Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Biotechnology

Applications of computational science for understanding enzymatic deconstruction of cellulose

Gregg T Beckham[1,2,3], Yannick J Bomble[4,5], Edward A Bayer[6], Michael E Himmel[4,5] and Michael F Crowley[4,5]

Understanding the molecular-level mechanisms that enzymes employ to deconstruct plant cell walls is a fundamental scientific challenge with significant ramifications for renewable fuel production from biomass. In nature, bacteria and fungi use enzyme cocktails that include processive and non-processive cellulases and hemicellulases to convert cellulose and hemicellulose to soluble sugars. Catalyzed by an accelerated biofuels R&D portfolio, there is now a wealth of new structural and experimental insights related to cellulases and the structure of plant cell walls. From this background, computational approaches commonly used in other fields are now poised to offer insights complementary to experiments designed to probe mechanisms of plant cell wall deconstruction. Here we outline the current status of computational approaches for a collection of critical problems in cellulose deconstruction. We discuss path sampling methods to measure rates of elementary steps of enzyme action, coarse-grained modeling for understanding macromolecular, cellulosomal complexes, methods to screen for enzyme improvements, and studies of cellulose at the molecular level. Overall, simulation is a complementary tool to understand carbohydrate-active enzymes and plant cell walls, which will enable industrial processes for the production of advanced, renewable fuels.

Addresses
[1] National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO, United States
[2] Department of Chemical Engineering, Colorado School of Mines, Golden, CO, United States
[3] Renewable and Sustainable Energy Institute, Boulder, CO, United States
[4] Biosciences Center, National Renewable Energy Laboratory, Golden, CO, United States
[5] BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, United States
[6] Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel

Corresponding authors: Himmel, Michael E (Mike.Himmel@nrel.gov) and Crowley, Michael F (Michael.Crowley@nrel.gov)

Current Opinion in Biotechnology 2011, **22**:231–238

This review comes from a themed issue on
Plant biotechnology
Edited by Adi Avni and Miguel Blazquez

Available online 17th December 2010

0958-1669/$ – see front matter

DOI 10.1016/j.copbio.2010.11.005

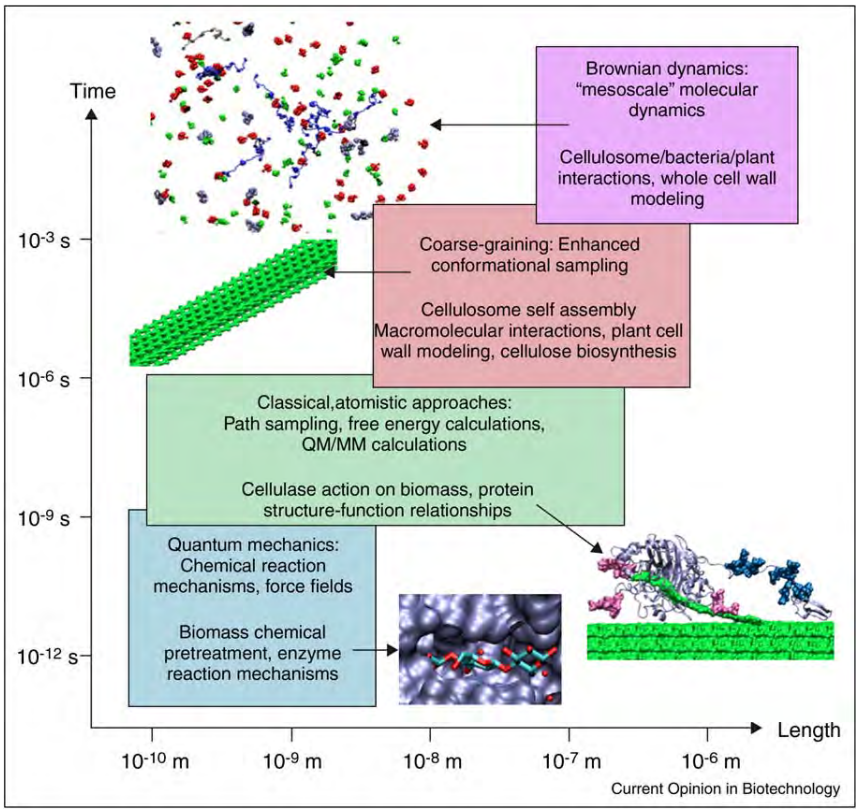
Introduction

Enzymatic deconstruction of plant cell walls to fermentable sugars is a primary, near-term option currently being pursued for the production of advanced biofuels. Driven by significant, international R&D efforts for biofuels, there exists a new wealth of experimental data about the chemistry and structure of plant cell walls and the mechanisms that cellulase and hemicellulase enzymes use to deconstruct cell wall polysaccharides [1]. It is noteworthy, however, that we still lack sufficient data to explain cell wall deconstruction, even though elucidation of these steps is crucial to develop enhanced conversion processes. Here, we discuss the challenges, opportunities, and early successes of theory and simulation to aid our understanding of mechanisms of plant cell wall deconstruction and in catalyst design. We also present, where appropriate, examples computational methods that can be applied to a given problem into the complex problem of cell wall deconstruction with the advent of improved simulation codes and computational power (Figure 1). We limit the scope to research of cellulose and cellulases from the last three years, but many of the discussions are extendable to hemicellulases, chitinases, and other carbohydrate-active enzymes [2•,3•]. We hope that this opinion illustrates to computational researchers from other fields that new, exciting opportunities exist in biomass conversion, and to biomass researchers that their work can be greatly enhanced by using computational science.

Plant cell wall polymer models

The study of cellulose requires: (i) reliable structural models and (ii) accurate potentials to describe cellulose. Structural models of plant cellulose have been proposed with 36 cellulose chains per elementary fibril (microfibril), although this has not been directly verified experimentally [4]. Further experimental characterization of plant microfibrils will aid in the construction of more accurate representations of microfibrils for simulations. Two popular atomistic potentials for carbohydrates are GLYCAM [5], which is consistent with the Amber force field [6] and C35 [7,8], which is consistent with the CHARMM force field [9]. C35 will likely be widely utilized because it can be applied with commonly used simulation molecular dynamics (MD) packages such as CHARMM [9], NAMD [10], and now, via the CHAMBER program [11] in Amber [6], and it is consistent with the large class of biological molecules (e.g., proteins, lipids, and nucleic acids) in the CHARMM force field [9]. Both force fields can also describe xylose and other carbohydrates, making them

Figure 1



Methods in computational science span multiple spatial and temporal scales from atomic resolution to the application of coarse-graining at multiple resolutions. Listed here are typical techniques applied at various scales with general questions of interest in enzymatic deconstruction of plant cell wall polysaccharides for which computational science can play a role.

appropriate to model other polysaccharides of the plant cell wall.

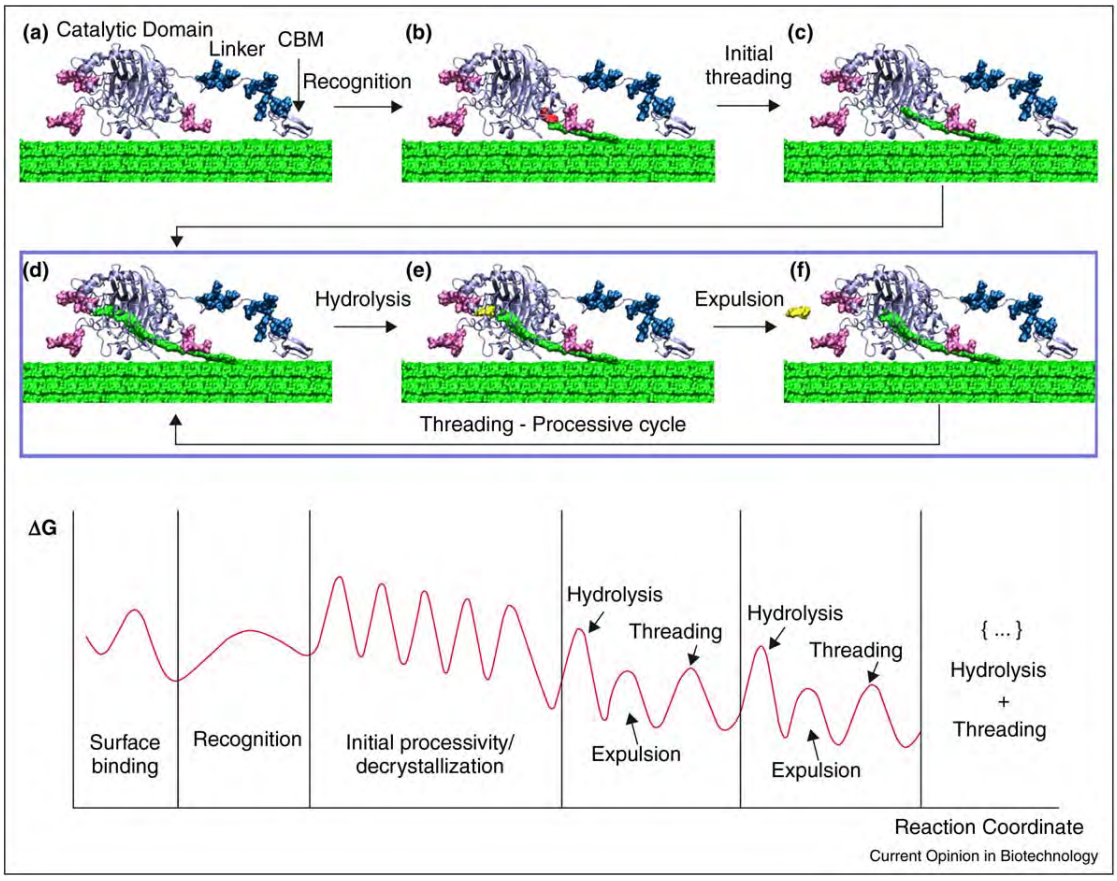
There is debate regarding the appropriate manner in which to simulate cellulose at atomistic resolution with MD, namely with infinite crystals or with finite crystals [12]. For finite crystals, the cellulose microfibrils will twist in MD simulations along the chain axis [4], whereas infinite crystals will prohibit twisting or large conformational changes because of the bonds across the periodic boundaries of a simulation cell [13]. Also, for finite crystals, it is likely that no cellulose simulations to date have been conducted long enough to equilibrate the structures adequately. Simulations up to the microsecond time scale are now readily accessible for crystals of the same size as in [4], and will likely be required to obtain equilibrated systems and to investigate the nature of microfibril twist at long times.

Petridis and Smith adapted the CHARMM potential to model lignin [14]. However, as the authors discuss, the structure and arrangement of lignin monomers are unknown experimentally. Thus, realistic MD simulations of lignin are not likely feasible until additional experimental characterization is conducted to ascertain relevant lignin chemistry and connectivity more directly. This will remain problematic if lignin is randomly assembled by free radical driven processes into complex, 3-dimensional structures. Additionally, modeling the 'whole' plant cell wall with MD necessitates more accurate descriptions of the plant cell wall at the molecular level, in terms of both the molecular connectivity and the 3-dimensional arrangement of the cell wall polymers.

Free cellulases

Free cellulases usually consist of one or more carbohydrate-binding modules (CBMs), one or more linkers, and catalytic domain (CD), as shown in Figure 2. The CDs of processive cellulases have tunnels for threading cellulose chains, whereas CDs of non-processive cellulases instead have clefts for binding to accessible chains [15]. Here we discuss the *Trichoderma reesei* Family 7 processive cellobiohydrolase (Cel7A), as it has been thoroughly characterized experimentally. Also, cellulases often exhibit *N*-glycosylation and *O*-glycosylation imparted by the expression host, which can affect activity [16•]. Figure 2 shows the steps

Figure 2



The *Trichoderma reesei* Family 7 cellobiohydrolase (Cel7A) acting on cellulose. Cel7A is comprised of a 36-amino acid CBM, a linker domain with *O*-glycan (dark blue), and a large catalytic domain with *N*-linked glycan (pink) and a 50-Å tunnel for processing cellulose chains (green). The cellobiose product is shown in yellow **(e)** and **(f)**. Here we show the putative steps that Cel7A takes to deconstruct biomass and the hypothesized free energy surface for each elementary step.

that Cel7A undergoes to deconstruct biomass with a hypothesized free energy landscape: (a) binding via the CBM to the hydrophobic face of cellulose [17], (b) surface diffusion to find a free, reducing chain end, (c) threading of the chain into the tunnel, (d) formation of the active complex, (e) hydrolysis of the glycosidic linkage to produce cellobiose, and (f) product expulsion, and threading of another cellobiose unit to re-form the active complex. As such, Cel7A processes a cellulose chain until it is fully hydrolyzed or until the enzyme is deactivated [18••]. The thermodynamics and kinetics of each step in Figure 2 represent a challenge to probe experimentally and computationally. Additionally, the roles of the CBM and linker at the molecular level are not fully characterized [19,20•,21–23]. Because of computational expense, it is likely to be more efficient to first examine individual components of cellulases.

To that end, the CBM on the hydrophobic cellulose face [17] has been studied using atomistic and coarse-grained models [20•,21], and it was demonstrated that the CBM exhibits regions of thermodynamic stability every cellobiose unit (~1 nm) along a cellulose chain. From the atomistic simulation, four amino acids were shown to be responsible for this behavior, which all form strategic hydrogen bonds every cellobiose unit [20•]. It is noteworthy that the CBM alone diffuses on the same critical length scale as the catalytic product of Cel7A, and the four residues responsible for this behavior are highly conserved in Family 1 CBMs.

Elucidating the reaction coordinate (RC), or mechanism, of threading of a cellobiose unit to form the active complex is an ideal problem for rare events simulation methods [24,25]. These methods have been applied in other fields to understand phase transitions and biomolecular conformational transitions [26–28]. Aimless shooting with likelihood maximization or forward-flux sampling with least squares-estimation could yield the RC to determine the threading mechanism. On a

related note, Eijsink *et al.* demonstrated the importance of the aromatic residues that line the tunnels of chitinases (cellulases and chitinases are structurally and functionally similar) by mutating aromatic residues to alanine, and measuring the processivity rates on crystalline chitin and soluble chitosan [29••,30••]. Using rare event simulation [24,25] to measure the free energy barriers to threading combined with free energy perturbation techniques [31,32] to mutate aromatic residues and polar residues in cellulase (and chitinase) tunnels will yield molecular-level insights into these intriguing experimental observations. These types of simulations will also likely yield insights into recent atomic force microscopy (AFM) data from Igarashi *et al.* on Cel7A [33••]. The effect of polar residues has not been probed experimentally, and thus offers opportunity for insights and predictions directly from simulation. Alchemical free energy calculations will also guide experimental efforts in tuning the binding free energy of the ligand in the CD, which may have significant ramifications for general cellulase deactivation, exemplified by processive enzymes becoming immobilized at certain points along a cellulose chain [18••].

Another key thermodynamic barrier to cellulose deconstruction is the decrystallization of a cellulose chain from the polymer crystal (Figure 2). Because different cellulose polymorphs exhibit different conversion rates, it is probable that the decrystallization free energy will depend on the crystal packing. This is an ideal question for free energy simulations (e.g., umbrella sampling [34,35]) over an RC designed to decrystallize a single cellulose chain. The RC for this process should not overconstrain the flexibility of the cellulose chain being decrystallized (e.g., as pulling from an end of the chain would), as entropic effects are likely to play an important role in the decrystallization process. For cellulose decrystallization, the only existing calculation was reported by Bergenstråhle *et al.* [13] who used steered MD to pull a cello-octaose chain placed randomly on the hydrophilic faces of a cellulose crystal. Similar calculations with relevance to enzymatic deconstruction, however, should be conducted with an improved RC from the hydrophobic cellulose face and crystal corners as these are the likely points of surface attack by cellulases [17]. These calculations would also aid in the design of organic or novel solvents for dissolving cellulose, for example, ionic liquids [36].

The hydrolysis reaction conducted by processive cellulases is likely to be a major contributor to the rate-limiting step. Simulations of reactions in enzymes with a hybrid quantum mechanics-molecular mechanics (QM/MM) approach [37,38] will be necessary to understand hydrolysis mechanisms. Barnett *et al.* used density functional theory free energy calculations to elucidate the preferred conformation needed to stabilize the transition state conformation in *T. reesei* Cel7A [39]. The next steps will be to elucidate the mechanism of the hydrolysis reaction with free energy methods, such as umbrella sampling or transition state search algorithms.

The product expulsion step is also an important thermodynamic parameter for developing models of cellulase action due to product inhibition. For cellobiose expulsion from cellulase tunnels, absolute binding free energy calculations [40,41] or steered MD [42] can be conducted to measure the work required to remove cellobiose from the product site. In these thermodynamics methods, care should be taken to sample appropriately the conformations of the protein and carbohydrate, and the effect of the *N*-glycosylation (Figure 2) should also be considered.

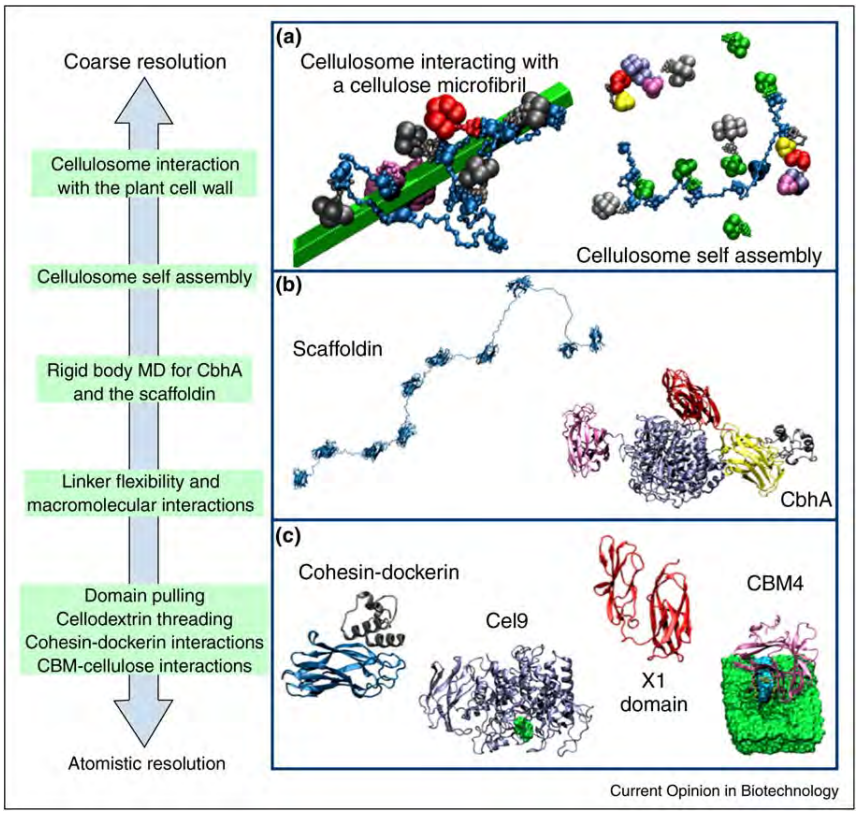
Relevant to understanding the role of the linker domain in cellulose hydrolysis, Ting *et al.* recently developed a master equation approach to calculate the steady-state hydrolysis rate of a processive cellulase like Cel7A [43]. The authors modeled the CBM and CD as random walkers coupled by a linker represented as a spring and found that the steady state hydrolysis rate was maximized at intermediate linker stiffness. The authors noted that a probability distribution of the Cel7A linker is not available. To that end, replica-exchange MD (REMD) [44] was recently used to construct the probability distribution for the Cel7A linker with the native glycosylation pattern, and the linker was shown to be an intrinsically disordered protein with a significant degree of flexibility [45,46]. Whereas some cellulase and/or cellulosomal linkers are inherently flexible [46,47], others are fixed integral components of the parent protein likely secure the conformations of the adjacent functional modules [48]. Because linker domains are prolific in both free and complexed cellulases, enhanced sampling methods for small proteins (i.e. on the order of 10 s of residues) like REMD offer a powerful approach to probe intrinsic flexibility and disorder.

Cellulosomal enzymes

Complexed enzymes are found in some bacteria and a few fungi where multiple carbohydrate-active enzymes are bound to protein scaffolds via cohesin–dockerin interactions to form a complex termed 'cellulosome' [49]. There are many open questions at multiple resolutions regarding the cellulosome structure and function for which simulations can offer valuable insight. Here, we review several open questions across multiple length and time scales, which are summarized in Figure 3.

At low resolution, self assembly of cellulosome components onto the scaffold is not well understood. To that end, a coarse-grained model was recently constructed consisting of three representative dockerin-containing

Figure 3



Multi-scale modeling can aid in the understanding of the cellulosomal complex and enzyme–cellulose interactions in the cellulosome. Here are several open questions at various degrees of resolution together with methods to probe each question. **(a)** A simple coarse-grained model has been developed to study self assembly of the entire cellulosome enzyme complex as a function of enzyme concentration and other relevant variables. **(b)** Rigid body MD enables calculation of solution behavior directly from simulation to compare with SAXS and FRET experiments of the large CbhA enzyme. **(c)** Multiple scientific questions exist at the atomistic scale that can be examined with methods such as rare event simulation to understand mechanisms of threading cellodextrin chains into cellulase tunnels, free energy perturbation methods for relative binding free energies and absolute binding free energies of carbohydrates to cellulases and CBMs, docking calculations to understand the non-covalent binding at the atomic scale, steered MD to understand the work to extend putatively flexible proteins, and REMD to understand intrinsic disorder.

enzymes and a scaffoldin, containing cohesin modules connected by flexible linkers, to study the potential effect of enzyme size, flexibility, and shape on binding affinity (Figure 3(a)) [50•]. It was observed that the large (~140 kDa) Family 9-containing, multi-modular enzyme, CbhA, binds with greater affinity (in the limit of microscopic diffusivity) because of significantly greater flexibility.

At slightly higher resolution is the solution behavior of CbhA (Figure 3(b)) [51]. CbhA consists of seven protein domains including CBMs, a processive CD, a domain of unknown function called the X1 domain, and an immunoglobulin-like domain connected by linkers. Small-angle X-ray scattering (SAXS) and rigid body MD or REMD can be used in concert to elucidate the solution behavior of this enzyme to better understand the interactions that might exist between these modules [52•].

At atomistic resolution (Figure 3(c)), there are many interesting questions about enzyme–substrate interactions in cellulosomal action. For example, the X1 domain in CbhA has been hypothesized either to disrupt cellulose or to be an extended linker in the presence of cellulose [53]. Steered MD can be used to study the extension mechanism of the X1 protein module by forced unfolding [54]. Valbuena *et al.* applied this method to test the mechanical stability of cohesin domains from cellulosomal scaffoldins [55•]. Similar to single-molecule pulling experiments, these experiments will provide an idea of the work needed for individual modules to unfold. These simulations can be compared to experimental techniques such as SAXS, fluorescence resonance energy transfer (FRET), and AFM pulling.

Additionally, the cohesin–dockerin interaction is one of the strongest non-covalent interactions known in proteins

[56]. Schueler-Furman *et al.* used RosettaDock to predict the structure of the cohesin–dockerin complex with high accuracy. Their work combined with more detailed thermodynamic methods makes possible the design of new cohesin–dockerin interactions to make designer cellulosomes with known components at specific locations.

CBM–cellulose interactions in cellulosomal enzymes have also been probed in a recent combined experimental–computational study [57]. Alahuhta *et al.* showed that a tryptophan residue located on a loop at the edge of the binding site cleft of CbhA CBM4 strongly interacts with a cellodextrin chain bound in the cleft. This result is significant because this tryptophan residue is unique to this CBM module, which hints that it might have a different function than other Family 4 CBMs.

Computational screening of cellulases for improved activity

The primary method used for improving cellulase activity to date is increasing protein thermal stability, for which computation offers significant benefits [58]. Heinzelman *et al.* computationally recombined 8 structural blocks from three wild-type Family 6 cellulases (Cel6A) to produce a library of Cel6A cellulases with improved thermal stability and activity [58]. High-throughput, computational screening tools like that used by Heinzelman *et al.* or with Rosetta [59] will aid the design of more thermally tolerant enzymes.

Conclusion

Knowledge of the elementary steps in cellulase action is essential for building enhanced models of cellulose deconstruction, which will in turn guide development of enhanced cellulase systems. Driven by the wealth of new experimental data on cellulases and cellulose, computer simulations are beginning to play an increasingly significant role in understanding the structure–function relationships of enzymatic cellulose deconstruction. By using a versatile portfolio of computational methods, simulations can offer insights that are complementary to experiments for understanding cellulase–cellulose interactions and for designing enhanced enzymes for biomass conversion. When possible, simulation results should be verified experimentally. In some cases, however, experimental approaches, for example, using recombinant technologies for expression of active and soluble, large multi-modular cellulases or cellulosomal components, or synthetic approaches, may be severely limited or impractical. In such cases, therefore, computational science may be the preferred or exclusive option.

Acknowledgements

We thank the DOE Office of the Biomass Program (Cel7A modeling), the DOE BER BioEnergy Science Center (cellulosome modeling), and the DOE ASCR SciDAC program for funding (cellulose modeling). EAB holds the Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry at the Weizmann Institute of Science.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Himmel ME, Ding SY, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD: **Biomass recalcitrance: engineering plants and enzymes for biofuels production**. *Science* 2007, **315**:804-807.

2. • Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics**. *Nucleic Acids Res* 2009, **37**:D233-D238.
The CAZy database at www.cazy.org is the seminal database for finding information about carbohydrate-active enzyme sequences, structural information, and references. The newly founded Cazypedia (www.cazypedia.org) as well contains background information, historical context, and the state of scientific knowledge for families of carbohydrate active enzymes.

3. • Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC: **CAZymes Analysis Toolkit (CAT): web-service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database**. *Glycobiology* 2010.
The authors developed a suite of search tools tailored to the CAZy database and metagenome studies for carbohydrate-active enzymes.

4. Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J, Himmel ME, Brady JW: **Computer simulations of microcrystalline cellulose Iβ**. *Carbohyd Res* 2006, **341**:138-152.

5. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, Woods RJ: **GLYCAM06: a generalizable biomolecular force field. Carbohydrates**. *J Comput Chem* 2008, **29**:622-655.

6. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ: **The Amber biomolecular simulation programs**. *J Comput Chem* 2005, **26**:1668-1688.

7. Guvench O, Greene SN, Kamath G, Brady JW, Venable RM, Pastor RW, Mackerell AD: **Additive empirical force field for hexopyranose monosaccharides**. *J Comput Chem* 2008, **29**:2543-2564.

8. Guvench O, Hatcher E, Venable RM, Pastor RW, MacKerell AD: **CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses**. *J Chem Theory Comput* 2009, **5**:2353-2370.

9. Brooks BR, Brooks CL 3rd, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S *et al.*: **CHARMM: the biomolecular simulation program**. *J Comput Chem* 2009, **30**:1545-1614.

10. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K: **Scalable molecular dynamics with NAMD**. *J Comput Chem* 2005, **26**:1781-1802.

11. Crowley MF, Williamson MJ, Walker RC: **CHAMBER: comprehensive support for CHARMM force fields within the AMBER software**. *Int J Quantum Chem* 2009, **109**:3767-3772.

12. Bellesia G, Asztalos A, Shen T, Langan P, Redondo A, Gnanakaran S: **In silico studies of crystalline cellulose and its degradation by enzymes**. *Acta Cryst D* 2010, **66**:1184-1188.

13. Bergenstråhle M, Thormann E, Nordgren N, Berglund LA: **Force pulling of single cellulose chains at the crystalline cellulose–liquid interface: a molecular dynamics study**. *Langmuir* 2009, **25**:4635-4642.

14. Petridis L, Smith JC: **A molecular mechanics force field for lignin**. *J Comput Chem* 2008, **30**:457-467.

15. Divne C, Stahlberg J, Teeri TT, Jones TA: **High-resolution crystal structures reveal how a cellulose chain is bound in the 50 angstrom long tunnel of cellobiohydrolase I from *Trichoderma reesei***. *J Mol Biol* 1998, **275**:309-325.

16. Adney WS, Jeoh T, Beckham GT, Chou YC, Baker JO, Michener W,
• Brunecky R, Himmel ME: **Probing the role of N-linked glycans in the stability and activity of fungal cellobiohydrolases by mutational analysis**. *Cellulose* 2009, **16**:699-709.
The work in this manuscript demonstrated that glycosylation can affect specific enzyme activity for Family 7 cellulases expressed in two different fungal hosts. The major outcomes from this work are (1) a demonstration that glycosylation can be tuned through biochemical removal of glycan sites via site-directed mutagenesis and (2) that the molecular-level details of glycosylation on the structure and function of cellulases is not yet well understood, which is an ideal problem for computational science.

17. Lehtiö J, Sugiyama J, Gustavsson M, Fransson L, Linder M, Teeri TT: **The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules**. *Proc Natl Acad Sci USA* 2003, **100**:484-489.

18. Jalak J, Valjamae P: **Mechanism of initial rapid rate retardation
•• in cellobiohydrolase catalyzed cellulose hydrolysis**. *Biotechnol Bioeng* 2010, **106**:871-883.
The authors developed rate expressions for Cel7A acting on cellulose that for the first time explicitly account for and explain enzyme deactivation as the reason for the leveling off of the conversion rate at long times. This manuscript, along with earlier papers from the chitinase literature [29••,30••], suggest that engineering the ligand binding free energy to be lower in the catalytic tunnels may be a viable strategy for engineering cellulases for biomass conversion. This study also offers a fitted parameter for enzyme deactivation that is consistent with experimental observations of 'restarting' conversion by adding fresh cellulase at long digestion times. The authors postulate that enzyme deactivation mechanism is primarily driven by surface blockage at multiple length scales (e.g., lignin coated on cellulose will not allow an enzyme to deconstruct the cellulose underneath it, or microfibrils of cellulose can aggregate).

19. Srisodsuk M, Reinikainen T, Penttila M, Teeri TT: **Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose**. *J Biol Chem* 1993, **268**:20756-20761.

20. Beckham GT, Matthews JF, Bomble YJ, Bu L, Adney WS,
• Himmel ME, Nimlos MR, Crowley MF: **Identification of amino acids responsible for processivity in a Family 1 carbohydrate-binding module from a fungal cellulase**. *J Phys Chem B* 2010, **114**:1447-1453.
The Family 1 CBM–cellulose interaction was thought to be driven primarily by hydrophobic interactions because of the 3 aromatic residues on the putative binding surface of the CBM. However, this study demonstrated that the CBM behavior on the hydrophobic cellulose face is primarily dictated by the formation of 4 hydrogen bonds that reproduce the cellobiose repeat distance on the surface. ITC or other types of thermodynamic experiments over a range of temperatures should be conducted to determine the contributions to the binding mode.

21. Bu L, Beckham GT, Crowley MF, Chang CH, Matthews JF, Bomble YJ, Adney WS, Himmel ME, Nimlos MR: **The energy landscape for the interaction of the Family 1 carbohydrate-binding module and the cellulose surface is altered by hydrolyzed glycosidic linkages**. *J Phys Chem B* 2009, **113**:10994-11002.

22. Mulakala C, Reilly PJ: ***Hypocrea jecorina* (*Trichoderma reesei*) Cel7A as a molecular machine: a docking study**. *Proteins-Struct. Func. Bioinform.* 2005, **60**:598-605.

23. Nimlos MR, Matthews JF, Crowley MF, Walker RC, Chukkapalli G, Brady JW, Adney WS, Clearyl JM, Zhong LH, Himmel ME: **Molecular modeling suggests induced fit of Family I carbohydrate-binding modules with a broken-chain cellulose surface**. *Protein Eng Design Select* 2007, **20**:179-187.

24. Peters B, Beckham GT, Trout BL: **Extensions to the likelihood maximization approach for finding reaction coordinates**. *J Chem Phys* 2007, **127**:034109.

25. Escobedo FA, Borrero EE, Araque JC: **Transition path sampling and forward flux sampling. Applications to biological systems**. *J Phys: Condens Mat* 2009, **21**:333101.

26. Beckham GT, Peters B, Variankaval N, Starbuck C, Trout BL: **Surface-mediated nucleation in the solid-state polymorph transformation of terephthalic acid**. *J Amer Chem Soc* 2007, **129**:4714-4724.

27. Juraszek J, Bolhuis PG: **Rate constant and reaction coordinate of Trp-cage folding in explicit water**. *Biophys J* 2008, **95**:4246-4257.

28. Vreede J, Juraszek J, Bolhuis PG: **Predicting the reaction coordinates of millisecond light-induced conformational changes in photoactive yellow protein**. *Proc Natl Acad Sci USA* 2010, **107**:2397-2402.

29. Horn SJ, Sikorski P, Cederkvist JB, Vaaje-Kolstad G, Sorlie M,
•• Synstad B, Vriend G, Varum KM, Eijsink VGH: **Costs and benefits of processivity in enzymatic degradation of recalcitrant polysaccharides**. *Proc Natl Acad Sci USA* 2006, **103**:18089-18094.
This study along with Ref. [30••] demonstrates, for 2 chitinases, that mutation of aromatic residues in the CD tunnels leads to increased rates of processivity on soluble chitosan but a decreased rate of conversion on crystalline chitin. The authors postulate that processivity of crystalline substrates comes at a significant loss in enzyme speed. An open question in these studies is if the mutations of the aromatic residues in the tunnel cause the enzyme to act as an endochitinase, although there is some evidence that the soluble chitosan chains are able to thread through the tunnels without a reaction occurring. Measuring the ligand binding free energy as a function of the sequence space in the CD tunnels of processive and non-processive cellulases will likely shed more light on the molecular-level phenomena occurring when aromatic (or polar) residues are removed from CD tunnels.

30. Zakariassen H, Aam BB, Horn SJ, Varum KM, Sorlie M,
•• Eijsink VGH: **Aromatic residues in the catalytic center of chitinase A from *Serratia marcescens* affect processivity, enzyme activity, and biomass converting efficiency**. *J Biol Chem* 2009, **284**:10610-10617.
See annotation to Ref. [29••].

31. Chipot C, Pohorille A: **Calculating free energy differences using perturbation theory**. In *Free Energy Calculations.* Edited by Chipot C, Pohorille A. Springer; 2007. . pp. 33–72. Springer Series in Chemical Physics, vol 1..

32. Knight JL, Brooks CL: **Lambda-dynamics free energy simulation methods**. *J Comput Chem* 2009, **30**:1692-1700.

33. Igarashi K, Koivula A, Wada M, Kimura S, Penttila M, Samejima M:
•• **High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose**. *J Biol Chem* 2009, **284**:36186-36190.
The authors used a novel high-speed atomic force microscope to study the *T. reesei* Cel7A WT enzyme, the CD only, the catalytically dead mutant, and a non-threading mutant (W40A) on cellulose. They found that the velocity of the CD and WT Cel7A enzyme is ~3.5 nm/s, which may correspond to the hydrolysis rate if the cellulases observed are productively engaged. The authors found that the catalytically dead mutant and W40A mutant would bind and not translate on the surface of cellulose, which remain to be explained at the molecular level. Conducting thermodynamic integration calculations or path sampling simulations for the W40A mutant, with a productively bound Cel7A on a cellulose crystal or a cellodextrin chain in solution would be an interesting application of simulation to these experimental results.

34. Torrie GM, Valleau JP: **Non-physical sampling distributions in Monte-Carlo free-energy estimation — umbrella sampling**. *J Comput Phys* 1977, **23**:187-199.

35. Kottalam J, Case DA: **Dynamics of ligand escape from the heme pocket of myoglobin**. *J Amer Chem Soc* 1988, **110**:7690-7697.

36. Liu HB, Sale KL, Holmes BM, Simmons BA, Singh S: **Understanding the interactions of cellulose with ionic liquids: a molecular dynamics study**. *J Phys Chem B* 2010, **114**:4293-4301.

37. Kamerlin SCL, Haranczyk M, Warshel A: **Progress in *ab initio* QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pK(a), redox reactions and solvation free energies**. *J Phys Chem B* 2009, **113**:1253-1272.

38. Gao J, Ma S, Major DT, Nam K, Pu J, Truhlar DG: **Mechanisms and free energies of enzymatic reactions**. *Chem Rev* 2006, **106**:3188-3209.

39. Barnett CB, Wilkinson KA, Naidoo KJ: **Pyranose ring transition state is derived from cellobiohydrolase I induced conformational stability and glycosidic bond polarization**. *J Amer Chem Soc* 2010, **132**:12800-12803.

40. Deng Y, Roux B: **Computations of standard binding free energies with MD simulations**. *J Phys Chem B* 2009, **113**:2234-2246.

41. Mobley DL, Chodera JD, Dill KA: **Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change**. *J Chem Theory Comput* 2007, **3**:1231-1235.

42. Park S, Khalili-Araghi F, Tajkhorshid E, Schulten K: **Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality**. *J Chem Phys* 2003, **119**:3559-3566.

43. Ting CL, Makarov DE, Wang ZG: **A kinetic model for the enzymatic action of cellulase**. *J Phys Chem B* 2009, **113**:4970-4977.

44. Sugita Y, Okamoto Y: **Replica-exchange molecular dynamics method for protein folding**. *Chem Phys Lett* 1999, **314**:141-151.

45. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK: **Intrinsic disorder and functional proteomics**. *Biophys J* 2007, **92**:1439-1456.

46. Beckham GT, Bomble YJ, Matthews JF, Taylor CB, Resch MG, Yarbrough JM, Decker SR, Bu L, Zhao X, McCabe CM, *et al.*: **The *O*-glycosylated linker from the *Trichoderma reesei* Family 7 cellulase is a flexible, disordered protein**. *Biophys J* 2010, in press, doi:10.1016/j.bpj.2010.10.032.

47. Noach I, Frolow F, Alber O, Lamed R, Shimon LJW, Bayer EA: **Intermodular linker flexibility revealed from crystal structures of adjacent cellulosomal cohesins of *Acetivibrio celluloyticus***. *J Mol Biol* 2009, **391**:86-97.

48. Noach I, Levy-Assaraf M, Lamed R, Shimon LJW, Frolow F, Bayer EA: **Modular arrangement of a cellulosomal scaffoldin subunit revealed from the crystal structure of a cohesin dyad**. *J Mol Biol* 2010, **399**:294-305.

49. Bayer EA, Belaich JP, Shoham Y, Lamed R: **The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides**. *Annu Rev Microbiol* 2004, **58**:521-554.

50. • Bomble YJ, Beckham GT, Matthews JF, Nimlos MR, Himmel ME, Crowley MF: **Modeling the self assembly of the cellulosome enzyme complex**. *J Biol Chem* 2010, in press, doi:10.1074/jbc.M.110.186031.

This work describes a newly developed coarse-grained model for three enzyme components of the cellulosome, Cel5B, CelS, and CbhA along with a model of the cellulosomal scaffoldin. It was found that the large multi-modular Family 9 enzyme, CbhA, binds statistically more often than the smaller enzymes, CelS and Cel5B. Also, the physical characteristics governing binding to the scaffoldin appear to be modularity and flexibility of the enzymes. The results can be directly compared to experimental binding studies conducted with purified scaffoldin and enzymes. This study can be used to produce designer cellulosomes with enhanced or decreased binding affinity for the scaffoldin.

51. Zverlov VV, Velikodvorskaya GV, Schwarz WH, Bronnenmeier K, Kellermann J, Staudenbauer WL: **Multidomain structure and cellulosomal localization of the *Clostridium thermocellum* cellobiohydrolase CbhA**. *J Bacteriol* 1998, **180**:3091-3099.

52. • Pelikan M, Hura GL, Hammel M: **Structure and flexibility within proteins as identified through small angle X-ray scattering**. *Gen Physiol Biophys* 2009, **28**:174-189.

The authors review the current state of the art in data analysis for SAXS experiments with flexible linker domains between rigid domains. These types of systems are ubiquitous in carbohydrate-active enzymes, and thus SAXS and MD simulation offer a powerful set of complementary tools to analyze the solution behavior of these multi-domain proteins. There are opportunities here for the application of advanced sampling methods, such as rigid body REMD wherein linkers are allowed to be flexible, and rigid portions of the protein are allowed to be flexible to converge to a statistically relevant ensemble of structures.

53. Kataeva IA, Seidel RD, Shah A, West LT, Li XL, Ljungdahl LG: **The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface**. *Appl Environ Microbiol* 2002, **68**:4292-4300.

54. Craig D, Krammer A, Schulten K, Vogel V: **Comparision of the early stages of forced unfolding of fibronectin type III modules**. *Proc Natl Acad Sci U S A* 2001, **98**:5590-5595.

55. • Valbuena A, Oroz J, Hervas R, Vera AM, Rodriguez D, Menendez M, Sulkowska JI, Cieplak M, Carrion-Vazquez M: **On the remarkable mechanostability of scaffoldins and the mechanical clamp motif**. *Proc Natl Acad Sci U S A* 2009, **106**:13791-13796.

The authors used single-molecule force spectroscopy, biochemical experiments, and multiple resolutions of computer simulation to demonstrate that a dual 'mechanical clamp' (a patch of backbone hydrogen bonds) on the cohesin from *C. thermocellum* is one of the strongest mechanical regions known in proteins. This type of approach can be used to probe protein structures in cellulosomal components with experiment and simulation.

56. Carvalho AL, Dias FMV, Prates JAM, Nagy T, Gilbert HJ, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA: **Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex**. *Proc Natl Acad Sci U S A* 2003, **100**:13809-13814.

57. Alahuhta M, Xu Q, Bomble YJ, Brunecky R, Adney WS, Ding SY, Himmel ME, Lunin VV: **The unique binding mode of cellulosomal CBM4 from *Clostridium thermocellum* cellobiohydrolase A**. *J Mol Biol* 2010, **402**:374-387.

58. Heinzelman P, Snow CD, Wu I, Nguyen C, Villalobos A, Govindarajan S, Minshull J, Arnold FH: **A family of thermostable fungal cellulases created by structure-guided recombination**. *Proc Natl Acad Sci U S A* 2009, **106**:5610-5615.

59. Korkegian A, Black ME, Baker D, Stoddard BL: **Computational thermostabilization of an enzyme**. *Science* 2005, **308**:857-860.