UC Berkeley
UC Berkeley Previously Published Works

CrossMark

Application of Molecular Simulations Toward Understanding Cellulase Mechanisms

Mansi Arora[1] · Ragothaman M. Yennamalli[1] · Taner Z. Sen[2,3]

Abstract

Computational approaches have provided new biological insights into the chemical mechanism of action of cellulases, which are used in the industrial production of bioethanol. Fine-grained methods, such as molecular dynamics and quantum mechanics, as well as coarse-grained methods, such as elastic network models, were used to investigate how the chemistry and structural dynamics of these enzymes contribute to their function. In this review, we highlight recent computational studies to understand this crucial biofuel enzyme class's chemistry and structural dynamics, as well as their significance in revealing enzymatic mechanism of action. Computational methods can complement and amplify the findings of experimental methods, which can be used in tandem to create more efficient industrial enzymes.

Keywords Cellulases · Molecular dynamics simulations · Multiscale modeling · Coarse-grained simulations · All-atom simulations · Biofuels

Introduction

Cellulases are a class of enzymes found in microbial life forms that break down cellulose and polysaccharides to obtain shorter (and sometimes, monomeric) polymer sugars. Cellulases play an important role in organisms, as they are a critical part of the metabolic pathways that confer the ability to obtain and use energy that sustain life. Without cellulases, life on Earth would not have existed.

✉ Ragothaman M. Yennamalli
ym.ragothaman@juit.ac.in; ragothaman@gmail.com

✉ Taner Z. Sen
taner.sen@ars.usda.gov

[1] Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, 2nd Floor, Vivekananda Bhavan, Waknaghat, Himachal Pradesh 173234, India

[2] Agricultural Research Service, Crop Improvement and Genetics Research Unit, U.S. Department of Agriculture, 800 Buchanan Street, Albany, CA 94710, USA

[3] Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA 50011, USA

In addition to their role in the cellular processes, driven by the search for new non-fossil-based alternative energy resources, cellulases are used in industrial ethanol production for fuel, as sugar molecules that are broken down by cellulases can be chemically converted into ethanol. Therefore, improving our understanding of the enzymatic mechanism and action of these amazing enzymes would not only satisfy our intellectual curiosity but also will help develop alternative energy resources for the humanity.

Cellulases and their associated proteins are, by itself, an exciting and active area of research. For example, the database Carbohydrate-Active enZymes (CAZy) is a niche database that has curated information for enzymes such as glycosyl hydrolases, glycosyl transferases, polysaccharide lyases, carbohydrate esterases, and auxiliary activity enzymes [1]. The role of cellulases is not limited to degradation of cellulose, but it also extends to plant energy storage, plant's life cycle, and others. Industrially, cellulases are widely used in paper and textile industries for processing recycled paper, enhancing softness of fabrics, and converting hardwood fibers to softer, malleable fibers for fine paper products [2]. In some detergents, cellulases are used as ingredients to make fabric appear brighter and relatively whiter. For these purposes, organisms such as *Trichoderma reesei* and *Clostridium thermocellum* are frequently used as a source for obtaining large quantities of biocatalysts, since their genome harbors multiple and diverse

cellulases, exo-glucanases, endoglucanases, and other genes of commercial importance [3].

In the case of biofuels, a diverse set of enzymes needs to be used during industrial processes, which act synergistically to break down the polymeric structure of cellulose to simple sugars. Because cellulolytic enzymes have different and unique mechanisms of action, a mix of enzymes is sometimes used industrially for the specific biomass to be degraded. With the second-, third-, and fourth-generation biomass for biofuel production, the determination of which enzymes and compositions will be most effective in converting biomass into simple sugars is a challenge, and computational methods were used to determine the optimal composition. Some of these computational methods include predicting the tertiary structure of enzyme and analyzing its structural dynamics to reveal its mechanism of action. Simulations of an enzyme structure, either by itself or in complex with substrate or another protein, are performed using a fine-grained (i.e., all-atom model) or a coarse-grained model. In either case, the trade-off is the level of details vs. computation time. There are distinct advantages and disadvantages of using each method, and researchers make their choice of appropriate details based on the model system and the hypothesis being tested.

Despite the importance of cellulases, however, our understanding into their structure, dynamics, and enzymatic action is limited, for this enzyme class. In this review, we will primarily focus on computational studies of this rich enzyme family, which complement experimental investigations and inform on molecular and structural mechanisms of enzymatic action. To accomplish this, we will first provide a brief description of computational methods followed by an introduction to cellulase enzymes, and then, we will review insights from computational studies in detail for specific cellulase families. The main objective of this review is to provide the overall landscape for the structural and dynamic studies of biofuel enzymes using computational methods to understand enzymatic action and demonstrate how the use of these methods advances our understanding of specific enzyme families.

Brief Review of Computational Methods Used in Cellulase Studies

Enzymes function optimally through physical and chemical arrangement of their structures, where the arrangement can be local, such as breaking of existing bonds to create new ones, or global, such as a conformational change causing distinct apoenzyme and holoenzyme structures. The local and global changes in a protein structure are usually dynamically coupled, and the exact coupling between these motions at different physical and temporal scales is an intense area of research. For example, some researchers argued that couplings between different dynamic modes are related to allosteric

behavior of proteins. Some recent computational studies expanded the classical definition of allostery, which is defined as a massive restructuring of protein shape upon ligand binding, to include coupled but localized motions of proteins, based on some experimental evidence, therefore arguing that allostery is not a behavior that is observed in some proteins, but it exists in various degrees in all proteins [4–10]. But, such theoretical interpretations often remain somewhat descriptive and lack physical and chemical specificity. In contrast, molecular dynamics simulations applied with a selective level of molecular details (fine-grained all-atom models vs. coarse-grained models) generate results that can be validated against experiments, such as single-molecule force spectroscopy [11] and small-angle X-ray scattering (SAXS) [12]. In this review, we will largely focus on computational methods related to molecular dynamics.

Coarse-Grained Vs. Fine-Grained Modeling

Granularity in computations is the extent to which an entity is divided into smaller groups of separable elements to facilitate computations. There are two main modeling approaches for biomolecular systems for granulation that incorporate different levels of atomistic details: coarse-grained and fine-grained.

In general, coarse-grained methods reduce computation time and can provide insights into molecular mechanisms at a longer time scales. However, coarse-grained approaches are limited in yielding details for subtle structural changes. For example, coarse-grained approaches can be used for studying cell components and processes for their interdependence, while atomically detailed simulations and molecular quantum mechanics can be applied to a region of few atoms.

A commonly used coarse-grained approach is the elastic network model (ENM) approach. These models use experimentally determined protein structures at equilibrium as input, and then, the fluctuations are calculated by simple mode decomposition using normal mode analysis. Although these models do not often provide atomistic details, their computational cost is usually much lower than molecular dynamics simulations [13]. In addition to elastic network models, other types of coarse-graining are applied to molecular dynamics in terms of energetics: for example, Martini force-field partition free energies between polar and apolar regions of chemical entities are applied to the studies of formation and fusion of vesicles, lamellar phase transformations, and membrane protein assemblies [14]; ELNEDIN, which is a physics-based coarse-grained model, uses a structural "scaffold" of protein that reduces the degree of freedom in the calculations [15].

Fine-grained methods include molecular dynamics (MD) simulations, quantum mechanics (QM), and quantum mechanics/molecular mechanics (QM/MM) methods. MD is

suitable to understand the detail of molecular and atomistic interactions that confer specificity to proteins. MD was applied to the simulations of polypeptide folding, biomolecular association, partitioning between solvents, membrane/micelle formation, chemical reactions and enzyme catalysis, photochemical reactions, and electron transfer. MD factors in degrees of freedom for motions, boundary conditions as to a system's temperature and pressure, and force field to generate dynamic trajectories by solving Newton's equations of motion to reveal structural mechanisms of enzymatic function [16].

In some cases, a combination of methods is used concurrently regardless of differences in time and space scales. These multiscale modeling approaches incorporate the application of both coarse-grained and fine-grained methods simultaneously, and use atomistic details of a small region (e.g., active site of a protein) and apply coarse-graining to the rest of the system to obtain the overall dynamics of a large complex to attempt to harness the advantages of both methods.

Constant-pH Molecular Dynamics

The constant-pH molecular dynamics (CpHMD) method involves determination of the protonation states of titratable sites in a protein at the specified pH. Detailed mechanistic studies of pH-dependent conformational processes use CpHMD to understand pH-coupled dynamical phenomena. The CpHMD method is capable of predicting experimental pK_a values and the pH-dependent conformational dynamics, but it is limited in modeling water titration and long-range interactions [17].

Thermodynamic Integration

The thermodynamic integration method computes free energy differences and other thermodynamic properties of the system between two distinct states/conformations. To measure the change from one state to another, parameters are slowly increased to maintain equilibrium at each stage in the trajectory [18]. The main requirement of the method is that the path should be reversible. Due to the chance of employing non-physical, but plausible paths (for example, allowing moves that increase free energy), the method confers great flexibility to molecular simulations. Both Monte Carlo and MD can be used with this method when only the equilibrium states along the path are needed to be simulated [18].

Metadynamics

Metadynamics is a sampling technique incorporating an additional bias potential that acts on multiple but selective number of degrees of freedom as *collective variables*. Methods that use this class include umbrella sampling and steered MD. The approach pushes the system away from local free energy minima and therefore allows exploring new reaction pathways. Moreover, no prior knowledge of the energetics of a system is needed to implement metadynamics techniques. The limitation is that because the model oscillates around the free energy rather than converging to it, it is difficult to decide as to when to stop the simulation. Second, identifying a set of collective variables for describing complex processes is very difficult to achieve [19].

Continuum-Molecular Dynamics

The continuum-molecular dynamics method generalizes simulated tempering to a continuous temperature space to provide a smooth transition from microscale to macroscale [20] by evaluating a conserved quantity that can be used to validate a simulation.

Quantum and Molecular Mechanics

The QM and MM are part of quantum chemistry toolbox [21]. Quantum mechanical descriptions are used to model accurate electronic rearrangements for those parts of a system that are involved during a chemical reaction, but quantum modeling is computationally expensive. Though less accurate, MM, on the other hand, is faster and computationally less costly. For simulations that do not involve a chemical reaction, the use of a simple MM force-field model is appropriate, which reduces the simulation time. To overcome the limitations of a full quantum mechanical or a full molecular mechanics description, the hybrid QM/MM methods are an option, in which the system is treated in part at the level of quantum chemistry (QM), retaining the computationally cheaper force field (MM) for the larger part [22].

Monte Carlo Methods

Monte Carlo methods are useful in the study of thermodynamics of the protein over its conformational space and also in searching for the low-energy conformations. The major limitation of Monte Carlo is that it is a data-intensive method [23]. A broad class of Monte Carlo algorithms, mean field particle methods, simulates a sequence of probability distribution for a non-linear evolution equation [24]. In contrast to Monte Carlo, this technique uses sequentially interacting samples, in a way that is similar to statistical Markov processes.

Simulated Annealing

Simulated annealing (also known as generalized simulated annealing) is often used when a large search space with many local minima needs to be studied by using a probabilistic approach to approximate the global optimum of a given function. With respect to enzymes, the approach involves the

system to be heated up to a high temperature then gradually cooling to find the global minima. When the energy landscape at the needed temperature is smooth, however, this method is limited, because it cannot identify an optimal solution [25].

Structural and Functional Description of Cellulase Families and Insights from Computational Studies

We have divided this section into various subsections in terms of the enzymes used for studying the protein dynamics. We have compiled and tabulated the MD parameters used in the studies cited below in Supplementary Table T1 that the reader will find useful.

Family 7 Cellobiohydrolase

Most of these enzymes cleave β-1,4-glycosidic bonds in cellulose and β-1,4-glucans, which are its substrates. According to Koshland [26], there are two classifications based on the mechanism of action of the enzymes, i.e., retaining and inverting, out of which the family 7 cellobiohydrolase is classified as retaining enzymes. The cellobiohydrolase's end product is cellobiose, a disaccharide, which is one step away from converting the cellulose polymer to glucose units (Figs. 1a, b and 2a). Computational studies on cellobiohydrolase (Cel7A, Cel7B) have focused on the overall structural dynamics either as a free enzyme or in complex with various substrates. For

example, the simulations for Cel7B from *Melanocarpus albomyces* were performed at a constant pH to analyze the dynamic fluctuations of the loop regions [27], because the loop regions of cellulases, in general, play a major role in enzymatic function [27]. The CpHMD works by computationally coupling the protonation states of some amino acids within the framework of classical MD and capturing the residue pK_a shifts and dynamic charge coupling. In the MD simulations of 70 ns with 2-ps steps across various pH levels, the loops showed differential fluctuations. At active pH, loop II showed increased flexibility compared to other pH levels. When CpHMD was performed for *T. reesei* Cel7A, a well-studied enzyme of industrial importance, the loop regions showed flexibility [28] that correlated with the neutron scattering experiments. Interestingly, whereas the presence of charged residues (Asp and His) in *Melanocarpus albomyces* is thought to contribute to the elevated pH profile compared to *T. reesei* Cel7A [28], these residues are not present in the loop region; therefore, the enzymatic function of this protein is an interplay of amino acid specificity and structural dynamics, possibly as a result of dynamics of loop regions on the protein surface. Separate comparisons of tunnel entrance at specific sites, loops near a specific subsite, loops near the catalytic center, and comparisons of product binding region with respect to other GH7 cellobiohydrolases revealed significant structural differences relating to differences in processivity, endo-initiation, and product inhibition [67]. Comparing *T. reesei*'s Cel7A with Cel7A of *Trichoderma harzianum* suggests that although they share high sequence homology (81%
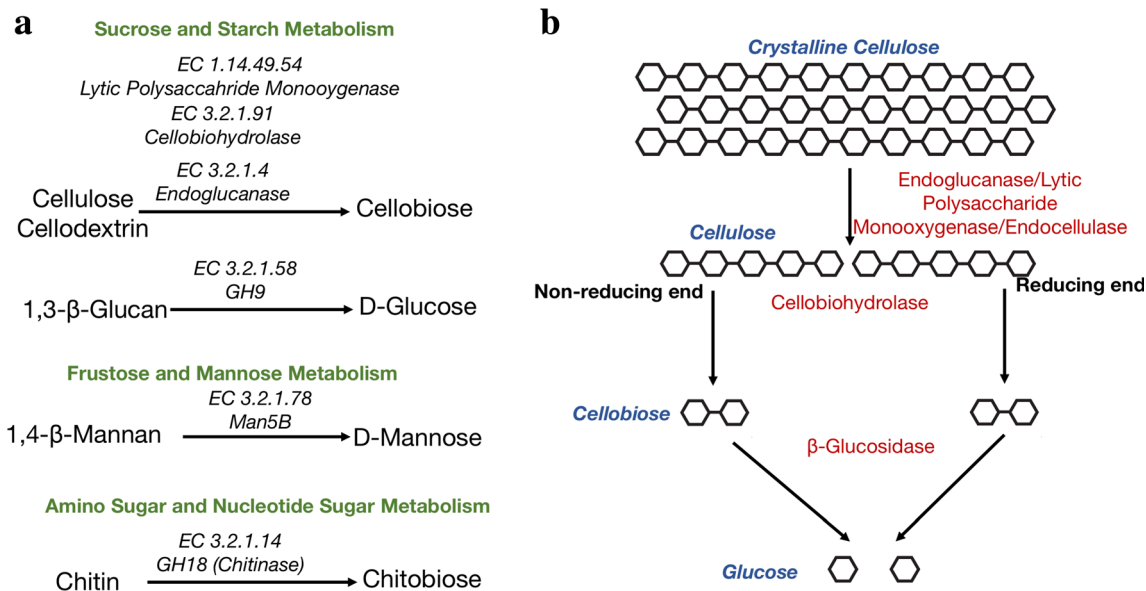

Fig. 1 Metabolic pathways and breakdown of crystalline cellulose. a The enzymes described in this study are present in KEGG as part of the sucrose and starch metabolic pathway, namely endoglucanase, cellobiohydrolase, and lytic polysaccharide monooxygenase. Also, GH9 is part of the same metabolic pathway but break downs 1,3-β-glucan. Similarly, Man5B is part of the fructose and mannose metabolic pathway, and GH18 is part of the amino sugar and nucleotide sugar metabolic pathway. **b** Schematic representation of enzymes involved in the conversion of cellulose to glucose

sequence identity), the short side chains of the adjacent residues in the catalytic tunnel create extra gaps at the side face of the catalytic tunnel [40].

The dynamics of a substrate-bound enzyme is significantly different than the free form. The Cel7A from *Geotrichum candidum* strain 3C (GcaCel7A) was subjected to MD simulation for 100 ns in three different setups: (1) in the free form, (2) in complex with a long substrate (cellononaose), and (3) in complex with the microfibril of cellulose [31]. In these three different MD simulations, GcaCel7A shows similar structural and functional characteristics to the industrially relevant HjeCel7A (from *Hypocrea jecorina*) (Table 1). On the other hand, the ligand-bound MD simulations on *T. reesei* Cel7A (TrCel7A) showed that there exists a competitive binding in the presence of lignin, which is a known inhibitor and a by-product of the plant biomass during the pre-processing step. Successful removal of lignin ensured high turnover for the enzyme catalyst. Using a mix of crystalline and non-crystalline fibers and in the presence of 468 lignin

molecules, the 1312-ns-long simulation indicated that the hydrophobic surface of the cellulose is the preferred binding site for both lignin and *Tr*Cel7A [29]. Lignin also was observed to bind to the hydrophobic patches of the carbohydrate-binding module (CBM) attached with *Tr*Cel7A to amplify its inhibitory activity. Similarly, MD simulation on the PfCBH1 (cellobiohydrolase belonging to GH7 family) from *Penicillium funiculosum* with microcrystalline cellulose revealed that the binding of substrate is relatively more accessible due to structural flexibility when compared with *T. reesei* and this might have a role in relatively faster product expulsion, thus a higher tolerance for product inhibition, i.e., cellobiose [30]. Study on Cel7A and Cel7B of *T. reesei* with the help of mutants suggests the structural differences between both the enzymes around the catalytic center, at active site tunnel entrances, and exits, all of which signify the processivity in GH7s [39]. Also, Cel7B catalytic domain of *T. reesei* with a cellulose microfibril revealed as to the domain's complexation on cellulose chains from a crystal surface [33].
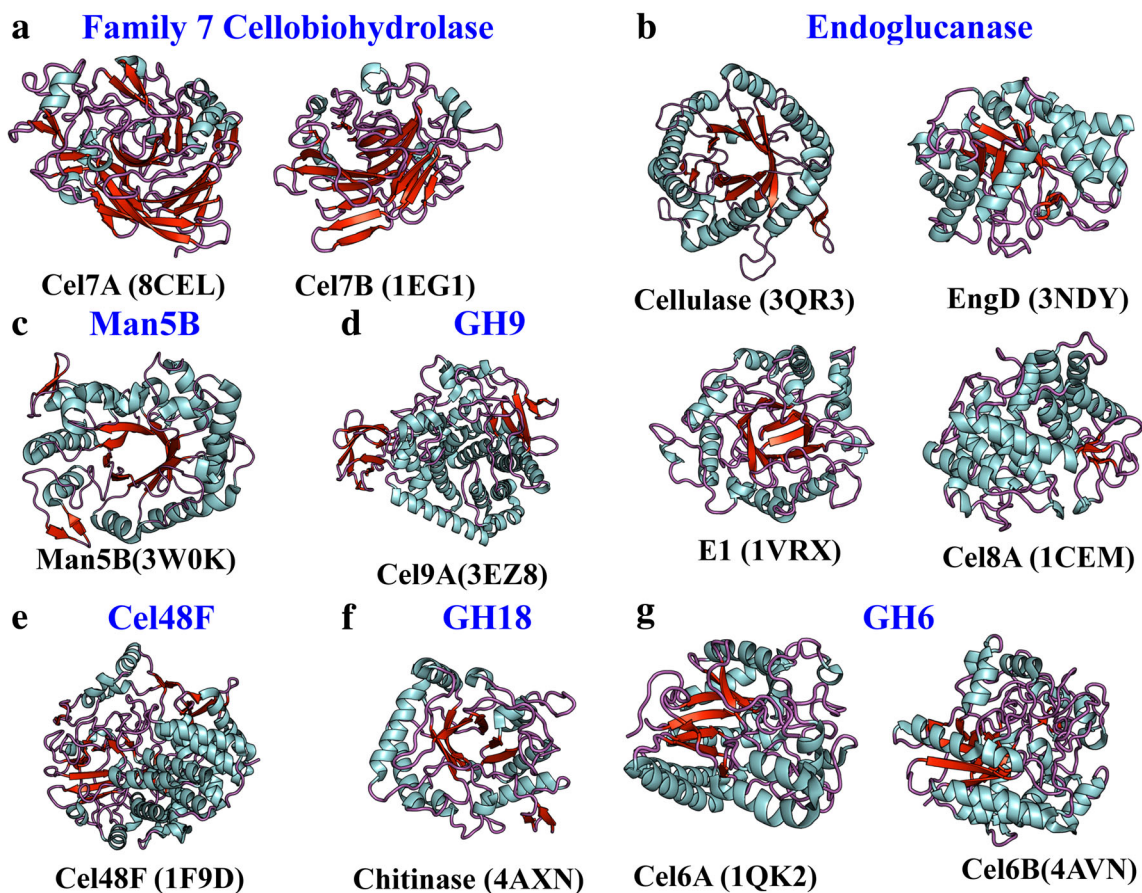

Fig. 2 Compilation of biofuel enzymes studied using computational methods. Cartoon representation of the structures mentioned in this review (Table 1), where the helices are colored cyan, beta-strands are colored red, and loops are colored magenta. **a** Family 7 cellobiohydrolase (pdb id: 8cel, 1eg1). The protein corresponding to pdb id: 8cel is prepared from homology modeling, and we have used the same coordinates of the modeled structures as used originally for carrying out the simulations. **b** Endoglucanase (pdb id: 3qr3 and 3ndy). **c** Man5B (pdb id: 3w0k). **d** GH9 (pdb id: 3ez8). **e** Cel48F (pdb id: 1f9d). **f** GH18 (4axn). **g** GH6 (pdb id: 1qk2 and 4avn)

Table 1 Enzymes used in biofuel application reviewed in this paper, detailing the MD simulation time

Family	Enzyme	PDB	Structure from	MD production run	References
Family 7 cellobiohydrolase	Cel7B	2RFW	Crystallography	70 ns	[27]
	Cel6A, Cel7A	8CEL	Theoretical model	200 ns	[28]
	Cel7A	8CEL	Theoretical model	1312 ns	[29]
	CBH1	–	Homology model from 8CEL	100 ns	[30]
	Cel7A	4ZZV	Crystallography	100 ns × 3	[31]
	Cel7A	1CBH and mutants	NMR spectroscopy	20 ns	[32]
	Cel7B	1EG1	Crystallography	100 ns	[33]
	Cel7A	8CEL 5CEL 4C4C	Theoretical model Crystallography Crystallography	500 ns	[34]
	Cel7A	8CEL	Theoretical	20–30 ns × 5 (1 for reach binding position)	[35]
	Cel7A (*Hir*Cel7A) Cel7A (*Hje*Cel7A) Cel7D (*Pch*Cel7D)	2YG1, 2XSP 8CEL 1GPI	Crystallography Theoretical model Crystallography	250 ns	[36]
	Cel7A Cel6A	8CEL 1QK2	Theoretical model	100 ns × 4	[37]
	Cel7A Cel7B	8CEL 1EG1	Theoretical model Crystallography	250 ns	[38]
	Cel7A Cel7B	8CEL 1EG1	Crystallography Crystallography	250 ns	[39]
	Cel7A (*Tr*Cel7A) Cel7A (ThCel7A)	8CEL 2YOK	Theoretical model Crystallography	60 ns	[40]
LPMO	Cel6A, Cel7A	1QK2	Crystallography	500 ns	[41]
	GH61D	–	Homology model from 3EII	100 ns	[42]
	ScLPMO10B ScLPMO10C	4OY6 4OY7	Crystallography Crystallography	8000 ns	[43]
	GH61D	4B5Q	Crystallography	10 ns	[44]
Cellulosome	Cellulosome	4 IU3	Crystallography	450 ns	[45]
	Cellulosome	3KCP	Crystallography	100 ps	[46]
	Type I cohesin–dockerin complex	1OHZ (D39N)	Crystallography	6 ns	[47]
	Type II cohesin–dockerin complex	2B59	Crystallography	20 ns	[48]
Cel48F	Cel48F	1KWF	Crystallography	10 ns	[49]
	Cel48F	1F9D mutant (Q55E)	Crystallography	5 ns	[50]
	Cel8A	1KWF	Crystallography	50 ns	[51]
Endoglucanase	Cellulase	3QR3 mutant (E54D) 3AMD 3AXX	Crystallography	500 ns	[52]
	E1	1VRX	Homology model from 1VRX	500 ns	[53]
	EngD	3NDY	Crystallography		[54]
	Endoglucanase 3	4H7M 1H8V	Crystallography	120 ns (each)	[55]
	Cel5A	–	Homology from 3QR3	100 ns	[56]
	Cel7B	1EG1	Crystallography	15 ns × 3	[57]
	Cel8A	1CEM (G283P)	Crystallography	4 ns	[58]
	Cel5A	3QR3	Crystallography	100 ns	[59]
	EG	1KS4	Crystallography	32 ns	[60]
GH46	Cellulase	5E0C 5E09	Crystallography	50 ns	[61]

Table 1 (continued)

Family	Enzyme	PDB	Structure from	MD production run	References
GH18	Chitinase	3IAN	Crystallography	250 ns	[62]
		4AXN	Crystallography		
GH9	Cel9A-68	1TF4	Crystallography	50 ns	[63]
	Cel9A	3EZ8	Crystallography	10 ns	[64]
GH6	Cel6B	4AVO	Crystallography	260 ns	[65]
		4AVN	Crystallography		
Man5B	Hexaose–Man5B complex	1GB1	NMR	40 ns	[66]
		1L2Y	NMR		
		3W0K	Crystallography		

Similar dynamics on substrate-bound form revealed the cellulose binding site was highly conserved in three enzymes when bound with cellononaose: Cel7A from *Heterobasidion irregulare* (*Hir*Cel7A), *H. jecorina* (*Hje*Cel7A), and Cel7D from *P. chrysosporium* (*Pch*Cel7D) [36]. HjeCel7A, this time in complex with a cellodextrin nanomer chain, was placed at five different positions around the binding site, and the cellodextrin chain was observed to spontaneously diffuse into the catalytic tunnel by a cellobiose unit [35]. Further, it was suggested by means of potential mean force calculations that the Cel7A recognizes the free cellulose-reducing chain end [35]. In one case, it was suggested that the glycosylation reaction serves as the rate-limiting step in cellulose degradation [34]. A two-step simulation protocol that was implemented to observe the binding and interaction of Cel7A's CBM with the cellulose Iβ fiber showed the binding preference of CBM toward the hydrophobic faces of the fiber rather than the hydrophilic ones via a 40-ns-long simulation [68]. In a related study, the flexible, glycosylated linkers of CBM bound to *T. reesei*'s TCel6A and TrCel7A were shown to bind non-specifically to the cellulose surface [37].

To study the interactions between the important residues of *T. reesei* Cel7A CBM and cellulose, a thermodynamic integration method was used to calculate the cellulose–Cel7A CBM binding free energy changes caused by Y5A, N29A, Y31A, Y32A, and Q34A mutations (pdb id: 1cbh) to demonstrate that interactions between residues and cellulose are dominated by the electrostatic changes [32]. The Cel7A from *T. reesei* was studied with MD to understand the structure–function relationships that glycosylation imparts to linkers. The enzyme is an intrinsically disordered protein most likely due to the absence of ordered secondary structure for the linker, as validated via 360-ns-long simulations. It was reported that the Cel7A linker is comparatively more disordered than other linkers in *T. reesei* cellulases [69].

MD simulations were used to examine the binding of cellobiose to the TrCel7A cellobiohydrolase and the effects of mutations that reduce cellobiose binding without affecting the structural integrity of the enzyme. The results showed that the binding site of the product demonstrates a specific flexibility that can hinder the cellobiose release sterically, though many point mutations can still maintain the structural integrity of the enzyme. It was suggested that there is a trade-off between inhibition of the product and the efficiency of the catalyst [70].

The unique Cel7B from the marine wood borer, *Limnoria quadripunctata*'s ability to operate in saline conditions was investigated with a 250-ns-long MD simulation that showed high flexibility of the exo-loops at the tunnel. The tolerance to high salt concentration is probably due to acidic charge distribution on its surface, and the aromatic residues at the entrance of the tunnel may be involved in substrate binding [38].

Endoglucanases

Endoglucanases were broadly studied with respect to two types: those that contain CBM domains and the ones that do not. Endoglucanase D (EngD) from *Clostridium cellulovorans* consists of a catalytic domain linked via a flexible linker to a CBM domain (Figs. 1a, b and 2b). While computational methods were not used to study the enzyme's structural dynamics, SAXS experiments revealed the flexibility of the linker that allows an extended conformation of EngD in the solution, which proved the importance of the CBM module. The cellotriose-bound EngD structure has an extended active-site cleft, which contains Trp162, a residue that is absent in few other variants of the enzyme with a significantly reduced activity [54]. Endoglucanase 3 of *T. reesei* (TrEG3) and *T. harzianum* (ThEG3), a member of GH12 enzymes, does not contain the CBM domain, and yet it catalyzes cellulose hydrolysis [55]. The tertiary structure of ThEG3 (at 2.07 Å resolution) was determined by X-ray crystallography, and then, MD simulations were used to investigate enzyme–substrate interactions to understand the role that certain aromatic residues play in recognizing and binding to the substrate. The study showed that due to the significant spatial distance of this *CBM-like cluster* region from the catalytic site, the productive substrate binding and catalytic efficiency require longer oligosaccharide chains to simultaneously bind to the catalytic triad and the aromatic CBM-like cluster for efficient hydrolysis. The study

highlighted the reason as to why shorter oligosaccharides have inefficient hydrolysis by Cel12A [55].

To understand how ionic liquids (ILs) interact with enzymes at the molecular scale, endoglucanase (E1) from *Acidothermus cellulolyticus* was simulated in aqueous 1-ethyl-3-methylimidazolium chloride ([Emim]Cl) to study potential inactivation mechanisms. The study showed that the utility of ionic liquids is highly restricted by the enzyme incompatibility and the interactions that are crucial to activation or inactivation of the results are unique. For example, [Emim]Cl interacts with higher specificity to E1's binding site and disrupts native hydrophobic contacts, leading to inactivation of E1 [53]. A similar study on GH5 family of endoglucanases from *Trichoderma viride*, *Thermogata maritima*, and *Pyrococcus horikoshii* in the presence of 1-ethyl-3-methyl-imidazolium acetate ([EMIM][OAc]) with water at various temperatures was carried out [52]. The results did show that structural changes that happen at a long time scale (500 ns) cause deactivation of the enzymes. For example, in GH5 of *T. viride*, the deformation of binding pocket is correlated with the deactivation at low concentrations of the IL. Similarly, the deactivation of GH5 of *T. maritima* is due to changes in secondary structures that correlate with experimental data. However, the GH5 of *P. horikoshii* did not show any deactivation at low concentrations of IL [52].

A trimodular endoglucanase (CelB) with a CBM46 domain and a rigid CBM_X domain sandwiched between them was identified in *Bacillus* sp. BG-CS10, where the CBM46 domain interacts both with the catalytic domain and CBM_X domain. The resulting structure is labeled as an L-shaped cellulase. MD simulations for 50 ns indicated that the loop regions of the catalytic domain that contain the aromatic residues involved in substrate binding undergo relatively large structural changes, facilitating product release [61].

The computational studies for the BGLI gene include homology modeling for prediction of its tertiary structure from multiple sequence alignment with three selected templates of β-glucosidase. MD simulations were performed on the docked structure of BGLI with cellulose ligand to identify the ligand-binding domain of the enzyme. Stable conformations that were observed were in agreement with the structural flexibility of the free enzyme [71].

A 3D model for VpEXPA2, an α-expansin involved in the softening of *Vasconcellea pubescens* fruit, was built by comparative modeling strategy based on the structure of Phlp1 as template, a β-allergen from Timothy grass (*Phleum pratense*). Docking studies were performed to predict the putative binding of different octasaccharides to the protein: two different hemicellulose octasaccharides and one cellodextrin 8-mer that resembles a water-soluble cellulose molecule. MD simulations were carried out for each substrate inside VpEXPA2 for 20 ns which showed a strong interaction to cellodextrin 8-mer polymer and, in contrast, a low interaction with hemicelluloses octasaccharide polymers. It is reasonably hypothesized that the function of domain D1 of VpEXPA2 is highly dependent on the binding of cellulose microfibril to domain D2 [72].

The atomistic simulations based on classical interaction potentials were used to examine the interactions of Cel5A with cellulose fibrils with amorphous-like and non-crystalline regions. The analysis of the catalytic domain suggests that the enzyme actually alters the cellulose structures and the charge around the catalytic cleft in the domain plays a significant role in enzymatic function [56].

Another structural study involved the prediction of the putative hydrogen bonds formed between the enzymes and cellohexaose using homology models of *Nf*EG12A from *Neosartorya fischeri* P1 and endoglucanase from *Aspergillus niger*. MD simulations were carried out to examine the effect of loop 3 on the catalytic efficiency of GH12 endoglucanases. Overall, the analysis through molecular mechanics/Poisson–Boltzmann and surface area continuum salvation (MM/PBSA) demonstrated that the hydrogen network interactions between protein and the substrate are enhanced by loop 3, resulting to an increase in the turnover rate, thereby improving the catalytic efficiency [60].

From a created library of consensus mutation using the sequence alignment of homologs of family 8 glycoside hydrolases, one of the mutants of Cel8A from *Clostridium thermocellum* showed a higher thermal stability without any loss of catalytic activity, possibly due to an increase in conformational rigidity of the protein backbone in unfolded state that was observed in an MD simulation of the enzyme's mutant (G283P) [58].

The sequence and structural comparison studies were performed for the CBM of three endoglucanases (EG1, EGO, and EGV) modeled from *T. reesei* cellobiohydrolase CBHI. All the structures were found to be similar in their cellulose-binding domains, and disulfide bridges seem to stabilize the polypeptide fold [73].

A *Tr*Cel5A cellulose complex was simulated with MD using *Tr*Cel5A-catalyzed phosphate acid-swollen cellulose (PASC) hydrolysis as a model system. The slowdown of hydrolysis was studied by kinetic measurements, and it was found that the hydrolysis slowdown is correlated with the adsorption. The simulations significantly helped in identifying the potential residues involved in binding. The results from the comparative analysis of the complex with the wild type further showed that catalytic ability affects the slowdown of endoglucanase [59].

The structure, dynamics, and behavior of the Cel7B from *Fusarium oxysporum* were analyzed at 80 °C through MD simulations. The dynamical factors for analysis involved hydrogen bonds and fluctuations in the turn regions, which influence the activity and stability of the enzyme [74].

Another study using MD simulations identified the regions in proteins that trigger the partial unfolding for denaturation (also known as "weak spots"). These regions were identified in *T. reesei* Cel7B by calculating the distances between Cα in contact and their capability of forming disulfide bonds [57].

Lytic Polysaccharide Monooxygenases

Lytic polysaccharide monooxygenases (LPMOs) are recently discovered enzymes (Fig. 1a) that show immense industrial application for degrading crystalline form of cellulose, as they boost the degradation process significantly [75]. Their importance and their relevance were described in detail in the following reviews [41, 76, 77]. It is now well established that LPMOs can be classified into three subtypes based on the site of attack, namely (1) LPMO1 when oxidation occurs at C1 carbon, (2) LPMO2 when oxidation occurs at C4 carbon, and (3) LPMO3 if either C1 or C4 carbons are attacked (Fig. 3). Additionally, there are four CAZy families to which LPMOs are classified as auxiliary activity enzymes (AA9, AA10, AA11, and AA13) on the basis of their potential abilities to help the originally classified enzymes, the glycoside hydrolases (GHs), the polysaccharide lyases (PLs), and the carbohydrate esterases (CEs), in gaining access to the carbohydrates encrusted in the plant cell wall [78].

Molecular insights into LPMO's mechanism of action were significantly improved by computational studies ranging from QM/MM models to multiscale models (Fig. 3). The QM/MM models that were built using the density functional theory on LPMO belonging to the AA9 family from *Thermoascus aurantiacus* shed light on the geometry and coordination chemistry of the reactive oxygen with Cu(II) atom. The results indicated that the formation of the complex (copper–oxyl reactive oxygen species) drives the catalytic activity with a rebound step for oxygen to complete the cycle [79].

Another study informs us about the four-coordinate tetragonal structure of *T. aurantiacus* in an oxidized state and a three-coordinate T-shaped structure in a reduced state [80]. The O_2 reactivity of the Cu(I) site was evaluated computationally using experimentally calibrated DFT calculations. To determine the number and type of coordinating ligands in Cu-AA9, extended X-ray absorption fine structure (EXAFS) experiments, which allows gathering information about atomic energy-level structures and the metallic center coordination, were performed on the oxidized and reduced enzyme forms to demonstrate that the structure of the enzyme site is suitable for rapid inner-sphere reductive activation of O_2 by Cu(II)–superoxide formation [80].

MD simulations of LPMO from the AA9 family revealed that the loop regions undergo conformational changes that make the enzyme flexible during substrate binding. These findings are in agreement with the QM/MM results where the distance between the active site copper and C1 carbon is around 5 Å, where a superoxide intermediate of the reaction (a product of the reactive oxygen with Cu(II) atom) can be easily accommodated. The tyrosines (Y28, Y75, and Y198) were computationally observed to form local hydrophobic interactions, stabilizing the active site during substrate binding [42].

Fig. 3 Schematic representation of C1 (type 1) and C4 (type 2) mechanism of action identified in lytic polysaccharide monooxygenases (LPMOs). LPMOs oxidize on either C1 or C4 carbon, giving rise to specific products, lactone or ketoaldose, respectively
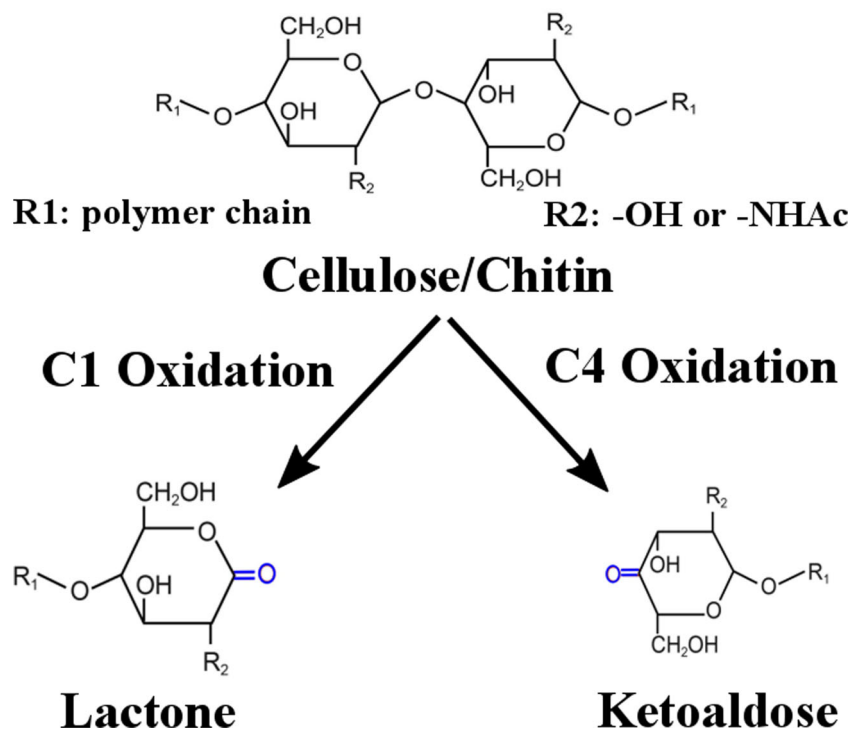
In order to accurately capture the enzyme dynamics using MD simulations, the use of accurate force fields for a given system is required. A recent study probed potential energy landscape for the AA9 family to create a specific set of force-field parameter [44]. The use of such accurate force fields that can represent metallo-proteins consists of single-point energy evaluations over a rectangular grid involving selected internal coordinates that incorporate the generation of energy profiles for the bond stretch, angle bend, and torsions for more realistic simulations. In recent years, the method of multiscale modeling was applied to study the large-scale dynamics of proteins and their interactions with substrates. The advantage of using multiscale modeling is that it gives the *big picture* of the interactions between the different components of a system. In the case of cellulases, the global level dynamics of cellulases on the surface of cellulose can shed light into how the complex synergistic activities of different enzymes help in degrading cellulose. A multiscale modeling of LPMO with Cel7A (non-reducing end specific exo-cellulase) and Cel7B (reducing end specific exo-cellulase) of *T. reesei* showed that LPMO decrystallizes the cellulose crystalline surface by forming new chain termini within the fibril, rather than at the ends of the fibril. It also has higher affinity to the reducing end of the fibril [41]. This multiscale modeling study emphasizes the possible synergistic interaction between LPMO and other enzymes for faster degradation of crystalline cellulose.

The reduction of the LPMO active site of AA9 enzyme from *T. aurantiacus* from states 1 (resting state) to 2 (reduced state) and two isomers of state 3 (copper–superoxide intermediate) was recently investigated (Fig. 4a) [81]. The results of combined QM/MM simulations provided evidence that the computational protocols that were followed in this study could reproduce the observed decrease in the coordination number when Cu(II) is reduced to Cu(I). Using QM for this system as opposed to full MM was a necessity because MM cannot model reactions. Among the two isomers that were observed in the Cu–superoxide complex, the multiscale modeling revealed that there is a preference for one isomer over the other for energetic stability. Further work on the enzyme–substrate complex from the same group led to the validation of four enzyme–substrate intermediate models based on bond-dissociation energy (BDE) [82]. BDE calculations are time consuming in an experimental setup, and thus, the alternative method of calculating BDEs from computational methods is quicker and sensitive. Specifically, in the LPMO studied (pdb id: 2yet [83]), the bond-dissociation energy for the four intermediates, $[Cu–OH]^{3+}$, $[Cu–OH]^{2+}$, $[Cu–O]^{2+}$, and $[Cu–O]^{+}$, is comparable; however, the intermediate $[Cu–OH]^{3+}$ is not favorable compared to the other three. The study also highlighted the non-dependency of the aromatic residue in the active site, as many LPMOs have either a Tyr or a Phe at the same position [82].

The MD simulations once again prove their worth in identifying key areas that deviate from the crystal structures of ScLPMO10B and ScLPMO10C LPMOs to identify surface charge modifications to increase stability in ILs. The MD was performed for 250 ns in three ILs at 0 wt%, 10 wt%, and 20 wt% in water. The IL effects of dynamic fluctuations for specific regions of the enzyme on exposure to ionic liquid, on enzyme's overall structure, as well as on the structure of enzyme's active site were comprehensively and comparatively studied for both the LPMOs. The results clearly indicate that they both show structural similarity, and the fluctuations in IL and water are nearly the same. Therefore, both the LPMOs are unaffected by the influence of ionic liquids [43].
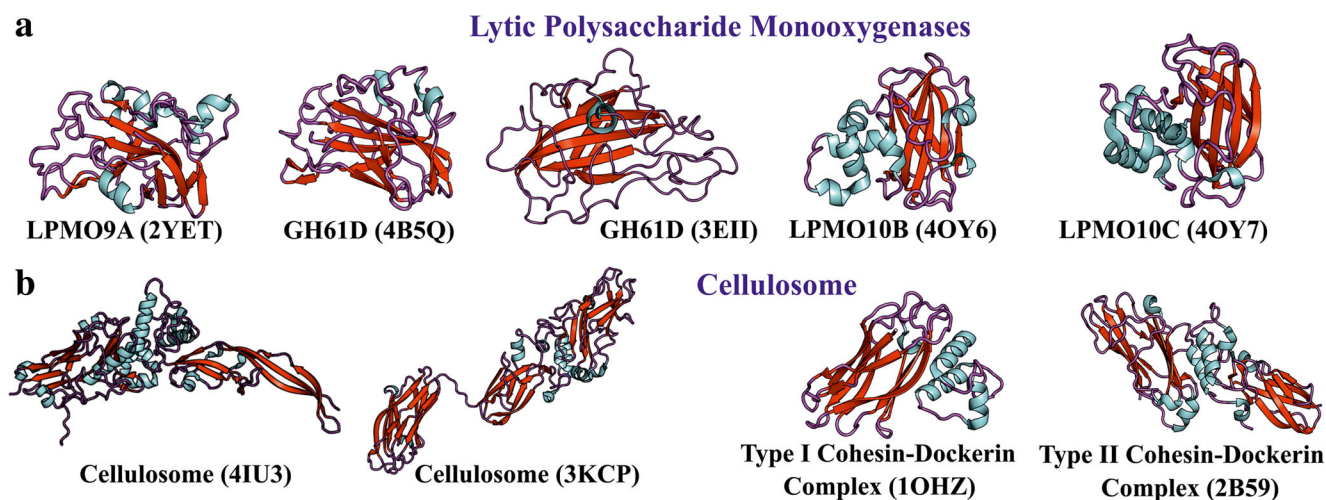
a Lytic Polysaccharide Monooxygenases

LPMO9A (2YET) GH61D (4B5Q) GH61D (3EII) LPMO10B (4OY6) LPMO10C (4OY7)

b Cellulosome

Cellulosome (4IU3) Cellulosome (3KCP)

Type I Cohesin-Dockerin Complex (1OHZ) Type II Cohesin-Dockerin Complex (2B59)

Fig. 4 **Compilation of biofuel enzymes studied using computational methods.** Cartoon representation of the structures mentioned in this review (Table 1), where the helices are colored cyan, beta-strands are colored red, and loops are colored magenta. **a** LPMO (pdb id: 2yet, 4b5q, 3eii, 4oy6, 4oy7). **b** Cellulosome (pdb id: 4iu3, 3kcp, 1ohz, 2b59). The protein corresponding to pdb id: 3EII is prepared from homology modeling, and we have used the same coordinates of the modeled structures as used originally for carrying out the simulations

To study the functional aspects of CBP21, a chitin-active member of carbohydrate-binding module family, NMR, and isothermal titration calorimetry were used to map surface binding based on pH dependency, which showed that CBP21 is a compact and rigid molecule except at its catalytic metal binding site. CBP21 depends on Cu ion for catalysis, and binding of cyanide to the metal indicates that it is involved in the oxidative cleavage of the substrate. The comparisons with GH61 LPMO further showed that their metal binding sites are significantly different despite the fact that both catalyze the same reaction. An approach that uses the pH dependency of both the chitin–CBP21 interaction and the 1H exchange rate led to the identification of the residues involved in binding CBP21 to the chitin surface based on the first NMR structure ever resolved for an LPMO [84].

Cellulosome

Cellulosomes are macromolecular complexes that are specialized in cellulose degradation. The flexible linkers that connect dockerins and cohesins in the cellulosome gained much attention due to their contribution to the structural dynamics of the enzyme. The cellulosome dynamics was studied by generalized simulated annealing (GSA) on a fragment of *C. thermocellum* CipA (Fig. 4b) scaffolding in complex with the SdbA type II cohesion module. The study revealed that the CohI9 module (CipA's ninth type I cohesion) has only two possible conformations (two thirds of occurrences of native form and one third of alternate form), despite the fact that the linker is highly flexible. Further MD simulation analysis showed that the small difference in the average potential energy between the two conformations can be overcome by the small changes in thermal energy, therefore affording the module the ability to easily switch between both conformations [46].

The X-modules-dockerin and cohesin complex (XMod-Doc:Coh) was studied to characterize the ligand–receptor complex responsible for substrate anchoring and interdomain stabilization in *Ruminococcus flavefaciens* whereby single-molecule force spectroscopy and steered molecular dynamics simulations examine the mechanical unbinding of the complex. The mechanical dissociation of XMod-Doc:Coh was probed by single-molecule force spectroscopy, the results of which show that xylanase fusion domain on XMod-Doc and CBM fusion domain on Coh show identifiable unfolding patterns. This allowed screening of large datasets of force-distant curves. The XMod-Doc:Coh ruptures reported there fell in a range from 600 to 750 pN at loading rates ranging from 10 to 100 nN s^{-1}, which were among the highest of their kind ever reported. The steered molecular dynamics results indicated that the force increased with distance continuously until the complex was ruptured. The analysis for the interacting residues and the contact surface area suggested that

the mechanism of such stability is remarkable while still allowing fast assembly and disassembly of the complex at equilibrium [45].

MD simulations were performed to probe both the type I and type II coh-Xdoc interactions in *C. thermocellum*. They involve the simulations and free energy calculations of both wild type and D39N mutant of the type I coh-Xdoc from the same organism, the results of which are a clear indication that comparatively, the mutant shows significant flexibility caused by the change in hydrogen-bonding network in the conserved loop regions. The energy differences demonstrate that though dynamic changes are small, the conformational changes persist [47].

In another study for the type II, *hot spots*, i.e., the amino acid residues responsible for drastic decrease in binding affinity upon mutation, were mutated to examine their effect on binding. The study concluded that the rigid cohesion–dockerin interface is maintained by means of bulky and hydrophobic residues and their contacts with the protein interface [48].

C. thermocellum was studied for capturing the physical characteristics of three cellulosomal enzymes (Cel5B, CelS, and CbhA) and the scaffoldin (CipA) by MD simulations. The results showed that shape and modularity dominate the cellulosomal enzyme complex. Comparative insights about the abovementioned enzymes indicated that CbhA binds more frequently to CipA than the other two because of its flexible nature multimodularity [85].

Coarse-grained and MD were performed on many cellulosomal linkers of different lengths and compositions, which indicated that the linker's stiffness depends on the length, and not the specific amino acid. The study showed that the short and stiff linkers are the cause of significant rearrangements in the folded domains of the mini-cellulosome composed of endoglucanase Cel8A in complex with scaffoldin ScafT (Cel8A-ScafT) of *C. thermocellum* as well as in a two-cohesin system derived from the scaffoldin ScaB of *Acetivibrio cellulolyticus* [86].

Man5B

In the MD computational analysis of Man5B (Figs. 1a and 2c), the enzyme from thermophilic bacteria *Caldanaerobius polysaccharolyticus*, molecular docking studies followed by principal component analysis were performed on the catalytic site bound with cellohexaose and mannohexaose to understand the mechanism by which Man5B hydrolyzes cello-oligosaccharide and manno-oligosaccharide substrates. The results showing Man5B binding to cellohexaose as tightly as mannohexaose were significant because the experimental assays showed that Man5B is relatively more efficient in hydrolyzing manno-oligosaccharides than on gluco-oligosaccharides [87].

Applying coarse-grained simulation on protein–oligosaccharide complex, where glucose is approximated to one bead (an approach similar to commonly used approximation of representing one amino acid as one bead), Poma et al. [66] constructed coarse-grained models for three different hexaoses and then tested it on a Man5B–hexaose complex. The predicted structural models correlated well with all-atom models reported earlier for the same system, and the analysis suggested that the interaction of Man5B with hexaose is four times stronger than the other oligosaccharides.

Another coarse-grained (CG) method application involved the use of Martini force field, which applied a mapping of four heavy atoms to one CG interaction site and was parameterized with the aim of reproducing thermodynamic properties. To overcome the barrier of unbreakable harmonic bonds controlling unfolding and folding processes, the ELNEDIN protein model [88] was based on the Martini CG force field, where the harmonic bonds were replaced with Lennard–Jones interactions on the contact map of the native protein structure as is done in Gō-like models. This model revealed the structural motion linked to a particular catalytic activity in the Man5B protein, the details of which agreed with those of all-atom simulations. The approach made use of the contact map, which identified the key pairs of contacts between residues required to preserve the native structure of the protein without the need for using adjustable parameters [88].

In another study, two coarse-grained models of three hexaoses were studied. One of the models was based on centers of mass and C4 atoms. The second one was based on $C\alpha$ atoms, and found more appropriate to analyze protein interactions. The corresponding stiffness constants were calculated by all-atom simulations and two statistical methods (Boltzmann inversion and energy-based). It was found that the energy-based method shows a better agreement with other theoretical and experimental determinations of non-bonded parameters. The contact energies were then calculated in the hexaose–Man5B complex, and the interactions of $C4–C_\alpha$ atoms were found to be stronger than the hydrogen bonds [66].

GH9

The cleavage of sugar chains from cellulose at high temperatures by the thermoresistant Cel9A-68 (Figs. 1a and 2d) from *Thermobifida fusca* is catalyzed by the cooperative action of two important domains of the cellulase: CBM and a catalytic domain connected by a Pro/Ser/Thr-rich linker. Based on this, the temperature dependence of the dynamics of Cel9A-68 was analyzed in detail at three temperatures: 300 K, 325 K (optimal temperature for activity), and 350 K. Using quasi-harmonic analysis, principal component analysis (PCA), and subsequent essential dynamics (ED) analysis, the conformational space and the collective motions were examined, and

the CBM domain was observed to be highly flexible than the catalytic domain as observed in experiments [63].

The Ig-like domain in GH9, if deleted, causes the loss of enzymatic activity, though there is no evidence of any direct relation with the active site. MD simulations were used to investigate the role of Ig-like domain in Cel9A. The results show that the residues of the domain are correlated dynamically with the residues of carbohydrate-binding pocket, with few of them being related to the important catalytic residues of Cel9A. Further, it was shown that the catalytic domain is significantly stabilized by the Ig-like domain, possibly enhancing the thermostability of Cel9A [64].

Cel48F

The processive endocellulase, Cel48F of *C. cellulolyticum*, was studied for its hydrolysis mechanism when it forms a complex with the sugar chains by MD simulations (Fig. 2e). The computational approach for examining the structure of Cel48F involved metadynamics, which computed free energies expeditiously and allowed the study of statistically rare events. Metadynamics simulations usually follow standard MD simulations to stabilize protein–polysaccharide structures. Metadynamics proved its utility in investigating the details of sugar chain I entering and chain II leaving the Cel48F tunnel through a 5-ns-long simulation [50].

For the study of *water control mechanism* in enzymatic hydrolysis of cellulose, MD simulations were carried out for the two conformations of the Cel48F [hydrolyzing (H) and sliding (S)]. These two conformations were compared after repeating the MD simulations thrice with the same starting conformations. The hydrolysis seemed to begin when a water molecule is present for every glycosidic linkage, suggesting a water control mechanism for hydrolysis. During the shifts between conformation from S to H, the water molecule that is initially bound to D230 in S (known as site a) turns to W417 and M414 (known as site b) in H and performs a nucleophilic attack on the anomeric carbon, causing the hydrolysis product to be excluded from the cleft and water control system to return to site a. The simulations revealed the roles of other certain key residues through their ability to form hydrogen bonds. Key residues involved the most probable candidates for inverting anomeric carbon, the ones that could help converting one conformation to the other and those that could provide a hydrophobic environment preventing the water molecules from entering the active site. It was proposed that in addition to Cel48F, the method can be applied to study the reaction mechanisms in other processing enzymes [49].

MD simulations were carried out for imidazole in an aqueous solution of glucose in order to investigate the interactions that take place between the two co-solutes and also between the neutral imidazole molecules. This study showed the role of

histidine side chains in the binding and hydrolysis in cellulases, including Cel48F [89].

The simulations showed the possible catalytic role of an unusual conserved water-filled pore structure in another member of the family, Cel48A from *T. fusca*, suggesting that the pore provides a pathway to the active site for the water molecules used in processive hydrolysis of the cellulose substrate [90].

Catalytic activity was studied in *C. cellulolyticum* by MD in combination with steered MD and binding free energy calculations, which gave insights about the important regions of the Cel48F that are involved in hydrolysis and the release process of the leaving group. The probable residues responsible for hydrolysis, which affect the catalytic activity significantly, were predicted [51].

A study for analyzing the mechanisms of cellulosomes used MD and normal mode analysis to refine the protein complex and investigated the dynamical differences between the domains. After determining the structure experimentally using SAXS, normal mode analysis confirmed that both the free dockerin and the dockerin–cohesin complexes undergo a rigid body motion with respect to the catalytic module [91].

GH18

A study reported the crystal structure and dynamics of the catalytic domain of the GH family 18 non-processive endochitinase, ChiC, from *Serratia marcescens* (Fig. 2f) with other processive enzymes, ChiA and ChiB from *S. marcescens* [62]. The study demonstrated that the dynamics of the processive enzymes is similar to that of a non-processive chitinase from *Lactococcus lactis* (pdb id: 3IAN; a structural homolog of ChiC's catalytic module). All four proteins were docked with the chitin substrate to study the processivity of the chitinase. Each simulation was run for 250 ns, for a total simulation time of 1 μs. The overall structure of ChiC2 (the catalytic domain of ChiC) was studied in terms of energy difference, hydrogen bonding, and root-mean-square distance.

The catalytic residue in ChiC2, i.e., E141, was observed to be in a different orientation in the crystal structure and not bonded with D139, which would be crucial for optimal catalytic activity. In order to understand the interactions of E141 and whether the side chain conformations change in the active site, MD simulations were performed and showed that E141 is indeed flexible and there are three distinct conformational states for E141 and D139. Similarly, between the processive and non-processive chitinases, there were structural dynamics differences shown through the free energy-based calculations to confirm the conformational flexibility of Glu141, the loop regions, and in the active site. These simulations showed that the ChiC2 is highly flexible, and the dynamic *on and off* ligand binding processes

associated with non-processive endochitinases correlate well to the experimentally derived processivity data in the *S. marcescens*'s chitinases [62].
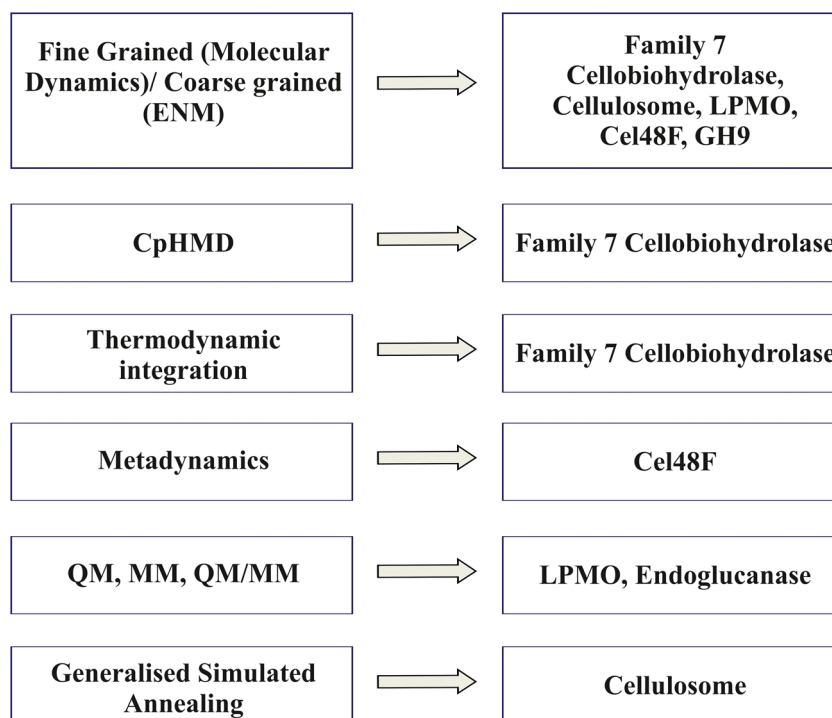
GH6

The simulations reveal new structural details of GH6 CBHs (Fig. 2g). Two mutant structures of *Thermobifida fusca*'s Cel6B were characterized that allowed the analysis of their hydrolysis mechanisms. Using the wild-type Cel6B structure, three complexes were constructed. Two complexes were bound to cellobiose, and the remaining with cellohexaose. In addition, a Cel6B complexed with crystalline cellulose Iβ was also constructed. These complexes were built to study the tunnel-shaped active site. Specifically, the process of product expulsion was identified by the dynamic action of two loops, i.e., the exit loop (residues 185–197) and the bottom loop (residues 501–510). The simulations suggest that driven by their flexibility, these loops open up to create space to expulse the product from the active site. These two loop regions fluctuate the largest and are not correlated with the fluctuations of other parts of the protein. Multiple SMD simulations showed the exit loop opening up to 14 Å and the bottom loop up to 16 Å, creating a large enough gap for allowing product transport. The simulations show the flexibility of the loops to open and allow release of the product with equal probability in solution or when bound to cellulose [65].

Discussion

Enzyme design is one of the most complex engineering areas in chemical engineering. Multiple steps are necessary for an enzyme to be modified in order to become an industrially useful product. In addition, temperature and pressure values in different processing stages influence each enzyme's stability, activity, and conversion rate. Each step is complicated enough to require extensive experimental testing, starting from small-scale experiments on the bench to testing in pilot plants. Success in the lab does not always translate into success in chemical plants. Many initially exciting products fail to reach to the production line. In multistep enzyme design and production stages, candidate enzymes are extensively and experimentally tested to demonstrate their superior performance against specific industrially meaningful conditions.

Given the importance of experiments in providing a realistic view of enzymatic performances, computational methods will not replace the experimental testing that provides dependable measures for enzymatic behavior in the near future. Computational methods, however, can be as useful as imperfect theoretical models which provide practical approximations that can guide experimental approaches. Although experiments are the ultimate touchstone to judge the performance of a novel

Fig. 5 Computational methods to understand dynamics in cellulases. Brief summary of the various fine-grained and coarse-grained methods employed to understand the structural dynamics of various families of cellulases as reviewed in this paper

Fine Grained (Molecular Dynamics)/ Coarse grained (ENM)	⟹	Family 7 Cellobiohydrolase, Cellulosome, LPMO, Cel48F, GH9
CpHMD	⟹	Family 7 Cellobiohydrolase
Thermodynamic integration	⟹	Family 7 Cellobiohydrolase
Metadynamics	⟹	Cel48F
QM, MM, QM/MM	⟹	LPMO, Endoglucanase
Generalised Simulated Annealing	⟹	Cellulosome

enzyme product, simulations and theoretical methods are important parts of engineers' toolsets to shape design approaches and optimize enzymatic behavior in industrial settings.

As we have discussed in this review, several computational methods are available to understand the molecular basis of enzymatic activity in cellulases. Different methods have different strengths and weaknesses. Given the computational cost of fine-grained atomistic simulations, coarse-grained approaches are preferentially used to speed up the simulations, but increased speed comes with a lower information content, sometimes as unrealistically as at the expense of lack of atomistic-level modeling of chemical reactions. Ultimately, before starting a research project, engineers and scientists need to determine critical parameters to optimize for, so that they can choose the most appropriate combination of tools from available experimental, theoretical, and computational methods to achieve their design goals.

When multiple approaches are used simultaneously, deciding on the effectiveness of each method individually is not an easy task. Obviously, experimental methods provide the most clear-cut product enhancements, especially for first few generations of enzymes. For example, randomly mutating enzymes and assessing their properties is a well-tested means for product development. For the development of later generation of enzymes, however, more sophisticated approaches are necessary, and a wider range of tools and a deeper understanding of enzymatic mechanisms are desired. Therefore, although computational methods often lack clear-cut success stories, their critical role by being a part of a larger engineering toolset when creating newer generation of enzymes is usually understood.

Conclusions

In this review, we surveyed how various computational approaches were used to understand how the structural dynamics and chemical specificity of cellulases contribute to enzymatic function. Complementary to experimental methods, computational methods, such as MD simulations and/or QM/MM method, and other methods were successfully used for many cellulases to understand the molecular underpinnings of their functions. By reviewing the recent scientific literature for the cellulase enzymes, we provided the latest computational research in structural and dynamics studies of these industrially important enzymes.

Numerous studies have demonstrated that computational methods are fast and reliable and provide the level of detail required to understand the enzymatic function of cellulases (Fig. 5). While the prerequisite for any such computational study is the availability of high-resolution structures, solved via NMR or X-ray crystallography, we do not perceive this as a limitation; because the speed at which even the low-resolution new protein structures are being deposited in structural databases, this would increasingly enable better computer simulations and analyses. As more computational studies will be performed in the future, a better understanding of mechanism of enzymatic action for cellulases will be developed, enabling scientist and engineers to make more informed design decisions for more efficient use of cellulases in biofuel applications.

Nomenclature AA, Auxiliary activity; BDE, Bond-dissociation energy; CBM, Carbohydrate-binding module; CE, Carbohydrate esterase; CG, Coarse-grained; CpHMD, Constant-pH molecular dynamics; EngD, Endoglucanase D; EXAFS, Extended X-ray absorption fine structure; GH, Glycoside hydrolases; LPMOs, Lytic polysaccharide monooxygenases; MD, Molecular dynamics; NMR, Nuclear magnetic resonance; PL, Polysaccharide lyase; QM/MM, Quantum mechanical/molecular mechanics; SAXS, Small-angle X-ray scattering; TrCel7A, *Trichoderma reesei* Cel7A

References

1. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42(Database issue):D490–D495. https://doi.org/10.1093/nar/gkt1178

2. Adav SS, Sze SK (2014) Trichoderma secretome: an overview. Biotechnol Biol Trichoderma:103–114. https://doi.org/10.1016/B978-0-444-59576-8.00008-4

3. Meenu K, Singh G, Vishwakarma RA (2014) Molecular mechanism of cellulase production systems in Trichoderma. Biotechnol Biol Trichoderma:319–324. https://doi.org/10.1016/B978-0-444-59576-8.00022-9

4. Collier G, Ortiz V (2013) Emerging computational approaches for the study of protein allostery. Arch Biochem Biophys US 538:6–15. https://doi.org/10.1016/j.abb.2013.07.025

5. Manley G, Loria JP (2012) NMR insights into protein allostery. Arch Biochem Biophys US 519:223–231. https://doi.org/10.1016/j.abb.2011.10.023

6. Manley G, Rivalta I, Loria JP (2013) Solution NMR and computational methods for understanding protein allostery. J Phys Chem B US 117:3063–3073. https://doi.org/10.1021/jp312576v

7. Strawn R, Stockner T, Melichercik M, Jin L, Xue W-F, Carey J et al (2011) Synergy of molecular dynamics and isothermal titration calorimetry in studies of allostery. Methods Enzymol US 492:151–188. https://doi.org/10.1016/B978-0-12-381268-1.00017-3

8. Chakravorty DK, Merz KMJ (2014) Studying allosteric regulation in metal sensor proteins using computational methods. Adv Protein Chem Struct Biol Netherlands 96:181–218. https://doi.org/10.1016/bs.apcsb.2014.06.009

9. Grutsch S, Bruschweiler S, Tollinger M (2016) NMR methods to study dynamic allostery. PLoS Comput Biol US 12:e1004620. https://doi.org/10.1371/journal.pcbi.1004620

10. Boulton S, Melacini G (2016) Advances in NMR methods to map allosteric sites: from models to translation. Chem Rev US 116:6267–6304. https://doi.org/10.1021/acs.chemrev.5b00718

11. Hughes ML, Dougan L (2016) The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. Reports Prog Phys [Internet]. IOP Publishing 79:76601. https://doi.org/10.1088/0034-4885/79/7/076601

12. Ehrenberg W, Franks A (1952) Small-angle X-ray scattering. Nature [Internet]. Nature Publishing Group 170:1076. https://doi.org/10.1038/1701076a0

13. Li G, Van Wynsberghe A, Demerdash ONA, Cui Q (2006) Normal mode analysis of macromolecules: from enzyme active sites to molecular machines. Norm Mode Anal Theory Appl Biol Chem Syst:65–90. https://www.crcpress.com/Normal-Mode-Analysis-Theory-and-Applications-to-Biological-and-Chemical/Cui-Bahar/p/book/9781584884729. Accessed 03 Nov 2017

14. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 111:7812–7824. https://doi.org/10.1021/jp071097f

15. Periole X, Cavalli M, Marrink SJ, Ceruso MA (2009) Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. J Chem Theory Comput 5:2531–2543. https://doi.org/10.1021/ct9002114

16. Hospital A, Goñi JR, Orozco M, Gelpi J (2015) Molecular dynamics simulations: advances and applications. Adv Appl Bioinforma Chem [Internet] 8:37–47. https://doi.org/10.2147/AABC.S70333

17. Chen W, Morrow BH, Shi C, Shen JK (2014) Recent development and application of constant pH molecular dynamics. Mol Simul 40:830–838. https://doi.org/10.1080/08927022.2014.907492

18. Callen HB (1985) Thermodynamics and an introduction to thermostatistics [Internet]. Wiley. https://cds.cern.ch/record/450289/files/0471862568_TOC.pdf. Accessed 03 Nov 2017

19. Barducci A, Bonomi M, Parrinello M (2011) Metadynamics. Wiley Interdiscip Rev Comput Mol Sci 1:826–843. https://doi.org/10.1002/wcms.31

20. Lenner N, Mathias G (2016) Continuous tempering molecular dynamics: a deterministic approach to simulated tempering. J Chem Theory Comput 12:486–498. https://doi.org/10.1021/acs.jctc.5b00751

21. Vesely FJ (1994) Quantum mechanical simulation. Comput Phys An Introd [Internet]. Springer US, Boston, pp 207–228. https://doi.org/10.1007/978-1-4757-2307-6_7

22. Van Der Kamp MW, Mulholland AJ (2013) Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. Biochemistry 52:2708–2728. https://doi.org/10.1021/bi400215w

23. Lotan I, Schwarzer F, Latombe J-C (2003) Efficient energy computation for Monte Carlo simulation of proteins. In: Benson G, RDM P (eds) Algorithms Bioinforma Third Int Work WABI 2003, Budapest, Hungary, Sept 15-20, 2003 Proc [Internet]. Springer Berlin Heidelberg, Berlin, pp 354–373. https://doi.org/10.1007/978-3-540-39763-2_26

24. Del Moral P (2013) Mean field simulation for Monte Carlo integration [Internet]. http://people.bordeaux.inria.fr/pierre.delmoral/NICE-INLN-2012-Del.Moral-Part-I.pdf. Accessed 03 Nov 2017

25. Tsallis C, Stariolo DA (1995) Generalized simulated annealing. Comput Optim Eng 233:395–406. https://doi.org/10.1016/S0378-4371(96)00271-3

26. Koshland DE Jr (1953) Stereochemistry and the mechanism of enzymatic reactions. Biol Rev [Internet]. Blackwell Publishing Ltd 28:416–436. https://doi.org/10.1111/j.1469-185X.1953.tb01386.x

27. Granum DM, Schutt TC, Maupin CM (2014) Computational evaluation of the dynamic fluctuations of peripheral loops enclosing the catalytic tunnel of a family 7 cellobiohydrolase. J Phys Chem B US 118:5340–5349. https://doi.org/10.1021/jp5011555

28. Bu L, Crowley MF, Himmel ME, Beckham GT (2013) Computational investigation of the pH dependence of loop flexibility and catalytic function in glycoside hydrolases. J Biol Chem US 288:12175–12186. https://doi.org/10.1074/jbc.M113.462465

29. Vermaas JV, Petridis L, Qi X, Schulz R, Lindner B, Smith JC (2015) Mechanism of lignin inhibition of enzymatic biomass deconstruction. Biotechnol Biofuels Engl 8:217

30. Ogunmolu FE, Jagadeesha NBK, Kumar R, Kumar P, Gupta D, Yazdani SS (2017) Comparative insights into the saccharification potentials of a relatively unexplored but robust Penicillium funiculosum glycoside hydrolase 7 cellobiohydrolase. Biotechnol Biofuels Engl 10:71. https://doi.org/10.1186/s13068-015-0379-8

31. Borisova AS, Eneyskaya EV, Bobrov KS, Jana S, Logachev A, Polev DE, Lapidus AL, Ibatullin FM, Saleem U, Sandgren M, Payne CM, Kulminskaya AA, Ståhlberg J (2015) Sequencing, biochemical characterization, crystal structure and molecular dynamics of cellobiohydrolase Cel7A from Geotrichum candidum 3C. FEBS J Engl 282:4515–4537. https://doi.org/10.1111/febs.13509

32. Li T, Yan S, Yao L (2012) The impact of Trichoderma reesei Cel7A carbohydrate binding domain mutations on its binding to a cellulose surface: a molecular dynamics free energy study. J Mol Model Germany 18:1355–1364. https://doi.org/10.1007/s00894-011-1167-4

33. Lin Y, Beckham GT, Himmel ME, Crowley MF, Chu J-W (2013) Endoglucanase peripheral loops facilitate complexation of glucan chains on cellulose via adaptive coupling to the emergent substrate structures. J Phys Chem B US 117:10750–10758. https://doi.org/10.1021/jp405897q

34. Knott BC, Crowley MF, Himmel ME, Stahlberg J, Beckham GT (2014) Carbohydrate-protein interactions that drive processive polysaccharide translocation in enzymes revealed from a computational study of cellobiohydrolase processivity. J Am Chem Soc US 136:8810–8819. https://doi.org/10.1021/ja504074g

35. Ghattyvenkatakrishna PK, Alekozai EM, Beckham GT, Schulz R, Crowley MF, Uberbacher EC et al (2013) Initial recognition of a cellodextrin chain in the cellulose-binding tunnel may affect cellobiohydrolase directional specificity. Biophys J US 104:904–912. https://doi.org/10.1016/j.bpj.2012.12.052

36. Momeni MH, Payne CM, Hansson H, Mikkelsen NE, Svedberg J, Engstrom A et al (2013) Structural, biochemical, and computational characterization of the glycoside hydrolase family 7 cellobiohydrolase of the tree-killing fungus Heterobasidion irregulare. J Biol Chem US 288:5861–5872. https://doi.org/10.1074/jbc.M112.440891

37. Payne CM, Resch MG, Chen L, Crowley MF, Himmel ME, Taylor LE 2nd et al (2013) Glycosylated linkers in multimodular lignocellulose-degrading enzymes dynamically bind to cellulose. Proc Natl Acad Sci U S A 110:14646–14651

38. Kern M, McGeehan JE, Streeter SD, Martin RNA, Besser K, Elias L et al (2013) Structural characterization of a unique marine animal family 7 cellobiohydrolase suggests a mechanism of cellulase salt tolerance. Proc Natl Acad Sci U S A 110:10189–10194. https://doi.org/10.1073/pnas.1301502110

39. Taylor CB, Payne CM, Himmel ME, Crowley MF, McCabe C, Beckham GT (2013) Binding site dynamics and aromatic-carbohydrate interactions in processive and non-processive family 7 glycoside hydrolases. J Phys Chem B US 117:4924–4933. https://doi.org/10.1021/jp401410h

40. Textor LC, Colussi F, Silveira RL, Serpa V, de Mello BL, Muniz JRC, Squina FM, Pereira N Jr, Skaf MS, Polikarpov I (2013) Joint X-ray crystallographic and molecular dynamics study of cellobiohydrolase I from Trichoderma harzianum: deciphering the structural features of cellobiohydrolase catalytic activity. FEBS J Engl 280:56–69. https://doi.org/10.1111/febs.12049

41. Vermaas JV, Crowley MF, Beckham GT, Payne CM (2015) Effects of lytic polysaccharide monooxygenase oxidation on cellulose structure and binding of oxidized cellulose oligomers to cellulases. J Phys Chem B US 119:6129–6143. https://doi.org/10.1021/acs.jpcb.5b00778

42. Wu M, Beckham GT, Larsson AM, Ishida T, Kim S, Payne CM, Himmel ME, Crowley MF, Horn SJ, Westereng B, Igarashi K, Samejima M, Ståhlberg J, Eijsink VGH, Sandgren M (2013) Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the Basidiomycota fungus Phanerochaete chrysosporium. J Biol Chem US 288:12828–12839. https://doi.org/10.1074/jbc.M113.459396

43. Sprenger KG, Choudhury A, Kaar JL, Pfaendtner J (2016) Lytic polysaccharide monooxygenases ScLPMO10B and ScLPMO10C are stable in ionic liquids as determined by molecular simulations. J Phys Chem B US 120:3863–3872. https://doi.org/10.1021/acs.jpcb.6b01688

44. Moses V, Tastan Bishop O, Lobb K (2017) The evaluation and validation of copper (II) force field parameters of the auxiliary activity family 9 enzymes. Chem Phys Lett 678:91–97. https://doi.org/10.1016/j.cplett.2017.04.022

45. Schoeler C, Malinowska KH, Bernardi RC, Milles LF, Jobst MA, Durner E, Ott W, Fried DB, Bayer EA, Schulten K, Gaub HE, Nash MA (2014) Ultrastable cellulosome-adhesion complex tightens under load. Nat Commun Engl 5:5635. https://doi.org/10.1038/ncomms6635

46. Bernardi RC, Melo MCR, Schulten K (2015) Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim Biophys Acta Netherlands 1850:872–877. https://doi.org/10.1016/j.bbagen.2014.10.019

47. Xu J, Crowley MF, Smith JC (2009) Building a foundation for structure-based cellulosome design for cellulosic ethanol: insight into cohesin-dockerin complexation from computer simulation. Protein Sci US 18:949–959. https://doi.org/10.1002/pro.105

48. Xu J, Smith JC (2010) Probing the mechanism of cellulosome attachment to the Clostridium thermocellum cell surface: computer simulation of the type II cohesin-dockerin complex and its variants. Protein Eng Des Sel Engl 23:759–768. https://doi.org/10.1093/protein/gzq049

49. Zhang H, Zhang J, Sun L, Niu X, Wang S, Shan Y (2014) Molecular dynamics simulation of the processive endocellulase Cel48F from Clostridium cellulolyticum: a novel "water-control mechanism" in enzymatic hydrolysis of cellulose. J Mol Recognit Engl 27:438–447. https://doi.org/10.1002/jmr.2364

50. Vital de Oliveira O (2014) Molecular dynamics and Metadynamics simulations of the Cellulase Cel48F. Enzyme Res US 2014:692738–692737. https://doi.org/10.1155/2014/692738

51. Qian M, Guan S, Shan Y, Zhang H, Wang S (2016) Structural and molecular basis of cellulase Cel48F by computational modeling: insight into catalytic and product release mechanism. J Struct Biol US 194:347–356. https://doi.org/10.1016/j.jsb.2016.03.012

52. Jaeger V, Burney P, Pfaendtner J (2015) Comparison of three ionic liquid-tolerant cellulases by molecular dynamics. Biophys J US 108:880–892. https://doi.org/10.1016/j.bpj.2014.12.043

53. Johnson LB, Snow CD (2017) Molecular dynamics simulations of cellulase homologs in aqueous 1-ethyl-3-methylimidazolium

chloride. J Biomol Struct Dyn Engl 35:1990–2002. https://doi.org/10.1080/07391102.2016.1204364

54. Bianchetti CM, Brumm P, Smith RW, Dyer K, Hura GL, Rutkoski TJ, Phillips GN Jr (2013) Structure, dynamics, and specificity of endoglucanase D from Clostridium cellulovorans. J Mol Biol Eng 425:4267–4285. https://doi.org/10.1016/j.jmb.2013.05.030

55. Prates ET, Stankovic I, Silveira RL, Liberato MV, Henrique-Silva F, Pereira NJ, Polikarpov I, Skaf MS (2013) X-ray structure and molecular dynamics simulations of endoglucanase 3 from Trichoderma harzianum: structural organization and substrate recognition by endoglucanases that lack cellulose binding module. PLoS One US 8:e59069. https://doi.org/10.1371/journal.pone.0059069

56. Orłowski A, Róg T, Paavilainen S, Manna M, Heiskanen I, Backfolk K, Timonen J, Vattulainen I (2015) How endoglucanase enzymes act on cellulose nanofibrils: role of amorphous regions revealed by atomistic simulations. Cellulose [Internet] 22:2911–2925. https://doi.org/10.1007/s10570-015-0705-0

57. Zhang S, Wang Y, Song X, Hong J, Zhang Y, Yao L (2014) Improving Trichoderma reesei Cel7B thermostability by targeting the weak spots. J Chem Inf Model US 54:2826–2833. https://doi.org/10.1021/ci500339v

58. Anbar M, Gul O, Lamed R, Sezerman UO, Bayer EA (2012) Improved thermostability of Clostridium thermocellum endoglucanase Cel8A by using consensus-guided mutagenesis. Appl Environ Microbiol US 78:3458–3464. https://doi.org/10.1128/AEM.07985-11

59. Shu Z, Wang Y, An L, Yao L (2014) The slowdown of the endoglucanase Trichoderma reesei Cel5A-catalyzed cellulose hydrolysis is related to its initial activity. Biochemistry US 53:7650–7658. https://doi.org/10.1021/bi501059n

60. Yang H, Shi P, Liu Y, Xia W, Wang X, Cao H et al (2017) Loop 3 of fungal endoglucanases of glycoside hydrolase family 12 modulates catalytic efficiency. Appl Environ Microbiol US 83. https://doi.org/10.1128/AEM.03123-16

61. Zhang H, Zhang G, Yao C, Junaid M, Lu Z, Zhang H, Ma Y (2015) Structural insight of a trimodular halophilic cellulase with a family 46 carbohydrate-binding module. PLoS One US 10:e0142107. https://doi.org/10.1371/journal.pone.0142107

62. Payne CM, Baban J, Horn SJ, Backe PH, Arvai AS, Dalhus B, Bjørås M, Eijsink VGH, Sørlie M, Beckham GT, Vaaje-Kolstad G (2012) Hallmarks of processivity in glycoside hydrolases from crystallographic and computational studies of the Serratia marcescens chitinases. J Biol Chem US 287:36322–36330. https://doi.org/10.1074/jbc.M112.402149

63. Batista PR, Costa MG, Pascutti PG, Bisch PM, de Souza W (2011) High temperatures enhance cooperative motions between CBM and catalytic domains of a thermostable cellulase: mechanism insights from essential dynamics. Phys Chem Chem Phys Engl 13:13709–13720. https://doi.org/10.1039/c0cp02697b

64. Liu H, Pereira JH, Adams PD, Sapra R, Simmons BA, Sale KL (2010) Molecular simulations provide new insights into the role of the accessory immunoglobulin-like domain of Cel9A. FEBS Lett Engl 584:3431–3435. https://doi.org/10.1016/j.febslet.2010.06.041

65. Wu M, Bu L, Vuong TV, Wilson DB, Crowley MF, Sandgren M, Ståhlberg J, Beckham GT, Hansson H (2013) Loop motions important to product expulsion in the Thermobifida fusca glycoside hydrolase family 6 cellobiohydrolase from structural and computational studies. J Biol Chem US 288:33107–33117. https://doi.org/10.1074/jbc.M113.502765

66. Poma AB, Chwastyk M, Cieplak M (2015) Polysaccharide-protein complexes in a coarse-grained model. J Phys Chem B US 119:12028–12041. https://doi.org/10.1021/acs.jpcb.5b06141

67. Momeni MH, Goedegebuur F, Hansson H, Karkehabadi S, Askarieh G, Mitchinson C, Larenas EA, Ståhlberg J, Sandgren M (2014) Expression, crystal structure and cellulase activity of the

thermostable cellobiohydrolase Cel7A from the fungus Humicola grisea var. thermoidea. Acta Crystallogr D Biol Crystallogr US 70:2356–2366. https://doi.org/10.1107/S1399004714013844

68. Strobel KL, Pfeiffer KA, Blanch HW, Clark DS (2015) Structural insights into the affinity of Cel7A carbohydrate-binding module for lignin. J Biol Chem US 290:22818–22826. https://doi.org/10.1074/jbc.M115.673467

69. Beckham GT, Bomble YJ, Matthews JF, Taylor CB, Resch MG, Yarbrough JM, Decker SR, Bu L, Zhao X, McCabe C, Wohlert J, Bergenstråhle M, Brady JW, Adney WS, Himmel ME, Crowley MF (2010) The O-glycosylated linker from the Trichoderma reesei Family 7 cellulase is a flexible, disordered protein. Biophys J US 99:3773–3781. https://doi.org/10.1016/j.bpj.2010.10.032

70. Silveira RL, Skaf MS (2015) Molecular dynamics simulations of family 7 cellobiohydrolase mutants aimed at reducing product inhibition. J Phys Chem B US 119:9295–9303. https://doi.org/10.1021/jp509911m

71. Wickramasinghe GHIM, Rathnayake PPAMSI, Chandrasekharan NV, Weerasinghe MSS, Wijesundera RLC, Wijesundera WSS (2017) Trichoderma virens beta-glucosidase I (BGLI) gene; expression in Saccharomyces cerevisiae including docking and molecular dynamics studies. BMC Microbiol Engl 17:137. https://doi.org/10.1186/s12866-017-1049-8

72. Gaete-Eastman C, Morales-Quintana L, Herrera R, Moya-Leon MA (2015) In-silico analysis of the structure and binding site features of an alpha-expansin protein from mountain papaya fruit (VpEXPA2), through molecular modeling, docking, and dynamics simulation studies. J Mol Model Germany 21:115. https://doi.org/10.1007/s00894-015-2656-7

73. Hoffren AM, Teeri TT, Teleman O (1995) Molecular dynamics simulation of fungal cellulose-binding domains: differences in molecular rigidity but a preserved cellulose binding surface. Protein Eng Engl 8:443–450. https://doi.org/10.1093/protein/8.5.443

74. Noorbatcha IA, Waesoho S, Salleh HM (2012) Structural and dynamics behavior of native endoglucanase from fusarium oxysporum. Aust J Basic Appl Sci [Internet] 6:89–92

75. Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sorlie M et al (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. Science US 330:219–222. https://doi.org/10.1126/science.1192231

76. Correa TLR, dos Santos LV, Pereira GAG (2016) AA9 and AA10: from enigmatic to essential enzymes. Appl Microbiol Biotechnol Germany 100:9–16. https://doi.org/10.1007/s00253-015-7040-0

77. Johansen KS (2016) Lytic polysaccharide monooxygenases: the microbial power tool for lignocellulose degradation. Trends Plant Sci Engl 21:926–936. https://doi.org/10.1016/j.tplants.2016.07.012

78. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels Engl 6:41. https://doi.org/10.1186/1754-6834-6-41

79. Kim S, Stahlberg J, Sandgren M, Paton RS, Beckham GT (2014) Quantum mechanical calculations suggest that lytic polysaccharide monooxygenases use a copper-oxyl, oxygen-rebound mechanism. Proc Natl Acad Sci U S A 111:149–154. https://doi.org/10.1073/pnas.1316609111

80. Kjaergaard CH, Qayyum MF, Wong SD, Xu F, Hemsworth GR, Walton DJ, Young NA, Davies GJ, Walton PH, Johansen KS, Hodgson KO, Hedman B, Solomon EI (2014) Spectroscopic and computational insight into the activation of O2 by the mononuclear Cu center in polysaccharide monooxygenases. Proc Natl Acad Sci U S A 111:8797–8802. https://doi.org/10.1073/pnas.1408115111

81. Hedegård ED, Ryde U (2017) Multiscale modelling of lytic polysaccharide monooxygenases. ACS Omega 2:536–545

82. Hedegard ED, Ryde U (2017) Targeting the reactive intermediate in polysaccharide monooxygenases. J Biol Inorg Chem Germany 22:1029–1037. https://doi.org/10.1007/s00775-017-1480-1

83. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen J-CN, Johansen KS, Krogh KBRM, Jorgensen CI, Tovborg M, Anthonsen A, Tryfona T, Walter CP, Dupree P, Xu F, Davies GJ, Walton PH (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. Proc Natl Acad Sci U S A 108:15079–15084. https://doi.org/10.1073/pnas.1105776108

84. Aachmann FL, Sorlie M, Skjak-Braek G, Eijsink VGH, Vaaje-Kolstad G (2012) NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. Proc Natl Acad Sci U S A 109: 18779–18784. https://doi.org/10.1073/pnas.1208822109

85. Bomble YJ, Beckham GT, Matthews JF, Nimlos MR, Himmel ME, Crowley MF (2011) Modeling the self-assembly of the cellulosome enzyme complex. J Biol Chem US 286:5614–5623. https://doi.org/10.1074/jbc.M110.186031

86. Rozycki B, Cazade P-A, O'Mahony S, Thompson D, Cieplak M (2017) The length but not the sequence of peptide linker modules exerts the primary influence on the conformations of protein domains in cellulosome multi-enzyme complexes. Phys Chem Chem Phys England 19:21414–21425. https://doi.org/10.1039/C7CP04114D

87. Bernardi RC, Cann I, Schulten K (2014) Molecular dynamics study of enhanced Man5B enzymatic activity. Biotechnol Biofuels Engl 7:83. https://doi.org/10.1186/1754-6834-7-83

88. Poma AB, Cieplak M, Theodorakis PE (2017) Combining the MARTINI and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins. J Chem Theory Comput US 13:1366–1374. https://doi.org/10.1021/acs.jctc.6b00986

89. Chen M, Bomble YJ, Himmel ME, Brady JW (2012) Molecular dynamics simulations of the interaction of glucose with imidazole in aqueous solution. Carbohydr Res Netherlands 349:73–77. https://doi.org/10.1016/j.carres.2011.12.008

90. Chen M, Kostylev M, Bomble YJ, Crowley MF, Himmel ME, Wilson DB, Brady JW (2014) Experimental and modeling studies of an unusual water-filled pore structure with possible mechanistic implications in family 48 cellulases. J Phys Chem B US 118:2306–2315. https://doi.org/10.1021/jp408767j

91. Hammel M, Fierobe H-P, Czjzek M, Finet S, Receveur-Brechot V (2004) Structural insights into the mechanism of formation of cellulosomes probed by small angle X-ray scattering. J Biol Chem US 279:55985–55994. https://doi.org/10.1074/jbc.M408979200