UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Multiscale Simulations: From Enzyme Kinetics to Fluctuating Hydrodynamics

Permalink
https://escholarship.org/uc/item/3fn8691d

Author
Shang, Barry Zhongqi

Publication Date
2013

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Multiscale Simulations:
From Enzyme Kinetics to Fluctuating Hydrodynamics

by

Barry Zhongqi Shang

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jhih-Wei Chu, Chair
Professor Phillip Geissler
Professor David Graves
Professor Berend Smit

Fall 2013

Multiscale Simulations:
From Enzyme Kinetics to Fluctuating Hydrodynamics

Copyright 2013
by
Barry Zhongqi Shang

Abstract

Multiscale Simulations:
From Enzyme Kinetics to Fluctuating Hydrodynamics

by

Barry Zhongqi Shang

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Jhih-Wei Chu, Chair

The development of multiscale methods for computational simulation of biophysical systems represents a significant challenge. Effective computational models that bridge physical insights obtained from atomistic simulations and experimental findings are lacking. An accurate passing of information between these scales would enable: (1) an improved physical understanding of structure-function relationships, and (2) enhanced rational strategies for molecular engineering and materials design. Two approaches are described in this dissertation to facilitate these multiscale goals.

In Part I, we develop a lattice kinetic Monte Carlo model to simulate cellulose decomposition by cellulase enzymes and to understand the effects of spatial confinement on enzyme kinetics. An enhanced mechanistic understanding of this reaction system could enhance the design of cellulose bioconversion technologies for renewable and sustainable energy. Using our model, we simulate the reaction up to experimental conversion times of days, while simultaneously capturing the microscopic kinetic behaviors. Therefore, the influence of molecular-scale kinetics on the macroscopic conversion rate is made transparent. The inclusion of spatial constraints in the kinetic model represents a significant advance over classical mass-action models commonly used to describe this reaction system. We find that restrictions due to enzyme jamming and substrate heterogeneity at the molecular level play a dominate role in limiting cellulose conversion.

We identify that the key rate limitations are the slow rates of enzyme complexation with glucan chains and the competition between enzyme processivity and jamming. We show that the kinetics of complexation, which involves extraction of a glucan chain end from the cellulose surface and threading through the enzyme active site, occurs slowly on the order of hours, while intrinsic hydrolytic bond cleavage occurs on the order of seconds. We also elucidate the subtle trade-off between processivity and jamming. Highly processive enzymes cleave a large fraction of a glucan chain during each processive run but are prone to jamming at obstacles. Less processive enzymes avoid jamming but cleave only a small fraction of a chain. Optimizing this trade-off maximizes the cellulose conversion rate.

We also elucidate the molecular-scale kinetic origins for synergy among cellulases in enzyme mixtures. In contrast to the currently accepted theory, we show that the ability of an endoglucanase to increase the concentration of chain ends for exoglucanases is insufficient for synergy to occur. Rather, endoglucanases must enhance the rate of complexation between exoglucanases and the newly created chain ends. This enhancement occurs when the endoglucanase is able to partially decrystallize the cellulose surface. We show generally that the driving forces for complexation and jamming, which govern the kinetics of pure exoglucanases, also control the degree of synergy in endo-exo mixtures.

In Part II, we focus our attention on a different multiscale problem. This challenge is the development of coarse-grained models from atomistic models to access larger length- and time-scales in a simulation. This problem is difficult because it requires a delicate balance between maintaining (1) physical simplicity in the coarse-grained model and (2) physical consistency with the atomistic model. To achieve these goals, we develop a scheme to coarse-grain an atomistic fluid model into a fluctuating hydrodynamics (FHD) model. The FHD model describes the solvent as a field of fluctuating mass, momentum, and energy densities. The dynamics of the fluid are governed by continuum balance equations and fluctuation-dissipation relations based on the constitutive transport laws. The incorporation of both macroscopic transport and microscopic fluctuation phenomena could provide richer physical insight into the behaviors of biophysical systems driven by hydrodynamic fluctuations, such as hydrophobic assembly and crystal nucleation.

To develop the FHD model, we map all-atom molecular dynamics trajectories onto mass, momentum, and energy density grids to generate a corresponding field trajectory. From the field statistics, the response functions and transport coefficients of the atomistic model are computed. These thermophysical properties are then used to parameterize an FHD model for the fluid that reproduces the hydrodynamic correlations underlying the atomistic model. We show that an FHD description of the fluid is preserved down to length scales of 5 Å, enabling application of this coarse-graining approach to molecular systems.

We further extend our coarse-graining method by developing an interfacial FHD model using information obtained from simulations of an atomistic liquid-vapor interface. We illustrate that a phenomenological Ginzburg-Landau free energy employed in the FHD model can effectively represent the attractive molecular interactions of the atomistic model, which give rise to phase separation. For argon and water, we show that the interfacial FHD model can reproduce the compressibility, surface tension, and capillary wave spectrum of the atomistic model. Via this approach, simulations that explore the coupling between hydrodynamic fluctuations and phase equilibria with molecular-scale consistency are now possible.

In both Parts I and II, the emerging theme is that the combination of bottom-up coarse graining and top-down phenomenology is essential for enabling a multiscale approach to remain physically consistent with molecular-scale interactions while simultaneously capturing the collective macroscopic behaviors. This hybrid strategy enables the resulting computational models to be both physically insightful and practically meaningful.

If I don't research,
I don't know what to write;
if I don't write,
I don't know what to research.

– Unknown

To Mom, Dad, and Joyce

Contents

List of Figures

List of Tables

Acknowledgements

It has been a long, yet rewarding journey, and with the help of many amazing people, I am finally here where I am today. It is difficult to trace back the countless support I've received over the years, but I will attempt to do so from the most recent to the very beginning.

First, I would like to thank my advisor, Professor Jhih-Wei Chu for the wonderful four years of guidance he has provided. Prof. Chu brings out the best in his students by giving them the flexibility to explore their own problems and approaches. As a result, I have had the opportunity to work on a wide variety of research problems and be exposed to many different ways of thinking. This breadth of experience has made these past four years very intellectually enjoyable.

Prof. Chu is much more than an advisor, however. He cares for his students not only as researchers, but as individuals as well. His joyful, enthusiastic attitude is one that alleviated much of the stress of graduate school research, and his cheerful personality is one that I will always remember. When I am inevitably faced with adversity in the future, I will be sure to think back to the examples he set and try to follow the positive, optimistic approach he takes when confronting challenges. However, I am not sure if I will be able to reproduce his trademark "Ha, Ha, Ha!" laughter.

In Berkeley, I've had the opportunity to work with many remarkable individuals. I first worked with Prof. Nikolaos Voulgarakis on fluctuating hydrodynamics. Nikos taught me many of the simulation techniques that I have used during my graduate studies. As a graduate student taking over the work he had done, he also wrote elegant code for which I am extremely grateful. I next worked with Prof. Rakwoo Chang on enzymatic cellulose decomposition. He introduced me to the null-event kinetic Monte Carlo (kMC) technique and was instrumental in initiating my work on cellulase kinetics. Without Rakwoo, the first four chapters of this dissertation would have been much harder to complete. I also thank Prof. Berend Smit for introducing the kMC technique in his graduate kinetics course. It is often believed (but not said) by many graduate students that graduate-level courses are usually irrelevant to their research, but this here was an exception. Without the classroom introduction to the kMC method, many more weeks would have been spent laboring through the more inaccessible kMC papers of the 1970's.

I am also grateful to have had some incredible lab mates, including Adam Gross, Kevin Haas, Dr. Yuchun Lin, and Dr. Hyung Min Cho. Adam and Kevin made sitting in a basement computer office all day as fun and exciting as it could be, and I enjoyed being their partners-in-crime during AIChE conferences. I also thank the support staff here at Berkeley, including Rocío Sanchez, Fred Deakin, Alegria Garcia, Carlet Altamirano, Drew Danielson, Marie-Josée Carlsen, Jon Forrest, and William Chen. It is safe to say that my communications with them increased asymptotically towards the graduation date. In Berkeley, I also met Jenny, whose love, support, and patience during this process has been unwavering. Once this dissertation process is completed, I will have to find a way to make up for these many missed nights and weekends.

Before coming to Berkeley, I was an undergraduate researcher in Prof. Ron Larson's lab at the University of Michigan, Ann Arbor. I joined Prof. Larson's lab in the manner that many other undergraduates find research positions – by emailing many professors and hoping for a response. I was very lucky to receive a reply from Prof. Larson, but remember being a little intimidated as he was chair of the department at the time. Little did I know that he was one of the nicest and most approachable people I have met. I will always remember his straight-forward style to research, solving problems by writing down a list of questions and going after the most important ones first. In the Larson lab, I worked with Dr. Zuowei Wang, who taught me all about molecular dynamics simulations and how to deal with temperamental UNIX clusters. Zuowei's guidance (and humor) was indispensable in enabling me to publish two papers as an undergraduate.

Lastly, I need to thank my family. Words cannot express how much love my parents have given me over these years. They are truly my role models for life. Even when I am not well-behaved, they are always understanding, encouraging, and loving. I am so grateful for the risk they took in coming to the United States when I was only three years old. Coming from a farming village in China, their journey is truly a rags-to-riches story. I am blessed by their support of my education. My dad's doctoral degree in engineering is inevitably a reason why I pursued my own doctoral degree. I just couldn't let him get one-up on me. My mom keeps me grounded and always watches out for me. I am naturally inclined to be book-smart, but she makes sure I am street-smart as well. My sister, Joyce, who is ten years my younger, has brought so much joy into my life and to our family. I am so proud to see her grow from a crawling, drooling baby to a mature young lady. She even introduced Miles, our 1-year old puppy, to our family. However, now that she is in her "independent" teenage years, I am no longer allowed to tease her as much as I used to. I am eagerly waiting for that opportunity to arise again.

Part I

Kinetic Monte Carlo Modeling of Enzymatic Cellulose Decomposition

Chapter 1

Introduction to Cellulase Kinetics

1.1 Overview

One of the most urgent challenges facing our generation is climate change due to rising levels of greenhouse gases such as carbon dioxide into the atmosphere [1,2]. A significant fraction of carbon emissions arises from the burning of fossil fuels. To reduce carbon emissions, an attractive technology is the conversion of cellulosic biomass into transportation fuels and chemicals, such as ethanol [3,4]. Cellulosic biomass sources include agricultural waste, such as corn stover and wood chips, and dedicated "energy" crops that could be grown on marginal lands. Among the most popular energy crops under consideration are *Miscanthus* grass species, switchgrass, and poplar trees. These crops have received attention due to their high yield per acre and low water usage [4].

The primary argument in support of cellulosic biofuel technologies is that cellulosic biomass, as the most abundant organic material on Earth [5], represents a potentially vast and renewable energy resource. According to a Department of Energy study, the United States has the capacity to produce 1.3 billion tons of biomass per year, which could displace up to 30% of the country's petroleum consumption [6,7]. In addition, unlike ethanol production from corn starch, cellulosic ethanol does not compete directly with food production and does not require displacement of fertile agricultural land [8]. Cleaner transportation fuels from biomass could also reduce net carbon emissions from sources which are more distributed, from which carbon capture would be more difficult [4].

Despite these advantages, significant challenges remain before cellulosic biofuel production at an industrial scale can become viable. The growth of energy crops will require large amounts of land and water, which could impact soil quality, biodiversity, and sustainability [4,9]. The low energy density of plants also implies that biomass resources will be highly distributed, which requires the construction of new transportation infrastructures to consolidate the biomass feedstocks [10]. In addition, life-cycle analyses that take into account the potential land-use changes from biomass harvesting suggest that under certain conditions, the net carbon emissions from biofuel production is increased [9,11,12]. Biofuel production is also far from becoming economically competitive with petroleum refining and other technologies [10]. The impetus for renewable biofuels to become a clean energy solution is further weakened by improvements in natural gas recovery within the past decade [13]. Climate change, however, will not be solved by a single "silver bullet" idea, but rather by a portfolio of technologies [1]. In this new "industrial revolution" [2], cellulosic biofuels can play an important role.

1.2 Conversion routes

A variety of routes exist for converting cellulosic biomass into fuels and chemicals. They can be divided into thermochemical and biochemical routes [4,14–16]. The most common thermochemical route is gasification of biomass to produce syngas, which can later be transformed into industrial chemicals such as alcohols, dimethyl ethers, and diesel products. The main advantage of gasification is the ability to utilize a variety of biomass feedstocks [4], while its main challenge is costs associated with removal of tar, ash, and sulfur that poison the catalysts used during fuel synthesis [16]. By 2014, at least ten commercial-scale biorefineries producing cellulosic biofuels are expected to begin full operations. Two of these ten commercial-scale plants are scheduled to utilize the gasification route [15]. Gasification is seen as the major competitor against biochemical conversion technologies.

Biochemical conversion of biomass involves the use of enzymes and microbes to deconstruct cellulose. This process is composed of four steps: (1) mechanical pretreatment, (2), chemical pretreatment, (3) enzymatic hydrolysis, and (4) sugar fermentation [4]. Mechanical pretreatment reduces the physical size of the biomass components, producing millimeter-sized pellets [17]. Chemical pretreatment increases the internal surface area of the pellets, rendering the crystalline microfibrils within cellulose more accessible to enzymes [17–19]. These crystalline microfibrils are embedded within a matrix of hemicellulose and lignin, which obstruct enzyme access to the microfibril surface [5,19]. Often, dilute sulfuric acid is used as the catalyst, which solubilizes hemicellulose, redistributes lignin, and exposes the crystalline surface of the cellulose microfibrils [17,19]. After pretreatment, a cocktail of different types of cellulase enzymes are added, which act cooperatively to decompose cellulose into free sugars [20]. Typically, a loading of around 20 mg enzymes per gram of cellulose is used [21,22]. For pretreated substrates, a conversion around 80% can be expected after 3 days [21–23]. The free sugars produced from enzymatic decomposition can be later converted into ethanol or butanol by microbes [24] or into alkanes via aqueous phase processing [25]. An illustration of the main steps in the biochemical route is shown in Fig. 1.1. In 2014, four of the ten scheduled commercial-scale plants are expected to utilize biochemical routes [15].

A primary challenge facing economically viable conversion of cellulosic biomass into liquid fuels is the slow rate of enzymatic decomposition of cellulose into free sugars [3,20,26,27]. Cellulose is highly recalcitrant to saccharification due to the crystalline structure of cellulose microfibrils, which consist of polysaccharide glucan chains ordered by an extensive hydrogen-bonding network [3,28]. The high enzyme loadings and hence, costs required to

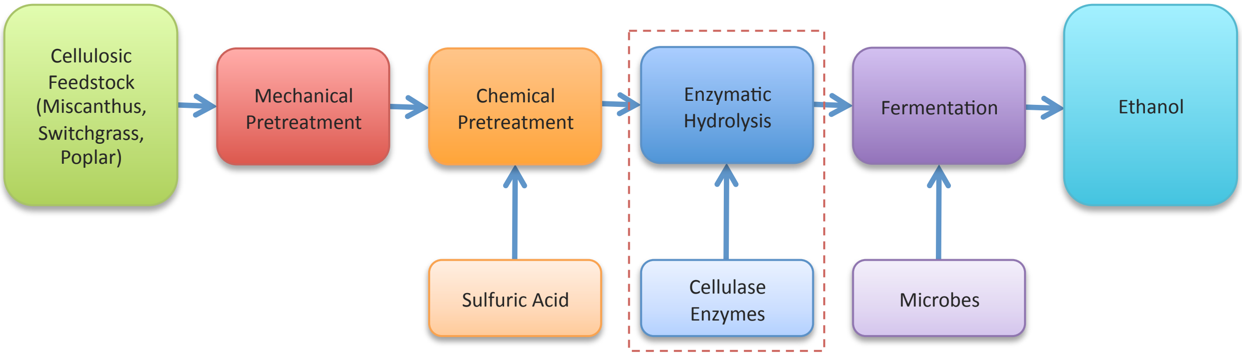


Figure 1.1. Outline of chemical process to convert cellulosic biomass into ethanol. The dotted red line encloses the process investigated in the dissertation work.

achieve appreciable cellulose conversion in industrial-scale processes limits economically viable production of cellulosic transportation fuels [20,21,29]. For example, compared to starch hydrolysis, up to 100-fold higher enzyme loading must be used [30]. Technoeconomic analyses show that the second largest component among raw materials costs is the cost of cellulases [21,31], which can be as high as \$1.47/gal ethanol [22]. Given current cellulosic ethanol production costs around \$3.50/gal [32], the enzyme cost represents a significant contribution. Current targets aim to reduce the production cost to around \$1.00/gal [33,34]. The ability to use fewer enzymes and/or improve their specific activities would aid in achieving more economically viable biofuel production. To this aim, an improved understanding of the enzymatic cellulose decomposition process is essential. In particular, a deeper mechanistic understanding of cellulase kinetics could facilitate more rational approaches to accelerate cellulose bioconversion.

1.3 Enzymatic decomposition of cellulose

The kinetic mechanism of enzymatic decomposition of cellulose is difficult to investigate primarily due to two features: (1) the complex structure of native plant cell walls, and (2) the complicated kinetic behaviors of cellulase enzymes. A majority of this work is aimed at understanding the latter. A brief overview of both the plant cell wall structure and enzyme kinetic behaviors will be provided next.

1.3.1 Plant cell structure

In the plant cell wall, the primary structural units are the cellulose microfibrils, which form a layered mesh within the cell wall to provide structural rigidity to the plant [28,35]. The microfibrils exist in a matrix of hemicellulose and lignin. Hemicellulose is a branched polysaccharide that links the microfibrils together, while lignin is complex network of aromatic car-

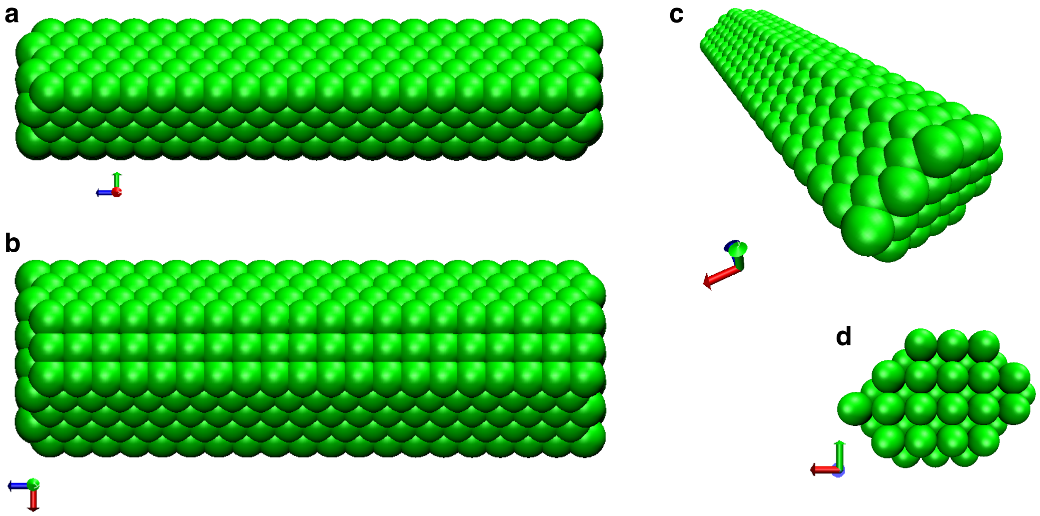


Figure 1.2. Illustration of geometry of a model microfibril. (**a**–**d**) Four different angles are shown. Each sphere represents a cellobiose residue (glucosidic dimer). The degree of polymerization of the microfibril above is 40 glucose residues.

bons that provides structural rigidity to woody plants [27,36]. The microfibrils consist of an ordered array of glucan chains [5,37,38], with length of 100 nm to microns and a diameter of 3 to 5 nm [35]. In higher plants, typically 36 glucan chains compose a microfibril [37]. Different views of a model microfibril are illustrated in Fig. 1.2.

The recalcitrance of cellulose is due to the extensive hydrogen bonding network and van der Waals interactions between the glucan chains, which impact high crystallinity to the microfibrils [3,38,39]. Therefore, a large amount of cellulase enzymes must be used to facilitate conversion into sugars. Besides cellulases, other types of plant cell wall degrading enzymes exist and are used for cell wall degradation. The most common are hemicellulases, which remove xylans associated with the microfibrils and increase cellulose surface accessibility to cellulases [36,40]. However, because the dominant component of the cell wall is cellulose (35-50%) [27,41], existing primarily as insoluble crystalline microfibrils, and because hemicellulose can be solubilized in water, cellulases are the most crucial polysaccharidase enzymes for overcoming recalcitrance of the plant cell wall.

1.3.2 Kinetic mechanisms of cellulases

The most industrially important cellulases are those produced by the fungus *Trichoderma reesei*. These cellulases are employed in current commercially available mixtures [26,30,42,43]. This fungus is important because it is able to secrete cellulases in large concentrations and with high activity [26]. Therefore, using these cellulases is economically advantageous. The

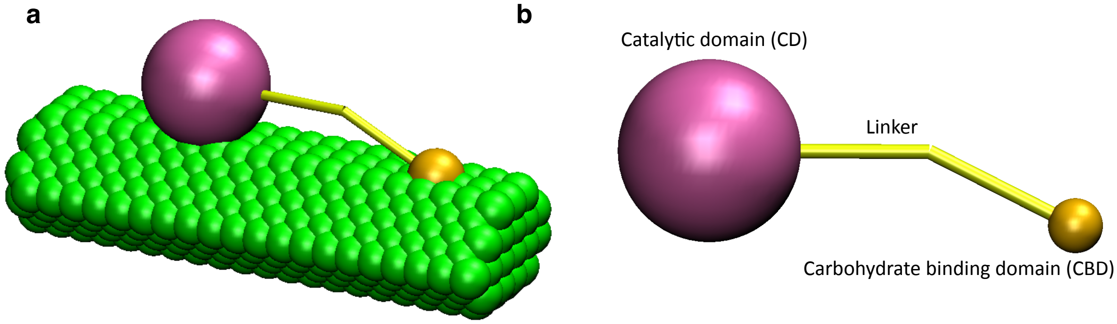


Figure 1.3. Illustration of cellulase as a ball-stick model. (**a**) A model *Tr*Cel7A cellulase adsorbed onto a microfibril. (**b**) Illustration of modular structure of cellulase.

fungus produces three major cellulase enzymes, *Tr*Cel7A, *Tr*Cel6A, and *Tr*Cel7B [41,44]. The first two are exoglucanases (or cellobiohydrolases), and the last is an endoglucanase. These two types of cellulases are distinguished by their kinetic mechanisms in decomposing cellulose, which will be described in Section 1.3.2. All three types of these enzymes possess a modular structure, consisting of a carbohydrate binding module (CBM) and a catalytic domain (CD) [45,46]. The CBM promotes adsorption of the cellulase onto insoluble substrates by hydrophobic interactions [47–49], while the CD possesses the active site responsible for hydrolytic bond cleavage [50]. A flexible linker connects both domains [51]. A schematic of the enzyme modules is shown in Fig. 1.3.

The crystal structures of the CBM for *Tr*Cel7A and *Tr*Cel7B have been resolved [52,53], while crystal structures of the CD for *Tr*Cel7A, *Tr*Cel6A, and *Tr*Cel7B have been resolved [50,54,55]. The CBM is much smaller; dimensions of the CBM of *Tr*Cel7A are $2\,\mathrm{nm} \times 1\,\mathrm{nm} \times 3\,\mathrm{nm}$ [52]. The CD is much larger, with dimensions for *Tr*Cel7A of $5\,\mathrm{nm} \times 4\,\mathrm{nm} \times 6\,\mathrm{nm}$ [50]. Therefore, the diameter of these enzymes is similar to the diameter of the microfibrils, and diffusion of the enzyme along the microfibril is effectively constrained to one dimension. The active sites of *Tr*Cel7A and *Tr*Cel7B accommodate 10 glucose residues [56,57], while the active site of *Tr*Cel6A accommodates 6 glucose residues [58,59]. The length of the linker connecting the CBM and CD is around 5 nm [51]. Removal of the CBM decreases adsorption onto insoluble substrates, lowering the cellulose conversion rate, but has little effect on conversion of soluble substrates, since substrate adsorption is no longer required [41].

The mechanism of enzymatic cellulose decomposition involves several elementary kinetic steps [27,60,61]. From solution, enzymes can adsorb onto the hydrophobic faces of the microfibril [49,62]. Once on the surface, enzymes can diffuse, desorb from the surface, or complex with a glucan chain. Complexation involves extraction of a glucan chain from its crystalline state in the microfibril and translocation into the active site of the CD [27,57]. Once complexed, the enzyme can decomplex or perform hydrolysis, which involves cleavage of the bonds connecting the glucose residues. These mechanisms will be further described in Chapter 2.

Exoglucanases can only complex with the free end of a glucan chain. *Tr*Cel7A, an exoglucanase, can only complex with the free reducing end of the chain, while *Tr*Cel6A, complexes with the free non-reducing end of the chain [46]. This difference means that these enzymes begin to cleave bonds from opposite ends of the chain. Complexation for exoglucanases involves threading of a chain into the active site of the CD, which resembles a tunnel [56]. The second bond from the end is cut by these enzymes, which releases a glucosidic dimer, cellobiose, into solution [45]. Exoglucanases are also processive [63,64]. After hydrolysis, the enzymes translate forward by the length of a cellobiose molecule, which moves the glucan chain within the tunnel into position to be cut again. The series of hydrolysis and translation steps occurs until the exoglucanase decomplexes from the chain.

Endoglucanases, such as *Tr*Cel7B, can complex anywhere along the chain, not only at chain ends [41,61]. In this case, the active site is shaped like a groove, rather than a tunnel [45,55], which allows interior portions of a glucan chain to be lifted up from the surface and into the active site. The open active site, however, also makes endoglucanases non-processive, since additional tunnel-forming loops that keep glucose residues locked into the catalytic domain are lacking [55]. This means that after hydrolysis, the enzyme decomplexes with the chain. Endoglucanases cut interior bonds within the glucan chain and typically do not release soluble glucan products into solution after cleavage. Rather, they act primarily to reduce the length, or degree of polymerization, of the glucan chains [41,65,66].

*Tr*Cel6A primarily acts as an exoglucanase, but displays some endo-like behavior [64,67–69], since its active site possesses more flexible loops that alternate between a tunnel-like and a groove-like structure [58,59]. Hence, *Tr*Cel6A can be considered to be an "endo-processive" enzyme, initiating processive hydrolysis within interior portions of a glucan chain. However, because the active site loops are also more flexible, *Tr*Cel6A is less processive than *Tr*Cel7A [70]. An illustration of the differences between exo- and endoglucanase mechanisms is shown in Fig. 1.4.

Synergy between exo- and endo-acting enzymes is well-studied [72,73], but its mechanistic origins remain unclear [26,73]. It refers to the observation under certain conditions that the net conversion achieved by a mixture of exo- and endoglucanases is greater than the sum of conversions obtained from their pure activities. Since additive conversion would appear if these enzymes acted completely independently, the presence of synergy indicates the existence of cooperation between the enzymes. To exploit this synergy, industrial formulations consist of a cocktail of enzymes, which typically include *Tr*Cel7A, *Tr*Cel6A, and *Tr*Cel7B from *Trichoderma reesei* [26,30]. Mechanistic theories explaining the origins of enzyme synergy will be further explored in Chapter 4.

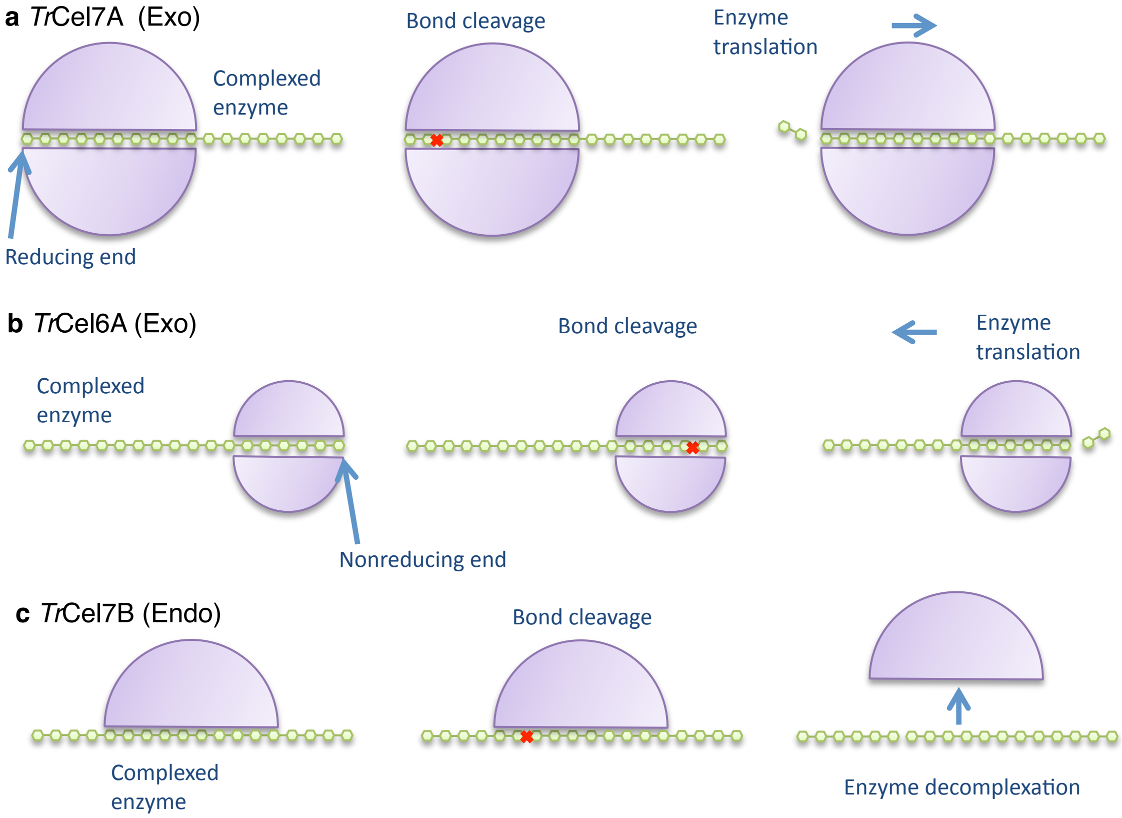


Figure 1.4. Similarities and differences between hydrolytic mechanisms of cellulases. (**a**) Hydrolytic mechanism of exoglucanase *Tr*Cel7A, showing tunnel-like active site environment holding 10 glucose residues [56]. The enzyme hydrolyzes from the reducing end and translates after hydrolysis. (**b**) Hydrolytic mechanism of exoglucanase *Tr*Cel6A, showing tunnel-like active site environment holding 6 glucose residues [59,71]. The enzyme hydrolyzes from the non-reducing end and translates after hydrolysis. (**c**) Hydrolytic mechanism of *Tr*Cel7B, showing active site holding 10 glucose residues [57]. The enzyme decomplexes after hydrolysis.

1.4 Limitations of current modeling techniques

To obtain a deeper mechanistic understanding of enzymatic cellulose decomposition, a variety of experimental [20,41,74] and modeling [60,75] approaches have been employed. The broad objectives in all of these studies are to identify features of the system that enhance conversion and features that limit conversion. By identifying and understanding these features, a more rational approach toward accelerating cellulose conversion could be pursued. Extensive reviews of experimental and modeling studies have appeared in the literature [20,41,60,73,76]. Here, a few of the limitations of current techniques will be described in order to motivate the development of a new kinetic model.

1.4.1 Spatial effects on cellulase kinetics

The key kinetic feature that many current techniques are unable to characterize well is that the decomposition process occurs on a spatially confined and heterogeneous surface. Surface enzymes are likely not well mixed and interfere with each other and with the substrate via excluded volume interactions [64,77–79]. Solution phase experiments are unable to account for the spatial heterogeneities associated with enzyme surface configuration and substrate morphology. In these experiments, the primary observables are the concentration of enzymes on the surface and in solution [80–84]. The spatial arrangement of surface enzymes cannot be directly measured. For surface confined species in crowded environments, however, reactant configuration, in addition to concentration, affects the conversion rate because the reactants are unable to fully explore the entire free surface area between consecutive reaction events [85]. Therefore, the classical mass-action assumption of a dilute, well-mixed system is not likely to hold. In this case, spatial configuration, in addition to surface concentration, would affect the conversion rate. Physical insights developed from measurements of only the surface enzyme concentration can provide only a partial explanation of the reaction kinetics.

In addition to solution-phase experiments, mass-action kinetic models also suffer from limitations of spatial averaging. These models assume dilute, well-mixed surface reactants, and therefore, model the reaction rate as a function of the reactant concentrations [60]. Since these models are often used to interpret solution-phase experimental data, it is unsurprising that they would share the same limitations. Many of these kinetic models, when best-fitted to experimental data, describe the corresponding experiment quite well. However, they have little transferability and give little insight into the kinetics of individual enzymes.

Recent high-speed atomic force microscopy (AFM) experiments have provided spatial resolution of the decomposition process at molecular scales and show cellulase enzymes are prone to forming “traffic jams” along the surface [64]. These jammed configurations are believed to contribute to lower cellulose conversion rates. They arise from surface obstacles that block the processive movement of complexed enzymes [86–91]. Surface obstacles are either native to the plant material or arise from gradual degradation of the microfibril. Mass-action kinetic models, however, fail to account for jamming and excluded volume interactions explicitly.

They also do not explicitly model the gradual surface erosion of the microfibril [63,92,93], which in turn can promote jamming and inhibit enzyme activity [78,82,83].

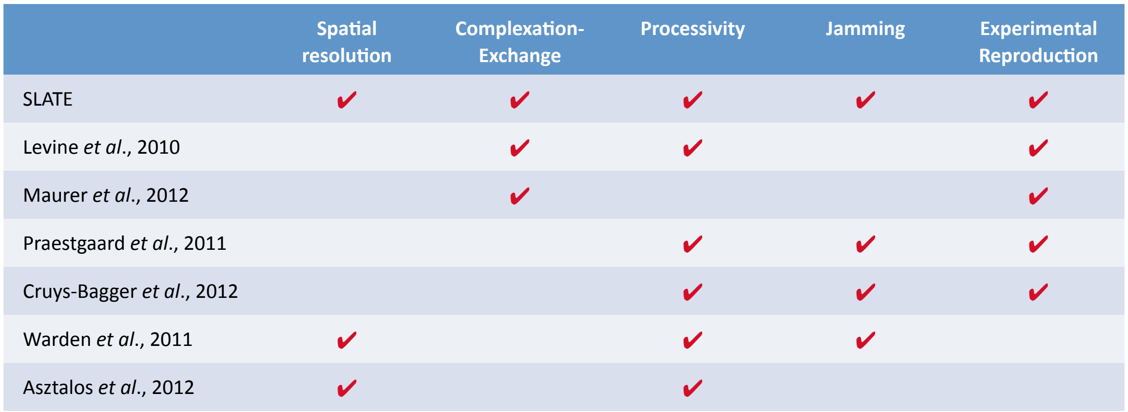
1.4.2 Modeling spatial effects

A strategy to incorporate spatial features implicitly is to preserve the mathematical structure of the mass-action model, but introduce time-dependent rate constants [77,78], which attempt to describe diffusional limitations that arise from spatial crowding. These rate constants decrease as a function of time and at the physical level, attempt to describe the gradual segregation of reactants as time progresses, which leads to a lower specific reaction rate [85]. However, spatial effects are treated implicitly, and the primary observable is again the species concentration. Without explicit modeling of the reactant configurations, it becomes difficult to determine if spatial effects are actually important, or if the observed trends arise as a result of the model's compensation for other neglected behaviors. In addition, with explicit spatial resolution, there is the obvious benefit of being able to transparently visualize and track the positions of individual reactants.

A few recent kinetic models have been developed to simulate enzymatic cellulose decomposition with explicit molecular resolution. Two of them are briefly discussed here. A cellular automaton model has been presented before for modeling cellulase kinetics on cellulose [94]. However, this model assumes complexation and decomplexation occur instantaneously and underestimates bulk and surface diffusion rates. The method also employs ad hoc probability parameters that are difficult to interpret in a physical context. A lattice-based coarse-grained model has also been used to study cellulase kinetics [95]. This model, however, employs lumped kinetic steps that obscure the rate-limiting behaviors. The model combines adsorption and diffusion with complexation, and combines decomplexation with desorption. By lumping the rates, it becomes difficult to identify which step is actually rate-limiting. The effects of complexation and decomplexation cannot be studied independently when their rates are lumped with others, although these processes are believed to govern conversion rates on insoluble crystalline substrates and require the most attention [57,96,97]. The model also only describes the degradation of a two-dimensional surface rather than a complete three-dimensional microfibril. Therefore, it does not capture the spatial constraints imposed by substrate morphology on enzyme configuration. In both models, the complexation-exchange kinetics cannot be studied as isolated processes, and the simulations do not quantitatively reproduce experimental conversion profiles.

1.5 Developing a comprehensive model

The development of kinetic models with spatial molecular resolution is essential for understanding the rate limitations that limit cellulose conversion. In this dissertation work, we develop a comprehensive kinetic model to account for the rich kinetic behaviors of cellulases acting on cellulose. We explicitly resolve the shape and structure of the enzymes and mi-



	Spatial resolution	Complexation-Exchange	Processivity	Jamming	Experimental Reproduction
SLATE	✓	✓	✓	✓	✓
Levine *et al.*, 2010		✓	✓		✓
Maurer *et al.*, 2012		✓			✓
Praestgaard *et al.*, 2011			✓	✓	✓
Cruys-Bagger *et al.*, 2012			✓	✓	✓
Warden *et al.*, 2011	✓		✓	✓	
Asztalos *et al.*, 2012	✓		✓		

Figure 1.5. Summary of features of various kinetic models for enzymatic cellulose decomposition. The newly developed model in the thesis work is called the Stochastic LATtice Enzyme model (SLATE). Reference numbers as as follows: Levine et al., 2010 [61]; Mauer et al., 2012 [97]; Praestgaard et al., 2011 [89]; Cruys-Bagger et al., 2012 [90]; Warden et al., 2011 [94]; and Asztalos et al., 2012 [95].

crofibrils in three dimensions. We explicitly track the locations of individual reactants with molecular-scale spatial resolution. We incorporate excluded volume interactions and capture jamming behavior among enzymes. We also isolate complexation and decomplexation of the enzymes with the surface as stand-alone steps. These advances enable us to achieve an enhanced kinetic understanding of the enzymatic decomposition process beyond current mechanistic theories [27,60].

The Stochastic LATtice Enzyme (SLATE) model developed here is a lattice kinetic Monte Carlo model [98] that simulates the spatiotemporal behavior of the enzyme reaction network with molecular-scale spatial resolution. A graphical illustration of the full suite of spatiotemporal enzymatic behaviors incorporated in the newly developed model and comparison to those of existing models is shown in Fig. 1.5. The insight gained from the model can be used to interpret solution-phase experiments and AFM studies. In addition, these insights can guide the focus of atomistic simulations, mass-action kinetic models, and single-molecule studies. Therefore, we believe SLATE is an integral part of a multiscale and multidisciplinary approach toward understanding and engineering enzymatic cellulose decomposition.

In SLATE, the shapes of individual *Tr*Cel7A enzymes and the microfibril glucan chains are resolved onto a three-dimensional lattice consisting of cubic grid cells with side length equal to the size of a glucose residue (5 Å). Excluded-volume interactions are imposed between the reactants to account for spatial confinement. A full suite of kinetic reactions is incorporated, including (1) adsorption, (2) desorption, (3) diffusion, (4) complexation, (5) decomplexation, and (6) hydrolysis. The system dynamics are governed by kinetic rate constants, which are obtained from the experimental literature.

1.6 Goals and objectives

The goals and objectives of Part I of this dissertation work are:

1. Identify the rate-limiting steps that hinder enzymatic cellulose decomposition.
2. Determine the role of complexation and decomplexation on cellulose conversion.
3. Investigate the extent to which spatial confinement and jamming limit conversion rate.
4. Understand the molecular and kinetic origins for synergism among different types of cellulases.
5. Develop a unified mechanistic understanding of enzymatic cellulose decomposition to facilitate rational engineering.

Chapter 2

Computational Methodologies

In this chapter, a brief overview of the kinetic Monte Carlo (kMC) method is provided and then the Stochastic LATtice Enzyme (SLATE) model for cellulases is described. The aim of the first section is to provide only a brief overview of the kMC method, as more complete and pedagogical explanations exist in publications and textbooks [98–100]. On the other hand, the SLATE model, which is unique to this dissertation work, will be described in more detail.

2.1 Kinetic Monte Carlo

2.1.1 Markov processes

The kinetic Monte Carlo (kMC) algorithm simulates a random walk in time through a reaction network. It is a stochastic simulation technique that generates sample paths following a Markov process [100]. A path refers to a sequence of states indexed by time. Two key ideas are that the dynamics is stochastic and Markovian. For chemical reactions occurring at molecular scales, the reaction kinetics behaves as a stochastic process due to the small number of molecules in the system and the choice to model the system with only a few key dynamical variables. The reaction kinetics is Markovian due to the large separation in time scales between the slow dynamical variables and fast neglected variables. The slow dynamical variables describe the state of the system, while the fast neglected variables are incorporated as model parameters. Under the Markov condition, the future state of the system depends only on its current state. The conditional dependence only on the current state refers to a process that lacks "memory". Mathematically, this condition is

$$\begin{aligned} P(X_n = x_n \,|\, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) &= P(X_n = x_n \,|\, X_{n-1} = x_{n-1}) \\ &\equiv P(x_n, t_n \,|\, x_{n-1}, t_{n-1}) \end{aligned} \tag{2.1}$$

where x_n is the system state at time t_n, x_{n-1} is the system state at time t_{n-1}, etc., and $t_n > t_{n-1} > t_{n-2} > \ldots$.

2.1.2 Transition probability

The connection between the kinetic Monte Carlo algorithm and Markov processes is the transition probability, which is the right-hand side of Eq. 2.1. The transition probability gives the probability for the system to be at state x_n at time t_n, given that it is in state x_{n-1} at time t_{n-1} [99].

In the kinetic Monte Carlo algorithm, the most important transition probability is the one that describes transitions between states separated by an infinitesimally small time difference Δt [99]. This transition probability is proportional to the time difference,

$$P(\boldsymbol{n}, t + d\tau \,|\, \boldsymbol{m}, t) = a_{m,n} d\tau \tag{2.2}$$

Here, $a_{m,n}$ is the rate constant, or transition probability per unit time, for a system at state $\boldsymbol{m}$ at time t to jump to state $\boldsymbol{n}$ at time $t+d\tau$. It has the units of a first-order rate constant. From the fundamental transition probability in Eq. 2.2, all other joint probability distributions, conditioned upon a given initial condition, can be constructed as a product of fundamental transition probabilities, due to the Markov property for conditional independence. The kMC algorithm uses this fundamental transition probability to exactly sample stochastic paths consistent with the joint distributions of the Markov process.

2.1.3 Gillespie algorithm

During a random walk through a reaction network, the system spends a certain amount of time in its current state and then instantaneously jumps to the next one. Therefore, to sample this process, we need to determine two quantities to describe a jump from one state to the next: (1) the amount of time spent in the current state, and (2) the location of the next state. These quantities behave as random variables whose joint probability distribution is described by the next-reaction probability $p(\tau, \boldsymbol{n} \,|\, \boldsymbol{m}, t)$ [99]. It is the probability that (1) nothing has happened during the interval $[t, t+\tau)$, and (2) the system jumps from state $\boldsymbol{m}$ to $\boldsymbol{n}$ at time $t + \tau + d\tau$ Hence, it the product of two probabilities

$$p(\tau, \boldsymbol{n} \,|\, \boldsymbol{m}, t) = P_0(\tau) a_{m,n} d\tau \tag{2.3}$$

The first is the probability that nothing has happened during the interval $[t, t+\tau)$, which is denoted $P_0(\tau)$. We can write this recursively as follows. Note that the probability that any reaction will occur in the time interval $[t, t+d\tau)$, given that the system is in state $\boldsymbol{m}$ at time t, is

$$a_m d\tau = \sum_n a_{m,n} d\tau \tag{2.4}$$

where the sum runs over all possible destination states $\boldsymbol{n}$. Therefore, the probability that nothing occurs in the time interval $[t, t+d\tau)$ is $1 - a_m d\tau$. The probability that nothing occurs in the time interval $[t, t+\tau+d\tau)$, $P_0(\tau + d\tau)$, can now be written recursively as $P_0(\tau + d\tau) = P_0(\tau)(1 - a_m d\tau)$. Hence, we have

$$P_0(\tau) = \exp(a_m \tau). \tag{2.5}$$

The emergence of the exponential function from an "infinitesimal rate parameter" is not unsurprising given the mathematical similarities here to continuously compounded interest.

From Eq. 2.3, the next-reaction probability becomes

$$\begin{aligned} p(\tau, \boldsymbol{n} \,|\, \boldsymbol{m}, t) &= \exp(a_m \tau) a_{m,n} d\tau \\ &= a_m \exp(a_m \tau) d\tau \times \frac{a_{m,n}}{a_m} \end{aligned} \tag{2.6}$$

We can see that the next-reaction probability is the product of the probabilities of two random variables: (1) the time to the next reaction (or waiting time), and (2) the location of the next state. The waiting time and next state are statistically independent.

We can sample the next-reaction probability by generating two uniform random numbers r_1 and r_2. To determine the next state, we first calculate $r_1 a_m$. We then accumulate a running sum of the rate constants starting from zero by looping through the possible reactions until the total exceeds $r_1 a_m$, at which point we choose the current reaction in our loop. To determine the time to the next reaction, we use the change of variables formula for probabilities [101]. We determine the function $\tau = f(r_2)$ such that τ has the desired exponential distribution. Using $\tau = (1/a_m)\ln(1/r_2)$ satisfies the change of variables formula. Therefore, we can sample the next-reaction probability distribution by only generating uniform random numbers.

After updating the system to its new state, we then enumerate the possible reactions from this new state to other states. We generate again two uniform random numbers and pick a reaction and waiting time. We repeat this process until a certain end condition is met, at which point the simulation is stopped. To sample all the different possible paths, we run different simulations from the same initial state, but with different random number seeds. The algorithm described above is called the direct method, or Gillespie algorithm [101], but alternative methods can also be used to sample the paths. Specifically, in SLATE simulations, the null-event algorithm is employed [98].

2.1.4 Null-event algorithm

The main difference between null-event kinetic Monte Carlo (kMC) and the Gillespie algorithm is that in the null-event simulation, sometimes, no reaction is chosen. Hence, a "null-event" occurs. At first glance, the null-event approach would seem to be computationally inefficient. We would generate random numbers but sometimes fail to advance the system forward. The speed up in the null-event algorithm lies in the fact that not all reactions need to be enumerated. In null-event kMC, we can save computational time by finding only a partial reaction list, at the cost of sometimes failing to advance the system. When the savings is larger than the cost, then the null-event algorithm can be more computationally efficient than the Gillespie algorithm. For spatially distributed systems, the enumeration of all the possible reactions from the current state can be very computationally expensive, because each location (or lattice site) within the system could participate in a reaction and the number of locations can be tens of thousands. Enumerating all the reactions here, as would be required in the Gillespie algorithm, would significantly slow down the computational speed. A review of the null-event algorithm is provided by Chatterjee & Vlachos [98].

In the null-event algorithm, we group the possible reactions from the current state into mutually exclusive and completely exhaustive families. Each reaction can only belong to one family. We generate a random number to choose a family. Each family has equal probability of being chosen. Once chosen, we only enumerate the reactions within the chosen family, and choose one of these reactions with probability $a_{m,n}/a_{\max}$. Therefore, like the Gillespie algorithm, the probability to choose a reaction is proportional to its rate.

The parameter $a_{\max}$ is a special value chosen to ensure that the all reactions are indeed selected with probabilities proportional to their rates. One potential way to do this is to have $a_{\max}$ equal to the sum of rate constants within the chosen family. However, this does not ensure that reactions belonging to different families would be chosen with probabilities in proportion to their rates, because the sum of rate constants within one family could be different than that within another family. $a_{\max}$ also cannot be smaller than the sum of rate constants within any one family, or else the reaction probabilities within that family are not properly normalized. Take the sum of rate constants within family f to be a_f. We have that $a_{\max}$ must be at least as large as the highest sum within the set $\{a_f\}$. For some families, the sum of rates within that family will be less than $a_{\max}$. Therefore, $1 - a_f/a_{\max}$ represents the probability that no reaction will occur, given that we have chosen family f.

Because sometimes no reaction will occur, during each iteration, the average time elapsed must be scaled to account for null-events. In the Gillespie algorithm, the mean time for a reaction is equal to $1/a_m$. Therefore, on average, time is advanced by $1/a_m$ per iteration. In the null-event algorithm, the average time elapsed per iteration is smaller than the corresponding time using Gillespie. This difference must occur because of the presence of null-events in the null-event algorithm. Because sometimes a reaction is not performed, we do not want to advance time by $1/a_m$ as in the Gillespie algorithm or else the system dynamics becomes artificially slow. To synchronize the dynamics, in the null-event algorithm, we advance time by $1/(N_f a_{\max})$ per iteration, where N_f is the number of families. Notice that $1/(N_f a_{\max})$ will at most be the value of $1/a_m$. It will equal $1/a_m$ if for all families, the sum of reactions within that family equals $a_{\max}$. A greater probability of null-events corresponds to a reduction in the average time elapsed per iteration. We can justify $1/(N_f a_{\max})$ by computing the average time elapsed before the next reaction takes place if we were to use the null-event algorithm. The probability that any reaction is chosen (or that a null-event is not chosen) is computed by adding up the probabilities of all ways that a reaction could be chosen, which gives

$$P(\text{rxn}) = \frac{1}{N_f a_{\max}} \sum_f a_f = \frac{a_m}{N_f a_{\max}}. \tag{2.7}$$

Therefore, the average number of iterations before a reaction is chosen is $1/P(\text{rxn})$. Hence, if we update time by $1/(N_f a_{\max})$ per iteration, the average time elapsed before a reaction is chosen is $1/a_m$, which is the Gillespie result.

Why is the null-event algorithm used in the SLATE model? It is used due to the large spatial extent of the system, which introduces numerous (1000+) possible reactions that

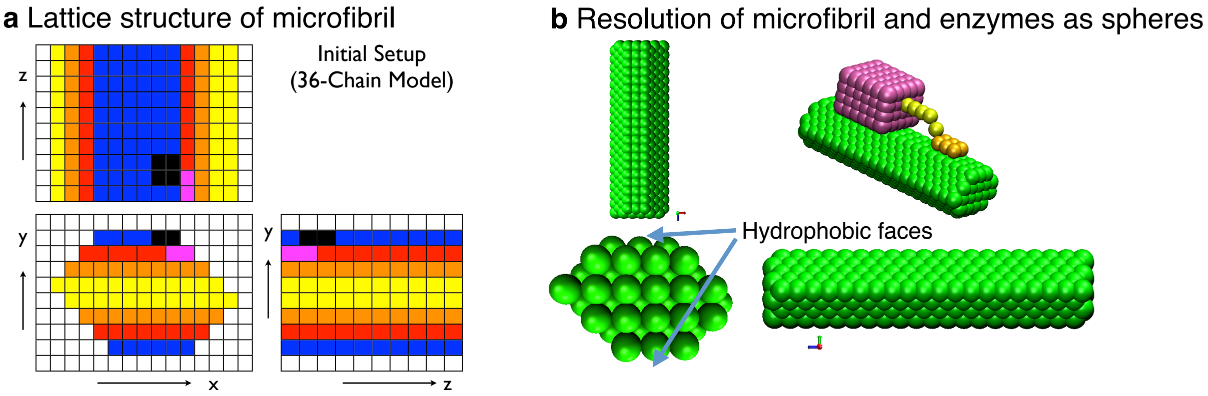


Figure 2.1. Lattice geometry of cellulose and cellulases in the Stochastic LATtice Enzyme (SLATE) model. (**a**) Lattice structure of microfibril from three different angles. The lines indicate the lattice spacing (5 Å). A cellobiose residue is shown as black. (**b**) Representation of microfibril and enzyme shape.

would require extensive computational power if a complete list of reactions must be updated during each iteration.

2.2 Stochastic LATice Enzyme (SLATE) model

The Stochastic LATice Enzyme (SLATE) model is based on the lattice kinetic Monte Carlo framework [98] to simulate enzyme dynamics along a spatiotemporal reaction network. In this Section, the system geometry and reactant components of SLATE will be described. In addition, a stochastic quasi-equilibrium method used to accelerate kMC simulations that face computational stiffness will be discussed [102]. This method is used in SLATE simulations because the reaction rate constants in the kMC model vary by over eight orders of magnitude.

2.2.1 System geometry and components

We use SLATE to simulate the degradation of a single microfibril by cellulase enzymes. Both the size and shape of the microfibril and enzymes are mapped onto a three-dimensional lattice. Their shapes are represented by space-filling cubes of 5 Å on this lattice, which is the size of a glucose residue [103]. An illustration of the lattice structure of the microfibril is shown in Fig. 2.1a. The primary building block is the cellobiose residue, which is modeled as a 1 nm × 0.5 nm × 1 nm (or 2 × 1 × 2 lattice units) rectangular prism [103]. The microfibril is composed of 36 glucan chains and each chain has a degree of polymerization of 1024 glucose residues (or 512 cellobiose residues). The structure and dimensions of the microfibril in the model are consistent with its native crystalline structure in the plant cell wall [35,37,39].

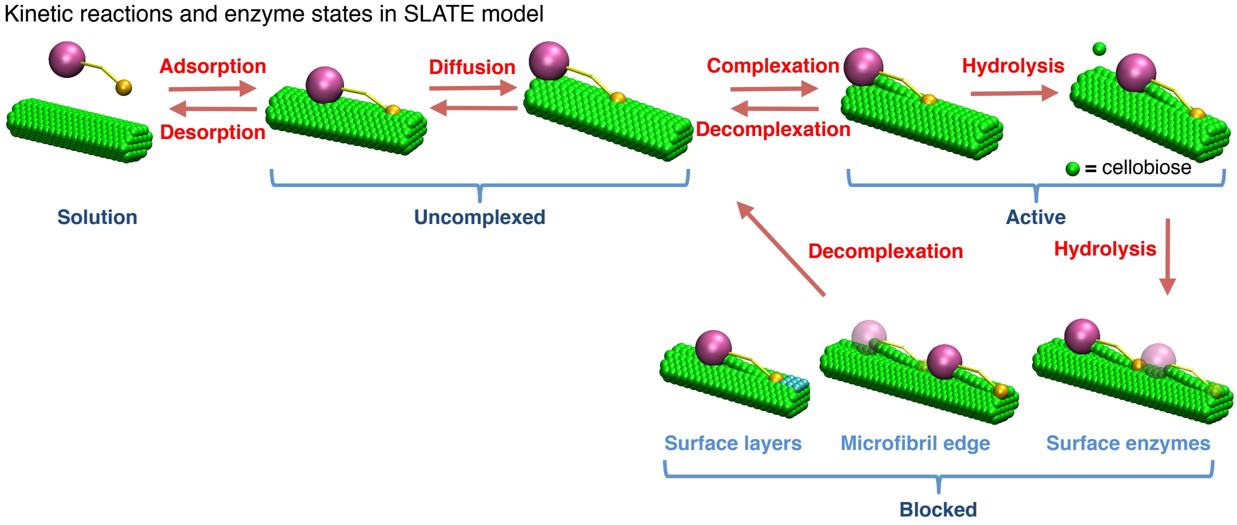


Figure 2.2. Kinetic reactions and enzyme states in SLATE model for *Tr*Cel7A. The kinetic reactions are labeled red, and enzyme states are labeled blue.

Three different types of enzymes from the fungus *Trichoderma reesei* are modeled: *Tr*Cel7A, *Tr*Cel6A, and *Tr*Cel7B. Their sizes and shapes are represented by space-filling cubes on a lattice. A lattice representation of the structures of the microfibril and enzymes is shown in Fig. 2.1b. The carbohydrate binding domain (CBD) of each enzyme is a $2\,\text{nm} \times 1\,\text{nm} \times 3\,\text{nm}$ rectangular prism, as determined from their crystal structures [52,53]. The catalytic domains for *Tr*Cel7A and *Tr*Cel7B are modeled as $5\,\text{nm} \times 4\,\text{nm} \times 6\,\text{nm}$ rectangular prisms [50,55], while for *Tr*Cel6A, the CD is a $5\,\text{nm} \times 4\,\text{nm} \times 3\,\text{nm}$ prism [54]. These dimensions are chosen to mimic the sizes and shapes of enzymes as determined from their crystal structures. For all three enzymes, a linker of length 5 nm connects the CBM to the CD [51].

2.2.2 Elementary kinetic steps for *Tr*Cel7A

Each enzyme can perform a number of elementary kinetic steps, consistent with current theories for their mechanisms [27,60,61,96]. The kinetic steps for the processive *Tr*Cel7A will be described below. The elementary kinetic steps for *Tr*Cel6A and *Tr*Cel7B are best discussed in comparison with those of *Tr*Cel7A and will be described afterward in Section 2.2.3.

In the SLATE model, *Tr*Cel7A can perform 7 elementary kinetic reactions: (1) adsorption, (2) desorption, (3) diffusion, (4) complexation, (5) decomplexation, and (7) hydrolysis. These are illustrated in Fig. 2.2 and described in more detail.

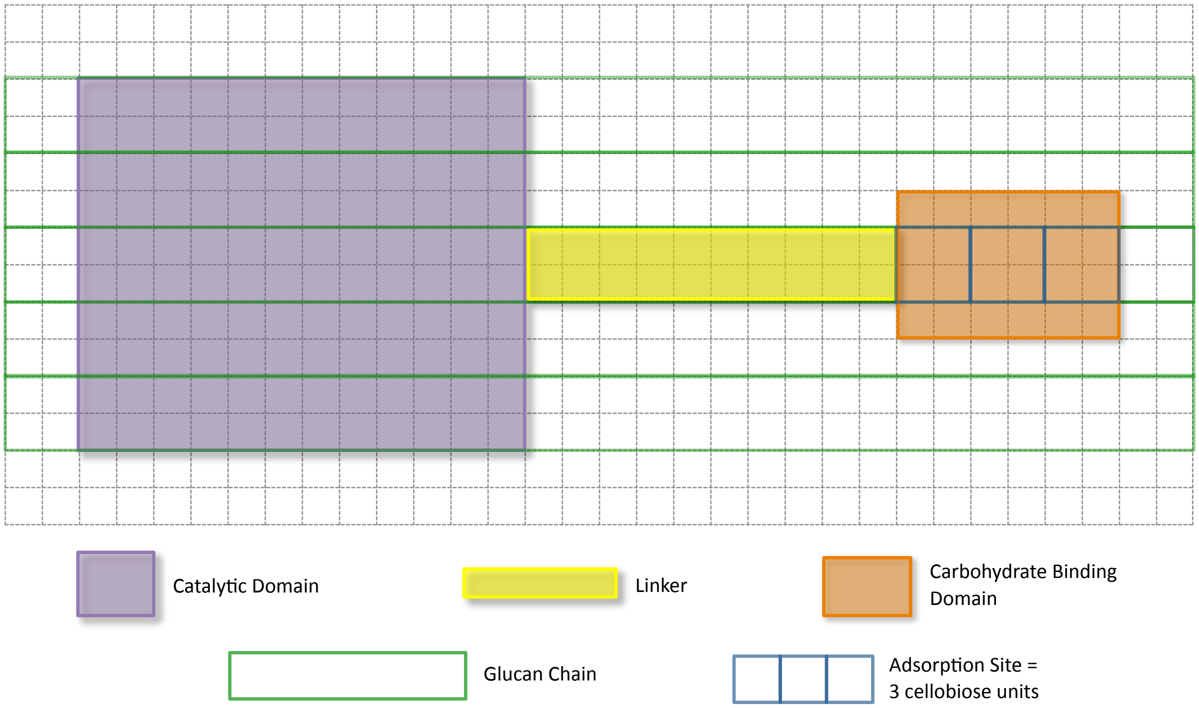


Figure 2.3. Top-down view of adsorption site and enzyme footprint. The grid lines indicate the lattice spacing (5 Å).

Adsorption and desorption

An enzyme in solution is considered well-mixed and its location is not considered. From solution, the enzyme can adsorb onto the hydrophobic microfibril faces, shown in Fig. 2.1. Since the enzyme is well-mixed in solution, adsorption follows classical mass-action kinetics, as the net adsorption rate is proportional to the product of the number of solution enzymes and available adsorption sites. An enzyme in solution may adsorb onto the surface if an adsorption site is available. An adsorption site consists of three consecutive cellobiose units at the hydrophobic surfaces of the microfibril. It is considered available if these cellobiose units are unoccupied (or not covered) by other enzymes and if the adsorbed enzyme does not overlap other enzymes. An illustration of an adsorption site and the enzyme "footprint" in relation to this site is shown in Fig. 2.3.

After adsorption, the lattice sites occupied by the enzyme are filled. No lattice site can be filled by more than one reactant, enforcing the excluded volume constraint. Enzymes are assumed to adsorb in the orientation in which they process. That is, *Tr*Cel7A adsorbs in an orientation such that its CBM is closer to the non-reducing end. Rotation of the enzyme is not explicitly modeled, as these extra degrees of freedom can be effectively sampled via translational diffusion.

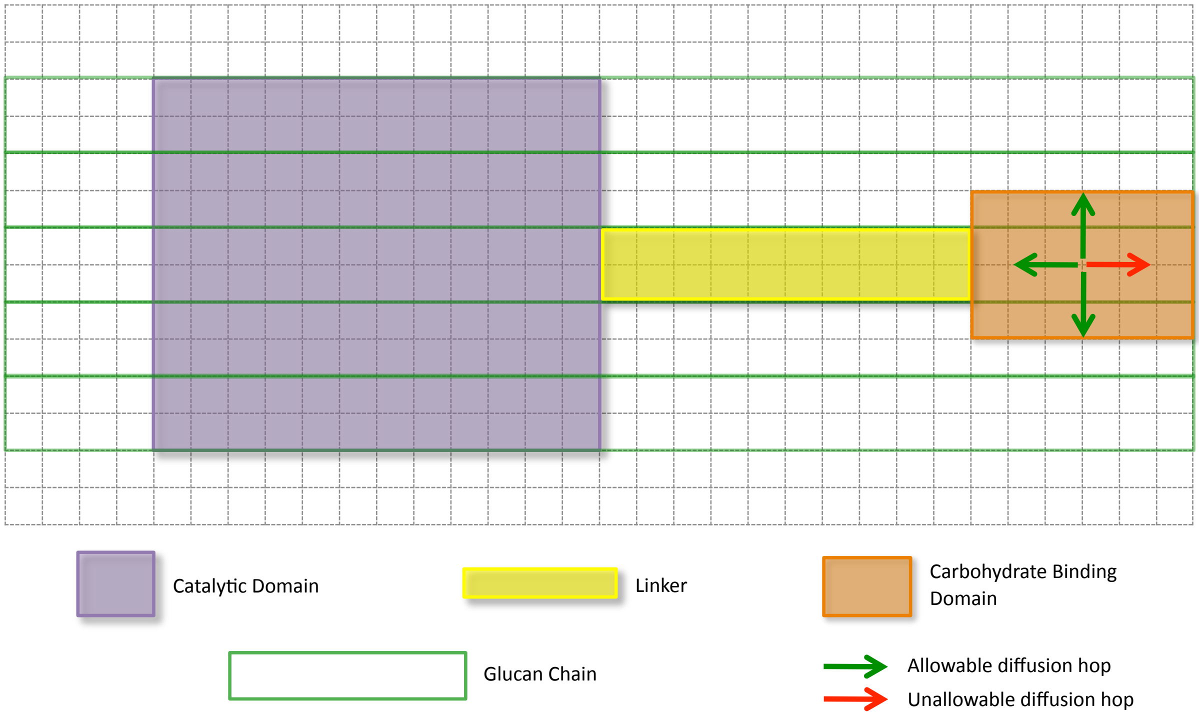


Figure 2.4. Possible diffusion moves for an enzyme sitting on the edge of the microfibril. It can only hop up, down, or to the left. Hopping to the right is not allowed.

An adsorbed, uncomplexed enzyme can always desorb. After desorption, the lattice sites that were once occupied by the enzyme are unfilled and the enzyme returns to solution.

Diffusion

An adsorbed uncomplexed enzyme can perform two-dimensional hops on the microfibril surface. The hopping distance is the length of a cellobiose residue (1 nm). An enzyme can hop if its translation does not overlap another enzyme or microfibril surface layer. An enzyme on the edge of the surface also cannot hop off the edge, as shown in Fig. 2.4, due to the high affinity of the CBM for the surface [49].

Complexation and decomplexation

In SLATE, the complexation process represents the extraction of a portion of the glucan chain from the surface and placement into the catalytic tunnel of the enzyme [27]. An adsorbed, uncomplexed enzyme can only complex with a glucan chain if it is appropriately aligned with the free reducing end of the chain. This configuration is shown in Fig. 2.5.

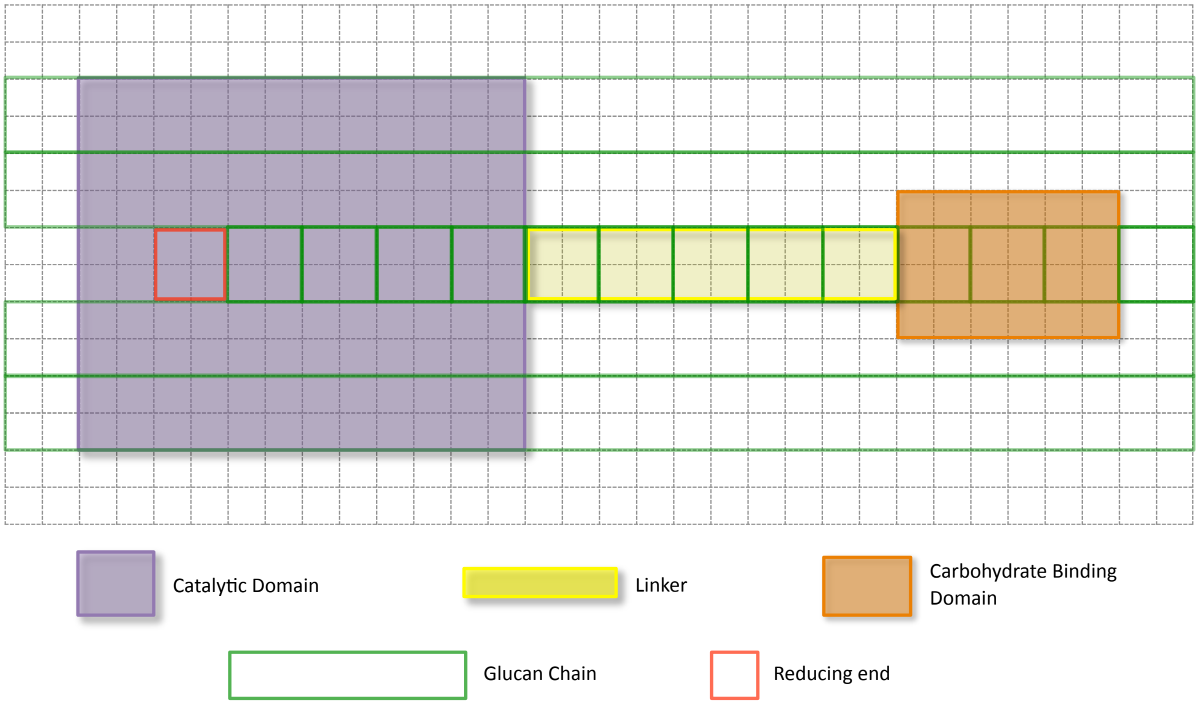


Figure 2.5. Top-down view of configuration allowable for complexation. The grid lines indicate the lattice spacing (5 Å). In this configuration, the active site of the enzyme covers the reducing end. Individual cellobiose residues in the middle glucan chain are outlined for clarity.

In this configuration, the active site of the *Tr*Cel7A enzyme covers the reducing end. After complexation, the portion of the glucan chain covered by the linker and catalytic domain is considered to be complexed with the enzyme and no longer in the crystalline state. This treatment ensures that cellobiose residues can only be complexed with one enzyme at a time. The change in location of complexed cellobiose residues due to lifting from the surface and threading through the active site is not explicitly modeled. Only the state of the cellobiose residue is changed (from "crystalline" to "complexed") after the complexation process. Section 2.2.4 describes in more detail the definition of system states.

A complexed enzyme can always decomplex with a chain. After decomplexation, the complexed portion of the glucan chain is returned to its "crystalline" state within the microfibril.

Hydrolysis

Hydrolysis consists of cleaving off a cellobiose unit from the chain and translation of the enzyme forward by a cellobiose residue distance. The catalytic tunnel consists of 10 glucose

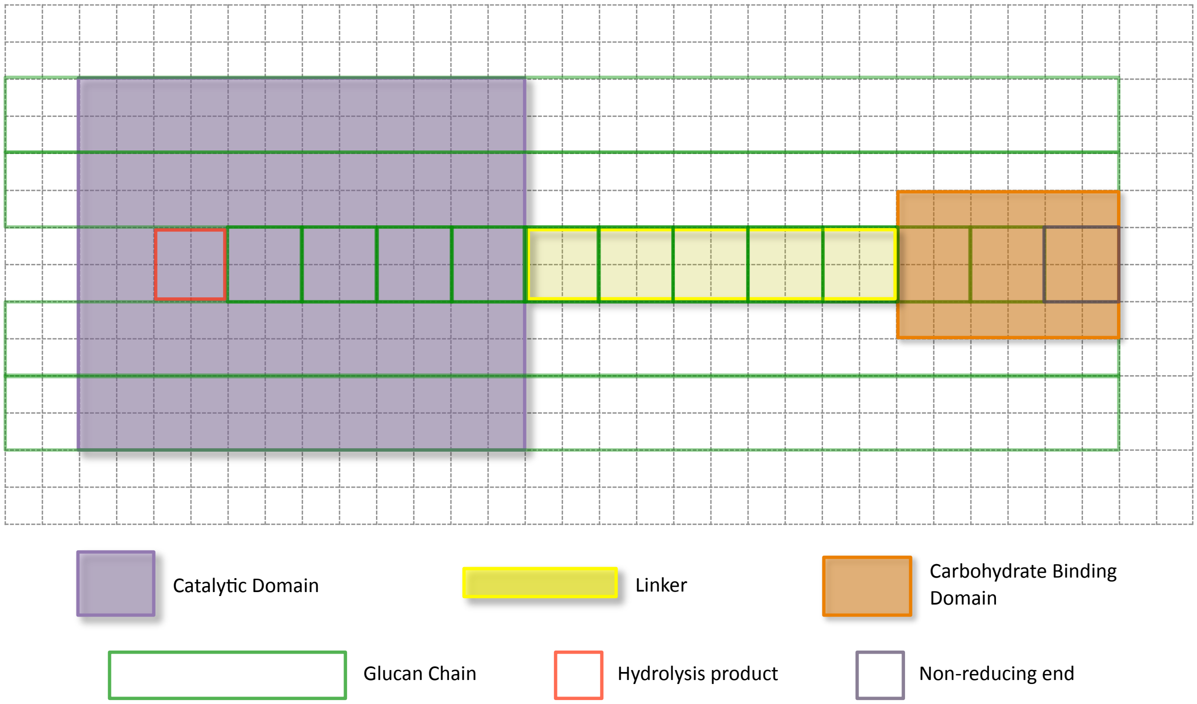


Figure 2.6. Top-down view of enzyme at the non-reducing end. The location of the hydrolysis product (cellobiose) is also shown. The grid lines indicate the lattice spacing (5 Å). Individual cellobiose residues in the middle glucan chain are outlined for clarity.

residues, consistent with crystal structure observations [56]. During hydrolysis, the cellobiose unit at the reducing end is removed from the glucan chain and an additional cellobiose unit becomes complexed as the cellulase slides forward. A complexed enzyme can hydrolyze if forward translation by a cellobiose distance (1 nm) is not blocked by the excluded volume of other enzymes or surface layers. At the non-reducing end, two cases are considered in this study. The "obstacle" case occurs when a complexed enzyme at the non-reducing end cannot hydrolyze any further and can only decomplex. The "obstacle-free" case occurs when this enzyme is able to hydrolyze. In this case, hydrolysis consists of removal of the remaining glucan chain from the microfibril and desorption of the enzyme. In Fig. 2.6, the location of the hydrolysis product (cellobiose) is shown, along with a complexed enzyme at the non-reducing end.

The default case in the SLATE model assumes that the non-reducing edge of the microfibril is an obstacle. Once a processing *Tr*Cel7A molecule reaches the non-reducing edge of the microfibril surface, the enzyme stops and can only become mobile again by decomplexing with its chain. If decomplexation occurs relatively slowly, complexed enzymes can remain blocked at the edge for a significant amount of time (hours). The non-reducing edge of a finite microfibril thus acts as an obstacle (Fig. 2.2) and can induce jamming. This treatment is based on the high affinity of *Tr*Cel7A to crystalline cellulose [49], which prevents processing enzymes from passing over the edge. Effectively, the microfibril length in the SLATE model defines a characteristic length for the obstacle-free region in cellulose. The impact of changing this boundary condition on simulation results is addressed in Section 3.6.3.

2.2.3 Kinetic steps for *Tr*Cel6A and *Tr*Cel7B

The processive *Tr*Cel6A performs a similar set of kinetic steps as does *Tr*Cel7A, except it complexes with the non-reducing end of the chain and moves processively in the opposite direction [46]. Hence, while *Tr*Cel7A can be blocked at the non-reducing end, *Tr*Cel6A can be blocked at the reducing end. Experimentally, it is known that *Tr*Cel6A also exhibits some endo-like activity [64,67]. This means it can complex with the interior portion of a chain and begin processive hydrolysis at this location. This behavior is also included in the model, and further described in Section 4.2.

The endoglucanase *Tr*Cel7B is not processive [41], and therefore, it acts differently in the SLATE model. Unlike *Tr*Cel7A and *Tr*Cel6A, it can complex anywhere with a chain. After complexation, it can only hydrolyze a bond, which creates two new glucan chains, one with a new reducing end and the other with a non-reducing end. Hydrolysis does not release cellobiose product into solution unless an isolated cellobiose residue on the surface is produced after the bond cleavage. After hydrolysis, the endoglucanase decomplexes immediately. Therefore, in the SLATE model, hydrolysis and decomplexation of endoglucanases are coupled steps.

2.2.4 Definition of system state

Here, we describe the set of variables that specify the system state in order to fully define the reaction network. The current system state determines the set of allowable reactions, and hence, the set of allowable destination states. The system state vector $\boldsymbol{s} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N; \boldsymbol{c}_1, \ldots, \boldsymbol{c}_M)$ is composed of vectors of dynamical variables characterizing the state and location of each enzyme and cellobiose residue.

The enzyme vector $\boldsymbol{e}_i = (s_i^e, l_i^e)$ can be further broken down into its state s_i^e and location l_i^e. For processive enzymes, the state variable s_i^e can hold the values “solution”, “unadsorbed”, “complexed”, and “blocked”. The latter three states occur on the microfibril surface. An uncomplexed enzyme can desorb, diffuse, or complex. Once complexed, an enzyme that is processive is either active or blocked. An active enzyme can decomplex or perform hydrolysis. A blocked enzyme can only decomplex since obstacles hinder the processive hydrolysis. For endoglucanases, the blocked state does not exist, since they are not processive and their hydrolysis is not impeded by surface obstacles.

As shown in Fig. 2.2, a processive *Tr*Cel7A enzyme can be in one of the four states: (1) in solution, (2) uncomplexed, (3) active, and (4) blocked. A blocked enzyme is one that is complexed but encounters obstacles that hinder the processive hydrolysis. Obstacles include

(a) uneven surface layers, (b) the non-reducing edge of the microfibril surface, and (c) other surface enzymes. An illustration of each blocking type is shown in Fig. 2.2. These states will be revisited again when analyzing simulation results in Chapter 3.

The location variable l_i^e holds the Cartesian coordinates of the enzyme and along with the enzyme shape, determines which lattice units are filled by the enzyme. The location variable also allows excluded volume constraints to be imposed and determines, for example, if a complexed enzyme is either active or blocked. The explicit structure and location of each enzyme is only mapped onto the lattice if the enzyme is on the surface. If the enzyme is in solution, its location is not explicitly tracked, since it is assumed to be well-mixed. However, once adsorption occurs, its location on the surface and its excluded volume contribution is mapped onto the lattice. When the enzyme desorbs into solution, the location coordinates are no longer tracked, and lattice sites become unfilled. Hence, the lattice does not fill the entire simulation "box". It is only introduced to track the structure of the microfibril and to map the shapes of the adsorbed enzymes. This "on-off" treatment of enzyme location is somewhat analogous to the treatment of particle location between the system and mass bath in a grand canonical Monte Carlo simulation [104].

The cellobiose vector $\boldsymbol{c}_i = (s_i^c, l_i^c)$ contains the state s_i^c and location l_i^c of the cellobiose residue. The state can be crystalline, complexed, or hydrolyzed. A cellobiose residue can only be complexed with one enzyme at the time. After hydrolysis, the location of the cellobiose residue is no longer tracked, since the residue is now in solution. In addition, the lattice units once filled by the cellobiose residue now become unfilled. Therefore, the microfibril gradually disappears as hydrolysis proceeds.

2.2.5 Model input parameters

The input parameters to the model include (1) the numbers of *Tr*Cel7A, *Tr*Cel6A, and *Tr*Cel7B enzymes, (2) the number of glucan chains and degree of polymerization, (3) the shapes and structures of the enzymes and microfibril, and (4) kinetic rate constants. These will be discussed in more detail in following Chapters when presenting specific applications. An `.xyz` file of the lattice coordinates of a *Tr*Cel7A enzyme is provided in Section 2.3.

2.3 Supplementary information

2.3.1 `.xyz` file of lattice *Tr*Cel7A enzyme

This file gives the precise set of lattice coordinates defining the shape of a *Tr*Cel7A enzyme. It can be uploaded into VMD 1.9.1 [105] for visualization of the lattice structure of the *Tr*Cel7A enzyme used in the SLATE model. Each line in the file represents the center of a rectangular prism with dimensions of a cellobiose residue ($1\,\text{nm} \times 0.5\,\text{nm} \times 1\,\text{nm}$). "A" denotes units belonging to the carbohydrate binding domain, "B" to the catalytic domain (CD), and "D" to the linker region. The *Tr*Cel7B shape is identical to that of *Tr*Cel7A. The

*Tr*Cel6A shape is different only in that it contains a smaller catalytic domain. The CD for *Tr*Cel7A is a 5 nm × 4 nm × 6 nm rectangular prism [50,55], while the CD for *Tr*Cel6A is a 5 nm × 4 nm × 3 nm prism [54].

```
95
TrCel7A lattice coordinates
A 32 38 6
A 32 38 8
A 32 38 10
A 30 38 6
A 30 38 8
A 30 38 10
B 35 38 22
B 35 38 24
B 35 38 26
B 35 38 28
B 35 38 30
B 35 38 32
B 33 38 22
B 31 38 22
B 29 38 22
B 27 38 22
B 27 38 24
B 27 38 26
B 27 38 28
B 27 38 30
B 27 38 32
B 29 38 32
B 31 38 32
B 33 38 32
B 35 40 22
B 35 40 24
B 35 40 26
B 35 40 28
B 35 40 30
B 35 40 32
B 33 40 22
B 31 40 22
B 29 40 22
B 27 40 22
B 27 40 24
B 27 40 26
```

```
B 27 40 28
B 27 40 30
B 27 40 32
B 29 40 32
B 31 40 32
B 33 40 32
D 31 40 12
D 31 42 14
D 31 42 16
D 31 42 18
D 31 42 20
B 35 42 22
B 35 42 24
B 35 42 26
B 35 42 28
B 35 42 30
B 35 42 32
B 33 42 22
B 31 42 22
B 29 42 22
B 27 42 22
B 27 42 24
B 27 42 26
B 27 42 28
B 27 42 30
B 27 42 32
B 29 42 32
B 31 42 32
B 33 42 32
B 35 44 22
B 35 44 24
B 35 44 26
B 35 44 28
B 35 44 30
B 35 44 32
B 33 44 22
B 31 44 22
B 29 44 22
B 27 44 22
B 27 44 24
B 27 44 26
B 27 44 28
```

```
B 27 44 30
B 27 44 32
B 29 44 32
B 31 44 32
B 33 44 32
B 33 44 24
B 33 44 26
B 33 44 28
B 33 44 30
B 31 44 24
B 31 44 26
B 31 44 28
B 31 44 30
B 29 44 24
B 29 44 26
B 29 44 28
B 29 44 30
```

Chapter 3

Identifying the Kinetic Limitations of Processive Cellulases on Cellulose

In this Chapter, we study the decomposition of a single cellulosic microfibril by *Tr*Cel7A enzymes using the Stochastic LATtice Enzyme (SLATE) model presented in Chapter 2. The physical insights obtained from these simulations will be useful when we investigate the effects of enzyme mixtures on the decomposition process.

Inter-protein and enzyme-substrate couplings in interfacial biocatalysis induce spatial correlations beyond the capabilities of classical mass-action principles in modeling reaction kinetics. Here, we develop a computational scheme to simulate the reaction network of enzymes with the structures of individual proteins and substrate molecules resolved in three dimensions. This methodology is applied to elucidate the rate-limiting mechanisms of a renowned system with significant technological importance: decomposition of crystalline cellulose by the processive exoglucanase *Tr*Cel7A. We illustrate that the primary kinetic bottlenecks are the slow complexation rate of glucan chains with the enzyme active site and excessive enzyme jamming along the crowded substrate. Jamming is alleviated by increasing the decomplexation rate, but at the expense of reduced processivity. We demonstrate that enhancing the apparent reaction rate requires a subtle balance between accelerating the complexation driving force while simultaneously avoiding inter-enzyme jamming. Via a spatiotemporal systems analysis, we develop a unified mechanistic framework that delineates the experimental conditions under which different sets of rate-limiting behaviors emerge. We find that optimization of the complexation-exchange kinetics is critical for overcoming kinetic barriers imposed by interfacial confinement and accelerating enzymatic cellulose conversion rates.

3.1 Motivation

Unraveling the importance of interfacial confinement on enzyme kinetics is a core problem for *in vivo* biology [106]. However, enzymatic behaviors at molecular scales and in heterogeneous environments are difficult to resolve. Here, we overcome this difficulty for a very important model system in the enzymatic decomposition of cellulose [3,20,27,41]. A model enzyme for this investigation is the most abundant exoglucanase produced by the fungus *Trichoderma reesei*, *Tr*Cel7A [46]. We devise a novel systems-level simulation method incorporating molecular-scale spatial resolution to illustrate how surface restriction induces non-classical kinetic behaviors among *Tr*Cel7A enzymes and limits the rate of converting cellulose into soluble sugars. An enhanced understanding of cellulase kinetics could help in developing strategies to reduce enzyme cost in cellulose bioconversion technologies [22,23].

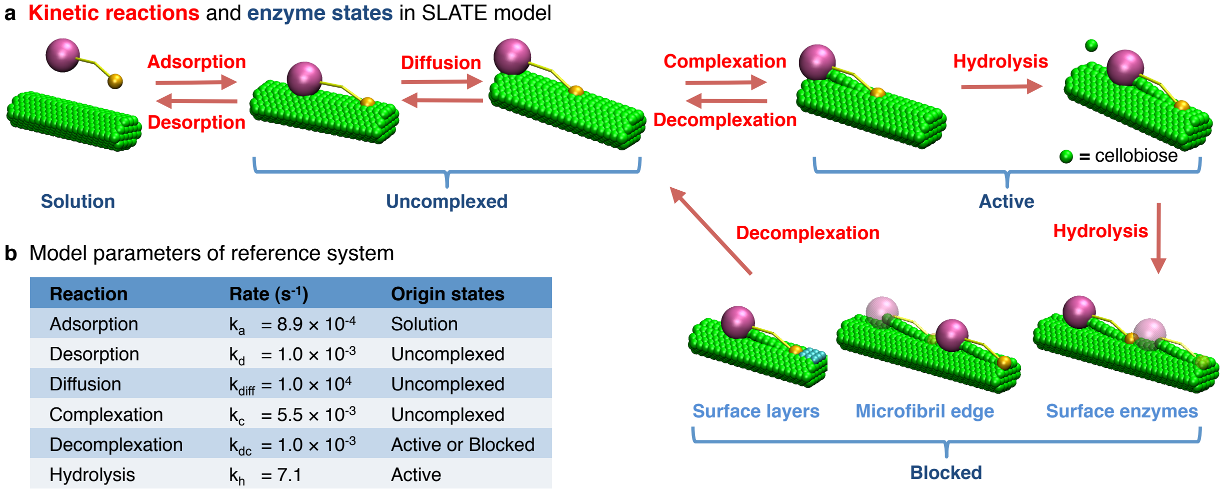
The elementary kinetic reactions performed by *Tr*Cel7A are shown in Fig. 3.1a [27,60]. The enzymes adsorb onto crystalline cellulose microfibrils, which are composed of linear glucan chains held tightly by hydrogen bonding and van der Waals forces [3]. On the microfibril surface, an adsorbed enzyme diffuses until it complexes with the free reducing end of a glucan chain [46]. Complexation involves extraction of the chain from the surface and threading through the active site tunnel [46]. After complexation, the enzyme processively hydrolyzes β-1,4-glycosidic linkages within the complexed chain and releases a cellobiose molecule into solution after each bond cleavage. The consecutive hydrolytic cleavage along a chain stops when the enzyme decomplexes or becomes blocked by surface obstacles [64,82,86].

In interfacial biocatalysis, the heterogeneous and crowded environments around insoluble substrates give rise to non-classical enzymatic behaviors which are difficult to model [60,85]. In enzymatic cellulose decomposition, the complexation-exchange kinetics of *Tr*Cel7A with glucan chains represent important interactions [27,60], but are difficult to investigate as stand-alone steps, since these processes occur at the solid-liquid interface. In addition, processive enzyme displacement is sensitive to surface obstacles that block processive motion and cause “traffic jams” among enzymes [60,64,82,86]. Similar phenomena exist in other examples of interfacial biocatalysis such as the functioning of motor proteins [107]. Substrate heterogeneity disqualifies classical mass-action (such as Michaelis-Menten) approaches for modeling kinetics, which assume a dilute, well-mixed environment and ignore excluded volume interactions [60,85]. Existing cellulase kinetic models, however, continue to employ these simplifying assumptions [60]. Lacking spatial resolution, the impact of surface-induced spatial constraints on cellulose conversion becomes difficult to identify.

3.2 Methods

To overcome these limitations, a novel kinetic model is developed to simulate the kinetic behavior of individual cellulases on cellulose microfibrils with molecular-scale spatial resolution. The method is a lattice kinetic Monte Carlo (kMC) model and is called the Stochastic LATtice Enzyme (SLATE) model, further discussed in Sections 2.2 and 3.5. In SLATE, the shapes of individual *Tr*Cel7A enzymes [50,52] and glucan chains [37] are spatially resolved on a three-dimensional lattice with 5 Å cubic grid cells [103] (See Sections 2.2 and 3.5.1). Excluded volume interactions are imposed between reactants to explicitly model spatial constraints arising from surface confinement. A full suite of kinetic reactions is incorporated

a Kinetic reactions and enzyme states in SLATE model



b Model parameters of reference system

Reaction	Rate (s^{-1})	Origin states
Adsorption	$k_a = 8.9 \times 10^{-4}$	Solution
Desorption	$k_d = 1.0 \times 10^{-3}$	Uncomplexed
Diffusion	$k_{diff} = 1.0 \times 10^{4}$	Uncomplexed
Complexation	$k_c = 5.5 \times 10^{-3}$	Uncomplexed
Decomplexation	$k_{dc} = 1.0 \times 10^{-3}$	Active or Blocked
Hydrolysis	$k_h = 7.1$	Active

Figure 3.1. Illustration of the Stochastic LATtice Enzyme (SLATE) model for *Tr*Cel7A. (**a**) Kinetic reactions (labeled red) and enzyme states (labeled blue) in the SLATE model. Cellulase enzymes can participate in one of the following reactions: adsorption, desorption, diffusion, complexation, decomplexation, and hydrolysis. Only hydrolysis is treated as an irreversible reaction. With these reactions, an enzyme can be in one of the four states: solution, uncomplexed, active, and blocked. These states are mutually exclusive and completely exhaustive. A blocked enzyme can be stalled by uneven surface layers, the non-reducing edge of the microfibril substrate, or the other surface enzymes. (**b**) Kinetic rates for the reference system. The possible originating states for each reaction are shown under "Origin states".

which includes complexation-exchange interactions and processive hydrolysis by *Tr*Cel7A, as shown in Fig. 3.1a.

A key feature of the SLATE model is the ability to track the states of individual enzymes. As shown in Fig. 3.1b, a *Tr*Cel7A enzyme can be in one of the four states: (1) solution, (2) uncomplexed, (3) active, and (4) blocked. An uncomplexed enzyme can desorb, diffuse, or complex. Once complexed, a *Tr*Cel7A molecule is either active or blocked. An active enzyme can decomplex or perform hydrolysis, while a blocked enzyme can only decomplex. Enzymes are blocked by surface obstacles, which include (a) uneven microfibril surface layers, (b) the non-reducing edge of the microfibril, and (c) other surface enzymes. An illustration of each obstacle type is shown in Fig. 3.1a. The non-reducing edge obstacle is based on the high affinity of the carbohydrate binding domain of *Tr*Cel7A to crystalline cellulose [49], which prevents processing enzymes from passing over the edge during hydrolysis (See Figure 2.6). This obstacle could also represent the presence of immobile structural obstructions such as hemicellulose and lignin [86,91]. The impact of removing the edge obstacle is addressed in Section 3.6.3.

The SLATE model is calibrated to reproduce the experimental conversion profile based on a reference set of reported kinetic rate constants, listed in Fig. 3.1b, and further described in Section 3.5.2. To estimate the complexation rate, the conversion profile is fit to the profile measured by Gusakov et al. [108]. The simulation system consists of a single microfibril composed of 36 glucan chains [37] with a representative degree of polymerization (DP) of 1024 glucose residues [41]. The averaged number of *Tr*Cel7A enzymes adsorbed on the microfibril is 18. This loading corresponds to a cellulose surface coverage of 25%, which is within the typical range employed in experiments [79]. This number is calculated as the intersection between the loading line from Gusakov et al. [108] and Langmuir isotherm [47]. The reported quantities are averaged over a sufficient number of independent trajectories for statistical uncertainties to become negligible. The kMC simulations are run for a reaction time of 200 hrs.

3.3 Results

3.3.1 Complexation is the slowest elementary step

In Fig. 3.2a, snapshots from a SLATE simulation illustrate the gradual erosion of the microfibril by *Tr*Cel7A. Since the processive cellulase decomposes glucan chains starting from their reducing ends [46], a more drastic thinning of the reducing edge develops with time as observed in experiments [62,63,109]. In Fig. 3.2b, quantitative agreement between the SLATE model and experimental conversion profile is illustrated.

The conversion profiles of several individual microfibrils are also shown in Fig. 3.2b; they exhibit substantial deviation from one another and to the averaged profile. The flat regions linked together by steep jumps are a signature of the single-microfibril kinetics that becomes obscured after spatial averaging. They are indicative of a complexation-limited mechanism. The flat regions represent the long waiting times required for uncomplexed enzymes to become active. As shown in Fig. 3.3a, increasing the complexation rate constant by 10-fold over the reference rate drastically enhances conversion, but the same increase in the hydrolysis rate hardly affects conversion.

To quantify the kinetic limitations at the single-molecule level, the accumulated time an enzyme spends in each state is recorded during the reaction. This is defined as the occupancy time for that state, further described in Section 3.6.1. States with high occupancy times are indicative of kinetic traps that prevent enzymes from becoming active. Fig. 3.3b plots the occupancy time distribution from the reference simulation and those from increasing the complexation or hydrolysis rate by 10-fold. The total reaction time is the time to reach 60% conversion. It is the sum of occupancy times over each state and labeled above the bars in Fig. 3.3b. The uncomplexed state clearly traps *Tr*Cel7A enzymes for the longest time in all cases, reflecting the high kinetic barrier for complexation. Fig. 3.3b shows that a 10-fold higher complexation rate can reduce the uncomplexed occupancy time by tens of hours. A same increase in hydrolysis rate only reduces the active state time by tens of seconds.

SLATE simulations affirm that complexation is the slowest step in crystalline cellulose

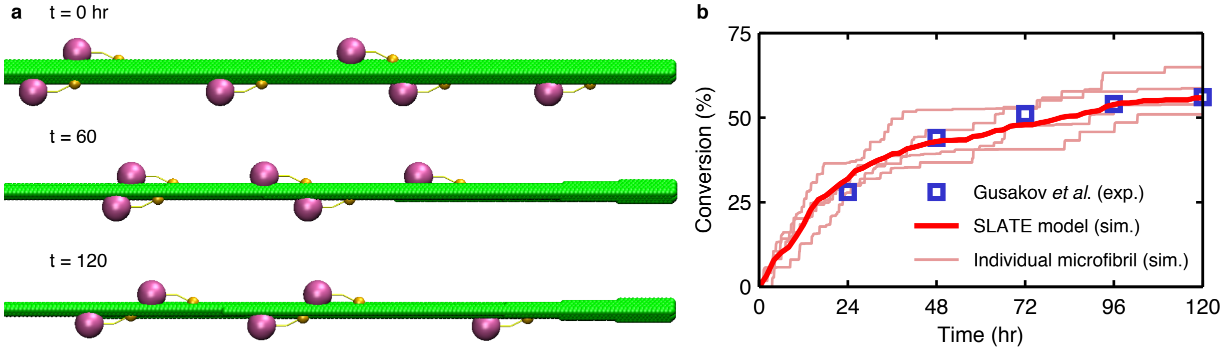


Figure 3.2. Illustration of kMC simulations with the SLATE model. (**a**) Snapshots from a simulated trajectory showing the degradation of a microfibril substrate over time. A quarter length of the microfibril in the model is shown; the non-reducing edge is located on the right hand side of the figure. The processing *Tr*Cel7A enzymes move from left to right after adsorbing onto the hydrophobic faces of the microfibril [49]. (**b**) The averaged substrate conversion over time in a SLATE simulation (dark red line) and the conversion from the study by Gusakov et al. [108] (blue squares) on the decomposition of Avicel by *Tr*Cel7A. The conversion profile from averaging over 8 independent simulations (dark red line) quantitatively agrees with the experimental data. Each simulation represents the decomposition of a single microfibril. Four of the eight simulations involved in computing the average are also shown for comparison as light red lines. The parameters used in the simulation are listed in Fig. 3.1b.

decomposition [96,97,110,111], rather than hydrolytic bond cleavage. The complexation time scale on the order of hours identified via SLATE simulations confirms recent experimental findings [97]. In Section 3.6.2, we show that the complexation-limiting behavior is robust to different assumptions in the model.

A key feature of Fig. 3.3b is that complexed enzymes spend more than 99% of their time in a blocked state, in which they cannot perform hydrolysis. Stalled *Tr*Cel7A molecules have been inferred from solution-phase "restart" experiments [82,86] and have been visualized by high-speed atomic force microscopy experiments as well [64]. Therefore, slow complexation and excessive blocking together make the active state short-lived and severely restrain the conversion rate. As illustrated in the following, acceleration of cellulose conversion can be achieved most effectively by increasing the complexation rate with coordinated removal of enzyme blocking.

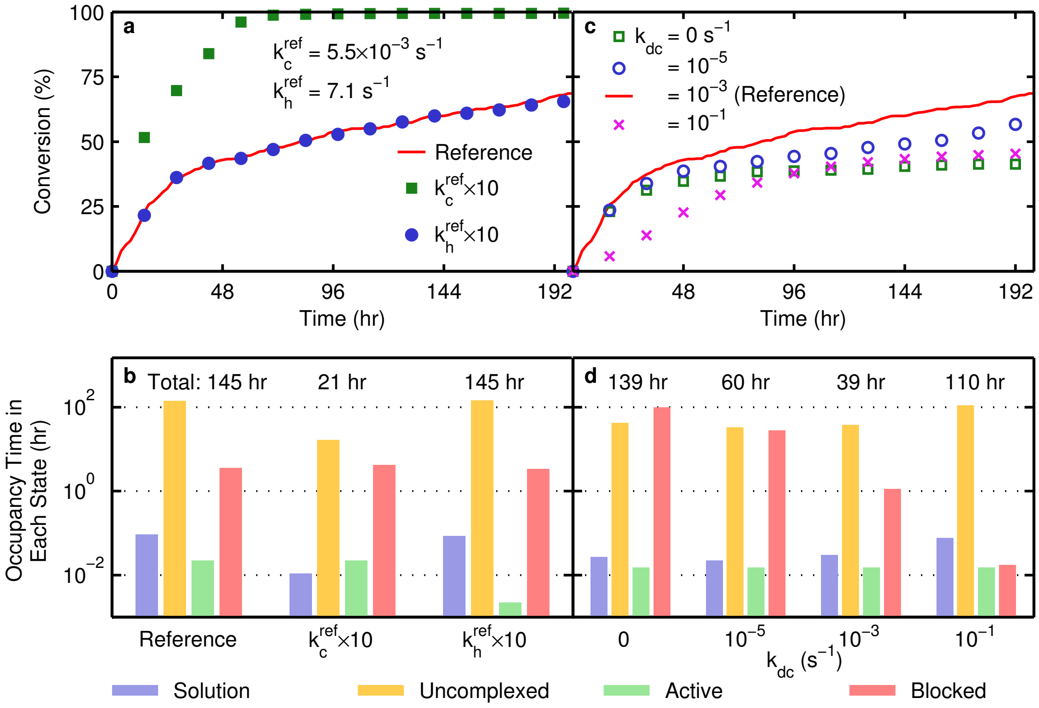


Figure 3.3. The effects of using different values of kinetic rate constants in SLATE simulations for illustration of the rate-limiting mechanisms. (**a**) The profile of the averaged substrate conversion over time using the reference parameters shown Fig. 3.1b (red line), a 10-fold higher value for the complexation rate constant k_c while keeping others unchanged (green squares), and a 10-fold higher value for the hydrolysis rate constant k_h while keeping others unchanged (blue circles). (**b**) The distributions of enzyme occupancy times in the following four states: solution (violet), uncomplexed (yellow), active (green), and blocked (red). Surface obstacles causing the blocked state include uneven surface layers, the non-reducing edge of the microfibril, and the other surface enzymes. The values shown above the bars are the total reaction times to reach 60% conversion. These values are the sum of the occupancy times over all of the four enzyme states. (**c**) The averaged profiles of substrate conversion over time using the reference parameters of Fig. 3.1b (red line) and varying the decomplexation rate constant k_{dc} from 0 to $10^{-1}\,\text{s}^{-1}$ while keeping others unchanged (symbols). (**d**) The distributions of enzyme occupancy times in the four states listed in (**b**) as a result of varying the decomplexation rate constant while keeping the others unchanged. The total reaction times above the bars represent the times to reach 40% conversion. For all of the data presented, the standard deviations are not shown for clarity, and they are within 5% of the averaged values.

3.3.2 Decomplexation plays dual and opposing roles in affecting conversion

Decomplexation is the primary mechanism by which an enzyme escapes from blocking, and therefore, the degree of blocking can be controlled via the decomplexation rate. Fig. 3.3c plots the SLATE conversion profiles from varying the decomplexation rate constant k_{dc} while fixing the other rates at their reference values. A maximum in conversion occurs at the reference value of $k_{dc} = 1 \times 10^{-3}\,\mathrm{s}^{-1}$, which is inferred from experiments [88–90,97]. In Fig. 3.3d, the occupancy time distributions at 40% conversion illustrate that two opposing forces occur from increasing k_{dc} that give rise to the maximum. A higher k_{dc} reduces the blocked state fraction, but also decreases the net complexation driving force k_c/k_{dc}, leading to an increase in the uncomplexed fraction.

To further dissect these opposing forces, the fractions of *Tr*Cel7A enzymes in each state are tracked over time at different values of the decomplexation rate. Fig. 3.4a indicates that at the reference value of $1 \times 10^{-3}\,\mathrm{s}^{-1}$, roughly 95% of surface enzymes are uncomplexed, since the complexation rate is low. Decreasing the decomplexation rate from this value lowers this uncomplexed fraction, which by itself would promote higher activity. However, the blocked fractions also increase, as seen in Fig. 3.4b-d, compensating for the higher complexation driving force. To reveal the molecular origins of these competing trends, the detailed mechanism of enzyme blocking on the microfibril substrate is analyzed next.

In Fig. 3.4b, the fraction of surface layer-blocked enzymes is illustrated. When k_{dc} is zero, this fraction is zero because the surface is initially smooth and the enzymes cannot decomplex to form uneven layers. When $k_{dc} = 1 \times 10^{-5}\,\mathrm{s}^{-1}$, surface layers begin to form, but the escape time from these layers is tens of hours, resulting in a large surface-layer blocked fraction. When $k_{dc} = 1 \times 10^{-3}\,\mathrm{s}^{-1}$, this fraction is low again because enzymes can decomplex within minutes when encountering these obstacles.

In Fig. 3.4c, the edge-blocked fractions are illustrated. When the decomplexation rate is low ($k_{dc} < 1 \times 10^{-3}\,\mathrm{s}^{-1}$), a steady fraction of edge-blocked enzymes appears, because enzymes performing processive hydrolysis are likely to reach the non-reducing edge before decomplexing. These enzymes then nucleate other lagging enzymes to become stalled, and form the head of the "traffic jam". The lagging enzymes are in the enzyme-blocked state, whose fractions are shown in Fig. 3.4d. An illustration of the traffic jam buildup on the microfibril surface is shown in Fig. 3.4e. Because the non-reducing edge represents an immobile obstacle, a majority of blocked enzymes are stalled within traffic jams, rather than near surface layers, which disappear once the layer-forming chain is processed.

The optimal k_{dc} value of $1 \times 10^{-3}\,\mathrm{s}^{-1}$ that maximizes the conversion in Fig. 3.3c represents the ideal balance in maintaining enzyme processivity while reducing enzyme jamming on the cellulose surface. Fig. 3.4f shows that if k_{dc} assumes a higher value, the *Tr*Cel7A enzymes would decomplex prematurely and fail to hydrolyze the glycosidic bonds near the non-reducing edge of the microfibril after each complexation event. The optimal k_{dc} thus needs to be sufficiently high for the surface enzymes to be free from jamming but adequately low for them to be as processive as possible. In this case, the decomplexation timescale is

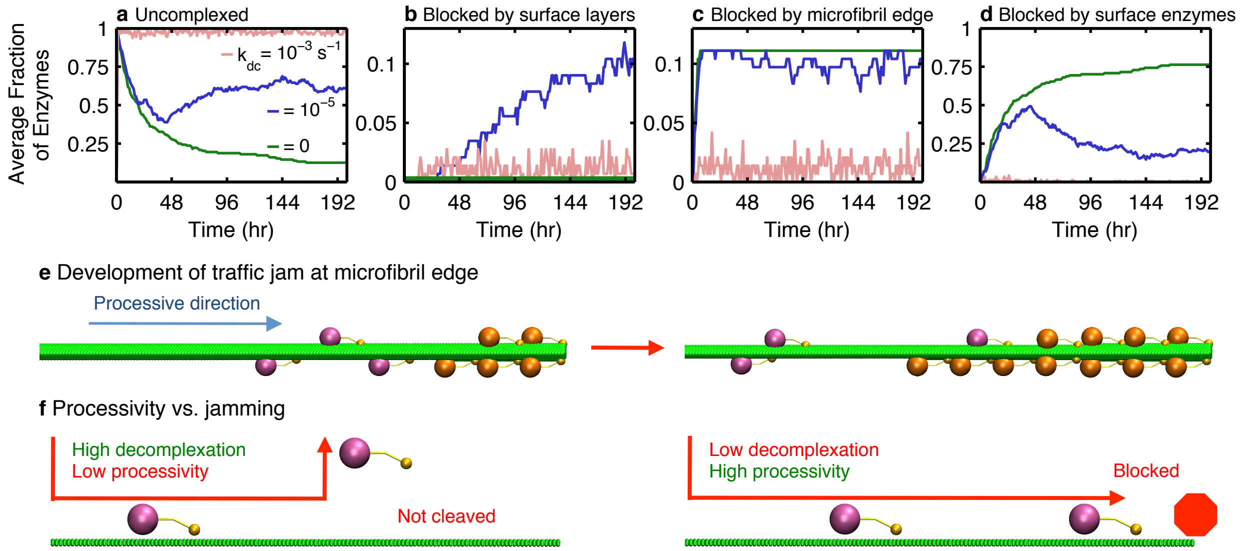


Figure 3.4. Fraction of surface enzymes in different states over time if the decomplexation rate k_{dc} (s^{-1}) were 0 (green), 10^{-5} (blue), and 10^{-3} (light red) while the other rate constants remain at the reference values of Fig. 3.1b. (**a**) The profiles of the fraction of uncomplexed enzymes. (**b**) The profiles of the fraction of enzymes blocked by uneven surface layers of glucan chains on the microfibril surface. (**c**) The profiles of the fraction of enzymes blocked by the non-reducing edge of the microfibril. (**d**) The profiles of the fraction of enzymes blocked by the other enzymes on the substrate surface. (**e**) An illustration of a "traffic jam" developing at the non-reducing edge of the microfibril when the decomplexation rate constant is low. Blocked enzymes are colored in orange. A quarter of the microfibril length is plotted with the non-reducing edge located on the right hand side. (**f**) An illustration of the interplay between enzyme processivity and jamming in affecting the reactivity of cellulose decomposition. A less processive enzyme tends to decomplex prematurely and the glycosidic bonds near the non-reducing ends of the glucan chains cannot be cleaved effectively. An overly processive enzyme cleaves the entire glucan chain after each complexation event but decomplexes very slowly and tends to be blocked at the non-reducing edge.

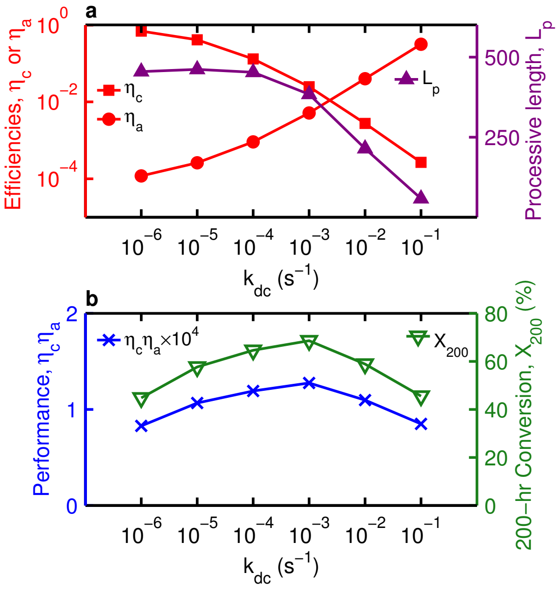


Figure 3.5. The efficiency measures, processive length, performance, and apparent activity of *Tr*Cel7A in cellulose decomposition as a function of the decomplexation rate k_{dc}. (**a**) The profiles of the complexation and activation efficiencies, η_C and η_A, respectively (red) and processive length, L_p (purple). The maximum processive length is 512, which is set by the average degree of polymerization of 1024 glucose residues per glucan chain. (**b**) The profiles of performance (blue) and 200-hour conversion (green).

commensurate with the average duration for the enzyme to process through a glucan chain on cellulose.

The conversion maximum is related to a leveling-off of the apparent processive length L_p with decreasing decomplexation rate. L_p is defined as the average number of glycosidic bonds cleaved per complexation event during the cellulose decomposition reaction [88]. When $k_{dc} < 1 \times 10^{-3}\,\mathrm{s}^{-1}$, Fig. 3.5a shows that *Tr*Cel7A enzymes cleave almost the entire glucan chain during each processive run. However, as illustrated in Fig. 3.4c-f, nearly complete cleavage comes at the expense of jamming. When $k_{dc} > 1 \times 10^{-3}\,\mathrm{s}^{-1}$, enzymes are likely to decomplex before reaching the non-reducing edge. Therefore, although jamming is avoided, L_p is severely reduced. The optimal decomplexation rate ($1 \times 10^{-3}\,\mathrm{s}^{-1}$) corresponds to the value where L_p begins to level off as k_{dc} is reduced. At this optimal k_{dc}, L_p is commensurate with the value set by the chain degree of polymerization (512 cellobiose residues), but jamming is minimized.

3.3.3 Kinetic efficiencies of enzymes quantify the performance of interfacial biocatalysts

The significant impact of *Tr*Cel7A jamming induced by surface heterogeneities and excluded volume interactions motivates the development of efficiency measures to characterize enzyme performance at an interface. These measures are further described in Section 3.6.1. The ratio of surface to total enzymes is defined as the surface-binding efficiency, η_S. The ratio of complexed to surface enzymes is the complexation efficiency, η_C. The ratio of active to complexed enzymes is the activation efficiency, η_A. As such, η_S and η_C are efficiency measures for adsorption and complexation processes, respectively. The effects of substrate heterogeneity and obstacle-induced jamming are then lumped into η_A. Fig. 3.5a plots η_C and η_A as a function of k_{dc} for a 200-hr decomposition and their dependences on the decomplexation rate are clearly opposing as discussed earlier. η_C also has a close correspondence with the apparent processivity L_p.

The overall conversion can be quantitatively related to the efficiencies through an enzyme "performance" metric that incorporates the competing requirements for effective reactivity on cellulosic substrates. In Section 3.6.1, the conversion after time T, denoted as X_T, is shown to be approximately proportional to $\eta_C \eta_A$. The opposing effects of k_{dc} on conversion can then be described via the enzyme performance on the substrate surface, $\eta_C \eta_A$. The $X_T \sim \eta_C \eta_A$ correspondence is clearly seen in Fig. 3.5b.

The effects of removing the edge obstacle and/or changing the substrate degree of polymerization can be anticipated based on the opposing forces of jamming and processivity. Further discussion is provided in Sections 3.6.3 and 3.6.4.

3.3.4 A unified mechanistic framework of cellulose decomposition via the productivity map

A surprising but important finding is that less than one percent of the adsorbed *Tr*Cel7A enzymes on cellulose are active. Because surface inactivity of *Tr*Cel7A is predominantly due to slow complexation and excessive blocking, enhancement of cellulose decomposition likely requires simultaneous consideration of these two factors. SLATE simulations can be used to answer the following questions: If k_c and k_{dc} could be varied independently via protein engineering or substrate pretreatment while keeping the other rate constants unchanged, to what extent can the cellulose conversion rate be increased? How do the optimal k_c and k_{dc} differ from those of native *Tr*Cel7A enzymes?

Fig. 3.6a-e plots the fractional occupancy times in different states for 100 hours of conversion as a function of both the k_c and k_{dc} values of *Tr*Cel7A. The competition between increasing the complexation driving force and reducing the fractions of blocked enzymes in maximizing the final conversion (productivity) can be seen in Fig. 3.6. The region of elevated occupancy in the active state in Fig. 3.6b can be reached via high complexation and intermediate decomplexation rates. In this scenario, the enzymes complex relatively quickly and process through entire glucan chains on cellulose without forming traffic jams. As seen

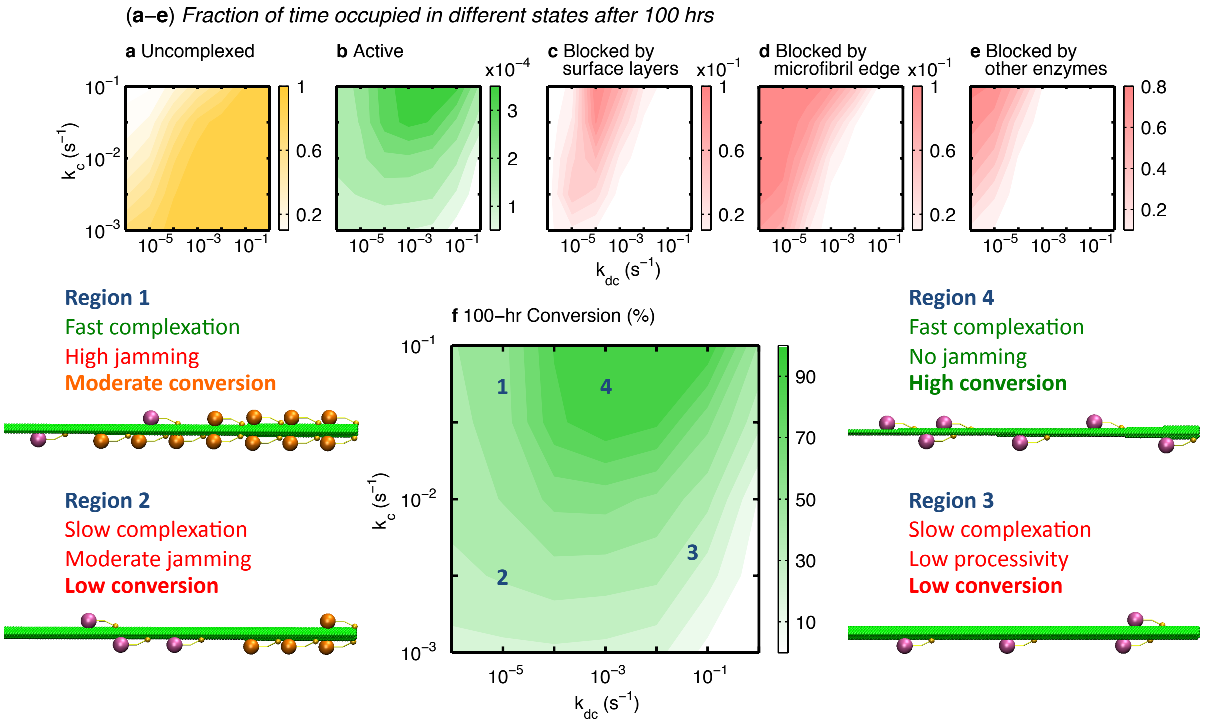


Figure 3.6. The maps of fractional occupancy times (**a-e**) and productivity of conversion (**f**) in the complexation-decomplexation ($k_c - k_{dc}$) plane. The fractional occupancy time for a state is its occupancy time divided by the total reaction time. Four kinetic regions are identified in the productivity map of (**f**), and a representative snapshot from a SLATE simulation is shown for each region. Enzymes in the blocked state are colored in orange. The kinetic characteristics of each region are also described.

in Fig. 3.6de, this high activity behavior corresponds to moderate occupancy fraction in the edge-blocked state, but minimal fraction in the enzyme-blocked state. The contour of the active-state fraction in Fig. 3.6b follows that of the 100-hr conversion in Fig. 3.6f due to the $X_T \sim \eta_C \eta_A$ correspondence discussed earlier.

The productivity map in Fig. 3.6f reveals the rich variety of cellulase behaviors on cellulose and unifies a diverse range of experimental observations. This comprehensive picture is made possible because the reaction kinetics is spatially resolved at the single-enzyme level. In Region 1, conversion is limited by excessive jamming rather than by slow complexation [88–90]. Hence, a decrease in processivity would enhance conversion. This prediction is corroborated in experiments with amorphous cellulose, on which complexation occurs quickly and jamming is prevalent [88,112]. In Region 2, both slow complexation and jamming are rate-limiting, which correspond to enzyme jamming on crystalline cellulose [64,96]. In Region 3, a decrease in processivity would lower conversion, in contrast to Region 1. This trend is confirmed by experiments with mutated cellulases on crystalline cellulose [112]. To achieve high productivity of Region 4, complexation must be enhanced along with a balance between processivity and jamming.

Lignocellulosic substrates in biomass feedstocks inevitably contain defects and obstacles arising from high contents of hemicellulose and lignin that would promote jamming [41,86]. A higher density of surface obstacles from non-cellulosic biomass components is mechanistically analogous to a lower decomplexation rate in the SLATE model. Therefore, kinetic behavior on lignocellulosic substrates likely falls in Region 2. In this regime, enhancing complexation or removing surface obstacles would increase productivity. In addition to protein engineering, these objectives can be achieved by substrate pretreatment technologies that remove lignin and decrystallize cellulose [17]. Therefore, the success of biomass pretreatment in promoting cellulose decomposition can be assessed via movement within the productivity map. The combined productivity and occupancy time analysis provides a framework for uncovering the rate limitations unique to each reaction system and identifying specific conversion strategies that would most effectively accelerate conversion.

3.4 Discussion

Kinetic simulations via the spatially resolved SLATE model elucidate the significant role of interfacial confinement on cellulose decomposition by *Tr*Cel7A. By incorporating molecular-scale spatial resolution, the factors that limit the rate of cellulose decomposition become explicit. For native *Tr*Cel7A, SLATE simulations underscore that the most important limiting factors are the slow rate of complexation and enzyme jamming. While complexation is expected to remain rate-limiting throughout the entire reaction, enzyme jamming develops further with reaction time and becomes important as the reaction proceeds. At later stages of the reaction, decomplexation can emerge as an additional rate-limiting step. Although this decomplexation-limited behavior has been inferred indirectly from experiments [88–90], SLATE simulations delineate the quantitative bounds for its occurence. In general, the SLATE model provides the conditions under which complexation [96,97], decomplexation [88–90], or both processes are rate-limiting. These findings underscore that a one-cause, time-independent explanation of enzyme inefficiency on crystalline cellulose provides an oversimplified and incomplete view. Such single-faceted approaches are likely inappropriate for understanding interfacial biocatalysis in general.

The flow of processive cellulases along microfibrils [64], as well as motor proteins along filaments [107], resembles in many respects vehicular travel along a highway. The reducing ends of glucan chains are analogous to narrow ramps restricting the entering flow of processing enzymes. The non-reducing edge of the microfibril acts as a “roadblock” that stops the flow of forward-moving enzymes and causes traffic jams. These jams can be avoided by introducing exit ramps via decomplexation. However, excessive decomplexation leads to inefficient detouring of traffic that reduces processive enzyme flow. SLATE simulations show that optimization of complexation-exchange kinetics can enhance the travel conditions along

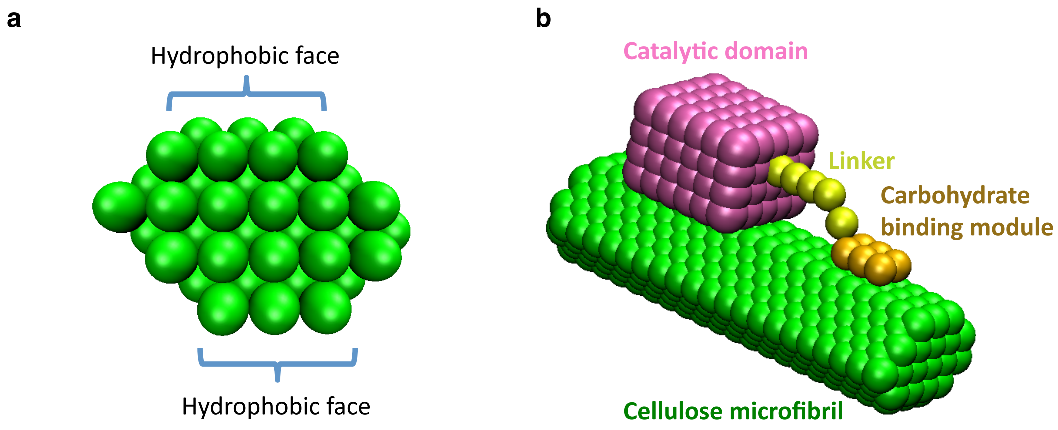


Figure 3.7. Microfibril and enzyme geometry in SLATE model. (**a**) End-on view of microfibril, showing hydrophobic faces available for adsorption. (**b**) A cellulase enzyme adsorbed onto a microfibril. The spheres represent a space-filling volume of 1 nm^3.

microfibrils and accelerate the rate of cellulose decomposition.

The presence of substrate heterogeneity in interfacial biocatalysis demands a spatially resolved approach to describe enzyme kinetics. For cellulose decomposition, the SLATE model illustrates that the inclusion of spatial confinement in the kinetic model is indispensable for uncovering the kinetic bottlenecks. Since crowded, inhomogeneous environments are ubiquitous in biological systems, the development of SLATE represents an important step toward a systems-level analysis of the spatiotemporal behaviors of enzyme reaction networks.

3.5 Supplementary methods

3.5.1 Structure of microfibril and cellulose

The structures of both the microfibril and cellulase are represented by space-filling cubes on a three-dimensional lattice. The cube size, or lattice spacing, is 5 Å.

The basic building block of the microfibril lattice is the cellobiose residue, which is represented as a rectangular prism with dimensions 1 nm × 0.5 nm × 1 nm (or 2 × 1 × 2 lattice units) [39,103]. The lattice structure of the microfibril is described in detail in Chang et al. [103]. Only the hydrophobic faces of the microfibril are available for adsorption [49,62]. These faces are illustrated in Fig. 3.7a.

The structure of the cellulase enzyme follows the crystal dimensions of *Tr*Cel7A. The carbohydrate binding module is modeled as a rectangular prism with dimensions 2 nm × 1 nm × 3 nm [52]. The catalytic domain is modeled as a rectangular prism with dimensions 5 nm × 4 nm × 6 nm [50]. A linker connects the two domains with length 5 nm, in accordance with computational studies [51]. An illustration of the enzyme and microfibril geometry is shown in Fig. 3.7b.

3.5.2 Determination of elementary kinetic rates

To calibrate the model with experiment, the elementary kinetic rates were determined from literature values to represent the conversion of Avicel by *Tr*Cel7A, as measured by Gusakov et al. [108]. From experiments [97,113–115], the desorption rate is estimated to be between $1 \times 10^{-3}\,\mathrm{s}^{-1} - 1 \times 10^{-2}\,\mathrm{s}^{-1}$; a value of $1 \times 10^{-3}\,\mathrm{s}^{-1}$ was chosen, although using $1 \times 10^{-2}\,\mathrm{s}^{-1}$ did not alter conversion profiles by more than 5%. To find the adsorption rate, the equilibrium adsorption constant K and system volume V are required. For *Tr*Cel7A on Avicel, a representative value for the equilibrium adsorption constant is $0.28\,\mathrm{L/\mu mol}$ [41,47]. The volume was chosen such that the cellulose concentration in the system is the same as the concentration used by Gusakov et al. [108], or $5\,\mathrm{mg/ml}$. The adsorption rate follows by transforming the phenomenological bimolecular adsorption rate $k_d K$ to a first-order association rate as $k_a = k_d K/(V N_{\mathrm{AV}} 10^{-6}) = 8.9 \times 10^{-4}\,\mathrm{s}^{-1}$ [116]. The factor $N_{\mathrm{AV}} 10^{-6}$ is used to convert µmol to number of molecules and N_{AV} is Avogadro's number.

The surface diffusion rate is related to the self-diffusion coefficient D_S from random walk theory as $k_{diff} = 4D_S/l_h^2$, where $l_h = 1\,\mathrm{nm}$ is the hopping distance [117]. An experimental diffusion coefficient of $1 \times 10^{-10}\,\mathrm{cm^2/s}$ [118] gives a diffusion rate of $1 \times 10^{4}\,\mathrm{s}^{-1}$. The hydrolysis rate is determined from the processive speed of *Tr*Cel7A, which is $7.1\,\mathrm{nm/s}$ [64]. Since a cellobiose unit is about $1\,\mathrm{nm}$ long, the corresponding hydrolysis rate is $7.1\,\mathrm{s}^{-1}$.

From experiments [88–90,97], the decomplexation rate is estimated to be within the range $1 \times 10^{-3}\,\mathrm{s}^{-1} - 1 \times 10^{-2}\,\mathrm{s}^{-1}$. In some studies [88–90], this rate is lumped together with desorption. As an initial estimate for SLATE simulations, the decomplexation rate was set to $1 \times 10^{-3}\,\mathrm{s}^{-1}$, but its value is extensively varied here to account for the uncertainty.

The complexation rate is difficult to directly measure in experiment, although studies suggest its value should be much lower than the hydrolysis rate [96,97,110,111]. This rate was determined by fitting the simulated conversion profile to the experimental conversion profile from Gusakov et al. [108]. In order to more precisely fit the experimental conversion, two different complexation rates were necessary. The top and bottom layers of the microfibril were assumed to be more loosely bound to the microfibril, and complexation with chains in these layers is assumed to occur 10 times faster than chains elsewhere. Since Avicel is known to be approximately 30% amorphous [119], the assumption that initial surface layers are more susceptible toward complexation is not unreasonable. The best-fit values for the fast and slow complexation rates are $5.5 \times 10^{-3}\,\mathrm{s}^{-1}$ and $5.5 \times 10^{-4}\,\mathrm{s}^{-1}$. Values for complexation rates presented in the main text refer to this faster value. Although two different complexation rates are introduced so that the experimental profile can be precisely fitted, the mechanistic trends and conclusions in this work do not depend on this assumption and hold when a single rate is used as well.

If only a single complexation rate were used, the sudden slow-down in conversion rate near 30%, as seen in the experimental profile, could not be captured by the model. Fig. 3.8

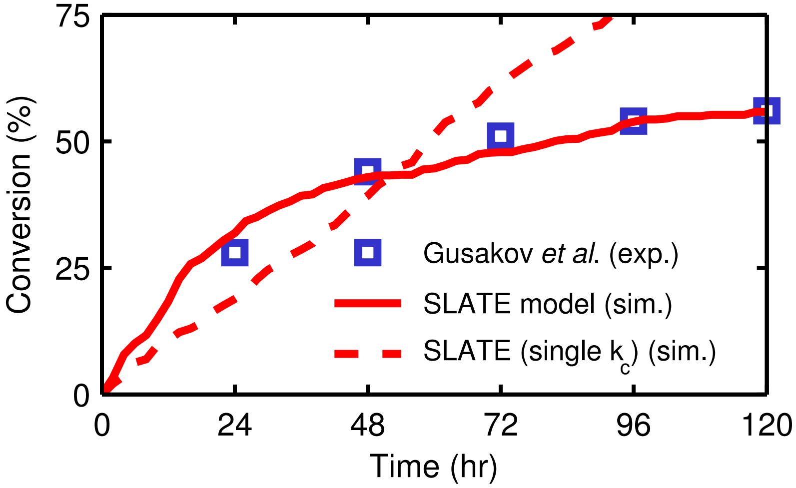


Figure 3.8. Effect of using single vs. dual complexation rates on shape of cellulose conversion profle. Conversion over time for SLATE model using two complexation rates k_c s^{-1} (5.5×10^{-3} and 5.5×10^{-4}) (solid red line), a single rate k_c (1.5×10^{-3}) (dotted red line), and from Gusakov et al. [108] (blue squares) for *Tr*Cel7A on Avicel. Conversions are averaged over 8 independent simulations.

shows an example of conversion profiles that result from using a single complexation rate ($1.5 \times 10^{-3}\,\mathrm{s}^{-1}$) and from using two different rates. Simulations results therefore suggest that this slow-down in rate can be attributed to the heterogeneous nature of the substrate [81,120]. It is not unreasonable to assume that naturally synthesized microfibrils would contain portions that are more easily decomposed and portions that are more recalcitrant. In particular, the initial surface layers of the microfibril should be more susceptible to complexation by enzymes. Therefore, the slow down in reaction rate can be attributed to the emergence of the more recalcitrant surface layers as decomposition proceeds [121].

It should be noted that the decline in conversion observed in Gusakov et al. [108] and in other experiments [78,80] could arise from factors such as product inhibition [122], enzyme thermal denaturation [123], or irreversible enzyme adsorption [124,125]. That is, an increasingly recalcitrant substrate may not be the only explanation for declining conversion rates. These additional behaviors are not included in the SLATE model, and it is beyond the scope of this work to evaluate their relative contributions on limiting conversion. The mechanistic trends and conclusions presented here are not expected to depend on the existence of these additional behaviors.

3.5.3 Lattice kinetic Monte Carlo simulations

The Stochastic LATtice Enzyme (SLATE) model applied here to simulate enzymatic cellulose decomposition is a lattice kinetic Monte Carlo (kMC) model [98]. These models have been used extensively to study surface kinetic phenomena occurring in, for example, heterogeneous catalysis [98] and motor protein transport [107]. Given the kinetic similarities that enzymatic cellulose decomposition shares with these systems, we apply kMC to model cellulose bioconversion. A cellular automaton model has been presented before for modeling cellulase kinetics on cellulose [94]. However, this model assumes complexation and decomplexation occur instantaneously, underestimates bulk and surface diffusion rates, and employs an ad hoc set of probability parameters that are difficult to interpret in a physical context. A lattice-based coarse-grained model has also been used to study cellulase kinetics [95]. This model, however, combines adsorption and diffusion processes with complexation, and desorption with decomplexation. Its limitation to two-dimensions also prevents spatial resolution of the microfibril degradation and enzyme-substrate coupling. In both models, the complexation-exchange kinetics cannot be studied as stand-alone steps, and the models do not quantitatively reproduce experimentally measured conversion profiles. The incorporation of a full suite of spatiotemporal behaviors in the SLATE model overcomes limitations in existing models that do not simultaneously account for spatial heterogeneity, complexation-exchange, processivity, and jamming.

To simulate the decomposition reaction, the null-event kMC algorithm is used [98], which is described further in Sections 2.1.4 and 3.5.4. The traditional lattice-based algorithm is modified to overcome computational limitations imposed by the high rate of diffusion. Because the diffusion rate is at least three orders of magnitude higher than the other rates, a naïve kMC simulation becomes computationally inefficient because the time step is severely reduced and most reactions become uninteresting diffusive jumps. Here, a significant speed up in computational time is achieved by treating diffusion adiabatically. A stochastic quasi-equilibrium approximation [102] is applied, and further described in Section 3.5.5. Conversion is found to be insensitive to diffusion if the diffusion rate is above $10\,\mathrm{s}^{-1}$ (See Section 3.5.5). Given an experimental rate of $1 \times 10^4\,\mathrm{s}^{-1}$, the quasi-equilibrium approach, which assumes infinitely fast diffusion, can be applied with high accuracy. Via this approach, the computational speed up is 50-fold, and conversion profiles up to 200 hours of reaction time can be simulated in less than five hours on a single CPU.

3.5.4 Null-event kinetic Monte Carlo algorithm

The kinetic Monte Carlo (kMC) algorithm is a method to generate stochastic trajectories consistent with an underlying master equation parameterized by the kinetic rate constants, or transition probabilities [116]. In particular, it involves choosing a reaction to perform during each iteration and then advancing the system time. The null-event kMC algorithm [98] is applied here. It consists of dividing the possible reactions from the current state into N_f families. A particular family is chosen, with each family being chosen with equal probability. Then, one particular reaction from that family, or no reaction is chosen. The probability for choosing a particular reaction is proportional to its kinetic rate constant. The time is then advanced by an amount determined by the number of families and kinetic rate constants. Further details are given in Section 2.1.4.

Specifically, in the Stochastic LATice Enzyme (SLATE) simulations, a subset of reactions is chosen by randomly selecting a cellobiose residue on the surface with equal probability. If the cellobiose residue is "linked" to an enzyme, then the possible reactions are enumerated for that enzyme, depending on its state and whether obstacles are present. The "linked" cellobiose residue is the one covered by the carbohydrate binding module (CBM) and closest to the non-reducing edge. If the cellobiose residue is "unlinked", then the possibility of enzyme adsorption onto the cellobiose site is tested. In this manner, three types of subsets can be chosen: (1) reactions involving an adsorbed enzyme, (2) reactions involving adsorption of any enzymes in solution to the chosen site, and (3) an empty set of reactions in the case that adsorption onto the chosen site is unallowable by excluded volume restrictions.

Once a subset of reactions is chosen, a reaction μ from that subset is chosen with probability $a_\mu/a_{\max}$. The parameter $a_{\max}$ must be above a certain threshold, as explained by Chatterjee & Vlachos [98] and Section 3.5.4. The time is advanced by an increment $1/(N_f a_{\max})$, so that the mean time for a reaction to occur is identical to that in the Gillespie algorithm [116]. The null-event algorithm is advantageous for spatially distributed systems, for which a full enumeration of all possible reactions (as required in the traditional Gillespie algorithm) would be too computationally expensive. The efficiency gained from only a partial enumeration of reactions outweighs the cost of a reduced time step.

3.5.5 Quasi-equilibrium for diffusion

The high rate of diffusion poses severe limitations to the traditional kinetic Monte Carlo (kMC) algorithm. First, because the probability to select a reaction is proportional to its rate, in the naïve algorithm, a large number of reactions will be diffusive hops. Second, because the time step is inversely proportional to the sum of the rates [98], the time step is severely reduced if diffusion is left untreated.

To overcome computational limitations posed by fast diffusion, the stochastic quasi-equilibrium approximation [102,126] is applied to the lattice kMC model. An outline of the algorithm is presented here. Reactions are partitioned into fast and slow reactions. Diffusion is the only fast reaction and all others are slow. After each slow reaction, a Metropolis Monte Carlo (MMC) sampling of the uncomplexed enzyme configurations using local diffusional hops of 1 nm is performed. After each MMC step, a set of slow reactions are tested to see if they can occur. This set corresponds to those slow reactions whose rates depend on the enzyme configurations (e.g. complexation and hydrolysis). If a slow reaction is allowed from this configuration, it is added to the reaction list and the enzyme configurations are saved. Therefore, after the MMC sampling, a new set of slow reactions are enumerated from the enzyme configurations accessible by diffusion. After the MMC sampling, a slow reaction (or null reaction) is selected. Following Samant & Vlachos [126], the probability of a slow reaction s occurring from a particular configuration x is proportional to $(N_x/N_{\text{total}})k_s(x)$, where N_x is the number of times configuration x is accessed by the MMC sampling, N_{total} is the total number of MMC steps, and $k_s(x)$ is the rate of slow reaction s as a function of x. We find that $N_{\text{total}} = 10{,}000$ provides sufficient sampling of the fast diffusive degrees of freedom.

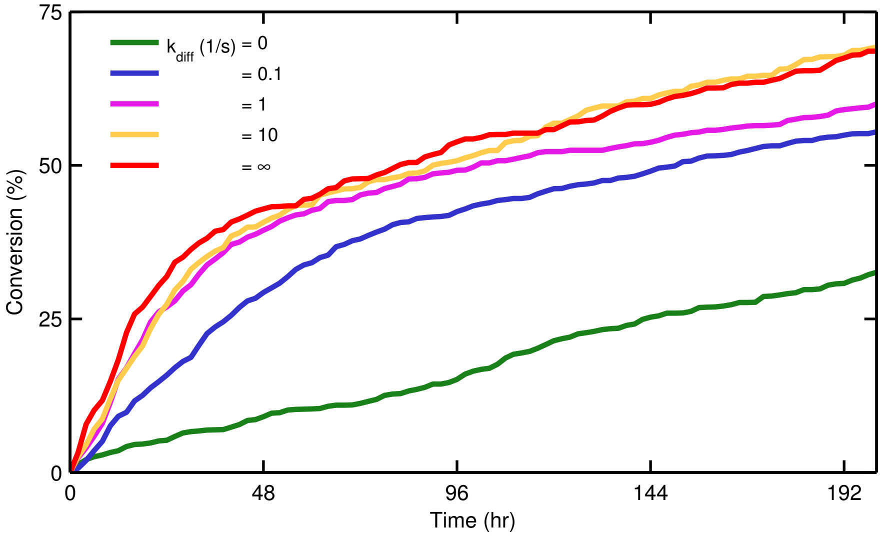


Figure 3.9. Conversion profile up to 200 hours for different diffusion rates k_{diff}. The quasi-equilibrium case is labeled as "∞".

Note that $k_s(x)$ is either 0 or a constant in our model. For example, if an uncomplexed enzyme is blocking the hydrolysis of another processing enzyme, the hydrolysis rate of the blocked enzyme from this configuration is $0\,\mathrm{s}^{-1}$. If the uncomplexed enzyme diffuses away after the next MMC step, then hydrolysis is no longer blocked and can proceed from this new configuration at a non-zero rate. After a slow reaction is selected and performed, a new cycle of MMC steps is initiated.

As a test of the algorithm, the conversion profile is plotted in Fig. 3.9 for diffusion rates (s^{-1}) of 0, 0.1, 1, 10, and ∞, where the latter represents the quasi-equilibrium assumption. We see that the conversion converges once the diffusion rate reaches 10. Experimental estimates of this rate range from $10\,\mathrm{s}^{-1}$ to $1 \times 10^4\,\mathrm{s}^{-1}$ [118,127] indicating that diffusion can be approximated as being infinitely fast (i.e. conversion is not diffusion-limited).

3.6 Supplementary information

3.6.1 Occupancy times, enzyme efficiency measures, and enzyme performance

Here, quantitative expressions for the occupancy times, enzyme efficiency measures, and enzyme performance are derived and related to the overall cellulose conversion. First, it is helpful to distinguish between the enzyme state and system state. The enzyme state refers to whether a particular enzyme is in solution, uncomplexed, active, or blocked. The system state contains information about the state and location of all enzymes and cellobiose residues in the system. The occupancy time for an enzyme state is the accumulated time an enzyme spends in its enzyme state over the course of the reaction, averaged over all enzymes in the system. In the kinetic Monte Carlo (kMC) simulation, the system jumps between system states via elementary kinetic steps. The system first spends a certain amount of time in the origin state, and then jumps instantaneously to the destination state. Over the course of the reaction, the system will explore different states, jumping from one to the next. Thus, the occupancy time for an enzyme state is a discrete sum over the system states encountered during the reaction, where the time spent in the system state is added to the occupancy time only if the enzyme is in the enzyme state of interest.

Let the system states encountered during the reaction be indexed by i, enzymes indexed by j, and the enzyme state of interest be labeled as M. Let S_{ij} be a function that returns the state of enzyme j when the system is in state i. The occupancy time for enzyme state M is

$$\tau_M = \frac{1}{N_{\text{Tot}}} \sum_i \sum_j \delta(S_{ij} - M) \Delta t_i, \tag{3.1}$$

where N_{Tot} is the total number of enzymes, $\delta(\cdots)$ is the Kronecker delta, and Δt_i is the length of time the system spends in state i. The occupancy time for state M is roughly equal to the number of times an enzyme encounters the state M multiplied by the mean reaction time (or mean first-passage time) from state M. Therefore, it incorporates two qualitative characteristics of a "kinetic trap": (1) that it is easily accessed, and (2) that it is difficult to escape from.

The fractional occupancy time, or fraction of time occupied in state M, is $x_M = \tau_M / \tau_{\text{Tot}}$, where $\tau_{\text{Tot}} = \sum_i \Delta t_i$ is the total reaction time. The quantity x_M can also be interpreted as the time-averaged fraction of enzymes in state M, which becomes more apparent by writing it as

$$x_M = \frac{1}{\tau_{\text{Tot}}} \sum_i \left(\frac{1}{N_{\text{Tot}}} \sum_j \delta(S_{ij} - M) \right) \tag{3.2}$$

The quantity in the outer parenthesis is the number of enzymes in state M when the system is in state i. The time-averaged number of enzymes in state M is $\widetilde{N}_M = x_M N_{\text{Tot}}$.

Enzyme efficiencies measure the degree to which certain kinetic processes are able to promote enzymes toward the active state. They quantify the fraction of an enzyme population that belong to a more productive population. As such, they are evocative of thermodynamic efficiencies that measure, for example, the fraction of heat converted into work. The enzyme populations, ordered from least productive to most productive, are the following: total, surface, complexed, and active. The total enzyme population consists of all the enzymes in the system. The surface population consists of uncomplexed, active, and blocked enzymes, which are all on the surface. The complexed population consists of active and blocked enzymes. The active population consists only of the enzymes in the active state. For each population, it is useful to consider the time-averaged number of enzymes in that population. It is the sum of the time-averaged number of enzymes over states belonging to that population. For example, the time-averaged number of enzymes on the surface is $\widetilde{N}_S = \widetilde{N}_U + \widetilde{N}_A + \widetilde{N}_B$, where

U, A, and B refer to the uncomplexed, active, and blocked states, respectively. The surface-binding efficiency is a measure of the fraction of all enzymes in the system that are bound to the surface, and given as $\eta_S = \widetilde{N}_S/N_{\mathrm{Tot}}$. The complexation efficiency is $\eta_C = \widetilde{N}_C/\widetilde{N}_S$, and the activation efficiency is $\eta_A = \widetilde{N}_A/\widetilde{N}_C$,

The enzyme efficiencies, and thus occupancy times, can be related to the cellulose conversion. The rate of cellobiose production is given by the first-order rate law: $dN_{CB}/dt = k_h N_A$, where N_{CB} is the number of cellobiose residues. Integration up to a total reaction time $T = \tau_{\mathrm{Tot}}$ gives $N_{CB}(T) = k_h \widetilde{N}_A \tau_{\mathrm{Tot}}$. The conversion X_T is $N_{CB}(T)/N_{CB,\mathrm{tot}}$, where $N_{CB,\mathrm{tot}}$ is the total number of cellobiose residues in the unreacted microfibril. Using the definition of the efficiencies, the conversion is $X_T = (1/N_{CB,\mathrm{tot}}) k_h N_{\mathrm{Tot}} \eta_S \eta_C \eta_A \tau_{\mathrm{Tot}}$. The maximum in conversion with decomplexation rate can be decomposed into the changes in efficiencies with decomplexation. $\widetilde{N}_S$ is nearly constant when changing the decomplexation rate, since its magnitude is mainly governed by the surface area and adsorption/desorption rates. Therefore, $X_T \sim \eta_C \eta_A$. The opposing trends of η_C and η_A with decomplexation rate quantify the competing requirements for maintaining processivity and avoiding blocking and explain the maximum in conversion. Their product $\eta_C \eta_A$ is performance metric that accounts for these competing requirements and is a clear reporter of the cellulose conversion.

The conversion is also proportional to the fractional occupancy time in the active state, or the time-averaged fraction of active enzymes. This follows from $N_{CB}(T) = k_h \widetilde{N}_A \tau_{\mathrm{Tot}}$ and $\widetilde{N}_A = x_A N_{\mathrm{Tot}}$. This relation $X_T \sim k_h x_A N_{\mathrm{Tot}} \tau_{\mathrm{Tot}}$ is intuitively pleasing since it says that conversion can be enhanced by maintaining a large pool of active enzyme over the course of the reaction. Additionally, from $x_A = \tau_A/\tau_{\mathrm{Tot}}$, the conversion is proportional to the occupancy time in the active state $X_T \sim k_h N_{\mathrm{Tot}} \tau_A$. This result is also intuitive; more time spent in the active state, from which hydrolysis directly proceeds, will increase conversion. The "lost" time spent in the inactive states reduces x_A and hence, decreases conversion.

3.6.2 Testing uncertainty in the complexation rate

The best-fit complexation rate obtained from the reference simulations could be underestimated if the enzyme loading in the simulations were too high. Using a lower enzyme loading, corresponding to 12% surface coverage, the complexation rate that best fits the experimental conversion by Gusakov et al. [108] is $1 \times 10^{-2}\,\mathrm{s}^{-1}$. This rate is still three orders of magnitude lower than the hydrolysis rate ($7.1\,\mathrm{s}^{-1}$). It is also interesting to determine if complexation would be rate-limiting if a higher conversion of 80% is achieved after 3 days of saccharification, a common industrial target [23]. To achieve this target, a complexation rate of $5 \times 10^{-2}\,\mathrm{s}^{-1}$ is required, still quite low compared to hydrolysis. An overestimation of the diffusion coefficient could also yield an artificially low complexation rate. Recent evidence suggests that the diffusion coefficient of $1 \times 10^{-10}\,\mathrm{cm}^2/\mathrm{s}$, as measured by Jervis et al. [118] and used in the reference simulation, may have been too high, and suggests the diffusion coefficient for mobile, uncomplexed cellulases may be as low as $1 \times 10^{-13}\,\mathrm{cm}^2/\mathrm{s}$ [127]. Simulations performed here indicate that the conversion is virtually unchanged even at this lower value, as shown in Section 3.5.5 for the diffusion rate of $10\,\mathrm{s}^{-1}$, which corresponds to a diffusion coefficient of $10^{-13}\,\mathrm{cm}^2/\mathrm{s}$ (See Section 3.5.2). This result indicates that diffusion is not rate-limiting and that complexation is the primary step limiting conversion. A recent experiment found that the lumped diffusion-complexation step is rate-limiting [96]; the isolation of complexation as rate-limiting further elucidates this finding. An underestimation of the decomplexation rate can also yield a best-fit complexation rate that is too low. However, even when a decomplexation rate of $1\,\mathrm{s}^{-1}$ was used, corresponding to the higher bound on literature estimates [114], the best-fit complexation rate was $1 \times 10^{-1}\,\mathrm{s}^{-1}$, still more than two orders of magnitude lower than hydrolysis.

3.6.3 Effect of non-reducing edge boundary condition on conversion

The possibility of jamming requires the presence of obstacles that can impede an enzyme's processive path. In the reference model, the primary obstacle that induces jamming is blocking at the non-reducing edge. Once a processing active enzyme reaches the non-reducing edge, it becomes blocked and can only decomplex. If decomplexation occurs slowly, then this enzyme remains inactive and the conversion rate is reduced due to loss of enzyme activity. An important question is if the conversion rate is enhanced when this obstacle, and hence the possibility of jamming, is removed. In the SLATE model, this is accomplished by coupling the hydrolysis of the remaining bonds at the non-reducing end with decomplexation and desorption of the enzyme. Because hydrolysis occurs more quickly than decomplexation, enzymes reaching the non-reducing edge are no longer jammed but return to solution.

To compare the effect of blocking at the non-reducing edge on enzyme jamming, the probability to find a surface enzyme at different positions along the microfibril is plotted for simulations with and without the non-reducing edge obstacle in Fig. 3.10a. A high density of enzymes near the non-reducing edge indicates the existence of jamming. This is seen when the obstacle is included and $k_{dc} = 1 \times 10^{-5}\,\mathrm{s}^{-1}$. Once k_{dc} is increased to $1 \times 10^{-3}\,\mathrm{s}^{-1}$, the jamming disappears. Removal of the obstacle eliminates jamming when $k_{dc} = 1 \times 10^{-5}\,\mathrm{s}^{-1}$, but has a negligible effect on enzyme density when $k_{dc} = 1 \times 10^{-3}\,\mathrm{s}^{-1}$, since jamming initially is absent.

The elimination of jamming yields an increase in conversion. In Fig. 3.10b, the conversion after 200 hours is compared for simulations with and without blocking at the non-reducing edge. For $k_{dc} < 1 \times 10^{-3}\,\mathrm{s}^{-1}$, removing the obstacle leads to a significant increase in conversion, since jamming is eliminated. At higher k_{dc}, obstacle removal has minimal effect on conversion. Under these conditions, processing enzymes are less likely to reach the non-reducing edge, and those enzymes that do reach it can decomplex quickly anyways. In these cases, the presence of an obstacle at the non-reducing edge is less important. One of the goals of cellulose pretreatment is to remove obstacles that limit substrate accessibility [17,128]. These results show that the response of conversion to pretreatment could depend not only on substrate modifications themselves, but also on the intrinsic enzyme

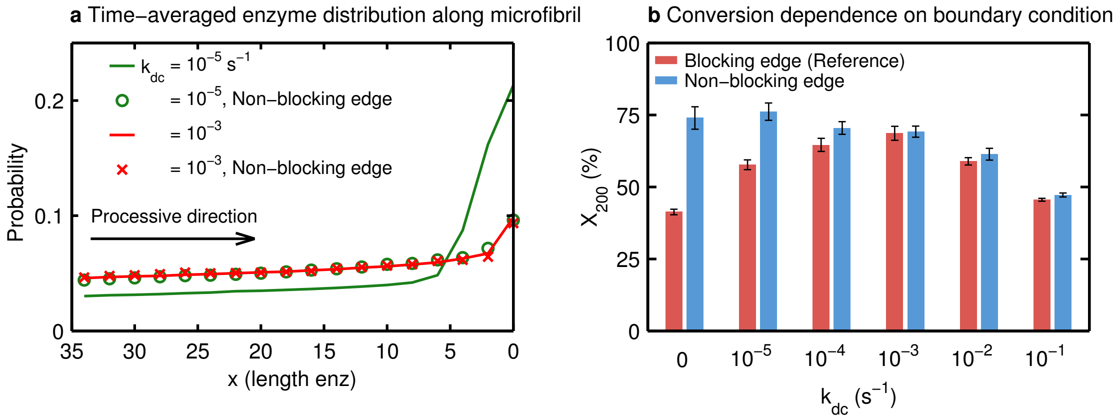


Figure 3.10. Effect of non-reducing edge boundary condition on cellulose conversion by *Tr*Cel7A. (**a**) Probability to find an enzyme at different locations along the microfibril averaged over 200 hours of reaction. The non-reducing edge is located at $x = 0$. The length of the enzyme is 14 nm. Two different decomplexation rates k_{dc} (s^{-1}) are shown: 10^{-5} (green) and 10^{-3} (red). Traffic jams, shown as a buildup of density near $x = 0$, disappear for $k_{dc} = 10^{-3}\,\mathrm{s}^{-1}$. (**b**) Conversion after 200 hours of reaction X_{200} for different decomplexation rates k_{dc}. In the reference case (red), the non-reducing edge is an obstacle. In the alternative case (blue), the non-reducing edge is not an obstacle. In this case, enzymes that reach the non-reducing edge decomplex and desorb immediately.

kinetics. Less processive enzymes in this case would be less sensitive to obstacle removal.

3.6.4 Effect of degree of polymerization on conversion

Lowering the degree of polymerization (DP) of glucan chains under fixed kinetic rate constants (s^{-1}) can either increase or decrease the conversion. The types of behavior can be understood by examining if lowering DP (or obstacle-free length of the microfibril) induces blocking at the non-reducing edge. A useful method to predict blocking is to compare the DP with the enzyme intrinsic processivity P_{intr}. This is the processive length for a freely processing enzyme in the absence of substrate limitations, which equals k_h/k_{dc} [88]. The reference value is 7000 according to the rates in Fig. 3.1b. Unlike the apparent processive length, which depends on the substrate DP, P_{intr} depends on the kinetic rate constants alone. As P_{intr} increases, the processing enzymes are more likely to become blocked at the non-reducing edge.

In Fig. 3.11, the dependence of conversion on DP is illustrated for varying k_{dc} (or equivalently, varying P_{intr}). Microfibril DPs from 256 to 1024 are explored, and the total number of enzymes in the system is adjusted to maintain a constant enzyme-to-substrate ratio. Simulations in Fig. 3.11a consider the non-reducing edge as an obstacle that blocks further hydrolysis. Therefore, DP here is also equal to the length of an obstacle-free path on the

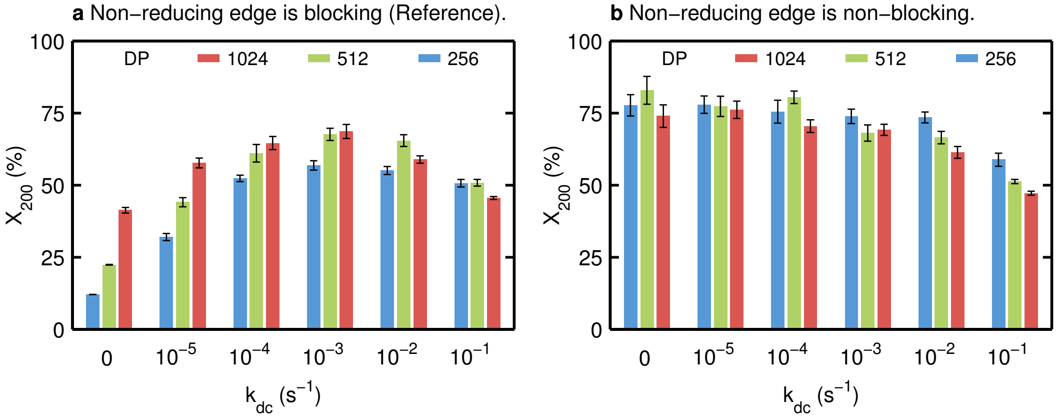


Figure 3.11. Effect of microfibril degree of polymerization (DP) on cellulose conversion by *Tr*Cel7A. Conversions after 200 hours of reaction X_{200} for different decomplexation rates k_{dc} and substrate degrees of polymerization (DPs) are shown. (**a**) The non-reducing edge is an obstacle, and therefore, DP is also equal to the length of an obstacle-free path. (**b**) The non-reducing edge is not an obstacle. In these simulations, enzymes that reach the non-reducing edge decomplex and desorb immediately.

initial substrate. The conversion is determined by the balance between P_{intr} and DP. When P_{intr} is high, a decrease in DP can create or induce further jamming because more enzymes reach the non-reducing edge and remain blocked for a longer period of time. This occurs for $k_{dc} < 1 \times 10^{-2}\,\mathrm{s}^{-1}$, and the conversion is lowered as DP is decreased. When P_{intr} is low, enzymes tend to decomplex prematurely, and bonds closer to the non-reducing edge become harder to hydrolyze. Therefore, a decrease in DP can make these bonds easier to hydrolyze by placing them closer to the complexation sites. This occurs for $k_{dc} = 1 \times 10^{-1}\,\mathrm{s}^{-1}$, and conversion is increased as DP is decreased.

An interesting transition is seen at $k_{dc} = 1 \times 10^{-2}\,\mathrm{s}^{-1}$. A decrease in DP from 1024 to 512 first increases the conversion by allowing bonds near the non-reducing end to be hydrolyzed more easily. A further decrease in DP from 512 to 128, however, begins to induce jamming, lowering conversion. A decomplexation rate around $1 \times 10^{-3}\,\mathrm{s}^{-1}$ – $1 \times 10^{-2}\,\mathrm{s}^{-1}$ based on experimental estimates [88–90,97] appears to give near optimal conversion for a range of substrate DPs. This result could indicate the outcome of an evolution toward the optimal k_{dc} to allow *Tr*Cel7A to cope with the diversity of cellulosic substrates encountered in nature.

Variations of the DP have also been performed under the case where the non-reducing edge is not an obstacle. As seen in Fig. 3.11b, by removing blocking at the non-reducing edge, a decrease in DP only enhances the conversion when P_{intr} is low. Under these conditions ($k_{dc} > 1 \times 10^{-3}\,\mathrm{s}^{-1}$), reduction in DP allows bonds closer to the non-reducing end to be reached by a processing enzyme. When P_{intr} is high, the conversion is relatively insensitive to reduction in DP, since blocking at the non-reducing edge is non-existent. Experimental studies regarding the effect of DP on conversion do not show any consistent trends [129–131], and the results here point to substrate- and enzyme-related features that can account for these inconsistencies. The non-trivial interplay between substrate features and enzyme kinetics observed here underscores that conversion enhancement requires achieving coordination between both substrate and enzyme interactions.

Chapter 4

Elucidating the Molecular Origins of Cellulase Synergy

4.1 Motivation

Harnessing the synergistic interactions among cellulases is one of the primary strategies for enhancing cellulose conversion rates [73,76]. Industrial cellulase preparations employ a cocktail of enzymes to exploit their synergistic interactions [26,30,42,76]. Among the most important types of enzymes are those secreted by the fungus *Trichoderma reesei*, which include two exoglucanases (or cellobiohydrolases) (*Tr*Cel7A, *Tr*Cel6A) and one endoglucanase (*Tr*Cel7B) [41]. These enzymes comprise the majority of cellulases secreted by this fungus [44,132]. Exoglucanases complex with the free ends of glucan chains and perform processive hydrolysis, releasing cellobiose into solution after each bond cleavage [46]. *Tr*Cel7A complexes with the reducing end, while *Tr*Cel6A complexes with the nonreducing end [46,133,134]. Complexation for exoglucanases involves extraction of the chain from the crystalline surface and threading through the active site tunnel [27,57]. Endoglucanases can complex anywhere within the glucan chain [46,66,67] by lifting it up from the surface and into its active site cleft [54,57]. These enzymes are not processive and decomplex after hydrolysis.

Endo-exo synergy among these fungal cellulases is well-studied [65,72,73,80,86,91,96,135–142], but its molecular origins remain an open question [26,73]. Synergy is mainly believed to arise from the ability of endoglucanases to create more free ends with which exoglucanases can complex [73,135]. Therefore, the slow rate of exoglucanase complexation can be partially overcome by exploiting this endo-exo synergy. A second explanation for endo-exo synergy is the putative ability of endoglucanases to remove surface obstacles that block processive hydrolysis of exoglucanases, thereby alleviating enzyme jamming [86,96,137,138]. A third explanation is that the internal cuts that endoglucanases create could promote decomplexation of processing exoglucanases to prevent them from becoming jammed at obstacles [91]. Among these theories, it is unclear which one is dominant and if these theories require further refinement.

Exo-exo synergy between *Tr*Cel7A and *Tr*Cel6A [64,67,72,82,136,143–146] is believed to arise due to the ability of *Tr*Cel6A to display some endo-like character [67,69,132]. The shorter flexible tunnel-forming loops around the active site can occasionally open to form a cleft-like topology, which would allow complexation with interior portions of the glucan chain [41,59,68]. An additional theory is that *Tr*Cel6A, which processes in the opposing direction of *Tr*Cel7A, could also play a role in clearing out isolated surface glucan chains left by incomplete degradation by *Tr*Cel7A [82]. It is currently unclear if one or both of these mechanisms operate within exo-exo mixtures.

Spatially-resolved kinetic models can provide molecular insight into the reaction kinetics which is inaccessible to solution-phase experiments. An enhanced understanding of the molecular origins of synergy could establish a more comprehensive framework for rational engineering of cellulase enzymes and their substrates. However, the kinetic mechanisms for synergy among cellulases have yet to be modeled comprehensively at the single-enzyme level. To the best of our knowledge, current models for enzyme synergy [61,95,114,131,136,147–150] do not simultaneously consider the effects of (1) spatial heterogeneity, (2) excluded volume interactions, (3) complexation/decomplexation kinetics, and (4) enzyme jamming on the degree of synergy among cellulases. These non-classical kinetic behaviors are difficult to capture with traditional mass-action kinetic models which assume a dilute, well-mixed reaction medium. Lacking a molecular understanding of enzyme synergy, a more rational approach towards bioconversion design becomes prohibitive. To achieve deeper molecular insight, we have developed a kinetic model that spatially resolves the enzyme and substrate configurations in three dimensions and simulates the kinetic reactions at the single-enzyme level. The Stochastic LATtice Enzyme (SLATE) model applied here captures these non-classical enzymatic behaviors mentioned above. It based on a lattice kinetic Monte Carlo framework further discussed in Section 2.2. In the SLATE model, the enzymes are able to perform a full suite of kinetic reactions: (1) adsorption, (2) desorption, (3) diffusion, (4) complexation, (5), decomplexation, and (6) hydrolysis. The locations and shapes of individual enzymes and the microfibril glucan chains are explicitly tracked.

In this Chapter, we elucidate the molecular origins for cellulase synergy via simulations with the SLATE model. We show that endo-exo synergy requires not only an increase in the number of free glucan chain ends produced by endoglucanases. Instead, for synergy to emerge, an increase in the complexation *rate* between exoglucanases and the newly created free ends is required. The physical significance of this complexation rate enhancement is that endoglucanases likely disrupt, or decrystallize, the surface during complexation and hydrolysis. Therefore, the synergistic interaction with respect to exoglucanases is to not only increase the chain end concentration but also the chain end susceptibility to complexation.

We also find that, during a reaction with an endo-exo mixture, the microfibril surface becomes rougher and induces blocking of processing *Tr*Cel7A enzymes. This anti-synergistic interaction is uncovered because substrate structure and excluded volume constraints are explicitly incorporated in the SLATE model. We show that this anti-synergistic interaction can be avoided by using less processive exoglucanases. The enhanced decomplexation rates of these enzymes alleviate the anti-synergistic effects of increasing surface roughness and blocking, since the activities of these enzymes are limited primarily by their processivity and not by the presence of obstacles, as discussed in Chapter 3. Under these conditions, the complexation rate enhancement far outweighs the increase in blocking. In general, we find that the endo-modulated competition between complexation and blocking of exoglucanases determines the degree of synergism.

Lastly, we illustrate that exo-exo synergy requires an endo-like behavior in at least one of the enzyme types. We show that anti-synergy occurs without this behavior. Gradual mutation of one of the exoglucanase enzymes in an exo-exo mixture into a more endo-like enzyme enables the exo-exo synergy to emerge.

4.2 Methods

Details of the SLATE model are provided in Section 2.2. These details include the kinetic Monte Carlo algorithm used to simulate the reaction, the shape and geometry of the enzymes and microfibril, and the elementary kinetic steps for all enzyme types in the reaction network. In addition, in Section 3.5, further details are provided for the simulation of *Tr*Cel7A alone, including the determination of a reference set of rate constants and the division of *Tr*Cel7A enzymes into different states used to analyze the kinetic bottlenecks. In this Section, the reference rate constants for *Tr*Cel7B and *Tr*Cel6A will be described.

4.2.1 Reference rate constants for *Tr*Cel7B

The adsorption and desorption rates of *Tr*Cel7B are the same as those of *Tr*Cel7A, similar to findings by Levine et al. [114]. The adsorption rate is $8.9 \times 10^{-4}\,\mathrm{s}^{-1}$, and the desorption rate is $1.0 \times 10^{-3}\,\mathrm{s}^{-1}$. The hydrolysis rate of *Tr*Cel7B is $9.2\,\mathrm{s}^{-1}$, adopted from work by Claeyssens et al. [151]. This value is consistent with the general observation that the intrinsic rate of bond cleavage by *Tr*Cel7B is higher than that of *Tr*Cel7A [88,91,95,114,151]. After hydrolysis, decomplexation of *Tr*Cel7B occurs immediately, since the enzyme is non-processive. Both hydrolysis and decomplexation are therefore treated as lumped kinetic steps for *Tr*Cel7B.

The complexation rate of *Tr*Cel7B is adjusted such that the cellulose conversion from *Tr*Cel7B alone is about half the value from *Tr*Cel7A alone. The lower conversion rate from *Tr*Cel7B is generally observed in experiments [41,65,72,80,86,91,137,152]. This best-fit value is $3.0 \times 10^{-3}\,\mathrm{s}^{-1}$. The finding that the complexation timescale for *Tr*Cel7B is on the order of hours is consistent with measurements by Maurer et al. [153]. After complexation, the enzyme is only allowed to hydrolyze a bond and then decomplex.

4.2.2 Reference rate constants for *Tr*Cel6A

The adsorption and desorption rates of *Tr*Cel6A are 10 times lower than those of *Tr*Cel7A to account for its less reversible binding [113,114,154,155]. The adsorption equilibrium constant is therefore unchanged, also consistent with experimental findings [41,44,143,145]. The decomplexation rate of *Tr*Cel6A is estimated to be $1.0\,\mathrm{s}^{-1}$. This value aims to capture the more flexible tunnel-forming loops of *Tr*Cel6A [59,68], which impart lower processivity [70] and partial endo-like character onto the enzyme [67,69]. The occasional opening of these loops would promote a faster decomplexation rate relative to that of *Tr*Cel7A, which possess more rigid loops [56]. The hydrolysis rate is $14\,\mathrm{s}^{-1}$, as measured experimentally [156,157].

The complexation rate of *Tr*Cel6A with free chain ends assumed to be the same as that of *Tr*Cel7A ($k_{c,7\mathrm{A}} = 5.5 \times 10^{-3}\,\mathrm{s}^{-1}$). This approximation is reasonable because both of these enzymes have active site tunnels [50,54], and therefore, complexation involves the similar physical steps of threading a chain into the active site. In addition to complexation with the non-reducing end of a glucan chain, *Tr*Cel6A can also complex with the interior portions of a chain. The complexation rate with these interior portions is taken to be equal to $\beta \times k_{c,6\mathrm{A}}$. Therefore, β is a measure of the ability of *Tr*Cel6A to create endo-like interior cuts. We vary β from 0 to 1 in the simulations to modulate this ability. Unlike *Tr*Cel7B, however, the enzyme does not decomplex immediately after hydrolysis, but can perform processive hydrolysis until it decomplexes or become blocked. Therefore, it exhibits an “endo-processive” behavior. In the simulations, the conversion from including this endo-processive activity for *Tr*Cel6A is typically higher than the conversion from *Tr*Cel7A alone. This observation is consistent with experiments, which generally find that the conversion from *Tr*Cel6A is higher [41,64,72,82,143,152,158–160]. The conversion is enhanced because *Tr*Cel6A can initiate processive hydrolysis from interior portions of a glucan chain, while *Tr*Cel7A cannot.

The diffusion rate for both *Tr*Cel7B and *Tr*Cel6A is the same as that for *Tr*Cel7A, $1.0 \times 10^4\,\mathrm{s}^{-1}$, which from random walk theory [117] and a hopping length of 1 nm corresponds to a diffusion coefficient of $1.0 \times 10^{-10}\,\mathrm{cm}^2/\mathrm{s}$, as measured experimentally for cellulases [118].

4.2.3 Scaling of *Tr*Cel7A complexation rate via α

An interior cut by *Tr*Cel7B or *Tr*Cel6A leaves a free glucan chain end that can be complexed by *Tr*Cel7A. Complexation of *Tr*Cel7A with these newly created ends occurs with rate $\alpha \times k_{c,7\mathrm{A}}$. We simulate the conversions using $\alpha = \{1, 10, 100\}$ to study this complexation rate enhancement effect. We also assume that complexation occurs 10 times more quickly with the initial top and bottom layers of the microfibril. Due to the heterogeneous structure of even crystalline cellulose [119], it is not unreasonable to assume certain parts of the native material are more susceptible to complexation than other parts. In the text, complexation rates refer to the higher value. Although we assume this structural heterogeneity to more precisely fit to experimental conversions (See Section 3.5.2), this treatment does not affect the mechanistic trends discussed in this Chapter.

Table 4.1. Reference rate constants for *Tr*Cel7A, *Tr*Cel7B, and *Tr*Cel6A used in SLATE simulations.

Reference rates (s^{-1})	*Tr*Cel7A	*Tr*Cel7B	*Tr*Cel6A
Adsorption, k_a	8.9×10^{-4}	8.9×10^{-4}	8.9×10^{-5}
Desorption, k_d	1.0×10^{-3}	1.0×10^{-3}	1.0×10^{-4}
Diffusion, k_{diff}	1.0×10^{4}	1.0×10^{4}	1.0×10^{4}
Complexation, k_c	5.5×10^{-3}	3.0×10^{-3}	5.5×10^{-3}
Decomplexation, k_{dc}	1.0×10^{-3}	—	1.0
Hydrolysis, k_h	7.1	9.2	14

Table 4.1 lists the reference rate constants used in the SLATE simulations. In the following simulations, when rate constants different from these are used, their values will be provided in the accompanying text.

4.3 Results

In the following, we identify the kinetic mechanisms that govern cellulase synergy. We also analyze the validity of current mechanistic theories for cellulase synergy via the SLATE model. First, endo-exo synergy is investigated. We find that the degree of synergy depends primarily on the competition between enhanced complexation and increased jamming of *Tr*Cel7A (exoglucanase) that occurs due to the addition of *Tr*Cel7B (endoglucanase). Second, exo-exo synergy is investigated. We find the ability of *Tr*Cel6A (exoglucanase) to create interior endo-like cuts within the glucan chain is essential for synergistic exo-exo interaction.

4.3.1 Endo-exo synergy depends on mixture composition, complexation rate enhancement, and reaction time

Here, the decomposition of a single microfibril by *Tr*Cel7A and *Tr*Cel7B is investigated. The total enzyme number is 18, corresponding to an initial surface coverage of 25%, within the range typically employed in experiments with enzyme mixtures [72,114,137]. This number is calculated as the intersection between the loading line [72] and Langmuir isotherm [47]. The numbers of *Tr*Cel7A and *Tr*Cel7B enzymes are varied between 0 and 18, while keeping the total enzyme number constant. We distinguish between rates of *Tr*Cel7A complexation with a pre-existing chain end and an endo-created chain end. Complexation with an endo-created chain end occurs at a rate of $\alpha \times k_c$, and α takes on values of 1, 10, and 100.

The effect of using $\alpha > 1$ is to physically model the disruption of interior portions of glucan chains after complexation and hydrolysis by *Tr*Cel7B. This complexation enhancement could occur by loosening of the hydrogen bonds between glucan chains by endoglucanases during complexation which would in turn increase the susceptibility of these ends toward *Tr*Cel7A. The effect of endoglucanases and other cellulases to disrupt the cellulose fiber has been suggested from experimental studies [92,161–167].

In Fig. 4.1, the conversion profiles for reactions with pure *Tr*Cel7A, pure *Tr*Cel7B, and their mixtures at different α values are shown. The enzyme ratios (Number of *Tr*Cel7A : Number of *Tr*Cel7B) are 2:16, 8:10, 10:8, and 16:2. The sum of the pure conversions is also shown, representing the null-synergy case. This null-synergy case corresponds to the conversion that would emerge if the two types of enzymes were to act completely independently on the surface. Synergy is defined as the mixture conversion divided by the sum of the conversions from the pure components.

For all enzyme ratios in Fig. 4.1, we observe that using $\alpha = 1$ produces null- or anti-synergy between the enzymes. Therefore, if the action of *Tr*Cel7B were to only increase the number of chain ends without affecting the *Tr*Cel7A complexation rate with these chain ends, no endo-exo cooperativity is achieved. In fact, at higher *Tr*Cel7B mole fractions (Fig. 4.1a-c), the interaction is anti-cooperative. For $\alpha = 10$ and 100, synergy is achieved at shorter times (< 40 hr) for all mole fractions, and is maintained at longer times when the *Tr*Cel7B mole fraction is low (Fig. 4.1d). At higher *Tr*Cel7B mole fractions (Fig. 4.1a-c), however, anti-synergy appears at later times (> 40 hrs). In addition, under high *Tr*Cel7B mole fractions, the conversion from $\alpha = 10$ begins to approach the conversion from $\alpha = 1$ at later times, while the conversion from $\alpha = 100$ drastically decreases, even falling below the conversion from $\alpha = 1$.

In Fig. 4.2, the conversion and synergy after selected times are shown for the pure enzyme and mixture reactions at different mixture compositions. For $\alpha = 1$, the synergy at all mole fractions is either non-existent (~ 1) or negative (< 1) throughout the reaction. For $\alpha > 1$, we see that the extents of synergy are around $1 - 3$, which is consistent with the values typically observed in experiments [41,65,72,80,86,91,137,142]. In addition, for all values of α and mole fractions, synergy decreases with reaction time, a behavior observed in experiments [65,80,86,91,137,141–143] but lacking a mechanistic explanation.

In general, Fig. 4.2 emphasizes the importance of the mixture composition, *Tr*Cel7A complexation enhancement (via α), and reaction time scale on the degree of endo-exo synergy. For example, at earlier times (< 24 hrs), increasing α increases the degree of synergy, while at later times (72 hrs), increasing α decreases synergy when the *Tr*Cel7B mole fraction is high, but increases synergy when the *Tr*Cel7B mole fraction is low. We now explain these observations by analyzing the enzyme state fractions to identify the mechanistic origins for these behaviors.

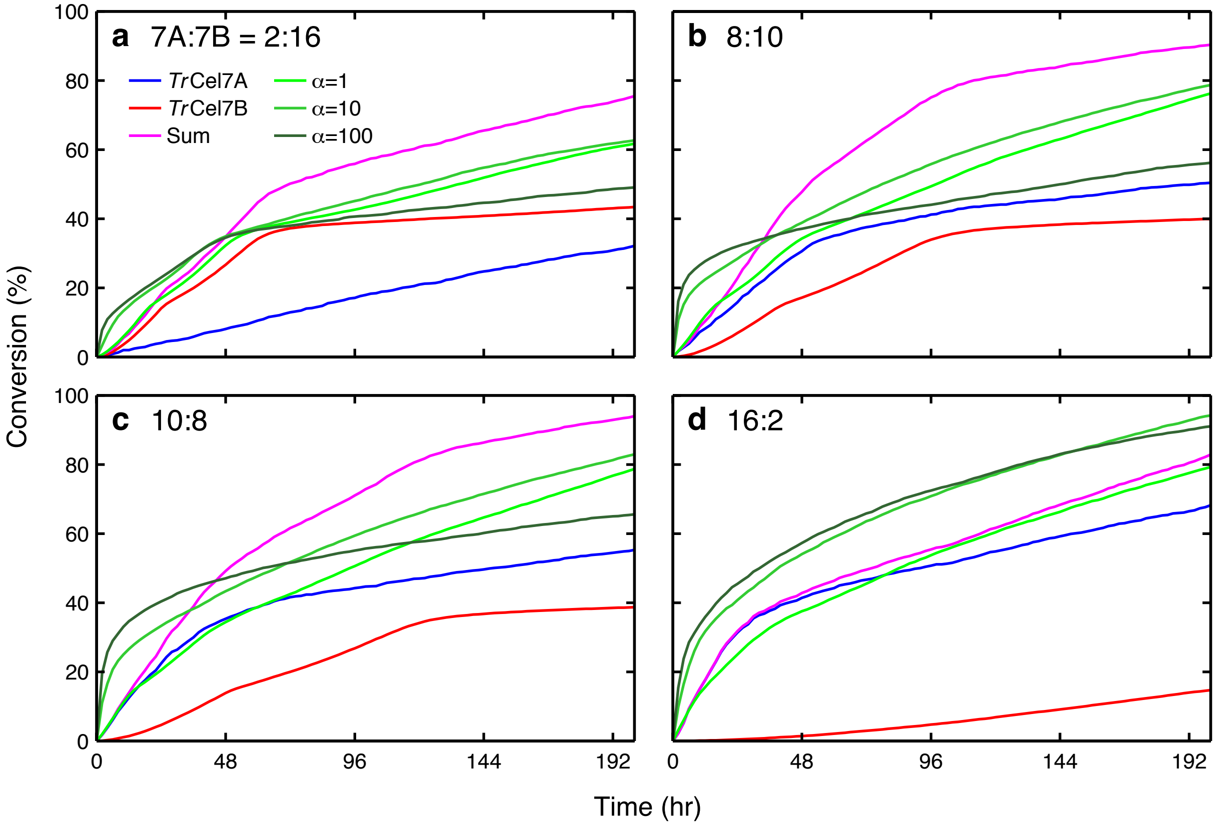


Figure 4.1. Conversion vs. time for mixtures of *Tr*Cel7A and *Tr*Cel7B at different mixture compositions and different values of α. Values for α are 1, 10, and 100. The blue line represents conversion from *Tr*Cel7A only. The red line represents conversion from *Tr*Cel7B only. The magenta line represents the sum of conversions from *Tr*Cel7A only and *Tr*Cel7B only and is the null-synergy conversion. The green lines indicate the actual mixture conversions for different values of α. The mixture composition is given as Number of *Tr*Cel7A : Number of *Tr*Cel7B. (**a**) 2:16, (**b**) 8:10, (**c**) 10:8, (**d**) 16:2.

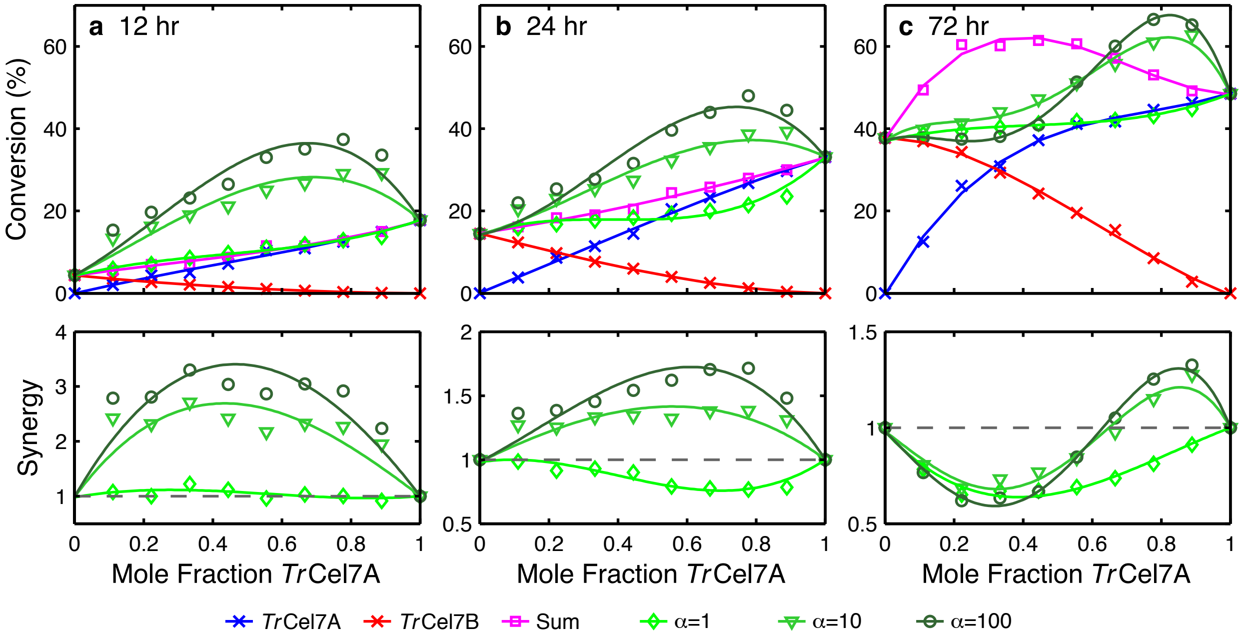


Figure 4.2. Conversion and synergy vs. mole fraction for mixtures of *Tr*Cel7A and *Tr*Cel7B enzymes at different values of α and reaction times. The blue crosses represent conversion from *Tr*Cel7A only. The red crosses represent conversion from *Tr*Cel7B only. The magenta squares represent the sum of conversions from *Tr*Cel7A only and *Tr*Cel7B only and is the null-synergy conversion. The green symbols indicate the actual mixture conversions and synergy at different values of α: 1 (diamond), 10 (inverted triangle), and 100 (circle). Synergy is defined as the mixture conversion divided by the sum of the individual conversions. The reaction times (hr) are (**a**) 12, (**b**) 24, and (**c**) 72.

4.3.2 Enzyme state analysis reveals kinetic bottlenecks that reduce synergy

To understand the trends observed in Figs. 4.1 and 4.2, the fractions of *Tr*Cel7A in different states in both their pure and mixture environments are averaged over selected time intervals. We ask the following question: how do the averaged fractions of *Tr*Cel7A in different states change when *Tr*Cel7B is added to the system? Therefore, we assess the effect of endoglucanase addition by comparing a system containing only *Tr*Cel7A enzymes with a system containing the same number of *Tr*Cel7A but in the presence of endoglucanase. This comparison allows us to understand how the addition of *Tr*Cel7B changes the distribution of *Tr*Cel7A over its different states.

The enzyme states of *Tr*Cel7A are described in more detail in Section 3.2. In short, a *Tr*Cel7A enzyme can be in one of the following states: (1) solution, (2) uncomplexed, (3) blocked by uneven microfibril surface layers (or surface-blocked), (4) blocked by other enzymes, (5) blocked by the microfibril edge, or (6) active. An enzyme in solution can only adsorb onto the surface. An uncomplexed enzyme is adsorbed, and can desorb, diffuse, or complex with a free chain end. Once complexed, an enzyme is active or blocked. An enzyme is active if no obstacles exist immediately in front of its processive path. A blocked enzyme can only decomplex, while an active enzyme can perform hydrolysis or decomplex. In endo-exo mixtures, we find that the largest fractions of inactive states are the uncomplexed and surface-blocked ones. Therefore, in the following, we consider the fractions in these two inactive states, along with the active fraction to investigate the observed trends in Figs. 4.1 and 4.2.

An analysis of the enzyme state fractions is insightful because it allows us to quantify the relative importance of kinetic bottlenecks that reduce active enzyme fractions. The active state fraction can be related to the mixture conversion rate using the first-order rate law for cellobiose production. Therefore, a high active state fraction promotes a high conversion rate. This relation is further discussed in Section 4.5. The active state fraction is reduced when enzyme occupation in inactive states is increased. The inactive states act as kinetic traps that reduce activity and conversion rate. The decomposition of inactive fractions into different states thus allows the most important kinetic traps to be identified. For example, a large fraction of surface-blocked enzymes indicates that processing enzymes often become stalled at surface defect sites and that these sites present kinetic traps on the surface.

Figs. 4.3, 4.4, and 4.5 illustrate the enzyme state fractions in the uncomplexed, surface-blocked, and active states, respectively, averaged from 0 to 12 hours and averaged from 12 to 48 hours. These two time intervals are used to capture the "biphasic" conversion behavior that appears under certain conditions, as illustrated in Figs. 4.1 and 4.2. The first interval captures the burst in the enzyme mixture conversion appearing for $\alpha > 1$. The second interval captures the time over which the mixture conversion slows down relative to the null-synergy rate.

We first focus the discussion within the first time interval. For $\alpha > 1$, the active fraction of *Tr*Cel7A in the mixture within the first interval ($0 - 12$ hr) is greater than that of *Tr*Cel7A in its pure environment, as seen in Fig. 4.5. This trend holds for all mixture compositions investigated and corresponds to the initial burst in mixture conversion observed when $\alpha > 1$. Under these conditions, the addition of *Tr*Cel7B enhances *Tr*Cel7A activity. It does so by not only creating new chain ends but by also enhancing the complexation rate of *Tr*Cel7A with these new chain ends, since $\alpha > 1$. Hence, the fractions of *Tr*Cel7A in the uncomplexed state are significantly reduced, as seen in Fig. 4.3.

However, the addition of *Tr*Cel7B also increases the surface-blocked fractions of *Tr*Cel7A within the first time interval, as seen in Fig. 4.4. This increase is more dramatic at higher mole fractions of *Tr*Cel7B and when $\alpha > 1$. Under these conditions, many chain ends are created and subsequently complexed by *Tr*Cel7A. A higher fraction of complexed *Tr*Cel7A shifts the enzyme distribution toward the active and blocked states.

The increase in *Tr*Cel7A blocking also arises from gradual roughening of the surface when *Tr*Cel7B is added. Because a *Tr*Cel7B creates random interior cuts within the chain, a processing *Tr*Cel7A is more likely to decomplex before reaching the end of the microfibril.

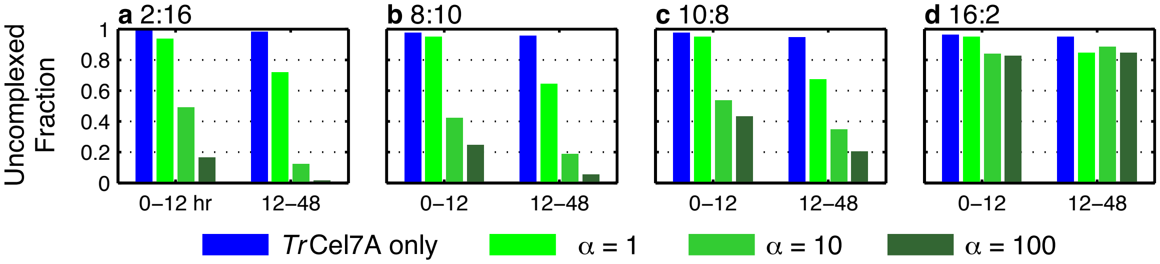


Figure 4.3. Time-averaged uncomplexed state fractions of *Tr*Cel7A at different mixture compositions. The two time intervals over which averaging is performed are 0 – 12 hr and 12 – 48 hr. The enzyme state fractions are computed for enzymes in their pure environments (blue) and mixture environments (green) at different values of α. The mixture compositions (Number of *Tr*Cel7A : Number of *Tr*Cel7B) are: (**a**) 2:16, (**b**) 8:10, (**c**) 10:8, and (**d**) 16:2.

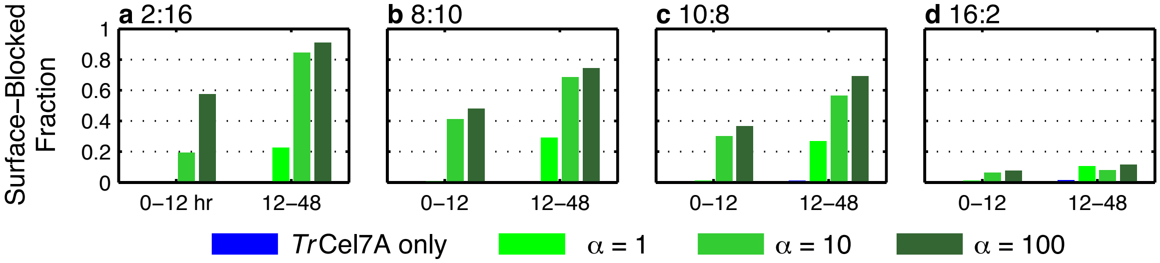


Figure 4.4. Time-averaged surface-blocked state fractions of *Tr*Cel7A at different mixture compositions. The two time intervals over which averaging is performed are 0 – 12 hr and 12 – 48 hr. The enzyme state fractions are computed for enzymes in their pure environments (blue) and mixture environments (green) at different values of α. The mixture compositions (Number of *Tr*Cel7A : Number of *Tr*Cel7B) are: (**a**) 2:16, (**b**) 8:10, (**c**) 10:8, and (**d**) 16:2.

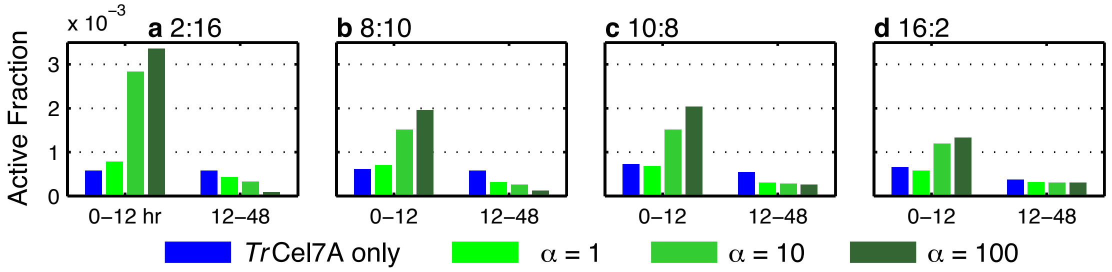


Figure 4.5. Time-averaged active state fractions of *Tr*Cel7A at different mixture compositions. The two time intervals over which averaging is performed are 0 – 12 hr and 12 – 48 hr. The enzyme state fractions are computed for enzymes in their pure environments (blue) and mixture environments (green) at different values of α. The mixture compositions (Number of *Tr*Cel7A : Number of *Tr*Cel7B) are: (**a**) 2:16, (**b**) 8:10, (**c**) 10:8, and (**d**) 16:2.

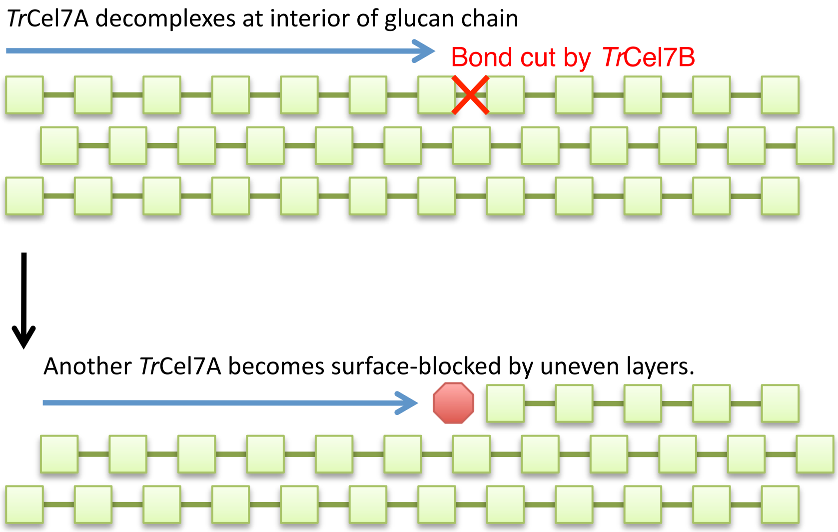


Figure 4.6. Surface roughening mechanism caused by combined actions of exoglucanase *Tr*Cel7A and endoglucanase *Tr*Cel7B. Each green square represents a cellobiose residue. A side view of the microfibril and three layers are shown. First, a *Tr*Cel7A enzyme decomplexes at the interior of a glucan chain once it reaches a bond that has been cut by *Tr*Cel7B. This leaves an uneven layer on the microfibril surface. Next, another *Tr*Cel7A enzyme complexes with a chain in a layer underneath and begins processive hydrolysis. However, it then becomes blocked by the uneven surface layer created previously. Therefore, the combined actions of endo- and exoglucanases can lead to "surface inhibition" of exoglucanases via increased surface roughness of the microfibril.

Each interior decomplexation event leaves an uneven layer (or step-like defect) on the microfibril surface. This rougher surface increases the probability that a processing *Tr*Cel7A encounters an uneven surface layer and becomes surface-blocked. This surface roughening mechanism is illustrated in Fig. 4.6. When a relatively large amount of *Tr*Cel7B is added, more random interior cuts are created, which lead to increased surface roughening and blocking of *Tr*Cel7A enzymes. At high α values, this surface roughening process is accelerated by the increasing numbers of *Tr*Cel7A that complex and then decomplex within the interior of the chain. Both high values of α and high mole fractions of *Tr*Cel7B hence accelerate the surface roughening process and increase "surface-inhibition" of the processing *Tr*Cel7A enzymes. The increased surface blocking of *Tr*Cel7A due to *Tr*Cel7B addition is shown in a simulation snapshot in Fig. 4.7.

The increased surface-blocking due to addition of *Tr*Cel7B explains why synergy is not observed for $\alpha = 1$. The complexation enhancement due to only the increase in the number of chain ends is insufficient to counterbalance the increased blocking of *Tr*Cel7A. Hence, as seen in Fig. 4.5, the active fractions of *Tr*Cel7A in the mixture are not significantly

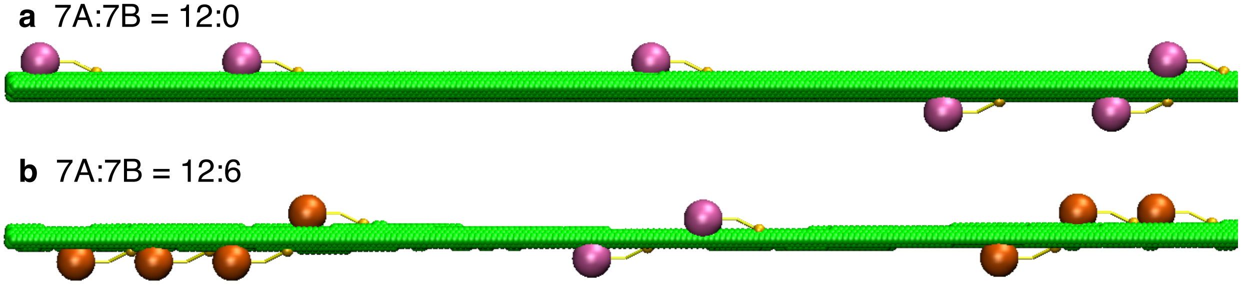


Figure 4.7. SLATE simulation snapshot illustrating the increasing surface roughness and surface blocking of *Tr*Cel7A enzymes in the presence of *Tr*Cel7B. Half of the microfibril is shown, with the reducing end on the left hand side. (**a**) Snapshot from simulation with 12 *Tr*Cel7A enzymes only. (**b**) Snapshot from simulation with 12 *Tr*Cel7A enzymes and 6 *Tr*Cel7B enzymes. For comparison to (**a**), only the *Tr*Cel7A enzymes on the microfibril are shown. The blocked *Tr*Cel7A enzymes are shown as orange. In (**b**), a larger fraction of the microfibril has been converted, due to the complexation enhancement provided by *Tr*Cel7B to *Tr*Cel7A. However, the surface is also increasingly rougher, which begins to inhibit *Tr*Cel7A activity and conversion rate as the reaction proceeds. As shown, a larger fraction of *Tr*Cel7A enzymes are blocked in (**b**). In both (**a**) and (**b**), the snapshot is taken after 24 hours of reaction time.

different from those in its pure environment. When $\alpha = 1$, the benefit of increased chain end number is erased by the costs of increased blocking. Only when *Tr*Cel7B also enhances the complexation rate of *Tr*Cel7A via surface disruption can the increased blocking be overcome to yield synergistic behavior.

As the mixture reaction proceeds, we observe in Figs. 4.4 and 4.5 generally that the activity decreases and blocking increases. In the second time interval (12 – 48 hr), the active fraction is lower relative to its value in the first interval for all values of α and mixture compositions. The drop off between time intervals is enhanced at higher values of α and *Tr*Cel7B mole fractions, because these conditions favor increased surface roughening and surface blocking. The drop off in the mixture active fraction between time intervals corresponds to the drop off in mixture conversion rate relative to the null-synergy rate, as illustrated in Fig. 4.1a-c. This lower activity leads to the appearance of anti-synergy at longer times for sufficiently high *Tr*Cel7B mole fractions, as illustrated in Figs. 4.1a-c and 4.2c. Under these conditions, both the high *Tr*Cel7B mole fraction and long reaction time contribute towards magnifying surface inhibition sufficiently to cause the mixture conversion to drop from the synergistic to anti-synergistic regime.

At high mole fractions of *Tr*Cel7B (Fig. 4.5), the active fraction within the second time interval actually decreases with α. This trend is seemingly counterintuitive, but occurs when complexation of *Tr*Cel7A tends to only lead to a blocked state, without any preceding activity. In this case, enhanced complexation acts as a trap that prevents the enzyme from finding a more productive free end that leads to processive activity. The number of these "complexation traps" is expected to increase with the mole fraction of *Tr*Cel7B, since more free ends could be created that could act as traps. These complexation traps explain why the mixture conversions for high *Tr*Cel7B mole fractions (Figs. 4.1a-c) and $\alpha = 100$ actually drops off below the mixture conversion for $\alpha = 1$. This drop off corresponds to the seemingly counterintuitive decrease in active fraction with increasing α within the second time interval, as observed in Fig. 4.5ab.

At low *Tr*Cel7B mole fractions (Fig. 4.5d), the drop off in activity with time is less dramatic and synergy is maintained at longer times as long as $\alpha > 1$, as seen in Fig. 4.2c. The surface-blocked fractions of *Tr*Cel7A remain at relatively low values, even when $\alpha = 100$. At these lower *Tr*Cel7B mole fractions, fewer chain ends are created, which retards the surface roughening process. As long as $\alpha > 1$, the number of new chain ends is sufficient to maintain synergistic conversion (Figs. 4.1d and 4.2c). Here, an optimal balance is struck between complexation enhancement and increased blocking. As a result, high mixture conversions can be achieved after 72 hours (Fig. 4.2c).

Our finding that a low *Tr*Cel7B mole fraction (< 0.2) is favorable for maximum conversion (Fig. 4.2) is consistent with the hypothesis by Henrissat et al. [72] that the optimal enzyme mixture should contain only a small fraction of endoglucanases, which is sufficient just to enhance complexation by *Tr*Cel7A. We find that adding excessive amounts of *Tr*Cel7B depletes the mixture of the more active *Tr*Cel7A enzymes and causes them to become surface-blocked. The conversion maximum favoring low endoglucanase mole fractions (~ 0.1) is also seen in experimental results [91,141,142].

4.3.3 Developing a mechanistic picture of endo-exo interaction

We can now outline a comprehensive mechanistic picture of the effects of mixture composition, *Tr*Cel7A complexation enhancement, and reaction time scale on the degree of endo-exo synergy.

1. **Mixture composition:** Addition of *Tr*Cel7B creates more free ends for *Tr*Cel7A but also increases surface blocking of *Tr*Cel7A by inducing surface roughening. The former effect dominates at early times, while the latter effect dominates as the reaction proceeds. These competing trends are reflected in Fig. 4.2, where we see a shift in the synergy vs. mole fraction curve over reaction time. As the reaction proceeds, the location of maximum synergy shifts toward lower *Tr*Cel7B mole fractions, where surface roughening is less severe. We are unaware of experiments probing the behavior of the synergy vs. mole fraction curve over time, but our model results presented here show such measurements could be useful for revealing the underlying kinetic behaviors.

2. ***Tr*Cel7A complexation enhancement:** The ability of *Tr*Cel7B to enhance the complexation rate of *Tr*Cel7A with the newly created ends is necessary for synergy to

occur. The rate enhancement is required to overcome the increased surface blocking introduced by *Tr*Cel7B. Increasing α generally enhances the mixture conversion rate. In some cases, however, increasing α lowers the active fraction of *Tr*Cel7A and conversion rate. Under these conditions (high *Tr*Cel7B mole fraction and later reaction times), complexation traps emerge, and increasing α induces *Tr*Cel7A enzymes to enter these inactive traps, lowering the active state fraction (Figs. 4.5ab). Eventually, the conversion from $\alpha = 100$ drops below the conversion from $\alpha = 10$ and even $\alpha = 1$ (Fig. 4.1a-c).

3. **Reaction time:** As the reaction proceeds, synergy gradually decreases due to increased roughening of the surface by the combined actions of *Tr*Cel7A and *Tr*Cel7B via the mechanism illustrated in Fig. 4.6. This rougher surface reduces the processivity of *Tr*Cel7A and increases its tendency to become surface-blocked. This emergent "surface-inhibition" leads to anti-synergy in different ways. When $\alpha = 1$, the surface inhibition dominates at all times, leading to anti-synergy for all mixture compositions (Fig. 4.2). When $\alpha > 1$, anti-synergy appears at longer times and high *Tr*Cel7B mole fractions. These conditions favor surface roughening and increased surface-blocking of *Tr*Cel7A. Hence, a transition from synergistic to anti-synergistic conversion occurs as the reaction proceeds (Fig. 4.2).

The effects of mixture composition, *Tr*Cel7A complexation enhancement, and reaction time scale on the degree of endo-exo synergy can be summarized graphically as a decision tree, shown in Fig. 4.8.

4.3.4 Understanding endo-exo synergy via an endo-exo productivity map

The productivity map of conversion introduced in Section 3.3.4 for cellulose conversion by pure *Tr*Cel7A can be helpful in understanding the effect of *Tr*Cel7B on the activity of *Tr*Cel7A. In Fig. 3.6, the conversion was illustrated to be affected by the complexation and decomplexation rates. The complexation rate represented the rate for complexation between *Tr*Cel7A and a pre-existing chain end. To understand the behavior of *Tr*Cel7A in a mixture with *Tr*Cel7B, we can consider more qualitatively the "complexibility" between *Tr*Cel7A and the surface. This measure is a *net* complexation rate per *Tr*Cel7A enzyme, which is determined by the number of pre-existing chain ends, the number of newly created chain ends, and the complexation rates with each type of end. With this measure in mind, we can represent the effect of *Tr*Cel7B on the location of *Tr*Cel7A within the productivity map. Increasing α and/or increasing reaction time leads to increased complexibility.

Similarly, an increase in surface roughness, and hence obstacle density, is mechanistically analogous to a lowering of the decomplexation rate, since both effects act to increase the blocking of processing *Tr*Cel7A enzymes. Therefore, rather than considering as before the effect of decomplexation rate on conversion, we consider at a more qualitative level the effect

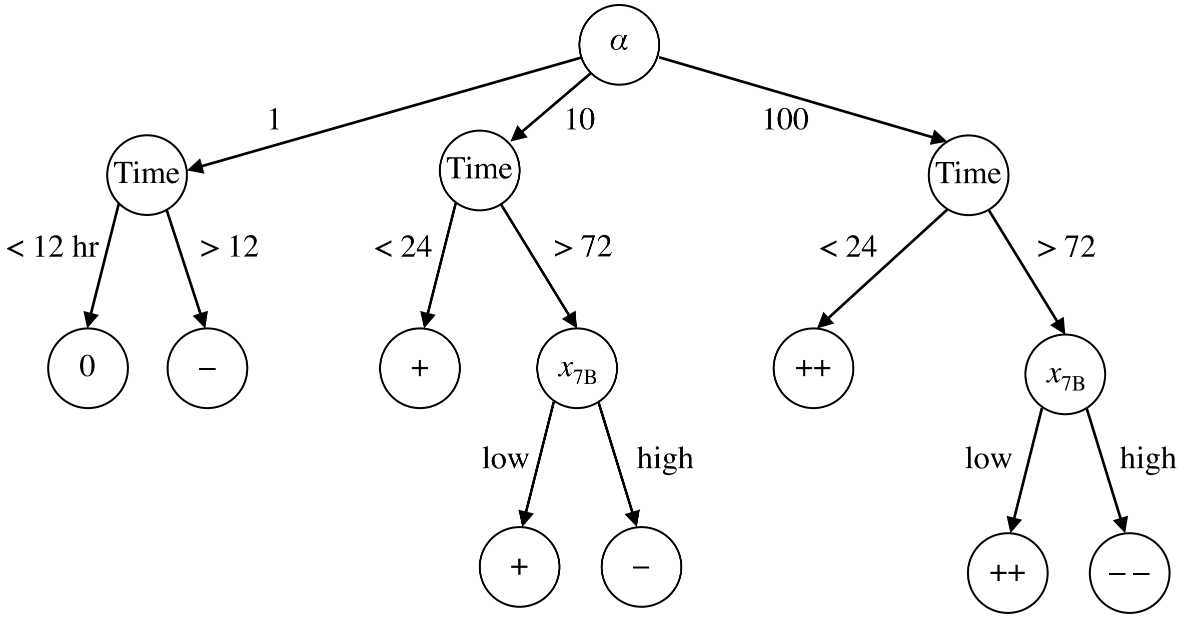


Figure 4.8. Decision tree representation of the effects of mixture composition, *Tr*Cel7A complexation enhancement, and reaction time scale on the degree of endo-exo synergy. The *Tr*Cel7B mole fraction is x_{7B}. Null-synergy is indicated as “0”. Anti-synergy and synergy are indicated as “–” and “+”, respectively, and high anti-synergy and high synergy are indicated as “– –” and “++”, respectively.

of the “blocking propensity” between *Tr*Cel7A and the surface on conversion. This property measures how likely a complexed *Tr*Cel7A would be blocked rather than active. Finally, instead of conversion, we can consider this new endo-exo productivity map to represent the mixture conversion rate instead, as the physical trends are still preserved upon this change. Therefore, we can use the same *Tr*Cel7A-only productivity map of Section 3.3.4 to illustrate the effect of *Tr*Cel7B addition on *Tr*Cel7A activity.

The addition of *Tr*Cel7B can now be represented as a change in the location of *Tr*Cel7A within this endo-exo productivity map. As shown in Fig. 4.9, the effect of *Tr*Cel7B on *Tr*Cel7A is to increase the complexibility and blocking propensity over time. For high α, in the initial stages, the mixture conversion rate is enhanced, as the increase in complexibility dominates over the increase in blocking. In the later stages, however, although the complexibility continues to increase, the high blocking propensity of the surface begins to dominate and cause a drop off in the conversion rate. This drop off in mixture conversion rate relative to the null-synergy rate due to surface roughening and inhibition is seen in Fig. 4.1. For low α, the complexibility only increases slightly with time and the increase in conversion rate is hence marginal. Eventually, it begins to decrease due to increased blocking induced by *Tr*Cel7B. The mixture conversion rate hence remains near or below its initial value, and the

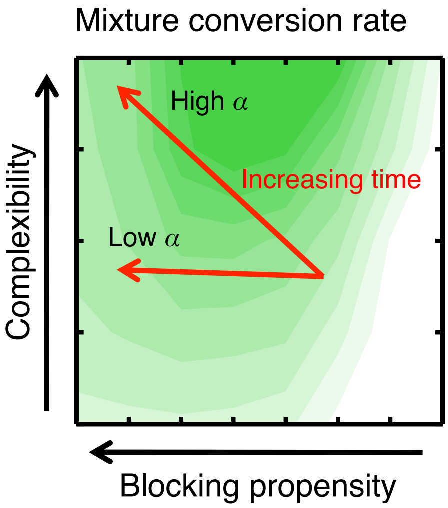


Figure 4.9. Endo-exo productivity map illustrating the qualitative trends of complexibility and blocking propensity on mixture conversion rate. The red arrows indicate the changes in mixture conversion rate over time at both high and low α values. Here, Fig. 3.6f from Section 3.3.4 has been adopted to illustrate these qualitative trends. Therefore, the influence of *Tr*Cel7B on *Tr*Cel7A activity can be effectively represented as a modulation of the complexation and decomplexation rates of *Tr*Cel7A.

endo-exo synergy is absent, as illustrated for low α in Fig. 4.2. Therefore, we have illustrated here that many of the mechanistic trends observed in the endo-exo mixture can be explained using the principles developed earlier in Chapter 3 for pure *Tr*Cel7A kinetics.

4.3.5 Physical significance of $\alpha > 1$

The requirement for complexation enhancement by *Tr*Cel7B to produce synergy arises from the increased surface blocking of *Tr*Cel7A induced by *Tr*Cel7B. In particular, when *Tr*Cel7B only creates free chain ends without complexation enhancement of *Tr*Cel7A, the surface blocking of *Tr*Cel7A dominates over the increase in chain end number. In this case, the mixture conversion is anti-synergistic. When *Tr*Cel7B, however, can enhance the *Tr*Cel7A complexation rate, the increased surface blocking caused by *Tr*Cel7B can be overcome to produce an endo-exo interaction that is synergistic.

Figs. 4.1 and 4.2 illustrate that for endo-exo synergy to occur, a mere increase in the number of free chain ends produced by *Tr*Cel7B is insufficient. Rather, the complexation rates of *Tr*Cel7A with these newly created chain ends must be enhanced over those with pre-existing chain ends. Physically, this enhancement arises if *Tr*Cel7B is able to disrupt the surface and loosen hydrogen bonds between glucan chains to make them more susceptible toward complexation by *Tr*Cel7A. Therefore, these simulations provide evidence for an ”amorphogenic” character of endoglucanases that not only cuts glucan chains, but weakens the local crystalline cellulose environment as well [163]. The fiber-disrupting behavior of cellulases has been suggested in experimental studies [92,161–167]. This work represents one of the first computational studies assessing the effect of surface disruption on cellulose conversion.

Kinetic models that do not account for surface blocking of *Tr*Cel7A have observed synergy when $\alpha = 1$ [61,95]. In these models, however, the ability of *Tr*Cel7B to promote surface roughening and surface blocking of *Tr*Cel7A is lacking. Hence, these models fail to uncover that a complexation rate enhancement by *Tr*Cel7B is required to produce synergy. Including spatial constraints in the kinetic model, as done in SLATE, not only uncovers this requirement, but also provides evidence for the amorphogenic character of endoglucanases. Therefore, we show that the addition of surface disrupting components to facilitate *Tr*Cel7A complexation is a viable strategy for accelerating cellulose conversion because addition of these components aims to enhance the complexation process, which is rate-limiting, as discussed in Section 3.3.1.

4.3.6 *In silico* protein engineering to enhance endo-exo synergy

Given an enhanced mechanistic understanding of endo-exo interaction, we perform “protein engineering” experiments on the computer to assess how changes in the kinetic rate constants of *Tr*Cel7A affect enzyme synergy and the mixture conversion. We make two predictions based on the mechanistic picture described above and confirm them via the SLATE simulations.

First, we predict that synergy will be increased if the intrinsic processivity of *Tr*Cel7A is decreased. The intrinsic processivity is a measure of the processive length in the absence of any surface obstacles. It is inversely proportional to the decomplexation rate, as explained further in Section 3.6.4. We expect an increase in synergy to occur from increasing the *Tr*Cel7A decomplexation rate because the enzyme becomes less sensitive to blocking by uneven surface layers created when *Tr*Cel7B is added. At a higher decomplexation rate, the conversion is more limited by low processivity than excessive blocking, and a processing enzyme is less likely to encounter an obstacle. If the enzyme does encounter an obstacle, the rate of escape is high due to its enhanced decomplexation rate. Therefore, the activity of the less processive enzyme will be less inhibited by the rougher surface, while it will still retain the benefits of enhanced complexation rate.

In Fig. 4.10, we plot the conversion and synergy at different mixture compositions and reaction times for a system in which the *Tr*Cel7A decomplexation rate is increased to $1.0 \times 10^{-1}\,\mathrm{s}^{-1}$. In these simulations, α is equal to 10. Compared to the cases where the decomplexation rate is $1.0 \times 10^{-3}\,\mathrm{s}^{-1}$ (Fig. 4.2), the degrees of synergy are significantly higher at all mixture compositions and reaction times. This high synergy occurs because the low processivity of *Tr*Cel7A reduces the sensitivity of the enzyme toward the blocking promoted by a rougher surface.

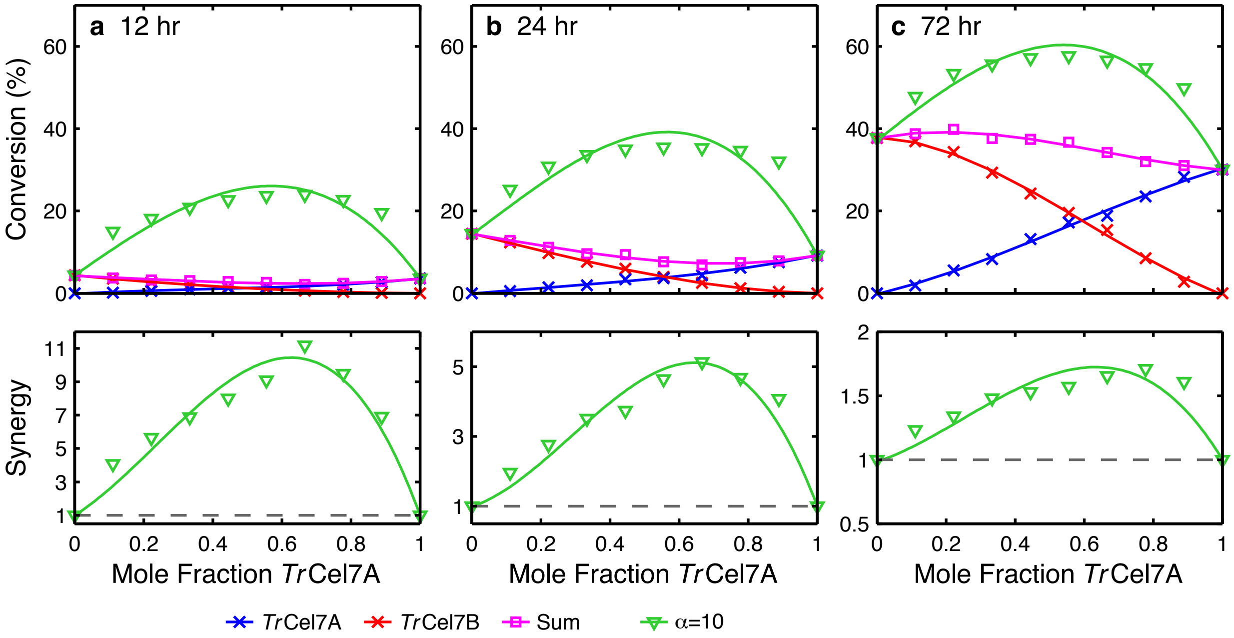


Figure 4.10. Conversion and synergy vs. *Tr*Cel7A mole fraction for mixtures of "mutated" *Tr*Cel7A and *Tr*Cel7B enzymes, where the *Tr*Cel7A decomplexation rate is $1.0 \times 10^{-1}\,\mathrm{s}^{-1}$. The blue crosses represent conversion from *Tr*Cel7A only. The red crosses represent conversion from *Tr*Cel7B only. The magenta squares represent the sum of conversions from *Tr*Cel7A only and *Tr*Cel7B only and is the null-synergy conversion. The green inverted triangles indicate the actual mixture conversions and synergy at $\alpha = 10$. Synergy is defined as the mixture conversion divided by the sum of the individual conversions. The reaction times (hr) are (**a**) 12, (**b**) 24, and (**c**) 72.

Conversely, synergy should be decreased if the intrinsic processivity of *Tr*Cel7A is increased. This is achieved here by decreasing the decomplexation rate to $1.0 \times 10^{-5}\,\mathrm{s}^{-1}$. In this case, the enzyme becomes more sensitive to surface obstacles because escape from these obstacles requires significantly more time. The effect of *Tr*Cel7B to increase surface roughness and blocking of *Tr*Cel7A dominates over the complexation enhancement it provides for *Tr*Cel7A. Hence, the mixture conversion is more anti-synergistic relative to the behavior at higher decomplexation rates, as shown in Fig 4.11.

One could ask the question: given a value of $\alpha = 10$, what is the optimal *Tr*Cel7A decomplexation rate and enzyme mixture composition that yields the highest conversion after 72 hours? As shown in Figs. 4.2, 4.10, and 4.11, the highest conversions are between 50 – 60%. This level can be achieved at high *Tr*Cel7A mole fractions when the decomplexation rate is $1.0 \times 10^{-3}\,\mathrm{s}^{-1}$. It can also be achieved for a broader range of *Tr*Cel7A mole fractions when the decomplexation rate is $1.0 \times 10^{-1}\,\mathrm{s}^{-1}$. At this higher decomplexation rate, a larger *Tr*Cel7B mole fraction can be afforded because sensitivity to surface blocking is minimized.

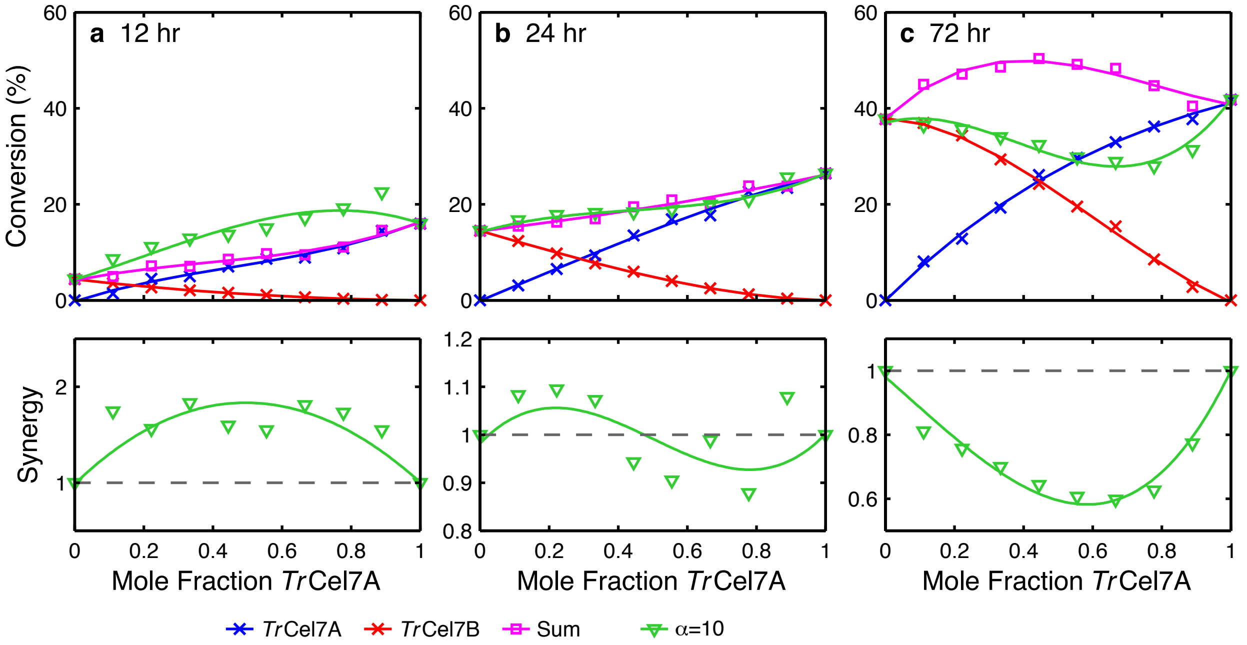
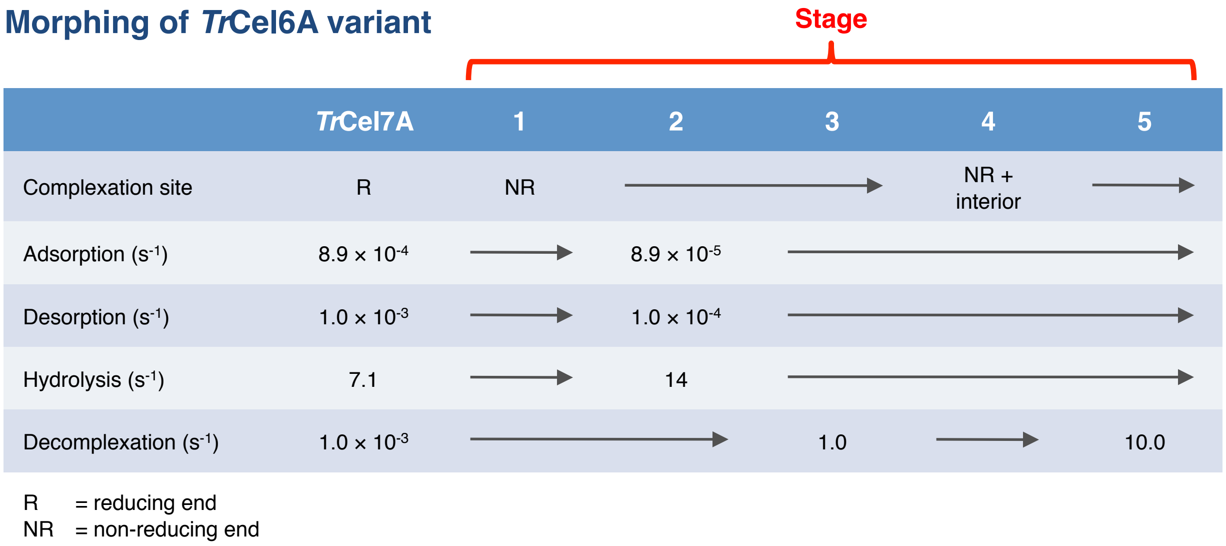


Figure 4.11. Conversion and synergy vs. *Tr*Cel7A mole fraction for mixtures of "mutated" *Tr*Cel7A and *Tr*Cel7B enzymes, where the *Tr*Cel7A decomplexation rate is $1.0 \times 10^{-5}\,\mathrm{s}^{-1}$. The blue crosses represent conversion from *Tr*Cel7A only. The red crosses represent conversion from *Tr*Cel7B only. The magenta squares represent the sum of conversions from *Tr*Cel7A only and *Tr*Cel7B only and is the null-synergy conversion. The green inverted triangles indicate the actual mixture conversions and synergy at $\alpha = 10$. Synergy is defined as the mixture conversion divided by the sum of the individual conversions. The reaction times (hr) are (**a**) 12, (**b**) 24, and (**c**) 72.

Therefore, for the enzyme mixtures investigated here, using a higher decomplexation rate (or less processive exoglucanase) allows for a wider concentration range of endoglucanases to be used in the mixture to provide conversions near 50 – 60%. Engineering the exoglucanase to become less processive in the endo-exo mixture overcomes its sensitivity to blocking and strengthens its synergistic interaction with endoglucanases over a broader range of mixture compositions.

4.3.7 Partial endo-activity of *Tr*Cel6A is essential for exo-exo synergy

In this section, synergy between *Tr*Cel7A and *Tr*Cel6A is investigated. To tease out the kinetic origins for synergy, we gradually "morph" the *Tr*Cel6A enzyme from *Tr*Cel7A-like to *Tr*Cel7B-like behavior. We show that synergy only emerges in our model if *Tr*Cel6A possesses sufficient endo-like behavior, as anticipated previously [67,69,132]. The morphing



Morphing of *Tr*Cel6A variant

Stage

	*Tr*Cel7A	1	2	3	4	5
Complexation site	R	NR	→	→	NR + interior	→
Adsorption (s^{-1})	8.9×10^{-4}	→	8.9×10^{-5}	→	→	→
Desorption (s^{-1})	1.0×10^{-3}	→	1.0×10^{-4}	→	→	→
Hydrolysis (s^{-1})	7.1	→	14	→	→	→
Decomplexation (s^{-1})	1.0×10^{-3}	→	→	1.0	→	10.0

R = reducing end
NR = non-reducing end

Figure 4.12. Morphing of *Tr*Cel6A in different stages from *Tr*Cel7A to endo-like behavior. At each stage, a variant *Tr*Cel6A enzyme is simulated along with *Tr*Cel7A to determine the degree of synergy. In Stage 1, only the complexation site for *Tr*Cel6A is changed from the "reducing" to "non-reducing" end. In Stage 2, the adsorption, desorption, and hydrolysis rates are changed. In Stage 3, the decomplexation rate is increased to reduce the *Tr*Cel6A processivity. In Stage 4, the *Tr*Cel6A variant is allowed to complex with and begin hydrolysis from interior portions of a glucan chain. In Stage 5, the decomplexation rate is further increased.

process is shown in Fig. 4.12. In Stage 1, the only difference between the two enzyme types is that *Tr*Cel6A complexes with the non-reducing end of a glucan chain and processes in the opposite direction of *Tr*Cel7A. Here, the kinetic rate constants for *Tr*Cel6A are the same as those of *Tr*Cel7A, listed in Table 4.1. In Stage 2, the adsorption, desorption, and hydrolysis rates for *Tr*Cel6A enzymes are changed to their reference values, listed in Table 4.1, while the decomplexation rate is unchanged. In Stage 3, the decomplexation rate for *Tr*Cel6A is increased to its reference value of $1.0\,\mathrm{s}^{-1}$. In Stage 4, the *Tr*Cel6A enzymes are allowed to complex and begin hydrolysis at interior portions of glucan chains, acting as an endo-processive enzyme. We set $\beta = 1$ to allow for this behavior, where the complexation rate between *Tr*Cel6A and interior portions of glucan chains is $\beta \times k_{c,6\mathrm{A}}$. An internal cut by *Tr*Cel6A creates a new free end that can be complexed by *Tr*Cel7A, with rate $\alpha \times k_{c,7\mathrm{A}}$. We set $\alpha = 10$ to model the increased susceptibility of the free ends created by *Tr*Cel6A. In Stage 5, we further increase the decomplexation rate $10.0\,\mathrm{s}^{-1}$, which reduces *Tr*Cel6A processivity and allows it to act more as a non-processive *Tr*Cel7B enzyme. Hence, in Stages 4 and 5, the *Tr*Cel6A enzyme adopts more endo-like behavior.

At each stage, we perform simulations with enzyme mixtures of *Tr*Cel7A and the *Tr*Cel6A

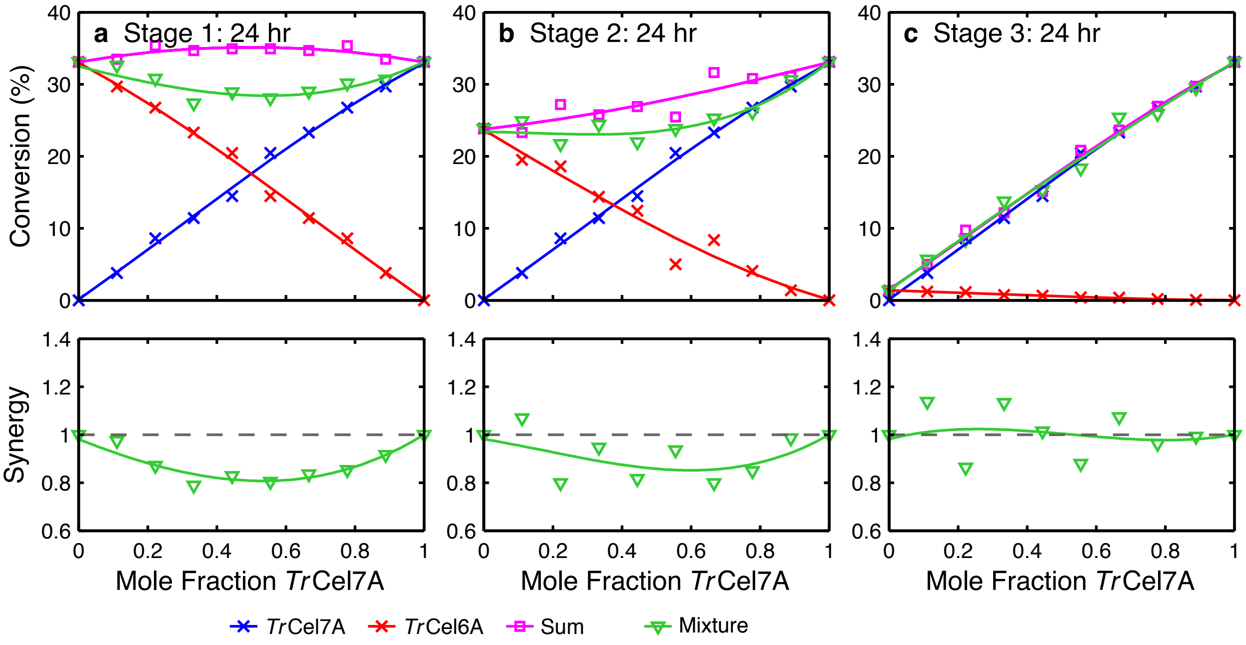


Figure 4.13. Conversion and synergy vs. *Tr*Cel7A mole fraction for mixtures of *Tr*Cel7A and *Tr*Cel6A enzymes for Stages 1, 2, and 3. The blue crosses represent conversion from *Tr*Cel7A only. The red crosses represent conversion from *Tr*Cel6A only. The magenta squares represent the sum of conversions from *Tr*Cel7A only and *Tr*Cel6A only and is the null-synergy conversion. The green inverted triangles indicate the actual mixture conversions and synergy at $\alpha = 10$. Synergy is defined as the mixture conversion divided by the sum of the individual conversions. The reaction time is 24 hours. The stage variants of *Tr*Cel6A are (**a**) 1, (**b**) 2, and (**c**) 3.

stage variant and calculate synergy as a function of mixture composition. In the simulations performed at each stage, the total enzyme number in the mixture is 18. The numbers of *Tr*Cel7A and *Tr*Cel6A enzymes are varied between 0 and 18, while keeping the total enzyme number constant.

Stages 1–3

Fig. 4.13 plots the conversion and synergy as a function of *Tr*Cel7A mole fraction for mixtures of *Tr*Cel7A and different variants of *Tr*Cel6A from Stages 1, 2, and 3. The Stage 1 *Tr*Cel6A enzyme is completely identical to the *Tr*Cel7A enzyme, except it complexes with the non-reducing end and processes in the opposite direction. Therefore, to determine the conversion by the pure *Tr*Cel6A in Stage 1, we simply use the conversion from pure *Tr*Cel7A, due to the symmetry of the microfibril and enzyme behaviors. The Stage 1 mixture can be constructed by taking a selected number of enzymes from a pure *Tr*Cel7A system with 18 enzymes and having them complex and process from the opposite end of the microfibril. Therefore, comparing conversions from the pure *Tr*Cel7A system with the Stage 1 mixture reveals the effect of changing the chain end specificity and processive direction of a selected number of enzymes in the mixture.

We see in Fig. 4.13a that the 24-hr mixture conversion is slightly lower than that from using 18 *Tr*Cel7A enzymes and that the exo-exo interaction is anti-synergistic. This behavior also holds at earlier (6 hrs) and later times (72 hrs). The lower conversion arises because a *Tr*Cel7A enzyme blocked at the microfibril edge prevents other *Tr*Cel6A enzymes from complexing with free ends at that edge, and vice versa. Although the presence of enzymes processing in opposite directions could promote further blocking, this limitation would only occur under high complexation rate conditions. For the low complexation rates used here (See Table 4.1), the existence of more than one enzyme processing along a microfibril face is extremely rare.

In Stage 2, the adsorption, desorption, and hydrolysis rate constants of *Tr*Cel6A are changed to their reference values listed in Table 4.1. We first examine the effect of changing these rates on the activity of *Tr*Cel6A alone. Changing the hydrolysis rate from $7.1\,\mathrm{s}^{-1}$ to $14\,\mathrm{s}^{-1}$ does not affect conversion, since the hydrolysis rates are already high. The decrease in conversion relative to Stage 1, as illustrated in Fig. 4.13, arises from the less reversible binding of *Tr*Cel6A. Since diffusion along the microfibril is single-file, enzymes cannot pass over one another by diffusion and may require desorption and re-adsorption to find a complexation site. Lowering the desorption rate would hence retard this process. As seen in Fig. 4.13b, using the Stage 2 *Tr*Cel6A enzyme in a mixture only gives rise to anti-synergy, much like the case for Stage 1.

In Stage 3, the decomplexation rate of *Tr*Cel6A is changed to its reference value of $1.0\,\mathrm{s}^{-1}$, listed in Table 4.1. At this higher decomplexation rate, the conversion from pure *Tr*Cel6A is nearly zero due to the low processivity of the enzyme, shown in Fig. 4.13c. Since *Tr*Cel6A has such low activity in this case, the resulting mixture conversion vs. *Tr*Cel7A mole fraction curve acts as if the *Tr*Cel6A is absent, and synergy does not emerge.

Physically, the high decomplexation rate emerges from the flexible tunnel-forming loops of *Tr*Cel6A that can occasionally open to form a cleft-like topology [41,59,68], resembling more the active site geometries of endoglucanases. Therefore, the ability of *Tr*Cel6A to complex with interior portions of glucan chains should be incorporated upon increasing its decomplexation rate. This ability is adopted in Stage 4. The rate of *Tr*Cel6A complexation with the interior regions is $\beta \times k_{c,6\mathrm{A}}$, where we set $\beta = 1$. In Fig. 4.14, we show the dependence of mixture conversion and synergy on mixture composition at 6 and 24 hours.

We see that by allowing *Tr*Cel6A to create interior cuts, synergy with *Tr*Cel7A is achieved at early times when the *Tr*Cel6A mole fraction is low. Under these conditions, the complexation enhancement afforded to *Tr*Cel7A outweighs the increase in surface-blocking of *Tr*Cel7A. When these conditions are not met, however, synergy disappears due to the increasing surface inhibition of *Tr*Cel7A, as discussed in Section 4.3.2. As shown in Fig. 4.14b, however, the absence of synergy does not imply the lack of an optimal mixture composition that gives maximum conversion. For this system, the optimum *Tr*Cel7A mole fraction is

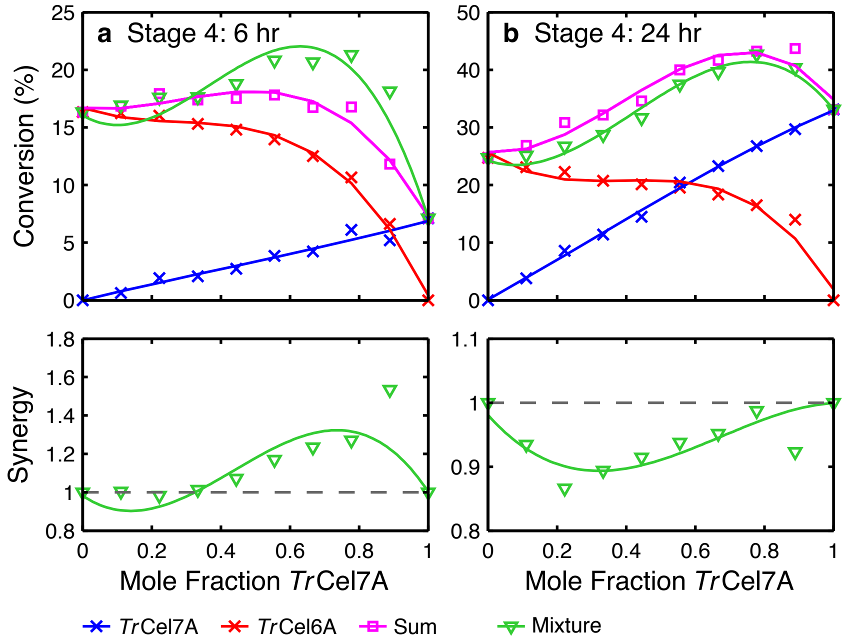


Figure 4.14. Conversion and synergy vs. *Tr*Cel7A mole fraction for mixtures of *Tr*Cel7A and *Tr*Cel6A enzymes for Stage 4. The blue crosses represent conversion from *Tr*Cel7A only. The red crosses represent conversion from *Tr*Cel6A only. The magenta squares represent the sum of conversions from *Tr*Cel7A only and *Tr*Cel6A only and is the null-synergy conversion. The green inverted triangles indicate the actual mixture conversions and synergy at $\alpha = 10$. Synergy is defined as the mixture conversion divided by the sum of the individual conversions. The reaction times (hr) are (**a**) 6 and (**b**) 24.

around 0.8.

In Stage 5, we further "engineer" the *Tr*Cel6A to behave more like *Tr*Cel7B by increasing the *Tr*Cel6A decomplexation rate to $10.0\,\mathrm{s}^{-1}$. At this higher decomplexation rate, the enzyme on average will perform about one cut before decomplexing, acting as if it were a non-processive *Tr*Cel7B. Comparing Fig. 4.15 with Fig. 4.14, we see that lowering the *Tr*Cel6A processivity further enhances the synergy. This synergy enhancement, however, largely arises due to the decrease in conversion from pure *Tr*Cel6A, as the mixture conversion also generally decreases after lowering the processivity. Therefore, the existence of high synergy does not always indicate high mixture conversion, because its calculation also considers the base line conversions from the pure components.

Comparing the 24-hr mixture conversions from all 5 stages illustrated in Figs. 4.13–4.15, we see that the highest conversion exists in Stage 4. Relative to the other stages, the Stage 4 *Tr*Cel6A enzyme plays a multitude of roles to accelerate conversion. First, unlike the Stage 1–3 enzymes, it is able to complex with interior portions of glucan chains. Second, unlike the Stage 5 enzyme, it can perform processive hydrolysis. Finally, since $\alpha = 10$, it enhances the

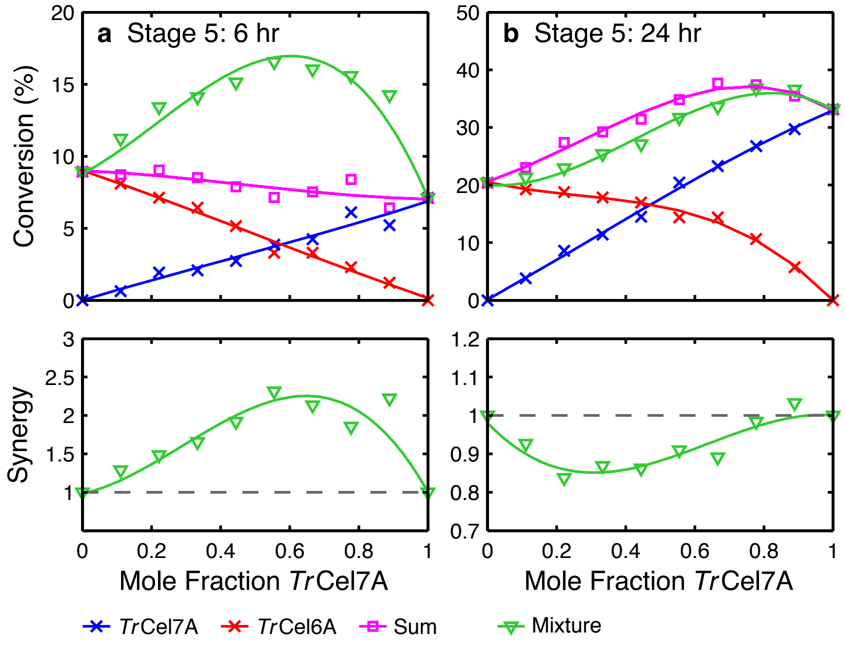


Figure 4.15. Conversion and synergy vs. *Tr*Cel7A mole fraction for mixtures of *Tr*Cel7A and *Tr*Cel6A enzymes for Stage 5. The blue crosses represent conversion from *Tr*Cel7A only. The red crosses represent conversion from *Tr*Cel6A only. The magenta squares represent the sum of conversions from *Tr*Cel7A only and *Tr*Cel6A only and is the null-synergy conversion. The green inverted triangles indicate the actual mixture conversions and synergy at $\alpha = 10$. Synergy is defined as the mixture conversion divided by the sum of the individual conversions. The reaction times (hr) are (**a**) 6 and (**b**) 24.

complexation of *Tr*Cel7A by creating free chain ends and locally disrupting the crystalline structure. This latter effect represents the cooperative interaction between *Tr*Cel6A and *Tr*Cel7A. Notably, the mixture behavior in Stage 4 is actually anti-synergistic over a wide range of *Tr*Cel7A mole fractions. However, due to the leveling-off, or concave shape of the pure *Tr*Cel6A conversion curve, an optimal mixture composition exists even though the mixture conversion is anti-synergistic. This example highlights that the traditional definition of synergy can in some cases be misleading and fail to capture underlying cooperative enzyme interactions that give rise to an optimal mixture composition for highest conversion.

4.4 Discussion

We have developed a kinetic Monte Carlo model to simulate the reaction kinetics of individual cellulases on a cellulose microfibril. This kinetic model is the first model to incorporate spatial constraints, complexation-exchange kinetics, and enzyme processivity into the reaction network. Using this Stochastic LATtice Enzyme (SLATE) model, we have elucidated the molecular-scale kinetic mechanisms that govern enzyme synergy. We delineate the conditions under which endo-exo interactions are not only cooperative, but also anti-cooperative. We find in general that addition of *Tr*Cel7B to the reaction system introduces two effects: (1) complexation enhancement for *Tr*Cel7A and (2) increased surface roughness and subsequent blocking of *Tr*Cel7A. The relative magnitudes of these two factors depend on the mixture composition, *Tr*Cel7A complexation enhancement, and reaction time scale. The balance between complexation enhancement and surface blocking determines if the enzyme mixture acts synergistically. Remarkably, the general theme that complexation rate and blocking control reaction rate, emergent from our investigations of pure *Tr*Cel7A exoglucanase systems, also holds for enzyme mixtures.

By explicitly incorporating spatial constraints on the enzyme mixture behavior, we extend and refine current mechanistic theories for cellulase synergy. The most widely accepted explanation for endo-exo synergy is that the creation of more free chain ends by endoglucanases enhances the complexation of exoglucanases [135]. Our simulations, however, show that an increase in only the number of free chain ends is insufficient for synergy to occur. In particular, when *Tr*Cel7B only creates free chain ends without enhancing the *Tr*Cel7A complexation rate with each of these endo-created chain ends, the mixture conversion is anti-synergistic. This occurs because the combined actions of *Tr*Cel7B and *Tr*Cel7A increase the roughness of the microfibril surface, which in turns leads to surface blocking of processing *Tr*Cel7A enzymes. Without complexation rate enhancement, the surface blocking of *Tr*Cel7A dominates over the increase in chain end number. When *Tr*Cel7B, however, can enhance the *Tr*Cel7A complexation rate with the endo-created chain ends, the increased surface blocking of *Tr*Cel7A can be overcome to produce a synergistic endo-exo interaction.

The major outcomes of this work are the identification of anti-synergistic factors that impede cellulose decomposition by cellulase mixtures and the requirement for complexation rate enhancement for synergistic conversion. These findings are uncovered because our SLATE model explicitly accounts for the complexation step and resolves the reactant configurations in three dimensions. A previous spatial kinetic model found that a complexation rate enhancement was not required to give rise to synergy [95]. However, in this model, the complexation step was not resolved explicitly, but lumped with adsorption and diffusion. In addition, the model was two-dimensional, and as such, surface-blocking of processing enzymes was not captured. Incorporation of an explicit slow complexation step and spatial constraints on processivity would likely eliminate the observed synergy in this model and require complexation rate enhancement to recover synergistic conversions. Classical mass-action models [60,61,131,147,150], which assume a dilute, well-mixed environment, also lack the capability to uncover the underlying anti-synergistic interactions that arise from spatial constraints on the reaction rate. In one study [61], the authors observed anti-synergy in their model under low surface area conditions and suggested it was due to competitive adsorption between endo- and exoglucanases, which decreased the endoglucanase ability to create new free ends. Although such an explanation for anti-synergy cannot be ruled out, the effect of complexation rate enhancement on recovering enzyme synergy was not explored. In addition, surface-blocking of processive exoglucanases was also not captured, which would decrease the degree of synergy.

In addition to resolving the effects of spatial confinement, our work also represents one of the first computational studies assessing the effect of surface disruption and substrate morphology on cellulose conversion. Physically, the complexation rate of *Tr*Cel7A with endo-created chain ends is enhanced if *Tr*Cel7B is able to disrupt and loosen hydrogen bonds between glucan chains on the surface, which makes them more susceptible toward complexation by *Tr*Cel7A. Therefore, our SLATE simulations provide evidence for an "amorphogenic" character of endoglucanases, which refers to their ability to not only create internal glucan chain cuts, but also weaken the local crystallinity and hydrogen bond structure of the surface as well [163]. The fiber-disrupting behavior of cellulases has been suggested in experimental studies [92,161–167]. Using these insights from the SLATE simulations, we show that the addition of surface disrupting components to facilitate *Tr*Cel7A complexation is a viable strategy for accelerating cellulose conversion. Addition of these "amorphogenic" components is advantageous because it enhances the rate of the slow complexation process.

The surface blocking induced by *Tr*Cel7B in our simulations could explain two general experimental trends regarding cellulase synergy. First, the optimal enzyme mixtures in many studies contain only a small fraction of endoglucanase [72,91,141,142]. In our simulations, we find that using high *Tr*Cel7B mole reactions is unfavorable because it leads to a rough surface that blocks the processive hydrolysis by *Tr*Cel7A. Surface-blocking of *Tr*Cel7A can only be avoided by keeping the amount of *Tr*Cel7B low (Fig. 4.4). Second, the decrease in synergy with reaction time (Fig. 4.2), as observed experimentally [65,80,86,91,137,141–143] could be due to increased surface-blocking. Previously, this decrease has been attributed to loss of reactive sites for endoglucanases over time as the cellulase surface becomes more recalcitrant [141,142]. We show an additional mechanism could be at play. The decrease in synergy could originate from a growing fraction of *Tr*Cel7A that remain blocked at uneven surface layers, illustrated in Fig. 4.7. The prevalent of surface-blocking, found here, which is an example of spatial confinement, also emphasizes the importance of spatial resolution in kinetic models to describe interfacial biocatalysis.

The increased surface blocking of *Tr*Cel7A introduced by *Tr*Cel7B is seemingly at odds with an auxiliary "obstacle-based" theory for endo-exo synergy [86,96,137,138]. In this theory, endoglucanases preferentially attack amorphous regions of cellulose, which can act as surface obstacles for processing exoglucanase enzymes. In this way, endoglucanases act to smoothen the surface rather than to roughen it. Processing exoglucanases therefore experience a decrease in surface blocking rather than an increase. In one experimental study [96], the processive length of an exoglucanase was found to increase in the presence of an endoglucanase, providing support for the reduced blocking of exoglucanases.

However, the assumption of endoglucanase specificity for amorphous cellulose regions must be treated with care. First, the observed high rate of sugar release from endoglucanase action on amorphous substrates does not mean that endoglucanases cannot adsorb or complex with more crystalline regions on cellulose. The higher rate of sugar release could be due in part to the fact that short isolated glucan chains on the surface are more likely to be released faster from an amorphous surface. These isolated chains could still be created by endoglucanases on crystalline surfaces but remain attached and experimentally undetected . Further, both exo- and endoglucanases from *Trichoderma reesei* contain family 1 carbohydrate binding domains, which display affinity for crystalline cellulose [168]. Therefore, endoglucanases likely do not solely adsorb onto amorphous cellulose regions. Clearly, a key outstanding question is the degree of specificity for endoglucanase activity on amorphous cellulose, which would further elucidate the validity of the auxiliary "obstacle-based" theory for endo-exo synergy.

For exo-exo synergy, our simulations show that a necessary requirement is an endo-like behavior for one of the exoglucanase enzymes [67,69,132]. We show that synergy does not emerge solely from having two types of enzymes with opposing chain end specificities. Therefore, our results do not support the mechanism by which synergy emerges from each type of enzyme clearing out isolated surface chains left by the other type [82]. Effectively, we find that exo-exo synergy essentially follows the mechanism of endo-exo synergy.

Endo-endo synergy [141,169–171] could also be explained in light of our findings. It could emerge if at least one of the endoglucanases displays some exo-like behavior, namely, that it can occasionally prefer to complex with free chain ends. For example, a mixture with a pure endoglucanase and a partial exo-like endoglucanase could display synergy in the following way. The pure endoglucanase can cut a bond in the interior glucan chain, thus creating two new free ends. The exo-like endoglucanase can now complex with interior glucan chain regions or with one of the new free ends. Therefore, synergy would manifest itself in this endo-endo mixture effectively as an endo-exo mechanism.

The possibility for shared mechanisms for endo-exo, exo-exo, and endo-exo synergy raises the question if such strict traditional classifications of cellulase synergy are useful for describing the underlying kinetic behaviors. In general, the division of cellulases into either endo- or exoglucanases is likely too severe, as each enzyme likely falls within a continuum of behaviors ranging from pure endo- to pure exoglucanase activity [46,67]. Further understanding of synergy likely requires consideration of the dual behaviors that some cellulases may exhibit and replacing the traditional synergy classifications with those more descriptive of the kinetic interactions among enzymes.

In this work, we show via spatially-resolved kinetic simulations that a more complete understanding of cellulase synergy requires consideration of both the synergistic and anti-synergistic interactions among enzymes. The balance between complexation rate enhancement and increased surface blocking of *Tr*Cel7A determines the degree of synergy. These interactions, as we have shown, are largely surface-mediated and require spatial resolution of the reaction system at the molecular scale in order to identify. Understanding the interplay between these factors enables a more rational approach toward designing optimal cellulase mixtures for accelerated cellulose conversion.

4.5 Supplementary information

4.5.1 Relation between active state fraction, conversion rate, and synergy

We will show here the relation between the fraction of active enzymes and the mixture conversion rate for an endo-exo mixture of *Tr*Cel7A and *Tr*Cel7B. This follows from the first-order rate law for conversion,

$$\frac{dN_{CB}}{dt} = k_{h,7A}x_{7A}N_{7A} + k_{h,7B}x_{7B,1}N_{7B} + 2k_{h,7B}x_{7B,2}N_{7B}. \tag{4.1}$$

In Eq. 4.1, N_{CB} is the number of cellobiose monomers produced, $k_{h,i}$ is the hydrolysis rate constant, x_i is the active state fraction, and N_i is the total number of enzymes of type i, where i is either 7A and 7B. Two active state fractions exist for *Tr*Cel7B because an endo-cut can create one or two cellobiose residues on the surface, which are assumed to be released quickly into solution.

In the analysis of enzyme states, we have focused only on the effect of *Tr*Cel7B addition on *Tr*Cel7A, but not the reverse. However, we find that the effect of *Tr*Cel7A on *Tr*Cel7B only causes a minimal change in the *Tr*Cel7B state fractions, since these endoglucanases are not processive and their activity is less sensitive to the presence of surface obstacles that appear once *Tr*Cel7A is added. Therefore, an analysis of only the changes in the *Tr*Cel7A state fractions between its pure and mixture environments is sufficient for analyzing synergy.

If both *Tr*Cel7A and *Tr*Cel7B were acting independently (null-synergy case), the expected conversion would be,

$$\frac{dN_{CB}^0}{dt} = k_{h,7A}x_{7A}^0N_{7A} + k_{h,7B}x_{7B,1}^0N_{7B} + 2k_{h,7B}x_{7B,2}^0N_{7B}. \tag{4.2}$$

N_{CB}^0 denotes the expected cellobiose number if null-synergy were to occur, and x_i^0 denotes the active fractions resulting from enzymes acting in their pure environments.

Since the quantities $x_{7B,1} - x_{7B,1}^0$ and $x_{7B,2} - x_{7B,2}^0$ are relatively low, the actual mixture rate is likely to be higher than the summed rate if $x_{7A} - x_{7A}^0$. Therefore, the quantity $x_{7A} - x_{7A}^0$ is a reliable reporter for the existence of synergy. This means that synergy typically occurs if the active fraction of *Tr*Cel7A is enhanced by the presence of *Tr*Cel7B. Specifically, synergy at time t is maintained as long as the active fraction x_{7A} averaged up to time t is higher than the fraction x_{7A}^0 averaged up to time t. Analysis of the changes in the active and inactive *Tr*Cel7A state fractions upon *Tr*Cel7B addition therefore isolates and reveals the individual driving forces for synergy and anti-synergy.

Part II

Fluctuating Hydrodynamics for Multiscale Modeling

Chapter 5

Bridging Fluctuating Hydrodynamics and Molecular Dynamics Simulations

Part II of this dissertation will present the development of hydrodynamic field models using information obtained from all-atom molecular dynamics (MD) simulations. The presented framework provides a practical bridge between atomistic and continuum descriptions of a fluid, and hence provides methods that could facilitate multiscale approaches toward modeling complex fluid systems. In this work, a fluctuating hydrodynamics (FHD) simulation model is developed for fluids from their all-atom models. An accurate passing of information to a more macroscopic scale will enable larger length- and time-scales to be accessed via simulation, while maintaining physical fidelity with properties of the molecular system.

Our multiscale approach is unique because we coarse-grain a particle representation of a fluid into a hydrodynamic field representation. This partice-to-field approach is advantageous over traditional particle-to-particle approaches for several reasons. First, the physical interpretation of the hydrodynamic model is clearer, because the dynamical variables are the mass, momentum, and energy densities commonly used to describe a fluid at the macroscopic level. Therefore, the transition from Newtonian mechanics used to describe molecular dynamics to continuum mechanics used to describe fluid dynamics is naturally followed by the coarse-graining method. In contrast, the "collective beads" in coarse-grain particle models are difficult to interpret in a physical context, especially if they represent the solvent. For example, one cannot easily trace the positions of a solvent bead from an atomistic trajectory of solvent molecules, and therefore, it becomes unclear to what extent these particles faithfully represent the atomistic fluid. Second, the hydrodynamical relaxation within the atomistic system can be captured by the hydrodynamic model by specifying the fluid transport coefficients, which are well-measured by experiments. In coarse particle models, these dynamics are governed by friction coefficients of a Langevin equation, which are not easily extracted from knowledge of the atomistic forcefield and do not have simple relations to the fluid transport coefficients.

In addition to this dissertation, the work presented in Part II is also discussed in the following publications.

- Chapter 5: Shang, B. Z., Voulgarakis, N. K. & Chu, J. W. Fluctuating hydrodynamics for multiscale modeling and simulation: energy and heat transfer in molecular fluids. *J. Chem. Phys.* **137,** 044117 (2012).
- Chapter 6: Shang, B. Z., Voulgarakis, N. K. & Chu, J. W. Fluctuating hydrodynamics for multiscale simulation of inhomogeneous fluids: mapping all-atom molecular dynamics to capillary waves. *J. Chem. Phys.* **135,** 044111 (2011).

The goals and objectives of Part II of this dissertation are:

1. Develop a mapping protocol to coarse-grain atomistic degrees of freedom into fluctuating hydrodynamic fields.
2. Determine and achieve metrics that establish thermophysical consistency between atomistic and hydrodynamic models.
3. Construct an interfacial fluctuating hydrodynamic model for water to study the coupling between thermal fluctuations and phase equilibrium in biophysical processes.

In Chapter 5, a general scheme is developed to construct a coarse-grained (CG) fluctuating hydrodynamic (FHD) field model from all-atom molecular dynamics (MD) simulations for a homogeneous fluid. The FHD model preserves the thermophysical properties of the all-atom model by reproducing the fluctuation and relaxation behavior of the atomistic density fields. We illustrate that fluctuating hydrodynamics simulations can be used to capture the thermodynamic and hydrodynamic responses of molecular fluids at the nanoscale, including those associated with energy and heat transfer.

Using all-atom molecular dynamics trajectories as the reference data, the atomistic coordinates of each snapshot are mapped onto mass, momentum, and energy density fields on Eulerian grids to generate a corresponding field trajectory. The molecular length-scale associated with finite molecule size is explicitly imposed during this coarse-graining by requiring that the variances of density fields scale inversely with the grid volume. From the fluctuations of field variables, the response functions and transport coefficients encoded in the all-atom MD trajectory are computed. By using the extracted fluid properties in FHD simulations, we show that the fluctuations and relaxation of hydrodynamic fields quantitatively match with those observed in the reference all-atom MD trajectory, hence establishing compatibility between the atomistic and field representations. We also show that inclusion of energy transfer in the FHD equations can more accurately capture the thermodynamic and hydrodynamic responses of molecular fluids. The results indicate that the proposed MD-to-FHD mapping with explicit consideration of finite molecule size provides a robust framework for coarse-graining the solution phase of complex molecular systems.

Chapter 5 is reprinted with permission from Shang, B. Z., Voulgarakis, N. K. & Chu, J. W. Fluctuating hydrodynamics for multiscale modeling and simulation: energy and heat transfer in molecular fluids. *J. Chem. Phys.* **137,** 044117 (2012). Copyright 2012, American Institute of Physics.

5.1 Motivation

Hydrodynamic fluctuations are important forces in affecting molecular transport at the nanometer to micrometer scales [172]. Self-assembly [173], mixing [174,175], capillary dynamics [176–178], and molecular motor actions [179] are only several examples that are of fundamental as well as technological importance [180]. A key signature at small length scales is the significant influence of thermal energy. To capture hydrodynamics and thermal fluctuations under the same footing, a powerful theoretical framework is fluctuating hydrodynamics (FHD) [172,181,182]. The FHD equations are stochastic partial differential equations that introduce stochastic fluxes into the conservation equations of momentum and energy transport to represent thermal noise. The magnitude of fluctuations depends on the grid size used to spatially discretize the system [172,182]. In the limit of large ($>$µm^3) volumes of grid cells, the deterministic conservation equations are recovered. In addition to solving the FHD equations directly in the Eulerian frame, the FHD equations have also been casted into different forms for developing mesoscale simulation methods such as the fluctuating Lattice-Boltzmann [183] technique and smoothed dissipative particle dynamics [184].

The explicit incorporation of thermal fluctuations into the conservation laws of continuum mechanics makes FHD an ideal framework for mesoscopic simulation. Indeed, some examples where FHD simulations have been applied include the modeling of Rayleigh-Taylor instability [174], diffusive mixing of miscible fluids [175], breakup of nanojets [176], capillary wave phenomena [177,178], droplet spreading [185], and dynamics of nanoparticles [186–188] and polymers [189,190]. Numerical techniques for simulating the FHD equations are also being developed [191–195] and assessed [192,196]. However, the potential for using FHD as a framework to coarse-grain (CG) atomistic systems has been less explored. A key issue is whether consistency between FHD and all-atom molecular dynamics (MD) can be established. Specific concerns include whether a small length-scale ($0.5 - 1\,\mathrm{nm}$) of discretization is physically appropriate for solving the FHD equations in order to match with the all-atom MD model and if the observables from FHD and MD simulations can be bridged to establish consistency in thermodynamic and hydrodynamic responses at different scales. These questions must be addressed before FHD can be applied as a framework of CG modeling for molecular fluids at the nanoscale.

To extend the accessible length- and time-scales of molecular simulation, an obvious strategy is to reduce the degrees of freedom in the model by coarse-graining [197]. Since the solvent part often constitutes the majority of computational cost in molecular simulation, especially for all-atom MD, proper coarse-graining of the solvent is critical for successful modeling and simulation at the mesoscopic scale. Ideally, a coarse-graining method should be simple and be able to retain the essential physics for predicting the behaviors of the system of interest. In the category of explicit-solvent modeling, the most common approach is to group one or more solvent molecules together into a bead [198]. An immediate difficulty of such particle-to-particle coarse-graining is that it is very difficult to control bulk and interfacial properties separately through the pair potentials between particles [198,201]. Using a field-theoretic framework to coarse-grain the solution phase, on the other hand, not only removes the computational cost of tracking the positions and neighbor list of the solvent beads but also allows interfacial properties to be tuned separately from the bulk equation of state [178]. Furthermore, for particle-based coarse-graining, connection to experimental transport coefficients such as viscosity is only implicit through the interaction potentials and friction coefficients in the equations of motion [198,199,202]. In field-theoretic coarse-graining using the FHD framework, on the other hand, the system dynamics are explicitly controlled by the equations of state and transport coefficients of the fluid, making the connection to experimental dynamics explicit [172,181,182].

Utilizing FHD as a framework of multiscale modeling is also advantageous in connecting the CG model with the atomistic-scale simulation. It has been shown that an all-atom MD trajectory can be mapped onto hydrodynamic fields of mass and momentum to establish common observables between MD and FHD simulations [195]. By matching the statistics of common observables, the FHD model can be parameterized to be consistent with the finer resolution all-atom model [178,195,203,204]. A key finding is that for such consistency to hold at the sub-nanometer length-scale, the finite sizes of atoms or molecules need to be explicitly considered in mapping the atomistic configurations and velocities onto field variables [178,195,203,204]. This consideration is necessary to ensure that thermodynamic and hydrodynamic responses of the fluid are preserved at the coarser scale.

In regards to performing FHD simulations, it has been established that a staggered finite volume scheme allows the FHD equations to be solved with high accuracy in satisfying the fluctuation-dissipation theorem at 300 K even if the grid size is as small as 5 Å [195]. The combination of mapping all-atom MD onto fields and constructing a corresponding FHD model has been applied to investigate the viscoelastic response of molecular fluids at nanoscales [203,204], but not yet to (1) the responses of energy and heat transfer [205], and (2) the interfacial properties that enable a fluid to phase separate and form interfaces [178]. The first of these two areas is investigated in this Chapter while the latter is explored in Chapter 6.

This Chapter aims to establish the applicability of FHD for coarse-graining the transport of energy via thermal fluctuations in all-atom MD simulations by generalizing the framework developed by Voulgarakis & Chu [195]. In particular, we couple the stochastic energy balance [172] with the equations of mass and momentum conservation according to FHD and identify the proper molecular length scales for coarse-graining all-atom MD onto the FHD model to capture the behaviors of energy and heat transfer.

In short, the atomistic coordinates, velocities, and internal energies calculated in MD simulations are mapped onto hydrodynamic variables on grid cells of nanometer scale. The finite size of an atom or molecule is explicitly considered during mapping, and is identified based on the criterion that density fluctuations within a grid cell should scale inversely with grid cell volume [182,195]. We illustrate that the fluid properties of argon and water extracted from this MD-to-FHD mapping are in quantitative agreement with the values determined from conventional methods that are mapping-free. Therefore, the results of this work establish that even at the sub-nanometer length-scale of grid cells, FHD can capture the thermodynamic and hydrodynamic responses of all-atom MD, establishing the applicability of FHD as a framework for multiscale modeling of systems that involve energy and heat transfer (EHT) of molecular fluids. Furthermore, we show that including EHT in the FHD equations more accurately captures the hydrodynamic responses encoded in all-atom MD simulations. We also investigate the effects of the non-linear terms in FHD equations on the thermodynamic and hydrodynamic responses of molecular fluids.

The rest of this Chapter is organized as follows. In Section 5.2, we provide the background on the FHD equations with EHT. In Section 5.3, we describe the methods used to calculate fluid properties from the statistics of field variables. In Section 5.4, we describe the mapping procedure to convert atomistic coordinates into hydrodynamic fields and the calculations of fluid properties from the field representation of an all-atom MD trajectory. In Section 5.5, we simulate the FHD model, using parameters obtained from MD, and demonstrate that hydrodynamic fluctuations in the FHD model are consistent with those of MD. In Section 5.6, we summarize the results of this work.

5.2 Fluctuating hydrodynamics with energy and heat transfer

5.2.1 Governing equations

The equations of fluctuating hydrodynamics (FHD) incorporate thermal fluctuations into fluid flow by modeling the time evolution of mass, momentum, and energy density fields via Langevin-type dynamics [172,181,182,192]. The governing equations for mass, momentum, and energy transfer are, respectively:

$$\frac{\partial}{\partial t}\rho = -\nabla\cdot(\rho\boldsymbol{v}), \tag{5.1}$$

$$\frac{\partial}{\partial t}\rho\boldsymbol{v} = -\nabla\cdot(\rho\boldsymbol{v}\otimes\boldsymbol{v}) + \nabla\cdot(\mathbf{P}) - \nabla\cdot(\boldsymbol{\Pi}+\boldsymbol{\delta\Pi}), \tag{5.2}$$

$$\frac{\partial}{\partial t}\rho e = -\nabla\cdot(\rho e\boldsymbol{v}) + \nabla\cdot(\mathbf{P}\cdot\boldsymbol{v}) - \nabla\cdot(\boldsymbol{\Pi}+\boldsymbol{\delta\Pi})\cdot\boldsymbol{v} - \nabla\cdot(\boldsymbol{q}+\boldsymbol{\delta q}). \tag{5.3}$$

The field variables include the mass density ρ, momentum density $\rho\boldsymbol{v}$, total energy density e, and streaming velocity $\boldsymbol{v}$. The reversible (pressure), dissipative, and fluctuating stress tensors are $\mathbf{P}$, $\boldsymbol{\Pi}$, and $\boldsymbol{\delta\Pi}$, respectively. For homogeneous fluids considered in this work, the pressure tensor is isotropic, $\mathbf{P} = -p\mathbf{I}$, where p is the thermodynamic pressure, and $\mathbf{I}$ is the unit tensor. The dissipative and fluctuating heat fluxes are $\boldsymbol{q}$ and $\boldsymbol{\delta q}$, respectively. The dissipative stress is assumed to be Newtonian,

$$\boldsymbol{\Pi} = -\eta_{\mathrm{S}}\big(\nabla\boldsymbol{v} + \nabla\boldsymbol{v}^{\mathrm{T}}\big) - \left(\eta_{\mathrm{B}} - \frac{2}{3}\eta_{\mathrm{S}}\right)(\nabla\cdot\boldsymbol{v})\mathbf{I}. \tag{5.4}$$

where η_{S} is the shear viscosity and η_{B} is the bulk viscosity. The dissipative heat flux is modeled by Fourier's Law to represent conductive heat transfer with thermal conductivity

λ,

$$\boldsymbol{q} = -\lambda \nabla T. \tag{5.5}$$

The fluctuating forces in momentum and energy transfer are described as Gaussian white noise with zero mean and covariance according to the fluctuation-dissipation theorem [172,181,182],

$$\langle \delta\Pi_{ij}(\boldsymbol{r},t)\delta\Pi_{kl}(\boldsymbol{r}',t')\rangle = 2k_{\mathrm{B}}T_0\left[\eta_{\mathrm{S}}(\delta_{il}\delta_{jk}+\delta_{ik}\delta_{jl}) + \left(\eta_{\mathrm{B}} - \frac{2}{3}\eta_{\mathrm{S}}\right)\delta_{ij}\delta_{kl}\right]\delta(\boldsymbol{r}-\boldsymbol{r}')\delta(t-t') \tag{5.6}$$

$$\langle \delta q_i(\boldsymbol{r},t)\delta q_j(\boldsymbol{r}',t')\rangle = 2\lambda k_{\mathrm{B}}T_0^2\delta_{ij}\delta(\boldsymbol{r}-\boldsymbol{r}')\delta(t-t') \tag{5.7}$$

In a numerical simulation, these stochastic fluxes are discretized in both space and time and therefore, care must be taken to scale the covariances among the discretized stochastic fluxes to preserve the time-dependent probability distribution of the field variables implied by the original partial differential equations. Because the resulting field variables are calculated as *sums* and *differences* of these discretized fluxes, their covariances are scaled by the grid volume V_{G} and time step Δt as [192],

$$\left\langle \delta\Pi_{ij}\left(\boldsymbol{n},\hat{t}\right)\delta\Pi_{kl}\left(\boldsymbol{n}',\hat{t'}\right)\right\rangle \sim 2k_{\mathrm{B}}T_0/(V_{\mathrm{G}}\Delta t), \tag{5.8}$$

$$\left\langle \delta q_i\left(\boldsymbol{n},\hat{t}\right)\delta q_j\left(\boldsymbol{n}',\hat{t'}\right)\right\rangle \sim 2\lambda k_{\mathrm{B}}T_0^2/(V_{\mathrm{G}}\Delta t). \tag{5.9}$$

The grid index is $\boldsymbol{n}$ and time index is $\hat{t}$. The magnitude of the stochastic fluxes is governed by the equilibrium temperature T_0. The dissipative and fluctuating forces represent the effective contribution from the microscopic degrees of freedom that have been coarse-grained away.

In this work, we consider liquids with fluctuations close to equilibrium and employ a linear equation of state [182,195],

$$p(\rho,T) = \frac{1}{\rho_0\kappa_T}\rho + \frac{\alpha_P}{\kappa_T}T. \tag{5.10}$$

This approximation will be verified in Section 5.5, since the fluctuations calculated from all-atom MD are essentially Gaussian. To close the system of equations, the energy balance is now written in terms of the temperature T as [172],

$$\frac{\partial}{\partial t}\rho c_V T = -\nabla\cdot(\rho c_V T\boldsymbol{v}) - \frac{T\alpha_P}{\kappa_T}\nabla\cdot\boldsymbol{v} - \nabla\cdot(\boldsymbol{q}+\boldsymbol{\delta q}). \tag{5.11}$$

The constant volume specific heat is c_V, the thermal expansivity is α_P, and the isothermal compressibility is κ_T. The temperature equation follows from the differential relation between the specific internal energy, temperature, and mass density.

This work generalizes the model presented by Voulgarakis & Chu [195] by incorporating energy and heat transfer (EHT) into the FHD equations. The FHD model with EHT is a more complete coarse-grained (CG) representation of the atomistic system because energy is constantly being exchanged at nanometer scales due to spontaneous fluctuations, even in systems at equilibrium. Therefore, the temperature degree of freedom is explicitly considered here. The additional parameters introduced by incorporating EHT into the FHD equations are the specific heat c_V, thermal expansivity α_P, and thermal conductivity λ.

5.2.2 Linearized FHD model

In this section, the linearized FHD equations are introduced [172,181,182,206,207], which will be later used to assist in parameterizing the full non-linear FHD model from MD simulations. The linear equations allow analytic expressions to be derived that relate response functions and transport coefficients to the fluctuations and time correlation functions of the hydrodynamic field variables. Therefore, the fluid properties of an atomistic model can be estimated by fitting the MD data to these analytical expressions as the starting point; the details of model parameterization will be discussed in Section 5.3. The linearized FHD equations are:

$$\frac{\partial}{\partial t}\delta\rho = -\rho_0\nabla\cdot\boldsymbol{\delta v}, \tag{5.12}$$

$$\rho_0\frac{\partial}{\partial t}\boldsymbol{\delta v} = -\nabla\delta p + \eta_{\mathrm{S}}\nabla^2\boldsymbol{\delta v} + \left(\eta_{\mathrm{B}} + \frac{1}{3}\eta_{\mathrm{S}}\right)\nabla(\nabla\cdot\boldsymbol{\delta v}) - \nabla\cdot\boldsymbol{\delta\Pi}, \tag{5.13}$$

$$\rho_0 c_V\frac{\partial}{\partial t}\delta T = -\frac{T_0\alpha_P}{\kappa_T}\nabla\cdot\boldsymbol{\delta v} + \lambda\nabla^2\delta T - \nabla\cdot\boldsymbol{\delta q}. \tag{5.14}$$

The fluctuations of field variables are $\delta\phi = \phi - \phi_0$, where $\phi = \{\rho, \boldsymbol{v}, T\}$, and ϕ_0 is the equilibrium average. The fluctuation of the local pressure δp is related to the density and temperature fluctuations by [172,206],

$$\delta p = \frac{1}{\rho_0\kappa_T}\delta\rho + \frac{\alpha_P}{\kappa_T}\delta T. \tag{5.15}$$

5.3 Calculation of fluid properties

To obtain parameters for the FHD equations, the response functions (κ_T, c_V, α_P) and transport coefficients (η_{S}, η_{B}, λ) are calculated from an all-atom MD trajectory by mapping atomistic coordinates onto density fields with explicit consideration of the finite sizes of atoms or molecules. The values of these fluid properties are compared with those obtained via mapping-free methods to assess the accuracy of the mapping procedure. Using the calculated fluid properties in an FHD simulation, we show later in Section 5.4 that the resulting fluctuations and relaxation of field variables quantitatively match with those observed in the reference all-atom MD trajectory, thereby establishing compatibility between FHD and MD models.

5.3.1 Response functions

With the MD-to-FHD mapping

To map atomistic coordinates onto density fields, the MD simulation box is discretized into cubic grids, and the atomistic coordinates in each snapshot in the MD trajectory are then used to calculate the mass and internal energy density of each grid. Details of this mapping will be described in Section 5.4. The all-atom MD trajectory is thus transformed into a field trajectory. Since each grid cell can be viewed as an open system that contains a grand canonical ensemble of microstates, response functions can be determined from the density fluctuations within each grid [208,209]. The isothermal compressibility is

$$\kappa_T = \frac{\left\langle (\delta\rho)^2 \right\rangle_{\mu VT} V_G}{\rho^2 k_{\mathrm{B}} T}, \tag{5.16}$$

the specific heat is

$$c_V = \frac{V_G}{\rho k_{\mathrm{B}} T^2} \left(\left\langle (\delta u)^2 \right\rangle_{\mu VT} - \frac{\langle \delta u \delta\rho \rangle_{\mu VT}^2}{\left\langle (\delta\rho)^2 \right\rangle_{\mu VT}} \right), \tag{5.17}$$

and the thermal expansivity is

$$\alpha_P = \frac{\kappa_T}{T}(P + \varepsilon) - \frac{\langle \delta\varepsilon \delta\rho \rangle_{\mu VT} V_G}{\rho k_{\mathrm{B}} T^2}. \tag{5.18}$$

In the above, u is the internal energy density, P is the pressure, and ε is the potential energy density. Thermal averages are taken over the grand canonical ensemble μVT distribution. Eqs. 5.16 and 5.17 also follow from the Gaussian stationary distribution satisfied by the linearized FHD equations [172,207,210]. The derivation of isothermal compressibility relation is commonly found in textbooks [211], but the derivations for the latter two response functions are less well-known. As a reference, their derivations are provided in Section 5.7.

Without the MD-to-FHD mapping

To assess the results obtained via mapping to grid cells, response functions are also computed using an alternate method that does not involve mapping. In particular, all-atom MD simulations of a molecular system are performed under both NVT and NPT conditions, and appropriate formulas [208] are used to determine the response functions. In particular, the compressibility can be calculated as

$$\kappa_T = \frac{\left\langle (\delta V)^2 \right\rangle_{NPT}}{V k_{\mathrm{B}} T}, \tag{5.19}$$

the specific heat can be calculated as

$$c_V = \frac{\left\langle (\delta U)^2 \right\rangle_{NVT}}{\rho V k_{\mathrm{B}} T^2}, \tag{5.20}$$

and the thermal expansivity can be calculated as

$$\alpha_P = \frac{\langle \delta V \delta(U + PV) \rangle_{NPT}}{V k_{\mathrm{B}} T^2}. \tag{5.21}$$

In the above, V is the volume of the MD simulation box and U is the total internal energy. Response functions calculated using Eqs. 5.16, 5.17, and 5.18 will be compared with those calculated from Eqs. 5.19, 5.20, and 5.21 to assess the accuracy of the mapping scheme.

5.3.2 Transport coefficients

With the MD-to-FHD mapping

Transport coefficients are calculated from all-atom MD using the mapping scheme followed by computing the time correlation functions of density and velocity fields in Fourier space. The results are then fit to analytical expressions derived from the linearized FHD equations. The density time correlation function from linearized FHD is [172,182,206]

$$C_\rho = \frac{\langle \rho(-k,0)\rho(k,t) \rangle}{\langle |\rho(k,0)|^2 \rangle} = \frac{\gamma - 1}{\gamma} \exp\left(-D_T k^2 t\right) + \frac{1}{\gamma} \cos\left(c_S k t\right) \exp\left(-\Gamma k^2 t\right). \tag{5.22}$$

The specific heat ratio is $\gamma = c_P/c_V$, the thermal diffusivity is $D_T = \lambda/\rho c_P$, the sound attenuation coefficient is $\Gamma = 1/2(D_V + (\gamma - 1)D_T)$, and the sound speed is $c_S = \left(\gamma/\rho\kappa_T\right)^{1/2}$. The longitudinal kinematic viscosity is $D_V = (4\eta_{\mathrm{S}}/3 + \eta_{\mathrm{B}})/\rho$, and k is the wave number. For transverse velocity, the correlation function is [182]

$$C_{v,t} = \frac{\langle v_t(-k,0) v_t(k,t) \rangle}{\langle |v_t(k,0)|^2 \rangle} = \exp\left(-\nu k^2 t\right). \tag{5.23}$$

where $\nu = \eta_{\mathrm{S}}/\rho$ is the kinematic shear viscosity, while for longitudinal velocity [182],

$$C_{v,l} = \frac{\langle v_l(-k,0) v_l(k,t) \rangle}{\langle |v_l(k,0)|^2 \rangle} = \cos\left(c_S k t\right) \exp\left(-\Gamma k^2 t\right), \tag{5.24}$$

which is valid for small k. In our analysis, we choose the wave vector to point in the z direction, i.e. $\boldsymbol{k} = (0, 0, k)$, so $v_t = v_{x,y}$ and $v_l = v_z$.

Without the MD-to-FHD mapping

As an independent check, transport coefficients can also be determined from all-atom MD data without mapping to grid cells by using the Green-Kubo relations [182,211,212]. The shear viscosity is related to the relaxation of the microscopic pressure as

$$\eta_{\mathrm{S}} = \frac{V}{k_{\mathrm{B}} T} \int_0^\infty \left\langle \tilde{P}_{\alpha\beta}(0) \tilde{P}_{\alpha\beta}(t) \right\rangle dt, \tag{5.25}$$

the bulk viscosity is

$$\eta_{\mathrm{B}} = \frac{V}{k_{\mathrm{B}}T} \int_{0}^{\infty} \left\langle \delta\tilde{P}(0)\delta\tilde{P}(t) \right\rangle dt, \tag{5.26}$$

where $\delta\tilde{P}(t) = (1/3)\sum_{\alpha}\tilde{P}_{\alpha\alpha} - \left\langle \tilde{P}_{\alpha\alpha} \right\rangle$, and the thermal conductivity is

$$\lambda = \frac{V}{k_{\mathrm{B}}T^2} \int_{0}^{\infty} \left\langle \delta\tilde{J}_{\alpha}(0)\delta\tilde{J}_{\alpha}(t) \right\rangle dt. \tag{5.27}$$

In the above, $\tilde{P}_{\alpha\beta}$ are the components of the microscopic pressure [182,208,211], and $\tilde{J}_{\alpha}$ are the components of the microscopic heat flux [182,211]. Transport coefficients calculated using Eqs. 5.22, 5.23, and 5.24 will be compared with those calculated from Eqs. 5.25, 5.26, and 5.27 to assess the accuracy of the mapping scheme.

To establish consistency between FHD and MD models, we next describe the mapping of atomistic coordinates from MD simulation onto density fields, taking into account finite molecule sizes, and analyze the field trajectory obtained from MD to calculate fluid properties, which are used later as input into FHD simulations.

5.4 Mapping an all-atom MD trajectory onto snapshots of field configurations

In this section, we describe the details of all-atom MD simulations and our proposed mapping scheme. In particular, we show that the finite size of an atom or molecule needs to be explicitly considered to ensure that the magnitude of density fluctuations is inversely proportional to the volume of grid cells, as illustrated in Eqs. 5.16, 5.17, and 5.18. Liquid argon and water are employed as illustrative examples.

5.4.1 Simulation details

For both argon and water, all-atom MD simulations were performed in the NVT ensemble using a Nosé-Hoover thermostat [213] in a 60 Å cubic box with periodic boundary conditions for 10 ns. Lennard-Jones (LJ) interactions were smoothly switched off from 10 to 12 Å. The time step is 2 fs, the neighbor list was updated every 10 integration steps, and a snapshot was saved every 0.1 ps. For argon, the LJ parameters are $\epsilon = 0.238\,\mathrm{kcal/mol}$, $\sigma = 3.405\,\text{Å}$ [214], and the atomic mass is 39.948 amu. Simulation conditions for argon were $T = 86.5\,\mathrm{K}$, and $\rho = 0.844\,\mathrm{amu}/\text{Å}^3$. For water, the TIP3P potential was used with particle-mesh Ewald [215] to compute electrostatic interactions and the SHAKE algorithm [216] was used to constrain bond lengths. Simulation conditions for water were $T = 300\,\mathrm{K}$ and $\rho = 0.603\,\mathrm{amu}/\text{Å}^3$. To compute response functions from fluctuation formulas, NPT simulations were also performed via the Langevin piston method [217]. To compute time correlation functions, simulations were performed under NVE conditions with a box size of 60 Å × 60 Å × 240 Å and a simulation time of 50 ns.

To map atomistic positions and velocities onto mass, momentum, and energy density fields, the simulation box was first discretized into an array of cubic grids varying in size from 5 to 20 Å. Each argon atom or water molecule was also discretized as a space-filling cube. For each grid, the total mass m, momentum $\boldsymbol{p}$, and internal energy U are calculated as

$$m = \sum_i \phi_i m_i, \tag{5.28}$$

$$\boldsymbol{p} = \sum_i \phi_i \boldsymbol{p}_i, \tag{5.29}$$

$$U = \sum_i \phi_i (U_{K,i} + U_{P,i}), \tag{5.30}$$

where the sum i runs over atoms (argon) or molecules (water). The volume fraction ϕ_i is the volume of an atom or molecule that overlaps with the grid divided by the total volume of the atom or molecule. The kinetic contribution to the internal energy is calculated as

$$U_{K,i} = (1/2) \sum_j m_j (\boldsymbol{v}_j - \boldsymbol{v})^2, \tag{5.31}$$

where the sum j runs over the atoms of molecule i, and the streaming velocity in a grid is $\boldsymbol{v} = \boldsymbol{p}/m$. The potential energy $U_{P,i}$ includes Lennard-Jones and the real-space electrostatic interactions (for water). To determine errors introduced by neglecting the long-range contributions, the real-space contribution was gradually increased when computing the potential energy until the response functions became independent of the cutoff. We found that this convergence occurs at a cutoff around 42 Å. To determine the volume fraction inside a grid, each argon atom or water molecule is modeled as a cube with length d_{mol}. Physically, this means the mass, momentum, and internal energy of an atom or molecule is spread uniformly over a cube with side length of d_{mol}. For water, the geometric center of the cube coincides with the molecule center of mass.

5.4.2 Fluid properties from MD

After the mapping with d_{mol}, fluctuations of mass and energy density fields in each grid are used to calculate the response functions encoded in an all-atom MD trajectory. For a grid size in the nanometer regime, the magnitude of fluctuations also depends on the molecular length scale d_{mol} [195]. A criterion based on statistical mechanics for determining d_{mol} is that the variances of density fluctuations within a grid should scale inversely with grid cell volume. The fluctuations of mass and internal energy densities mapped from all-atom MD trajectories are plotted as a function of cubic grid length d_G and molecule size d_{mol} for argon

Table 5.1. Response functions of liquid argon and water calculated by mapping with d_{mol} and without mapping. Values calculated with the MD-to-FHD mapping were calculated using Eqs. 5.16, 5.17, and 5.18, while values without the MD-to-FHD mapping were calculated using Eqs. 5.19, 5.20, and 5.21. The isothermal compressibility is $\kappa_T\,[\mathrm{atm}^{-1}]$, the constant volume specific heat capacity is $c_V\,[\mathrm{cal/mol\,K}]$, and the thermal expansivity is $\alpha_P\,[\mathrm{K}^{-1}]$. Standard errors are determined from results using three independent MD simulations of 1 ns each.

Argon	With mapping	Without mapping
$\kappa_T \times 10^{-4}$	1.7 ± 0.2	2.1 ± 0.2
c_V	2.5 ± 0.5	2.3 ± 0.1
$\alpha_P \times 10^{-3}$	3.6 ± 0.2	3.6 ± 0.1
Water	**With mapping**	**Without mapping**
$\kappa_T \times 10^{-5}$	5.3 ± 0.1	5.6 ± 0.6
c_V	16.7 ± 0.2	16.8 ± 0.4
$\alpha_P \times 10^{-4}$	7.4 ± 0.2	7.5 ± 0.5

and water in Figs. 5.1 and 5.2, respectively. For argon, the optimal d_{mol} based on the criterion of size consistent scaling of variances is 2.4 Å for mass density and 0.4 Å for energy density. For water, the optimal d_{mol} is 2.2 Å for mass density and 1.8 Å for energy density. For the covariance of mass and energy density, the optimal d_{mol} is 1.6 Å for argon and 2.0 Å for water. In general, density fluctuations gradually become overestimated with decreasing d_{mol}, as the molecule mass is over-concentrated in a small volume. On the other hand, density fluctuations gradually become underestimated with increasing d_{mol}, since volumes of different molecules overlap and cause artificial correlation. For larger grid sizes, the effects of d_{mol} become less significant, as the open system of a given grid cell is closer to the thermodynamic limit. The response functions are computed from the covariances via Eqs. 5.16, 5.17, and 5.18 using the optimal values of d_{mol} and a grid size of 10 Å, the same grid size that is used in FHD simulations. In Table 5.1, the response functions obtained via mapping to grids are listed, which agree well with results obtained from *NVT* and *NPT* simulations, affirming the self-consistency of mapping with d_{mol}. The inverse scaling of variances with grid volume down to 5 Å-sized grids, observed in Figs. 5.1 and 5.2, also underscores that the thermodynamic limit is satisfied even at sub-nanometer length scales, ensuring that an FHD model employing nanometer spatial discretization can be valid.

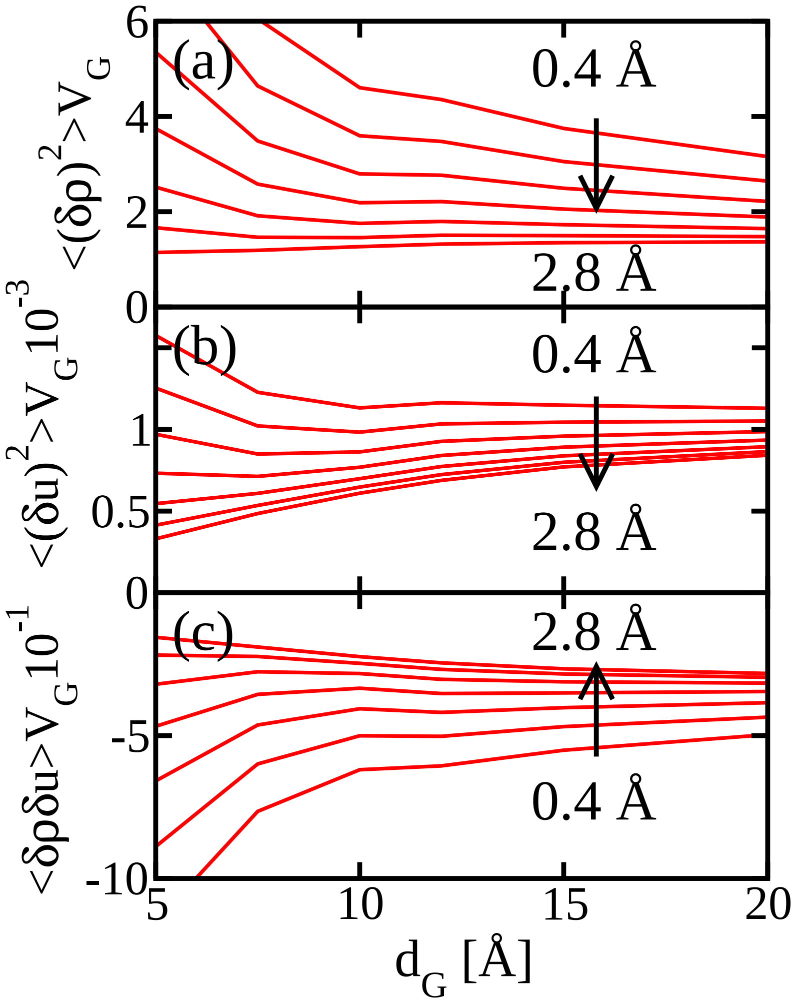


Figure 5.1. Variation of d_{mol} for argon as a function of cubic grid size d_G. Conversion of the snapshots in an all-atom MD simulation into a trajectory of field variables by the MD-to-FHD mapping with d_{mol} using different length-scales for coarse-graining the solvent molecules. Variances of the fluctuations of (**a**) mass density ρ and (**b**) internal energy density u of argon as a function of the length of cubic grid cells, d_G, and the molecular length scale of coarse-graining, d_{mol}. (**c**) Covariance of ρ and u fluctuations. The volume of the grid cell is $V_G = d_G^3$. Lines are plotted for d_{mol} ranging from 0.4 to 2.8 Å with 0.4 Å increments. Units are ρ [amu/Å^3], u [amu Å^2/ps^2], V_G [Å^3], and d_{mol} [Å].

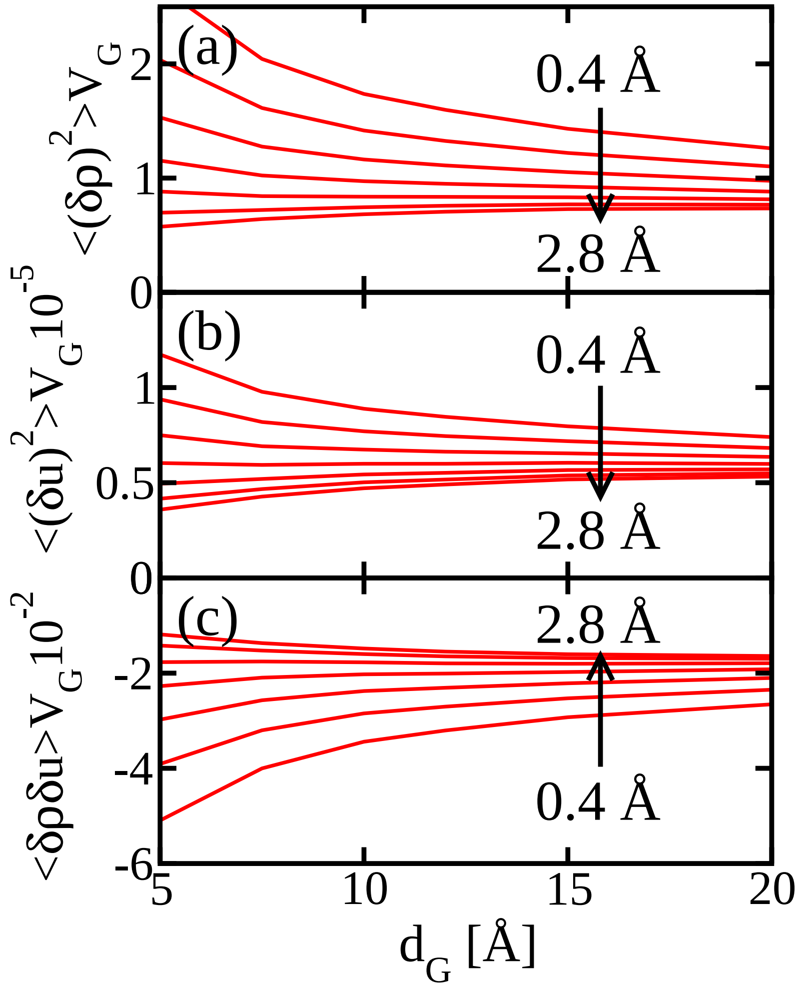


Figure 5.2. Variation of d_{mol} for water as a function of cubic grid size d_G. Conversion of the snapshots in an all-atom MD simulation into a trajectory of field variables by the MD-to-FHD mapping with d_{mol} using different length-scales for coarse-graining the solvent molecules. Variances of the fluctuations of (**a**) mass density ρ and (**b**) internal energy density u of water as a function of the length of cubic grid cells, d_G, and the molecular length scale of coarse-graining, d_{mol}. (**c**) Covariance of ρ and u fluctuations. The volume of the grid cell is $V_G = d_G^3$. Lines are plotted for d_{mol} ranging from 0.4 to 2.8 Å with 0.4 Å increments. Units are ρ [amu/Å^3], u [amu Å^2/ps^2], V_G [Å^3], and d_{mol} [Å].

Table 5.2. Transport coefficients of liquid argon and water calculated by mapping with d_{mol} and without mapping. Values calculated with the MD-to-FHD mapping were obtained via best-fitting to Eqs. 5.22, 5.23, and 5.24, while values without the MD-to-FHD mapping were obtained via the Green-Kubo relations in Eqs. 5.25, 5.26, and 5.27. The shear viscosity is η_{S} [cP], the bulk viscosity is η_{B} [cP], and the thermal conductivity is λ [W/m K]. Standard errors are determined from results using three independent MD simulations of 50 ns each (mapping) and 10 ns each (without mapping).

Argon	With mapping	Without mapping
η_{S}	0.27 ± 0.02	0.28 ± 0.02
η_{B}	0.09 ± 0.02	0.10 ± 0.01
λ	0.13 ± 0.03	0.12 ± 0.01
Water	**With mapping**	**Without mapping**
η_{S}	0.31 ± 0.01	0.32 ± 0.01
η_{B}	0.77 ± 0.04	0.77 ± 0.05
λ	1.04 ± 0.03	0.99 ± 0.03

The observation that the thermodynamic limit holds even at molecular length scales has also been supported by calculations of local stress tensor fluctuations in LJ liquids [218], which exhibit an inverse scaling with volume down to length-scales on the order of the size of the atom. These calculations also show that spatial covariances of the local stress tensor become uncorrelated at distances less than two LJ diameters, confirming the delta function assumption made in fluctuating stress in Eq. 5.6. In these studies, the temporal correlations also appear to decay rapidly, justifying the use of white noise in the FHD equations.

To determine transport coefficients, the density and velocity time correlation functions were computed from the mapped coordinates and fitted to Eqs. 5.22, 5.23. and 5.24, using both transport coefficients and response functions as adjustable parameters. The correlation functions and resulting fit are shown for argon in Fig. 5.3 and for water in Fig. 5.4. The best-fit transport coefficients listed in Table 5.2 agree well with the values obtained from independent Green-Kubo calculations. These results underscore the importance of accounting for finite molecule size, such as using d_{mol}, upon coarse-graining atomistic coordinates to density fields. This procedure ensures that the values of fluid properties are preserved upon mapping atomistic positions, velocities, and interaction energies onto field variables.

We note that in prior work [219–221], others have computed mass density and velocity

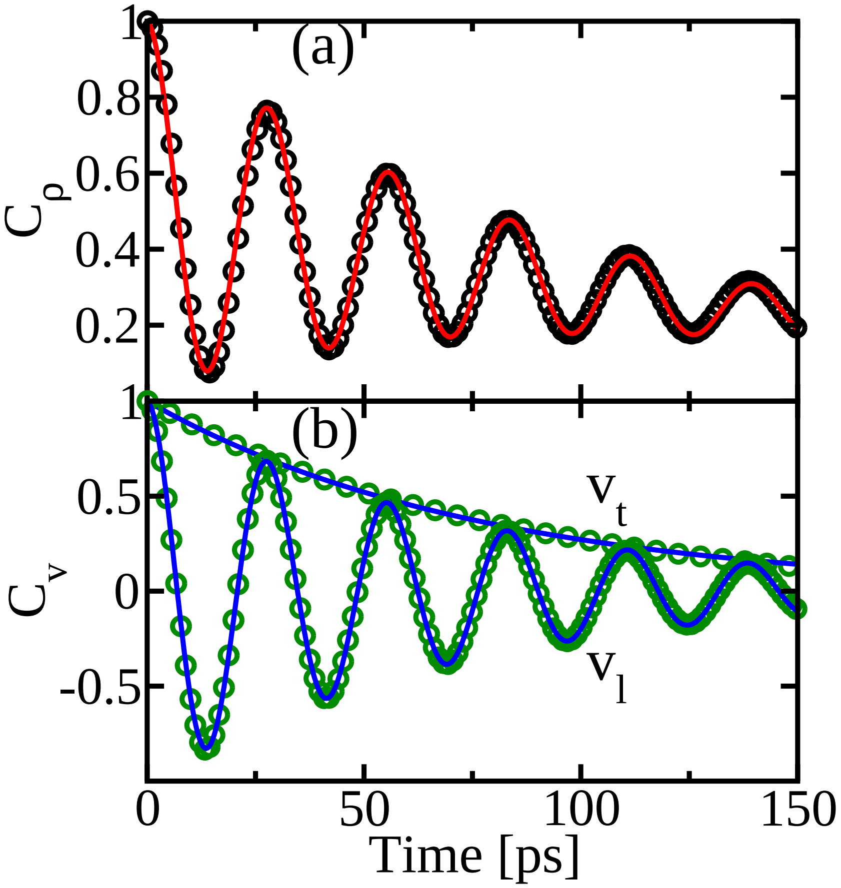


Figure 5.3. Time correlation functions of (**a**) mass density and (**b**) transverse v_t and longitudinal v_l velocity of argon at 86.5 K. The wave vector is $\boldsymbol{k} = (0, 0, k)$, where $k = 2\pi/240\,\text{Å}^{-1}$. All-atom MD results are shown as circles, while the best-fit curves of Eqs. 5.22, 5.23, and 5.24 are shown as solid lines.

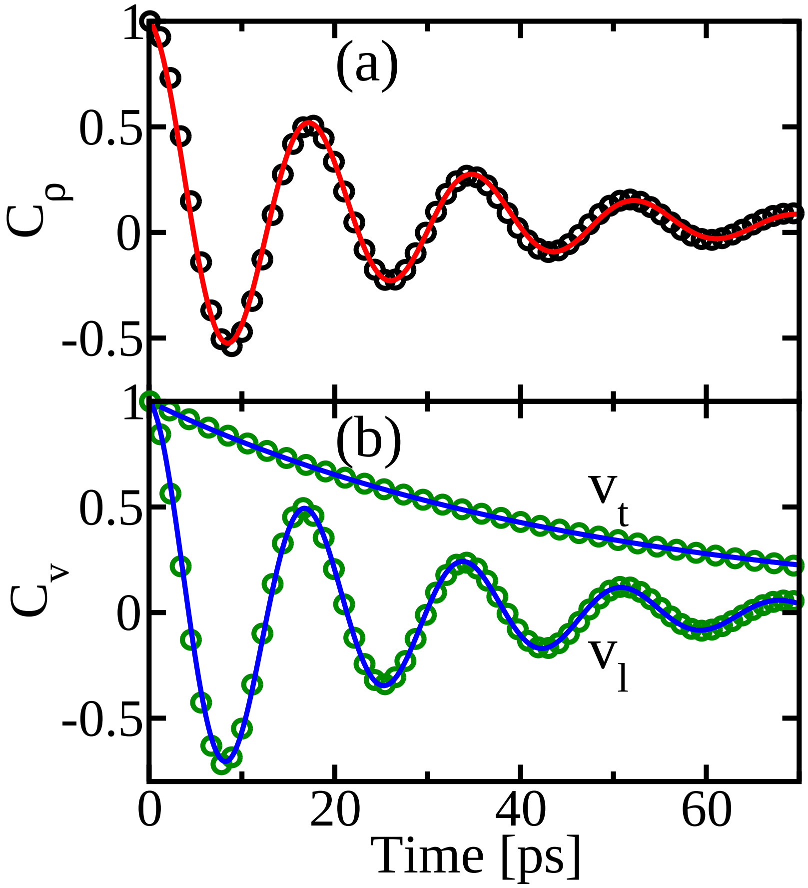


Figure 5.4. Time correlation functions of (**a**) mass density and (**b**) transverse v_t and longitudinal v_l velocity of water at 300 K. The wave vector is $\boldsymbol{k} = (0, 0, k)$, where $k = 2\pi/240\,\text{Å}^{-1}$. All-atom MD results are shown as circles, while the best-fit curves of Eqs. 5.22, 5.23, and 5.24 are shown as solid lines.

time correlation functions and fit these results to analytical predictions from linearized hydrodynamics to obtain fluid properties. The distinction of this work from prior analyses is that our MD-to-FHD mapping concerns the coarse-graining from the atomistic phase space onto density fields represented by grid cells, The earlier works, on the other hand, computed the time correlation functions by off-lattice Fourier transform and were primarily interested in assessing the dependence of fluid properties on wave number. As a result, for modeling complex systems with spatial heterogeneity, our coarse-graining method can provide the required generality.

5.5 Simulation with FHD equations

In this section, we present results of FHD simulations with energy and heat transfer using the response functions and transport coefficients of argon and water calculated from the MD-to-FHD mapping with d_{mol}. We illustrate that the density, velocity, and temperature fluctuations observed in FHD and MD simulations quantitatively agree with each other. We also show that density and velocity time correlation functions between the FHD and MD simulations are matched. These results demonstrate that the FHD model accurately describes the emergent hydrodynamic fluctuations of molecular fluids and is a promising tool for coarse-graining the solvent in multiscale simulations.

5.5.1 Simulation details

FHD simulations were performed for argon and water in a $60\,\text{Å}$ cubic box replicated periodically and discretized by $10\,\text{Å}$ cubic grids. The total simulation time was $10\,\mathrm{ns}$, the time step was $5\,\mathrm{fs}$, and a snapshot was saved every $0.1\,\mathrm{ps}$. For argon, simulation conditions were $T = 86.5\,\mathrm{K}$, and $\rho = 0.844\,\mathrm{amu}/\text{Å}^3$, and for water, $T = 300\,\mathrm{K}$ and $\rho = 0.603\,\mathrm{amu}/\text{Å}^3$. These conditions are identical to those used in all-atom MD simulations. To compute time correlation functions, simulations were performed using a box size of $60\,\text{Å} \times 60\,\text{Å} \times 240\,\text{Å}$ and a simulation time of $50\,\mathrm{ns}$. The FHD equations are solved using a staggered finite volume scheme in space and second-order Runge-Kutta integration in time [195]. Details for the random number generation for the stochastic fluxes can be found in Voulgarakis et al. [203]. The FHD input parameters are the response functions (κ_T, c_V, α_P) and transport coefficients (η_{S}, η_{B}, λ), which were determined previously from the MD-to-FHD mapping.

5.5.2 FHD-MD consistency in the statistics of fluctuations

We compare the fluctuations of mass density, streaming velocity, and temperature within a grid cell observed in MD and FHD simulations. Before doing so, we first establish the connection of the fluctuating temperature to molecular mechanical variables. From thermodynamic fluctuation theory [182,210,222], the fluctuating temperature can be estimated by expanding the internal energy density fluctuation to linear order in temperature and mass density,

$$\delta u = \left(\frac{\partial u}{\partial T}\right)_\rho \delta T + \left(\frac{\partial u}{\partial \rho}\right)_T \delta \rho. \tag{5.32}$$

In the above, $(\partial u/\partial T)_\rho = \rho c_V$, where c_V is found from Eq. 5.17. The temperature and mass density fluctuations are uncorrelated [182,210,222], which implies [222]

$$\left(\frac{\partial u}{\partial \rho}\right)_T = \frac{\langle \delta u \delta \rho \rangle}{\left\langle (\delta \rho)^2 \right\rangle}. \tag{5.33}$$

Eqs. 5.17, 5.32, and 5.33 allow the temperature within each grid to be determined from the mass and internal energy density fluctuations in the MD simulation. After mapping with d_{mol}, the fluctuating temperature fields in each grid cell represent the part of the internal energy density fluctuation that is uncorrelated with mass density.

In Figs. 5.5 and 5.6, the probability distributions for mass density, streaming velocity, and temperature from MD and FHD simulations are compared for argon and water, respectively. The results from linear and non-linear FHD simulations are both shown to illustrate the effects of non-linear advection on these fluctuations. As expected, solving the linearized FHD equations with the response functions obtained from MD using the analytic formulas gives rise to statistics that quantitatively agree with those extracted from MD directly. This result illustrates that the FHD equations can be solved accurately with nanometer-scale discretization. The presence of non-linear advection appears to affect the equilibrium fluctuations only slightly. Therefore, under the conditions of MD simulation conducted in this work, advection does not have an appreciable effect on the equilibrium distributions and hence the equation of state. The agreement between the atomistic and CG scales indicates that by mapping atomistic trajectories onto hydrodynamic fields with scaling consistency by optimizing d_{mol}, an FHD model can be developed to capture the thermodynamic responses of the molecular fluid.

5.5.3 FHD-MD consistency in time correlation functions

The time correlation functions obtained from the MD simulation and the linear and non-linear FHD simulations are plotted in Fig. 5.7 for argon and in Fig. 5.8 for water. Since the response functions and transport coefficients used in FHD simulations are determined by analyzing the field representation of all-atom MD trajectories with linearized hydrodynamics, the correlation functions calculated from the linearized FHD simulations indeed match with those calculated from all-atom MD. Therefore, solving the FHD equations with nanometer-scale discretization also gives an accurate representation of the dynamics of molecular fluids. As seen in Figs. 5.7 and 5.8, the primary difference between the linear and non-linear correlation functions for both argon and water is that the oscillation period is about 10% less due to the additional flux from advection. Fitting the MD results to the linear correlation function thus slightly underestimates the fluid compressibility, and can be easily corrected by iteration through simulations using the non-linear FHD model. Since this underestimation

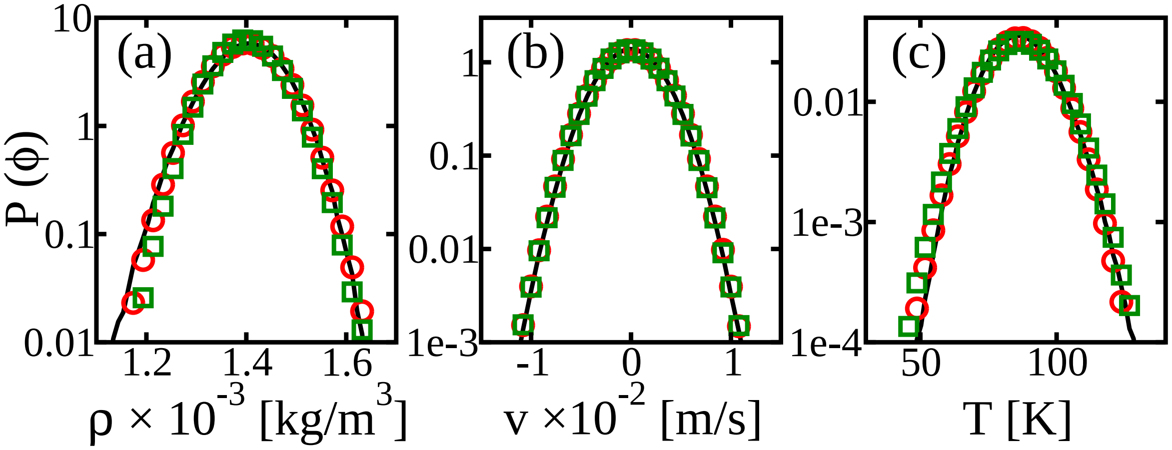


Figure 5.5. Probability distribution functions of (**a**) mass density, (**b**) streaming velocity, and (**c**) temperature of liquid argon at 86.5 K. The streaming velocity is defined in Section 5.4, and the temperature is defined in Eq. 5.32. Results of the all-atom MD simulation are shown as black curves. Results of the linear and non-linear FHD simulations are shown as red circles and green squares, respectively. $\phi = \rho$, v, or T

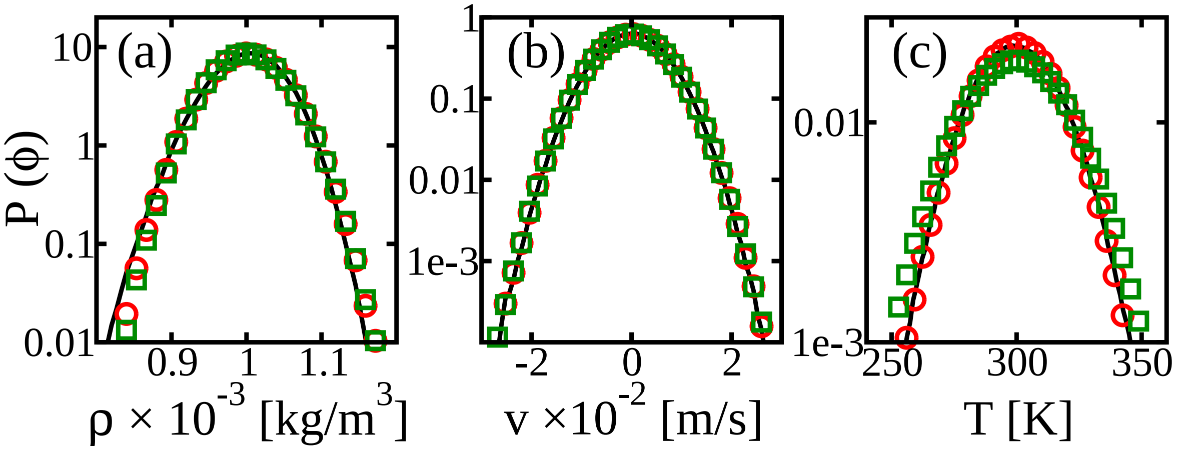


Figure 5.6. Probability distribution functions of (**a**) mass density, (**b**) streaming velocity, and (**c**) temperature of liquid water at 300 K. The streaming velocity is defined in Section 5.4, and the temperature is defined in Eq. 5.32. Results of the all-atom MD simulation are shown as black curves. Results of the linear and non-linear FHD simulations are shown as red circles and green squares, respectively. $\phi = \rho$, v, or T

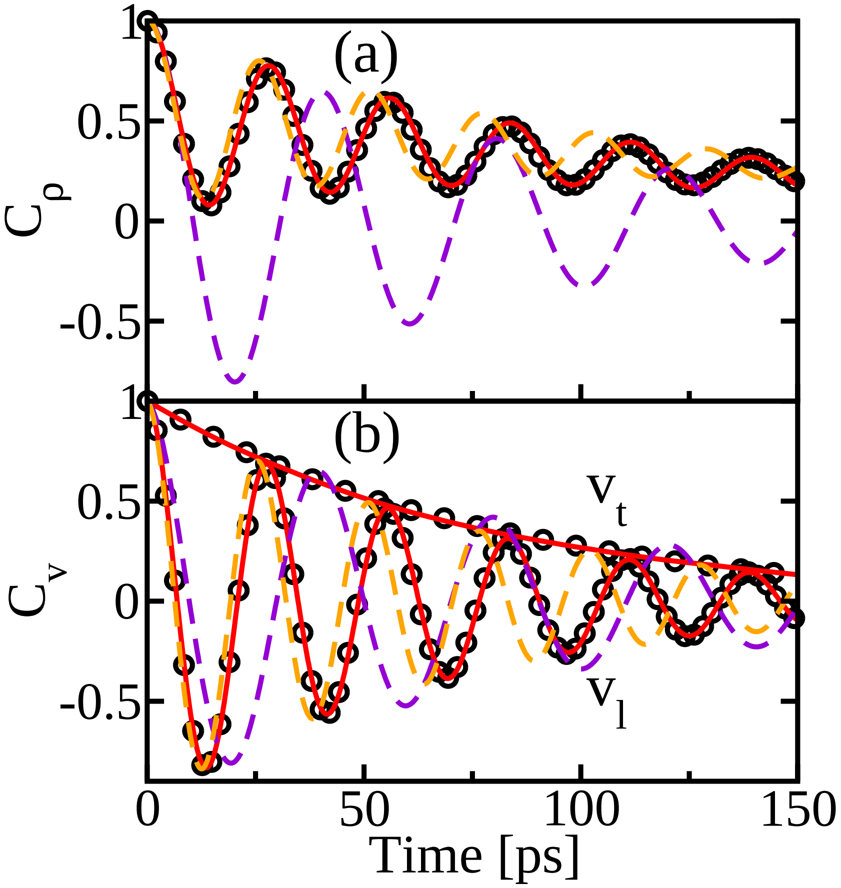


Figure 5.7. Comparison of time correlation functions for liquid argon at 86.5 K between MD and FHD simulations. Time correlation functions of (**a**) mass density and (**b**) transverse v_t and longitudinal v_l velocity are shown. The wave vector is $\boldsymbol{k} = (0, 0, k)$, where $k = 2\pi/240\,\text{Å}^{-1}$. Results of the all-atom MD simulation are shown as black circles. Results of the linear and non-linear FHD simulations with energy transfer are shown as solid red and dotted orange curves, respectively. The dotted purple curve shows the result of FHD simulations without energy transfer.

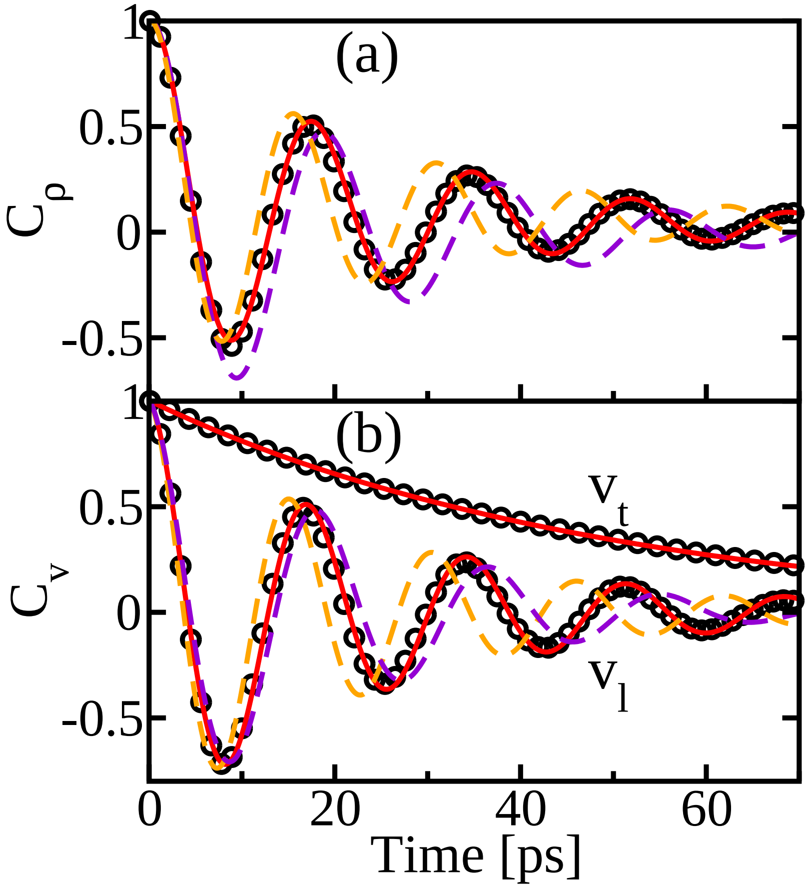


Figure 5.8. Comparison of time correlation functions for liquid water at 300 K between MD and FHD simulations. Time correlation functions of (**a**) mass density and (**b**) transverse v_t and longitudinal v_l velocity are shown. The wave vector is $\boldsymbol{k} = (0, 0, k)$, where $k = 2\pi/240\,\text{Å}^{-1}$. Results of the all-atom MD simulation are shown as black circles. Results of the linear and non-linear FHD simulations with energy transfer are shown as solid red and dotted orange curves, respectively. The dotted purple curve shows the result of FHD simulations without energy transfer.

is within the statistical uncertainties of the MD results presented in Table 5.1, satisfactory agreement between MD and non-linear FHD simulations is achieved for both fluids even without iteration. With the MD-to-FHD mapping, the time scales in the all-atom MD and FHD models can be transparently compared and synchronized, which remains difficult in current particle-based CG models [198,199,202].

In addition, Figs. 5.7 and 5.8 also show the time correlation function obtained by solving the FHD equations without EHT. In this case, the effects of the advection term on the time correlation functions becomes unnoticeable, indicating that the primary contribution of non-linearity lies in the advection of energy directly, rather than momentum through fluid flow. Clear deviation of the results using the simpler model without EHT from the all-atom MD results is observed. The difference is larger for argon due to its higher specific heat ratio (2.1 vs. 1.1 for water), which enhances the dissipative contribution from the thermal diffusivity, as seen in Eq. 5.22. Therefore, to accurately capture the dynamics of a molecular fluid in a CG representation, it is important to include the energy transport equation to capture the coupling between mass and energy transfer. The results show that the FHD equations with energy transfer – along with a self-consistent MD-to-FHD mapping – can be used to coarse-grain a molecular fluid to preserve both the thermodynamic and hydrodynamic responses.

5.6 Discussion

In this work, the FHD equations are shown to accurately capture the thermodynamic responses and hydrodynamic relaxation of molecular fluids, including those associated with energy and heat transfer. The results thus endorse FHD as a promising framework to coarse-grain the liquid phase environment in complex molecular systems. An essential component in connecting MD and FHD is a mapping procedure that converts atomistic positions, masses, velocities, and interaction energies into density fields of mass, momentum, and energy on Eulerian grid cells. For nanometer grids, the finite size of an atom or molecule needs to be explicitly considered to ensure that the variances of density fluctuations scale inversely with grid volume, and that the density fluctuations are consistent with the response functions of the fluid. An important finding is that for different hydrodynamic variables, different molecule sizes are required in order to achieve the correct scaling of the fluctuations with grid size. Remarkably, the mass and energy density fluctuations obey this inverse scaling for grid sizes down to 5 Å, affirming the applicability of FHD modeling at the nanoscale.

By using fluid properties extracted from all-atom MD simulations to simulate the FHD equations, the resulting fluctuations of mass density, streaming velocity, and temperature are shown to quantitatively agree with those observed in all-atom MD simulations. This result indicates that the FHD equations with EHT can be solved accurately with nanometer-scale discretization. Furthermore, dynamical relaxations in FHD simulations also synchronize with those observed in all-atom MD simulations. Therefore, the proposed procedure of converting and matching common observables between MD and FHD simulations is capable of establishing thermodynamic and hydrodynamic compatibility between atomistic and hydrodynamic representations of a fluid. The results of this work indicate that FHD modeling can be used as an accurate and robust framework for coarse-graining the solution phase of complex molecular systems.

5.7 Supplementary information

In this Section, the fluctuation formulas for specific heat capacity (Eq. 5.17) and thermal expansivity (Eq. 5.18) will be derived. Although the relation in Eq. 5.16 is well-known, the latter two are found less commonly in the literature [208,209], and their derivations are harder to find. As a reference, they will be derived in this Section. These derivations assume the reader is familiar with material in standard texts [211,223].

5.7.1 Derivation of specific heat capacity in grand canonical ensemble

Our strategy is to relate the variances in Eq. 5.17 to thermodynamic derivatives. The energy is E, the number of particles is N, the volume is V, the inverse temperature is $\beta = 1/k_\mathrm{B}T$, and the chemical potential is μ. In the following, the brackets indicate expectations taken over the grand canonical ensemble distribution. First, we would like to show

$$\left\langle (\delta E)^2 \right\rangle = -\left(\frac{\partial \langle E\rangle}{\partial \beta}\right)_{\beta\mu,V}. \tag{5.34}$$

The expectation $\langle E\rangle$ is

$$\langle E\rangle = \sum_\nu P_\nu E_\nu = \frac{\sum_\nu E_\nu \exp(-\beta E_\nu + \beta\mu N_\nu)}{\Xi}. \tag{5.35}$$

where P_ν is the probability of microstate ν and Ξ is the grand canonical partition function $\Xi = \sum_\nu \exp(-\beta E_\nu + \beta\mu N_\nu)$. Differentiation of Eq. 5.35 gives

$$\begin{aligned}\left(\frac{\partial \langle E\rangle}{\partial \beta}\right)_{\beta\mu,V} &= \frac{\left(\sum_\nu E_\nu^2 \exp(-\beta E_\nu + \beta\mu N_\nu)\right)\Xi}{\Xi^2} - \frac{\left(\sum_\nu E_\nu \exp(-\beta E_\nu + \beta\mu N_\nu)\right)^2}{\Xi^2},\\ &= \left\langle E^2\right\rangle - \langle E\rangle^2,\\ &= \left\langle (\delta E)^2\right\rangle.\end{aligned} \tag{5.36}$$

Following similar manipulations, we can show

$$\langle \delta N \delta E\rangle = \left(\frac{\partial \langle E\rangle}{\partial \beta\mu}\right)_{\beta,V} = \left(\frac{\partial \langle N\rangle}{\partial \beta}\right)_{\beta\mu,V}, \tag{5.37}$$

$$\left\langle (\delta N)^2\right\rangle = \left(\frac{\partial \langle N\rangle}{\partial \beta\mu}\right)_{\beta,V}. \tag{5.38}$$

The key step now is to use the expansion rule [224] to rewrite $(\partial\langle E\rangle/\partial\beta)_{\beta\mu,V}$. First, consider the energy E as a function of β, N, and V, and write its total derivative,

$$dE = \left(\frac{\partial E}{\partial\beta}\right)_{N,V} d\beta + \left(\frac{\partial E}{\partial N}\right)_{\beta,V} dN + \left(\frac{\partial E}{\partial V}\right)_{\beta,N} dV \tag{5.39}$$

Dividing both sides by $(\partial\beta)_{\beta\mu,V}$, we have

$$\begin{aligned}\left(\frac{\partial E}{\partial\beta}\right)_{\beta\mu,V} &= \left(\frac{\partial E}{\partial\beta}\right)_{N,V} + \left(\frac{\partial E}{\partial N}\right)_{\beta,V}\left(\frac{\partial N}{\partial\beta}\right)_{\beta\mu,V}, \\ &= \left(\frac{\partial E}{\partial\beta}\right)_{N,V} + \left(\frac{\partial E}{\partial\beta\mu}\right)_{\beta,V}\left(\frac{\partial\beta\mu}{\partial N}\right)_{\beta,V}\left(\frac{\partial N}{\partial\beta}\right)_{\beta\mu,V},\end{aligned} \tag{5.40}$$

where we have used the chain rule on the second line. The specific heat c_V is defined as

$$c_V = \frac{1}{M}\left(\frac{\partial E}{\partial T}\right)_{N,V}, \tag{5.41}$$

where M is the total mass of the system. Therefore,

$$\begin{aligned}\left(\frac{\partial E}{\partial\beta}\right)_{N,V} &= \left(\frac{\partial E}{\partial T}\right)_{N,V}\left(\frac{\partial T}{\partial\beta}\right)_{N,V}, \\ &= Mc_V k_{\mathrm{B}}T^2.\end{aligned} \tag{5.42}$$

Plugging Eqs. 5.34, 5.37, 5.38, and 5.42 into Eq.5.40, we obtain

$$\left\langle(\delta E)^2\right\rangle = k_{\mathrm{B}}T^2 Mc_V + \frac{\langle\delta N\delta E\rangle^2}{\left\langle(\delta N)^2\right\rangle}. \tag{5.43}$$

Defining the mass density $\rho = \langle M\rangle/V$ and energy density $u = \langle E\rangle/V$ and rearranging Eq. 5.43, we arrive at Eq. 5.17.

5.7.2 Derivation of thermal expansivity in grand canonical ensemble

Following the similar procedure used to derive Eq. 5.34, we can show

$$\langle\delta N\delta(E-\mu N)\rangle = -\left(\frac{\partial\langle N\rangle}{\partial\beta}\right)_{\mu,V}. \tag{5.44}$$

Let ρ_N be the number density $\langle N\rangle/V$ and the volume per particle $v = 1/\rho_N$. We therefore have

$$-\left(\frac{\partial\langle N\rangle}{\partial\beta}\right)_{\mu,V} = -V\left(\frac{\partial\rho_N}{\partial\beta}\right)_{\mu} = -\langle N\rangle k_{\mathrm{B}}T^2\frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_{\mu}. \tag{5.45}$$

Let us now consider v as a function of T and P and write out its total differential

$$dv = \left(\frac{\partial v}{\partial T}\right)_P dT + \left(\frac{\partial v}{\partial P}\right)_T dP. \tag{5.46}$$

Dividing both sides by $v(dT)_\mu$, we have

$$\frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_\mu = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_P + \frac{1}{v}\left(\frac{\partial v}{\partial P}\right)_T \left(\frac{\partial P}{\partial T}\right)_\mu. \tag{5.47}$$

We will now identify each of the derivatives in Eq. 5.47 above. The thermal expansivity α_P and isothermal compressibility κ_T are defined as

$$\alpha_P = \frac{1}{v}\left(\frac{\partial v}{\partial T}\right)_P, \tag{5.48}$$

$$\kappa_T = -\frac{1}{v}\left(\frac{\partial v}{\partial P}\right)_T. \tag{5.49}$$

The entropy per volume s is

$$s = \frac{\langle E\rangle}{TV} + \frac{P}{T} - \frac{\mu\langle N\rangle}{TV}, \tag{5.50}$$

which from the Gibbs-Duhem equation $dP = s dT + \rho_N d\mu$ is also

$$s = \left(\frac{\partial P}{\partial T}\right)_\mu. \tag{5.51}$$

Plugging in Eqs. 5.44, 5.45, 5.48, 5.49, 5.50, 5.51 into Eq. 5.47, and rearranging, we have

$$\alpha_P = \kappa_T\left(\frac{\langle E\rangle}{TV} + \frac{P}{T} - \frac{\mu\langle N\rangle}{TV}\right) - \frac{1}{\langle N\rangle k_{\mathrm{B}} T^2}\left(\langle \delta N \delta E\rangle - \mu\left\langle (\delta N)^2\right\rangle\right). \tag{5.52}$$

The term $\left\langle (\delta N)^2\right\rangle$ can be related to the compressibility κ_T through the relation

$$\kappa_T = \frac{\left\langle (\delta N)^2\right\rangle V}{\langle N\rangle^2 k_{\mathrm{B}} T}, \tag{5.53}$$

so we can cancel the terms involving $\kappa_T \mu\langle N\rangle/TV$ in Eq. 5.52 and arrive at

$$\alpha_P = \kappa_T\left(\frac{\langle E\rangle}{TV} + \frac{P}{T}\right) - \frac{\langle \delta N \delta E\rangle}{\langle N\rangle k_{\mathrm{B}} T^2}. \tag{5.54}$$

We now break down the total internal energy E into the potential energy U and kinetic energy K,

$$\alpha_P = \kappa_T\left(\frac{\langle U\rangle}{TV} + \frac{\langle K\rangle}{TV} + \frac{P}{T}\right) - \frac{\langle \delta N \delta U\rangle}{\langle N\rangle k_{\mathrm{B}} T^2} - \frac{\langle \delta N \delta K\rangle}{\langle N\rangle k_{\mathrm{B}} T^2}. \tag{5.55}$$

The fluctuation $\langle \delta N \delta K\rangle = \langle NK\rangle - \langle N\rangle\langle K\rangle$ in Eq. 5.55 will now be further manipulated. The kinetic energy is $K = \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i$. For classical fluids with position $\boldsymbol{r}_i$ and momentum $\boldsymbol{p}_i$ degrees of freedom, the quantity $\langle NK\rangle$ is

$$\langle NK\rangle = \frac{\sum_N N e^{\beta\mu N} \int d\boldsymbol{r}_{1\ldots N} e^{-\beta U(\boldsymbol{r}_1\ldots\boldsymbol{r}_N)} \int d\boldsymbol{p}_{1\ldots N} \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i \exp\Big(-\beta \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i\Big)}{\sum_N e^{\beta\mu N} \int d\boldsymbol{r}_{1\ldots N} e^{-\beta U(\boldsymbol{r}_1\ldots\boldsymbol{r}_N)} \int d\boldsymbol{p}_{1\ldots N} \exp\Big(-\beta \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i\Big)}. \tag{5.56}$$

Let us consider the quantity

$$\begin{aligned} &\int d\boldsymbol{p}_{1\ldots N} \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i \exp\left(-\beta \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i\right), \\ =&\int d\boldsymbol{p}_{1\ldots N} \left(\frac{\boldsymbol{p}_1^2}{2m_1} e^{-\beta\sum_i \boldsymbol{p}_i^2/2m_i} + \cdots + \frac{\boldsymbol{p}_N^2}{2m_N} e^{-\beta\sum_i \boldsymbol{p}_i^2/2m_i}\right). \end{aligned} \tag{5.57}$$

This expression is a sum of integrals. To begin, let us evaluate the integral for particle 1

$$\begin{aligned} &\int d\boldsymbol{p}_{1\ldots N} \frac{\boldsymbol{p}_1^2}{2m_1} e^{-\beta\sum_i \boldsymbol{p}_i^2/2m_i}, \\ =&\int d\boldsymbol{p}_1 \frac{\boldsymbol{p}_1^2}{2m_1} e^{-\beta \boldsymbol{p}_1^2/2m_1} \int d\boldsymbol{p}_{2\ldots N} e^{-\beta\sum_{i=2}^{N} \boldsymbol{p}_i^2/2m_i}, \\ =&\frac{\int d\boldsymbol{p}_1 (\boldsymbol{p}_1^2/2m_1) e^{-\beta \boldsymbol{p}_1^2/2m_1}}{\int d\boldsymbol{p}_1 e^{-\beta \boldsymbol{p}_1^2/2m_1}} \int d\boldsymbol{p}_{1\ldots N} e^{-\beta\sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i}, \\ =&\frac{3}{2} k_{\mathrm{B}} T \int d\boldsymbol{p}_{1\ldots N} e^{-\beta\sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i}. \end{aligned} \tag{5.58}$$

In Eq. 5.57, we have N terms identical to the expression in Eq. 5.58. Therefore, Eq. 5.56 becomes

$$\begin{aligned} \langle NK\rangle &= \frac{\frac{3}{2} k_{\mathrm{B}} T \sum_N N^2 e^{\beta\mu N} \int d\boldsymbol{r}_{1\ldots N} e^{-\beta U(\boldsymbol{r}_1\ldots\boldsymbol{r}_N)} \int d\boldsymbol{p}_{1\ldots N} \exp\Big(-\beta \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i\Big)}{\sum_N e^{\beta\mu N} \int d\boldsymbol{r}_{1\ldots N} e^{-\beta U(\boldsymbol{r}_1\ldots\boldsymbol{r}_N)} \int d\boldsymbol{p}_{1\ldots N} \exp\Big(-\beta \sum_{i=1}^{N} \boldsymbol{p}_i^2/2m_i\Big)}, \\ &= \langle NN\rangle \frac{3}{2} k_{\mathrm{B}} T. \end{aligned} \tag{5.59}$$

Likewise, following the same procedure as above, we have

$$\langle K\rangle = \frac{3}{2} \langle N\rangle k_{\mathrm{B}} T. \tag{5.60}$$

For the fluctuation $\langle \delta N \delta K\rangle$, we have, using Eqs. 5.59 and 5.60,

$$\begin{aligned} \langle \delta N \delta K\rangle &= \langle NK\rangle - \langle N\rangle\langle K\rangle, \\ &= \big\langle (\delta N)^2 \big\rangle \frac{3}{2} k_{\mathrm{B}} T. \end{aligned} \tag{5.61}$$

Plugging Eqs. 5.60 and 5.61 into Eq. 5.55, we have

$$\alpha_P = \frac{\kappa_T}{T}\left(\frac{\langle U\rangle}{V} + P\right) + \kappa_T \frac{\frac{3}{2}\langle N\rangle k_\mathrm{B}}{V} - \frac{\langle \delta N \delta U\rangle}{\langle N\rangle k_\mathrm{B} T^2} - \frac{\left\langle (\delta N)^2\right\rangle \frac{3}{2} k_\mathrm{B} T}{\langle N\rangle k_\mathrm{B} T^2}. \tag{5.62}$$

Using the isothermal compressibility relation in Eq. 5.53, we can simplify further by canceling the second and fourth terms

$$\alpha_P = \frac{\kappa_T}{T}\left(\frac{\langle U\rangle}{V} + P\right) - \frac{\langle \delta N \delta U\rangle}{\langle N\rangle k_\mathrm{B} T^2}. \tag{5.63}$$

Defining the potential energy density as $\varepsilon = \langle U\rangle/V$, we arrive at Eq. 5.18.

Chapter 6

Developing an Interfacial Fluctuating Hydrodynamics Model from Molecular Dynamics

The fluctuating hydrodynamic model developed in Chapter 5 is extended in this Chapter to describe the physics of phase separation and equilibria within the solvent. We introduce a multiscale framework to simulate inhomogeneous fluids by coarse-graining an all-atom molecular dynamics trajectory onto sequential snapshots of hydrodynamic fields. We show that the field representation of an atomistic trajectory is quantitatively described by a dynamic field-theoretic model that couples hydrodynamic fluctuations with a Ginzburg-Landau free energy. For liquid-vapor interfaces of argon and water, the parameters of the field model can be adjusted to reproduce the bulk compressibility and surface tension calculated from the positions and forces of atoms in an MD simulation. These optimized parameters also enable the field model to reproduce the static and dynamic capillary wave spectra calculated from atomistic coordinates at the liquid-vapor interface. In addition, we show that a density-dependent gradient coefficient in the Ginzburg-Landau free energy enables bulk and interfacial fluctuations to be controlled separately. For water, this additional degree of freedom is necessary to capture both the bulk compressibility and surface tension emergent from the atomistic trajectory. The proposed multiscale framework illustrates that bottom-up coarse-graining and top-down phenomenology can be integrated with quantitative consistency to simulate the interfacial fluctuations in nanoscale transport processes.

Chapter 6 is reprinted with permission from Shang, B. Z., Voulgarakis, N. K. & Chu, J. W. Fluctuating hydrodynamics for multiscale simulation of inhomogeneous fluids: mapping all-atom molecular dynamics to capillary waves. *J. Chem. Phys.* **135,** 044111 (2011).
Copyright 2011, American Institute of Physics.

6.1 Motivation

In biology and nanotechnology, emergent phenomena in the nanometer to micrometer regimes – such as hydrophobic self-assembly [173], dynamics of Brownian motors [179], and breakup of fluid nanojets [176] – are governed by the interplay between molecular interactions and hydrodynamic fluctuations [172,174,180]. The analysis of these processes via computer simulation inevitably requires a multiscale approach to link molecular and continuum mechanics [180]. Although all-atom molecular dynamics (MD) simulations can provide a detailed description of the interactions among individual atoms, the range of accessible length- and time-scales is severely limited. While particle-based coarse-grained (CG) methods [200,225,226] have been developed to reduce the computational cost of atomistic simulations and to capture hydrodynamic behaviors approximately [184,227], it remains difficult to derive formal equations of motion for these CG methods and establish their consistency with atomistic simulations [202,228–230]. Consequently, most adopted approaches employ phenomenological Langevin-type models [202,225,228]. Meanwhile, continuum and field-theoretic models have found only limited application in CG methods [195], yet they conveniently describe the thermodynamic and transport behaviors of a fluid at the hydrodynamic level. Therefore, an inevitable mapping in multiscale simulation is the coarse graining of all-atom molecular dynamics onto hydrodynamics.

To map atomistic coordinates onto hydrodynamic fields, we recently introduced a strategy [195,203–205] to coarse-grain the positions and velocities of atoms onto hydrodynamic field variables in the spirit of the Irving-Kirkwood procedure for identifying expressions of thermodynamic forces and fluxes in terms of molecular degrees of freedom [231,232]. Further details are also provided in Chapter 5. From the field representation of atomistic coordinates, we determine the equation of state and transport properties of the fluid by comparing the spectra of density and momentum field fluctuations calculated from all-atom MD with the spectra predicted by solving the equations of fluctuating hydrodynamics (FHD) [172,206,233]. The FHD equations introduced by Landau and Lifshitz [181] are employed to describe thermal fluctuations at the hydrodynamic level by adding stochastic stress to the equation for momentum conservation. Mapping atomistic positions and velocities onto field variables and computing the transport and thermodynamic properties from the spectra of field fluctuations enables a systematic and self-consistent coarse-graining of a molecular fluid into a hydrodynamic description.

We have shown previously in Chapter 5 that mapping atomistic coordinates onto Eulerian grids with nanometer size is non-trivial, and that an additional length-scale, namely the excluded volume of the molecule, must be considered to reproduce density field fluctuations with size scaling consistent with statistical mechanics [195,205]. By considering the additional molecular length-scale, we showed that density and momentum field fluctuations in all-atom MD simulations can be matched to those obtained from solving the FHD equations on Eulerian grids [195,205]. The emergent viscoelastic response of molecular fluids at the nanoscale can also be captured quantitatively by the equations of generalized fluctuating hydrodynamics [203,204]. For both argon and water, we found that a hydrodynamic description using the FHD equations remains valid with a grid size as small as 5 Å, affirming their applicability to modeling nanoscale transport processes [195,205].

However, the conventional FHD equations only describe the spectra of fluctuations for bulk, homogeneous fluids [181,195,205], while nanoscale transport processes ubiquitously involve phenomena occurring at interfaces [173,180,234]. For example, material self-assembly depends upon interfacial thermodynamics and kinetics to drive the separation and partition of individual molecular components [235]. To describe the dynamics of proteins and the self-assembly of hydrophobic solutes, it is necessary to consider inhomogeneities within the fluid induced by interfaces [173,180]. Therefore, field-theoretic coarse-graining must go beyond homogeneous fluids and capture the behaviors of surface tension and phase separation from molecular interactions and dynamics.

In this work, we generalize our multiscale framework to inhomogeneous fluids by coupling the FHD equations with a Ginzburg-Landau (GL) free energy functional [236–238], which introduces surface tension and bending rigidity into the hydrodynamic model. The Ginzburg-Landau free energy has been used to study a wide range of interfacial problems including critical phenomena, spinodal decomposition, and nucleation [237]. Surface tension is introduced in the model by expanding the free energy in terms of density gradients. We validate the combined FHD-GL equations by simulating the liquid-vapor and liquid-liquid interfaces of model inhomogeneous systems. We show that the fluctuation-dissipation theorem is satisfied and that the computed capillary wave (CW) spectra agree quantitatively with the predictions of CW theory [239]. To establish consistency with all-atom MD simulations, we present a systematic method to determine the Ginzburg-Landau free energy functional of argon and water by matching the hydrodynamic fluctuations calculated from an all-atom MD trajectory of a liquid-vapor interface with those observed in the FHD-GL field simulations. We show that the field simulations can reproduce the bulk compressibility, surface tension, and static and dynamic capillary wave spectra extracted from the positions, velocities, and forces of atoms recorded in a MD trajectory. By using a density-dependent gradient coefficient in the Ginzburg-Landau free energy, density fluctuations in the bulk can be decoupled from height fluctuations at the interface, allowing the field model to explore a wider range of fluid behaviors. The results in this work thus illustrate that the attractive and unbalanced molecular interactions at interfaces in an all-atom MD simulation can be mapped effectively and self-consistently onto the coupling between hydrodynamic fields.

The rest of this Chapter is organized as follows. In Section 6.2, we describe the governing FHD-GL equations. In Section 6.3, we validate our equations for liquid-vapor and liquid-liquid interfaces of model systems. In Section 6.4, we determine the Ginzburg-Landau free energy for argon and water and establish consistency between hydrodynamic fluctuations observed in field and MD simulations. In Section 6.5, we provide concluding remarks and directions for future work.

6.2 Coupled fluctuating hydrodynamics and Ginzburg-Landau model

We describe the time evolution of density and momentum fields using the equations of fluctuating hydrodynamics (FHD) [181]. The mass and momentum balances are:

$$\frac{\partial \rho}{\partial t} = - \nabla \cdot (\rho \boldsymbol{v}), \tag{6.1}$$

$$\frac{\partial \boldsymbol{g}}{\partial t} = - \nabla \cdot (\boldsymbol{g} \otimes \boldsymbol{v}) + \nabla \cdot (\mathbf{P}) - \nabla \cdot (\mathbf{\Pi} + \boldsymbol{\delta\Pi}), \tag{6.2}$$

In Eqs. 6.1 and 6.2, ρ is the mass density, $\boldsymbol{g} \equiv \rho \boldsymbol{v}$ is the momentum density, and $\boldsymbol{v}$ is the fluid velocity. The reversible, dissipative, and fluctuating stress tensors are $\mathbf{P}$, $\mathbf{\Pi}$, and $\boldsymbol{\delta\Pi}$, respectively. For a homogeneous fluid, $\mathbf{P} = -p_0 \mathbf{I}$, where p_0 is the thermodynamic pressure and $\mathbf{I}$ is the unit tensor. For inhomogeneous fluids, $\mathbf{P}$ can be derived from the free energy functional, as will be explained later. In this work, we consider the Newtonian dissipative stress,

$$\mathbf{\Pi} = -\eta_{\mathrm{S}} \left(\nabla \boldsymbol{v} + \nabla \boldsymbol{v}^{\mathrm{T}} \right) - \left(\eta_{\mathrm{B}} - \frac{2}{3} \eta_{\mathrm{S}} \right) (\nabla \cdot \boldsymbol{v}) \mathbf{I}. \tag{6.3}$$

The fluctuating counterpart $\boldsymbol{\delta\Pi}$ is modeled as Gaussian white noise with zero mean and covariance determined by the fluctuation-dissipation theorem [207,240],

$$\left\langle \delta\Pi^{ij}_{\boldsymbol{n}}(t) \delta\Pi^{kl}_{\boldsymbol{n}'}(t') \right\rangle = \frac{2 k_{\mathrm{B}} T}{V_{\mathrm{G}}} \left[\eta_{\mathrm{S}} (\delta_{il} \delta_{jk} + \delta_{ik} \delta_{jl}) + \left(\eta_{\mathrm{B}} - \frac{2}{3} \eta_{\mathrm{S}} \right) \delta_{ij} \delta_{kl} \right] \delta(t - t') \delta(\boldsymbol{n} - \boldsymbol{n}'). \tag{6.4}$$

Here, $\delta\Pi^{ij}_{\boldsymbol{n}}(t)$ is the ij component of the stress tensor $\boldsymbol{\delta\Pi}$ associated with a grid of index vector $\boldsymbol{n}$ at time t, k_{B} is Boltzmann's constant, T is the fluid temperature, V_{G} is the volume of a grid in which the density and momentum fields are defined, η_{S} is the shear viscosity, and η_{B} is the bulk viscosity.

6.2.1 Two-phase fluid

For an isothermal two-phase fluid, we describe the total Helmholtz free energy (or effective Hamiltonian) $F[\rho(\boldsymbol{r})]$ by the Ginzburg-Landau functional,

$$F[\rho(\boldsymbol{r})] = \int \left(\psi_0(\rho) + \frac{m}{2} |\nabla \rho|^2 + \frac{\alpha}{2} \left(\nabla^2 \rho \right)^2 \right) d^3 r, \tag{6.5}$$

where ψ_0 is the local free energy density, m is the density gradient coefficient, and α is the density curvature coefficient. From the least-action principle [241], the reversible stress tensor is related to the free energy by $\nabla \cdot \mathbf{P} = -\rho \nabla (\delta F / \delta \rho)$, where $\mathbf{P} = -p_0 \mathbf{I} + \mathbf{P}_m + \mathbf{P}_\alpha$, and

$$\mathbf{P}_m = \left[m\rho\nabla^2\rho + \frac{(m+m'\rho)|\nabla\rho|^2}{2} \right]\mathbf{I} - m\nabla\rho\otimes\nabla\rho, \tag{6.6}$$

$$\begin{aligned}\mathbf{P}_\alpha = &\left[-\rho\nabla^2\big(\alpha\nabla^2\rho\big) - \frac{(\alpha+\alpha'\rho)\big(\nabla^2\rho\big)^2}{2} - \nabla\big(\alpha\nabla^2\rho\big)\cdot\nabla\rho \right]\mathbf{I} \\ &+ \nabla\rho\otimes\nabla\big(\alpha\nabla^2\rho\big) + \nabla\big(\alpha\nabla^2\rho\big)\otimes\nabla\rho.\end{aligned} \tag{6.7}$$

The primed variables denote differentiation with respect to density ρ. In Section 6.6.1, we describe in detail the derivation of Eqs 6.6 and 6.7. The thermodynamic pressure is related to the free energy functional by $p_0 = \mu_0\rho - \psi_0$, and $\mu_0 \equiv \partial\psi_0/\partial\rho$ is the local chemical potential. We model the local free energy density as a double-well potential $\psi_0(\rho) = A(\rho-\rho_l)^2(\rho-\rho_\mathrm{v})^2/2$, with minima at liquid ρ_l and vapor ρ_v densities. The coefficient A is a measure of the barrier height $\psi_{0,\max}$ separating the two phases through the relation $\psi_{0,\max} = A(\Delta\rho/2)^4/2$, where $\Delta\rho = \rho_l - \rho_\mathrm{v}$.

6.2.2 Binary fluid

For a binary fluid, we have the additional equation for component mass conservation [172],

$$\frac{\partial(\rho c)}{\partial t} + \nabla\cdot(\rho c\boldsymbol{v}) = -\nabla\cdot(\boldsymbol{j} + \boldsymbol{\delta j}), \tag{6.8}$$

where $c \equiv c_2 - c_1$ is the mass fraction difference between components 1 and 2, $\boldsymbol{j} = \boldsymbol{j}_2 - \boldsymbol{j}_1 = 2\boldsymbol{j}_2$ is the diffusive component mass flux, and $\boldsymbol{\delta j}$ is the component mass flux due to thermal fluctuations. The diffusive flux $\boldsymbol{j}$ is given by the constitutive equation, $\boldsymbol{j} = -\lambda\nabla\mu$, where λ is the mobility coefficient, and $\mu \equiv \mu_2 - \mu_1$ is the generalized chemical potential difference. The fluctuating component mass flux $\boldsymbol{\delta j}$ is modeled as Gaussian white noise with zero mean and covariance determined by the fluctuation-dissipation theorem [242],

$$\left\langle \delta j^i_{\boldsymbol{n}}(t)\delta j^j_{\boldsymbol{n}'}(t') \right\rangle = \frac{2\lambda k_\mathrm{B}T}{V_\mathrm{G}}\delta_{ij}\delta(t-t')\delta(\boldsymbol{n}-\boldsymbol{n}'). \tag{6.9}$$

We model the total Helmholtz free energy $F[\rho(\boldsymbol{r}), c(\boldsymbol{r})]$ of a binary fluid by

$$F[\rho(\boldsymbol{r}), c(\boldsymbol{r})] = \int \left(\psi_0(\rho, c) + \frac{m}{2}|\nabla\rho|^2 + \frac{\alpha}{2}\big(\nabla^2\rho\big)^2 + \frac{k}{2}|\nabla c|^2 + \frac{\beta}{2}\big(\nabla^2 c\big)^2 \right) d^3r, \tag{6.10}$$

where k is the concentration gradient coefficient and β is the concentration curvature coefficient. From the least-action principle [241,243], we have

$$\nabla\cdot\mathbf{P} = -\rho\nabla(\delta F/\delta\rho)_c + (\delta F/\delta c)_\rho\nabla c, \tag{6.11}$$

where $\mathbf{P} = -p_0\mathbf{I} + \mathbf{P}_m + \mathbf{P}_\alpha + \mathbf{P}_k + \mathbf{P}_\beta$, and

$$\mathbf{P}_k = \left[\frac{k}{2}|\nabla c|^2\right]\mathbf{I} - k\nabla c \otimes \nabla c, \tag{6.12}$$

$$\begin{aligned}\mathbf{P}_\beta = &\left[-\frac{\beta}{2}\left(\nabla^2 c\right)^2 - \nabla\left(\beta\nabla^2 c\right)\cdot\nabla c\right]\mathbf{I} \\ &+ \nabla c \otimes \nabla\left(\beta\nabla^2 c\right) + \nabla\left(\beta\nabla^2 c\right)\otimes\nabla c.\end{aligned} \tag{6.13}$$

Similar to the case of two-phase fluids, $p_0 = \rho g_0 - \psi_0$, where $g_0 \equiv \left(\partial\psi_0/\partial\rho\right)_\mathrm{c}$ is the local specific Gibbs free energy. The generalized chemical potential difference is defined by $\mu = \left(\delta F/\delta c\right)_\rho/\rho$. The local free energy density is a double-well potential

$$\psi_0(\rho, c) = A(\rho-\rho_l)^2(\rho-\rho_v)^2/2 + B\left(c - c_1^{\min}\right)^2\left(c - c_2^{\min}\right)^2/2, \tag{6.14}$$

with minima at ρ_l, ρ_v, $c_1^{\min}$, and $c_2^{\min}$. The double-well potential requires the concentration in each cell to remain near the potential minima, while the gradient and curvature coefficients constrain the neighboring cells to have similar concentrations, giving rise to phase separation. The governing FHD-GL equations describe compressible, viscous, isothermal flow of an inhomogeneous fluid, and are similar to the conventional Model H [236] and diffuse-interface [237] equations used to investigate dynamic critical phenomena and two-phase flows. We solve these equations using a staggered finite volume scheme in space [195,196], and second-order Runge-Kutta integration in time. In the staggered scheme, the system volume is discretized into an array of cubic grids. The mass density fields are placed at grid centers, while the momentum density fields are placed at grid faces. We found that the staggered scheme is advantageous because it prevents unphysical non-local transport of mass, which may occur in conventional non-staggered schemes [195]. Compared to all-atom MD simulations of water, the FHD-GL field simulations using a grid size of 5 Å yield a total speed up of 50 times and allows an increase in time step by at least a factor of two. Furthermore, the computational cost of FHD simulations scales linearly with system size.

6.3 Liquid-vapor and liquid-liquid interfaces of model systems

To validate the FHD-GL approach for modeling multiphase fluids, we perform simulations of fluctuating liquid-vapor and liquid-liquid interfaces in model systems. We explore four distinct cases, and the fluid parameters for each are listed in Table 6.1. Models A1 and A2 are two-phase fluids, while Models B1 and B2 are binary fluids. For Models A1 and B1, the curvature coefficients are set to zero, while for Models A2 and B2, the gradient coefficients are set to zero. This selection allows us to investigate separately the effects of gradients and curvature on the spectra of interfacial fluctuations. Simulations for the model fluids were performed for a system of volume 240 Å × 240 Å × 240 Å, replicated periodically

Table 6.1. Model fluid parameters of Ginzburg-Landau free energy. The grid size d used for solving the FHD equations is 5 Å, the density ratio is $\rho_\mathrm{v}/\rho_l = 1/3$, the temperature T is 300 K, and k_B is Boltzmann's constant. For all model systems, the shear and bulk viscosities are $\eta_\mathrm{S} = 4.0\,(\rho_l k_\mathrm{B} T/d)^{1/2}$ and $\eta_\mathrm{B} = 8.0\,(\rho_l k_\mathrm{B} T/d)^{1/2}$, respectively. For binary fluids, the mass fraction differences are $c_1^\mathrm{min} = -1$, $c_2^\mathrm{min} = 1$, and the mobility is $\lambda = 18.0\,(\rho_l^3 d^5/k_\mathrm{B}T)^{-1/2}$. Units of parameters listed in the table are as follows: $A\,[k_\mathrm{B}T/\rho_l^4 d^3]$, $B\,[k_\mathrm{B}T/d^3]$, $m\,[k_\mathrm{B}T/\rho_l^2 d]$, $k\,[k_\mathrm{B}T/d]$, $\alpha\,[k_\mathrm{B}Td/\rho_l^2]]$, and $\beta\,[k_\mathrm{B}Td]$.

Model	A	B	m	k	α	β
A1	406	—	90.3	—	0.0	—
A2	406	—	0.0	—	72.2	—
B1	406	5.01	90.3	10.0	0.0	0.0
B2	406	5.01	0.0	0.0	72.2	8.02

in all directions and discretized by cubic grids of length 5 Å. We initially prepare a slab of thickness 120 Å of either liquid or a pure component and create two independent fluid surfaces. We use a time step of 5 fs and perform field simulations for 10 ns, using the last 9 ns for analysis. In Fig. 6.1a, we show a snapshot for system A1, and in Fig. 6.1b, we plot its kinetic temperature. The averaged temperature is 300.03 K, identical to the input temperature 300.0 K in the fluctuating stress, indicating that the fluctuation-dissipation theorem is numerically satisfied in our solution of the FHD-GL equations. Results for the binary fluid models are similar.

To determine if the field model samples interfacial fluctuations consistent with an equilibrium distribution, we compute the capillary wave (CW) spectra for each of the model systems and compare with the predictions of mean-field [238] and CW [239] theories. For a planar liquid-vapor interface, mean-field theory (MF) gives

$$\sigma_\mathrm{MF} = \int m\left(\frac{d\rho_\mathrm{MF}}{dz}\right)^2 + 2\alpha\left(\frac{d^2\rho_\mathrm{MF}}{dz^2}\right)^2 dz, \tag{6.15}$$

and

$$\kappa_\mathrm{MF} = \int \alpha\left(\frac{d\rho_\mathrm{MF}}{dz}\right)^2 dz, \tag{6.16}$$

where σ_MF is the macroscopic surface tension, κ_MF is the bending rigidity, and $\rho_\mathrm{MF} = \rho_\mathrm{MF}(z)$ is the density profile that minimizes $F[\rho(z)]$. For a liquid-liquid interface, σ_MF and κ_MF follow the form of Eqs. 6.15 and 6.16 after making the substitution $\rho_\mathrm{MF} \to c_\mathrm{MF}$, where $c_\mathrm{MF} = c_\mathrm{MF}(z)$ is the concentration profile that minimizes $F[\rho_l, c(z)]$. In the absence of bending rigidity, the intrinsic mean-field density profile $\rho_\mathrm{MF}(z)$ for a liquid-vapor interface described by a

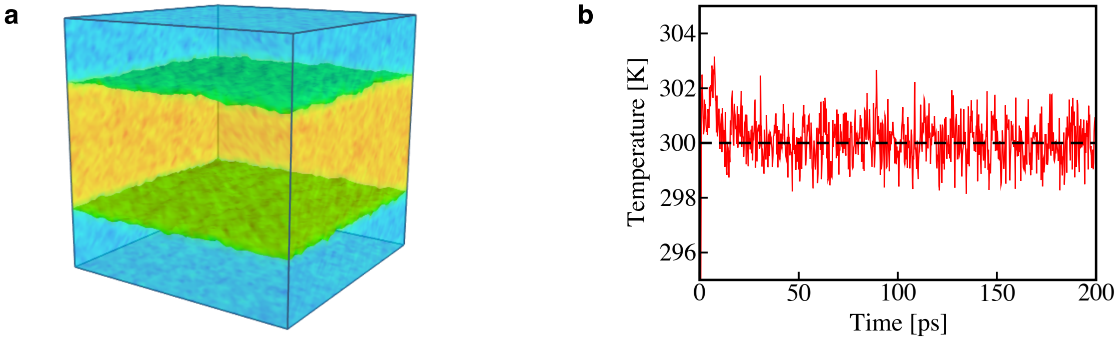


Figure 6.1. Illustration of snapshot from an interfacial FHD simulation and temperature equilibration. (**a**) A snapshot of the model fluid at liquid-vapor coexistence. (**b**) Kinetic temperature, $\hat{T}$, (solid line) as a function of time. $\hat{T}$ is calculated as $(1/2)\sum_{\boldsymbol{n}} \rho_{\boldsymbol{n}} \boldsymbol{v}_{\boldsymbol{n}}^2 = (3/2)Nk_{\mathrm{B}}\hat{T}$, where $\boldsymbol{n}$ denotes the grid index, and N is the number of grids. The input temperature T appearing in Eq. 6.4 is shown as the dashed line.

double-well potential is given by [244],

$$\rho_{\mathrm{MF}}(z) = \frac{(\rho_l + \rho_{\mathrm{v}})}{2} + (\rho_l - \rho_{\mathrm{v}})\frac{\tanh(z/2\xi)}{2}, \tag{6.17}$$

where the interfacial width is

$$\xi = 2^{-5/2}(\rho_l - \rho_{\mathrm{v}})\sqrt{m/\psi_{0,\mathrm{max}}}, \tag{6.18}$$

and the surface tension is

$$\sigma_{\mathrm{MF}} = (2/3)(\rho_l - \rho_{\mathrm{v}})\sqrt{2m\psi_{0,\mathrm{max}}}. \tag{6.19}$$

In the presence of bending rigidity, we find the mean-field density profile numerically, from which the surface tension and bending rigidity are calculated by quadrature. Capillary wave theory provides an alternate route to obtain the surface tension. For a smooth, nearly flat interface described by a height function $h(\boldsymbol{r}_\ell)$, where $\boldsymbol{r}_\ell = (x, y)$ is the lateral coordinate, CW theory predicts the spectrum [244],

$$\left\langle |h(\boldsymbol{q})|^2 \right\rangle = \frac{k_{\mathrm{B}}T}{L^2 \gamma_{\mathrm{CW}}(q) q^2}. \tag{6.20}$$

In Eq. 6.20, $L = L_x = L_y$ is the lateral system length, and

$$h(\boldsymbol{q}) = \frac{1}{L^2}\int h(\boldsymbol{r}_\ell)\exp(i\boldsymbol{q}\cdot\boldsymbol{r}_\ell)d^2\boldsymbol{r}_\ell. \tag{6.21}$$

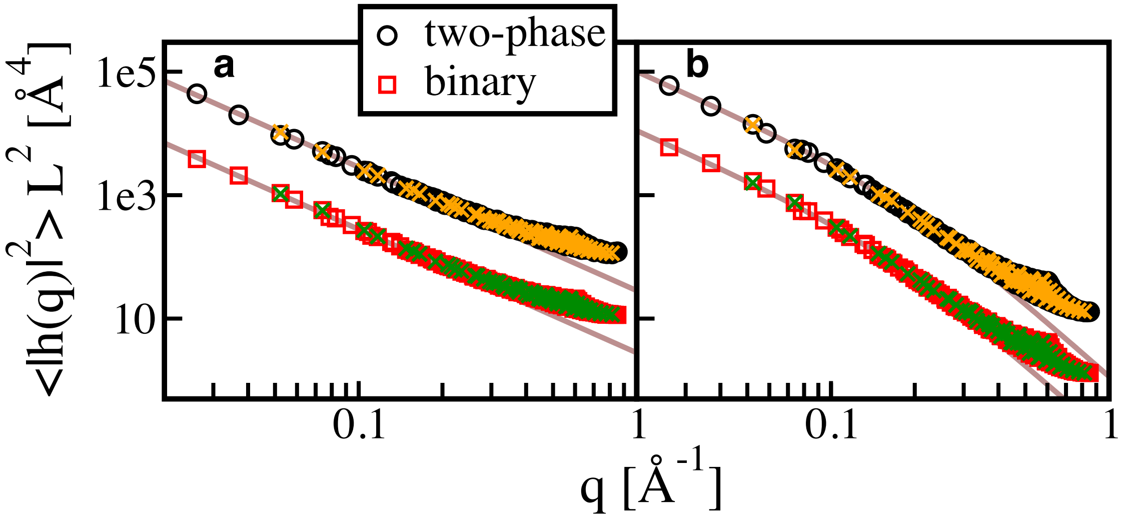


Figure 6.2. Capillary wave spectra of the model fluids used in FHD simulations. (**a**) Models A1 and B1, and (**b**) Models A2 and B2. Two-phase fluids are shown as black circles and binary fluids as red squares. Solid lines correspond to Eq. 6.20, which is used to best fit σ_{CW} and κ_{CW} values. Data for two-phase fluids are multiplied by ten for clarity. Black and red symbols correspond to systems with lateral size $L = 240\,\text{Å}$. Orange and green crosses indicate results for $L = 120\,\text{Å}$.

is the Fourier transform of $h(\boldsymbol{r}_\ell)$. Also, the wave vector-dependent surface tension is

$$\gamma_{\text{CW}}(q) = \sigma_{\text{CW}} + \kappa_{\text{CW}} q^2. \tag{6.22}$$

In the limit of small fluctuations about the interface, one expects $\sigma_{\text{MF}} = \sigma_{\text{CW}}$ and $\kappa_{\text{MF}} = \kappa_{\text{CW}}$. From the results of simulations, we identify the height of the interface $h(\boldsymbol{r}_\ell)$ as an isosurface of density $\rho_{\text{s}} = (\rho_\ell + \rho_{\text{v}})/2$, or concentration $c_{\text{s}} = \left(c_1^{\text{min}} + c_2^{\text{max}}\right)/2$, from which $h(\boldsymbol{q})$ is found by discrete Fourier transform.

In Fig. 6.2a, we show the CW spectra for systems A1 and B1, and in Fig. 6.2b, the spectra for systems A2 and B2, along with best-fit curves. For all model systems, the CW spectra obtained from the field simulations agree closely with the predictions by CW theory at wavelengths larger than about five times the grid size, providing evidence that in the hydrodynamic limit, the field model samples interfacial fluctuations in accordance with an equilibrium distribution. At smaller wavelengths, deviations from the theory emerge due to compressibility of the fluid [245]. From the best-fit values for σ_{CW} and κ_{CW}, we find that $\sigma_{\text{CW}} \simeq \sigma_{\text{MF}}$ and $\kappa_{\text{CW}} \simeq \kappa_{\text{MF}}$, as shown in Table 6.2. The mean-field approximation provides an accurate estimate for $\gamma_{\text{CW}}(q)$, indicating that thermal fluctuations in our model systems do not significantly perturb the mean-field density profile [238,239]. In Fig. 6.2, we also show the CW spectra for a lateral system size $L = 120\,\text{Å}$, illustrating that the scaling of

Table 6.2. The capillary wave (CW) and mean-field (MF) surface tension $\sigma\,[k_\mathrm{B}T/\mathrm{nm}^2]$ and bending rigidity $\kappa\,[k_\mathrm{B}T]$ of the model fluid interfaces.

Model	σ_CW	σ_MF	κ_CW	κ_MF
A1	35 ± 2	37	—	—
A2	25 ± 2	24	8 ± 1	8.4
B1	36 ± 1	37	—	—
B2	22 ± 2	24	8 ± 1	8.4

the spectra with the projected area L^2, as predicted by Eq. 6.20, is achieved in our field simulations.

6.4 Consistency between FHD and MD simulations for argon and water

We have shown that the field FHD simulations for model systems satisfy the fluctuation-dissipation theorem and sample the spectrum of interfacial fluctuations in accordance with CW theory. We now explore to what extent the field model captures the fluctuations and correlations observed in a field representation of an all-atom MD simulation. In order to transparently compare field and atomistic simulations, we recently developed a scale-consistent mapping scheme to coarse-grain atomistic configurations sampled in a MD trajectory onto sequential snapshots of field coordinates. This mapping enables the comparison of atomistic and field-theoretic models under a unified and self-consistent framework [195,203,204]. In the following, we present a method to determine the Ginzburg-Landau free energy parameters for argon and water by comparing field trajectories obtained from MD and field simulations. We show that a relatively simple Ginzburg-Landau model can reproduce the bulk compressibility, surface tension, and static and dynamic CW spectra observed in MD simulations of argon and water.

The main parameters of the Ginzburg-Landau free energy are the double-well barrier height $\psi_{0,\mathrm{max}}$ and the gradient coefficient m. We set the curvature coefficient α to zero, since the bending rigidity κ is nearly zero, as will be inferred from the capillary wave spectra of argon and water. We determine the Ginzburg-Landau parameters by requiring that the bulk compressibility and surface tension computed from a field simulation match with those computed from the field representation of a MD simulation. The compressibility can be measured by fluctuations in liquid density $\left\langle(\delta\rho)^2\right\rangle$, while from CW theory, the surface tension can be measured by fluctuations in interfacial height $\left\langle (\delta h)^2 \right\rangle$ (where $\delta\rho \equiv \rho - \langle\rho\rangle$ and likewise for h). Since thermal fluctuations are important driving forces for nanoscale processes [172–174,176,179,180], it is important to preserve these fluctuations quantitatively upon coarse-graining from a molecular model to ensure that the field model accurately describes the physical mechanisms responsible for transport at the nanoscale. The fluctuations can be computed directly from field statistics obtained from MD and field simulations with the mapping procedure described by Voulgarakis & Chu [195] and in Chapter 5. In short, the mass of each fluid molecule is associated with a cubic volume of side length d_{mol}, which corresponds to the excluded volume radius inferred from the radial distribution function. The contribution of each fluid molecule to the mass density of an Eulerian grid cell is then calculated as the mass density of the fluid molecule weighted by the volume fraction of the molecule in the grid cell [195]. In this work, we focus on developing field-theoretic models for liquid-vapor interfaces of argon and water from atomistic simulations.

6.4.1 Argon

A slab of liquid argon ($\sigma_{\mathrm{LJ}} = 3.8\,\text{Å}$ and $\epsilon_{\mathrm{LJ}} = 0.238\,\mathrm{kcal/mol}$) of thickness 60 Å is placed in the center of the MD simulation cell (120 Å × 120 Å × 120 Å) to create two interfaces. The positions and velocities of argon atoms in MD simulation were saved every 2 ps for mapping onto the corresponding field trajectory. The objective is to illustrate if field simulations can match both the fluctuation patterns of density fields and interfacial height observed in MD simulations. Both MD and field simulations of the argon slab were performed at 76 K, and the field trajectories from both MD and field simulations were spatially discretized using cubic grids of length 5 Å. Both simulations were performed for 10 ns, using the last 9 ns for analysis. The time step is 2 fs in MD simulations and 5 fs in field simulations. The shear and bulk viscosities in the FHD-GL equations are calculated by matching the momentum time correlation functions computed from field simulations to the correlation functions from MD simulations, as described by Voulgarakis & Chu [195] and in Chapter 5. This procedure yields $\eta_{\mathrm{S}} = 13\,\mathrm{amu/(\text{Å}\,ps)}$ and $\eta_{\mathrm{B}} = 17\,\mathrm{amu/(\text{Å}\,ps)}$ for argon. The double-well minima for $\psi_0(\rho)$ are determined by the axial density profile computed from MD simulations. For argon, we found $\rho_l = 0.92\,\mathrm{amu/\text{Å}^3}$ and set $\rho_{\mathrm{v}} = 10^{-3}\rho_l$. In the field simulations, we avoid the unphysical appearance of negative densities caused by fluctuating stresses by introducing a flux correction scheme satisfying mass and momentum conservation; further details are described in Section 6.6.2. Bulk density fluctuations are computed for grids inside a liquid slab of thickness 15 Å placed around the center of mass of the liquid phase. The statistics of density and height fluctuations of interfaces calculated from MD simulation are used to determine the values of $\psi_{0,\max}$ and m in the FHD-GL equations.

In Fig. 6.3, we show the contour curves on a plot of $\psi_{0,\max}$ vs. m that represent the loci satisfying $\left\langle (\delta\rho)^2 \right\rangle_{\mathrm{FHD}} = \left\langle (\delta\rho)^2 \right\rangle_{\mathrm{MD}}$ and $\left\langle (\delta h)^2 \right\rangle_{\mathrm{FHD}} = \left\langle (\delta h)^2 \right\rangle_{\mathrm{MD}}$. We observe that $\left\langle (\delta\rho)^2 \right\rangle_{\mathrm{FHD}}$ decreases with increasing $\psi_{0,\max}$ and m. This can be understood from a harmonic

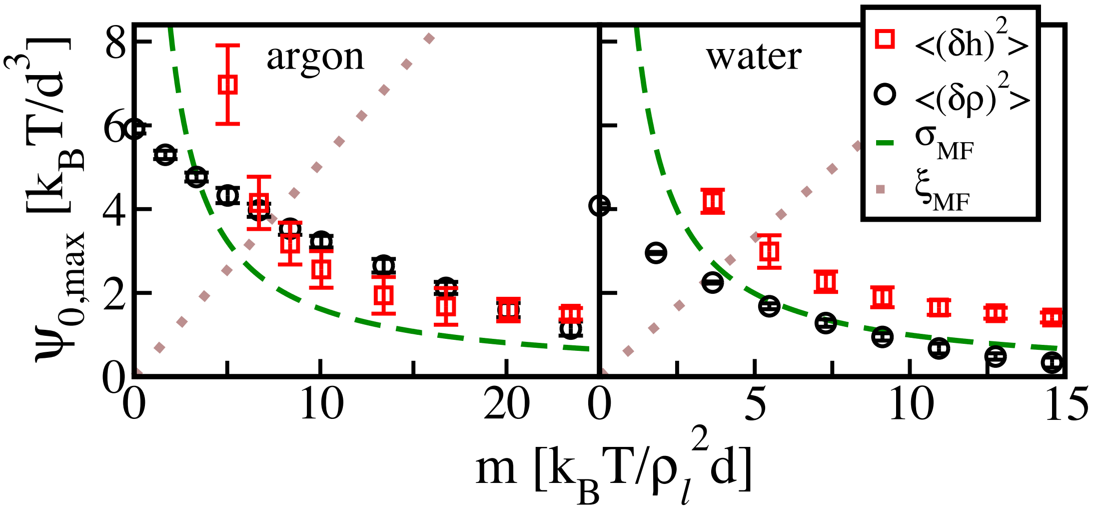


Figure 6.3. The $\psi_{0,\mathrm{max}}$ vs. m parameter diagram for the bulk density and interfacial height fluctuations of argon and water. Black circles represent the contour of $\langle(\delta\rho)^2\rangle_{\mathrm{FHD}} = \langle(\delta\rho)^2\rangle_{\mathrm{MD}}$, red squares represent $\langle(\delta h)^2\rangle_{\mathrm{FHD}} = \langle(\delta h)^2\rangle_{\mathrm{MD}}$, dashed lines represent $\sigma_{\mathrm{MF}} = \sigma_{\mathrm{CW}}^{\mathrm{MD}}$, and dotted lines represent $\xi_{\mathrm{FHD}} = \xi_{\mathrm{MD}}$, where for constant m, $\sigma_{\mathrm{CW}}^{\mathrm{MD}} = (2/3)(\rho_l - \rho_{\mathrm{v}})\sqrt{2m\psi_{0,\mathrm{max}}}$ and $\xi_{\mathrm{FHD}} = 2^{-5/2}(\rho_l - \rho_{\mathrm{v}})\sqrt{m/\psi_{0,\mathrm{max}}}$. For argon, $\sigma_{\mathrm{CW}}^{\mathrm{MD}} = 14.7 \pm 0.4\,k_{\mathrm{B}}T/\mathrm{nm}^2$ and $\xi_{\mathrm{MD}} = 1.2\,\mathrm{\AA}$, and for water, $\sigma_{\mathrm{CW}}^{\mathrm{MD}} = 12.6 \pm 0.5\,k_{\mathrm{B}}T/\mathrm{nm}^2$ and $\xi_{\mathrm{MD}} = 1.1\,\mathrm{\AA}$. Values for $\sigma_{\mathrm{CW}}^{\mathrm{MD}}$ are obtained by best fit to the capillary wave spectrum.

approximation to the Ginzburg-Landau free energy around ρ_l, which gives

$$\langle(\delta\rho)^2\rangle \sim \sum_{\boldsymbol{q}>0}\left(mq^2 + A(\rho_l - \rho_{\mathrm{v}})^2\right)^{-1}, \tag{6.23}$$

where A is proportional to the double-well barrier height, discussed in Section 6.2.1. Along similar lines, $\langle(\delta h)^2\rangle_{\mathrm{FHD}}$ decreases with increasing $\psi_{0,\mathrm{max}}$ and m, since from CW theory [244], we expect $\langle(\delta h)^2\rangle \sim \sigma^{-1}$ and from a mean-field estimate, $\sigma \sim \sqrt{m\psi_{0,\mathrm{max}}}$. Therefore, along the curves of constant $\langle(\delta\rho)^2\rangle_{\mathrm{FHD}}$ and $\langle(\delta h)^2\rangle_{\mathrm{FHD}}$, $\psi_{0,\mathrm{max}}$ decreases with increasing m. One can anticipate this trend upon physical grounds by noting that m penalizes the appearance of density gradients and $\psi_{0,\mathrm{max}}$ determines the local curvature at the free energy minimum. For argon, two sets of values for $\psi_{0,\mathrm{max}}$ and m satisfy the constraints on bulk and interfacial fluctuations. However, they correspond to different interfacial widths ξ appearing in the intrinsic density profile. The best-fit parameters can be determined by requiring additionally that ξ is matched between field and MD simulations. The interfacial width observed in MD simulations ξ_{MD} is obtained by fitting the mean-field expression for the Ginzburg-Landau density profile in Eq. 6.17 to the intrinsic density profile found from MD using a local Gibbs dividing surface method [245]. Setting $\xi_{\mathrm{MD}} = \xi_{\mathrm{FHD}}$, where ξ_{FHD} is given by Eq. 6.18,

Table 6.3. The optimized Ginzburg-Landau free energy parameters from all-atom liquid-vapor simulations of argon at 76 K and water at 300 K. The liquid density ρ_l at phase coexistence is 0.92 amu/Å^3 for argon and 0.603 amu/Å^3 for water. The grid size d in the field representation is 5 Å for both fluids. Units are m $[k_\mathrm{B}T/\rho_l^2 d]$ and $\psi_{0,\mathrm{max}}$ $[k_\mathrm{B}T/d^3]$.

Fluid	m	m_bulk	m_int	$\psi_{0,\mathrm{max}}$
Argon	7.2	—	—	3.8
Water	—	2.8	12	1.8

Table 6.4. Comparison of bulk density ρ and interfacial height h fluctuations computed from all-atom MD and FHD simulations of argon and water. The FHD simulations are performed with the Ginzburg-Landau free energy parameters listed in Table 6.3.

	$\langle(\delta\rho)\rangle^{1/2}/\rho_l$	$\langle(\delta h)\rangle^{1/2}$
Argon (MD)	0.086	3.4
Argon (FHD)	0.084	3.3
Water (MD)	0.12	3.8
Water (FHD)	0.13	3.9

introduces an additional constraint on $\psi_{0,\mathrm{max}}$ and m. From the parameter diagram shown in Fig. 6.3, we choose the intersecting point that most closely satisfies this additional constraint as optimal parameters, which are listed in Table 6.3. Using the optimal values in a field simulation, the relative differences in density and height fluctuations compared to those obtained from MD are less than 5% (Table 6.4).

6.4.2 Water

The results of an argon liquid-vapor interface indicates that a simple Ginzburg-Landau free energy with constant coefficients is sufficient to reproduce the fluctuations of bulk density and interfacial height occurring in all-atom MD. For polyatomic fluids, on the other hand, the intrinsic molecular structure at an interface tends to be distinct from that in the bulk due to the imbalance of intermolecular forces across the phase boundary. In a field representation, it is thus expected that fluid densities have different extents of spatial correlation in the bulk and at an interface. In this section, we show that this is indeed the case for all-atom MD simulations of a water slab. To capture the density and height fluctuations occurring in atomistic simulations, the Ginzburg-Landau free energy must be amended with a density-dependent gradient coefficient $m = m(\rho)$.

The simulation geometry of a water slab with two interfaces is the same as the argon simulation. From the MD simulation using a TIP3P potential, the calculated water density is $\rho_l = 0.603\,\mathrm{amu}/\text{Å}^3$ and we set $\rho_\mathrm{v} = 10^{-3}\rho_l$. The calculated viscosities using the same procedure described earlier are $\eta_\mathrm{S} = 21\,\mathrm{amu}/(\text{Å}\,\mathrm{ps})$ and $\eta_\mathrm{B} = 45\,\mathrm{amu}/(\text{Å}\,\mathrm{ps})$ [195]. In contrast to argon, the $\psi_{0,\max} - m$ contour curves of $\left\langle(\delta\rho)^2\right\rangle_\mathrm{FHD} = \left\langle(\delta\rho)^2\right\rangle_\mathrm{MD}$ and $\left\langle(\delta h)^2\right\rangle_\mathrm{FHD} = \left\langle(\delta h)^2\right\rangle_\mathrm{MD}$ of water do not intersect when a constant value for m is used (Fig. 6.3). Therefore, if height fluctuations are matched, the bulk density fluctuations of field simulations are smaller than those in all-atom MD; if density fluctuations are matched instead, the height fluctuations of field simulations are exaggerated. This result indicates that the field statistics of a Ginzburg-Landau free energy with constant m are inconsistent with the field statistics calculated from all-atom MD simulations of a liquid-vapor interface of water.

A potential solution for this inconsistency is a density-dependent gradient coefficient in the Ginzburg-Landau functional for water, i.e., $m = m(\rho)$. As suggested in Fig. 6.3, $m(\rho)$ should be higher at interfacial densities and lower elsewhere. This modification is expected to suppress interfacial fluctuations without significantly restricting density fluctuations in the bulk liquid. We model $m(\rho)$ with an auxiliary Gaussian function centered at the interfacial density ρ_s to represent the increase in $m(\rho)$ at the interface,

$$m(\rho) = m_\mathrm{bulk} + m_\mathrm{int}\exp\bigl(-(\rho/\rho_\mathrm{s} - 1)^2/w\bigr). \tag{6.24}$$

This decomposition of $m(\rho)$ allows the separate control of bulk and interfacial fluctuations by m_bulk and m_int, respectively, where w regulates the degree of decoupling. In Section 6.6.3, we show that using $w = 0.1$ sets the largest degree of decoupling. The parameters m_bulk, m_int, and $\psi_{0,\max}$ are determined by three constraints: $\left\langle(\delta\rho)^2\right\rangle_\mathrm{FHD} = \left\langle(\delta\rho)^2\right\rangle_\mathrm{MD}$, $\left\langle(\delta h)^2\right\rangle_\mathrm{FHD} = \left\langle(\delta h)^2\right\rangle_\mathrm{MD}$, and $\xi_\mathrm{FHD} = \xi_\mathrm{MD}$, with the interfacial width approximated as Eq. 6.18 based on the mean-field theory. The resulting Ginzburg-Landau parameters for water are listed in Table 6.3. The results of a field simulation using these parameters are compared with those of all-atom MD in Table 6.4. The relative differences in density and height fluctuations are within 5%, illustrating that the density-dependent gradient coefficient in the Ginzburg-Landau free energy functional for water reestablishes consistency between the fluctuation patterns of field and MD simulations.

The unique molecular structure of water causes it to form a specific hydrogen-bonding network in the liquid state. At an interface, the imbalance of intermolecular interactions causes water molecules to adopt a hydrogen-bonding network different from that in the bulk [173,246–248] with suppressed interfacial density fluctuations as observed in the all-atom MD simulation. At the field-theoretic level, this phenomenon can be effectively represented by a higher interfacial gradient coefficient.

We note that in practice, the Ginzburg-Landau parameters for a molecular fluid depend on the grid size used for solving the FHD-GL equations. However, if capturing the field fluctuations mapped from all-atom MD with the field model is employed as a constraint, the allowable range of grid sizes is limited. In the case of water, for example, using 10 Å grids severely suppresses density fluctuations and capillary waves at the liquid-vapor interface [195]. Using a grid size smaller than the length scale of a single water molecule, on the other hand, challenges the local thermodynamic approximation employed in the FHD equations. A heuristic analysis of the van der Waals equation of state of water showed that the square-gradient free energy becomes valid when the mean molecular density field is coarse-grained over a length scale larger than 4 Å [173], questioning the applicability of a grid size below this length scale. In addition, when using grid sizes less than 6 Å to convert atomistic configurations into a field trajectory, we found that the observed mean-square interfacial fluctuations are invariant under spatial translation of atomistic coordinates, indicating that potential artifacts due to periodicity of the lattice are not present. Therefore, a grid size of 5 Å appears to be an ideal coarse-graining length scale for capturing hydrodynamic fluctuations, since this size resolves the stochastic behavior of density and momentum fields while ensuring the validity of a hydrodynamic description.

6.4.3 Capillary wave spectra

To go beyond averaged fluctuations, we examine in this section if using the Ginzburg-Landau parameterization in field simulations for argon and water can capture the static and dynamic capillary wave spectra at the liquid-vapor interface observed in all-atom MD simulations. Since the temporal and spatial correlation at interfaces are key properties in governing the transport processes at the nanoscale, it is important for a coarse-grained field model to describe such fluctuation patterns.

The static CW spectrum given in Eq. 6.20 describes the equal-time correlations between Fourier modes of a liquid-vapor interface. In Fig. 6.4, we plot the static CW spectra obtained from MD and field simulations using the optimized Ginzburg-Landau parameters for argon and water. Close agreement between the MD and field simulations are clear, indicating that the coarse-grained field models for argon and water preserve equal-time interfacial correlations and samples fluctuation modes in accordance with CW theory. From the CW spectra of MD simulations, we have also extracted the best-fit values for surface tension and compared these values with virial calculations of the surface tension [249]. As seen in Table 6.5, these values are in close agreement. Since the virial surface tension is directly computed from atomistic coordinates, the agreement demonstrates that our mapping scheme [195] is scale-consistent when applied to interfacial configurations. Furthermore, although the Ginzburg-Landau model is typically used to describe fluids with diffuse interfacial profiles (such as those near the critical point) [236,237], we illustrate that consistency with CW theory can be obtained accurately even for cold fluids with relatively sharp interfacial profiles.

As shown in Table 6.5, applying the mean-field prediction with Ginzburg-Landau parameters to water and argon overestimates the surface tension as compared with the value computed from the CW spectra. This discrepancy can be rationalized by examining the

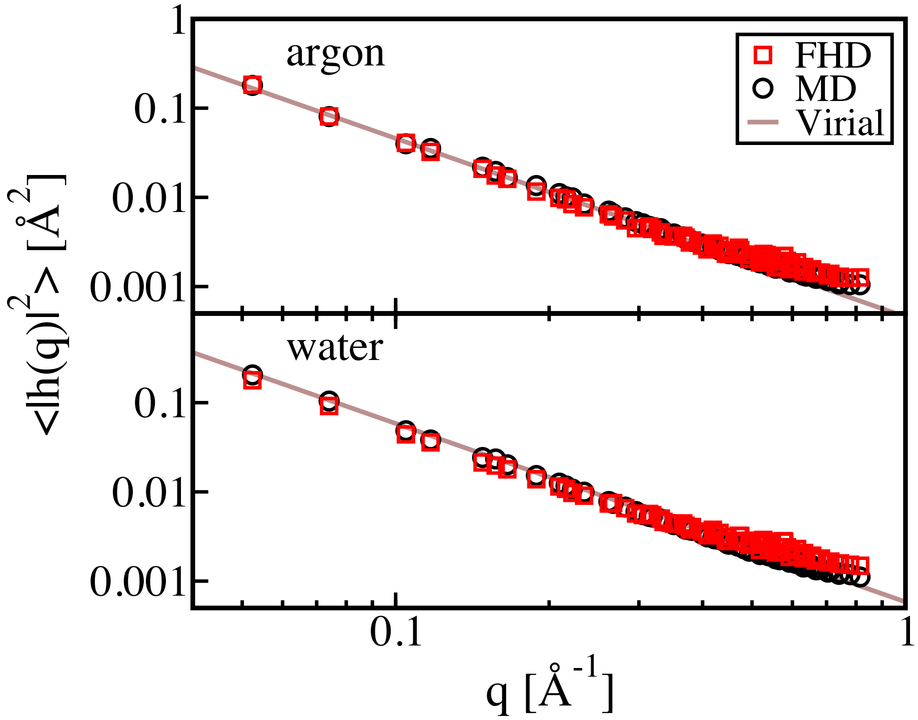


Figure 6.4. Static capillary wave spectra computed from FHD (red squares) and all-atom MD (black circles) simulations for argon and water. Solid lines represent Eq. 6.20 using the surface tension obtained from the virial tensor.

Table 6.5. Comparison of surface tensions $[k_\mathrm{B}T/\mathrm{nm}^2]$ computed from the capillary wave (CW) spectra of MD and FHD simulations with virial and mean-field calculations.

	Argon	Water
CW (MD)	14.7 ± 0.4	12.6 ± 0.5
CW (FHD)	15 ± 1	14.1 ± 0.5
Virial	15.2 ± 0.3	11.9 ± 0.4
Mean-field	19.1	28.5

ratio of thermal to surface energy, $\theta \equiv k_{\mathrm{B}}T/(\sigma_{\mathrm{MF}}\mathrm{\AA}^2)$. For the model fluids discussed earlier in which the mean-field prediction is accurate, this ratio ($\theta = 2.6$) is about three times lower than the ratio for argon ($\theta = 6.8$) and water ($\theta = 7.9$). Therefore, at higher temperatures, the mean-field density profile is disrupted by thermal fluctuations and the mean-field estimate no longer provides an accurate prediction for the surface tension calculated from the CW spectra. For the model fluids, we found that this behavior occurs when $\theta = 6$.

In addition to the equilibrium spectra of fluctuations, we also probe the relaxation of interfacial fluctuations by comparing the dynamic CW spectra obtained from field and MD simulations of argon and water. This comparison allows us to assess to what extent the coarse-grained field model preserves the relaxation time-scales at the liquid-vapor interfaces observed in all-atom MD simulations. The dynamic CW spectra, $S(\boldsymbol{q},\omega)$, is defined as the Fourier transform of the time correlation function of interfacial Fourier modes,

$$S(\boldsymbol{q},\omega) = \int_{-\infty}^{+\infty} dt \exp(i\omega t)\langle h(\boldsymbol{q},0)h(-\boldsymbol{q},t)\rangle. \tag{6.25}$$

In Fig. 6.5, we plot the normalized dynamic CW spectra $S_0(\boldsymbol{q},\omega)$ for field and MD simulations of argon and water, where $\boldsymbol{q}_1 = (2\pi/L, 0, 0)$ corresponds to the longest wavelength along the interface. The close agreement in $S_0(\boldsymbol{q},\omega)$ between field and MD simulations is clear. It is also important to emphasize that agreement in $S_0(\boldsymbol{q},\omega)$ was not imposed when parameterizing the Ginzburg-Landau functional. The input viscosity values in the FHD-GL equations were determined by matching the bulk momentum time correlation functions between field and MD simulations. The Ginzburg-Landau free energy was determined by matching density and interfacial fluctuations. Since $S_0(\boldsymbol{q},\omega)$ depends uniquely on the transport coefficients and free energy, a robust field-theoretic model for a molecular fluid is expected to capture this dynamical observable without explicit fitting. Therefore, the additional agreement we find with $S_0(\boldsymbol{q},\omega)$ underscores that the Ginzburg-Landau model coupled with hydrodynamic fluctuations is capable of capturing the spatial and temporal correlations among field variables observed in MD simulations.

We also compare the dynamic CW spectra obtained from MD and field simulations with predictions by linearized hydrodynamics. For a fluctuating liquid-vapor interface for an incompressible fluid [250–252], $S(\boldsymbol{q},\omega)$ is expected to follow,

$$S(\boldsymbol{q},\omega) = -2k_{\mathrm{B}}T\frac{q}{L^2\rho\omega}\frac{\operatorname{Im} D(\boldsymbol{q},\omega)}{|D(\boldsymbol{q},\omega)|^2}, \tag{6.26}$$

where $D(\boldsymbol{q},\omega)$ is given by,

$$D(\boldsymbol{q},\omega) = \sigma q^3/\rho - \left(\omega + 2i\nu q^2\right)^2 - 4\nu^2 q^4\left(1 - i\omega/\left(\nu q^2\right)\right)^{1/2}, \tag{6.27}$$

and $\nu = \eta_{\mathrm{S}}/\rho$ is the kinematic shear viscosity. In Fig. 6.5, we plot best-fit curves described by Eq. 6.25, using the shear viscosity η_{S} as an adjustable parameter. In the long wavelength

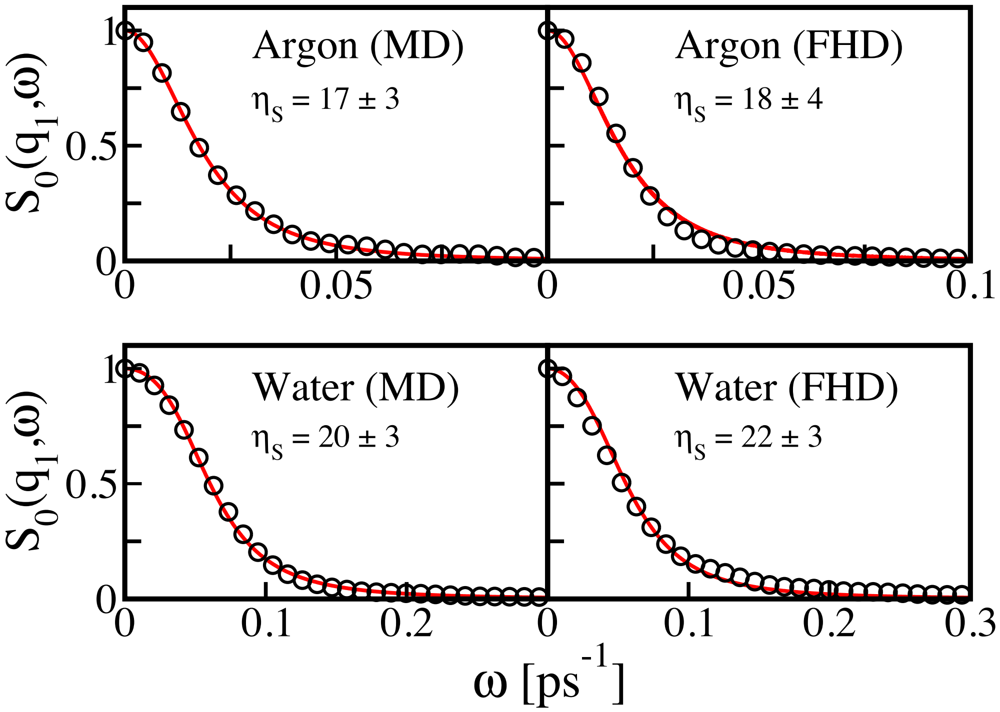


Figure 6.5. Dynamic capillary wave spectra computed from FHD and all-atom MD (black circles) simulations for argon and water. Here, $\boldsymbol{q}_1 = (2\pi/L, 0, 0)$ corresponds to the longest wavelength along the interface. Best-fit curves using Eq. 6.25 are shown as solid curves, along with the best-fit values for shear viscosity η_S [amu/(Å ps)]. FHD and MD simulations were performed for a total of 50 ns with box size 120 Å × 120 Å × 120 Å

.

limit, we find Eq. 6.25 is an accurate model for $S_0(\boldsymbol{q}, \omega)$ computed from simulation, as the effects due to density compressibility are expected to be small. We compare the best-fit values for interfacial viscosity shown in Fig. 6.5 using Eq. 6.25 to the values for shear viscosity determined from the bulk momentum time correlation functions [195,204] and used as input for the field simulations. For both argon and water, the best-fit interfacial viscosities agree with the input viscosities to within statistical error. We observe, however, that time correlation functions for long wavelength modes converges slowly in both MD and field simulations, and longer simulation times up to one microsecond may be required to reduce errors in estimates for interfacial viscosity. Nevertheless, the agreement among the dynamic CW spectra obtained from MD and field simulations and predicted by theory is satisfactory. This result underscores that the mapping of atomistic simulations onto a coupled Ginzburg-Landau fluctuating hydrodynamics field model may be a practically useful strategy for coarse-graining atomistic simulations of fluid interfaces.

6.5 Discussion

In this work, we develop a coarse-grained (CG) field-theoretic model for inhomogeneous molecular fluids by coupling a Ginzburg-Landau free energy with hydrodynamic fluctuations. We show that fluctuating hydrodynamic (FHD) simulations of liquid-vapor and liquid-liquid interfaces of model systems satisfy the fluctuation-dissipation theorem accurately and describe interfacial fluctuations in accordance with capillary wave theory. To achieve consistency with molecular dynamics (MD) simulations, we use a scale-consistent mapping scheme to coarse-grain atomistic configurations into a field representation, enabling FHD and MD simulations to be compared transparently. By optimizing the transport coefficients and free energy functional in FHD simulations for argon and water, we show that a Ginzburg-Landau model coupled to hydrodynamic fluctuations can capture the compressibility, surface tension, and static and dynamic capillary wave spectra observed in MD simulations. To describe these fluctuation patterns, a grid size of 5 Å was found to be ideal for solving the FHD equations to balance the requirements of resolving the density profile of an interface and approaching the local thermodynamic approximation in a CG field-theoretic model. The results of this work thus illustrate that combining bottom-up coarse-graining with top-down phenomenology provides a practical framework for constructing field-theoretic models from the hydrodynamic fluctuations and correlations that emerge from the underlying molecular interactions.

6.6 Supplementary information

6.6.1 Derivation of reversible stress tensor

We describe the derivation of the reversible stress tensor $\mathbf{P}$ from the free energy functional $F[\rho(\boldsymbol{r})]$ by using the least-action principle [241]. We note that an alternative route, involving entropy production, can also be used and gives identical results [253].

We consider an inviscid, isothermal fluid with total Helmholtz free energy

$$F[\rho(\boldsymbol{r})] = \int \Big(\psi_0(\rho) + \frac{m}{2}|\nabla\rho|^2 + \frac{\alpha}{2}\big(\nabla^2\rho\big)^2\Big)\, d^3r, \tag{6.28}$$

where ρ is the mass density, ψ_0 is the local free energy density, m is the density gradient coefficient, and α is the density curvature coefficient. The dynamical action is

$$S = \int dt \int d^3r\, \frac{1}{2}\rho\boldsymbol{v}^2 - F[\rho(\boldsymbol{r};t)], \tag{6.29}$$

where $\boldsymbol{v}$ is the fluid velocity. We impose that the continuity equation is satisfied and that advected quantities χ, such as fluid particle identities, are carried with the flow,

$$S = \int dt \int d^3r\, \frac{1}{2}\rho\boldsymbol{v}^2 - F[\rho(\boldsymbol{r};t)] + \phi\Big(\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{v})\Big) + \lambda\Big(\frac{\partial\rho\chi}{\partial t} + \nabla\cdot(\rho\chi\boldsymbol{v})\Big), \tag{6.30}$$

where ϕ and λ are Lagrange multipliers. Taking the extremum of action S with respect to variations in $\boldsymbol{v}$, ρ, and χ gives the following relations:

$$\begin{aligned}
\delta\boldsymbol{v}:&\quad \boldsymbol{v} = \nabla\phi + \chi\nabla\lambda,\\
\delta\rho:&\quad \frac{1}{2}\boldsymbol{v}^2 - \left(\frac{\delta F}{\delta\rho}\right) - \left(\frac{\partial\phi}{\partial t} + \boldsymbol{v}\cdot\nabla\phi\right) - \chi\left(\frac{\partial\lambda}{\partial t} + \boldsymbol{v}\cdot\nabla\lambda\right) = 0, && (6.31)\\
\delta\chi:&\quad \frac{\partial\lambda}{\partial t} + \boldsymbol{v}\cdot\nabla\lambda = 0, && (6.32)
\end{aligned}$$

where we have used integration by parts and allowed the boundary conditions to vanish. Taking the gradient of the second line in Eq. 6.32 and utilizing the other two equations, we obtain

$$\rho\frac{D\boldsymbol{v}}{Dt} = -\rho\nabla\left(\frac{\delta F}{\delta\rho}\right). \tag{6.33}$$

The reversible stress tensor $\mathbf{P}$ (up to a divergence-free contribution) is defined through the expression,

$$\nabla\cdot\mathbf{P} = -\rho\nabla\left(\frac{\delta F}{\delta\rho}\right). \tag{6.34}$$

The functional derivative of F is given by the Euler-Lagrange equation,

$$\frac{\delta F}{\delta\rho} = \frac{\partial\psi_0}{\partial\rho} + \frac{1}{2}m'|\nabla\rho|^2 - \nabla\cdot(m\nabla\rho) + \frac{1}{2}\alpha'\big(\nabla^2\rho\big)^2 + \nabla^2\big(\alpha\nabla^2\rho\big), \tag{6.35}$$

where primed variables denote differentiation with respect to density ρ. Substitution of Eq. 6.35 into Eq. 6.34 gives

$$\begin{aligned}
\nabla\cdot\mathbf{P} = &-\rho\nabla\left(\frac{\partial\psi_0}{\partial\rho}\right) && (6.36)\\
&-\rho\nabla\left(\frac{1}{2}m'|\nabla\rho|^2\right) + \rho\nabla(\nabla\cdot(m\nabla\rho))\\
&-\rho\nabla\left(\frac{1}{2}\alpha'\big(\nabla\rho^2\big)^2\right) - \rho\nabla\big(\nabla^2\big(\alpha\nabla^2\rho\big)\big).
\end{aligned}$$

The first term on the right hand side is

$$-\rho\nabla\left(\frac{\partial\psi_0}{\partial\rho}\right) = -\rho\nabla\mu_0 = -\nabla p_0, \tag{6.37}$$

by the Gibbs-Duhem equation, where p_0 is the thermodynamic pressure. We divide $\mathbf{P}$ into gradient and curvature contributions,

$$\nabla\cdot\mathbf{P}_m = -\rho\nabla\left(\frac{1}{2}m'|\nabla\rho|^2\right) + \rho\nabla(\nabla\cdot(m\nabla\rho)), \tag{6.38}$$

$$\nabla\cdot\mathbf{P}_\alpha = -\rho\nabla\left(\frac{1}{2}\alpha'\big(\nabla^2\rho\big)^2\right) - \rho\nabla\big(\nabla^2\big(\alpha\nabla^2\rho\big)\big). \tag{6.39}$$

First, we derive the expression for $\nabla \cdot \mathbf{P}_m$. The first and second terms on the right hand side of Eq. 6.38 need to be manipulated into the form of a divergence of a tensor. The first term can be rewritten using the product rule,

$$-\rho\nabla\left(\frac{1}{2}m'|\nabla\rho|^2\right) = -\nabla\left(\frac{1}{2}m'\rho|\nabla\rho|^2\right) + \frac{1}{2}m'\nabla\rho|\nabla\rho|^2. \tag{6.40}$$

The second term can be identified using the following manipulations,

$$\begin{aligned}-\nabla\cdot(m\nabla\rho\otimes\nabla\rho) =& - m\nabla\rho\cdot\nabla\nabla\rho - (\nabla\cdot m\nabla\rho)\nabla\rho,\\ =& - m\nabla\rho\cdot\nabla\nabla\rho - (\nabla\cdot m\nabla\rho)\nabla\rho - \rho\nabla(\nabla\cdot m\nabla\rho) + \rho\nabla(\nabla\cdot m\nabla\rho),\\ =& - m\nabla\rho\cdot\nabla\nabla\rho - \nabla(\rho\nabla\cdot m\nabla\rho) + \rho\nabla(\nabla\cdot m\nabla\rho),\\ =& - \nabla\left(\frac{1}{2}m|\nabla\rho|^2\right) + \frac{1}{2}m'\nabla\rho|\nabla\rho|^2\\ & - \nabla(\rho\nabla\cdot m\nabla\rho) + \rho\nabla(\nabla\cdot m\nabla\rho).\end{aligned} \tag{6.41}$$

In the first, third, and fourth lines, we have applied the product rule, and in the second line, added zero. Note that the last term in Eq. 6.41 is equal to the second term on the right hand side in Eq. 6.38. Combining Eqs. 6.38, 6.40, and 6.41, we obtain the gradient contribution to the reversible stress,

$$\mathbf{P}_m = \left[m\rho\nabla^2\rho + \frac{(m+m'\rho)|\nabla\rho|^2}{2}\right]\mathbf{I} - m\nabla\rho\otimes\nabla\rho. \tag{6.42}$$

The curvature contribution $\mathbf{P}_\alpha$ is found by manipulating the first and second terms on the right hand side in Eq. 5.30. The second term can be rewritten as,

$$\begin{aligned}-\rho\nabla\big(\nabla^2(\alpha\nabla^2\rho)\big) &= -\nabla\big(\rho\nabla^2(\alpha\nabla^2\rho)\big) + \nabla^2(\alpha\nabla^2\rho)\nabla\rho,\\ &= -\nabla\big(\rho\nabla^2(\alpha\nabla^2\rho)\big) + \nabla^2(\alpha\nabla^2\rho)\nabla\rho\\ &\quad +\nabla\Big(\frac{\alpha}{2}(\nabla^2\rho)^2\Big) - \nabla\Big(\frac{\alpha}{2}(\nabla^2\rho)^2\Big),\\ &= -\nabla\big(\rho\nabla^2(\alpha\nabla^2\rho)\big) + \nabla^2(\alpha\nabla^2\rho)\nabla\rho\\ &\quad +\alpha\nabla^2\rho\nabla^2(\nabla\rho) + \frac{\alpha'}{2}\nabla\rho(\nabla^2\rho)^2 - \nabla\Big(\frac{\alpha}{2}(\nabla^2\rho)^2\Big),\end{aligned} \tag{6.43}$$

where we have used the product rule in the first and fifth lines and added zero on the third line. We introduce the following identities,

$$\nabla\cdot\big(\nabla\rho\otimes\nabla(\alpha\nabla^2\rho)\big) = \nabla\rho\cdot\nabla\nabla(\alpha\nabla^2\rho) + \nabla(\alpha\nabla^2\rho)\nabla^2\rho, \tag{6.44}$$

$$\nabla\cdot\big(\nabla(\alpha\nabla^2\rho)\otimes\nabla\rho\big) = \nabla(\alpha\nabla^2\rho)\cdot\nabla\nabla\rho + \nabla^2(\alpha\nabla^2\rho)\nabla\rho. \tag{6.45}$$

Adding Eqs. 6.44 and 6.44 gives

$$\begin{aligned}\nabla\cdot\big(\nabla\rho\otimes\nabla(\alpha\nabla^2\rho)\big) + \nabla\cdot\big(\nabla(\alpha\nabla^2\rho)\otimes\nabla\rho\big) =& \nabla\big(\nabla(\alpha\nabla^2\rho)\cdot\nabla\rho\big)\\ &+ \nabla(\alpha\nabla^2\rho)(\nabla^2\rho) + \nabla^2(\alpha\nabla^2\rho)\nabla\rho.\end{aligned} \tag{6.46}$$

Combining Eqs. 6.39, 6.43 and 6.47, we have

$$\begin{aligned}\mathbf{P}_\alpha =& \left[-\rho\nabla^2\big(\alpha\nabla^2\rho\big) - \frac{(\alpha+\alpha'\rho)(\nabla^2\rho)^2}{2} - \nabla\big(\alpha\nabla^2\rho\big)\cdot\nabla\rho\right]\mathbf{I} \\ &+ \nabla\rho\otimes\nabla\big(\alpha\nabla^2\rho\big) + \nabla\big(\alpha\nabla^2\rho\big)\otimes\nabla\rho.\end{aligned} \tag{6.47}$$

Expressions for reversible stress for a two-component fluid can be derived similarly by introducing the inviscid component mass balance as a constraint in the expression for the action, and taking additionally an extremum of action S with the associated Lagrange multiplier ξ,

$$S = \int dt \int d^3r\,\frac{1}{2}\rho\boldsymbol{v}^2 - F[\rho(\boldsymbol{r};t)] + \phi\left(\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\boldsymbol{v})\right) + \lambda\left(\frac{\partial\rho\chi}{\partial t} + \nabla\cdot(\rho\chi\boldsymbol{v})\right) \tag{6.48}$$

$$+\,\xi\left(\frac{\partial\rho c}{\partial t} + \nabla\cdot(\rho c\boldsymbol{v})\right). \tag{6.49}$$

The result after following similar manipulations described above is Eq. 6.11. For viscous fluids, the dissipative and fluctuating stress are added in parallel alongside the reversible stress in the momentum balance.

Why should we go through the trouble of rewriting the momentum balance as the divergence of tensors? It is advantageous for numerical simulation because it allows the identification of the momentum flux, which provides a straightforward procedure to ensure that the integration scheme conserves momentum. The momentum added to one grid cell can then be exactly subtracted (up to floating point precision) from the other cell.

6.6.2 Low density flux correction scheme

We solve the FHD-GL equations using the staggered discretization scheme described by Voulgarakis & Chu [195]. Briefly, the total mass density ρ is located at grid centers, while the momentum density $\boldsymbol{g}$ and mass flux $\boldsymbol{j}$ are located at grid faces. We associate each grid with an index vector $\boldsymbol{n} = (n_x, n_y, n_z)$, with the integers n_x, n_y, and n_z corresponding to the grid index in the x, y, and z directions, respectively. The unit vector along each direction is denoted as $\boldsymbol{i}_\alpha$, where $\alpha = x$, y, or z. The total mass density associated with the center of the grid at $\boldsymbol{n}$ and time t is $\rho^t_{\boldsymbol{n}}$. The component of mass flux associated with the grid face between grid centers with index $\boldsymbol{n}$ and $\boldsymbol{n}+\boldsymbol{i}_\alpha$ is $j^t_{a,\boldsymbol{n}}$. The discretized continuity equation is given by

$$\rho^{t+\Delta t}_{\boldsymbol{n}} = \rho^t_{\boldsymbol{n}} - \frac{\Delta t}{d}\sum_\alpha\big(j^t_{\alpha,\boldsymbol{n}} - j^t_{\alpha,\boldsymbol{n}-\boldsymbol{i}_\alpha}\big), \tag{6.50}$$

where d is the length of the cubic grid and Δt is the time step.

Since the fluctuations in density are approximately Gaussian, the discretized continuity equation does not prevent the appearance of negative densities especially in the presence of fluctuating stresses and small grid volumes. The magnitude of density fluctuation $\delta\rho \equiv \rho - \langle\rho\rangle$

in each cell scales with grid size as $\langle(\delta\rho)\rangle^{1/2} \sim d^{-3/2}$, as shown in Chapter 5. If no corrections are introduced in the identity $j^t_{\alpha,\boldsymbol{n}} = g^t_{\alpha,\boldsymbol{n}}$, then unphysical negative densities can occur even for bulk water simulations at average density $\langle\rho\rangle = \rho_{\mathrm{water}} = 10^3\,\mathrm{kg/m^3}$ and temperature $T = 300\,\mathrm{K}$ using grid size $d = 5\,\text{Å}$. In addition, positive, near-zero values of $\rho^t_{\boldsymbol{n}}$ introduce a singularity in the calculation of fluid velocity from momentum density $v^t_{\alpha,\boldsymbol{n}} = g^t_{\alpha,\boldsymbol{n}}/\rho^t_{\alpha,\boldsymbol{n}}$, and disrupt the fluctuation-dissipation balance in a fluid. The density $\rho^t_{\alpha,\boldsymbol{n}}$ is the average of the densities at grid centers $\boldsymbol{n}$ and $\boldsymbol{n} + \boldsymbol{i}_\alpha$.

To circumvent these problems, we use a flux correction scheme that avoids the appearance of negative densities and satisfies total mass and momentum conservation. First, the velocity is calculated as $v^t_{\alpha,\boldsymbol{n}} = g^t_{\alpha,\boldsymbol{n}}/\rho_{\mathrm{vel}}$, where $\rho_{\mathrm{vel}} = \max\big(\rho^t_{\alpha,\boldsymbol{n}}, \rho^*_l\big)$, and ρ^*_l specifies a lower density threshold, which removes the singularity in the velocity calculation. The mass flux $j^t_{\alpha,\boldsymbol{n}}$ is then determined by $j^t_{\alpha,\boldsymbol{n}} = \rho_j v^t_{\alpha,\boldsymbol{n}}$, where

$$\rho_j = \begin{cases} \rho^t_{\boldsymbol{n}} & \text{if } j^t_{\alpha,\boldsymbol{n}} > 0 \text{ and } (\rho^t_{\boldsymbol{n}} < \rho^*_2 \text{ or } \rho^t_{\boldsymbol{n}+\boldsymbol{i}_\alpha} < \rho^*_2), \\ \rho^t_{\boldsymbol{n}+\boldsymbol{i}_\alpha} & \text{if } j^t_{\alpha,\boldsymbol{n}} < 0 \text{ and } (\rho^t_{\boldsymbol{n}} < \rho^*_2 \text{ or } \rho^t_{\boldsymbol{n}+\boldsymbol{i}_\alpha} < \rho^*_2), \\ \rho^t_{\alpha,\boldsymbol{n}} & \text{otherwise.} \end{cases} \tag{6.51}$$

Physically, this assignment has the effect of suppressing the amount of mass exchanged between two grids when mass flows from a lower density grid to a higher one, and exaggerating the mass exchange when mass flows from a higher density grid to a lower one. Therefore, the change in density of a grid with low density at time t is biased in the direction of higher density at time $t + \Delta t$. For argon and water, we use $\rho^*_1 = 0.25\rho_l$ and $\rho^*_2 = 0.15\rho_l$, where ρ_l is the liquid density. Although the choice for these values is somewhat arbitrary, we find that variation of these values between the physically reasonable range of $0.1\rho_l < \{\rho^*_1, \rho^*_2\} < 0.3\rho_l$ does not affect values of the observables $\big\langle(\delta\rho)^2\big\rangle_{\mathrm{FHD}}$ and $\big\langle(\delta h)^2\big\rangle_{\mathrm{FHD}}$. Note that when corrections to the mass flux are unnecessary, $j^t_{\alpha,\boldsymbol{n}} = g^t_{\alpha,\boldsymbol{n}}$ is recovered. We caution that our FHD-GL model does not attempt to describe the dynamics of rarified vapor. Rather, the flux correction scheme is a computationally efficient method to allow a background vapor phase to exist in simulation and enables us to solve the FHD-GL equations throughout the entire volume of the system without having to explicitly track regions of low density. In addition, our scheme ensures that simulations of bulk water remain numerically stable when small grid volumes are used, allowing for transient cavitation within a grid but avoiding the numerical singularities associated with the appearance of low density.

Conservation of mass and momentum is ensured because the exchanges of these quantities obey “equal and opposite”. The amount of mass or momentum leaving one cell is equal to the amount entering the neighboring cell. We note that alternative strategies are possible but they do not conserve mass and/or momentum. One may consider using a cut-off scheme, which resets the mass in a cell to zero, if it were to become negative. However, this modification does not conserve mass. One may also consider scaling the momentum densities – assumed here to be equal to mass fluxes – to increase (decrease) the mass entering (leaving) a low-density cell. While this scheme conserves mass, it does not always conserve momentum. By decoupling momentum density and mass flux for low density cells, both mass and momentum are conserved. We find such conservation is crucial for the computed capillary wave spectra for argon and water to agree with predictions by capillary wave theory.

6.6.3 Effect of density-dependent gradient coefficient

We show that a density-dependent gradient coefficient $m = m(\rho)$ in the Ginzburg-Landau free energy allows fluctuations in bulk liquid density and interfacial height to be decoupled. To match the fluctuations in TIP3P water, we use the function,

$$m(\rho) = m_{\mathrm{bulk}} + m_{\mathrm{int}} \exp\bigl(-(\rho/\rho_{\mathrm{s}} - 1)^2/w\bigr), \tag{6.52}$$

where $\rho_{\mathrm{s}} = (\rho_l + \rho_{\mathrm{v}})/2$. Here, we explore the effects of the width w on decoupling bulk and interfacial fluctuations. In Fig. 6.6ab, we show the effects of using different values of w on interfacial and bulk fluctuations of fluids, respectively. Note that for $w = \infty$, we have constant $m = m_{\mathrm{bulk}} + m_{\mathrm{int}}$. The optimal $w = w^*$ should be large enough such that $\left\langle (\delta h)^2 \right\rangle$ computed using w^* is close to the value computed using $w = \infty$, for given values of m_{bulk} and m_{int}. Also, w^* should be small enough such that $\left\langle (\delta \rho)^2 \right\rangle$ computed using w^* is close to the value computed using $m_{\mathrm{int}} = 0$, for a given value of m_{bulk}. An objective function quantifying the degree of decoupling reads

$$\chi(m_{\mathrm{bulk}}, m_{\mathrm{int}}, w) = \frac{\mathrm{abs}\left(\left\langle (\delta h)^2 \right\rangle - \left\langle (\delta h)^2 \right\rangle \Big|_{w=\infty} \right)}{\left\langle (\delta h)^2 \right\rangle \Big|_{w=\infty}} + \frac{\mathrm{abs}\left(\left\langle (\delta \rho)^2 \right\rangle - \left\langle (\delta \rho)^2 \right\rangle \Big|_{m_{\mathrm{int}}=0} \right)}{\left\langle (\delta \rho)^2 \right\rangle \Big|_{m_{\mathrm{int}}=0}}. \tag{6.53}$$

In Fig. 6.6, we plot χ for $m_{\mathrm{bulk}} = 2.8\, k_{\mathrm{B}}T/\rho_l^2 d$, corresponding to the value for water, and different values of m_{int} and w. In the range where the optimal m_{int} for water lies ($m_{\mathrm{int}} = 12\, k_{\mathrm{B}}T/\rho_l^2 d$), we find χ is smallest for $w = 0.1$. At higher values, bulk fluctuations become too suppressed as m_{int} increases, while at lower values, height fluctuations become less sensitive to changes in m_{int}. By decoupling bulk and interfacial fluctuations using a density-dependent gradient coefficient, a wider range of fluid behavior can be explored within the coupled Ginzburg-Landau fluctuating hydrodynamics model.

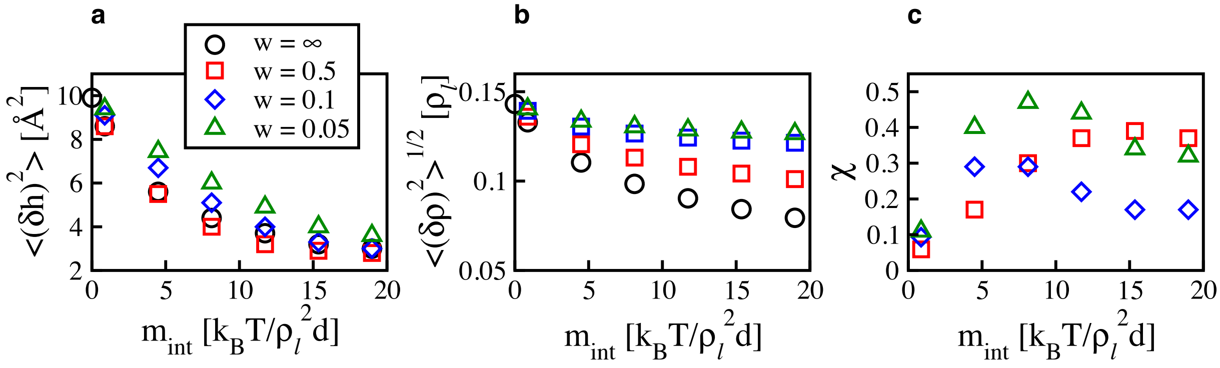


Figure 6.6. Parameterization of the density-dependent square gradient coefficient in the Ginzburg-Landau free energy functional for fluid interfaces. (**a**) Fluctuations in interfacial height $\left\langle(\delta h)^2\right\rangle$ and (**b**) fluctuations in liquid density $\left\langle(\delta \rho)^2\right\rangle$ determined from FHD simulations for different values of m_{int} and w. (**c**) Values of the objective function χ calculated from Eq. 6.53. For all data points, we use $m_{\text{bulk}} = 2.8\, k_{\text{B}}T/\rho_l^2 d$, corresponding to the value for water.

Bibliography

1. Pacala, S. & Socolow, R. H. Stabilization wedges: solving the climate problem for the next 50 years with current technologies. *Science* **305,** 968–972 (2004).
2. Chu, S. & Majumdar, A. Opportunities and challenges for a sustainable energy future. *Nature* **488,** 294–303 (2012).
3. Himmel, M. E. *et al.* Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **315,** 804–807 (2007).
4. Carroll, A. & Somerville, C. Cellulosic biofuels. *Annu. Rev. Plant. Biol.* **60,** 165–182 (2009).
5. O'Sullivan, A. C. Cellulose: the structure slowly unravels. *Cellulose* **4,** 173–207 (1997).
6. Perlack, R. D. *et al. Biomass as feedstock for a biomass and bioproducts industry: the technical feasibility of a 1 billion ton annual feedstock supply* tech. rep. ORNL/TM-2005/66 (Oak Ridge National Laboratory, Oak Ridge, TN, USA, 2005).
7. Perlack, R. D. & Stokes, B. J. *U. S. billion-ton update: biomass supply for a bioenergy and bioproducts industry* tech. rep. ORNL/TM-2011/224 (Oak Ridge National Laboratory, Oak Ridge, TN, USA, 2011).
8. Tilman, D. *et al.* Beneficial biofuels – the food, energy, and environment trilemma. *Science* **325,** 270–271 (2009).
9. Williams, P., Inman, D., Aden, A. & Heath, G. A. Environmental and sustainability factors associated with next-generation biofuels in the US: what do we really know? *Environ. Sci. Technol.* **43,** 4764–4775 (2009).
10. Banholzer, W. F. Practical limitations and recognizing hype. *Environ. Sci. Technol.* **5,** 5478–5480 (2012).
11. Searchinger, T. *et al.* Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **319,** 1238–1240 (2008).
12. Fargione, J., Hill, J., Tilman, D., Polasky, S. & Hawthorne, P. Land clearing and the biofuel carbon debt. *Science* **319,** 1235–1238 (2008).
13. Montgomery, C. T. & Smith, M. B. Hydraulic fracturing: history of an enduring technology. *J. Petrol. Technol.* **62,** 26–32 (2010).

14. McKendry, P. Energy production from biomass (part 2): conversion technologies. *Bioresource Technol.* **83,** 47–54 (2002).
15. Brown, T. R. & Brown, R. C. A review of cellulosic biofuel commercial-scale projects in the United States. *Biofuels, Bioprod. Bioref.* **7,** 235–245 (2013).
16. Ragauskas, A. J. *et al.* The path forward for biofuels and biomaterials. *Science* **311,** 484–489 (2006).
17. Hendriks, A. T. W. M. & Zeeman, G. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technol.* **100,** 10–18 (2009).
18. Mosier, N. *et al.* Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresource Technol.* **96,** 673–686 (2005).
19. Yang, B. & Wyman, C. E. Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioprod. Bioref.* **2,** 26–40 (2007).
20. Yang, B., Dai, Z., Ding, S.-Y. & Wyman, C. E. Enzymatic hydrolysis of cellulosic biomass. *Biofuels* **2,** 421–450 (2011).
21. Humbird, D. *et al. Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol.* tech. rep. NREL/TP-5100-47764 (National Renewable Energy Laboratory, Golden, CO, USA, 2011).
22. Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A. & Blanch, H. W. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnol. Bioeng.* **109,** 1083–1087 (2012).
23. Lynd, L. R. *et al.* How biotech can transform biofuels. *Nat. Biotechnol.* **26,** 169–172 (2008).
24. Hahn-Hagerdal, B., Karhumaa, K., Jeppsson, M & Gorwa-Grausland, M. F. Metabolic engineering for pentose utilization in Saccharomyces cerevisiae. *Adv. Biochem. Eng. Biotechnol.* **108,** 147–177 (2007).
25. Huber, G. W., Chheda, J. N. & Barrett, J. A. C. J.and Dumesic. Production of liquid alkanes by aqueous phase processing of biomass-derived carbohydrates. *Science* **308,** 1446–1450 (2005).
26. Wilson, D. B. Cellulases and biofuels. *Curr. Opin. Biotechnol.* **20,** 295–299 (2009).
27. Chundewat, S. P. S., Beckham, G. T., Himmel, M. E. & Dale, B. E. Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* **2,** 121–145 (2011).
28. Somerville, C. *et al.* Towards a systems approach to understanding plant cell walls. *Science* **306,** 2206–2211 (2004).
29. Wyman, C. E. What is (and is not) vital to advancing cellulosic ethanol. *Trends Biotechnol.* **25,** 153–157 (7).

30. Merino, S. T. & Cherry, J. Progress and challenges in enzyme development for biomass. *Adv. Biochem. Eng. Biot.* **108,** 95–120 (2007).

31. Aden, A. & Foust, T. Technoeconomic analysis of the dilute sulfuric acid and enzymatic hydrolysis process for the conversion of corn stover to ethanol. *Cellulose* **16,** 535–545 (2009).

32. Downing, L. *Clean fuel from trash, crop waste to match corn-ethanol by 2016* <`http://www.bloomberg.com/news/2013-03-12/clean-fuel-from-trash-crop-waste-to-match-corn-ethanol-by-2016.html`> (2013).

33. Aden, A., Ruth, M. & Ibsen, K. *Lignocellulosic biomass to ethanol process design and economics utilizing co-current dilute acid prehydrolysis and enzymatic hydrolysis for corn stover.* tech. rep. NREL/TP-510-32438 (National Renewable Energy Laboratory, Golden, CO, USA, 2002).

34. Goldemberg, J. Ethanol for a sustainable energy future. *Science* **315,** 808–810 (2007).

35. Moon, R. J., Martini, A., Nairn, J., Simonsen, J. & Youngblook, J. Cellulose nanomaterials review: structure, properties and nanocomposites. *Chem. Soc. Rev.* **40,** 3941–3994 (2011).

36. Jordan, D. B. *et al.* Plant cell walls to ethanol. *Biochem. J.* **442,** 241–252 (2012).

37. Ding, S.-Y. & Himmel, M. E. The maize primary cell wall microfibril: a new model derived from direct visualization. *J. Agric. Food Chem.* **54,** 597–606 (2006).

38. Gross, A. & Chu, J. W. On the molecular origins of biomass recalcitrance: the interaction network and solvation structures of cellulose microfibril. *J. Phys. Chem. B* **114,** 13333–13341 (2010).

39. Nishiyama, Y., Langan, P. & Chanzy, H. Crystal structure and hydrogen-bonding system in cellulose Iβ from synchrotron X-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* **124,** 9074–9082 (2002).

40. Bhat, M. K. Cellulases and related enzymes in biotechnology. *Biotechnol. Adv.* **18,** 355–383 (2000).

41. Zhang, Y.-H. P. & Lynd, L. R. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems. *Biotechnol. Bioeng.* **88,** 797–824 (2004).

42. Nieves, R. A., Ehrman, C. I., Adney, W. S., Elander, R. T. & Himmel, M. E. Technical communication: survey and analysis of commercial cellulase preparations suitable for biomass conversion to ethanol. *World. J. Microbiol. Biotechnol.* **14,** 301–304 (1998).

43. Geddes, C. C., Nieves, I. U. & Ingram, L. O. Advances in ethanol production. *Curr. Opin. Biotech.* **22,** 312–319 (2011).

44. Nidetzky, B. & Claeyssens, M. Specific quantification of Trichoderma reesei cellulases in reconstituted mixtures and its application to cellulase-cellulose binding studies. *Biotechnol. Bioeng.* **44,** 961–966 (1994).

45. Davies, G. & Henrissat, B. Structure and mechanisms of glycosyl hydrolases. *Structure* **3,** 853–859 (1995).

46. Teeri, T. T. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends Biotechnol.* **15,** 160–167 (1997).

47. Ståhlberg, J., Johansson, G. & Pettersson, G. A new model for enzymatic hydrolysis of cellulose based on the two-domain structure of cellobiohydrolase I. *Nat. Biotechnol.* **9,** 286–290 (1991).

48. Linder, M. *et al.* Identification of functionally important amino acids in the cellulose-binding domain of Trichoderma reesei cellobiohydrolase I. *Protein Sci.* **4,** 1056–1064 (1995).

49. Lehtiö, J. *et al.* The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc. Natl. Acad. Sci. USA* **110,** 484–489 (2003).

50. Divne, C. *et al.* The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from Trichoderma reesei. *Science* **265,** 524–528 (1994).

51. Beckham, G. T. *et al.* The O-glycosylated linker from the Trichoderma reesei family 7 cellulase is a flexible, disordered protein. *Biophys. J.* **99,** 3773–3781 (2010).

52. Kraulis, P. J. *et al.* Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from Trichoderma reesei. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. *Biochemistry* **28,** 7241–7257 (1989).

53. Mattinen, M.-L., Linder, M., Drakenberg, T. & Annila, A. Solution structure of the cellulose-binding domain of endoglucanase I from Trichoderma reesei and its interaction with cello-oligosaccharides. *Eur. J. Biochem.* **256,** 279–286 (1998).

54. Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J. K. C. & Jones, T. A. Three-dimensional structure of cellobiohydrolase II from Trichoderma reesei. *Science* **249,** 380–386 (1990).

55. Kleywegt, G. J. *et al.* The crystal structure of the catalytic core domain of endoglucanase I from Trichoderma reesei at 3.6 Å resolution, and a comparison with related enzymes. *J. Mol. Biol.* **272,** 383–397 (1997).

56. Divne, C., Ståhlberg, J., Teeri, T. T. & Jones, T. A. High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from Trichoderma reesei. *J. Mol. Biol.* **275,** 309–325 (1998).

57. Lin, Y. *et al.* Protein allostery at the solid-liquid interface: endoglucanase attachment to cellulose affects glucan clenching in the blinding cleft. *J. Am. Chem. Soc.* **133,** 16612–16624 (2011).

58. Varrot, A., Schülein, M. & Davies, G. J. Structural changes of the active site tunnel of Humicola insolens cellobiohydrolase, Cel6A, upon oligosaccharide binding. *Biochemistry* **38,** 8884–8891 (1999).

59. Zou, J.-y. *et al.* Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from Trichoderma reesei. *Structure* **7,** 1035–1045 (1999).

60. Bansal, P., Hall, M., Realff, M. J. R., Lee, J. H. & Bommarius, A. S. Modeling cellulase kinetics on lignocellulosic substrates. *Biotech. Adv.* **27,** 833–848 (2009).

61. Levine, S. E., Fox, J. M., Blanch, H. W. & Clark, D. S. A mechanistic model for the enzymatic hydrolysis of cellulose. *Biotechnol. Bioeng.* **107,** 37–51 (2010).

62. Liu, Y.-S. *et al.* Cellobiohydrolase hydrolyzes crystalline cellulose on hydrophobic faces. *J. Biol. Chem.* **286,** 11195–11201 (2011).

63. Imai, T., Boisset, C., Samejima, M., Igarashi, K. & Sugiyama, J. Unidirectional processive action of cellobiohydrolase Cel7A on Valonia cellulose microcrystals. *FEBS Letters* **432,** 113–116 (1998).

64. Igarashi, K. *et al.* Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333,** 1279–1282 (2011).

65. Srisodsuk, M., Kleman-Leyer, K., Keränen, S., Kirk, T. K. & Teeri, T. T. Modes of action on cotton and bacterial cellulose of a homologous endoglucanase-exoglucanase pair from Trichoderma reesei. *Eur. J. Biochem.* **251,** 885–892 (1998).

66. Kleman-Leyer, K. M., Siika-Aho, M., Teeri, T. T. & Kirk, T. K. The cellulases endoglucanase I and cellobiohydrolase II of Trichoderma reesei act synergistically to solubilize native cotton cellulose but not to decrease its molecular size. *Appl. Environ. Microbiol.* **62,** 2883–2887 (1996).

67. Boisset, C., Fraschini, C., Schülein, M., Henrissat, B. & Chanzy, H. Imaging the enzymatic digestion of bacterial cellulose ribbons reveals the endo character of the cellobiohydrolase Cel6A from Humicola insolens and its mode of synergy with cellobiohydrolase Cel7A. *Appl. Environ. Microbiol.* **66,** 1444–1452 (2000).

68. Meinke, A. *et al.* Enhancement of the endo-beta-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. *J. Biol. Chem.* **270,** 4383–4386 (1995).

69. Kyriacou, A., MacKenzie, C. R. & Neufeld, R. J. Detection and characterization of the specific and nonspecific endoglucanase of Trichoderma reesei: evidence demonstrating endoglucanase activity by cellobiohydrolase II. *Enzyme Microb. Technol.* **9,** 25–32 (1987).

70. Lantz, S. E. *et al.* Hypocrea jecorina CEL6A protein engineering. *Biotechnol. Biofuels* **3,** 20 (2010).

71. Varrot, A., Frandsen, T. P., Driguez, H. & Davies, G. J. Structure of the Humicola insolens cellobiohydrolase Cel6A D416A mutant in complex with a non-hydrolysable substrate analogue, methyl cellobiosyl-4-thio-β-cellobioside, at 1.9 Å. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58,** 2201–2204 (2002).

72. Henrissat, B., Driguez, H., Viet, C. & Schülein, M. Synergism of cellulases from Trichoderma reesei in the degradation of cellulose. *Nat. Biotechnol.* **3,** 722–726 (1985).

73. Kostylev, M. & Wilson, D. Synergistic interactions in cellulose hydrolysis. *Biofuels* **3,** 61–70 (2013).

74. Mansfield, S. D., Mooney, C. & Saddler, J. N. Substrate and enzyme characteristics that limit cellulose hydrolysis. *Biotechnol. Prog.* **15,** 804–816 (1999).

75. Sousa Jr., R., Carvalho, M. L., Giordano, R. L. C. & Giordano, R. C. Recent trends in the modeling of cellulose hydrolysis. *Braz. J. Chem. Eng.* **28,** 545–564 (2011).

76. Van Dyk, J. S. & Pletschke, B. I. A review of lignocellulosic bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes – factors affecting enzymes, conversion and synergy. *Biotechnol. Adv.* **30,** 1458–1480 (2012).

77. Xu, F. & Ding, H. A new kinetic model for heterogeneous (or spatially confined) enzymatic catalysis: Contributions form the fractal and jamming (overcrowding) effects. *Appl. Catal. A-Gen.* **317,** 70–81 (2007).

78. Bommarius, A. S. *et al.* Cellulase kinetics as a function of cellulose pretreatment. *Metab. Eng.* **10,** 379–381 (2008).

79. Igarashi, K., Wada, M., Hori, R. & Samejima, M. Surface density of cellobiohydrolyase on crystalline celluloses A critical parameter to evaluate enzymatic kinetics at a solid-liquid interface. *FEBS J.* **273,** 2869–2878 (2006).

80. Medve, J., Karlsson, J., Lee, D. & Tjerneld, F. Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from Trichoderma reesei: adsorption, sugar production pattern, and synergism of the enzymes. *Biotechnol. Bioeng.* **59,** 621–634 (1998).

81. Nidetzky, B. & Steiner, W. A new approach for modeling cellulase-cellulose adsorption and the kinetics of the enzymatic hydrolysis of microcrystalline cellulose. *Biotechnol. Bioeng.* **42,** 669–679 (1993).

82. Väljamäe, P., Sild, V., Petterson, G. & Johannsson, G. The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model. *Eur. J. Biochem.* **253,** 469–475 (1998).

83. Hong, J., Xinhao, Y. & Zhang, Y.-H. P. Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. *Langmuir* **23,** 12535–12540 (2007).

84. Jeoh, T. *et al.* Cellulose digestibility of pretreated biomass is limited by cellulose accessibility. *Biotechnol. Bioeng.* **98,** 112–122 (2007).

85. Berry, H. Monte Carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. *Biophys. J.* **83,** 1891–1901 (2002).

86. Eriksson, T., Karlsson, J. & Tjerneld, F. A model explaining the declining rate in hydrolysis of lignocellulose substrates with cellobiohydrolase I (Cel7A) and endoglucanase I (Cel7B) of Trichoderma reesei. *Appl. Biochem. Biotech.* **101,** 41–60 (2002).
87. Jalak, J. & Väljamäe, P. Mechanism of initial rapid rate retardation in cellobiohydrolase catalyzed cellulose hydrolysis. *Biotechnol. Bioeng.* **106,** 871–883 (2010).
88. Kurašin, M. & Väljamäe, P. Processivity of cellobiohydrolases is limited by the substrate. *J. Biol. Chem.* **286,** 169–177 (2011).
89. Praestgaard, E. *et al.* A kinetic model for the burst phase of processive cellulases. *FEBS J.* **278,** 1547–1560 (2011).
90. Cruys-Bagger, N. *et al.* Pre-steady state kinetics for the hydrolysis of insoluble cellulose by cellobiohydrolase Cel7A. *J. Biol. Chem.* **287,** 18551–18458 (2012).
91. Karlsson, J., Medve, J. & Tjerneld, F. Hydrolysis of steam-pretreated lignocellulose. *Appl. Biochem. Biotech.* **82,** 243–258 (1999).
92. Chanzy, H., Henrissat, B., Vuong, R. & Schülein, M. The action of 1,4-β-D-glucan cellobiohydrolase on Valonia cellulose microcrystals. *FEBS Letters* **153,** 113–118 (1983).
93. Wang, L. *et al.* Changes in the structural properties and rate of hydrolysis of cotton fibers during extended enzymatic hydrolysis. *Biotechnol. Bioeng.* **93,** 443–456 (2006).
94. Warden, A. C., Little, B. A. & Haritos, V. S. A cellular automaton model of crystalline cellulose hydrolysis by cellulases. *Biotechnol. Biofuels* **4,** 39 (2011).
95. Asztalos, A. *et al.* A coarse-grained model for synergistic action of multiple enzymes on cellulose. *Biotechnol. Biofuels* **5,** 55 (2012).
96. Fox, J. M., Levine, S. E., Clark, D. S. & Blanch, H. W. Initial- and processive-cut products reveal cellobiohydrolase rate limitations and the role of companion enzymes. *Biochemistry* **51,** 442–452 (2012).
97. Maurer, S. A., Bedbrook, C. N. & Radke, C. J. Cellulase adsorption and reactivity on a cellulose surface from flow ellipsometry. *Ind. Eng. Chem. Res.* **394,** 498–508 (2012).
98. Chatterjee, A. & Vlachos, D. G. An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *J. Computer-Aided Mater. Des.* **14,** 253–308 (2007).
99. Gillespie, D. T. A rigorous derivation of the chemical master equation. *Physica A* **188,** 404–425 (1992).
100. Gardiner, C. *Handbook of stochastic methods: for physics, chemistry and the natural sciences* (Springer, 2004).
101. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22,** 403–434 (1976).
102. Cao, Y., Gillespie, D. T. & Petzold, L. R. The slow-scale stochastic simulation algorithm. *J. Chem. Phys.* **122,** 014116 (2005).

103. Chang, R., Gross, A. S. & Chu, J. W. Degree of polymerization of glucan chains shapes the structure fluctuations and melting thermodynamics of a cellulose microfibril. *J. Phys. Chem. B* **116,** 8074–8083 (2012).

104. Frenkel, D. & Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications* 2nd ed. (Academic, 2002).

105. Humphrey, W., Dalke, A. & Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **14,** 33–38 (1996).

106. Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37,** 375–397 (2008).

107. Leduc, C. *et al.* Molecular crowding creates traffic jams of kinesin motors on microtubules. *Proc. Natl. Acad. Sci. USA* **109,** 6100–6105 (2012).

108. Gusakov, A. V. *et al.* Design of highly efficient cellulase mixtures for enzymatic hydrolysis of cellulose. *Biotechnol. Bioeng.* **97,** 1028–1038 (2007).

109. Quirk, A. *et al.* Direct visualization of the enzymatic digestion of a single fiber of native cellulose in an aqueous environment by atomic force microscopy. *Langmuir* **26,** 5007–5013 (2007).

110. Zhang, S. & Wilson, D. B. Surface residue mutations which change the substrate specificity of Thermomonospora fusca endoglucanase E2. *J. Biotechnol.* **57,** 101–113 (1997).

111. Koivula, A. *et al.* Tryptophan 272: an essential determinant of crystalline cellose degradation by Trichoderma reesei cellobiohydrolase Cel6A. *FEBS Letters* **429,** 341–346 (1998).

112. Von Ossowski, I. *et al.* Engineering the exo-loop of Trichoderma reesei cellobiohydrolase, Cel7A. A comparison with Phanerochaete chrysosporium Cel7D. *J. Mol. Biol.* **333,** 817–829 (2003).

113. Linder, M. & Teeri, T. T. The cellulose-binding domain of the major cellobiohydrolase of Trichoderma reesei exhibits true reversibility and a high exchange rate on crystalline cellulose. *Proc. Natl. Acad. Sci. USA* **93,** 12251–12255 (1996).

114. Levine, S. E., Fox, J. M., Blanch, H. W. & Clark, D. S. A mechanistic model for the rational design of optimal cellulase mixtures. *Biotechnol. Bioeng.* **108,** 2561–2570 (2011).

115. Wang, Q. Q., Zhu, J. Y., Hunt, C. G. & Zhang, H. Y. Kinetics of adsorption, desorption, and re-adsorption of a commercial endoglucanase in lignocellulosic suspensions. *Biotechnol. Bioeng.* **109,** 1965–1975 (2012).

116. Gillespie, D. T. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58,** 35–55 (2007).

117. Smit, B. & Maesen, T. L. M. Molecular simulations of zeolites: adsorption, diffusion, and shape selectivity. *Chem. Rev.* **108,** 4125–4184 (2008).

118. Jervis, E. J., A., H. C. & Kilburn, D. G. Surface diffusion of cellulases and their isolated binding domains on cellulose. *J. Biol. Chem.* **272,** 24016–24023 (1997).

119. Park, S., Baker, J. O., Himmel, M. E., Parilla, P. A. & Johnson, D. K. Cellulose crystallinity index: measurements techniques and their impact on interpreting cellulase performance. *Biotechnol. Biofuels* **3,** 10 (2010).

120. Zhang, S., Wolfgang, D. E. & Wilson, D. B. Substrate heterogeneity causes the nonlinear kinetics of insoluble cellulose hydrolysis. *Biotechnol. Bioeng.* **66,** 35–41 (1999).

121. Lee, Y. H. & Fan, L. T. Kinetic studies of enzymatic-hydrolysis of insoluble cellulase. (2). Analysis of extended hydrolysis times. *Biotechnol. Bioeng.* **25,** 939–966 (1983).

122. Holtzapple, M., Cognata, M., Shu, Y. & Hendrickson, C. Inhibition of Trichoderma reesei cellulase by sugars and solvents. *Biotechnol. Bioeng.* **36,** 275–287 (1990).

123. Caminal, G., Lopez-Santin, J. & Sola, C. Kinetic modeling of the enzymic hydrolysis of pretreated cellulose. *Biotechnol. Bioeng.* **27,** 1282–1290 (1985).

124. Converse, A. O., Matsuno, R., Tanaka, M. & Taniguchi, M. A model of enzyme adsorption and hydrolysis of microcrystalline cellulose with slow deactivation of the adsorbed enzyme. *Biotechnol. Bioeng.* **32,** 38–45 (1988).

125. Ye, Z. & Berson, R. E. Kinetic modeling of cellulose hydrolysis with first order inactivation of adsorbed cellulase. *Bioresource Technol.* **153,** 103–115 (2009).

126. Samant, A. & Vlachos, D. G. Overcoming stiffness in stochastic simulation stemming from partial equilibrium: a multiscale stochastic algorithm. *J. Chem. Phys.* **123,** 144114 (2005).

127. Moran-Mirabal, J. M., Bolewski, J. C. & Walker, L. P. Thermobifida fusca cellulases exhibit limited surface diffusion on bacterial micro-crystalline cellulose. *Biotechnol. Bioeng.* **110,** 47–56 (2012).

128. Yang, B. & Wyman, C. E. Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioprod. Bioref.* **2,** 26–40 (2008).

129. Puri, V. P. Effect of crystallinity and degree of polymerization of cellulose on enzymatic saccharification. *Biotechnol. Bioeng.* **26,** 1219–1222 (1984).

130. Sinitsyn, A. P., Gusakov, A. V. & Vlasenko, E. Y. Effect of structural and physicochemical features of cellulosic substrates on the efficiency of enzymatic hydrolysis. *Appl. Biochem. Biotechnol.* **30,** 43–59 (1991).

131. Zhang, Y.-H. P. & R., L. L. A functionally based model for hydrolysis of cellulose by fungal cellulase. *Biotechnol. Bioeng.* **94,** 888–898 (2006).

132. Goyal, A., Ghosh, B. & Eveleigh, D. Characteristics of fungal cellulases. *Bioresource Technol.* **36,** 37–50 (1991).

133. Vrsanská, M. & Biely, P. The cellobiohydrolase-I from Trichoderma-reesei QM-9414 - action of cello-oligosaccarides. *Carbohydr. Res.* **227,** 19–27 (1992).

134. Barr, B. K., Hsieh, Y.-L., Ganem., B. & Wilson, D. B. Identification of two functionally different classes of exocellulases. *Biochemistry* **35,** 586–592 (1996).

135. Wood, T. M. & McCrae, S. I. Synergism between enzymes involved in the solubilization of native cellulose. *Adv. Chem. Ser.* **181,** 181–209 (1979).

136. Nidetzky, B., Steiner, W., Hayn, M. & Claeyssens, M. Cellulose hydrolysis by the cellulases from Trichoderma reesei: a new model for synergistic interaction. *Biochem. J* **298,** 705–710 (1994).

137. Väljamäe, P., Sild, V., Pettersson, G. & Johansson, G. Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I. *Eur. J. Biochem.* **266,** 327–334 (1999).

138. Jalak, J., Kurašin, M., Teugjas, H. & Väljamäe, P. Endo-exo synergism in cellulose hydrolysis revisted. *J. Biol. Chem.* **287,** 28802–28815 (2012).

139. Ganner, T. *et al.* Dissecting and reconstructing synergism: in situ visualization of cooperativity among cellulases. *J. Biol. Chem.* **287,** 43215–43222 (2012).

140. Andersen, N. *et al.* Hydrolysis of cellulose using mono-component enzymes shows synergy during hydrolysis of phosphoric acid swollen cellulose (PASC) but competition on Avicel. *Enzyme Microb. Tech.* **42,** 362–370 (2008).

141. Jeoh, T., Wilson, D. B. & Walker, L. P. Effect of cellulase mole fraction and cellulose recalcitrance on synergism in cellulose hydrolysis and binding. *Biotechnol. Prog.* **22,** 270–277 (2006).

142. Boisset, C., Pétrequin, C., Chanzy, H., Henrissat, B. & Schülein, M. Optimized mixtures of recombinant Humicola insolens cellulases for the biodegradation of crystalline cellulose. *Biotechnol. Bioeng.* **72,** 339–345 (2001).

143. Medve, J., Ståhlberg, J. & Tjerneld, F. Adsorption and synergism of cellobiohydrolase I and II from Trichoderma reesei during hydrolysis of microcrystalline cellulose. *Biotechnol. Bioeng.* **44,** 1064–1073 (1994).

144. Fagerstam, L. G. & Petterson, L. G. The 1,4–β–glucan cellobiohydrolases of Trichoderma reesei QM9414. *FEMS Lett.* **119,** 97–101 (1980).

145. Tomme, P., Heriban, V. & Claeyssens, M. Adsorption of two cellobiohydrolases from Trichoderma reesei to Avicel: evidence for exo-exo synergism and possible loose complex formation. *Biotechnol. Lett.* **12,** 525–530 (1990).

146. Wood, T. M. & García-Campayo, V. Enzymology of cellulose degradation. *Biodegradation* **1,** 147–161 (1990).

147. Okazaki, M. & Moo-Young, M. Kinetics of enzymatic hydrolysis of cellulose: analytical description of a mechanistic model. *Biotechnol. Bioeng.* **20,** 637–663 (1987).

148. Dean, I. S. W. & Rollings, J. E. Analysis and quantification of a mixed exo-acting and endo-acting polysaccharide depolymerization system. *Biotechnol. Bioeng.* **39,** 968–976 (1992).

149. Fenske, J. J., Penner, M. H. & Bolt, J. P. A simple individual-based model of insoluble polysaccharide hydrolysis: the potential for autosynergism with dual-activity glycosidases. *J. Theor. Biol.* **199,** 113–118 (1999).

150. Converse, A. O. & Optekar, J. D. A synergistic kinetics model for enzymatic cellulose hydrolysis compared to degree-of-synergism experimental results. *Biotechnol. Bioeng.* **42,** 145–148 (1993).

151. Claeyssens, M., van Tilbeurgh, H., Tomme, P., Wood, T. M. & McCrae, S. I. Fungal cellulase systems: comparison of the specificities of the cellobiohydrolases isolated from Penicillium pinophilium and Trichoderma reesei. *Biochem J.* **261,** 819–825 (1989).

152. Baker, J. O., Ehrman, C. I., Adney, W. S., Thomas, S. R. & Himmel, M. E. Hydrolysis of cellulose using ternary mixtures of purified cellulases. *Appl. Biochem. Biotechnol.* **70–72,** 395–403 (1998).

153. Maurer, S. A., Bedbrook, C. N. & Radke, C. J. Surface kinetics for cooperative fungal cellulase digestion of cellulose from quartz crystal microgravimetry. *J. Colloid Interf. Sci.* **51,** 11389–11400 (2013).

154. Palonen, H., Tenkanen, M. & Linder, M. Dynamic interaction of Trichoderma reesei cellobiohydrolases Cel6A and Cel7A and cellulose at equilibrium and during hydrolysis. *Appl. Environ. Microbiol.* **65,** 5229–5233 (1999).

155. Carrard, G. & Linder, M. Widely different off rates of two closely related cellulose-binding domains from Trichoderma reesei. *Eur. J. Biochem.* **262,** 637–643 (1999).

156. Harjunpää, V. *et al.* Cello-oligosaccharide hydrolysis by cellobiohydrolase II from Trichoderma reesei: association and rate constants derived from an analysis of progress curves. *Eur. J. Biochem.* **240,** 584–591 (1996).

157. Koivula, A. *et al.* The active site of cellobiohydrolase Cel6A from Trichoderma reesei: the roles of aspartic acids D221 and D175. *J. Am. Chem. Soc.* **124,** 10015–10024 (2002).

158. Nidetzky, B., Steiner, W. & Claeyssens, M. Cellulose hydrolysis by the cellulases from Trichoderma reesei: adsorptions of two cellobiohydrolases, two endocellulases and their core proteins on filter paper and their relation to hydrolysis. *Biochem. J.* **303,** 817–823 (1994).

159. Tomme, P. *et al.* Studies of the cellulolytic systems of Trichoderma reesei QM 9414: analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur. J. Biochem.* **170,** 575–581 (1988).

160. Van Tilbeurgh, H., Bhikhabhai, R., Pettersson, L. G. & Claeyssens, M. Separation of endo- and exo- type cellulases using a new affinity chromatography method. *FEBS Lett.* **204,** 215–218 (1984).

161. Joseffson, P., Henriksson, G. & Wågberg, L. The physical action of cellulases revealed by a quartz crystal microbalance study using ultra thin cellulose films and pure cellulases. *Biomacromolecules* **9,** 249–254 (2008).

162. Wang, J. *et al.* Real-time observation of the swelling and hydrolysis of a single crystalline cellulose fiber catalyzed by cellulase 7B from Trichoderma reesei. *Langmuir* **28,** 9664–9672 (2012).

163. Arantes, V. & Saddler, J. N. Access to cellulose limits the efficiency of enzymatic hydrolysis: the role of amorphogenesis. *Biotechnol. Biofuels* **3,** 4 (2010).

164. Din, N. *et al.* C1-Cx revisited: intramolecular synergism in a cellulase. *Proc. Natl. Acad. Sci. USA* **91,** 11383–11387 (1994).

165. Gilkes, N. R. *et al.* Visualization of the adsorption of a bacterial endo-β-1,4-glucanase and its isolated cellulose-binding domain to crystalline cellulose. *Int. J. Biol. Macromol.* **15,** 347–351 (1993).

166. Cheng, G. *et al.* Neutron reflectometry and QCM-D study of the interaction of cellulases with films of amorphous cellulose. *Biomacromolecules* **12,** 2216–2224 (2011).

167. Gao, P. J., Chen, G. J., Wang, T. H., Zhang, Y. S. & Liu, J. Non-hydrolytic disruption of crystalline structure of cellulose by cellulose binding domain and linker sequence of cellobiohydrolase I from Penicillium janthinellum. *Shengwu Huaxue Yu Shengwu Wuli Xuebao* **33,** 13–18 (2001).

168. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J* **382,** 769–781 (2004).

169. Zhou, S. & Ingram, L. G. Synergistic hydrolysis of carboxymethyl cellulose and acid-swollen cellulose by two endoglucanases (CelZ and CelY) from Erwina chrysanthemi. *J. Biotechnol.* **182,** 5676–5682 (2000).

170. Walker, L. P., Wilson, D. B., Irwin, D. C., McQuire, C. & Price, M. Fragmentation of cellulose by the major Thermomonospora fusca cellulases, Trichoderma reesei CBH1, and their mixtures. *Biotechnol. Bioeng.* **40,** 1019–1026 (1992).

171. Fox, J. M. *et al.* A single-molecule analysis reveals morphological targets for cellulase synergy. *Nat. Chem. Bio.* **9,** 356–361 (2013).

172. Zarate, J. O. d. & Sengers, J. *Hydrodynamic Fluctuations in Fluids and Fluid Mixtures* (Elsevier, 2006).

173. Lum, K., Chandler, D. & Weeks, J. D. Hydrophobicity at small and large lengthscales. *J. Phys. Chem. B* **103,** 4570–4577 (1999).

174. Kadau, K. *et al.* The important of fluctuations in fluid mixing. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 7741–7745 (2007).

175. Donev, A., de la Fuente, A., Bell, J. B. & Garcia, A. L. Diffusive transport by thermal velocity fluctuations. *Phys. Rev. Lett.* **106,** 204501 (2011).

176. Moseler, M. & Landman, U. Formation, stability, and breakup of nanojets. *Science* **289,** 1165–1169 (2000).

177. Delgado-Buscalioni, R., Chacon, E. & Tarazona, P. Hydrodynamics of nanoscopic capillary waves. *Phys. Rev. Lett.* **101,** 106102 (2008).

178. Shang, B. Z., Voulgarakis, N. K. & Chu, J. W. Fluctuating hydrodynamics for multiscale simulation of inhomogeneous fluids: mapping all-atom molecular dynamics to capillary waves. *J. Chem. Phys.* **135,** 044111 (2011).

179. Hänggi, P. & Marchesoni, F. Artificial Brownian motors: Controlling transport on the nanoscale. *Rev. Mod. Phys.* **81,** 387–442 (2009).

180. Bocquet, L. & Charlaix, E. Nanofluidics, from bulk to interfaces. *Chem. Soc. Rev.* **39,** 1073–1095 (2010).

181. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* (Pergamon, 1959).

182. Hansen, J. P. & McDonald, I. R. *Theory of Simple Liquids* 3rd ed. (Academic Press, 2006).

183. Dünweg, B., Schiller, U. D. & Ladd, A. J. C. Statistical mechanics of the fluctuating lattice Boltzmann equation. *Phys. Rev. E* **76,** 036704 (2007).

184. Español, P. & Revenga, M. Smoothed dissipative particle dynamics. *Phys. Rev. E* **67,** 026705 (2003).

185. Davidovitch, B., Moro, E. & Stone, H. A. Spreading of viscous fluid drops on a solid substrate assisted by thermal fluctuations. *Phys. Rev. Lett.* **95,** 244505 (2005).

186. Sharma, N. & Patankar, N. A. Direct numerical simulation of the Brownian motion of particles by using fluctuating hydrodynamic equations. *J. Comput. Phys.* **201,** 466–486 (2004).

187. Donev, A., Bell, J. B., Garcia, A. L. & Alder, B. J. A hybrid particle-continuum method for hydrodynamics of complex fluids. *Multiscale Model. Simul.* **8,** 871–911 (2010).

188. Uma, B., Swaminathan, T. N., Radhakrishnan, R., Eckmann, D. M. & Ayyaswamy, P. S. Brownian motion and hydrodynamic interactions in the presence of flow fields. *Phys. Fluids* **23,** 073602 (2011).

189. Ahlrichs, P. & Dünweg, B. Simulation of a single polymer chain in solution by combining lattice Boltzmann and molecular dynamics. *J. Chem. Phys.* **111,** 8225 (1999).

190. Giupponi, G., de Fabritiis, G. & Coveney, P. V. Hybrid method coupling fluctuating hydrodynamics and molecular dynamics for the simulation of macromolecules. *J. Chem. Phys.* **126,** 154903 (2007).

191. Garcia, A. L., Malek Mansour, M., Lie, G. & Clementi, E. Numerical integration of the fluctuating hydrodynamic equations. *J. Stat. Phys.* **47,** 209–228 (1987).

192. Bell, J. B., Garcia, A. L. & Williams, S. A. Numerical methods for the stochastic Landau-Lifshitz Navier-Stokes equations. *Phys. Rev. E* **76,** 016708 (2007).

193. De Fabritiis, G., Serrano, M., Delgado-Buscalioni, R. & Coveney, P. V. Fluctuating hydrodynamic modeling of fluids at the nanoscale. *Phys. Rev. E* **75,** 026307 (2007).

194. Delgado-Buscalioni, R. & de Fabritiis, G. Embedding molecular dynamics within fluctuating hydrodynamics in multiscale simulations of liquids. *Phys. Rev. E* **76,** 036709 (2007).

195. Voulgarakis, N. K. & Chu, J. W. Bridging fluctuating hydrodynamics and molecular dynamics simulations of fluids. *J. Chem. Phys.* **130,** 134111 (2009).

196. Donev, A., Vanden-Eijnden, E., Garcia, A. L. & Bell, J. B. On the accuracy of finite-volume schemes for fluctuating hydrodynamics. *Communications in Applied Mathematics and Computational Science* **5,** 149–197 (2010).

197. *Coarse-Graining of Condensed Phase and Biomolecular Systems* (ed Voth, G. A.) (CRC Press, 2009).

198. Groot, R. D. & Warren, P. B. Dissipative particle dynamics: bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.* **107,** 4423 (1997).

199. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & Vries, A. H. d. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111,** 7812–7824 (2007).

200. Noid, W. G. *et al.* The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.* **128,** 244114 (2008).

201. He, W. X.and Shinoda, DeVane, R. & Klein, M. L. Exploring the utility of coarse-grained water models for computational studies of interfacial systems. *Mol. Phys* **108,** 2007–2020 (2010).

202. Izvekov, S. & Voth, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *J. Chem. Phys.* **125,** 151101 (2006).

203. Voulgarakis, N. K., Satish, S. & Chu, J. W. Modeling the nanoscale viscoelasticity of fluids by bridging non-Markovian fluctuating hydrodynamics and molecular dynamic simulations. *J. Chem. Phys.* **131,** 234115 (2009).

204. Voulgarakis, N. K., Satish, S. & Chu, J. W. Modeling the viscoelasticity and thermal fluctuations of fluids at the nanoscale. *Mol. Simulat.* **36,** 552–559 (2010).

205. Shang, B. Z., Voulgarakis, N. K. & Chu, J. W. Fluctuating hydrodynamics for multiscale modeling and simulation: energy and heat transfer in molecular fluids. *J. Chem. Phys.* **137,** 044117 (2012).

206. Berne, B. J. & Pecora, R. *Dynamic Light Scattering* (Wiley, 1976).

207. Fox, R. F. & Uhlenbeck, G. E. Contributions to non-equilibrium thermodynamics. I. Theory of hydrodynamical fluctuations. *Phys. Fluids* **13,** 1893–1902 (1970).

208. Allen, M. P. & Tildesley, D. J. *Computer Simulation of Liquids* (Clarendon Press, 1987).

209. Adams, D. J. Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid. *Mol. Phys.* **29,** 307–311 (1975).

210. Rubí, J. M. & Mazur, P. Nonequilibrium thermodynamics and hydrodynamic fluctuations. *Physica A* **276,** 477–488 (2000).

211. McQuarrie, D. *Statistical Mechanics* (University Science Books, 2000).

212. Zwanzig, R. Time-correlation functions and transport coefficients in statistical mechanics. *Annu. Rev. Phys. Chem.* **16,** 67–102 (1965).

213. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697 (1985).

214. Rowley, L. A., Nicholson, D. & Parsonage, N. G. Monte Carlo grand canonical ensemble calculation in a gas-liquid transition region for 12-6 argon. *J. Comput. Phys.* **17,** 401–414 (1975).

215. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

216. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23,** 327–341 (1977).

217. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103,** 4613–4621 (1995).

218. Schindler, M. A numerical test of stress correlations in fluctuating hydrodynamics. *Chem. Phys.* **373,** 327–336 (2010).

219. De Schepper, I. M. *et al.* Hydrodynamic time correlation functions for a Lennard-Jones fluid. *Phys. Rev. A* **38,** 271–287 (1988).

220. Palmer, B. J. Transverse-current autocorrelation-function calculations of the shear viscosity for molecular liquids. *Phys. Rev. E* **49,** 359–366 (1994).

221. Bertolini, D. & Tani, A. Generalized hydrodynamics and the acoustic models of water: theory and simulation results. *Phys. Rev. E* **51,** 1091–1118 (1995).

222. Schofield, P. Wave-length dependent fluctuations in classical fluids: I. the long wavelength limit. *Proc. Phys. Soc.* **88,** 149–170 (1966).

223. Chandler, D. *Introduction to Modern Statistical Mechanics* (Oxford University Press, 1987).

224. Sandler, S. I. *Chemical, Biochemical, and Engineering Thermodynamics* 4th ed. (John Wiley & Sons, Inc., 2006).

225. Hoogerbrugge, P. J. & Koelman, J. M. V. A. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhys. Lett.* **19,** 155–160 (1992).

226. Monaghan, J. J. Smoothed particle hydrodynamics. *Annu. Rev. Astron. Astrophys.* **30,** 543–574 (1992).

227. Español, P. Hydrodynamics from dissipative particle dynamics. *Phys. Rev. E* **52,** 1734 (1995).

228. Flekkøy, E. G. & Coveney, P. V. From molecular dynamics to dissipative particle dynamics. *Phys. Rev. Lett.* **83,** 1775–1778 (1999).

229. Grmela, M. & Öttinger, H. C. Dynamics and thermodynamics of complex fluids. I. Development of a general formalism. *Phys. Rev. E* **56,** 6620 (1997).

230. Öttinger, H. C. & Grmela, M. Dynamics and thermodynamics of complex fluids. I. Illustrations of a general formalism. *Phys. Rev. E* **56,** 6633 (1997).

231. Irving, J. H. & Kirkwood, J. The statistical mechanical theory of transport processes. IV. The equations of hydrodynamics. *J. Chem. Phys.* **18,** 817 (1950).

232. Evans, D. J. & Morriss, G. P. *Statistical Mechanics of NonEquilibrium Liquids* (Academic, 1990).

233. Boon, J. P. & Yip, S. *Molecular Hydrodynamics* (Dover, 1991).

234. Napoli, M., Eijkel, J. C. T. & Pennathur, S. Nanofluidic technology for biomolecular applications: a critical review. *Lab Chip* **10,** 957–985 (2010).

235. Whitelam, S., Feng, E. H., Hagan, M. F. & Geissler, P. L. The role of collective motion in examples of coarsening and self-assembly. *Soft Matter* **5,** 1251–1262 (2009).

236. Hohenberg, P. C. & Halperin, B. I. The theory of dynamic critical phenomena. *Rev. Mod. Phys.* **49,** 435–479 (1977).

237. Anderson, D. M., McFadden, G. B. & Wheeler, A. A. Diffuse-interface methods in fluid mechanics. *Annu. Rev. Fluid Mech.* **30,** 139–165 (1998).

238. Gompper, G. & Zschocke, S. Elatstic properties of interfaces in a Ginzburg-Landau theory of swollen micelles, droplet crystals and lamellar phases. *Europhys. Lett.* **16,** 731–736 (1991).

239. Davis, H. T. Capillary waves and the mean field theory of interfaces. *J. Chem. Phys.* **67,** 3636 (1977).

240. Bixon, M. & Zwanzig, R. Boltzmann-Langevin equation and hydrodynamic fluctuations. *Phys. Rev.* **187,** 267–272 (1969).

241. Salmon, R. Hamiltonian fluid mechanics. *Annu. Rev. Fluid Mech.* **20,** 225–256 (1988).

242. Cohen, C., Sutherland, J. W. H. & Deutch, J. M. Hydrodynamic correlation functions for binary mixtures. *Phys. Chem. Liquids* **2,** 213–235 (1971).

243. Lamura, A., Gonnella, G. & Yeomans, J. M. A lattice Boltzmann model of ternary fluid mixtures. *Europhys. Lett.* **45,** 314–320 (1999).

244. Rowlinson, J. S. & Widom, B. *Molecular Theory of Capillarity* (Clarendon, 1982).

245. Sedlmeier, F., Horinek, D. & Netz, R. R. Nanoroughness, intrinsic density profile, and rigidity of the air-water interface. *Phys. Rev. Lett.* **103,** 136102 (2009).

246. Du, Q., Freysz, E. & Shen, Y. R. Surface vibrational spectroscopic studies of hydrogen bonding and hydrophobicity. *Science* **264,** 826–828 (1994).

247. Taylor, R. S., Dang, L. X. & Garrett, B. C. Molecular dynamics simulations of the liquid/vapor interface of SPC/E water. *J. Phys. Chem.* **100,** 11720–11725 (1996).

248. Raymond, E. A., Tarbuck, T. L., Brown, M. G. & Richmond, G. L. Hydrogen-bonding interactions at the vapor/water interface investigated by vibrational sum-frequency spectroscopy of HOD/H20/D20 mixtures and molecular dynamics simulations. *J. Phys. Chem. B* **107,** 546–556 (2003).

249. Vega, C. & de Miguel, E. Surface tension of the most popular models of water by using the test-area simulation method. *J. Chem. Phys.* **126,** 154707 (2007).

250. Jäckle, J. & Kawasaki, K. Intrinsic roughness of glass surfaces. *J. Phys.: Condens. Matter* **7,** 4351–4358 (1995).

251. Grant, M. & Desai, R. C. Fluctuating hydrodynamics and capillary waves. *Phys. Rev. A* **27,** 2577–2584 (1983).

252. Harden, J. L., Pleiner, H. & Pincus, P. A. Hydrodynamic surface modes on concentrated polymer solutions and gels. *J. Chem. Phys.* **94,** 5208 (1991).

253. Lowengrub, J. & Truskinovsky, L. Quasi-incompressible Cahn-Hilliard fluids and topological transitions. *Proc. R. Soc. Lond. A* **454,** 2617–2654 (1998).