A biophysical perspective on the cellulosome: new opportunities for biomass conversion
Shi-You Ding1, Qi Xu1, Michael Crowley1, Yining Zeng1, Mark Nimlos2, Raphael Lamed3, Edward A Bayer4 and Michael E Himmel1

The cellulosome is a multiprotein complex, produced primarily by anaerobic microorganisms, which functions to degrade lignocellulosic materials. An important topic of current debate is whether cellulosomal systems display greater ability to deconstruct complex biomass materials (e.g. plant cell walls) than nonaggregated enzymes, and in so doing would be appropriate for improved, commercial bioconversion processes. To sufficiently understand the complex macromolecular processes between plant cell wall polymers, cellulolytic microbes, and their secreted enzymes, a highly concerted research approach is required. Adaptation of existing biophysical techniques and development of new science tools must be applied to this system. This review focuses on strategies likely to permit improved understanding of the bacterial cellulosome using biophysical approaches, with emphasis on advanced imaging and computational techniques.

Address
1 Chemical and Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA
2 National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO 80401, USA
3 Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv 69978, Israel
4 Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Corresponding author: Ding, Shi-You (shi_you_ding@nrel.gov)


This review comes from a themed issue on Energy Biotechnology
Edited by Lee R. Lynd

Available online 29th May 2008

0958-1669/$ – see front matter
Published by Elsevier Ltd.

DOI 10.1016/j.copbio.2008.04.008

Introduction
Biological conversion of lignocellulosic materials has been proposed as a sustainable and renewable route for the production of liquid transportation fuels [1**]. Current technology for biomass conversion to biofuels, primarily bioethanol, involves the integration of three major unit operations (steps): particle size reduction and pretreatment, enzymatic hydrolysis, and fermentation of the lignocellulosic sugars. Pretreatment of biomass feedstocks produces materials that are more amenable to enzymatic digestion, often involving chemical treatment at a temperature range of 120–200 °C. Inoculation of the biomass during storage with microbial communities has also been proposed as a means of reducing required pretreatment severity [2]. In such processes, thermal/chemical pretreatment hydrolyzes easily available hemicelluloses, rendering the feedstock accessible to cellulases and hemicellulases, which catalyze enzymatic hydrolysis to soluble sugars. These sugars are subjected to fermentation for bioethanol production in a myriad of varied processing schemes. Enzymatic hydrolysis is often considered the feasibility-limiting step, because of the high cost and limited performance of currently available enzyme preparations. Indeed, current processing strategies have been derived empirically, with little knowledge of the fine structure of the feedstocks and even less information about the molecular processes involved in biomass conversion. Substantial progress toward cost-effective conversion of biomass to fuels would be fostered by fundamental breakthroughs in our current understanding of the chemical and structural properties that have evolved in the plant cell walls, which prevent its easy disassembly, collectively known as ‘biomass recalcitrance.’

Recently, new strategies in biotechnology have been pursued to reduce the cost of the cellulases used for biomass conversion. Most actual improvements in processing cost have come from work to improve enzyme productivity, not enzyme performance. Improvements in cellulase performance have been incremental, when reported, include engineering enzyme component mixtures (i.e. for superior synergism), enzyme robustness (usually assured when enzymes from thermophiles are used), and processing options designed to be synergistic, that is, simultaneous saccharification and fermentation (SSF).

In nature, there are currently two major types of cellulolytic systems recognized, those based on ‘free’ enzymes that are discretely acting cellulases typically produced by aerobic fungi and bacteria and those based on complexes of cellulolytic enzymes or ‘cellulosomes’ produced by some anaerobic bacteria. An important concept currently debated is whether or not cellulosomal systems display greater ability to deconstruct complex biomass materials, such as the plant cell walls, than do noncomplexed enzymes. For example, some evidence suggests that
specific hydrolysis rates on model cellulosic substrates are higher for the complexed cellulase system of Clostridium thermocellum than for the noncomplexed cellulase system of Trichoderma reesei, and are also higher for metabolically active cultures of C. thermocellum than for the cellulosome when it is not attached to the surface of a fermentative microorganism [3]. Although these results suggest potential and provide impetus for further study, there are also several important questions outstanding. These questions include how rates compare on pretreated plant cell walls rather than model substrates, as well as performance under conditions representative of an industrial process. If indeed cellulosomes exhibit superior biomass-degrading activities, they would provide an opportunity for improving current bioconversion technologies.

The present communication does not attempt to provide another extensive review of the cellulosome literature that has been the subject of numerous publications [4,5,6,7–9]. Instead, we focus on recently developed biophysical techniques that could potentially allow characterization of the cellulosomal system at the molecular level and thus generate the level of understanding required to assess optimal paths forward for cellulase research.

Native and engineered cellulosomes

The cellulosome was first described in the early 1980s using the thermophilic anaerobe, C. thermocellum [10]. Since then, several other microbial species, primarily anaerobic bacteria, have been reported to produce cellulosomes [4,11,12,70]. In addition, the subunit composition of an extracellular ‘xylanosome’ secreted by Streptomyces olivaceoviridis E-86 was determined [13].

In general, the cellulosome is composed of two major types of subunit: the noncatalytic scaffoldin(s) and the catalytically active components. The assembly of the cellulosome is facilitated by the high-affinity recognition between cohesin modules of the scaffoldin subunit and enzyme-borne dockerin modules. The scaffoldin often contains multiple cohesin modules, thereby enabling numerous different enzymes to be assembled into the cellulosome complex. In addition, multiple scaffoldins have been found in some species, which lends a higher level of complexity to the cellulosome assembly. Theoretically, nearly 100 components can be assembled into an individual cellulosome of A. cellulolyticus (Figure 1) [14,15,16,17]. Another important cellulosomal component is the cellulose-specific carbohydrate-binding module (CBM), which functions as the major binding factor for specific recognition of the polysaccharide substrates. CBMs can reside either in the cellulosomal scaffoldin or enzyme subunits.

Biochemical studies have determined different types of cohesins and dockerins in different microbial species, and the recognition between cohesin and dockerin is typespecific and species-specific. Designer cellulosomes have thus been proposed to engineer components of the native cellulosomes for controlled inclusion of selected enzymes into desired positions of an artificial complex [18–20,21]. Hypotheses that researchers propose to investigate using such designer cellulosomes

Figure 1

The Acetivibrio cellulolyticus cellulosome. (a) Scanning electron micrograph showing an A. cellulolyticus cell bound to cellulosic material [17]. (b) Schematic structure of the A. cellulolyticus cellulosome. Four scaffoldin subunits (ScaA, ScaB, ScaC, and ScaD) have been identified, which contain different types of cohesins. Three divergent types of cohesin–dockerin pairs are shown in red, green, and blue. Dockerin-containing enzymes are incorporated into the type I ScaA cohesins and one of the ScaD cohesins (coded in blue). The ScaA dockerin binds to the ScaB or ScaD type II cohesins (coded in red), and the ScaB dockerin binds to the ScaC cohesins (coded in green). ScaC and ScaD are anchored to the cell surface by their S-layer homologs (SLH) modules [14–17].
include: first, whether the proximity of different cellulolytic enzymes may provide synergistic action on the crystalline substrate, which would perform better than their free forms when positioned in a designed pattern; and second, whether enzymes from different species that have superior activities on given substrates can be assembled into a single complex to create designer cellulosomes for such substrates. Indeed, several researchers have constructed artificial cellulosomes by using a truncated scaffoldin that contains cohesins of equivalent specificity [23] or a chimeric scaffoldin that contains divergent cohesins [24], and by adding dockerins to free enzymes [24]. The key feature of the native cellulosome is the synergy function, in which an intricate macromolecular complex can degrade a heterogenous recalcitrant insoluble substrate in an efficient manner. The postulated plasticity of the quaternary structure of the cellulosome is the main rationale for the synergy [25,26].

Although significant progress in understanding cellulosomes has been achieved, exploring their mechanism of synergistic action is still a major challenge. The structures of single cellulosome-related proteins or simple complexes thereof have been determined by conventional technology, such as mutagenesis and structural biology. In addition, electron microscopy has been used preliminarily to analyze the macromolecular structure of the cellulosome as well as interactions between the cellulosome, microbial cell, and cellulosic materials [27–33]. However, these strategies are inadequate for determining the structure–function relationships crucial for understanding cellulosome action. Specifically, we need to understand the intricate protein–protein and protein–carbohydrate interactions of the massive, supramolecular cellulosome complex, in which it is believed that the synergistic actions between cellulosomal components, microbial cell surfaces, and the plant cell wall are the key. Experimental results have shown that the presence of cellulosome-producing microbial cells enhance the digestibility of cellulosomal cellulases that act on cellulose [34]; however, these studies were conducted with traditional tools attempting to analyze ensemble results. It would be useful to develop new tools to monitor the macromolecular dynamics in the cellulosomal ‘reaction zone’ in real-time. New biophysical techniques, such as advanced imaging approaches, are now being applied to cellulosome investigation, and we feel that these tools will provide insight into the general nature of cellulosome structure and function, with emphasis on the synergistic action among cellulosome components.

**Advanced imaging approaches**

The cellulosomal plant cell wall degrading system involves microbial cells and biomacromolecules in the solid and liquid states. Techniques designed to characterize such intricate systems must meet several criteria: nanometer scale resolution, minimal sample preparation, and natural environments (i.e. the use of buffer systems in ambient or controlled temperatures). Such requirements exclude most high-resolution electron microscopy techniques, yet are completely amenable to atomic force microscopy (AFM) and recently developed optical microscopy methods, such as nonlinear imaging and single-molecule methods. These techniques are particularly well suited to elucidating enzymatic actions on plant cell walls at the molecular level, that is, determining what they do, how they work individually, and how they work collectively (synergism).

**Nanoscale imaging of the plant cell walls using AFM**

AFM is one type of scanning probe microscopy (SPM), which is based on the principle of the scanned-proximity probe. Generally, AFM uses a microcantilever with a nanoscale tip to scan the sample surface. The distance between tip and sample is adjusted by a feedback mechanism using a piezoelectric scanner. Deflections generated by the forces between the tip and sample surface are measured by photodiodes through a reflected laser spot. AFM has been used increasingly for characterizing biomolecules, because imaging can be accomplished at atomic resolution and under aqueous solutions [35]. In practice, AFM measures attractive or repulsive forces that are sensitive to the structure and chemistry of both the tip and the sample. It is widely recommended to use a well-characterized tip and to understand as much as possible about the imaging surface to avoid experimental artifacts. Indeed, systematic artifacts encountered during early AFM studies were often hotly debated and eventually became fairly well understood [35]. Today, the AFM technique has been significantly improved using sharper and better characterized tips and improved tool control systems developed by the manufacturers [36**]. In the case of the cellulosome system, AFM can be used to characterize the surface structure of the plant cell walls, as well as the binding and assembly of the cellulosome complex.

Cell walls of various plant species, both from the Monocotyledonae and the Dicotyledonae, have been imaged using AFM [36**,37–41]. These studies have revolutionized our understanding of the molecular structure of plant cell walls, particularly the primary cell walls. For example, in never-dried plant cell walls, microfibrils were found to be smaller, uniformly distributed, and highly parallel, whereas in dried cell walls, the microfibrils were aggregated, disorganized, and twisted [42*,43]. These AFM measurements suggested that the dehydration processes could significantly affect the structure and arrangement of primary cell wall microfibrils. Furthermore, on the basis of AFM observations of maize parenchyma cell walls, we recently proposed a new model for the elementary cellulose fibril, as well as a potential biosynthesis pathway of cellulose and hemicelluloses [36**]. In brief, cellulose is synthesized by a protein assembly or rosette...
Nonlinear optical microscopy

In theory, when light is injected into nonlinear media, the dielectric polarization $P$ responds nonlinearly to the electric field $E$ of the incident light as described by

$$P = \chi^{(1)} E + \chi^{(2)} E^2 + \chi^{(3)} E^3 + \cdots$$

In principle, the coefficient $\chi^{(n)}$ is the $n$th order susceptibilities of the medium, and higher order frequency mixing could be extended for processes if $\chi^{(n)}$ is nonzero. In practice, second-order and third-order frequency mixing is usually applied. Nonlinear optical microscopes have recently been developed. Examples include multiphoton excited fluorescence (MPEF), second harmonic generation (SHG) and third harmonic generation (THG), and coherent anti-Stokes Raman scattering (CARS). These nonlinear optical techniques have been found to have several advantages for imaging biological systems, which include the following features: first, no need for extrinsic labels; second, sensitivity and specificity; and third, fast integration times. Moreover, nondestructive, long wavelengths are used, which exhibit only low levels of absorption to biomolecules, thereby significantly reducing the photodamage to biological samples [44**].

CARS has matured to be a highly sensitive, label-free nonlinear vibrational imaging technique [44**,45–47]. In CARS, two high power pulsed-laser beams, the pump ($\omega_p$) and Stokes ($\omega_s$) beams, stimulate the sample through a four-wave mixing process to generate an anti-Stokes field at frequency $\omega_m = 2\omega_p - \omega_s$. The induced nonlinear polarization by the anti-Stokes field reaches a maximum when the frequency difference between the pump and Stokes beam $\omega_p - \omega_s$ coincides with the frequency of a molecular vibration. Because the induced dipoles have a well-defined phase relationship, the molecular vibrators are stimulated to vibrate coherently when $\omega_p - \omega_s$, also called the beating frequency, matches their Raman-active vibrational frequency, thus providing a constructive coherent radiation many orders of magnitude stronger than normal Raman scattering. The much higher signal strength from CARS therefore allows for much higher data acquisition rate as compared with traditional Raman spectroscopy. CARS microscopy provides a highly efficient, nondamaging, and label-free approach to selectively imaging a certain chemical component, for example, lignin, in complexed biomass without the need for pretreatment.

Single-molecule spectroscopy

Single-molecule spectroscopy refers to a set of spectroscopic approaches capable of investigating the dynamics and kinetics of each molecule in an assembly of molecules. The great advantage of single-molecule spectroscopy over conventional ensemble approaches lies in the fact that this approach captures transient intermediates and provides direct information on the distribution of physical properties of a single molecule in a highly
heterogeneous system, thereby avoiding the problem of ensemble averaging.

The common approach to single-molecule spectroscopy involves analyzing the spectrum (fluorescence, Raman) under the manipulation (AFM, optical, and magnetic tweezers) of a specific key molecule that resides in the sample [48,49]. Single-molecule fluorescence has been widely applied to biological systems, and recently to probe transcription in living cells [50–52]. Basically, the illumination volume is minimized either by using a tightly focused laser beam through a high numerical-aperture (NA) objective or by utilizing the narrow evanescent wave generated along the coverglass surface from total internal reflection (TIR). The spatial resolution of a single molecule is diffraction limited to several hundred nanometers by the NA of the collection objective. However, the centroid of the spot, considered to define the position of the molecule, can be determined by fitting the spot profile with a point spread function (PSF) to surpass the diffraction limit. Recently, several microscopic techniques have been developed to achieve 2D resolution to <50 nm, such as stimulated emission depletion (STED), reversible saturable optical linear fluorescence transitions (RESOLFT) microscopy, saturated structured illumination microscopy (SSIM), and stochastic optical reconstruction microscopy (STORM). When two chromophores, composed of a Förster resonance energy transfer (FRET) pair, are tagged on a single molecule or two different molecules, the intermolecular or intramolecular distance change can be monitored by the FRET efficiency. Single-molecule FRET can be used as a molecular ruler, which can measure distance changes on the order of several nanometers.

Single-molecule approaches have recently been explored for studying issues in biomass conversion. Ding et al. [53] have used fluorescently tagged CBMs to probe the surface of carbohydrate-containing materials. In this case, CBMs were labeled with (CdSe)ZnS quantum dots or fluorescent proteins. Using this approach, single CBM molecules could be detected directly through their fluorescent tags. It is believed that different CBMs bind selectively to different cell wall constituents; this technique thus allows mapping the distribution of cell wall polymers at the molecular level of resolution. This approach can also be extended to correlative imaging of cellulose and hemicelluloses, as depicted in Figure 3. By measuring FRET efficiency, the distance between cellulose and hemicelluloses can be calculated. Furthermore, detailed polymer structures of the plant cell walls can be mapped by selecting combinations of different CBMs with different binding specificities, and different fluorescence proteins with different excitation wavelengths. Similarly, cellulosomal components can be labeled, and single-molecule approaches can be used to track their behavior. It is currently unknown how the native cellulosome is assembled. However, it is known that the cellulosome content changes during bacterial growth, and an individual cellulosome complex is dynamic and distinct in its composition [54–57]. In the future, employment of single-molecule analyses would be useful to better understand cellulosome biosynthesis and action.

**Computational approaches**

Computational modeling will be essential for developing a full understanding of the function of the cellulosome and its ability to disassemble plant cell walls into constituent sugars and oligosaccharides. Dynamics calculations can be used to explore enzyme/substrate conformational space in order to understand structural properties and limitations of cellulosomal structures and to determine energetics and reaction kinetics. Computation...
Computational modeling is a powerful tool to generate new hypotheses for experimental validation, and such new ideas are especially crucial for the case of cellulosome mechanism. When combined with experimental measurements and mutational biochemistry, dynamics modeling is crucial for understanding the mechanism of cellulosome action.

A major challenge for computational modeling will be developing robust methods that are capable of treating the large cellulosome system and its interaction with carbohydrate microfibrils. Molecular dynamics computational technology contains a rich assortment of tools and approaches that are capable of investigating and describing enzyme/substrate/water systems that contain up to $10^7$ to $10^8$ atoms. However, the size of the cellulosome with its substrate and water molecules is likely to exceed $10^8$ atoms and is thus out of the reach of these atomistic models. Even when the methods are scaled up to $10^8$ atoms, the simulation times necessary for meaningful results are beyond the capabilities of even the largest supercomputers for years to come [58], so it is unrealistic to model cellulosomes with atomistic models. In general, the high frequency motions characteristic of atoms are not the significant characteristic of large molecular structures that describe their important behaviors as a composite structure. Instead, large-scale molecular motions and structural properties describe this behavior and can be described by the interactions of structural components like alpha helices, beta sheets, and other structural moieties. Therefore, models to describe cellulosome systems will use a coarse-grain approach, where groups of atoms are treated as single flexible units or units with structural properties. Modeling at this mesoscale will significantly simplify the calculations, but if the model is chosen carefully and is well parameterized, accurate information can be gained about the structure and function of the cellulosome.

The level of granularity needed for these calculations will depend on the properties of the cellulosome that are being investigated. Figure 1 shows an example of a model section of a hypothetical cellulosome at different grain sizes. Determination of the binding interactions of the different domains of a cellulosome will initially require atomistic calculations (Figure 4a). However, these will probably have to be restricted to a pair of domains at a time, such as the dockerin–cohesin complex, cohesin-linker-cohesin subdomains, or the enzyme-linker-dockerin complex. Replacing residues or groups of residues with individual ‘beads’ [59,60] (Figure 4b) will allow assembly of the components of a simple cellulosomal scaffoldin with cohesins, dockerins, catalytic domains, and binding domains. This model will allow the investigation of the conformational range of the cellulosome in water solution and near a carbohydrate substrate. Cooperative effects on account of adjacent catalytic domains on the scaffoldin can also be investigated, as can translational restrictions imposed by the scaffoldin structure and the effect of interactions between subdomains of adjacent components on cellulosome structure and structural and mechanical properties. Higher granularity [61,62] (Figure 4c) will allow the assembly of many cellulosomes, which will allow the study of bulk interactions between the microorganism and the cell wall substrate. Study of the interactions with substrate will require coarse-grain
modeling of cellulose and other cell wall components [63] and their interactions with the individual components of the cellulosome. Studies of the assembly of the cellulosome by attaching enzyme-dockerins to cohesins on the scaffoldin can be used to determine the effects of different enzyme concentrations, the ordering of enzymes on the scaffoldin, and conformational changes accompanying binding of enzymes. This modeling/experimental engineering approach may lead to human-designed cellulosomes that can selfassemble into desired compositions and shapes, ‘tailored’ for effectiveness on plant cell walls of interest.

The use of coarse-grain techniques for modeling the cellulosome will require the development of force fields that describe the interactions of larger grains (such as ‘beads’). Typically, these interactions are quantified based upon calculations done at a finer scale. For instance, the force parameters for atomistic calculations are often determined from electronic structure calculations. Likewise, parameters for bead calculations will be based on simulations done at the atomistic scale. For proteins, there has been a great deal of progress made in the development of mesoscale models [59,60–62,64] and these advances can help the development of models for the cellulosome.

Once the model and force field has been developed, much of the machinery that has been developed for atomistic calculations can be used for mesoscale models. Software suites such as CHARMM [65], Amber [66], NAMD [67], Gromacs [68], and LAMMPS [69] contain numerous molecular dynamics tools and analysis tools that can be used for bead dynamics modeling. These programs will allow the evaluation of structures, conformational space, and free energies. These tools will be important drivers for the development of a detailed understanding of cellulosomes.

**Conclusion**

In natural ecosystems, recycling of plant biomass is a relatively slow process [6**]. Communities of plant cell wall-degrading microbes have evolved in these ecosystems, and the species therein vary, depending on the substrates and conditions of the environment. It is easy to understand why such complex ecosystems are required to degrade plant biomass, considering the intricate nature and recalcitrance of the different polymers that comprise plant cell walls. In order to deconstruct such complex natural structures, the degradative enzymes that contribute to these processes often function synergistically and the cellulosome is indeed a multiple enzyme aggregate. However, it may be a mistake to consider the cellulosome as simply another form of glycosyl hydrolases, presented to the cell wall in a spatially defined and compact format. One must consider that the cellulosome functions in a highly defined context, trapped closely between the plant cell wall (substrate) and the microbial cell wall. Moreover, the entire microbial cell, as well as consortia of complementary cells, are also crucial to the efficient degra-
dation of plant cell wall material in nature [34]. An added challenge is that biomass substrates are continuously modified during natural senescence, storage, and thermal/chemical pretreatment.

The molecular interactions among the enzymes, enzyme–cell, enzyme–substrate, and cell–substrate remain poorly understood today. The cellulosomal system provides an ideal and much needed tool with which to study the cellular and molecular interactions during microbial biomass conversion. However, before a practical approach can be implemented for its use in the efficient deconstruction of plant cell wall biomass, a deeper understanding of cellulosomal functionality would be useful. Application of the biophysical methodologies described here may reveal important new clues about the functionality of this multi-enzyme, multidimensional scaffold, which enables biomass deconstruction by many microorganisms.

Acknowledgements

This work was supported partly by the DOE Office of Biological and Environmental Research-Genomes to Life Program through the BioEnergy Science Center (BESC). This work was also supported by the Israel Science Foundation (Grant Nos 422/05 and 159/07), and by grants from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. The authors also wish to thank Lee Lynd for helpful comments made during the writing of this work.

References and recommended reading

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

- of special interest
- of outstanding interest


21. Fierobe HP, Mingardon F, Mechaly A, Belaich A, Rincon MT, Pages S, Lamed R, Tartic D, Belaich JP, Bayer EA: Action of designer cellulosomes on homogeneous versus complex substrates — controlled incorporation of three distinct enzymes into a defined trifunctional scaffold. J Biol Chem 2006, 280:16325-16334. This study investigates the designer cellulosomes on various substrates. Enhanced activities on complex cellulose substrate were found when both cellulases and hemicellulases are incorporated.


