MONASH University

Modelling Consolidated Bioprocessing of Cellulose via Population Balances Coupled with Cybernetic Models

Firnaaz Ahamed
B. Eng. (Hons.) in Chemical Engineering

A thesis submitted for the degree of Doctor of Philosophy at Monash University

September 2020

Copyright Notice

Abstract

Owing to its abundance and renewability, the direct conversion of cellulosic biomass to biochemicals via consolidated bioprocessing (CBP) is attractive. The convoluted dynamics of enzymatic saccharification, fermentation of soluble sugars through cellular metabolism, and concurrent cellulase secretion in CBP eludes a full mechanistic understanding of the underlying process fundamentals, thus, rational process and microbial strain designs and optimizations are challenging. The foregoing limitations are best resolved with theory-based dynamic modelling frameworks, yet, existing frameworks that rely on semi-mechanistic approaches at most, are not adequate to represent all critical nuances of CBP. This study proposes the coupling of two complementing frameworks — population balances and cybernetic models, to simulate the enzymatic saccharification and fermentation in CBP, respectively. The population balances, which lack analytical solutions, are solved numerically over a discrete-continuous mesh to embody the discrete bond-breaking phenomena of enzymatic saccharification while economizing on the computational load. Through an assessment of a variety of numerical solutions over several performance indicators, the Fixed Pivot Technique (FPT) demonstrated increased superiority, where FPT offers at least 7% better accuracy and 63% more efficiency than the second-best contender. Thus, the FPT is used as the basis for the development of population balances in this work. To represent the enzymatic saccharification of the complex distributed-heterogeneous polymeric cellulose, a new modelling platform termed the Multi-Layered Population Balance Model (ML-PBM) is developed, which enables various key aspects of the saccharification occurring over the entire breakdown process to be captured. As a core component for predicting the well-known slowdown phenomenon, the ML-PBM properly accounts for heterogeneity in cellulose crystallinity and chain lengths across the structural layers of cellulose particles with different morphologies. Beyond a decent quantitative fit to highly nonlinear dynamic experimental data collected across different conditions, the ML-PBM reveals that the rate slowdown phenomenon is po-

tentially due to heterogeneity in cellulose properties coupled with cellulose morphology. Subsequently, the ML-PBM is coupled with a metabolic network-based Lumped Hybrid Cybernetic Model (L-HCM) to form a new framework, named the Unified Cybernetic-Population Balance Model (UC-PBM), to simulate CBP of cellulose. To account for the interdependent dynamics of saccharification and fermentation in CBP, the UC-PBM features a closed-loop control mechanism between global regulation of extracellular cellulases and local intracellular regulation of fermentation at the network level. Through a case study on *Clostridium thermocellum*, the UC-PBM reveals that the cellulolytic microorganism actively reciprocates via flux re-distribution to changes in the environment through the onset of famine/feast conditions, which is dependent on the saccharification dynamics. Despite good fit with most exometabolomic data, the model overpredicts cellulosome secretion by more than three times in a nutrient-rich condition, while preserving the prediction of cellulase-coupled biomass growth with less than 6% error margin. This result suggests that future consideration of a new group of cellulase-decoupled growth metabolic pathways will allow for a more rigorous representation of the microorganism's metabolic behaviour. The foregoing insights on the nature of the cellulolytic microorganism would not have been possible with existing frameworks devoid of the closed-loop regulation. In overall, the UC-PBM is a rational framework with the potential to facilitate sound cellulose bioprocessing.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: Firnaaz Ahamed **Date:** 16 September 2020

Publications During Enrolment

- **F. Ahamed**, H.-S. Song, Y.K. Ho. 2020. Modelling Coordinated Enzymatic Control of Saccharification and Fermentation by *Clostridium thermocellum* During Consolidated Bioprocessing of Cellulose. *Manuscript in preparation for publication.*

- **F. Ahamed**, M. Singh, H.-S. Song, P. Doshi, C.W. Ooi, Y.K. Ho. 2020. On the use of sectional techniques for the solution of depolymerization population balances: Results on a discrete-continuous mesh. *Adv. Powder Technol.*, 31, 2669–2679.

- **F. Ahamed**, H.-S. Song, C.W. Ooi, Y.K. Ho. 2019. Modelling heterogeneity in cellulose properties predicts the slowdown phenomenon during enzymatic hydrolysis. *Chem. Eng. Sci.*, 206, 118–133.

Acknowledgements

The journey of a Ph.D. student is not an easy one and certainly not without unique challenges. Yet, I have learned to relish the experience, which has undoubtedly moulded me to become a better researcher in all aspects. This research journey would not have been memorable and educative, if not for the inspirations, guidance, and support that I have received from several notable individuals in my life. I would like to take this opportunity to express my gratitude to all those individuals.

First and foremost, I would like to express my deepest gratitude to my lead academic advisor/supervisor, Dr. Joseph Ho Yong Kuen, for his invaluable guidance throughout my candidature. The meticulousness, confidence, and enthusiasm towards scientific knowledge and research that he exudes have certainly moulded my outlook on research and inspired me to pursue impactful and worthwhile contributions to the research community. I am always indebted to his patience in training me in computational research from the very beginning, and continuous support until the completion of my Ph.D. adventure. Dr. Joseph is indeed a fine mentor one could ever ask for, and I am looking forward to many more years of working with him in research.

I am also profoundly grateful to my research collaborator, Prof. Hyun-Seob Song from the University of Nebraska-Lincoln (formerly affiliated to Pacific Northwest National Laboratory, PNNL), for his guidance in research across opposite time zones, often via video conferences or emails. I had the precious opportunity to learn the concepts of metabolic modelling first-hand from Prof. Song, one of the leading researchers in the field, whilst also refining my writing skills by producing manuscripts alongside an avid author like him. I also owe special thanks to my academic co-advisor, Prof. Edward Ooi Chien Wei, for his advice and support towards the research.

I would also like to acknowledge Monash University for the financial support with full-

stipend and merit scholarships to pursue graduate studies. I also acknowledge the award of Vice-Chancellor's International Inter-Campus Ph.D. Travel Grant to visit Monash University Clayton Campus in Australia to initiate collaborative research discussions and knowledge-sharing as well as to expand connections with our research counterpart.

I am also overwhelmingly indebted to my parents and my sister for their impartial and endless mental support, as well as their patience and understanding while I take up on one of the most challenging adventures, that is, the Ph.D. research. I am also thankful to my friends and colleagues for their relentless support in my research works. Finally, all praises, thanks and glory be to the Almighty for imparting me with the strength to face the challenges of this world.

So comes the end of my Ph.D. journey, but a research journey never truly ends. I look forward to more challenging adventures ahead in my pursuit of research that transcends across boundaries and disciplines in the future.

Firnaaz Ahamed,
September 2020.

Contents

Copyright Notice i

Abstract iii

Declaration v

Publications During Enrolment vii

Acknowledgements ix

List of Tables xix

List of Figures xxiii

Nomenclature xxxiii

1 Introduction 1

1.1 Background and Motivation 1

1.2 Objectives 5

1.3 Organization of the Thesis 6

2 Literature Review **9**

2.1 Modelling Consolidated Bioprocessing of Cellulose 9

2.1.1 Population Balance Modelling for Depolymerization Processes . . . 10

2.1.2 Modelling Frameworks for Enzymatic Hydrolysis of Cellulose 12

2.1.3 Modelling Frameworks for Metabolic Systems of Cellulolytic Microorganisms 19

2.2 Efforts to Address Research Gaps 21

2.2.1 Employing Sectional Techniques to Solve Depolymerization Population Balances over a Hybridized Discrete-Continuous Mesh 21

2.2.2 Introducing Multi-Layered Population Balance Model for Enzymatic Hydrolysis of Cellulose 23

2.2.3 Introducing Unified Cybernetic-Population Balance Model for Consolidated Bioprocessing of Cellulose 24

3 Solving Depolymerization Population Balances Using Sectional Techniques on a Discrete-Continuous Mesh **29**

3.1 Preface 29

3.2 Theoretical Framework 30

3.2.1 Discrete-Continuous Mesh 30

3.2.2 General Discretized PBE Framework for Sectional Techniques . . . 31

3.2.3 Fixed Pivot Technique (FPT) . . . 33

3.2.4 Cell Average Technique (CAT) . . . 34

3.2.5 Finite Volume Scheme (FVS) . . . 35

3.3 Case Study . . . 36

3.3.1 Preamble . . . 36

3.3.2 FPT and CAT Predicts Chain-end Scission while FVS Fails . . . 39

3.3.3 Modified Sectional Technique Solves the Discrepancy at the Discrete-Continuous Boundary for Chain-End Scission . . . 46

3.3.4 Alternative Meshing Strategy to Alleviate Discrepancy at the Discrete-Continuous Boundary for Chain-End Scission . . . 49

3.3.5 Prediction of Chain-End Scission over Uniform Mesh in the Continuous Region . . . 50

3.3.6 New Modified Stoichiometric Kernel for Random Scission Improves the Prediction of All Sectional Techniques . . . 51

3.3.7 Prediction of Random Scission over Uniform Mesh in the Continuous Region . . . 56

3.3.8 Summary . . . 57

3.4 Concluding Remarks . . . 58

4 Multi-Layered Population Balance Model Predicts the Dynamics of Enzymatic Hydrolysis of Cellulose 59

4.1 Preface 59

4.2 Model Framework 60

4.2.1 Conceptual Description of Heterogeneity in Cellulose Properties . . 60

4.2.2 Hydrolysis Reaction Mechanisms 63

4.2.3 Population Balance Modelling for Enzymatic Hydrolysis Process . . 65

4.2.4 Multi-Layered Population Balance Model (ML-PBM) 67

4.3 Results and Discussion 69

4.3.1 Heterogeneity in Cellulose Distribution 69

4.3.2 Prediction of Rate Slowdown and Effective Enzyme Footprints . . . 71

4.3.3 Cellulose Heterogeneity: A Key Modelling Component for Predicting Slowdown 75

4.3.4 Further Validations of the ML-PBM and Morphological Interpretations 76

4.3.5 Effect of Heterogeneous Cellulose Properties on Cellulose Distribution During Enzymatic Hydrolysis 80

4.3.6 ML-PBM Predicts the Enzyme Tunnelling Effect 83

4.3.7 ML-PBM Predicts the Dynamics of Fed-Batch Hydrolysis 86

4.4 Concluding Remarks 88

5 Unified Cybernetic-Population Balance Model Predicts the Dynamics of Consolidated Bioprocessing of Cellulose 91

5.1 Preface 91

5.2 Model Framework 92

5.2.1 Conceptual Description of Modelling Scheme 92

5.2.2 General Formulation of Model Equations 94

5.3 Case Study 96

5.3.1 Motivations to Examine Cellulose Utilization by *Clostridium thermocellum* 96

5.3.2 Abstraction of Core Metabolism from Genome-Scale Metabolic Network 97

5.3.3 Potential Trade-Offs between Cellulosome-Secreting and Growth-Coupled Metabolic Pathways 99

5.3.4 EFV Lumping Scheme in the UC-PBM 101

5.3.5 Model Equations 103

5.3.6 Metabolic Objective of *C. thermocellum* 106

5.4 Results and Discussion 108

5.4.1 UC-PBM Predicts Coupled Dynamics of Saccharification and Fermentation Regulated by *C. thermocellum* 108

5.4.2 Cellulose Loading is a Major Driving Force of CBP 113

5.4.3 Growth-Coupled Pathways Dominate the Metabolism of *C. thermocellum* 114

5.4.4 UC-PBM Suggests Possible Cellulosome-Decoupled Growth Metabolic Diversity 115

5.4.5 UC-PBM Guides to Metabolic Engineering 118

5.5 Concluding Remarks 120

6 Conclusions and Future Scopes **121**

6.1 Conclusions and Contributions of the Research 121

6.2 Future Scopes 123

Bibliography **124**

Appendices **151**

A Solution of Depolymerization Population Balances **151**

A.1 Fully Discrete (Exact) ODEs for Depolymerization Population Balances . . 151

A.2 FPT ODEs for Depolymerization Population Balances 152

A.3 CAT ODEs for Depolymerization Population Balances 154

A.4 Failure of the FVS for Chain-End Scission 155

A.5 CAT and FPT Solutions are Identical for Chain-End Scission 157

A.6 FVS and FPT Solutions are Identical for Random Scission 159

B Formulation of Multi-Layered Population Balance Model **163**

B.1 Formulation of Multi-Layered Distributed Cellulose Hydrolysis 163

B.2 Sterically Hindered Competitive Enzyme Adsorption 165

B.3 Fixed Pivot Discretization of ML-PBM Equations 166

B.4 Modelling Initial Conditions 172

B.5 Modelling Simplified Homogeneous Model 173

B.6 Sensitivity Analysis and Parameter Identifiability 174

B.7 Model Parameter Estimation 178

C Formulation of Unified Cybernetic-Population Balance Model 181

C.1 Reduced Core Metabolic Network of *C. thermocellum* 181

C.2 EFV Lumping Scheme 188

C.3 UC-PBM Equations 189

C.3.1 ML-PBM Equations 189

C.3.2 L-HCM Equations 193

C.4 Cellobiose Phosphorylation Kinetics 195

C.5 Initial and Maximum Intracellular Enzyme Levels 195

C.6 Summary of Power Law Forms of the ML-PBM Rate Coefficients 197

C.7 Model Parameter Estimation 198

D MATLAB Codes 199

D.1 Simulation of Chain-End Monomer Scission 199

D.2 Simulation of Random Scission 210

D.3 Simulation of Enzymatic Hydrolysis of Cellulose (ML-PBM) 221

D.4 Simulation of Consolidated Bioprocessing of Cellulose (UC-PBM) 234

List of Tables

2.1 Review of existing models for the enzymatic hydrolysis of celluloses, benchmarked against the core features of the ML-PBM developed in this research. 16

3.1 Average global errors ($\bar{\varepsilon}$), and moments errors (ξ_0, ξ_1, ξ_2) of the FPT and the CAT for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$, over a uniform mesh with progressively refined mesh. For a non-uniform geometric mesh with $[p, q] = [100, 500]$ and a common ratio, $r = 1.0109$, the $\bar{\varepsilon}$ is 0.047, ξ_0 and ξ_1 are 3.53×10^{-5}, and ξ_2 is 0.028. 51

3.2 Average global errors ($\bar{\varepsilon}$) and moments errors (ξ_0, ξ_1, ξ_2) for random scission with a linear rate kernel, $k^{\alpha}(v) = v$. 55

3.3 Average global errors ($\bar{\varepsilon}$), and moments errors (ξ_0, ξ_1, ξ_2) of the FPT, the FVS and the CAT for random scission with the new improved stoichiometric kernel and linear rate kernel, $k^{\alpha}(v) = v$, over a uniform mesh with progressively refined mesh. 56

4.1 Reaction mechanisms of enzymatic hydrolysis of celluloses. Here, E_F^m are the free enzymes, E_S^m are the surface-adsorbed enzymes, $E_B^m P_S(i)$ are the enzyme-polymer complexes, where m = CBH, EG, BG or EG-CBH, $P_S(i)$ are the insoluble surface polymers and $P(i)$ are the soluble oligomers, i (or j) denotes the DP and N represents the maximum cellulose DP. 64

4.2 Parameter values used in ML-PBM predictions in this study. 72

5.1 Number of EFVs enumerated from reduced metabolic network ($N_{20\%}$) of *C. thermocellum*. Thermodynamically infeasible EFVs with no substrate uptakes are omitted. 101

5.2 Key UC-PBM equations. 103

5.3 Values of model parameters used in this study. Conceptualization of the ML-PBM parameters are presented in Chapter 4/Appendix B. 109

B.1 Summary of rate coefficients and corresponding power law form. 172

C.1 List of participating reactions in reduced core metabolic network of *C. thermocellum* under anaerobic conditions. 182

C.2 List of participating metabolites in reduced core metabolic network of *C. thermocellum* under anaerobic conditions. 186

C.3 Estimated values of tuning coefficients, a_i^{J} (i = eth, lac, form, ace; J = I, II) and the corresponding model-estimated lumped yields ($Y_{i,\mathrm{J}}$) of metabolites secreted by *C. thermocellum* compared with experimental steady-state yield data. The yields are presented in mole basis of cellobiose. Here, the yields of biomass are evaluated by discounting the cellulase component from the yields of dry cell weights from the metabolic networks. 189

C.4 Summary of rate coefficients and corresponding power law form. 197

List of Figures

1.1 Chart illustrating the thesis structure. 7

2.1 A schematic representation of the ML-PBM and the UC-PBM concepts. The closed-loop regulatory mechanism in the metabolism of cellulolytic microorganisms utilizing cellulose is realized through the coupling between the ML-PBM and the L-HCM. 25

3.1 A graphical representation of a typical discrete-continuous mesh, where p denotes the number of grid points in the discrete region and q is the number of grid points in the continuous region. Alternative notations are displayed in the parentheses. Here, $x_{p+q} = x_{p+q+1/2} = N$, where N is the maximum DP of the polymer distribution and $v_m = 1$ is the DP of a monomer. 30

3.2 A comparison of the population density for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$. 40

3.3 A comparison of (a) temporal evolutions of oligomers with DP = 2 and 10, (b) zeroth moment, (c) first moment, and (d) second moment, for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$. 41

3.4 A comparison of (a) the total molar concentrations of DP 2 to N and the monomer molar concentration, and (b) the rate of change in the monomer molar concentration as given by the RHS of Eq. (3.3), for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$. 42

3.5 A comparison of the total mass of polymers with DP 2 to N and the mass of monomer, for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$. . . 43

3.6 Predictions of the zeroth moment for chain-end scission with: (a) a constant rate kernel, $k^{\gamma}(v) = v$, and (b) a quadratic rate kernel, $k^{\gamma}(v) = v^2$. 45

3.7 The discrepancy in the population density for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$. The discrepancy occurs at the discrete-continuous boundary, i.e., $x_{p+1} = 101$. FPT 1, FPT 2 and FPT 3 represent the original fixed pivot solution, the modified fixed pivot solution and the alternative meshing strategy to alleviate the discrepancy, respectively. . . . 47

3.8 Temporal evolution of the dimensionless population density of $x_{p+1} = 101$ for FPT 1 (original formulation), FPT 2 (modified formulation) and FPT 3 (alternative meshing strategy). Here, chain-end scission is simulated with a linear rate kernel, $k^{\gamma}(v) = v$. 49

3.9 The performance of different sectional techniques for random scission with a linear rate kernel, $k^{\alpha}(v) = v$, using: (a) $b_1^{\alpha}(v, x_j) = 2/(x_j - 1)$, and (b) $b_2^{\alpha}(v, x_j) = 2/(x_{j-1})$. 53

3.10 The performance of the sectional techniques in predicting the zeroth and second moments along with temporal evolution of oligomers using: (a) $b_1^{\alpha}(v, x_j) = 2/(x_j - 1)$, and (b) $b_2^{\alpha}(v, x_j) = 2/(x_{j-1})$, for pure random scission with a linear rate kernel, $k^{\alpha}(v) = v$. 54

4.1 The depiction of the multi-layered distributed cellulose ultrastructure hypothesis. A cylindrical shape is assumed for microfibrils **(A)**, where the cellulose chains are organized in two-dimensional layered lattices. $R(t)$ is the microfibril radius, $R_0 = 1$ nm is the thickness of a single layer (i.e., the diameter of a single glucan molecule) and L is the length of the microfibril. The individual layers are denoted as l, where $l = 1, 2, 3, \ldots, l_{\max}$ and $l_{\max} = R(0)/R_0$. The native celluloses are made up of homogeneous polymer distribution across the layers **(B)** and the pre-treatment isolates the distribution into the penetration and internal zones **(C)**. Microfibril aggregates and the nature of the aggregates are traced based on the model-predicted initial particle radius, $R(0)$ **(D)**. Densely packed microfibril aggregates can be visualized macroscopically as a multi-layered large cylinder encompassing numerous microfibrils analogous to a single microfibril. . . . 61

4.2 Fitting to experimental molecular weight distributions of celluloses (**A**, **B** and **C**) and subsequent prediction of probable distributions by the ML-PBM (**D**, **E** and **F**). The experimental distribution data represents: (**A**) Avicel from Engel et al. [137], (**B**) α-cellulose and (**C**) HCC from Nag et al. [85]. The probable distributions were predicted for: (**D**) Avicel hydrolysis data in Bezerra et al. [87], Bezerra and Dias [89], and Medve et al. [86], (**E**) α-cellulose and (**F**) HCC hydrolysis data in Nag et al. [85]. 70

4.3 Model fitting to cellobiose concentration and surface-adsorbed enzyme concentration for: (**A**) Avicel hydrolysis by CBH and (**B**) Avicel hydrolysis by CBH and EG. For both cases, the Avicel loading is 10 g/L, CBH loading is 11.2 mg/g cellulose, and EG loading is 7.7 mg/g cellulose. Enzyme-centric variables are evaluated from this fitting and used for subsequent predictions. The ML-PBM is represented by the solid lines and the experimental data from Medve et al. [86] are represented by the markers. The rest are the results of the simplified homogeneous model, where dashed lines $(---)$ represent the prediction with the same set of kinetic parameters from ML-PBM, dotted lines $(\cdots)$ represent the prediction with recalibrated kinetic parameters to the entire data set and dash-dotted line $(- \cdot -)$ represent prediction with recalibrated kinetic parameters to the initial data points only. The overall initial cellulose distribution is common for all cases. . . . 73

4.4 ML-PBM calibration and validation to cellobiose concentration for Avicel hydrolysis by CBH. In (**A**), the ML-PBM is calibrated with 50 g/L Avicel and CBH loading of 0.84 mg/g cellulose, and subsequently, validated with the other two data sets, where the experimental data are obtained from Bezerra et al. [87]. Similarly in (**B**), the ML-PBM is calibrated with 50 g/L Avicel and CBH loading of 0.2 mg/g cellulose and validated with the remaining data sets, where the experimental data are from Bezerra and Dias [89]. The solid lines are the fitted ML-PBM, dashed lines are the predictions and the markers are experimental data. For Avicel loading of 2.5 g/L, the model predicts poorly (blue-dotted lines) but the prediction improved when substrate-centric parameters are re-fitted while retaining other parameters (cf. parameter values in Table 4.2). Av in the legend refers to Avicel. 77

4.5 ML-PBM calibration and validation to conversion for: (**A**) α-cellulose hydrolysed by CBH+BG and CBH+BG+EG enzyme mixtures, and (**B**) HCC hydrolysed by CBH+BG and CBH+BG+EG enzyme mixtures. The loadings for α-cellulose and HCC are 20.13 g/L and 24.19 g/L, respectively. Enzyme loadings for CBH+BG mixture are 30 mg/g cellulose and 2 mg/g cellulose, respectively, whereas enzyme loadings for CBH+BG+EG mixture are 25 mg/g cellulose, 2 mg/g cellulose and 5 mg/g cellulose, respectively. The lines are ML-PBM predictions while the markers denote experimental data from Nag et al. [85]. The parameter values are consistent over all four cases. 79

4.6 Temporal evolution of number-average DP ($\bar{M}_{\mathrm{n}}$), weight-average DP ($\bar{M}_{\mathrm{w}}$) and particle radius of insoluble celluloses for hydrolysis of: (**A**) α-cellulose by CBH+BG+EG mixture and (**B**) HCC by CBH+BG+EG mixture. ML-PBM predictions are for the cases presented in Figure 4.5. The ML-PBM demonstrates an initial increase in mean DPs during the hydrolysis of the penetration zone due to the solubilization of shorter chains while preserving the longer chains underneath in the internal zone. The mean DPs level off and subsequently decreases when the hydrolysis advances into the internal zone. 81

4.7 Temporal evolution of insoluble cellulose distributions for hydrolysis of: (**A**) α-cellulose by CBH+BG+EG mixture, and (**B**) HCC by CBH+BG+EG mixture. ML-PBM predictions are for the cases presented in Figure 4.5. The mass concentration densities for DP < 7 are not shown as it represents soluble oligomers liberated into the reaction medium. The distributions generally shift towards the larger DP and the polydispersity increases over the course of the hydrolysis. 82

4.8 Demonstration of the effect of enzyme crowding on the surface of celluloses for hydrolysis of: (**A**) α-cellulose by CBH+BG, and (**B**) HCC by CBH+BG. In both cases, the loading of CBH is increased from enzyme-limiting condition to saturation point (onset of enzyme crowding effect) and finally to substrate-limiting condition (in excess of enzymes). BG loading is maintained at 2 mg/g cellulose for all cases. ML-PBM predictions are represented by the lines and the markers denote experimental data from Nag et al. [85]. 84

4.9 Illustration of enzyme tunnelling effect. 85

4.10 ML-PBM prediction of glucose and cellobiose concentrations for fed-batch hydrolysis of SPS by CBH. Initial loadings of SPS and CBH are 10 g/L and 9.53 mg/g, respectively. The ML-PBM was calibrated with the fully batch data and validated with the fed-batch data. The solid line is the fitted ML-PBM, dashed lines are the predictions and the markers are experimental data from Eriksson et al. [127]. 87

5.1 Illustration of CBP of cellulose as conceptualized by the UC-PBM. Here, the utilization of cellulose by *C. thermocellum* is shown as an example, where the UC-PBM can be similarly adapted for any other cellulolytic microorganisms. The blue arrows display a closed-loop system that includes the intracellular and the extracellular regulations that govern the CBP. The individual stages (**A** — **D**) of CBP are detailed in the text. 93

5.2 Distributions of EFV lengths of *C. thermocellum* for different groups classified according to secretion of representative products. The length refers to the number of participating metabolic reactions in each individual EFV. Here, $N_{20\%}$ is used for the enumeration of the EFVs. 100

5.3 A flowchart depicting the extraction of essential biological information for implementation of the UC-PBM. Here, a meaningful core metabolic network encompassing the secretion of cellulosomes, metabolites of interest and the biomass growth is obtained from the genome-scale metabolic network of *C. thermocellum* (iAT601) [155]. EFVs are classified into two principal families – $\mathbf{F}_{\mathrm{I}}$ that contributes to inducive secretion of cellulosomes and $\mathbf{F}_{\mathrm{II}}$ that allows biomass growth and growth-associated cellulosome secretion. Both families contribute to the secretion of all metabolites of interest. 102

5.4 Model prediction of metabolism of *C. thermocellum* ATCC 27405 grown on 10.67 g/L Avicel as the sole carbon source. The markers represent experimental data from Zhang and Lynd [163] and the lines are model predictions. The solid markers represent data used for parameter estimation, while the model predicts the data represented by open markers. The grey-shaded regions denote periods of severe famine conditions and the drop-down arrow indicates the time for complete cellulose solubilization. The fractions of cellulase secreted through $\mathbf{F}_{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{II}}$ are given in the figure, where the remaining fraction is contributed by the constitutive secretion. The total cellobiose uptakes through the respective EFV families are 7.9% and 92.1%. Here, $\mathrm{N}_{20\%}$ is used to demonstrate the metabolism. 110

5.5 Model prediction of metabolism of C. thermocellum ATCC 27405 grown on 4.59 g/L Avicel (5.10 g glu eq/L) as the sole carbon source. The lines are model predictions and the markers are experimental data from Holwerda et al. [172]. Solid markers represent data used for parameter estimation, while the model predicts the data represented by open markers. The biomass growth is expressed in terms of biosynthate dry weight that constitutes cell mass, as well as cell-bound and supernatant protein (i.e., cellulosomes), whereas residual Avicel is presented in glucose equivalents following the data source. The grey-shaded regions denote periods of severe famine conditions and the drop-down arrows indicate the time for complete cellulose solubilization. The fractions of metabolites secreted through $\mathbf{F}_{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{II}}$ are given in the figure, whereas the total cellobiose uptakes through the respective EFV families are 12.7% and 87.3%. The residual fraction of cellulase is contributed by the constitutive secretion. Here, $\mathrm{N}_{20\%}$ is used to demonstrate the metabolism. 111

5.6 Model prediction of metabolism of *C. thermocellum* ATCC 27405 grown on 10 g/L pure cellobiose as the sole carbon source. The lines are model predictions and the markers are experimental data from Zhang and Lynd [163]. Estimated model parameter values from previous cases are used for predictions in this case. The fractions of metabolites secreted through $\mathbf{F}_{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{II}}$ are given in the figure, whereas the total cellobiose uptakes through the respective EFV families are 1.3% and 98.7%. Here, both $\mathrm{N}_{20\%}$ and $\mathrm{N}_{2\%}$ metabolisms are compared, where the grey-shaded region for famine conditions, the fractions of cellulase secretion and cellobiose uptakes as well as the relative intracellular enzyme levels in the bottom-right panel are provided for the latter. 116

5.7 Projections of three-dimensional convex hull of ethanol, cellulosome and biomass yields of *C. thermocellum* onto two-dimensional surfaces. The dots represent the yield points of individual EFVs that span the entirety of the convex hull, whereas the red lines denote the metabolic trajectories that the microorganism adopts to respond to perturbations in the environmental conditions as predicted by the UC-PBM. The EFVs are also classified into the individual families, where the vertical axes obscures the $\mathbf{F}_{\mathrm{I}}$ EFVs (with zero biomass yields) in the last two panels. The change in the cellulosome-biomass yield space when $N_{2\%}$ is imposed is shown in the last panel. The inset in the last panel depicts a hypothetical yield space, should a new unrecognized group of cellulosome-decoupled growth pathways be considered. The yields are presented in mole basis of cellobiose. 117

A.1 Illustration of (a) failure in FVS formulation in polymer birth allocation during chain-end scission, and (b) alternative polymer birth allocations based on the FPT and the CAT formulations for chain-end scission. 156

B.1 Illustration of multi-layered cellulose ultrastructure undergoing progressive hydrolysis. 164

B.2 Relative sensitivities of ML-PBM outputs (cellobiose production and conversion) to various parameters. The sensitivities are evaluated for hydrolysis of: (**A**) Avicel by CBH based on the data given in Medve et al. [86], (**B**) α-cellulose by CBH and BG, and (**C**) α-cellulose by CBH, BG and EG based on the data given in Nag et al. [85]. For (**C**), the EG-related parameters were evaluated whilst fixing the CBH and BG-related parameters. A +20% perturbation was enforced for all parameters except for P_{zone}, where an increment of 1 was employed. For case (**A**), results for the conversion have a similar trend to that of cellobiose production. 175

B.3 Overview of inputs and outputs for ML-PBM calibration. 179

Nomenclature

Latin Symbols

a_i^{J}	Tuning coefficients for adjustment of fluxes to EFVs in Eq. (C.2)	–
$b(v, x_j)$	Breakage stoichiometric kernel	mol/L
$c(v,t)$	Continuous molar concentration density	mol L^{-1} DP^{-1}
$\widetilde{c}(v,t)$	Dimensionless population density	–
c_i	Molar concentration density of DP = i	mol L^{-1} DP^{-1}
C_i	Molar concentration of of soluble polymers with DP = i, or molar concentration of metabolites, i = eth, lac, form, ace, biom, csm, cb, glu	mol/L
$C_{\mathrm{S},i}$	Molar concentration of surface polymers	mol/L
$C_{\mathrm{int},i}$	Molar concentration of internal polymers	mol/L
$C_{\mathrm{int},il}$	Molar concentration of internal polymers in individual layer l	mol/L
$C_{\mathrm{B},i}^{\mathrm{m}}$	Molar concentration of enzyme-polymer complexes, m = CBH, EG, EG-CBH	mol/L

$C_{\mathrm{F}}^{\mathrm{m}}$	Molar concentration of free enzymes, m = CBH, EG, BG	mol/L
$C_{\mathrm{S}}^{\mathrm{m}}$	Molar concentration of surface-adsorbed enzymes, m = CBH, EG	mol/L
$C_{\mathrm{T}}^{\mathrm{m}}$	Molar concentration of total enzymes, m = CBH, EG, BG, exo, endo	mol/L
$C_{\mathrm{sat}}^{\mathrm{m}}$	Enzyme saturation concentration, m = CBH, EG	mg/g cellulose
$\boldsymbol{\mathcal{C}}$	Vector of enzyme-polymer complexes	–
d	Common difference of arithmetic progression	–
$\mathbf{D}$	Operator $\mathbf{D}$ converts the vectors into diagonal matrices	–
$e_{\mathrm{J}}, e_{\mathrm{J}}^{\mathrm{rel}}, e_{\mathrm{J}}^{\mathrm{max}}$	Actual, relative and maximum intracellular enzyme levels	–
$\mathbf{e}$	Vector of intracellular enzyme levels	–
$\boldsymbol{\mathcal{E}}$	Vector of extracellular enzyme species	–
$\mathrm{E}_{\mathrm{F}}^{\mathrm{m}}$	Free enzyme species, m = CBH, EG, BG	–
$\mathrm{E}_{\mathrm{S}}^{\mathrm{m}}$	Surface-adsorbed enzyme species, m = CBH, EG	–
$\mathrm{E}_{\mathrm{B}}^{\mathrm{m}}\mathrm{P}_{\mathrm{S}}(i)$	Enzyme-polymer complex species, m = CBH, EG, EG-CBH	–
$\mathbf{F}_{\mathrm{J}}$	EFV matrix of J-th EFV family	–
i	Index representing DP of polymers and various metabolites	–
I	Total number of sections of discretized DP domain	–
j	Index representing DP of polymers ($j \geq i$) or individual EFVs	–
J	Index representing EFV families	–

k_i	Breakage rate kernel of DP $= i$	1/s
$k_{\mathrm{a}}^{\mathrm{m}}, k_{\mathrm{d}}^{\mathrm{m}}$	Adsorption and desorption rate constants, m = CBH, EG, exo, endo	L mol^{-1} s^{-1}, 1/s
$k_{\mathrm{f},i}^{\mathrm{m}}$	Complexation rate coefficients, $k_{\mathrm{f}}^{\mathrm{m}}(v) = k_{\mathrm{p,f}}^{\mathrm{m}}/v$; m = CBH, EG, exo	L mol^{-1} s^{-1}
$k_{\mathrm{f}}^{\mathrm{endo}}$	Complexation rate constant of endo-enzymes	L mol^{-1} s^{-1}
$k_{\mathrm{e}}^{\mathrm{m}}$	Decomplexation rate constants, m = CBH, EG, exo, endo	1/s
$k_{\mathrm{h}}^{\mathrm{m}}$	Hydrolysis rate constants, m = CBH, exo, endo	1/s
$k_{\mathrm{h},i}^{\mathrm{EG}}$	EG hydrolysis rate coefficient, $k_{\mathrm{h}}^{\mathrm{EG}}(v) = k_{\mathrm{p,h}}^{\mathrm{EG}}/v$	1/s
$k_{\mathrm{h}'}^{\mathrm{m}}$	Second order hydrolysis rate constants, m = EG, BG, endo	L mol^{-1} s^{-1}
$k_{\mathrm{p,f}}^{\mathrm{m}}$	Complexation rate constants, m = CBH, EG, exo	DP L mol^{-1} s^{-1}
$k_{\mathrm{p,h}}^{\mathrm{EG}}$	EG hydrolysis rate constant	DP/s
$k_{\mathrm{J}}^{\mathrm{max}}$	Maximum substrate uptake rates	mol gDW^{-1} s^{-1}
$k_{\mathrm{e,J}}$	Maximum intracellular enzyme synthesis rates	1/s
$K_{\mathrm{e,J}}, K_{\mathrm{J}}$	MM constants for synthesis of intracellular enzymes and substrate uptake	mol/L
$K_{\mathrm{inh-m}}$	Ethanol (m = eth) or glucose (m = glu) inhibition constants	mol/L
$\boldsymbol{k}_\gamma, \boldsymbol{k}_\lambda$	Vector of rate coefficients/constants of ML-PBM and L-HCM	–
$\boldsymbol{\mathcal{K}}$	Vector of MM constants	–

l	Structural layer number in cellulose particles	–
L	Length of cellulose particles	nm
m_j^i	Integral limit of Eq. (3.7)	–
M_i	Total mass of particles entering i-th section	–
$\bar{M}_\mathrm{n}, \bar{M}_\mathrm{w}$	Number- and weight-average DP	–
$M_\mathrm{T}, M_\mathrm{P}, M_\mathrm{I}$	Initial mass concentrations of total, penetration and internal zone cellulose	g/L
M_m	Molecular weight, m = CBH, EG, csm	g/mol or Da
n	Number of cellulose particles per unit volume of reaction medium	$1/\mathrm{m}^3$
n_eta	Parameter that controls that sensitivity of adjustment of fluxes to EFVs in Eq. (C.1)	–
n_r	Number of metabolic reactions	–
n_z	Number of EFVs	–
N	Maximum DP of polymer/cellulose	–
$\mathrm{N}_{20\%}$	Metabolic network of *C. thermocellum* when grown on cellulose	–
$\mathrm{N}_{2\%}$	Metabolic network of *C. thermocellum* when grown on cellobiose	–
$\boldsymbol{\mathcal{N}}$	Vector of fermentable sugars	–
p	Number of sections/grid points in discrete region	–
$\mathrm{P_S}(i)$	Surface insoluble polymer species	–
$\mathrm{P}(i)$	Soluble polymer species	–

P_{zone}	Number of layers in the penetration zone	–
$\boldsymbol{\mathcal{P}}$	Vector of secreted metabolites of interest	–
q	Number of sections/grid points in continuous region	–
r	Common ratio of geometric progression	–
$r_{\text{mass,P}}$	Ratio of cellulose mass in the penetration zone	–
$r_{\text{exp},i}$	Rate of exposure of internal polymers	mol L^{-1} s^{-1}
$r_{\text{loss},i}$	Rate of loss of surface polymers	mol L^{-1} s^{-1}
R	Radius of cellulose particles	nm
R_0	Thickness of a single layer in cellulose particles	nm
$\mathbf{r}$	Vector of substrate uptake fluxes through EFV families	mol gDW^{-1} s^{-1}
$\mathbf{r}^{\text{kin}}$	Vector of unregulated substrate uptake fluxes through EFV families	mol gDW^{-1} s^{-1}
$\mathbf{r}_{\text{e}}^{\text{kin}}$	Vector of unregulated inducive intracellular enzyme synthesis rates	1/s
S	Specific surface area of cellulose particles	m^2/m^3
$\boldsymbol{\mathcal{S}}$	Vector of all polymer species from monomer to max DP (N)	–
$\mathbf{U}$	Vector of cybernetic variables controlling synthesis of intracellular enzymes	–
v	Continuous DP of polymers	–
v_m	DP of a monomer = 1	–
$\bar{v}_i$	Average particle size entering i-th section	–

V_i	Total number of particles entering i-th section	–
$\mathbf{V}$	Vector of cybernetic variables controlling activities of intracellular enzymes	–
t	Time	s or h
$t_{99\%}$	Time taken for 99% of monomer production	s
x_i	Grid point/representative point of i-th section	–
$x_{i-1/2}, x_{i+1/2}$	Boundary points of i-th section	–
X	Mass concentration of biomass	g/L
$Y_{i,j}$	Yields of metabolites of interest of individual EFVs	mol/mol or g/mol
$Y_{i,\mathrm{J}}$	Model-estimated lumped yields of metabolites of interest of J-th EFV family	mol/mol or g/mol
$\mathbf{z}_\mathrm{J}$	Lumped EFV vector of J-th EFV family	–
$\mathbf{Z}$	EFV matrix	–
$\boldsymbol{\mathcal{Z}}$	Matrix containing stoichiometric information	–

Greek Symbols

$\boldsymbol{\alpha}$	Vector of constitutive intracellular enzyme synthesis rates	1/s
α_{csm}	Constitutive cellulosome synthesis rate	g gDW^{-1} s^{-1}
$\boldsymbol{\beta}$	Vector of intracellular enzyme decay rates	1/s
β_{csm}	Cellulosome decay rate	1/s
χ	Correction factor implemented in Eq. (3.23)	–

$\delta(x)$	Dirac-delta function	–
Δx_i	Width of i-th section	–
ε_{D}	Error in the initial mass of polymer due to discretization	–
ε_i	Global error in the population density of $i = [2, p]$	–
$\varepsilon_j^{\mathrm{J}}$	Correction factor for adjustment of fluxes to EFVs in Eq. (C.2)	–
ϕ	Parameter of gamma distribution defined by Eqs. (3.15) and (4.15)	–
$\mathbf{\Phi}$	Vector of ROIs	–
φ_i	Weighted allocation function for death terms	–
$\Gamma(x)$	Gamma function	–
η_{ij}	Particle allocation function for birth terms	–
κ_l	Mass concentration of celluloses in individual layers of cellulose particles	g/L
λ	Parameter of gamma distribution defined by Eqs. (3.15) and (4.15)	–
μ	Biomass growth rate	1/h
θ	Dimensionless time	–
ϑ, ϑ'	Fraction of occupied and unoccupied cellulose surface area	–
ρ	Density of cellulose	g/L
σ_{m}	Enzyme footprints, m = CBH, EG, exo, endo	mol/m^2
ω_i^{b}	Weight allocation for birth terms of i-th section	–

ω_i	Mass fractions, i = DW, biom, csm, cls	–
$\Omega(v)$	Parameter of gamma distribution defined by Eqs. (3.15) and (4.15)	–
ξ_k	Maximum relative error in moment of order k	–
ζ_k	Moment of order k	–

Abbreviations

αC	α-cellulose
ace	Acetate
Av	Avicel
BG	β-glucosidase
biom	Biomass
CAT	Cell average technique
cb	Cellobiose
CBH	Cellobiohydrolase
CBP	Consolidated bioprocessing
cls	Cellulase
csm	Cellulosome
DFBA	Dynamic flux balance analysis
DP	Degree of polymerization
DRUM	Dynamic reduction of unbalanced metabolism

DW	Dry weight
EFM	Elementary flux mode
EFV	Elementary flux vector
EG	Endoglucanase
ELISA	Enzyme-linked immunosorbent assay
EMA	Elementary mode analysis
eth	Ethanol
FBA	Flux balance analysis
form	Formate
FPT	Fixed pivot technique
FVS	Finite volume scheme
glu	Glucose
GPC	Gel permeation chromatography
HCC	Highly crystalline cellulose
lac	Lactate
L-HCM	Lumped hybrid cybernetic model
MBM	Macroscopic bioreaction model
ML-PBM	Multi-layered population balance model
MM	Michaelis-Menten

ODE	Ordinary differential equation
PBE	Population balance equation
PBM	Population balance model
RHS	Right hand side
ROI	Return-on-investment
SEC	Size exclusion chromatography
SPS	Steam-pre-treated spruce
SSF	Simultaneous saccharification and fermentation
UC-PBM	Unified cybernetic-population balance model

Chapter 1

Introduction

1.1 Background and Motivation

Production of biochemicals from renewable resources is a promising alternative to fossil fuels that pose various adverse effects on the environment [1–3]. Cellulosic plant biomass is considered an excellent resource for biochemical production, primarily because they are not derived from crops, but from forest or agricultural residues, including switchgrass [4], wheat straw [5], corn stover [6], rice straw [7], and poplar [3], among many others. The use of cellulosic biomass, therefore, helps to alleviate some issues associated with disposal of solid wastes, increase of trade deficits and declining energy security [2, 8, 9]. In this regard, consolidated bioprocessing (CBP) provides routes for direct conversion of cellulosic biomass to biochemicals using microbial cell factories, as the consolidation of cellulase synthesis, cellulose saccharification and fermentation of released sugars in a single unit

operation vastly enhances the process economics. Despite the outstanding potential, CBP is encumbered by the absence of native microorganism strains with crucial traits for efficient bioconversion of cellulose under industrial conditions, such as high titre and yield of target biochemicals, tolerant to toxic compounds and secretion of cellulases at high levels without compromising the fermentation capacity [10–14].

To that end, metabolic engineering is a potential strategy to develop an ideal CBP strain by integrating all the required biological functionalities into a host strain. CBP strain developments through metabolic intervention strategies could benefit from an integrated quantitative understanding of the metabolic capabilities of the organism but, new target organisms are often less well characterised [14], especially against diverse and sophisticated cellulosic properties. Besides, a trial-and-error type approach for strain developments with an incomplete understanding of the metabolic mechanisms consumes substantial resources and time, and may lead to haphazard process designs [15]. Such limitations in CBP strain developments call for a holistic *in silico* approach to examine the metabolic characteristics of the target host organisms supported with a predictive capacity of dynamic shifts in the metabolism to various environmental and genetic perturbations. *In silico* tools which can be used for this purpose should be able to capture the process dynamics to enable accurate prediction of productivity as well as yield [1, 16–19]. However, development of reliable dynamic CBP models is challenging due to increased complexity arising from coupled dynamics of saccharification and fermentation.

The enzymatic saccharification or hydrolysis of cellulose is riddled with its own set of challenges. Rate of conversion of cellulose to valuable monomeric and oligomeric products is often significantly slowed down after an initial rapid but short-lived phase. A mechanistic understanding of the underlying mechanisms and key driving factors has yet

to be fully established due to complexities in cellulose structure and properties, as well as hydrolysis reactions, impeding the development of effective processes [20–23]. The sources of complexity include recalcitrant nature and ultrastructural organization of celluloses, mixture of insoluble celluloses and dissolved enzymes, multiple elementary steps for substrate-enzyme interactions, concerted action of enzymes with different modes of scission, and polydispersity of cellulose polymer chains. Owing to the interplay between the complex process dynamics above, fine control of the enzymatic hydrolysis of cellulose is challenging, which can be effectively addressed by employing predictive mathematical models. Despite past efforts, most models developed thus far lack the ability to predict the foremost critical aspect of cellulose enzymatic hydrolysis, which controls the productivity of the process, i.e., the rate slowdown, and as a result, enzymatic hydrolysis is often poorly represented in existing models.

On the other hand, fermentation of soluble sugars by microorganisms is aided by intricate cellular metabolism across vast networks of metabolic reactions. Moreover, in a CBP setting, both saccharification and fermentation of celluloses are highly interdependent as cellulolytic microorganisms consume fermentable sugars for the energy required to secrete cellulases, which in return facilitates saccharification to maintain a constant supply of fermentable sugars for metabolism. Consequently, without a fundamental understanding of all critical aspects of the process, a rational reaction engineering of enzymatic hydrolysis and subsequent fermentation process, as well as any metabolic interventions to engineer ideal CBP strains, is difficult. Such limitations call for predictive modelling to promote a mechanistic understanding of CBP of cellulose. However, existing models do not represent the interdependent dynamics of saccharification and fermentation in CBP in a mechanistic manner.

This work proposes the use of two contrasting but complementing modelling tools – population balances and cybernetic models, to simulate saccharification and fermentation, respectively, in CBP of cellulose. Like the usual partial differential equations of conservation of mass, momentum and energy, population balances deliver the fourth governing equation that conserves the transport of dispersed phase entities [24, 25]. Population balance model (PBM) is an elegant mathematical tool to describe the temporal evolution of dispersed phase entities and specific ways through which the dispersed entities change their states, such as polymeric substrates and specific modes of enzymatic scission, respectively. Therefore, PBM is used as the basis for the development of the Multi-Layered Population Balance Model (ML-PBM) in this research to simulate the enzymatic hydrolysis of cellulose, where all critical features of the process are mechanistically and independently reflected, based on first principles wherever possible. The ML-PBM is specially formulated to manifest and explain the potential primary mechanisms of the cellulose hydrolysis rate slowdown. In tandem, the use of the PBM necessitates assessment and customization of numerical solution techniques specifically to solve depolymerization population balance equations (PBEs) because analytical solutions are often limited to simplistic circumstances. Conversely, cybernetic models view the microorganisms as sentient beings that encompass specialized regulatory mechanisms used to optimally control the allocation of finite internal resources to steer the metabolism to fulfil certain objectives, which often gravitate towards ensuring the survival of the microorganisms [26]. Developed across many iterations over the past three decades [17], cybernetic models evolved from imposing minimalistic perspective on cellular metabolism via single lumped reaction routes from substrates to products [27–32], to a more advanced metabolic network-based cybernetic models [33–38]. The latter is particularly attractive for the application in this work, which not only aims to benefit bioprocess optimizations through reaction engineer-

ing, but also to aid metabolic interventions to engineer optimal strains. Lumped Hybrid Cybernetic Model (L-HCM) [36–38] specifically stands out as it allows accurate dynamic metabolic modelling with high resolution at the network level, yet it remains computationally efficient and requires minimal experimental measurements for implementation.

Inspired by the works of Ho et al. [39, 40], population balances (ML-PBM) and cybernetic models (L-HCM) are coupled to establish a comprehensive framework to represent the dynamics of CBP of cellulose. The resulting framework, named the Unified Cybernetic-Population Balance Model (UC-PBM), equipped with the unification of various process features, allows a more fundamental analysis of CBP of cellulose. Unlike any existing frameworks, the coupling of the ML-PBM and the L-HCM presents a unique closed-loop network-level control mechanism to portray the interdependent dynamics of CBP of cellulose. In overall, accurate, detailed network-based simulation of complex dynamics of CBP enabled by the UC-PBM offers itself as a promising tool for both strain engineering and bioprocess optimizations.

1.2 Objectives

Alluding to the motivations in the preceding section, the main objectives of this research are prescribed chronologically as follows:

1. To adapt and apply various potential solution techniques to solve depolymerization PBEs, followed by a detailed performance assessment of the techniques. This includes the exploration of the inherent characteristics of each technique and attempts to improve solution accuracy specific to the nature of this work. The best-suited

solution technique chosen from this study will be used for subsequent works.

2. To develop a population balance modelling framework for enzymatic hydrolysis of celluloses inclusive of all the inherent critical characteristics of the process. This includes detailed parametric analysis and validation of the model against various experimental data from the literature. The resulting model will be used to determine the potential rate slowdown mechanism and other insightful process fundamentals. The model developed here will be used to fulfil subsequent objective.

3. To develop an overall comprehensive framework for CBP of cellulose by coupling population balances and cybernetic models. This includes detailed parametric analysis and validation of the model against various experimental data from the literature. The resulting model will be used for a detailed analysis of CBP at a fundamental level to reveal new and useful findings on the metabolism of a chosen CBP host strain.

1.3 Organization of the Thesis

The thesis is organized as shown in Figure 1.1 and the details are as follows. In Chapter 1, a brief background on CBP of cellulose and the need for modelling frameworks is established. A summary of the research scheme and the breakdown of objectives that must be fulfilled to arrive at the thesis is also explained. In Chapter 2, a thorough review of the challenges associated with modelling CBP of cellulose, existing modelling frameworks available in the literature and their limitations are presented. This is followed by a detailed exposition of the efforts undertaken in this work to address the research gaps.

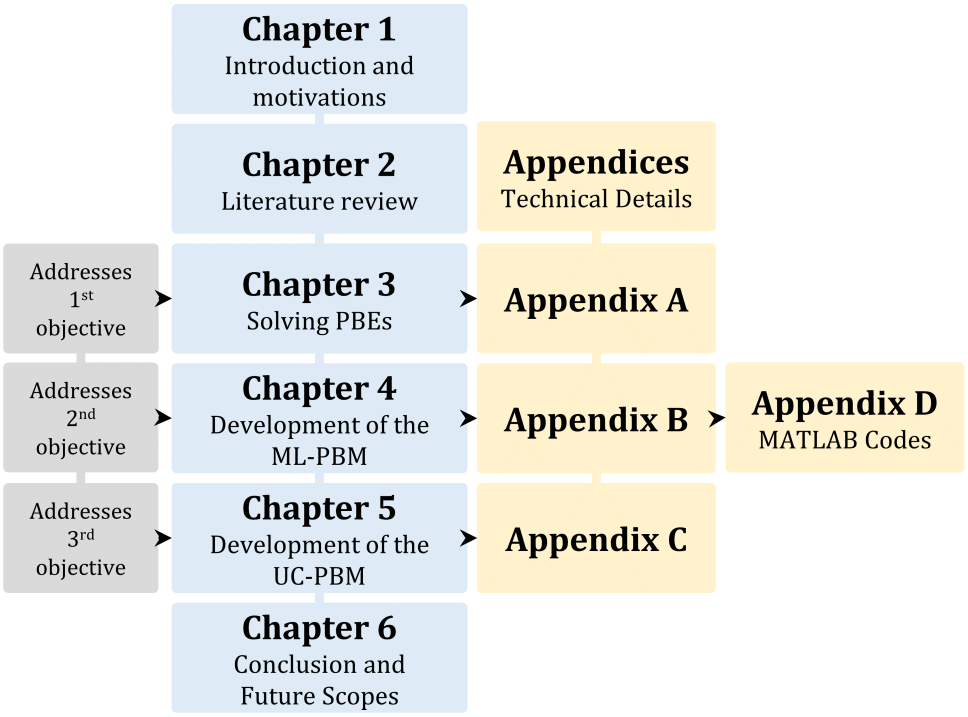

Figure 1.1: Chart illustrating the thesis structure.

Chapter 3 presents the study of various sectional techniques to solve depolymerization PBEs. Subsequently, the exhaustive development of the ML-PBM to model enzymatic hydrolysis of cellulose is presented in Chapter 4. In Chapter 5, the formulation of the comprehensive UC-PBM to model CBP of cellulose is deliberated. Overall, Chapter 3 – 5 incrementally demonstrates the fulfilment of the three research objectives. Lastly, Chapter 6 presents an overall conclusion to the research and recommendations for future works.

Due to the computational nature of this work that involves mathematical formulation of model frameworks, the research methodologies are embedded in the respective chapters of the thesis to enhance readability. Highly technical details of the model formulations are relegated to Appendices, where Appendix A contains details pertaining to the numerical

solution techniques to solve depolymerization population balances, Appendix B includes information concerning the ML-PBM, Appendix C organises materials related to the UC-PBM and Appendix D consolidates all MATLAB codes written for this research.

Chapter 2

Literature Review

2.1 Modelling Consolidated Bioprocessing of Cellulose

Optimal design of microbial strains and process configurations to maximize the productivity and the economic viability of CBP can be greatly facilitated by employing predictive mathematical modelling frameworks. CBP of cellulose by microbial cell factories is best resolved by models that integrate process fundamentals of both saccharification of distributed heterogeneous cellulose and regulation of fermentation at the metabolic network level. However, such models are often overlooked because of the associated difficulty in maintaining a reasonable computational efficiency and mechanistic accuracy of dynamic biological systems. Such limitations call for a new modelling approach that economizes on computational intensity without compromising the representation of critical factors of

the process.

In this research, the requisites for mechanistic modelling of CBP are satisfied by addressing the saccharification and the fermentation components separately using population balances and cybernetic models, respectively. Their integration thereof delivers an all-encompassing framework that closely resembles the dynamics of CBP. In the forthcoming subsections, the challenges of modelling the separate components of CBP of cellulose, as well as the shortcomings of past efforts to address the challenges, are extensively reviewed. Subsequently, the research scheme of this work to address the knowledge gaps are deliberated in Section 2.2.

2.1.1 Population Balance Modelling for Depolymerization Processes

In principle, saccharification of cellulosic polymers or depolymerization is a form of breakage process. The fundamental understanding of the breakage phenomenon is of critical importance, not only in polymer science, but also in diverse fields including cell biology, geology, agriculture engineering, etc. To model the particle breakage phenomenon, the Population Balance Equation (PBE) that describes the spatial and temporal dependence of the number density for a dispersed phase entity, is the fundamental governing equation [24, 25]. However, the reliability of the solution techniques employed is an essential criterion to model the breakage phenomenon with adequate accuracy. Due to the complexity exhibited by the breakage PBE in the form of a linear partial integro-differential equation, generalized analytical solutions are a rare find and exist only for very specific cases as elucidated by Ziff and McGrady [41, 42]. In this regard, numerical solutions are

often necessary [43–45].

To enable utilization of the PBE for a wider range of applications, many numerical solution techniques had been devised to date. The three main classes of solution techniques, apart from the conventional finite difference [46] and the finite element techniques [47], are the method of successive approximations [25], the method of moments [48–51] and the sectional methods [44, 52–59]. Although all these methods have been demonstrated to be successful in solving the PBE, the choice of the solution technique should support the necessary requisites of a specific application. While it is important to accurately predict the particle distribution along the internal coordinate of choice, it is equally vital to preserve the fundamental moment-related properties at the same time. For breakage phenomena that involve particulate systems, the fundamental moment-related properties are the total number of particles and the total mass [44, 55]. For this reason, the sectional technique is an apt and convenient choice as this technique is devised to preserve a number of selected properties while at the same time yields the complete number density at a reasonable accuracy [44]. Although the method of moments is comparatively more computationally efficient, it is not capable of predicting the transients of the complete number density accurately. Additionally, the sectional techniques meet mid-way between maintaining the accuracy of the predictions and the computational efficiency [52], thus rendering it an attractive choice for practical day-to-day usage in the industry. Alternatively, Monte Carlo methods could also be used to solve PBEs involving depolymerization processes [60–62]. The use of Monte Carlo methods becomes inevitable when the associated model is highly complex [63]. However, the high computational intensity associated with the Monte Carlo simulation renders it unsuitable for general usage and model-based optimizations [62].

The foregoing discussion primarily focused on the general breakage phenomenon

that can be entirely approximated by continuous distribution kinetics. However, several important breakage phenomena are inherently discrete in nature, e.g., polymer degradation [64–66] and dynamics of molecular clusters in the aerosol domain [67]. In this case, a fully discrete PBE is established by forming a set of Ordinary Differential Equations (ODEs) for each integer multiples of a single structural unit of the particulate system [67]. The solution of the fully discrete set of ODEs is often not practical due to the typical wide range of particle size in many particulate systems, which translates to extremely large number of ODEs to be solved numerically [43, 45, 54, 67, 68]. Although the fully discrete solutions are impractical to be applied directly, they serve as a benchmark for the performance of other short-cut solution techniques. The impracticality of numerically solving the fully discrete PBE for large systems calls for the development of solution strategies that strike a balanced compromise between solution accuracy and computational efficiency.

2.1.2 Modelling Frameworks for Enzymatic Hydrolysis of Cellulose

A full mechanistic understanding of enzymatic hydrolysis of cellulose has been elusive, owing to the effects of various confounding factors on the process dynamics resulting in excessive variability, which cannot be readily discerned through conventional experimental approaches. As an alternative, models to elucidate the fundamentals of enzymatic hydrolysis of cellulose should necessarily be physicochemical-based mechanistic models that independently account for the key factors of the process.

The modelling of enzymatic hydrolysis of cellulose is arduous due to the complex

structural arrangements of cellulosic biomass and its disposition to any form of degradation. For instance, cellulosic biomass encompasses structural arrangements spanning across multiple scales. At the microscopic scale, cellulose is a linear, homopolymer of anhydrous glucose molecules bound together by β-1,4 glycosidic bonds, with cellobiose (disaccharide) as the basic repeating unit. Subsequently, the cellulosic polymers are assembled into structures of increasing scales, starting from elementary fibrils, followed by microfibril or fibril aggregates and lastly microfibril aggregates or macrofibrils [8, 69]. Moreover, high insolubility of cellulose particles in water forces the enzymes to operate at the liquid-solid interface to depolymerize cellulose chains [3], which renders the process many orders of magnitude slower as the cellulose accessibility by the enzymes is limited to the surface polymers [69,70]. Although cellulose hydrolysis is generally viewed as a surface ablation process, the possibility of enzymes diffusing through the pores and hydrolysing internal polymers as discussed by past studies [71–73] or analogous manifestations should not be completely ruled out.

Furthermore, strong hydrogen bonds between cellulose chains form crystalline regions that are more impervious to degradation than amorphous regions, but the possible arrangement of the two regions in the cellulose particles in unison with the observed process dynamics remains elusive. In addition, the hydrolysis of cellulose is catalysed by three major classes of enzymes, namely endoglucanases (EG) that randomly cleave any β-1,4 glycosidic bonds, exoglucanases or cellobiohydrolases (CBH) that processively cleave cellobiose dimers from chain-ends and β-glucosidases (BG) that hydrolyse cellobiose and other soluble oligomers (DP < 7) to glucose monomers [8, 69]. Given that each enzyme has a specialized mode of scission, the concerted action of the enzymes exhibits strong synergism, which is only evident when the complete cellulose chain length

distributions are considered. Lastly, pre-treatment processes to improve the accessibility of enzymes to the cellulose fractions in lignocellulosic biomass prior to hydrolysis are often performed at harsh conditions (elevated temperatures and pressures, presence of acidic/alkaline medium, etc.) [2, 8, 20], which can severely disrupt the properties of the native celluloses, and its effects on the enzymatic hydrolysis remain poorly understood.

To aid sound cellulose bioprocessing, models for enzymatic hydrolysis should essentially uncover insights into the critical, yet uninformed process mechanisms. Various modelling approaches for simulating enzymatic hydrolysis of celluloses have been developed so far. However, currently available models do not embody all the essential aspects of the process as discussed above. For example, a so-called enzyme-centric approach decomposed the enzyme-substrate interactions into multiple elementary steps, but neglected the distributed nature of the cellulose polymer chains [72, 74–78]. The complete distribution of cellulose chains and different modes of scission were considered in other more structured models, which however, did not account for substrate morphology to demonstrate the heterogeneity of the insoluble celluloses [79–81]. A few sophisticated models have incorporated detailed substrate-enzyme interactions, substrate morphology and cellulose chain distributions [43, 82–85]. Despite some promising results, these recent models also lack the ability to accurately capture the complex and delicate aspect of system dynamics over the entire period of enzymatic hydrolysis process, especially the sharp decline in rate after an initial rapid phase [69, 85–90]. For instance, Levine et al. [84] utilized a complex random sequential arrangement to model the adsorption of the enzymes, and while the model fit to the initial rapid phase is commendable, the rate slowdown at the late phase of hydrolysis was not predicted adequately. Similarly, the ability of the model by Huron et al. [83] to predict the rate slowdown is doubtful as validation with experi-

mental data is lacking for late phases of the process. The mechanistic model developed by Griggs et al. [43, 82] which was later calibrated and validated with the experimental data by Nag et al. [85] assigned different rates of hydrolysis to two separate cellulose populations, i.e., amorphous and crystalline celluloses, to simulate the effect of different enzyme accessibility to the amorphous and crystalline regions. Although it was not their intention to explicitly model the amorphous and crystalline celluloses, doing so was empirically sufficient to match the rate slowdown to a certain extent. To the best of the author's knowledge, none of the currently available models provides a mechanistic basis for predicting this slowdown phenomenon in the enzymatic hydrolysis of celluloses. An up-to-date review of the existing kinetic models of enzymatic hydrolysis that highlights the limitations of the models and motivation of this work is given in Table 2.1.

Table 2.1: Review of existing models for the enzymatic hydrolysis of celluloses, benchmarked against the core features of the ML-PBM developed in this research.

Ref.	Model features				Core rate slowdown mechanism	Remarks
	Distributed cellulose chain lengths	Multiple enzyme-substrate interaction	Multiple modes of enzymatic scission	Insoluble substrate morphology		
Kadam et al. [91]	n.a.*	n.a.	n.a.	n.a.	Product inhibition	Michaelis-Menten (MM) based kinetics
Peri et al. [90]	n.a.	n.a.	n.a.	n.a.	Adsorption, product inhibition	MM-based kinetics
Levine et al. [84]	√	√	√	√	Accessible substrate surface area and cellulose fraction	Fails to trace the rate slowdown adequately at late phases
Bezerra et al. [87, 88], Bezerra and Dias [89]	n.a.	n.a.	n.a.	n.a.	Product inhibition	MM-based kinetics
Praestgaard et al. [77]	√	√	n.a.	n.a.	Decomplexation, enzyme inactivation	–
Hosseini and Shah [79, 80]	√	n.a.	√	n.a.	n.a.	Prediction of rate slowdown is not apparent
Gao et al. [76]	n.a.	√	n.a.	n.a.	Decomplexation	–
Maurer et al. [75]	n.a.	√	n.a.	n.a.	Complexation	–
Shang et al. [74]†	n.a.	√	n.a.	n.a.	Complexation	–
Kumar and Murthy [92]†	√	n.a.	√	√	n.a.	Fails to trace the rate slowdown adequately at late phases

Continued on next page

Table 2.1 – (continued)

Ref.	Model features				Core rate slowdown mechanism	Remarks
	Distributed cellulose chain lengths	Multiple enzyme-substrate interaction	Multiple modes of enzymatic scission	Insoluble substrate morphology		
Griggs et al. [43, 82], Nag et al. [85]	√	√	√	√	Different reactivity to amorphous & crystalline cellulose populations	Empirical assignment of differential rates of hydrolysis to separate cellulose populations
Luterbacher et al. [72, 78]	n.a.	√	n.a.	√	Accessible substrate surface area and cellulose fraction	–
Lebaz et al. [81, 93, 94]	√	n.a.	√	n.a.	Product inhibition	Model only validated for initial phase ($t \leq 7$ h)
Huron et al. [83]	√	n.a.	√	√	Product inhibition, enzyme deactivation	Unable to predict rate slowdown without enzyme deactivation
Eibinger et al. [63]†	√	√	√	√	Decomplexation of CBH	Model only validated for initial phase ($t \leq 7$ h)
Tervasmäki et al. [95]	n.a.	n.a.	n.a.	n.a.	Empirical decrease in enzyme adsorption/activity, product inhibition	Empirical rate-decreasing factors are exclusive to each fed-batch sub-population
Liang et al. [96]	n.a.	√	n.a.	n.a.	Enzyme deactivation and empirical decrease in adsorption	Empirical assignment of decreasing rate coefficients with respect to conversion

Continued on next page

Table 2.1 – (continued)

Ref.	Model features				Core rate slowdown mechanism	Remarks
	Distributed cellulose chain lengths	Multiple enzyme-substrate interaction	Multiple modes of enzymatic scission	Insoluble substrate morphology		
Caro et al. [97]	n.a.	n.a.	√	√	Empirical setting of parameter to reflect the hydrolysable portion of cellulose	Highly empirical MM-based kinetics
Lischeske and Stickel [98]	n.a.	n.a.	n.a.	n.a.	Adsorption with varying degree of accessibility to facile and recalcitrant glucans	Phenomenological assignment of adsorption coefficients to mimic varying degree of accessibility
Nill and Jeoh [99]	n.a.	√	n.a.	n.a.	Depletion of productive enzyme binding sites on cellulose	Phenomenological use of separate exponential decay rates for depletion of productive binding sites during the initial and the late phase
This work: ML-PBM [100]	√	√	√	√	Heterogeneity in cellulose DP & crystallinity, substrate morphology	Predicts the full transient of the process mechanistically, substrate morphology can be inferred

*n.a. – not available.

†Stochastic model solved via Monte Carlo simulation involves high computational intensity,unsuitable for general usage and model-based optimizations [62].

2.1.3 Modelling Frameworks for Metabolic Systems of Cellulolytic Microorganisms

As microorganisms cannot directly assimilate celluloses, cellulose polymer chains are first broken down to simple sugars through enzymatic hydrolysis. Here, commercial depolymerizing enzymes could either be dosed externally or cellulases could be synthesized *in situ* by cellulolytic microorganisms. The latter, where microorganisms are responsible for synthesizing the cellulases in the abiotic environment and concurrently ferment the simple substrates into bioproducts, is referred to as CBP of cellulose. Unlike the microbial utilization of soluble substrates, the biochemistry of the metabolism of cellulolytic microorganisms, or CBP strains, utilizing cellulose is rather elaborate. As the cellulase synthesis poses a substantial metabolic burden, CBP strains generally encompass multiple phenotypic functions that induce regulation of carbon sources and internal cellular resources between cell growth and cellulase synthesis. Moreover, exponential cell growth is not very apparent during cellulose utilization as the fermentable substrates are not present in excess but are produced and consumed instantaneously over the full span of conversions [101]. Understandably, a dynamic regulation of resources exists in the metabolism of CBP strains to maintain the balance between secreting cellulases to produce fermentable substrates and to ensure survival through cell growth.

Despite the development of various approaches to model metabolic systems, currently existing models do not embody all the essential traits for modelling metabolism of CBP strains. For instance, unstructured kinetic models haphazardly truncate the metabolic pathways into a single-step reaction extending from substrates to products [16,36], leading to the loss of vital quantitative information on flux distribution. Moreover, most of the

existing unstructured models [102–107] that rely on grossly simplified Michaelis-Menten (MM) and Monod kinetics, only cater to simultaneous saccharification of fermentation (SSF) of cellulose by non-cellulolytic microorganisms with exogenous addition of fungal cellulases for saccharification. Constraint-based approaches such as the flux balance analysis (FBA) [108] and elementary mode analysis (EMA) [109] do not reflect dynamic shifts in the metabolism as response to external stimuli, whereas the dynamic flux balance analysis (DFBA) [110] and the macroscopic bioreaction models (MBM) [111] include the metabolic responses to external perturbations by varying substrate uptake rates but not the intracellular flux distribution. Moreover, DFBA and FBA manifest the metabolism through a single optimal solution of flux distribution among many equally possible alternate solutions [16]. Conversely, the Dynamic Reduction of Unbalanced Metabolism (DRUM) [112] implicitly accounts for the dynamic metabolic shift by dividing metabolic networks into subnetworks, which are then represented by separate simple proportional kinetics. Nevertheless, all the models above disregard the regulatory mechanism of cellular resources and metabolic burden, which are essential factors that influence the metabolism of CBP strains. In addition, all the existing models for metabolic systems above do not include provisions to trace the effects of dynamically evolving polymeric environment of the cellulose during CBP. As CBP strains are known to regulate the extracellular cellulase levels as required based on the level and nature of the encountered substrates [14], one can expect that the varying polymeric environment during the hydrolysis of cellulose to actively dictate the metabolism. In return, the regulation of extracellular cellulases is governed by the intracellular regulation at network level, subject to the metabolic burden. Therefore, all the limiting features of the existing models above are deemed crucial to capture the metabolic nuances of CBP strains.

2.2 Efforts to Address Research Gaps

2.2.1 Employing Sectional Techniques to Solve Depolymerization Population Balances over a Hybridized Discrete-Continuous Mesh

This research rests on the idea of addressing the saccharification of cellulosic polymers using the population balances, which require a fit-for-purpose, tailored numerical solution strategy. As raised in Section 2.1.1, breakage PBEs for large systems require a solution strategy that offers adequate accuracy whilst economizing on the computational load. As far as the sectional techniques are concerned, one of the possibilities is to employ a discrete-continuous mesh by splitting the particle size domain (or in the polymer context, degree of polymerization; DP) into a discrete domain at the small size range and a continuous domain at large size range [54, 55, 67, 68]. Doing so allows the user to predict the populations of the monomers and oligomers at the small size range distinctly using the discrete mesh and, at the same time, economize on the computational load by approximating the large size range via a continuous mesh. The former has great experimental significance as the individual concentrations and the properties of small oligomers may differ considerably and they may elute as separate peaks when performing analytical chromatography [113]. Furthermore, in cases where the recombination of monomers and small oligomers occurs [65,114], the use of a discrete-continuous mesh is inevitable to capture the detailed molecular dynamics at the small size region. Although Kostoglou [68] gave the analytical solutions to the discrete-continuous model for chain-end scission (by approximating the continuous region using a first-order Taylor series expansion of the con-

tinuous fundamental model [115]), this approach lacks versatility as it cannot be readily extended to account for more complex molecular phenomena, e.g., the simultaneous occurrence of breakage and aggregation. As the sectional techniques were originally developed for solving continuous distribution kinetics, implementing a fully continuous formulation for problems with a semi-continuous/semi-discrete nature would pale in accuracy and efficiency. Nevertheless, applications that are not inherently discrete (e.g., emulsions, granulations, crystallization) could still benefit from the fully continuous implementation of the sectional techniques.

As the use of the sectional techniques is more versatile and practical, this work examines in detail the implementation of the sectional techniques for solving the two most commonly encountered depolymerization phenomena, i.e., chain-end and random scissions, on a discrete-continuous mesh. Three state-of-the-art sectional techniques are studied here, i.e., the Fixed Pivot Technique (FPT) developed by Kumar and Ramkrishna [55] and further explored by Ho et al. [54], the Cell Average Technique (CAT) established by Kumar et al. [44], and the Finite Volume Scheme (FVS) devised by Saha et al. [53]. In this work, the CAT and the FVS are examined for the first time on the discrete bond-breaking processes of depolymerization. The prevailing shortcomings in the existing implementation for both chain-end and random scissions that hitherto have not been uncovered are highlighted and modifications are proposed to improve the performance of the sectional techniques. Particularly, the issue of an abrupt under-prediction in the number density for chain-end scission at the discrete-continuous boundary is revealed and solved for the first time. Not only that, a continuous analogue of the discrete random scission stoichiometric kernel that yields a far superior prediction is proposed. The detailed analysis of the sectional techniques are deliberated in Chapter 3. Suggestions for

the choice and the implementation of sectional techniques for various circumstances in a depolymerization context are also provided.

The foregoing efforts, which are intended to adapt and assess various available solution techniques to solve depolymerization PBEs, address the first research objective. This research finds the FPT that offers adequate accuracy, good computational efficiency, and ease of implementation, is the best choice to solve depolymerization PBEs. Subsequently, the FPT is used as the basis for the development of an elaborate modelling framework for enzymatic hydrolysis of cellulose, which is introduced in the following section.

2.2.2 Introducing Multi-Layered Population Balance Model for Enzymatic Hydrolysis of Cellulose

Using the population balance modelling framework and the solution strategy emphasised in the previous section, a new approach for modelling the enzymatic hydrolysis of cellulose is developed. Various mechanistic aspects of cellulose hydrolysis as discussed in Section 2.1.2 are explicitly incorporated, including multiple elementary steps in enzyme-substrate interactions, different modes of enzymatic scission and complete cellulose chain distribution. More importantly, the new model has an elaborate hierarchical scheme to account for non-uniform distributions of cellulose properties across the structural layers of cellulose particles and the concomitant morphologies. Unlike existing models, the resulting model, which is termed the Multi-Layered Population Balance Model (ML-PBM) in this work, not only enables an accurate prediction of the rate slowdown with a single set of model parameters, but also yields mechanistic insights into the sophisticated aspects of the process, i.e., the extent of microfibril aggregation and the resulting cellulose particle

morphology, arrangement of probable cellulose chain distribution, physicochemical effects of pre-treatment steps in cellulose preparations on cellulose properties and its subsequent impact on hydrolysis.

The core elements of the ML-PBM are illustrated in Figure 2.1 and the motivation of the ML-PBM is reflected in Table 2.1, where its core features are compared to those of major cellulose hydrolysis models from the literature. In line with global efforts to promote the growth of technologies based on renewable resources, the ML-PBM proposed in this work is not only useful to provide guided and rational optimizations to enzymatic hydrolysis of celluloses as well as to guide the preparation of commercial celluloses and biosynthesis of celluloses for various applications, but it also serves as an essential element to be integrated in other computational frameworks to simulate cellulose-related processes (e.g., fermentation process). The development of the ML-PBM addresses the second research objective and is explored in detail in Chapter 4.

2.2.3 Introducing Unified Cybernetic-Population Balance Model for Consolidated Bioprocessing of Cellulose

The elaborate metabolic systems of cellulolytic microorganisms accompanied by the complex cellulose characteristics renders the development of CBP model highly challenging. To address the inherent limitations of existing modelling approaches as seen in Section 2.1.3, this work draws inspiration from the studies of Ho et al. [39, 40], where population balances and cybernetic models were interlinked to simultaneously account for both the distributed dynamics of polymeric saccharification and regulation of cellular metabolism during fermentation. Here, this research addresses the limitations of the previous works

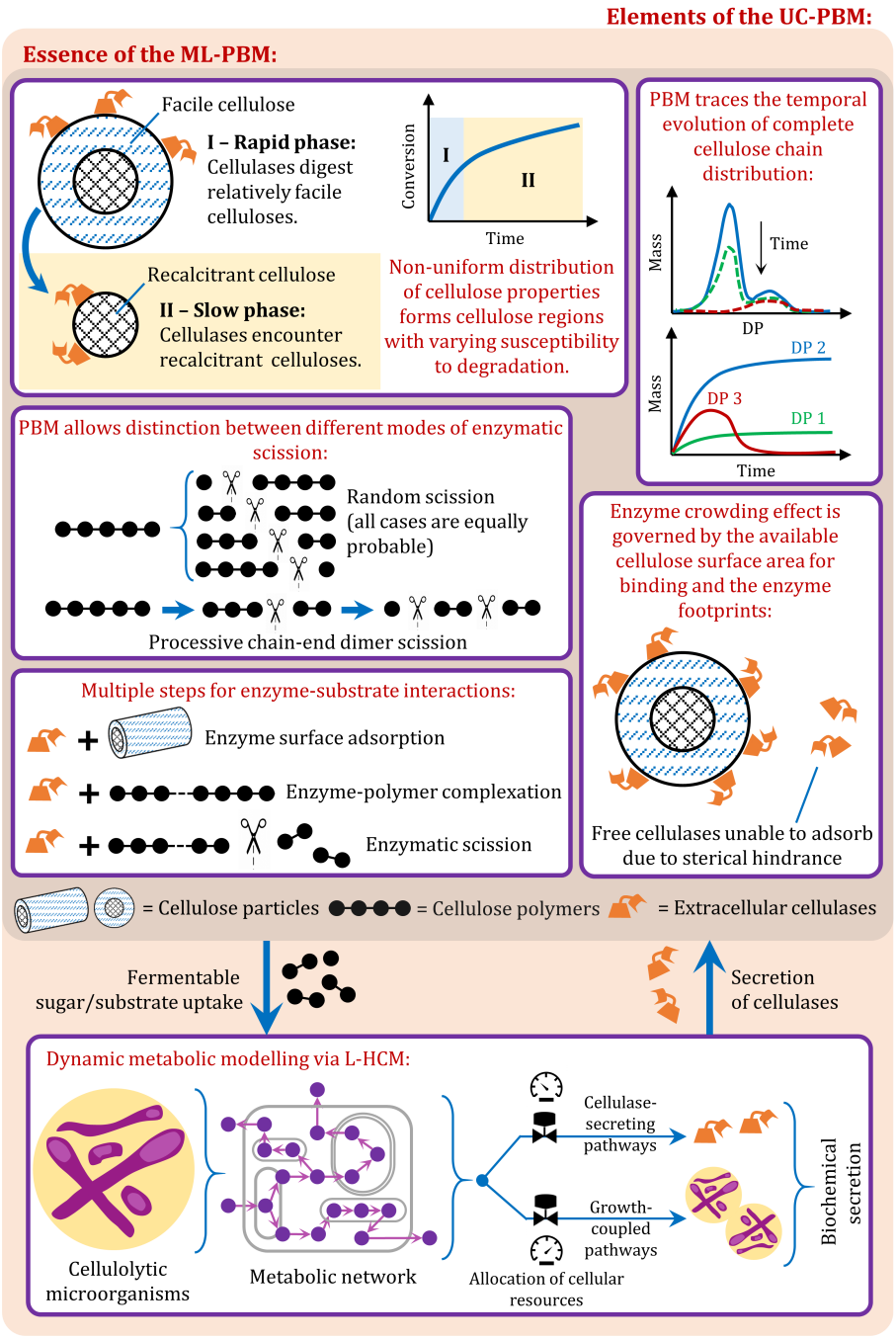

Figure 2.1: A schematic representation of the ML-PBM and the UC-PBM concepts. The closed-loop regulatory mechanism in the metabolism of cellulolytic microorganisms utilizing cellulose is realized through the coupling between the ML-PBM and the L-HCM.

that relied on a simplified unstructured lumped cybernetic model (LCM), only catered to the hydrolysis of soluble natural polymers (i.e., starch) devoid of extensive recalcitrance, and formulated network-independent extracellular enzyme balances. For CBP of cellulose, the ML-PBM [100] introduced in the previous section is employed, where the model offers a mechanistic interpretation for the recalcitrance of cellulose against degradation, which is crucial to gauge the efficacy of CBP strains to consume cellulose. Moreover, the ML-PBM also discerns the dynamic changes in the properties of residual celluloses and the composition of released fermentable sugars and other soluble oligomers, which are the primary impetus that triggers the regulatory mechanisms in CBP strains. For the fermentation component, the Lumped Hybrid Cybernetic Model (L-HCM) [36–38] is utilized, where the L-HCM offers good estimates of diverse metabolic behaviours with minimal experimental data for implementation and includes dynamic cellular regulations and flux distributions at the network level, without considerably adding on to the computational intensity of the overall framework. By coupling the ML-PBM and the L-HCM, a closed-loop control mechanism is featured between global regulation of extracellular cellulases and local intracellular regulation of fermentation at the network level. Unlike any existing models, the resulting model, which is termed the Unified Cybernetic-Population Balance Model (UC-PBM), closely mimics the interdependent dynamics of CBP through the closed-loop interactions. The features of the UC-PBM are illustrated in Figure 2.1. Notably, the UC-PBM allows detailed analysis of the dynamic substrate-cellulase-microorganism interactions and is particularly useful for initial studies on new emerging CBP strains.

The advent of the UC-PBM tackles the third research objective, where the effectiveness of the UC-PBM is demonstrated through a case study of *Clostridium thermocellum*. This microorganism has been researched for about 70 years [116] for its potential as a

promising CBP strain, with new findings continue to be reported [117, 118], and yet a complete understanding of its physiology remains elusive. The UC-PBM not only provides a better understanding of the nature of the flux regulation and the excellent cellulolytic ability in *C. thermocellum*, but also reveals the possibility of previously unrecognized cellulosome-decoupled growth metabolic diversity that explains growth-independent down-regulation of cellulases under nutrient-rich conditions, knowledge of which is equally important for rational strain design for improved performance. Given the potentials of this unique framework, the UC-PBM must not be ignored in future investigations of ambiguous metabolic traits of *C. thermocellum* (e.g., premature growth cessation, overflow metabolism, etc.), or other promising CBP strains, under diverse environmental conditions. In general, the UC-PBM is not only useful to examine the metabolic characteristics of microorganisms, but it is also imperative for sound metabolic interventions for effective strain developments, bioprocess designs and optimizations. The detailed exploration of the UC-PBM is presented in Chapter 5.

Chapter 3

Solving Depolymerization Population Balances Using Sectional Techniques on a Discrete-Continuous Mesh

3.1 Preface

As alluded to in Section 2.2.1, this chapter presents a comparative assessment of sectional techniques (FPT, CAT and FVS) to solve depolymerization population balances over a discrete-continuous mesh. Moreover, the performance of the CAT and the FVS for discrete bond-breaking depolymerization processes is explored for the first time in this

work. In addition, the inherent shortcomings of the sectional techniques are identified, and solutions are proposed accordingly, where relevant. A major portion of this chapter is published in Ahamed et al. [119].

3.2 Theoretical Framework

3.2.1 Discrete-Continuous Mesh

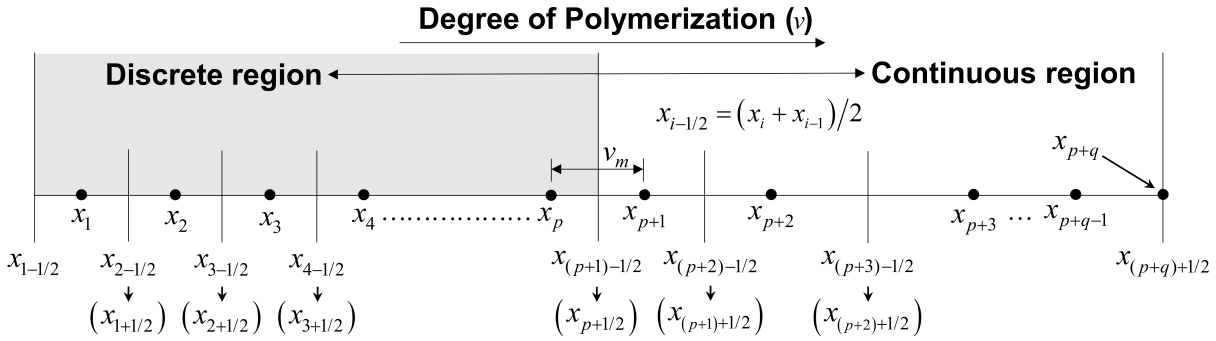

Figure 3.1: A graphical representation of a typical discrete-continuous mesh, where p denotes the number of grid points in the discrete region and q is the number of grid points in the continuous region. Alternative notations are displayed in the parentheses. Here, $x_{p+q} = x_{p+q+1/2} = N$, where N is the maximum DP of the polymer distribution and $v_m = 1$ is the DP of a monomer.

A typical discrete-continuous mesh is illustrated in Figure 3.1, where p and q are the number of grid points in the discrete and the continuous region, respectively. The value of p can be chosen such that it fulfils the requisites of the application [54]. The total number of discretized sections throughout the entire domain is denoted by $I = p + q$, where each section is represented by the grid points, x_i. Taking a cue from Kumar and Ramkrishna [55], Kumar et al. [44] and Saha et al. [53], the grid points are encompassed by the boundary points $[x_{i-1/2}, x_{i+1/2}]$, where the boundary points are the midpoints between

the adjacent grid points, i.e., $x_{i-1/2} = (x_i + x_{i-1})/2$, and the width of each section is given as $\Delta x_i = x_{i+1/2} - x_{i-1/2}$. In the discrete region, the grid points are spaced one monomer unit apart and $x_1 = v_m$, where $v_m = 1$ is the DP of a monomer. Likewise, the last grid point is $x_{p+q} = x_{(p+q)+1/2} = N$, where N is the maximum DP of the polymer distribution. To maintain the width of each section in the discrete region to be one monomer unit, the lower boundary of the first section is chosen as $x_{1-1/2} = 0.5$ in this work. The width of the sections in the continuous region can be chosen arbitrarily, where it can take on a uniform or a non-uniform grid depending on the nature of the application [44], with one logical constraint that the grid points should be spaced more than v_m apart. In this work, geometric grids with a common ratio of $r = (x_{p+q}/x_{p+1})^{1/(q-1)}$ are employed for the continuous region to cater for a broad distribution of DP, which is common for natural polymers. In this case, given the values of p, N and v_m, the value of q which ensures that the grid points in the continuous region are spaced at least v_m unit apart is given by Ho et al. [54]:

$$q < 1 + \left\{ \ln\left(\frac{N}{p + v_m}\right) \Big/ \ln\left(1 + \frac{v_m}{p + v_m}\right) \right\} \tag{3.1}$$

This meshing strategy is implemented for all the sectional techniques in this work.

3.2.2 General Discretized PBE Framework for Sectional Techniques

Few terminologies employed here require clarification to avoid ambiguity in the subsequent texts. In the polymer context, the terms 'population density' and 'molar concentration density' are used interchangeably as the latter can be associated with the number of molecules through the use of the Avogadro number [54]. Likewise, 'particle size' and

'DP' as well as 'particle' and 'polymer' are also denoted interchangeably throughout this chapter.

To facilitate the representation of different sectional techniques on a uniform framework, it is convenient to introduce the following general discretized PBE:

$$\frac{dC_i}{dt} = \overbrace{\sum_{a}\sum_{j=a}^{I}\eta_{ij}^{\text{sec}}k_jC_j}^{\text{Birth of DP } = x_i} - \overbrace{\varphi_i^{\text{sec}}k_iC_i}^{\substack{\text{Death of}\\ \text{DP } = x_i}}; \quad i = 2,3,\ldots,I \tag{3.2}$$

Here, C_i is the molar concentration of the polymer in the i-th section, I is the final section in the DP domain, η_{ij}^{sec} is the allocation function for polymers entering the i-th section due to the breakage of DP $= j$ where $j \geq i$, φ_i^{sec} is the weighted allocation function for polymers leaving the i-th section due to the breakage of DP $= i$, and finally k_i is the rate kernel for the breakage of polymers in the i-th section. In Eq. (3.2), the first summation over the domain a is included to maintain the generality of the equation for all three techniques, where a denotes the lower limit of the second summation. For FPT and FVS, $a = i$ only and therefore, the first summation disappears, whereas for CAT, $a = i-1, i$, and $i+1$. The superscript "sec" denotes sectional techniques. The temporal evolution of the monomer concentration is computed separately for all the sectional techniques for chain-end and random scission, respectively, as follows:

$$\frac{dC_1}{dt} = 2k_2^{\gamma}C_2 + \sum_{j=3}^{I}k_j^{\gamma}C_j \tag{3.3}$$

$$\frac{dC_1}{dt} = \sum_{j=2}^{I}\left(\frac{2}{x_{j-1}}\right)k_j^{\alpha}C_j \tag{3.4}$$

In this work, for simplicity, $C_i(t)$ is written as C_i, but it shall be understood that the molar concentration is an explicit function of time. Where the molar concentration density c_i is required instead, mean value theorem is applied on $C_i = \int_{x_{i-1/2}}^{x_{i+1/2}} c(v,t)\ dv$ to arrive at $c_i = C_i/(x_{i+1/2} - x_{i-1/2})$, and for the discrete region, $c_i = C_i$. One can apply Eq. (3.2) to the discrete-continuous mesh introduced in Section 3.2.1 by using the appropriate η_{ij}^{sec} and φ_i^{sec} given in the subsequent subsections for each sectional technique. Readers interested in the final discrete-continuous ODEs which can be readily integrated with suitable commercial ODE solvers are referred to Appendix A.1 – A.3.

3.2.3 Fixed Pivot Technique (FPT)

One of the most state-of-the-art sectional techniques for solving the breakage PBEs is the FPT [52]. The FPT was devised to conserve two important properties of the distribution [55]. In this regard, the total number of particles and the total mass of the system, characterized by the zeroth and the first moments, respectively, are usually chosen [44, 52–55]. The simplicity of the FPT lies in its polymer allocation function, where the polymers entering the i-th section due to the breakage of the parent polymers are directly allocated to two adjacent grid points in such a way that it preserves the two moments. Referring to the general discretized PBE in Eq. (3.2), for the FPT, $a = i$ only and the polymer allocation function is as follows:

$$\eta_{ij}^{\text{FPT}} = \int_{x_i}^{x_{i+1}} \left(\frac{x_{i+1} - v}{x_{i+1} - x_i}\right) b(v, x_j)\ dv + \int_{x_{i-1}}^{x_i} \left(\frac{v - x_{i-1}}{x_i - x_{i-1}}\right) b(v, x_j)\ dv \tag{3.5}$$

For the FPT, $\varphi_i^{\text{FPT}} = 1$. Here, x's are the grid points, v is the continuous DP and $b(v, x_j)$ is the stoichiometric kernel for the formation of daughter polymers with DP $= v$ due to

the breakage of parent polymers with DP $= x_j$. The stoichiometric kernels are exclusive for different modes of scission or breakage and are discussed in Section 3.3.

3.2.4 Cell Average Technique (CAT)

The CAT is another sectional technique that also conserves the total number of particles and the total mass of the system. In fact, the CAT was proposed to alleviate the FPT's issue with over-prediction at regions with steep number density variations [44]. Unlike the FPT, the CAT computes the average size of all the polymers that enter the i-th section due to the breakage of larger parent polymers. If this average size is larger than the representative size x_i, the polymers are shared between x_i and x_{i+1}, and if the average size is smaller than x_i, then the polymers are shared between x_{i-1} and x_i [44,52]. Although the CAT was shown to perform better than the FPT [44], Kostoglou and Karabelas [52] stated that this result may not be general as the performance of the sectional techniques is sensitive to the choice of the pivot sizes. Nevertheless, it is noteworthy to comprehend that the CAT was only tested for simple continuous particle breakage processes and never tested for an inherently discrete bond-breaking depolymerization process and that further substantiates the purpose of this present work. Referring to Eq. (3.2), for the CAT, $a = i-1, i$, and $i+1$, and the polymer allocation functions are given as:

$$\eta_{ij}^{\mathrm{CAT}} = \begin{cases} p_{i-1,j}^{(0)} \left(\dfrac{\bar{v}_{i-1} - x_{i-1}}{x_i - x_{i-1}} \right) H(\bar{v}_{i-1} - x_{i-1}); & \text{for } a = i-1 \\ p_{ij}^{(0)} \left[\left(\dfrac{\bar{v}_i - x_{i-1}}{x_i - x_{i-1}} \right) H(x_i - \bar{v}_i) + \left(\dfrac{\bar{v}_i - x_{i+1}}{x_i - x_{i+1}} \right) H(\bar{v}_i - x_i) \right]; & \text{for } a = i \\ p_{i+1,j}^{(0)} \left(\dfrac{\bar{v}_{i+1} - x_{i+1}}{x_i - x_{i+1}} \right) H(x_{i+1} - \bar{v}_{i+1}); & \text{for } a = i+1 \end{cases} \tag{3.6}$$

and $\varphi_i^{\mathrm{CAT}} = 1$, where

$$p_{ij}^{(s)} = \int_{x_{i-1/2}}^{m_j^i} v^s b(v, x_j)\, dv \tag{3.7}$$

The term $p_{ij}^{(s)}$ in Eq. (3.7) should not be confused with the meshing parameter, p, that denotes the number of grid points in the discrete region. Here, the upper limit of the integral is $m_j^i = x_i$ when $j = i$, otherwise $m_j^i = x_{i+1/2}$. $H(x)$ is the Heaviside step function that is defined as:

$$H(x) = \begin{cases} 1; & x > 0 \\ 1/2; & x = 0 \\ 0; & x < 0 \end{cases} \tag{3.8}$$

The average polymer size entering the i-th section is simply:

$$\bar{v}_i = \frac{M_i}{V_i} \tag{3.9}$$

where the total mass and the total number of particles entering the i-th section are expressed, respectively, as:

$$M_i = \sum_{j=i}^{I} p_{ij}^{(1)} k_j C_j \tag{3.10}$$

$$V_i = \sum_{j=i}^{I} p_{ij}^{(0)} k_j C_j \tag{3.11}$$

3.2.5 Finite Volume Scheme (FVS)

The mass conserving and number preserving FVS established by Saha et al. [53] is considered in this work, where two weight allocations are introduced to the birth and the death terms of the PBE to ensure the conservation of the zeroth and first moments. The

FVS was shown to predict the properties well for simple breakage processes even at a very coarse mesh [53]. Here, the performance of the FVS is examined for an inherently discrete bond-breaking depolymerization process on a discrete-continuous mesh. Referring to Eq. (3.2), for the FVS, $a = i$ only and the polymer allocation functions are given as:

$$\eta_{ij}^{\mathrm{FVS}} = \omega_j^{\mathrm{b}} p_{ij}^{(0)} \tag{3.12}$$

$$\varphi_i^{\mathrm{FVS}} = \frac{\omega_i^{\mathrm{b}}}{x_i} \sum_{j=1}^{i} x_j p_{ji}^{(0)} \tag{3.13}$$

where the weight allocation for preserving the total number of particles and the total mass is defined as:

$$\omega_j^{\mathrm{b}} = \frac{x_j \left(\sum_{i=1}^{j} p_{ij}^{(0)} - 1 \right)}{\sum_{i=1}^{j} (x_j - x_i) p_{ij}^{(0)}} \tag{3.14}$$

For $j = i = 1$, the weights ω_1^{b} and φ_1^{FVS} are zero [53], as the monomers do not undergo any further breakage into smaller sizes.

3.3 Case Study

3.3.1 Preamble

To demonstrate the performance of the sectional techniques, here, a polymer with a realistically broad DP distribution employed by Ho et al. [54] is chosen, where the polymer is assumed to be starch with glucose as the monomer and the initial distribution is given in the form of the gamma distribution:

$$c(v,0) = \frac{M_{\mathrm{T}}\Omega(v)}{\int_1^N \Omega(v)\left[(162v+18)\right]dv}; \quad \Omega(v) = \frac{v^{\lambda-1}\exp\left(-v/\phi\right)}{\phi^{\lambda}\Gamma(\lambda)} \tag{3.15}$$

where M_{T} is the initial total mass concentration of the polymer, $\lambda = \bar{M}_{\mathrm{n}}/(\bar{M}_{\mathrm{w}} - \bar{M}_{\mathrm{n}})$, $\phi = \bar{M}_{\mathrm{w}} - \bar{M}_{\mathrm{n}}$ and $\Gamma(\lambda)$ is the gamma function. The parameters for the above distribution are obtained from Breuninger et al. [120], where the number-average DP is $\bar{M}_{\mathrm{n}} = 4100$ and weight-average DP is $\bar{M}_{\mathrm{w}} = 5430$. The initial mass concentration, $M_{\mathrm{T}} = 10$ g/L and maximum DP, $N = 22496$, are chosen for this study.

As alluded to previously, both chain-end and random scissions are investigated and the solutions via the sectional techniques are benchmarked against the fully discrete (exact) solution. The final ODEs for fully discrete solutions for both scissions are presented in Appendix A.1. For clarity, superscripts γ and α are used to denote chain-end scission and random scission, respectively. The system of ODEs is integrated using the 'ode15s' sub-routine of MATLAB® R2016a. Non-negativity of the solution vector is ensured by invoking the 'NonNegative' option of the solver. The solutions are computed on a workstation equipped with an Intel® Core™ i5-7200U processor with a clock speed of 2.50 GHz and 16 GB of installed memory (RAM). Following Kostoglou and Karabelas [52], the molar concentration density is represented as the dimensionless population density:

$$\widetilde{c}(v,t) = \frac{x_0 c(v,t)}{\zeta_0(0)} \tag{3.16}$$

where $x_0 = \zeta_1(0)/\zeta_0(0)$ and the k-th order moment of the polymer population is given as:

$$\zeta_k(t) = \int_0^{\infty} v^k c(v,t)\, dv = \sum_{i=1}^{I} x_i^k C_i \tag{3.17}$$

In addition, the results are reported in dimensionless time $\theta = t/t_{99\%}$, where $t_{99\%}$ refers to the time required for 99% monomer production [54]. Likewise, the moments are normalized against the initial moments of identical order, i.e., $\zeta_k(t)/\zeta_k(0)$. Additionally, one can track the progress of depolymerization through the normalized version of the zeroth moment that represents the extent of the breakage [53].

In this study, the sectional techniques are assessed for their performance in predicting the population density, as well as the zeroth, first and second moments. The second moment is included due to its significance in polymer applications, i.e., for predicting the polydispersity index (PDI) [54]. The PDI describes the degree of non-uniformity in the distribution of the polymer and is expressed as $\mathrm{PDI}(t) = \bar{M}_{\mathrm{w}}(t)/\bar{M}_{\mathrm{n}}(t)$, where $\bar{M}_{\mathrm{w}}(t) = \zeta_2(t)/\zeta_1(t)$ and $\bar{M}_{\mathrm{n}}(t) = \zeta_1(t)/\zeta_0(t)$ [121].

To quantify the prediction error, a global error indicator similar to that employed by Ho et al. [54] is adopted here:

$$\varepsilon_i = \frac{1}{F}\sum_{j=2}^{F}\left|\frac{C_i^{\mathrm{exact}}(t_j) - C_i^{\mathrm{sec}}(t_j)}{C_i^{\mathrm{exact}}(t_j)}\right|; \quad t_F = t_{99\%} \tag{3.18}$$

It should be noted that in Eq. (3.18) above, the errors are collectively computed from t_2 to $t_{99\%}$, where F is the total number of time steps between the two limits. Here, the initial time, $t_1 = 0$ is excluded to discount the error accrued in the initial mass due to the discretization of the DP domain. The global errors, ε_i, are computed for the polymers in the discrete region excluding the monomer, i.e., $i = [2, p]$, as their accurate predictions are of practical importance in the polymer context. In addition, ε_i is also a good reflection of the solution accuracy at the continuous region as the errors in this region are inevitably propagated to the discrete region through the polymer birth and breakage events. Errors

for the monomer densities ($i = 1$) are excluded from the error computations as they are usually of several orders of magnitude larger compared to the other DPs. In addition to the global error, the relative error for the moments [122] is computed as follows:

$$\xi_k = \max_t \left| \frac{\zeta_k^{\text{exact}}(t) - \zeta_k^{\text{sec}}(t)}{\zeta_k^{\text{exact}}(t)} \right| \tag{3.19}$$

The monomer evolution follows the same trend as the zeroth moment and thus, the errors in monomer predictions are reflected in the zeroth moment error. The error in the initial mass of the polymer due to the discretization of the DP domain is separately computed as:

$$\varepsilon_{\text{D}} = \left| \frac{\left[\sum_{i=1}^{I} C_i^{\text{sec}}(0) \times (162x_i + 18) \right] - M_{\text{T}}}{M_{\text{T}}} \right| \tag{3.20}$$

3.3.2 FPT and CAT Predicts Chain-end Scission while FVS Fails

In this work, chain-end monomer scission is examined, which implies the removal of one monomer unit from the end of the polymer chain. Here, the discretized PBEs are formulated for all sectional techniques based on the "continuous fundamental model" for chain-end scission by McCoy and Madras [115], which satisfies the conservation of number and mass. The stoichiometric kernel for this mode of scission is given separately for the monomer and the rest of the oligomers as [54, 115, 123, 124]:

$$b^{\gamma}(v, x_j) = \begin{cases} \delta(v - v_m); & \text{for DP} = 1 \\ \delta(v - [x_j - v_m]); & \text{for DP} > 1 \end{cases} \tag{3.21}$$

To test the performance of the sectional techniques, a linear rate kernel, $k^{\gamma}(v) = v$, is considered. For all the results of chain-end scission that follows, meshing configuration of $[p, q] = [100, 500]$ and non-uniform grids with a common ratio, $r = 1.0109$, are employed for the continuous region, unless stated otherwise. The error in the initial mass due to discretization based on this meshing configuration is $\varepsilon_D \sim \mathrm{O}(10^{-4})$. The numerical solutions for this case are shown in Figure 3.2 for the late phase of depolymerization, i.e., $\theta = 1$, where all the numerical errors would have accrued significantly.

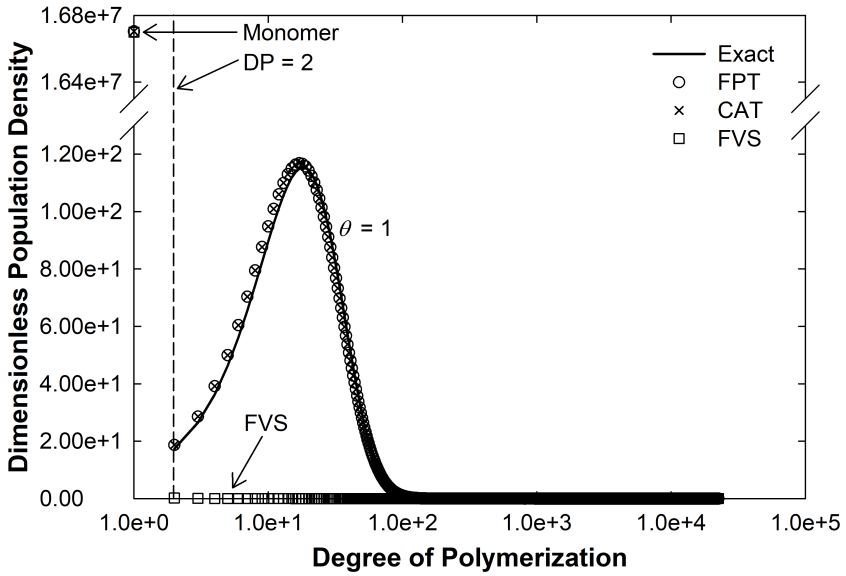

Figure 3.2: A comparison of the population density for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$.

From Figure 3.2, it is evident that the FVS fails to predict the population density (excluding the monomer). Unlike the FPT and the CAT, particles that fall in the i-th section due to chain-end scission do not get assigned to the adjoining pivot points in the FVS, thereby resulting in a zero-birth section as the interval widths are stretched apart with the increasing DP. This failure in the population density prediction consequently results in incorrect predictions of the molar concentration transients of individual oligomers, as

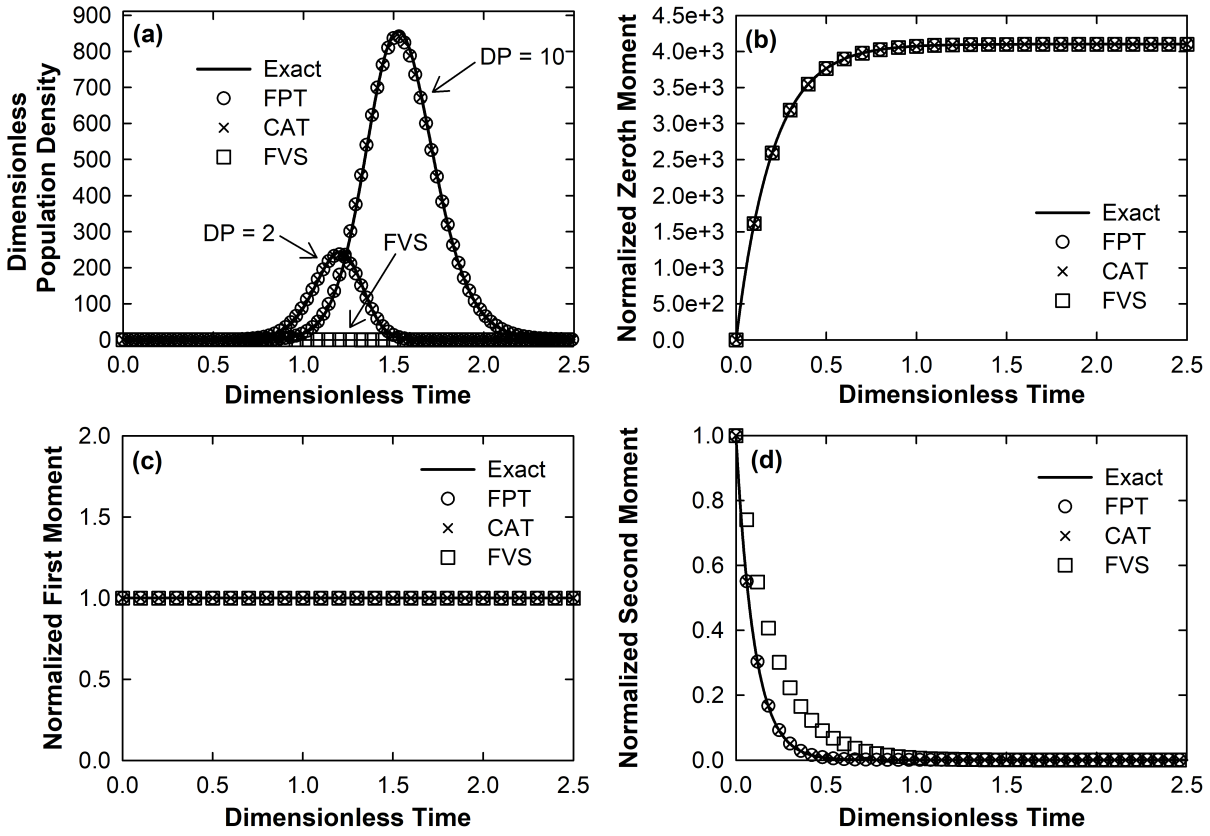

Figure 3.3: A comparison of (a) temporal evolutions of oligomers with DP = 2 and 10, (b) zeroth moment, (c) first moment, and (d) second moment, for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$.

shown in Figure 3.3(a). Because of the zero-birth section, the breakage of larger polymers that lie to the right of this section via chain-end scission does not result in the birth of smaller polymers to the left. This results in a net loss of mass and is reflected by the near-zero number densities of the FVS at the small size region, which are attained at a much faster rate as compared to other sectional methods and the exact solution. This is an inherent weakness of the FVS in dealing with chain-end scission and further details on the failure of the FVS for chain-end scission are provided in Appendix A.4.

Despite this poor prediction of the population density, Figure 3.3(b) and (c) show

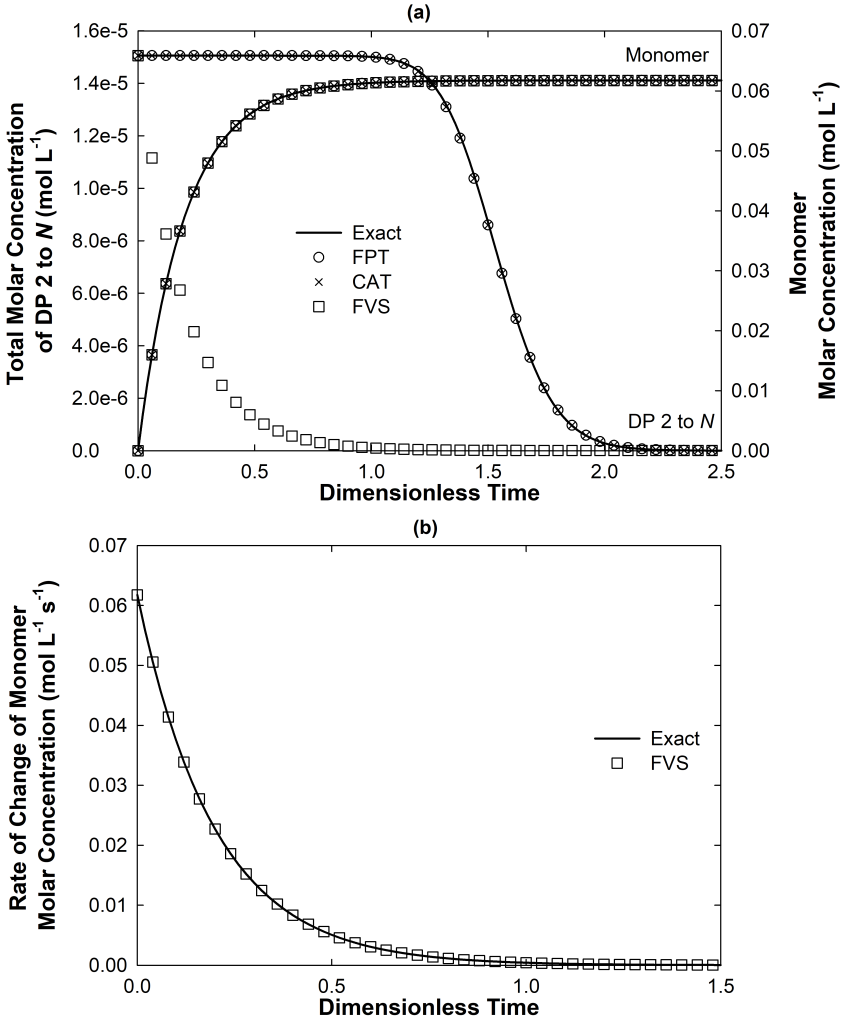

Figure 3.4: A comparison of (a) the total molar concentrations of DP 2 to N and the monomer molar concentration, and (b) the rate of change in the monomer molar concentration as given by the RHS of Eq. (3.3), for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$.

that the FVS unexpectedly predicts the zeroth and the first moments well. The performance of the FVS, however, is unsatisfactory for the prediction of the second moment (Figure 3.3(d)). Additionally, the FVS is also successful in predicting the temporal evolution of the monomer density (Figure 3.4(a)), although it fails to predict the rest of the population densities. To understand this counter-intuitive result, the rate of change in

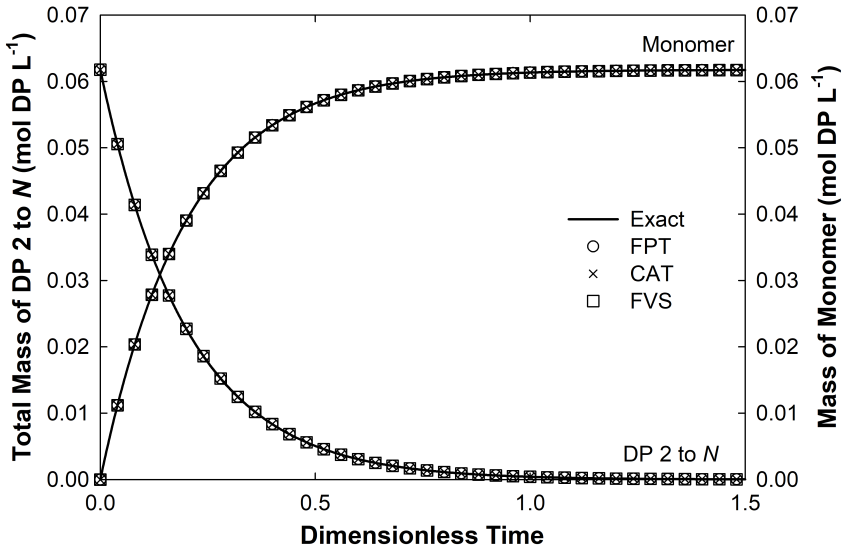

Figure 3.5: A comparison of the total mass of polymers with DP 2 to N and the mass of monomer, for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$.

the monomer molar concentration is obtained by computing the Right Hand Side (RHS) of Eq. (3.3) at every time step using the incorrectly predicted population densities of the FVS. As shown in Figure 3.4(b), the computed rate closely matches that of the exact solution, which allows one to rationalize the good prediction of the monomer concentration transient. Consequently, good prediction of the zeroth moment is also attained as the bulk of the zeroth moment is contributed by the monomer, which is of many orders of magnitude larger than the total molar concentration of the rest of the DPs (Figure 3.4(a)). Essentially, the effect of the incorrect population densities for the size region of DP $= 2$ to N is shrouded by the accurate prediction of the monomer concentration by the FVS. For the first moment, it is not difficult to rationalize the accurate prediction by the FVS, as the total mass of DP $= 2$ to N (computed using the incorrect population density transient) and the total mass of the monomer at any particular time match the exact solution, as shown in Figure 3.5. Despite failing to predict the population densities for chain-end scission, the FVS strongly retains its ability to preserve the moments. From

the investigation (not shown), this counter-intuitive behaviour of the FVS is independent of the type of rate kernels, initial distribution and meshing convention. As improving the formulation of the solution technique is not within the scope of this research, further future research to circumvent this issue by the FVS community is certainly encouraged.

In contrast to the performance of the FVS, Figure 3.2 and Figure 3.3 show that the FPT and the CAT predict the population density, zeroth, first and the second moments almost flawlessly. Moreover, Figure 3.4(a) and Figure 3.5 also indicate that the FPT and the CAT closely predict the temporal evolution of the monomer density along with the transient of the complete population density. Interestingly, the solution for the CAT coincides with the solution for the FPT. The mathematical rationale for this behaviour is given in Appendix A.5. In dealing with the chain-end scission using the FPT, Ho et al. [54] showed that the FPT was able to predict the monomer transient using a very coarse mesh, i.e., $[p, q] = [10, 30]$, but a much finer mesh of $[p, q] = [100, 500]$ was necessary to maintain an accurate prediction of the complete population density. Nevertheless, the latter only makes up to about 2.7% of the maximum DP, i.e., $(p + q)/N \approx 2.7\%$. The need for a finer mesh to capture the transient of the complete population density is also true for the CAT when dealing with chain-end scission due to the identical set of final equations. For guidelines on meshing, readers are referred to the study by Ho et al. [54]. Ultimately, chain-end scission is the most challenging breakage phenomenon to be approximated by sectional techniques and the performance of the FPT and the CAT demonstrated in this work is no simple feat. To substantiate the observations above, the ε_i of the DPs in the discrete region for both the FPT and the CAT are $\sim \mathrm{O}(10^{-2})$, while for the FVS it ranges between $\sim \mathrm{O}(10^{-1})$ and $\sim \mathrm{O}(10^{0})$. As for the zeroth and the first moments, the errors (ξ_0 and ξ_1) of all three sectional techniques are negligible, at about $\sim \mathrm{O}(10^{-5})$. Lastly,

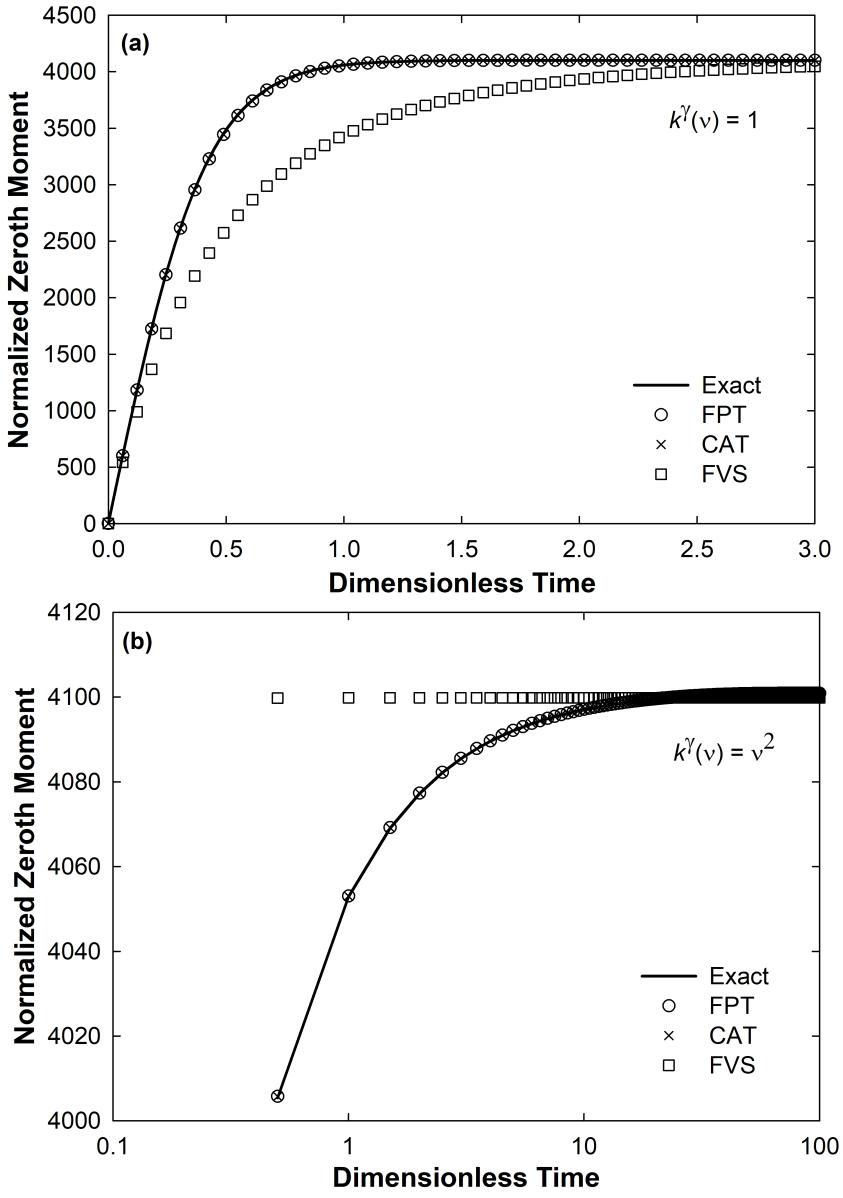

Figure 3.6: Predictions of the zeroth moment for chain-end scission with: (a) a constant rate kernel, $k^{\gamma}(v) = v$, and (b) a quadratic rate kernel, $k^{\gamma}(v) = v^2$.

the second moment errors, ξ_2, for both the FPT and the CAT are 0.028, whereas for the FVS, the error is significantly larger at 34.575.

The performance of the sectional techniques is further explored for chain-end scission

by employing a constant rate kernel, $k^{\gamma}(v) = 1$, and a quadratic rate kernel, $k^{\gamma}(v) = v^2$, while retaining the rest of the test conditions. A similar trend is observed for both the FPT and the CAT, where both techniques predict the population densities, zeroth, first and second moments exceptionally well. Surprisingly, for the FVS, the predictions of the monomer evolution and the zeroth moment (Figure 3.6) deteriorate. Consequently, the zeroth moment error, ξ_0, of the FVS increases to the range of $\sim \mathrm{O}(10^{-2})$ to $\sim \mathrm{O}(10^{-1})$.

3.3.3 Modified Sectional Technique Solves the Discrepancy at the Discrete-Continuous Boundary for Chain-End Scission

The results thus far suggest that the FPT and the CAT can complacently tackle the discrete-continuous implementation for chain-end scission. However, upon further scrutiny, a sudden drop in the number density is observed at the discrete-continuous boundary with the existing implementation. Such an inconsistency, to the best of the author's knowledge, has not been reported as few have explored the possibility of using sectional techniques for solving chain-end scission on a discrete-continuous mesh. Here, attempt to resolve this issue is presented. The discrepancy at the discrete-continuous boundary, i.e., at x_{p+1} is shown in Figure 3.7. This discrepancy is observed to be apparent at times with steep number density variation, e.g., at $\theta = 0.70$. In Figure 3.7, even with a fine mesh configuration of $[p, q] = [100, 500]$, the discrepancy is clearly observed. A coarser mesh at the continuous region or a narrow polymer distribution further amplifies this discrepancy, which further impairs the prediction of population densities in the discrete region. Additionally, the need to resolve such discrepancy becomes crucial when the DP of choice at

the discrete-continuous boundary is of great importance.

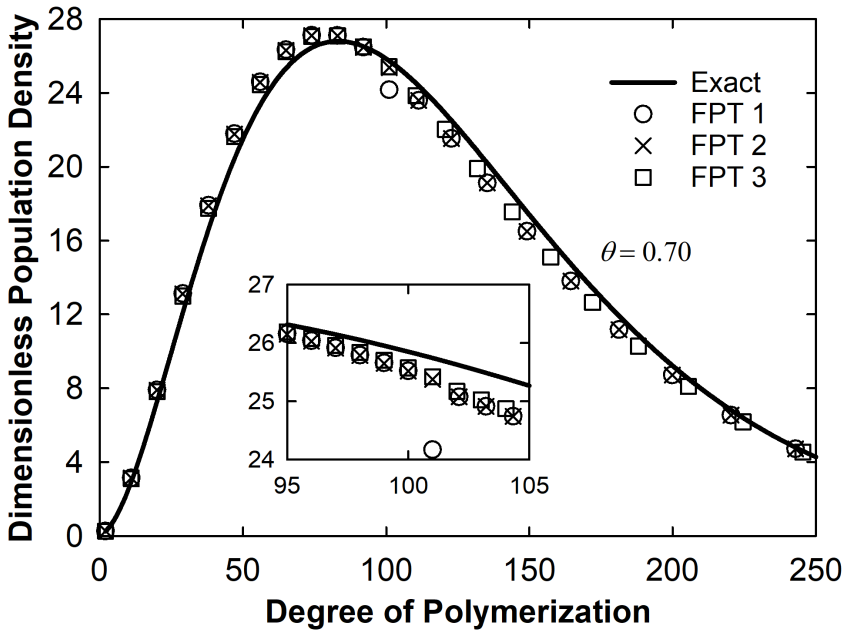

Figure 3.7: The discrepancy in the population density for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$. The discrepancy occurs at the discrete-continuous boundary, i.e., $x_{p+1} = 101$. FPT 1, FPT 2 and FPT 3 represent the original fixed pivot solution, the modified fixed pivot solution and the alternative meshing strategy to alleviate the discrepancy, respectively.

To resolve this discrepancy, the reason for this discrepancy is first delved. In the original FPT implementation on a discrete-continuous mesh, the grid points x_p and x_{p+1} lie at the boundary between the discrete and the continuous region. Imposing the chain-end scission stoichiometric kernel on Eq. (3.5), the polymer allocation function at x_{p+1} is:

$$\eta^{\mathrm{FPT}}_{p+1,p+2} = \frac{v_m}{x_{p+2} - x_{p+1}} \tag{3.22}$$

where $\eta^{\mathrm{FPT}}_{p+1,j} = 0, \forall\ j \mid j \neq p+2$. Thus, x_{p+1} receives polymer allocations solely from the splitting of polymers from x_{p+2} and there is no intra-interval polymer birth despite the fact that x_{p+1} lies in the fully continuous region. This is the cause of the observed discrep-

ancy, where the population density of x_{p+1} is always under-predicted. This discrepancy is corrected by modifying the limits of the integral and by introducing a correction factor, $\chi = x_{p+1}(r-1)/2$ (or $\chi = d/2$, if an arithmetic uniform mesh is employed for the continuous region, where d is the common difference between the grid points) to the Dirac delta term of the chain-end scission stoichiometric kernel to calculate $\eta^{\mathrm{FPT}}_{p+1,p+1}$:

$$\begin{aligned}\eta^{\mathrm{FPT}}_{p+1,p+1} &= \int_{x_{(p+1)-1/2}}^{x_{(p+1)+1/2}} \left[\frac{v - x_{(p+1)-1/2}}{x_{(p+1)+1/2} - x_{(p+1)-1/2}}\right] \delta\left(v - [x_{p+1} - (v_m - \chi)]\right) dv \\ &= \frac{x_{p+1} - v_m + \chi - x_{(p+1)-1/2}}{x_{(p+1)+1/2} - x_{(p+1)-1/2}} \\ &= 1 - \frac{v_m}{x_{(p+1)+1/2} - x_{(p+1)-1/2}}\end{aligned} \tag{3.23}$$

Due to this additional birth at the section represented by x_{p+1}, which is now shared between the grid points x_p and x_{p+1}, an equal amount of particle allocation should be discounted from x_p to conserve mass. Hence, instead of $\eta^{\mathrm{FPT}}_{p,p+1} = 1$, it is:

$$\eta^{\mathrm{FPT}}_{p,p+1} = \frac{v_m}{x_{(p+1)+1/2} - x_{(p+1)-1/2}} \tag{3.24}$$

The correction above is found to be adequate in resolving the issue of discrepancy at the discrete-continuous region boundary, cf. FPT 2 in Figure 3.7. This is achieved while retaining the accuracy of the original formulation. Moreover, the modified formulation works well for any type of grids and rate kernels. The temporal evolutions of the population density for FPT 1 and FPT 2 at x_{p+1} are shown in Figure 3.8, where the FPT 2 exhibits an improved performance compared to the FPT 1. Specifically, the FPT 2 minimizes the inherent under-prediction at regions near the peak, where there is a drastic variation in the population density. The modifications above can also be similarly applied

to the CAT.

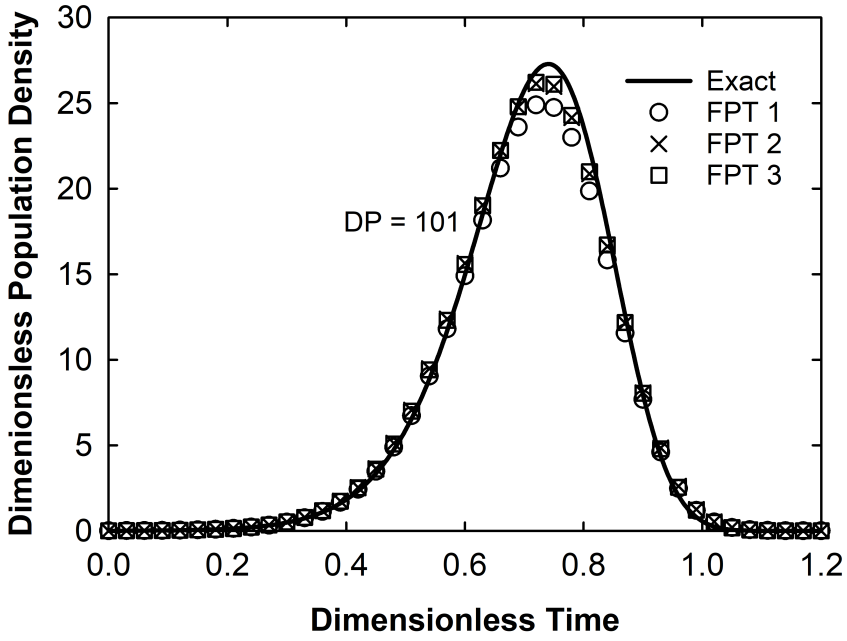

Figure 3.8: Temporal evolution of the dimensionless population density of $x_{p+1} = 101$ for FPT 1 (original formulation), FPT 2 (modified formulation) and FPT 3 (alternative meshing strategy). Here, chain-end scission is simulated with a linear rate kernel, $k^{\gamma}(v) = v$.

3.3.4 Alternative Meshing Strategy to Alleviate Discrepancy at the Discrete-Continuous Boundary for Chain-End Scission

Alternatively, it is found that one can constrain the number of grid points in the continuous region, q, based on Eq. (3.1), to offer a smooth transition of number densities between the discrete and the continuous region:

$$q = \left\lfloor 1 + \left\{ \ln\left(\frac{N}{p + v_m}\right) \Big/ \ln\left(1 + \frac{v_m}{p + v_m}\right) \right\} \right\rfloor \tag{3.25}$$

where floor function $\lfloor Z \rfloor$ denotes the nearest lower integer to Z. In this work, for $p = 100$ and $N = 22496$, the constraint in Eq. (3.25) sets $q = 549$ (as opposed to the original setting of $[p, q] = [100, 500]$). The alternative meshing strategy above, represented by FPT 3 in Figure 3.7 and Figure 3.8, demonstrates equally decent improvement to the discrepancy at the discrete-continuous boundary as the modified sectional technique (FPT 2) introduced in the previous section. Nevertheless, the alternative meshing strategy is advocated here as it offers a simpler and intuitive solution to the issue of discrepancy at the discrete-continuous boundary. Moreover, the new meshing strategy applies for all sectional techniques and works well for any type of grids and rate kernels.

3.3.5 Prediction of Chain-End Scission over Uniform Mesh in the Continuous Region

To divulge the performance of the sectional techniques in solving chain-end scission under different scenarios, the capability of the solution techniques is further stretched by employing a uniform mesh. One main drawback of utilizing a uniform mesh over a broadly distributed system is that the common difference between the pivots or the section widths in the continuous region would be invariably large. Table 3.1 gives the average global errors and moment errors of the FPT and the CAT over a uniform mesh with progressively refined mesh. The results show that for a uniform mesh, the mesh has to be refined to $(p + q)/N \approx 9.3\%$ to attain similar (albeit still poorer) accuracy to that exhibited by utilizing a non-uniform geometric mesh, i.e., $(p + q)/N \approx 2.7\%$. Therefore, in general, for a widely distributed particulate system, a non-uniform geometric mesh is more appropriate, whereas for a small and narrowly distributed system, a uniform mesh is a better choice.

This further suggests the need for a prudent mesh design to cater to the requisites of the application.

Table 3.1: Average global errors ($\bar{\varepsilon}$), and moments errors (ξ_0, ξ_1, ξ_2) of the FPT and the CAT for chain-end scission with a linear rate kernel, $k^{\gamma}(v) = v$, over a uniform mesh with progressively refined mesh. For a non-uniform geometric mesh with $[p, q] = [100, 500]$ and a common ratio, $r = 1.0109$, the $\bar{\varepsilon}$ is 0.047, ξ_0 and ξ_1 are 3.53×10^{-5}, and ξ_2 is 0.028.

$[p, q]$	Common difference, d	$(p+q)/N$	$\bar{\varepsilon}$	ξ_0	ξ_1	ξ_2
$[100, 500]$	44.9	2.7%	0.6376	0.0054	0.0054	0.3130
$[100, 1000]$	22.4	4.9%	0.3367	0.0026	0.0026	0.1580
$[100, 2000]$	11.2	9.3%	0.1697	0.0012	0.0012	0.0769

3.3.6 New Modified Stoichiometric Kernel for Random Scission Improves the Prediction of All Sectional Techniques

Random scission in the polymer context implies that each bond within a polymer chain is equally likely to break at any given time [64, 66, 125]. In view of the fact that a polymer chain with DP $= j$ contains only $j - 1$ bonds that are susceptible to depolymerization, Kumar and Ramkrishna [55] and Ho et al. [39] attempted to approximate the inherently discrete nature of depolymerization by employing the following random scission kernel:

$$b_1^{\alpha}(v, x_j) = \frac{2}{x_j - 1} \tag{3.26}$$

This stoichiometric kernel was tested in this work and is found to perform well for all three sectional techniques using a relatively fine mesh. However, considerable over-predictions

are evident when a coarse mesh is imposed. In applying Eq. (3.26), the use of a continuous PBE with a continuous stoichiometric kernel to predict the fully discrete PBE is, at best, a forceful approximation. In this work, it is proposed that this approximation gap due to the discretization error can be closed by employing a continuous analogue of the inherently discrete bond-breaking process. Notably, to imitate the discrete process, the continuous stoichiometric kernel should be free from intra-interval interactions, leading to:

$$b_2^{\alpha}(v, x_j) = \frac{2}{x_{j-1}} \tag{3.27}$$

In tandem with Eq. (3.27), all intra-interval birth contributions at the continuous region are eliminated. For all the results of random scission that follows, meshing configuration of $[p, q] = [10, 50]$ and non-uniform grids with a common ratio, $r = 1.1683$, are employed for the continuous region, unless stated otherwise. The error in the initial mass due to the associated discretization is $\varepsilon_D \sim \mathrm{O}(10^{-3})$. The numerical solutions for all three sectional techniques using $b_1^{\alpha}(v, x_j)$ and $b_2^{\alpha}(v, x_j)$ are demonstrated in Figure 3.9(a) and (b), respectively. From the results, the latter is found to be superior, where even at a very coarse mesh of $[p, q] = [10, 50]$ (as compared to 22496 ODEs for the exact solution), the considerable deviation observed with the use of $b_1^{\alpha}(v, x_j)$ reduces markedly.

It should be noted that the mesh applied in this case is extremely coarse, i.e., $(p + q)/N \approx 0.3\%$. The over-predictions exhibited by the sectional techniques, particularly for the CAT, are greatly reduced by imposing the new $b_2^{\alpha}(v, x_j)$ kernel. However, the FVS exhibits marginal improvements, as both cases show good prediction accuracy. This observation is congruent with the findings of Saha et al. [53], where the FVS was shown to adequately predict the densities for a wide range of rate kernels at coarse meshes for random scission. Despite the improvements for both the FPT and the CAT, slight de-

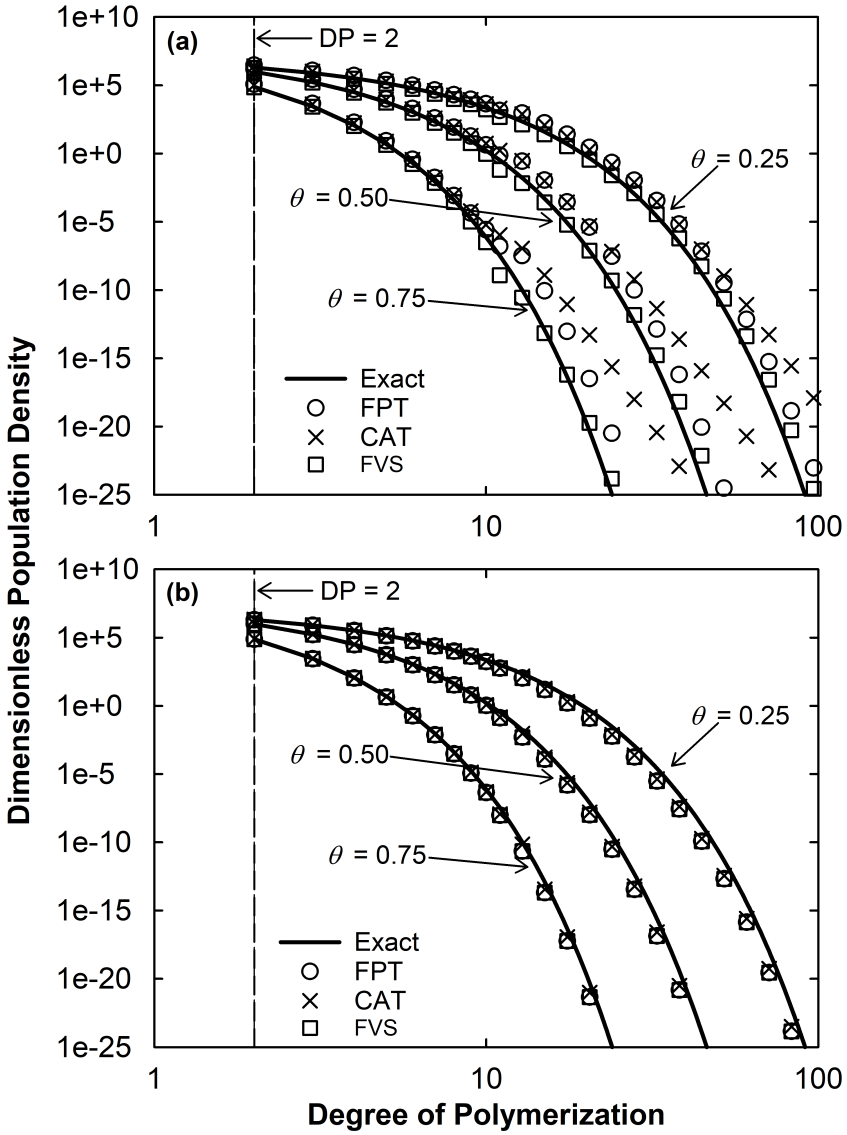

Figure 3.9: The performance of different sectional techniques for random scission with a linear rate kernel, $k^{\alpha}(v) = v$, using: (a) $b_1^{\alpha}(v, x_j) = 2/(x_j - 1)$, and (b) $b_2^{\alpha}(v, x_j) = 2/(x_{j-1})$.

viations at regions of steep number density variations persist, similar to the observations reported by Kumar and Ramkrishna [55], where the accuracy can be further improved by mesh refinement. Intriguingly, the solution of the FVS is observed to coincide with the solution of the FPT when the new $b_2^{\alpha}(v, x_j)$ kernel is imposed (see Appendix A.6).

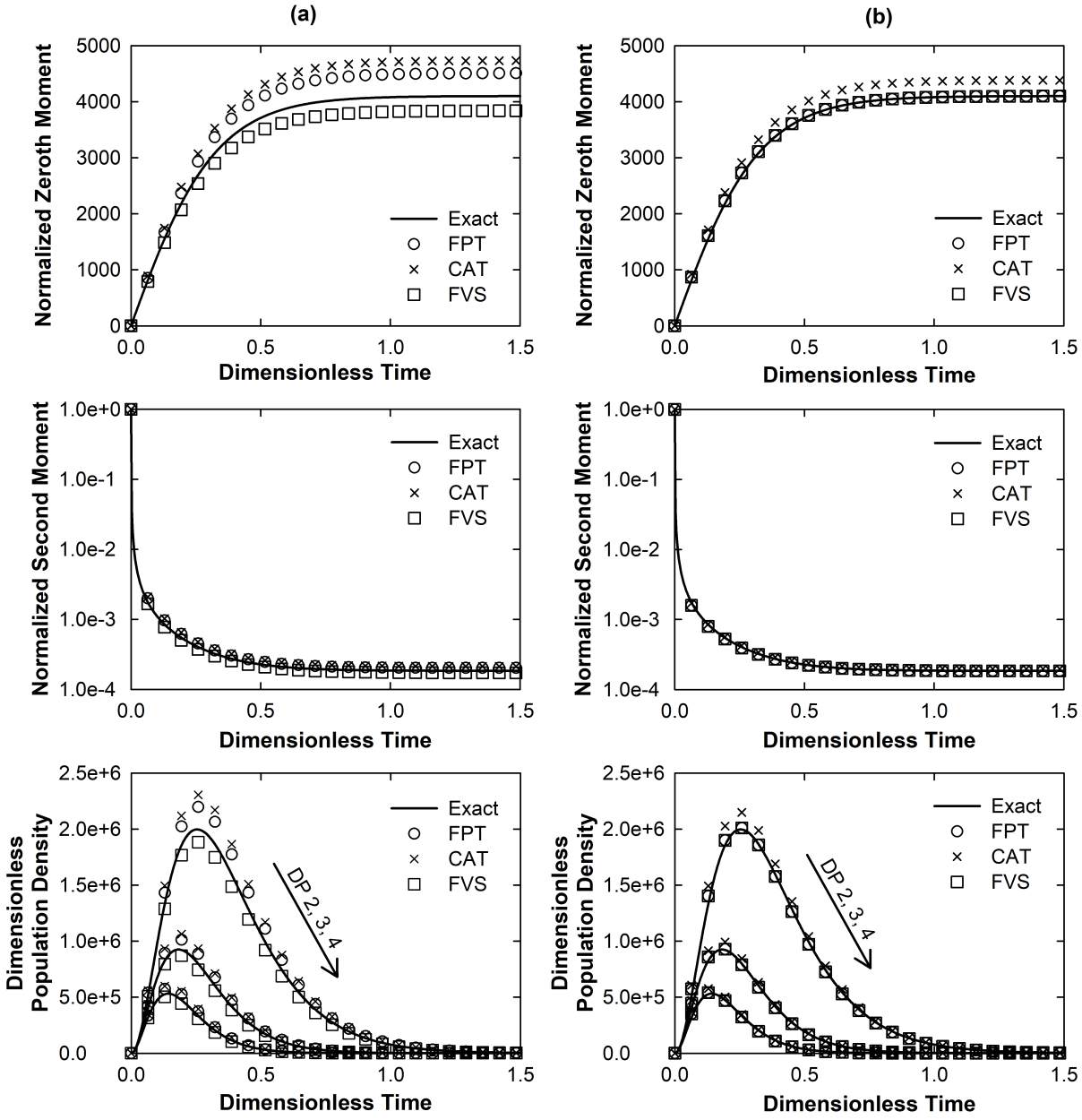

Figure 3.10: The performance of the sectional techniques in predicting the zeroth and second moments along with temporal evolution of oligomers using: (a) $b_1^{\alpha}(v, x_j) = 2/(x_j - 1)$, and (b) $b_2^{\alpha}(v, x_j) = 2/(x_{j-1})$, for pure random scission with a linear rate kernel, $k^{\alpha}(v) = v$.

The enhancements brought about by the kernel is not limited to the prediction of the population densities, as a substantial increase in accuracy is also seen in the moments and temporal evolution of the oligomer predictions, as shown in Figure 3.10. Although all the sectional techniques exhibit considerable improvements in the prediction of all the properties, the CAT pales in comparison as there are still notable deviations in the zeroth moment and temporal evolution of the oligomers with the improved kernel.

Table 3.2: Average global errors ($\bar{\varepsilon}$) and moments errors (ξ_0, ξ_1, ξ_2) for random scission with a linear rate kernel, $k^\alpha(v) = v$.

Errors	FPT		CAT		FVS	
	$b_1^\alpha(v, x_j)$	$b_2^\alpha(v, x_j)$	$b_1^\alpha(v, x_j)$	$b_2^\alpha(v, x_j)$	$b_1^\alpha(v, x_j)$	$b_2^\alpha(v, x_j)$
$\bar{\varepsilon}$	0.2390	0.1414	0.3299	0.1019	0.1679	0.1414
ξ_0	0.1041	0.0045	0.1589	0.1740	0.0598	0.0045
ξ_1	0.1041	0.0041	0.1589	0.0734	0.0598	0.0041
ξ_2	0.1489	0.1178	0.2101	0.0734	0.0844	0.1178

To quantify the performance, the average global errors and moment errors for all three sectional techniques are tabulated in Table 3.2. The decline in the moment errors is significant, especially for the FPT. The errors exhibited by the FPT and the FVS for the modified kernel are identical (and lower compared to that of the CAT) as the final equations using this kernel for both techniques coincide. Ultimately, the modified stoichiometric kernel proposed here preserves the "discrete" nature of the bond-breaking process. The random scission case using the modified stoichiometric kernel is also examined for a constant and a quadratic rate kernel, and positive outcomes are similarly observed.

3.3.7 Prediction of Random Scission over Uniform Mesh in the Continuous Region

To further illustrate the behaviour of the sectional techniques under different types of mesh, the performance of the sectional techniques for random scission over a uniform mesh with progressively refined mesh is shown in Table 3.3. For the FPT and the FVS, the uniform mesh has to be refined to $(p+q)/N \approx 0.9\%$ to attain similar prediction accuracy to that demonstrated by the non-uniform geometric mesh at $(p+q)/N \approx 0.3\%$, whereas for the CAT, the uniform mesh has to be refined to $(p+q)/N \approx 1.8\%$ in order to achieve similar accuracy to the non-uniform counterpart (cf. errors in Table 3.2 for $b_2^{\alpha}(v, x_j)$ kernel). The performance of the sectional techniques over a uniform mesh follows a similar trend to that of a non-uniform mesh, where the CAT is found to pale in comparison to the FPT and the FVS in terms predictions of zeroth moment and temporal evolution of oligomers. In general, random scission is more tolerant to the implementation of the uniform mesh than chain-end scission.

Table 3.3: Average global errors ($\bar{\varepsilon}$), and moments errors (ξ_0, ξ_1, ξ_2) of the FPT, the FVS and the CAT for random scission with the new improved stoichiometric kernel and linear rate kernel, $k^{\alpha}(v) = v$, over a uniform mesh with progressively refined mesh.

Sectional techniques	$[p, q]$	Common difference, d	$(p+q)/N$	$\bar{\varepsilon}$	ξ_0	ξ_1	ξ_2
FPT/FVS	$[100, 100]$	226.2	0.9%	0.1020	0.0675	1.08×10^{-5}	0.0744
CAT				0.7750	0.9460	0.8800	0.8800
FPT/FVS	$[100, 200]$	112.5	1.3%	0.0805	0.0513	6.17×10^{-6}	0.0591
CAT				0.1288	0.2520	0.2035	0.2035
FPT/FVS	$[100, 300]$	74.9	1.8%	0.0671	0.0412	6.14×10^{-6}	0.0462
CAT				0.0473	0.1354	0.0969	0.0969

3.3.8 Summary

Based on the findings in the preceding sections, a few points regarding the use of sectional techniques to solve depolymerization PBEs on a discrete-continuous mesh can be gleaned. For chain-end scission, the use of the FVS is impractical as the formulation is inherently incapable of dealing with a discontinuous stoichiometric kernel, leading to failure in the number density prediction (despite being able to preserve the moments). Conversely, the FPT and the CAT give an identical performance for any type of rate kernels and mesh configurations. The computation time taken by the CAT, however, is considerably longer ($t_{\mathrm{CAT}} = 19.0$ s and $t_{\mathrm{FPT}} = 7.0$ s for the numerical solutions shown in Figure 3.2), thus rendering the FPT to be the superior choice. For random scission, the implementation of the modified stoichiometric kernel enables accurate modelling with a very coarse mesh using all three sectional techniques. Among the three techniques assessed, the CAT is found to be marginally inferior in the prediction of the zeroth moment and temporal evolution of oligomers. The FPT and the FVS which yield an identical set of final ODEs, perform well when tested using a coarse mesh for all types of classical power-law rate kernels. Although the computation times for both the FPT and the FVS are comparable ($t_{\mathrm{FVS}} = 1.2$ s and $t_{\mathrm{FPT}} = 0.8$ s for the numerical solutions shown in Figure 3.9), the FPT is advocated here as the FVS eventually concurs to the mechanism of particle allocation of the FPT and forms identical ODEs.

3.4 Concluding Remarks

Three state-of-the-art sectional techniques assessed in this study reveal varying degrees of capabilities in solving the discrete bond-breaking depolymerization PBEs on a discrete-continuous mesh. The study reveals that a discrete-continuous mesh, along with a prudent choice of mesh types and parameters, is essential to successfully model discrete bond-breaking depolymerization processes. From the comparison of the sectional techniques, chain-end scission can only be tackled by the FPT and the CAT, whereas for random scission, all three sectional techniques — FPT, CAT and FVS, demonstrate excellent performance when intra-interval terms are disregarded, which includes the use of a new continuous analogue of the fully discrete random scission stoichiometric kernel. In terms of modelling simultaneous chain-end and random scissions on a discrete-continuous mesh, only the FPT and the CAT are the feasible choices through a superposition of the PBEs for various breakage phenomena. Due to this additive response, the performance-enhancing modifications proposed in this work also apply to the simultaneous breakage phenomena. Nevertheless, the assessment suggests that the added implementation complexity in the CAT offers no additional advantage over the FPT in all performance indicators assessed. In addition, the FPT is relatively simpler to implement and is computationally more efficient. Thus, the first research objective is achieved as the FPT is the best choice for dealing with the depolymerization PBEs on a discrete-continuous mesh and is used as the basis for the development of a new modelling framework for enzymatic hydrolysis of cellulose, which is detailed in the next chapter.

Chapter 4

Multi-Layered Population Balance Model Predicts the Dynamics of Enzymatic Hydrolysis of Cellulose

4.1 Preface

As alluded to in Section 2.2.2, this chapter introduces the Multi-Layered Population Balance Model (ML-PBM) to simulate enzymatic hydrolysis of cellulose. Here, special mechanistic considerations are made in the model to account for the rate slowdown often encountered in the enzymatic hydrolysis of cellulose. The numerical solution technique (FPT) explored in the previous chapter forms the foundation for the development of the ML-PBM in this chapter. The work presented here is published in Ahamed et al. [100].

4.2 Model Framework

4.2.1 Conceptual Description of Heterogeneity in Cellulose Properties

Cylindrical shapes are assumed for cellulose microfibrils [43, 82–84], where the cellulose chains are arranged laterally in layers across the radial direction (Figure 4.1A). The cellulose chains in each layer are confined and do not crossover to the adjacent layers. The hydrolysis of celluloses is restricted to the polymers on the surface and the polymers on the subsequent layers are exposed as the surface polymers are gradually solubilized by the enzymes. The microfibril is modelled to shrink in the radial direction as the cellulose mass is lost through solubilization, whereas the length of the cylinder is assumed to remain unchanged as it is usually several orders of magnitude larger than the radius, following the maximum DP of the cellulose distribution [126].

Provided that the enzymes work their way into the cellulose particles radially from the exterior and that enzyme deactivation and product inhibition are not the rate-limiting factors [89, 127–129], there must be changes in the substrate-related factors that dictate the rate slowdown. Potential factors attributed to the rate slowdown in the configuration described in Figure 4.1 include crystallinity, cellulose DP, and particle size, as widely discussed in the literature. In this work, it is hypothesized that the pre-treatment prior to hydrolysis reactions causes a non-uniform distribution of these key factors.

The native celluloses have a near homogeneous distribution of cellulose chains and properties across the multi-layered microfibril (Figure 4.1B). Pre-treatments upset this

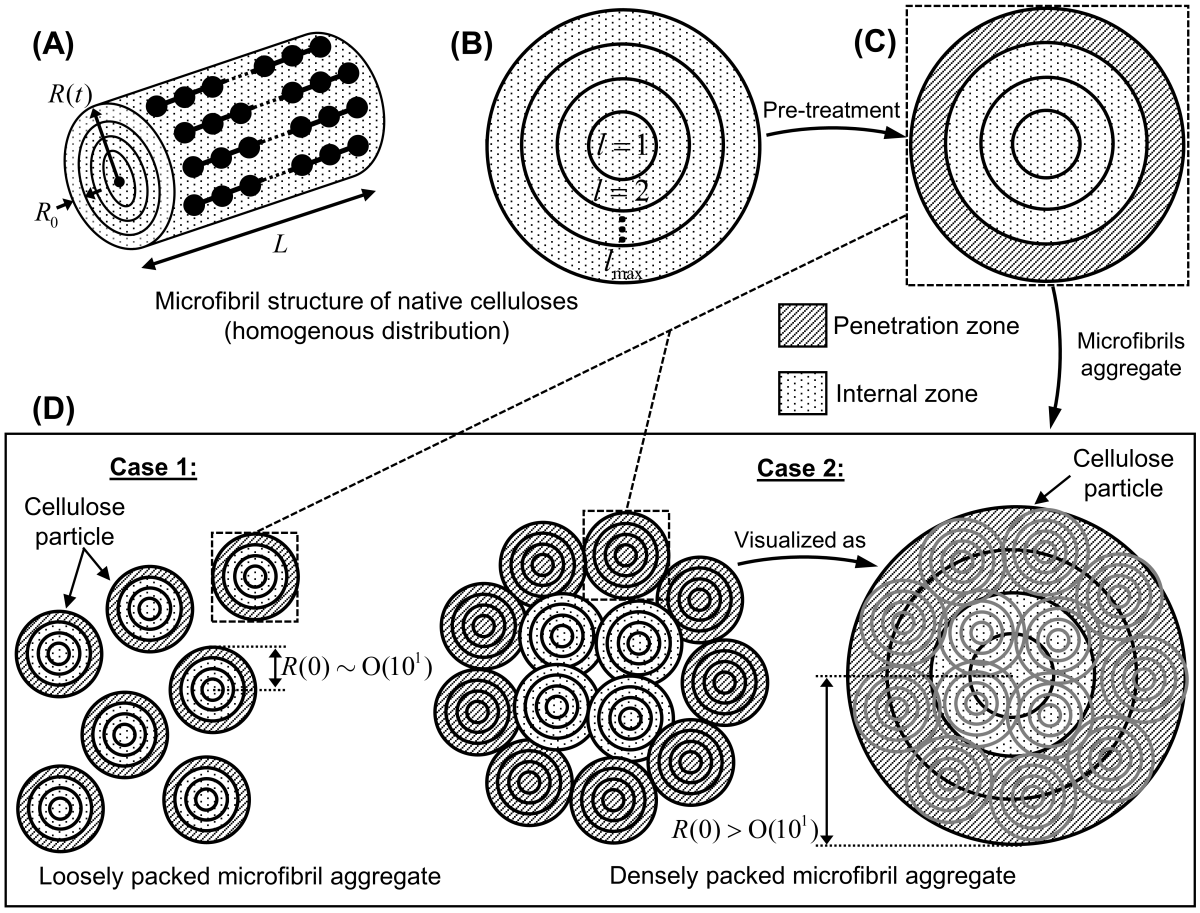

Figure 4.1: The depiction of the multi-layered distributed cellulose ultrastructure hypothesis. A cylindrical shape is assumed for microfibrils **(A)**, where the cellulose chains are organized in two-dimensional layered lattices. $R(t)$ is the microfibril radius, $R_0 = 1$ nm is the thickness of a single layer (i.e., the diameter of a single glucan molecule) and L is the length of the microfibril. The individual layers are denoted as l, where $l = 1, 2, 3, \ldots, l_{\max}$ and $l_{\max} = R(0)/R_0$. The native celluloses are made up of homogeneous polymer distribution across the layers **(B)** and the pre-treatment isolates the distribution into the penetration and internal zones **(C)**. Microfibril aggregates and the nature of the aggregates are traced based on the model-predicted initial particle radius, $R(0)$ **(D)**. Densely packed microfibril aggregates can be visualized macroscopically as a multi-layered large cylinder encompassing numerous microfibrils analogous to a single microfibril.

uniformity, leading to the formation of two distinct zones with different properties and distributions (Figure 4.1C). The outer zone (significantly affected by pre-treatment) is

characterized by a larger portion of celluloses with amorphous structure [126,130,131] and lower DP [132], which is defined as the "penetration zone" at the vicinity of the surface, while leaving the "internal zone" relatively untouched. The severity of pre-treatments will understandably affect the extent of the penetration zone in the microfibrils. The variation in the properties of the zones delivers different working conditions to the enzymes, which in this work is hypothesized to result in the rate slowdown.

The amorphous region of celluloses is hydrolysed much more rapidly as compared to the crystalline region due to the reduction in the enzyme accessibility to the latter owing to the strong hydrogen bonds [69,70,85,133]. The pre-treatments cause partial hydrolysis or degradation of cellulose chains, resulting in shorter cellulose chains on average in the penetration zone, as compared to the internal zone. Although the action by endoglucanases (EG) is greatly affected by the extent of cellulose crystallinity [134], crystallinity barely affects the cellobiohydrolases (CBH) [134–136]. Conversely, the rate of processive cleaving of cellobiose from chain ends by CBH is faster when short chains are encountered as chain ends provide ample binding sites for enzymatic action, but the cellulose DP has little to no effect on the random cleaving action of EG [69]. Therefore, the rate of EG and CBH actions will in general decline naturally as the enzymes progress from the penetration zone to the internal zone, but on the account of different mechanisms.

Here, the formation of two distinct zones (penetration and internal zones) composed of celluloses with different properties is included as a core element of modelling to predict the slowdown phenomenon. This multi-layered modelling concept applies to the following two different situations affected by the extent of microfibril aggregation. The microfibril aggregates can conform to one of two possibilities as shown in Figure 4.1D. If loosely packed, all microfibrils are prone to enzymatic or chemical (pre-treatment) attack from all

directions and the ML-PBM traces the particle radius of a single microfibril. Conversely, densely packed microfibril aggregates limit the enzymatic and chemical attack to the microfibrils at the exterior, and this can be visualized in the same manner as a single microfibril, i.e., a macroscopic multi-layered cylindrical structure. In the latter, the ML-PBM considers the aggregate as the cellulose particle and traces the macroscopic particle radius. In the subsequent text, the microfibril or the microfibril aggregates are collectively referred as cellulose particles for brevity, where the two are distinguished from each other based on the two possible cases (Figure 4.1D) given by the order of magnitude of the model-predicted initial particle radius, $R(0)$. The detailed technical formulation of the multi-layered distributed cellulose hydrolysis is further expounded in Appendix B.1.

4.2.2 Hydrolysis Reaction Mechanisms

In this work, celluloses are generally treated as insoluble populations contained within the cellulose particles, while polymers with DP < 7 are regarded as soluble oligomers, which are liberated into the reaction medium [8, 69]. Three distinct enzymatic actions on the hydrolysis of celluloses — processive cellobiose cleaving from the chain ends by CBH, random cleaving by EG and conversion of cellobiose to glucose by β-glucosidases (BG), are included in the ML-PBM. Reaction mechanisms of enzymatic hydrolysis of celluloses by CBH, EG and BG are established in Table 4.1. Note here that as the extent and mechanisms of CBH processivity remain uncertain [22], a complete processivity is assumed, where the CBH remains engaged to the cellulose chains until the polymer is reduced to a tetramer or a trimer, as shown in Eqs. (4.7) and (4.8). Due to the processivity of the CBH, the possibility of EG binding to existing CBH-polymer complexes is also included, as given by Eq. (4.5).

Table 4.1: Reaction mechanisms of enzymatic hydrolysis of celluloses. Here, E_F^m are the free enzymes, E_S^m are the surface-adsorbed enzymes, $E_B^m P_S(i)$ are the enzyme-polymer complexes, where m = CBH, EG, BG or EG-CBH, $P_S(i)$ are the insoluble surface polymers and $P(i)$ are the soluble oligomers, i (or j) denotes the DP and N represents the maximum cellulose DP.

Adsorption of enzymes:

$$E_F^{CBH} \underset{k_d^{CBH}}{\overset{k_a^{CBH}}{\longleftrightarrow}} E_S^{CBH} \tag{4.1}$$

$$E_F^{EG} \underset{k_d^{EG}}{\overset{k_a^{EG}}{\longleftrightarrow}} E_S^{EG} \tag{4.2}$$

Complexation of enzymes:

$$E_S^{CBH} + P_S(i) \underset{k_e^{CBH}}{\overset{k_{f,i}^{CBH}}{\longleftrightarrow}} E_B^{CBH} P_S(i); \quad i = 7, 8, \ldots, N \tag{4.3}$$

$$E_S^{EG} + P_S(i) \underset{k_e^{EG}}{\overset{k_{f,i}^{EG}}{\longleftrightarrow}} E_B^{EG} P_S(i); \quad i = 7, 8, \ldots, N \tag{4.4}$$

$$E_S^{EG} + E_B^{CBH} P_S(i) \underset{k_e^{EG}}{\overset{k_{f,i}^{EG}}{\longleftrightarrow}} E_B^{EG-CBH} P_S(i); \quad i = 4, 5, \ldots, N \tag{4.5}$$

Hydrolysis of insoluble celluloses:

$$E_B^{CBH} P_S(i) \xrightarrow{k_h^{CBH}} E_B^{CBH} P_S(i-2) + P(2); \quad i = 5, 6, \ldots, N \tag{4.6}$$

$$E_B^{CBH} P_S(4) \xrightarrow{k_h^{CBH}} 2P(2) + E_F^{CBH} \tag{4.7}$$

$$E_B^{CBH} P_S(3) \xrightarrow{k_h^{CBH}} P(2) + P(1) + E_F^{CBH} \tag{4.8}$$

$$E_B^{EG} P_S(j) \xrightarrow{k_{h,j}^{EG}} P_S(j-i) + P_S(i) + E_F^{EG}; \quad j = 7, 8, \ldots, N; \quad i = 1, 2, \ldots, N-1 \tag{4.9}$$

$$E_B^{EG-CBH} P_S(j) \xrightarrow{k_{h,j}^{EG}} P_S(j-i) + P_S(i) + E_F^{EG} + E_F^{CBH}; \quad j = 4, 5, \ldots, N; \quad i = 1, 2, \ldots, N-1 \tag{4.10}$$

Continued on next page

Table 4.1 – (continued)

where

$$\mathrm{P_S}(i) = \begin{cases} \mathrm{P}(i); & 1 \leq i < 7 \\ \mathrm{P_S}(i); & i \geq 7 \end{cases}$$

Hydrolysis of soluble oligomers:

$$\mathrm{E_F^{EG}} + \mathrm{P}(j) \xrightarrow{k_{\mathrm{h}'}^{\mathrm{EG}}} \mathrm{P}(j-i) + \mathrm{P}(i) + \mathrm{E_F^{EG}}; \quad j = 3, 4, \ldots, 6; \quad i = 1, 2, \ldots, 5 \tag{4.11}$$

$$\mathrm{E_F^{BG}} + \mathrm{P}(2) \xrightarrow{k_{\mathrm{h}'}^{\mathrm{BG}}} 2\mathrm{P}(1) + \mathrm{E_F^{BG}} \tag{4.12}$$

4.2.3 Population Balance Modelling for Enzymatic Hydrolysis Process

Here, a dynamic hydrolysis process model based on the reaction mechanisms shown in Table 4.1 is provided. While the cleaving of cellobiose into glucose by BG can be described by simple kinetics, other hydrolysis reactions, are conveniently represented using a general discretized breakage PBE as follows:

$$\frac{dC_i}{dt} = \sum_{j=i}^{N} \eta_{ij}^{\mathrm{m}} k_j^{\mathrm{m}} C_j - k_i^{\mathrm{m}} C_i; \quad \mathrm{m} = \mathrm{CBH}, \mathrm{EG} \tag{4.13}$$

where the polymer allocation function η_{ij}^{m} for FPT is given as:

$$\eta_{ij}^{\mathrm{m}} = \int_{x_i}^{x_{i+1}} \left(\frac{x_{i+1} - v}{x_{i+1} - x_i} \right) b_{\mathrm{m}}(v, x_j)\ dv + \int_{x_{i-1}}^{x_i} \left(\frac{v - x_{i-1}}{x_i - x_{i-1}} \right) b_{\mathrm{m}}(v, x_j)\ dv; \quad \mathrm{m} = \mathrm{CBH}, \mathrm{EG} \tag{4.14}$$

As celluloses are linear homopolymers, one-dimensional PBM with DP as the sole internal coordinate is employed in this work. Alternatively, two-dimensional PBM [62] can be adapted for the hydrolysis of branched polymers such as hemicelluloses. Here, k_i^{m}, C_i and x_i are the hydrolysis rate kernel, molar concentration and representative grids for polymers with DP $= i$, respectively, whereas, v is the continuous DP and b_{m}^{v,x_j} is the stoichiometric kernel that describes the distinct modes of scission by different enzymes. Note that the molar concentration in Eq. (4.13) can be interchanged with molar concentration densities, $c(v,t)$, by applying the mean value theorem, that is, $C_i(t) = \int_{x_{i-1/2}}^{x_{i+1/2}} c(v,t)\, dv$, where it forms the basis for derivation of ML-PBM equations according to the hydrolysis reaction mechanisms given in Table 4.1. Eq. (4.13) is numerically solved over a hybridized discrete-continuous mesh as described in Chapter 3. Here, the monomer and oligomers with great experimental significance are modelled discretely with high accuracy while approximating the rest of the polymer sizes in a continuous domain to reduce the computational load. The complete form of all equations is provided in Appendix B.3.

While the stoichiometric kernel in the PBE specifies the mode of enzymatic scission, the hydrolysis rate kernel specifies the functional relationship between the rate of scission and the DP (cf. Table B.1 in Appendix B.3). In the context of the reaction mechanisms and PBE presented above, it is worth mentioning that apparent rate coefficients are chosen for complexation and hydrolysis for EG, where both are taken as inversely proportional to the cellulose DP, i.e., $k_{\mathrm{f}}^{\mathrm{EG}}(v) = k_{\mathrm{p,f}}^{\mathrm{EG}}/v$ and $k_{\mathrm{h}}^{\mathrm{EG}}(v) = k_{\mathrm{p,h}}^{\mathrm{EG}}/v$, where v is the cellulose DP, $k_{\mathrm{p,f}}^{\mathrm{EG}}$ and $k_{\mathrm{p,h}}^{\mathrm{EG}}$ are rate constants. This is an implicit attempt to relate the rate coefficients to the crystallinity, where the variation in the crystallinity is accompanied by a change in the average cellulose DP, as deliberated in Section 4.2.1. The complexation and hydrolysis by EG are challenged by the limited navigation and cleavage through densely packed

crystalline regions. The proposition here allows the enzymatic action of EG to experience a rate slowdown as it progresses from the penetration zone to the internal zone with increasing degree of crystallinity (accompanied by the increase in average cellulose DP). This scheme is adequate until better explicit relationships between the crystallinity and the cellulose DP or other substrate-centric properties emerge.

4.2.4 Multi-Layered Population Balance Model (ML-PBM)

The preceding sections provided an overall outlook of the conceptual model and the accompanying reaction mechanisms. Here, the approach undertaken to account for the heterogeneity in the cellulose properties in line with the ML-PBM is presented. Experimental molecular weight distribution from Size Exclusion Chromatography (SEC) of various celluloses are studied to identify any indication of heterogeneity in the distributions. Subsequently, the experimental distributions are fitted using two superimposed gamma distributions. This allows the de-convolution of the bimodality often observed in the experimental distributions of celluloses into two individual distributions. In line with the multi-layered distributed cellulose hypothesis, the two individual distributions with smaller and larger average DP are assigned to the penetration zone and the internal zone, respectively. In general, the molar concentration density in the form of the gamma distribution is given as:

$$c(v,t) = \frac{M\Omega(v)}{\int_1^N \Omega(v)\left[(162v+18)\right]dv}; \quad \Omega(v) = \frac{v^{\lambda-1}\exp\left(-v/\phi\right)}{\phi^{\lambda}\Gamma(\lambda)} \tag{4.15}$$

where $\lambda = \bar{M}_{\mathrm{n}}/(\bar{M}_{\mathrm{w}} - \bar{M}_{\mathrm{n}})$, $\phi = \bar{M}_{\mathrm{w}} - \bar{M}_{\mathrm{n}}$, $\bar{M}_{\mathrm{n}}$ is the number-average DP, $\bar{M}_{\mathrm{w}}$ is the weight-average DP, $\Gamma(\lambda)$ is the gamma function and M is the mass concentration of the

celluloses. In tandem with the multi-layered modelling concept, Eq. (4.15) can be used to identify the cellulose distribution in individual layers, in the penetration/internal zone or the overall distribution by imposing the corresponding value to the parameters λ, ϕ and M. The variation of the cellulose distribution across layers of cellulose particles provides the necessary heterogeneity in the cellulose properties to exhibit the hydrolysis rate slowdown. Further details on formulating initial multi-layered cellulose distribution is explored in Appendix B.4.

Subsequently, the heterogeneous cellulose distribution characterized above is integrated with the cellulose structure shown in Figure 4.1. The overall mass concentration of the bimodal cellulose distribution is linked to the cellulose structure as:

$$M_{\mathrm{T}} = n\rho\pi L R(0)^2 \tag{4.16}$$

where n denotes the total number of particles per unit volume of the reaction medium and $\rho = 1500$ g/L is the density of the celluloses. The overall mass concentration is a sum of the mass concentration of the distribution of the penetration and internal zones, $M_{\mathrm{T}} = M_{\mathrm{P}} + M_{\mathrm{I}}$, where the mass concentration contained in the penetration zone is calculated based on the particle geometry as $M_{\mathrm{P}} = 2n\rho\pi L R_0 R(0) P_{\mathrm{zone}}$. Here, P_{zone} is the number of layers in the penetration zone. The total number of layers in a single cellulose particle is thus:

$$l_{\max} = \frac{R(0)}{R_0} = \mathrm{round}\left(\frac{2P_{\mathrm{zone}}}{r_{\mathrm{mass,P}}}\right) \tag{4.17}$$

where $r_{\mathrm{mass,P}} = M_{\mathrm{P}}/M_{\mathrm{T}}$ is the mass ratio of the penetration zone. With the above, P_{zone} and $r_{\mathrm{mass,P}}$, along with other model parameters, are fitted to the experimental hydrolysis data, where the resulting initial particle radius, $R(0)$ reveals the nature of the microfibril

aggregation (Figure 4.1D). The foregoing quantification of the experimental distribution allows one to identify the nature of the distributions in distinct zones in terms of number and weight-average DPs as well the concomitant cellulose morphology.

4.3 Results and Discussion

4.3.1 Heterogeneity in Cellulose Distribution

The experimental cellulose distributions from the literature and the corresponding model fits are shown in Figure 4.2A – C. Here, the model was fitted to the raw experimental data from the respective literature, where the distributions were originally normalized to the maximal values. An inverse procedure was employed where the fitted distributions (Figure 4.2A — C) were used as starting points. This was done because cellulose analytics often do not reflect the nature and the distribution of the real cellulosic starting material [138]: (a) experimental molecular weight distributions are known to vary extensively between batches of celluloses, cellulose preparation methods and calibration standards used [137, 139], (b) the activation treatments employed prior to SEC or Gel Permeation Chromatography (GPC) with the use of standard solvents, e.g., N,N-dimethylacetamide containing lithium chloride (DMAc/LiCl), are known to exhibit pronounced cellulose degradation resulting in the loss of average DP [138, 139], and (c) the low molecular weight polymers are also often substantially lost in the precipitation and purification steps of the analytical procedures [138]. Here, the average DPs of the distinct zones are retained while calibrating the mass portions of the zones, dictated by the $r_{\mathrm{mass,P}}$ to the experimental hydrolysis data. By doing so, a more accurate distribution of the celluloses

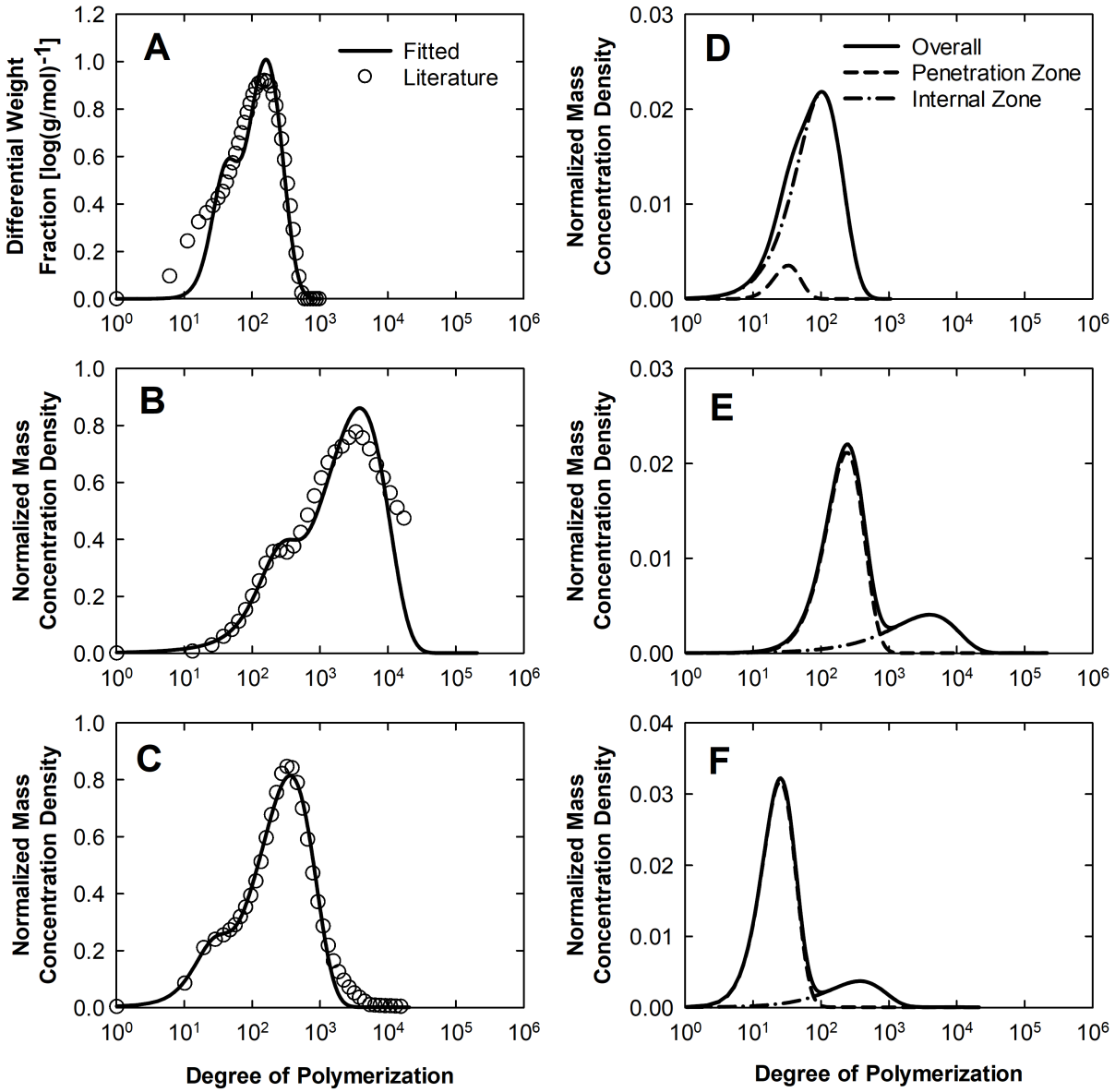

Figure 4.2: Fitting to experimental molecular weight distributions of celluloses (**A**, **B** and **C**) and subsequent prediction of probable distributions by the ML-PBM (**D**, **E** and **F**). The experimental distribution data represents: (**A**) Avicel from Engel et al. [137], (**B**) α-cellulose and (**C**) HCC from Nag et al. [85]. The probable distributions were predicted for: (**D**) Avicel hydrolysis data in Bezerra et al. [87], Bezerra and Dias [89], and Medve et al. [86], (**E**) α-cellulose and (**F**) HCC hydrolysis data in Nag et al. [85].

can be revealed given that the reliability of SEC/GPC remains elusive. Additionally, this strategy also divulges the possible arrangement of cellulose chains in the particles for various celluloses.

The predicted distributions of celluloses for various experimental hydrolysis data are shown in Figure 4.2D — F, where the mass concentration densities are normalized by the total mass of the celluloses. Here, the predicted distributions of Avicel across three different sources [86,87,89] are similar (Figure 4.2D), owing to only minor differences in the estimated parameter values that govern the cellulose distribution, further corroborating the viability of the ML-PBM.

4.3.2 Prediction of Rate Slowdown and Effective Enzyme Footprints

The values of the model parameters used to generate the simulation results shown in this work are collated in Table 4.2. The model parameters are divided into three major classes — (1) kinetic parameters comprising the complexation/decomplexation and the hydrolysis rate constants, (2) enzyme-centric parameters including adsorption/desorption rate constants and enzyme footprints and finally, (3) substrate-centric parameters that determine the cellulose particle structures and cellulose chain distributions. The ensuing simulation results are presented as follows.

The transients of cellobiose concentration, conversion and surface-adsorbed enzymes for the fitted models and experimental data [86] are shown in Figure 4.3. Some interesting observations could be gleaned. As far as the ML-PBM is concerned, the hydrolysis of the celluloses in the penetration zone aptly traced the initial rapid phase and the migration

Table 4.2: Parameter values used in ML-PBM predictions in this study.

Parameters	Unit	Experimental Data Source				
		Medve et al. [86]	Bezerra et al. [87]	Bezerra and Dias [89]	Nag et al. [85]	Eriksson et al. [127]
Cellulose	–	Avicel	Avicel	Avicel	α-cellulose, HCC	SPS
Enzyme	–	CBH I, EG II	CBH I	CBH I	CBH I, EG I, BG	CBH I
Kinetic parameters						
k_{h}^{CBH}	1/s	4.491	2.367	9.966	0.025	1.001
$k_{e}^{CBH}/k_{p,f}^{CBH}$	mol L^{-1} DP^{-1}	2.31×10^{-4}	3.19×10^{-4}	1.00×10^{-4}	7.58×10^{-9}	9.13×10^{-5}
$k_{p,h}^{EG}$	DP/s	21	n.a.*	n.a.	3.619	n.a.
$k_{h'}^{EG}$	L mol^{-1} s^{-1}	3	n.a.	n.a.	3	n.a.
$k_{e}^{EG}/k_{p,f}^{EG}$	mol L^{-1} DP^{-1}	2.50×10^{-4}	n.a.	n.a.	1.00×10^{-8}	n.a.
$k_{h'}^{BG}$	L mol^{-1} s^{-1}	n.a.	n.a.	n.a.	3.26×10^{3}	n.a.
Enzyme-centric parameters						
k_{d}^{CBH}/k_{a}^{CBH}	mol/L	1.36×10^{-8}	1.36×10^{-8}	1.36×10^{-8}	1.36×10^{-8}	1.36×10^{-8}
σ_{CBH}	mol/m^2	2.22×10^{-8} (6.80×10^{-9})†	2.22×10^{-8}	2.22×10^{-8}	2.22×10^{-8} (6.80×10^{-9})†	2.22×10^{-8}
k_{d}^{EG}/k_{a}^{EG}	mol/L	1.41×10^{-8}	n.a.	n.a.	1.41×10^{-8}	n.a.
σ_{EG}	mol/m^2	(8.04×10^{-8})†	n.a.	n.a.	(8.04×10^{-8})†	n.a.
Substrate-centric parameters						
P_{zone}	–	1	1	1	1 (30)‡	2
$r_{mass,P}$	%	2.80	2.77 (18.84)§	3.13 (14.00)§	16.00 (27.02)‡	4.10
$R(0)$	nm	71	72 (11)§	64 (14)§	13 (222)‡	98
$\bar{M}_n, \bar{M}_w$	DP	93, 154	93, 154 (71, 136)§	92, 153 (76, 141)§	1143, 7038 (79, 505)‡	90, 152

*n.a. - not applicable.
†Parameter values when both CBH and EG co-exist in the reaction medium.
‡Parameter values associated with HCC.
§Parameter values when cellulose is added in a very minute amount to the reaction medium (2.5 g/L).

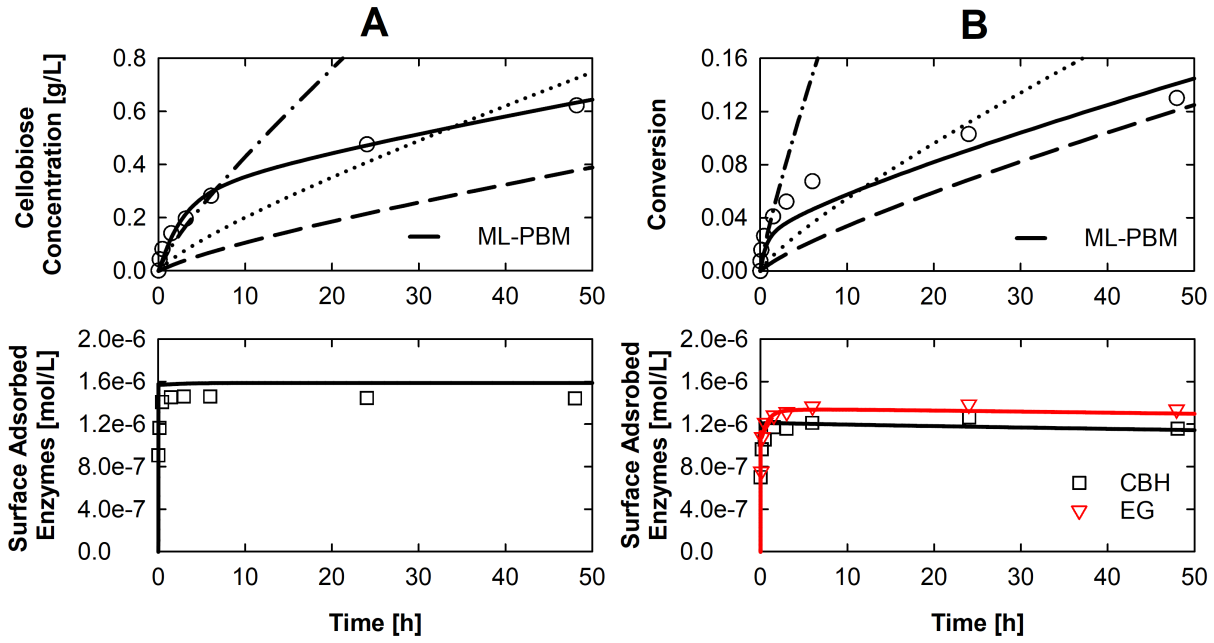

Figure 4.3: Model fitting to cellobiose concentration and surface-adsorbed enzyme concentration for: (**A**) Avicel hydrolysis by CBH and (**B**) Avicel hydrolysis by CBH and EG. For both cases, the Avicel loading is 10 g/L, CBH loading is 11.2 mg/g cellulose, and EG loading is 7.7 mg/g cellulose. Enzyme-centric variables are evaluated from this fitting and used for subsequent predictions. The ML-PBM is represented by the solid lines and the experimental data from Medve et al. [86] are represented by the markers. The rest are the results of the simplified homogeneous model, where dashed lines (− − −) represent the prediction with the same set of kinetic parameters from ML-PBM, dotted lines (· · ·) represent the prediction with recalibrated kinetic parameters to the entire data set and dash-dotted line (− · −) represent prediction with recalibrated kinetic parameters to the initial data points only. The overall initial cellulose distribution is common for all cases.

of the enzymatic action into the internal zone resulted in an adequate rate slowdown. Although the ML-PBM fit for the Avicel hydrolysis by CBH is decent, the fit is less than satisfactory when the EG action was included. One can infer that this disparity stems from the fact that EG II was employed in the experiments, where Medve et al. [86] inferred that the enzyme may have the tendency to be processive in action and has a preference towards the chain-ends. This was later confirmed by Cohen et al. [140] as EG II was found to produce mainly cellobiose as the soluble product. Understandably, the ML-PBM does

not trace the mode of action of EG II as the endoglucanases in this model are strictly assumed to carry out non-processive random cleaving.

On the contrary, the ML-PBM fit to the transients of the surface-adsorbed enzymes is qualitatively and quantitatively ($< 10\%$ error margin) acceptable for both cases of hydrolysis by CBH alone and by both CBH and EG (Figure 4.3). Here, the crowding effect is included in the ML-PBM, where the adsorbed enzymes occupy spaces and reduce the available surface area for further adsorption of free enzymes [43,82,84]. The surface coverage of the enzymes, which is governed by the physical structure of the enzymes, is reflected through the enzyme footprints, σ_{m}, where m = CBH and EG (cf. Appendix B.2). From the model fitting, the enzyme footprint for CBH alone is 2.22×10^{-8} mol/m^2, equivalent to 299 glucan units (assuming that each glucan unit spans 0.25 nm^2 [84]), whereas the enzyme footprints of CBH and EG together are 6.80×10^{-9} and 8.04×10^{-8} mol/m^2, equivalent to 977 and 82 glucan units, respectively. Generally, the footprints of CBH gathered from the ML-PBM are greater than the values predicted by the random sequential adsorption model of Levine et al. [84]. Although physical structure is not assumed for the enzymes, the ML-PBM traces the effective enzyme footprints. The arrangement of the adsorbed enzymes on the cellulose surface could be regarded as a "structured packing" when only one type of enzyme is present or as a "random packing" when more than one type of enzyme co-exist. This explains the sharp increase in the effective footprint of CBH when it is present together with EG. In addition, the larger CBH footprint could be attributed to its elaborate and developed structure comprising a separate carbohydrate-binding module and tunnel-shaped catalytic domain conjoined by a linker [69]. The estimated enzyme-centric parameters for CBH and EG were used for the calibration and validation with subsequent experimental data, provided that the experimental conditions are similar.

Here, it is assumed that the enzyme-centric parameters of EG II could be extended to EG I as both have relatively similar molecular weights and protein structures [69].

4.3.3 Cellulose Heterogeneity: A Key Modelling Component for Predicting Slowdown

To demonstrate the role of heterogeneity in cellulose properties in the ML-PBM, here the performance of the ML-PBM is benchmarked against a homogenous model by discounting the heterogeneity of cellulose properties in the structural layers of the cellulose particles. The homogeneous model assumes that the bimodality of cellulose distribution is uniformly dispersed throughout the cellulose particles, negating the need for multi-layers as well as the distinct penetration and internal zones. Generally, the homogeneous model (cf. Appendix B.5) is analogous to all the preceding modelling attempts that do not account for the heterogeneity of cellulose properties.

Figure 4.3 shows that the homogeneous models pale in comparison against the ML-PBM. As the homogeneous model generally lacks the slowdown "bend" in the predictions of the hydrolysis profiles, the fits to the full experimental transient are always inadequate — regardless of whether the kinetic parameters are retained from the ML-PBM or re-calibrated. Although the homogeneous model can be selectively fitted adequately to the initial data points, the over-prediction at the later stages is disastrous. The disparities in the model fitting between the homogeneous model and the ML-PBM are notably substantial despite them sharing the same overall initial cellulose distribution. Therefore, the selective deposition of cellulose chains at different regions in the cellulose particles, i.e., heterogeneity in cellulose properties across the structural layers of cellulose particles,

becomes a potential enzymatic hydrolysis rate slowdown mechanism. The inclusion of this critical element in the ML-PBM provides a fresh perspective into the slowdown of enzymatic hydrolysis of celluloses, possibly allowing one to manipulate and control the rate-limiting phenomenon.

4.3.4 Further Validations of the ML-PBM and Morphological Interpretations

To substantiate the predictive capabilities of the ML-PBM, the model is calibrated and validated to multiple data sets for Avicel hydrolysis by CBH [87, 89] as shown in Figure 4.4. The model validations are performed by retaining the same parameter values from prior calibrations to showcase the predictive capability of the ML-PBM over a range of operating conditions. The ML-PBM shows reasonable agreement with all experimental data sets except when the initial substrate concentration is very minute (2.5 g/L Avicel, blue-dotted lines in Figure 4.4). There has been a report of substrate concentrations imposing apparent influence on model parameters [141], but from the modelling point of view, the estimated model parameters (kinetic and enzyme-centric parameters) should be global and should validate multiple data sets, at least from the same source. In this regard, one can infer that there must have been a variation in the consistency of cellulose distribution or structure when the initial cellulose concentration is very small, such that it prevents the mechanistic model from validating the data set. By merely re-fitting the substrate-centric parameters while retaining the rest of the model parameters (blue-dashed lines in Figure 4.4), the ML-PBM correctly predicts the experimental transient. This suggests that the reported influence of dilute substrate concentrations on hydrolysis

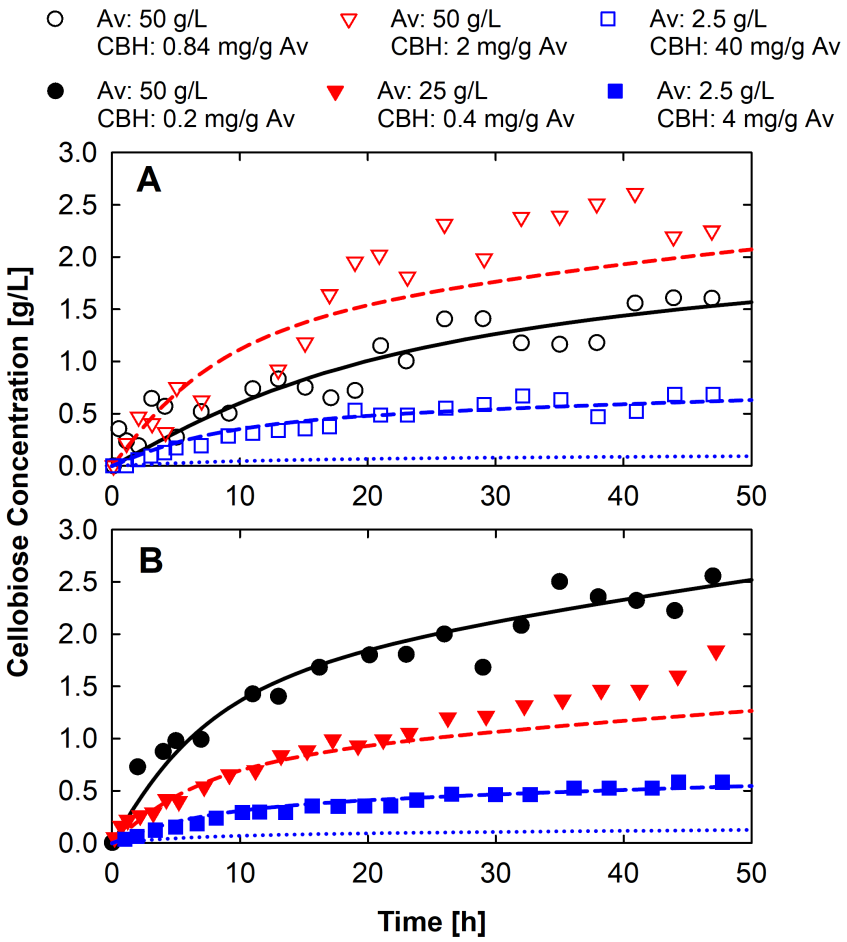

Figure 4.4: ML-PBM calibration and validation to cellobiose concentration for Avicel hydrolysis by CBH. In (**A**), the ML-PBM is calibrated with 50 g/L Avicel and CBH loading of 0.84 mg/g cellulose, and subsequently, validated with the other two data sets, where the experimental data are obtained from Bezerra et al. [87]. Similarly in (**B**), the ML-PBM is calibrated with 50 g/L Avicel and CBH loading of 0.2 mg/g cellulose and validated with the remaining data sets, where the experimental data are from Bezerra and Dias [89]. The solid lines are the fitted ML-PBM, dashed lines are the predictions and the markers are experimental data. For Avicel loading of 2.5 g/L, the model predicts poorly (blue-dotted lines) but the prediction improved when substrate-centric parameters are re-fitted while retaining other parameters (cf. parameter values in Table 4.2). Av in the legend refers to Avicel.

profile may not be due to the changes in the underlying hydrolysis mechanism, but rather due to the variations of cellulose distribution and structure as a result of inconsistent

sampling from the bulk at dilute loadings.

To further extend the ML-PBM to the hydrolysis of different types of celluloses, the model is calibrated and validated to multiple data sets from Nag et al. [85] that demonstrates the hydrolysis of α-cellulose and highly crystalline cellulose (HCC) by CBH+BG and CBH+BG+EG enzyme cocktails, as shown in Figure 4.5. All the model parameters except the substrate-centric parameters that differentiate the α-cellulose and HCC, were maintained for all data sets in this series. The ML-PBM generally shows excellent agreement to experimental data for CBH+BG enzyme cocktails, but a minor overestimation in conversion is observed at large times when EG is included in the enzyme cocktail. Nevertheless, the ML-PBM traces the rate slowdown adequately and captured the increase in rate and conversion due to synergism when EG is included in the enzyme cocktail. The minor overestimation may be the result of oversimplification of the EG mode of action, where EG might have the tendency to perform a "mixed" mode of scission, as opposed to a pure non-processive random cleaving [142]. Alternatively, the implicit relationship between the hydrolysis and complexation rates with the crystallinity via cellulose DP may not be valid at large times or inadequate to represent the intrinsic relationship between the rates and the crystallinity. The decline in cellulose chain lengths over time would indicate that the crystallinity is declining as well according to the scheme proposed in this study, but in reality, crystallinity showed no appreciable change over the course of hydrolysis [85, 133]. Thus, it is understandable that the ML-PBM registers a slight overestimation in conversion over a supposed decrease in the crystallinity, in contrast to the experimental data. These areas are worthy of further study and may assist in further enhancing the viability of the ML-PBM in the future.

The cellulose morphology is also expounded by the ML-PBM. The predicted initial

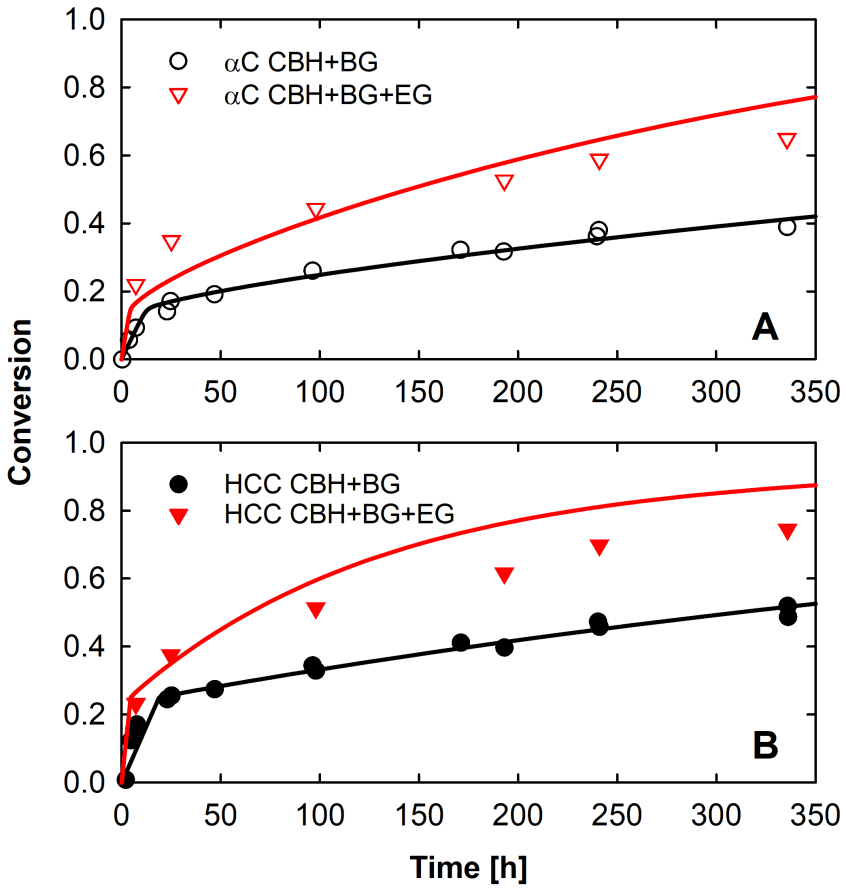

Figure 4.5: ML-PBM calibration and validation to conversion for: (**A**) α-cellulose hydrolysed by CBH+BG and CBH+BG+EG enzyme mixtures, and (**B**) HCC hydrolysed by CBH+BG and CBH+BG+EG enzyme mixtures. The loadings for α-cellulose and HCC are 20.13 g/L and 24.19 g/L, respectively. Enzyme loadings for CBH+BG mixture are 30 mg/g cellulose and 2 mg/g cellulose, respectively, whereas enzyme loadings for CBH+BG+EG mixture are 25 mg/g cellulose, 2 mg/g cellulose and 5 mg/g cellulose, respectively. The lines are ML-PBM predictions while the markers denote experimental data from Nag et al. [85]. The parameter values are consistent over all four cases.

particle radius (see Table 4.2) for each cellulose is such that α-cellulose $<$ Avicel $<$ HCC. Understandably, α-cellulose with the least degree of crystallinity on average is predicted to form the smallest particles suggesting that the microfibrils are loosely aggregated (Case 1 of Figure 4.1D). Conversely, HCC which has the highest degree of crystallinity expectedly forms the largest particles indicating that the microfibrils are densely aggregated (Case

2 of Figure 4.1D). Avicel, a microcrystalline cellulose, conforms to an intermediary point in between the two extreme cases. The ability of the ML-PBM to sensibly infer the substrate morphologies according to the cellulose types with minimal data inputs (base-fitted cellulose distribution and hydrolysis data only) is certainly a remarkable feat.

4.3.5 Effect of Heterogeneous Cellulose Properties on Cellulose Distribution During Enzymatic Hydrolysis

To understand the intrinsic mechanisms of the ML-PBM, here, the evolution of some of the important properties that accompany the hydrolysis of cellulose are divulged. The ML-PBM restricts the enzymatic hydrolysis to a layer-by-layer attack on the cellulose chains, often dubbed as a slow surface ablation process [3,20]. Ultimately, the ML-PBM effectively replicates the recalcitrant nature of celluloses, where enzyme accessibility is often limited to the surface of the cellulose particles. The insoluble celluloses beneath the surface are only made accessible as the surface celluloses are progressively "peeled" off by the enzymes. Figure 4.6 clearly demonstrates this mechanism where the average DPs of the insoluble celluloses show a sharp increase (as opposed to the expected decrease due to the chain cleaving) during the solubilization of celluloses in the penetration zone as the longer chains remain untouched in the internal zone. Once the enzymatic hydrolysis reaches the internal zone as indicated by the decrease in the particle radius, the number-average DPs level off and subsequently decrease as it should at large times (results not shown). The decrease in the number-average DP of HCC is more pronounced as it generally has a smaller average DP, leading to a faster formation of shorter chains. Conversely, the weight-average DP of α-cellulose continues to rise while for HCC, it plateaus indefinitely.

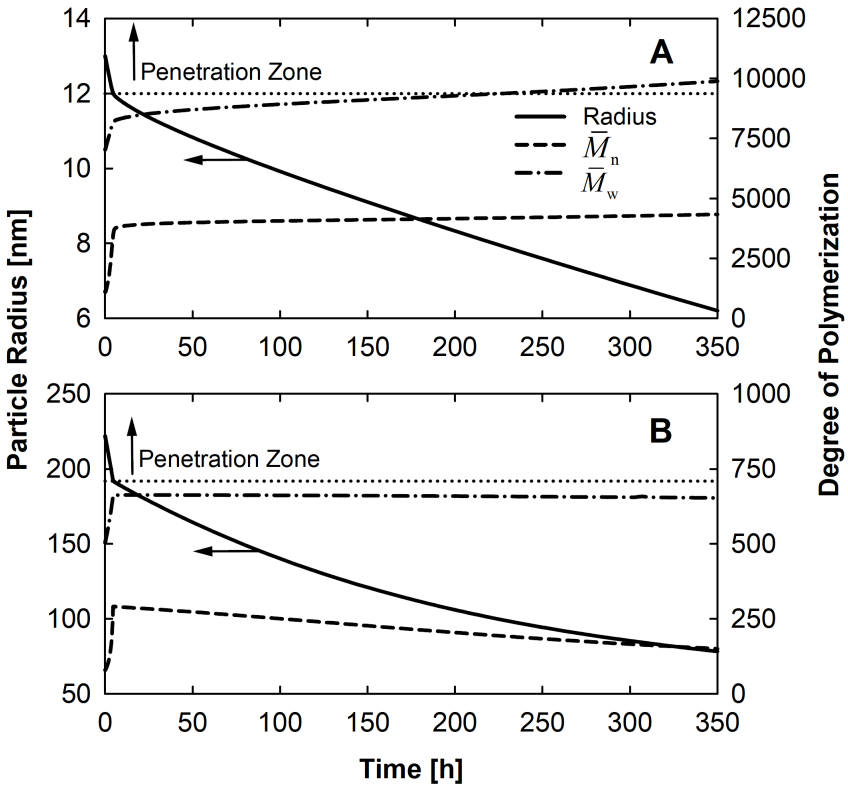

Figure 4.6: Temporal evolution of number-average DP ($\bar{M}_n$), weight-average DP ($\bar{M}_w$) and particle radius of insoluble celluloses for hydrolysis of: (**A**) α-cellulose by CBH+BG+EG mixture and (**B**) HCC by CBH+BG+EG mixture. ML-PBM predictions are for the cases presented in Figure 4.5. The ML-PBM demonstrates an initial increase in mean DPs during the hydrolysis of the penetration zone due to the solubilization of shorter chains while preserving the longer chains underneath in the internal zone. The mean DPs level off and subsequently decreases when the hydrolysis advances into the internal zone.

The changes in the average DPs showed here are reflected by the evolution of the insoluble cellulose distributions shown in Figure 4.7. The insoluble cellulose distributions broaden and shift towards the larger DP as time progresses.

The ML-PBM predictions shown in Figure 4.6 and 4.7 are in close agreement with the findings of Levi and Sellen [132], where the authors studied the changes in the average DPs in cellulose fibre layers by decaying preferential regions using a fungal species. Similar

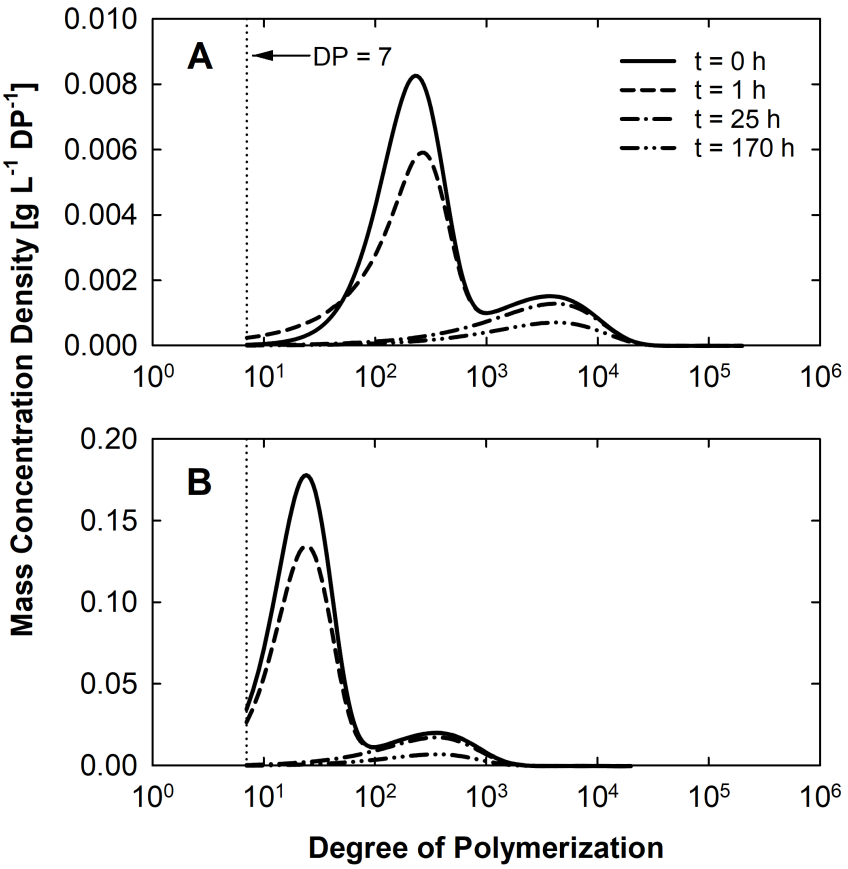

Figure 4.7: Temporal evolution of insoluble cellulose distributions for hydrolysis of: (**A**) α-cellulose by CBH+BG+EG mixture, and (**B**) HCC by CBH+BG+EG mixture. ML-PBM predictions are for the cases presented in Figure 4.5. The mass concentration densities for DP < 7 are not shown as it represents soluble oligomers liberated into the reaction medium. The distributions generally shift towards the larger DP and the polydispersity increases over the course of the hydrolysis.

to the ML-PBM predictions, the study reported that the long chains were untouched until the very advanced stages of decay, indicating that the larger DPs are isolated in the inner regions. The study also showed an increase in the breadth of the distribution and a shift towards the larger DP over the course of the decay, induced by the increase in the weight-average DP and the decrease in the number-average DP. Provided that the majority of the residual insoluble celluloses consist mainly of large molecules and the weight of the

small molecules is sufficiently small, the weight-average DP increases. Conversely, the number-average DP is highly sensitive to the number of molecules and tends to decrease as a large portion of small molecules are lost to degradation [132]. Nevertheless, the ML-PBM predictions shown in Figure 4.6 and 4.7 are contradictory to the temporal evolution of the experimental molecular weight distributions reported by Nag et al. [85], where the reported experimental distributions remained relatively unchanged over the course of hydrolysis. This discrepancy is attributed to the shortcomings associated with the SEC/GPC. Thus, tracing the temporal evolution of cellulose molecular weight distribution via the conventional SEC/GPC systems is not ideal and the advent of an alternative reliable method remains to be seen.

4.3.6 ML-PBM Predicts the Enzyme Tunnelling Effect

Reactive surface area available to enzymes has always been regarded as one of the major determinants of cellulose hydrolysis [71, 72, 82, 84]. By avoiding the rapid equilibrium assumption, ML-PBM traces the dynamic interactions between the enzymes and the cellulose particles with the use of physically sound enzyme footprints to characterize the area coverage. Noticing that the enzyme-centric parameters are not sensitive to hydrolysis profile, model was fitted with both hydrolysis and adsorption data for a sensible parameter estimation. The enzyme-centric parameters are also retained for both penetration and internal zones as the enzyme properties are dictated by its genetic imprint and should not be governed by external factors. With the above, the unrealistic conventions that have limited earlier efforts in representing the enzyme adsorption kinetics are avoided. Here, the ML-PBM clearly demonstrates the enzyme crowding effect governed by the cellulose surface area available to enzyme binding and the corresponding enzyme footprints in

Figure 4.8.

Given that the ML-PBM predicts the enzyme footprints and the substrate morphology (i.e., cellulose particle radius), finding the enzyme concentration that induces the enzyme crowding effect is straightforward (see Appendix B.2). Any enzyme concentrations that are below or above this saturation point signify enzyme-limiting and

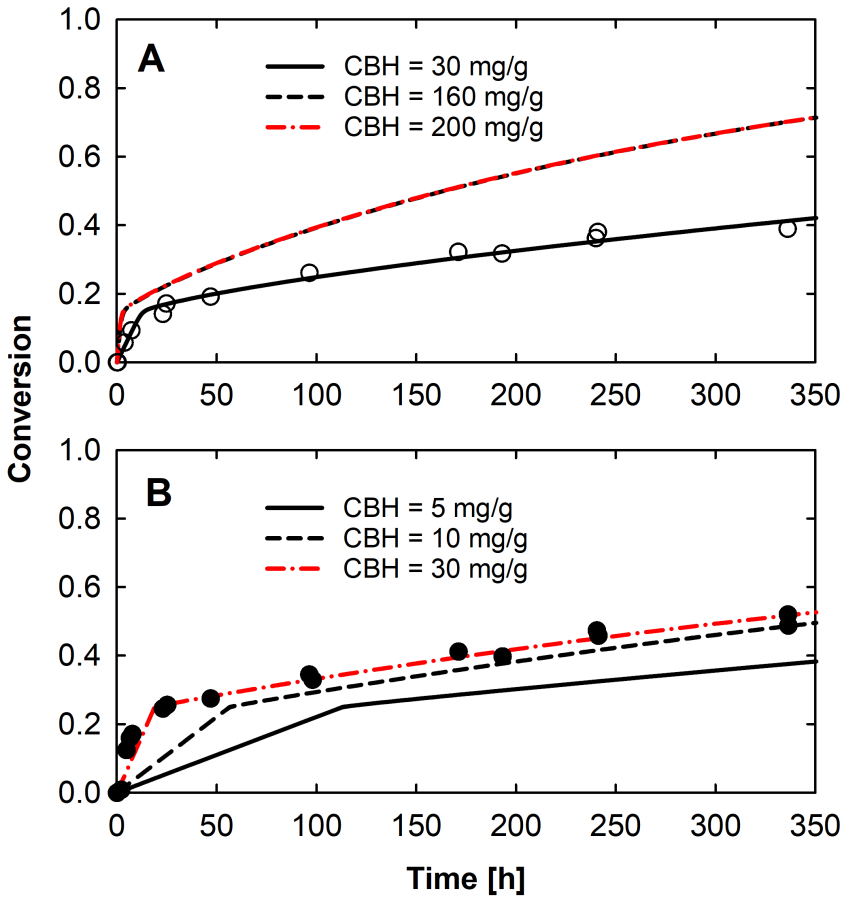

Figure 4.8: Demonstration of the effect of enzyme crowding on the surface of celluloses for hydrolysis of: (**A**) α-cellulose by CBH+BG, and (**B**) HCC by CBH+BG. In both cases, the loading of CBH is increased from enzyme-limiting condition to saturation point (onset of enzyme crowding effect) and finally to substrate-limiting condition (in excess of enzymes). BG loading is maintained at 2 mg/g cellulose for all cases. ML-PBM predictions are represented by the lines and the markers denote experimental data from Nag et al. [85].

substrate-limiting conditions, respectively. Figure 4.8A shows that for relatively small α-cellulose particles ($R(0) = 13$ nm), the saturation point is attained at CBH concentration of 160 mg/g cellulose and any increase in enzyme concentration beyond this value do not present any improvement over the conversion or the rate. On the contrary, Figure 4.8B indicates that for large HCC particles ($R(0) = 222$ nm), the saturation point is achieved at a mere CBH concentration of 10 mg/g cellulose. Unexpectedly, the ML-PBM predicts a substantial increase in the initial rate when the CBH concentration is increased beyond the saturation point, despite without a significant increase in conversion. Upon further scrutiny, the ML-PBM is describing a phenomenon which is termed "enzyme tunnelling effect" in this work, where the surface-adsorbed enzymes continue to complex with the newly exposed cellulose chains from the successive underlying layers, effectively freeing up the cellulose surface area and allowing more free enzymes to adsorb to the cellulose surface (see Figure 4.9). However, the enzyme tunnelling effect is only observable in the penetration zone as the rate of solubilization and the corresponding rate of exposure of underlying celluloses are relatively slower in the internal zone where the large molecules are prevalent. This also explains the reason for the lack of enzyme tunnelling effect on α-cellulose which is predicted to only have a single layer in its penetration zone as compared

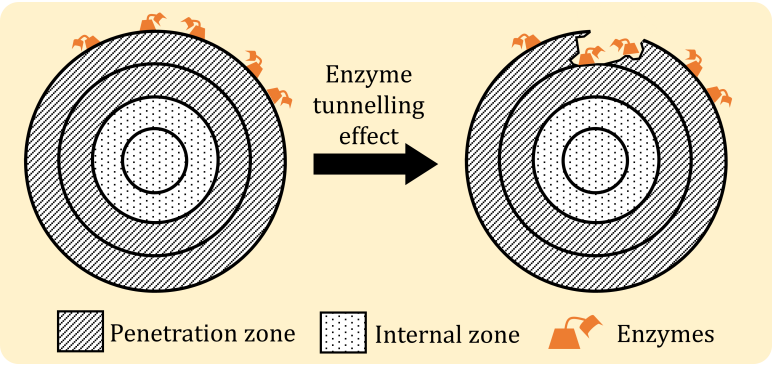

Figure 4.9: Illustration of enzyme tunnelling effect.

to 30 layers for HCC in this study.

Although enzyme tunnelling effect is yet to be corroborated with experimental findings, Grethlein [71], Luterbacher et al. [72] and Tanaka et al. [73] have explored the possibility of enzymes diffusing through the pores in the cellulose particles and hydrolysing the internal polymers. Though porous structure was not assumed for the cellulose particles in this study, the ability of the ML-PBM to trace analogous phenomena demonstrates the robustness of this model and expands the prospect of future model development.

4.3.7 ML-PBM Predicts the Dynamics of Fed-Batch Hydrolysis

In order to extend the relevance of the ML-PBM to industrial practices, the performance of the model is assessed in handling fed-batch hydrolysis. Figure 4.10 demonstrates the performance of the ML-PBM in predicting a fed-batch hydrolysis of steam-pre-treated spruce (SPS) cellulose by CBH. The ML-PBM rationally predicts the increase in product concentration when fresh SPS was instantaneously added at the 24^{th} h mark. Remarkably, the ML-PBM predicts a reduced rate with the addition of fresh SPS as compared to the rapid initial rate at the beginning of hydrolysis. The ML-PBM reveals that the decomplexation of CBH from old cellulose particles becomes limiting when fresh SPS was added to the existing reaction broth. As almost all CBH remained adsorbed to the substrate or remained complexed with cellulose chains after 24 h of hydrolysis, a sudden addition of fresh substrate poses an enzyme-limiting condition to the new cellulose populations because the existing CBH enzymes have to decomplex, desorb and subsequently re-adsorb to the newly added substrate.

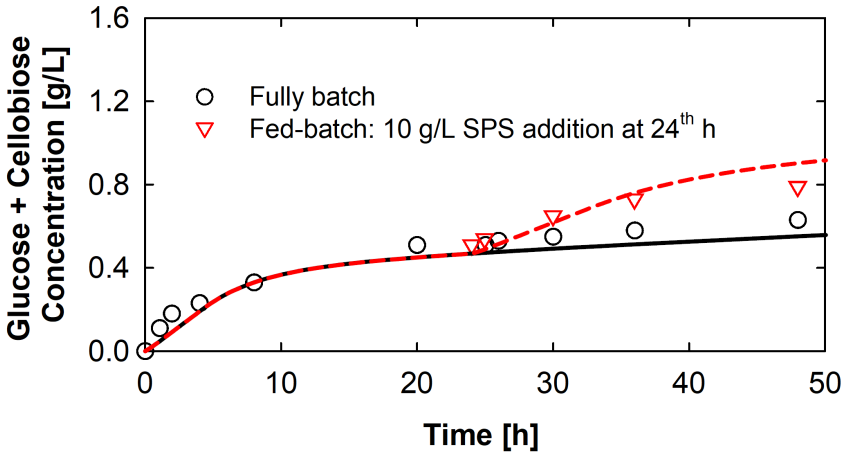

Figure 4.10: ML-PBM prediction of glucose and cellobiose concentrations for fed-batch hydrolysis of SPS by CBH. Initial loadings of SPS and CBH are 10 g/L and 9.53 mg/g, respectively. The ML-PBM was calibrated with the fully batch data and validated with the fed-batch data. The solid line is the fitted ML-PBM, dashed lines are the predictions and the markers are experimental data from Eriksson et al. [127].

The encouraging results imply that the ML-PBM can be potentially used for model-based optimizations and design of control systems for industrial scale processes which often employ fed-batch as an economical process scheme [95]. Although the main thrust of this work is fundamental, the ML-PBM serves as a future basis to model more challenging hydrolysis conditions encountered in the industry with high solid cellulose loadings and concomitant mass transfer limitations. Readers are referred to the studies by Pino et al. [143, 144] on the effect of bioreactor design and operation as well as the studies by Chakraborty et al. [145], Gaikwad [146], Gaikwad and Chakraborty [147] on the effects of mixing on enzymatic hydrolysis of celluloses under high solid loadings.

4.4 Concluding Remarks

When a population of cellulose particles is subjected to enzymatic hydrolysis, the feasibility of achieving rapid and efficient breakdown to valuable monomeric raw materials is frequently impeded by the significant slowdown of rates after the short-lived initial rapid phase. The proposed ML-PBM framework that has the capacity to account for heterogeneity in cellulose properties (cellulose crystallinity and DP) across structural layers of cellulose particles, is explored to potentially explain the underlying reason for this rate slowdown. Benchmarked against the homogenous model, the ML-PBM demonstrates far superior prediction where the full transient of enzymatic hydrolysis process including the rate slowdown phenomenon can be successfully traced using a global parameter set. This shows that the slowdown phenomenon in enzymatic hydrolysis of cellulose is potentially attributed to structural heterogeneity and knowing this allows one to rationally steer the process between the enzyme and substrate limiting regions. In addition, by accounting for the structural differences across the cellulose cross-section through the concept of penetration and internal zones, coupled with cellulose morphology, a notable merit of the ML-PBM is the ability to predict the extent of microfibril aggregation, and the probable arrangement of the cellulose chain distribution from limited data. Possible applications of this modelling tool include, but not limited to: (a) exploring the effects of existing pre-treatment methods on substrate-centric properties, (b) exploring new alternative pre-treatment methods that maximize the hydrolysis yield/rate, (c) exploring new techniques to produce commercial celluloses or biosynthesis of celluloses for various applications, and more. The introduction of the ML-PBM for a reasonable modelling of enzymatic hydrolysis of cellulose here fulfils the second research objective and sets the stage for the

subsequent development of CBP model by integrating modelling of metabolic systems, which is explored in the following chapter.

Chapter 5

Unified Cybernetic-Population Balance Model Predicts the Dynamics of Consolidated Bioprocessing of Cellulose

5.1 Preface

As indicated in Section 2.2.3, this chapter introduces the Unified Cybernetic-Population Balance Model (UC-PBM) to simulate CBP of cellulose. The UC-PBM features a closed-loop system that integrates the ML-PBM [100] developed in the previous chapter and the L-HCM developed by Song and Ramkrishna [36–38]. This chapter presents a thorough

examination on the metabolism of *Clostridium thermocellum* as a case study to showcase the effectiveness of the UC-PBM. The contents of this chapter are part of a manuscript in preparation for publication under the following designation:

- **F. Ahamed**, H.-S. Song, Y.K. Ho. 2020. Modelling Coordinated Enzymatic Control of Saccharification and Fermentation by *Clostridium thermocellum* During Consolidated Bioprocessing of Cellulose

5.2 Model Framework

5.2.1 Conceptual Description of Modelling Scheme

As cellulose is a non-permeating precursory substrate, CBP leads with the saccharification to produce fermentable sugars (Figure 5.1A). Commonly, cellulolytic microorganisms secrete basal levels of constitutive cellulases for initial degradation of celluloses to produce a small amount of permeable sugars that act as inducers to trigger the metabolism [148–150]. Consequently, the cellulases are produced at an adequate level (Figure 5.1B) to facilitate the saccharification of cellulose to maintain a constant supply of fermentable nutrients for the cells. The metabolic fluxes of the ingested nutrients are distributed to metabolic reactions spanning across vast metabolic networks (Figure 5.1C), where individual key intracellular enzymes catalyse the reactions. As saccharification and fermentation reactions are enzymatic reactions catalysed by extracellular and intracellular enzymes that are energetically expensive to synthesize, microorganisms optimally control the allocation of finite internal cellular resources among these two metabolic options to optimize their metabolic objective in varying environments. Finite cellular resources are allocated to reg-

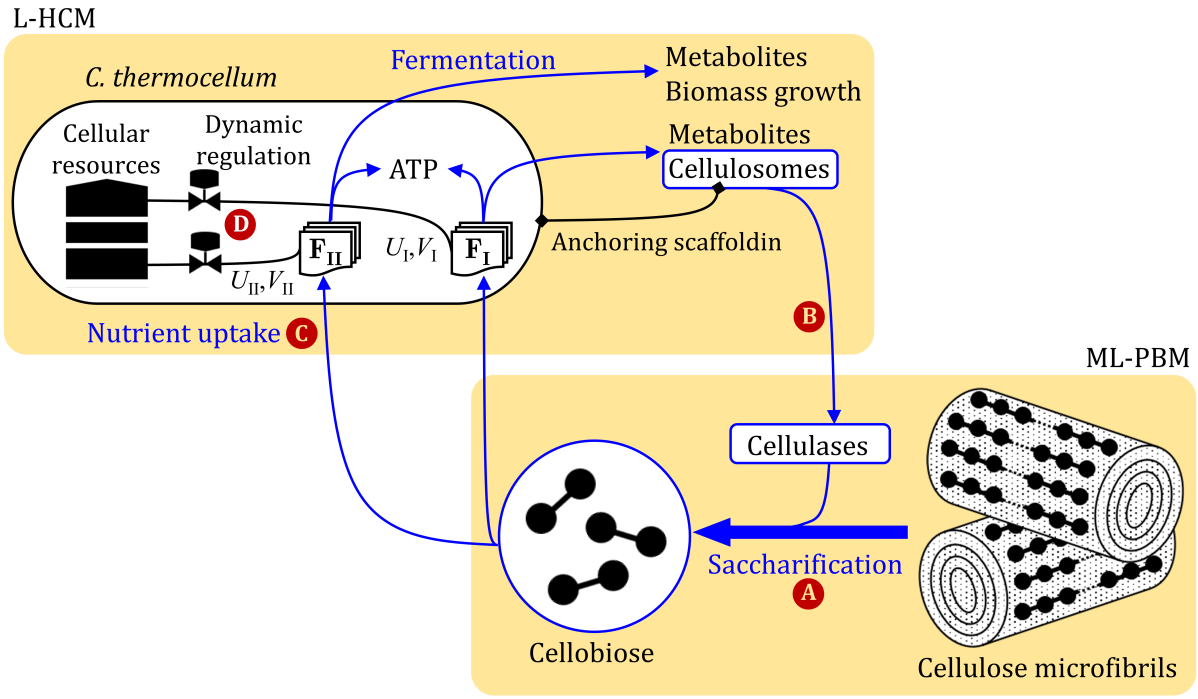

Figure 5.1: Illustration of CBP of cellulose as conceptualized by the UC-PBM. Here, the utilization of cellulose by *C. thermocellum* is shown as an example, where the UC-PBM can be similarly adapted for any other cellulolytic microorganisms. The blue arrows display a closed-loop system that includes the intracellular and the extracellular regulations that govern the CBP. The individual stages (**A** — **D**) of CBP are detailed in the text.

ulate the key intracellular enzymes (i.e., regulation at transcription and translation level, as well as the catalytic activity), which in turn governs the metabolic flux distribution.

In general, the combined saccharification and fermentation process proceeds through a closed-loop regulation, where CBP strains rely on extracellular cellulase regulation to produce an adequate amount of fermentable nutrients from the saccharification of cellulose to sustain the metabolism and in return, the extracellular cellulase regulation is governed by the intracellular flux distribution. Thus, a modelling framework that features a closed-loop system is crucial to portray the interdependent dynamics of CBP. In the formulation of the UC-PBM, the enzymatic saccharification of cellulose is modelled using the ML-

PBM and the intracellular flux distribution is defined by the L-HCM, where the latter dynamically allocates the cellular resources to families of basic elementary metabolic pathways consisting of a minimal sequence of metabolic reactions (Figure 5.1D) [26,36–38]. Ultimately, the coupling of the ML-PBM and the L-HCM features a closed-loop system via the extracellular cellulase secretion (Figure 5.1B) and the fermentable nutrient uptake pathways (Figure 5.1C) to describe the CBP.

5.2.2 General Formulation of Model Equations

In the foregoing section, a conceptual description of the CBP of cellulose as accounted for in the UC-PBM is provided. Here, the mathematical linkage between the ML-PBM and the L-HCM in the generalized UC-PBM formulation is demonstrated. The linkage is expressed through the balances of fermentable sugars, enzymes, metabolites and biomass as follows:

$$\frac{d\boldsymbol{\mathcal{N}}}{dt} = \underbrace{\boldsymbol{f}(\boldsymbol{k}_\gamma, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{C}}, R)}_{\substack{\text{Saccharification of} \\ \text{bulk cellulose} \\ \text{(ML−PBM)}}} - \underbrace{\boldsymbol{g}(\boldsymbol{k}_\lambda, \boldsymbol{\mathcal{K}}, \mathbf{e}, \mathbf{V}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{N}}, \boldsymbol{\mathcal{P}}, X)}_{\substack{\text{Microbial uptake of} \\ \text{fermentable sugars} \\ \text{(L−HCM)}}} \tag{5.1}$$

$$\frac{d}{dt}\begin{bmatrix} \boldsymbol{\mathcal{E}} \\ \boldsymbol{\mathcal{P}} \\ X \end{bmatrix} = \underbrace{\boldsymbol{h}(\boldsymbol{k}_\lambda, \boldsymbol{\mathcal{K}}, \mathbf{e}, \mathbf{V}, \boldsymbol{\mathcal{Z}}, \boldsymbol{\mathcal{N}}, \boldsymbol{\mathcal{P}}, X)}_{\substack{\text{Secretion of cellulases,} \\ \text{metabolites, and biomass} \\ \text{growth (L−HCM)}}} \tag{5.2}$$

Here, $\boldsymbol{\mathcal{S}}$ is the vector of all polymer species extending from monomer to polymer with maximum DP, $\boldsymbol{\mathcal{N}} \subseteq \boldsymbol{\mathcal{S}}$ is the vector of fermentable sugars, $\boldsymbol{\mathcal{E}}$ is the vector of extracellular cellulase species, $\boldsymbol{\mathcal{C}}$ is the vector of enzyme-polymer complexes, $\mathbf{e}$ is the vector of intracellular enzymes that catalyse the families of elementary metabolic pathways, $\boldsymbol{\mathcal{P}}$ is the vector of metabolites of interest secreted by the microorganisms and X is the biomass.

Subsequently, the model parameter $\boldsymbol{k}$ is the vector of rate coefficients, where the subscripts γ and λ are used to distinguish between the coefficients for the ML-PBM and the L-HCM, respectively, R is the average radius of cellulose particles, $\boldsymbol{\mathcal{K}}$ is the vector of MM-type saturation constants, $\mathbf{V}$ is the vector of cybernetic variables that control the activities of the enzymes $\mathbf{e}$, and matrix $\boldsymbol{\mathcal{Z}}$ supplies the stoichiometric information of the metabolic reactions. Except for $\boldsymbol{k}$, $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{\mathcal{Z}}$, all variables are functions of time.

The first term in Eq. (5.1) is directly adopted from the ML-PBM, where it depicts the saccharification of parent polymer chains which produces soluble sugars, and the second term is added to account for the consumption of the soluble sugars by microorganisms. On the other hand, Eq. (5.2), given by the L-HCM, expresses the balances of the extracellular cellulases, metabolites and biomass. Note that Eq. (5.1) is dependent on $\boldsymbol{\mathcal{E}}$ and the whole of second term, both of which are defined by the L-HCM, whereas Eq. (5.2) is a function of $\boldsymbol{\mathcal{N}}$ defined by the ML-PBM, thus, an interdependent closed-loop system is established. Readers are referred to Chapter 4/Appendix B and to the studies by Song and Ramkrishna [36–38] for the conceptualization, derivation and solution of the ML-PBM and the L-HCM, respectively.

The generalized UC-PBM formulation presented here is amenable to suit various scenarios. For example, the UC-PBM can be used to model the metabolism of any cellulolytic microorganisms provided that adequate biological information (i.e., metabolic network, exometabolomic data, etc.) exists. Moreover, modelling of the metabolism of co-cultures with competing/complementing affinity over common fermentable sugars is also a possibility. SSF process with exogenous cellulase addition also can be modelled by eliminating the cellulase secretion by microorganism and limiting the extracellular cellulase species $\boldsymbol{\mathcal{E}}$ to a time-independent total concentration, where the closed-loop system

ceases to exist. Lastly, various cellulosic substrates (e.g., Avicel, α-cellulose) and cellulases (endo- and/or exo-acting enzymes) can be modelled by varying substrate-centric and enzyme-centric parameters of the ML-PBM accordingly.

5.3 Case Study

5.3.1 Motivations to Examine Cellulose Utilization by *Clostridium thermocellum*

Using the generalized expressions in the preceding section, the UC-PBM is formulated for cellulose utilization by *Clostridium thermocellum* as a case study. *C. thermocellum* is an attractive anaerobic and thermophilic bacterium for CBP owing to its high efficiency in consuming recalcitrant celluloses [11, 109, 148, 151]. Supramolecular multienzyme protein complexes, called cellulosomes, are able to degrade cellulose to mainly fermentable cellobiose [151–153] with high efficiency, which is then consumed to produce various end metabolites including ethanol, lactate, formate, acetate, hydrogen and carbon dioxide [148].

The primary goal in strain development efforts of *C. thermocellum* is the maximization of fermentation capacity to produce target biochemicals at industrially relevant levels without compromising the rapid cell growth and cellulolytic ability needed to utilize cellulose efficiently. Despite extensive research, the complete physiology of *C. thermocellum* remains elusive as most computational studies are limited to steady-state constraint-based approaches [109, 154–156] and simplified unstructured kinetics [157–160]. Therefore, it is

incumbent upon this research to employ the UC-PBM to examine the role and significance of cellulosome-secreting metabolic pathways in the overall metabolism of *C. thermocellum*. However, cellulosome is simply an intermediate pseudo-metabolite that solubilizes cellulose into fermentable cellobiose, whereas the microorganism ultimately steers the metabolism towards ensuring survival, i.e., biomass growth. Hence, the relative importance and contention for cellular resources between cellulosome-secreting and growth-coupled pathways is examined in this work. Ultimately, the role of the diverse metabolic pathways of *C. thermocellum* under varying environmental conditions and the characteristics of the in-built cellular regulatory mechanism in navigating the diverse metabolic pathways is investigated. To this end, the structured UC-PBM that actively accounts for the dynamics, cellular regulation, and closed-loop system will certainly be beneficial to corroborate the existing inferences and to uncover possible unrecognized metabolic diversity.

5.3.2 Abstraction of Core Metabolism from Genome-Scale Metabolic Network

The core metabolic network of *C. thermocellum* is extracted from the genome-scale metabolic network (iAT601) developed by Thompson et al. [155] to avoid losing the crucial cellulosome secretion pathways, which are often omitted in central metabolic networks of the microorganism in the literature [109]. To avoid the combinatorial explosion of the number of elementary metabolic pathways due to the sheer size of the genome-scale metabolic network, a network reduction algorithm called the *NetworkReducer* [161] is utilized to condense the genome-scale network into a meaningful core network encompassing the

cellulosome secretion, the fermentation capacity and the biomass growth. All essential metabolites and reactions in the genome-scale network that also occur in the core network [109] are flagged as protected to retain them in the reduced network, in addition to the cellulosome and cellulosome-secreting reactions. The full set of participating reactions and metabolites in the reduced network are given in Appendix C.1. The resulting core network is used for subsequent enumeration of elementary metabolic pathways.

Thompson et al. [155] present two separate metabolic networks for *C. thermocellum* by varying cellulosome fractions in the abiotic phase to 20% and 2% of the dry cell weight when grown on cellulose and cellobiose, respectively, based on empirical observation from the study by Zhang and Lynd [162, 163]. As both syntheses of cellulosome and cell mass proteins share the same amino acids, the basis of fixing only 20% (or 2%) of the total cellulosomes as extracellular secretion compared to 100% of cell mass for biomass growth is unclear. Moreover, Zhang and Lynd [162,163] made clear distinctions between cellulase and non-cellulase protein in their testing protocol using enzyme-linked immunosorbent assay (ELISA), thus, one can infer that the total cellulase that made up to 20% (or 2%) of total cell mass is pure cellulase exclusive of non-cellulase components of cellulosomes. Due to the inconsistencies, the whole of cellulosomes is treated as an enabling extracellular pseudo-metabolite in this work, and the distinction between cellulase and non-cellulase components of cellulosome is made based on the cellulosomal compositions reported in the literature [151, 152]. Accordingly, the yields of biomass are calculated by discounting the cellulase component from the yields of dry cell weights obtained from the metabolic network, i.e.,

$$Y_{\text{biom}}\omega_{\text{DW}} = Y_{\text{DW}}\left[\omega_{\text{biom}} + \omega_{\text{csm}}(1 - \omega_{\text{cls}})\right] \tag{5.3}$$

where Y_{biom} is the yield of biomass discounting the cellulase proteins, Y_{DW} is the yield of

dry cell weights inclusive of cell mass and cellulosome derived directly from the metabolic networks, and the mass fractions of dry cell weight, biomass, cellulosome and cellulase are $[\omega_{\mathrm{DW}}, \omega_{\mathrm{biom}}, \omega_{\mathrm{csm}}, \omega_{\mathrm{cls}}] = [1.20, 1.00, 0.20, 0.50]$ or $[1.02, 1.00, 0.02, 0.39]$ when *C. thermocellum* is grown on cellulose or cellobiose, respectively. In subsequent texts, the metabolic networks for *C. thermocellum* grown on cellulose and cellobiose are referred to as $\mathrm{N}_{20\%}$ and $\mathrm{N}_{2\%}$, respectively, and the former are primarily used as the main focus is the CBP of cellulose in this work.

5.3.3 Potential Trade-Offs between Cellulosome-Secreting and Growth-Coupled Metabolic Pathways

Cybernetic models typically use elementary flux modes (EFMs) to describe the metabolic pathways, where the EFMs form unbounded flux cone through homogenous constraints arising from the assumption of steady-state concentrations for internal metabolites and reaction irreversibilities [164]. While the EFM analysis is a powerful tool to study metabolic networks, the study on metabolic engineering interventions, such as rate/yield maximizations are futile, as the solutions are unbounded maximums [165]. In this work, elementary flux vectors (EFVs) are utilized instead, where a bounded polyhedron is formed by imposing maximum cellobiose uptake rate (13 mmol/gDW/h) [154] and non-negative fluxes for extracellular metabolite-secreting reactions as additional inhomogeneous constraints. The EFVs are enumerated using *efmtool* [166] included in *CellNetAnalyzer* toolbox [167, 168].

The distributions of lengths of the enumerated EFVs in Figure 5.2 show that the cellulosome secretion imposes a significant metabolic burden on the metabolism of the *C. thermocellum*, as the length of an EFV directly corresponds to the demand of cellular

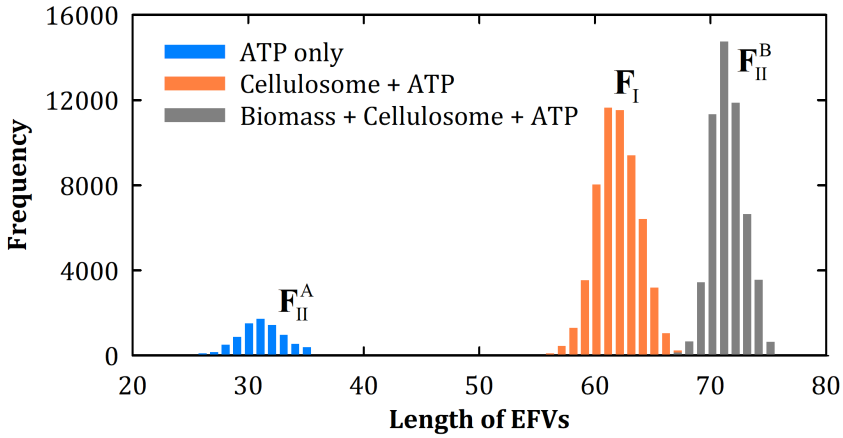

Figure 5.2: Distributions of EFV lengths of *C. thermocellum* for different groups classified according to secretion of representative products. The length refers to the number of participating metabolic reactions in each individual EFV. Here, $N_{20\%}$ is used for the enumeration of the EFVs.

resources to instil metabolic flux along the pathway [169]. Moreover, the distributions of the EFVs show multiple peaks, which indicate multiple essential pathways around which metabolic deviations can occur. The estimation of ATP requirements in *C. thermocellum* (57 mmol ATP/g cellulosome/h vs. 13.5 mmol ATP/gDW/h) also indicates that the synthesis of cellulosome is more energy-intensive than biomass growth [155]. As *C. thermocellum* is known to also down-regulate the expression of cellulases in the adequate presence of fermentable cellobiose [151, 163], at this point, one can infer that there are potential metabolic trade-offs in investing internal resources for cellulosome secretion and biomass growth. However, all growth-coupled EFVs also co-secrete cellulosomes (Figure 5.2), albeit at lower yields. Such an overlap in the metabolic functions in the EFVs is counter-intuitive, and the roles of the EFVs on the metabolism of *C. thermocellum* is investigated by modelling with the UC-PBM.

To study the nature of the in-built regulatory mechanism of *C. thermocellum* in steering the metabolic fluxes and to account for the potential contention of finite cellu-

Table 5.1: Number of EFVs enumerated from reduced metabolic network ($N_{20\%}$) of *C. thermocellum*. Thermodynamically infeasible EFVs with no substrate uptakes are omitted.

EFV family	Representative product	Cellulosome- & ATP-producing EFVs, n_{z1}	Biomass- & ATP-producing EFVs, n_{z2B}	ATP-producing EFVs, n_{z2A}	Total EFVs
$\mathbf{F}_{\mathrm{I}}$	Cellulosome	57,030	-	-	57,030
$\mathbf{F}_{\mathrm{II}}$	Biomass	-	53,184	8,481	61,665
				Total EFVs, n_z	118,695

lar resources between the inherent energy-intensive functions of the microorganism, the EFVs are grouped into two principal families, namely $\mathbf{F}_{\mathrm{I}}$ that represents growth-decoupled cellulosome secretion and $\mathbf{F}_{\mathrm{II}}$ which contributes to biomass growth and growth-coupled constitutive cellulosome secretion. As a result of this grouping strategy, production of other metabolites of interest are shared between both families. Following the protocol of the L-HCM [37], the $\mathbf{F}_{\mathrm{II}}$ is further split into biomass- and ATP-producing ($\mathbf{F}_{\mathrm{II}}^{\mathrm{B}}$) and only ATP-producing ($\mathbf{F}_{\mathrm{II}}^{\mathrm{A}}$) sub-groups. Similar classification is not performed to $\mathbf{F}_{\mathrm{I}}$ as all EFVs in the family co-produces cellulosomes and ATP. The summary of the number of EFVs in each family and sub-group is presented in Table 5.1.

5.3.4 EFV Lumping Scheme in the UC-PBM

The procedures employed to lump the EFVs in the UC-PBM are illustrated in Figure 5.3. While the cellular resources (and by extension, the metabolic fluxes) to the two EFV families are regulated according to the environmental conditions and the competing metabolic functions of the microorganism, the subsequent flux distribution to the individual EFVs is solely based on the stoichiometry. The latter is reflected in the lumping of the EFVs

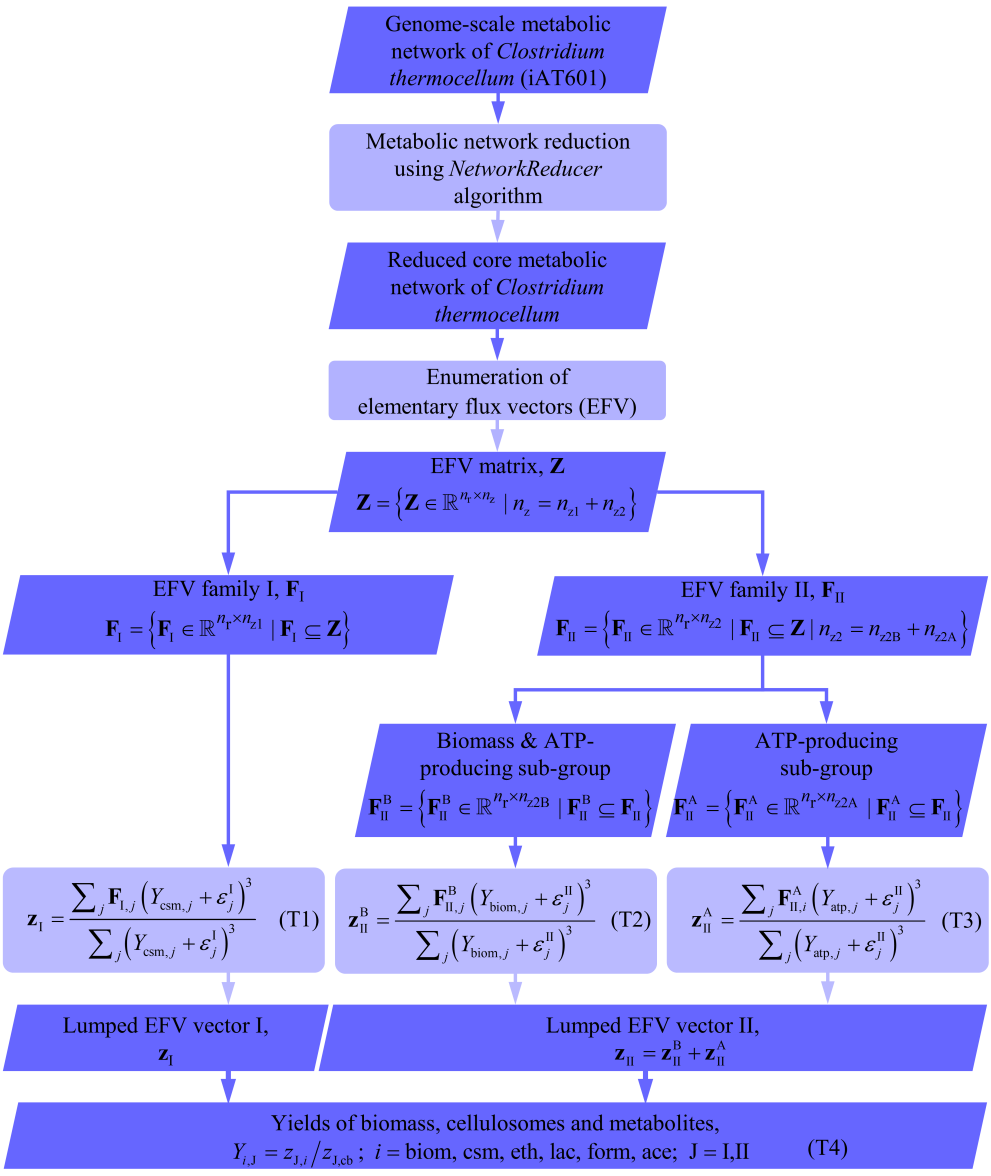

Figure 5.3: A flowchart depicting the extraction of essential biological information for implementation of the UC-PBM. Here, a meaningful core metabolic network encompassing the secretion of cellulosomes, metabolites of interest and the biomass growth is obtained from the genome-scale metabolic network of *C. thermocellum* (iAT601) [155]. EFVs are classified into two principal families – $\mathbf{F}_I$ that contributes to inducive secretion of cellulosomes and $\mathbf{F}_{II}$ that allows biomass growth and growth-associated cellulosome secretion. Both families contribute to the secretion of all metabolites of interest.

by taking a weighted-average (Eq. (T1) – (T3) in Figure 5.3). The EFV lumping scheme based on the concept of the L-HCM is detailed in Appendix C.2.

5.3.5 Model Equations

The preceding sections deliberated the aim of the study, special considerations made to model the metabolism of *C. thermocellum* as well as the enumeration, classification and lumping of the metabolic pathways. Here, the key UC-PBM equations that trace the cellulose utilization by *C. thermocellum* are presented in Table 5.2.

Table 5.2: Key UC-PBM equations.

Cellobiose:

$$\frac{dC_{\mathrm{cb}}}{dt} = f(\boldsymbol{k}_{\gamma}, \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{C}}, R) - X \sum_{\mathrm{J}} r_{\mathrm{J}} \tag{5.4}$$

Intracellular enzymes:

$$\frac{d\mathbf{e}}{dt} = \boldsymbol{\alpha} + \mathbf{D}(\mathbf{U})\mathbf{r}_{\mathrm{e}}^{\mathrm{kin}} - [\mathbf{D}(\boldsymbol{\beta}) + \mu\mathbf{I}]\mathbf{e} \tag{5.5}$$

Cellobiose uptake fluxes:

$$\mathbf{r} = \mathbf{V}\mathbf{e}^{\mathrm{rel}}\mathbf{r}^{\mathrm{kin}} \tag{5.6}$$

Metabolites and biomass:

$$\frac{d}{dt}\begin{bmatrix} X & C_{\mathrm{eth}} & C_{\mathrm{lac}} & C_{\mathrm{form}} & C_{\mathrm{ace}} \end{bmatrix}^{\mathrm{T}} = X\mathbf{Y}\mathbf{r} \tag{5.7}$$

Cellulosome:

$$\frac{dC_{\mathrm{csm}}}{dt} = \alpha_{csm} X + X \sum_{\mathrm{J}} Y_{\mathrm{csm,J}} r_{\mathrm{J}} - \beta_{\mathrm{csm}} C_{\mathrm{csm}} \tag{5.8}$$

First, the balance for the fermentable cellobiose in Eq. (5.4) is obtained by re-writing the general Eq. (5.1), where r_{J} represents the cellobiose uptake flux through the J-th EFV family. For brevity and to avoid clutter, the first term of Eq. (5.4) which originates from the ML-PBM formulation is given in the general form. The full set of the ML-PBM equations are given in Appendix C.3.1. Although uptake of long-chained cellodextrins (DP 3 to 6) by *C. thermocellum* is also suggested in the literature [101,148], the extent of their assimilation relative to cellobiose remains unclear. Moreover, as the hydrolysis of cellodextrins is rapid in the presence of adequate cellulases, cellodextrins do not accumulate and cellobiose is often the only product of hydrolysis available for cell uptake. While there is the possibility of contention for cellodextrins for cell uptake as well as for subsequent hydrolysis to cellobiose, the metabolic strategy of the microorganism in choosing the growth substrate remains elusive, though it is generally thought to be highly dependent on the environmental conditions [101]. Therefore, due to the lack of definitive evidence on cellodextrins uptake, this study assumes complete hydrolysis of cellulose to cellobiose as the sole intermediate product of hydrolysis available for subsequent cell uptake. Next, the balances of intracellular enzymes and cellobiose uptake fluxes in Eqs. (5.5) and (5.6), respectively, are based on the concept of the L-HCM [36–38], which are detailed in Appendix C.3.2. Subsequently, the balances of the biomass and metabolites of interests are presented in Eq. (5.7), where $\mathbf{Y}$ is the matrix of model-estimated lumped yields (Eq. (T4) in Figure 5.3) with the configuration $Y_{i,\mathrm{J}}$ (i = biom, eth, lac, form, ace; J = I, II).

Lastly, the balance of cellulosome secretion is given separately in Eq. (5.8). Out of the total cellulosome secretions, the cellulases are assumed to make up about 50 wt% and 39 wt% when grown on cellulose and cellobiose, respectively, based on the cellu-

losomal compositions [151, 152]. Moreover, the cellulosomes of *C. thermocellum* consist of a multitude of catalytic units, thus, accounting for their catalytic actions individually in the model will lead to overparameterization. Therefore, the enzymatic actions of the catalytic units are consolidated into exo- and endo-acting enzyme groups performing chain-end dimeric and random scission, respectively. Although the molar ratio of exo- to endo-enzymes in the cellulosome can be discerned from the literature [151, 152], the ratio is suspected to be higher in actuality, as some endo-acting catalytic units are found to adapt to the hydrolytic mechanism of exo-acting catalytic units due to evolutionary implications [170]. To identify the most probable composition, the molar ratio is empirically adjusted while maintaining the total moles until the production ratio of cellobiose to glucose matched the reported values in the literature. As a result, the total molar concentrations of exo- and endo-enzymes used in this work are presented as follows:

$$
\begin{aligned}
C_{\mathrm{T}}^{\mathrm{exo}}(t) &= \begin{cases} 10.97\left[C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}\right]; & \text{when } \textit{C. thermocellum} \text{ grown on cellulose} \\ 9.39\left[C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}\right]; & \text{when } \textit{C. thermocellum} \text{ grown on cellobiose} \end{cases} \\
C_{\mathrm{T}}^{\mathrm{endo}}(t) &= \begin{cases} 0.11\left[C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}\right]; & \text{when } \textit{C. thermocellum} \text{ grown on cellulose} \\ 0.09\left[C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}\right]; & \text{when } \textit{C. thermocellum} \text{ grown on cellobiose} \end{cases}
\end{aligned}
\tag{5.9}
$$

where, $M_{\mathrm{csm}} \approx 1.73$ MDa and 1.89 MDa are the molecular weights of the cellulosomes when *C. thermocellum* is grown on cellulose and cellobiose, respectively [151, 152].

Additionally, the complexed cellulase system of *C. thermocellum* modelled here for CBP is fundamentally different from the non-complex fungal cellulases (CBH, EG and BG) used for the design of the ML-PBM framework in Chapter 4. Although the complexed cellulase system differs in interactions against bulk celluloses, they share similar modes of enzymatic scissions with their non-complex counterparts. Moreover, a substantial portion of the catalytic units from cellulosomes are found disassociated and non-complexed in

reaction broth as CBP proceeds [163]. Given the certain degree of similarities between the two cellulase systems, the original design of the ML-PBM is amenable for both CBP by complexed cellulases and enzymatic hydrolysis by non-complex cellulases with minor adjustments, as detailed in Appendix C.6.

5.3.6 Metabolic Objective of *C. thermocellum*

Here, the postulation on the metabolic objective of *C. thermocellum* is presented, where the objective binds the allocation of cybernetic variables $\mathbf{U}$ and $\mathbf{V}$ in Eqs. (5.5) and (5.6), respectively. As alluded to in the previous sections, one can posit that *C. thermocellum* bears multiple contending metabolic functions that allow efficient cellulose utilization alongside rapid biomass growth [151]. Unlike simpler microorganisms, *C. thermocellum* not only safeguards its survival through biomass growth but also ensures a constant and adequate supply of fermentable nutrients by secreting cellulosomes to break down cellulose. The foregoing metabolic functions are reflected in the classification of the EFVs employed in this work (cf. Figure 5.3 and Section 5.3.3). Given that cellulosome secretion and biomass growth are both energy-intensive, the two metabolic functions are potentially tailored for different environmental conditions and cannot be optimized simultaneously.

Essentially, the flux distribution is manoeuvred towards optimizing the metabolic objective of the microorganisms. The common postulates for the metabolic objectives are the maximization of the biomass growth rate or the substrate uptake rate, where both are pertinent to the survival of the microorganisms [27, 36, 171]. Although maximization of the biomass growth rate and substrate uptake rate are used as metabolic objective interchangeably in the metabolic models of simple microorganisms in the past, the same

cannot be assumed for CBP strains as the substrates can be consumed for other purposes than the biomass growth (such as growth-independent cellulosome secretion). Motivated by the above, maximization of cellobiose uptake rate is selected as the metabolic objective of *C. thermocellum* and the dynamic allocation of the cybernetic variables $\mathbf{U}$ and $\mathbf{V}$ are established through the Matching and Proportional Laws [171], respectively:

$$\mathbf{U} = \frac{\mathbf{\Phi}}{\|\mathbf{\Phi}\|_1}; \quad \mathbf{V} = \frac{\mathbf{\Phi}}{\|\mathbf{\Phi}\|_\infty} \tag{5.10}$$

where $\mathbf{\Phi} = \mathbf{e}^{\text{rel}}\mathbf{r}^{\text{kin}}$ is the return-on-investments (ROIs) of cellular resources which represent the cellobiose uptake rate through the EFV families.

Based on the generalized cybernetic laws, the control above implies that the microorganism immediately reaps the benefit of the imposed control actions [19,26,171], i.e., maximization of cellobiose uptake rate. However, the hydrolysis of cellulose is rate-limiting in CBP and the production of cellobiose is not instantaneous following the secretion of cellulosomes. Nevertheless, typical CBP by *C. thermocellum* entails a condition where the cellobiose concentration is maintained sufficiently low [172] and the estimated saturation constant of cellobiose is adequately small such that the quasi-instantaneous attainment of the ROI can be established [171].

5.4 Results and Discussion

5.4.1 UC-PBM Predicts Coupled Dynamics of Saccharification and Fermentation Regulated by *C. thermocellum*

The values of model parameters used to generate simulation results in this work are collated in Table 5.3. Here, the model parameters that display poor sensitivity or identifiability (not shown) following the methodology outlined by Brun et al. [173] are either fixed to values estimated for the ML-PBM, obtained from the literature or set to a reasonable order of magnitude. All endo-enzyme-related parameters fall in this category as they only make up a small portion of the total cellulases. Substrate-centric parameter values (P_{zone} and $r_{\text{mass,P}}$) are directly adopted from Chapter 4, where the values were consistent and universal for Avicel across multiple data sets. The ensuing simulation results are presented as follows.

Figure 5.4 and 5.5 show quantitatively decent model fits and predictions of the metabolism of *C. thermocellum* with Avicel as the sole carbon source. Notably, by only using the transient data sets of residual Avicel (denoted by solid markers) for parameter estimation, the UC-PBM predicts other crucial descriptors (denoted by open markers) such as biomass growth, cellulase and metabolite secretions. This is only possible with the UC-PBM that features a closed-loop system, where the rate of hydrolysis of Avicel dictates the production of fermentable cellobiose and the consumption of the cellobiose controls the cellulosome (cellulase) secretion, which in turn ultimately limits the rate of hydrolysis. The only exception to the decent predictions is the overprediction observed

Table 5.3: Values of model parameters used in this study. Conceptualization of the ML-PBM parameters are presented in Chapter 4/Appendix B.

Parameters	Unit	Value	Remarks
ML-PBM			
$k_{\mathrm{h}}^{\mathrm{exo}}$	1/s	2.15	Fitted to data.
$k_{\mathrm{e}}^{\mathrm{exo}}/k_{\mathrm{p,f}}^{\mathrm{exo}}$	mol L^{-1} DP^{-1}	3.33×10^{-7}	Fitted to data.
σ_{exo}	$\mathrm{mol/m^2}$	2.22×10^{-8}	Estimated value from Chapter 4 for exo-enzyme.
$k_{\mathrm{d}}^{\mathrm{exo}}/k_{\mathrm{a}}^{\mathrm{exo}}$	mol/L	1.36×10^{-8}	Estimated value from Chapter 4 for exo-enzyme.
$k_{\mathrm{h}}^{\mathrm{endo}}$	1/s	4.00	Estimated value from Chapter 4 for endo-enzyme.
$k_{\mathrm{h'}}^{\mathrm{endo}}$	L mol^{-1} s^{-1}	3.00	Estimated value from Chapter 4 for endo-enzyme.
$k_{\mathrm{e}}^{\mathrm{endo}}/k_{\mathrm{f}}^{\mathrm{endo}}$	mol/L	1.00×10^{-8}	Estimated value from Chapter 4 for endo-enzyme.
σ_{endo}	$\mathrm{mol/m^2}$	8.04×10^{-9}	Estimated value from Chapter 4 for endo-enzyme.
$k_{\mathrm{d}}^{\mathrm{endo}}/k_{\mathrm{a}}^{\mathrm{endo}}$	mol/L	1.41×10^{-8}	Estimated value from Chapter 4 for endo-enzyme.
P_{zone}	–	1	Estimated value from Chapter 4 for Avicel.
$r_{\mathrm{mass,P}}$	%	2.80	Estimated value from Chapter 4 for Avicel.
L-HCM			
$\alpha_{\mathrm{I}} = \alpha_{\mathrm{II}}$	1/s	2.78×10^{-5}	Following Song and Ramkrishna [36].
$\beta_{\mathrm{I}} = \beta_{\mathrm{II}}$	1/s	5.56×10^{-5}	Following Song and Ramkrishna [36].
$k_{\mathrm{e,I}} = k_{\mathrm{e,II}}$	1/s	2.78×10^{-4}	Following Song and Ramkrishna [36].
$K_{\mathrm{e,I}}$	mol/L	1.00×10^{-6}	Assumed equal to K_{I} for simplicity.
$K_{\mathrm{e,II}}$	mol/L	9.00×10^{-6}	Assumed equal to K_{II} for simplicity.
$k_{\mathrm{I}}^{\mathrm{max}}$	mol gDW^{-1} s^{-1}	1.30×10^{-6}	Fitted to data.
$k_{\mathrm{II}}^{\mathrm{max}}$	mol gDW^{-1} s^{-1}	1.90×10^{-6}	Fitted to data.
K_{I}	mol/L	1.00×10^{-6}	Fitted to data.
K_{II}	mol/L	9.00×10^{-6}	Fitted to data.
$K_{\mathrm{inh-eth}}$	mol/L	1.74	Value reported by Hogsett et al. [174] for 50% inhibition by ethanol.
$K_{\mathrm{inh-glu}}$	mol/L	1.22×10^{-3}	Value reported by Hogsett [175] for 50% inhibition by glucose.
α_{csm}	g gDW^{-1} s^{-1}	3.00×10^{-7}	Set sufficiently small to avoid haphazard accumulation of cellulosomes.

for the steady-state secretion of acetate in Figure 5.5. Unlike other metabolites, acetate is consumed in vivo for the production of the highly versatile acetyl-CoA, which is the

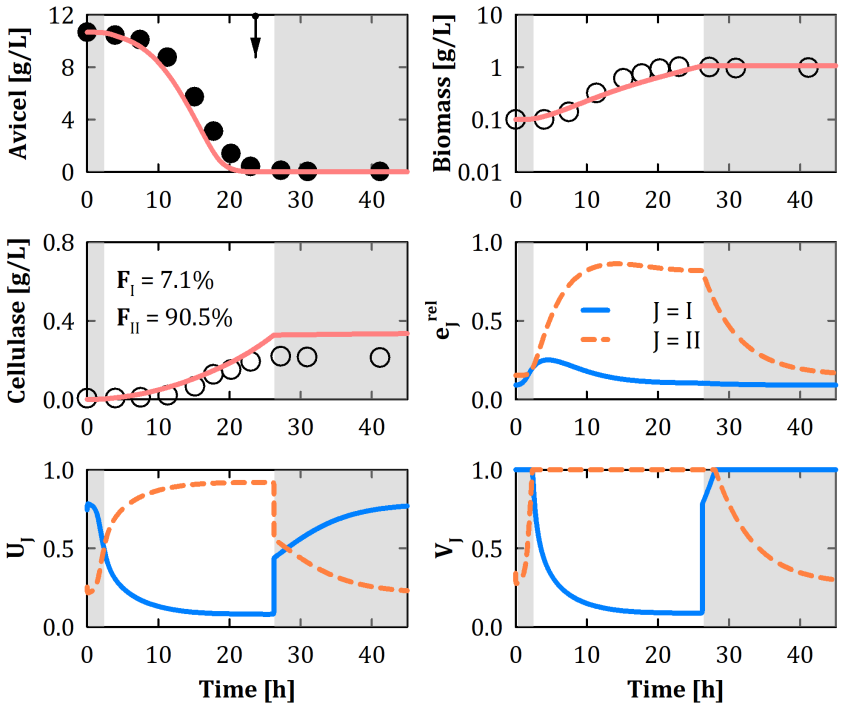

Figure 5.4: Model prediction of metabolism of *C. thermocellum* ATCC 27405 grown on 10.67 g/L Avicel as the sole carbon source. The markers represent experimental data from Zhang and Lynd [163] and the lines are model predictions. The solid markers represent data used for parameter estimation, while the model predicts the data represented by open markers. The grey-shaded regions denote periods of severe famine conditions and the drop-down arrow indicates the time for complete cellulose solubilization. The fractions of cellulase secreted through $\mathbf{F}_{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{II}}$ are given in the figure, where the remaining fraction is contributed by the constitutive secretion. The total cellobiose uptakes through the respective EFV families are 7.9% and 92.1%. Here, $\mathrm{N}_{20\%}$ is used to demonstrate the metabolism.

starting metabolite for many major metabolic pathways, including the TCA cycle and biosynthesis of fatty acids [176]. Therefore, the complex metabolic pathways involving acetate may be haphazardly truncated during the metabolic network reduction routine.

Apart from the quantitative fit to the experimental data, some interesting qualitative observations could be gleaned from the UC-PBM outputs. In the beginning of the CBP, where the concentration of cellobiose is minimal, $e_{\mathrm{I}}^{\mathrm{rel}}$ is swiftly induced to allow majority

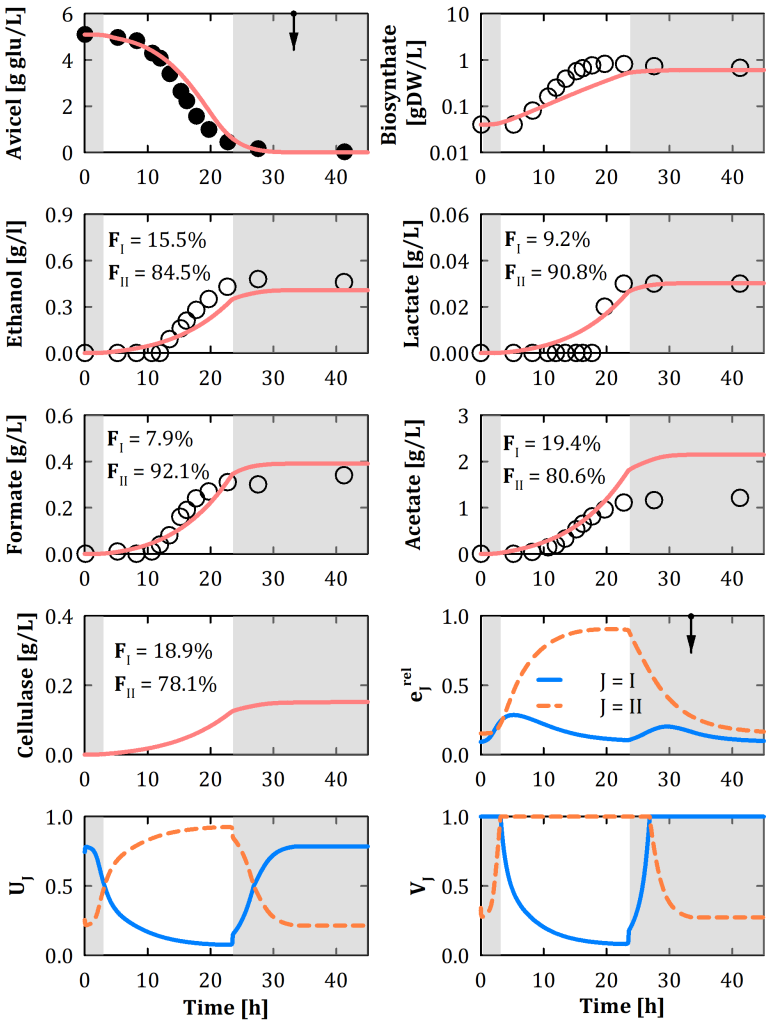

Figure 5.5: Model prediction of metabolism of C. thermocellum ATCC 27405 grown on 4.59 g/L Avicel (5.10 g glu eq/L) as the sole carbon source. The lines are model predictions and the markers are experimental data from Holwerda et al. [172]. Solid markers represent data used for parameter estimation, while the model predicts the data represented by open markers. The biomass growth is expressed in terms of biosynthate dry weight that constitutes cell mass, as well as cell-bound and supernatant protein (i.e., cellulosomes), whereas residual Avicel is presented in glucose equivalents following the data source. The grey-shaded regions denote periods of severe famine conditions and the drop-down arrows indicate the time for complete cellulose solubilization. The fractions of metabolites secreted through $\mathbf{F}_{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{II}}$ are given in the figure, whereas the total cellobiose uptakes through the respective EFV families are 12.7% and 87.3%. The residual fraction of cellulase is contributed by the constitutive secretion. Here, $\mathrm{N}_{20\%}$ is used to demonstrate the metabolism.

of the fluxes to flow through $\mathbf{F}_{\mathrm{I}}$. This can be viewed as a pragmatic strategy that *C. thermocellum* adopts for rapid secretion of cellulosomes at high yield ($Y_{\mathrm{csm,I}} > 2Y_{\mathrm{csm,II}}$, cf. Table C.3 in Appendix C.2) to promote the production of additional fermentable cellobiose without apportioning limited cellular resources and energy for biomass growth. Once cellobiose accumulates to an adequate level, $e_{\mathrm{I}}^{\mathrm{rel}}$ is repressed and $e_{\mathrm{II}}^{\mathrm{rel}}$ is instead induced to direct the fluxes through $\mathbf{F}_{\mathrm{II}}$ to promote biomass growth. Evidently, the metabolic shift from $\mathbf{F}_{\mathrm{I}}$ to $\mathbf{F}_{\mathrm{II}}$ marks the transition from the initial lag phase to exponential growth phase. The model predictions suggest that the growth-decoupled cellulosome-secreting pathways ($\mathbf{F}_{\mathrm{I}}$) are devoted for famine environments, whereas the growth-coupled pathways ($\mathbf{F}_{\mathrm{II}}$) dictate the metabolism in feast environments. The secretion of other metabolites such as ethanol, lactate, formate and acetate are, however, shared between $\mathbf{F}_{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{I}}$, albeit at different rates and yields.

The foregoing model predictions concur with the inferences that cellulosome secretions are biomass growth rate-dependent and/or regulated by the amount of available cellobiose through catabolite repression [151,163,177–179]. Moreover, the model also closely agrees with the rationale that *C. thermocellum* synchronizes cellobiose uptake with hydrolysis, where cellulosome secretion is quickly slowed down by shifting the metabolic fluxes to $\mathbf{F}_{\mathrm{II}}$ before cellobiose accumulates to a level that can inhibit cellulase activity [153]. In all cases examined in this work, cellobiose does not accumulate significantly and is maintained at a concentration of at least an order of magnitude lower than 2% which induces near-complete inhibition of cellulases [175,180–182].

5.4.2 Cellulose Loading is a Major Driving Force of CBP

The UC-PBM demonstrates that the starting cellulose concentration is a major driving force for the metabolism of *C. thermocellum* during the CBP. Comparing the metabolism shown in Figure 5.4 and 5.5, where the starting Avicel concentrations are 10.67 g/L and 4.59 g/L, respectively, the former secretes more than double the amount of cellulase over the same period and achieves complete cellulose solubilization 30% faster than the latter, despite starting with a greater amount of cellulose. The UC-PBM predictions here are in close agreement with the findings of Holwerda et al. [183], where the rate of cellulose solubilization reportedly increases almost linearly with the cellulose loadings.

Moreover, cellulose loading also plays an important but indirect role in dictating the flux distributions in *C. thermocellum*. When the starting cellulose concentration is abundant, the amount of cellulosomes secreted is adequate to sustain hydrolysis until complete cellulose solubilization before famine environment sets, which induces the stationary phase of metabolism (Figure 5.4). Conversely, when the starting cellulose concentration is sufficiently small, famine condition sets in before complete cellulose solubilization as the cellulosome secretion is inadequate to provide a constant supply of cellobiose (Figure 5.5). The UC-PBM predicts that *C. thermocellum* responds by re-inducing $e_{\mathrm{I}}^{\mathrm{rel}}$ to promote rapid secretion of cellulosomes to drive hydrolysis to complete solubilization. This flux re-distribution closely predicts the premature cessation of biomass growth before total cellulose consumption in the experimental data [172]. The flux re-distribution in the late stage of CBP evidently results in a much greater fraction of cellobiose uptake and metabolite secretion through the growth-decoupled cellulosome-secreting pathways.

While the premature cessations of biomass growth prior to total consumption of

cellulose under low solid loadings [184, 185] are often attributed to reduced accessible surface area of cellulose particles at extended stages of hydrolysis [172], the UC-PBM suggests the possibility that growth stops due to flux re-distribution to induce growth-decoupled cellulosome secretion. The reasoning is, in part, supported by the findings of Kothari et al. [117], where the cellulose utilization by *C. thermocellum* is unaffected by cellulose accessibility and often attains complete cellulose solubilization with little to no hindrances. Under high solid loadings, although overflow metabolism is believed to cause the premature halt in biomass growth [183], the possibility of flux re-distributions to growth-decoupled cellulosome-secreting pathways, the induction of famine environments and mass transfer limitations due to high solid loadings should not be overlooked. To this end, the UC-PBM is methodologically ideal for future extension to study the metabolism of *C. thermocellum* under high solid loadings.

5.4.3 Growth-Coupled Pathways Dominate the Metabolism of *C. thermocellum*

Overall, the contribution of growth-decoupled cellulosome-secreting pathways is inferior to growth-coupled metabolism in *C. thermocellum* during CBP (Figure 5.4 and 5.5), which is expected given that the latter also constitutively secretes cellulosomes, albeit at lower yields. If the starting cellulose loading is sufficiently high, the majority of metabolism will be dictated by the growth-coupled pathways (except during the initial lag phase) as the growth-coupled cellulosome secretion is adequate to drive CBP to completion. This is indicative of the capability of *C. thermocellum* to retain its fermentation capacity and cellulolytic ability even if the growth-decoupled cellulosome-secreting pathways are elimi-

nated. However, it remains to be seen if the growth-decoupled cellulosome-secreting pathways are associated with other vital metabolic functionalities, which may be detrimental to the overall metabolism of *C. thermocellum* if removed. Herein lies a major question: If the cellulosome secretion and biomass growth are heavily coupled, how does *C. thermocellum* down-regulate the expression of cellulase independently of biomass growth under nutrient-rich conditions? This is discussed in the following section.

5.4.4 UC-PBM Suggests Possible Cellulosome-Decoupled Growth Metabolic Diversity

To further substantiate the predictive capabilities of the UC-PBM, the model is validated by predicting the metabolism of *C. thermocellum* on pure cellobiose as the sole carbon source by retaining the same parameter values from the previous cases. Figure 5.6 clearly shows that the model prediction closely matches the residual cellobiose and the biomass growth, but the cellulase secretion is vastly overestimated by the $N_{20\%}$ metabolism, which is meant for *C. thermocellum* grown on cellulose. While it is not ideal, switching to $N_{2\%}$ metabolism that caters to *C. thermocellum* grown on cellobiose greatly improves the prediction of the cellulase secretion and no considerable changes are seen in the prediction of residual cellobiose or biomass growth. The improvement in the cellulase secretion is possible due to the drastic change in the cellulosome-biomass yield space when the metabolic networks are switched (Figure 5.7), while the rest of the yield spaces remain relatively unchanged. Here, much lower yields of cellulosomes are possible for the same biomass yields as before because the lower boundary of cellulosome-biomass yield space is closer to the horizontal axis. In spite of that, a slight degree of overestimation persists

in the cellulase secretion (Figure 5.6).

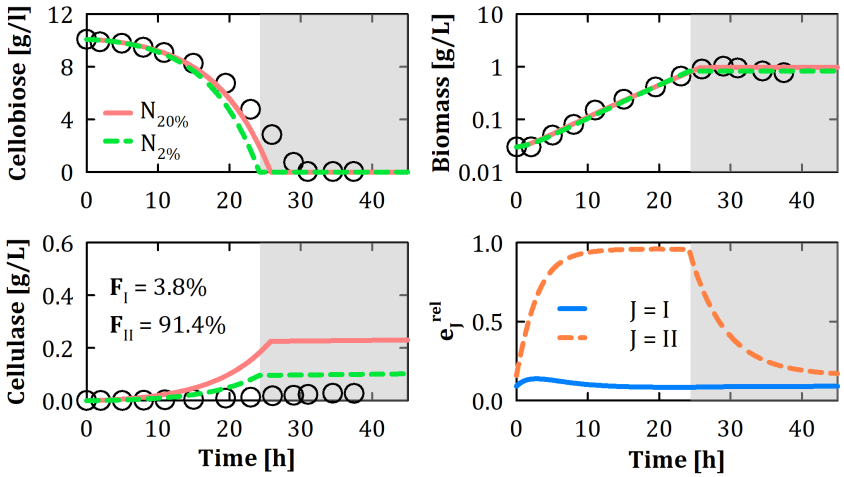

Figure 5.6: Model prediction of metabolism of *C. thermocellum* ATCC 27405 grown on 10 g/L pure cellobiose as the sole carbon source. The lines are model predictions and the markers are experimental data from Zhang and Lynd [163]. Estimated model parameter values from previous cases are used for predictions in this case. The fractions of metabolites secreted through $\mathbf{F}_{\mathrm{I}}$ and $\mathbf{F}_{\mathrm{II}}$ are given in the figure, whereas the total cellobiose uptakes through the respective EFV families are 1.3% and 98.7%. Here, both $N_{20\%}$ and $N_{2\%}$ metabolisms are compared, where the grey-shaded region for famine conditions, the fractions of cellulase secretion and cellobiose uptakes as well as the relative intracellular enzyme levels in the bottom-right panel are provided for the latter.

The growth-decoupled cellulosome-secreting pathways are primarily repressed, but the growth-coupled cellulosome secretion is, nonetheless, unrestrained (Figure 5.6). Therefore, the extent of growth-coupled cellulosome secretion is hypothesized to be much lower in actuality, which indicates the possibility of previously unrecognized cellulosome-decoupled growth metabolic pathways. The cellulosome and biomass are heavily coupled in the metabolic network [155] as the syntheses of both share the same amino acids. If the contribution of protein synthesis to cellulosomes and biomass growth can be distinguished, hypothetical cellulosome-biomass yield space as shown in the inset of the third panel in Figure 5.7 will be obtained, if indeed cellulosome-decoupled growth pathways

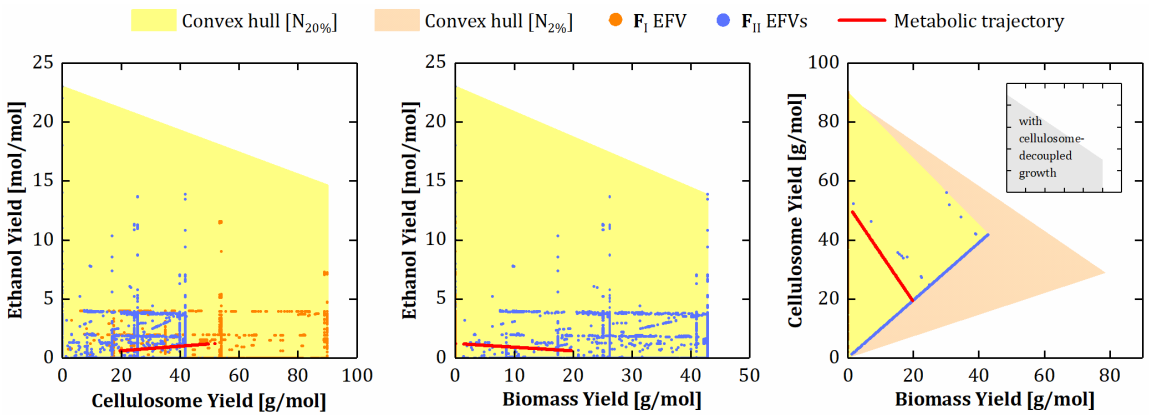

Figure 5.7: Projections of three-dimensional convex hull of ethanol, cellulosome and biomass yields of *C. thermocellum* onto two-dimensional surfaces. The dots represent the yield points of individual EFVs that span the entirety of the convex hull, whereas the red lines denote the metabolic trajectories that the microorganism adopts to respond to perturbations in the environmental conditions as predicted by the UC-PBM. The EFVs are also classified into the individual families, where the vertical axes obscures the $\mathbf{F}_{\mathrm{I}}$ EFVs (with zero biomass yields) in the last two panels. The change in the cellulosome-biomass yield space when $N_{2\%}$ is imposed is shown in the last panel. The inset in the last panel depicts a hypothetical yield space, should a new unrecognized group of cellulosome-decoupled growth pathways be considered. The yields are presented in mole basis of cellobiose.

existed. In the case where the metabolic network of *C. thermocellum* comprises of growth-decoupled cellulosome-secreting pathways, growth-coupled pathways and cellulosome-decoupled growth pathways, it may certainly be possible to model the metabolism under cellulose- and cellobiose-rich conditions without manually switching metabolic networks. Further investigation into this area should not be overlooked, given the prospect of examining the metabolic system of *C. thermocellum* under different substrate conditions with a consolidated metabolic network.

Nevertheless, multiple experimental metabolic data sets [163,172] are predicted using a single set of global parameter values (cf. Table 5.3) even though the metabolism of the microorganism is often affected by several factors including culture methods, properties

and the amount of substrate used. This observation is attributed to the tractability of the UC-PBM in independently accounting for various crucial mechanistic features of the CBP and the robustness of *C. thermocellum* that remains relatively impervious to the properties of the imposed substrates [117] and operating conditions. Lastly, as the UC-PBM predicts the metabolism of *C. thermocellum* across different settings without extensive preconditioning of the metabolic network structure, the identified biological nuances are, to a certain extent, the result of the evolutionary development of regulatory mechanism necessary for the survival of the microorganism.

5.4.5 UC-PBM Guides to Metabolic Engineering

Here, some of the potential uses of the UC-PBM in metabolic engineering are explored. First, metabolic interventions like gene deletions can be targeted to a specific metabolic function based on the EFV classifications performed in the UC-PBM. However, care must be taken to ensure that the deletion of certain EFVs specific to a metabolic function does not impair other associated crucial metabolic functions of the microorganism. Second, in this work, the role of cellulosome-secreting and growth-coupled pathways of *C. thermocellum* is examined by accordingly classifying the EFVs, and similarly, one can modify the classification of the EFVs to study other crucial metabolic functions of microorganisms.

Furthermore, the analysis of yield spaces of essential metabolites is highly useful in aiding alterations to metabolic networks for strain designs. Unlike the conventional representation of the yield spaces using the steady-state constraint-based approaches, the UC-PBM discerns the underlying nature of the EFVs that form the yield spaces. Figure 5.7 demonstrates the yield spaces formed between two essential metabolites of *C. ther-*

mocellum — cellulosome and biomass, as well as ethanol, a most abundantly produced fermentation commodity from renewable biomass [176]. For metabolic interventions that rely on the visual cues of the yield spaces of the native strains, such as the yield optimizations and maximizations, the UC-PBM provides a more focused approach as it discerns the areas of yield spaces populated by EFVs of specific metabolic functions.

While the FBA demonstrates stationary optimal flux state of a microorganism under a prescribed environmental condition, the UC-PBM depicts the possible trajectories that the microorganism may utilize to navigate the metabolism across dynamic changes in the environmental conditions. The metabolic trajectories are represented by the red lines that connect the areas densely populated by EFVs of distinct metabolic functions (Figure 5.7). The metabolic trajectories are essentially the result of the dynamic distribution of fluxes to the two EFV families, and in extension, to the individual EFVs in the families, where the slope of the trajectories is determined by the EFV lumping scheme.

Lastly, detailed analysis of the metabolic network of *C. thermocellum* to optimize the ethanol-cellulosome-biomass yields would be interesting, but it is beyond the scope of this work that serves as a proof of concept of the UC-PBM and to showcase its potential applications. Nevertheless, some interesting observations could be gleaned. In general, the ethanol-associated yield spaces of *C. thermocellum* in Figure 5.7 form Pareto front curves that indicate the possibility of improving ethanol yield without equivalent penalty in cellulosome or biomass yield. Although the convex hulls extend to a large yield range of ethanol, the ethanol yield maximization to these high levels is principally difficult as the majority of the EFVs lie in the lower yield range, and it remains to be seen if it is constrained by other critical metabolites.

5.5 Concluding Remarks

When a cellulolytic microorganism is subjected to CBP of cellulose, an interdependent relationship between the dynamics of the saccharification and the cellular metabolism emerges. Using the proposed UC-PBM framework, which has the capacity to account for the interdependency through a closed-loop control mechanism between the global regulation of extracellular cellulases and local intracellular regulation of fermentation at the network level, this work explores the nature of the flux regulation that gives the cellulolytic microorganism the excellent CBP capabilities. Through a case study on *C. thermocellum*, the UC-PBM reveals that the growth-coupled pathways govern the overall metabolism, while growth-decoupled cellulosome-secreting pathways are only activated in famine conditions. Although cellulose is a non-permeating precursory substrate, the UC-PBM shows that the cellulose loadings strongly influence the flux distribution, where *C. thermocellum* reciprocates to changes in the environment to steer off famine conditions, and knowing this allows one to rationally guide the metabolism by manipulating the famine/feast settings. Moreover, this study reckons the existence of cellulosome-decoupled growth pathways, which not only explains the growth-independent down-regulation of cellulases under nutrient-rich conditions, but also could eliminate the need for separate metabolic networks to account for metabolism of *C. thermocellum* under different carbon sources (cellulose and cellobiose). The realization of modelling and prediction of the metabolic behaviours of a cellulolytic microorganism utilizing cellulose via the UC-PBM concludes the third and the last research objective. The outcome of this research undoubtedly paves the way forward for a more exhaustive analysis of the potentials of cellulose bioprocessing.

Chapter 6

Conclusions and Future Scopes

6.1 Conclusions and Contributions of the Research

This research presents a new theory-based framework, named the Unified Cybernetic-Population Balance Model (UC-PBM), to simulate CBP of cellulose to biochemicals. The framework couples the tenets of the well-known population balances and cybernetic models, to simulate the concurrent enzymatic saccharification of distributed-heterogeneous polymeric cellulose and fermentation of soluble sugars via cellular metabolism, respectively.

To arrive at the ultimate aim of this research, a series of objectives were formulated and fulfilled. First, to tailor solution strategy to solve depolymerization population balances, a set of sectional techniques were implemented over a hybridized discrete-continuous mesh and assessed across various performance indicators. Fixed Pivot Technique (FPT)

was identified as the best pick for subsequent model development for enzymatic saccharification. Second, in efforts to appease to the complexities associated with the enzymatic saccharification of cellulose, a new framework is formed, termed the Multi-Layered Population Balance Model (ML-PBM), which not only delivers various mechanistic insights regarding the process, but also reasons that the heterogeneity in cellulose properties across the structural layers of cellulose particles to cause the rate slowdown. Third, the ML-PBM, coupled with the metabolic network-based Lumped Hybrid Cybernetic Model (L-HCM), delivers the UC-PBM that minutely accounts for the interplay between the saccharification at the macroscopic level and the fermentation via cellular metabolism at the metabolic network-level.

The overall framework features critical process factors such as tracing of complete cellulose chain distribution, a hierarchical scheme to account for non-uniform distributions of cellulose chain lengths and crystallinity across the structural layers of cellulose particles, cellulose particle morphologies, multiple elementary steps in enzyme-substrate interactions, distinct modes of enzymatic scission, classification of elementary metabolic pathways based on *a priori* knowledge of metabolic functions, and regulation of internal cellular resources and metabolic flux distributions to aid distinct metabolic functions according to the changes in the environment subject to the metabolic burden. Overall, the UC-PBM provides a robust framework for prediction of cellular metabolism of cellulolytic microorganisms across diverse conditions.

In the grand scheme of things, this study reveals the necessity for a closed-loop control mechanism across multiple scales of critical process features to represent the dynamics of CBP of cellulose at an adequate resolution to benefit bioprocess and microbial strain designs and optimizations, which is only possible through the coupling of population

balances and cybernetic models as elucidated in this work. Thus, this confirms the thesis of this research.

6.2 Future Scopes

Significant contributions of this research aside, there is much more to explore ahead. Several areas that are worthy of further study are explored here.

Firstly, high cellulose (solids) loadings greatly transform the process dynamics of the enzymatic hydrolysis due to the elevated fluid viscosity, which introduces significant mass transfer limitations. While the ML-PBM in this research alludes to the heterogeneity in cellulose properties as the rate-limiting phenomenon under dilute cellulose concentrations, the same may not be true at high solid loadings. The implication of the mass transfer limitations requires the account of both spatial and temporal changes in the process dynamics with special consideration of the concomitant effects of mixing and hydrodynamics. Therefore, the extension of the ML-PBM with the inclusion of computational fluid dynamics to investigate the enzymatic hydrolysis at industrially relevant high cellulose concentrations is an absolute necessity. Moreover, the extension of the UC-PBM for high solid loadings will also benefit the study of unresolved metabolic behaviour such as overflow metabolism and premature growth halt in *C. thermocellum*.

Second, although this work does not consider the evolution of particle size distribution and assumes that the cellulose particles are monodispersed at an average size, the unique ML-PBM framework that includes heterogeneity in cellulose properties serves as an important foundation that must not be ignored in future extensions to study the effects

of particle size distribution on the process dynamics and to model particle breakage due to hydrodynamic shear during mixing (i.e., each cellulose particle should possess heterogeneous properties that affects the degree of tendency to undergo physical breakage or enzymatic degradation). Another possible future extension of this work is model-based optimizations of operating conditions (e.g., enzyme/cellulose loadings, enzyme cocktail compositions, enzyme/cellulose dosing strategy) to maximize the productivity of the process. Moreover, the current ML-PBM and UC-PBM is only designed for batch and, at most, fed-batch processes, thus, extensions to accommodate chemostat processes is also a strong possibility.

Lastly, the use of the UC-PBM to predict the dynamic metabolic behaviour of engineered strains using metabolic data of the respective parent strains, without extensive post-conditioning of the framework, remains to be seen. Furthermore, the use of co-cultures to combine the biological functionalities of multiple strains to attain efficient CBP of cellulose is an attractive alternative to metabolic engineering, and so, the use of the UC-PBM to study the synergistic/competitive interactions between multiple strains will also be interesting.

These foregoing aspects, which expand the toolset for more in-depth fundamental investigations of cellulose bioprocessing, should be the focus of future studies. Nevertheless, it is expected that the subsequent model developments would be built on the key concepts proposed in the ML-PBM and the closed-loop system instituted in the UC-PBM.

Bibliography

[1] H.-S. Song, J. A. Morgan, and D. Ramkrishna, "Towards Increasing the Productivity of Lignocellulosic Bioethanol: Rational Strategies Fueled by Modeling," in *Bioethanol* (M. A. P. Lima, ed.), ch. 8, pp. 173–190, InTech, 2 2012.

[2] S. E. Jacobsen and C. E. Wyman, "Cellulose and hemicellulose hydrolysis models for application to current and novel pretreatment processes," *Applied biochemistry and biotechnology*, vol. 84-86, pp. 81–96, 2000.

[3] S. P. S. Chundawat, G. T. Beckham, M. E. Himmel, and B. E. Dale, "Deconstruction of lignocellulosic biomass to fuels and chemicals," *Annual Review of Chemical Biomolecular Engineering*, vol. 2, pp. 121–45, 2011.

[4] D. R. Keshwani and J. J. Cheng, "Switchgrass for bioethanol and other value-added applications: A review," *Bioresource Technology*, vol. 100, no. 4, pp. 1515–1523, 2009.

[5] F. Talebnia, D. Karakashev, and I. Angelidaki, "Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation," *Bioresource Technology*, vol. 101, no. 13, pp. 4744–4753, 2010.

[6] K. Kadam and J. McMillan, "Availability of corn stover as a sustainable feedstock for bioethanol production," *Bioresource Technology*, vol. 88, no. 1, pp. 17–25, 2003.

[7] P. Binod, R. Sindhu, R. R. Singhania, S. Vikram, L. Devi, S. Nagalakshmi, N. Kurien, R. K. Sukumaran, and A. Pandey, "Bioethanol production from rice straw: An overview," *Bioresource Technology*, vol. 101, no. 13, pp. 4767–4774, 2010.

[8] C. Wyman, S. Decker, M. Himmel, J. Brady, C. Skopec, and L. Viikari, "Hydrolysis of Cellulose and Hemicellulose," in *Polysaccharides: Structural Diversity and Functional Versatility* (S. Dumitriu, ed.), ch. 43, pp. 994–1033, Taylor & Francis Group, 2nd ed., 2004.

[9] L. R. Lynd, J. H. Cushman, R. J. Nichols, and C. E. Wyman, "Fuel Ethanol from Cellulosic Biomass," *Science*, vol. 251, pp. 1318–1323, 1991.

[10] J. J. Minty and X. N. Lin, "Engineering Synthetic Microbial Consortia for Consolidated Bioprocessing of Ligonocellulosic Biomass into Valuable Fuels and Chemicals," in *Direct Microbial Conversion of Biomass to Advanced Biofuels*, pp. 365–381, Elsevier B.V., 2015.

[11] L. R. Lynd, W. H. Van Zyl, J. E. McBride, and M. Laser, "Consolidated bioprocessing of cellulosic biomass: An update," *Current Opinion in Biotechnology*, vol. 16, no. 5, pp. 577–583, 2005.

[12] D. G. Olson, J. E. McBride, A. Joe Shaw, and L. R. Lynd, "Recent progress in consolidated bioprocessing," *Current Opinion in Biotechnology*, vol. 23, no. 3, pp. 396–405, 2012.

[13] U. Ábrego, Z. Chen, and C. Wan, "Consolidated Bioprocessing Systems for Cellulosic Biofuel Production," in *Advances in Bioenergy* (Y. Li and X. Ge, eds.), vol. 2, ch. 3, pp. 143–182, Elsevier Inc., 2017.

[14] R. den Haan, E. van Rensburg, S. H. Rose, J. F. Görgens, and W. H. van Zyl, "Progress and challenges in the engineering of non-cellulolytic microorganisms for

consolidated bioprocessing," *Current Opinion in Biotechnology*, vol. 33, pp. 32–38, 2015.

[15] K. R. Choi, W. D. Jang, D. Yang, J. S. Cho, D. Park, and S. Y. Lee, "Systems Metabolic Engineering Strategies: Integrating Systems and Synthetic Biology with Metabolic Engineering," *Trends in Biotechnology*, vol. 37, no. 8, pp. 817–837, 2019.

[16] H.-S. Song, F. DeVilbiss, and D. Ramkrishna, "Modeling metabolic systems: The need for dynamics," *Current Opinion in Chemical Engineering*, vol. 2, no. 4, pp. 373–382, 2013.

[17] D. Ramkrishna and H. S. Song, "Dynamic models of metabolism: Review of the cybernetic approach," *AIChE Journal*, vol. 58, no. 4, pp. 986–997, 2012.

[18] J. I. Kim, H. S. Song, S. R. Sunkara, A. Lali, and D. Ramkrishna, "Exacting predictions by cybernetic model confirmed experimentally: Steady state multiplicity in the chemostat," *Biotechnology Progress*, vol. 28, no. 5, pp. 1160–1166, 2012.

[19] J. D. Young, K. L. Henne, J. A. Morgan, A. E. Konopka, and D. Ramkrishna, "Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control," *Biotechnology and Bioengineering*, vol. 100, no. 3, pp. 542–559, 2008.

[20] T. Jeoh, M. J. Cardona, N. Karuna, A. R. Mudinoor, and J. Nill, "Mechanistic kinetic models of enzymatic cellulose hydrolysis—A review," *Biotechnology and Bioengineering*, vol. 114, no. 7, pp. 1369–1385, 2017.

[21] P. Bansal, M. Hall, M. J. Realff, J. H. Lee, and A. S. Bommarius, "Modeling cellulase kinetics on lignocellulosic substrates," *Biotechnology Advances*, vol. 27, no. 6, pp. 833–848, 2009.

[22] J. Nill, N. Karuna, and T. Jeoh, "The impact of kinetic parameters on cellulose hydrolysis rates," *Process Biochemistry*, vol. 74, pp. 108–117, 2018.

[23] H. Niu, N. Shah, and C. Kontoravdi, "Modelling of amorphous cellulose depolymerisation by cellulases, parametric studies and optimisation," *Biochemical Engineering Journal*, vol. 105, pp. 455–472, 2016.

[24] D. Ramkrishna and M. R. Singh, "Population Balance Modeling: Current Status and Future Prospects," *Annu. Rev. Chem. Biomol. Eng*, vol. 5, pp. 123–46, 2014.

[25] D. Ramkrishna, *Population Balances: Theory and Applications to Particulate Systems in Engineering*. San Diego, California, USA: Academic Press, 2000.

[26] D. Ramkrishna and H.-S. Song, *Cybernetic Modeling for Bioreaction Engineering*. Cambridge University Press, 2019.

[27] D. S. Kompala, D. Ramkrishna, and G. T. Tsao, "Cybernetic modeling of microbial growth on multiple substrates," *Biotechnology and Bioengineering*, vol. 26, no. 11, pp. 1272–1281, 1984.

[28] P. Dhurjati, D. Ramkrishna, M. C. Flickinger, and G. T. Tsao, "A cybernetic view of microbial growth: Modeling of cells as optimal strategists," *Biotechnology and Bioengineering*, vol. 27, no. 1, pp. 1–9, 1985.

[29] D. S. Kompala, D. Ramkrishna, N. B. Jansen, and G. T. Tsao, "Investigation of bacterial growth on mixed substrates: Experimental evaluation of cybernetic models," *Biotechnology and Bioengineering*, vol. 28, no. 7, pp. 1044–1055, 1986.

[30] S. Baloo and D. Ramkrishna, "Metabolic regulation in bacterial continuous cultures: I," *Biotechnology and Bioengineering*, vol. 38, no. 11, pp. 1337–1352, 1991.

[31] S. Baloo and D. Ramkrishna, "Metabolic regulation in bacterial continuous culture: II," *Biotechnology and Bioengineering*, vol. 38, no. 11, pp. 1353–1363, 1991.

[32] R. Ramakrishna, D. Ramakrishna, and A. E. Konopka, "Cybernetic modeling of growth in mixed, substitutable substrate environments: Preferential and simulta-

neous utilization," *Biotechnology and Bioengineering*, vol. 52, no. 1, pp. 141–151, 1996.

[33] J. Young, K. Henne, J. Morgan, A. Konopka, and D. Ramkrishna, "Cybernetic modeling of metabolism: Towards a framework for rational design of recombinant organisms," *Chemical Engineering Science*, vol. 59, no. 22-23, pp. 5041–5049, 2004.

[34] J. I. Kim, J. D. Varner, and D. Ramkrishna, "A hybrid model of anaerobic E. coli GJT001: Combination of elementary flux modes and cybernetic variables," *Biotechnology progress*, vol. 24, no. 5, pp. 993–1006, 2008.

[35] H. S. Song, J. A. Morgan, and D. Ramkrishna, "Systematic development of hybrid cybernetic models: Application to recombinant yeast co-consuming glucose and xylose," *Biotechnology and Bioengineering*, vol. 103, no. 5, pp. 984–1002, 2009.

[36] H.-S. Song and D. Ramkrishna, "Prediction of metabolic function from limited data: Lumped Hybrid Cybernetic Modeling (L-HCM)," *Biotechnology and Bioengineering*, vol. 106, no. 2, pp. 271–284, 2010.

[37] H.-S. Song and D. Ramkrishna, "Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function," *Biotechnology and Bioengineering*, vol. 108, no. 1, pp. 127–140, 2011.

[38] H.-S. Song and D. Ramkrishna, "Prediction of dynamic behavior of mutant strains from limited wild-type data," *Metabolic Engineering*, vol. 14, no. 2, pp. 69–80, 2012.

[39] Y. K. Ho, P. Doshi, H. K. Yeoh, and G. C. Ngoh, "Why Are Two Enzymes Better than One for the Efficient Simultaneous Saccharification and Fermentation (SSF) of Natural Polymers? Hints from Inside and Outside a Yeast," *Industrial and Engineering Chemistry Research*, vol. 54, no. 42, pp. 10228–10244, 2015.

[40] Y. K. Ho, P. Doshi, H. K. Yeoh, and G. C. Ngoh, "Interlinked population balance and cybernetic models for the simultaneous saccharification and fermentation of

natural polymers," *Biotechnology and Bioengineering*, vol. 112, no. 10, pp. 2084–2105, 2015.

[41] R. M. Ziff and E. D. McGrady, "The kinetics of cluster fragmentation and depolymerisation," *Journal of Physics A: Mathematical and General*, vol. 18, no. 15, pp. 3027–3037, 1985.

[42] R. M. Ziff and E. D. McGrady, "Kinetics of polymer degradation," *Macromolecules*, vol. 19, pp. 2513–2519, 10 1986.

[43] A. J. Griggs, J. J. Stickel, and J. J. Lischeske, "A mechanistic model for enzymatic saccharification of cellulose using continuous distribution kinetics I: Depolymerization by EG I and CBH I," *Biotechnology and Bioengineering*, vol. 109, no. 3, pp. 665–675, 2012.

[44] J. Kumar, M. Peglow, G. Warnecke, and S. Heinrich, "An efficient numerical technique for solving population balance equation involving aggregation, breakage, growth and nucleation," *Powder Technology*, vol. 182, no. 1, pp. 81–104, 2008.

[45] M. Vanni, "Approximate Population Balance Equations for Aggregation-Breakage Processes.," *Journal of colloid and interface science*, vol. 221, no. 2, pp. 143–160, 2000.

[46] V. John and C. Suciu, "Direct discretizations of bi-variate population balance systems with finite difference schemes of different order," *Chemical Engineering Science*, vol. 106, pp. 39–52, 2014.

[47] S. Rigopoulos and A. G. Jones, "Finite-element scheme for solution of the dynamic population balance equation," *AIChE Journal*, vol. 49, pp. 1127–1139, 5 2003.

[48] M. Kostoglou and A. J. Karabelas, "An assessment of low-order methods for solving the breakage equation," *Powder Technology*, vol. 127, no. 2, pp. 116–127, 2002.

[49] M. Kostoglou and A. J. Karabelas, "Optimal low order methods of moments for solving the fragmentation equation," *Powder Technology*, vol. 143-144, pp. 280–290, 2004.

[50] E. Ghadirian, J. Abbasian, and H. Arastoopour, "CFD simulation of particle size change during the coal char gasification process using the population balance model with FCMOM," *Powder Technology*, vol. 323, pp. 128–138, 2018.

[51] D. L. Marchisio and R. O. Fox, "Solution of population balance equations using the direct quadrature method of moments," *Journal of Aerosol Science*, vol. 36, no. 1, pp. 43–73, 2005.

[52] M. Kostoglou and A. J. Karabelas, "On sectional techniques for the solution of the breakage equation," *Computers and Chemical Engineering*, vol. 33, no. 1, pp. 112–121, 2009.

[53] J. Saha, J. Kumar, A. Bück, and E. Tsotsas, "Finite volume approximations of breakage population balance equation," *Chemical Engineering Research and Design*, vol. 110, pp. 114–122, 2016.

[54] Y. K. Ho, P. Doshi, H. K. Yeoh, and G. C. Ngoh, "Modeling chain-end scission using the Fixed Pivot technique," *Chemical Engineering Science*, vol. 116, pp. 601–610, 2014.

[55] S. Kumar and D. Ramkrishna, "On the solution of population balance equations by discretization—I. A fixed pivot technique," *Chemical Engineering Science*, vol. 51, no. 8, pp. 1311–1332, 1996.

[56] E. Tsotsas, G. Kaur, and J. Kumar, "An accurate and efficient discrete formulation of aggregation population balance equation," *Kinetic and Related Models*, vol. 9, no. 2, pp. 373–391, 2016.

[57] M. Singh, J. Kumar, A. Bück, and E. Tsotsas, "A volume-consistent discrete formulation of aggregation population balance equations," *Mathematical Methods in the Applied Sciences*, vol. 39, no. 9, pp. 2275–2286, 2016.

[58] Y. K. Ho, P. Doshi, and H. K. Yeoh, "Modelling Simultaneous Chain-End and Random Scissions Using the Fixed Pivot Technique," *The Canadian Journal of Chemical Engineering*, vol. 96, pp. 800–814, 8 2018.

[59] Y. K. Ho, C. Kirse, H. Briesen, M. Singh, C.-H. Chan, and K.-W. Kow, "Towards improved predictions for the enzymatic chain-end scission of natural polymers by population balances: The need for a non-classical rate kernel," *Chemical Engineering Science*, vol. 176, pp. 329–342, 2018.

[60] M. Salami-Kalajahi, M. Najafi, and V. Haddadi-Asl, "Application of Monte Carlo Simulation Method to Polymerization Kinetics over Ziegler–Natta Catalysts," *International Journal of Chemical Kinetics*, vol. 41, no. 1, pp. 45–56, 2009.

[61] H. Zhao, A. Maisels, T. Matsoukas, and C. Zheng, "Analysis of four Monte Carlo methods for the solution of population balances in dispersed systems," *Powder Technology*, vol. 173, no. 1, pp. 38–50, 2007.

[62] C. Kirse and H. Briesen, "Numerical solution of mixed continuous-discrete population balance models for depolymerization of branched polymers," *Computers and Chemical Engineering*, vol. 73, pp. 154–171, 2015.

[63] M. Eibinger, T. Zahel, T. Ganner, H. Plank, and B. Nidetzky, "Cellular automata modeling depicts degradation of cellulosic material by a cellulase system with single-molecule resolution," *Biotechnology for Biofuels*, vol. 9, no. 1, p. 56, 2016.

[64] J. E. J. Staggs, "Discrete bond-weighted random scission of linear polymers," *Polymer*, vol. 47, no. 3, pp. 897–906, 2006.

[65] J. E. J. Staggs, "Modelling end-chain scission and recombination of linear polymers," *Polymer Degradation and Stability*, vol. 85, no. 2, pp. 759–767, 2004.

[66] J. E. J. Staggs, "Modelling random scission of linear polymers," *Polymer Degradation and Stability*, vol. 76, no. 1, pp. 37–44, 2002.

[67] F. Gelbard and J. H. Seinfeld, "The general dynamic equation for aerosols. Theory and application to aerosol formation and growth," *Journal of Colloid And Interface Science*, vol. 68, no. 2, pp. 363–382, 1979.

[68] M. Kostoglou, "Mathematical analysis of polymer degradation with chain-end scission," *Chemical Engineering Science*, vol. 55, no. 13, pp. 2507–2513, 2000.

[69] D. Kumar and G. S. Murthy, *Enzymatic Hydrolysis of Cellulose for Ethanol Production: Fundamentals, Optimal Enzyme Ratio, and Hydrolysis Modeling*. Elsevier B.V., 2016.

[70] T. Jeoh, C. I. Ishizawa, M. F. Davis, M. E. Himmel, W. S. Adney, and D. K. Johnson, "Cellulase digestibility of pretreated biomass is limited by cellulose accessibility," *Biotechnology and Bioengineering*, vol. 98, no. 1, pp. 112–122, 2007.

[71] H. E. Grethlein, "The effect of pore size distribution on the rate of enzymic hydrolysis of cellulosic substrates," *Nature Biotechnology*, vol. 3, no. 2, pp. 155–160, 1985.

[72] J. S. Luterbacher, J. Y. Parlange, and L. P. Walker, "A pore-hindered diffusion and reaction model can help explain the importance of pore size distribution in enzymatic hydrolysis of biomass," *Biotechnology and Bioengineering*, vol. 110, no. 1, pp. 127–136, 2013.

[73] M. Tanaka, M. Ikesaka, R. Matsuno, and A. O. Converse, "Effect of pore size in substrate and diffusion of enzyme on hydrolysis of cellulosic materials with cellulases," *Biotechnology and Bioengineering*, vol. 32, no. 5, pp. 698–706, 1988.

[74] B. Z. Shang, R. Chang, and J. W. Chu, "Systems-level modeling with molecular resolution elucidates the rate-limiting mechanisms of cellulose decomposition by cellobiohydrolases," *Journal of Biological Chemistry*, vol. 288, no. 40, pp. 29081–29089, 2013.

[75] S. A. Maurer, N. W. Brady, N. P. Fajardo, and C. J. Radke, "Surface kinetics for cooperative fungal cellulase digestion of cellulose from quartz crystal microgravimetry," *Journal of Colloid and Interface Science*, vol. 394, no. 1, pp. 498–508, 2013.

[76] D. Gao, S. P. S. Chundawat, A. Sethi, V. Balan, S. Gnanakaran, and B. E. Dale, "Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis," *Proceedings of the National Academy of Sciences*, vol. 110, no. 27, pp. 10922–10927, 2013.

[77] E. Praestgaard, J. Elmerdahl, L. Murphy, S. Nymand, K. C. McFarland, K. Borch, and P. Westh, "A kinetic model for the burst phase of processive cellulases," *FEBS Journal*, vol. 278, no. 9, pp. 1547–1560, 2011.

[78] J. S. Luterbacher, J. M. Moran-Mirabal, E. W. Burkholder, and L. P. Walker, "Modeling enzymatic hydrolysis of lignocellulosic substrates using confocal fluorescence microscopy I: Filter paper cellulose," *Biotechnology and Bioengineering*, vol. 112, no. 1, pp. 21–31, 2015.

[79] S. A. Hosseini and N. Shah, "Modelling enzymatic hydrolysis of cellulose part I: Population balance modelling of hydrolysis by endoglucanase," *Biomass and Bioenergy*, vol. 35, no. 9, pp. 3841–3848, 2011.

[80] S. A. Hosseini and N. Shah, "Enzymatic hydrolysis of cellulose part II: Population balance modelling of hydrolysis by exoglucanase and universal kinetic model," *Biomass and Bioenergy*, vol. 35, no. 9, pp. 3830–3840, 2011.

[81] N. Lebaz, A. Cockx, M. Spérandio, and J. Morchain, "Population balance approach for the modelling of enzymatic hydrolysis of cellulose," *The Canadian Journal of Chemical Engineering*, vol. 93, no. FEBRUARY, pp. 276–284, 2015.

[82] A. J. Griggs, J. J. Stickel, and J. J. Lischeske, "A mechanistic model for enzymatic saccharification of cellulose using continuous distribution kinetics II: Cooperative enzyme action, solution kinetics, and product inhibition," *Biotechnology and Bioengineering*, vol. 109, no. 3, pp. 676–685, 2012.

[83] M. Huron, D. Hudebine, N. Lopes Ferreira, and D. Lachenal, "Mechanistic modeling of enzymatic hydrolysis of cellulose integrating substrate morphology and cocktail composition," *Biotechnology and Bioengineering*, vol. 113, no. 5, pp. 1011–1023, 2016.

[84] S. E. Levine, J. M. Fox, H. W. Blanch, and D. S. Clark, "A mechanistic model of the enzymatic hydrolysis of cellulose," *Biotechnology and Bioengineering*, vol. 107, no. 1, pp. 37–51, 2010.

[85] A. Nag, M. A. Sprague, A. J. Griggs, J. J. Lischeske, J. J. Stickel, A. Mittal, W. Wang, and D. K. Johnson, "Parameter determination and validation for a mechanistic model of the enzymatic saccharification of cellulose-Iβ," *Biotechnology Progress*, vol. 31, no. 5, pp. 1237–1248, 2015.

[86] J. Medve, J. Karlsson, D. Lee, and F. Tjerneld, "Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from Trichoderma reesei: Adsorption, sugar production pattern, and synergism of the enzymes," *Biotechnology and Bioengineering*, vol. 59, no. 5, pp. 621–634, 1998.

[87] R. M. F. Bezerra, A. A. Dias, I. Fraga, and A. N. Pereira, "Cellulose hydrolysis by cellobiohydrolase Cel7A Shows mixed hyperbolic product inhibition," *Applied Biochemistry and Biotechnology*, vol. 165, no. 1, pp. 178–189, 2011.

[88] R. M. F. Bezerra, A. A. Dias, I. Fraga, and A. N. Pereira, "Simultaneous ethanol and cellobiose inhibition of cellulose hydrolysis studied with integrated equations assuming constant or variable substrate concentration," *Applied Biochemistry and Biotechnology*, vol. 134, no. 1, pp. 27–38, 2006.

[89] R. M. F. Bezerra and A. A. Dias, "Discrimination among eight modified michaelis-menten kinetics models of cellulose hydrolysis with a large range of substrate/enzyme ratios: Inhibition by cellobiose," *Applied Biochemistry and Biotechnology*, vol. 112, no. 1, pp. 173–184, 2004.

[90] S. Peri, S. Karra, Y. Y. Lee, and M. N. Karim, "Modeling intrinsic kinetics of enzymatic cellulose hydrolysis," *Biotechnology Progress*, vol. 23, no. 3, pp. 626–637, 2007.

[91] K. L. Kadam, E. C. Rydholm, and J. D. McMillan, "Development and validation of a kinetic model for enzymatic saccharification of lignocellulosic biomass," *Biotechnology Progress*, vol. 20, no. 3, pp. 698–705, 2004.

[92] D. Kumar and G. S. Murthy, "Stochastic molecular model of enzymatic hydrolysis of cellulose for ethanol production," *Biotechnol Biofuels*, vol. 6, no. 1, p. 63, 2013.

[93] N. Lebaz, A. Cockx, M. Spérandio, A. Liné, and J. Morchain, "Application of the Direct Quadrature Method of Moments for the modelling of the enzymatic hydrolysis of cellulose: I. Case of soluble substrate," *Chemical Engineering Science*, vol. 149, pp. 306–321, 2016.

[94] N. Lebaz, A. Cockx, M. Spérandio, A. Liné, and J. Morchain, "Application of the Direct Quadrature Method of Moments for the modelling of the enzymatic hydrolysis of cellulose: II. Case of insoluble substrate," *Chemical Engineering Science*, vol. 149, pp. 322–333, 2016.

[95] P. Tervasmäki, V. Sotaniemi, J. Kangas, S. Taskila, H. Ojamo, and J. Tanskanen, "A discretized model for enzymatic hydrolysis of cellulose in a fed-batch process," *Bioresource Technology*, vol. 227, pp. 112–124, 2017.

[96] C. Liang, C. Gu, J. Raftery, M. N. Karim, and M. Holtzapple, "Development of modified HCH-1 kinetic model for long-term enzymatic cellulose hydrolysis and comparison with literature models," *Biotechnology for Biofuels*, vol. 12, no. 1, pp. 1–13, 2019.

[97] I. Caro, A. Blandino, A. B. Díaz, and C. Marzo, "A kinetic model considering the heterogeneous nature of the enzyme hydrolysis of lignocellulosic materials," *Biofuels, Bioproducts and Biorefining*, vol. 13, no. 4, pp. 1044–1056, 2019.

[98] J. J. Lischeske and J. J. Stickel, "A two-phase substrate model for enzymatic hydrolysis of lignocellulose: application to batch and continuous reactors," *Biotechnology for Biofuels*, vol. 12, 2019.

[99] J. D. Nill and T. Jeoh, "The Role of Evolving Interfacial Substrate Properties on Heterogeneous Cellulose Hydrolysis Kinetics," *ACS Sustainable Chemistry & Engineering*, 2020.

[100] F. Ahamed, H.-S. Song, C. W. Ooi, and Y. K. Ho, "Modelling heterogeneity in cellulose properties predicts the slowdown phenomenon during enzymatic hydrolysis," *Chemical Engineering Science*, vol. 206, pp. 118–133, 5 2019.

[101] L. R. Lynd, P. J. Weimer, W. H. V. Zyl, and I. S. Pretorius, "Microbial celullosa utilizaton: Fundamentals and Biotechnology," *Bioresource Technology*, vol. 66, no. 3, pp. 506–577, 2002.

[102] G. P. Philippidis, D. D. Spindler, and C. E. Wyman, "Mathematical modeling of cellulose conversion to ethanol by the simultaneous saccharification and fermentation

process," *Applied Biochemistry and Biotechnology*, vol. 34-35, no. 1, pp. 543–556, 1992.

[103] P. O. Pettersson, R. Eklund, and G. Zacchi, "Modeling simultaneous saccharification and fermentation of softwood," *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology*, vol. 98-100, pp. 733–746, 2002.

[104] J. Shen and F. A. Agblevor, "Modeling semi-simultaneous saccharification and fermentation of ethanol production from cellulose," *Biomass and Bioenergy*, vol. 34, no. 8, pp. 1098–1107, 2010.

[105] J. M. van Zyl, E. van Rensburg, W. H. van Zyl, T. M. Harms, and L. R. Lynd, "A Kinetic Model for Simultaneous Saccharification and Fermentation of Avicel With Saccharomyces cerevisiae," *Biotechnology and Bioengineering*, vol. 108, no. 4, pp. 924–933, 2011.

[106] K. Sakimoto, M. Kanna, and Y. Matsumura, "Kinetic model of cellulose degradation using simultaneous saccharification and fermentation," *Biomass and Bioenergy*, vol. 99, pp. 116–121, 2017.

[107] J. Shadbahr, F. Khan, and Y. Zhang, "Kinetic modeling and dynamic analysis of simultaneous saccharification and fermentation of cellulose to bioethanol," *Energy Conversion and Management*, vol. 141, pp. 236–243, 2017.

[108] J. D. Orth, I. Thiele, and B. O. Palsson, "What is flux balance analysis?," *Nature Biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.

[109] R. A. Thompson, D. S. Layton, A. M. Guss, D. G. Olson, L. R. Lynd, and C. T. Trinh, "Elucidating central metabolic redox obstacles hindering ethanol production in Clostridium thermocellum," *Metabolic Engineering*, vol. 32, pp. 207–219, 2015.

[110] R. Mahadevan, J. S. Edwards, and F. J. Doyle, "Dynamic Flux Balance Analysis of diauxic growth in Escherichia coli," *Biophysical Journal*, vol. 83, no. 3, pp. 1331–1340, 2002.

[111] A. Provost, G. Bastin, S. N. Agathos, and Y. J. Schneider, "Metabolic design of macroscopic bioreaction models: Application to Chinese hamster ovary cells," *Bioprocess and Biosystems Engineering*, vol. 29, no. 5-6, pp. 349–366, 2006.

[112] C. Baroukh, R. Muñoz-Tamayo, J. P. Steyer, and O. Bernard, "DRUM: A new framework for metabolic modeling under non-balanced growth. Application to the carbon metabolism of unicellular microalgae," *PLoS ONE*, vol. 9, no. 8, 2014.

[113] J. J. Stickel and A. J. Griggs, "Mathematical modeling of chain-end scission using continuous distribution kinetics," *Chemical Engineering Science*, vol. 68, no. 1, pp. 656–659, 2012.

[114] J. E. J. Staggs, "A continuous model for vapourisation of linear polymers by random scission and recombination," *Fire Safety Journal*, vol. 40, no. 7, pp. 610–627, 2005.

[115] B. J. McCoy and G. Madras, "Discrete and continuous models for polymerization and depolymerization," *Chemical Engineering Science*, vol. 56, no. 8, pp. 2831–2836, 2001.

[116] R. H. McBee, "The anaerobic thermophilic cellulolytic bacteria.," *Bacteriological reviews*, vol. 14, no. 1, pp. 51–63, 1950.

[117] N. Kothari, S. Bhagia, M. Zaher, Y. Pu, A. Mittal, C. G. Yoo, M. E. Himmel, A. J. Ragauskas, R. Kumar, and C. E. Wyman, "Cellulose hydrolysis by Clostridium thermocellum is agnostic to substrate structural properties in contrast to fungal cellulases," *Green Chemistry*, vol. 21, no. 10, pp. 2810–2822, 2019.

[118] E. K. Holwerda, D. G. Olson, N. M. Ruppertsberger, D. M. Stevenson, S. J. Murphy, M. I. Maloney, A. A. Lanahan, D. Amador-Noguez, and L. R. Lynd, "Metabolic and

evolutionary responses of Clostridium thermocellum to genetic interventions aimed at improving ethanol production," *Biotechnology for Biofuels*, vol. 13, no. 40, 2020.

[119] F. Ahamed, M. Singh, H.-S. Song, P. Doshi, C. W. Ooi, and Y. K. Ho, "On the use of sectional techniques for the solution of depolymerization population balances: Results on a discrete-continuous mesh," *Advanced Powder Technology*, 2020.

[120] W. F. Breuninger, K. Piyachomkwan, and K. Sriroth, "Tapioca/Cassava Starch: Production and Use," in *Starch*, pp. 541–568, Elsevier Inc., 3rd ed., 2009.

[121] D. J. Pollock and R. F. Kratz, "2. Polymer Molecular Weights," in *Methods in Experimental Physics* (R. Fava, ed.), ch. 2, pp. 13–72, Elsevier Inc., 1980.

[122] M. Singh, D. Ghosh, and J. Kumar, "A comparative study of different discretizations for solving bivariate aggregation population balance equation," *Applied Mathematics and Computation*, vol. 234, pp. 434–451, 2014.

[123] M. Wang, J. M. Smith, and B. J. McCoy, "Continuous kinetics for thermal degradation of polymer in solution," *AIChE J.*, vol. 41, no. 6, pp. 1521–1533, 1995.

[124] G. Madras, J. M. Smith, and B. J. McCoy, "Thermal degradation of poly(α-methylstyrene) in solution," *Polymer Degradation and Stability*, vol. 52, no. 3, pp. 349–358, 1996.

[125] A. A. Martens, N. A. M. Besseling, S. Rueb, E. J. R. Sudhölter, H. P. Spaink, and L. C. P. M. De Smet, "Random scission of polymers: Numerical simulations, and experiments on hyaluronan hydrolosis," *Macromolecules*, vol. 44, no. 8, pp. 2559–2567, 2011.

[126] A. C. O'Sullivan, "Cellulose: the structure slowly unravels," *Cellulose*, vol. 4, no. 3, pp. 173–207, 1997.

[127] T. Eriksson, J. Karlsson, and F. Tjerneld, "A model explaining declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglu-

canase I (Cel7B) of Trichoderma reesei," *Applied Biochemistry and Biotechnology - Part A Enzyme Engineering and Biotechnology*, vol. 101, no. 1, pp. 41–60, 2002.

[128] S. Zhang, D. E. Wolfgang, and D. B. Wilson, "Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis," *Biotechnology and Bioengineering*, vol. 66, no. 1, pp. 35–41, 1999.

[129] A. S. Bommarius, A. Katona, S. E. Cheben, A. S. Patel, A. J. Ragauskas, K. Knudson, and Y. Pu, "Cellulase kinetics as a function of cellulose pretreatment," *Metabolic Engineering*, vol. 10, no. 6, pp. 370–381, 2008.

[130] L. Zhang, L. Fu, H. F. Wang, and B. Yang, "Discovery of cellulose surface layer conformation by nonlinear vibrational spectroscopy," *Scientific Reports*, vol. 7, no. October 2016, pp. 1–7, 2017.

[131] P. T. Larsson, K. Wickholm, and T. Iversen, "A CP/MAS carbon-13 NMR investigation of molecular ordering in celluloses," *Carbohydrate Research*, vol. 302, no. 1-2, pp. 19–25, 1997.

[132] M. P. Levi and D. B. Sellen, "The degree-of-polymerization and polydispersity of cellulose in beechwood, and its radial variation across fibre cell-walls," *Carbohydrate Research*, vol. 5, no. 3, pp. 351–355, 1967.

[133] M. Hall, P. Bansal, J. H. Lee, M. J. Realff, and A. S. Bommarius, "Cellulose crystallinity - A key predictor of the enzymatic hydrolysis rate," *FEBS Journal*, vol. 277, no. 6, pp. 1571–1582, 2010.

[134] M. Samejima, J. Sugiyama, K. Igarashi, and K.-E. L. Eriksson, "Enzymatic hydrolysis of bacterial cellulose," *Carbohydrate Research*, vol. 305, no. 2, pp. 281–288, 1997.

[135] K. Igarashi, A. Koivula, M. Wada, S. Kimura, M. Penttilä, and M. Samejima, "High speed atomic force microscopy visualizes processive movement of Trichoderma

reesei cellobiohydrolase I on crystalline cellulose," *Journal of Biological Chemistry*, vol. 284, no. 52, pp. 36186–36190, 2009.

[136] T. T. Teeri, "Crystalline cellulose degradation: New insight into the function of cellobiohydrolases," *Trends in Biotechnology*, vol. 15, no. 5, pp. 160–167, 1997.

[137] P. Engel, L. Hein, and A. C. Spiess, "Derivatization-free gel permeation chromatography elucidates enzymatic cellulose hydrolysis," *Biotechnology for Biofuels*, vol. 5, no. 1, pp. 1–8, 2012.

[138] T. Rosenau, A. Potthast, and P. Kosma, "Trapping of reactive intermediates to study reaction mechanisms in cellulose chemistry," *Advances in Polymer Science*, vol. 205, pp. 153–197, 2006.

[139] R. Evans, R. H. Wearne, and A. F. A. Wallis, "Molecular weight distribution of cellulose as its tricarbanilate by high performance size exclusion chromatography," *Journal of Applied Polymer Science*, vol. 37, no. 12, pp. 3291–3303, 1989.

[140] R. Cohen, M. R. Suzuki, and K. E. Hammel, "Processive Endoglucanase Active in Crystalline Cellulose Hydrolysis by the Brown Rot Basidiomycete Gloeophyllum trabeum," *Applied and environmental microbiology*, vol. 71, no. 5, pp. 2412–2417, 2005.

[141] N. Cruys-Bagger, H. Tatsumi, G. R. Ren, K. Borch, and P. Westh, "Transient kinetics and rate-limiting steps for the processive cellobiohydrolase Cel7A: Effects of substrate structure and carbohydrate binding domain," *Biochemistry*, 2013.

[142] L. Olsson, H. Jørgensen, K. B. R. Krogh, and C. Roca, "Bioethanol Production from Lignocellulosic Material," in *Polysaccharides: Structural Diversity and Functional Versatility* (S. Dumitriu, ed.), ch. 42, pp. 957–993, Marcel Dekker, 2004.

[143] M. S. Pino, R. M. Rodríguez-Jasso, M. Michelin, and H. A. Ruiz, "Enhancement and modeling of enzymatic hydrolysis on cellulose from agave bagasse hydrothermally

pretreated in a horizontal bioreactor," *Carbohydrate Polymers*, vol. 211, no. January, pp. 349–359, 2019.

[144] M. S. Pino, R. M. Rodríguez-Jasso, M. Michelin, A. C. Flores-Gallegos, R. Morales-Rodriguez, J. A. Teixeira, and H. A. Ruiz, "Bioreactor design for enzymatic hydrolysis of biomass under the biorefinery concept," *Chemical Engineering Journal*, vol. 347, pp. 119–136, 2018.

[145] S. Chakraborty, Aniket, and A. Gaikwad, "Mixing effects in cellulase-mediated hydrolysis of cellulose for bio-ethanol production," *Industrial and Engineering Chemistry Research*, vol. 49, no. 21, pp. 10818–10825, 2010.

[146] A. Gaikwad, "Interactions of mixing and reaction kinetics of depolymerization of cellulose to renewable fuels," *Chemical Engineering Communications*, vol. 205, no. 1, pp. 47–81, 2018.

[147] A. Gaikwad and S. Chakraborty, "Mixing effects on the kinetics of enzymatic hydrolysis of Avicel for batch production of cellulosic ethanol," *Industrial and Engineering Chemistry Research*, vol. 52, no. 11, pp. 3988–3999, 2013.

[148] A. L. Demain, M. Newcomb, and J. H. D. Wu, "Cellulase, Clostridia, and Ethanol," *Microbiology and Molecular Biology Reviews*, vol. 69, no. 1, pp. 124–154, 2005.

[149] M. Suto and F. Tomita, "Induction and catabolite repression mechanisms of cellulase in fungi," *Journal of Bioscience and Bioengineering*, vol. 92, no. 4, pp. 305–311, 2001.

[150] Y. Zhang, B. Tang, and G. Du, "Self-induction system for cellulase production by cellobiose produced from glucose in Rhizopus stolonifer," *Scientific Reports*, vol. 7, no. 1, pp. 1–9, 2017.

[151] B. Raman, C. Pan, G. B. Hurst, M. Rodriguez, C. K. McKeown, P. K. Lankford, N. F. Samatova, and J. R. Mielenz, "Impact of Pretreated Switchgrass and Biomass

Carbohydrates on Clostridium thermocellum ATCC 27405 Cellulosome Composition: A Quantitative Proteomic Analysis," *PLoS ONE*, vol. 4, no. 4, 2009.

[152] K. Hirano, M. Kurosaki, S. Nihei, H. Hasegawa, S. Shinoda, M. Haruki, and N. Hirano, "Enzymatic diversity of the Clostridium thermocellum cellulosome is crucial for the degradation of crystalline cellulose and plant biomass," *Scientific Reports*, vol. 6, no. October, pp. 1–9, 2016.

[153] S. Yoav, Y. Barak, M. Shamshoum, I. Borovok, R. Lamed, B. Dassa, Y. Hadar, E. Morag, and E. A. Bayer, "How does cellulosome composition influence deconstruction of lignocellulosic substrates in Clostridium (Ruminiclostridium) thermocellum DSM 1313?," *Biotechnology for Biofuels*, vol. 10, no. 1, pp. 1–16, 2017.

[154] S. B. Roberts, C. M. Gowen, J. P. Brooks, and S. S. Fong, "Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production," *BMC Systems Biology*, vol. 4, 2010.

[155] R. A. Thompson, S. Dahal, S. Garcia, I. Nookaew, and C. T. Trinh, "Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome," *Biotechnology for Biofuels*, vol. 9, no. 1, pp. 1–20, 2016.

[156] S. Dash, A. Khodayari, J. Zhou, E. K. Holwerda, D. G. Olson, L. R. Lynd, and C. D. Maranas, "Development of a core Clostridium thermocellum kinetic metabolic model consistent with multiple genetic perturbations," *Biotechnology for Biofuels*, vol. 10, no. 1, pp. 1–16, 2017.

[157] Y.-H. P. Zhang and L. R. Lynd, "Kinetics and Relative Importance of Phosphorolytic and Hydrolytic Cleavage of Cellodextrins and Cellobiose in Cell Extracts of Clostridium thermocellum," *Applied and Environmental Microbiology*, vol. 70, no. 3, pp. 1563–1569, 2004.

[158] J. P. Olsen, K. Alasepp, J. Kari, N. Cruys-Bagger, K. Borch, and P. Westh, "Mechanism of product inhibition for cellobiohydrolase Cel7A during hydrolysis of insoluble cellulose," *Biotechnology and Bioengineering*, vol. 113, no. 6, pp. 1178–1186, 2016.

[159] J. L. Linville, M. Rodriguez, J. R. Mielenz, and C. D. Cox, "Kinetic modeling of batch fermentation for Populus hydrolysate tolerant mutant and wild type strains of Clostridium thermocellum," *Bioresource Technology*, vol. 147, pp. 605–613, 2013.

[160] P. Zhang, B. Wang, Q. Xiao, and S. Wu, "A kinetics modeling study on the inhibition of glucose on cellulosome of Clostridium thermocellum," *Bioresource Technology*, vol. 190, pp. 36–43, 2015.

[161] P. Erdrich, R. Steuer, and S. Klamt, "An algorithm for the reduction of genome-scale metabolic network models to meaningful core models," *BMC Systems Biology*, vol. 9, no. 1, pp. 1–12, 2015.

[162] Y. Zhang and L. R. Lynd, "Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: Development of an enzyme-linked immunosorbent assay-based method with application to Clostridium thermocellum batch cultures," *Analytical Chemistry*, vol. 75, no. 2, pp. 219–227, 2003.

[163] Y.-H. P. Zhang and L. R. Lynd, "Regulation of cellulase synthesis in batch and continuous cultures of Clostridium thermocellum," *Journal of Bacteriology*, vol. 187, no. 1, pp. 99–106, 2005.

[164] S. Klamt, G. Regensburger, M. P. Gerstl, C. Jungreuthmayer, S. Schuster, R. Mahadevan, J. Zanghellini, and S. Müller, "From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints," *PLoS Computational Biology*, vol. 13, no. 4, pp. 1–22, 2017.

[165] S. Klamt, S. Müller, G. Regensburger, and J. Zanghellini, "A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering," *Metabolic Engineering*, vol. 47, pp. 153–169, 2018.

[166] M. Terzer and J. Stelling, "Large-scale computation of elementary flux modes with bit pattern trees," *Bioinformatics*, vol. 24, no. 19, pp. 2229–2235, 2008.

[167] A. von Kamp, S. Thiele, O. Hädicke, and S. Klamt, "Use of CellNetAnalyzer in biotechnology and metabolic engineering," *Journal of Biotechnology*, vol. 261, pp. 221–228, 2017.

[168] S. Klamt, J. Saez-Rodriguez, and E. D. Gilles, "Structural and functional analysis of cellular networks with CellNetAnalyzer," *BMC Systems Biology*, vol. 1, no. 2, 2007.

[169] S. Klamt and J. Stelling, "Two approaches for metabolic pathway analysis?," *Trends in biotechnology*, vol. 21, no. 2, pp. 64–69, 2003.

[170] B. G. Guimarães, H. Souchon, B. L. Lytle, J. H. David Wu, and P. M. Alzari, "The crystal structure and catalytic mechanism of cellobiohydrolase celS, the major enzymatic component of the Clostridium thermocellum cellulosome," *Journal of Molecular Biology*, vol. 320, no. 3, pp. 587–596, 2002.

[171] J. D. Young and D. Ramkrishna, "On the matching and proportional laws of cybernetic models," *Biotechnology Progress*, vol. 23, no. 1, pp. 83–99, 2007.

[172] E. K. Holwerda, L. D. Ellis, and L. R. Lynd, "Development and evaluation of methods to infer biosynthesis and substrate consumption in cultures of cellulolytic microorganisms," *Biotechnology and Bioengineering*, vol. 110, no. 9, pp. 2380–2388, 2013.

[173] R. Brun, P. Reichert, and H. R. Künsch, "Practical identifiability analysis of large environmental simulation models," *Water Resources Research*, vol. 37, no. 4, pp. 1015–1030, 2001.

[174] D. A. L. Hogsett, H.-J. Ahn, T. D. Bernardez, C. R. South, and L. R. Lynd, "Direct Microbial Conversion," *Applied Biochemistry and Biotechnology*, vol. 34/35, pp. 527–541, 1992.

[175] D. A. L. Hogsett, *Cellulose Hydrolysis and Fermentation by Clostridium thermocellum for the Production of Ethanol*. PhD thesis, Dartmouth College, 1995.

[176] S. Y. Lee, H. U. Kim, T. U. Chae, J. S. Cho, J. W. Kim, J. H. Shin, D. I. Kim, Y. S. Ko, W. D. Jang, and Y. S. Jang, "A comprehensive metabolic map for production of bio-based chemicals," *Nature Catalysis*, vol. 2, no. 1, pp. 18–33, 2019.

[177] L. Artzi, E. A. Bayer, and S. Moraïs, "Cellulosomes: Bacterial nanomachines for dismantling plant polysaccharides," *Nature Reviews Microbiology*, vol. 15, no. 2, pp. 83–95, 2017.

[178] T. W. Dror, E. Morag, A. Rolider, E. A. Bayer, R. Lamed, and Y. Shoham, "Regulation of the cellulosomal celS (cel48A) gene of Clostridium thermocellum is growth rate dependent," *Journal of Bacteriology*, vol. 185, no. 10, pp. 3042–3048, 2003.

[179] T. W. Dror, A. Rolider, E. A. Bayer, R. Lamed, and Y. Shoham, "Regulation of major cellulosomal endoglucanases of Clostridium thermocellum differs from that of a prominent cellulosomal xylanase," *Journal of Bacteriology*, vol. 187, no. 7, pp. 2261–2266, 2005.

[180] R. Lamed, R. Kenig, E. Morgenstern, J. F. Calzada, F. De Micheo, and E. A. Bayer, "Efficient cellulose solubilization by a combined cellulosome-β-glucosidase system," *Applied Biochemistry and Biotechnology*, vol. 27, no. 2, pp. 173–183, 1991.

[181] R. Lamed, R. Kenig, E. Setter, and E. A. Bayer, "Major characteristics of the cellulolytic system of Clostridium thermocellum coincide with those of the purified cellulosome," *Enzyme and Microbial Technology*, vol. 7, no. 1, pp. 37–41, 1985.

[182] G. Gefen, M. Anbar, E. Morag, R. Lamed, and E. A. Bayer, "Enhanced cellulose degradation by targeted integration of a cohesin-fused β-glucosidase into the Clostridium thermocellum cellulosome," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, no. 26, pp. 10298–10303, 2012.

[183] E. K. Holwerda, P. G. Thorne, D. G. Olson, D. Amador-Noguez, N. L. Engle, T. J. Tschaplinski, J. P. Van Dijken, and L. R. Lynd, "The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading," *Biotechnology for Biofuels*, vol. 7, no. 1, pp. 1–11, 2014.

[184] D. I. C. Wang, G. C. Avgerinos, I. Biocic, S.-D. Wang, and H.-Y. Fang, "Ethanol from cellulosic biomass," *Philosophical Transactions of the Royal Society of London B*, vol. 300, no. 1100, pp. 85–95, 1983.

[185] E. K. Holwerda, K. D. Hirst, and L. R. Lynd, "A defined growth medium with very low background carbon for culturing Clostridium thermocellum," *Journal of Industrial Microbiology and Biotechnology*, vol. 39, no. 6, pp. 943–947, 2012.

[186] A. Tanaka, C. Nakagawa, K. Kodaira, K. Senoo, and H. Obata, "Basic subsite theory assumptions may not be applicable to hydrolysis of cellooligosaccharides by almond β-glucosidase," *Journal of Bioscience and Bioengineering*, vol. 88, no. 6, pp. 664–666, 1999.

[187] G. Sin, A. S. Meyer, and K. V. Gernaey, "Assessing reliability of cellulose hydrolysis models to support biofuel process design-Identifiability and uncertainty analysis," *Computers and Chemical Engineering*, vol. 34, no. 9, pp. 1385–1392, 2010.

[188] S. G. Desai and A. O. Converse, "Substrate reactivity as a function of the extent of reaction in the enzymatic hydrolysis of lignocellulose," *Biotechnology and Bioengineering*, vol. 56, no. 6, pp. 650–655, 1997.

[189] Y. H. P. Zhang and L. R. Lynd, "Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems," *Biotechnology and Bioengineering*, vol. 88, no. 7, pp. 797–824, 2004.

[190] T. K. Ng and J. G. Zeikus, "Comparison of extracellular cellulase activities of Clostridium thermocellum LQRI and Trichoderma reesei QM9414," *Applied and Environmental Microbiology*, vol. 42, no. 2, pp. 231–240, 1981.

Appendix A

Solution of Depolymerization Population Balances

A.1 Fully Discrete (Exact) ODEs for Depolymerization Population Balances

The final set of ODEs for the fully discrete solutions of bond-breakage process of depolymerization is given as follows.

Chain-End Monomer Scission:

Following Kostoglou [68], the fully discrete equations for chain-end scission for the case of depolymerization are:

$$\frac{dc_1}{dt} = 2k_2^{\gamma} c_2 + \sum_{j=3}^{N} k_j^{\gamma} c_j \tag{A.1}$$

$$\frac{dc_i}{dt} = k_{i+1}^{\gamma} c_{i+1} - k_i^{\gamma} c_i; \quad i = 2, 3, \ldots, N-1 \tag{A.2}$$

$$\frac{dc_N}{dt} = -k_N^{\gamma} c_N \tag{A.3}$$

Random Scission:

Based on the studies by Ziff and McGrady [41] as well as by Kumar and Ramkrishna [55], the fully discrete equations for random scission for the case of depolymerization are:

$$\frac{dc_1}{dt} = \sum_{j=2}^{N} \left(\frac{2}{j-1} \right) k_j^{\alpha} c_j \tag{A.4}$$

$$\frac{dc_i}{dt} = \sum_{j=i+1}^{N} \left(\frac{2}{j-1} \right) k_j^{\alpha} c_j - k_i^{\alpha} c_i; \quad i = 2, 3, \ldots, N-1 \tag{A.5}$$

$$\frac{dc_N}{dt} = -k_N^{\alpha} c_N \tag{A.6}$$

Here, $c_i = C_i$ for the fully discrete solution.

A.2 FPT ODEs for Depolymerization Population Balances

The final set of ODEs for the FPT solutions of depolymerization population balances over a discrete-continuous mesh is as follows.

Chain-End Monomer Scission:

$$\frac{dC_1}{dt} = 2k_2^{\gamma}C_2 + \sum_{j=3}^{p+q} k_j^{\gamma}C_j \tag{A.7}$$

$$\frac{dC_i}{dt} = k_{i+1}^{\gamma}C_{i+1} - k_i^{\gamma}C_i; \quad i = 2, 3, \ldots, p \tag{A.8}$$

$$\frac{dC_{p+1}}{dt} = \left(\frac{v_m}{x_{p+2} - x_{p+1}}\right) k_{p+2}^{\gamma}C_{p+2} - k_{p+1}^{\gamma}C_{p+1} \tag{A.9}$$

$$\begin{aligned}\frac{dC_i}{dt} &= \left(1 - \frac{v_m}{x_i - x_{i-1}}\right) k_i^{\gamma}C_i + \left(\frac{v_m}{x_{i+1} - x_i}\right) k_{i+1}^{\gamma}C_{i+1} - k_i^{\gamma}C_i; \\ i &= p+2, p+3, \ldots, p+q-1\end{aligned} \tag{A.10}$$

$$\frac{dC_{p+q}}{dt} = \left(1 - \frac{v_m}{x_{p+q} - x_{p+q-1}}\right) k_{p+q}^{\gamma}C_{p+q} - k_{p+q}^{\gamma}C_{p+q} \tag{A.11}$$

The alternative meshing strategy is used here to alleviate the discrepancy at the discrete-continuous boundary (see Section 3.3.4).

Random Scission:

$$\frac{dC_1}{dt} = \sum_{j=2}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_j^{\alpha}C_j \tag{A.12}$$

$$\frac{dC_i}{dt} = \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_j^{\alpha}C_j - k_i^{\alpha}C_i; \quad i = 2, 3, \ldots, p \tag{A.13}$$

$$\frac{dC_i}{dt} = \sum_{j=i+1}^{p+q} \left(\frac{x_{i+1} - x_{i-1}}{x_{j-1}}\right) k_j^{\alpha}C_j - k_i^{\alpha}C_i; \quad i = p+1, p+2, \ldots, p+q-1 \tag{A.14}$$

$$\frac{dC_{p+q}}{dt} = -k_{p+q}^{\alpha}C_{p+q} \tag{A.15}$$

A.3 CAT ODEs for Depolymerization Population Balances

The final set of ODEs for the CAT solutions of depolymerization population balances over a discrete-continuous mesh is given as follows.

Chain-End Monomer Scission:

Final ODEs of the CAT are identical to the FPT for chain-end scission. Refer to Appendix A.2.

Random Scission:

$$\frac{dC_1}{dt} = \sum_{j=2}^{p+q} \left(\frac{2}{x_{j-1}} \right) k_j^{\alpha} C_j \tag{A.16}$$

$$\begin{aligned}
\frac{dC_i}{dt} = & \sum_{j=i-1}^{p+q} \left[\frac{2(x_{i+1/2} - x_{i-1/2})}{x_{j-1}} \right] \left(\frac{\bar{v}_{i-1} - x_{i-1}}{x_i - x_{i-1}} \right) H(\bar{v}_{i-1} - x_{i-1}) k_j^{\alpha} C_j \\
& + \sum_{j=i}^{p+q} \left[\frac{2(x_{i+1/2} - x_{i-1/2})}{x_{j-1}} \right] \left[\frac{\bar{v}_i - x_{i-1}}{x_i - x_{i-1}} H(x_i - \bar{v}_i) + \frac{\bar{v}_i - x_{i+1}}{x_i - x_{i+1}} H(\bar{v}_i - x_i) \right] k_j^{\alpha} C_j \\
& + \sum_{j=i+1}^{p+q} \left[\frac{2(x_{i+1/2} - x_{i-1/2})}{x_{j-1}} \right] \left(\frac{\bar{v}_{i+1} - x_{i+1}}{x_i - x_{i+1}} \right) H(x_{i+1} - \bar{v}_{i+1}) k_j^{\alpha} C_j \\
& - k_i^{\alpha} C_i; \quad i = 2, 3, \ldots, p+q-1
\end{aligned} \tag{A.17}$$

$$\frac{dC_{p+q}}{dt} = -k_{p+q}^{\alpha} C_{p+q} \tag{A.18}$$

For Eq. (A.17), the average particle size is $\bar{v}_i = (x_{i+1/2} + x_{i-1/2})/2$.

A.4 Failure of the FVS for Chain-End Scission

To direct future efforts in refining the FVS for solving depolymerization PBEs, the failure of the FVS for chain-end scission is clarified here. For breakage, the final discretized form of the FVS is given as:

$$\frac{dC_i}{dt} = \sum_{j=i}^{I} \omega_j^{\mathrm{b}} p_{ij}^{(0)} k_j^{\gamma} C_j - \left(\frac{\omega_i^{\mathrm{b}}}{x_i} \sum_{j=1}^{i} x_j p_{ji}^{(0)} \right) k_i^{\gamma} C_i \tag{A.19}$$

where $p_{ij}^{(0)} = \int_{x_{i-1/2}}^{m_j^i} \delta\left(v - [x_j - v_m]\right) dv$ for chain-end monomer scission. As the FVS does not distribute the birth of daughter polymers to adjoining pivot points like the FPT and the CAT, the FVS simply does not have the resolution required to distinguish between the size of the resulting daughter polymers $(x_j - 1)$ from the parent polymers (x_j), especially when the size of parent polymers are large. This is clearly illustrated in Figure A.1(a).

In the discrete region, where all the grid points are spaced one monomer unit apart, $p_{ii}^{(0)} = 0,\ p_{i,i+1}^{(0)} = 1,$ and $p_{ij}^{(0)} = 0,\ \forall\ j \mid j > i+1$. This holds true in general when a non-uniform mesh is employed for the continuous region, where $p_{i,i+1}^{(0)} = 1$ (Region 1 in Figure A.1(a)). However, as the width of each section increases with the DP, the removal of monomers from the parent polymers will result in daughter polymers with a size that falls within the same section as the parent polymers, i.e., $p_{ii}^{(0)} = 1,\ p_{i,i+1}^{(0)} = 0$ and $p_{ij}^{(0)} = 0, \forall\ j \mid j > i+1$ (Region 2 in Figure A.1(a)). The transition from the former to the latter results in zero-birth contribution to a single section that falls in the intermediary, i.e., $p_{ij}^{(0)} = 0, \forall\ j$ (shaded in Figure A.1(a)). This occurs when the inequality $x_{i+1} - v_m > x_{i+1/2}$ is fulfilled at the first instance. In addition, with a geometric mesh, one can establish that $x_{i+1} = r x_i,\ x_{i+1/2} = (x_i + x_{i+1})/2$ and $x_i = x_{p+1} r^{i-(p+1)}$. Substi-

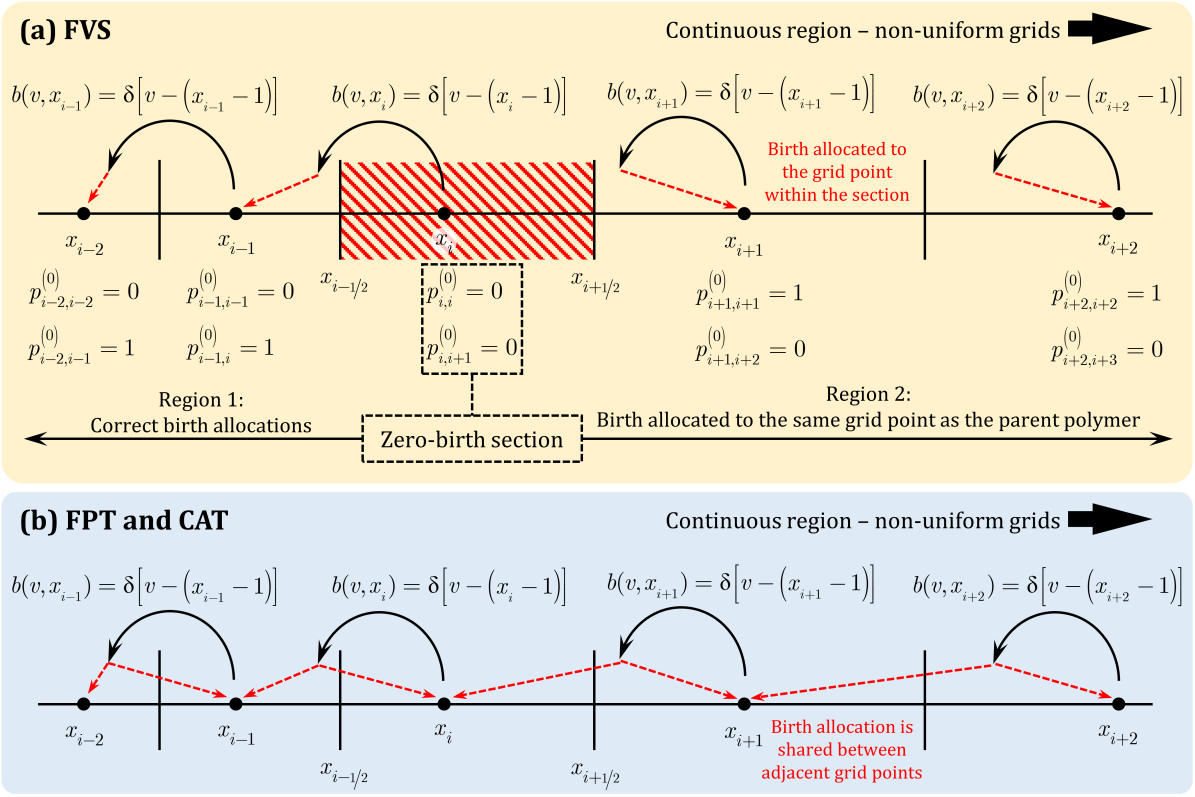

Figure A.1: Illustration of (a) failure in FVS formulation in polymer birth allocation during chain-end scission, and (b) alternative polymer birth allocations based on the FPT and the CAT formulations for chain-end scission.

tuting these into the inequality, the i-th section that has zero birth contribution can be found as:

$$i = \left\lceil \left\{ \left[\ln\left(\frac{2v_m}{r-1}\right) - \ln\left(p+1\right) \right] \Big/ \ln\left(r\right) \right\} + p + 1 \right\rceil ; \quad r \leq 1.02 \tag{A.20}$$

where the ceiling operator $\lceil Z \rceil$ denotes the nearest higher integer to Z. For $r > 1.02$, this "loss of information" occurs at the first section in the continuous region. For the polymer case examined in this work with a mesh configuration of $[p, q] = [100, 500]$ and $r = 1.0109$, this loss of information occurs at Δx_{157}. Theoretically, if the bulk of the initial polymer

mass lies at regions below this zero-birth interval, this loss of information can be minimized for the case of pure depolymerization. For an arbitrary initial polymer distribution, theoretically this will require the i-th section that experiences the loss of information to fulfil $i \to \infty$ in order for the bulk of the initial polymer mass to be adequately contained. With the current FVS formulation, this is impractical as the condition $i \to \infty$ requires r to approach unity, i.e., $\lim_{r \to 1} i = \infty$. In this case, the FVS has no computational advantage over the fully discrete solution.

A.5 CAT and FPT Solutions are Identical for Chain-End Scission

This section shows that the CAT and the FPT yield an identical set of final equations for chain-end scission. Recall that the polymer allocation function for the CAT is expressed as:

$$\eta_{ij}^{\mathrm{CAT}} = \begin{cases} p_{i-1,j}^{(0)} \left(\dfrac{\bar{v}_{i-1} - x_{i-1}}{x_i - x_{i-1}} \right) H(\bar{v}_{i-1} - x_{i-1}); & \text{for } a = i-1 \\ p_{ij}^{(0)} \left[\left(\dfrac{\bar{v}_i - x_{i-1}}{x_i - x_{i-1}} \right) H(x_i - \bar{v}_i) + \left(\dfrac{\bar{v}_i - x_{i+1}}{x_i - x_{i+1}} \right) H(\bar{v}_i - x_i) \right]; & \text{for } a = i \\ p_{i+1,j}^{(0)} \left(\dfrac{\bar{v}_{i+1} - x_{i+1}}{x_i - x_{i+1}} \right) H(x_{i+1} - \bar{v}_{i+1}); & \text{for } a = i+1 \end{cases} \tag{A.21}$$

In addition, the average polymer size that enters the i-th section is computed as:

$$\bar{v}_i = \frac{M_i}{V_i} = \sum_{j=i}^{I} \frac{p_{ij}^{(1)}}{p_{ij}^{(0)}} \tag{A.22}$$

where $p_{ij}^{(s)} = \int_{x_{i-1/2}}^{m_j^i} v^s \delta\left(v - [x_j - v_m]\right) dv$ for chain-end scission. In the discrete region, where all the grid points are spaced one monomer unit apart, $p_{ii}^{(0)} = 0$, $p_{i,i+1}^{(0)} = 1$ and $p_{ij}^{(0)} = 0, \forall\, j \mid j > i{+}1$, whereas $p_{ii}^{(1)} = 0$, $p_{i,i+1}^{(1)} = x_i$ and $p_{ij}^{(1)} = 0, \forall\, j \mid j > i{+}1$. Therefore, this renders the average polymer size, $\bar{v}_i = x_i$. Similarly, $\bar{v}_{i-1} = x_{i-1}$ and $\bar{v}_{i+1} = x_{i+1}$. Imposing these average polymer sizes to Eq. (A.21), the polymer allocation function reduces to $\eta_{ij}^{\mathrm{CAT}} = p_{ij}^{(0)}$ for $a = i$. It is known that $p_{i,i+1}^{(0)} = 1$ and is zero for every other values of j and consequently, the final ODE for the discrete region becomes:

$$\frac{dC_i}{dt} = k_{i+1}^{\gamma} C_{i+1} - k_i^{\gamma} C_i; \quad i = 2, 3, \ldots, p \tag{A.23}$$

Similarly, for the continuous region with a logical constraint that the grid points are spaced more than v_m units apart, the average polymer sizes can be computed following the approach shown above as $\bar{v}_{i-1} = x_i - v_m$, $\bar{v}_i = x_{i+1} - v_m$ and $\bar{v}_{i+1} = x_{i+2} - v_m$. Following this, the polymer allocation function assumes the following form:

$$\eta_{ij}^{\mathrm{CAT}} = \begin{cases} p_{i-1,j}^{(0)} \left(1 - \dfrac{v_m}{x_i - x_{i-1}}\right); & \text{for } a = i - 1 \\ p_{ij}^{(0)} \left(\dfrac{v_m}{x_{i+1} - x_i}\right); & \text{for } a = i \end{cases} \tag{A.24}$$

Here, the $p_{i-1,i}^{(0)} = 1$, $p_{i,i+1}^{(0)} = 1$ and are zero for every other values of j. Eventually, the final ODE for the continuous region is given by the following expression:

$$\begin{aligned} \frac{dC_i}{dt} &= \left(1 - \frac{v_m}{x_i - x_{i-1}}\right) k_i^{\gamma} C_i + \left(\frac{v_m}{x_{i+1} - x_i}\right) k_{i+1}^{\gamma} C_{i+1} - k_i^{\gamma} C_i; \\ & i = p + 2, p + 3, \ldots, p + q - 1 \end{aligned} \tag{A.25}$$

For the last section of the discretized DP domain, the final ODE in Eq. (A.25) reduces

to:

$$\frac{dC_{p+q}}{dt} = \left(1 - \frac{v_m}{x_{p+q} - x_{p+q-1}}\right) k_{p+q}^{\gamma} C_{p+q} - k_{p+q}^{\gamma} C_{p+q} \tag{A.26}$$

The final set of ODEs for the CAT for chain-end scission presented above in Eqs. (A.23), (A.25), and (A.26) are identical to that of the FPT as presented in Appendix A.2. The monomer equation is solved separately and is identical across all techniques, as given in Eq. (3.3).

A.6 FVS and FPT Solutions are Identical for Random Scission

As alluded to in Section 3.3.6, the FVS and the FPT yield identical set of equations for random scission using a modified stoichiometric kernel that is a continuous analogue of the discrete bond-breaking process, i.e., $b_2^{\alpha}(v, x_j) = 2/x_{j-1}$. Here, the proof is presented. For the FVS, in conjunction with the new kernel, the intra-interval terms (i.e., $i = j$) are eliminated and thus, the final form of the discretized FVS reduces to:

$$\frac{dC_i}{dt} = \sum_{j=i+1}^{I} \omega_j^{\mathrm{b}} p_{ij}^{(0)} k_j^{\alpha} C_j - \varphi_i^{\mathrm{FVS}} k_i^{\alpha} C_i; \quad i = 2, 3, \ldots, I \tag{A.27}$$

where the birth weight allocation is:

$$\omega_j^{\mathrm{b}} = \frac{x_j \left(\sum_{i=1}^{j-1} p_{ij}^{(0)} - 1\right)}{\sum_{i=1}^{j-1} (x_j - x_i) p_{ij}^{(0)}} \tag{A.28}$$

Here, $p_{ij}^{(0)} = \int_{x_{i-1/2}}^{x_{i+1/2}} 2/x_{j-1}\, dv$. Further, it can be showed that:

$$\begin{aligned}
\sum_{i=1}^{j-1} p_{ij}^{(0)} &= \sum_{i=1}^{j-1} \int_{x_{i-1/2}}^{x_{i+1/2}} 2/x_{j-1}\, dv \\
&= \int_{x_{1-1/2}}^{x_{(j-1)+1/2}} 2/x_{j-1}\, dv \\
&= 2\left(\frac{x_{(j-1)+1/2} - x_{1-1/2}}{x_{j-1}}\right) \\
&\approx 2
\end{aligned} \tag{A.29}$$

In arriving at the final form of Eq. (A.29), for the discrete region, $x_{(j-1)+1/2} - x_{1-1/2} = x_{j-1}$, whereas for the continuous region, the terms $x_{(j-1)+1/2}/x_{j-1} \approx 1$ and $x_{1-1/2}/x_{j-1} \approx 0$. By the same reasoning, it is possible to write the following:

$$\begin{aligned}
\sum_{i=1}^{j-1} x_i p_{ij}^{(0)} &= \sum_{i=1}^{j-1} x_i \int_{x_{i-1/2}}^{x_{i+1/2}} 2/x_{j-1}\, dv \\
&= \int_{x_{1-1/2}}^{x_{(j-1)+1/2}} 2v/x_{j-1}\, dv \\
&= x_{(j-1)+1/2}\left(\frac{x_{(j-1)+1/2}}{x_{j-1}}\right) - x_{1-1/2}\left(\frac{x_{1-1/2}}{x_{j-1}}\right) \\
&\approx x_{(j-1)+1/2}
\end{aligned} \tag{A.30}$$

Substituting the results from Eqs. (A.29) and (A.30) into Eq. (A.28), it can be shown that for any values of j, the following holds:

$$\begin{aligned}\omega_j^{\mathrm{b}} &= \frac{x_j}{2x_j - x_{(j-1)+1/2}} \\ &= \frac{1}{2 - x_{(j-1)+1/2}/x_j} \approx 1; \quad \because (x_{(j-1)+1/2}/x_j \approx 1)\end{aligned} \tag{A.31}$$

Similarly, the death weight allocation also collapses to 1 for any values of i:

$$\begin{aligned}\varphi_i^{\mathrm{FVS}} &= \frac{\omega_i^{\mathrm{b}}}{x_i} \sum_{j=1}^{i} x_j p_{ji}^{(0)} \\ &\approx \frac{1}{x_i}(x_{i+1/2}) \approx 1\end{aligned} \tag{A.32}$$

From the above, it follows that the final form of the discretized FVS is:

$$\frac{dC_i}{dt} = \sum_{j=i+1}^{I} p_{ij}^{(0)} k_j^{\alpha} C_j - k_i^{\alpha} C_i; \quad i = 2, 3, \ldots, I \tag{A.33}$$

Written for the discrete and the continuous region and invoking $p_{ij}^{(0)} = \int_{x_{i-1/2}}^{x_{i+1/2}} 2/x_{j-1}\, dv = 2(x_{i+1/2} - x_{i-1/2})/x_{j-1}$, the final equations (excluding the monomer case) become:

$$\frac{dC_i}{dt} = \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_j^{\alpha} C_j - k_i^{\alpha} C_i; \quad i = 2, 3, \ldots, p \tag{A.34}$$

$$\begin{aligned}\frac{dC_i}{dt} &= \sum_{j=i+1}^{p+q} \frac{2}{x_{j-1}} \left[x_{i+1/2} - x_{i-1/2}\right] k_j^{\alpha} C_j - k_i^{\alpha} C_i \\ &= \sum_{j=i+1}^{p+q} \frac{2}{x_{j-1}} \left[\frac{x_i + x_{i+1}}{2} - \frac{x_{i-1} + x_i}{2}\right] k_j^{\alpha} C_j - k_i^{\alpha} C_i \\ &= \sum_{j=i+1}^{p+q} \left[\frac{x_{i+1} + x_{i-1}}{x_{j-1}}\right] k_j^{\alpha} C_j - k_i^{\alpha} C_i; \\ i &= p+1, p+2, \ldots, p+q-1\end{aligned} \tag{A.35}$$

$$\frac{dC_{p+q}}{dt} = -k^{\alpha}_{p+q} C_{p+q} \tag{A.36}$$

The final set of ODEs presented above is identical to the FPT as given in Appendix A.2. The monomer equation is solved separately and is identical across all techniques, as given in Eq. (3.4).

Appendix B

Formulation of Multi-Layered Population Balance Model

B.1 Formulation of Multi-Layered Distributed Cellulose Hydrolysis

As alluded to in Section 4.2.1, the particle size decreases as the cellulose polymers on the surface are solubilized by the enzymes. At the same time, the cellulose polymers in the adjacent layer are exposed to enzymes for hydrolysis. The rate of exposure of underlying polymers and the rate of loss of polymers on the surface are interrelated through the surface area across which the mass transfer occurs, i.e.,

$$\frac{r^{(1)}_{\text{exp},i}(t)}{r^{(1)}_{\text{loss},i}(t)} = \frac{A_2(t)}{A_1(t)} \tag{B.1}$$

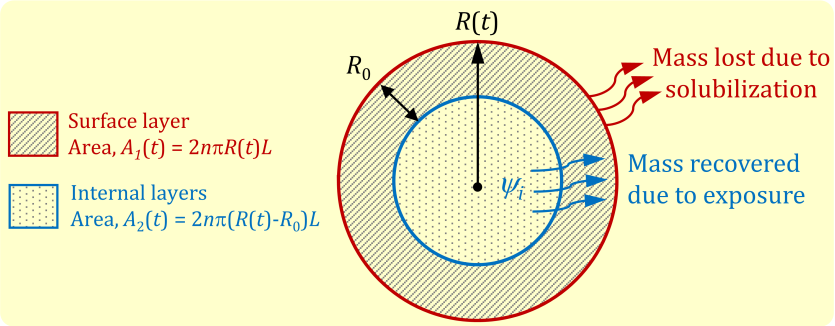

Figure B.1: Illustration of multi-layered cellulose ultrastructure undergoing progressive hydrolysis.

where $r^{(1)}_{\text{exp},i}(t)$ is the total mass rate of exposure of underlying polymers and $r^{(1)}_{\text{loss},i}(t)$ is the total mass rate of loss of polymers on the surface, whereas $A_1(t)$ and $A_2(t)$ are the surface areas as represented in Figure B.1. Given the above, the total mass rate of exposure of underlying polymers is:

$$r^{(1)}_{\text{exp},i}(t) = -\left[\frac{R(t) - R_0}{R(t)}\right] r^{(1)}_{\text{loss},i}(t) \tag{B.2}$$

In the above, the negative sign is added to reverse the sign of $r^{(1)}_{\text{loss},i}(t)$ in order to represent the mass leaching into the surface layer from the underlying layer. Subsequently, Eq. (B.2) can be re-written to express the molar rate of exposure of individual polymer species with DP i as follows:

$$r_{\text{exp},i}(t) = -\left[\frac{R(t) - R_0}{R(t)}\right] r^{(1)}_{\text{loss},i}(t)\psi_i \tag{B.3}$$

where ψ_i is the ratio of initial molar concentration of polymer species with DP i to the total initial mass of polymers in the underlying layer. If the cellulose polymers are evenly distributed across the layers, the ψ_i is taken as a time-independent constant. However, in the case of heterogenous cellulose distributions across the layers (i.e., penetration and internal zones), the ψ_i is formulated as a piecewise function of layers, l, which in turn is

a function of the particle radius, $R(t)$. The formulation of the ψ_i is further explored in Appendix B.3 and B.5.

B.2 Sterically Hindered Competitive Enzyme Adsorption

Provided that the adsorption/desorption of enzymes can be expressed in its elementary form as:

$$\mathrm{E}_{\mathrm{F}}^{\mathrm{m}} + \vartheta' \underset{k_{\mathrm{d}}^{\mathrm{m}}}{\overset{k_{\mathrm{a}}^{\mathrm{m}}}{\leftrightarrows}} \vartheta; \quad \mathrm{m} = \mathrm{CBH}, \mathrm{EG} \tag{B.4}$$

where ϑ and ϑ' are the fraction of occupied and unoccupied surface area, respectively, the material balance for the fraction of occupied surface area is:

$$\frac{d\vartheta}{dt} = k_{\mathrm{a}}^{\mathrm{m}}(1-\vartheta)C_{\mathrm{F}}^{\mathrm{m}} - k_{\mathrm{d}}^{\mathrm{m}}\vartheta; \quad \mathrm{m} = \mathrm{CBH}, \mathrm{EG} \tag{B.5}$$

Here, $C_{\mathrm{F}}^{\mathrm{m}}$ is the molar concentration of free enzymes, $k_{\mathrm{a}}^{\mathrm{m}}$ and $k_{\mathrm{d}}^{\mathrm{m}}$ are the adsorption and desorption rate constants, respectively. The molar concentration of surface-adsorbed enzymes can be derived as $C_{\mathrm{S}}^{\mathrm{m}}(t) = \sigma_{\mathrm{m}}\vartheta(t)S(t)$, where $S(t) = 2n\pi LR(t)$ is the total specific surface area of cellulose particles. Following this, Eq. (B.5) can be re-written to express the mole balance of surface-adsorbed enzymes as:

$$\frac{dC_{\mathrm{S}}^{\mathrm{m}}}{dt} = k_{\mathrm{a}}^{\mathrm{m}}C_{\mathrm{F}}^{\mathrm{m}}[\sigma_{\mathrm{m}}S(t) - C_{\mathrm{S}}^{\mathrm{m}}] - k_{\mathrm{d}}^{\mathrm{m}}C_{\mathrm{S}}^{\mathrm{m}}; \quad \mathrm{m} = \mathrm{CBH}, \mathrm{EG} \tag{B.6}$$

With the above, the ML-PBM portrays the effect of enzyme crowding when the enzyme loading exceeds the available surface area. Accordingly, the enzyme saturation concen-

tration that induces enzyme crowding can be expressed in mg/g cellulose as:

$$C_{\text{sat}}^{\text{m}} = \frac{S(t)\sigma_{\text{m}}M_{\text{m}}}{M_{\text{T}}}; \quad \text{m} = \text{CBH}, \text{EG} \tag{B.7}$$

where M_{m} are the molecular weights of the enzymes. Similarly, Eq. (B.6) also portrays the effect of competitive adsorption when more than one type of enzyme is employed with differing footprints and adsorption rate constants.

B.3 Fixed Pivot Discretization of ML-PBM Equations

As discussed in Chapter 3, FPT is employed to discretize the PBEs over a discrete-continuous mesh. In the discrete-continuous domain implementation, p and q represents the number of discrete and continuous points, respectively, producing a total of $p + q$ FPT-discretized breakage PBEs as follows:

$$\begin{aligned} \frac{dC_i}{dt} &= \sum_{j=i}^{p+q} \eta_{ij}^{\text{m}} k_j^{\text{m}} C_j - k_i^{\text{m}} C_i; \quad \text{m} = \text{CBH}, \text{EG}; \\ i &= 1, 2, \ldots, p, p+1, p+2, \ldots, p+q \end{aligned} \tag{B.8}$$

where

$$\begin{aligned} \eta_{ij}^{\text{m}} &= \int_{x_i}^{x_{i+1}} \left(\frac{x_{i+1} - v}{x_{i+1} - x_i} \right) b_{\text{m}}(v, x_j)\, dv + \int_{x_{i-1}}^{x_i} \left(\frac{v - x_{i-1}}{x_i - x_{i-1}} \right) b_{\text{m}}(v, x_j)\, dv; \\ \text{m} &= \text{CBH}, \text{EG} \end{aligned} \tag{B.9}$$

Following Eq. (3.21), for the chain-end cellobiose scission of CBH, the stoichiometric kernel is:

$$b_{\mathrm{CBH}}(v, x_j) = \begin{cases} \delta(v-2); & \text{for DP} = 2 \\ \delta(v-[x_j-2]); & \text{for DP} > 2 \end{cases} \tag{B.10}$$

Similarly, for the action of the EG, the random scission stoichiometric kernel, as defined in Section 3.3.6, is:

$$b_{\mathrm{EG}}(v, x_j) = \frac{2}{x_{j-1}} \tag{B.11}$$

In the discrete region, the grid points are spaced one monomer unit apart where $x_1 = 1$ is the DP of a monomer and the last grid point in the continuous region, $x_{p+q} = N$. In this work, non-uniform geometric grids with a common ratio of $r = (x_{p+q}/x_{p+1})^{1/(q-1)}$ are employed for the continuous region to cater to a broad distribution of DP. In order to have a smooth transition between the discrete and the continuous region, the choice of q was constrained to:

$$q = \left\lfloor 1 + \left\{ \ln\left(\frac{N}{p+v_m}\right) \Big/ \ln\left(1+\frac{v_m}{p+v_m}\right) \right\} \right\rfloor \tag{B.12}$$

where the floor function $\lfloor Z \rfloor$ denotes the nearest lower integer to Z. Throughout this work, $p = 20$ was used, where the total number of grids did not exceed $p+q = 217$ needed for the largest $N = 2 \times 10^5$ for α-cellulose. Accounting for all the species traced in the ML-PBM, the total number of equations was no more than $5(p+q) - 15 = 1070$ ODEs. Further information on the breakage PBM, derivation of FPT solutions and discrete-continuous meshing convention are given by Ho et al. [54, 58, 59], as well as in Chapter 3/Appendix A. Here, the final implementable form of ML-PBM equations is presented:

Soluble products:

$$\begin{aligned}\frac{dC_1}{dt} &= k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},3}^{\mathrm{CBH}} + \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG}} + \sum_{j=4}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG-CBH}} \\ &\quad + \sum_{j=3}^{6} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}'}^{\mathrm{EG}} C_{\mathrm{F}}^{\mathrm{EG}} C_j + 2 k_{\mathrm{h}'}^{\mathrm{BG}} C_{\mathrm{F}}^{\mathrm{BG}} C_2\end{aligned} \tag{B.13}$$

$$\begin{aligned}\frac{dC_2}{dt} &= k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},3}^{\mathrm{CBH}} + 2 k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},4}^{\mathrm{CBH}} + \sum_{5}^{p+q} k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B,j}}^{\mathrm{CBH}} + \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG}} \\ &\quad + \sum_{j=4}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG-CBH}} + \sum_{j=3}^{6} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}'}^{\mathrm{EG}} C_{\mathrm{F}}^{\mathrm{EG}} C_j - k_{\mathrm{h}'}^{\mathrm{BG}} C_{\mathrm{F}}^{\mathrm{BG}} C_2\end{aligned} \tag{B.14}$$

$$\begin{aligned}\frac{dC_i}{dt} &= \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG}} + \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG-CBH}} \\ &\quad + \sum_{j=i+1}^{6} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}'}^{\mathrm{EG}} C_{\mathrm{F}}^{\mathrm{EG}} C_j - k_{\mathrm{h}'}^{\mathrm{EG}} C_{\mathrm{F}}^{\mathrm{EG}} C_i; \quad i = 3, 4, 5\end{aligned} \tag{B.15}$$

$$\frac{dC_6}{dt} = \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG}} + \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG-CBH}} - k_{\mathrm{h}'}^{\mathrm{EG}} C_{\mathrm{F}}^{\mathrm{EG}} C_6 \tag{B.16}$$

CBH-polymer complex species:

$$\frac{dC_{\mathrm{B},3}^{\mathrm{CBH}}}{dt} = k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},5}^{\mathrm{CBH}} - k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},3}^{\mathrm{CBH}} \tag{B.17}$$

$$\begin{aligned}\frac{dC_{\mathrm{B},i}^{\mathrm{CBH}}}{dt} &= k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},i+2}^{\mathrm{CBH}} + k_{\mathrm{e}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG-CBH}} - k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B,i}}^{\mathrm{CBH}} - k_{\mathrm{f},i}^{\mathrm{EG}} C_{\mathrm{S}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{CBH}}; \\ &\quad i = 4, 5, 6\end{aligned} \tag{B.18}$$

$$\begin{aligned}\frac{dC_{\mathrm{B},i}^{\mathrm{CBH}}}{dt} &= k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},i+2}^{\mathrm{CBH}} + k_{\mathrm{e}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG-CBH}} + k_{\mathrm{f},i}^{\mathrm{CBH}} C_{\mathrm{S}}^{\mathrm{CBH}} C_{\mathrm{S},i} - k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},i}^{\mathrm{CBH}} \\ &\quad - k_{\mathrm{e}}^{\mathrm{CBH}} C_{\mathrm{B},i}^{\mathrm{CBH}} - k_{\mathrm{f},i}^{\mathrm{EG}} C_{\mathrm{S}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{CBH}}; \quad i = 7, 8, \ldots, p-1\end{aligned} \tag{B.19}$$

$$\frac{dC_{\mathrm{B},i}^{\mathrm{CBH}}}{dt} = \sum_{j=i}^{p+q} \eta_{ij}^{\mathrm{CBH}} k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},j}^{\mathrm{CBH}} + k_{\mathrm{e}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG-CBH}} + k_{\mathrm{f},i}^{\mathrm{CBH}} C_{\mathrm{S}}^{\mathrm{CBH}} C_{\mathrm{S},i} - k_{\mathrm{h}}^{\mathrm{CBH}} C_{\mathrm{B},i}^{\mathrm{CBH}} - k_{\mathrm{e}}^{\mathrm{CBH}} C_{\mathrm{B},i}^{\mathrm{CBH}} - k_{\mathrm{f},i}^{\mathrm{EG}} C_{\mathrm{S}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{CBH}}; \quad i = p, p+1, \ldots, p+q \tag{B.20}$$

EG-polymer complex species:

$$\frac{dC_{\mathrm{B},i}^{\mathrm{EG}}}{dt} = k_{\mathrm{f},i}^{\mathrm{EG}} C_{\mathrm{S}}^{\mathrm{EG}} C_{\mathrm{S},i} - k_{\mathrm{e}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG}} - k_{\mathrm{h},i}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG}}; \quad i = 7, 8, \ldots, p+q \tag{B.21}$$

EG-CBH-polymer complex species:

$$\frac{dC_{\mathrm{B},i}^{\mathrm{EG-CBH}}}{dt} = k_{\mathrm{f},i}^{\mathrm{EG}} C_{\mathrm{S}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{CBH}} - k_{\mathrm{e}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG-CBH}} - k_{\mathrm{h},i}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG-CBH}}; \; i = 4, 5, \ldots, p+q \tag{B.22}$$

Surface-accessible polymers:

$$\frac{dC_{\mathrm{S},i}}{dt} = C_{\mathrm{loss},i} + r_{\mathrm{exp},i}; \quad i = 7, 8, \ldots, p+q \tag{B.23}$$

where

$$C_{\mathrm{loss},i}(t) = \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG}} + \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG-CBH}} + k_{\mathrm{e}}^{\mathrm{CBH}} C_{\mathrm{B},i}^{\mathrm{CBH}} + k_{\mathrm{e}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG}} - k_{\mathrm{f},i}^{\mathrm{CBH}} C_{\mathrm{S}}^{\mathrm{CBH}} C_{\mathrm{S},i} - k_{\mathrm{f},i}^{\mathrm{EG}} C_{\mathrm{S}}^{\mathrm{EG}} C_{\mathrm{S},i}; \quad i = 7, 8, \ldots, p \tag{B.24}$$

$$C_{\mathrm{loss},i}(t) = \sum_{j=i+1}^{p+q} \eta_{ij}^{\mathrm{EG}} k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG}} + \sum_{j=i+1}^{p+q} \eta_{ij}^{\mathrm{EG}} k_{\mathrm{h},j}^{\mathrm{EG}} C_{\mathrm{B},j}^{\mathrm{EG-CBH}} + k_{\mathrm{e}}^{\mathrm{CBH}} C_{\mathrm{B},i}^{\mathrm{CBH}} + k_{\mathrm{e}}^{\mathrm{EG}} C_{\mathrm{B},i}^{\mathrm{EG}} - k_{\mathrm{f},i}^{\mathrm{CBH}} C_{\mathrm{S}}^{\mathrm{CBH}} C_{\mathrm{S},i} - k_{\mathrm{f},i}^{\mathrm{EG}} C_{\mathrm{S}}^{\mathrm{EG}} C_{\mathrm{S},i}; \quad i = p+1, p+2, \ldots, p+q \tag{B.25}$$

$$r_{\mathrm{exp},i}\left(t\right) = -\left[\frac{R\left(t\right)-R_0}{R\left(t\right)}\right]\left[\sum_{j=1}^{p+q}\left(162x_j+18\right)\, r_{\mathrm{loss},j}\left(t\right)\right]\psi_i\left(l\right); \\ i=7,8,\ldots,p+q; \quad l=1,2,\ldots,l_{\max}-1; \quad R\left(t\right)\geq R_0 \tag{B.26}$$

$$r_{\mathrm{loss},i}\left(t\right) = C_{\mathrm{loss},i} + \frac{dC_{\mathrm{B},i}^{\mathrm{CBH}}}{dt} + \frac{dC_{\mathrm{B},i}^{\mathrm{EG}}}{dt} + \frac{dC_{\mathrm{B},i}^{\mathrm{EG-CBH}}}{dt}; \quad i=3,4,\ldots,p+q \tag{B.27}$$

$$\psi_i\left(l\right) = \frac{C_{\mathrm{int},il}\left(0\right)}{\sum_{j=1}^{p+q}\left(162x_j+18\right)C_{\mathrm{int},jl}\left(0\right)}; \\ i=7,8,\ldots,p+q; \quad l=1,2,\ldots,l_{\max}-1 \tag{B.28}$$

Here, $r_{\mathrm{exp},i}(t)=0$ for $R(t)<R_0$. For Eq. (B.26) and (B.28), the $l=\lceil R(t)/R_0\rceil-1$ where the ceiling operator $\lceil Z\rceil$ denotes the nearest higher integer to Z.

Internal polymers:

$$\frac{dC_{\mathrm{int},i}}{dt} = r_{\mathrm{exp},i}; \quad i=7,8,\ldots,p+q \tag{B.29}$$

Surface-adsorbed enzymes:

$$\frac{dC_{\mathrm{S}}^{\mathrm{CBH}}}{dt} = k_{\mathrm{a}}^{\mathrm{CBH}}C_{\mathrm{F}}^{\mathrm{CBH}}\left(\sigma_{\mathrm{CBH}}S\left(t\right)-C_{\mathrm{S}}^{\mathrm{CBH}}\right)+\sum_{i=7}^{p+q}k_{\mathrm{e}}^{\mathrm{CBH}}C_{\mathrm{B},i}^{\mathrm{CBH}}-k_{\mathrm{d}}^{\mathrm{CBH}}C_{\mathrm{S}}^{\mathrm{CBH}} \\ -\sum_{i=7}^{p+q}k_{\mathrm{f},i}^{\mathrm{CBH}}C_{\mathrm{S}}^{\mathrm{CBH}}C_{\mathrm{S},i} \tag{B.30}$$

$$\frac{dC_{\mathrm{S}}^{\mathrm{EG}}}{dt} = k_{\mathrm{a}}^{\mathrm{EG}}C_{\mathrm{F}}^{\mathrm{EG}}\left(\sigma_{\mathrm{EG}}S\left(t\right)-C_{\mathrm{S}}^{\mathrm{EG}}\right)+\sum_{i=7}^{p+q}k_{\mathrm{e}}^{\mathrm{EG}}C_{\mathrm{B},i}^{\mathrm{EG}}+\sum_{i=4}^{p+q}k_{\mathrm{e}}^{\mathrm{EG}}C_{\mathrm{B},i}^{\mathrm{EG-CBH}} \\ -k_{\mathrm{d}}^{\mathrm{EG}}C_{\mathrm{S}}^{\mathrm{EG}}-\sum_{i=7}^{p+q}k_{\mathrm{f},i}^{\mathrm{EG}}C_{\mathrm{S}}^{\mathrm{EG}}C_{\mathrm{S},i}-\sum_{i=4}^{p+q}k_{\mathrm{f},i}^{\mathrm{EG}}C_{\mathrm{S}}^{\mathrm{EG}}C_{\mathrm{B},i}^{\mathrm{CBH}} \tag{B.31}$$

$$C_{\mathrm{T}}^{\mathrm{CBH}} = C_{\mathrm{F}}^{\mathrm{CBH}}+C_{\mathrm{S}}^{\mathrm{CBH}}+\sum_{i=3}^{p+q}C_{\mathrm{B},i}^{\mathrm{CBH}}+\sum_{i=7}^{p+q}C_{\mathrm{B},i}^{\mathrm{EG-CBH}} \tag{B.32}$$

$$C_{\mathrm{T}}^{\mathrm{EG}} = C_{\mathrm{F}}^{\mathrm{EG}} + C_{\mathrm{S}}^{\mathrm{EG}} + \sum_{i=7}^{p+q} C_{\mathrm{B},i}^{\mathrm{EG}} + \sum_{i=7}^{p+q} C_{\mathrm{B},i}^{\mathrm{EG-CBH}} \tag{B.33}$$

$$C_{\mathrm{T}}^{\mathrm{BG}} = C_{\mathrm{F}}^{\mathrm{BG}} \tag{B.34}$$

For Eq. (B.25), the simplified form of the particle allocation function is given as $\eta_{ij}^{\mathrm{EG}} = (x_{i+1} - x_{i-1})/x_{j-1}$. Here, the molar concentrations of all the species are traced, where C_i are the soluble oligomers, $C_{\mathrm{B},i}^{\mathrm{CBH}}$ are the CBH-enzyme complexes, $C_{\mathrm{B},i}^{\mathrm{EG}}$ are the EG-enzyme complexes, $C_{\mathrm{B},i}^{\mathrm{EG-CBH}}$ are the EG-CBH-enzyme complexes, $C_{\mathrm{S},i}$ are the surface polymers, $C_{\mathrm{int},i}$ are the total internal polymers beneath the surface, $C_{\mathrm{int},il}$ are the internal polymers in individual layers (specified by l) beneath the surface, $r_{\mathrm{loss},i}$ is the rate of loss of surface polymers due to solubilization and $r_{\mathrm{exp},i}$ is the rate of exposure of internal polymers to the surface. Subsequently, $C_{\mathrm{S}}^{\mathrm{m}}$ are the surface-adsorbed enzymes, $C_{\mathrm{F}}^{\mathrm{m}}$ are the free enzymes in the solution and $C_{\mathrm{T}}^{\mathrm{m}}$ are the total enzymes fed to the system, where m = CBH, EG or BG. As alluded to in Section 4.2.1, during hydrolysis, the particle radius reduces with time as the insoluble cellulose particle mass is lost through solubilization by the enzymes. The ML-PBM equations above are conveniently associated with the cellulose structure as:

$$R(t) = \sqrt{\frac{M_{\mathrm{T}} - \sum_{i=1}^{N}(162x_i + 18)C_i(t)}{n\rho\pi L}} \tag{B.35}$$

The rate coefficients in the equations above and the corresponding power law form assumed in this work are summarized in Table B.1.

Table B.1: Summary of rate coefficients and corresponding power law form.

Rate coefficients	Power law form	Remarks
$k_{\mathrm{a}}^{\mathrm{m}}, k_{\mathrm{d}}^{\mathrm{m}}$	Constant	m = CBH, EG
$k_{\mathrm{f}}^{\mathrm{CBH}}$	$k_{\mathrm{f}}^{\mathrm{CBH}}(v) = k_{\mathrm{p,f}}^{\mathrm{CBH}}/v$	Longer time is needed to locate the chain-ends of longer chains [43].
$k_{\mathrm{h}}^{\mathrm{CBH}}$	Constant	Subsite theory may not be applicable to the hydrolysis of celluloses [186]. Also, there is a lack of definitive evidence relating the intrinsic rate of hydrolysis to the cellulose DP.
$k_{\mathrm{f}}^{\mathrm{EG}}$ $k_{\mathrm{h}}^{\mathrm{EG}}$	$k_{\mathrm{f}}^{\mathrm{EG}}(v) = k_{\mathrm{p,f}}^{\mathrm{EG}}/v$ $k_{\mathrm{h}}^{\mathrm{EG}}(v) = k_{\mathrm{p,h}}^{\mathrm{EG}}/v$	An implicit attempt to relate the rate coefficients to the crystallinity via cellulose DP.
$k_{\mathrm{h}'}^{\mathrm{m}}$	Constant	m = EG, BG

B.4 Modelling Initial Conditions

The ML-PBM equations collated in the preceding section in the form of ODEs can be easily solved using appropriate commercial ODE solvers, given the appropriate initial conditions. For ease of readability, the gamma distribution to represent the molar concentration density for each layer l is reproduced here:

$$c_l(v,0) = \frac{\kappa_l \Omega(v)}{\int_1^N \Omega(v)\left[(162v+18)\right] dv}; \quad \Omega(v) = \frac{v^{\lambda-1}\exp\left(-v/\phi\right)}{\phi^{\lambda}\Gamma(\lambda)} \tag{B.36}$$

where $\lambda = \bar{M}_{\mathrm{n}}/(\bar{M}_{\mathrm{w}} - \bar{M}_{\mathrm{n}})$ and $\phi = \bar{M}_{\mathrm{w}} - \bar{M}_{\mathrm{n}}$. The mass concentration of celluloses in the individual layers is given as:

$$\kappa_l = n\rho\pi L[(2l-1)R_0^2]; \quad l = 1, 2, \ldots, l_{\max} \tag{B.37}$$

For surface polymers expressed by Eq. (B.23), the initial distribution is computed for $l = l_{\max}$ and the parameters λ and ϕ follow the distribution of celluloses in the penetration

zone. Likewise, for internal polymers expressed by Eq. (B.29), the initial distribution is the sum of distribution of all the remaining layers in the penetration and the internal zone, i.e., $\sum_{l=1}^{l_{\max}-1} c_l(v,0)$. Here, the parameters λ and ϕ follow the distribution of celluloses in the zone of the respective layers. The molar concentration density given here are easily converted to molar concentration for use with the ODEs in the previous section by applying the mean value theorem, i.e., $C_i(t) = \int_{x_{i-1/2}}^{x_{i+1/2}} c(v,t)\ dv$, where $C_i(t) = c_i(t)$ in the discrete region.

B.5 Modelling Simplified Homogeneous Model

Appropriate changes can be made to the ML-PBM equations to form a simplified homogeneous model. First, Eq. (B.26) is altered to:

$$\begin{gathered} r_{\text{exp},i}\,(t) = -\left[\frac{R\,(t) - R_0}{R\,(t)}\right]\left[\sum_{j=1}^{p+q} (162x_j + 18)\ r_{\text{loss},j}\,(t)\right]\psi_i; \\ i = 7, 8, \ldots, p+q; \quad l = 1, 2, \ldots, l_{\max} - 1; \quad R\,(t) \geq R_0 \end{gathered} \tag{B.38}$$

Unlike the ML-PBM, here ψ_i is no longer a function of l, but instead a time-independent constant is assumed, given by:

$$\psi_i = \frac{C_{\text{int},i}\,(0)}{\sum_{j=1}^{p+q} (162x_j + 18)\, C_{\text{int},j}\,(0)}; \quad i = 7, 8, \ldots, p+q \tag{B.39}$$

In tandem with the changes above in the model equations, the overall bimodal initial distribution (obtained by superimposing two gamma distribution) is uniformly dispersed throughout the cellulose particles. The initial mass concentration of the surface polymers (Eq. (B.23)) and internal polymers (Eq. (B.29)) is derived from the geometry as

$M_S = 2n\rho\pi LR_0R(0)$ and $M_I = M_T - M_S$, respectively.

B.6 Sensitivity Analysis and Parameter Identifiability

Sensitivity and parameter identifiability analyses were done prior to parameter estimation. Specifically, the methodology outlined by Brun et al. [173] was employed in this work. Sensitivity analysis enables the rating of the parameters in the order of importance in influencing the model outputs, whereas the collinearity indices (index of 20 or lower was considered identifiable) were then used to narrow down the identifiable parameter subsets with respect to the available data to avoid the effects of confounding parameters during model calibration. Based on the two calculations, the model parameters with the least sensitivities and poor identifiability were omitted from the model calibration exercise with rational assumptions. This approach was also recommended by Sin et al. [187].

Figure B.2 shows the relative sensitivities of ML-PBM outputs to perturbation of model parameters for different celluloses and enzyme cocktails. Specifically, cellobiose concentration and overall conversion were chosen for this analysis as they are the most common enzymatic hydrolysis descriptors in the literature. Throughout this work, the conversion was evaluated based on the remaining amount of insoluble celluloses. In all cases examined, $r_{\text{mass,P}}$ and P_{zone} significantly dictate the model outputs in addition to the kinetic parameters. Obviously, the idea of a multi-layered cellulose structure with distinct regions of different distributions and properties play a major role in influencing the trend of enzymatic hydrolysis. In general, an increase in $r_{\text{mass,P}}$ positively affects

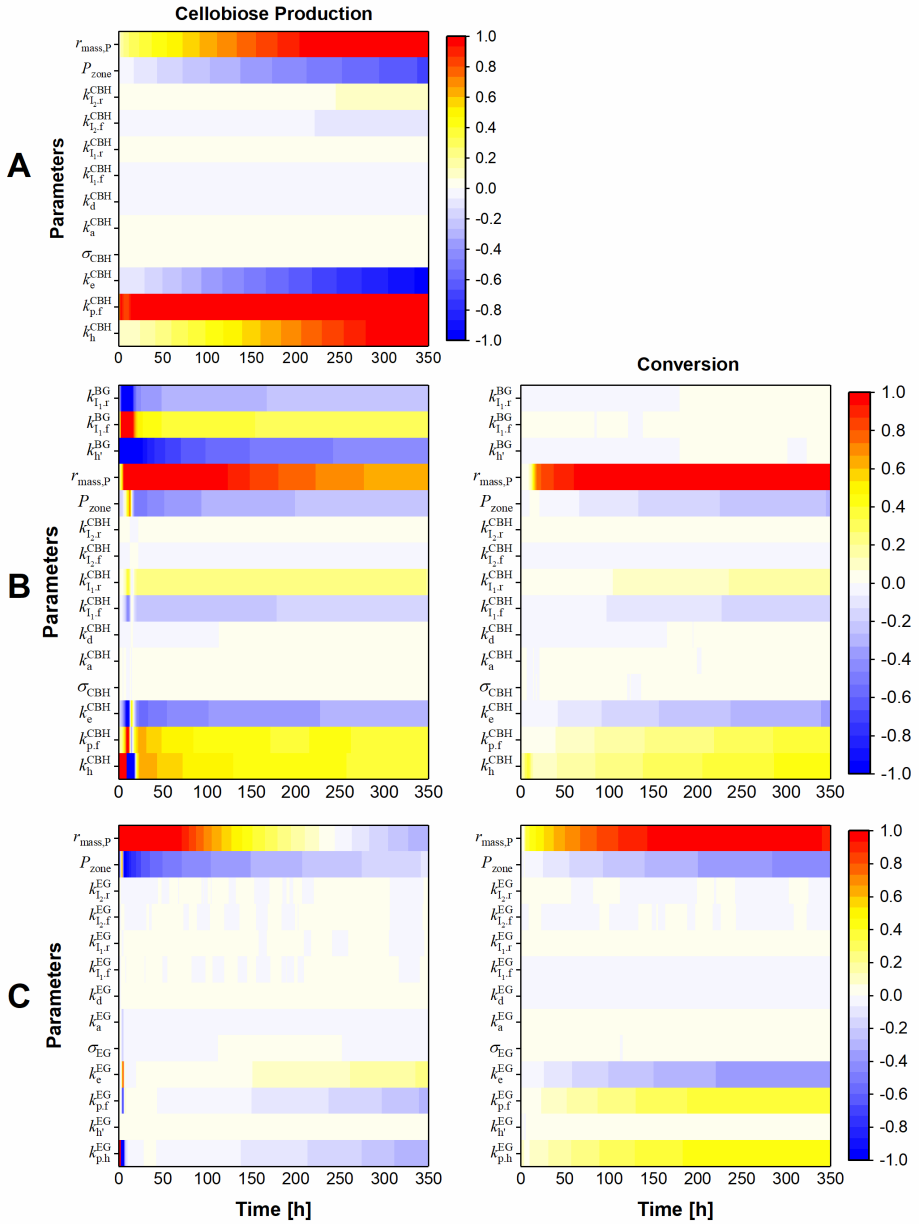

Figure B.2: Relative sensitivities of ML-PBM outputs (cellobiose production and conversion) to various parameters. The sensitivities are evaluated for hydrolysis of: (**A**) Avicel by CBH based on the data given in Medve et al. [86], (**B**) α-cellulose by CBH and BG, and (**C**) α-cellulose by CBH, BG and EG based on the data given in Nag et al. [85]. For (**C**), the EG-related parameters were evaluated whilst fixing the CBH and BG-related parameters. A +20% perturbation was enforced for all parameters except for P_{zone}, where an increment of 1 was employed. For case (**A**), results for the conversion have a similar trend to that of cellobiose production.

the rate and the extent of hydrolysis, whereas an increase in P_{zone} reduces the rate and extent of hydrolysis. An increase in $r_{\text{mass,P}}$ indicates that the portion of short-chained and more amorphous cellulose polymers is larger in the penetration zone where the enzymatic actions are rapid, thus, leading to a faster and greater overall hydrolysis conversion. The foregoing observation is also best explained in tandem with Eq. (4.17). A larger $r_{\text{mass,P}}$ also indicates that the particle size is smaller resulting in numerous particles in the reaction medium to account for the same mass, offering ample surface area for enzymatic adsorption. Conversely, the increase in P_{zone} renders an increase in the initial particle size, indicating a greater degree of microfibril aggregation (see Figure 4.1D), which effectively reduces the surface area for enzymatic adsorption. The effect of the increase in P_{zone} on conversion is only noticeable in later stages of the process after all the available surface area has been crowded by the surface-adsorbed enzymes and the rate of exposure of new polymers from the layers underneath is slower. On the contrary, when BG is included in the enzyme cocktails, the effect of the increase in P_{zone} on cellobiose production is more prominent at the peak of the cellobiose concentration profile.

Also in Figure B.2, $k^{\text{m}}_{\text{I}_1,\text{f}}$ and $k^{\text{m}}_{\text{I}_2,\text{f}}$ are the forward inhibition rate constants, whereas $k^{\text{m}}_{\text{I}_1,\text{r}}$ and $k^{\text{m}}_{\text{I}_2,\text{r}}$ are the reverse inhibition rate constants for glucose and cellobiose, respectively, where m = CBH, EG or BG. Well-known studies by Eriksson et al. [127] and Zhang et al. [128] rejected product inhibition as the main reason for the drastic rate slowdown. The enzymatic hydrolysis rate was observed to precipitously decline even when the strong cellobiose inhibition was eliminated by adding excess BG enzymes [21,101,128,129]. Desai and Converse [188] also observed that the enzymatic hydrolysis rate continued to decline in the absence of product inhibition where hydrolysis was interrupted multiple times and restarted with fresh buffer free from any soluble products from before. Therefore, one can

deduce that product inhibition could not possibly be the main rate-limiting phenomenon during enzymatic hydrolysis of celluloses. Here, the inhibition constants were included in the analysis to show that they do not play a major role in dictating the model outputs. The glucose inhibition on BG characterized by $k_{\mathrm{I_1,f}}^{\mathrm{BG}}$ and $k_{\mathrm{I_1,r}}^{\mathrm{BG}}$ are the exception, where a considerable influence on cellobiose production is observed. However, the lack of transient data for cellobiose in the literature [85] renders the inhibition rate constants poorly identifiable and thus, they were set to zero when calibrating against the overall conversion. As the experimental data used in this study displays either hydrolysis with less than 10% conversion (Figure 4.3, Figure 4.4, Figure 4.10) or hydrolysis without strong cellobiose inhibition (by adding BG, Figure 4.5), the assumption to neglect product inhibition is rationalized. In order to improve the identifiability of the inhibition rate constants in the future, the model is best calibrated with experimental data of specific inhibition studies with varying initial product concentrations in the reaction medium.

Similarly, enzyme-centric parameters (σ_{m}, $k_{\mathrm{a}}^{\mathrm{m}}$ and $k_{\mathrm{d}}^{\mathrm{m}}$, where m = CBH, EG) have little to no influence on the cellobiose concentration or conversion but their effect is more noticeable in the concentration profile of surface-adsorbed enzymes at larger perturbations (not shown). The enzyme-centric parameters of CBH and EG were estimated by calibrating the ML-PBM to the experimental adsorption data provided by Medve et al. [86], where the estimated parameter values were retained for the rest of the ML-PBM predictions with similar experimental conditions in this work.

B.7 Model Parameter Estimation

Parameter estimation was performed systematically in stages. Enzyme-centric parameters were estimated once with the hydrolysis data that comes equipped with enzyme adsorption data such as in Figure 4.3. Here, the enzyme adsorption data is crucial for the estimation of the enzyme-centric parameters as they are not sensitive to other model outputs such as cellobiose concentration or overall conversion. Following this, the enzyme-centric parameter values obtained were retained for other experimental cases without re-fitting. Provided that the experimental conditions are similar, enzyme-centric parameters such as enzyme footprints and adsorption/desorption rate constants should be global across various data sets as they refer to the general structure of the enzymes and their adsorption kinetics. On the other hand, when multiple data sets with different enzyme cocktails are present, parameter estimation and validation was also done in stages.

First, the ML-PBM was calibrated to the experimental data from Bezerra et al. [87], Bezerra and Dias [89], and Medve et al. [86], Nag et al. [85] and Eriksson et al. [127]. Subsequently, in order to demonstrate the predictive capability, the ML-PBM was validated against multiple data sets from the same literature with different initial cellulose and enzyme loadings or with different types of celluloses. Cross-literature validation was not performed as the physical properties of the commercial celluloses may vary between different sources and batches [92]. The enzyme activities are also found to greatly vary according to the source, degree and method of purification employed [189]. As for the data provided by Nag et al. [85], the ML-PBM was first calibrated to the hydrolysis of α-cellulose by CBH+BG enzyme cocktail. Subsequently, for the hydrolysis of α-cellulose by CBH+BG+EG enzyme cocktail, only the EG-related parameters were calibrated while

retaining the previously estimated parameters. Similarly, for the hydrolysis of HCC by both enzyme cocktails above, only the substrate-related parameters of HCC were calibrated while fixing all the previously estimated parameters. For the results in Section 4.3.7, the ML-PBM was calibrated with the fully batch data from Eriksson et al. [127] and validated with the fed-batch data from the same source. In general, a large number of ML-PBM parameters were calibrated, validated and evaluated systematically in stages, with one identifiable parameter subset at a time, leaving very little room for doubt on the parameter compensation or confounding effects.

For clarity, the inputs and outputs for the ML-PBM calibration are summarized in Figure B.3. The parameters were estimated by minimizing the objective function in the

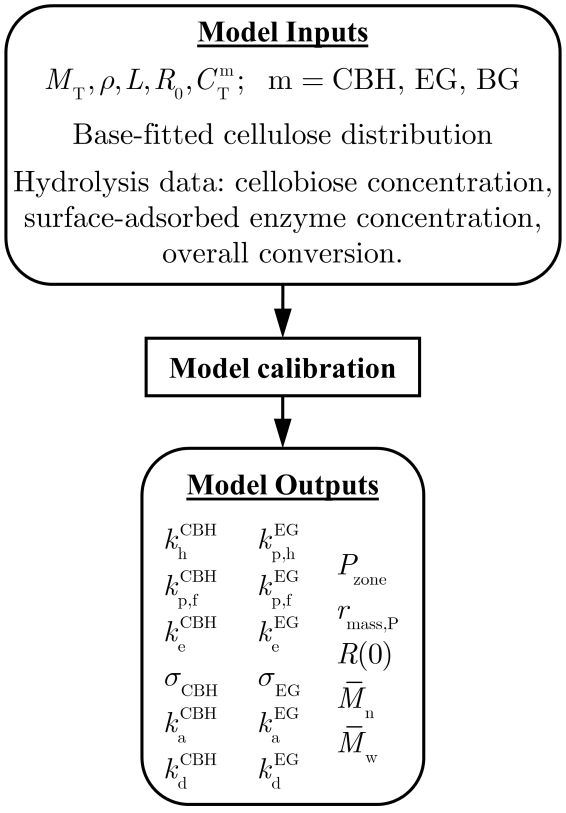

Figure B.3: Overview of inputs and outputs for ML-PBM calibration.

form of a weighted sum of squares of residuals, given as:

$$J_{\text{opt}} = \sum_{i} W_i \left\{ \left(\widehat{\mathbf{y}}_i - \mathbf{y}_i \right)^{\text{T}} \left(\widehat{\mathbf{y}}_i - \mathbf{y}_i \right) \right\} \tag{B.40}$$

where i = cellobiose concentration, conversion and/or surface-adsorbed enzyme concentration, $\widehat{\mathbf{y}}_i$ is the vector of predicted model output, $\mathbf{y}_i$ is the vector of experimental data and W_i is the weight allocated to different model outputs. The minimization of the objective function was performed using the Genetic Algorithm of MATLAB® R2016a on a workstation equipped with an Intel® Core™ i5-7200U processor with a clock speed of 2.50 GHz and 16 GB of installed memory (RAM). The ODEs are solved using the "ode15s" sub-routine over a time span that covers the experimental data.

Appendix C

Formulation of Unified Cybernetic-Population Balance Model

C.1 Reduced Core Metabolic Network of *C. thermocellum*

The genome-scale metabolic network of *C. thermocellum* provided by Thompson et al. [155] is reduced to a manageable core network while retaining the crucial cellulosome secretion component. The list of participating metabolic reactions and metabolites in the reduced network are given in the following tables.

Table C.1: List of participating reactions in reduced core metabolic network of *C. thermocellum* under anaerobic conditions.

No.	Reaction ID	Equation*	Lower bound flux†	Upper bound flux†	Subsystem
1	R_R00004_c	2 Orthophosphate[C_c] + 2 H+[C_e] =>H2O[C_c] + Diphosphate[C_c] + 2 H+[C_c]	0	1000	-
2	R_R00014_c	Pyruvate[C_c] + Thiamin diphosphate[C_c] + 2 H+[C_c] =>CO2[C_c] + 2-(alpha-Hydroxyethyl)thiamine diphosphate[C_c]	0	1000	Glycolysis / Gluco-neogenesis
3	R_R00019_c	2 H+[C_c] + 2 Reduced ferredoxin[C_c] =>2 Oxidized ferredoxin[C_c] + Hydrogen[C_c]	0	1000	Redox metabolism
4	R_R00206_c	Diphosphate[C_c] + AMP[C_c] + Phosphoenolpyruvate[C_c] =>ATP[C_c] + Orthophosphate[C_c] + Pyruvate[C_c] + H+[C_c]	0	1000	Pyruvate metabolism
5	R_R00212_c	CoA[C_c] + Pyruvate[C_c] =>Acetyl-CoA[C_c] + Formate[C_c]	0	1000	Pyruvate metabolism
6	R_R00216_c	NADP+[C_c] + (S)-Malate[C_c] =>NADPH[C_c] + CO2[C_c] + Pyruvate[C_c]	0	1000	Pyruvate metabolism
7	R_R00217_c	Oxaloacetate[C_c] + H+[C_c] =>CO2[C_c] + Pyruvate[C_c]	0	1000	Pyruvate metabolism
8	R_R00228_c	NADH[C_c] + Acetyl-CoA[C_c] + H+[C_c] =>NAD+[C_c] + CoA[C_c] + Acetaldehyde[C_c]	0	1000	Pyruvate Metabolism
9	R_R00235_c	ATP[C_c] + CoA[C_c] + Acetate[C_c] =>Diphosphate[C_c] + AMP[C_c] + Acetyl-CoA[C_c]	0	1000	Pyruvate Metabolism
10	R_R00236_c	AMP[C_c] + Acetyl-CoA[C_c] + H+[C_c] =>CoA[C_c] + Acetyl adenylate[C_c]	0	1000	Pyruvate metabolism
11	R_R00267_c	NADP+[C_c] + Isocitrate[C_c] =>NADPH[C_c] + CO2[C_c] + 2-Oxoglutarate[C_c] + H+[C_c]	0	1000	Citrate cycle (TCA cycle)
12	R_R00268_c	H+[C_c] + Oxalosuccinate[C_c] =>CO2[C_c] + 2-Oxoglutarate[C_c]	0	1000	Citrate cycle (TCA cycle)
13	R_R00315_c	ADP[C_c] + Acetyl phosphate[C_c] =>ATP[C_c] + Acetate[C_c]	0	1000	Pyruvate Metabolism
14	R_R00316_c	Diphosphate[C_c] + Acetyl adenylate[C_c] =>ATP[C_c] + Acetate[C_c] + H+[C_c]	0	1000	Pyruvate metabolism
15	R_R00344_c	ATP[C_c] + CO2[C_c] + Pyruvate[C_c] =>ADP[C_c] + Orthophosphate[C_c] + Oxaloacetate[C_c] + H+[C_c]	0	1000	Citrate cycle (TCA cycle)
16	R_R00351_c	H2O[C_c] + Acetyl-CoA[C_c] + Oxaloacetate[C_c] =>CoA[C_c] + H+[C_c] + Citrate[C_c]	0	1000	Citrate cycle (TCA cycle)
17	R_R00431_c	CO2[C_c] + GDP[C_c] + Phosphoenolpyruvate[C_c] =>Oxaloacetate[C_c] + GTP[C_c]	0	1000	Glycolysis / Gluco-neogenesis
18	R_R00451_c	H+[C_c] + meso-2,6-Diaminoheptanedioate[C_c] =>CO2[C_c] + L-Lysine[C_c]	0	1000	Lysine biosynthe-sis
19	R_R00658_c	2-Phospho-D-glycerate[C_c] <=>H2O[C_c] + Phosphoenolpyruvate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
20	R_R00703_c	NADH[C_c] + Pyruvate[C_c] + H+[C_c] =>NAD+[C_c] + (S)-Lactate[C_c]	0	1000	Glycolysis / Gluco-neogenesis
21	R_R00710_c	H2O[C_c] + NAD+[C_c] + Acetaldehyde[C_c] =>NADH[C_c] + Acetate[C_c] + H+[C_c]	0	1000	Glycolysis / Gluco-neogenesis
22	R_R00711_c	H2O[C_c] + NADP+[C_c] + Acetaldehyde[C_c] =>NADPH[C_c] + Acetate[C_c] + H+[C_c]	0	1000	Glycolysis / Gluco-neogenesis
23	R_R00754_c	NADH[C_c] + H+[C_c] + Acetaldehyde[C_c] =>NAD+[C_c] + Ethanol[C_c]	0	1000	Glycolysis / Gluco-neogenesis

Continued on next page

Table C.1 – (continued)

No.	Reaction ID	Equation	Lower bound flux	Upper bound flux	Subsystem
24	R_R00948_c	ATP[C_c] + H+[C_c] + D-Glucose 1-phosphate[C_c] =>Diphosphate[C_c] + ADP-glucose[C_c]	0	1000	Glycolysis / Gluco-neogenesis
25	R_R00952_c	Orthophosphate[C_c] + Cellobiose[C_c] =>D-Glucose[C_c] + D-Glucose 1-phosphate[C_c]	0	1000	Cellulose Metabolism
26	R_R00959_c	D-Glucose 1-phosphate[C_c] <=>alpha-D-Glucose 6-phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
27	R_R01015_c	D-Glyceraldehyde 3-phosphate[C_c] <=>Glycerone phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
28	R_R01049_c	ATP[C_c] + D-Ribose 5-phosphate[C_c] =>AMP[C_c] + H+[C_c] + 5-Phospho-alpha-D-ribose 1-diphosphate[C_c]	0	1000	Pentose phosphate pathway
29	R_R01056_c	D-Ribose 5-phosphate[C_c] <=>D-Ribulose 5-phosphate[C_c]	-1000	1000	Pentose phosphate pathway
30	R_R01057_c	alpha-D-Ribose 1-phosphate[C_c] <=>D-Ribose 5-phosphate[C_c]	-1000	1000	Pentose phosphate pathway
31	R_R01061_c	NAD+[C_c] + Orthophosphate[C_c] + D-Glyceraldehyde 3-phosphate[C_c] <=>NADH[C_c] + H+[C_c] + 3-Phospho-D-glyceroyl phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
32	R_R01082_c	(S)-Malate[C_c] <=>H2O[C_c] + Fumarate[C_c]	-1000	1000	Citrate cycle (TCA cycle)
33	R_R01196_c	CoA[C_c] + Pyruvate[C_c] + 2 Oxidized ferredoxin[C_c] =>CO2[C_c] + Acetyl-CoA[C_c] + 2 H+[C_c] + 2 Reduced ferredoxin[C_c]	0	1000	Glycolysis / Gluco-neogenesis
34	R_R01197_c	CoA[C_c] + 2-Oxoglutarate[C_c] + 2 Oxidized ferredoxin[C_c] =>CO2[C_c] + 2 H+[C_c] + Succinyl-CoA[C_c] + 2 Reduced ferredoxin[C_c]	0	1000	Citrate cycle (TCA cycle)
35	R_R01324_c	Citrate[C_c] <=>Isocitrate[C_c]	-1000	1000	-
36	R_R01512_c	ATP[C_c] + 3-Phospho-D-glycerate[C_c] =>ADP[C_c] + 3-Phospho-D-glyceroyl phosphate[C_c]	0	1000	Glycolysis / Gluco-neogenesis
37	R_R01518_c	2-Phospho-D-glycerate[C_c] <=>3-Phospho-D-glycerate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
38	R_R01529_c	D-Ribulose 5-phosphate[C_c] <=>D-Xylulose 5-phosphate[C_c]	-1000	1000	Pentose phosphate pathway
39	R_R01600_c	GTP[C_c] + beta-D-Glucose[C_c] =>GDP[C_c] + beta-D-Glucose 6-phosphate[C_c]	0	1000	Glycolysis / Gluco-neogenesis
40	R_R01899_c	NADP+[C_c] + Isocitrate[C_c] =>NADPH[C_c] + H+[C_c] + Oxalosuccinate[C_c]	0	1000	Citrate cycle (TCA cycle)
41	R_R02073_c	Diphosphate[C_c] + beta-D-Fructose 6-phosphate[C_c] =>Orthophosphate[C_c] + beta-D-Fructose 1,6-bisphosphate[C_c]	0	1000	Fructose and mannose metabolism
42	R_R02739_c	alpha-D-Glucose 6-phosphate[C_c] <=>beta-D-Glucose 6-phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
43	R_R02740_c	alpha-D-Glucose 6-phosphate[C_c] =>beta-D-Fructose 6-phosphate[C_c]	0	1000	Glycolysis / Gluco-neogenesis
44	R_R03321_c	beta-D-Glucose 6-phosphate[C_c] <=>beta-D-Fructose 6-phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
45	R_R04779_c	ATP[C_c] + beta-D-Fructose 6-phosphate[C_c] =>ADP[C_c] + beta-D-Fructose 1,6-bisphosphate[C_c]	0	1000	Glycolysis / Gluco-neogenesis

Continued on next page

Table C.1 – (continued)

No.	Reaction ID	Equation	Lower bound flux	Upper bound flux	Subsystem
46	R_R08572_c	ATP[C_c] + D-Glycerate[C_c] =>ADP[C_c] + 2-Phospho-D-glycerate[C_c]	0	1000	Pentose phosphate pathway
47	R_R05875_c	NAD+[C_c] + H+[C_c] + 2 Reduced ferredoxin[C_c] <=>NADH[C_c] + 2 Oxidized ferredoxin[C_c]	-1000	1000	Redox metabolism
48	R_R07181_c	NADP+[C_c] + Hydrogen[C_c] <=>NADPH[C_c] + H+[C_c]	-1000	1000	Redox metabolism
49	R_R09508_c	NADH[C_c] + 5 H+[C_c] + 2 Reduced ferredoxin[C_c] <=>NAD+[C_c] + 2 Oxidized ferredoxin[C_c] + 2 Hydrogen[C_c]	-1000	1000	Redox metabolism
50	R_R_OPM9r	NADH[C_c] + 2 NADP+[C_c] + 2 Reduced ferredoxin[C_c] <=>NAD+[C_c] + 2 NADPH[C_c] + 2 Oxidized ferredoxin[C_c]	-1000	1000	Redox metabolism
51	R_R01070_c	beta-D-Fructose 1,6-bisphosphate[C_c] <=>Glycerone phosphate[C_c] + D-Glyceraldehyde 3-phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
52	R_R00342_c	NAD+[C_c] + (S)-Malate[C_c] <=>NADH[C_c] + Oxaloacetate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
53	R_R01517_c	GTP[C_c] + 3-Phospho-D-glycerate[C_c] =>GDP[C_c] + 3-Phospho-D-glyceroyl phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
54	R_R01641_c	D-Ribose 5-phosphate[C_c] + D-Xylulose 5-phosphate[C_c] <=>D-Glyceraldehyde 3-phosphate[C_c] + Sedoheptulose 7-phosphate[C_c]	-1000	1000	Glycolysis / Gluco-neogenesis
55	R_R01829_c	Sedoheptulose 7-phosphate[C_c] <=>Glycerone phosphate[C_c] + D-Erythrose 4-phosphate[C_c]	-1000	1000	-
56	R_R01830_c	D-Xylulose 5-phosphate[C_c] + D-Erythrose 4-phosphate[C_c] <=>D-Glyceraldehyde 3-phosphate[C_c] + beta-D-Fructose 6-phosphate[C_c]	-1000	1000	-
57	R_Glycogen_Cycle_1	ADP-glucose[C_c] =>ADP[C_c] + H+[C_c] + Glycogen[C_c]	0	1000	-
58	R_R_DCW	0.5285 Protein[C_c] + 0.026 DNA[C_c] + 0.0655 RNA[C_c] + 0.076 Lipid[C_c] + 0.2242 Cell Wall[C_c] + 0.00494 Solute Pool[C_c] + 0.0304 Total_LTA[C_c] + 0.2 Cellulosome term[C_c] =>Cell Dry Weight no GAM[C_c]	0‡	1000‡	Biomass + Protein (no ATP)
59	R_R_BIO	H2O[C_c] + ATP[C_c] + Cell Dry Weight no GAM[C_c] =>ADP[C_c] + Orthophosphate[C_c] + Biomass[C_c]	0‡	1000‡	Biomass + Protein (w ATP)
60	R_R_MAINT	H2O[C_c] + ATP[C_c] =>ADP[C_c] + Orthophosphate[C_c] + H+[C_c]	3.27	3.27	Maintenance
61	R_R_Cellulosome	43.78 H2O[C_c] + 43.78 ATP[C_c] + 0.4317 L-Glutamate[C_c] + 0.6373 Glycine[C_c] + 0.5581 L-Alanine[C_c] + 0.441 L-Lysine[C_c] + 0.5523 L-Aspartate[C_c] + 0.2365 L-Arginine[C_c] + 0.1871 L-Glutamine[C_c] + 0.4942 L-Serine[C_c] + 0.1469 L-Methionine[C_c] + 0.1718 L-Tryptophan[C_c] + 0.2851 L-Phenylalanine[C_c] + 0.3235 L-Tyrosine[C_c] + 0.0665 L-Cysteine[C_c] + 0.5329 L-Leucine[C_c] + 0.1099 L-Histidine[C_c] + 0.4823 L-Proline[C_c] + 0.5523 L-Asparagine[C_c] + 0.506 L-Valine[C_c] + 0.5572 L-Threonine[C_c] + 0.4739 L-Isoleucine[C_c] =>43.78 ADP[C_c] + 43.78 Orthophosphate[C_c] + Cellulosome term[C_c]	0§	1000§	Cellulosome Secretion
62	T_c_to_e_C00011_c	CO2[C_c] <=>CO2[C_e]	-1000	1000	CO2 out
63	T_e_to_c_C00185_c	H2O[C_c] + ATP[C_c] + Cellobiose[C_e] =>ADP[C_c] + Orthophosphate[C_c] + Cellobiose[C_c]	0	1000	Cellobiose in
64	T_c_to_e_C00469_c	H+[C_c] + Ethanol[C_c] <=>Ethanol[C_e] + H+[C_e]	-1000	1000	Ethanol out
65	T_c_to_e_C00033_c	Acetate[C_c] + H+[C_c] <=>Acetate[C_e] + H+[C_e]	-1000	1000	Acetate out
66	T_c_to_e_C00058_c	Formate[C_c] + H+[C_c] <=>Formate[C_e] + H+[C_e]	-1000	1000	Formate out
67	T_c_to_e_C00186_c	H+[C_c] + (S)-Lactate[C_c] <=>(S)-Lactate[C_e] + H+[C_e]	-1000	1000	Lactate out
68	T_c_to_e_C00282_c	Hydrogen[C_c] <=>Hydrogen[C_e]	-1000	1000	H2 out
69	T_c_to_e_C00022_c	Pyruvate[C_c] + H+[C_c] <=>Pyruvate[C_e] + H+[C_e]	-1000	1000	Pyruvate out
70	T_c_to_e_C00042_c	Succinate[C_c] + H+[C_c] <=>Succinate[C_e] + H+[C_e]	-1000	1000	Succinate out
71	T_c_to_e_m85	Biomass[C_c] =>Biomass[C_e]	0	1000	Biomass + Protein out
72	T_c_to_e_C14710_c	Isobutanol[C_c] <=>Isobutanol[C_e]	-1000	1000	Isobutanol out
73	R_EXC_OUT_m90	Cellulosome term[C_c] =>	0	1000	Cellulosome out

Continued on next page

Table C.1 – (continued)

No.	Reaction ID	Equation	Lower bound flux	Upper bound flux	Subsystem
74[¶]	R_Lumped_1	H2O[C_c] + ATP[C_c] + NADH[C_c] + NADPH[C_c] + NH3[C_c] + Acetyl-CoA[C_c] + Glycine[C_c] + L-Alanine[C_c] + GTP[C_c] + L-Lysine[C_c] + L-Aspartate[C_c] + L-Arginine[C_c] + L-Glutamine[C_c] + L-Serine[C_c] + L-Methionine[C_c] + Phosphoenolpyruvate[C_c] + L-Tryptophan[C_c] + L-Phenylalanine[C_c] + H+[C_c] + L-Tyrosine[C_c] + L-Cysteine[C_c] + D-Glucose 1-phosphate[C_c] + Glycerone phosphate[C_c] + 5-Phospho-alpha-D-ribose 1-diphosphate[C_c] + L-Leucine[C_c] + L-Histidine[C_c] + Oxidized ferredoxin[C_c] + 3-Methyl-2-oxobutanoic acid[C_c] + L-Proline[C_c] + L-Asparagine[C_c] + L-Valine[C_c] + L-Threonine[C_c] + L-Isoleucine[C_c] + meso-2,6-Diaminoheptanedioate[C_c] + beta-D-Fructose 6-phosphate[C_c] + Ca2+[C_e] + Mg2+[C_e] + K+[C_e] + Orthophosphate[C_e] + Fe3+[C_e] + Nicotinate D-ribonucleotide[C_e] =>NAD+[C_c] + NADP+[C_c] + ADP[C_c] + Orthophosphate[C_c] + CoA[C_c] + CO2[C_c] + Diphosphate[C_c] + AMP[C_c] + L-Glutamate[C_c] + GDP[C_c] + Fumarate[C_c] + Reduced ferredoxin[C_c] + Protein[C_c] + DNA[C_c] + RNA[C_c] + Lipid[C_c] + Cell Wall[C_c] + Solute Pool[C_c] + Total_LTA[C_c]	0	1000	-
75	T_c_to_e_C00183_c	H+[C_c] + L-Valine[C_c] <=>L-Valine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
76	T_c_to_e_C00082_c	H+[C_c] + L-Tyrosine[C_c] <=>L-Tyrosine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
77	T_c_to_e_C00078_c	L-Tryptophan[C_c] + H+[C_c] <=>L-Tryptophan[C_e] + H+[C_e]	-1000	1000	Transport between c and e
78	T_c_to_e_C00188_c	H+[C_c] + L-Threonine[C_c] <=>L-Threonine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
79	T_c_to_e_C00065_c	L-Serine[C_c] + H+[C_c] <=>L-Serine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
80	T_c_to_e_C00148_c	H+[C_c] + L-Proline[C_c] <=>L-Proline[C_e] + H+[C_e]	-1000	1000	Transport between c and e
81	T_c_to_e_C00079_c	L-Phenylalanine[C_c] + H+[C_c] <=>L-Phenylalanine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
82	T_c_to_e_C00073_c	L-Methionine[C_c] + H+[C_c] <=>L-Methionine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
83	T_c_to_e_C00123_c	H+[C_c] + L-Leucine[C_c] <=>L-Leucine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
84	T_c_to_e_C00135_c	H+[C_c] + L-Histidine[C_c] <=>L-Histidine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
85	T_c_to_e_C00037_c	Glycine[C_c] + H+[C_c] <=>Glycine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
86	T_c_to_e_C00064_c	L-Glutamine[C_c] + H+[C_c] <=>L-Glutamine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
87	T_c_to_e_C00097_c	H+[C_c] + L-Cysteine[C_c] <=>L-Cysteine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
88	T_c_to_e_C00049_c	L-Aspartate[C_c] + H+[C_c] <=>L-Aspartate[C_e] + H+[C_e]	-1000	1000	Transport between c and e
89	T_c_to_e_C00152_c	H+[C_c] + L-Asparagine[C_c] <=>L-Asparagine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
90	T_c_to_e_C00062_c	L-Arginine[C_c] + H+[C_c] <=>L-Arginine[C_e] + H+[C_e]	-1000	1000	Transport between c and e
91	T_c_to_e_C00014_c	H2O[C_c] + ATP[C_c] + NH3[C_e] <=>ADP[C_c] + Orthophosphate[C_c] + NH3[C_c]	-1000	1000	Transport between c and e
92	R_Glycogen_Cycle_2	Orthophosphate[C_c] + Glycogen[C_c] =>D-Glucose 1-phosphate[C_c]	0	1000	-

Continued on next page

Table C.1 – (continued)

No.	Reaction ID	Equation	Lower bound flux	Upper bound flux	Subsystem
93¶	R_Lumped_2	NADPH[C_c] + H+[C_c] + Oxidized ferredoxin[C_c] + 3-Methyl-2-oxobutanoic acid[C_c] =>NADP+[C_c] + CO2[C_c] + Reduced ferredoxin[C_c] + Isobutanol[C_c]	0	1000	-
94¶	R_Lumped_3	NADH[C_c] + Pyruvate[C_c] + L-Serine[C_c] + H+[C_c] =>NAD+[C_c] + L-Alanine[C_c] + D-Glycerate[C_c]	-1000	1000	-
95¶	R_Lumped_4	NADPH[C_c] + Pyruvate[C_c] + L-Glutamate[C_c] + H+[C_c] + L-Aspartate 4-semialdehyde[C_c] =>H2O[C_c] + NADP+[C_c] + 2-Oxoglutarate[C_c] + meso-2,6-Diaminoheptanedioate[C_c]	0	1000	-
96¶	R_Lumped_5	NADPH[C_c] + L-Glutamate[C_c] + H+[C_c] + Succinyl-CoA[C_c] + L-Aspartate 4-semialdehyde[C_c] + 2-(alpha-Hydroxyethyl)thiamine diphosphate[C_c] =>NADP+[C_c] + CoA[C_c] + NH3[C_c] + 2-Oxoglutarate[C_c] + Succinate[C_c] + Thiamin diphosphate[C_c] + L-Isoleucine[C_c]	0	1000	-
97¶	R_Lumped_6	H2O[C_c] + ATP[C_c] + 5-Phospho-alpha-D-ribose 1-diphosphate[C_c] =>ADP[C_c] + Orthophosphate[C_c] + Diphosphate[C_c] + alpha-D-Ribose 1-phosphate[C_c]	0	1000	-
98¶	R_Lumped_7	ATP[C_c] + NADPH[C_c] + L-Aspartate[C_c] + H+[C_c] =>NADP+[C_c] + ADP[C_c] + Orthophosphate[C_c] + L-Aspartate 4-semialdehyde[C_c]	0	1000	-
99	R_R01214_c	2-Oxoglutarate[C_c] + L-Valine[C_c] <=>L-Glutamate[C_c] + 3-Methyl-2-oxobutanoic acid[C_c]	-1000	1000	Valine, leucine and isoleucine degradation
100	R_R00230_c	Orthophosphate[C_c] + Acetyl-CoA[C_c] <=>CoA[C_c] + Acetyl phosphate[C_c]	-1000	1000	Taurine and hypotaurine metabolism

Table C.2: List of participating metabolites in reduced core metabolic network of *C. thermocellum* under anaerobic conditions.

No.	Metabolites	Compartment*	No.	Metabolites	Compartment*
1	H2O	C_c	67	D-Erythrose 4-phosphate	C_c
2	ATP	C_c	68	Hydrogen	C_c
3	NAD+	C_c	69	Isocitrate	C_c
4	NADH	C_c	70	L-Isoleucine	C_c
5	NADPH	C_c	71	L-Aspartate 4-semialdehyde	C_c
6	NADP+	C_c	72	Ethanol	C_c
7	ADP	C_c	73	ADP-glucose	C_c
8	Orthophosphate	C_c	74	alpha-D-Ribose 1-phosphate	C_c
9	CoA	C_c	75	2-Phospho-D-glycerate	C_c
10	CO2	C_c	76	alpha-D-Glucose 6-phosphate	C_c
11	Diphosphate	C_c	77	meso-2,6-Diaminoheptanedioate	C_c
12	NH3	C_c	78	beta-D-Glucose 6-phosphate	C_c
13	AMP	C_c	79	2-(alpha-Hydroxyethyl)thiamine diphosphate	C_c
14	Pyruvate	C_c	80	beta-D-Fructose 6-phosphate	C_c
15	Acetyl-CoA	C_c	81	beta-D-Fructose 1,6-bisphosphate	C_c
16	L-Glutamate	C_c	82	Oxalosuccinate	C_c
17	2-Oxoglutarate	C_c	83	Acetyl adenylate	C_c
18	Acetate	C_c	84	Isobutanol	C_c

Continued on next page

*Compartments: [C_c] = Intracellular (cytosol), [C_e] = Extracellular
†Flux (mmol gDW^{-1} h^{-1})
‡Biomass growth rate (1/h)
§Cellulosome secretion rate (g gDW^{-1} h^{-1})
¶Lumped reactions generated by *NetworkReducer* [161]

Table C.2 – (continued)

No.	Metabolites	Compartment	No.	Metabolites	Compartment
19	GDP	C_c	85	Sedoheptulose 7-phosphate	C_c
20	Oxaloacetate	C_c	86	Protein	C_c
21	Glycine	C_c	87	DNA	C_c
22	L-Alanine	C_c	88	RNA	C_c
23	Succinate	C_c	89	Lipid	C_c
24	GTP	C_c	90	Cell Wall	C_c
25	L-Lysine	C_c	91	Solute Pool	C_c
26	L-Aspartate	C_c	92	Biomass	C_c
27	Formate	C_c	93	Cell Dry Weight no GAM	C_c
28	L-Arginine	C_c	94	Total_LTA	C_c
29	L-Glutamine	C_c	95	Cellulosome term	C_c
30	L-Serine	C_c	96	Glycogen	C_c
31	Thiamin diphosphate	C_c	97	NH3	C_e
32	L-Methionine	C_c	98	CO2	C_e
33	Phosphoenolpyruvate	C_c	99	Cellobiose	C_e
34	L-Tryptophan	C_c	100	Ca2+	C_e
35	L-Phenylalanine	C_c	101	Mg2+	C_e
36	H+	C_c	102	K+	C_e
37	L-Tyrosine	C_c	103	Orthophosphate	C_e
38	Acetaldehyde	C_c	104	Fe3+	C_e
39	Succinyl-CoA	C_c	105	Nicotinate D-ribonucleotide	C_e
40	L-Cysteine	C_c	106	L-Arginine	C_e
41	D-Glucose 1-phosphate	C_c	107	L-Asparagine	C_e
42	Glycerone phosphate	C_c	108	L-Aspartate	C_e
43	D-Ribose 5-phosphate	C_c	109	L-Cysteine	C_e
44	D-Glyceraldehyde 3-phosphate	C_c	110	L-Glutamine	C_e
45	5-Phospho-alpha-D-ribose 1-diphosphate	C_c	111	Glycine	C_e
46	Fumarate	C_c	112	L-Histidine	C_e
47	L-Leucine	C_c	113	L-Leucine	C_e
48	L-Histidine	C_c	114	L-Methionine	C_e
49	Reduced ferredoxin	C_c	115	L-Phenylalanine	C_e
50	Oxidized ferredoxin	C_c	116	L-Proline	C_e
51	3-Methyl-2-oxobutanoic acid	C_c	117	L-Serine	C_e
52	L-Proline	C_c	118	L-Threonine	C_e
53	(S)-Malate	C_c	119	L-Tryptophan	C_e
54	L-Asparagine	C_c	120	L-Tyrosine	C_e
55	Citrate	C_c	121	L-Valine	C_e
56	L-Valine	C_c	122	Ethanol	C_e
57	Cellobiose	C_c	123	Acetate	C_e
58	(S)-Lactate	C_c	124	Formate	C_e
59	L-Threonine	C_c	125	(S)-Lactate	C_e
60	3-Phospho-D-glycerate	C_c	126	Hydrogen	C_e
61	D-Ribulose 5-phosphate	C_c	127	Pyruvate	C_e
62	beta-D-Glucose	C_c	128	Succinate	C_e
63	Acetyl phosphate	C_c	129	Biomass	C_e
64	D-Xylulose 5-phosphate	C_c	130	Isobutanol	C_e
65	3-Phospho-D-glyceroyl phosphate	C_c	131	H+	C_e
66	D-Glycerate	C_c			

C.2 EFV Lumping Scheme

The EFVs are lumped into primary EFV families by taking a weighted-average, where the weightage is given as [37]:

$$\left(Y_{i,j} + \varepsilon_j^{\mathrm{J}}\right)^{3n_{\mathrm{eta}}}; \quad i = \mathrm{csm, biom, atp}; \quad \mathrm{J} = \mathrm{I, II} \tag{C.1}$$

Here, the indexes j and J are used to denote the individual EFVs and the principal families of EFVs, respectively, and $Y_{i,j}$ is the yield of i-th metabolite of the j-th EFV. The weightage prioritizes the EFVs with greater yield of reference product in the final average, where the reference products for $\mathbf{F}_{\mathrm{I}}$, $\mathbf{F}_{\mathrm{II}}^{\mathrm{B}}$ and $\mathbf{F}_{\mathrm{II}}^{\mathrm{A}}$ are the cellulosome, biomass and ATP, respectively, on the basis of the grouping. The $\varepsilon_j^{\mathrm{J}}$ is the correction factor to adjust the flux distribution to the individual EFVs, whereas the n_{eta} controls the sensitivity of the adjustment. In this work, the n_{eta} is set to the default value of unity (cf. Eq. (T1) – (T3) in Figure 5.3) as there was adequate sensitivity to the adjustments. The correction factor is given as [37]:

$$\varepsilon_j^{\mathrm{J}} = \sum_i a_i^{\mathrm{J}} Y_{i,j}; \quad i = \mathrm{eth, lac, form, ace}; \quad \mathrm{J} = \mathrm{I, II} \tag{C.2}$$

Here, a_i^J are the tuning coefficients for metabolites of interest (i = ethanol, lactate, formate and acetate) that are estimated by minimizing the error between the model-estimated lumped yields (Eq. (T4) in Figure 5.3) and the steady-state experimental yield data from the literature. Table C.3 presents the optimal values of the tuning coefficients estimated in this work.

Table C.3: Estimated values of tuning coefficients, a_i^{J} (i = eth, lac, form, ace; J = I, II) and the corresponding model-estimated lumped yields ($Y_{i,\mathrm{J}}$) of metabolites secreted by *C. thermocellum* compared with experimental steady-state yield data. The yields are presented in mole basis of cellobiose. Here, the yields of biomass are evaluated by discounting the cellulase component from the yields of dry cell weights from the metabolic networks.

EFV family		Metabolites, i					
		eth†	lac†	form†	ace†	biom‡	csm‡
$\mathbf{F}_{\mathrm{I}}(\mathrm{N}_{20\%})$	a_i^{I}	–1.5441	1.5699	4.7301	−8.0586	n.a.*	n.a.
	$Y_{i,\mathrm{I}}$	1.2267	0.0269	0.5948	6.3832	0	51.8544
$\mathbf{F}_{\mathrm{II}}(\mathrm{N}_{20\%})$	a_i^{II}	–0.0020	–0.0079	−0.0091	0.0058	n.a.	n.a.
	$Y_{i,\mathrm{II}}$	0.5995	0.0239	0.6224	2.3739	21.8979	19.2621
$\mathbf{F}_{\mathrm{I}}(\mathrm{N}_{2\%})$	a_i^{I}	–4.4672	4.5776	13.7881	–23.7054	n.a.	n.a.
	$Y_{i,\mathrm{I}}$	1.1855	0.0223	0.6044	6.3454	0	50.6977
$\mathbf{F}_{\mathrm{II}}(\mathrm{N}_{2\%})$	a_i^{II}	3.6440	–2.6866	–9.4710	13.0979	n.a.	n.a.
	$Y_{i,\mathrm{II}}$	1.3920	0.0222	0.4708	5.9060	22.1441	7.9861
exp§	Y_i^{exp}	0.7040	0.0239	0.5424	1.4517	20.9276¶ 18.7429	2.3987¶ 13.1085

C.3 UC-PBM Equations

C.3.1 ML-PBM Equations

In the ML-PBM, the FPT is employed to discretize the PBEs over a discrete-continuous mesh. In this work, for Avicel cellulose population of relatively short polymer chains, $p = 20$ was used and q was constrained to a function dependent on p and N. Detailed

*n.a. - not applicable.
†Metabolites for which the tuning parameters are fitted to experimental yields.
‡Yields of biomass and cellulosome are expressed as g/mol, while the rest are mol/mol.
§Experimental steady-state yields are extracted from Holwerda et al. [172] and Zhang and Lynd [163].
¶Experimental steady-state yield when cellobiose is the sole carbon source.

information on the solution technique and the ML-PBM can be found elsewhere (Chapter 3/Appendix A and Chapter 4/Appendix B). The final implementable form of the ML-PBM equations are given here:

Soluble products:

$$\begin{aligned}\frac{dC_1}{dt} &= k_{\mathrm{h}}^{\mathrm{exo}} C_{\mathrm{B},3}^{\mathrm{exo}} + \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo}} + \sum_{j=4}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo-exo}} \\ &\quad + \sum_{j=3}^{6} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}'}^{\mathrm{endo}} C_{\mathrm{F}}^{\mathrm{endo}} C_j\end{aligned} \tag{C.3}$$

$$\begin{aligned}\frac{dC_2}{dt} &= k_{\mathrm{h}}^{\mathrm{exo}} C_{\mathrm{B},3}^{\mathrm{exo}} + 2k_{\mathrm{h}}^{\mathrm{exo}} C_{\mathrm{B},4}^{\mathrm{exo}} + \sum_{5}^{p+q} k_{\mathrm{h}}^{\mathrm{exo}} C_{\mathrm{B,j}}^{\mathrm{exo}} + \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo}} \\ &\quad + \sum_{j=4}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo-exo}} + \sum_{j=3}^{6} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}'}^{\mathrm{endo}} C_{\mathrm{F}}^{\mathrm{endo}} C_j - X \sum_{\mathrm{J}} r_{\mathrm{J}}\end{aligned} \tag{C.4}$$

Here, the molar concentrations of glucose and cellobiose are represented by $C_1 = C_{\mathrm{glu}}$ and $C_2 = C_{\mathrm{cb}}$, respectively.

$$\begin{aligned}\frac{dC_i}{dt} &= \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo}} + \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo-exo}} \\ &\quad + \sum_{j=i+1}^{6} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}'}^{\mathrm{endo}} C_{\mathrm{F}}^{\mathrm{endo}} C_j - k_{\mathrm{h}'}^{\mathrm{endo}} C_{\mathrm{F}}^{\mathrm{endo}} C_i; \quad i = 3, 4, 5\end{aligned} \tag{C.5}$$

$$\frac{dC_6}{dt} = \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo}} + \sum_{j=7}^{p+q} \left(\frac{2}{x_{j-1}}\right) k_{\mathrm{h}}^{\mathrm{endo}} C_{\mathrm{B},j}^{\mathrm{endo-exo}} - k_{\mathrm{h}'}^{\mathrm{endo}} C_{\mathrm{F}}^{\mathrm{endo}} C_6 \tag{C.6}$$

Exo-enzyme-polymer complex species:

$$\frac{dC_{\mathrm{B},3}^{\mathrm{exo}}}{dt} = k_{\mathrm{h}}^{\mathrm{exo}} C_{\mathrm{B},5}^{\mathrm{exo}} - k_{\mathrm{h}}^{\mathrm{exo}} C_{\mathrm{B},3}^{\mathrm{exo}} \tag{C.7}$$

$$\frac{dC^{\text{exo}}_{\text{B},i}}{dt} = k^{\text{exo}}_{\text{h}} C^{\text{exo}}_{\text{B},i+2} + k^{\text{endo}}_{\text{e}} C^{\text{endo-exo}}_{\text{B},i} - k^{\text{exo}}_{\text{h}} C^{\text{exo}}_{\text{B,i}} - k^{\text{endo}}_{\text{f}} C^{\text{endo}}_{\text{S}} C^{\text{endo}}_{\text{B},i}; \quad i = 4,5,6 \tag{C.8}$$

$$\begin{aligned}\frac{dC^{\text{exo}}_{\text{B},i}}{dt} &= k^{\text{exo}}_{\text{h}} C^{\text{exo}}_{\text{B},i+2} + k^{\text{endo}}_{\text{e}} C^{\text{endo-exo}}_{\text{B},i} + k^{\text{exo}}_{\text{f},i} C^{\text{exo}}_{\text{S}} C_{\text{S},i} - k^{\text{exo}}_{\text{h}} C^{\text{exo}}_{\text{B},i} - k^{\text{exo}}_{\text{e}} C^{\text{exo}}_{\text{B},i} \\ &\quad - k^{\text{endo}}_{\text{f}} C^{\text{endo}}_{\text{S}} C^{\text{exo}}_{\text{B},i}; \quad i = 7,8,\ldots,p-1\end{aligned} \tag{C.9}$$

$$\begin{aligned}\frac{dC^{\text{exo}}_{\text{B},i}}{dt} &= \sum_{j=i}^{p+q} \eta^{\text{exo}}_{ij} k^{\text{exo}}_{\text{h}} C^{\text{exo}}_{\text{B},j} + k^{\text{endo}}_{\text{e}} C^{\text{endo-exo}}_{\text{B},i} + k^{\text{exo}}_{\text{f},i} C^{\text{exo}}_{\text{S}} C_{\text{S},i} - k^{\text{exo}}_{\text{h}} C^{\text{exo}}_{\text{B},i} - k^{\text{exo}}_{\text{e}} C^{\text{exo}}_{\text{B},i} \\ &\quad - k^{\text{endo}}_{\text{f}} C^{\text{endo}}_{\text{S}} C^{\text{exo}}_{\text{B},i}; \quad i = p, p+1, \ldots, p+q\end{aligned} \tag{C.10}$$

Endo-enzyme-polymer complex species:

$$\frac{dC^{\text{endo}}_{\text{B},i}}{dt} = k^{\text{endo}}_{\text{f}} C^{\text{endo}}_{\text{S}} C_{\text{S},i} - k^{\text{endo}}_{\text{e}} C^{\text{endo}}_{\text{B},i} - k^{\text{endo}}_{\text{h}} C^{\text{endo}}_{\text{B},i}; \quad i = 7,8,\ldots,p+q \tag{C.11}$$

Endo-exo-enzyme-polymer complex species:

$$\frac{dC^{\text{endo-exo}}_{\text{B},i}}{dt} = k^{\text{endo}}_{\text{f}} C^{\text{endo}}_{\text{S}} C^{\text{exo}}_{\text{B},i} - k^{\text{endo}}_{\text{e}} C^{\text{endo-exo}}_{\text{B},i} - k^{\text{endo}}_{\text{h}} C^{\text{endo-exo}}_{\text{B},i}; \quad i = 4,5,\ldots,p+q \tag{C.12}$$

Surface-accessible polymers:

$$\frac{dC_{\text{S},i}}{dt} = C_{\text{loss},i} + r_{\text{exp},i}; \quad i = 7,8,\ldots,p+q \tag{C.13}$$

where

$$\begin{aligned}C_{\text{loss},i}(t) &= \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k^{\text{endo}}_{\text{h}} C^{\text{endo}}_{\text{B},j} + \sum_{j=i+1}^{p+q} \left(\frac{2}{x_{j-1}}\right) k^{\text{endo}}_{\text{h}} C^{\text{endo-exo}}_{\text{B},j} + k^{\text{exo}}_{\text{e}} C^{\text{exo}}_{\text{B},i} \\ &\quad + k^{\text{endo}}_{\text{e}} C^{\text{endo}}_{\text{B},i} - k^{\text{exo}}_{\text{f},i} C^{\text{exo}}_{\text{S}} C_{\text{S},i} - k^{\text{endo}}_{\text{f}} C^{\text{endo}}_{\text{S}} C_{\text{S},i}; \quad i = 7,8,\ldots,p\end{aligned} \tag{C.14}$$

$$\begin{aligned} C_{\text{loss},i}(t) = &\sum_{j=i+1}^{p+q} \eta_{ij}^{\text{endo}} k_{\text{h}}^{\text{endo}} C_{\text{B},j}^{\text{endo}} + \sum_{j=i+1}^{p+q} \eta_{ij}^{\text{endo}} k_{\text{h}}^{\text{endo}} C_{\text{B},j}^{\text{endo-exo}} + k_{\text{e}}^{\text{exo}} C_{\text{B},i}^{\text{exo}} \\ &+ k_{\text{e}}^{\text{endo}} C_{\text{B},i}^{\text{endo}} - k_{\text{f},i}^{\text{exo}} C_{\text{S}}^{\text{exo}} C_{\text{S},i} - k_{\text{f}}^{\text{endo}} C_{\text{S}}^{\text{endo}} C_{\text{S},i}; \\ &i = p+1, p+2, \ldots, p+q \end{aligned} \tag{C.15}$$

$$\begin{aligned} r_{\text{exp},i}(t) = -\left[\frac{R(t)-R_0}{R(t)}\right] &\left[\sum_{j=1}^{p+q} (162x_j + 18)\; r_{\text{loss},j}(t)\right] \psi_i(l); \\ &i = 7, 8, \ldots, p+q; \quad l = 1, 2, \ldots, l_{\max} - 1; \quad R(t) \geq R_0 \end{aligned} \tag{C.16}$$

$$r_{\text{loss},i}(t) = C_{\text{loss},i} + \frac{dC_{\text{B},i}^{\text{exo}}}{dt} + \frac{dC_{\text{B},i}^{\text{endo}}}{dt} + \frac{dC_{\text{B},i}^{\text{endo-exo}}}{dt}; \quad i = 3, 4, \ldots, p+q \tag{C.17}$$

$$\psi_i(l) = \frac{C_{\text{int},il}(0)}{\sum_{j=1}^{p+q} (162x_j + 18)\, C_{\text{int},jl}(0)}; \quad i = 7, 8, \ldots, p+q; \quad l = 1, 2, \ldots, l_{\max} - 1 \tag{C.18}$$

Here, $r_{\text{exp},i}(t) = 0$ for $R(t) < R_0$. For Eq. (C.16) and (C.18), the $l = \lceil R(t)/R_0 \rceil - 1$ where the ceiling operator $\lceil Z \rceil$ denotes the nearest higher integer to Z

Internal polymers:

$$\frac{dC_{\text{int},i}}{dt} = r_{\text{exp},i}; \quad i = 7, 8, \ldots, p+q \tag{C.19}$$

Surface-adsorbed enzymes:

$$\frac{dC_{\text{S}}^{\text{exo}}}{dt} = k_{\text{a}}^{\text{exo}} C_{\text{F}}^{\text{exo}} \left(\sigma_{\text{exo}} S(t) - C_{\text{S}}^{\text{exo}}\right) + \sum_{i=7}^{p+q} k_{\text{e}}^{\text{exo}} C_{\text{B},i}^{\text{exo}} - k_{\text{d}}^{\text{exo}} C_{\text{S}}^{\text{exo}} - \sum_{i=7}^{p+q} k_{\text{f},i}^{\text{exo}} C_{\text{S}}^{\text{exo}} C_{\text{S},i} \tag{C.20}$$

$$\begin{aligned} \frac{dC_{\text{S}}^{\text{endo}}}{dt} = &k_{\text{a}}^{\text{endo}} C_{\text{F}}^{\text{endo}} \left(\sigma_{\text{endo}} S(t) - C_{\text{S}}^{\text{endo}}\right) + \sum_{i=7}^{p+q} k_{\text{e}}^{\text{endo}} C_{\text{B},i}^{\text{endo}} + \sum_{i=4}^{p+q} k_{\text{e}}^{\text{endo}} C_{\text{B},i}^{\text{endo-exo}} \\ &- k_{\text{d}}^{\text{endo}} C_{\text{S}}^{\text{endo}} - \sum_{i=7}^{p+q} k_{\text{f}}^{\text{endo}} C_{\text{S}}^{\text{endo}} C_{\text{S},i} - \sum_{i=4}^{p+q} k_{\text{f}}^{\text{endo}} C_{\text{S}}^{\text{endo}} C_{\text{B},i}^{\text{exo}} \end{aligned} \tag{C.21}$$

where $S(t) = 2n\pi LR(t)$ and $R(t) = \left\{ \left[M_{\mathrm{T}} - \sum_{i=1}^{N} (162x_i + 18)\, C_i(t) \right] \Big/ n\rho\pi L \right\}^{1/2}$.

Free enzymes:

$$C_{\mathrm{F}}^{\mathrm{exo}} = C_{\mathrm{T}}^{\mathrm{exo}} - C_{\mathrm{S}}^{\mathrm{exo}} - \sum_{i=3}^{p+q} C_{\mathrm{B},i}^{\mathrm{exo}} - \sum_{i=7}^{p+q} C_{\mathrm{B},i}^{\mathrm{endo-exo}} \tag{C.22}$$

$$C_{\mathrm{F}}^{\mathrm{endo}} = C_{\mathrm{T}}^{\mathrm{endo}} - C_{\mathrm{S}}^{\mathrm{endo}} - \sum_{i=7}^{p+q} C_{\mathrm{B},i}^{\mathrm{endo}} - \sum_{i=7}^{p+q} C_{\mathrm{B},i}^{\mathrm{endo-exo}} \tag{C.23}$$

Total enzymes:

$$\begin{aligned} C_{\mathrm{T}}^{\mathrm{exo}}(t) &= \begin{cases} 10.97 C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}; \text{ when } \textit{C. thermocellum} \text{ grown on cellulose} \\ 9.39 C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}; \text{ when } \textit{C. thermocellum} \text{ grown on cellobiose} \end{cases} \\ C_{\mathrm{T}}^{\mathrm{endo}}(t) &= \begin{cases} 0.11 C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}; \text{ when } \textit{C. thermocellum} \text{ grown on cellulose} \\ 0.09 C_{\mathrm{csm}}(t)/M_{\mathrm{csm}}; \text{ when } \textit{C. thermocellum} \text{ grown on cellobiose} \end{cases} \end{aligned} \tag{C.24}$$

where, $M_{\mathrm{csm}} \approx 1.73$ MDa and 1.89 MDa are the molecular weights of the cellulosomes when *C. thermocellum* is grown on cellulose and cellobiose, respectively.

C.3.2 L-HCM Equations

Intracellular enzymes:

$$\frac{de_{\mathrm{J}}}{dt} = \alpha_{\mathrm{J}} + U_{\mathrm{J}} r_{\mathrm{e,J}}^{\mathrm{kin}} - (\beta_{\mathrm{J}} + \mu)\, e_{\mathrm{J}}; \quad \mathrm{J} = \mathrm{I}, \mathrm{II} \tag{C.25}$$

Cellobiose uptake fluxes:

$$r_{\mathrm{J}} = V_{\mathrm{J}} e_{\mathrm{J}}^{\mathrm{rel}} r_{\mathrm{J}}^{\mathrm{kin}}; \quad \mathrm{J} = \mathrm{I}, \mathrm{II} \tag{C.26}$$

Metabolites:

$$\frac{dC_i}{dt} = X \sum_{\mathrm{J=I}}^{\mathrm{II}} Y_{i,\mathrm{J}} r_{\mathrm{J}}; \quad i = \mathrm{eth, lac, form, ace}; \quad \mathrm{J = I, II} \tag{C.27}$$

Biomass:

$$\frac{dX}{dt} = Y_{\mathrm{biom,II}} r_{\mathrm{II}} X \tag{C.28}$$

Cellulosome:

$$\frac{dC_{\mathrm{csm}}}{dt} = \alpha_{\mathrm{csm}} X + X \sum_{\mathrm{J}} Y_{\mathrm{csm,J}} r_{\mathrm{J}} - \beta_{\mathrm{csm}} C_{\mathrm{csm}} \tag{C.29}$$

Here, the balance of intracellular enzymes that catalyse the uptake of cellobiose to each EFV family is given in Eq. (C.25), where α_{J} are the constitutive intracellular enzyme synthesis rates, U_{J} are the cybernetic variables that control the levels of e_{J}, $r_{\mathrm{e,J}}^{\mathrm{kin}}$ are the unregulated inducive intracellular enzyme synthesis rates and β_{J} are the enzyme decay rates. The biomass growth rate is given as $\mu = Y_{\mathrm{biom,II}} r_{\mathrm{II}}$ considering only $\mathbf{F}_{\mathrm{II}}$ contributes to biomass growth. Subsequently, the cellobiose uptake fluxes through the EFV families, facilitated by the enzymes e_{J} is expressed in Eq. (C.26), where $e_{\mathrm{J}}^{\mathrm{rel}} = e_{\mathrm{J}} / e_{\mathrm{J}}^{\mathrm{max}}$ are the relative intracellular enzyme levels, $e_{\mathrm{J}}^{\mathrm{max}}$ are the maximum intracellular enzyme levels and $r_{\mathrm{J}}^{\mathrm{kin}}$ are the unregulated cellobiose uptake fluxes. The derivation of the unregulated $r_{\mathrm{e,J}}^{\mathrm{kin}}$ and $r_{\mathrm{J}}^{\mathrm{kin}}$ rates, as well as the derivation of the initial and maximum intracellular enzyme levels necessary for implementation of the equations above are provided in Appendix C.4 and Appendix C.5, respectively.

C.4 Cellobiose Phosphorylation Kinetics

The unregulated rates of cellobiose uptake and inducive intracellular enzyme synthesis are derived using MM equation. Supported by vast majority of experimental findings in the literature [148, 174, 175, 190], the inhibitions of cellobiose phosphorylation by ethanol and glucose are also included:

$$r_{\mathrm{J}}^{\mathrm{kin}} = k_{\mathrm{J}}^{\max}\left(\frac{C_{\mathrm{cb}}}{K_{\mathrm{J}}+C_{\mathrm{cb}}}\right)\left(\frac{1}{1+C_{\mathrm{glu}}/K_{\mathrm{inh-glu}}}\right)\left(\frac{1}{1+C_{\mathrm{eth}}/K_{\mathrm{inh-eth}}}\right); \quad \mathrm{J}=\mathrm{I},\mathrm{II} \quad \text{(C.30)}$$

$$r_{\mathrm{e,J}}^{\mathrm{kin}} = k_{\mathrm{e,J}}\left(\frac{C_{\mathrm{cb}}}{K_{\mathrm{e,J}}+C_{\mathrm{cb}}}\right)\left(\frac{1}{1+C_{\mathrm{glu}}/K_{\mathrm{inh-glu}}}\right)\left(\frac{1}{1+C_{\mathrm{eth}}/K_{\mathrm{inh-eth}}}\right); \quad \mathrm{J}=\mathrm{I},\mathrm{II} \quad \text{(C.31)}$$

Here, $k_{\mathrm{J}}^{\max}$ and $k_{\mathrm{e,J}}$ are the maximum cellobiose uptake and enzyme synthesis rates, respectively, $K_{\mathrm{e,J}}$ and K_{J} are the MM-type saturation constants, $K_{\mathrm{inh-glu}}$ is the glucose inhibition constant, and $K_{\mathrm{inh-eth}}$ is the ethanol inhibition constant. As $k_{\mathrm{e,J}}$ and $K_{\mathrm{e,J}}$ exhibit minimal sensitivity to the model outputs in the parametric analysis, it is assumed that $k_{\mathrm{e,I}} = k_{\mathrm{e,II}}$ and $K_{\mathrm{e,J}} = K_{\mathrm{J}}$ for simplicity.

C.5 Initial and Maximum Intracellular Enzyme Levels

For ease of readability, the intracellular enzyme level balance is reproduced here:

$$\frac{d\mathbf{e}}{dt} = \boldsymbol{\alpha} + \mathbf{D}(\mathbf{U})\mathbf{r}_{\mathrm{e}}^{\mathrm{kin}} - [\mathbf{D}(\boldsymbol{\beta}) + \mu\mathbf{I}]\mathbf{e} \quad \text{(C.32)}$$

The maximum intracellular enzyme levels are derived by imposing a steady-state condition over Eq. (C.32), where U_{J} is at maximum value of unity, the carbon source is abundant ($C_{\mathrm{cb}} \gg K_{\mathrm{e,J}}$), inhibiting products are at the minimal level ($C_{\mathrm{glu}} \ll K_{\mathrm{inh-glu}}$, $C_{\mathrm{eth}} \ll K_{\mathrm{inh-eth}}$), and the biomass growth rate is solely contributed by the J-th family ($\mu = \mu_{\mathrm{J}}$). Imposing the above, $r_{\mathrm{e,J}}^{\mathrm{kin}} \approx k_{\mathrm{e,J}}$, and thus:

$$e_{\mathrm{J}}^{\max} = \frac{\alpha_{\mathrm{J}} + k_{\mathrm{e,J}}}{\beta_{\mathrm{J}} + \mu_{\mathrm{J}}}; \quad \mathrm{J} = \mathrm{I}, \mathrm{II} \tag{C.33}$$

Here, $\mu_{\mathrm{I}} = 0$ as $\mathbf{F}_{\mathrm{I}}$ does not contribute to biomass growth and $\mu_{\mathrm{II}} = Y_{\mathrm{biom,II}} r_{\mathrm{II}} \approx Y_{\mathrm{biom,II}} k_{\mathrm{II}}^{\max}$.

Subsequently, the initial intracellular enzyme levels are calculated based on the pre-culture history of the microorganism. Again, a steady-state condition is imposed over Eq. (C.32), where no species other than the pre-culture carbon source exists:

$$e_{\mathrm{J}}(0) = \frac{\alpha_{\mathrm{J}} + U_{\mathrm{J}} k_{\mathrm{e,J}} \left(\dfrac{C_{\mathrm{cb-pc}}}{K_{\mathrm{e,J}} + C_{\mathrm{cb-pc}}} \right)}{\beta_{\mathrm{J}} + \mu_{\mathrm{J}}}; \quad \mathrm{J} = \mathrm{I}, \mathrm{II} \tag{C.34}$$

Here, $\mu_{\mathrm{I}} = 0$, whereas $\mu_{\mathrm{II}} = Y_{\mathrm{biom,II}} r_{\mathrm{II}} \approx Y_{\mathrm{biom,II}} k_{\mathrm{II}}^{\max} (C_{\mathrm{cb-pc}} / K_{\mathrm{II}} + C_{\mathrm{cb-pc}})$, $C_{\mathrm{cb-pc}}$ is the pre-culture molar concentration of cellobiose and the cybernetic variable U_{J} is chosen according to the pre-culture history:

$$\mathbf{U} = \begin{cases} [1, 0]; & \text{if pre} - \text{culture substrate is pure cellulose} \\ [0, 1]; & \text{if pre} - \text{culture substrate is pure cellobiose} \end{cases} \tag{C.35}$$

Assuming that the pre-culture is at the incipient log phase, the condition above is schemed such that the $\mathbf{F}_{\mathrm{I}}$ is fully activated in the presence of pure cellulose to allow rapid secretion of cellulosomes to facilitate the saccharification or such that $\mathbf{F}_{\mathrm{II}}$ is fully activated to allow direct biomass growth in the presence of pure fermentable cellobiose. All the experimental

data employed in this study [163, 172] uses pure cellulose (Avicel) as the pre-culture substrate for *C. thermocellum*.

C.6 Summary of Power Law Forms of the ML-PBM Rate Coefficients

Power law form for the ML-PBM rate coefficients representing various enzyme-substrate interactions should be chosen according to the characteristics of the specific cellulases in question. As the complexed cellulase concoction secreted by *C. thermocellum* is vastly different than the non-complex fungal cellulases (which was used as the basis for the ML-PBM development), the power law form of the ML-PBM rate coefficients are modified accordingly, as shown in Table C.4.

Table C.4: Summary of rate coefficients and corresponding power law form.

Rate coefficients	Power law form	Remarks
$k_{\mathrm{a}}^{\mathrm{m}}, k_{\mathrm{d}}^{\mathrm{m}}$	Constant	m = exo, endo
$k_{\mathrm{f}}^{\mathrm{exo}}$	$k_{\mathrm{f}}^{\mathrm{exo}}(v) = k_{\mathrm{p,f}}^{\mathrm{exo}}/v$	Longer time is needed to locate the chain-ends of longer chains [43].
$k_{\mathrm{h}}^{\mathrm{exo}}$ $k_{\mathrm{f}}^{\mathrm{endo}}$ $k_{\mathrm{h}}^{\mathrm{endo}}$ $k_{\mathrm{h}'}^{\mathrm{endo}}$	Constant	Unlike fungal enzymes, hydrolysis of cellulose by cellulases of *C. thermocellum* is unaffected by cellulose structural properties like cellulose DP, crystallinity and accessible surface area [117]. Also, *C. thermocellum* is observed to completely solubilize various cellulosic substrates with little to no hindrance [148, 163, 172].

C.7 Model Parameter Estimation

The model parameters were estimated by minimizing the objective function in the form of a weighted sum of squares of residuals:

$$J_{\text{opt}} = \sum_{i} W_i \left\{ \left(\widehat{\mathbf{y}}_i - \mathbf{y}_i \right)^{\text{T}} \left(\widehat{\mathbf{y}}_i - \mathbf{y}_i \right) \right\} \tag{C.36}$$

Here, i denotes the data sets, $\widehat{\mathbf{y}}_i$ is the vector of model-predicted output, $\mathbf{y}_i$ is the vector of experimental data and W_i is the weightage allocated to each data set. The sum of squares of residuals are normalized by setting the weightage W_i to the inverse of the order of magnitude formed at the initial setting. The minimization of the objective function was performed using the Genetic Algorithm of MATLAB® R2016a on a workstation equipped with an Intel® Core™ i5-7200U processor with a clock speed of 2.50 GHz and 16 GB of installed memory (RAM). The ODEs are solved using the "ode15s" sub-routine over a time span that covers the experimental data.

Appendix D

MATLAB Codes

D.1 Simulation of Chain-End Monomer Scission

Main file:

```
close all; clear all; clc;

% General pre-simulation preparations

% Distribution settings

N = 22496;                          % Max DP
Ms_t0 = 10;                         % Initial mass concentration, g/L
Mn = 4100;                          % Number-average DP
Mw = 5430;                          % Weight-average DP
alpha = Mn/(Mw-Mn);
beta = Mw-Mn;

vm = 1; % Size of monomer

% Discrete-continuous mesh

p = 100;                            % No. of pivots in discrete region
qmax = 1+(log(N/(p+1))/log(1+(vm/(p+1))));
```

```
20 q = floor(qmax);                 % No. of pivots in continuous region
21 ratio = (N/(p+1))^(1/(q-1));     % Ratio of geometric progression
22 % ratio = (N-(p+1))/(q-1);       % Common diff of arithmetic progression
23
24 % Pivots for discrete-continuous mesh
25
26 x_piv = zeros(p+q,1);
27
28 x_piv(1:p) = 1:p;
29
30 for i = p+1:p+q
31     x_piv(i) = (p+1)*ratio^(i-(p+1));       % Geometric mesh
32 %     x_piv(i) = (p+1)+(ratio*(i-(p+1)));   % Arithmetic mesh
33 end
34
35 % Boundary points for discrete-continuous mesh
36
37 x_bound = zeros(p+q+1,1);
38
39 x_bound(1) = 0.5;
40
41 for i=2:p+q
42     x_bound(i) = (x_piv(i)+x_piv(i-1))/2;
43 end
44
45 x_bound(p+q+1)=x_piv(p+q)+(x_piv(p+q)-x_bound(p+q));
46
47 % Fully discrete mesh
48
49 DP = 1:N;
50
51 % Initial distribution
52
53 P = @(x) ((((x-1)./beta).^(alpha-1)).*exp(-(x-1)./beta))...
54     ./(beta*gamma(alpha)).*(((x.*162)+18)-180);
55
56 denom = integral(P,1,N);
57
58 c_in = @(x) (Ms_t0*((((x-1)./beta).^(alpha-1)).*exp(-(x-1)./beta))...
59     ./(beta*gamma(alpha)))./denom;
60
61 % Initial molar concentrations - sectional techniques
62
63 C_in_mol = zeros(p+q,1);
64
65 for i=1:p+q
66     C_in_mol(i) = c_in(x_piv(i))*(x_bound(i+1)-x_bound(i));
67 end
68
69 % Initial molar concentrations - fully discrete
70
71 c_in_mol_disc = zeros(N,1);
72
73 for i = 1:N;
74    c_in_mol_disc(i) = c_in(i);
75 end
76
77 % Rate kernel
78
```

```
kp = 1;                 % Rate constant
m = 1;                  % m = 0 (constant), m = 1 (linear), m = 2 (quadratic)
k_sec = kp*(x_piv.^m);
k_disc = kp*(DP'.^m);

% Solving ODEs

tstart = 0;
tfinal = 20;
tstep = 0.1;
tgrid = tstart:tstep:tfinal;

[t,C_fpt,C_cat,C_fvs,c_disc] = ChainEnd...
    (p,q,N,vm,x_piv,x_bound,k_sec,k_disc,C_in_mol,c_in_mol_disc,tgrid);

% Post-processing

% Molar concentration density

c_fpt = zeros(size(C_fpt));
c_fvs = zeros(size(C_fvs));
c_cat = zeros(size(C_cat));

for j = 1:length(x_piv)
    c_fpt(:,j) = C_fpt(:,j)/(x_bound(j+1)-x_bound(j));
    c_cat(:,j) = C_cat(:,j)/(x_bound(j+1)-x_bound(j));
    c_fvs(:,j) = C_fvs(:,j)/(x_bound(j+1)-x_bound(j));
end

% Moments

mom0_fpt = zeros(length(t),1);
mom0_cat = zeros(length(t),1);
mom0_fvs = zeros(length(t),1);
mom0_disc = zeros(length(t),1);

mom1_fpt = zeros(length(t),1);
mom1_cat = zeros(length(t),1);
mom1_fvs = zeros(length(t),1);
mom1_disc = zeros(length(t),1);

mom2_fpt = zeros(length(t),1);
mom2_cat = zeros(length(t),1);
mom2_fvs = zeros(length(t),1);
mom2_disc = zeros(length(t),1);

for i = 1:length(t)
    mom0_fpt(i) = sum(C_fpt(i,:));
    mom1_fpt(i) = sum(x_piv.*C_fpt(i,:)');
    mom2_fpt(i) = sum((x_piv.^2).*C_fpt(i,:)');

    mom0_cat(i) = sum(C_cat(i,:));
    mom1_cat(i) = sum(x_piv.*C_cat(i,:)');
    mom2_cat(i) = sum((x_piv.^2).*C_cat(i,:)');

    mom0_fvs(i) = sum(C_fvs(i,:));
    mom1_fvs(i) = sum(x_piv.*C_fvs(i,:)');
    mom2_fvs(i) = sum((x_piv.^2).*C_fvs(i,:)');

```

```
138     mom0_disc(i) = sum(c_disc(i,:));
139     mom1_disc(i) = sum(DP'.*c_disc(i,:)');
140     mom2_disc(i) = sum((DP'.^2).*c_disc(i,:)');
141 end
142 
143 % Normalized moments
144 
145 norm_mom0_fpt = mom0_fpt./sum(C_fpt(1,:));
146 norm_mom0_cat = mom0_cat./sum(C_cat(1,:));
147 norm_mom0_fvs = mom0_fvs./sum(C_fvs(1,:));
148 norm_mom0_disc = mom0_disc./sum(c_disc(1,:));
149 
150 norm_mom1_fpt = mom1_fpt./sum(x_piv.*C_fpt(1,:)');
151 norm_mom1_cat = mom1_cat./sum(x_piv.*C_cat(1,:)');
152 norm_mom1_fvs = mom1_fvs./sum(x_piv.*C_fvs(1,:)');
153 norm_mom1_disc = mom1_disc./sum(DP'.*c_disc(1,:)');
154 
155 norm_mom2_fpt = mom2_fpt./sum((x_piv.^2).*C_fpt(1,:)');
156 norm_mom2_cat = mom2_cat./sum((x_piv.^2).*C_cat(1,:)');
157 norm_mom2_fvs = mom2_fvs./sum((x_piv.^2).*C_fvs(1,:)');
158 norm_mom2_disc = mom2_disc./sum((DP'.^2).*c_disc(1,:)');
159 
160 % Moments error
161 
162 error_mom0_fpt = zeros(length(t),1);
163 error_mom1_fpt = zeros(length(t),1);
164 error_mom2_fpt = zeros(length(t),1);
165 
166 error_mom0_cat = zeros(length(t),1);
167 error_mom1_cat = zeros(length(t),1);
168 error_mom2_cat = zeros(length(t),1);
169 
170 error_mom0_fvs = zeros(length(t),1);
171 error_mom1_fvs = zeros(length(t),1);
172 error_mom2_fvs = zeros(length(t),1);
173 
174 for i = 1:length(t)
175     error_mom0_fpt(i) = max(abs((mom0_disc(i)-mom0_fpt(i))/mom0_disc(i)));
176     error_mom1_fpt(i) = max(abs((mom1_disc(i)-mom1_fpt(i))/mom1_disc(i)));
177     error_mom2_fpt(i) = max(abs((mom2_disc(i)-mom2_fpt(i))/mom2_disc(i)));
178 
179     error_mom0_cat(i) = max(abs((mom0_disc(i)-mom0_cat(i))/mom0_disc(i)));
180     error_mom1_cat(i) = max(abs((mom1_disc(i)-mom1_cat(i))/mom1_disc(i)));
181     error_mom2_cat(i) = max(abs((mom2_disc(i)-mom2_cat(i))/mom2_disc(i)));
182 
183     error_mom0_fvs(i) = max(abs((mom0_disc(i)-mom0_fvs(i))/mom0_disc(i)));
184     error_mom1_fvs(i) = max(abs((mom1_disc(i)-mom1_fvs(i))/mom1_disc(i)));
185     error_mom2_fvs(i) = max(abs((mom2_disc(i)-mom2_fvs(i))/mom2_disc(i)));
186 end
187 
188 % Error in initial mass due to discretization
189 
190 E_d = (sum(C_in_mol.*((162*x_piv)+18))-Ms_t0)/Ms_t0;
191 
192 % Non-dimensionalization
193 
194 % ft = find(round(c_disc(:,1),3)==round(0.99*max(c_disc(:,1)),3));
195 % t99 = t(ft(1));           % Time when monomer reaches 99% of total monomers
196 t99 = 5;
```

```
td = t./t99;                % Dimensionless time

N0 = sum(c_in_mol_disc);
x0 = sum(c_in_mol_disc.*DP')/sum(c_in_mol_disc);

c_fpt_d = c_fpt.*(x0/N0);   % Dimensionless population density
c_cat_d = c_cat.*(x0/N0);
c_fvs_d = c_fvs.*(x0/N0);
c_disc_d = c_disc.*(x0/N0);
```

Function file:

```
function [t,C_fpt,C_cat,C_fvs,c_disc] = ChainEnd...
    (p,q,N,vm,x_piv,x_bound,k_sec,k_disc,C_in_mol,c_in_mol_disc,tgrid)

[t,C_fpt] = fpt_chain_end(p,q,vm,x_piv,x_bound,k_sec,C_in_mol,tgrid);
[t,C_cat] = cat_chain_end(p,q,vm,x_piv,x_bound,k_sec,C_in_mol,tgrid);
[t,C_fvs] = fvs_chain_end(p,q,x_piv,x_bound,k_sec,C_in_mol,tgrid);
[t,c_disc] = disc_chain_end(N,k_disc,c_in_mol_disc,tgrid);

end

%%-------------------------------------------------------------------------

function [t,C_fpt] = fpt_chain_end(p,q,vm,x_piv,x_bound,k_sec,C_in_mol,...
    tgrid)

% Fixed pivot technique

% Particle allocation function

n = zeros(p+q,p+q);

n(1,2) = 2; n(1,3:end) = 1;

for i = 2:p+q
    for j = i:p+q
        if j==i && x_piv(i)-vm>x_piv(i-1)
            n(i,j) = 1-(vm/(x_piv(i)-x_piv(i-1)));
        elseif j==i && x_piv(i)-vm==x_piv(i-1)
            n(i,j) = 0;
        elseif j~=i && x_piv(j)-vm>x_piv(i) && x_piv(j)-vm<x_piv(i+1)
            n(i,j) = vm/(x_piv(i+1)-x_piv(i));
        elseif j~=i && x_piv(j)-vm==x_piv(i)
            n(i,j) = 1;
        end
    end
    n(i,i) = n(i,i)-1;
end

% ODE Solver

tic;

```

```
43  options_sec = odeset('AbsTol',1e-16,'RelTol',1e-6,'NonNegative',1:(p+q),...
44      'Refine',8);
45
46  [t,C_fpt] = ode15s(@(t,C) fpt_ce_ode(p,q,k_sec,n,C),...
47      tgrid,C_in_mol,options_sec);
48
49  time_sim = '\nTotal time elapsed for FPT simulation is %.1f seconds.\n';
50
51  fprintf(time_sim,toc)
52
53  end
54
55  %%-----------------------------------------------------------------------
56
57  function diff_func = fpt_ce_ode(p,q,k,n,C)
58
59  dCdt = zeros(length(n(:,1)),1);
60
61  dCdt(1) = n(1,2)*k(2)*C(2)+sum(n(1,3:end)'.*k(3:(end)).*C(3:(end)));
62
63  for i=2:p+q;
64      dCdt(i) = sum(n(i,:)'.*k.*C);
65  end
66
67  % Output
68
69  diff_func = dCdt;
70
71  end
72
73  %%-----------------------------------------------------------------------
74
75  function [t,C_cat] = cat_chain_end(p,q,vm,x_piv,x_bound,k_sec,C_in_mol,...
76      tgrid)
77
78  % Cell average technique
79
80  % Bi matrix - number of particles entering i
81
82  Bi = zeros(p+q,p+q);
83  Bi(1,2) = 2; Bi(1,3:end) = 1;
84
85  for i = 2:p+q-1
86      for j = i+1:p+q
87          if x_piv(i)-vm>x_bound(i)
88              Bi(i,i) = 1;
89          end
90          if x_piv(j)-1>x_bound(i) && x_piv(j)-1<x_bound(i+1)
91              Bi(i,j) = 1;
92          end
93      end
94  end
95
96  if x_piv(p+q)-vm>x_bound(p+q)
97      Bi(p+q,p+q) = 1;
98  end
99
100 % Vi matrix - mass of particles entering i
101
```

```
Vi = zeros(p+q,p+q);

for i = 2:p+q-1
    for j = i+1:p+q
        if x_piv(i)-vm>x_bound(i)
            Vi(i,i) = x_piv(i)-vm;
        end
        if x_piv(j)-vm>x_bound(i) && x_piv(j)-vm<x_bound(i+1)
            Vi(i,j) = x_piv(j)-vm;
        end
    end
end

if x_piv(p+q)-vm>x_bound(p+q)
    Vi(p+q,p+q) = x_piv(p+q)-1;
end

% Average volume, v_i

v_i = zeros(p+q,1);

for i = 1:p+q
    if sum(Bi(i,:))==0||sum(Vi(i,:))==0
        v_i(i) = 0;
    else
        v_i(i) = sum(Vi(i,:))/sum(Bi(i,:));
    end
end

% ODE Solver

tic;

options_sec = odeset('AbsTol',1e-16,'RelTol',1e-6,'NonNegative',1:(p+q),...
    'Refine',8);

[t,C_cat] = ode15s(@(t,C) cat_ce_ode(p,q,Bi,v_i,k_sec,C,x_piv),...
    tgrid,C_in_mol,options_sec);

time_sim = '\nTotal time elapsed for CAT simulation is %.1f seconds.\n';

fprintf(time_sim,toc)

end

%%-------------------------------------------------------------------------

function diff_func = cat_ce_ode(p,q,Bi,v_i,k_sec,C,x_piv)

dCdt = zeros(p+q,1);

dCdt(1) = 2*k_sec(2)*C(2)+sum(k_sec(3:end).*C(3:end));

for i = 2:p-1
    dCdt(i) = sum(Bi(i-1,:)'.*k_sec.*C).*(forw(v_i(i-1),x_piv,i)...
        *heav(v_i(i-1),x_piv(i-1)))+...
        sum(Bi(i,:)'.*k_sec.*C).*(forw(v_i(i),x_piv,i)...
        *heav(x_piv(i),v_i(i)))+...
        sum(Bi(i,:)'.*k_sec.*C).*(back(v_i(i),x_piv,i)...
```

```
161         *heav(v_i(i),x_piv(i)))+...
162         sum(Bi(i+1,:)'.*k_sec.*C).*(back(v_i(i+1),x_piv,i)...
163         *heav(x_piv(i+1),v_i(i+1)))-...
164         k_sec(i)*C(i);
165 end
166
167 for i = p
168     dCdt(i) = sum(Bi(i-1,:)'.*k_sec.*C).*(forw(v_i(i-1),x_piv,i)...
169         *heav(v_i(i-1),x_piv(i-1)))+...
170         sum(Bi(i,:)'.*k_sec.*C).*(forw(v_i(i),x_piv,i)...
171         *heav(x_piv(i),v_i(i)))+...
172         sum(Bi(i,:)'.*k_sec.*C).*(back(v_i(i),x_piv,i)...
173         *heav(v_i(i),x_piv(i)))+...
174         sum(Bi(i+1,:)'.*k_sec.*C).*(back(v_i(i+1),x_piv,i)...
175         *heav(x_piv(i+1),v_i(i+1)))-...
176         k_sec(i)*C(i);
177 end
178
179 for i = p+1
180     dCdt(i) = sum(Bi(i-1,:)'.*k_sec.*C).*(forw(v_i(i-1),x_piv,i)...
181         *heav(v_i(i-1),x_piv(i-1)))+...
182         sum(Bi(i,:)'.*k_sec.*C).*(forw(v_i(i),x_piv,i)...
183         *heav(x_piv(i),v_i(i)))+...
184         sum(Bi(i,:)'.*k_sec.*C).*(back(v_i(i),x_piv,i)...
185         *heav(v_i(i),x_piv(i)))+...
186         sum(Bi(i+1,:)'.*k_sec.*C).*(back(v_i(i+1),x_piv,i)...
187         *heav(x_piv(i+1),v_i(i+1)))-...
188         k_sec(i)*C(i);
189 end
190
191 for i = p+2:p+q-1
192     dCdt(i) = sum(Bi(i-1,:)'.*k_sec.*C).*(forw(v_i(i-1),x_piv,i)...
193         *heav(v_i(i-1),x_piv(i-1)))+...
194         sum(Bi(i,:)'.*k_sec.*C).*(forw(v_i(i),x_piv,i)...
195         *heav(x_piv(i),v_i(i)))+...
196         sum(Bi(i,:)'.*k_sec.*C).*(back(v_i(i),x_piv,i)...
197         *heav(v_i(i),x_piv(i)))+...
198         sum(Bi(i+1,:)'.*k_sec.*C).*(back(v_i(i+1),x_piv,i)...
199         *heav(x_piv(i+1),v_i(i+1)))-...
200         k_sec(i)*C(i);
201 end
202
203 dCdt(p+q) = sum(Bi(p+q-1,:)'.*k_sec.*C).*(forw(v_i(p+q-1),x_piv,p+q)...
204     *heav(v_i(p+q-1),x_piv(p+q-1)))+...
205         sum(Bi(p+q,:)'.*k_sec.*C).*(forw(v_i(p+q),x_piv,p+q)...
206         *heav(x_piv(p+q),v_i(p+q)))-...
207         k_sec(p+q)*C(p+q);
208
209 % Output
210
211 diff_func = dCdt;
212
213 end
214
215 %%-------------------------------------------------------------------------
216
217 function heav_func = heav(v,x)
218
219 if v-x > 0
```

```
220     output = 1;
221 elseif v-x == 0
222     output = 0.5;
223 else
224     output = 0;
225 end
226
227 % Output
228
229 heav_func = output;
230
231 end
232
233 %%-------------------------------------------------------------------------
234
235 function forward = forw(v,x,i)
236
237 % Output
238
239 forward = (v-x(i-1))/(x(i)-x(i-1));
240
241 end
242
243 %%-------------------------------------------------------------------------
244
245 function backward = back(v,x,i)
246
247 % Output
248
249 backward = (v-x(i+1))/(x(i)-x(i+1));
250
251 end
252
253 %%-------------------------------------------------------------------------
254
255 function [t,C_fvs] = fvs_chain_end(p,q,x_piv,x_bound,k_sec,C_in_mol,tgrid)
256
257 % Finite volume scheme
258
259 % Integrals, p_ij = p_ji
260
261 int_1 = zeros(p+q,p+q);
262 int_1(1,2:end) = 1;
263
264 for i = 1:p+q
265     for j = i:p+q
266         if i==j
267             if x_piv(i)-1>x_bound(i)
268                 int_1(i,j) = 1;
269             end
270         else
271             if x_piv(j)-1>x_bound(i) && x_piv(j)-1<x_bound(i+1)
272                 int_1(i,j) = 1;
273             end
274         end
275     end
276 end
277
278 % Birth weight
```

```
279
280 w_b = zeros(p+q,1);
281
282 for i = 1:p+q
283     j = 1:i;
284     mat = int_1(1:i,i);
285     xxi = zeros(length(j),1); xxi(:) = x_piv(i);
286     w_b(i) = (x_piv(i)*(sum(mat(:))-1))/sum((xxi-x_piv(j))'*mat);
287 end
288
289 % Death weight
290
291 w_d = zeros(p+q,1);
292
293 for i = 1:p+q
294     j = 1:i;
295     mat = int_1(1:i,i);
296     w_d(i) = (w_b(i)*sum(x_piv(j)'*mat))/x_piv(i);
297 end
298
299 w_b(1) = 0; w_d(1) = 0;
300
301 % ODE solver
302
303 tic;
304
305 options_sec = odeset('AbsTol',1e-16,'RelTol',1e-6,'NonNegative',1:(p+q),...
306     'Refine',8);
307
308 [t,C_fvs] = ode15s(@(t,C) fvs_ce_ode(p,q,k_sec,C,w_b,int_1,w_d),...
309     tgrid,C_in_mol,options_sec);
310
311 time_sim = '\nTotal time elapsed for FVS simulation is %.1f seconds.\n';
312
313 fprintf(time_sim,toc)
314
315 end
316
317 %%-------------------------------------------------------------------------
318
319 function diff_func = fvs_ce_ode(p,q,k_sec,C,w_b,int_1,w_d)
320
321 dCdt = zeros(p+q,1);
322
323 dCdt(1) = 2*k_sec(2)*C(2)+sum(k_sec(3:end).*C(3:end));
324
325 for i = 2:p+q
326     dCdt(i) = sum(w_b(i:end).*k_sec(i:end).*C(i:end).*int_1(i,i:end)')...
327         -w_d(i)*k_sec(i)*C(i);
328 end
329
330 % Output
331
332 diff_func = dCdt;
333
334 end
335
336 %%-------------------------------------------------------------------------
337
```

```
338 function [t,c_disc] = disc_chain_end(N,k_disc,c_in_mol_disc,tgrid)
339
340 % Fully discrete solution
341
342 % Jacobian matrix
343
344 jacobian = zeros(N,N);
345 jacobian(1,2) = 2*k_disc(2);
346 jacobian(1,3:end) = k_disc(3:end);
347 jacobian(end,end) = -k_disc(N);
348
349 for i = 2:N-1
350     jacobian(i,i) = -k_disc(i);
351     jacobian(i,i+1) = k_disc(i+1);
352 end
353
354 jacobian = sparse(jacobian);
355
356 % ODE solver
357
358 tic;
359
360 options_disc = odeset('AbsTol',1e-16,'RelTol',1e-6,'NonNegative',1:N,...
361     'Refine',8,'jacobian',jacobian);
362
363 [t,c_disc] = ode15s(@(t,C) disc_ce_ode(N,C,k_disc),tgrid,...
364     c_in_mol_disc,options_disc);
365
366 time_sim =...
367    '\nTotal time elapsed for fully discrete simulation is %.1f seconds.\n';
368
369 fprintf(time_sim,toc)
370
371 end
372
373 %%-------------------------------------------------------------------------
374
375 function diff_func = disc_ce_ode(N,C,k_disc)
376
377 dCdt = zeros(N,1);
378
379 dCdt(1)= 2*k_disc(2)*C(2)+sum(k_disc(3:N).*C(3:N));
380
381 for i = 2:(N-1);
382     dCdt(i) = k_disc(i+1)*C(i+1)-k_disc(i)*C(i);
383 end
384
385 dCdt(N) = -k_disc(N)*C(N);
386
387 % Output
388 diff_func = dCdt;
389
390 end
```

D.2 Simulation of Random Scission

Main file:

```
close all; clear all; clc;

% General pre-simulation preparations

% Distribution settings

N = 22496;                  % Max DP
Ms_t0 = 10;                 % Initial mass concentration in g/L
Mn = 4100;                  % Number-average DP
Mw = 5430;                  % Weight-average DP
alpha = Mn/(Mw-Mn);
beta = Mw-Mn;

% Discrete-continuous mesh

p = 10;                         % No. of pivots in discrete region
q = 50;                         % No. of pivots in continuous region
ratio = (N/(p+1))^(1/(q-1));    % Ratio of geometric progression
% ratio = (N-(p+1))/(q-1);      % Common diff of arithmetic progression

vm = 1; % Size of monomer

% Pivots for discrete-continuous mesh

x_piv = zeros(p+q,1);

x_piv(1:p) = 1:p;

for i = p+1:p+q
    x_piv(i) = (p+1)*ratio^(i-(p+1));       % Geometric mesh
%     x_piv(i) = (p+1)+(ratio*(i-(p+1)));   % Arithmetic mesh
end

% Boundary points for discrete-continuous mesh

x_bound = zeros(p+q+1,1);

x_bound(1) = 0.5;

for i=2:p+q
    x_bound(i) = (x_piv(i)+x_piv(i-1))/2;
end

x_bound(p+q+1)=x_piv(p+q)+(x_piv(p+q)-x_bound(p+q));

% Fully discrete mesh

DP = 1:N;
```

```
49
50  % Initial distribution
51
52  P = @(x) ((((x-1)./beta).^(alpha-1)).*exp(-(x-1)./beta))...
53      ./(beta*gamma(alpha)).*(((x.*162)+18)-180);
54
55  denom = integral(P,1,N);
56
57  c_in = @(x) (Ms_t0*((((x-1)./beta).^(alpha-1)).*exp(-(x-1)./beta))...
58      ./(beta*gamma(alpha)))./denom;
59
60  % Initial molar concentrations - sectional techniques
61
62  C_in_mol = zeros(p+q,1);
63
64  for i=1:p+q
65      C_in_mol(i) = c_in(x_piv(i))*(x_bound(i+1)-x_bound(i));
66  end
67
68  % Initial molar concentrations - fully discrete
69
70  c_in_mol_disc = zeros(N,1);
71
72  for i = 1:N;
73     c_in_mol_disc(i) = c_in(i);
74  end
75
76  % Rate kernel
77
78  kp = 1;                % Rate constant
79  m = 1;                 % m = 0 (constant), m = 1 (linear), m = 2 (quadratic)
80  k_sec = kp*(x_piv.^m);
81  k_disc = kp*(DP'.^m);
82
83  % Solving ODEs
84
85  tstart = 0;
86  tfinal = 10;
87  tstep = 0.1;
88  tgrid = tstart:tstep:tfinal;
89
90  [t,C_fpt,C_cat,C_fvs,c_disc] = RandomScission...
91      (p,q,N,x_piv,x_bound,k_sec,k_disc,C_in_mol,c_in_mol_disc,tgrid);
92
93  % Post-processing
94
95  % Molar concentration density
96
97  c_fpt = zeros(size(C_fpt));
98  c_fvs = zeros(size(C_fvs));
99  c_cat = zeros(size(C_cat));
100
101 for j = 1:length(x_piv)
102     c_fpt(:,j) = C_fpt(:,j)/(x_bound(j+1)-x_bound(j));
103     c_cat(:,j) = C_cat(:,j)/(x_bound(j+1)-x_bound(j));
104     c_fvs(:,j) = C_fvs(:,j)/(x_bound(j+1)-x_bound(j));
105 end
106
107 % Moments
```

```
108
109 mom0_fpt = zeros(length(t),1);
110 mom0_cat = zeros(length(t),1);
111 mom0_fvs = zeros(length(t),1);
112 mom0_disc = zeros(length(t),1);
113
114 mom1_fpt = zeros(length(t),1);
115 mom1_cat = zeros(length(t),1);
116 mom1_fvs = zeros(length(t),1);
117 mom1_disc = zeros(length(t),1);
118
119 mom2_fpt = zeros(length(t),1);
120 mom2_cat = zeros(length(t),1);
121 mom2_fvs = zeros(length(t),1);
122 mom2_disc = zeros(length(t),1);
123
124 for i = 1:length(t)
125     mom0_fpt(i) = sum(C_fpt(i,:));
126     mom1_fpt(i) = sum(x_piv.*C_fpt(i,:)');
127     mom2_fpt(i) = sum((x_piv.^2).*C_fpt(i,:)');
128
129     mom0_cat(i) = sum(C_cat(i,:));
130     mom1_cat(i) = sum(x_piv.*C_cat(i,:)');
131     mom2_cat(i) = sum((x_piv.^2).*C_cat(i,:)');
132
133     mom0_fvs(i) = sum(C_fvs(i,:));
134     mom1_fvs(i) = sum(x_piv.*C_fvs(i,:)');
135     mom2_fvs(i) = sum((x_piv.^2).*C_fvs(i,:)');
136
137     mom0_disc(i) = sum(c_disc(i,:));
138     mom1_disc(i) = sum(DP'.*c_disc(i,:)');
139     mom2_disc(i) = sum((DP'.^2).*c_disc(i,:)');
140 end
141
142 % Normalized moments
143
144 norm_mom0_fpt = mom0_fpt./sum(C_fpt(1,:));
145 norm_mom0_cat = mom0_cat./sum(C_cat(1,:));
146 norm_mom0_fvs = mom0_fvs./sum(C_fvs(1,:));
147 norm_mom0_disc = mom0_disc./sum(c_disc(1,:));
148
149 norm_mom1_fpt = mom1_fpt./sum(x_piv.*C_fpt(1,:)');
150 norm_mom1_cat = mom1_cat./sum(x_piv.*C_cat(1,:)');
151 norm_mom1_fvs = mom1_fvs./sum(x_piv.*C_fvs(1,:)');
152 norm_mom1_disc = mom1_disc./sum(DP'.*c_disc(1,:)');
153
154 norm_mom2_fpt = mom2_fpt./sum((x_piv.^2).*C_fpt(1,:)');
155 norm_mom2_cat = mom2_cat./sum((x_piv.^2).*C_cat(1,:)');
156 norm_mom2_fvs = mom2_fvs./sum((x_piv.^2).*C_fvs(1,:)');
157 norm_mom2_disc = mom2_disc./sum((DP'.^2).*c_disc(1,:)');
158
159 % Moments error
160
161 error_mom0_fpt = zeros(length(t),1);
162 error_mom1_fpt = zeros(length(t),1);
163 error_mom2_fpt = zeros(length(t),1);
164
165 error_mom0_cat = zeros(length(t),1);
166 error_mom1_cat = zeros(length(t),1);
```

```
error_mom2_cat = zeros(length(t),1);

error_mom0_fvs = zeros(length(t),1);
error_mom1_fvs = zeros(length(t),1);
error_mom2_fvs = zeros(length(t),1);

for i = 1:length(t)
    error_mom0_fpt(i) = max(abs((mom0_disc(i)-mom0_fpt(i))/mom0_disc(i)));
    error_mom1_fpt(i) = max(abs((mom1_disc(i)-mom1_fpt(i))/mom1_disc(i)));
    error_mom2_fpt(i) = max(abs((mom2_disc(i)-mom2_fpt(i))/mom2_disc(i)));

    error_mom0_cat(i) = max(abs((mom0_disc(i)-mom0_cat(i))/mom0_disc(i)));
    error_mom1_cat(i) = max(abs((mom1_disc(i)-mom1_cat(i))/mom1_disc(i)));
    error_mom2_cat(i) = max(abs((mom2_disc(i)-mom2_cat(i))/mom2_disc(i)));

    error_mom0_fvs(i) = max(abs((mom0_disc(i)-mom0_fvs(i))/mom0_disc(i)));
    error_mom1_fvs(i) = max(abs((mom1_disc(i)-mom1_fvs(i))/mom1_disc(i)));
    error_mom2_fvs(i) = max(abs((mom2_disc(i)-mom2_fvs(i))/mom2_disc(i)));
end

% Error in initial mass due to discretization

E_d = (sum(C_in_mol.*((162*x_piv)+18))-Ms_t0)/Ms_t0;

% Non-dimensionalization

% ft = find(round(c_disc(:,1),3)==round(0.99*max(c_disc(:,1)),3));
% t99 = t(ft(1));           % Time when monomer reaches 99% of total monomers
t99 = 3.1;
td = t./t99;                % Dimensionless time

N0 = sum(c_in_mol_disc);
x0 = sum(c_in_mol_disc.*DP')/sum(c_in_mol_disc);

c_fpt_d = c_fpt.*(x0/N0);   % Dimensionless population density
c_cat_d = c_cat.*(x0/N0);
c_fvs_d = c_fvs.*(x0/N0);
c_disc_d = c_disc.*(x0/N0);
```

Function file:

```
function [t,C_fpt,C_cat,C_fvs,c_disc] = RandomScission...
    (p,q,N,x_piv,x_bound,k_sec,k_disc,C_in_mol,c_in_mol_disc,tgrid)

[t,C_fpt] = fpt_rand(p,q,x_piv,k_sec,C_in_mol,tgrid);
[t,C_cat] = cat_rand(p,q,x_piv,x_bound,k_sec,C_in_mol,tgrid);
[t,C_fvs] = fvs_rand(p,q,x_piv,x_bound,k_sec,C_in_mol,tgrid);
[t,c_disc] = disc_rand(N,k_disc,c_in_mol_disc,tgrid);

end

%%-------------------------------------------------------------------------

function [t,C_fpt] = fpt_rand(p,q,x_piv,k_sec,C_in_mol,tgrid)
```

```
14
15 % Fixed pivot technique
16
17 % Particle allocation function
18
19 n = zeros(p+q,p+q);
20
21 for j = 2:p+q
22     n(1,j) = 2/(x_piv(j-1));
23 end
24
25 for i = 2:p
26     for j = i+1:p+q
27         n(i,i) = -1;
28         n(i,j) = 2/(x_piv(j-1));
29     end
30 end
31
32 for i = p+1:p+q
33     for j = i+1:p+q
34         n(i,j) = (x_piv(i+1)-x_piv(i-1))/(x_piv(j-1));
35     end
36     n(i,i) = -1;
37 end
38
39 % ODE Solver
40
41 tic;
42
43 options_sec = odeset('AbsTol',1e-16,'RelTol',1e-6,'NonNegative',1:(p+q),...
44     'Refine',8);
45
46 [t,C_fpt] = ode15s(@(t,C) fpt_rand_ode(p,q,k_sec,n,C),...
47     tgrid,C_in_mol,options_sec);
48
49 time_sim = '\nTotal time elapsed for FPT simulation is %.1f seconds.\n';
50
51 fprintf(time_sim,toc)
52
53 end
54
55 %%-------------------------------------------------------------------------
56
57 function diff_func = fpt_rand_ode(p,q,k_sec,n,C)
58
59 dCdt = zeros(length(n(:,1)),1);
60
61 for i = 1:p+q;
62     dCdt(i) = sum(n(i,:)'.*k_sec(:).*C(:));
63 end
64
65 % Output
66
67 diff_func = dCdt;
68
69 end
70
71 %%-------------------------------------------------------------------------
72
```

```
function [t,C_cat] = cat_rand(p,q,x_piv,x_bound,k_sec,C_in_mol,tgrid)

% Cell average technique

% Bi matrix — number of particles entering i

Bi = zeros(p+q,p+q);

for i = 1:p
    for j = i+1:p+q
        Bi(i,j) = 2/(x_piv(j-1));
    end
end

for i = p+1:p+q
    for j = i+1:p+q
        Bi(i,j) = (2/(x_piv(j-1)))*(x_bound(i+1)-x_bound(i));
    end
end

% Vi matrix — mass of particles entering i

Vi = zeros(p+q,p+q);

for i = 1:p
    for j = i+1:p+q
        Vi(i,j) = (2*x_piv(i))/(x_piv(j-1));
    end
end

for i = p+1:p+q
    for j = i+1:p+q
        Vi(i,j) = ((x_bound(i+1)^2)-(x_bound(i)^2))/(x_piv(j-1));
    end
end

% Average volume, v_i

v_i = zeros(p+q,1);

for i = 1:p+q
    if sum(Vi(i,:))==0||sum(Bi(i,:))==0
        v_i(i) = 0;
    else
        v_i(i) = sum(Vi(i,:))/sum(Bi(i,:));
    end
end

% ODE Solver

tic;

options_sec = odeset('AbsTol',1e-16,'RelTol',1e-6,'NonNegative',1:(p+q),...
    'Refine',8);

[t,C_cat] = ode15s(@(t,C) cat_rand_ode(p,q,Bi,v_i,k_sec,C,x_piv),...
    tgrid,C_in_mol,options_sec);

time_sim = '\nTotal time elapsed for CAT simulation is %.1f seconds.\n';
```

```
132
133 fprintf(time_sim,toc)
134
135 end
136
137 %%-------------------------------------------------------------------------
138
139 function diff_func = cat_rand_ode(p,q,Bi,v_i,k_sec,C,x_piv)
140
141 dCdt = zeros(p+q,1);
142
143 for i = 1
144     dCdt(1) = sum(Bi(i,:)'.*k_sec.*C);
145 end
146
147 for i = 2:p+q-1
148     dCdt(i) = sum(Bi(i-1,:)'.*k_sec.*C).*(forw(v_i(i-1),x_piv,i)...
149         *heav(v_i(i-1),x_piv(i-1)))+...
150         sum(Bi(i,:)'.*k_sec.*C).*(forw(v_i(i),x_piv,i)...
151         *heav(x_piv(i),v_i(i)))+...
152         sum(Bi(i,:)'.*k_sec.*C).*(back(v_i(i),x_piv,i)...
153         *heav(v_i(i),x_piv(i)))+...
154         sum(Bi(i+1,:)'.*k_sec.*C).*(back(v_i(i+1),x_piv,i)...
155         *heav(x_piv(i+1),v_i(i+1)))-...
156         k_sec(i)*C(i);
157 end
158
159 dCdt(p+q) = sum(Bi(p+q-1,:)'.*k_sec.*C).*(forw(v_i(p+q-1),x_piv,p+q)...
160     *heav(v_i(p+q-1),x_piv(p+q-1)))+...
161         sum(Bi(p+q,:)'.*k_sec.*C).*(forw(v_i(p+q),x_piv,p+q)...
162         *heav(x_piv(p+q),v_i(p+q)))-...
163         k_sec(p+q)*C(p+q);
164
165 % Output
166
167 diff_func = dCdt;
168
169 end
170
171 %%-------------------------------------------------------------------------
172
173 function heav_func = heav(v,x)
174
175 if v-x > 0
176     output = 1;
177 elseif v-x == 0
178     output = 0.5;
179 else
180     output = 0;
181 end
182
183 % Output
184
185 heav_func = output;
186
187 end
188
189 %%-------------------------------------------------------------------------
190
```

```
191 function forward = forw(v,x,i)
192
193 % Output
194
195 forward = (v-x(i-1))/(x(i)-x(i-1));
196
197 end
198
199 %%-------------------------------------------------------------------------
200
201 function backward = back(v,x,i)
202
203 % Output
204
205 backward = (v-x(i+1))/(x(i)-x(i+1));
206
207 end
208
209 %%-------------------------------------------------------------------------
210
211 function [t,C_fvs] = fvs_rand(p,q,x_piv,x_bound,k_sec,C_in_mol,tgrid)
212
213 % Finite volume scheme
214
215 % Integrals, p_ij = p_ji
216
217 int_1 = zeros(p+q,p+q);
218
219 for i = 1:p
220     for j = i+1:p+q
221         int_1(i,j) = 2/(x_piv(j-1));
222     end
223 end
224
225 for i = p+1:p+q
226     for j = i+1:p+q
227         int_1(i,j) = (2*(x_bound(i+1)-x_bound(i)))/(x_piv(j-1));
228     end
229 end
230
231 int_1(1,1) = 0;
232
233 % Birth weight
234
235 w_b = zeros(p+q,1);
236
237 for i = 1:p+q
238     j = 1:i;
239     mat = int_1(1:i,i);
240     xxi = zeros(length(j),1); xxi(:) = x_piv(i);
241     w_b(i) = (x_piv(i)*(sum(mat(:))-1))/sum((xxi-x_piv(j))'*mat);
242 end
243
244 % Death weight
245
246 w_d = zeros(p+q,1);
247
248 for i = 1:p+q
249     j = 1:i;
```

```
250     mat = int_1(1:i,i);
251     w_d(i) = (w_b(i)*sum(x_piv(j)'*mat))/x_piv(i);
252 end
253
254 w_b(1) = 0; w_d(1) = 0;
255
256 % Particle allocation function for discrete region
257
258 n = zeros(p+q,p+q);
259
260 for j = 2:p+q
261     n(1,j) = 2/(x_piv(j-1));
262 end
263
264 for i = 2:p
265     for j = i+1:p+q
266         n(i,j) = 2/(x_piv(j-1));
267         n(i,i) = -1;
268     end
269 end
270
271 % ODE solver
272
273 tic;
274
275 options_sec = odeset('AbsTol',1e-16,'RelTol',1e-6,'NonNegative',1:(p+q),...
276     'Refine',8);
277
278 [t,C_fvs] = ode15s(@(t,C) fvs_rand_ode(p,q,n,k_sec,C,w_b,int_1,w_d),...
279     tgrid,C_in_mol,options_sec);
280
281 time_sim = '\nTotal time elapsed for FVS simulation is %.1f seconds.\n';
282
283 fprintf(time_sim,toc)
284
285 end
286
287 %%-------------------------------------------------------------------------
288
289 function diff_func = fvs_rand_ode(p,q,n,k_sec,C,w_b,int_1,w_d)
290
291 dCdt = zeros(p+q,1);
292
293 dCdt(1) = sum(n(1,:)'.*k_sec.*C);
294
295 for i = 2:p+q;
296     dCdt(i) = sum(w_b(i:end).*k_sec(i:end).*C(i:end).*int_1(i,i:end)')...
297         -w_d(i)*k_sec(i)*C(i);
298 end
299
300 % Output
301
302 diff_func = dCdt;
303
304 end
305
306 %%-------------------------------------------------------------------------
307
308 function [t,c_disc] = disc_rand(N,k_disc,c_in_mol_disc,tgrid)
```

```
309
310 % Fully discrete solution
311
312 % Partitioning of matrices for solving large ODE system
313
314 n = zeros(N,N);
315
316 for j = 2:N
317     n(1,j) = (2/(j-1))*k_disc(j);
318 end
319
320 for i = 2:N-1
321     for j = i+1:N
322         n(i,i) = -k_disc(i);
323         n(i,j) = (2/(j-1))*k_disc(j);
324     end
325 end
326
327 n(end,end) = -k_disc(N);
328
329 div = 19;       % No. of sub-matrices
330 mat = N/div;    % Square matrix size = 1184
331
332 G = cell(div,div);
333
334 for i = 1:div
335     G{i,i} = zeros(mat,mat);
336 end
337
338 G{1,1} = n(1:mat,1:mat);
339
340 for i = 2:div
341     ix = (i-1)*mat+1;
342     G{i,i} = n(ix:ix+mat-1,ix:ix+mat-1);
343 end
344
345 F = cell(div,div);
346
347 for i = 1:div
348     for j = i+1:div
349         ix = (i-1)*mat+1;
350         ij = (j-1)*mat+1;
351         F{i,j} = zeros(mat,mat);
352         F{i,j} = n(ix:ix+mat-1,ij:ij+mat-1);
353     end
354 end
355
356 % Jpattern
357
358 jp = triu(ones(mat,mat));
359
360 % ODE solver
361
362 tt = [];
363 c = [];
364
365 tic;
366
367 options_disc = odeset('RelTol',1e-6,'AbsTol',1e-16,'NonNegative',1:mat,...
```

```
368      'Refine',8,'jpattern',jp);
369
370  for i = 1:div
371      ii = div-(i-1);
372      low_end = N-(mat*i)+1;
373      high_end = low_end+mat-1;
374
375      FF = [];
376      if ii<div;
377          for k = ii+1:div;
378              FF = [FF F{ii,k}];
379          end
380      end
381
382      [t2,c_exact] = ode15s(@(t2,c_exact) disc_rand_ode(G{ii,ii},FF,c,...
383          c_exact,mat,tt,t2,ii,div),[tgrid(1) tgrid(end)],...
384          c_in_mol_disc(low_end:high_end),options_disc);
385
386      tt{ii} = t2;
387      c{ii} = c_exact;
388  end
389
390  % Overall number density
391
392  c_disc = [];
393
394  for ii = 1:div
395      cc = interp1(tt{ii},c{ii},tgrid,'pchip');
396      c_disc = [c_disc cc];
397  end
398
399  t = tgrid;
400
401  time_sim =...
402     '\nTotal time elapsed for fully discrete simulation is %.1f seconds.\n';
403
404  fprintf(time_sim,toc)
405
406  end
407
408  %%-------------------------------------------------------------------------
409
410  function diff_func = disc_rand_ode(G,FF,c,c_exact,mat,tt,t2,ii,div)
411
412  dCdt = zeros(mat,1);
413
414  if ii<div
415      c_exact_full = [];
416
417      for i = ii+1:div
418          c_exact_int = interp1(tt{i},c{i},t2,'linear');
419          c_exact_full = [c_exact_full c_exact_int];
420      end
421
422      dCdt = G*c_exact+FF*c_exact_full';
423  else
424      dCdt = G*c_exact;
425  end
426
```

```
427  % Output
428
429  diff_func = dCdt;
430
431  end
```

D.3 Simulation of Enzymatic Hydrolysis of Cellulose (ML-PBM)

Main file:

```
1   close all; clear all; clc;
2
3   % Setting input parameters
4
5   % Cellulose and enzymes
6
7   mss = 10;           % Initial concentration of cellulose, g/L
8
9   L_CBH = 11.2;       % CBH concentration, mg enzyme/g cellulose
10  L_EG = 0;           % EG concentration, mg enzyme/g cellulose
11  L_BG = 0;           % BG concentration, mg enzyme/g cellulose
12
13  MW_CBHi = 70000;    % CBH molecular weight, g/mol
14  MW_EGi = 50000;     % EG molecular weight, g/mol
15  MW_BGi = 135000;    % BG molecular weight, g/mol
16
17  % Mesh setting
18
19  p = 20;             % Number of discrete pivots
20
21  % Time span
22
23  t_end = 1.8e5;      % Time span of simulation, seconds
24
25  % Initial distribution - Base fitted cellulose distribution.
26
27  load('Raw_Avicel_calibrated_initialDist_Engel_2012_final','MnL','MwL',...
28      'MnH','MwH','N','r_massL');
29
30  % List of model paramaters
31
32  % CBH-related parameters
33
34  % 1  kp_h_CBH - Rate contant of CBH hydrolysis, 1/s
35  % 2  kp_f_CBH - Rate constant of CBH complexation, DP.L/mol.s
```

```
% 3  kp_e_CBH — Rate constant of CBH decomplexation, 1/s
% 4  tau_CBH — CBH enzyme footprint/area coverage, mol/m2
% 5  k_ads_CBH — Rate constant of CBH adsorption, L/mol.s
% 6  k_des_CBH — Rate constant of CBH desoprtion, 1/s
% 7  k_If_CBH(1) — Forward glucose inhibition rate constant, L/mol.s
% 8  k_If_CBH(2) — Forward cellobiose inhibition rate constant, L/mol.s
% 9  k_Ir_CBH(1) — Reverse glucose inhibition rate constant, 1/s
% 10 k_Ir_CBH(2) — Reverse cellobiose inhibition rate constant, 1/s

% Substrate—related parameters

% 11 pnt_zone — number of layers in penetration zone
% 12 r_massL — mass ratio of cellulose in penetration zone

% BG—related parameters

% 13 k_h_BG — Rate constant of BG hydrolysis, L/mol.s
% 14 k_If_BG — Forward glucose inhibition rate constant, L/mol.s
% 15 k_Ir_BG — Reverse glucose inhibition rate constant, 1/s

% EG—related parameters

% 16 kp_h_EG — Rate constant of EG hydrolysis (insoluble cellulose), DP/s
% 17 kp_hs_EG — Rate constant of EG hydrolysis (soluble cellulose), L/mol.s
% 18 kp_f_EG — Rate constant of EG complexation, DP.L/mol.s
% 19 kp_e_EG — Rate constant of EG decomplexation, 1/s
% 20 tau_EG — EG enzyme footprint/area coverage, mol/m2
% 21 k_ads_EG — Rate constant of EG adsorption, L/mol.s
% 22 k_des_EG — Rate constant of EG desorption, 1/s
% 23 k_If_EG(1) — Forward glucose inhibition rate constant, L/mol.s
% 24 k_If_EG(2) — Forward cellobiose inhibition rate constant, L/mol.s
% 25 k_Ir_EG(1) — Reverse glucose inhibition rate constant, 1/s
% 26 k_Ir_EG(2) — Reverse cellobiose inhibition rate constant, 1/s

% Input the values for model parameters below:

prm(1) = 4.491;
prm(2) = 4.32e4;
prm(3) = 10;
prm(4) = 2.22e—8;
prm(5) = 7.3722e5;
prm(6) = 0.01;
prm(7) = 0;
prm(8) = 0;
prm(9) = 0;
prm(10) = 0;

prm(11) = 1;
prm(12) = 0.028;

prm(13) = 3550;
prm(14) = 0;
prm(15) = 0;

prm(16) = 21;
prm(17) = 3;
prm(18) = 4e5;
prm(19) = 10;
prm(20) = 8.04e—9;
```

```
prm(21) = 7.11e5;
prm(22) = 0.01;
prm(23) = 0;
prm(24) = 0;
prm(25) = 0;
prm(26) = 0;

% Run simulation

[x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,pdiT,C,CXS,CNS,CNXS,CFS,CS,CI,...
    CT,CC,E_S_EG,E_S_CBH,E_F_EG,E_F_CBH,E_F_BG,R,conv_rem,t] = MLPBM...
    (N,p,prm,mss,MnL,MwL,MnH,MwH,L_CBH,L_EG,L_BG,MW_CBHi,MW_EGi,MW_BGi,...
    t_end);

% List of Outputs

% x_piv     % DP mesh
% R_in      % Initial particle radius
% layers    % Total no. of layers in cellulose particles
% C_T       % Total initial cellulose distribution
% C_S       % Surface initial cellulose distribution, mol/L
% C_INT     % Internal initial cellulose distribution, mol/L
% MnT       % Overall initial number-average DP
% MwT       % Overall initial weight-average DP
% pdiT      % Overall initial polydispersity index

% t         % Time, s

% Time,DP-dependent variables

% C         % Soluble products, mol/L
% CXS       % CBH-bound complex, mol/L
% CNS       % EG-bound complex, mol/L
% CNXS      % CBH-EG-bound complex, mol/L
% CFS       % Free un-bound surface polymers, mol/L
% CS        % Total surface polymers, mol/L
% CI        % Internal polymers, mol/L
% CT        % Total polymers, mol/L
% CC        % Insoluble polymers, mol/L

% Time-dependent variables

% E_S_EG    % Surface-adsorbed EG, mol/L
% E_S_CBH   % Surface-adsorbed CBH, mol/L
% E_F_EG    % Free EG, mol/L
% E_F_CBH   % Free CBH, mol/L
% E_F_BG    % Free BG, mol/L

% R         % Transient of particle radius, m

% conv_rem  % Overall conversion
```

Function file:

```
function [x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,pdiT,C,CXS,CNS,CNXS,...
    CFS,CS,CI,CT,CC,E_S_EG,E_S_CBH,E_F_EG,E_F_CBH,E_F_BG,R,conv_rem,t]...
    = MLPBM(N,p,prm,mss,MnL,MwL,MnH,MwH,L_CBH,L_EG,L_BG,MW_CBHi,MW_EGi,...
    MW_BGi,t_end)

[x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,pdiT,q,vm,R0,ro,L,n,exp_ratio,...
    E_T_CBH,E_T_EG,E_T_BG] = MLPBM_setting(p,N,prm,mss,MnL,MwL,MnH,MwH,...
    L_CBH,L_EG,L_BG,MW_CBHi,MW_EGi,MW_BGi);

[t,y] = ODE_setting(p,q,C_S,C_INT,prm,x_piv,vm,t_end,E_T_EG,E_T_CBH,...
    E_T_BG,R0,ro,L,n,exp_ratio);

[C,CXS,CNS,CNXS,CFS,CS,CI,CT,CC,E_S_EG,E_S_CBH,E_F_EG,E_F_CBH,E_F_BG,R,...
    conv_rem] = post_processing(t,y,p,q,x_piv,ro,L,n,mss,E_T_EG,E_T_CBH,...
    E_T_BG);

end

%%-------------------------------------------------------------------------

function [x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,pdiT,q,vm,R0,ro,L,n,...
    exp_ratio,E_T_CBH,E_T_EG,E_T_BG] = MLPBM_setting(p,N,prm,mss,MnL,MwL,...
    MnH,MwH,L_CBH,L_EG,L_BG,MW_CBHi,MW_EGi,MW_BGi)

% General settings

vm = 1;   % Size of monomer

qmax = 1+(log(N/(p+1))/log(1+(vm/(p+1))));  % Maximum pivots in cont region
q = floor(qmax);                            % No. of pivots in cont region
ratio = (N/(p+1))^(1/(q-1));                % Ratio of geometric progr

% Pivots for discrete-continuous mesh

x_piv = zeros(p+q,1);
x_piv(1:p) = 1:p;

for i = p+1:p+q
    x_piv(i) = (p+1)*ratio^(i-(p+1));       % Geometric mesh
end

% Boundary points for discrete-continuous mesh

x_bound = zeros(p+q+1,1);
x_bound(1) = 0.5;

for i=2:p+q
    x_bound(i) = (x_piv(i)+x_piv(i-1))/2;
end

x_bound(p+q+1) = x_piv(end);

% Initial distribution

```

```
55  c_in = @(x,alpha,beta,pin) pin*gampdf(x,alpha,beta); % Gamma distribution
56
57  % Physical properties
58
59  layers = round(2/(((mss*prm(12))/prm(11))/mss));     % Total no. of layers
60  ro = 1500;                    % Cellulose density, g/L
61  L = (N/2)*10.38e-10;          % Length of microfibril, m
62  R0 = 1e-9;                    % Width of single layer, m
63  R_in = layers*R0;             % Total radius of particle, m
64  n = mss/(ro*pi*(R_in^2)*L);   % No. of particles per unit vol, 1/m3
65  p_zone = prm(11);             % No. of layers in penetration zone
66  i_zone = (R_in/R0)-p_zone;    % No. of layers in internal zone
67
68  % Radiuses of discrete layers
69
70  R_set = sort(0:R0:R_in,'descend')';
71
72  % Mass of polymers in each layer
73
74  mass_l = zeros(length(R_set)-1,1);
75
76  for i = 2:length(R_set)
77      mass_l(i-1) = (n*ro*pi*(R_set(i-1)^2)*L)-(n*ro*pi*(R_set(i)^2)*L);
78  end
79
80  % Initial cellulose distribution in each layer
81
82  pin_l = zeros(length(mass_l),1);
83  c_in_tl = zeros(length(mass_l),p+q); % Molar concentration density
84  C_in_tl = zeros(length(mass_l),p+q); % Molar concentration
85
86  optionsfmincon = optimoptions('fsolve','display','none');
87
88  for i = 1:p_zone
89      pin_l(i) = fsolve(@(x) sum(c_in(x_piv,MnL/(MwL-MnL),MwL-MnL,x)...
90          .*(x_bound(2:end)-x_bound(1:p+q)).*(162*x_piv+18))-mass_l(i),0,...
91          optionsfmincon);
92      c_in_tl(i,:) = c_in(x_piv,MnL/(MwL-MnL),MwL-MnL,pin_l(i));
93  end
94
95  for i = p_zone+1:p_zone+i_zone
96      pin_l(i) = fsolve(@(x) sum(c_in(x_piv,MnH/(MwH-MnH),MwH-MnH,x)...
97          .*(x_bound(2:end)-x_bound(1:p+q)).*(162*x_piv+18))-mass_l(i),0,...
98          optionsfmincon);
99      c_in_tl(i,:) = c_in(x_piv,MnH/(MwH-MnH),MwH-MnH,pin_l(i));
100 end
101
102 for i = 1:p+q
103     C_in_tl(:,i) = c_in_tl(:,i).*(x_bound(i+1)-x_bound(i));
104 end
105
106 % Overall initial cellulose distribution
107
108 C_T = zeros(p+q,1);
109
110 for i = 1:p+q
111     C_T(i) = sum(C_in_tl(:,i));
112 end
113
```

```
114 % Surface initial cellulose distribution
115
116 C_S = zeros(p+q,1);
117
118 for i = 1:p+q
119     C_S(i) = sum(C_in_tl(1,i));
120 end
121
122 % Internal initial cellulose distribution
123
124 C_INT = zeros(p+q,1);
125
126 for i = 1:p+q
127     C_INT(i) = sum(C_in_tl(2:end,i));
128 end
129
130 % Ratio of molar concentration to total mass in each layer
131
132 C_ratio = zeros(length(R_set)-1,p+q);
133
134 for i = 1:length(R_set)-1
135     for j = 1:p+q
136         C_ratio(i,j) = C_in_tl(i,j)/sum(C_in_tl(i,:)'.*(162*x_piv+18));
137     end
138 end
139
140 % Piece-wise function for ratio of molar concentration to total mass as a
141 % function of particle radius
142
143 exp_ratio = cell(p+q,1);
144
145 for i = 1:p+q
146     exp_ratio{i} = mkpp(flipud(R_set),flipud(C_ratio(:,i)));
147 end
148
149 % General properties of initial cellulose distribution
150
151 MnT = sum(C_T.*x_piv)/sum(C_T);             % Overall number-averaged DP
152 MwT = sum(C_T.*(x_piv.^2))/sum(C_T.*x_piv); % Overall mass-averaged DP
153 pdiT = MwT/MnT;                             % Overall poly-dispersity index
154
155 % Enzyme loadings, mol/L
156
157 E_T_CBH = ((L_CBH/1000)*sum(C_T.*(162*x_piv+18)))/MW_CBHi;
158 E_T_EG = ((L_EG/1000)*sum(C_T.*(162*x_piv+18)))/MW_EGi;
159 E_T_BG = ((L_BG/1000)*sum(C_T.*(162*x_piv+18)))/MW_BGi;
160
161 end
162
163 %%-------------------------------------------------------------------------
164
165 function [t,y] = ODE_setting(p,q,C_S,C_INT,prm,x_piv,vm,t_end,E_T_EG,...
166     E_T_CBH,E_T_BG,R0,ro,L,n,exp_ratio)
167
168 % Initial conditions
169
170 init = zeros(6*p+6*q+9,1);
171 init(1:p+q) = 0;                    % 1 to p+q --> free liberated polymers
172 init(p+q+1:2*p+2*q) = 0;            % p+q+1 to 2p+2q --> CBH-bound polymers
```

```
173 init(2*p+2*q+1:3*p+3*q) = 0;        % 2p+2q+1 to 3p+3q ---> EG-bound polymers
174 init(3*p+3*q+1:4*p+4*q) = 0;        % 3p+3q+1 to 4p+4q ---> EG-CBH-bound polymers
175 init(4*p+4*q+1:5*p+5*q) = C_S;       % 4p+4q+1 to 5p+5q ---> surface-accessible polymer
176 init(5*p+5*q+1:6*p+6*q) = C_INT;     % 5p+5q+1 to 6p+6q ---> internal inaccessible polymer
177 init(6*p+6*q+1) = 0;                % 6p+6q+1 ---> surface adsorbed EG
178 init(6*p+6*q+2) = 0;                % 6p+6q+2 ---> surface adsorbed CBH
179 init(6*p+6*q+3:6*p+6*q+4) = 0;       % 6p+6q+3 to 6p+6q+4 ---> inhibited EG
180 init(6*p+6*q+5:6*p+6*q+6) = 0;       % 6p+6q+5 to 6p+6q+6 ---> inhibited free CBH
181 init(6*p+6*q+7:6*p+6*q+8) = 0;       % 6p+6q+7 to 6p+6q+8 ---> inhibited adsorbed CBH
182 init(6*p+6*q+9) = 0;                % 6p+6q+9 ---> inhibited BG
183
184 % Rate kernels
185
186 % CBH
187
188 kp_h_CBH = prm(1);                    % Rate constant of hydrolysis, 1/s
189 m_h_CBH = 0;
190 k_h_CBH = kp_h_CBH*(x_piv.^m_h_CBH);    % Rate coefficients of hydrolysis, 1/s
191
192 kp_f_CBH = prm(2);                    % Rate constant of complexation, DP.L/mol.s
193 m_f_CBH = -1;
194 k_f_CBH = kp_f_CBH*(x_piv.^m_f_CBH);    % Rate coefficients of complexation (L/mol.s)
195
196 kp_e_CBH = prm(3);                    % Rate constant of decomplexation, 1/s
197 m_e_CBH = 0;
198 k_e_CBH = kp_e_CBH*(x_piv.^m_e_CBH);    % Rate coefficients of decomplexation, 1/s
199
200 tau_CBH = prm(4);                     % Enzyme footprint, mol/m2
201 k_ads_CBH = prm(5);                   % Adsorption rate constant, L/mol.s
202 k_des_CBH = prm(6);                   % Desorption rate constant, 1/s
203
204 k_If_CBH = [prm(7) prm(8)];           % Forward glucose/cellobiose inhibition, L/mol.s
205 k_Ir_CBH = [prm(9) prm(10)];          % Reverse glucose/cellobiose inhibition, 1/s
206
207 % EG
208
209 kp_h_EG = prm(16);                    % Rate constant of hydrolysis (insoluble cellulose), DP/s
210 m_h_EG = -1;
211 k_h_EG = kp_h_EG*(x_piv.^m_h_EG);     % Rate coefficients of hydrolysis (insoluble cellulose),1/s
212
213 kp_hs_EG = prm(17);                   % Rate constant of hydrolysis (soluble), L/mol.s
214 m_hs_EG = 0;
215 k_hs_EG = kp_hs_EG*(x_piv.^m_hs_EG);    % Rate coefficients of hydrolysis (soluble), L/mol.s
216 k_hs_EG(2) = 0;                       % EG unable to hydrolyze cellobiose
217
218 kp_f_EG = prm(18);                    % Rate constant of complexation, DP.L/mol.s
219 m_f_EG = -1;
220 k_f_EG = kp_f_EG*(x_piv.^m_f_EG);     % Rate coefficients of complexation, L/mol.s
221
222 kp_e_EG = prm(19);                    % Rate constant of decomplexation, 1/s
223 m_e_EG = 0;
224 k_e_EG = kp_e_EG*(x_piv.^m_e_EG);     % Rate coefficients of decomplexation, 1/s
225
226 tau_EG = prm(20);                     % Enzyme footprint, mol/m2
227 k_ads_EG = prm(21);                   % Adsorption rate constant, L/mol.s
228 k_des_EG = prm(22);                   % Desorption rate constant, 1/s
229
230 k_If_EG = [prm(23) prm(24)];          % Forward glucose/cellobiose inhibition, L/mol.s
231 k_Ir_EG = [prm(25) prm(26)];          % Reverse glucose/cellobiose inhibition, 1/s
```

```
232
233 % BG
234
235 k_h_BG = prm(13);                        % Rate constant of hydrolysis, L/mol.s
236
237 k_If_BG = prm(14);                       % Forward glucose inhibition, L/mol.s
238 k_Ir_BG = prm(15);                       % Reverse glucose inhibition, 1/s
239
240 % Fixed pivot technique
241
242 [n_exo,n_endo] = fp(p,q,vm,x_piv);       % Particle allocation functions
243
244 % ODE solver
245
246 tstart = 0;
247 tfinal = t_end;
248
249 tic;
250
251 options = odeset('AbsTol',1e-8,'RelTol',1e-6,...
252     'NonNegative',1:(6*p+6*q+9));
253
254 [t,y] = ode15s(@(t,y) MLPBM_ode(p,q,n_exo,n_endo,E_T_EG,...
255     E_T_CBH,E_T_BG,k_h_CBH,k_h_EG,k_hs_EG,k_h_BG,k_f_CBH,k_f_EG,k_e_CBH,...
256     k_e_EG,k_ads_EG,k_des_EG,k_ads_CBH,k_des_CBH,tau_EG,tau_CBH,k_If_CBH,...
257     k_Ir_CBH,k_If_EG,k_Ir_EG,k_If_BG,k_Ir_BG,x_piv,R0,ro,L,n,...
258     exp_ratio,y,t),[tstart tfinal],init,options);
259
260 time_fp_sim =...
261     '\nTotal time elapsed for ML-PBM simulation is %.1f seconds.\n';
262
263 fprintf(time_fp_sim,toc)
264
265 end
266
267 %%-------------------------------------------------------------------------
268
269 function [n_exo,n_endo] = fp(p,q,vm,x_piv)
270
271 % Particle allocation function - fixed pivot technique
272
273 % Chain-end dimer scission
274
275 n_exo = zeros(p+q,p+q);
276 n_exo(1,3) = 1;
277 n_exo(2,3) = 1; n_exo(2,4) = 2; n_exo(2,5:end) = 1;
278
279 for i = 3:p+q
280     for j = i:p+q
281         if j==i && x_piv(j)-2*vm>x_piv(i-1)
282             n_exo(i,j) = ((x_piv(j)-2*vm)-x_piv(i-1))/(x_piv(i)-x_piv(i-1));
283         elseif j==i && x_piv(j)-2*vm==x_piv(i-1)
284             n_exo(i,j) = 0;
285         elseif j~=i && x_piv(j)-2*vm>x_piv(i-1) && x_piv(j)-2*vm<x_piv(i)
286             n_exo(i,j) = ((x_piv(j)-2*vm)-x_piv(i-1))/(x_piv(i)-x_piv(i-1));
287         elseif j~=i && x_piv(j)-2*vm>x_piv(i) && x_piv(j)-2*vm<x_piv(i+1)
288             n_exo(i,j) = (x_piv(i+1)-(x_piv(j)-2*vm))/(x_piv(i+1)-x_piv(i));
289         elseif j~=i && x_piv(j)-2*vm==x_piv(i)
290             n_exo(i,j) = 1;
```

```
        elseif j≠i && x_piv(j)-2*vm==x_piv(i-1)
            n_exo(i,j) = 0;
        else
            n_exo(i,j) = 0;
        end
    end
    n_exo(i,i) = n_exo(i,i)-1;
end

% Random scission

n_endo = zeros(p+q,p+q);

for j = 2:p+q
    n_endo(1,j) = 2/(x_piv(j-1));
end

for i = 2:p
    for j = i+1:p+q
        n_endo(i,i) = -1;
        n_endo(i,j) = 2/(x_piv(j-1));
    end
end

for i = p+1:p+q
    for j = i+1:p+q
        n_endo(i,j) = (x_piv(i+1)-x_piv(i-1))/(x_piv(j-1));
    end
    n_endo(i,i) = -1;
end

end

%%------------------------------------------------------------------------

function ode = MLPBM_ode(p,q,n_exo,n_endo,E_T_EG,E_T_CBH,...
    E_T_BG,k_h_CBH,k_h_EG,k_hs_EG,k_h_BG,k_f_CBH,k_f_EG,k_e_CBH,k_e_EG,...
    k_ads_EG,k_des_EG,k_ads_CBH,k_des_CBH,tau_EG,tau_CBH,k_If_CBH,...
    k_Ir_CBH,k_If_EG,k_Ir_EG,k_If_BG,k_Ir_BG,x_piv,R0,ro,L,n,...
    exp_ratio,y,t)

dydt = zeros(6*p+6*q+9,1);

% Enzymes

E_F_EG = E_T_EG-y(6*p+6*q+1)-sum(y(2*p+2*q+1:4*p+4*q))...
    -sum(y(6*p+6*q+3:6*p+6*q+4));
E_F_CBH = E_T_CBH-y(6*p+6*q+2)-sum(y(p+q+1:2*p+2*q))...
    -sum(y(3*p+3*q+1:4*p+4*q))-sum(y(6*p+6*q+5:6*p+6*q+8));
E_F_BG = E_T_BG - y(6*p+6*q+9);

% Particle radius

x3 = [x_piv;x_piv;x_piv];
x5 = [x_piv;x_piv;x_piv;x_piv;x_piv];
R = sqrt(sum(y(p+q+1:6*p+6*q).*(162*x5+18))/(ro*pi*L*n));

if R-R0≤0;
    R_ratio = 0;
```

```
350 else
351     R_ratio = (R-R0)/R;
352 end
353
354 % Total surface area
355
356 As = (n*2*pi*R*L)/1000;
357
358 % Soluble polymers
359
360 dydt(1) = sum(n_exo(1,:)'.*k_h_CBH.*y(p+q+1:2*p+2*q))...
361     +sum(n_endo(1,7:end)'.*k_h_EG(7:end).*y(2*p+2*q+7:3*p+3*q))...
362     +sum(n_endo(1,4:end)'.*k_h_EG(4:end).*y(3*p+3*q+4:4*p+4*q))...
363     +sum(E_F_EG*n_endo(1,2:6)'.*k_hs_EG(2:6).*y(2:6))...
364     +2*k_h_BG*E_F_BG*y(2)+k_Ir_CBH(1)*y(6*p+6*q+5)+k_Ir_CBH(1)*y(6*p+6*q+7)...
365     +k_Ir_EG(1)*y(6*p+6*q+3)+k_Ir_BG*y(6*p+6*q+9)...
366     -k_If_CBH(1)*E_F_CBH*y(1)-k_If_CBH(1)*y(6*p+6*q+2)*y(1)...
367     -k_If_EG(1)*E_F_EG*y(1)-k_If_BG*E_F_BG*y(1);
368
369 dydt(2) = sum(n_exo(2,:)'.*k_h_CBH.*y(p+q+1:2*p+2*q))...
370     +sum(n_endo(2,7:end)'.*k_h_EG(7:end).*y(2*p+2*q+7:3*p+3*q))...
371     +sum(n_endo(2,4:end)'.*k_h_EG(4:end).*y(3*p+3*q+4:4*p+4*q))...
372     +sum(E_F_EG*n_endo(2,3:6)'.*k_hs_EG(3:6).*y(3:6))...
373     -k_hs_EG(2)*E_F_EG*y(2)-k_h_BG*E_F_BG*y(2)...
374     +k_Ir_CBH(2)*y(6*p+6*q+6)+k_Ir_CBH(2)*y(6*p+6*q+8)+k_Ir_EG(2)*y(6*p+6*q+4)...
375     -k_If_CBH(2)*E_F_CBH*y(2)-k_If_CBH(2)*y(6*p+6*q+2)*y(2)...
376     -k_If_EG(2)*E_F_EG*y(2);
377
378 for i = 3:5;
379     dydt(i) = sum(n_endo(i,7:end)'.*k_h_EG(7:end).*y(2*p+2*q+7:3*p+3*q))...
380         +sum(n_endo(i,i+1:end)'.*k_h_EG(i+1:end).*y(3*p+3*q+i+1:4*p+4*q))...
381         +sum(E_F_EG*n_endo(i,i+1:6)'.*k_hs_EG(i+1:6).*y(i+1:6))...
382         -k_hs_EG(i)*E_F_EG*y(i);
383 end
384
385 for i = 6;
386     dydt(i) = sum(n_endo(i,i+1:end)'.*k_h_EG(i+1:end).*y(2*p+2*q+i+1:3*p+3*q))...
387         +sum(n_endo(i,i+1:end)'.*k_h_EG(i+1:end).*y(3*p+3*q+i+1:4*p+4*q))...
388         -k_hs_EG(i)*E_F_EG*y(i);
389 end
390
391 % CBH-bound surface polymers
392
393 dydt(p+q+3) = sum(n_exo(3,3:end)'.*k_h_CBH(3:end).*y(p+q+3:2*p+2*q));
394
395
396 for i = p+q+4:p+q+6;
397     dydt(i) = sum(n_exo(i-p-q,:)'.*k_h_CBH.*y(p+q+1:2*p+2*q))...
398         -(k_f_EG(i-p-q)*y(6*p+6*q+1)*y(i))...
399         +(k_e_EG(i-p-q)*y(i-p-q+3*p+3*q));
400
401 end
402
403 for i = p+q+7:2*p+2*q-1;
404     dydt(i) = sum(n_exo(i-p-q,:)'.*k_h_CBH.*y(p+q+1:2*p+2*q))...
405         +(y(6*p+6*q+2)*k_f_CBH(i-p-q)*y(i-p-q+4*p+4*q))...
406         -(k_e_CBH(i-p-q)*y(i))...
407         -(k_f_EG(i-p-q)*y(6*p+6*q+1)*y(i))...
408         +(k_e_EG(i-p-q)*y(i-p-q+3*p+3*q));
```

```
409
410 end
411
412 for i = 2*p+2*q;
413     dydt(i) = sum(n_exo(i-p-q,:)'.*k_h_CBH.*y(p+q+1:2*p+2*q))...
414         +(y(6*p+6*q+2)*k_f_CBH(i-p-q)*y(i-p-q+4*p+4*q))...
415         -(k_e_CBH(i-p-q)*y(i))...
416         -(k_f_EG(i-p-q)*y(6*p+6*q+1)*y(i))...
417         +(k_e_EG(i-p-q)*y(i-p-q+3*p+3*q));
418 end
419
420 % EG-bound surface polymers
421
422 for i = 2*p+2*q+7:3*p+3*q;
423     dydt(i) = (k_f_EG(i-2*p-2*q)*y(6*p+6*q+1)*y(i-2*p-2*q+4*p+4*q))...
424         -(k_e_EG(i-2*p-2*q)*y(i))...
425         -(k_h_EG(i-2*p-2*q)*y(i));
426 end
427
428 % CBH-EG-bound surface polymers
429
430 for i = 3*p+3*q+4:4*p+4*q;
431     dydt(i) = (k_f_EG(i-3*p-3*q)*y(6*p+6*q+1)*y(i-3*p-3*q+p+q))...
432         -(k_e_EG(i-3*p-3*q)*y(i))...
433         -(k_h_EG(i-3*p-3*q)*y(i));
434 end
435
436 % Surface polymers
437
438 m = zeros(p+q,1);
439
440 for i = 4*p+4*q+7:5*p+5*q-1;
441     m(i-4*p-4*q) = ...
442         (sum(n_endo(i-4*p-4*q,i-4*p-4*q+1:end)'.*k_h_EG(i-4*p-4*q+1:end)...
443         .*y(i-4*p-4*q+2*p+2*q+1:3*p+3*q))...
444         +sum(n_endo(i-4*p-4*q,i-4*p-4*q+1:end)'.*k_h_EG(i-4*p-4*q+1:end)...
445         .*y(i-4*p-4*q+1+3*p+3*q:4*p+4*q))...
446         -(k_f_CBH(i-4*p-4*q)*y(6*p+6*q+2)*y(i))...
447         +(k_e_CBH(i-4*p-4*q)*y(i-4*p-4*q+p+q))...
448         -(k_f_EG(i-4*p-4*q)*y(6*p+6*q+1)*y(i))...
449         +(k_e_EG(i-4*p-4*q)*y(i-4*p-4*q+2*p+2*q)));
450 end
451
452 m(p+q) = (-(k_f_CBH(p+q)*y(6*p+6*q+2)*y(5*p+5*q))...
453     +(k_e_CBH(p+q)*y(2*p+2*q))...
454     -(k_f_EG(p+q)*y(6*p+6*q+1)*y(5*p+5*q))...
455     +(k_e_EG(p+q)*y(3*p+3*q)));
456
457 for i = 4*p+4*q+7:5*p+5*q;
458     dydt(i) = m(i-4*p-4*q)...
459         -(sum(m.*(162*x_piv+18))+sum(dydt(p+q+1:4*p+4*q).*(162*x3+18)))...
460         *R_ratio*ppval(exp_ratio{i-4*p-4*q},R-R0);
461 end
462
463 % Internal inaccessible polymer
464
465 for i = 5*p+5*q+7:6*p+6*q;
466     dydt(i) = (sum(m.*(162*x_piv+18))+sum(dydt(p+q+1:4*p+4*q)...
467         .*(162*x3+18)))*R_ratio*ppval(exp_ratio{i-5*p-5*q},R-R0);
```

```
468 end
469
470 % Surface-adsorbed EG
471
472 dydt(6*p+6*q+1) = (k_ads_EG*E_F_EG*(As*tau_EG-y(6*p+6*q+1)))...
473     -(k_des_EG*y(6*p+6*q+1))...
474     -sum(y(6*p+6*q+1)*k_f_EG(7:end).*y(4*p+4*q+7:5*p+5*q))...
475     +sum(k_e_EG(7:end).*y(2*p+2*q+7:3*p+3*q))...
476     -sum(y(6*p+6*q+1)*k_f_EG(4:end).*y(p+q+4:2*p+2*q))...
477     +sum(k_e_EG(4:end).*y(3*p+3*q+4:4*p+4*q));
478
479 % Surface-adsorbed CBH
480
481 dydt(6*p+6*q+2) = (k_ads_CBH*E_F_CBH*(As*tau_CBH-y(6*p+6*q+2)))...
482     -(k_des_CBH*y(6*p+6*q+2))...
483     -sum(y(6*p+6*q+2)*k_f_CBH(7:end).*y(4*p+4*q+7:5*p+5*q))...
484     +sum(k_e_CBH(7:end).*y(p+q+7:2*p+2*q))...
485     +k_Ir_CBH(1)*y(6*p+6*q+7)+k_Ir_CBH(2)*y(6*p+6*q+8)...
486     -y(6*p+6*q+2)*sum(k_If_CBH'.*y(1:2));
487
488 % Inhibited EG
489
490 for i = 6*p+6*q+3:6*p+6*q+4;
491     dydt(i) = k_If_EG(i-6*p-6*q-3+1)*E_F_EG*y(i-6*p-6*q-3+1)...
492         -k_Ir_EG(i-6*p-6*q-3+1)*y(i);
493 end
494
495 % Inhibited CBH
496
497 for i = 6*p+6*q+5:6*p+6*q+6;
498     dydt(i) = k_If_CBH(i-6*p-6*q-5+1)*E_F_CBH*y(i-6*p-6*q-5+1)...
499         -k_Ir_CBH(i-6*p-6*q-5+1)*y(i);
500 end
501
502 for i = 6*p+6*q+7:6*p+6*q+8;
503     dydt(i) = k_If_CBH(i-6*p-6*q-7+1)*y(6*p+6*q+2)*y(i-6*p-6*q-7+1)...
504         -k_Ir_CBH(i-6*p-6*q-7+1)*y(i);
505 end
506
507 % Inhibited BG
508
509 for i = 6*p+6*q+9;
510     dydt(i) = k_If_BG*E_F_BG*y(i-6*p-6*q-9+1)...
511         -k_Ir_BG*y(i);
512 end
513
514 % Output
515
516 ode = dydt;
517
518 end
519
520 %%-------------------------------------------------------------------------
521
522 function [C,CXS,CNS,CNXS,CFS,CS,CI,CT,CC,E_S_EG,E_S_CBH,E_F_EG,E_F_CBH,...
523     E_F_BG,R,conv_rem] = post_processing(t,y,p,q,x_piv,ro,L,n,mss,E_T_EG,...
524     E_T_CBH,E_T_BG)
525
526 % Molar concentrations, mol/L
```

```
527
528  C = y(:,1:p+q);                      % Soluble products
529  CXS = y(:,p+q+1:2*p+2*q);            % CBH-bound complex
530  CNS = y(:,2*p+2*q+1:3*p+3*q);        % EG-bound complex
531  CNXS = y(:,3*p+3*q+1:4*p+4*q);       % CBH-EG-bound complex
532  CFS = y(:,4*p+4*q+1:5*p+5*q);        % Free un-bound surface polymers
533  CS = CXS+CNS+CNXS+CFS;               % Total surface polymers
534  CI = y(:,5*p+5*q+1:6*p+6*q);         % Internal polymers
535  CT = C+CS+CI;                        % Total polymers
536  CC = CS+CI;                          % Insoluble polymers
537
538  E_S_EG = y(:,6*p+6*q+1);             % Surface-adsorbed EG
539  E_S_CBH = y(:,6*p+6*q+2);            % Surface-adsorbed CBH
540
541  CNI(:,1) = y(:,6*p+6*q+3);           % Inhibited EG
542  CNI(:,2) = y(:,6*p+6*q+4);
543
544  CXI(:,1) = y(:,6*p+6*q+5)+y(:,6*p+6*q+7);   % Inhibited CBH
545  CXI(:,2) = y(:,6*p+6*q+6)+y(:,6*p+6*q+8);
546
547  CBI = y(:,6*p+6*q+9);                % Inhibited BG
548
549  E_F_EG = zeros(length(t),1);         % Free EG
550  E_F_CBH = zeros(length(t),1);        % Free CBH
551  E_F_BG = zeros(length(t),1);         % Free BG
552
553  for i = 1:length(t)
554      E_F_EG(i) = E_T_EG-E_S_EG(i)-sum(CNS(i,:))-sum(CNXS(i,:))-sum(CNI(i,:));
555      E_F_CBH(i) = E_T_CBH-E_S_CBH(i)-sum(CXS(i,:))-sum(CXI(i,:))-sum(CNXS(i,:));
556      E_F_BG(i) = E_T_BG-sum(CBI(i));
557  end
558
559  % Transient particle radius, m
560
561  R = zeros(length(t),1);
562
563  for i = 1:length(t)
564      R(i) = sqrt(sum(CC(i,:)'.*(162*x_piv+18))/(ro*pi*L*n));
565  end
566
567  % Mass of insoluble polymers, g/L
568
569  mss_CC = zeros(length(t),1);
570
571  for i = 1:length(t)
572      mss_CC(i) = sum(CC(i,:)'.*(162*x_piv+18));
573  end
574
575  % Conversion
576
577  conv_rem = zeros(length(t),1);
578
579  for i = 1:length(t)
580      conv_rem(i) = (mss-mss_CC(i))/mss;
581  end
582
583  end
```

D.4 Simulation of Consolidated Bioprocessing of Cellulose (UC-PBM)

Main file:

```
close all; clear all; clc;

% Setting input parameters

% Cellulose, cellobiose, biomass loadings

mss = 4.59;                 % Initial concentration of cellulose, g/L
cellobiose = 1e-4;          % Initial concentration of cellobiose, g/L
biomass = 0.04;             % Initial concentration of biomass, g/L

Av_pc = 10;                 % Pre-culture Avicel concentration, g/L
CB_pc = 0;                  % Pre-culture Cellobiose concentration, g/L

% Cellulosome, cellulase, biomass settings (if biomass grown on cellulose)

MW_cellome = 1.73e6;        % Molecular weight of cellulosomes, g/mol
mass_frac_cellulase = 0.50; % Fraction of cellulase in cellulosome
mol_exo = 11.08*0.99;       % Moles of exo-enzymes per mole of cellulosome
mol_endo = 11.08-mol_exo;   % Moles of endo-enzymes per mol of cellulosome
biom_DCW_prtn = 1/1.2;      % Fraction of biomass in DW (N20%)

% Cellulosome, cellulase, biomass settings (if biomass grown on cellobiose)

% MW_cellome = 1.89e6;
% mass_frac_cellulase = 0.39;
% mol_exo = 9.48*0.99;
% mol_endo = 9.48-mol_exo;
% biom_DCW_prtn = 1/1.02;   % (N2%)

% Mesh setting

p = 20;                     % Number of discrete pivots

% Time Span

t_end = 1.62e5;             % Time span of simulation, seconds

% Initial Distribution - base fitted cellulose distribution

load('Raw_Avicel_calibrated_initialDist_Engel_2012_final','MnL','MwL',...
    'MnH','MwH','N','r_massL');

% EFVs from metabolic network

load EFM_NewReduced9aa_CTh_DSM1313_iAT601_Thompson_2016.mat;    % (N20%)
```

```
% load EFM_NewReduced10_CTh_DSM1313_iAT601_Thompson_2016.mat;   % (N2%)

% List of Model Paramaters

% ML-PBM parameters

% Exo-enzyme-related parameters

% 1  kp_h_exo - Rate contant of exo-enzyme hydrolysis, 1/s
% 2  kp_f_exo - Rate constant of exo-enzyme complexation, DP.L/mol.s
% 3  kp_e_exo - Rate constant of exo-enzyme decomplexation, 1/s
% 4  tau_exo - Exo-enzyme footprint/area coverage, mol/m2
% 5  k_ads_exo - Rate constant of exo-enzyme adsorption, L/mol.s
% 6  k_des_exo - Rate constant of exo-enzyme desoprtion, 1/s
% 7  k_If_exo(1) - Forward glucose inhibition rate constant (exo-enzyme), L/mol.s
% 8  k_If_exo(2) - Forward cellobiose inhibition rate constant(exo-enzyme), L/mol.s
% 9  k_Ir_exo(1) - Reverse glucose inhibition rate constant(exo-enzyme), 1/s
% 10 k_Ir_exo(2) - Reverse cellobiose inhibition rate constant(exo-enzyme), 1/s

% Substrate-related parameters

% 11 pnt_zone - Number of layers in penetration zone
% 12 r_massL - Mass ratio of cellulose in penetration zone

% Endo-enzyme-related parameters

% 13 kp_h_endo - Rate constant of endo-enzyme hydrolysis (insoluble cellulose), DP/s
% 14 kp_hs_endo - Rate constant of endo-enzyme hydrolysis (soluble cellulose), L/mol.s
% 15 kp_f_endo - Rate constant of endo-enzyme complexation, DP.L/mol.s
% 16 kp_e_endo - Rate constant of endo-enzyme decomplexation, 1/s
% 17 tau_endo - Endo-enzyme enzyme footprint/area coverage, mol/m2
% 18 k_ads_endo - Rate constant of endo-enzyme adsorption, L/mol.s
% 19 k_des_endo - Rate constant of endo-enzyme desorption, 1/s
% 20 k_If_endo(1) - Forward glucose inhibition rate constant (endo-enzyme), L/mol.s
% 21 k_If_endo(2) - Forward cellobiose inhibition rate constant (endo-enzyme), L/mol.s
% 22 k_Ir_endo(1) - Reverse glucose inhibition rate constant (endo-enzyme), 1/s
% 23 k_Ir_endo(2) - Reverse cellobiose inhibition rate constant (endo-enzyme), 1/s

% L-HCM parameters

% 24 aF(1) - Constitutive intracellular enzyme synthesis rate (F1), 1/s
% 25 aF(2) - Constitutive intracellular enzyme synthesis rate (F2), 1/s
% 26 bF(1) - Intracellular enzyme decay rate (F1),1/s
% 27 bF(2) - Intracellular enzyme decay rate (F2),1/s
% 28 kE(1) - Inducive intracellular enzyme synthesis rate (F1), 1/s
% 29 kE(2) - Inducive intracellular enzyme synthesis rate (F2), 1/s
% 30 KE(1) - MM constant for intracellular enzyme synthesis (F1), mol/L
% 31 KE(2) - MM constant for intracellular enzyme synthesis (F2), mol/L
% 32 k_inh_G - Glucose inhibition constant, mol/L
% 33 k_inh_etoh - Ethanol inhibition constant, mol/L
% 34 kmax(1) - Maximum substrate uptake rate (F1), mol/g-biom.s
% 35 kmax(2) - Maximum substrate uptake rate (F2), mol/g-biom.s
% 36 K(1) - MM constant for substrate uptake (F1), mol/L
% 37 K(2) - MM constant for substrate uptake (F2), mol/L
% 38 aE - Constitutive cellulase synthesis rate, g/g-biom.s

% Input the values for model parameters below:

prm(1) = 2.15;
```

```
105 prm(2) = 3e7;
106 prm(3) = 10;
107 prm(4) = 2.22e-8;
108 prm(5) = 7.3722e5;
109 prm(6) = 0.01;
110 prm(7) = 0;
111 prm(8) = 0;
112 prm(9) = 0;
113 prm(10) = 0;
114
115 prm(11) = 1;
116 prm(12) = 0.028;
117
118 prm(13) = 4;
119 prm(14) = 3;
120 prm(15) = 1e9;
121 prm(16) = 10;
122 prm(17) = 8.04e-9;
123 prm(18) = 7.11e5;
124 prm(19) = 0.01;
125 prm(20) = 0;
126 prm(21) = 0;
127 prm(22) = 0;
128 prm(23) = 0;
129
130 prm(24) = 0.1/3600;
131 prm(25) = 0.1/3600;
132 prm(26) = 0.2/3600;
133 prm(27) = 0.2/3600;
134 prm(28) = 1/3600;
135 prm(29) = 1/3600;
136 prm(32) = 1.22e-3;
137 prm(33) = 1.74;
138 prm(34) = 1.3e-6;
139 prm(35) = 1.9e-6;
140 prm(36) = 1e-6;
141 prm(37) = 9e-6;
142 prm(30) = prm(36);
143 prm(31) = prm(37);
144 prm(38) = 3e-7;
145
146 % Run simulation
147
148 [Z,idx_cb,idx_atp,idx_csm,idx_biom,idx_etoh,idx_lac,idx_form,idx_ace,...
149     F1,F1A,F1C,F2,F2A,F2B,idx_met,ZF1,ZF2,Y1_met_lump,Y2_met_lump,Y1_E,...
150     Y2_E,Y2_B,x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,pdiT,t,C,CXS,CNS,...
151     CNXS,CFS,CS,CI,CT,CC,E_T_exo,E_T_endo,CB,etoh,lac,form,ace,cellome,...
152     biom,cellulase_tot,cellu,cell_gluc_eq,eF_rel,R,conv_rem,conv_cb,uF,vF,...
153     r_up,mu] = UCPBM(p,N,prm,mss,MnL,MwL,MnH,MwH,t_end,CB_pc,cellobiose,...
154     biomass,MW_cellome,mol_exo,mol_endo,biom_DCW_prtn,ems,ems_idx,reac,...
155     mass_frac_cellulase);
156
157 % List of outputs
158
159 % x_piv     % DP mesh
160 % R_in      % Initial particle radius
161 % layers    % Total no. of layers in cellulose particles
162 % C_T       % Total initial cellulose distribution
163 % C_S       % Surface initial cellulose distribution
```

```
164 % C_INT     % Internal initial cellulose distribution
165 % MnT       % Overall initial number-average DP
166 % MwT       % Overall initial weight-average DP
167 % pdiT      % Overall initial polydispersity index
168
169 % Indexes of key metabolic rxns
170
171 % idx_cb,idx_atp,idx_csm,idx_biom,idx_etoh,idx_lac,idx_form,idx_ace
172
173 % Metabolic information
174
175 % Z                                          % EFV matrix
176 % F1,F1A,F1C,F2,F2A,F2B                      % Number of EFVs in EFV families
177 % ZF1,ZF2                                    % Lumped EFV families
178 % Y1_met_lump,Y2_met_lump,Y1_E,Y2_E,Y2_B     % Lumped yields
179
180 % t         % Time, s
181
182 % Time,DP-dependent variables
183
184 % C         % Soluble products, mol/L
185 % CXS       % CBH-bound complex, mol/L
186 % CNS       % EG-bound complex, mol/L
187 % CNXS      % CBH-EG-bound complex, mol/L
188 % CFS       % Free un-bound surface polymers, mol/L
189 % CS        % Total surface polymers, mol/L
190 % CI        % Internal polymers, mol/L
191 % CT        % Total polymers, mol/L
192 % CC        % Insoluble polymers, mol/L
193
194 % Time-dependent variables
195
196 % E_T_endo          % Total endo-enzymes, mol/L
197 % E_T_exo           % Total exo-enzymes, mol/L
198 % CB                % Cellobiose, g/L
199 % etoh              % Ethanol, g/L
200 % lac               % Lactate, g/L
201 % form              % Formate, g/L
202 % ace               % Acetate, g/L
203 % cellome           % Cellulosome, g/L
204 % biom              % Biomass, g/L
205 % cellulase_tot     % Total cellulase, g/L
206 % cellu             % Remaining insoluble cellulose, g/L
207 % cell_gluc_eq      % Remaining insoluble cellulose, g glu/L
208 % eF_rel           % Relative intracellular enzyme levels
209 % R                % Transient of particle radius, m
210 % conv_rem         % Overall cellulose conversion
211 % conv_cb          % Overall cellobiose conversion
212 % uF,vF            % Cybernetic variables
213 % r_up             % Uptake fluxes through EFV families, mol/g-biom.s
214 % mu               % Specific biomass growth rate, 1/s
```

Function file:

```
function [Z,idx_cb,idx_atp,idx_csm,idx_biom,idx_etoh,idx_lac,idx_form,...
    idx_ace,F1,F1A,F1C,F2,F2A,F2B,idx_met,ZF1,ZF2,Y1_met_lump,...
    Y2_met_lump,Y1_E,Y2_E,Y2_B,x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,...
    pdiT,t,C,CXS,CNS,CNXS,CFS,CS,CI,CT,CC,E_T_exo,E_T_endo,CB,etoh,lac,...
    form,ace,cellome,biom,cellulase_tot,cellu,cell_gluc_eq,eF_rel,R,...
    conv_rem,conv_cb,uF,vF,r_up,mu] = UCPBM(p,N,prm,mss,MnL,MwL,MnH,MwH,...
    t_end,CB_pc,cellobiose,biomass,MW_cellome,mol_exo,mol_endo,...
    biom_DCW_prtn,ems,ems_idx,reac,mass_frac_cellulase)

[Z,idx_cb,idx_atp,idx_csm,idx_biom,idx_etoh,idx_lac,idx_form,idx_ace,F1,...
    F1A,F1C,F2,F2A,F2B,idx_met,ZF1,ZF2,Y1_met_lump,Y2_met_lump,Y1_E,...
    Y2_E,Y2_B] = EFV_setting(ems,ems_idx,biom_DCW_prtn,reac);

[x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,pdiT,q,vm,R0,ro,L,n,exp_ratio]...
    = MLPBM_setting(p,N,prm,mss,MnL,MwL,MnH,MwH);

[t,y,eF_max,kmax,K,K_inh_G,K_inh_etoh] = ODE_setting(p,q,C_S,C_INT,prm,...
    x_piv,vm,t_end,Y1_met_lump,Y2_met_lump,Y1_E,Y2_E,Y2_B,CB_pc,...
    cellobiose,biomass,MW_cellome,mol_exo,mol_endo,R0,ro,L,n,exp_ratio);

[C,CXS,CNS,CNXS,CFS,CS,CI,CT,CC,E_T_exo,E_T_endo,CB,etoh,lac,form,ace,...
    cellome,biom,cellulase_tot,cellu,cell_gluc_eq,eF_rel,R,conv_rem,...
    conv_cb,uF,vF,r_up,mu] = post_processing(t,y,p,q,MW_cellome,mol_endo,...
    mol_exo,mass_frac_cellulase,x_piv,eF_max,ro,L,n,mss,cellobiose,kmax,...
    K,K_inh_G,K_inh_etoh,Y2_B);

end

%%-------------------------------------------------------------------------

function [Z,idx_cb,idx_atp,idx_csm,idx_biom,idx_etoh,idx_lac,idx_form,...
    idx_ace,F1,F1A,F1C,F2,F2A,F2B,idx_met,ZF1,ZF2,Y1_met_lump,...
    Y2_met_lump,Y1_E,Y2_E,Y2_B] = EFV_setting(ems,ems_idx,biom_DCW_prtn,...
    reac)

% EFV setting

n_eta = 1;                      % Sensitivity setting to tuning parameters

% Tuning parameters for N20%

a_tune1 = [-1.544065229170075 1.569902586838220...
    4.730079923633626 -8.058648461794135];
a_tune2 = [-0.002019303084171 -0.007859218375312...
    -0.009091575956450 0.005826206931968];

% Tuning parameters for N2%

% a_tune1 = [-4.467191370597475 4.577593783957661...
%     13.788127542564130 -23.705397747646195];
% a_tune2 = [3.644003789716016 -2.686629088044898...
%     -9.470963287592038 13.097851990509195];

Z = ems';                                       % EFV matrix, Z
```

```
55  rxn = reac(ems_idx);                              % List of metabolic rxns
56
57  % Indexes for key metabolic rxns
58
59  idx_cb = find(¬cellfun(@isempty,strfind(rxn,'T_e_to_c_C00185_c')));
60  idx_atp = find(¬cellfun(@isempty,strfind(rxn,'R_R_MAINT')));
61  idx_biom = find(¬cellfun(@isempty,strfind(rxn,'T_c_to_e_m85')));
62  idx_csm = find(¬cellfun(@isempty,strfind(rxn,'R_EXC_OUT_m90')));
63  idx_etoh = find(¬cellfun(@isempty,strfind(rxn,'T_c_to_e_C00469_c')));
64  idx_lac = find(¬cellfun(@isempty,strfind(rxn,'T_c_to_e_C00186_c')));
65  idx_form = find(¬cellfun(@isempty,strfind(rxn,'T_c_to_e_C00058_c')));
66  idx_ace = find(¬cellfun(@isempty,strfind(rxn,'T_c_to_e_C00033_c')));
67
68  % Classification of EFV families
69
70  F = find(Z(idx_cb,:)≠0)';
71  F1 = intersect(F,setdiff(find(Z(idx_csm,:)≠0),find(Z(idx_biom,:)≠0)));
72  F2 = setdiff(F,F1);
73
74  e1C = intersect(F1,find(Z(idx_csm,:)≠0));
75  e1A = intersect(F1,find(Z(idx_atp,:)≠0));
76
77  e2B = intersect(F2,find(Z(idx_biom,:)≠0));
78  e2A = intersect(F2,find(Z(idx_atp,:)≠0));
79
80  F1C = e1C;
81  F1A = setdiff(e1A,e1C);
82  F1 = unique([F1C;F1A]);
83
84  F2B = e2B;
85  F2A = setdiff(e2A,e2B);
86  F2 = unique([F2B;F2A]);
87
88  z1C = Z(:,F1C);                % z-matrix of F1
89  z2B = Z(:,F2B);                % z-matrix of F2,biom
90  z2A = Z(:,F2A);                % z-matrix of F2,atp
91
92  % Yields of metabolites in individual EFVs before lumping
93
94  Y1C_csm = (z1C(idx_csm,:)./z1C(idx_cb,:));
95  Y1C_etoh = z1C(idx_etoh,:)./z1C(idx_cb,:);
96  Y1C_lac = z1C(idx_lac,:)./z1C(idx_cb,:);
97  Y1C_form = z1C(idx_form,:)./z1C(idx_cb,:);
98  Y1C_ace = z1C(idx_ace,:)./z1C(idx_cb,:);
99
100 Y2B_biom = (z2B(idx_biom,:)./z2B(idx_cb,:));
101 Y2B_etoh = z2B(idx_etoh,:)./z2B(idx_cb,:);
102 Y2B_lac = z2B(idx_lac,:)./z2B(idx_cb,:);
103 Y2B_form = z2B(idx_form,:)./z2B(idx_cb,:);
104 Y2B_ace = z2B(idx_ace,:)./z2B(idx_cb,:);
105
106 Y2A_atp = (z2A(idx_atp,:)./z2A(idx_cb,:));
107 Y2A_etoh = z2A(idx_etoh,:)./z2A(idx_cb,:);
108 Y2A_lac = z2A(idx_lac,:)./z2A(idx_cb,:);
109 Y2A_form = z2A(idx_form,:)./z2A(idx_cb,:);
110 Y2A_ace = z2A(idx_ace,:)./z2A(idx_cb,:);
111
112 % Metabolite order: eth,lac,form,ace
113
```

```
idx_met = [idx_etoh;idx_lac;idx_form;idx_ace];  % Index of metabolite secreting rxns

Y1C_met = [Y1C_etoh;Y1C_lac;Y1C_form;Y1C_ace];
Y2B_met = [Y2B_etoh;Y2B_lac;Y2B_form;Y2B_ace];
Y2A_met = [Y2A_etoh;Y2A_lac;Y2A_form;Y2A_ace];

% EFV lumping scheme

wt = 1;     % Arbitrary weightage

eta1C = zeros(length(F1C),1);

for i = 1:length(F1C)
    eta1C(i) = wt*(Y1C_csm(i)+sum(a_tune1'.*Y1C_met(:,i)));
end

eta2B = zeros(length(F2B),1);

for i = 1:length(F2B)
    eta2B(i) = wt*(Y2B_biom(i)+sum(a_tune2'.*Y2B_met(:,i)));
end

eta2A = zeros(length(F2A),1);

for i = 1:length(F2A)
    eta2A(i) = wt*(Y2A_atp(i)+sum(a_tune2'.*Y2A_met(:,i)));
end

% Lumped EFV matrices

ZF1C = zeros(length(rxn),1);
ZF2B = zeros(length(rxn),1);
ZF2A = zeros(length(rxn),1);

for i = 1:length(rxn)
    ZF1C(i) = sum(z1C(i,:)'.*(eta1C.^(3*n_eta)))/sum(eta1C.^(3*n_eta));
    ZF2B(i) = sum(z2B(i,:)'.*(eta2B.^(3*n_eta)))/sum(eta2B.^(3*n_eta));
    ZF2A(i) = sum(z2A(i,:)'.*(eta2A.^(3*n_eta)))/sum(eta2A.^(3*n_eta));
end

ZF1 = ZF1C;
ZF2 = ZF2B+ZF2A;

% Final yields of metabolites after EFV lumping

Y1_met_lump = zeros(length(idx_met),1);                 % mol/mol
Y2_met_lump = zeros(length(idx_met),1);

for i = 1:length(idx_met)
    Y1_met_lump(i) = (ZF1(idx_met(i))/ZF1(idx_cb));
    Y2_met_lump(i) = (ZF2(idx_met(i))/ZF2(idx_cb));
end

Y1_E = (ZF1(idx_csm)/ZF1(idx_cb))*1000;                 % g/mol
Y2_E = (ZF2(idx_csm)/ZF2(idx_cb))*1000;

Y2_B = biom_DCW_prtn*(ZF2(idx_biom)/ZF2(idx_cb))*1000;  % g/mol

end
```

```
173
174 %%-------------------------------------------------------------------------
175
176 function [x_piv,R_in,layers,C_T,C_S,C_INT,MnT,MwT,pdiT,q,vm,R0,ro,L,n,...
177     exp_ratio] = MLPBM_setting(p,N,prm,mss,MnL,MwL,MnH,MwH)
178
179 % General settings
180
181 vm = 1;   % Size of monomer
182
183 qmax = 1+(log(N/(p+1))/log(1+(vm/(p+1))));  % Maximum pivots in cont region
184 q = floor(qmax);                            % No. of pivots in cont region
185 ratio = (N/(p+1))^(1/(q-1));                % Ratio of geometric progr
186
187 % Pivots for discrete-continuous mesh
188
189 x_piv = zeros(p+q,1);
190 x_piv(1:p) = 1:p;
191
192 for i = p+1:p+q
193     x_piv(i) = (p+1)*ratio^(i-(p+1));       % Geometric mesh
194 end
195
196 % Boundary points for discrete-continuous mesh
197
198 x_bound = zeros(p+q+1,1);
199 x_bound(1) = 0.5;
200
201 for i=2:p+q
202     x_bound(i) = (x_piv(i)+x_piv(i-1))/2;
203 end
204
205 x_bound(p+q+1) = x_piv(end);
206
207 % Initial distribution
208
209 c_in = @(x,alpha,beta,pin) pin*gampdf(x,alpha,beta); % Gamma distribution
210
211 % Physical properties
212
213 layers = round(2/(((mss*prm(12))/prm(11))/mss));     % Total no. of layers
214 ro = 1500;                    % Cellulose density, g/L
215 L = (N/2)*10.38e-10;          % Length of microfibril, m
216 R0 = 1e-9;                    % Width of single layer, m
217 R_in = layers*R0;             % Total radius of particle, m
218 n = mss/(ro*pi*(R_in^2)*L);   % No. of particles per unit vol, 1/m3
219 p_zone = prm(11);             % No. of layers in penetration zone
220 i_zone = (R_in/R0)-p_zone;    % No. of layers in internal zone
221
222 % Radiuses of discrete layers
223
224 R_set = sort(0:R0:R_in,'descend')';
225
226 % Mass of polymers in each layer
227
228 mass_l = zeros(length(R_set)-1,1);
229
230 for i = 2:length(R_set)
231     mass_l(i-1) = (n*ro*pi*(R_set(i-1)^2)*L)-(n*ro*pi*(R_set(i)^2)*L);
```

```
232 end
233
234 % Initial cellulose distribution in each layer
235
236 pin_l = zeros(length(mass_l),1);
237 c_in_tl = zeros(length(mass_l),p+q); % Molar concentration density
238 C_in_tl = zeros(length(mass_l),p+q); % Molar concentration
239
240 optionsfmincon = optimoptions('fsolve','display','none');
241
242 for i = 1:p_zone
243     pin_l(i) = fsolve(@(x) sum(c_in(x_piv,MnL/(MwL-MnL),MwL-MnL,x)...
244         .*(x_bound(2:end)-x_bound(1:p+q)).*(162*x_piv+18))-mass_l(i),0,...
245         optionsfmincon);
246     c_in_tl(i,:) = c_in(x_piv,MnL/(MwL-MnL),MwL-MnL,pin_l(i));
247 end
248
249 for i = p_zone+1:p_zone+i_zone
250     pin_l(i) = fsolve(@(x) sum(c_in(x_piv,MnH/(MwH-MnH),MwH-MnH,x)...
251         .*(x_bound(2:end)-x_bound(1:p+q)).*(162*x_piv+18))-mass_l(i),0,...
252         optionsfmincon);
253     c_in_tl(i,:) = c_in(x_piv,MnH/(MwH-MnH),MwH-MnH,pin_l(i));
254 end
255
256 for i = 1:p+q
257     C_in_tl(:,i) = c_in_tl(:,i).*(x_bound(i+1)-x_bound(i));
258 end
259
260 % Overall initial cellulose distribution
261
262 C_T = zeros(p+q,1);
263
264 for i = 1:p+q
265     C_T(i) = sum(C_in_tl(:,i));
266 end
267
268 % Surface initial cellulose distribution
269
270 C_S = zeros(p+q,1);
271
272 for i = 1:p+q
273     C_S(i) = sum(C_in_tl(1,i));
274 end
275
276 % Internal initial cellulose distribution
277
278 C_INT = zeros(p+q,1);
279
280 for i = 1:p+q
281     C_INT(i) = sum(C_in_tl(2:end,i));
282 end
283
284 % Ratio of molar concentration to total mass in each layer
285
286 C_ratio = zeros(length(R_set)-1,p+q);
287
288 for i = 1:length(R_set)-1
289     for j = 1:p+q
290         C_ratio(i,j) = C_in_tl(i,j)/sum(C_in_tl(i,:)'.*(162*x_piv+18));
```

```
    end
end

% Piece-wise function for ratio of molar concentration to total mass as a
% function of particle radius

exp_ratio = cell(p+q,1);

for i = 1:p+q
    exp_ratio{i} = mkpp(flipud(R_set),flipud(C_ratio(:,i)));
end

% General properties of initial cellulose distribution

MnT = sum(C_T.*x_piv)/sum(C_T);
MwT = sum(C_T.*(x_piv.^2))/sum(C_T.*x_piv);
pdiT = MwT/MnT;

end

%%-------------------------------------------------------------------------

function [t,y,eF_max,kmax,K,K_inh_G,K_inh_etoh] = ODE_setting(p,q,C_S,...
    C_INT,prm,x_piv,vm,t_end,Y1_met_lump,Y2_met_lump,Y1_E,Y2_E,Y2_B,...
    CB_pc,cellobiose,biomass,MW_cellome,mol_exo,mol_endo,R0,ro,L,n,...
    exp_ratio)

% Parameter setting

% ML-PBM parameters

% Exo-enzyme-related parameters

kp_h_exo = prm(1);                      % Rate constant of hydrolysis, 1/s
m_h_exo = 0;
k_h_exo = kp_h_exo*(x_piv.^m_h_exo);    % Rate coefficients of hydrolysis, 1/s

kp_f_exo = prm(2);                      % Rate constant of complexation, DP.L/mol.s
m_f_exo = -1;
k_f_exo = kp_f_exo*(x_piv.^m_f_exo);    % Rate coefficients of complexation, L/mol.s

kp_e_exo = prm(3);                      % Rate constant of decomplexation, 1/s
m_e_exo = 0;
k_e_exo = kp_e_exo*(x_piv.^m_e_exo);    % Rate coefficients of decomplexation, 1/s

tau_exo = prm(4);              % Enzyme footprint, mol/m2
k_ads_exo = prm(5);            % Adsorption constant, L/mol.s
k_des_exo = prm(6);            % Desorption constant, 1/s

k_If_exo = [prm(7) prm(8)];    % Forward glucose/cellobiose inhibition, L/mol.s
k_Ir_exo = [prm(9) prm(10)];   % Reverse glucose/cellobiose inhibition, 1/s

% Endo-enzyme-related parameters

kp_h_endo = prm(13);                    % Rate constant of hydrolysis (insoluble cellulose), 1/s
m_h_endo = 0;                           % Endo-enzymes in C.th unaffected by crystallinity
k_h_endo = kp_h_endo*(x_piv.^m_h_endo); % Rate coefficients of hydrolysis(insoluble cellulose), 1/s

kp_hs_endo = prm(14);           % Rate constant of hydrolysis (soluble cellulose), 1/s
```

```
m_hs_endo = 0;
k_hs_endo = kp_hs_endo*(x_piv.^m_hs_endo);
k_hs_endo(2) = 0;               % Rate coefficients of hydrolysis (soluble cellulose), L/mol.s
                                % Endo-enzymes in C.th do not hydrolyze cellobiose

kp_f_endo = prm(15);                    % Rate constant of complexation, L/mol.s
m_f_endo = 0;                           % Endo-enzymes in C.th unaffected by crystallinity
k_f_endo = kp_f_endo*(x_piv.^m_f_endo); % Rate coefficients of complexation, L/mol.s

kp_e_endo = prm(16);                    % Rate constant of decomplexation, 1/s
m_e_endo = 0;
k_e_endo = kp_e_endo*(x_piv.^m_e_endo); % Rate coefficients of decomplexation, 1/s

tau_endo = prm(17);                     % Enzyme footprint, mol/m2
k_ads_endo = prm(18);                   % Adsorption constant, L/mol.s
k_des_endo = prm(19);                   % Desorption constant, 1/s

k_If_endo = [prm(20) prm(21)];          % Forward glucose/cellobiose inhibition, L/mol.s
k_Ir_endo = [prm(22) prm(23)];          % Reverse glucose/cellobiose inhibition, 1/s

% L-HCM parameters

aF(1) = prm(24);        % Constitutive intracellular enzyme synthesis rate (F1), 1/s
aF(2) = prm(25);        % Constitutive intracellular enzyme synthesis rate (F2), 1/s
bF(1) = prm(26);        % Intracellular enzyme decay rate (F1),1/s
bF(2) = prm(27);        % Intracellular enzyme decay rate (F2),1/s
kE(1) = prm(28);        % Inducive intracellular enzyme synthesis rate (F1), 1/s
kE(2) = prm(29);        % Inducive intracellular enzyme synthesis rate (F2), 1/s
KE(1) = prm(30);        % MM constant for intracellular enzyme synthesis (F1), mol/L
KE(2) = prm(31);        % MM constant for intracellular enzyme synthesis (F2), mol/L
K_inh_G = prm(32);      % Glucose inhibition constant, mol/L
K_inh_etoh = prm(33);   % Ethanol inhibition constant, mol/L
kmax(1) = prm(34);      % Maximum substrate uptake rate (F1), mol/g-biom.s
kmax(2) = prm(35);      % Maximum substrate uptake rate (F2), mol/g-biom.s
K(1) = prm(36);         % MM constant for substrate uptake (F1), mol/L
K(2) = prm(37);         % MM constant for substrate uptake (F2), mol/L
aE = prm(38);           % Constitutive cellulase synthesis rate, g/g-biom.s

% Maximum intracellular enzyme levels

eF_max(1) = (aF(1)+kE(1))/bF(1);
eF_max(2) = (aF(2)+kE(2))/(bF(2)+(Y2_B*kmax(2)));

% Initial intracellular enzyme levels

[init_e] = int_enz(aF,kE,bF,Y2_B,kmax,CB_pc,K,KE);

% Fixed pivot technique

[n_exo,n_endo] = fp(p,q,vm,x_piv);      % Particle allocation functions

% Initial conditions

init = zeros(6*p+6*q+17,1);
init(1:p+q) = 0;                        % 1 to p+q --> free liberated polymers
init(2) = cellobiose/342;               % 2 --> cellobiose
init(p+q+1:2*p+2*q) = 0;                % p+q+1 to 2p+2q --> CBH-bound polymers
init(2*p+2*q+1:3*p+3*q) = 0;            % 2p+2q+1 to 3p+3q --> EG-bound polymers
init(3*p+3*q+1:4*p+4*q) = 0;            % 3p+3q+1 to 4p+4q --> EG-CBH-bound polymers
```

```
409 init(4*p+4*q+1:5*p+5*q) = C_S;          % 4p+4q+1 to 5p+5q --> surface-accessible polymer
410 init(5*p+5*q+1:6*p+6*q) = C_INT;        % 5p+5q+1 to 6p+6q --> internal inaccessible polymer
411 init(6*p+6*q+1) = 0;                    % 6p+6q+1 --> surface adsorbed EG
412 init(6*p+6*q+2) = 0;                    % 6p+6q+2 --> surface adsorbed CBH
413 init(6*p+6*q+3:6*p+6*q+4) = 0;          % 6p+6q+3 to 6p+6q+4 --> inhibited EG
414 init(6*p+6*q+5:6*p+6*q+6) = 0;          % 6p+6q+5 to 6p+6q+6 --> inhibited free CBH
415 init(6*p+6*q+7:6*p+6*q+8) = 0;          % 6p+6q+7 to 6p+6q+8 --> inhibited adsorbed CBH
416 init(6*p+6*q+9) = 0;                    % 6p+6q+9 --> inhibited BG (placeholder for future ...
        extension)
417 init(6*p+6*q+10) = init_e(1)*eF_max(1); % 6p+6q+10 --> intracellular enzyme (F1)
418 init(6*p+6*q+11) = init_e(2)*eF_max(2); % 6p+6q+11 --> intracellular enzyme (F2)
419 init(6*p+6*q+12) = 0;                   % 6p+6q+12 --> etoh
420 init(6*p+6*q+13) = 0;                   % 6p+6q+13 --> lac
421 init(6*p+6*q+14) = 0;                   % 6p+6q+14 --> form
422 init(6*p+6*q+15) = 0;                   % 6p+6q+15 --> ace
423 init(6*p+6*q+16) = 0;                   % 6p+6q+16 --> cellome
424 init(6*p+6*q+17) = biomass;             % 6p+6q+17 --> biomass
425
426 % ODE solver
427
428 tstart = 0;
429 tfinal = t_end;
430
431 tic;
432
433 options = odeset('AbsTol',1e-10,'RelTol',1e-10,'NonNegative',1:(6*p+6*q+17));
434
435 [t,y] = ode15s(@(t,y) UCPBM_ode(p,q,n_exo,n_endo,k_h_exo,k_h_endo,k_hs_endo,...
436     k_f_exo,k_f_endo,k_e_exo,k_e_endo,k_ads_endo,k_des_endo,k_ads_exo,...
437     k_des_exo,tau_endo,tau_exo,k_If_exo,k_Ir_exo,k_If_endo,k_Ir_endo,...
438     x_piv,R0,ro,L,n,exp_ratio,kmax,K,kE,K_inh_G,K_inh_etoh,eF_max,...
439     Y2_B,aF,bF,Y1_met_lump,Y2_met_lump,aE,Y1_E,...
440     Y2_E,MW_cellome,mol_exo,mol_endo,y,t),[tstart tfinal],init,options);
441
442 time_fp_sim = '\nTotal time elapsed for UC-PBM simulation is %.1f seconds.\n';
443
444 fprintf(time_fp_sim,toc)
445
446 end
447
448 %%-----------------------------------------------------------------------
449
450 function [init_e] = int_enz(aF,kE,bF,Y2_B,kmax,CB_pc,K,KE)
451
452 e_rel_lb(1) = aF(1)/(aF(1)+kE(1));
453 e_rel_lb(2) = (aF(2)/(bF(2)+Y2_B*kmax(2)*((CB_pc/342)/(K(2)+(CB_pc/342)))))*...
454     ((bF(2)+Y2_B*kmax(2))/(aF(2)+kE(2)));
455 e_rel_ub(1) = (aF(1)+kE(1)*((CB_pc/342)/(KE(1)+(CB_pc/342))))/(aF(1)+kE(1));
456 e_rel_ub(2) = ((aF(2)+kE(2)*((CB_pc/342)/(KE(2)+(CB_pc/342))))/...
457     (bF(2)+Y2_B*kmax(2)*((CB_pc/342)/(K(2)+(CB_pc/342)))))*...
458     ((bF(2)+Y2_B*kmax(2))/(aF(2)+kE(2)));
459
460 if CB_pc == 0
461     init_e(1) = e_rel_ub(1);
462     init_e(2) = e_rel_lb(2);
463 else
464     init_e(1) = e_rel_lb(1);
465     init_e(2) = e_rel_ub(2);
466 end
```

```
467
468 end
469
470 %%-------------------------------------------------------------------------
471
472 function [n_exo,n_endo] = fp(p,q,vm,x_piv)
473
474 % Particle allocation function - fixed pivot technique
475
476 % Chain-end dimer scission
477
478 n_exo = zeros(p+q,p+q);
479 n_exo(1,3) = 1;
480 n_exo(2,3) = 1; n_exo(2,4) = 2; n_exo(2,5:end) = 1;
481
482 for i = 3:p+q
483     for j = i:p+q
484         if j==i && x_piv(j)-2*vm>x_piv(i-1)
485             n_exo(i,j) = ((x_piv(j)-2*vm)-x_piv(i-1))/(x_piv(i)-x_piv(i-1));
486         elseif j==i && x_piv(j)-2*vm==x_piv(i-1)
487             n_exo(i,j) = 0;
488         elseif j~=i && x_piv(j)-2*vm>x_piv(i-1) && x_piv(j)-2*vm<x_piv(i)
489             n_exo(i,j) = ((x_piv(j)-2*vm)-x_piv(i-1))/(x_piv(i)-x_piv(i-1));
490         elseif j~=i && x_piv(j)-2*vm>x_piv(i) && x_piv(j)-2*vm<x_piv(i+1)
491             n_exo(i,j) = (x_piv(i+1)-(x_piv(j)-2*vm))/(x_piv(i+1)-x_piv(i));
492         elseif j~=i && x_piv(j)-2*vm==x_piv(i)
493             n_exo(i,j) = 1;
494         elseif j~=i && x_piv(j)-2*vm==x_piv(i-1)
495             n_exo(i,j) = 0;
496         else
497             n_exo(i,j) = 0;
498         end
499     end
500     n_exo(i,i) = n_exo(i,i)-1;
501 end
502
503 % Random scission
504
505 n_endo = zeros(p+q,p+q);
506
507 for j = 2:p+q
508     n_endo(1,j) = 2/(x_piv(j-1));
509 end
510
511 for i = 2:p
512     for j = i+1:p+q
513         n_endo(i,i) = -1;
514         n_endo(i,j) = 2/(x_piv(j-1));
515     end
516 end
517
518 for i = p+1:p+q
519     for j = i+1:p+q
520         n_endo(i,j) = (x_piv(i+1)-x_piv(i-1))/(x_piv(j-1));
521     end
522     n_endo(i,i) = -1;
523 end
524
525 end
```

```
526
527 %%-------------------------------------------------------------------------
528
529 function ode = UCPBM_ode(p,q,n_exo,n_endo,k_h_exo,k_h_endo,k_hs_endo,...
530     k_f_exo,k_f_endo,k_e_exo,k_e_endo,k_ads_endo,k_des_endo,k_ads_exo,...
531     k_des_exo,tau_endo,tau_exo,k_If_exo,k_Ir_exo,k_If_endo,k_Ir_endo,...
532     x_piv,R0,ro,L,n,exp_ratio,kmax,K,kE,K_inh_G,K_inh_etoh,eF_max,...
533     Y2_B,aF,bF,Y1_met_lump,Y2_met_lump,aE,Y1_E,...
534     Y2_E,MW_cellome,mol_exo,mol_endo,y,t)
535
536 dydt = zeros(6*p+6*q+17,1);
537
538 % Enzymes
539
540 E_T_endo = mol_endo*(y(6*p+6*q+16)/MW_cellome);
541 E_T_exo = mol_exo*(y(6*p+6*q+16)/MW_cellome);
542
543 E_F_endo = E_T_endo-y(6*p+6*q+1)-sum(y(2*p+2*q+1:4*p+4*q))...
544     -sum(y(6*p+6*q+3:6*p+6*q+4));
545 E_F_exo = E_T_exo-y(6*p+6*q+2)-sum(y(p+q+1:2*p+2*q))...
546     -sum(y(3*p+3*q+1:4*p+4*q))-sum(y(6*p+6*q+5:6*p+6*q+8));
547
548 E_F_BG = 0; k_h_BG = 0; k_If_BG = 0; k_Ir_BG = 0; % Placeholder for future extension
549
550 % Particle radius
551
552 x3 = [x_piv;x_piv;x_piv];
553 x5 = [x_piv;x_piv;x_piv;x_piv;x_piv];
554 R = sqrt(sum(y(p+q+1:6*p+6*q).*(162*x5+18))/(ro*pi*L*n));
555
556 % Total surface area
557
558 As = (n*2*pi*R*L)/1000;
559
560 if R-R0<=0
561     R_ratio = 0;
562 else
563     R_ratio = (R-R0)/R;
564 end
565
566 % L-HCM
567
568 rF_kin = zeros(2,1);     % Unregulated cellobiose uptake flux, mol/g-biom.s
569 rFE_kin = zeros(2,1);    % Unregulated inducive intracellular enzyme synthesis rates, 1/s
570 eF_rel = zeros(2,1);     % Relative intracellular enzyme levels
571 ROI = zeros(2,1);        % ROIs
572
573 for i = 1:2
574     rF_kin(i) = kmax(i)*(y(2)/(K(i)+y(2)))*(1/(1+(y(1)/K_inh_G)))...
575         *(1/(1+(y(6*p+6*q+12)/K_inh_etoh)));
576     rFE_kin(i) = kE(i)*(y(2)/(K(i)+y(2)))*(1/(1+(y(1)/K_inh_G)))...
577         *(1/(1+(y(6*p+6*q+12)/K_inh_etoh)));
578     eF_rel(i) = y(i+6*p+6*q+10-1)/eF_max(i);
579     ROI(i) = eF_rel(i)*rF_kin(i);
580 end
581
582 uF = zeros(2,1);         % Cybernetic variables
583 vF = zeros(2,1);         % Cybernetic variables
584 rF = zeros(2,1);         % Regulated cellobiose uptake flux, mol/g-biom.s
```

```
585
586 for i = 1:2
587     uF(i) = ROI(i)/sum(ROI);
588     vF(i) = ROI(i)/max(ROI);
589     rF(i) = vF(i)*eF_rel(i)*rF_kin(i);
590 end
591
592 mu = Y2_B*rF(2);        % Specific biomass growth rate, 1/s
593
594 % Soluble polymers
595
596 dydt(1) = sum(n_exo(1,:)'.*k_h_exo.*y(p+q+1:2*p+2*q))...
597     +sum(n_endo(1,7:end)'.*k_h_endo(7:end).*y(2*p+2*q+7:3*p+3*q))...
598     +sum(n_endo(1,4:end)'.*k_h_endo(4:end).*y(3*p+3*q+4:4*p+4*q))...
599     +sum(E_F_endo*n_endo(1,2:6)'.*k_hs_endo(2:6).*y(2:6))...
600     +2*k_h_BG*E_F_BG*y(2)+k_Ir_exo(1)*y(6*p+6*q+5)+k_Ir_exo(1)*y(6*p+6*q+7)...
601     +k_Ir_endo(1)*y(6*p+6*q+3)+k_Ir_BG*y(6*p+6*q+9)...
602     -k_If_exo(1)*E_F_exo*y(1)-k_If_exo(1)*y(6*p+6*q+2)*y(1)...
603     -k_If_endo(1)*E_F_endo*y(1)-k_If_BG*E_F_BG*y(1);
604
605 dydt(2) = sum(n_exo(2,:)'.*k_h_exo.*y(p+q+1:2*p+2*q))...
606     +sum(n_endo(2,7:end)'.*k_h_endo(7:end).*y(2*p+2*q+7:3*p+3*q))...
607     +sum(n_endo(2,4:end)'.*k_h_endo(4:end).*y(3*p+3*q+4:4*p+4*q))...
608     +sum(E_F_endo*n_endo(2,3:6)'.*k_hs_endo(3:6).*y(3:6))...
609     -k_hs_endo(2)*E_F_endo*y(2)-k_h_BG*E_F_BG*y(2)...
610     +k_Ir_exo(2)*y(6*p+6*q+6)+k_Ir_exo(2)*y(6*p+6*q+8)+k_Ir_endo(2)*y(6*p+6*q+4)...
611     -k_If_exo(2)*E_F_exo*y(2)-k_If_exo(2)*y(6*p+6*q+2)*y(2)-k_If_endo(2)*E_F_endo*y(2)...
612     -sum(rF)*y(6*p+6*q+17);
613
614 for i = 3:5
615     dydt(i) = sum(n_endo(i,7:end)'.*k_h_endo(7:end).*y(2*p+2*q+7:3*p+3*q))...
616         +sum(n_endo(i,i+1:end)'.*k_h_endo(i+1:end).*y(3*p+3*q+i+1:4*p+4*q))...
617         +sum(E_F_endo*n_endo(i,i+1:6)'.*k_hs_endo(i+1:6).*y(i+1:6))...
618         -k_hs_endo(i)*E_F_endo*y(i);
619 end
620
621 for i = 6
622     dydt(i) = sum(n_endo(i,i+1:end)'.*k_h_endo(i+1:end).*y(2*p+2*q+i+1:3*p+3*q))...
623         +sum(n_endo(i,i+1:end)'.*k_h_endo(i+1:end).*y(3*p+3*q+i+1:4*p+4*q))...
624         -k_hs_endo(i)*E_F_endo*y(i);
625 end
626
627 % Exo-enzyme-bound surface polymers
628
629 dydt(p+q+3) = sum(n_exo(3,3:end)'.*k_h_exo(3:end).*y(p+q+3:2*p+2*q));
630
631 for i = p+q+4:p+q+6
632     dydt(i) = sum(n_exo(i-p-q,:)'.*k_h_exo.*y(p+q+1:2*p+2*q))...
633         -(k_f_endo(i-p-q)*y(6*p+6*q+1)*y(i))...
634         +(k_e_endo(i-p-q)*y(i-p-q+3*p+3*q));
635 end
636
637 for i = p+q+7:2*p+2*q-1
638     dydt(i) = sum(n_exo(i-p-q,:)'.*k_h_exo.*y(p+q+1:2*p+2*q))...
639         +(y(6*p+6*q+2)*k_f_exo(i-p-q)*y(i-p-q+4*p+4*q))...
640         -(k_e_exo(i-p-q)*y(i))...
641         -(k_f_endo(i-p-q)*y(6*p+6*q+1)*y(i))...
642         +(k_e_endo(i-p-q)*y(i-p-q+3*p+3*q));
643
```

```
644 end
645
646 for i = 2*p+2*q
647     dydt(i) = sum(n_exo(i-p-q,:)'.*k_h_exo.*y(p+q+1:2*p+2*q))...
648         +(y(6*p+6*q+2)*k_f_exo(i-p-q)*y(i-p-q+4*p+4*q))...
649         -(k_e_exo(i-p-q)*y(i))...
650         -(k_f_endo(i-p-q)*y(6*p+6*q+1)*y(i))...
651         +(k_e_endo(i-p-q)*y(i-p-q+3*p+3*q));
652 end
653
654 % Endo-enzyme-bound surface polymers
655
656 for i = 2*p+2*q+7:3*p+3*q
657     dydt(i) = (k_f_endo(i-2*p-2*q)*y(6*p+6*q+1)*y(i-2*p-2*q+4*p+4*q))...
658         -(k_e_endo(i-2*p-2*q)*y(i))...
659         -(k_h_endo(i-2*p-2*q)*y(i));
660 end
661
662 % Exo-endo-bound surface polymers
663
664 for i = 3*p+3*q+4:4*p+4*q
665     dydt(i) = (k_f_endo(i-3*p-3*q)*y(6*p+6*q+1)*y(i-3*p-3*q+p+q))...
666         -(k_e_endo(i-3*p-3*q)*y(i))...
667         -(k_h_endo(i-3*p-3*q)*y(i));
668 end
669
670 % Surface polymers
671
672 m = zeros(p+q,1);
673
674 for i = 4*p+4*q+7:5*p+5*q-1
675     m(i-4*p-4*q) = (sum(n_endo(i-4*p-4*q,i-4*p-4*q+1:end)'...
676         .*k_h_endo(i-4*p-4*q+1:end).*y(i-4*p-4*q+2*p+2*q+1:3*p+3*q))...
677         +sum(n_endo(i-4*p-4*q,i-4*p-4*q+1:end)'...
678         .*k_h_endo(i-4*p-4*q+1:end).*y(i-4*p-4*q+1+3*p+3*q:4*p+4*q))...
679         -(k_f_exo(i-4*p-4*q)*y(6*p+6*q+2)*y(i))...
680         +(k_e_exo(i-4*p-4*q)*y(i-4*p-4*q+p+q))...
681         -(k_f_endo(i-4*p-4*q)*y(6*p+6*q+1)*y(i))...
682         +(k_e_endo(i-4*p-4*q)*y(i-4*p-4*q+2*p+2*q)));
683 end
684
685 m(p+q) = (-(k_f_exo(p+q)*y(6*p+6*q+2)*y(5*p+5*q))...
686     +(k_e_exo(p+q)*y(2*p+2*q))...
687     -(k_f_endo(p+q)*y(6*p+6*q+1)*y(5*p+5*q))...
688     +(k_e_endo(p+q)*y(3*p+3*q)));
689
690 for i = 4*p+4*q+7:5*p+5*q
691     dydt(i) = m(i-4*p-4*q)-(sum(m.*(162*x_piv+18))...
692         +sum(dydt(p+q+1:4*p+4*q).*(162*x3+18)))*R_ratio...
693         *ppval(exp_ratio{i-4*p-4*q},R-R0);
694 end
695
696 % Internal inaccessible polymer
697
698 for i = 5*p+5*q+7:6*p+6*q
699     dydt(i) = (sum(m.*(162*x_piv+18))...
700         +sum(dydt(p+q+1:4*p+4*q).*(162*x3+18)))*R_ratio...
701         *ppval(exp_ratio{i-5*p-5*q},R-R0);
702 end
```

```
703
704 % Surface-adsorbed endo-enzymes
705
706 dydt(6*p+6*q+1) = (k_ads_endo*E_F_endo*(As*tau_endo-y(6*p+6*q+1)))...
707     -(k_des_endo*y(6*p+6*q+1))...
708     -sum(y(6*p+6*q+1)*k_f_endo(7:end).*y(4*p+4*q+7:5*p+5*q))...
709     +sum(k_e_endo(7:end).*y(2*p+2*q+7:3*p+3*q))...
710     -sum(y(6*p+6*q+1)*k_f_endo(4:end).*y(p+q+4:2*p+2*q))...
711     +sum(k_e_endo(4:end).*y(3*p+3*q+4:4*p+4*q));
712
713 % Surface-adsorbed exo-enzymes
714
715 dydt(6*p+6*q+2) = (k_ads_exo*E_F_exo*(As*tau_exo-y(6*p+6*q+2)))...
716     -(k_des_exo*y(6*p+6*q+2))...
717     -sum(y(6*p+6*q+2)*k_f_exo(7:end).*y(4*p+4*q+7:5*p+5*q))...
718     +sum(k_e_exo(7:end).*y(p+q+7:2*p+2*q))...
719     +k_Ir_exo(1)*y(6*p+6*q+7)+k_Ir_exo(2)*y(6*p+6*q+8)...
720     -y(6*p+6*q+2)*sum(k_If_exo'.*y(1:2));
721
722 % Inhibited endo-enzymes
723
724 for i = 6*p+6*q+3:6*p+6*q+4
725     dydt(i) = k_If_endo(i-6*p-6*q-3+1)*E_F_endo*y(i-6*p-6*q-3+1)...
726         -k_Ir_endo(i-6*p-6*q-3+1)*y(i);
727 end
728
729 % Inhibited exo-enzymes
730
731 for i = 6*p+6*q+5:6*p+6*q+6
732     dydt(i) = k_If_exo(i-6*p-6*q-5+1)*E_F_exo*y(i-6*p-6*q-5+1)...
733         -k_Ir_exo(i-6*p-6*q-5+1)*y(i);
734 end
735
736 for i = 6*p+6*q+7:6*p+6*q+8
737     dydt(i) = k_If_exo(i-6*p-6*q-7+1)*y(6*p+6*q+2)*y(i-6*p-6*q-7+1)...
738         -k_Ir_exo(i-6*p-6*q-7+1)*y(i);
739 end
740
741 % Inhibited BG (placeholder for future extension)
742
743 for i = 6*p+6*q+9;
744     dydt(i) = k_If_BG*E_F_BG*y(i-6*p-6*q-9+1)...
745         -k_Ir_BG*y(i);
746 end
747
748 % Intracellular enzymes
749
750 for i = 6*p+6*q+10:6*p+6*q+11
751     dydt(i) = aF(i-6*p-6*q-10+1)+uF(i-6*p-6*q-10+1)...
752         *rFE_kin(i-6*p-6*q-10+1)-bF(i-6*p-6*q-10+1)*y(i)-mu*y(i);
753 end
754
755 % Metabolites
756
757 for i = 6*p+6*q+12:6*p+6*q+15
758     dydt(i) = (Y1_met_lump(i-6*p-6*q-12+1)*rF(1)...
759         +Y2_met_lump(i-6*p-6*q-12+1)*rF(2))*y(6*p+6*q+17);
760 end
761
```

```
762 % Cellulosome
763
764 dydt(6*p+6*q+16) = aE*y(6*p+6*q+17)+((Y1_E*rF(1)+Y2_E*rF(2))*y(6*p+6*q+17));
765
766 % Biomass
767
768 dydt(6*p+6*q+17) = mu*y(6*p+6*q+17);
769
770 % Output
771
772 ode = dydt;
773
774 end
775
776 %%-------------------------------------------------------------------------
777
778 function [C,CXS,CNS,CNXS,CFS,CS,CI,CT,CC,E_T_exo,E_T_endo,CB,etoh,lac,...
779     form,ace,cellome,biom,cellulase_tot,cellu,cell_gluc_eq,eF_rel,R,...
780     conv_rem,conv_cb,uF,vF,r_up,mu] = post_processing(t,y,p,q,MW_cellome,...
781     mol_endo,mol_exo,mass_frac_cellulase,x_piv,eF_max,ro,L,n,mss,...
782     cellobiose,kmax,K,K_inh_G,K_inh_etoh,Y2_B)
783
784 % Molar concentrations, mol/L
785
786 C = y(:,1:p+q);                     % Soluble products
787 CXS = y(:,p+q+1:2*p+2*q);           % CBH-bound complex
788 CNS = y(:,2*p+2*q+1:3*p+3*q);       % EG-bound complex
789 CNXS = y(:,3*p+3*q+1:4*p+4*q);      % CBH-EG-bound complex
790 CFS = y(:,4*p+4*q+1:5*p+5*q);       % Free un-bound surface polymers
791 CS = CXS+CNS+CNXS+CFS;              % Total surface polymers
792 CI = y(:,5*p+5*q+1:6*p+6*q);        % Internal polymers
793 CT = C+CS+CI;                       % Total polymers
794 CC = CS+CI;                         % Insoluble polymers
795
796 E_T_endo = mol_endo*(y(:,6*p+6*q+16)/MW_cellome);   % Total endo-enzymes
797 E_T_exo = mol_exo*(y(:,6*p+6*q+16)/MW_cellome);     % Total exo-enzymes
798
799 E_S_endo = y(:,6*p+6*q+1);  % Surface-adsorbed endo-enzymes
800 E_S_exo = y(:,6*p+6*q+2);   % Surface-adsorbed exo-enzymes
801
802 CNI(:,1) = y(:,6*p+6*q+3);  % Inhibited endo-enzymes
803 CNI(:,2) = y(:,6*p+6*q+4);
804
805 CXI(:,1) = y(:,6*p+6*q+5)+y(:,6*p+6*q+7);   % Inhibited exo-enzymes
806 CXI(:,2) = y(:,6*p+6*q+6)+y(:,6*p+6*q+8);
807
808 E_F_endo = zeros(length(t),1);  % Free endo-enzymes
809 E_F_exo = zeros(length(t),1);   % Free exo-enzymes
810
811 for i = 1:length(t)
812     E_F_endo(i) = E_T_endo(i)-E_S_endo(i)-sum(CNS(i,:))-sum(CNXS(i,:))-sum(CNI(i,:));
813     E_F_exo(i) = E_T_exo(i)-E_S_exo(i)-sum(CXS(i,:))-sum(CXI(i,:))-sum(CNXS(i,:));
814 end
815
816 % Molecular weight of metabolites, g/mol
817
818 mw_met = [46.07;90.08;45.017;59.044];
819
820 % Mass Concentrations, g/L
```

```
821
822 CB = y(:,2).*342;                          % Cellobiose
823 cellome = y(:,6*p+6*q+16);                 % Cellulosome
824 biom = y(:,6*p+6*q+17);                    % Biomass
825 etoh = y(:,6*p+6*q+12).*mw_met(1);         % Ethanol
826 lac = y(:,6*p+6*q+13).*mw_met(2);          % Lactate
827 form = y(:,6*p+6*q+14).*mw_met(3);         % Formate
828 ace = y(:,6*p+6*q+15).*mw_met(4);          % Acetate
829
830 cellulase_tot = y(:,6*p+6*q+16).*mass_frac_cellulase;   % Total cellulase
831
832 cellu = zeros(length(t),1);     % Remaining insoluble cellulose
833
834 for i = 1:length(t)
835     cellu(i) = sum(CC(i,:)'.*(162*x_piv+18));
836 end
837
838 cell_gluc_eq = cellu.*(180/162); % Remaining insoluble cellulose (g gluc/L)
839
840 % Relative intracellular enzyme levels
841
842 eF_rel(:,1) = y(:,6*p+6*q+10)/eF_max(1);
843 eF_rel(:,2) = y(:,6*p+6*q+11)/eF_max(2);
844
845 % Transient particle radius, m
846
847 R = zeros(length(t),1);
848
849 for i = 1:length(t)
850     R(i) = sqrt(sum(CC(i,:)'.*(162*x_piv+18))/(ro*pi*L*n));
851 end
852
853 % Conversion
854
855 conv_rem = zeros(length(t),1);  % Conversion of cellulose
856
857 for i = 1:length(t)
858     conv_rem(i) = (mss-cellu(i))/mss;
859 end
860
861 conv_cb = zeros(length(t),1);  % Conversion of cellobiose
862 for i = 1:length(t);
863     conv_cb(i) = (cellobiose-CB(i))/cellobiose;
864 end
865
866 % Metabolic regulations
867
868 ROI = zeros(length(t),2);       % ROIs
869 rF_kin = zeros(length(t),2);    % Unregulated cellobiose uptake flux, mol/g-biom.s
870
871 for i = 1:length(t)
872     rF_kin(i,1) = (kmax(1)*(y(i,2)/(K(1)+y(i,2)))...
873         *(1/(1+(y(i,1)/K_inh_G)))*(1/(1+(y(i,6*p+6*q+12)/K_inh_etoh))));
874     rF_kin(i,2) = (kmax(2)*(y(i,2)/(K(2)+y(i,2)))...
875         *(1/(1+(y(i,1)/K_inh_G)))*(1/(1+(y(i,6*p+6*q+12)/K_inh_etoh))));
876     ROI(i,1) = eF_rel(i,1)*rF_kin(i,1);
877     ROI(i,2) = eF_rel(i,2)*rF_kin(i,2);
878 end
879
```

```
880  uF = zeros(length(t),2);     % Cybernetic variables
881  vF = zeros(length(t),2);
882
883  for i = 1:length(t)
884      uF(i,1) = ROI(i,1)/sum(ROI(i,:));
885      uF(i,2) = ROI(i,2)/sum(ROI(i,:));
886      vF(i,1) = ROI(i,1)/max(ROI(i,:));
887      vF(i,2) = ROI(i,2)/max(ROI(i,:));
888  end
889
890  r_up = zeros(length(t),2);  % Regulated cellobiose uptake/fluxes, mol/g-biom.s
891
892  for i = 1:length(t)
893      r_up(i,1) = vF(i,1)*eF_rel(i,1)*rF_kin(i,1);
894      r_up(i,2) = vF(i,2)*eF_rel(i,2)*rF_kin(i,2);
895  end
896
897  mu = Y2_B.*r_up(:,2);       % Specific biomass growth rate, 1/s
898
899  end
```