4CL3 in mixture. Finally, the third modified model uses a 3 to 1 ratio of 4CL3 and 4CL5, and becomes a better-fit model. Even when we only show three steps for building the model on the iteration process here, more diverse assumptions and conditions applied and analyzed in many times. The applied assumptions or conditions were associated chiefly with the enzyme complex and the effects of the enzyme complex on the rate of product formation, e.g., the change of ratios of both enzymes comprising of the enzyme complex. The results of each supported that the final model in the third row of Fig, 3.12 is the best.

Evaluation The final optimized model for 4CLs enzymatic reaction is evaluated against goodness-of-fit and biological relevance. For goodness-of-fit, visual comparison and numerical analysis are used. Visual display evaluation considers how well the developed model can present the nonlinear characteristics of the multi-enzymatic system. Specifically, we focus on the ability to represent the characteristics that the experiment rates for the 40nM:10nM (4CL3:4CL5) case and 40nM:40nM (4CL3:4CL5) case significantly drop and increase respectively. RMSE is used as a numerical indicator of the accuracy of the fits of model to data.

The important biological relevancy issue in the 4CLs multi-enzymatic reaction modeling,
Figure 3.12: Model Optimization Process for 4CLs Enzymatic Reactions
which requires validation, concerns the 4CL3/4CL5 enzyme complex. To prove the interaction of the two 4CLs, four diverse methods are used: (1) Bimolecular Fluorescence Complementation (BiFC) to test the interaction between the enzymes [41], (2) chemical crosslinking [51] and (3) Co-Immunoprecipitation (Co-IP) to verify the existence of the enzyme complex [14] (4) mass spectrometry (MS) [71] to identify 4CL3 : 4CL5 ratio in the enzyme complex [15].

3.4 Results

3.4.1 Case 1: CAld5Hs Enzymatic Reaction with Functional Redundancy

CAld5H Multiple Enzymatic Reaction in Lignin Biosynthesis using Coniferaldehyde and Coniferyl alcohol respectively, as Substrates

Model Structure and mathematical expression A plausible mechanistic description of functional redundancy in Fig. 3.5 is shown in Fig. 3.8 and used for the initial model for 4CLs enzymatic reaction, for coniferaldehyde and coniferyl alcohol. A mathematical model representing the rate of total product formation associated with CAld5H1 and CAld5H2, was derived from Fig. 3.8 and formed as the sum of Michaelis-Menten kinetics of the individual enzyme (3.4). The kinetic parameters associated with CAld5H1 and CAld5H2, according to each substrate, are the same as the observed kinetics displayed in Wang et al. (2012) [80].

Simulation Results Fig. 3.13 and Fig. 3.14 display these simulated values along with the experimentally measured rates for the different substrates. The model fits the data well as shown by the scatter (experimentally measured rates) and line (simulated rates) plots. RMSE and $R^2$ value used to describe the goodness of fit of the model also show the accuracy of the model prediction: RMSE of 0.0600 and $R^2$ of 99.5372 for coniferaldehyde (Fig. 3.13) and RMSE of 0.1146 and $R^2$ of 99.3738 for coniferyl alcohol (Fig. 3.14). Overall, the model provides an adequate representation of the metabolic rate involving CAld5H1 and CAld5H2 for two different substrates, coniferaldehyde and coniferyl alcohol.

CAld5H Multiple Enzymatic Reaction in Lignin Biosynthesis using Multiple Substrates with Coniferyl alcohol as a Main Substrate

Model Structure and Mathematical Expression A plausible mechanistic description of functional redundancy in Fig. 3.5 is shown in Fig. 3.15, for 5-hydroxylation of coniferyl alcohol with inhibition by coniferaldehyde. A mathematical model representing the rate of total product formation associated with CAld5H1 and CAld5H2, was derived from Fig. 3.15 and formed as the sum of Michaelis-Menten kinetics of the individual enzyme (3.24):
Figure 3.13: Simulation Results along with the Experimental Rates using Coniferaldehyde as Substrate.

Figure 3.14: Simulation Results along with the Experimental Rates using Coniferyl Alcohol as Substrate.
\[ v = v_{\text{CAld}5\text{H}1} + v_{\text{CAld}5\text{H}2} = \frac{k_{\text{cat}1}[E_1][S]}{k_{M1} \cdot (1 + \frac{[I]}{k_{1\text{IC}}}) + [S] \cdot (1 + \frac{[I]}{k_{1\text{IU}}})} + \frac{k_{\text{cat}2}[E_2][S]}{k_{M2} \cdot (1 + \frac{[I]}{k_{2\text{IC}}}) + [S]}, \] (3.24)

where \( k_{1\text{IC}} \) and \( k_{2\text{IC}} \) are the competitive inhibition kinetics of CAld5H1 enzyme and CAld5H2 enzyme by coniferaldehyde on coniferyl alcohol 5-hydroxylation, respectively, and \( k_{1\text{IU}} \) is the uncompetitive inhibition kinetic of CAld5H1 enzyme by coniferaldehyde on coniferyl alcohol 5-hydroxylation. The inhibition kinetic parameters related to coniferaldehyde are the same as the observed kinetics displayed in Wang et al. (2012) [80].

![Model Structure Development for CAld5Hs Enzymatic Reaction with Inhibition](image)

**Simulation Results**  Fig. 3.16 displays these simulated values along with the experimentally measured rates. Numerical analysis of the model also describes the accuracy of the model prediction: RMSE of 0.1801 and \( R^2 \) of 98.5193 for 5-hydroxylation of coniferyl alcohol inhibited by coniferaldehyde (Fig. 3.16). The modeling results show that the assumption of functional redundancy can provide an adequate representation of the metabolic rate involving CAld5H1 and CAld5H2 in the case of multiple substrates.
3.4.2 Case 2: 4CLs Enzymatic Reaction with Enzyme Complex

4CL Multiple Enzymatic Reaction in Lignin Biosynthesis using 4-coumaric Acid as Substrate

**Structure**  Fig. 3.17 represents the proposed interactions of 4CL3 and 4CL5 and their effects on product formation. The formation of the 4CL3/4CL5 complex with a 3:1 ratio leads to decreasing amounts of free enzymes and causes a rate reduction (Inhibition Path A and Inhibition Path B). The 4CL3/4CL5 complex is then involved in product formation and results in an increased rate (Activation Path). A plausible mechanistic description of the inhibition and activation concept in Fig. 3.17 is shown in Fig. 3.18 for 4CA. Inhibition occurs due to interaction between free 4CL3 and 4CL5 that leads to the formation of the 4CL3/4CL5 tetramer. All interactions involving free 4CL3, free 4CL5, and other enzyme-complex intermediates are assumed to occur at rate $1/k$. Alternative causes of inhibition involving the interaction of free enzymes with enzyme-substrate complexes are not supported by other model structures. A mechanistic model for activation in enzymatic reactions in Saboury, 2009 and Fontes et al., 2000 was used here to describe the activation effect of the 4CL3/4CL5 complex (Fig. 3.18) [63] [29]. The kinetic parameters associated with 4CL3 ($E_1$) and 4CL5 ($E_2$) are the same as the observed kinetics displayed in Chen et al., 2013 and Chen et al., 2014 [16] [15]. Given that the results...
described above and in Chen et al. (2013) indicated dominant 4CL5 kinetics, we set the kinetic parameters associated with the 4CL3/4CL5 complex to $1/k_{m2}$ and $\gamma \cdot 1/k_{cat2}$ respectively.

Figure 3.18: Model Structure of 4CLs Enzymatic Reaction using 4-coumaric Acid as Substrate
Mathematical Model  A mathematical model representing the rate of total product formation associated with 4CL3, 4CL5, and the 4CL3/4CL5 complex, was derived from Fig. 3.18. The derivation was based on the Michaelis-Menten assumption of quasi-equilibrium where the association and disassociation of enzymes, enzyme complexes, and their intermediates are in binding equilibrium. The quasi-equilibrium assumption is adequate given that the individual enzymes follow Michaelis-Menten kinetics. Further, this quasi-equilibrium assumption allows several simplifications based on inequality bounds on enzyme complex intermediates (Appendix A). These derivations lead to a combined rate equation for the 4-coumaric acid substrate (3.25).

\[ v = \frac{k_{cat1} \cdot [E_1] \cdot [S]}{k_{M1} + [S] + \frac{3k_{M1}[E_1][S]^2}{k_1^3} \cdot (1 + \frac{[S]}{k_{M2}})} + \frac{k_{cat2} \cdot [E_2t] \cdot [S] \cdot (1 + \gamma \cdot \left(\frac{[E_1]}{k_2}\right)^2)}{(k_{M2} + [S]) \cdot (1 + \left(\frac{[E_1]}{k_2}\right)^2) + k_{M1} \cdot [S] + \frac{3k_{M1}[E_1][S]^2}{k_1^3} \cdot (1 + \frac{[S]}{k_{M2}})} \]  (3.25)

The assumption of quasi-equilibrium allows us to describe the concentrations of free 4CL3, free 4CL5, and all enzyme complex intermediates in terms of proportional amounts of known total 4CL3 and 4CL5 concentrations. The proportion of 4CL3 (\(\alpha\)) and the proportion of 4CL5 (\(\beta\)) in each binding associated with tetramer and intermediates, however, are unknown and have not yet been measured experimentally. Thus, these unknowns are combined with the parameter \(k\) to form \(k_1 = (k/\alpha^2\beta)^{1/3}\) and \(k_2 = k/\alpha\), leading to three unknown values in (3.25); \(k_1\), \(k_2\) and \(\gamma\).

Experimental rates of product formation under different total 4CL3 and 4CL5 concentrations about experiment design scenarios were used to optimize \(k_1\), \(k_2\) and \(\gamma\). An objective function based on Least Mean Square Error was used to assess the goodness of fit between the experimental data and the mathematical model in (3.25) [82]. This objective function, in combination with a Hybrid optimization approach was used to estimate the values of \(k_1\), \(k_2\) and \(\gamma\) in (3.25) [84]. The Hybrid optimization algorithm is based on a search routine that utilizes both global optimization and local optimization to efficiently optimize large-scale problems within a complex search space. Genetic Algorithms [31], were used as the global optimization and Fmincon (Mathworks. Optimization Toolbox, Version 3, Users Guide (2007)) was used for the local optimization. 100 runs using the Hybrid optimization scheme were performed with random initial conditions over the range \([0 - 0.1]\) for \(k_1\), \([0 - 0.1]\) for \(k_2\), and \([1 - 3]\) for \(\gamma\). Wider ranges for these parameter values did not lead to any significant change in the optimized value.

The resulting mean values and standard errors of the optimized parameters are shown in Table 3.1. These mean values were incorporated in (3.25) and used to produce simulated total product formation rates along with two different experiment design scenarios.
Table 3.1: Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using 4-coumaric Acid as Substrate

<table>
<thead>
<tr>
<th>Unknown Parameter</th>
<th>Optimized Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_1 )</td>
<td>0.0213 ± 0.0190</td>
</tr>
<tr>
<td>( k_2 )</td>
<td>0.0326 ± 0.0115</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>2.0429 ± 0.3644</td>
</tr>
</tbody>
</table>

Simulation Results Fig. 3.19 displays these simulated values along with the experimentally measured rates for two different experimental scenarios as we mentioned in Section 3.3.2. The solid lines represent simulation based on the equation and the optimized parameters in Table 3.1. Dots represent experimental data with 4-coumaric acid as a main substrate. Colors represent different substrate concentrations. The model fits the data well as shown by the scatter (experimentally measured rates) and line (simulated rates) plots, with a mean squared error of 0.0062, \( R^2 \) value of 0.81 and RMSD of 0.13 used to describe the goodness of fit of the model also show the accuracy of the model prediction (Fig. 3.19). The model describes the prominent dip that occurs due to the inhibition of the product formation rate at high levels of 4CL3 and low levels of 4CL5. The model also reflects the activation at high levels of both 4CL3 and 4CL5. Overall, the model provides an adequate representation of the metabolic rate involving 4CL3, 4CL5, and the 4CL3/4CL5 complex for 4-coumaric acid as substrate.

![Simulation Results](image)

Figure 3.19: Simulation Results along with the Experimental Rates using 4-coumaric Acid as Substrate
4CL Multiple Enzymatic Reaction in Lignin Biosynthesis using Caffeic Acid as Substrate

**Structure**  The two most important substrates in 4CL ligation are 4-coumaric acid and caffeic acid. The mechanistic model derived for 4-coumaric acid cannot be applied directly to caffeic acid because caffeic acid exhibits substrate self-inhibition in the 4CL5 reaction [16]. 4-coumaric acid does not show self-inhibition with either 4CL3 or 4CL5. A description of the effect of self-inhibition of caffeic acid with 4CL5 is displayed in Fig. 3.20. Self-inhibition acts in an uncompetitive mode with excess caffeic acid reducing the rate associated with 4CL5 by binding to the enzyme-substrate complex. The complex was assumed to take on the same inhibitory characteristics as 4CL5 because 4CL5 takes a dominant role in the metabolic flux [16]. The self-inhibition function leads to excess caffeic acid binding to the 4CL3/4CL5-substrate complex, inhibiting product formation associated with the complex. The inhibition-binding rate in each case is $k_{is}$ [16].

![Diagram of 4CLs Enzymatic Reaction using Caffeic Acid as Substrate](image-url)

**Figure 3.20:** Model Structure of 4CLs Enzymatic Reaction using Caffeic Acid as Substrate
Table 3.2: Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using Caffeic Acid as Substrate

<table>
<thead>
<tr>
<th>Unknown Parameter</th>
<th>Optimized Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>$0.0242 \pm 0.0101$</td>
</tr>
<tr>
<td>$k_2$</td>
<td>$0.0544 \pm 0.0082$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$5.1893 \pm 1.0668$</td>
</tr>
</tbody>
</table>

Mathematical Model A mathematical model quantifying the rate of total product formation associated with caffeic acid as a substrate was also derived based on the Michaelis-Menten assumptions of quasi-equilibrium used previously for 4-coumaric acid. The inclusion of self-inhibition in the combined mechanistic model for caffeic acid modifies (3.25) slightly, resulting in (3.26).

\[
v = \frac{k_{\text{cat}1} \cdot [E_1] \cdot [S]}{[S] + \frac{3k_{M1} [E_1]^2 [E_2]}{k_1^2} \cdot (1 + \frac{[S]}{k_{M2}^2 k_{IS}})} + \frac{k_{\text{cat}2} \cdot [E_2] \cdot [S] \cdot (1 + \gamma \cdot (\frac{[E_1]}{k_2})^3)}{(k_{M2} + [S] + \frac{[S]^2}{k_{IS}}) \cdot (1 + (\frac{[E_1]}{k_2})^3)},
\]

(3.26)

The hybrid optimization approach performed for 4-coumaric acid was also used to assess the goodness of fit of (3.26) with measured product formation rates for varying amounts of total 4CL3, 4CL5, and caffeic acid concentrations. A total of 100 runs for the optimization routine was performed with random initial conditions over the range $[0 - 0.1]$ for $k_1$, $[0 - 0.1]$ for $k_2$, and $[1 - 6]$ for $\gamma$ to assess variation in estimated parameters $k_1$, $k_2$ and $\gamma$. While the range of values for $k_1$ and $k_2$ did not change from 4-coumaric acid to caffeic acid, it was expected that the range for $\gamma$ may change due to differences between enzyme substrate interactions. Larger ranges did not reveal any significant difference in the optimized results. The resulting mean values of the optimized parameters are shown in Table 3.2.

Simulation Results As with the 4-coumaric acid substrate, both experimental scenarios were simulated using the mean values of $k_1$, $k_2$ and $\gamma$ (Table 3.2). Visual presentations of similarities and differences between observed and simulated data for the models show a good fit (Fig. 3.21). Both simulated graphs fit the data well as indicated by a minimum mean square error of 0.0034. The proposed models in Figs 3.18 and 3.20 provide an important characteristic that differentiates the experimentally measured product formation rates for 4-coumaric acid and caffeic acid, respectively. Experimental rates in Fig. 3.19 show that product formation rates associated with 4-coumaric acid increase steadily as the concentration of 4-coumaric
acid increases. Experimental rates with caffeic acid in Fig. 3.21, however, show that increased concentrations of caffeic acid result in a slow down or in some cases, a decrease of the product formation rate, due to the self inhibition exhibited by caffeic acid and 4CL5. The model of total product formation in (3.26) is able to capture this reduction in the rate, specifically for high concentrations of caffeic acid (18.75uM and 25uM). This consistency of the model and the experimental results emphasize that self-inhibition must be included when modeling and predicting CoA ligation rates for the caffeic acid substrate.

Figure 3.21: Simulation Results along with the Experimental Rates using Caffeic Acid as Substrate)

3.4.3 Case 3 : 4CLs Enzymatic Reaction with Enzyme Complex considering Inhibitions by Multiple Substrates

4CL Multiple Enzymatic Reaction in Lignin Biosynthesis using Multiple Substrates with 4-Coumaric Acid as a Main Substrate

Structure Fig. 3.17 represents the proposed interactions of 4CL3 and 4CL5 with multiple inhibitors and the effects on product formation. The free 4CL3 enzymatic reaction has only competitive inhibition and the 4CL5 enzymatic reaction has competitive and uncompetitive inhibition. The enzymatic reaction of the 4CL3/4CL5 complex with a 3:1 ratio also considers competitive and uncompetitive inhibition.

Fig. 3.23 illustrates a mechanistic model of interactions between 4-coumaric acid, caffeic
Figure 3.22: Mechanistic Description for 4CLs Enzymatic Reaction with Multiple Substrates

acids, ferulic acid, 4CL3, 4CL5, the 4CL3/4CL5 complex, and the corresponding intermediates with 4-coumaric acid being the main substrate. All kinetic constants and associated inhibition constants except $k$ and $\gamma$ are obtained from the in vitro kinetics of 4CL3 and 4CL5 in Chen et al. (2013) and Chen et al. (2014) [16] [15].

Mathematical Model The quasi-equilibrium assumption was again used to derive a mathematical model of the total product formation rate associated with 4-coumaric acid in the presence of 4CL3, 4CL5, the 4CL3/4CL5 complex, caffceic acid, and ferulic acid in (3.27). The detail derivation process for the final equation is displayed in Appendix A. Measured product formation rates under conditions of varying the concentrations of total 4CL3, 4CL5, 4-coumaric acid, caffeic acid, and ferulic acid were used in combination with the aforementioned hybrid optimization approach to estimate $k_1$, $k_2$ and $\gamma$. A total of 100 runs using Hybrid optimization were performed with random initial conditions over the range [0 - 0.1] for $k_1$, [0 - 0.1] for $k_2$, and [1 - 3] for $\gamma$. Optimized values for $k_1$, $k_2$, and $\gamma$ are shown in Table 3.3, which includes the resulting mean values and their standard errors. Both experimental scenarios (4CL3 fixed: 4CL5 varied; 4CL5 fixed: 4CL3 varied) were simulated based on (3.27) using the mean values of $k_1$, $k_2$, and $\gamma$. 

43
Figure 3.23: Model Structure of 4CLs Enzymatic Reaction using 4-coumaric Acid as a Main Substrate with Inhibition
Table 3.3: Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using 4-coumaric Acid as a Main Substrate with Inhibition

<table>
<thead>
<tr>
<th>Unknown Parameter</th>
<th>Optimized Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_1 )</td>
<td>0.0202 ± 0.0091</td>
</tr>
<tr>
<td>( k_2 )</td>
<td>0.0331 ± 0.0143</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>2.2902 ± 0.3808</td>
</tr>
</tbody>
</table>

\( v = \frac{k_{cat1} [E1]_1 [S]}{k_{M1} (1 + \frac{[I1]}{k_{IC1}1} + \frac{[I2]}{k_{IC2}1} + [S]) + \frac{2 k_{M1} [E1]_1 [E2]_1}{k_1} \cdot \frac{[I1]}{k_{IC1}1} \cdot \frac{[I2]}{k_{IC2}1} + \frac{[S]}{k_M^2} \cdot (1 + \frac{[I1]}{k_{IC1}2} + \frac{[I2]}{k_{IC2}2})} + \frac{k_{cat2} [E2]_2 [S] (1 + \gamma + \frac{[E1]_1}{k_{E1}})^3}{(k_{M2} (1 + \frac{[I1]}{k_{IC1}2} + \frac{[I2]}{k_{IC2}2} + [S]) (1 + \frac{[I1]}{k_{IC1}2} + \frac{[I2]}{k_{IC2}2})) (1 + \frac{[E1]_1}{k_{E1}})^3}. \)  

(3.27)

**Simulation Results**  
Fig. 3.24 shows that the model fits the data well, with the mean squared error equal to 0.0077. We test the plausibility of the multiple substrate mechanistic model and resulting mathematical model in (3.27). We need to assess whether the additional complexity (addition of inhibition characteristics) is required to describe changes in the total rate of product formation. We perform this test by fitting equation in (3.25) without integrated inhibition to experimental rates in Fig. 3.21 to identify whether characterization of inhibition is critical for describing changes in product formation rates in the presence of multiple hydroxycinnamic acids. The best result after multiple runs of the hybrid optimization approach provided estimates of \( k_1 \), \( k_2 \), and \( \gamma \) that yield a mean square error of 2.4. This mean square error, using the equation in (3.25), is four orders of magnitude higher than the mean square error of 0.0059 produced using the model that includes substrate inhibitions. The mechanistic description of the rate of product formation using 4-coumaric acid does not adequately describe situations that are closer to in vivo conditions without considering interactions between caffeic acid, and ferulic acid, 4CL3, and 4CL5.

**4CL Multiple Enzymatic Reaction in Lignin Biosynthesis using Multiple Substrates and Caffeic Acid as the Main Substrate**

**Structure**  
The known substrate self-inhibition characteristics exhibited by caffeic acid on the 4CL5 reaction must be incorporated into the mechanistic model to correctly predict the rate of product formation resulting from caffeic acid in the presence of multiple substrates. As before, we model the self-inhibition on 4CL5 as uncompetitive. We maintain the assumption
that the enzyme complex takes on the same characteristics due to the dominance of the 4CL5. Fig. 3.25 illustrates the mechanistic configuration for substrate self-inhibition of caffeic acid in the presence of 4-coumaric acid and ferulic acid. Again, $k_i$s shown in the yellow box in Fig. 3.25 is the self-inhibition binding rate for both 4CL5 and the enzyme complex.

Mathematical Model  The mathematical model of the total product formation rate associated with caffeic acid derived from the mechanistic configuration in Fig. 3.25 is shown in (3.28). The detail derivation process for the final equation is displayed in Appendix A. Experimental rates and optimized values of $k_1$, $k_2$ and $\gamma$ were obtained as described in previous sections, with random initial conditions of $k_1$, $k_2$ and $\gamma$ extending over $[0 - 0.1]$, $[0 - 0.1]$, and $[1 - 6]$, respectively. The optimized values for $k_1$, $k_2$ and $\gamma$ are shown in Table 3.4

$$
\nu = \frac{k_{cat1} \cdot [E1] \cdot [S]}{k_{M1} \cdot (1 + \frac{[I1]}{k_{IC1}} + \frac{[I2]}{k_{IC2}}) + [S] + \frac{3 \cdot k_{M1} \cdot [E1] \cdot [S]}{k_1} \cdot \frac{[I1]}{k_{IC1}} \cdot (1 + \frac{[I1]}{k_{IC1}} + \frac{[I2]}{k_{IC2}}) + \frac{[S]}{k_{M2}} \cdot (1 + \frac{[I1]}{k_{IC1}} + \frac{[I2]}{k_{IC2}}) + (1 + \frac{[E1]}{k_{lo}})^3} \\
+ \frac{k_{cat2} \cdot [E2] \cdot [S] \cdot (1 + \gamma \cdot (\frac{[E1]}{k_{lo}})^3)}{(k_{M2} \cdot (1 + \frac{[I1]}{k_{IC1}} + \frac{[I2]}{k_{IC2}}) + [S] \cdot (1 + \frac{[I1]}{k_{IC1}} + \frac{[I2]}{k_{IC2}}) \cdot (1 + \frac{[E1]}{k_{lo}})^3)}. 
$$  (3.28)

Simulation Results  Fig. 3.26A and 3.26B illustrates both the experimental scenarios with mean values of $k_1$, $k_2$, and $\gamma$. We calculate a mean square error of 0.0233, showing that the model adequately fits the experimental product formation rates.
Figure 3.25: Model Structure of 4CLs Enzymatic Reaction using Caffeic Acid as a Main Substrate with Inhibition

Table 3.4: Parameter Optimization for Mathematical Model of 4CLs Enzymatic Reaction using Caffeic Acid as a Main Substrate with Inhibition

<table>
<thead>
<tr>
<th>Unknown Parameter</th>
<th>Optimized Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_1$</td>
<td>0.0256 ± 0.0069</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.0716 ± 0.0138</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>3.3806 ± 1.1059</td>
</tr>
</tbody>
</table>
Validation We now assess the plausibility of this caffeic acid model by asking if the mechanistic description of the multi-substrate inhibitions of 4-coumaric acid and ferulic acid and substrate self-inhibition of caffeic acid was needed to adequately describe the experimental rates. We explore the plausibility of \( (3.28) \) by looking at the following characteristics independently: 1) inhibitions associated with the presence of 4-coumaric acid and ferulic acid substrates and 2) substrate self-inhibition associated with caffeic acid and 4CL5. We assess point 1) by identifying whether the multiple enzyme / single substrate mechanistic model for caffeic acid that lacked multi-substrate inhibitions in \( (3.26) \) presents an equally plausible model to describe the data shown in Fig. 3.26. The fit of \( (3.28) \) to these data resulted in a mean square error of 0.1076, which was substantially greater than the mean square error of 0.0233 associated with the multiple enzyme / multiple substrate model for caffeic acid in \( (3.28) \). This higher mean square error highlights the need to mechanistically describe substrate inhibitions when exploring the impact of multiple enzymes and multiple substrates on the rate of product formation.

We assess the second point 2) by investigating the impact of self-inhibition on the experimental rates. Visually comparing the experimental rates in Fig. 3.24 (4-coumaric acid : no known self-inhibition) and Fig. 3.26 (caffeic acid: known self-inhibition), we see more tightly grouped measurements for experimental rates in Fig. 3.26 as the substrate concentration is increased for each 4CL3:4CL5 combination. Similar characteristics were visualized in Fig. 3.21, which we attributed to substrate self-inhibition. We assess whether a mechanistic description of substrate self-inhibition better describes these data by fitting a multiple enzyme / multiple
substrate model for caffeic acid without self-inhibition to the data in Fig. 3.26. The fit resulted in a mean square error of 0.0741, which is higher than the mean square error of 0.0233 of the model that incorporates substrate self-inhibition into the multiple enzyme / multiple substrate model (3.28). Based on these results, we conclude that both a mechanistic description of 4-coumaric acid and ferulic acid inhibition, along with a mechanistic description of caffeic acid self-inhibition, better describes CoA ligation associated with caffeic acid in the presence of 4-coumaric acid and ferulic acid.

3.5 Discussion

3.5.1 Analysis

We have proposed mathematical models derived from experimentally verified mechanistic interactions among enzymes and substrates that participate in CoA ligation of 4-coumaric acid and caffeic acid substrates in P. trichocarpa. These models (3.25 and 3.26) were developed based on conditions assumed to be an approximation of the enzymes in vivo. These rate equations provide the same benefits as Michaelis-Menten rate kinetics, which have been used to describe the product rate formation for simpler enzymatic reactions [68]. Conceptually, (3.25) and (3.26) can be illustrated using a nonlinear function block model whose output (product formation rate) is dependent upon the inputs (initial substrate concentrations, total enzyme concentrations), the functional form of the rate equation, and internal kinetic constants (Fig. 3.27). Such a convenient mathematical construct enables us to predict how variation in these inputs, in this case 4-coumaric acid, caffeic acid, and ferulic acid substrates and total initial concentrations of 4CL3 and 4CL5, influence the product formation rate for a given substrate.

Knowledge of these and other individual reaction rates in the lignin biosynthesis pathway provides a basis for predicting how the interplay between multiple metabolites impact pathway-wide substrate consumption and formation. The development of specific reaction kinetics, such as this model of CoA ligation, is the initial step to constructing a multi-reaction pathway model consisting of a system of equations that can predict changes in metabolite products and their substrate intermediates in response to changes in enzyme regulation, inhibition characteristics, and other enzymatic properties critical to metabolic engineering.

3.5.2 Prediction

The mathematical models presented here allow us to assess how the reaction rate changes for each of the individual enzymes or the complex (free 4CL3, free 4CL5, and 4CL3/4CL5 complex) under two experimental scenarios. This free enzyme and enzyme complex pathway analysis for 4CL allows us to assess the transition of enzyme activity and identify plausible mechanistic
interactions that impact the overall reaction rate through individual enzyme states. This ability to analyze the contribution of the individual reaction rates to the total rate is currently difficult due to the inability to isolate and purify the 4CL3-4CL5 complex. Fig. 3.28 displays the results of using (3.25) to compute the fraction of the total rate that can be attributed to each of the enzyme entities (4CL3, 4CL5, and the enzyme complex), for any combination of total enzyme concentration of 4CL3 and 4CL5.

In Fig. 3.28A, we see that as 4CL5 is introduced, the rate associated with 4CL3 immediately starts to decrease. This reduction in rate associated with free 4CL3 can be attributed to the recruitment of 4CL3 by 4CL5 into the enzyme complex, thus reducing the amount of free 4CL3 that is available to bind to the substrate. This free 4CL3 rate reduction may lead to the initial slow-down or inhibition that was observed with the total rate when small amounts of 4CL5 are introduced to 40nM of 4CL3. As more 4CL5 is introduced, the rate associated with 4CL3 continues to decrease rapidly, while the simulated rate associated with the 4CL3-4CL5 complex increases rapidly. Under this model, free 4CL5 and the 4CL3/4CL5 complex are the primary drivers of the total rate and are responsible for the increased activation seen at equal levels of 4CL3 and 4CL5. The reduction in rate associated with 4CL3 in the model fits well with the experimental results.

In Fig. 3.28B, we see a similar effect. As we introduce small amounts of 4CL3 to 40nM of 4CL5, we see that the simulated rate associated with the 4CL3-4CL5 complex starts to increase slowly. This slow increase allows us to propose that high concentrations of 4CL5 and low concentrations of 4CL3 result in relatively low concentrations of the 4CL3-4CL5 complex. We
also notice that the simulated rate associated with free 4CL3 does not increase, likely meaning that available 4CL3 has been recruited into the 4CL3-4CL5 complex. As the initial concentration of 4CL3 increases, the simulated rate associated with 4CL3 remains relatively low. This low free 4CL3 rate would suggest that the 4CL3 enzyme is not available to bind to the substrate but instead, continues to bind to the free 4CL5, resulting in a greater concentration of the 4CL3-4CL5 complex and an increased rate associated with the complex. We do see a decrease in the simulated rate associated with 4CL5 in scenario 2 but not as significant as the decrease seen with 4CL3 in scenario 1. Several factors can be inferred from the simulation in Fig. 3.28. First, CoA ligation can be manipulated in the plant by a nonlinear control that is not proportional to the individual expression of 4CL3 and 4CL5. Second, we can also infer that any significant control of CoA ligation would likely come from the manipulation of 4CL5 concentration when both enzymes are present. Finally, we can infer that any metabolic engineering of CoA ligation should be conducted via the manipulation of 4CL5 because it appears to be the primary controller of the total reaction rate.

We further investigated how inhibitions impact the individual rates associated with 4CL3, 4CL5, and the 4CL3-4CL5 complex by analyzing the model in (3.28). Fig. 3.29 displays the simulated and experimental total reaction rates for caffeic acid in the presence of 4-coumaric acid and ferulic acid, along with the simulated individual rates associated with each enzyme entity (4CL3, 4CL5, and enzyme complex).

In Fig. 3.29A, we see that there is a reduction 4CL3 rate when small amounts of 4CL5 are introduced. This reduction, however, is not as drastic as shown in Fig. 3.29. We infer that the
inhibition (multiple substrate inhibition of 4-coumaric acid and ferulic acid and/or substrate self-inhibition of caffeic acid) impacts the recruitment of 4CL3 by 4CL5 to form the enzyme complex, increasing the amount of free 4CL3 and hence free 4CL5 that are available to bind to the substrate. We also see that the simulated rate associated with 4CL5 is greater than the simulated rate associated with the 4CL3-4CL5 complex, which is different from the results seen in Fig. 3.29A. This change in the major contributor to the total 4CL pathway supports the mechanistic hypothesis that the inhibitions identified in the 4CL pathway either 1) reduce the concentration of available 4CL3-4CL5 complex, or 2) reduce the reaction rate associated with the complex significantly or both. Further experiments would be needed to verify this hypothesis. The simulated individual rates shown in Fig. 3.29B confirm that the introduction of increasing initial concentrations of 4CL3 to 40nM of 4CL5 results in increased rates associated with the 4CL3-4CL5 complex. The simulated rates of free 4CL3 and 4CL5 are still relatively large compared to those in Fig. 3.29B, supporting the hypothesis that more free 4CL3 and 4CL5 enzymes are available. Thus, altering the inhibition characteristics of the enzymes may mitigate the rate of product formation from the 4CL3/4CL5 enzyme complex. This change in priority of flux through the pathway under inhibition provides another potential avenue for engineering nonlinear control of CoA ligation.
3.6 Chapter Summary

A mechanistic model for multi-enzymatic reactions can lead to the more accurate prediction of relationships between enzymes, which makes it possible to regulate product we want. We proposed the approach for developing the mechanistic model. The framework included the parameter and structure optimization process considering the uncertainty in multi-enzymatic reactions. The development modeling framework applied to two multiple enzymatic reactions with different characteristics, CAld5Hs enzymatic reactions with functional redundancy and 4CLs enzymatic reactions with enzyme complex, in lignin biosynthesis. The modeling process gave us the plausible mechanistic enzyme-enzyme interaction, the mathematical models based on the interaction, the reproduction and prediction of experimental reaction data. Additionally, simulation using the developed model showed the ability to analyze the contribution of individual enzyme components to the total reaction, which is hard to estimate in a general experiment. Thus, the proposed modeling framework provides the useful process for mechanistic models of complex biological systems with unknown structures and parameters, and the developed models yield the same benefits as Michaelis-Menten kinetics. Nevertheless, time consumption and complexity, which are problems typical of mechanistic modeling, are still issues to be overcome.
Chapter 4

Mechanistic Modeling Approach based on Rule-Based Modeling and Evolutionary Computation

4.1 Motivation

Modeling multi-enzymatic reactions for biomass regulation enables more comprehensive analysis of biosynthesis. This modeling can identify the impact of transgenic perturbations on systemic characteristics of metabolic pathways. It would be helpful to have an approach that could guide more complex experimentation which could then be used to verify putative interactions. However, there is significant difficulty involved in establishing mechanistic representations of reactions.

Chapter 3 proposed the mechanistic modeling approach based on theoretical analysis, including profound chemical knowledge. The modeling approach in Chapter 3 gave us the predicted mechanistic enzyme-enzyme interaction and the detailed mathematical models. The developed models in Chapter 3 reproduced experimental data of reaction rates, for two important multi-enzymatic reactions in lignin biosynthesis. But, the time-consuming and complicated process of searching for the optimized model persists as the future work.

In Chapter 4, we suggest a new mechanistic modeling approach for resolving the troublesome issues of general mechanistic modeling. The proposed modeling approach utilizes computational methods for accelerating the model optimization process and reducing the unnecessary complexity. Rule-based modeling is very attractive for representing complex interactions in biochemical networks, which can build up the model with the topology of interactions between numerous enzymes and substrates that influence the reaction rate. Evolutionary computation is a computational optimization approach, which is inspired by the biological mechanisms of evolution.
and adaptation. The algorithmic evolutionary processes automatically searches and optimizes the best solution for complex nonlinear problems. The proposed modeling approach, based on rule-based modeling and evolutionary computation, provides information related to plausible interactions, the core interactions among the involved components, and intuitive model structures more than the models seen in Chapter 3. Such an advantage of the proposed modeling approach is discovered in three different reactions that have important roles in lignin biosynthesis.

This chapter is outlined as follows. Section 4.2 proposes the new modeling approach which describes the modeling concept, framework, and important algorithms. Section 4.3 implements the modeling methods using rule-based modeling and an evolutionary computation algorithm. As a proof of concept, this method is applied to three different multi-enzymatic reactions, such as CAld5Hs enzymatic reactions, 4CLs enzymatic reaction, and 4CLs/HCTs enzymatic reactions in lignin biosynthesis, in Section 4.4. Section 4.5 summarize of the performance of the modeling method.

4.2 Proposed Modeling Approach

4.2.1 Proposed Modeling Concepts

The proposed modeling approach is suggested as a method for guiding experimental specifications. The essential concept of the new modeling approach is based on simple structure representing the flow of enzymatic regulation for metabolic reaction rates. The proposed modeling makes it possible to avoid the excessive physical and chemical complexities of modeling and suggests a concise model including potential core components and interactions in multi-enzymatic reactions. Fig. 4.1 shows the difference among a mechanistic model, an empirical model and the proposed model for the multi-enzymatic reactions with two enzymes and one substrate. The example has the characteristics of a nonlinear system presumed to be to the results of enzyme-enzyme interaction. Each figure in Fig. 4.1 represents the expected structure of the model based on mass action law (Fig. 4.1A), simple machine learning (Fig. 4.1B) that is one type black box modeling where nonlinear relationships are connected without mechanistic functions, and the proposed concept (Fig. 4.1C). Fig. 4.1A is comprised of an example of a mechanistic model description of a multi-enzymatic reaction with an enzyme complex. Through the detailed structure, we could develop a biochemically realistic representation. But we must address issues, such as necessary physical and chemical information, numerous reactions among components and long development time, arising from excessively complicated descriptions. While, the model of Fig. 4.1B focuses only on a phenomenological relationship between inputs and outputs, from which it is very difficult to extract meaningful relationships inside the system. The proposed model structure in Fig. 4.1C presents many configurations.
using combinations of simpler known interactions. This concise expression not only makes the structure simple but also helps us recognize the enzymatic regulation flow and the effects on their product in metabolic pathways.

Figure 4.1: Model Structures for Multi-Enzymatic Reactions - A. Mechanistic Model B. Empirical Model C. Proposed Model (E1, E2: Single Enzyme, Ec: Enzyme Complex, S: Substrate, P: Product)

The modeling approach for the proposed model concepts allows us to rule out interactions or topologies that do not fit the data quickly, and reduces the available space of experimental specifications needed to help build the model. This proposed modeling approach is comprised of roughly two parts; important biochemical participants such as enzyme-enzyme complexes and connections that represent the results of the relationships among participants. This modeling requires the ability to find the participants based on first principles knowledge such as experimental data or information from previous studies. Also, the modeling must have the ability to find and represent the relationships, e.g., independence and combination, between nodes without detailed mechanistic knowledge of the interactions of components associated with the reaction flux, shown in Fig. 4.1A. The reactions are expressed by simplified characteristics extracted from chemical reaction methods like mass action law and information retrieval from diverse references. In order to implement these two potentials, this research presents the modeling framework with new approach algorithms in the next subsection.

4.2.2 Proposed Modeling Framework

The purpose of this research is to model the effects of involved enzymes on reaction rates in multi-enzymatic reactions for regulated metabolic pathways. This modeling focuses on two pur-
poses, as mentioned in the previous subsection; finding and keeping core interactions among components and overcoming the unknown and the complexity in multi-enzymatic reactions. The proposed framework is comprised of a rule-based algorithm and an agent-based evolutionary computation. A rule-based algorithm is suitable for representing the core factors specified in the form of reaction rules [28]. A rule-based algorithm will consider various relationships revealed among the involved enzymes and substrates and suggest potential components and interactions in this model. These potential forms will be constructed by the rules made with first principles interaction knowledge; possible biochemical reaction rule between enzymes, information obtained from the analysis of experimental data, and the knowledge gathered from previous literature. Potential components and interactions from rule-based modeling will be considered as agents and relationship between agents, in agent-based algorithms. An agent-based algorithm formulates the mathematical representation of the model by selecting required agents and considering interactions among agents [62]. Through mixing and matching potential forms, the most suitable forms will be selected and will produce a better starting point for model development for multiple enzyme systems. This agent-based modeling uses evolutionary computation algorithm to optimize unknown parameters and structure in the proposed process. Evolutionary computation will predict multi-enzymatic reaction structure and parameters through performing algorithmic recombination, algorithmic mutation, algorithmic selection and fitness functions on the mixing and matching of agents iteratively. The process works with the limited information about individual interactions. Algorithmic recombination and algorithmic mutation will follow possible interaction rules applied in a rule-based algorithm. Configuration fitness and selection is based on analyzing the difference between the simulation results of the proposed model and experimental data related to reaction rate.

The proposed process flowchart for building the model of multiple enzymatic reactions is displayed in Fig. 4.2. This flowchart roughly consists of two algorithms, rule-based modeling and evolutionary computation. We begin the process by collecting experimental data. Every experiment collects two types of data which includes input data, the concentrations of substrates and the amounts of enzymes, and, the related output reaction rate. The collected data could be rearranged to reveal interactions among the involved components effectively, such as enzymes and substrates. Next, the candidates, as the factors comprising the model, are generated by the rules based on experimental data analysis and first principle knowledge, such as chemical laws, information from earlier literature, and kinetic information. An evolutionary computational method then optimizes the structure and the unknown parameters of the multi-enzymatic reaction model. The optimization process uses an objective function which represents the difference between the experimental data and the predicted data.

Fig. 4.3 shows the flows of the modeling process through the block diagram using the example reaction with two enzymes and one substrate. The main input and output variables in
Figure 4.2: Proposed Modeling Process Flowchart with Rule-based Modeling and Evolutionary Computation
the model are the initial enzyme and substrate concentrations and the reaction rate related to the inputs, respectively.

Figure 4.3: Proposed Rule-based Modeling and Evolutionary Computation for Multi-Enzymatic Reaction Modeling Framework

A rule-based algorithm suggests the topology of interactions between numerous enzymes and substrates that influence the reaction rate, such as enzyme-substrate complex or enzyme-enzyme complex, which yields possible components and interactions. The formed components will be agents and the possible interaction will give direction to the perturbation process, including algorithmic recombination and algorithmic mutation. The agent-based evolutionary computation optimizes the actions and interactions of autonomous agents using an evolutionary optimization process. The optimization process in agent-based evolutionary computation performs selection, variation and fitness evaluation which are operators of evolutionary algorithm in this framework iteratively. Algorithmic selection chooses the agents that can participate in the reaction rate. The agents change by algorithmic recombination and algorithmic mutation in variation. Fitness evaluation tests whether the proposed model is in need of change. In this process, if necessarily, we can estimate unknown parameters by diverse estimation methods [31] [45]. Finally, meaningful enzyme components and interactions would be selected and developed for representing the reaction rate.
4.2.3 Rule-Based Modeling

The rule-based algorithm in the proposed modeling process will suggest the topology of interactions between enzymes and substrates that influence the reaction. The functions representing the topology have symbolized expressions for a model structure and mathematical expressions for a model equations.

![Diagram](image)

**First Principal Knowledge**
- Single enzyme kinetic
- Physical and chemical principle
- Experiment and analysis
- Previous literature

**Topology of Potential Interactions**

$$v = \frac{k_{cat}[E][S]}{k_M + [S]}$$

- Mass Action Law
- Total enzyme amount conservation
- Enzyme Complex

$$C_1 \cdot [E_1] = v_1 \quad C_2 \cdot [E_2] = v_2$$

$$f([E_1],[E_2])$$

Figure 4.4: The effects of rule-based modeling

Fig. 4.4 shows the roles of a rule based algorithm for multi-enzymatic reaction modeling. The foundation of this process can be built in various ways. First, we can use diverse expression methods representing enzyme reaction including mass action law. Next, we can be informed through analyzing experimental data sets applying varied inputs. In addition, previous literatures dealing with similar applications can be a source of this process. The important information associated with interactions among involved components in the specific multi-enzymatic reaction is selected, e.g., Michaelis-Menten kinetics representing single enzymatic reaction, kinetic information from experiment, or enzyme complex formation. The information provides the topology of interactions in multi-enzymatic reaction as symbolic expression forms and mathematical expression forms. For example, the product formation of a single enzyme for a specific substrate is represented as the first diagram in the symbolic expression box in Fig. 4.4 and a linear function between the enzyme and the reaction rate mathematically, based on Michaelis-
Menten kinetics. In addition, the enzyme complex formation of two enzymes is represented as the second diagram in the symbolic expression box and as a mathematic function with complicated nonlinear characteristics between two enzymes and the reaction rate. The detailed process of development will be dealt with in Section 4.3.

As we mentioned previously in Section 4.2, the proposed modeling approach attempts to test many configurations using combination of simpler interaction forms and find the most potential configuration. In this process, a rule-based algorithm is an effective algorithm in charge of providing the foundation of basic interactions forms.

### 4.2.4 Evolutionary Computation

Agent-based evolutionary computation has been used often for the modeling of diverse biological systems [54] [49]. The results of this algorithm are composed of the functions of each agent and the relationships between agents. The reason this algorithm is attractive for biological systems is that it represents complex behavior patterns and provides biological information about a real system. In multi-enzymatic reaction modeling, agent-based evolutionary optimization will complete the proposed modeling based on the results of a rule-based algorithm with enzyme inputs.

![Agent-based Evolutionary Computation for Multi-Enzymatic Reactions](image)

Figure 4.5: Agent-based Evolutionary Computation for Multi-Enzymatic Reactions

Fig. 4.5 shows the process of an agent-based evolutionary computation for the proposed
multi-enzymatic reaction modeling. The optimization process of agent-based evolutionary computation is comprised of algorithmic selection, algorithmic recombination, algorithmic mutation and fitness evaluation. Each process works based on given information, which is the reason that it can keep biologically significant issues such as enzyme-enzyme interaction and enzyme-substrate interaction. First, algorithmic selection chooses appropriate components from a pool of possible enzyme and substrate components. Next, algorithmic recombination and algorithmic mutation apply the expected interaction among enzymes to the chosen enzyme components. Algorithmic recombination denotes that enzyme complex is made or divided and mutation reflects inhibition or activation effects among enzymes. A generated multi-enzymatic reaction based on the changed components is tested in the fitness evaluation. Known single enzyme kinetics will be use as a starting point to build the mathematical representation to describe the overall flux response. The simulation results of the generated multi-enzymatic reaction model compared with experimental data about reaction rates. According to the error difference, we evaluate whether the generated multi-enzymatic reaction model is acceptable.

4.3 Model Implementation for Multiple Enzymatic Reactions

4.3.1 Topology Generation based on First Principle Knowledge

Topology of Interactions in Multi-Enzymatic Reactions with Enzyme Independence

The classic form representing single enzymatic reactions is Michaelis-Menten kinetics \[53\] (4.1). This form has focused on representing the relationship between the substrate concentration \([S]\) and the reaction rate \(v\) in metabolic pathway. In (4.1), \(V_{\text{max}}\), the product of \(k_{\text{cat}}\) and \([E]\), and \(K_M\) are main kinetic parameters with constant values. \([E]\) in \(V_{\text{max}}\), the total initial enzyme concentration, is treated as a constant. However, we focus on the relationship between the reaction rate \(v\) and the enzyme concentration \([E]\) instead of \([S]\). Equation (4.1) can be described as linear with respect to change in \([E]\) as shown in (4.2), where \(C_1(S)\) is a function of kinetic constants, such as \(k_M\) and \(k_{\text{cat}}\). Thus for any specific initial substrate concentration \([S]\), \(C_1\) can be considered constant.

\[
\begin{align*}
v &= \frac{V_{\text{max}} \cdot [S]}{K_M + [S]}, & V_{\text{max}} &= k_{\text{cat}} \cdot [E] \\
&= C_1 \cdot [E], & C_1 &= \frac{k_{\text{cat}} \cdot [S]}{K_M + [S]}
\end{align*}
\]

Equation (4.2) can be used as the mathematical description for specific independent reaction

62
rules such as functional redundancy and functional specificity.

Table 4.1: Multi-Enzymatic Reaction Models with the Linear Relationship between the Reaction Rate and the Enzymes

<table>
<thead>
<tr>
<th>Enzyme Interaction</th>
<th>Model Structure</th>
<th>Productive Components</th>
<th>Model Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Functional Redundancy</td>
<td>$E_1 \rightarrow E_2$</td>
<td>$v = v_1 + v_2$</td>
<td>$C_1[E_1] + C_2[E_2]$</td>
</tr>
<tr>
<td>(2) Functional Specificity 1</td>
<td>$E_1 \rightarrow P$</td>
<td>$v = v_1$</td>
<td>$C_1[E_1]$</td>
</tr>
<tr>
<td>(3) Functional Specificity 2</td>
<td>$E_2 \rightarrow P$</td>
<td>$v = v_2$</td>
<td>$C_2[E_2]$</td>
</tr>
</tbody>
</table>

**Topology of Interactions in Multi-Enzymatic Reactions with Enzyme Complex**

Some interactions with network-level enzymatic control, such as producing enzyme complex forms in Fig. 3.17, need other rules supporting nonlinear dependence on changing enzyme concentration. It is very complicated work to derive the mass action equation for the multiple enzymatic reaction associated with the enzyme complex as shown in Chap 3. We revealed that the derivation requests to follow the assumptions, such as total enzyme conservation, quasi-steady state, and chemical equilibrium, and to deal with unknown additory parameters, such as enzyme-enzyme association/disassociation rates and the proportion of each enzyme involved in an enzyme complex in Chap 3. Equation (4.3) shows the equation for product reaction rate in the presence of interacting enzymes derived in [15] and is represented as the sum of the reaction rate for free $E_1$, $v_1$, the reaction rate for free $E_2$, $v_2$, and the reaction rate for the enzyme complex, $v_3$.
\[
    v = v_1 + v_2 + v_3 \\
    = \frac{k_{\text{cat}1} \cdot [E_1] \cdot [S]}{K_{M1} + [S] + m \cdot \frac{K_{M1} \cdot [E_1]^{m-1} \cdot [E_2]^n}{k_1^{m+n-1}} (1 + \frac{[S]}{K_{M3}})} + \frac{k_{\text{cat}2} \cdot [E_2] \cdot [S]}{K_{M2} + [S] + n \cdot \frac{K_{M2} \cdot [E_1]^{m-1} \cdot [E_2]^n}{k_2^{m+n-1}} (1 + \frac{[S]}{K_{M3}})} + \frac{\gamma \cdot k_{\text{cat}3} \cdot [E_1]^m \cdot [E_2]^n \cdot [S]}{k_3^{m+n-1} (K_{M3} + [S])},
\]

where the assumed ratio of \(E_1\) to \(E_2\) in the enzyme complex is \(m:n\), \(k_{\text{cat}1}\), \(k_{\text{cat}2}\), \(K_{M1}\), and \(K_{M2}\) are the measured data including kinetic information of the single enzymatic reaction respectively, \(k_{\text{cat}3}\) and \(K_{M3}\) are median values of two enzyme kinetic constants for general cases in lieu of specific information and assume the basic kinetic characteristics of the produced enzyme complex, \([S]\) is the controllable variable representing a substrate concentration, and \([E_1]\) and \([E_2]\) are also the controllable variables representing the total initial enzyme concentrations respectively. In addition, (4.3) includes some unknown parameters, \(k_1\), \(k_2\), and \(\gamma\). \(k_1\) contains enzyme-enzyme interaction rate and the proportion of \(E_2\) which is involved in enzyme complex. \(k_2\) consists of enzyme-enzyme interaction rate and the proportion of \(E_1\) in enzyme complex. \(k_3\) also reflect the impact of the enzyme interaction rate and the unknown amounts of the enzyme complex. \(\gamma\) is the activation constant and adjusts the productivity level by the enzyme complex. The derivation of (4.3) and the mathematic expressions and meanings of unknown parameters are explained in Appendix B.

Equation (4.3) informs us that the formulation of enzyme complex leads to the nonlinear effects on the reaction rate, depending on the amount of enzymes. The created enzyme complex would not only increase the product, but also inhibit the reactions of two free enzymes. We want to extract the general forms, representing the nonlinear effects by the enzyme complex, from (4.3) and then simplify the forms to focus only on the relationship between the reaction rate and enzyme components.

**Inhibition Effects Associated with Enzyme Complex**  The enzyme complex generated in multi-enzymatic reactions inhibits the reactions of two free enzymes as shown Section 3.4. These inhibition effects result from the change of the denominators of Michaelis-Menten Kinetics for pure free enzymatic reactions, which are represented in the first and second terms of (4.3). My proposed algorithm requires that these characteristics are represented as simpler and more general forms. This approximation process focuses on representing the inhibition effect as the simple multiplication form of the pure free enzymatic reactions.

The first step of this approximation for the inhibition effects of the free \(E_1\) reaction rate is
shown in (4.4)

\[
\frac{[E_1]^m \cdot [E_2]^n}{k_1^{m+n-1}} \approx x_1 \cdot \frac{[E_{c1}]}{[E_{cm}]},
\]

(4.4)

\(k_1\) in the left term in (4.4) includes the unknown proportion of each enzyme involved in the enzyme complex and the unknown equilibrium constant of enzyme-enzyme interaction where the equilibrium constant depends on the enzyme complex amounts. The numerator in the left term in (4.4) shows that \(E_2\) enzyme concentration amount involved in the enzyme complex plays a more prominent role in inhibition effects than \(E_1\). All these characteristics considered, the simplified form of the first step is comprised of three variables, \(x_1\), \([E_{c1}]\) and \([E_{cm}]\). \(x_1\) includes all unknown proportional constants. \([E_{c1}]\) represents the maximum value of the \(E_2\) that can be involved in the enzyme complex, \(\text{min}([E_1], [E_2]/n)\). \([E_{cm}]\) represents the maximum value, \(\text{min}(E_1, E_2)\), that the enzyme complex may have. The same approximation is applied for the reaction rate of the other enzyme, and represented as

\[
\frac{[E_1]^m \cdot [E_2]^{n-1}}{k_2^{m+n-1}} \approx x_2 \cdot \frac{[E_{c2}]}{[E_{cm}]},
\]

(4.5)

\(k_2\) in the left term in (4.5) also includes the unknown proportion of each enzyme involved in the enzyme complex and the unknown equilibrium constant of enzyme-enzyme interaction where the equilibrium constant depends on the enzyme complex amounts. The nominator in the left term in (4.5) shows that \(E_1\) enzyme concentration amount involved in enzyme complex plays a more prominent role in inhibition effects than \(E_2\). Therefore, the simplified form of the first step is comprised of three variables, \(x_2\), \([E_{c2}]\) and \([E_{cm}]\). \(x_2\) also includes unknown proportional constants. \([E_{c2}]\) represents the maximum value of the \(E_2\) that can be involved in the enzyme complex, \(\text{min}([E_1]/m, [E_2])\). \([E_{cm}]\) has the same value in (4.4).

The second step for the simple expression is the approximation with the calibration parameters, \(y_1\) and \(y_2\) [9], which is represented as

\[
k_{M1} \cdot (1 + \frac{[S]}{k_{M3}}) \approx y_1 \cdot (k_{M1} + [S])
\]

(4.6)

and

\[
k_{M2} \cdot (1 + \frac{[S]}{k_{M3}}) \approx y_2 \cdot (k_{M2} + [S]),
\]

(4.7)

where \(y_1\) and \(y_2\) are unknown parameters. These two approximations yield

\[
m \cdot \frac{k_{M1} \cdot [E_1]^m \cdot [E_2]^n}{k_1^{m+n-1}} (1 + \frac{[S]}{k_{M3}}) \approx z_1 \cdot m \cdot (k_{M1} + [S]) \cdot \frac{[E_{c1}]}{[E_{cm}]},
\]

(4.8)
where $z_1$ is the one unknown parameter, the product of $x_1$ and $y_1$, and
\[
\frac{n \cdot k_{M2} \cdot [E_1]^m \cdot [E_2]^n}{k_2^{m+n-1}} \left(1 + \frac{[S]}{k_{M3}}\right) \approx z_2 \cdot n \cdot (k_{M2} + [S]) \frac{[E_2]}{[E_{cm}]} \]
(4.9)
where $z_2$ is the one unknown parameter, the product of $x_2$ and $y_2$.

**Activation Effects Associated with Enzyme Complex**  The activation effect on the total reaction rate is due to the assumption that the enzyme complex may contribute to the product formation. The mathematical expression is represented by the function of the concentration of the enzyme complex including the unknown kinetic parameters related to the enzyme complex. The form, presenting the concentration of the enzyme complex in (4.3), can be approximated by
\[
\frac{[E_1]^m \cdot [E_2]^n}{k_3^{m+n-1}} \approx \delta \cdot [E_{cm}],
\]
(4.10)
The left term in (4.10) is the mathematical expression of the enzyme complex generated by the repeated interaction process based on binding equilibrium. Therefore this form can be represented as the proportion of the maximum value of which the enzyme complex is capable in (4.10) where $[E_{cm}]$ represents the maximum value which the enzyme complex is capable and $\delta$ is an unknown parameter representing the proportion of the enzyme complex in $[E_{cm}]$. The unknown parameter, $z_3$, includes $\gamma$ representing the uncertainty of kinetic constants of the enzyme complex in (4.3) and $\delta$ in (4.10).

**Reaction Rate Expressions for Nonlinear Rules**  All approximation equations from (4.4) to (4.5) helps the complicated mathematical expressions in mechanistic modeling, (B.8), to be the simple forms used in the proposed modeling algorithm. The reaction rate associated with $E1$ including the inhibition effect of the enzyme complex, $v'_1$, can be represented by
\[
\frac{k_{cat1} \cdot [E_1] \cdot [S]}{K_{M1} + [S] + m \cdot \frac{K_{M1} \cdot [E_1] \cdot [E_2]^{n-1}}{k_1^{m+n-1}} \left(1 + \frac{[S]}{K_{M3}}\right)} \approx \frac{C_1 \cdot [E_1]}{1 + z_1 \cdot m \cdot \frac{[E_2]}{[E_{cm}]}}
\]
(4.11)
where $C_1$ is the value of $\frac{k_{cat1} \cdot [S]}{k_{M1} + [S]}$, comprised of the known information of previous knowledge, such as kinetic information of $E1$ and the concentration of substrate. The reaction rate associated with $E2$ including the inhibition effect of the enzyme complex, $v'_2$, can be represented by
\[
\frac{k_{cat2} \cdot [E_2] \cdot [S]}{K_{M2} + [S] + n \cdot \frac{K_{M2} \cdot [E_1] \cdot [E_2]^{n-1}}{k_2^{m+n-1}} \left(1 + \frac{[S]}{K_{M3}}\right)} \approx \frac{C_2 \cdot [E_2]}{1 + z_2 \cdot n \cdot \frac{[E_2]}{[E_{cm}]}}
\]
(4.12)
where $C_2$ is the value of $\frac{k_{cat2}[S]}{k_{M2}+[S]}$, comprised of the known information of previous knowledge, such as kinetic information of $E2$ and the concentration of substrate. The reaction rate associated with the enzyme complex leading to activation effect on the total reaction rate, $v'_3$, can be represented by

$$v'_3 = \frac{z_1 \cdot k_{cat3} \cdot [E1]_t^m [E2]_t^n \cdot [S]}{k_3^{m+n-1} \cdot (k_{M3} + [S])} \approx z_3 \cdot C_3 \cdot [E_{cm}] \tag{4.13}$$

where $C_3$ is the value of $\frac{k_{cat3}[S]}{k_{M3}+[S]}$, comprised of the known information of previous knowledge. When we assume that all enzyme components, the involved free enzymes and the enzyme complex, work for the product in the multiple enzymatic reaction, as (4.3), the total reaction rate can be expressed as:

$$v' = v'_1 + v'_2 + v'_3 = \frac{C_1 \cdot [E1]_t}{1 + z_1 \cdot m \cdot \frac{[E_{cm}]}{[E_{cm}]}} + \frac{C_2 \cdot [E2]_t}{1 + z_2 \cdot n \cdot \frac{[E_{cm}]}{[E_{cm}]}} + z_3 \cdot C_3 \cdot [E_{cm}] \tag{4.14}$$

Each term in (4.14) will be used as the candidates which consist of the model for the multi-enzymatic reactions with specific interactions such as the enzyme complex with the same ratio (Table 4.2 (1)) and the different ratio (Table 4.2 (2)).

The rule based modeling, from the previous knowledge and literature, provides the potential topologies, representing reaction rates catalyzed by diverse enzyme components, such as $v_1$ and $v_2$ in Table 4.1 and $v'_1$, $v'_2$, and $v'_3$ in Table 4.2. Additional first principle knowledge and other important assumptions could modify and extend the topology of interactions generated in this research. Next, the proposed agent based evolutionary optimization process with the given factors would result in a suitable multi-enzymatic reaction model.
4.3.2 Model Optimization based on Genetic Programming and Genetic Algorithm

The optimization algorithm involving evolutionary computing techniques is applied to build the model. The optimization process in this modeling requires not only searching the model structure but also estimating unknown parameters due to newly assumed interactions in the model. Genetic Programming (GP) is one of the evolutionary algorithms, which has the ability to approximate complicated systems with undefined functions as a compact representation and develop a mathematical expression based on input-output data [5]. GP automatically evolves the functional form, that fits a set of experimental data points. The form is traditionally represented as tree structures [48]. The tree consists of terminal nodes and functional nodes. The terminal nodes have operands, such as independent variables or constants, generally. The functional nodes have operators, such as addition, subtraction, multiplication, and division, generally. The tree is optimized using algorithmic methods inspired by biological evolution, such as mutation, recombination, and selection by survival of the fittest as shown in Fig. 4.6. Every iteration in the optimization process for multi-enzymatic reaction modeling accompanies parameter optimization as needed. The Genetic Algorithm (GA) is used as the algorithm for parameter estimation belongs to the field of evolutionary algorithms, using techniques inspired by natural evolution [52] and is used as one of the best global optimization methods [31].

![Figure 4.6: The Optimization Example using Tree Structures of GP](attachment:image)

The optimization process in this multi-enzymatic reaction modeling executes these two steps in order: GP with the operands provided by the topology of interactions associated with enzyme independence and GP with operands generated by the topology of interactions associated with the enzyme complex as shown in Table 4.3.

The first step searches the model structure under the topology that has a linear relationship between each enzyme and the reaction rate. The operands in the step are not single chemical
Table 4.3: Operands and Operators in GP for multi-enzymatic reaction modeling with two enzymes

<table>
<thead>
<tr>
<th>Enzyme Interaction</th>
<th>Operand</th>
<th>Operator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme Independence</td>
<td>$v_1, v_2$</td>
<td>addition (+)</td>
</tr>
<tr>
<td>Enzyme Complex</td>
<td>$v_1', v_2', v_3'$</td>
<td>addition (+)</td>
</tr>
</tbody>
</table>

$v_1$ and $v_2$ are the functional forms based on (4.1) with kinetic information of $E_1$ and $E_2$, respectively. $v_1'$, $v_2'$, and $v_3'$ are the functional forms as (4.11), (4.12), and (4.13) respectively.

components but functional forms, representing the impact each enzyme component on the productive reaction rate. The tree structure of the operand is represented in Fig. 4.7A [18]. The operator in the steps is addition only, because the reaction rate can be represented by the sum of each reaction rate associated with enzyme components considered in multi-enzymatic reactions. Other operators, however, are used as the internal operators in the functional operands representing the nonlinear relationship between substrate, $[S]$, and the reaction rate, $v$. Based on the operands and operators, the initial population is randomly generated. Evolution processes optimize the initial tree structures. Algorithmic mutation brings about the change of initial tree structure. The algorithmic selection method following the survival of fitness replace the generation iteratively, where the fitness is calculated by the mean squared error (MSE), known as one of general estimators of the goodness of fit. In this fitness calculation, unknown parameters used for preventing overfitting of the optimization would require parameter estimation by GA. The second step searches the model structure including the nonlinear relationship between each enzyme and the reaction rate. The operands in the step are the functional forms, representing the impact of each enzyme component on the productive reaction rate. The tree structure of the operand is represented in Fig. 4.7B and C. In Fig. 4.7B, $C_{\{i\}}$ is the same as in Fig. 4.7A, and $f_{\{i\}}(E_{all}, p)$ affirms the denominator of (4.11). In Fig. 4.7C, $C_3$ has the same structure of $C_{\{i\}}$ and includes kinetic parameters associated with enzyme complex, and $E_{cm}$ is the variable related to the amount of enzyme complex, which is defined in Appendix B.

The operator in the steps is also addition, for the same reasons in the first GP process. Other operators, are used to represent the nonlinear relationship between substrate $[S]$ and the reaction rate $v$ and enzyme-enzyme interactions with other operators (Fig. 4.7). Evolution process is the same as the first step of the optimization. Unknown parameters, required for the parameter estimation for the fitness calculation, would mainly reflect the inhibition or activation effects by the enzyme complex. The final models from the first and second optimization steps are evaluated with numerical analysis, such as RMSE, and by visual comparison between simulation results and experimental data in the next subsection. The overall optimization process yields a
A : The tree structure of the functional operands, $v_1$ and $v_2$ based on (4.1), in GP considering enzyme independence
B : The tree structure of the functional operands associated with free enzymes, $v'_1$ as (4.11), $v'_2$ as (4.12), in GP considering enzyme complex
C : The tree structure of the functional operands associated with enzyme complex, $v'_3$ as (4.13), in GP considering enzyme complex

4.3.3 Evaluation using Numerical Methods

The optimized model is evaluated by three different analysis methods: numerical analysis, visual comparison, and biological relevance. First, the proposed modeling process uses three different numerical methods, such as Root Mean Squared Error (RMSE), R-squared ($R^2$), and Bayesian Information Criterion (BIC). The quantitative analysis has the advantage of making possible to compare the fitness of the selected candidate models relatively. Root Mean Squared Error (RMSE) is used to measure the difference between the predicted values and actual values and is represented as (3.5). RMSE is used not only as the objective function for selection strategy in the evolutionary optimization process, but also as the one of the fitness comparison indexes between the optimization steps with different rules. Second, R-squared ($R^2$) demonstrates how well actual values are replicated by the model [24] and is represented as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(v_i - \bar{v})^2}{\sum_{i=1}^{n}(v_i - \bar{v})^2}$$  \hspace{1cm} (4.15)$$

where $\bar{v}$ is the mean value of the measured reaction rates. The closer $R^2$ value is to 1 the better the data fit the curve. Third, Bayesian Information Criterion (BIC) is the method that
reflects the model complexity by using the penalty term for the number of parameters \[67\] and is represented as

\[BIC = n \cdot \ln\left(\frac{\sum_{i=1}^{n} v_i - \hat{v}_i}{n}\right) + k \cdot \ln(n)\]  

(4.16)

where \(k\) is the number of free parameters. The proposed modeling process uses BIC as the index for selecting the model with the least complexity under similar conditions, such as insignificant RMSE values and competent \(R^2\) values.

Visual comparison evaluates the similarities and differences in the graph shapes represented by the model simulation results and experiment data respectively. The visual display evaluation is used to make a decision to proceed with the next optimization step. The visual method can prevent oversimplification or over-fitting, which sometimes happens on the model structure optimization process, focusing only on the numerical evaluation. Otherwise, the visual evaluation helps to expose the characteristics of multi-enzymatic reactions intuitively. Although, we developed the model with the best goodness-of-fit, the model could be in conflict with biological rationale. Even though the components involved in building the model are already satisfied with first principle knowledge, such as physical and chemical theories, we need to validate the biological relevance of the assumptions used for unknown interactions among the components.

4.4 Simulations and Results

4.4.1 Application 1: CAld5H enzymes in Lignin Biosynthesis

Model Optimization Process

The proposed algorithm decides the potential enzyme-enzyme interactions and the mathematical equations representing the effects of the enzyme components on the metabolic reaction using information from CAld5H1 and CAld5H2 kinetics based on Michaelis-Menton kinetics\[80\] and the observed input-output data. My algorithm produces the basic functional forms (operands), proposed in Section 4.3, using CAld5H1 and CAld5H2 enzyme kinetic information. Next, the resultant functional forms (operands) are mixed and matched during the optimization process, which uses the results of the numerical analysis, such as RMSE, \(R^2\) and BIC, for deciding the best of various potential models. The optimization process was performed using the following conditions: the number of generations = 100, population size = 20, the number of parameter optimization runs = 50.

Fig. 4.8 shows the numerical analysis results, RMSE and BIC, of the optimized models in the different enzyme-enzyme interaction conditions, where 'NC' assumes that there is no enzyme complex among the involved enzymes (CAld5H1 and CAld5H2) and the ratio form shown as 'm
Figure 4.8: Numerical analysis (MSE and BIC) in optimization process

A : The reaction with coniferaldehyde
B : The reaction with coniferyl alcohol
C : The reaction with coniferyl alcohol inhibited by coniferaldehyde

$n'$ represents a CAld5H1 : CAld5H2 enzyme amount ratio in the assumed enzyme complex. The optimized model under the enzyme independence (‘NC’) is functional redundancy. The optimized models under the diverse ratios of enzyme complex have an impact of all enzyme components, such as free enzymes and enzyme complex, on the total reaction rate. Other structures, considering enzyme complexes with ratios not included in Fig. 4.8, were also considered in the optimization process, but may not be candidates for the model structure. The smaller values of both RMSE and BIC represent the better model. Similarly, the changes of RMSE and BIC values according to enzyme-enzyme interaction assumptions follow a very similar trend in different substrate conditions. In each graph in Fig. 4.8, all RMSE values show no large difference, it may even be a little bit better to assume the presence of the enzyme complex in A and C graphs. This may be a case where there is an overfitting issue. Therefore, BIC reflecting the complexity of the model indicates that functional redundancy with low complexity, compared to other models with enzyme complexes, represents the CAld5Hs multi-enzymatic reaction model. All $R^2$ values are over 0.97, which means that the experimental data reflection ability is sufficiently satisfied and $R^2$ does not have an effect on the final model selection in this case.

Results

As we see in the model selection process in Fig. 4.8, regardless of the different kinds of substrates (A and B sub-cases) and the presence of an inhibitor (C sub-case), the developed models by the proposed algorithm suggest there is no complex form between CAld5H1 and CAld5H2.
The diagram forms of the models are shown in Table 4.4. In addition, the mathematical expressions of the developed models are generated by the optimized combination of the given operands, where \([E_1]\) and \([E_2]\) represent the enzyme amounts of CAld5H1 and CAld5H2, \(C_1\) and \(C_2\) are the known values comprised of each piece of single enzyme kinetic information, substrate concentration and the mathematical expression in (4.2). In Table 4.4, the unknown parameters \(p\) and \(q\), indicating which enzyme is dominant, are optimized in the modeling process. All variables follow the different conditions given in each sub-case, respectively. Both optimized values of \(p\) and \(q\) are close to 1 in all sub-cases, which means two reactions by each enzyme are not affected by the other enzyme.

Fig. 4.9 shows the simulation results of the optimized model with experimental data in the different conditions respectively (A, B and C). In these figures, x-axis represents the main substrate concentration values and y-axis represents the reaction rate, where the amounts of both enzymes are fixed to 1nM. RMSE and \(R^2\) evaluate the goodness-of-fit numerically. The simulation graphs show how well the simulation results approximate the measured data visually. In other words, the simulation results can provide supportive evidence to suggest that the multiple enzymatic reactions would be a linear combination of each enzymatic reaction represented by Michaelis-Menten kinetics. These results was supported by functional redundancy of CAld5H1 and CAld5H2 enzymatic reactions in Wang et al. (2012) [80].

4.4.2 Application 2: 4CL enzymes in Lignin Biosynthesis

Model Optimization Process

The data collected under the special experiment conditions defined in Section 3.3 can be useful in increasing modeling accuracy. The experiment designed for discovering potential enzyme-enzyme interactions in 4CLs enzymatic reactions measures the rate of product formation when the 4CL3 or 4CL5 molar concentration is held constant at 40nM while the other enzyme molar concentration is gradually changed between 0nM and 40nM, by 10nM. The proposed algorithm is carried out with 4CL3 and 4CL5 enzyme amounts as inputs, and the experiment data measured the reaction rate as outputs. The basic operands produced with linear rules and nonlinear rules are selected via the agent based evolutionary algorithm in the modeling process, which gives us plausible interactions among enzymes and substrates that can describe the data. BIC and RMSE objective metric are used for this search/optimization process.

Fig. 4.10 shows the numerical analysis results (RMSE and BIC) of the optimized models in the different enzyme-enzyme interaction conditions, where 'NC' assumes that there is no interaction among the involved enzymes (4CL3 and 4CL5) and the ratio form displayed as 'm : n' represents a 4CL3: 4CL5 enzyme amount ratio in the assumed enzyme complex. The smaller values of both RMSE and BIC represent the better model. A, C and D graphs in Fig. 4.10 have
Figure 4.9: Simulation Results of CAld5Hs enzymatic reaction models

A : the reaction with coniferaldehyde, B : the reaction with coniferyle alcohol, C : the reaction with coniferyl alcohol inhibited by coniferaldehyde
Table 4.4: The optimized models of the reaction rates for CAld5H1 and CAld5H2

<table>
<thead>
<tr>
<th>Components</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enzyme</strong></td>
<td>CAld5H1/CAld5H2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Substrate</strong></td>
<td>Coniferaldehyde</td>
<td>Coniferyl alcohol</td>
<td>Coniferyl alcohol</td>
</tr>
<tr>
<td><strong>Inhibition</strong></td>
<td>-</td>
<td>-</td>
<td>Coniferaldehyde</td>
</tr>
</tbody>
</table>

**Optimized Models**

<table>
<thead>
<tr>
<th>Enzyme interaction</th>
<th>Equation</th>
<th>Parameter</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$v = p \cdot C_1 [E_1 t]$ + $q \cdot C_2 [E_2 t]$</td>
<td>$p$</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>$v = p \cdot C_1 [E_1 t]$ + $q \cdot C_2 [E_2 t]$</td>
<td>$q$</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>$v = p \cdot C_1 [E_1 t]$</td>
<td></td>
<td>1.05</td>
</tr>
</tbody>
</table>

A: $C_1 = \frac{k_{cat1} \cdot [S]}{K_{M1} + [S]}$ and $C_2 = \frac{k_{cat2} \cdot [S]}{K_{M2} + [S]}$, where $[S]$ is the concentration of Coniferaldehyde, $k_{cat1}$ and $K_{M1}$ are CAld5H1 enzyme kinetic parameters for Coniferaldehyde reaction, and $k_{cat2}$ and $K_{M2}$ are CAld5H2 enzyme kinetic parameters for Coniferaldehyde reaction.

B: $C_1 = \frac{k_{cat1} \cdot [S]}{K_{M1} + [S]}$ and $C_2 = \frac{k_{cat2} \cdot [S]}{K_{M2} + [S]}$, where $[S]$ is the concentration of Coniferyl alcohol, $k_{cat1}$ and $K_{M1}$ are CAld5H1 enzyme kinetic parameters for Coniferyl alcohol reaction, and $k_{cat2}$ and $K_{M2}$ are CAld5H2 enzyme kinetic parameters for Coniferyl alcohol reaction.

C: $C_1 = \frac{k_{cat1} \cdot [S]}{K_{M1} (1 + \frac{[I]}{K_{IC1}}) + [S] (1 + \frac{[I]}{K_{IU1}})}$ and $C_2 = \frac{k_{cat2} \cdot [S]}{K_{M2} (1 + \frac{[I]}{K_{IC2}}) + [S] (1 + \frac{[I]}{K_{IU2}})}$, where $[S]$ is the concentration of Coniferyl alcohol, $[I]$ is the concentration of Coniferaldehyde, $k_{cat1}$, $K_{M1}$, $K_{IC1}$, and $K_{IU1}$ are CAld5H1 enzyme kinetic parameters for Coniferyl alcohol reaction with Coniferaldehyde inhibitor, and $k_{cat2}$, $K_{M2}$, $K_{IC2}$, and $K_{IU2}$ are CAld5H2 enzyme kinetic parameters for Coniferyl alcohol reaction with Coniferaldehyde inhibitor.
Figure 4.10: Numerical analysis (RMSE and BIC) in optimization process

A : The reaction with 4-coumaric acid, B : The reaction with caffeic acid, C : The reaction with 4-coumaric acid inhibited by caffeic acid and feruic acid, D : The reaction with caffeic acid inhibited by 4-coumaric acid, feruic acid
the same trend that the model considering the enzyme complex with the ratio of $m : 1$ ($m > 1$, $m$ is integer) are better than the linear model and the nonlinear model with the assumption of the enzyme complex with the ratio of $1 : n$ ($n > 1$, $n$ is integer). As a result, the model with the enzyme complex of 2 : 1 or 3 : 1 ratio is the best model of each condition (A. 3 : 1 enzyme complex, C. 2 : 1 enzyme complex, and D. 2 : 1 enzyme complex).

Fig. 4.10B indicates that the optimization models, considering 1 : 2, 1 : 1, or 2 : 1 enzyme complex, are better than others. Specifically, the models including enzyme complex with 1 : 2 ratio or 2 : 1 ratio can be considered as the best models. In Fig. 4.10C, the numerical analysis graph has a symmetrical pattern with respect to 1 : 1 ratio, which yields several optima. This phenomenon is due to the similar kinetic constants of the two involved enzymes, unlike other cases related to 4CL enzymes [15]. Because, the proposed algorithm is composed of the optimized model using the topology based on the kinetic information of mainly single enzymatic reactions. These similar activities of two enzymes make it difficult to determine which enzyme is dominant in enzyme complex. In the future, additional first principle knowledge might solve issues like this and improve the modeling accuracy. In the second sub-case in application 2, considering the optimization results of other sub-cases, we use the optimized model with 2 : 1 ratio enzyme complex as the final model.

Results

This algorithm predicted that an enzyme complex exists in 4CL3 and 4CL5 enzymatic reactions. The diagram forms of the models are displayed in Table 4.5. The mathematical equations of the models are comprised of the given operands, where $[E_1]$ and $[E_2]$ represent the enzyme amounts of 4CL3 and 4CL5, $C_1$ and $C_2$ are the known values comprised of each single enzyme kinetic information and substrate concentration, $C_3$ uses the mean values of the two kinetic values, such as $K_M$ values and the $k_{cat}$ values, of 4CL3 and 4CL5 enzymes, $m$ and $n$ are the ratio between the enzyme amounts of 4CL3 and 4CL5 involved in the enzyme complex, $E_{C_1}$, $E_{C_2}$ and $E_{C_3}$ are the function values generated with the total enzyme amounts of 4CL3 and 4CL5 by the proposed rules, and the unknown parameters, such as $z_1$, $z_2$ and $z_3$, are optimized in the modeling process. All variables follow the different conditions given in each sub-case, respectively. The values of $z_1$ and $z_2$ make possible the relative comparison of the inhibition effects of the enzyme complex on the reactions associated with free 4CL3 and 4CL5 enzymes, respectively. For example, in sub-case A, the $z_1$ optimized value, 3.2006, is greater than the optimized $z_2$ value, 0.6401, which suggests that the reaction for free 4CL3 might be more significantly reduced by the enzyme complex, more than the reaction for free 4CL5. On the contrary, in sub-case D, we can infer that the reaction for free 4CL5 might be more significantly reduced by the enzyme complex, than the reaction for free 4CL3. The optimized value of $z_3$
shows the productivity of the potential reaction caused by the enzyme complex. In sub-case A and C with a 4-coumaric acid as a main substrate, the reaction activate effects by the enzyme complex are greater than the effects in sub-case B and D with caffeic acid.

Fig. 4.11 - 4.14 show the simulation results of the optimized models with enzyme complexes of 2 : 1 or 3 : 1 ratio in the different sub-cases respectively. In these figures, x-axis represents the concentration of 4CL3 and 4CL5, and y-axis represents the reaction rate, where the lines have fixed concentrations of substrates, respectively. RMSE and $R^2$ evaluate the goodness-of-fit numerically. The simulation graphs show how well the simulation results approximate the measured data visually. Specifically, each graph has the gray dashed line as a baseline, which is simulated under the assumption that 4CL3 and 4CL5 enzymes work for the reaction independently without any interaction, such as interference or combination. The baselines are used to analyze the characteristics of the experimental data and the predicted modeling simulation line, black in color, represents the lowest and highest substrate values (160uM and 21.26uM in Fig. 4.11, 25uM and 5.93uM in Fig. 4.12, 100uM and 12.80uM in Fig 4.13, and 56.25uM and 23.73uM in Fig. 4.14). These baselines show that the reaction rate is linearly increased by gradually increasing 4CL5 enzyme amount from 0uM to 40uM when 4CL3 enzyme amount is fixed and decreased by gradually decreasing 4CL3 enzyme amount from 40uM to 0uM. On the other hand, the change of experimental data between 40:0 (4CL3:4CL5) case and 40:10 (4CL3:4CL5) case are increased less or more decreased than the baseline and the change of experimental data between 40:10 (4CL3:4CL5) case and 40:40 (4CL3:4CL5) case are increased more than the baseline. The period, from 40:40 (4CL3:4CL5) case to 40:0 (4CL3:4CL5) case, also has a similar change. This phenomenon is well reproduced through the proposed algorithm assuming the nonlinear effects, such as inhibition and activation, by the enzyme complex on the reaction. Other substrate simulations also follow the same trend.

4.4.3 Application 3 : 4CL-HCT enzymes in Lignin Biosynthesis

Model Selection Process

The multiple enzymatic reactions in the applications 1 and 2 have involved the two different enzymes as isoforms, which have different amino sequences but catalyze the same chemical reaction. The modeling for 4CLs-HCTs enzymatic reactions considers the specific characteristics that 4CL enzymes work for the reaction associated with 4-coumaric acid independently but HCT enzymes cannot control the reaction for themselves. The experiment designed for discovering potential enzyme-enzyme interactions in 4CLs-HCTs enzymatic reactions measures the rate of product formation, when the 4CL3 or 4CL5 molar concentration is held constant at 10nM and the HCT1 or HCT6 molar concentration is changed between 0nM and 40nM. The proposed algorithm is carried out with the selected experiment data, and only 4CL3 and 4CL5 enzyme
### Components

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>4CL3/4CL5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate</td>
<td>4-coumaric Acid</td>
</tr>
<tr>
<td>Inhibition</td>
<td>-</td>
</tr>
</tbody>
</table>

### Optimized Models

<table>
<thead>
<tr>
<th>Enzyme interaction</th>
<th>Equation</th>
<th>Parameter</th>
<th>m:n ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>S - E</td>
<td>( v = \frac{C_1 \cdot [E1]}{1 + \frac{S}{K_{M1}}} + \frac{C_2 \cdot [E2]}{1 + \frac{2 \cdot [E2]}{K_{C1}} + \frac{3 \cdot [E2]}{K_{C2}} + \frac{z_3 \cdot [E_{Cm}]}{K_{C1}} + \frac{z_3 \cdot [E_{Cm}]}{K_{C2}} + \frac{z_3 \cdot [E_{Cm}]}{K_{C3}}} )</td>
<td>( z_1 )</td>
<td>3:1</td>
</tr>
<tr>
<td>S - E</td>
<td>( v = \frac{C_1 \cdot [E1]}{1 + \frac{S}{K_{M1}}} + \frac{C_2 \cdot [E2]}{1 + \frac{2 \cdot [E2]}{K_{C1}} + \frac{3 \cdot [E2]}{K_{C2}} + \frac{z_3 \cdot [E_{Cm}]}{K_{C1}} + \frac{z_3 \cdot [E_{Cm}]}{K_{C2}}} )</td>
<td>( z_2 )</td>
<td>2:1</td>
</tr>
<tr>
<td>S - E</td>
<td>( v = \frac{C_1 \cdot [E1]}{1 + \frac{S}{K_{M1}}} + \frac{C_2 \cdot [E2]}{1 + \frac{2 \cdot [E2]}{K_{C1}} + \frac{3 \cdot [E2]}{K_{C2}}} )</td>
<td>( z_3 )</td>
<td>2:1</td>
</tr>
<tr>
<td>S - E</td>
<td>( v = \frac{C_1 \cdot [E1]}{1 + \frac{S}{K_{M1}}} + \frac{C_2 \cdot [E2]}{1 + \frac{2 \cdot [E2]}{K_{C1}}} )</td>
<td>( z_4 )</td>
<td>2:1</td>
</tr>
</tbody>
</table>

### Notes

\( A = \frac{k_{cat1}}{[S]} \cdot \frac{[S]}{K_{M1} + [S]} \) and \( C_3 = k_{cat3} \cdot [S] \), where \([S]\) is the concentration of 4-coumaric Acid, \( k_{cat1} \) and \( K_{M1} \) are 4CL3 enzyme kinetic parameters for 4-coumaric Acid reaction, and \( k_{cat3} \) is \( (k_{cat1} + k_{cat2})/2 \)

\( B = \frac{k_{cat1}}{K_{M1} + [S]} \) and \( C_2 = \frac{k_{cat2}}{K_{M2} + [S]} \), where \([S]\) is the concentration of Caffeic Acid, \( k_{cat1} \) and \( K_{M1} \) are 4CL3 enzyme kinetic parameters for Caffeic Acid reaction, and \( k_{cat2} \) is \( (k_{cat1} + k_{cat2})/2 \)

\( C = \frac{k_{cat1}}{K_{M1} + [S]} \) and \( C_2 = \frac{k_{cat2}}{K_{M2} + [S]} \), where \([S]\) is the concentration of 4-coumaric Acid, \([I]\) is the concentration of Feruic Acid, \( k_{cat1} \), \( K_{M1} \), \( K_{IC1} \), and \( K_{IC12} \) are 4CL3 enzyme kinetic parameters for 4-coumaric Acid reaction with Caffeic Acid and Feruic Acid inhibitors, \( k_{cat2} \), \( K_{M2} \), \( K_{IC21} \), \( K_{IC22} \), \( K_{IU21} \), and \( K_{IU22} \) are 4CL5 enzyme kinetic parameters for 4-coumaric Acid reaction with Caffeic Acid and Feruic Acid inhibitors, and \( k_{cat3} \) is \( (k_{cat1} + k_{cat2})/2 \)

\( D = \frac{k_{cat1}}{K_{M1} + [S]} \) and \( C_2 = \frac{k_{cat2}}{K_{M2} + [S]} \), where \([S]\) is the concentration of Caffeic Acid, \([I]\) is the concentration of 4-coumaric Acid, \( k_{cat1} \), \( K_{M1} \), \( K_{IC1} \), and \( K_{IC12} \) are 4CL3 enzyme kinetic parameters for Caffeic Acid reaction with 4-coumaric Acid and Feruic Acid inhibitors, \( k_{cat2} \), \( K_{M2} \), \( K_{IC21} \), \( K_{IC22} \), \( K_{IU21} \), and \( K_{IU22} \) are 4CL5 enzyme kinetic parameters for Caffeic Acid reaction with 4-coumaric Acid and Feruic Acid inhibitors, and \( k_{cat3} \) is \( (k_{cat1} + k_{cat2})/2 \)
Figure 4.11: Simulation Results of the reaction with 4-coumaric acid (A)

Figure 4.12: Simulation Results of the reaction with caffeic acid (B)
Figure 4.13: Simulation Results of the reaction with 4-coumaric acid inhibited by caffeic acid and feruic acid (C)

Figure 4.14: Simulation Results of the reaction with caffeic acid inhibited by 4-coumaric-acid and feruic acid (D)
kinetics in Table 4.6 - single enzymatic kinetic information related to HCT1 and HCT6 do not exist. The basic operands produced with linear rules and nonlinear rules are mixed and matched in the modeling process, which gives us potential enzyme-enzyme interactions and potential models. The mixed and matched process depends on numerical analyses such as RMSE and BIC for the optimized model decision.

Table 4.6: Michaelis-Menten Kinetics of 4CL3 and 4CL5 using 4-coumaric Acid as a Substrate

<table>
<thead>
<tr>
<th></th>
<th>4CL3</th>
<th>4CL5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_M$</td>
<td>6.44 uM</td>
<td>146.06 uM</td>
</tr>
<tr>
<td>$k_{cat}$</td>
<td>301.3575 min$^{-1}$</td>
<td>106.83 min$^{-1}$</td>
</tr>
</tbody>
</table>

Fig. 4.15 shows the numeral analysis results (RMSE and BIC) of the optimized models in the different enzyme-enzyme interaction conditions, where ‘NC’ assumes that there is no interaction among the involved enzymes (4CL3 and 4CL5) and the ratio form displayed as ‘m : n’ represents a 4CL3: 4CL5 enzyme amount ratio in the assumed enzyme complex. The smaller values of both RMSE and BIC represent the better model.

All graphs in Fig. 4.15 have the same trend that the model considering the enzyme complex with the ratio of 1 : n ($n > 1$, $n$ is integer) are better than the linear model and the nonlinear model with the assumption of the enzyme complex with the ratio of $m : 1$ ($m > 1$, $m$ is integer). RMSE results according to the ratio of the enzyme complex represent that the nonlinear models including enzyme complex with the ratio of 1 : n are not at all improved by the change of ‘n’ value at all. On the other hand, the nonlinear models including the enzyme complex with the ratio of $m : 1$ have different values. Specifically, the model with the enzyme complex of 1 : 2 or 1 : 3 ratio is the best model of each condition (A, B, and D. 1 : 2 enzyme complex, C. 1 : 3 enzyme complex). BIC results considering the complexity of the model amplify the analysis of RMSE.

Results

The developed models by the proposed algorithm assume that one of the 4CL enzymes and one of the HCT enzymes make an enzyme complex. The diagram forms of the models are displayed in Table 4.7. The mathematical equations of the models are comprised of the given operands, where $[E_1]$ and $[E_2]$ represent the enzyme amounts of 4CL3 or 4CL5, and HCT1 or HCT6, $C_1$ and $C_2$ are the known values comprised of single enzyme kinetic information and substrate concentration respectively, $C_3$ uses the half values of the kinetic values, such as $K_M$ values.
Figure 4.15: Numerical analysis (RMSE and BIC) in optimization process

A : The reaction for 4CL3 and HCT1, B : The reaction for 4CL3 and HCT6, C : The reaction for 4CL5 and HCT1, D : The reaction for 4CL5 and HCT6
Table 4.7: The optimized models of the reaction rates for 4CLs and HCTs

<table>
<thead>
<tr>
<th>Components</th>
<th>Enzyme</th>
<th>Enzyme</th>
<th>Enzyme</th>
<th>Enzyme</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4CL3/</td>
<td>4CL3/</td>
<td>4CL5/</td>
<td>4CL5</td>
</tr>
<tr>
<td></td>
<td>HCT1</td>
<td>HCT6</td>
<td>HCT1</td>
<td>HCT6</td>
</tr>
<tr>
<td>Substrate</td>
<td>4-coumaric Acid</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Inhibition</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Optimization

<table>
<thead>
<tr>
<th>Step</th>
<th>Nonlinear</th>
<th>Nonlinear</th>
<th>Nonlinear</th>
<th>Nonlinear</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme interaction</td>
<td><img src="image" alt="Diagram" /></td>
<td><img src="image" alt="Diagram" /></td>
<td><img src="image" alt="Diagram" /></td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Equation</td>
<td>( v = \frac{C_1[E_1]}{1 + \frac{z_1 \cdot [E_{Cm}]}{K_{M1}}} + z_3 \cdot \frac{C_3 \cdot [E_{Cm}]}{[S]} )</td>
<td>( v = \frac{C_1[E_1]}{1 + \frac{z_1 \cdot [E_{Cm}]}{K_{M1}}} + z_3 \cdot \frac{C_3 \cdot [E_{Cm}]}{[S]} )</td>
<td>( v = \frac{C_1[E_1]}{1 + \frac{z_1 \cdot [E_{Cm}]}{K_{M1}}} + z_3 \cdot \frac{C_3 \cdot [E_{Cm}]}{[S]} )</td>
<td>( v = \frac{C_1[E_1]}{1 + \frac{z_1 \cdot [E_{Cm}]}{K_{M1}}} + z_3 \cdot \frac{C_3 \cdot [E_{Cm}]}{[S]} )</td>
</tr>
<tr>
<td>Parameter</td>
<td>( z_1 )</td>
<td>0.1219</td>
<td>0.1087</td>
<td>0.0528</td>
</tr>
<tr>
<td></td>
<td>( z_3 )</td>
<td>0.5390</td>
<td>0.4315</td>
<td>0.2706</td>
</tr>
<tr>
<td>ratio m:n</td>
<td>1:2</td>
<td>1:2</td>
<td>1:3</td>
<td>1:2</td>
</tr>
</tbody>
</table>

A and B: \( C_1 = \frac{k_{cat1} \cdot [S]}{K_{M1} + [S]} \) and \( C_3 = k_{cat1} \cdot [S] \) where \([S]\) is the concentration of 4-coumaric Acid, and \( k_{cat1} \) and \( K_{M1} \) are 4CL3 enzyme kinetic parameters for 4-coumaric Acid reaction

C and D: \( C_1 = \frac{k_{cat1} \cdot [S]}{K_{M1} + [S]} \) and \( C_3 = k_{cat1} \cdot [S] \) where \([S]\) is the concentration of 4-coumaric Acid, and \( k_{cat1} \) and \( K_{M1} \) are 4CL5 enzyme kinetic parameters for 4-coumaric Acid reaction

and the \( k_{cat} \) values, of 4CL enzymes, \( m \) and \( n \) are the ratio between the enzyme amounts of 4CL3 or 4CL5 and HCT1 or HCT6 involved in the enzyme complex, \( E_{C1}, E_{C2} \) and \( E_{C3} \) are the function values generated with the total enzyme amounts of the involved enzymes by the proposed rules, and the unknown parameters, such as \( z_1 \) and \( z_3 \), are optimized in the modeling process. All variables follow the different conditions given in each sub-case, respectively. The values of \( z_1 \) and \( z_3 \) make it possible to relatively estimate the inhibition and activation effects of the enzyme complex on the reactions associated with free 4CL3 or 4CL5 enzymes. First, through the optimization parameter values in sub-case A with HCT1 enzyme and sub-case B with HCT6 enzyme, we can compare the effects of the type of the added HCT enzymes on the reaction catalyzed by 4CL3 enzyme under the same conditions in other experiments. The \( z_1 \) value in sub-case A, 0.1219, is greater than the \( z_1 \) value in sub-case B, 0.1087, which implies that the reaction catalyzed by 4CL3 would be reduced more by the enzyme complex formed by
4CL3 and HCT1, than by the enzyme complex formed by 4CL3 and HCT6. In addition, the $z_3$ value in sub-case A, 0.5390, is greater than the $z_3$ value in sub-case B, 0.4315, which suggests that the reaction for 4CL3 would be increased more by the enzyme complex formed by 4CL3 and HCT1, than by HCT6, according to increasing the involved amount of HCT enzyme in the reaction. These predictions from the optimization values in the developed model are explained in detail in Fig.s 4.16 and 4.17.

Second, through the optimization parameter values in sub-case C with HCT1 enzyme and sub-case D with HCT6 enzyme, we analyze the effects of the type of the added HCT enzymes on the reaction catalyzed by 4CL5 enzyme. The $z_1$ value in sub-case C, 0.0528, is less than the $z_1$ value in sub-case D, 0.2127, which suggests that the reaction catalyzed by 4CL5 would be more significantly reduced by the enzyme complex formed by 4CL5 and HCT6, than by the enzyme complex formed by 4CL5 and HCT1. In addition, the $z_3$ value in sub-case A, 0.2706, is less than the $z_3$ value in sub-case B, 0.4073, which implies that the reaction for 4CL3 would be increased more by the enzyme complex formed by 4CL5 and HCT6, than by HCT1, according to the increase of the involved amount of HCT enzyme in the reaction. These predictions from the optimization values in the developed model can be explained in detail in Fig.s 4.18 and 4.19.

Fig. 4.16 - 4.19 show the goodness-of-fit of the optimized model with experimental data under the respective different conditions. In the figures, x-axis represents the concentration of HCT1 or HCT6 enzyme, and y-axis represents the reaction rate, where the lines have the fixed concentrations of substrates of 50 $\mu$M. RMSE and $R^2$ evaluate the goodness-of-fit numerically. The simulation graphs show how well the simulation results visually approximate the measured data. Specifically, each graph has the gray dashed line as a baseline, which is simulated under the assumption that the involved 4CL enzyme and HCT enzyme in each reaction work independently with no interaction, such as interference or combination. The baseline is used to analyze the characteristics of the experimental data and the predicted modeling simulation line. The baseline shows that the reaction rate is unchanged when HCT enzyme amounts gradually increase from 0 $\mu$M to 40 $\mu$M and 4CL enzyme amount is fixed. On the other hand, in the experiment, the measured reaction rates by the increase of HCT enzyme concentration amounts have diverse changes, which appear on all sub-cases (Fig 4.16 - 4.19), even though the characteristics of change are somewhat different than each other.

In Fig. 4.16, the characteristics of the change of the reaction rates by the HCT1 enzyme concentration amounts can be roughly divided into three parts. First, when 4CL3 enzyme concentration amount is fixed as 10 $nM$, the experimental data point associated with HCT1 enzyme amount, 2.5 $nM$, is located under the baseline, which means that small HCT1 enzyme amounts cause the reaction rate reduction. Next, according to increasing the HCT1 enzyme amounts above 5 $nM$ in the same 4CL3 condition, the experimental data points are not only
located the baseline but also increased gradually by the increasing amount of HCT1 enzyme. The experimental points after the HCT1 enzyme concentration amounts of 20 nM, however, are not increased any more, but have some limits. The characteristics may suggest that some interactions between 4CL3 and HCT1 enzymes would exist and the interactions would generate the inhibition effects in the beginning term, the activation effects in the middle term, and the activation would be influenced by certain conditions, even though HCT1 enzyme amounts are continuously increased, in the end term. Fig. 4.17, also show that the characteristics of the experimental data for 4CL3-HCT6 enzymatic reactions are similar to the 4CL3-HCT1 enzymatic reaction characteristics, although the degree of rate rise and decline of HCT6 enzyme effects on the 4CL3 enzymatic reaction is slightly less than that of HCT1 enzyme effects. These phenomena of both sub-cases A and B are well reproduced through the proposed algorithm assuming the nonlinear effects by the enzyme complex on the reaction in Fig. 4.16 and Fig. 4.17. In addition, the optimization parameters, $z_1$ and $z_3$, in the developed model are well used as indicators representing the characteristics of these multiple enzymatic reactions, when we compare two sub-cases A and B in application 3.

The measured data of 4CL5-HCT1 and 4CL5-HCT6 multiple enzymatic reactions also have similar aspects to the characteristics, shown in 4CL3-HCT1 and 4CL3-HCT6 multiple enzymatic reaction data, i.e., that the reaction rate is first inhibited and then activated, according to the increase of the HCT1 enzyme concentration amounts. But there are two appreciable differences between the reactions related to 4CL3 enzyme and 4CL5 enzyme. First, the 4CL5 enzymatic reactions, unlike 4CL3 enzyme, are influenced more by the HCT6 enzyme than by the HCT1 enzyme. In Fig 4.18, when small HCT1 enzyme amounts are added in the 4CL5 enzymatic reaction, the experimental data points associated with HCT1 enzyme amount, from 0 nM to 10 nM, are little changed and stayed around the baseline. According to increasing the HCT1 enzyme amounts above 10 nM, the experimental data points are increased gradually but slowly, when compared with the increase slope in the period from 10 nM to 20 nM of HCT6 enzyme amounts in Fig 4.19. Second, unlike other multiple enzyme combinations, the end part in Fig. 4.18, from 20 nM to 40 nM of HCT1 enzyme, show that there is still the increase of the reaction rate, even though the slope of the increase is reduced. The proposed algorithm implements the potential model, representing this characteristic, as giving variety to the ratio of 4CL5 and HCT1 enzyme concentration amounts involved in the enzyme complex. Finally, the developed model uses the ratio of 1 : 3 instead of 1 : 2, used in other sub-cases A, B, and D, in the enzyme complex. These phenomena of both sub-cases C and D are well reproduced through the model simulation generated the proposed algorithm in Fig. 4.18 and Fig. 4.19. In addition, the optimization parameters, $z_1$ and $z_3$, in the developed model are well used as indicators representing the characteristics of these multiple enzymatic reactions, when we compare two sub-cases C and D in application 3. These results can be used to guide future experiments to
verify the results.

![Graph](image)

Figure 4.16: Simulation Results of the reaction for 4CL3 and HCT1 (A)

### 4.5 Chapter Summary

In this research, we develop the new approach, using a rule-based algorithm and evolutionary computation methods, for modeling multiple enzymatic reactions in metabolic pathways. The developed modeling approach improves the modeling process with the complexity and the uncertainty issues due to potential enzyme interactions, such as enzyme independent and enzyme complex. More specifically, the rule-based modeling method provided diverse potential components and interactions in multiple enzymatic reactions, using first principle knowledge. Genetic programming, as one of evolutionary computational methods, revealed the model comprised of the optimized combination of the given components and interactions. This approach is a tool that can be used in the early stages of modeling.

In this chapter, the developed approach is applied to the core multiple enzymatic reactions in lignin biosynthesis, which discovers the advantages of the proposed modeling. First, the developed modeling method provides not only the optimized enzyme-enzyme interaction considering information but also overall relationship among the involved components, in each
Figure 4.17: Simulation Results of the reaction with 4CL3 and HCT6 (B)

Figure 4.18: Simulation Results of the reaction with 4CL5 and HCT1 (C)
Figure 4.19: Simulation Results of the reaction with 4CL5 and HCT6 (D)

application. Furthermore, the mathematical equation from the optimized model structure represents experimental data in simulation. These advantages help the biologist to infer the potential mechanisms in uncovered biological systems.

In CAld5H1 and CAld5H2 enzymatic reaction modeling, the proposed modeling approach predicted that two enzymes are a functional redundancy and provided mathematical equations representing these reactions. Biological relevance of the results was verified by experimental techniques in Wang et al. (2012) [80]. In 4CL3 and 4CL5 enzymatic reaction modeling, the proposed modeling approach predicted that two enzymes generated an enzyme complex, with the ratio of 2 : 1 or 3 : 1, and provided mathematical equations representing these reactions. Biological relevance of the results was verified by experimental techniques in Chen et al. (2013) and Chen et al. (2014) [16] [15]. In multi-enzymatic reaction modeling with 4CL enzymes and HCT enzymes, the modeling approach predicted that two enzymes generate an enzyme complex, with the ratio of 1 : 2 or 1 : 3, and free 4CL enzyme and the enzyme complex contribute to the rate of product formation. This would, in effect, help biologists to guide the subsequent experiments needed to verify any relationships predicted by this algorithm.
Chapter 5

Conclusions and Future Work

5.1 Conclusions

We proposed mechanistic modeling approaches for multi-enzymatic reactions in metabolic pathways. The proposed modeling approaches predicted the interactions of multiple enzymes and assessed their influence on the overall reaction rates. The results of these modeling approaches provided us with model structures and mathematical expressions that reveal mechanisms of multi-enzymatic reactions. The simulation using the developed models reproduced the experimental data and predicted the influence of unknown internal components on the reaction rate, which is generally not accessible with typical experimental technology. The developed approaches were applied to the core multi-enzymatic reactions in lignin biosynthesis, which showed the advantages of the proposed modeling.

The first approach in Chapter 3 is a mechanistic modeling approach based on mass action law. This approach includes model optimization process with the specific steps for predicting and representing the unknown inner components and interactions in multi-enzymatic reactions. This optimization process was carried out through analysis and assumption based on intimate knowledge of chemistry. The developed modeling framework was applied to two multi-enzymatic reactions with different characteristics, CAld5Hs enzymatic reactions with functional redundancy and 4CLs enzymatic reactions with enzyme complex, in lignin biosynthesis. The modeling process gave us a plausible mechanistic enzyme-enzyme interaction, the mathematical models based on the interaction, reproduction and prediction of experimental reaction data. In addition, simulation using the developed model showed the capability to analyze the contribution of individual enzyme components to the total reaction. The proposed modeling framework provides a useful process for mechanistic models of complex biological systems with unknown structures and parameters. Nevertheless, it is a time-consuming process, which is problem typical of mechanistic modeling and there are still issues to be overcome.
The second approach in Chapter 4 is the mechanistic modeling approach based on computational methodologies, such as a rule-based algorithm and evolutionary computation methods. The rule-based algorithm provided diverse potential components and interactions in multi-enzymatic reactions, using first principle knowledge. Genetic programming, as one of the evolutionary computational methods, revealed a model comprised of the optimized combination of the given components and interactions. These algorithms contributed to general insights of the mechanism of multi-enzyme reaction and to decreasing model development time. This modeling approach represented, as a simple structure, the key relationships among the enzymes associated with the reaction rate, and, at the same time, corresponding mathematical equations, without any specific process. This approach can be used at the early stages of detailed mechanistic modeling. In addition, the modeling results can point to the directions for future experiment. This modeling approach was applied to three multi-enzymatic reactions with different characteristics in lignin biosynthesis, i.e., CAld5Hs enzymatic reactions with functional redundancy, 4CLs enzymatic reactions with enzyme complex and 4CLs/HCTs enzymatic reaction with an enzyme complex. The developed models and their simulation results identified potential enzyme-enzyme interactions and the important influence of the identified components on the overall reaction rate. In the future, additional first principle knowledge will enable improvement in the modeling accuracy and will extend the area of application.

A mechanistic model for multi-enzymatic reactions leads to more comprehensive analysis of biosynthesis and makes possible the use of transgenic perturbation strategies to control each reaction rate for the product. I claim that the mechanistic modeling approaches proposed in this dissertation can reveal the unknown interaction mechanism in multi-enzymatic reactions and provide appropriate mechanistic models. The computational methods, which have generally been used in engineering, led to the success of these approaches. The research perspective and its results, shown in this dissertation, could be an attractive topic in systems biology.

5.2 Future Work

We expect that the modeling approach using a rule-based algorithm and evolutionary computation can have improved model accuracy through diverse methods and information. Additional first principle knowledge about chemical reactions can generate improved rules, representing the topology of interactions among involved components, in a rule-based algorithm. It would be an important research topic how we can present newly found biological phenomena as effective mathematical expressions for a rule-based algorithm.

In addition, the abilities of the modeling approach based on these computational methods, managing complexity and providing flexibility, would enable to expand the size of applicable multi-enzymatic reactions, such as reactions with more than two enzymes, furthermore, to
model biochemical pathways considering multi-level/multi-scale.
REFERENCES


APPENDICES
Appendix A

Derivation of Model Equations based on Mass Action Law

A.1 Enzymatic Reaction Equations of the Interaction of 4CL3 and 4CL5 using 4-coumaric Acid or Caffeic Acid as a Substrate

A.1.1 Assumptions

The enzymatic reactions via Michaelis-Menten kinetics are constructed based on quasi-equilibrium assumptions and the conservation of the total enzyme [26]. Terms, with *, representing the inclusion of self-inhibition in each equation are applied for the model utilizing caffeic acid as the main substrate only.

**Quasi-equilibrium Assumptions**

Quasi-equilibrium assumes that the association and disassociation between the components related to enzymes and substrates are in binding equilibrium. Enzyme-substrate complexes are defined as (3.8) and (3.9). The complex \([E2(S)2]\) considering self-inhibition is written as

\[
[E2(S)2]^* = \frac{1}{k_{IS}}[E2S][S] = \frac{1}{k_{M2} \cdot k_{IS}}[E2][S]^2,
\]

where \(k_{IS}\) is the self-inhibition constant. The enzyme-complexes and their intermediates are represented as

\[
[E1E2] = \frac{1}{k}[E1][E2]
\]
\[ ((E_1)_2E_2) = \frac{1}{k^2} [E_1E_2][E_1] = \frac{1}{k^2} [E_1]^2 [E_2] \quad (A.3) \]

\[ ((E_1)_3E_2) = \frac{1}{k} [(E_1)_2E_2][E_1] = \frac{1}{k^3} [E_1]^3 [E_2], \quad (A.4) \]

where \( k \) is the interaction equilibrium constant between 4CL3 and 4CL5, which is defined as: dissociation constant \( (k_d) \) / association constant \( (k_a) \) of two enzymes.

The binding between enzyme-complexes and substrates yields

\[ [(E_1)_3E_2S] = \frac{1}{k_{M_2}} [(E_1)_3E_2][S] = \frac{1}{k^3 \cdot k_{M_2}} [E_1]^3 [E_2][S], \quad (A.5) \]

\[ [(E_1)_3E_2S_2]^* = \frac{1}{k_{IS}} [(E_1)_3E_2S][S] = \frac{1}{k^3 \cdot k_{M_2} \cdot k_{IS}} [E_1]^3 [E_2][S]^2, \quad (A.6) \]

**Conservation of Total Enzyme**

Dimers and trimers were not included in the derivation of (S9) and (S10) because they were undetected for both recombinant proteins and in SDX when both 4CL3 and 4CL5 are present. This model was built under the simplifying assumption that dimer and trimer formation are transient in the reversible reaction.

The total \((t)\) 4CL3 enzyme concentration is equal to the sum of species with 4CL3:

\[ [E_{1t}] = [E_1] + [E_1S] + 3 \cdot [(E_1)_3E_2] + 3 \cdot [(E_1)_3E_2S] + 3 \cdot [(E_1)_3E_2(S)_2]^* \quad (A.7) \]

The total \((t)\) 4CL5 enzyme concentration is equal to the sum of species with 4CL5:

\[ [E_{2t}] = [E_1] + [E_2S] + [E_2(S)_2]^* + [(E_1)_3E_2] + [(E_1)_3E_2S] + [(E_1)_3E_2(S)_2]^* \quad (A.8) \]

**A.1.2 Derivation of Rate Equation**

**Definition of Unknown Parameters**: \( k_1, k_2 \) and \( \gamma \)

- \( k_1 \) definition

We assume that each enzyme component, involved in enzyme-envelope interactions, can be represented by constant proportions of total enzyme concentrations of 4CL3 or 4CL5: \([E_1] = \alpha \cdot [E_{1t}]\), \([E_2] = \beta \cdot [E_{2t}]\), where \( \alpha \) and \( \beta \) are constants between 0 and 1. These proportions
are constant under the simplifying assumption that other components, such as substrates, and enzyme-enzyme intermediates do not significantly impact the proportions of 4CL3 and 4CL5 in the tetrameric complex. These conditions yield

\[
\frac{1}{k_3} [E_1]^2 [E_2] = \frac{\alpha^2 \beta}{k_3^3} [E_1t]^2 [E_2t] = \frac{1}{(k_1)^3} [E_1t]^2 [E_2t], \quad 0 < \alpha^2 \beta < 1 \quad (A.9)
\]

where \(k\) is the unknown equilibrium constant between 4CL3 and 4CL5, \(\alpha\) is the proportion of 4CL3 in each enzyme-enzyme binding in terms of the total 4CL3 enzyme concentration, and \(\beta\) is the proportion of 4CL5 in each enzyme-enzyme binding in terms of the total 4CL5 enzyme concentration. \(k_1\) is a new constant involving the effects of \(\alpha\), \(\beta\), and \(k\) on the rate equation. It is defined as

\[
\frac{1}{k_1^3} = \frac{\alpha^2 \beta}{k^3} \quad (A.10)
\]

- \(k_2\) definition

The enzyme concentration of 4CL3 involved in each enzyme-enzyme binding can be represented as the constant proportion of total enzyme concentrations of 4CL3: \([E_1] = \alpha * [E_1t]\), based on the conditions mentioned as above in \(k_1\) definition. These conditions yield

\[
\frac{1}{k_3} [E_1]^3 = \frac{\alpha}{k_3^3} [E_1t]^3 = \frac{1}{(k_2)^3} [E_1t]^2, \quad 0 < \alpha^3 < 1 \quad (A.11)
\]

where the definitions of \(k\) and \(\alpha\) are the same as the definitions of those constants in the \(k_1\) definition. \(k_2\) is a new constant involving the effects of \(k\) and \(\alpha\) on the rate equation. It is defined as

\[
\frac{1}{k_2^3} = \frac{\alpha^3}{k^3} \quad (A.12)
\]

- \(\gamma\) definition

\(\gamma\) represents the reactions rate constant for the enzyme-complex.

\([E_1]\) Derivation

Equation (3.8), (A.4), (A.5), and (A.6) are substituted in (A.7), which is then written as:

\[
[E_1t] = [E_1] + \frac{1}{k_{M1}} [E_1][S] + \frac{3}{k_3^3} [E_1]^3 [E_2] + \frac{3}{k_3^3 \cdot k_{M2}} [E_1]^3 [E_2][S] + \frac{3}{k_3^3 \cdot k_{M2} \cdot k_{IS}} [E_1]^3 [E_2][S]^2 \quad (A.13)
\]
This yields

$$[E_1] = \frac{[E_1 t]}{(1 + \frac{[S]}{k_{M1}} + \frac{3}{k^3}[E_1]^2[E_2] \cdot (1 + \frac{[S]}{k_{M2}} + \frac{[S]^2}{k_{M2}^2 k_{IS}}) \cdot (1 + (\frac{[E_1]}{k})^3}}$$  \hspace{1cm} (A.14)$$

Using (A.9), The expression of $[E_1]$ is written as:

$$[E_1] = \frac{[E_1 t]}{(1 + \frac{[S]}{k_{M1}} + \frac{3}{k^3}[E_1]^2[E_2] \cdot (1 + \frac{[S]}{k_{M2}} + \frac{[S]^2}{k_{M2}^2 k_{IS}}) \cdot (1 + (\frac{[E_1]}{k})^3}}$$  \hspace{1cm} (A.15)$$

$[E_2]$ Derivation

Equation (3.9), (A.1), (A.4), (A.5), and (A.6) are substituted in (A.8), which is then written as

$$[E_2t] = [E_2] + \frac{1}{k_{M2}} [E_2][S] + \frac{1}{k_{M2} \cdot k_{IS}} [E_2][S]^2 \cdot \frac{1}{k^3}[E_1]^3[E_2] + \frac{3}{k^3 \cdot k_{M2}} [E_1]^3[E_2][S]$$
$$+ \frac{3}{k^3 \cdot k_{M2} \cdot k_{IS}} [E_1]^3[E_2][S]^2 \cdot (\frac{[E_1]}{k})^3$$  \hspace{1cm} (A.16)$$

This yields

$$[E_2] = \frac{[E_2t]}{(1 + \frac{[S]}{k_{M2}} + \frac{[S]^2}{k_{M2}^2 k_{IS}} \cdot (1 + (\frac{[E_1]}{k})^3)}$$  \hspace{1cm} (A.17)$$

Using (A.11), The expression of $[E_2]$ is written as

$$[E_2] = \frac{[E_2t]}{(1 + \frac{[S]}{k_{M2}} + \frac{[S]^2}{k_{M2}^2 k_{IS}} \cdot (1 + (\frac{[E_1]}{k})^3)}$$  \hspace{1cm} (A.18)$$

Rate Equation

In our models, the product is formed by three enzymatic reactions: 1) free 4CL3, 2) free 4CL5, and 3) the 4CL3-4CL5 complex. Dimer and trimer formation are transient in the reversible reaction and do not contribute significantly to product formation. The total rate equation is represented by the sum of the reactions of enzyme-substrate by free 4CL3, free 4CL5, and the reaction of enzyme-complex-substrate by 4CL3-4CL5, which is written as

$$v = k_{cat1}[E_1S] + k_{cat2}[E_2S] + \gamma \cdot k_{cat2}([E_1]_3[E_2S]).$$  \hspace{1cm} (A.19)$$

Equation (A.19) can be changed using (3.8), (3.9) and (A.5) to
\[
v = \frac{k_{\text{cat}1}}{k_{M1}}[E1][S] + \frac{k_{\text{cat}2}}{k_{M2}}[E2][S] + \frac{\gamma \cdot k_{\text{cat}2}}{k^3 \cdot k_{M2}}[E1]^3[E2][S]. \tag{A.20}
\]

Considering the derived expressions of \([E1]\) and \([E2]\) and after some algebraic manipulation (See (A.15) and (A.18) above), the final rate equation is

\[
v = \frac{k_{\text{cat}1}[E1t][S]}{k_{M1} + [S] + \frac{k_{\text{cat}1}[E1]^2[E2t]}{k^1_{I1}}(1 + \frac{[S]}{k_{M2}^2k_{I2}^*})} + \frac{k_{\text{cat}1}[E2t][S](1 + \gamma \cdot \frac{(E1t)^3}{k^2_{I2}})}{(k_{M2} + [S] + \frac{[S]^2}{k_{I2}^*}) \cdot (1 + \frac{(E1t)^3}{k^2_{I2}})} \tag{A.21}
\]

\section*{A.2 Enzymatic Reaction Equations of the Interaction of 4CL3 and 4CL5 with Multiple Substrates Inhibition using 4-coumaric Acid or Caffeic Acid as a Main Substrate}

\subsection*{A.2.1 Assumptions}

The enzymatic reactions involved in the interaction of 4CL3 and 4CL5 with multiple substrates also follow the general Michaelis-Menten kinetic assumptions as the process in Section 1. Terms, with \(^*\), represent the self-inhibition effect in each equation, and are applied in the model with caffeic acid as a primary substrate only.

\textit{Quasi-equilibrium Assumptions}

The association and disassociation between the components, related to enzymes, substrates, and inhibitors, are in binding equilibrium. The equations representing enzyme-substrate or enzyme-enzyme complexes are the same as the expressions, (3.8), (3.9) and (A.1)-(A.6), in Section 1. The competitive inhibitions of 4CL3 are represented by

\[
[E11] = \frac{1}{k^1_{I1}C1}[E1][I1] \tag{A.22}
\]

\[
[E12] = \frac{1}{k^1_{I2}C2}[E1][I2] \tag{A.23}
\]

where \(k^1_{I1}C1\) and \(k^1_{I2}C2\) are the competitive inhibition constants for 4CL3. The competitive and uncompetitive inhibitions of 4CL5 are represented by
\[ [E2I1] = \frac{1}{k_{2IC1}}[E2][I1] \]  
(A.24)

\[ [E2I2] = \frac{1}{k_{2IC2}}[E2][I2] \]  
(A.25)

\[ [E2SI1] = \frac{1}{k_{2IU1}}[E2S][I1] = \frac{1}{k_{M2} \cdot k_{2IU1}}[E2][S][I1] \]  
(A.26)

\[ [E2SI2] = \frac{1}{k_{2IU2}}[E2S][I2] = \frac{1}{k_{M2} \cdot k_{2IU2}}[E2][S][I2] \]  
(A.27)

where \( k_{2IC1} \) and \( k_{2IC2} \) are the competitive inhibition constants for 4CL5 and \( k_{2IU1} \) and \( k_{2IU2} \) are the uncompetitive inhibition constants for 4CL5. The competitive and uncompetitive inhibitions of the enzyme-complex are represented by

\[ [(E1)_3E2I1] = \frac{1}{k_{2IC1}}[(E1)_3E2][I1] = \frac{1}{k_{3} \cdot k_{2IC1}}[E1]^3[E2][I1] \]  
(A.28)

\[ [(E1)_3E2I2] = \frac{1}{k_{2IC2}}[(E1)_3E2][I2] = \frac{1}{k_{3} \cdot k_{2IC2}}[E1]^3[E2][I2] \]  
(A.29)

\[ [(E1)_3E2SI1] = \frac{1}{k_{2IU1}}[(E1)_3E2S][I1] = \frac{1}{k_{3} \cdot k_{M2} \cdot k_{2IU1}}[E1]^3[E2][S][I1] \]  
(A.30)

\[ [(E1)_3E2SI2] = \frac{1}{k_{2IU2}}[(E1)_3E2S][I2] = \frac{1}{k_{3} \cdot k_{M2} \cdot k_{2IU2}}[E1]^3[E2][S][I2] \]  
(A.31)

where we assume that the inhibition constants for the enzyme-complex use the same values as the inhibition constants for 4CL5.

**Conservation of Total Enzyme**

The total (t) 4CL3 enzyme concentration is equal to the sum of species with 4CL3

\[ [E1t] = [E1] + [E1S] + [E1I1] + [E1I2] + 3 \cdot [(E1)_3E2] + 3 \cdot [(E1)_3E2S] + 3 \cdot [(E1)_3E2I1] + 3 \cdot [(E1)_3E2SI1] + 3 \cdot [(E1)_3E2SI2] + 3 \cdot [(E1)_3E2(S)2]^* \]  
(A.32)
The total \((t)\) 4CL5 enzyme concentration is equal to the sum of species with 4CL5

\[
[E2t] = [E2] + [E2S] + [E2I1] + [E2I2] + [E2(S)_2]^* + [E2SI1] + [E2SI2]
+ [(E1)_3E2] + [(E1)_3E2S] + [(E1)_3E2I1] + [(E1)_3E2I2] + [(E1)_3E2SI1]
+ [(E1)_3E2SI2] + [(E1)_3E2(S)_2]^* \tag{A.33}
\]

A.2.2 Derivation of Rate Equation

**Definition of Unknown Parameters :** \(k_1, k_2\) and \(\gamma\)

The same definition is applied as that in Section 1.

**[E1] Definition**

Equation (3.8), (A.4), (A.5), (A.6), (A.22), (A.23), (A.28), (A.29), (A.30), and (A.31) are substituted in (A.32), which is written as

\[
[E1] = [E1t] = \frac{[E1]}{D_{E1}}, \tag{A.35}
\]

where

\[
D_{E1} = (1 + \frac{[S]}{k_{M1}} + \frac{[I1]}{k_{1IC1}} + \frac{[I2]}{k_{1IC2}})
+ \frac{3}{k^3} [E1]^2 [E2] \cdot (1 + \frac{[I1]}{k_{2IC1}} + \frac{[I2]}{k_{2IC2}} + \frac{[S]}{k_{M1}} \cdot (1 + \frac{[I1]}{k_{2IU1}} + \frac{[I2]}{k_{2IU2}} + \frac{[S]^*}{k_{IS}})) \tag{A.36}
\]
Using (A.9), the expression of \([E1]\) is written as

\[
[E1] = \frac{[E1t]}{D_{E1t}},
\]  
(A.37)

where

\[
D_{E1t} = (1 + \frac{[S]}{k_{M1}} + \frac{[I1]}{k_{1IC1}} + \frac{[I2]}{k_{1IC2}})
\]

\[
+ \frac{3}{k_{3}} [E1t]^{2}[E2t] \cdot (1 + \frac{[I1]}{k_{2IC1}} + \frac{[I2]}{k_{2IC2}} + \frac{[S]}{k_{M1}} \cdot (1 + \frac{[I1]}{k_{2IU1}} + \frac{[I2]}{k_{2IU2}} + \frac{[S]}{k_{IS}}))
\]  
(A.38)

\[E2\] Definition

Equation (3.9), (A.1), (A.4), (A.5), (A.6), (A.24), (A.26), (A.27), (A.28), (A.29), (A.30), and (A.31) are substituted in (A.33), which is written as

\[
[E2t] = [E2] + \frac{1}{k_{M2}} [E2][S] + \frac{1}{k_{2IC1}} [E2][I1] + \frac{1}{k_{2IC2}} [E2][I2] + \frac{1}{k_{M2} \cdot k_{IS}} [E2][S]^{2} \ast
\]

\[
+ \frac{1}{k_{M2} \cdot k_{2IU1}} [E2][S][I1] + \frac{1}{k_{M2} \cdot k_{2IU2}} [E2][S][I2] + \frac{3}{k_{3}} [E1^{3}] [E2]
\]

\[
+ \frac{1}{k_{3} \cdot k_{M2}} [E1^{3}] [E2][S] + \frac{1}{k_{3} \cdot k_{2IC1}} [E1^{3}] [E2][I1] + \frac{1}{k_{3} \cdot k_{2IC2}} [E1^{3}] [E2][I2]
\]

\[
+ \frac{1}{k_{3} \cdot k_{M2} \cdot k_{2IU1}} [E1^{3}] [E2][S][I1] + \frac{1}{k_{3} \cdot k_{M2} \cdot k_{2IU2}} [E1^{3}] [E2][S][I2]
\]

\[
+ \frac{1}{k_{3} \cdot k_{M2} \cdot k_{IS}} [E1^{3}] [E2][S]^{2} \ast
\]  
(A.39)

This yields

\[
[E2] = \frac{[E2t]}{D_{E2}},
\]  
(A.40)

where

\[
D_{E2} = (1 + \frac{[I1]}{k_{1IC1}} + \frac{[I2]}{k_{1IC2}} + \frac{[S]}{k_{M2}} \cdot (1 + \frac{[I1]}{k_{2IU1}} + \frac{[I2]}{k_{2IU2}} + \frac{[S]}{k_{IS}})) \cdot (1 + \frac{[E1]}{k})
\]  
(A.41)

Using (A.11), the expression of \([E2]\) is written as
\[ [E2] = \frac{[E2t]}{D_{E2t}}, \]  
(A.42)

where

\[
D_{E2t} = (1 + \frac{[I1]}{k_{1IC1}} + \frac{[I2]}{k_{1IC2}} + \frac{[S]}{k_{M2}} \cdot (1 + \frac{[I1]}{k_{2IU1}} + \frac{[I2]}{k_{2IU2}} + \frac{[S]}{k_{IS}})) \cdot (1 + \frac{[E1t]}{k_2}) \]  
(A.43)

**Rate Equation**

The total rate equation is written as (A.20). Considering the derived expressions of \([E1]\) and \([E2]\) above, the final rate equation is

\[
v = \frac{k_{cat1}[E1t][S]}{D_{E1t}} + \frac{k_{cat2}[E2t][S](1 + \gamma \cdot \frac{[E1t]}{k_2})^3}{D_{E2t}} \]  
(A.44)
Appendix B

Derivation of Details of mechanistic model equation considering the enzyme complex

In the modeling of the two enzymatic reaction considering enzyme complex, we assume that the rate equation can be represented by the sum of each reaction by enzyme components involved in reaction [15], which is defined as:

\[ v = v_1 + v_2 + v_c, \]  
(B.1)

where \( v \) represents the total reaction rate and \( v_1, v_2, \) and \( v_c \) represent the reaction rates with free \( E_1, \) free \( E_2, \) and the enzyme complex composed of \( E_1 \) and \( E_2, \) respectively. With kinetic constants used in Michaelis-Menton Kinetic, the reaction rates associated with each enzyme component can be explained as:

\[ v_1 = k_{cat1}[E_1S] = \frac{k_{cat1}[E_1]}{k_{M1}}[E_1][S], \]  
(B.2)

\[ v_2 = k_{cat2}[E_2S] = \frac{k_{cat2}[E_2]}{k_{M2}}[E_2][S], \]  
(B.3)

\[ v_c = k_{cat3}[E_{1,m}E_{2,n}S] = \frac{z_1 \cdot k_{cat3}}{k_{M3}}[E_{1,m}E_{2,n}][S]. \]  
(B.4)

In (B.2), \( k_{cat1} \) and \( K_{M1} \) are kinetic information from the measured data, \( [E_1] \) represents the free \( E_1 \) concentration amount and \( [S] \) represents the substrate concentration. In (B.3), \( k_{cat2} \) and \( K_{M2} \) are also kinetic information from the measured reaction rate data, \( [E_2] \) represents the free \( E_2 \) concentration amount and \( [S] \) represents the substrate concentration. In (B.4),
the unknown kinetic information of the reaction rate associated with the enzyme complex are represented by \( z_1 \cdot k_{\text{cat}3} \) and \( K_{M3} \), where \( k_{\text{cat}3} \) and \( K_{M3} \) are the median values of two kinetic constants of \( E1 \) and \( E2 \), and \( z_1 \) is the unknown parameter inferring the effects of the enzyme complex on the reaction relatively. \([E1_mE2_n]\) represents the enzyme complex with the ratio of \( m : n \).

Referring to the model equation derivation for the two enzymatic reaction considering the specific enzyme complex with the ratio of 3:1, in Supplement Methods in [15], the general expressions of the concentrations of free \( E1 \), \([E1]\), free \( E2 \), \([E2]\), and the enzyme complex, \([E1_mE2_n]\), can be displayed in (B.5), (B.6), and (B.7):

\[
[E1] = \frac{[E1_t]}{(1 + \frac{[S]}{K_{M1}}) + m \cdot \frac{[E1_t]^{m-1}[E2_t]^n}{k_1^{m+n-1}(1 + \frac{[S]}{K_{M3}})}},
\]

\[
[E2] = \frac{[E2_t]}{(1 + \frac{[S]}{K_{M2}}) + n \cdot \frac{[E1_t]^{m-1}[E2_t]^n}{k_2^{m+n-1}(1 + \frac{[S]}{K_{M3}})}},
\]

\[
[E1_mE2_n] = \frac{[E1_t][E2_t]^n}{k_3^{m+n-1}(1 + \frac{[S]}{K_{M3}})}.
\]

where the assumed ratio of \( E1 \) to \( E2 \) in the enzyme complex is \( m : n \). The unknown parameters, such as \( k_1 \), \( k_2 \) and \( k_3 \), are involved in the unknown equilibrium constant between \( E1 \) and \( E2 \) in enzyme-enzyme interactions, and the unpredictable proportions of \( E1 \) and \( E2 \) in each enzyme-enzyme binding in terms of the total enzyme concentration, respectively. The detail mathematical definition of these unknown parameters can be referred to Supplement Methods in [15]. Considering all equations from (B.1) to (B.7), the final rate equation in the detail mechanistic modeling can be represented by

\[
v = \frac{k_{\text{cat}1} \cdot [E1_t] \cdot [S]}{K_{M1} + [S] + m \cdot \frac{K_{M1}[E1_t]^{m-1}[E2_t]^n}{k_1^{m+n-1}(1 + \frac{[S]}{K_{M3}})}} + \frac{k_{\text{cat}2} \cdot [E2_t] \cdot [S]}{K_{M2} + [S] + n \cdot \frac{K_{M2}[E1_t]^{m-1}[E2_t]^n}{k_2^{m+n-1}(1 + \frac{[S]}{K_{M3}})}} + \frac{\gamma \cdot k_{\text{cat}3} \cdot [E1_t][E2_t]^n \cdot [S]}{k_3^{m+n-1}(K_{M3} + [S])},
\]

which is duplicated to (4.3) in the main text.