

Nanoscale Engineering of Designer Cellulosomes

Melissabye Gunnoo, Pierre-André Cazade, Albert Galera-Prat, Michael A. Nash, Mirjam Czjzek, Marek Cieplak, Beatriz Alvarez, Marina Aguilar, Alon Karpol, Hermann Gaub, Mariano Carrión-Vázquez, Edward A. Bayer, and Damien Thompson*

Biocatalysts showcase the upper limit obtainable for high-speed molecular processing and transformation. Efforts to engineer functionality in synthetic nanostructured materials are guided by the increasing knowledge of evolving architectures, which enable controlled molecular motion and precise molecular recognition. The cellulosome is a biological nanomachine, which, as a fundamental component of the plant-digestion machinery from bacterial cells, has a key potential role in the successful development of environmentally-friendly processes to produce biofuels and fine chemicals from the breakdown of biomass waste. Here, the progress toward so-called “designer cellulosomes”, which provide an elegant alternative to enzyme cocktails for lignocellulose breakdown, is reviewed. Particular attention is paid to rational design via computational modeling coupled with nanoscale characterization and engineering tools. Remaining challenges and potential routes to industrial application are put forward.

organisms including plants and algae convert, through photosynthesis, inorganic carbon into organic carbon that can be processed by heterotrophic organisms.^[1] Nowadays, the major source of carbon and energy in nature is fiber, i.e., the plant cell wall formed by polysaccharides, mainly cellulose and hemicellulose (Figure 1). These two compounds are the first and second most abundant organic molecules on Earth, respectively (plants produce almost 200 billion tons of cellulose per year globally), and offer a renewable, virtually inexhaustible feedstock not only for the production of biofuels but also for a variety of fine chemicals. However, the secondary cell wall (produced after the cell has stopped growing) is strengthened by polymeric (non-polysaccharide) lignin,

covalently cross-linked to hemicellulose (Figure 1B). This forms a recalcitrant, i.e., difficult to degrade, carbohydrate material called lignocellulose. Since biofuel production relies on the transformation of simple sugars into ethanol by yeast, a large

1. Introduction

The synthesis of organic carbon is a major biological process and the primary source of energy for life. Using sunlight,

M. Gunnoo, Dr. P.-A. Cazade, Dr. D. Thompson
Materials and Surface Science Institute and
Department of Physics and Energy
University of Limerick
Limerick, Ireland
E-mail: Damien.Thompson@ul.ie

A. Galera-Prat, Prof. M. Carrión-Vázquez
Instituto Cajal
Consejo Superior de Investigaciones Científicas (CSIC)
IMDEA Nanociencias and CIBERNED
Madrid, Spain

Dr. M. A. Nash, Prof. H. Gaub
Lehrstuhl für Angewandte Physik and Center for Nanoscience
Ludwig-Maximilians-University
80799 Munich, Germany

Dr. M. Czjzek
Sorbonne Universités
UPMC
Université Paris 06, and Centre National de la Recherche Scientifique
UMR 8227, Integrative Biology of Marine Models
Station Biologique, de Roscoff, CS 90074
F-29688, Roscoff cedex, Bretagne, France

Prof. M. Cieplak
Laboratory of Biological Physics
Institute of Physics
Polish Academy of Sciences
Warsaw, Poland

Dr. B. Alvarez
Biopolis S.L.
Parc Científic de la Universitat de Valencia
Edificio 2, C/ Catedrático Agustín Escardino 9
46980 Paterna (Valencia), Spain

Dr. M. Aguilar
Abengoa, S.A.
Palmas Altas
Calle Energía Solar nº 1
41014 Seville, Spain

Dr. A. Karpol
Designer Energy Ltd.
2 Bergman St.
Tamar Science Park
Rehovot 7670504, Israel

Prof. E. A. Bayer
Department of Biological Chemistry
Weizmann Institute of Science
Rehovot 76100, Israel



DOI: 10.1002/adma.201503948

worldwide effort is currently focused on the identification^[2] and engineering^[3,4] of enzymes capable of efficiently degrading this feedstock to simple sugars.^[5] Cellulose is a linear homopolymer with (1→4) β -linked D-glucose, which is present in plant tissue primarily as an insoluble crystalline matrix of parallel glucan chains (Figure 2). On the other hand, “hemicellulose” is a polymer formed by a variety of compounds (e.g., xylans, xyloglucans, arabinoxylans and mannans) in complex branched structures with a wide range of substituents (e.g., acetyl and feruloyl groups). Hemicelluloses are usually bound to cellulose and to other hemicelluloses via hydrogen bonding and hydrophobic interactions (Figure 3), which help stabilize the cell wall matrix. The high complexity of this material requires cascades of highly specialized enzymes to catalyze its breakdown.

Life is based on the nanoworld, the intracellular milieu of protein molecules (made from information encoded in DNA) that recognize and bind other molecules. Many complex biological processes are catalyzed by self-assembled cascades of multiple enzymes that act in a coordinated manner. To this end, organisms have developed protein scaffolds that anchor the corresponding enzymes so that their activities can be spatiotemporally coordinated in an efficient manner. Not surprisingly, these scaffolding proteins are increasingly attracting the attention of materials scientists.^[8] Nature has provided a complex but remarkably efficient means of breaking down fibers and releasing sugars using the cellulosome nanomachine (see below), a process with large net benefits for bacteria and fungi. These specialized organisms have evolved complex nanomaterials solutions to enable their growth on plant material rich in cellulose, with enzyme cascades ensuring that the work required to break down fibers is less than the energy gain associated with consuming the sugars.

The exploitation of the enzymatic conversion of crystalline polysaccharides, often called “saccharification”, is crucial for the production of biofuels and fine chemicals through environmentally benign means. However, in order to be cost effective, industry requires efficient and safe methods for the breakdown of biomass waste (residues that generate the so-called “second generation” biofuels and other products). In nature, microorganisms that produce these enzymes typically require weeks, months or even years to decompose feed stocks, e.g., fallen tree trunks, in contrast to the putative enzymatic systems sought by industry that could do the job on a much shorter scale, of days or even hours, given pre-treatment. For this reason, establishing an efficient enzymatic conversion process for crystalline polysaccharides is an active area of research. The purpose of this article is to critically review recent efforts to re-engineer the natural cellulosome enzymes to make catalytic materials for the renewable energy sector.

A specialized sub-class of anaerobic bacteria performs enzymatic conversion of crystalline polysaccharides using a range of biological nanomachines called cellulosomes (Figure 4). Enzymes, more specifically cellulases, are the proteins embedded in the cellulosome complex that catalyze cellulose hydrolysis. Detailed knowledge of the nature of the different enzymes and their interactions is essential for the understanding of the biological reactions at play during cellulosome breakdown. The structure and function of the cellulosome has been greatly clarified in recent years using a variety of



Melissabye Gunnoo received her BSc degree in physics with biomedical sciences from Dublin City University in 2013. She is currently studying for a doctoral degree in Dr. D. Thompson's group at the University of Limerick, with a particular interest in modeling the mechanical stability of proteins.



Pierre-André Cazade graduated from the University of Pau, France, with a PhD degree in physical chemistry in 2008 and then performed postdoctoral studies at the University of Montpellier and the University of Basel. He is currently a postdoctoral researcher in Dr. D. Thompson's group at the University of Limerick, studying the molecular dynamics of flexible linkers in multi-protein assemblies.



Damien Thompson graduated from the University of Limerick with a PhD degree in computational chemistry in 2003 and then performed postdoctoral studies in biophysics at École Polytechnique France from 2003 to 2005, before joining the Materials Modelling for Devices group at Tyndall National Institute. From 2013, he has been a Lecturer in Physics and Principal Investigator at the Materials and Surface Science Institute at University of Limerick. His main research interests are in the areas of self-assembled monolayers and protein re-engineering.

experimental methods,^[9] but several crucial questions remain open. Some of these include: How transferable are cellulosome components between species? Do the disordered segments that link components (Figure 5) play a key functional role or are they evolutionary holdovers akin to non-coding “junk” DNA? What role does the balance between the mechanostability conferred by the nanoarchitecture and the flexibility required for functionality play in maintain the integrity and steering the

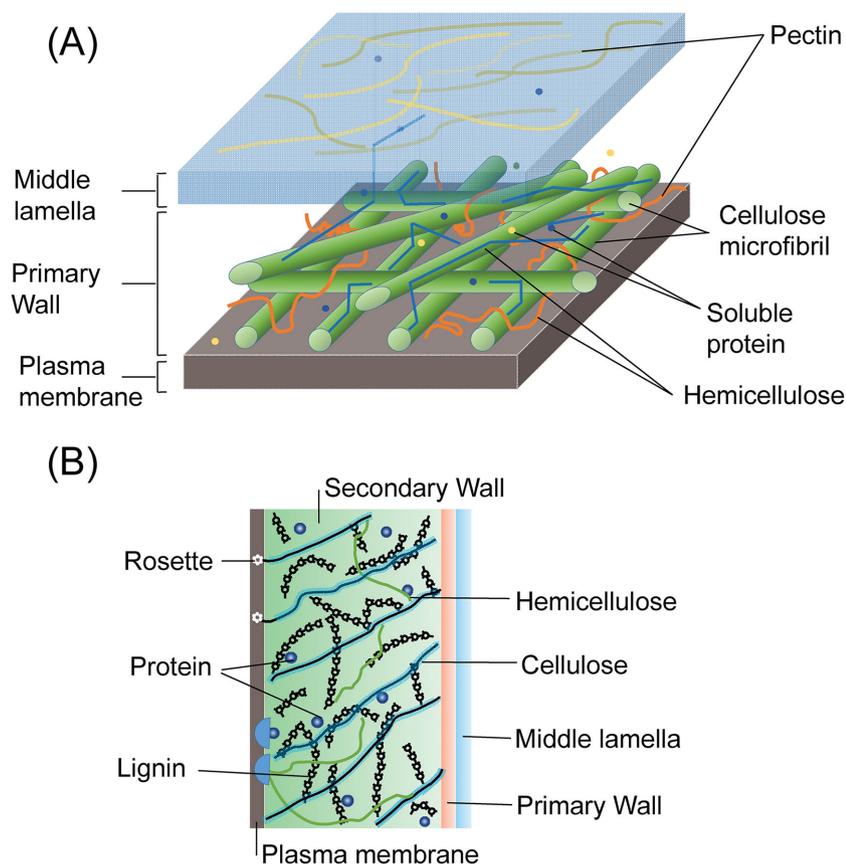


Figure 1. A) Plant cell wall-bound cellulose and hemicellulose, a potential feedstock for production of biofuels and fine chemicals. Shown is a section of a plant cell wall, made up of cellulose microfibrils (green), hemicellulose (blue), pectin (orange) and soluble proteins. The secondary wall is omitted for clarity. B) Cross-section of the secondary cell wall showing the cellulose synthase enzyme rosette complexes that float in the plasma membrane. Advanced bionanocatalyst materials based on the cellulosome, a biological nanomachine, are attractive candidates for efficient breakdown of lignocellulose into useful chemicals. Cellulosomes are multi-enzyme complexes whose building block is the intermodular cohesin-dockerin pair (Figures 3 and 4). Adapted with permission.^[6] Copyright 2008, Nature Publishing Group.

collective dynamics of the cohesin, dockerin, carbohydrate binding modules (CBMs) and enzymatic subunits (Figure 4) that make up the cellulosome? Can site-directed mutagenesis provide a route to re-engineer the catalytic units and modify the specificity in order to boost turnover rates? We will attempt to shed light on these questions in the search for a reasonable approximation to the cellulosome “minimal unit”, which must be first identified in order to fully realize and enable technology transfer of suitably re-engineered “designer cellulosome” materials.^[10,11]

Following the saccharification step, a variety of microorganisms can be used to ferment the products of hydrolysis of polysaccharides and yield desirable end products, including ethanol and longer chain alcohols. Agricultural residues, forest wood, herbaceous crops and municipal solid wastes have so far been considered as feed stocks for ethanol production.^[13] These materials primarily consist of cellulose, hemicellulose and lignin. Once the cellulose is converted to glucose, this compound is easily fermented to ethanol by yeast, a process well developed by industry. Conversion of cellulosic stocks into ethanol has the

advantages of readily available resources, the carbon neutrality of the plant fiber, the benign environmental impact of the alcohol product as a biofuel and the avoidance of incineration and transportation of waste products.

In addition to the economic and environmental benefit of converting biomass into useful chemicals, the cellulosome provides materials scientists with a fully worked out example of an advanced functional nanostructured material. Other evolved nanomachines that are beginning to be understood at the molecular level include DNA topoisomerase, RNA polymerase and the genetic code translation machinery (involving a large range of components and active sites ranging from DNA- and RNA-binding proteins to complete ribosomes). The structure and workings of some of these bionanomachines are briefly summarized in Section 2, and they provide rich inspiration for rational design of nanomaterials; the interested reader is directed to other sources for excellent reviews of the more general chemistry and physics of nanostructured materials.^[9,11,14–23]

2. Biocatalysts

Techniques such as atomic resolution crystallography (and, increasingly, electron microscopy and X-ray scattering) have revealed a wealth of mechanisms that biological nanomachines use to perform their tasks. Notable examples include DNA topoisomerase, RNA polymerase, ribosome, and kinesin.^[14,24] We provide here a brief overview of the current state of the art in molecular-level knowledge of how these structures function, to provide con-

text and to illustrate the broad lessons available to materials science from (evolved) biological nanostructures.^[25] See, for example, the self-assembled biosensor described by Han et al.^[26] These molecular machines operate at scales between 0.1 and 100 nanometers, exploiting nanoscale physics, in particular Brownian motion and van der Waals sticking forces, to manipulate individual atoms in order to process (and manufacture) materials from the bottom up. Particular attention is paid to modeling studies that complement experimental characterization and serve as templates for simulating the dynamics and energetics of million-plus atom biological nanomachines.

2.1. DNA Topoisomerase

Chemical analyzes together with X-ray crystallography, supported by bioinformatics and molecular modeling studies, have provided key insights into the nanoscale mechanisms used by topoisomerase to mediate the replication, transcription and recombination of DNA.^[27] The X-ray data are fragmentary but represent a broad range of enzyme classes. Combined with

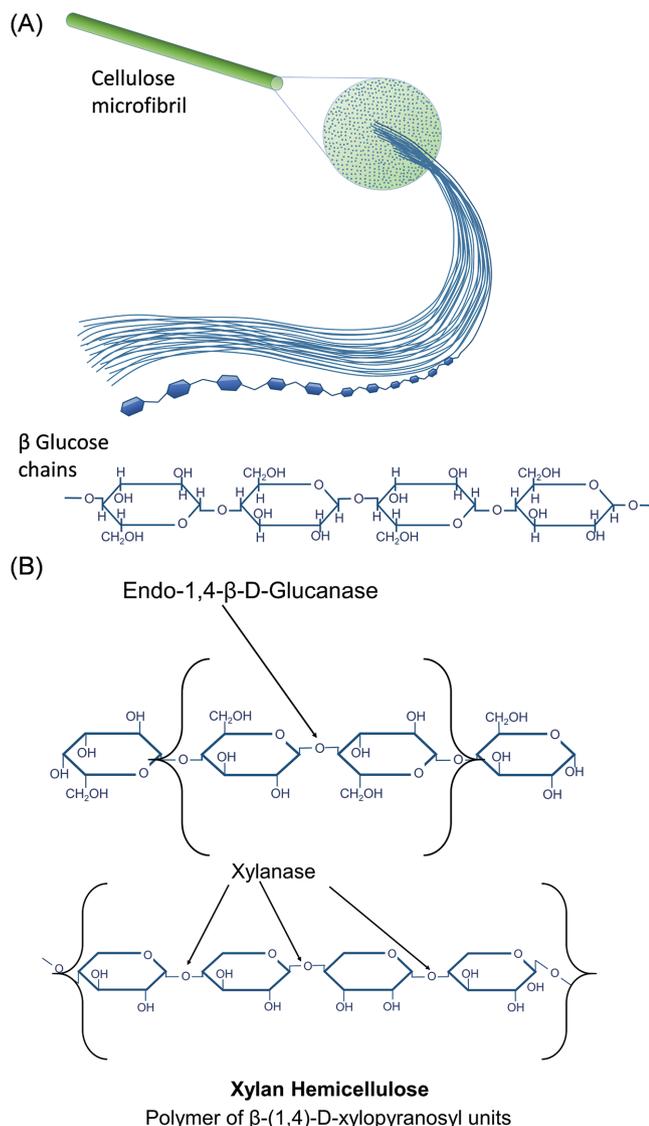


Figure 2. A) Cellulose microfibrils are composed of strings of cellulose units. Long polypeptide chains of β -D-glucose are held together by hydrogen bonds to form a bundle of microfibrils. The β -D-glucose units are linked by (1 \rightarrow 4) glycosidic bonds. B) Cellulases are enzymes that catalyze the rupture of 1,4- β -D-glycosidic linkages in cellulose β -D-glucan chains. Endo-1,4- β -D-glucanase breaks bonds in non-terminal regions, producing oligosaccharides. Analogously to glucanase, xylanase attacks 1,4- β -D-xylosidic linkages in xylan hemicellulose, a polymer of β -(1,4)-D-xylopyranosyl units, yielding various xylooligosaccharides including xylan, xyloglucan (1-6-linked xyloses on glucose chains), and arabinoxytan.

bioinformatics, they allow general lessons to be extrapolated analogously to the way in which fossils and genomic data are combined in paleontology to piece together overall architectures and corresponding mechanisms. The same crystal-centered approach has driven much of the research into cellulosome architecture and function, as described later in this review. The topoisomerase enzymes manage the rupture and remaking of DNA by creating temporary breaks in the DNA structure, and then threading a single strand of the DNA through a break in the opposing strand before healing (type-I subfamily) or by

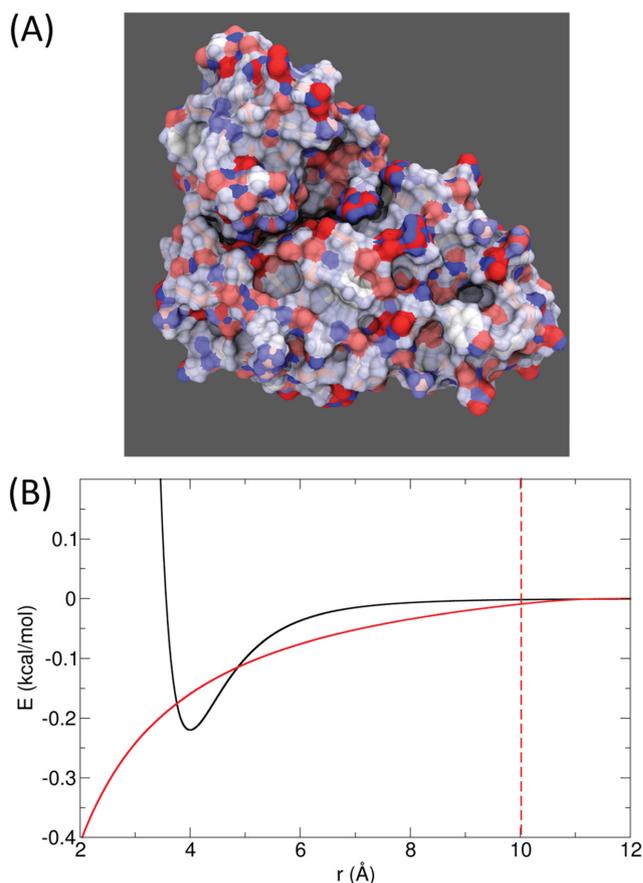


Figure 3. A) An example of a tightly coupled cohesin-dockerin bio-interface (marked by the black shadow) held together by H-bonds and hydrophobic interactions. The electrostatic potential surface is overlaid on the structure (PDB code 1OHZ) with sites colored according to atomic charge. This and all molecular graphics in this review were produced using visual molecular dynamics (VMD) software.^[7] B) The corresponding electrostatic and van der Waals's potentials that quantify the sub-microscopic driving forces behind the assembly of nanostructured (bio)materials. These Lennard-Jones (black) and electrostatic (red) potentials are used in molecular dynamics simulations. The parameters used here correspond to a positive (+1e) and a negative (-1e) interacting charges for the electrostatic component, while typical carbon-carbon separations are used for the LJ potential. The switching procedure is used to dampen the electrostatic potential between 10 and 12 Å so that it goes to zero at the cut-off (12 Å).

cutting and reknitting both strands of the DNA helix to heal DNA tangles (type-II subfamily). The type-II family uses the hydrolysis of ATP to fuel these interactions, unlike the type-I topoisomerases that rely on DNA interactions with the protein (e.g., **Figure 6**) to drive the protein conformational changes that accompany each step in DNA relaxation.

Molecular dynamics simulations^[29] have used available crystal data as the starting point to access kinetic and thermodynamic information not easily obtained using experiments alone, including: domain opening/closing by conformational shifts modeled using steered molecular dynamics, identification of key functional residues (and candidates for topoisomerase re-engineering via site-directed mutagenesis) by calculation of binding free energies, rational drug design

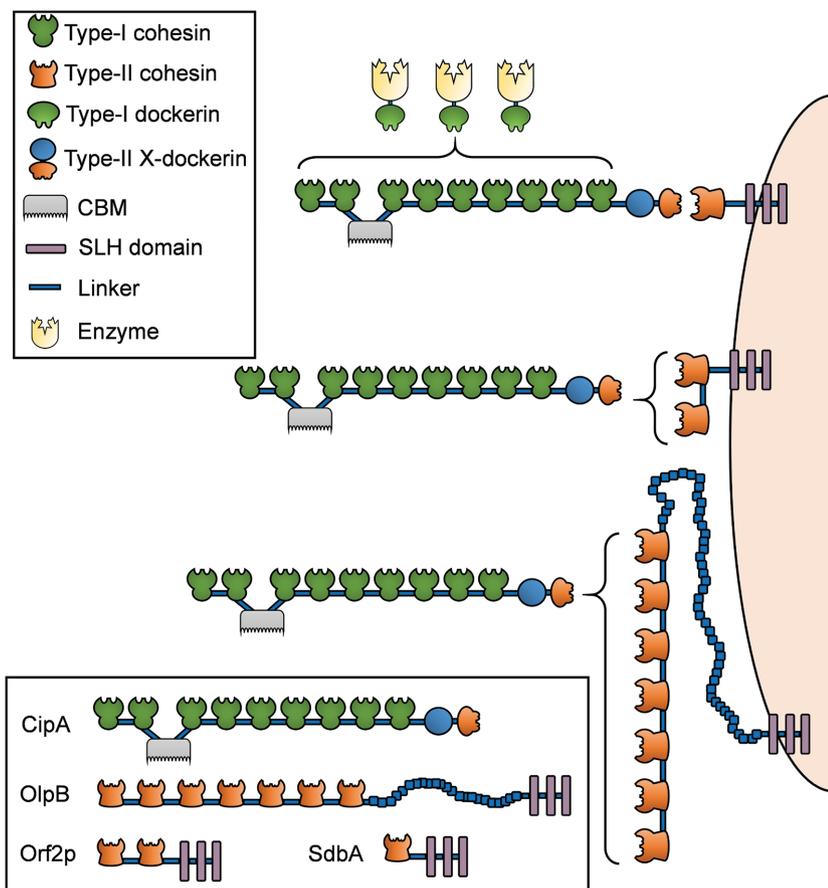


Figure 4. The cellulosome architecture postulated for *C. thermocellum*. The component units and assemblies are described in the text. Acronyms: CBM is the carbohydrate binding module, SLH is S-layer homology domain, CipA is a scaffolding protein and anchoring proteins OlpB, Orf2p, and SdbA refer to the outer-layer protein, a scaffolding containing two cohesin modules (which is thus capable of binding two CipA molecules), and a scaffolding dockerin binding protein, respectively. The schematic representation of the *C. thermocellum* cellulosome on the bacterial cell wall shows the dockerin-containing enzymes incorporated into the primary scaffoldin (CipA) via interaction with type-I cohesins. The *C. thermocellum* scaffolding CipA contains nine type-I cohesins. The type-II cohesins of the anchoring scaffoldin (OlpB, Orf2p or SdbA) bind specifically to the C-terminal type-II dockerin domain of CipA. One single CBM in the primary scaffoldin targets the cellulosome complex (and consequently the entire cell) to the cellulose substrate. The specificities of the different cohesin-dockerin pairs are color-coded.

by calculation of topoisomerase-inhibitor binding profiles to explain known sequence selectivity, and the structural stability of ternary complexes. modeling techniques are described in detail in Section 4.

These studies of structure-function relationships in DNA topoisomerase have highlighted the key role of controlled DNA substrate motion by specific interactions with the topoisomerase scaffold. The evolution of such functional interfaces is also a key component of the cellulosome, which require not only precise control of sticking interactions but also controlled molecular motion. The kinesin nanomachine that harnesses Brownian motion will be described below, but we conclude this brief summary of topoisomerase by mentioning recent simulation studies that revealed how DNA-protein interactions can exploit two evolved design principles that allow ultra-fast sampling of the protein architecture by the DNA substrate.^[30]

DNA-binding proteins work by combining three-dimensional diffusion in solution with one-dimensional sliding along DNA. The need to form a stable protein-DNA complex at the target site poses a challenge because, similar to cellulosome, the architecture contains many identical and near-identical binding sites. The evolved workaround is termed “frustration”; a compromise between rapid sliding and rapid formation of an active protein-DNA complex,^[30] and this provides the right balance between motion and sticking.^[25] Based primarily on bioinformatics and coarse-grained molecular dynamics simulations (which have been validated against all-atom models of topoisomerase^[31], another important insight from these studies^[30] in the context of designer cellulosomes is the finding that the amino acid sequence of the flexible tails of DNA-binding proteins were evolutionarily selected to be long and positively charged to facilitate DNA search. Hence these disordered regions, which resemble the flexible linkers that join cohesins in the scaffolding and tether dockerins to catalytic modules (Figures 4 and 5), can play very specific roles in the mechanism of substrate-enzyme association and recognition. Structural disorder does not necessarily equate with non-function in bionanomaterials, hence the need for chemical analyzes and informatics/simulation tools to complement structure determinations.

2.2. RNA Polymerase

Cells use RNA polymerase to make RNA chains based on DNA gene templates in a process called transcription that is used by all living organisms and many viruses. Similar to the DNA-binding proteins described above, RNA polymerases use random Brownian motion to sample a range of possible binding sites on DNA. The high binding energy of complementary pairing fuels the overall transcription process.^[32] The RNA molecule then acts as a template for producing polypeptides at the ribosome (see below), completing the production line from gene to protein. The crystal structures reveal a dynamic, highly-evolved architecture, featuring folding of a “clamp” over the DNA substrate and RNA product and five “switch” regions. These molecules shift and reshape to fit the DNA–RNA hybrid together with a ratchet-like mechanism. Concerted loop and bridge motions maintain a ratio of active and inactive states of the elongation complex and provide directed motion at the molecular scale to propel transcription forward. One representative crystal structure is shown in **Figure 7**, which captures the polymerase machine “in the act” of unravelling and transcribing DNA into the RNA strand.

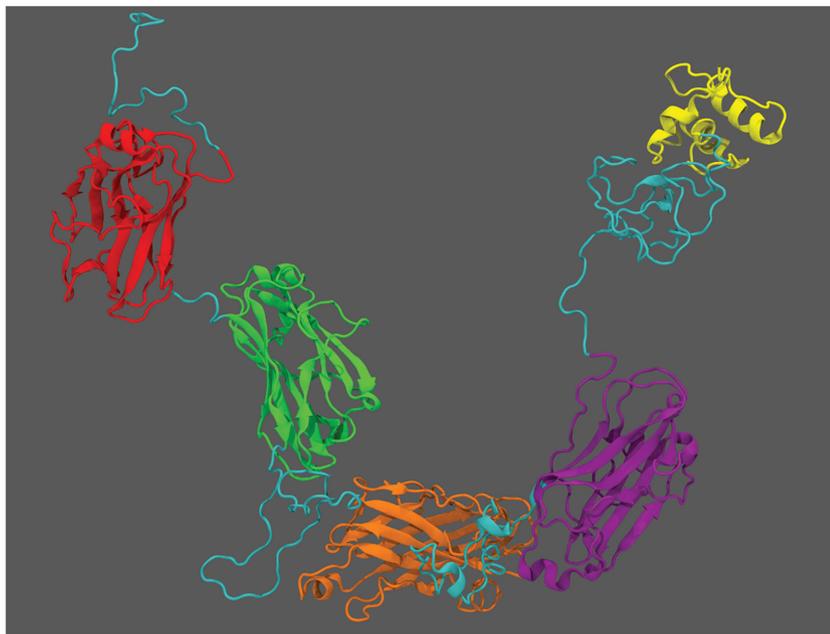


Figure 5. Disordered “linkers” join functional units to make a first approximation of a molecular model of a “scaffoldin” architecture. The scaffoldin coordinates a variety of polysaccharide-degrading enzymes into the highly polymorphic cellulosome complex. This multi-enzyme complex has been shown to be much more efficient in degrading plant cell wall derived polysaccharides than the sum of the component enzymes alone.^[12] The representation shown is of the adaptor scaffoldin ScaB from *A. cellulolyticus*. The four CohII (type-II cohesin) units are shown in red, green, orange, and purple. The dockerin module is colored yellow, while cyan is used for the linkers.

Molecular simulations have served as a valuable ally to experimental studies of RNA polymerase. Details from calculations of dynamics and energetics provide leads for polymerase re-engineering^[33,34] and production of synthetic analogues of RNA polymerase, e.g., rotaxane-based nanomachines.^[35] In particular, recent millisecond simulations^[34] revealed unexpected intermediate stages in the processing of DNA by RNA polymerase. These simulations used a multi-scale approach that extrapolates from conventional sub-microsecond molecular dynamics models using Markov models^[36] to approach the real timescale for translocation (at least tens of microseconds). They also represent an important step forward in the state of the art for using modeling to describe features that are important for rational design but which are difficult if not impossible to measure experimentally. The method works by scanning for local minima in translocation pathways and initiating a series of molecular dynamics simulations along the path. Then combining the trajectories and piecing together a Markov state model provides a model for translocation at atomic resolution and millisecond timescales.

2.3. Ribosome

The ribosome represents a very complex biological nanomachine, optimized by evolution to read RNA and translate these instructions into the correct sequence of amino acids to form a polypeptide chain. The ribosome is essentially the protein-production factory of the cell.^[38]

Despite the daunting complexity of the ribosome, it has been exploited very successfully in nanobiomaterials synthesis, specifically protein arrays, where a technique known as “ribosome display” is used to tether selected proteins to their corresponding mRNA. Thus, instead of random deposition of proteins as in dip-pen nanolithography, one can control the position and orientation of proteins on the surface of biochips, to maximize the surface coverage of active protein.^[39] The available ribosome crystal structures have highlighted that the interaction of RNA with protein is sufficient to perform all the ribosomal tasks^[38,40] of decoding mRNA and adding the amino acids one at a time to a growing polypeptide chain, at a rate of fifteen peptide bonds per second, which provides a target upper speed limit for “bottom-up” materials assembly.

Molecular simulations have identified several key features of the RNA-protein interaction that were not clear from the crystal structures alone.^[41–44] For example:

- (i) the key functional role of ordered cations that strengthen the binding of tRNA molecules (and which are important for the design of tetracycline antibiotics to inhibit bacterial ribosomes),^[41]
- (ii) how a tunnel of universally conserved ribosomal RNA bases acts as a gate to mediate access of tRNA into the ribosome,^[44]
- (iii) the mechanism of assembly of the small ribosomal subunit and polypeptide folding in the ribosome tunnel,^[42] and
- (iv) the enthalpy-entropy balance in coupled protein/RNA dynamics.^[43,45]

2.4. Kinesin

We conclude this brief survey of biocatalysts with a description of the kinesin motor proteins^[46] that “walk” along microtubules found in the cytoplasm of cells.^[47] They serve as a striking example of directed motion along a scaffold and share common features with the cellulosome, for example in targeting a specific substrate to a scaffold composed of a host of repeat units with defined architecture. Recent applications in materials science include the use of kinesin in developing nanoscopic models of wear in molecular materials,^[48] and the use of an active unit in autonomous dust sensing^[49] and optical gear-shifting motors.^[50] A range of synthetic analogues of kinesin walkers have been developed including DNA-based and small-molecule walkers (e.g., **Figure 8**). Given that widespread industrial uptake of nanotechnology requires inexpensive, easy and robust solutions that allow manipulation of matter at the smallest scales, a key enabling feature will be the ability to move material around molecule by molecule. The research described in Perl et al.^[51] enabled

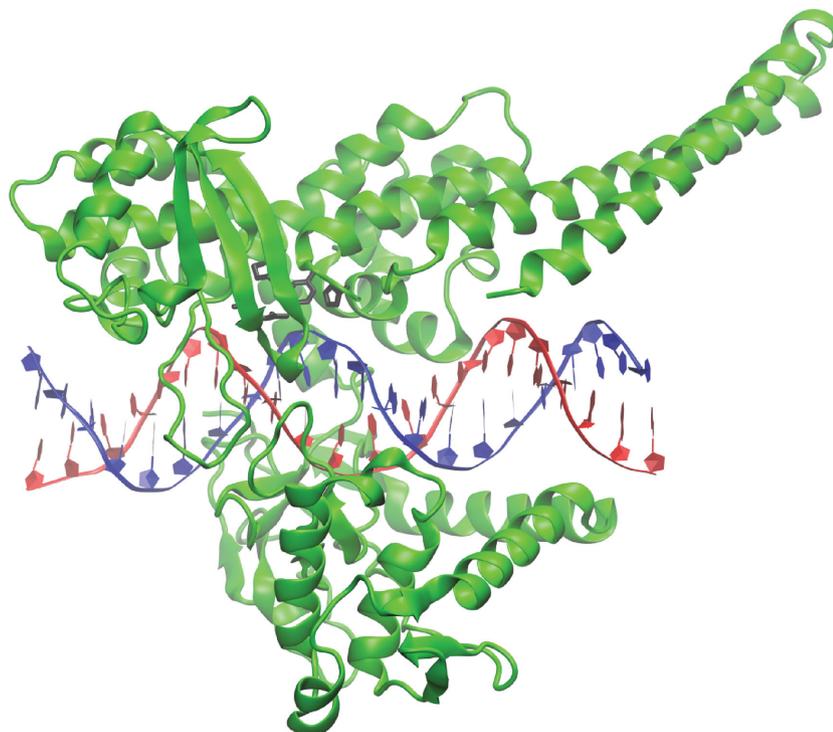


Figure 6. Structure of a type-I topoisomerase complex with DNA. PDB ID = 1A36.^[28] The protein is shown in cartoon format and colored green. DNA is shown as ribbons with the two strands colored red and blue. Water molecules are omitted for clarity. The Arg488, Arg590, His632 and Tyr723 sidechains that attack the ribose phosphate to cleave DNA are shown as black sticks. The coiled 77-residue linker at the far right of the figure is thought to interact with rotating DNA during topoisomerization and play a role in winding DNA.

a greater understanding of how two-legged dendrimer molecules move along patterned surfaces, in a kind of molecular hopscotch reminiscent of kinesin walkers and cellular chemotaxis.

The dendrimer molecules (Figure 8) have multiple legs, and display a surprisingly rich behavior at the surface, beyond simply attaching/detaching. Using atomic-resolution computer simulations (Figure 8) complementing fluorescence microscopy measurements and Monte Carlo and numerical models, the different mechanisms by which the molecules move could be shown. The motion switches from walking to hopping to flying, as the environment changes, with one computed “hopping” mode shown in Figure 8. In this biomimetic complex, a major obstacle for molecular-level materials design is the very different physics that operate at the scale of atoms and molecules. This is surmounted by restricting the motion to the most controlled “walking” regime simply by “sweetening” the solution with the right amount of sugar molecules. These results will allow better control of molecules in bottom-up nanofabrication, and could serve as test beds for development of more potent drugs that block the attachment of viruses to cells. We note that the application of more advanced electrochemically fabricated gradients for high-throughput deposition and device development has been recently reviewed.^[52]

3. Bacterial Breakdown of Crystalline Cellulose

Eons of microbial evolution have produced multi-enzyme cellulosome complexes that efficiently degrade plant cell wall polysaccharides, enabling simpler sugars to be produced from crystalline cellulose.^[18,20] The modular nature of these polysaccharide-degrading enzymes in cellulosomes was discovered in the 1980s, and their ongoing atomic characterization is crucial to understand their structure-function relationships.^[53,54] A large effort in comparative genomics was able to identify highly-conserved, functionally-indispensable modules and residues.^[55,56] Since the heterogeneity and flexibility of the full-length cellulases and cellulosomes are impediments to crystallization, small angle X-ray scattering (SAXS) measurements have helped to confirm the overall architecture of the cellulosome.^[57] Recent advances in instrumentation and automation (in particular, the routine use of synchrotron radiation, exploitation of anomalous diffraction methods in protein crystallography, and adaptation of beam lines to perform SAXS on proteins in solution) provide more sophisticated experimental techniques to study these challenging and intricate cellulolytic assemblies.^[9,58–61]

Despite these advances in genomics and structural characterization, the establishment of standard enzymatic assays on cellulosic substrates has been and continues to be a serious issue in assessing the catalytic abilities of cellulosome materials. There is a long history of research on cellulases and related enzymes (e.g., hemicellulases, pectinases, carbohydrate esterases, etc.)^[4,5,19,62,63] and a vast range of assays have been established (which are critical to their study).^[63,64] Yet, at the time of writing in mid-2015, there is no simple standard assay to monitor enzymatic activity during the degradation of crystalline cellulose and complex cellulosic substrates (e.g., natural cellulosic material such as wheat straw). Approaches to quantify enzymatic activity include the use of absorbing dyes that react with reducing carbohydrate chain ends, along with enzymatic reaction cascades (e.g., glucose-oxidase/horseradish peroxidase), or chromatographic analysis of saccharification products. More recently, spatially localized hydrogel polymerization was used as a method to quantify and image reactions at the interface between crystalline cellulose and aqueous enzyme solutions.^[65] Time-lapse atomic force microscopy (AFM) imaging has also been used to correlate substrate digestibility with topological features.^[66]

Optimization of cellulosomes requires the establishment of reference assays to quantify the efficiency of the different enzymes and their mutants.^[63] Several factors complicate the achievement of this goal:

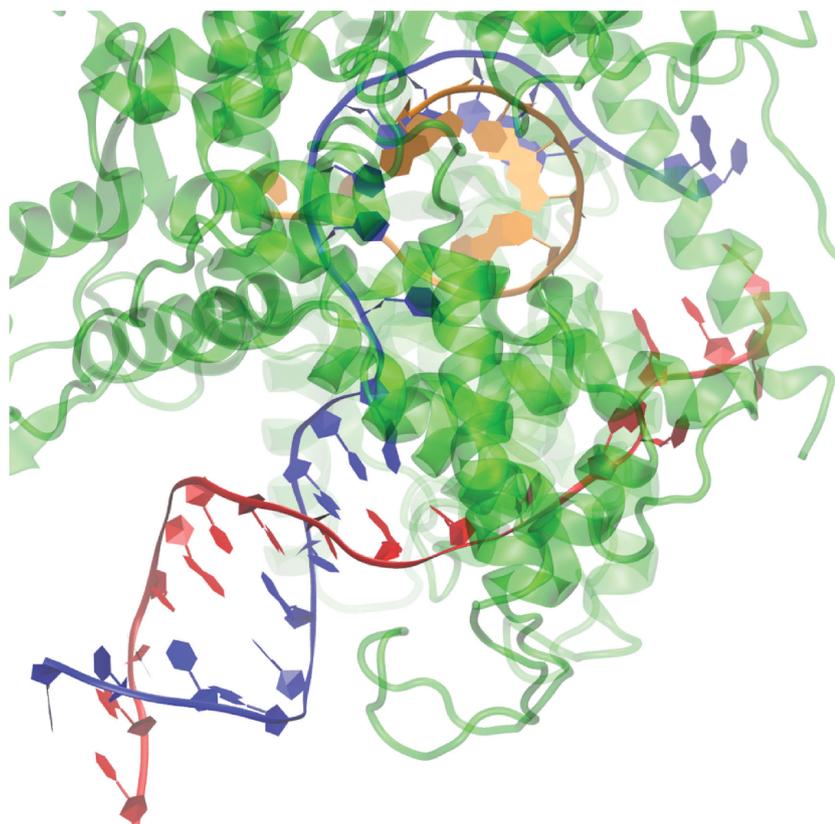


Figure 7. Crystal structure showing T7 RNA polymerase (green) producing an mRNA molecule (gold) from a DNA template (blue/red). PDB ID 1MSW.^[37]

- (i) the multiplicity of the different types and modes of action of cellulases and the wide variety of cellulose breakdown products (Figure 2),
- (ii) the contrast between the difficulty in the degradation of crystalline cellulose and the relative ease of degradation of non-crystalline substrates, and

- (iii) β -glucosidases, which degrade a wide variety of cellulose and cellobiose from the non-reducing ends and release β -D-glucose. Among β -glucosidases there is a large variety of β -1,4-glucosidases which specifically attack β -1,4 linkages. A particular case is glucan β -1,4-glucosidase which degrades 1,4- β -D-glucan but not cellobiose.

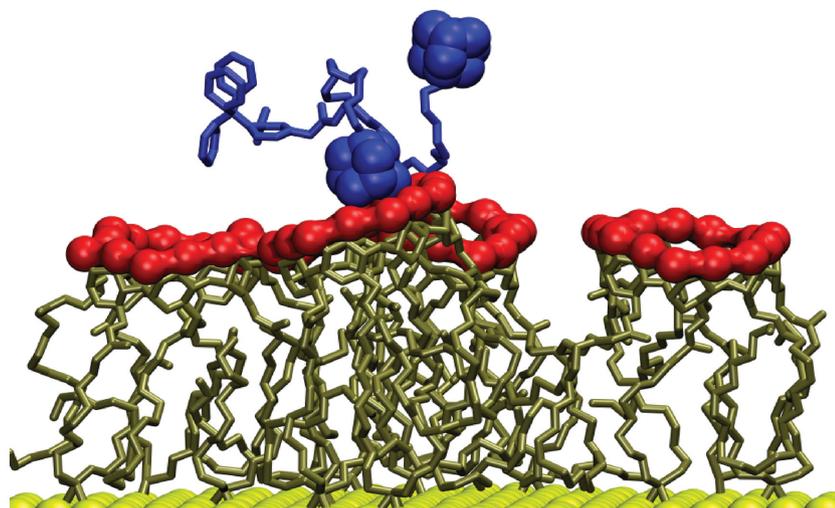


Figure 8. Molecular dynamics model of a multivalent dendrimer molecule (in blue) "hopping" on a β -cyclodextrin (in red) receptor-functionalized surface. The underlying gold-bound monolayer is colored green.^[51] Hydrogen atoms and solvent molecules are omitted for clarity.

- (iii) the wide variety of hemicellulases and their tremendous array of oligosaccharide products (Figure 2).

Thus, it is not surprising that currently there is no efficient high-throughput activity assay relating the enzymes and substrates. This constitutes a serious impediment to advancing research towards the realization of designer cellulosomes. Furthermore, the crystallinity of cellulose and its occurrence within the solid and compact cell wall of plants means that the catalytic reaction must take place in a heterogeneous (solid/liquid) phase.

Three types of hydrolytic enzymes (glycoside hydrolases) are commonly used to break down cellulose (Figure 2), namely:

- (i) endoglucanases, endo-acting enzymes which attack the cellulose by randomly breaking β -1,4 linkages. They prefer soluble amorphous forms of substrates and their affinity decreases with decreasing degree of polymerization. They are not active on short fragments (typically, the di-saccharide cellobiose). Endoglucanases generally exhibit open active site clefts;^[67]
- (ii) processive exo-acting enzymes or cellobiohydrolases which degrade single cellulose chains from the non-reducing end and processively release cellobiose. They prefer crystalline forms of substrate. They typically exhibit closed tunnels;^[68]

A recently discovered fourth category is the selective oxidative enzymes that cleave the chains in crystalline regions.^[69] The discovery of auxiliary modules (also known as accessory modules) such as the CBMs^[70] revealed the first rational explanations of how these enzymes are bound to and diffuse on the surface of the solid substrate. Several roles have been attributed to CBMs. These are:

- (i) bringing the catalytic domain in close proximity to the substrate,
- (ii) enhanced hydrolysis of insoluble substrates through polysaccharide structure disruption,
- (iii) feeding a single cellulose chain into the active site of an endoglucanase thereby converting the enzyme to a processive endoglucanase, and
- (iv) anchoring to cell surface proteins.

In parallel, the discovery of the macromolecular self-assembled machinery called cellulosome secreted by anaerobic bacteria^[71,72] revealed an extremely complex multi-modular arrangement for cellulose degradation. In this context, it is believed that the anaerobic environment induces a selective pressure for the evolution of enzymes which assemble in effective machineries for the extracellular degradation of polymeric substrates.^[73] These macromolecular assemblies present a serious challenge to biochemists to understand the synergistic interplay of all the components. They also defy structural biologists to paint a complete 3D picture^[74,75] of how the architectural arrangement of this large number of modules and proteins allows the efficiency of the ensemble to match^[76] and surpass (up to fifty times faster)^[77] that of extracellular cellulases.

4. Cellulosome Structure

4.1. The Daunting Variety of Components and Assemblies in Cellulosomes

Cellulosome was first identified in the anaerobic thermophilic cellulolytic bacterium *Clostridium thermocellum*.^[71,72] It is a multi-enzyme complex which was later found in other cellulolytic bacteria,^[18,78] including *C. cellulovorans*,^[79,80] *C. cellulolyticus*,^[81] *C. josui*,^[82] *C. acetobutylicum*,^[83] *C. papyrosolvans*,^[84] *C. clariflavum*,^[85,86] *Acetivibrio cellulolyticus*,^[87] *Bacteroides cellulosolvans*,^[88] *Ruminococcus albus*,^[89] and *R. flavofaciens*.^[90,91]

Cellulosomes are multi-enzyme complexes whose building block is the intermodular cohesin-dockerin pair (Figure 4). They are produced by numerous anaerobic bacteria to decompose cellulose into simple sugars. This results in a large variety of embedded enzymes in terms of size, number and architecture of constituent units in the cohesin-based scaffolds.^[21] As sophisticated enzyme arrays, cellulosomes involve a large range of biological units including cellulases, hemicellulases (e.g., xylanases, mannanases and arabinanases), pectine lyases, and carbohydrate esterases. These enzymes are then assembled on an intricate protein scaffold (scaffoldin). The CBM targets cellulosome to its cellulosic substrate. However, cellulosomes are not indiscriminate multi-enzyme nanocatalysts. Rather, they have evolved to generate enzymatic arrays of high structural and catalytic complexity to process the various lignocellulosic substrates. In the cellulosome, enzymes and structural subunits interconnect through specific non-chemical but high-affinity cohesin-dockerin interactions.^[92]

Thus, cellulosomes are complex machineries composed of numerous non-catalytic and catalytic subunits. These subunits are the various hierarchically organized scaffoldins, and a large variety of catalytically active proteins, mainly cellulases. Scaffoldins are very large proteins composed of a series of functional domains or modules, each with a distinct role. Typically cohesin domains are separated by linker segments (Figures 4 and 5). There exist various types of domains: cohesin, dockerin, CBM, X-module, and S-layer homology module (SLH) (Figure 4). For example, one very well characterized scaffoldin, namely CipA from *C. thermocellum*, contains nine type-I cohesin domains, a

single CBM module, and a terminal X-dockerin modular dyad (i.e., X-module and type-II dockerin).^[21] Cohesin (≈ 150 protein residues) and dockerin ($\approx 60\text{--}70$ residues) domains are of particular importance as their selective and complementary interaction is the basic building-block for the formation of the cellulosome complex. Indeed, cohesin domains are the main constituent of the non-hydrolytic scaffoldin protein which hosts the hydrolytic cellulase enzymes that are the actual catalytic units of the cellulosome (Figure 4). The cohesin-dockerin interaction is a very strong non-covalent protein receptor-ligand bond, with equilibrium $K_D \approx 10^{-9} - 10^{-10}$ M.^[93,94] Under non-equilibrium mechanical tension, certain cohesin-dockerin pairs can reach half of the mechanical rupture strength of a covalent bond.^[95] Cohesin-dockerin interactions from the thermophilic *C. thermocellum* are also very thermally stable in vivo. These receptor-ligand bonds are therefore primed for high stability under wide ranging and adverse conditions. There are three types of cohesins (I, II, III) and consequently three types of dockerin counterparts (Figure 9), but in fact this simple nomenclature is of limited use due to the broad sequence divergence of many recently discovered cohesin-dockerin pairs that do not fit into this classification scheme.

Type-I cohesin (CohI, panel I in Figure 9A) exhibits a jelly-roll topology that folds into a nine-stranded β -sandwich.^[96] Type-II cohesin (CohII, panel II in Figure 9A) exhibits the same nine-stranded β -sandwich but differs in few specific points. First, there is a short α -helix between two of the β -strands and two other strands are each disrupted by a " β -flap" that interrupts the structure. Type-III cohesin (CohIII, panel III in Figure 9A) has the core nine-stranded jellyroll cohesin topology with two type-II like β -flaps but also displays a unique N-terminal loop and dominant α -helix region.^[97-100]

Type-I dockerin (DocI, panel I in Figure 9B) is made of two 22-residue duplicated sequences, separated by a linker of 9-16 residues. Each sequence contains two well-conserved 12-residue loops which bind to calcium ions and two α -helices. The structure is a variation and subtype of the so-called "EF-hand" helix-loop-helix motif.^[101] Type-II dockerin (DocII, panel II in Figure 9B) is made from two loop-helix motifs, termed F-hand motifs, separated by a 14-residue linker region. Like in type-I dockerin, Ca^{2+} ions are bound to a well-conserved 12-residue loop.^[93] Type-III dockerin (DocIII, panel III in Figure 9B) diverges from the other two. It exhibits two F-hand motifs but the second motif misses the 12-residue sequence found in the Ca^{2+} -binding loop.^[99] It exhibits a total of five α -helices, unlike the other types which contain only three helices. Furthermore, the linker between the repeats is much longer as are the helices.^[98]

The X-module exhibits a β -stranded structure (Figure 9C). Its function is still not fully understood but it is often bound to a CBM (Figure 9 D) and recent studies suggest a protection of the cohesin-dockerin interface against mechanical stress.^[95] SLH motifs are found at the N-terminus of many S-layer proteins (the external constituent of many bacteria cell walls) and at the C-terminus of some exo-proteins of Gram-positive bacteria.^[102] Three types of SLH domains have been identified according to their origin in S-layer proteins, exo-proteins, and porins.^[103] Usually, proteins contain three repeats of SLH motifs (Figure 9 E), each consisting of 50 to 60 amino acids. SLH domain binds to secondary cell wall polymers.^[104]

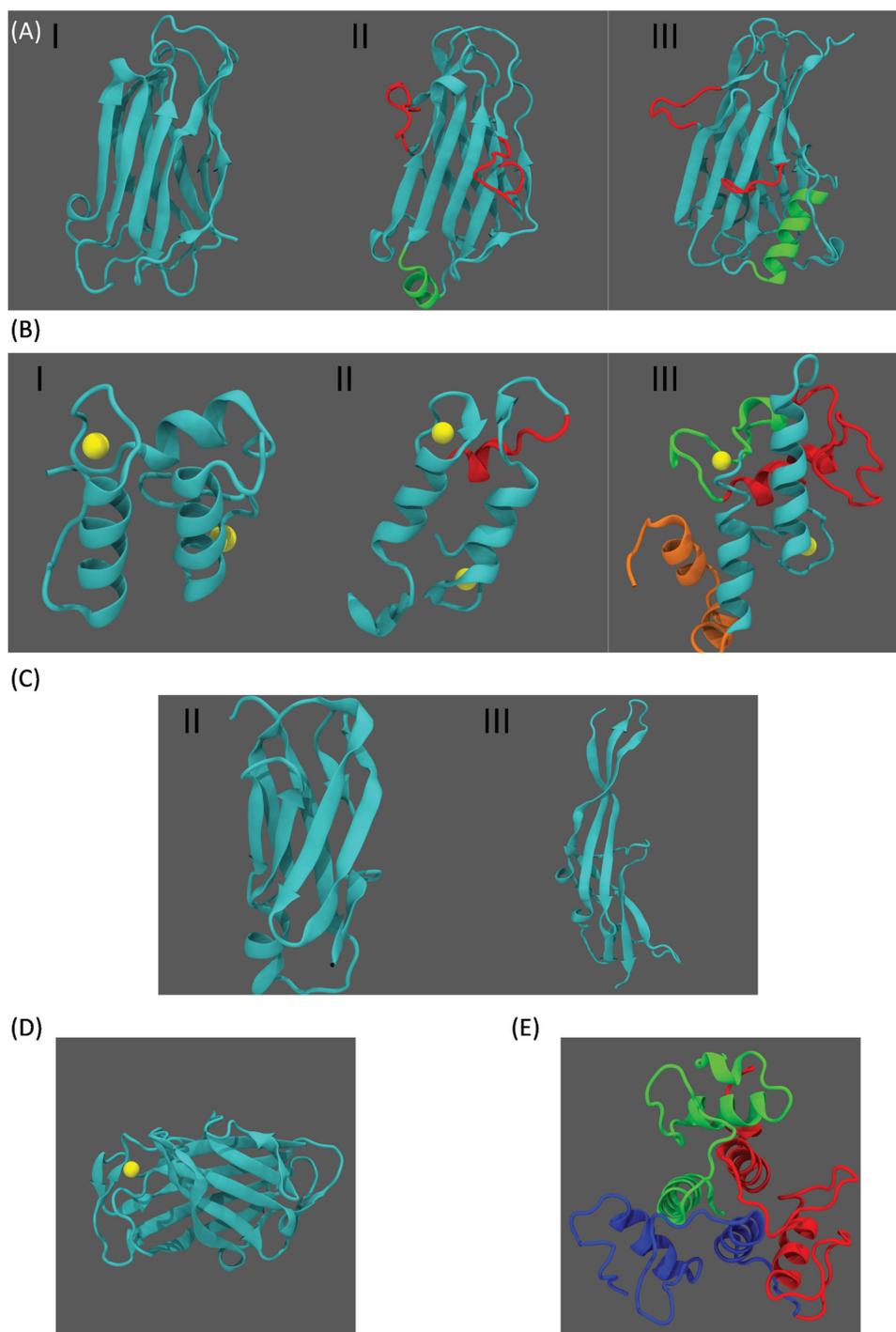


Figure 9. Classification and tertiary structures of cellulosomal components. A) shows type-I, II and III cohesins. CohII (PDB code 3KCP) and CohIII (2ZF9) are compared to CohI (3KCP) and differences are highlighted in red and green. The β -flaps on both CohII and CohIII are shown in red. The additional α -helix on CohII and on CohIII is shown in green. B) shows the corresponding dockerins. DocII (3KCP) and DocIII (4IU2) are compared to DocI (4FL4), and differences are highlighted in red, green, and orange. In red, the linker and α -helix between the two “EF-hand” helix-loop-helix structural domains found in many calcium-binding proteins are shown. The absence on DocIII of a well-ordered loop around the calcium ion is shown in green. Two additional α -helices found on DocIII are shown in orange. C) Two types of X-modules (3KCP, 4IU2). D) A carbohydrate-binding module (CBM) (1NBC). E) An SLH domain (3PYW). The three repeats of the same sequence are highlighted in blue, red and green.

All the scaffoldin domains and modules are attached one to the other by a large variety of linkers (Figure 5, and discussed in detail below). Despite the large number of moving

parts in the cellulosome, it is important to note that the various cohesins carried by scaffoldins do not differ significantly in terms of their recognition of the enzyme-borne dockerins.

This implies that the incorporation of the enzyme modules within the cellulosome framework is managed by the bacterial cell apparatus rather than by the sequence and nature of the cohesin units. The CipA scaffoldin (Figure 4) sequence is thus: 2 type-I cohesins, a CBM, 7 type-I cohesins, and an X-dockerin (type-II) module. This CipA module binds via its C-terminal X-dockerin module to the type-II cohesin of another scaffoldin, SdbA. The latter scaffoldin is strongly anchored to the cell wall through its C-terminal SLH.^[87,105–108] This cellulosome is the paradigm from *C. thermocellum* to which all other cellulosomes are compared. The detailed structures of the various scaffoldins found in the species listed above have already been thoroughly reviewed, along with genes coding them.^[9,18,88,91,97,109–111] We provide a brief overview here to emphasise the evolved specialization that tailors each cellulosome for its particular environment through small but key changes in the number, arrangement and molecular structure of the domains.

Recent studies indicate that not only is there a great diversity of cellulosomes associated with the various bacterial species but also that individual species can produce a range of different cellulosomes. Differences lie in the number and nature of modules involved in a given scaffoldin, and understanding their functional roles is daunting but can aid the development of blue-prints for optimization of synthetic nanostructured materials.^[112] Compounding the complexity, there are variations in each module as mentioned above (e.g., three types of cohesin and dockerin, Figure 9). For example, the primary scaffoldin of *C. thermocellum* involves nine type-I cohesins with a single type-II cohesin anchoring scaffoldin. In contrast, *B. cellulovorans* exhibits eleven type-II cohesins in the primary scaffoldin and ten type-I cohesins in the anchoring scaffoldin.^[88,109] CipA binds directly to an anchoring scaffoldin (SdbA), but ScaA from *A. cellulolyticus* binds to an intermediate adaptor scaffoldin (ScaB) made of four type-II cohesins which binds via a type-I dockerin to three specialized cohesins of the anchoring scaffoldin (ScaC).^[107,108] Intriguingly, the cellulosome system of *C. clariflavum* closely emulates the *A. cellulolyticus* system, wherein ScaA harbours eight type-I cohesins, ScaB bears five type-II cohesins and ScaC presents four specialized cohesins.^[86] Some primary scaffoldins (*C. cellulovorans*, *C. cellulolyticum*, *C. josui* and *C. acetobutyllum*) may not be attached to the bacterial cell wall and are free in the extra-cellular environment.^[20,80] The primary scaffoldin of *A. cellulolyticus* also differs from the CipA paradigm by an extra enzyme.^[87] Unlike CipA, the major *R. flavefaciens* scaffoldin (ScaB) does not contain any CBM^[113] and not only accommodates enzymes but also, via a type-III cohesin-dockerin interface,^[114] a smaller scaffoldin (ScaA), which itself bears both cellulases and at least one scaffoldin (ScaC).^[115] This makes the *R. flavefaciens* cellulosome complex one of the most elaborate cellulosomes found so far.^[91,110] Finally, we note also that at a higher organizational level, two models of gene arrangements were identified. In the first, genes coding for multiple scaffoldins form a cluster on the chromosome while enzyme genes are scattered in different locations over the chromosome.^[105,116] In the second, a single scaffoldin and multiple enzyme coding genes are found in sequence on the same chromosome.^[18,117]

Different cohesins and dockerins have been identified over the years. The type-I cohesin-dockerin pair (Figure 9) is the

building block of the primary scaffoldins with the cohesins making the framework on which the enzymes bind via the dockerin module. Primary scaffoldin ends in the so called “X-module” attached by a linker to the last type-I cohesin, e.g., the 9th cohesin in CipA of *C. thermocellum* (Figure 4). Another linker connects the X module with a type-II dockerin that binds to a type-II cohesin. Type-II cohesins make the skeleton of a second type of scaffoldin, the anchoring scaffoldin. The latter is attached to the cell wall and binds the primary scaffoldin through the type-II cohesin-dockerin interface to ensure the catalysts remain in the vicinity of the bacterial wall. The linkers attaching the different modules within the scaffoldin framework are very flexible, providing the scaffoldin the necessary plasticity that allows the synergetic work between all its constituents, but the cohesin-dockerin pairs are relatively rigid. The interface between the terminal part of a primary scaffoldin and the head of an anchoring one can therefore be studied by means of X-ray crystallography and SAXS.^[60]

The head of the primary scaffoldin, the N-terminal part, is also of much interest as the CBM module is found there, e.g., between Coh2 and Coh3 in CipA of *C. thermocellum*. The combination of experimental X-ray techniques with molecular modeling allows the characterization of crystals of larger sections of the scaffoldin. Indeed, if detailed knowledge of each component of the cellulosome is fundamental, understanding their collective behavior within the full architecture is essential.^[97] Thus, crystallization of the three first cohesins, including the CBM, and of Coh3-Coh4-Coh5 of CipA is a step forward in the comprehension of cooperativity, and possible synergy, between cellulosome components.^[59] The N-terminus is highly dynamic and can adopt both elongated and compact arrangements.^[59] This suggests that the two binding modes of type-I dockerin (DocI) have little effect on the enzyme functionality. However, the central section Coh3-Coh4-Coh5, although relatively flexible, exhibits preference for a compact arrangement with the enzymes on the outside.^[118] Here, the scaffoldin may benefit from the two binding modes of DocI as they make it possible to precisely orient the enzymes. A representative high-resolution X-ray structure of the interface between two scaffoldins is shown in Figure 10.

4.2. Linkers – Functional Actors or Passive Spectators in Cellulose Breakdown?

An integrated experimental/simulation co-design approach is required to fully understand the processes involved in the catalytic breakdown of lignocellulose in plant cell walls into simple sugars. Molecular-level techniques such as X-ray diffraction/scattering, atomic force microscopy, non-linear spectroscopy and single-molecule spectroscopy provide a wealth of information, particularly when allied with models for all-atom and coarse-grained molecular dynamics that can provide general design rules for rational re-engineering of cellulosome components.^[17,119] Detailed knowledge about the various constituents of the cellulosome is essential, in particular for the enzymes that bind to the scaffoldin. Characterization of these proteins by means of crystallisation and X-ray scattering provides the enzyme structure and architecture which helps in identifying

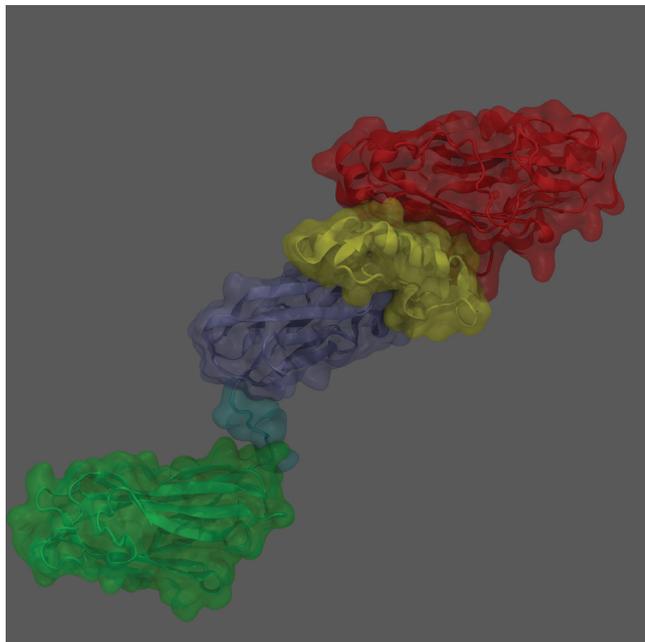


Figure 10. Volume isosurface representation of a Cohesin/X-Doc/Dockerin/Cohesion interface; PDB code 3KCP.^[75] Coh9 (green), linked to an X-module (indigo), and dockerin (yellow) belong to the CipA scaffoldin from *C. thermocellum*. The cohesin at the interface (red) belongs to the ScbA anchoring scaffoldin.

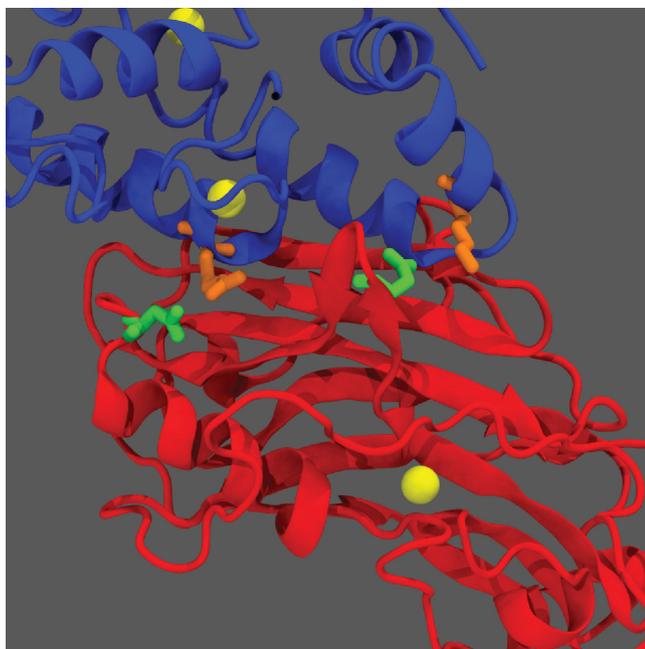


Figure 11. Interface between the Ctta X-Dockerin module (blue) and CohE (red) of ScaE of *R. flavefaciens*. The protein secondary structures are shown in cartoon representation with Asp153 and Asp165 sidechains of CohE shown as green sticks and Arg131 and Lys219 sidechains of X-Doc colored orange. The bound calcium ions are shown as yellow spheres. PDB access codes 4IU3 and 4IU2.^[98]

its biological role in the catalytic processes, and can provide physically realistic starting configurations for molecular modeling.^[120] SAXS experiments combined with molecular modeling simulations potentially allow the study and characterization of a whole cellulosome. Results to date underscore that the plasticity of the complex through its linkers is an essential feature for its catalytic activity.^[61]

One major, eminently re-engineerable source of variety then in cellulosomes is the difference in the nature of the linkers connecting the protein modules within a given scaffoldin subunit.^[9,22,59,121] Within the CipA scaffoldin of *C. thermocellum*, the linkers are found to have between 13 and 49 amino acid residues, with 24 residues being the average length.^[9,21,23] However, the linkers can be as short as 5 residues in the CbpA scaffoldin of *C. cellulovorans* or as long as 721 residues for the exceptional linker binding the last cohesin module with the SLH of an anchoring scaffoldin (OlpB) of *C. thermocellum*.^[9,18,20,21,77,91,107,113,122,123] This 721-residue linker is rather peculiar and may in fact be an as-yet-uncharacterised module as its sequence structure does not correspond to a typical “linker”. Yet despite the broad distribution of linker lengths, linkers of different length do not appear randomly in scaffoldins but are found at definite locations between the various domains. The type of amino acids involved in the linkers is diverse but with a greater-than-average incidence of proline and threonine residues. The non-random evolution of linkers is supported by the fact that they are highly glycosylated,^[71] with recent studies suggesting that the glycosylation occurs on threonine residues and to a lesser extent on serine residues.^[124] Indeed, glycosylation of peptides is a post-processing mechanism by which glycans (polymers made of different sugars) are attached to a protein. This is a complex mechanism that involves numerous enzymatic steps. Glycosylation is essential for some proteins to fold properly. For others it considerably increases their stability, though it does not affect folding, and may also facilitate cell adhesion. There are two main mechanisms:

- (i) N-linked glycosylation where glycans are covalently attached to a nitrogen atom in the side-chain of asparagine or arginine. This occurs in eukaryotes and widely in archaea but rarely in bacteria;
- (ii) O-linked glycosylation where glycans are covalently attached to the hydroxyl oxygen of serine, threonine, tyrosine, hydroxylated lysine, or hydroxylated proline side-chains. This type of glycosylation is found among eukaryotes, archaea and bacteria.

Besides the mechanism, the type of glycans is also characteristic of a given species. Scaffoldin linkers contain oligosaccharides including N-acetylglucosamine, galactopyranose and the rare saccharide, galactofuranose.^[124,125] The fact that linkers are O-glycosylated and are found, for a given length, at specific locations within the scaffoldin, supports the hypothesis of non-random evolution of these linkers and suggest an important, but as yet poorly understood, role for linkers in scaffoldin dynamics, substrate adhesion, and/or catalytic activity. It has been proposed that the diversity and the length of the inter-modular linkers helps maintain the flexibility of scaffoldins and prevent “jamming” of the cellulosome nanomachine even

with a large number of enzymes bonded to the cohesins.^[61,126] Indeed, maintaining scaffoldin flexibility has been shown to be crucial for designer cellulosomes (DC) to obtain a high catalytic activity.^[61,126]

Due to the many possible enzyme combinations and combinations of the other functional units (interlinked cohesin, dockerin, CBM, X-module, SLH, enzymes; Figures 4, 5 and 9) in the scaffoldin subunits, there is great potential to generate a rich structural and catalytic diversity. As a result, a single bacterium can contain a range of cellulosomal enzyme combinations, which may act collectively to access and degrade plant cell wall polysaccharides. The flexible architecture and non-covalent connectivity of the cellulosome potentially allows bacteria to rapidly adapt to changes in their substrate environment. Different cellulosome-producing bacteria have been identified in multiple environments at a wide range of temperatures between 30 °C and 60 °C.^[123] The understanding of cellulosome efficiency and synergy is still in its infancy, which constitutes the main barrier towards the industrial exploitation of its capabilities. Producing high levels of cellulosomal proteins in expression strains suitable for large scale industrial bio-refining is also a major hurdle.

We conclude this section on physicochemical/mechanical characterization of cellulosome by describing the remarkable cellulosome produced by *Ruminococcus flavefaciens*. It contains a large variety of scaffoldins; five of the scaffoldins are well-characterized and a larger number has been suggested.^[18] These scaffoldins range in size from a very short single cohesin anchoring scaffoldin ScaE and single cohesin ScaC and ScaD to the nine-cohesin ScaB that hosts both catalytically active enzymes and other short catalytic secondary scaffoldins ScaA. ScaE contains a third type of cohesin (CohE)^[114] that binds both the X-dockerin of ScaB and the X-dockerin of a cellulose binding protein called CttA. This CttA protein contains two CBM units and mediates a strong attachment of the bacteria to the substrate. X-ray scattering experiments provide a detailed view of the structure of the XDoc-CohE interface.^[98] The structure reveals an atypical calcium-binding loop containing a 13-residue insert on the dockerin. Results shows that two charged residues on CohE (Asp153 and Asp165, Figure 11) form specific electrostatic interactions with Lys219 and Arg131 of the dockerin module. Moreover, Isothermal Titration Calorimetry (ITC) and Differential Scanning Calorimetry (DSC) experiments provide the cohesin-dockerin complex dissociation constant: $K_d = 20.83$ nM. This interaction was also found to be among the most mechanically stable non-covalent interactions reported to date.^[95] ScaE is also found to bind not only scaffoldins but also an autonomous cohesin (CohG). It can attach to dockerin-bearing scaffoldin on one side while keeping its own dockerin module available for binding to ScaE. In this instance, CohG would serve as a shuttle to deliver scaffoldin to the bacterial wall.^[127]

4.3. The Role of Mechanostability in Nanoscale Function

Substantial advances have been made in recent years^[16,59,60,75,95,98,99,127–130] and models that are generally accepted (at least in gross features) exist for cellulosome structure and function, but the fine details of the molecular

mechanisms are still poorly understood. For example, even appended to a CBM, it is not clear how an endoglucanase would be capable of pulling a single polysaccharide chain out of its crystalline environment and forcing the chain productively into its active site cleft. So far, crystal structures^[131] and sequence alignments^[132] suggest that the shape and charge balance in the pocket induces a local distortion of a few units of a cellulose chain and it is postulated that this distortion may be essential for the catalysis to take place. Single-molecule techniques have permitted the study and manipulation of individual biomolecules, providing unprecedented insights into their function and dynamics. In particular, single-molecule force spectroscopy (SMFS, Figure 12), applied to proteins, allowed their mechanical stability to be studied,^[133] a property unrelated to thermodynamic stability but highly relevant to proteins that mediate adhesion between different elements, like the cellulosome.^[119] Depending on the goal of the experiment, cohesin and dockerin can be used as mechanostable “handles” to study unfolding of other domains of interest,^[134] or the mechanical properties of the cohesin-dockerin interaction can be studied directly.^[95,135]

As an adhesion system, cellulosomes might be expected to be subjected to mechanical stress.^[119] Hence, their mechanical properties may have been a key evolutionary constraint conditioning their architecture. Indeed, there are three key lines of evidence that strongly support the idea of scaffoldin being subjected to mechanical stress:^[119]

- (i) Detailed inspection of the scaffoldin architecture of different natural cellulosomes reveals the presence of “connecting” cohesin modules between anchoring points of the system (CBM and SLH modules (Figures 4 and 10) that mediate binding between the bacterial cell and its cellulosic substrate) and “hanging” modules outside the anchoring points (not expected to be subject to mechanical stress). Initial SMFS studies on cohesins revealed that modules located in the “connecting” region of scaffoldin are extremely mechanically stable, much more so than those in the hanging region;^[119]
- (ii) SMFS and Steer Molecular Dynamics simulations of the cohesin modules showed the existence of a mechanical resistance region, the so-called “mechanical clamp” motif,^[119,136]
- (iii) The anchoring points of the system, CBM-substrate and scaffoldin-cell complexes, have extremely high affinity binding constants with the cell anchoring being covalent in at least one case.

All these findings support the mechanical hypothesis of the cellulosome by which cellulosomes are expected to be very mechanically stable in order to maintain the structural integrity of their scaffold and retain the catalytic units docked to it.^[119]

The question remains as to the role of mechanical force in the cellulosome (and nanobiomaterials in general). Is it something that evolution designed around (perhaps in the presence of external environmental constraints that necessitated a stiff architecture), or can it constitute an essential and useful ingredient for materials design? To this end, it would be useful to understand the molecular mechanisms by which mechanical forces could contribute to the enzymatic activity of the complex. Mechanical forces can conceivably play a very important role

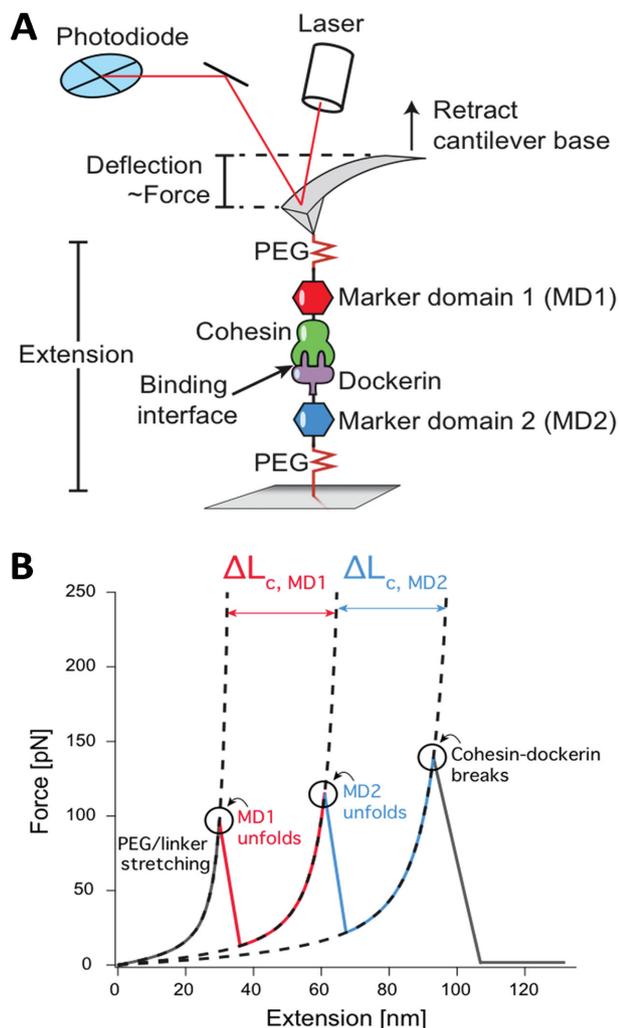


Figure 12. Schematics of AFM-based single-molecule force spectroscopy (SMFS) with mechanostable Cohesin-Dockerin handles. A) Cohesin and dockerin modules are expressed as fusions with marker domains 1 and 2 (MD1 & MD2), respectively. The marker domains provide site specific attachment points as well as contour length (L_c) increments of known length. The L_c increments from the marker domains are used as a filter to screen large AFM-SMFS data sets replete with unusable traces showing non-specific or multiple interactions. Observing both MD1 and MD2 length increments in a single trace ensures only behavior of single molecules is analyzed. The MD1-Cohesin and MD2-Dockerin fusion proteins are attached site-specifically to a silicon cantilever and a glass surface. Upon contact between the cantilever and glass surface, a specific and non-covalent bond is formed that is strong enough to withstand unfolding forces of the marker domains. Prudent choice of marker domains is crucial in this context. B) Force-extension traces from such an experiment show polyethylene glycol PEG- and protein-linker stretching, unfolding and stretching of MD1 and MD2 in sequence, followed by rupture of the Cohesin-dockerin complex. Dotted lines represent entropic spring models (e.g., worm-like chain, freely rotating chain) that are used to calculate the contour length changes.

in the cellulosome physiology; in the first instance the bacteria is a mesoscopic object (“microscopic” in biological jargon) and as such is not only subjected to Brownian motion, like molecules in solution, but will additionally feel longer-range hydrodynamic forces (e.g., in a turbulent cow stomach or

capillary flow gradients in moist soil).^[56] Under these conditions maximum holding strength and maximum catalytic efficiency seems to be needed. It is therefore not surprising that the mechanical strengths measured in cellulosome studies are among the highest found so far in biological systems.^[95,134,135]

5. Insights from Molecular Modeling Studies

Characterization of the atomic-scale structure and function of the cellulosome remains an intense area of research, one that continues to reveal interesting, unexpected and useful insights into nanostructured materials. We focus in this review on recent advances in atomic-scale understanding of cellulosome that feed directly into materials engineering. According to experimental and computational results to date on cellulosome subunits, it appears that cellulosome is a cellulose degrading factory, rather than an optimized machinery (from an engineering point of view) solely dedicated to cellulose breakdown. Indeed, all the proteins and modules involved in cellulosomes can be found in other biological systems and may not have been specifically developed, through evolution, for the cellulosome complex. Dockerin modules are found in a wide range of proteins, not all of which are enzymes. For example, bioinformatics screens have found serine protease inhibitors and other domains of unknown function containing homologous Doc sequences^[97] that presumably bind to scaffolds using strong Coh-Doc interactions. CBMs are found in many enzymes working alone on carbohydrate degradation, as found in various fungi. Following the factory analogy, each subunit appears as a worker hired from other plants or workshops to work synergistically on a specific task (cellulose degradation) along a production line (primary scaffolds). This suggests that alternative configurations, not found by nature, could make better materials for renewable energy, particularly those formed by mixing and matching components extracted from different species. In order to improve cellulosome efficiency and design artificial versions, it is essential to understand each subunit not only in the cellulosome but in other contexts as well. Computational techniques are suitable to such an exploratory task as they cover modeling of chemical reactions at the quantum level to models of the mesoscopic architectures using less detailed “coarse grained” models. These methods can generate a qualitative description of structures and dynamics at experimental scales and be used to study chemical reactivity (catalytic turnover rates).

Another aspect which requires attention and where molecular modeling can bring answers is the raw substrate material treated in the cellulosome factory, i.e., cellulose. Enzymatic hydrolysis of crystalline cellulose, even when cellulosome is involved, is a slow process because the polymer is insoluble and difficult to decrystallize.^[137] This insolubility is not obvious as glucose is a very soluble organic molecule. However, polysaccharides consisting of more than six glucose units are typically insoluble, and a cellulose chain usually contains between 2000 and 15 000 glucose units, making it highly insoluble in aqueous buffers. Furthermore, cellulose chains tend to pack together into fibrils ranging from the 36-chain elementary fibril to the 1200-chain macrofibril.^[138] Besides fibrils, cellulose can form other types of complex crystals. Seven allomorphs belonging to two

different groups identified by the polarity of the polymer have been proposed. The first group corresponds to naturally occurring allomorphs where the cellulose chains pack in a parallel fashion whereas the second group corresponds to artificial allomorphs with antiparallel chains. Among the first group, there are two main allomorphs, a triclinic form 1α which is easier to hydrolyse, and a monoclinic 1β which has a lower energy.^[139] Each crystal of these two allomorphs exhibits hydrophobic and hydrophilic surfaces due to all hydroxyl groups being equatorial. 1β form is dominant within higher plants while algae contain more 1α . However, both phases can be encountered within the same fibril. This shows the challenge for molecular modeling to accurately account for the diverse properties of cellulose, which is essential to fully understand the catalytic mechanisms used by cellulosome to degrade cellulose.

5.1. Coarse Grained and Mesoscale Modeling of Scaffolds

Due to the size of the cellulosomal subunits, coarse grained (CG) approached have been used to perform simulations of models of nine-cohesin scaffoldins ($\approx 35\,000$ atoms without solvent)^[140] as well as of 380-residue cohesin-dockerin complexes.^[141] This method provides a computationally less-expensive means of approximating meso scale properties by representing groups of atoms (from $-\text{CH}_3$ groups to entire molecules) as beads. This decreases the number of degrees of freedom and therefore the number of calculations needed. However microsecond all-atom simulations of complex systems involving millions of atoms are now feasible, if not yet routine, using parallel computing on high-performance computing (HPC) platforms. CG models are usually simpler than all-atom force fields, where the force field describes the potential energy landscape in which molecules sit and interact with each other through non-bonded van der Waals and electrostatic interactions.^[142] The solvent is generally described using an implicit model. Popular coarse grained force fields include the Go^[143] and MARTINI^[144] models, along with purpose-built variants.^[119,141]

The approximations described above lack precise atomic and femtosecond resolution but permit computationally feasible, qualitative meso scale simulations. In such CG simulations, energy components and forces are faster to estimate and larger integration steps can be used. Coarse-grained molecular dynamics (CG-MD) have been used to study mechanostability of cohesins^[119] by means of steered molecular dynamics methods^[145] that model protein-unfolding in atomic force microscopy/single-molecule force spectroscopy (AFM/SMFS) experiments. In particular, CG-MD predicted the mechanostability of a first set of 7500 proteins^[146] and subsequently one of 17 134 proteins,^[147] showing, after further improvements,^[119] that type-I cohesins in the primary scaffoldin of *C. thermocellum* are among the 30 most mechanostable proteins in the Protein Data Bank (PDB). Certain proteins (growth factors) with disulfide bonds and the cystine knot topology (a protein structural motif containing three disulfide bridges formed from pairs of cysteine molecules) have been predicted to yield mechanostability forces still larger than the cohesins.^[147,148] This method is then well-placed to make predictions and develop leads for site-directed mutagenesis experiments to re-engineer existing

cohesin modules towards improved thermostability of designer cellulosomes while preserving cohesin mechanostability.

Simulations are also used to validate new experimental strategies to perform SMFS (Figure 12), such as mechanically protecting the molecule of interest, the guest, by grafting it into a mechanically highly-stable host protein with so-called “single-molecule markers” flanking both sides of the host.^[149] These simulations can simplify the experimental design; the guest will be the last to unfold provided that it is more mechanostable than the markers.^[149] Furthermore, simulations can be used to test the validity and limitations of specific experimental designs like those used in biological unfoldases.^[150] On larger systems such as cohesin-dockerin complexes with a bound X-module, CG-MD provides information on the different unfolding pathways available to these complexes upon pulling and before the anisotropic final breaking of cohesin-dockerin interfacial bonds.^[141] In the studied complex from *Ruminococcus flavefaciens* (PDB code 4IU3, CohIII-X-DocII), unfolding usually involves many shear points within the cohesin and within the dockerin-X-module complex, with the exact unfolding pathway depending on the direction of pulling. It was shown that the X-module strengthens the dockerin mechanostability resulting in unfolding steps with maximum force of 300 pN for dockerin shear mechanism and of 440 pN for cohesin shear mechanism.^[151] By contrast, cohesin-dockerin rupture is a tensile-based mechanism^[151] with maximum force between 90 pN and 150 pN. On the other hand, an alternative structure from *A. cellulolyticus* (PDB code 2B59, CohII-X-DocII) shows a more isotropic behavior between the unfolding and cohesin-dockerin rupture stages. Both steps exhibit maximum rupture force between 100 pN and 200 pN which can be explained by a reduced number of contacts compared with the 4IU3 structure. The unfolding mechanism is either shear-based or tensile-based depending on the pulling direction. Another complex 1OHZ, which does not contain an X-module, showed force profiles similar to 4IU3 but with maximum forces between 100 and 200 pN, in one set of pulling directions. However, in the alternative set of pulling directions, profiles are much more similar to those of 2B59. This unfolding mechanism involves both shear and tensile ruptures but the cohesin-dockerin breaking is always tensile-based.^[141] These results illustrate the richness of mechanisms that underlie the experimentally measured maximum pulling forces, and provide leads for future mutation experiments.

Two binding modes for cohesin-dockerin complex have been identified in X-ray structure codes 1OHZ and a mutant 2CCL. They are related by a rotation of the dockerin by an angle of 180° . CG-MD docking simulations show that dockerin attaches to the cohesin binding site within 100 nanoseconds and that the full binding mode is achieved within 1 microsecond.^[152] This strong binding clamp involves a close packing of hydrophobic residues and H-bonds along two α -helices of the dockerin. The electrostatic pattern of the cohesin surface exhibits a negatively charged area related to the binding site and is complementary to a positively charged area on the dockerin surface.^[152]

Further coarse graining that combines elements of native-based topology, Go-like models and off-lattice protein simulations can be used to model entire scaffoldins and study cellulosome assembly through enzyme binding.^[140] Beads in

such a model encompass up to hundreds of residues, making these models very approximate. The simulations show that the stretched scaffoldin CipA (Figure 4) from *C. thermocellum* tends to adopt a more compact structure, as proposed from electron microscopy data.^[153] Regarding the enzyme binding to the scaffoldin, the main driving force is the specific protein architecture while the mass and volume of the enzyme are less important. High modularity and flexibility of the enzyme is a key feature for high binding affinity with scaffoldins; for example, multimodular enzymes where different functional domains are attached to the core of the enzyme by linkers provide great flexibility to the enzyme and facilitate access to the binding sites. Large enzymes diffuse more slowly than smaller proteins, which increases the residence time of larger secreted enzymes around the cohesins. Furthermore, the high flexibility introduced by inter-module linkers facilitates cohesin-dockerin recognition as the dockerin module is able to sample a large number of conformations and approach configurations with minimal rotation of the overall enzyme-dockerin-cohesion-scaffoldin complex.

Finally, it is worth mentioning in this section alternative techniques that are also useful and complementary to coarse grained models: namely, homology modeling, structure prediction methods based on scoring functions like Rosetta,^[154] and regression methods such as Quantitative Structure-Activity Relationship (QSAR). These techniques are particularly useful to provide protein structure from their sequence when no X-ray or nuclear magnetic resonance (NMR) data is available. Indeed, discovery of new proteins is much faster and generally less expensive than their characterization which can take months. They also provide an efficient way to estimate the effects of a mutation on a set of properties. This could be useful to identify a small set of mutations to test using more rigorous simulations and then follow up with experiments on these cellulose subunits. For example, Rosetta can be used to estimate the contribution of different residues to the overall binding free energy of a cohesin-dockerin complex and help design a mutant with higher specificity.^[128]

5.2. Coarse Grained and Mesoscale Modeling of Carbohydrate Binding Module and Cellulose

The CBM plays a crucial role in adhesion of cellulosomes to cellulose substrates. Two copies of this module are also found in the CttA protein from *R. flavefaciens* and CBM binds the entire bacteria to the plant cell wall. As mentioned, CBMs are suggested to play three main roles: proximity effects,^[155] substrate targeting,^[156,157] and microcrystallite disruption.^[158] Indeed, CBMs keep the enzymes embedded within the scaffoldin close to the cellulose surface which promotes enzyme-cellulose associations and increases enzyme local concentration. Furthermore, these modules have selective affinities for the different forms of cellulose. Finally, some CBMs disturb the local structure of the cellulose to make it more sensitive to enzymatic action. modeling the attachment mechanism of different CBMs and their dynamics at the substrate surface is then crucial to prepare efficient designer cellulosomes in the future.

Interconversion between cellulose fibril allomorphs was studied with a CG model including additional inter-chain interactions to mimic H-bonds.^[159] Although interconversion of single chains is feasible with such a model, it lacks the ability to reproduce the overall bending mode that drives the fibril transition. On the other hand, models with larger number of beads and using point-charge electrostatics have been developed.^[160] A systematic comparison^[161] between 1, 2, 3 and 4 bead CG models shows that the 1 and 2 bead strategy is best suited for microsecond simulations required to observe crystalline to amorphous transition of cellulose. Both perform equally well regarding scanning of inter-chain separation in conformational space (RDFs). The 2-bead models allow the glucose residue to rotate around the 1,4 glycosidic bond which is crucial for calculating properties where free rotation and orientation of the glucose play a role. This is confirmed by comparison between the 3 and 4 bead models, which both describe well the intra-chain dynamics and conformational change in glucose units, but the 4-bead, which allows free rotation of the glucose unit around the 1,4 glycosidic bond, samples fewer unphysical states than the 3-bead model.

Both all-atom force field (CHARMM) and CG models (Go) have been used to model the structure and dynamics of CBM units.^[162] (Recent unpublished work from our group on all-atom forced unfolding of CBM is shown for illustrative purposes in Figure 13). Upon adhesion to the cellulose surface, the CBM undergoes a conformational change with a tyrosine residue moving from its internal location to form a contact with the cellulose. This is observed in both all-atom and CG MD simulations. Subsequently, three tyrosine residues are found at the interface with the cellulose, each facing a different (but neighboring) glucose unit. Conformational change in CBMs is thought to play a crucial role in cellulose recognition, according to results for CBM units modeled to date. On perfectly crystalline cellulose, CBM displacement is due to random diffusion. However, when accidents occur like broken cellulose chains, CBMs migrate away from the reducing end or the non-reducing end by successive jumps, depending on the catalytic mechanism involved in their attached enzyme. For example, CBM1 from cellobiohydrolase I (CBH I) recognizes the reducing end and moves away whereas CBM2 from CBH II recognizes the non-reducing end. This motion is relatively fast: about 0.2 nm per ns.^[162]

Both cellulose and scaffoldin units (cohesin, dockerin, catalytic peptides, etc.) have been extensively studied by atomistic modeling, both to provide parameters for more coarse-grained meso scale models and to directly provide an accurate description of the phenomena at play in cellulose breakdown by cellulosome, as described in the next section. The atomistic models involve quantum mechanical (QM) calculations of electronic structures and chemical reactions, classical molecular dynamics (MD) simulations of solvated macromolecules and large-scale van der Waals and electrostatic interfaces, and combinations of quantum and molecular mechanics (QM/MM). We point the interested reader to selected recent studies of atomistic simulations of other evolved machines for degradation of macromolecules besides cellulose-binding cellulosome, including machines to breakdown DNA (restriction modification of endonucleases),^[163] RNA (exosome),^[164] protein (proteasome)^[165]

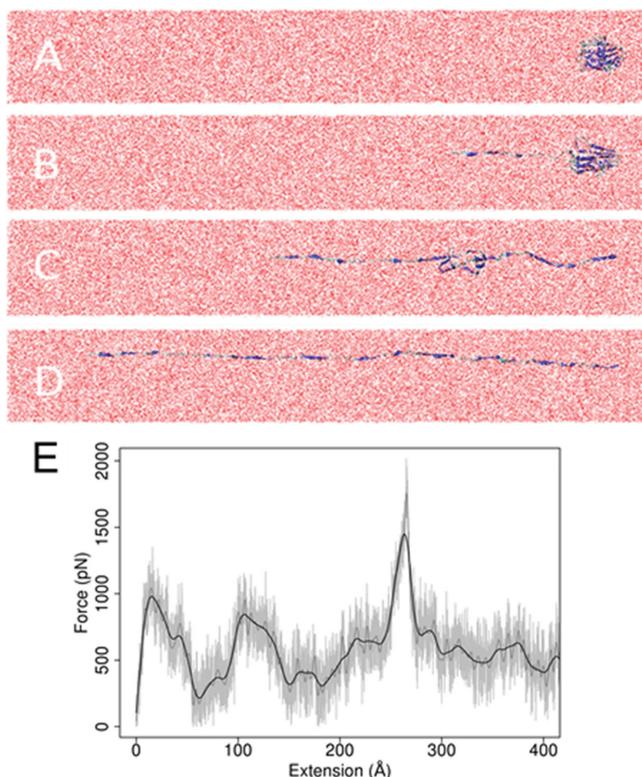


Figure 13. Representative MD structural snapshots showing (A–D) the forced unravelling and (E) corresponding computed force profile obtained in modeling the mechanostability of cellulosome component the carbohydrate binding module family 3 (CBM with PDB code 1NBC), in water (unpublished data); the results of these simulations can be compared with and used to guide AFM-SMFS experiments (Figure 12). The N terminal of the protein is pulled harmonically at constant velocity of 0.05 \AA ps^{-1} in the simulations with the C terminal kept fixed. Snapshots correspond to extended lengths of A) 0 \AA , B) $\approx 74 \text{ \AA}$, C) 350 \AA , D) $\approx 560 \text{ \AA}$, showing one of the pathways of protein unfolding, and E) the computed force-extension profile.

together with polysaccharides other than cellulose, e.g. chitin (chitinase),^[166] mannan (mannanase).^[167]

5.3. Atomistic Simulations using Quantum Mechanical Methods

Chemical reactions are complex dynamical processes which involve the formation and rupture of covalent bonds via electron transfer, a quantum phenomenon best modeled using so-called “ab initio” first principles methods.

Ab initio molecular dynamics is then the leading technique for studying chemical reactions. Density functional theory (DFT) is the preferred QM method for these simulations, particularly when combined with the efficient Car-Parrinello algorithm (CPMD).^[168] This approach can treat hundreds of atoms and timescales of $\approx 100 \text{ ps}$ with reasonable computational time on parallel computing clusters. For larger systems and more complex energy landscapes, many of the accelerated sampling techniques developed for classical molecular dynamics can be transferred and exploited in ab initio MD, such as umbrella sampling^[169] and metadynamics.^[170] Together these techniques

can be used to elucidate the mechanisms involved in the experimental pre-treatment prior to any enzymatic hydrolysis of cellulose and then fermentation of the simple sugars into ethanol. Particularly for potential industrial applications, the preparation of the biomass is a crucial step and the most costly one. Use of sulfuric acid at relatively high temperature is one of the most efficient approaches.^[4] However it requires a strict control of the thermodynamics and pH conditions to prevent further degradation of glucose and xylose which decrease the conversion rate. In both gas phase and in bulk water, the two step dissociation of xylose (protonation of the ether oxygen and CO bond breaking) involves a single transition state which takes place during the protonation step.^[171] However, it is shown from ab initio molecular dynamics modeling^[172] that water plays a crucial role during this chemical reaction as the computed free energy barrier is at least two times larger in bulk water than in gas phase. The dissociation is not favorable from both a kinetic and thermodynamic point of view. This confirms the necessity of an enzymatic catalysis for the hydrolysis to take place.

For larger systems of more than a few hundred atoms, QM/MM models can be useful with a QM region, where the reaction takes place, surrounded by a larger MM region accounting for the environment. QM/MM studies of the catalysis of glycosylation and deglycosylation of cellulose by the cellulase enzymes Cel2A and Cel5A provide the free energy profiles with identification of an intermediate state and a transition state.^[173] The latter resembles an oxocarbenium ion.^[174,175] The thermophilic enzyme TmCel12A exhibits free energy barriers for glycosylation and deglycosylation of 22.5 and $24.5 \text{ kcal mol}^{-1}$, respectively.^[174] The second barrier is found to decrease with temperature. At $85 \text{ }^\circ\text{C}$, a very stable hydrogen-bonded network of three glutamates and an ordered active-site water molecule is shown to hold the cellulose in a favorable, reaction-ready orientation even at this elevated temperature. Similar H-bonding networks appear to play a key functional role in maintaining the reactive site of the cellulases.^[173,174,176] Similarly, the study of the cellulose glycosidic bond cleavage provides a free energy barrier of 19 kcal mol^{-1} , showing that the enzyme is responsible for the acid and base action through Glu87 and Asp255 respectively. The location of a highly-ordered active-site water molecule which is responsible for the nucleophilic attack on cellulose is regularly observed,^[174,176,177] and simulations underline the crucial role of conformational changes of cellulose during the glycosylation and deglycosylation reactions.^[173,174,176,177]

5.4. Atomistic Simulations using Classical Molecular Dynamics

The classical simulation techniques provide a means of extensively sampling the potential energy surface (PES) of macromolecules and their interaction complexes. Structures are generated using Monte Carlo (MC) or Molecular Dynamics (MD) methods. For MC, a new configuration is generated and either kept or rejected based on the Metropolis sampling algorithm; when many configurations are generated, properly Boltzmann-weighted averages can be obtained for structural and thermodynamic properties. MD relies on Newton's second law of motion to determine new atomic positions and velocities at each time step. In both cases, the total energy (MC) and

forces (MD) are generated using a potential energy surface (also known as a force field, FF) that determines the quality of the simulation. Force field parameters are generally determined from quantum mechanical calculations, scaled where necessary against known experimental thermodynamics quantities, e.g. heats of solvation.

In the same manner as for proteins, standard force fields for sugars have been developed over the last decade like CHARMM,^[178] GROMOS,^[179] and GLYCAM.^[180] Due to the variety of properties these force fields are required to account for, none is able to simultaneously describe small sugars and large cellulose aggregates with a high degree of accuracy. Comparisons^[181] show that CHARMM and GLYCAM are among the most useful. Internal structure is relatively well described by these two force fields, particularly by CHARMM, using which the structure converges ten times faster than with GLYCAM and provides a closer match to experimental lattice parameters. These force fields are suitable to be used in the study of hydrolyase activity at the cellulose surface.^[182] Possible calculations involve the various binding and solvation free energy methods such as Poisson Boltzmann calculations of electrostatic binding free energies and alanine scanning.^[183] Results emphasize the importance of active site hydrophobic residues in the adhesion of enzymes on the cellulose surface.^[182]

Due to the size of scaffoldins, modeling of cellulosomes has been mainly performed using coarse grained (CG) models.^[140] However, attachment of scaffoldins and enzymes on cellulosome is mediated by a dedicated module, the carbohydrate binding module CBM. The latter is of a moderate size and exhibits a recognition mechanism towards the hydrophobic surface of cellulose due to specific interactions. All-atom molecular dynamics is then the tool of choice to study such a system. MD simulations confirm results already found by means of the CG and QM/MM simulations while bringing further insight into the clamping mechanism. MD results indicate that adhesion of the CBM is directed by aromatic residues among which a series of tyrosine sidechains play a crucial role,^[184,185] together with a few polar residues (asparagine and glutamine) that form H-bonds to cellulose.^[186] More precisely, the tyrosine network induces a distortion on the cellulose chain which is instrumental in guiding the polysaccharide towards the charged cleft^[187] of four aspartic acid groups that can coordinate the carboxyl and hydroxyl groups and the glycosidic oxygens. The structure of the cleft confirms that CBMs can only bind a single chain of at least four subunits.^[187] Furthermore, energy minima are found for CBM conformations corresponding to the length of the degradation unit. The adsorption on the hydrophobic surface of cellulose is enthalpically driven; favorable van der Waals's interactions between tyrosine residues and glucose units initiate binding with the electrostatic binding in the cleft dominating the overall binding energetics.^[185]

Simulations have revealed an intriguing mechanism of molecular motion whereby CBM-enzyme complexes diffuse on cellulose by successive leaps.^[184] CBM is found on two preferential sites on the cellulose surface with (anti-)parallel binding depending on the enzyme terminus of attachment (C- or N-terminal).^[188] This also affects the catalytic mechanism of attack on the cellulose chain. Some cellulosomal enzymes, despite the CBM module present on the scaffoldin, also contain a similar

module that shares most of the features as in free enzymes. In this context, we note finally here that molecular simulations have revealed that some cellobiohydrolase A (CbhA) complexes with cellobiose have unusually-extended binding pockets in which tryptophan residues on loops at the vicinity of the binding site help bind non-crystalline cellodextrin of five or more subunits.^[189] Such functional conformational shifts are becoming more apparent as a key enabling design principle for nanobiomaterials, and molecular dynamics simulations can reveal information on minor states and relative populations and roles of open/intermediate/closed states that are not easily accessible from crystal structures alone.^[190]

Molecular recognition of substrate by the enzyme binding site is key to specificity. Nevertheless it is also necessary to process molecules in and out of the binding pocket efficiently. To this end, some cellulose degrading enzymes have evolved transport mechanisms to harness Brownian motion and ensure a steady flow of reactant molecules into, and products away from, the active site. For example, Cel7A, a cellobiohydrolase found in *T. reesei*, contains a tunnel out of its binding pocket which removes the two main products cellobiose and glucose and frees the enzyme to degrade the next section of cellulose. Experimental results show that lingering cellobiose, but not glucose, has an inhibition effect on cellobiohydrolase.^[191] Binding free energies obtained by means of steered MD simulations and alchemical free energy calculations are 14.4 kcal mol⁻¹ and 10.9 kcal mol⁻¹ for cellobiose and glucose, respectively,^[192] and this 3.5 kcal mol⁻¹ difference may explain why cellobiose is found to be an inhibitor in experiments while glucose is not. The simulations showed that cellobiose is stabilized in the binding pocket by five residues including three charged groups, two negatively-charged aspartic acids and one positively-charged arginine. Single mutations of each of these five residues led to a sizable decrease of the binding free energy, down to 6 kcal mol⁻¹ for one of the aspartic acid residues. This shows that there is room for improvement in terms of catalytic efficiency of the enzymes involved in cellulosomes by, for example, increasing the speed of the diffusion of the reactants and products in and out of the reactive site. To this end, mutation of residues in the second shell beyond the immediate ligand binding pocket may provide a means of altering the charge balance (and so ligand binding specificity) in the binding pocket without significantly decreasing overall binding affinity and turnover rates in the (bio)catalyst.

The key "glue" in cellulosome, cohesin-dockerin binding, relies on an extensive H-bonding network and hydrophobic interactions. Upon binding and unbinding, cohesin and dockerin undergo conformational changes. A single mutation of an interfacial aspartic acid into an asparagine results in an increased flexibility of the cohesin and conformational changes around the loops.^[193] These structural modifications reduce the magnitude of the binding free energy by about 5 kcal mol⁻¹ relative to the -17 kcal mol⁻¹ for the native complex. Combined AFM experiments and SMD simulations on the X-module-dockerin-cohesin complex that binds CttA to the bacterial wall of *R. flavefaciens* show one of the most mechanostable protein-protein interfaces with a rupture force of 600–750 pN at the experimental loading rate used. CttA is the two-CBM module that mediates bacterial adhesion on cellulose.^[95] The amplitude

of the force peak indicates a complexation energy that is at least half the strength of a covalent bond. MD simulations show this strong interaction is driven by a core of hydrophobic residues surrounded and protected by hydrophilic (polar and charged) residues on both cohesin and dockerin. The contact area is increased upon pulling which increases cohesin-dockerin binding resulting in a very high mechanostability by so-called “catch bonds”. Further simulations where pulling is focused on the X-module show that it has an unfolding strength similar to and slightly higher than the cohesin-dockerin interface, which is supported by experimental results and suggests a force-shielding role of the X-module. We note that *R. flavefaciens* also exhibits an unusual dockerin module which binds scaffoldin A on the primary scaffoldin B. Experiments and simulations have probed the details of the cohesion-dockerin interface in various species including very recently in the cellulose-degrading bacterium *R. flavefaciens*; a particular valine-alanine motif was identified as important but not the sole director of interface formation, with mutation studies showing that the interface can in some cases compensate and form alternative, near iso-energetic interfaces^[16] Similar results were found for the type-II cohesin-dockerin complex of *C. thermocellum*, and highlighted again the large effect the X-module can have on the cohesin-dockerin interaction^[194]

Multimodular enzymes, enzymes which embed various functional domains connected by linkers, are responsible for a majority of cellulose hydrolysis in nature,^[4] in particular cellobiohydrolases TrCel6A and TrCel7A from *T. reesei*. Linkers in such enzymes are highly glycosylated,^[195] e.g., in Cel6A and Cel7A, linkers (41 and 27 residues, respectively) connect the CBM to the rest of the protein which are mainly O-glycosylated with also some N-linked glycans (Cel6). MD simulations using codes such as NAMD^[196] and CHARMM,^[197] and subsequent experiments revealed that the linkers do not only serve to tether the modules but also assist the CBM adhesion on the cellulose surface. The simulations predict that the glycans interact favorably with the hydrophobic surface of cellulose, inducing conformation changes in the linkers around the CBM, and, crucially, the peptide chains also participate in substrate adhesion. Experiments confirm these findings with an adhesion rate an order of magnitude larger when glycosylated linkers are involved: $K_a = 0.8 \times 10^{-6} \text{ M}^{-1}$, against $K_a = 0.08 \times 10^{-6} \text{ M}^{-1}$ without linkers.^[198]

Glycoside Hydrolases (GH) are composed of cellobiohydrolases and endoglucanases and they hydrolyse cellulose via two mechanisms, namely “inverting” and “retaining” mechanisms. The latter has been recently elucidated in the case of cellobiohydrolases by means of classical and QM/MM MD simulations, combined with umbrella sampling and transition path sampling techniques.^[199,200] The simulations elucidated the processive two-step catalysis starting at the end of a cellulose chain. The first step involves glycosylation of the cellulose chain by proton transfer via a glutamate residue (Glu217). Simultaneously, the nearby nucleophilic residue (Glu212) attacks the anomeric carbon to form a bond and consequently the glycosyl-enzyme intermediate (GEI) where a cellobiose unit is attached to the enzyme. The second step –deglycosylation – is promoted by a conformational change of the intermediate which allows the approach of a water molecule which attacks the anomeric carbon. This attack breaks the bond between the enzyme and

the cellobiose and release a proton which is transferred towards the acid/base residue, thus regenerating the reactive site. Beside these two (de)glycosylation steps, this catalysis involves many other steps from uptake of the cellulose chain through an aromatic tunnel of tryptophan residues to the processive translocation of the enzyme on the cellulose chain. Other steps include the conformational change of a glucose ring from the non-catalytically active armchair configuration — anomeric carbon sits $\approx 7 \text{ \AA}$ from the nucleophile — into a twisted “envelope” or “half-chair” configuration, which primes the glycosidic bond for catalysis, and cellobiose release. The twisted conformation of the sugar ring forms a Michaelis complex with three residues of the enzyme (Asp173, Glu212 and Glu217), two of which take part in the subsequent glycosylation step. The nucleophile Glu212 is stabilized by forming H-bonds with two other residues: Ser174 and Asp214. Asp 259 plays a crucial role in the processive step and in cellobiose release as it forms with the latter a long lasting H-bond.^[201] Simulations quantify the thermodynamics of the entire process with free energy barriers associated with each step, showing that neither the activation step ($\Delta G = 2.9 \text{ kcal mol}^{-1}$) nor cellobiose release ($\Delta G = 11.8 \text{ kcal mol}^{-1}$) are the limiting step but that glycosylation is ($\Delta G = 15.5 \text{ kcal mol}^{-1}$).^[200] Such insights are key to speeding up rational design of nanobiomaterials; structures and reaction schemes which look good on paper are often misleading, meaning rigorous, generally multi-scale, modeling is required to accelerate progress in materials discovery, optimization and re-engineering.^[202]

6. Progress Towards Engineered Designer Cellulosome Nanocatalysts

6.1. A new Paradigm for Realization of Multi-Component Functional Architectures

Designer cellulosome (DC) is a term used to describe artificial cellulosome complexes with defined arrangement. The diversity and modularity of natural cellulosomes have served as a base towards the idea of engineering these designer enzymes^[203] for an enhanced degradation of recalcitrant cellulosic substrates.^[204,205] DCs are composed of a chimeric scaffoldin (Figure 14), designed to contain multiple copies of cohesins of different specificities. The exact positions of each of these enzymes can be pre-programmed by incorporating complementary dockerin-bearing enzymes into the complex. As cohesin-dockerin recognition appears to be relatively non-specific in their native state, a superior control over the organization of desired components in the cellulosome complex is possible with chimeric scaffoldin.^[19] Thus, specific combinations of enzymes can be integrated at precise positions in the complex, producing homogeneous preparations of these “nanocatalysts” with cellulosomal and extra-cellulosomal combinations.

In order for the field of nanocatalysis to advance beyond the trial-and-error of classical catalysis, chemical reactions need to be controlled by rationally modifying the size, shape, chemical composition and morphology of the catalyst to achieve the desired substrate discrimination capacity and turnover rates. This approach aims to drive dynamics towards the desired reaction

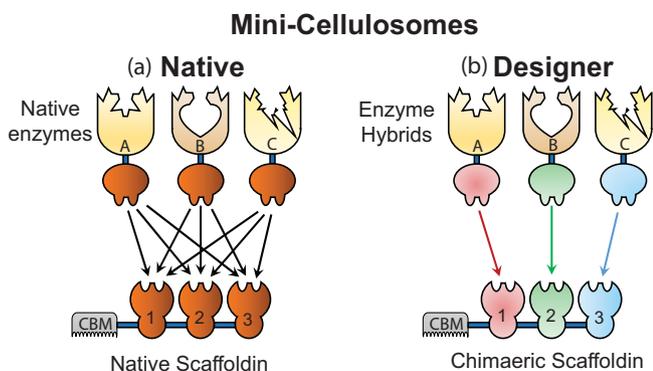


Figure 14. Native vs designer cellulosome architectures. The left hand panel shows the uniform specificity between cohesins and dockerins in the native cellulosome. By contrast, chimeric scaffoldin and hybrid enzymes (right hand panel) allow for selective interaction.

product, which opens up new avenues for atom-by-atom design of nanocatalysts with distinct and tunable chemical activity, specificity and selectivity.^[203] However, this discipline is highly focused on inorganic catalysts (typically expensive and pollutant rare-earths or precious metals) and still relies to some extent on brute force, trial-and-error methods for discovery and improvement of useful catalyst compositions.^[206]

By contrast, enzymes are readymade nanomachines refined by millions of years of natural selection, such that some of them are considered “catalytically perfect”, with k_{cat}/K_M specificity constants approaching the 10^8 – 10^9 $\text{M}^{-1}\text{s}^{-1}$ diffusion limit. Thus, enzymes provide efficient nanocatalysis and serve as the ideal starting point from which to set out on rational design for industrial applications. There are several advantages in using enzymes as nanocatalysts in industry. First, they are ideal catalysts with extremely high activity, selectivity and specificity. Second, the control of the particle size is extremely precise and reproducible, resulting in a very homogeneous population of nanocatalysts. Achieving homogeneity in particle size is one of the challenges in inorganic nanocatalysis. Third, they can be produced and re-engineered very precisely, easily and cheaply using established biotechnological processes (including site-directed mutagenesis, protein engineering and directed evolution) to make them more suitable for industrial applications. Fourth, they can be easily deactivated (by denaturation, cross-linking or hydrolysis) and are biodegradable, which further reduces their already minor environmental impact and makes them extremely safe.^[207]

The concept of designer cellulosome (DC) was established in 1994 by Bayer et al and has since become a popular notion^[19,208] as a potential solution to improve cellulose degradation. Originally, three main different approaches were suggested to create these designer cellulosomes.^[204] The first method proposed making designer cellulosome using the method of cross-linking to the CBM units to attach additional enzymes or other molecules producing supercellulosomes from natural cellulosomes.^[204] Another type of DC, heterocellulosomes, were proposed, created from enzyme pools doped with specific enzyme–dockerin complexes in order to create cellulosomes with new enzyme arrangements. Finally, chimeras (Figure 14), employing recombinant cellulosomal modules, were suggested.

The concept behind this type of designer cellulosome was based on the specific interaction between cohesin and dockerin modules.

In 2001, Fierobe et al,^[12] were the first to report their attempt to construct a small artificial cellulosome. They engineered a divalent DC using cellulosomes from two well characterized cellulosome systems, *C. cellulolyticum* and *C. thermocellum*. The activity of chimeric cellulosomes on microcrystalline cellulose was monitored from the amount of soluble sugars released (μM) after 24 h incubation at 37 °C, with 4 ml of Avicel, a type of crystalline cellulose.^[12] Cellulosome chimeras were found to be about 2–3-fold more active than simple mixtures of free enzyme pairs. However, it is interesting to observe how in the absence of CBM, the enhanced cellulolytic activity was only about 1.5-fold. The presence of targeting CBM therefore, greatly contributes to the overall enzyme activity. As an extension to this work published the following year, Fierobe et al constructed a library of 75 chimeric cellulosomes.^[209] The catalytic properties of these chimeras were tested on various cellulose sources and a number of synergistic effects were observed from varying enzymes and substrate arrangements. Crucially, the study revealed that two defined factors, the proximity of cellulosomal enzymes and the presence of scaffoldin-borne CBM, play major roles in the cellulolytic activity enhancement. To further investigate the effect of CBM on cellulose degradation, scaffoldins containing two CBMs were compared to single-CBM scaffoldins. As expected, the double-CBM complexes showed lower activities mainly due to extra interaction with the substrate, therefore restricting mobility of the DC complex across the substrate. This observation is consistent with the observation that cellulosomal scaffoldins never contain more than one CBM.^[210] In conjunction with this experiment, the activities of wild-type cellulosomes purified from *C. cellulolyticum* and one of the most efficient ternary complexes in the study were evaluated on two types of cellulose substrates. The cellulosomes were found to be 10- and 3-fold more active than the DC on Avicel and bacterial cellulose, respectively.^[209] Bayer et al showed in 2005 that DCs can be assembled that exploit cooperation between cellulases and a hemicellulase from different microorganisms, with enhanced activity on straw.^[211]

Other works reported DCs composed of truncated scaffoldin from *C. cellulovorans* and copies of recombinant cellulosomal endoglucanase (Figure 2) enzymes. Murashima et al reported increased cellulolytic action on substrate^[212] and then later, Cha et al demonstrated a chimeric scaffoldin (with up to four cohesins) from *C. cellulovorans* combined with both endoglucanase and xylanase, had cellulose degradation activity 1.1- to 1.8-fold higher than wild type.^[213] However, in this study, the cohesins were not of divergent species like previously demonstrated by Fierobe et al.^[12,209,211] In an interesting sidenote, Doi and colleagues reported that *C. cellulovorans* native scaffoldin contained cohesin-dockerin interactions that were more selective than random, as reported earlier for *C. thermocellum*^[21] and *C. cellulolyticum*^[214] whereby cohesins of the same species recognize all dockerins in the same manner.^[215] The study showed hydrolytic activity differences as significant as 2.2-fold versus 3.9-fold for mini-DC with 2 different pairs of cohesins from the same *C. cellulovorans* scaffoldin.^[215]

Now that the combination of enzymes from different organisms in one single scaffold has been realized, the protocol for construction of DCs is quite malleable, therefore allowing scientists to further investigate questions about enzyme synergism and CBM function. A 2007 study of cellulosome geometry compared numerous shapes of DCs.^[216] Novel atypical DC geometries and their activities were compared with the corresponding cellulosome structure-function properties in nature.^[216] The new cellulosome geometries were found to be about 20–25% less active than the “conventional” hybrid cellulosome, which was ascribed to multiple cohesin-dockerin interactions restricting mobility of catalytic modules. The mobility of catalytic subunits and the use of a single CBM to target substrates were shown to be essential factors in these novel systems.^[216] These properties indeed match wild-type cellulosome complexes from the clostridia family, but probably do not apply to elaborate cellulosomes generated by *A. cellulolyticus* or ruminal bacteria with complex organisations of interacting scaffoldins.^[108,109,113]

Anaerobic cellulolytic bacteria such as *C. thermocellum* produce cellulases in two alternative states: “cellulosomal” cellulases and “free” cellulases (Figure 15). Unlike cellulosomal cellulases, a free cellulase usually contains a CBM to guide the catalytic module to the substrate, instead of a full dockerin.^[157,217] Vazana and coworkers created a set of wild type and converted DCs to study and compare the action of combinations of cellulosomal and non-cellulosomal cellulases from *C. thermocellum* on crystalline cellulose.^[130] Cellulosomal enzymes were converted to non-cellulosomal enzymes by swapping the dockerin domain of the cellulosomal enzyme with a CBM and vice versa. CBM-bearing enzymes (both wild type and converted) were the most efficient for solubilizing microcrystalline cellulose. It was therefore made clear once again that targeting the enzyme to cellulose is a crucial factor responsible for increased activity among the enzyme combinations.^[130]

Alternatively, cellulolytic enzymes from non-cellulosome producing microbes can be engineered and incorporated into DCs. To this end, the usage of a free-cellulase system of *Thermobifida fusca* as a DC was explored.^[218–221] The main objective of this research was to entirely convert a free cellulase system to a cellulosomal system by binding the cellulases to a chimeric scaffoldin. The aerobic bacterium *T. fusca* was chosen as it contains a limited number of highly active non-cellulosomal cellulases. Early research concentrated on converting the free cellulase systems from *T. fusca* to cellulosomal mode.^[218,219] *T. fusca* enzymes in cellulosomal mode were then incorporated into scaffoldin which served to boost the cellulose hydrolysis compared to the free wild-type enzymes. As part of this series

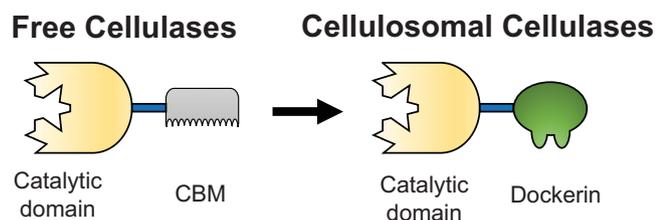


Figure 15. Free cellulases in DCs are converted to cellulosomal systems by replacing the enzyme-bound CBM module with a dockerin module.

of papers, the effect on enzymatic activity of linker length and dockerin position (relative to the catalytic module) was systematically analysed.^[220] The data indicated that positioning of the dockerin on opposite sides of the enzyme resulted in an enhanced synergistic response but changing linker lengths (in the range of 9 to 22 residues) between the catalytic module and the dockerin had little, if any, effect on the activity.^[220,222] By contrast, Vazana and colleagues later investigated the spatial organization of scaffoldin subunits using a synthetic biology approach. They reported that DCs assembled using scaffoldins containing longer linkers achieved higher levels of activity, compared to scaffoldins containing shorter linkers (ranging from 5 to 31 residues) or no linker.^[223] In this study, the linkers connected four cohesin modules in a tetravalent chimeric scaffold. On comparing the two studies (Vazana 2013 and Caspi 2009), it can be concluded that the high intrinsic flexibility of linkers on the chimeric scaffoldin is a principal factor in the activity of cellulase. It may facilitate substrate targeting which is not the case for linkers present on the cellulase domain.

All the DCs created so far are mini-cellulosomes. As they require a low level of enzymes, the expression of mini-cellulosomes is relatively easy. To date, the largest scaffoldin-based DC reported is the hexavalent scaffoldin subunit which forms a homogeneous multi-enzyme complex.^[208] Morais and colleagues reported that the hexavalent DC produced higher quantities of all sugars from untreated wheat straw, compared to the free enzymes. Until now, mini-cellulosomes with defined subunit compositions have been constructed in vivo.^[12,210,211,218,220] These minicellulosomes were built using cellulosomal cellulases that have a single catalytic module in each individual enzyme. Xu et al presented a novel DC demonstrating how multi-catalytic cellulases increased the hydrolysis of cellulose, due to high effective concentration of enzymatic sites.^[224] In general although more efficient than free enzymes, DCs are still not as effective as wild-type cellulosomes on cellulose substrates. For the first time in vitro, instead of a mini chimeric scaffoldin, a full active cellulosome structure has recently been reproduced from *Clostridium thermocellum*. This complex contains the full scaffoldin CipA and reached an overall activity of 80% of that of the native cellulosome, in the hydrolysis of crystalline cellulose.^[225]

Designer cellulosome studies have provided rich insights into the nanoscale mechanisms of cellulose hydrolysis, but the cellulosome structure and function still remains to be completely clarified. By combining innovative techniques such as small angle X-ray scattering (SAXS), cryo-EM and computational methods, Smith and Bayer have reported more comprehensive molecular understanding of the conformational event involved in the assembly of cellulase on the scaffoldin subunit.^[97] This recent development in the field will bring significant impact in the development of novel forms of artificial cellulosomes.

Finally, we note that a recent series of simulation papers have made an impressive leap forward in terms of model size and complexity, and revealed new, potentially useful, features of cellulosome assembly and function. Smith et al.^[226–229] recently studied the impediment effects of lignin on cellulose degradation by TrCel7A from *T. reesei*, through microsecond scale simulations of lignocellulosic systems containing up to

5 million atoms.^[226] Such massive simulation cells allow modeling of 36 chains of cellulose with 80 monomers per chain and 52 molecules of lignin each made of 61 monomers. modeling of solvent using a Reaction Field method made it possible to parallelise the calculations over 30 000 computing cores, which resulted in a production rate of 30 nanoseconds of dynamics per day on this extremely large system, making microsecond simulations accessible in a few-week timeframe. Such model sizes and time scales allow direct comparison with experiments in reactor-like environments, providing insights that relate directly to industrial needs. For example, the assembly of lignin aggregates of 25 molecules (with a diameter around 84 Å) was studied in simulations and used to rationalize small-angle neutron scattering (SANS) data.^[227] The simulations revealed the size-invariant fractal nature of the surface of the aggregates, the high penetration of water in them and their complex dynamics as the core of the aggregates is rigid while the surface is fluid. This lignin model was also used for simulation of lignin polymers at different temperatures.^[228] Taking a total of 17.5 μs of simulations arising from tens of trajectories, Smith et al.^[228] studied in detail the structural and dynamical changes lignin and its solvation water undergo as temperature increases. Lignin was shown to evolve form a collapse structure with glassy dynamics to an extended shape, with faster dynamics for temperature above 420 K. These changes are thermodynamically driven by water in the solvation shell. At low temperature, the collapsed lignin structure is driven by unfavorable translational entropy of water while the extended structure is enthalpically driven. More recently, Smith et al. focused on precipitation of lignin onto cellulose fiber after pretreatment of lignocellulosic biomass and studied how it impedes subsequent cellulose breakdown.^[229] Building large cellulose fibers (36 chains each composed of 160 monomers) in crystalline and non-crystalline forms with 52 lignin molecules at different initial locations, they performed a total of 3 μs of non-equilibrium simulation of 3 million atoms. Their simulation data shows that hydrophilic regions of cellulose tend to attract less lignin precipitate than hydrophobic and crystalline cellulose regions. Biomass pretreatment should then maximize hydrophilic non-crystalline cellulose.

are still restricted to the laboratory level in which researchers concentrate on improving cellulosomes using biophysical and biochemical properties that can be measured and compared using traditional techniques. However, bulk techniques are very limited in their description of the system and there is a clear need to obtain more detailed descriptions from these nanoscale systems including properties that can only be measured at this level (e.g., mechanostability or exceedingly high affinity interactions). Thus, the remarkable mechanostability and the extraordinarily high affinity of some of the interactions of the cellulosomal constituents remain to be fully explained and exploited. Allied with experiments, current computational multi-scale methods allow a first description of this system for verification and prediction purposes.

In recent years several groups have reported the use of cellulosomes or related composite catalysts claiming the possibility of efficient hydrolysis of biomass. For example, cellulosomes from the mesophilic bacterium *C. cellulovorans* were able to completely degrade soft biomass (rice straw).^[231] In addition, *C. thermocellum* cells were used to degrade the same substrate in an efficient manner.^[232] Others used cell-surface display of mini-cellulosomes on bacteria and yeasts to degrade cellulosic substrates.^[233,234] A synthetic cellulosome mimic for cellulose ethanol production was also constructed,^[235] based on a similar previously published approach.^[236]

By combining the cohesin-dockerin interactions and advancements in protein re-engineering, new types of synthetic cellulosomes have been generated, an example of which is the thermostable group II chaperonin called a rosettazyme.^[235,237] The rosettazyme is an 18-subunit protein complex (Figure 16) that can be genetically engineered to bind dockerin-containing enzymes and function like a cellulosome. The rosettasome linked with a cohesin can function as a scaffold for a maximum of four *C. thermocellum* cellulases.^[235] The rosettazyme complex showed improved enzymatic activity when at least two scaffolds were combined.^[235] A similar approach showed that the cohesin module fused to each of the 12 subunits of a protein complex from aspen trees (SP1) could bind to a single dockerin-containing cellulase from *Thermobifida fusca*.^[236] Up to ≈10 enzyme units were introduced on a Coh-SP1 through

6.2. Convergence of Designer Cellulosomes with Existing and Emerging Strategies

In the market of second generation biofuels and more generally lignocellulose derived chemical products there is an increasing need for more efficient and cost-effective saccharification processes. While cellulosome provides a potential complete solution from biology that is (to paraphrase a 19th Century Irish amateur botanist and process engineer, in keeping with the theme of this special issue) a rather unique association of the useful, the profitable, and agreeable,^[230] most industrial efforts concentrate on the use of fungal cellulases and fungal-based enzyme cocktails. Bacterial cellulosomes

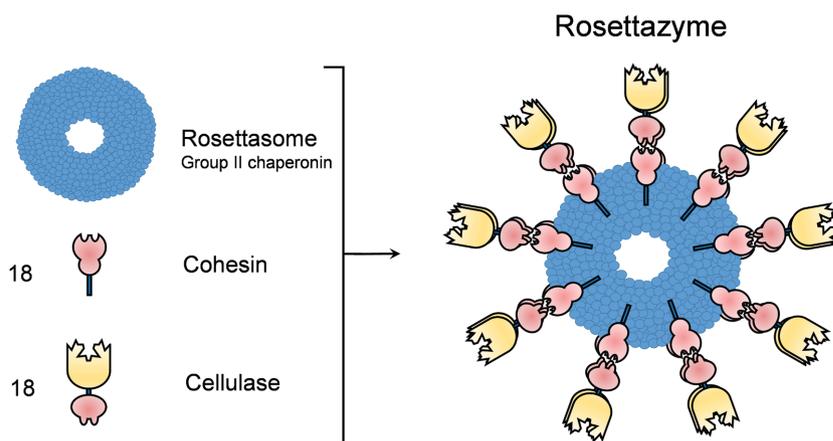


Figure 16. Schematic representation of the recombinant proteins used in rosettazyme construction. The rosettasome is made up of 18 subunits which are each combined with a cohesin module from CipA to make rosettazyme.

dockerin-cohesin interaction. In a complementary study, a total incorporation of dockerin-bearing exoglucanase was achieved on the SP1 complex.^[238] The incorporation of endoglucanase resulted in a significant enhancement of enzyme activity. The main advantage with such structures is that one single particle can host a larger number of enzymes.

A more recent study reported a DC constructed from a scaffold of ankyrin proteins (from the Greek word for “fused”, these are adaptor molecules that help attach cells to tissues) combined with endocellulase catalytic domains. Unlike the “rozetta” particle used in refs,^[235,237] the arrays are single polypeptide chains with cellulase domains internal to the scaffoldin protein. The results show ankyrin arrays to be a promising scaffold for constructing DCs, preserving or enhancing the enzymatic activity while retaining thermostability.^[237]

It is clear that the construction of designer cellulosomes is quite challenging, and it is essential to build designer cellulosomes with functions specific to the industrial application of interest (e.g., bio-refineries). Along with the improvements in the atomic scale design, there is the need, from the point of view of biotechnology, to produce high quantities of DCs through microbial fermentation.^[73] Regardless of the scale of fermentation or the producer microorganisms, the DCs must be produced with consistent activity as well as stable structures. An understanding of the cellulosome complexes and the specific properties associated with their components has allowed the design of specialized microorganisms with controlled expression of enzymes and scaffoldings that permitted the in vivo construction of the first minicellulosomes.^[12,238] Recombinant minicellulosome formation first used multiple bacterial systems to produce the required protein. However, the latest advances in expression techniques allow unprecedented control over protein motion and localisation, which in turn allows heterogeneous minicellulosome components to be produced within single organisms in *E. coli*, *B. subtilis* and others. The MiniCbpA scaffoldin from *C. cellulovorans*, was successfully expressed in *B. subtilis*, *S. cerevisiae* and *Corynebacterium glutamicum*.^[234,239]

Parallel to these activities in the production of minicellulosomes, Nanosized platforms, based on a variety of different types of materials, are currently being assessed as potential cellulosomal scaffolds. Kim and colleagues^[240] used the high affinity streptavidin-biotin binding pair to form clusters of organic nanoparticles with expressed CBM modules and cellulosome enzymatic domains. In some cases, streptavidin conjugated cadmium selenide nanospheres were used as a cellulosomal scaffold. Their high surface area and low diffusion properties make nanospheres attractive candidates for the immobilization of cellulases. Cellulose hydrolysis activity was boosted with an increase in the number of CBM domains on the scaffold. A near ten-fold increase in activity was observed for the CBM and enzyme complex on streptavidin-conjugated CdSe nanoparticles compared to free enzyme. Also recently, cellulases were clustered on nanospheres of CdSe-ZnS core shell quantum dots^[241] which resulted in more uniform nanoparticles. The evaluation of nanoparticle size and enzyme proximity indicated that enzyme proximity was more important than particle size.

Chemical conjugation, as opposed to non-covalent binding complexes, is an alternative technique to immobilize cellulases,

used recently to deposit cellulases onto polystyrene nanospheres.^[242] The results showed that both recalcitrant wood biomass and crystalline cellulose were efficiently degraded, with lower efficiency obtained in degradation of soluble carboxyl methyl cellulose.

Another recent, promising advance in the design of artificial cellulosomes is a novel approach based on the use of DNA^[243], in which the protein cross-linking enzyme, microbial transglutaminase, is used. DNA comes with in-built attractive features of mechanical flexibility, easy manipulation and expression, and is also both physically and chemically stable. The first DNA-cellulase cluster (with Cel5A endoglucanase from *Thermobifida fusca*) exhibited a five-fold increase in efficiency compared to free enzyme, on Avicel in the presence of free endoglucanase and β -glucosidase. In another more recent study using DNA as a scaffold,^[244] a series of artificial cellulosomes based on Zinc finger protein (ZFP)-guided assemblies were constructed. CelA and CBM, both from *C. thermocellum*, were assembled into a bifunctional cellulosome complex on DNA for two-fold enhanced cellulose hydrolysis.

6.3. Prospects for Near-Term Biofuel Production using Designer Cellulosomes

A reliable supply of sustainable energy is critical for the healthy, wealthy and peaceful future of our planet. Energy is the world's largest market, with a political and strategic impact that is unmatched by any other sector. Most countries, including European nations, are currently highly dependent on the finite and non-renewable resources of fossil fuels for their energy needs. This allows countries rich in such resources to become major players in world politics, frequently at the expense of countries that lack them. Biofuels constitute a major alternative to face this problem.^[207] Among all the candidate catalysts, nanocatalysts are very attractive ones as they greatly increase the surface-to-volume ratio compared to bulk materials. Thus, they hold promise for dramatically faster, cheaper, less toxic and environmentally friendly biofuel production. Recent advances in nanocatalysis have prompted a persistent shift in the economic and political balance of the fossil fuels market. As for all technological shifts, the control of the direction and magnitude of the change is in the hands of developers. Furthermore, there are alternative energy sources (some of them renewable and sustainable) that remain to be exploited. One of them is fiber, the non-edible plant cell wall cellulosic biomass, which is the source for second generation biofuels.

As described in this review, recycling the photosynthetically fixed carbon present in plant fiber is a relatively inefficient biological process due to the chemical and physical complexity of plant cell walls, which restricts their accessibility to enzyme attack so that only a restricted number of microorganisms (remarkably, some bacteria) have acquired the capacity to deconstruct these structural carbohydrates that are extremely recalcitrant to degradation. Thus the degradation of polysaccharides to fermentable sugars (saccharification)^[54] is the bottleneck for biofuel production from waste feed stocks. Since the needs of organisms are not necessarily the same as our industrial needs, a possible approach consists in reverse engineering

of available cellulosomes, exploiting the many different solutions which have appeared through evolution, in order to integrate them into an artificial design such as the “rosettazyme” and other constructs described above.

European directives impose that by the year 2020, 20% of the fuel consumption in each of the member states should be obtained from renewable sources (including biofuels, which are potential major contributors).^[245] Increasing bioethanol production using standard technology would imply a massive investment and would have a strong impact on both food resources and environment. Currently in mid-2015, a large majority of the bioethanol produced in EU and in the world is obtained from plant storage polysaccharides (i.e., starch). However, these are a major food source, and its potential stored energy represents less than 10% of that stored in the plant structural polysaccharide cellulose. Thus, the potential to produce so-called second generation biofuels from plant lignocellulosic biomass is enormous. Our assessment of the current literature indicates that the major impediment to technology transfer of DCs remains the lack of knowledge of nanoscale mechanics, dynamics and recognition, bottlenecks that are being systematically dismantled using a combination of bioinformatics, protein re-engineering, force microscopy, X-ray diffraction/scattering and multi-scale molecular dynamics modeling, by a growing community of researchers worldwide.

Acknowledgments

M.G. and P.-A.C. contributed equally to this work. The authors gratefully acknowledge financial support from the European Union's 7th Framework Programme (FP7/2007–2013) under grant number 604530–2 (CellulosomePlus) and ERA-NET Grant ERA-IB number EIB.12.022 (FiberFuel). D.T. and P.-A.C. thank Science Foundation Ireland (SFI) for additional financial support under Grant Number 11/SIRG/B2111. E.A.B. appreciates additional funding from the following agencies: Grant No. 1349 from the Israel Science Foundation (ISF), the Israel Strategic Alternative Energy Foundation (I-SAEF), and the Israeli Center of Research Excellence (I-CORE Center No. 152/11). M.A.N. acknowledges funding from the Society in Science/Branco Weiss Fellowship program administered through the ETH-Zurich. M.C. acknowledges support under the ERA-NET-IB/06/2013 Grant FiberFuel funded by the National Centre for Research and Development in Poland.

Received: August 13, 2015

Revised: October 1, 2015

Published online: January 7, 2016

- [1] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Essential Cell Biology*, Garland Science, New York 2013.
- [2] a) L. R. Lynd, P. J. Weimer, W. H. van Zyl, I. S. Pretorius, *Microbiol. Mol. Biol. Rev.* **2002**, *66*, 506; b) Y. H. P. Zhang, L. R. Lynd, *Biotechnol. Bioeng.* **2006**, *94*, 888.
- [3] Y. H. P. Zhang, L. R. Lynd, *Biotechnol. Bioeng.* **2004**, *88*, 797.
- [4] M. E. Himmel, S.-Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, T. D. Foust, *Science* **2007**, *315*, 804.
- [5] C. M. Payne, B. C. Knott, H. B. Mayes, H. Hansson, M. E. Himmel, M. Sandgren, J. Stahlberg, G. T. Beckham, *Chem. Rev.* **2015**, *115*, 1308.
- [6] M. B. Sticklen, *Nat. Rev. Genet.* **2008**, *9*, 433.
- [7] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.
- [8] a) J. K. Park, J. H. Shim, K. S. Kang, J. Yeom, H. S. Jung, J. Y. Kim, K. H. Lee, T. H. Kim, S. Y. Kim, D. W. Cho, S. K. Hahn, *Adv. Funct. Mater.* **2011**, *21*, 2906; b) K. N. L. Huggins, A. P. Schoen, M. A. Arunagirinathan, S. C. Heilshorn, *Adv. Funct. Mater.* **2014**, *24*, 7737; c) U. G. K. Wegst, H. Bai, E. Saiz, A. P. Tomsia, R. O. Ritchie, *Nat. Mater.* **2015**, *14*, 23.
- [9] E. Bayer, S. Smith, I. Noach, O. Alber, J. Adams, R. Lamed, L. Shimon, F. Frolow, *Biotechnology of Lignocellulose Degradation and Biomass Utilization*, (Eds: K. Sakka, S. Karita, T. Kimura, M. Sakka, H. Matsui, H. Miyake, A. Tanaka), Ito Print Publishing Division, Tokyo, Japan **2009**, pp. 183–205.
- [10] Y. Vazana, S. Morais, Y. Barak, R. Lamed, E. A. Bayer, in *Cellulases*, Vol. 510 (Ed: H. J. Gilbert), **2012**, 429.
- [11] E. Bayer, *FEBS J.* **2010**, *277*, 6.
- [12] H. P. Fierobe, A. Mechaly, C. Tardif, A. Belaich, R. Lamed, Y. Shoham, J. P. Belaich, E. A. Bayer, *J. Biol. Chem.* **2001**, *276*, 21257.
- [13] A. Limayem, S. C. Ricke, *Prog. Energy Combust. Sci.* **2012**, *38*, 449.
- [14] R. A. L. Jones, *Soft Machines: Nanotechnology and Life*, Oxford University Press, Oxford **2007**.
- [15] a) J. Israelachvili, M. Ruths, *Langmuir* **2013**, *29*, 9605; b) G. M. Whitesides, B. Grzybowski, *Science* **2002**, *295*, 2418.
- [16] J. Y. Weinstein, M. Slutzki, A. Karpol, Y. Barak, O. Gul, R. Lamed, E. A. Bayer, D. B. Fried, *J. Mol. Recognit.* **2015**, *28*, 148.
- [17] S.-Y. Ding, Q. Xu, M. Crowley, Y. Zeng, M. Nimlos, R. Lamed, E. A. Bayer, M. E. Himmel, *Curr. Opin. Biotechnol.* **2008**, *19*, 218.
- [18] E. A. Bayer, R. Lamed, B. A. White, H. J. Flint, *Chem. Rec.* **2008**, *8*, 364.
- [19] E. A. Bayer, R. Lamed, M. E. Himmel, *Curr. Opin. Biotech.* **2007**, *18*, 237.
- [20] E. A. Bayer, J. P. Belaich, Y. Shoham, R. Lamed, *Annu. Rev. Microbiol.* **2004**, *58*, 521.
- [21] E. A. Bayer, L. J. W. Shimon, Y. Shoham, R. Lamed, *J. Struct. Biol.* **1998**, *124*, 221.
- [22] E. A. Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham, in *Carbohydrases from Trichoderma Reesei and other Microorganisms*, The Royal Society of Chemistry, London **1998**, 39.
- [23] R. Lamed, E. A. Bayer, in *Genetics, Biochemistry and Ecology of Lignocellulose Degradation*, (Ed: K. Shimada, S. Hoshino, K. Ohmiya, K. Sakka, Y. Kobayashi, S. Karita), Uni Publishers Co., Ltd., Tokyo, Japan **1993**, 1.
- [24] A. S. Bommarius, B. R. Riebel, *Biocatalysis*, Wiley-VCH, Weinheim, Germany, **2004**.
- [25] G. M. Whitesides, *Sci. Am.* **1995**, *273*, 146.
- [26] J. E. Hyeon, D. H. Kang, S. O. Han, *Analyst* **2014**, *139*, 4790.
- [27] a) J. J. Champoux, *Annu. Rev. Biochem.* **2001**, *70*, 369; b) C. Mayer, Y. L. Janin, *Chem. Rev.* **2014**, *114*, 2313.
- [28] L. Stewart, M. R. Redinbo, X. Y. Qiu, W. G. J. Hol, J. J. Champoux, *Science* **1998**, *279*, 1534.
- [29] a) B. Xiong, D. L. Burk, J. H. Shen, X. M. Luo, H. Liu, J. K. Shen, A. M. Berghuis, *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1984; b) N. N. Wei, A. Hamza, C. Hao, Z. L. Xiu, C. G. Zhan, *Theor. Chem. Acc.* **2013**, *132*; c) F. M. Siu, Y. Pommier, *Nucleic Acids Res.* **2013**, *41*, 10010; d) F. M. Siu, C. M. Che, *J. Am. Chem. Soc.* **2008**, *130*, 17928.
- [30] a) D. Vuzman, Y. Levy, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 21004; b) A. Marcovitz, Y. Levy, *J. Phys. Chem. B* **2013**, *117*, 13005; c) A. Marcovitz, Y. Levy, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17957; d) N. Khazanov, A. Marcovitz, Y. Levy, *Biochemistry* **2013**, *52*, 5335; e) A. Bhattacharjee, Y. Levy, *Nucleic Acids Res.* **2014**, *42*, 12415.
- [31] O. Szklarczyk, K. Staron, M. Cieplak, *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 420.

- [32] a) A. L. Gnat, P. Cramer, J. H. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **2001**, 292, 1876; b) P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **2001**, 292, 1863; c) G. Bar-Nahum, V. Epshtein, A. E. Ruckenstein, R. Rafikov, A. Mustae, E. Nudler, *Cell* **2005**, 120, 183.
- [33] a) R. O. J. Weinzierl, *Chem. Rev.* **2013**, 113, 8350; b) L. T. Da, F. P. Avila, D. Wang, X. H. Huang, *PLoS Comput. Biol.* **2013**, 9, e1003020.
- [34] D. A. Silva, D. R. Weiss, F. P. Avila, L. T. Da, M. Levitt, D. Wang, X. H. Huang, *Proc. Natl. Acad. Sci. USA* **2014**, 111, 7665.
- [35] a) V. Serreli, C. F. Lee, E. R. Kay, D. A. Leigh, *Nature* **2007**, 445, 523; b) A. M. Rijs, N. Sandig, M. N. Blom, J. Oomens, J. S. Hannam, D. A. Leigh, F. Zerbetto, W. J. Buma, *Angew. Chem., Int. Ed.* **2010**, 49, 3896; c) M. R. Panman, P. Bodis, D. J. Shaw, B. H. Bakker, A. C. Newton, E. R. Kay, A. M. Brouwer, W. J. Buma, D. A. Leigh, S. Woutersen, *Science* **2010**, 328, 1255; d) S. Smolarek, A. M. Rijs, J. S. Hannam, D. A. Leigh, M. Drabbels, W. J. Buma, *J. Am. Chem. Soc.* **2009**, 131, 12902.
- [36] M. R. Perket, M. F. Hagan, *J. Chem. Phys.* **2014**, 140.
- [37] Y. W. Yin, T. A. Steitz, *Science* **2002**, 298, 1387.
- [38] A. Yonath, *Angew. Chem., Int. Ed.* **2010**, 49, 4340.
- [39] a) P. Jonkheijm, D. Weinrich, H. Schroder, C. M. Niemeyer, H. Waldmann, *Angew. Chem., Int. Ed.* **2008**, 47, 9618; b) J. D. Kim, D. G. Ahn, J. W. Oh, W. J. Park, H. I. Jung, *Adv. Mater.* **2008**, 20, 3349.
- [40] M. Laurberg, H. Asahara, A. Korostelev, J. Y. Zhu, S. Trakhanov, H. F. Noller, *Nature* **2008**, 454, 852.
- [41] A. Aleksandrov, T. Simonson, *J. Am. Chem. Soc.* **2008**, 130, 1114.
- [42] a) J. Trylska, *Q. Rev. Biophys.* **2009**, 42, 301; b) P. M. Petrone, C. D. Snow, D. Lucent, V. S. Pande, *Proc. Natl. Acad. Sci. USA* **2008**, 105, 16549.
- [43] R. Brandman, Y. Brandman, V. S. Pande, *PLoS One* **2012**, 7, e30022.
- [44] K. Y. Sanbonmatsu, S. Joseph, C. S. Tung, *Proc. Natl. Acad. Sci. USA* **2005**, 102, 15854.
- [45] K. Y. Sanbonmatsu, *Curr. Opin. Struct. Biol.* **2012**, 22, 168.
- [46] G. D. Bachand, N. F. Bouxsein, V. VanDelinder, M. Bachand, *Wiley Interdiscip. Rev.: Nanomed. Nanobiotechnol.* **2014**, 6, 163.
- [47] R. D. Vale, T. S. Reese, M. P. Sheetz, *Cell* **1985**, 42, 39.
- [48] E. L. P. Dumont, C. Do, H. Hess, *Nat. Nanotechnol.* **2015**, 10, 166.
- [49] T. Fischer, A. Agarwal, H. Hess, *Nat. Nanotechnol.* **2009**, 4, 162.
- [50] M. Nakamura, L. Chen, S. C. Howes, T. D. Schindler, E. Nogales, Z. Bryant, *Nat. Nanotechnol.* **2014**, 9, 693.
- [51] A. Perl, A. Gomez-Casado, D. Thompson, H. H. Dam, P. Jonkheijm, D. N. Reinhoudt, J. Huskens, *Nat. Chem.* **2011**, 3, 317.
- [52] S. O. Krabbenborg, J. Huskens, *Angew. Chem., Int. Ed.* **2014**, 53, 9152.
- [53] a) G. Vaaje-Kolstad, B. Westereng, V. G. H. Eijsink, S. J. Horn, M. Sorlie, F. Zerbetto, *WO/2012/01915* **2011**; b) A. Amore, V. Faraco, *Renewable Sustainable Energy Rev.* **2012**, 16, 3286; c) E. A. Bayer, G. Gefen, M. Anbar, E. A. Bayer, G. Gefen, M. Anbar, *WO2013114362-A1* **2013**, *US2014335573-A1* **2014**.
- [54] L. Viikari, Koivula Vehmaanperä, *Biomass Bioenergy* **2012**, 46, 13.
- [55] A. Peer, S. P. Smith, E. A. Bayer, R. Lamed, I. Borovok, *FEMS Microbiol. Lett.* **2009**, 291, 1.
- [56] B. Dassa, I. Borovok, V. Ruimy-Israeli, R. Lamed, H. J. Flint, S. H. Duncan, B. Henrissat, P. Coutinho, M. Morrison, P. Mosoni, C. J. Yeoman, B. A. White, E. A. Bayer, *PLoS One* **2014**, 9, e99221.
- [57] M. Czjzek, H. P. Fierobe, V. Receveur-Brechot, in *Cellulases*, Vol. 510 (Ed: H. J. Gilbert), Elsevier Academic Press Inc, San Diego **2012**, 183.
- [58] a) M. Mueller, S. Jenni, N. Ban, *Curr. Opin. Struct. Biol.* **2007**, 17, 572; b) I. Noach, F. Frolow, O. Alber, R. Lamed, L. J. W. Shimon, E. A. Bayer, *J. Mol. Biol.* **2009**, 391, 86.
- [59] M. A. Currie, K. Cameron, F. M. V. Dias, H. L. Spencer, E. A. Bayer, C. M. G. A. Fontes, S. P. Smith, Z. Jia, *J. Biol. Chem.* **2013**, 288, 7978.
- [60] M. A. Currie, J. J. Adams, F. Faucher, E. A. Bayer, Z. Jia, S. P. Smith, *J. Biol. Chem.* **2012**, 287, 26953.
- [61] M. Hammel, H. P. Fierobe, M. Czjzek, V. Kurkal, J. C. Smith, E. A. Bayer, S. Finet, V. Receveur-Brechot, *J. Biol. Chem.* **2005**, 280, 38562.
- [62] a) C. McCabe, X. Zhao, W. S. Adney, M. E. Himmel, in *Computational Modeling in Lignocellulosic Biofuel Production*, Vol. 1052 (Eds: M. R. Nimlos, M. F. Crowley), **2010**, 119; b) M. E. Himmel, M. F. Ruth, C. E. Wyman, *Curr. Opin. Biotechnol.* **1999**, 10, 358.
- [63] Y. H. P. Zhang, M. E. Himmel, J. R. Mielenz, *Biotechnol. Adv.* **2006**, 24, 452.
- [64] a) S. K. Kracun, J. Schuckel, B. Westereng, L. G. Thygesen, R. N. Monrad, V. G. H. Eijsink, W. G. T. Willats, *Biotechnol. Biofuels* **2015**, 8, 1; b) Q. Li, A. M. Coffman, L. K. Ju, *Enzyme Microb. Technol.* **2015**, 72, 42; c) M. Maki, K. T. Leung, W. Qin, *Int. J. Biol. Sci.* **2009**, 5, 500; d) Y. H. P. Zhang, J. Hong, X. Ye, in *Biofuels: Methods and Protocols*, Vol. 581 (Ed: J. R. Mielenz), **2009**, 213; e) S. Tian, Z. Wang, Z. Fan, L. Zuo, J. Wang, *Afr. J. Biotechnol.* **2011**, 10, 7122; f) M. Dashtban, M. Maki, K. T. Leung, C. Mao, W. Qin, *Crit. Rev. Biotechnol.* **2010**, 30, 302.
- [65] a) K. H. Malinowska, T. Verdorfer, A. Meinhold, L. F. Milles, V. Funk, H. E. Gaub, M. A. Nash, *ChemSusChem* **2014**, 7, 2825; b) K. H. Malinowska, T. Rind, T. Verdorfer, H. E. Gaub, M. A. Nash, *Anal. Chem.* **2015**, 87, 7133.
- [66] a) P. Bubner, J. Dohr, H. Plank, C. Mayrhofer, B. Nidetzky, *J. Biol. Chem.* **2012**, 287, 2759; b) P. Bubner, H. Plank, B. Nidetzky, *Biotechnol. Bioeng.* **2013**, 110, 1529.
- [67] G. J. Kleywegt, J. Y. Zou, C. Divne, G. J. Davies, I. Sinning, J. Stahlberg, T. Reinikainen, M. Srisodsuk, T. T. Teeri, T. A. Jones, *J. Mol. Biol.* **1997**, 272, 383.
- [68] W. Ubhayasekera, I. G. Munoz, A. Vasella, J. Stahlberg, S. L. Mowbray, *FEBS J.* **2005**, 272, 1952.
- [69] a) G. Vaaje-Kolstad, B. Westereng, S. J. Horn, Z. Liu, H. Zhai, M. Sorlie, V. G. H. Eijsink, *Science* **2010**, 330, 219; b) J. A. Langston, T. Shaghasi, E. Abbate, F. Xu, E. Vlasenko, M. D. Sweeney, *Appl. Environ. Microbiol.* **2011**, 77, 7007; c) C. M. Phillips, W. T. Beeson, J. H. Cate, M. A. Marletta, *ACS Chem. Biol.* **2011**, 6, 1399; d) R. J. Quinlan, M. D. Sweeney, L. Lo Leggio, H. Otten, J.-C. N. Poulsen, K. S. Johansen, K. B. R. M. Krogh, C. I. Jorgensen, M. Tovborg, A. Anthonsen, T. Tryfona, C. P. Walter, P. Dupree, F. Xu, G. J. Davies, P. H. Walton, *Proc. Natl. Acad. Sci. USA* **2011**, 108, 15079; e) B. Westereng, J. W. Agger, S. J. Horn, G. Vaaje-Kolstad, F. L. Aachmann, Y. H. Stenstrom, V. G. H. Eijsink, *J. Chromatogr. A* **2013**, 1271, 144.
- [70] P. Tomme, H. Vantilbeurgh, G. Pettersson, J. Vandamme, J. Vandekerckhove, J. Knowles, T. Teeri, M. Claeysens, *Eur. J. Biochem.* **1988**, 170, 575.
- [71] a) R. Lamed, E. Setter, R. Kenig, E. A. Bayer, *Biotechnol. Bioeng.* **1983**, 13, 163; b) R. Lamed, E. Setter, E. A. Bayer, *J. Bacteriol.* **1983**, 156, 828.
- [72] E. A. Bayer, R. Kenig, R. Lamed, *J. Bacteriol.* **1983**, 156, 818.
- [73] C. M. G. A. Fontes, H. J. Gilbert, in *Annu. Rev. Biochem.*, Vol. 79 (Eds: R. D. Kornberg, C. R. H. Raetz, J. E. Rothman, J. W. Thorner), **2010**, 655.
- [74] M. Hammel, H. P. Fierobe, M. Czjzek, J. C. Smith, V. Receveur-Brechot, *Eur. Biophys. J.* **2005**, 34, 723.
- [75] J. J. Adams, M. A. Currie, S. Ali, E. A. Bayer, Z. Jia, S. P. Smith, *J. Mol. Biol.* **2010**, 396, 833.
- [76] T. K. Ng, J. G. Zeikus, *Appl. Environ. Microbiol.* **1981**, 42, 231.
- [77] A. L. Demain, M. Newcomb, J. H. D. Wu, *Microbiol. Mol. Biol. Rev.* **2005**, 69, 124.

- [78] R. Lamed, J. Naimark, E. Morgenstern, E. A. Bayer, *J. Bacteriol.* **1987**, *169*, 3792.
- [79] R. H. Doi, Y. Tamaru, *Chem. Rec.* **2001**, *1*, 24.
- [80] R. H. Doi, A. Kosugi, K. Murashima, Y. Tamaru, S. O. Han, *J. Bacteriol.* **2003**, *185*, 5907.
- [81] J. P. Belaich, C. Tardif, A. Belaich, C. Gaudin, *J. Biotechnol.* **1997**, *57*, 3.
- [82] M. Kakiuchi, A. Isui, K. Suzuki, T. Fujino, E. Fujino, T. Kimura, S. Karita, K. Saki, K. Ohmiya, *J. Bacteriol.* **1998**, *180*, 4303.
- [83] a) J. Nolling, G. Breton, M. V. Omelchenko, K. S. Makarova, Q. D. Zeng, R. Gibson, H. M. Lee, J. Dubois, D. Y. Qiu, J. Hitti, Y. I. Wolf, R. L. Tatusov, F. Sabathe, L. Doucette-Stamm, P. Soucaille, M. J. Daly, G. N. Bennett, E. V. Koonin, D. R. Smith, G. T. C. S. C. P. Finishing, *J. Bacteriol.* **2001**, *183*, 4823; b) F. Sabathe, A. Belaich, P. Soucaille, *FEMS Microbiol. Lett.* **2002**, *217*, 15; c) F. Sabathe, P. Soucaille, *J. Bacteriol.* **2003**, *185*, 1092.
- [84] a) M. Pohlschroder, S. B. Leschine, E. Canaleparola, *J. Bacteriol.* **1994**, *176*, 70; b) M. Pohlschroder, E. Canaleparola, S. B. Leschine, *J. Bacteriol.* **1995**, *177*, 6625.
- [85] L. Artzi, E. Morag, Y. Barak, R. Lamed, E. A. Bayer, *mBio* **2015**, *6*, e00411.
- [86] L. Artzi, B. Dassa, I. Borovok, M. Shamshoum, R. Lamed, E. A. Bayer, *Biotechnol. Biofuels* **2014**, *7*, 1.
- [87] S. Y. Ding, E. A. Bayer, D. Steiner, Y. Shoham, R. Lamed, *J. Bacteriol.* **1999**, *181*, 6720.
- [88] S. Y. Ding, E. A. Bayer, D. Steiner, Y. Shoham, R. Lamed, *J. Bacteriol.* **2000**, *182*, 4915.
- [89] a) H. Ohara, S. Karita, T. Kimura, K. Sakka, K. Ohmiya, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 254; b) H. Ohara, J. Noguchi, S. Karita, T. Kimura, K. Sakka, K. Ohmiya, *Biosci. Biotechnol. Biochem.* **2000**, *64*, 80.
- [90] J. Kirby, J. C. Martin, A. S. Daniel, H. J. Flint, *FEMS Microbiol. Lett.* **1997**, *149*, 213.
- [91] S. Y. Ding, M. T. Rincon, R. Lamed, J. C. Martin, S. I. McCrae, V. Aurilia, Y. Shoham, E. A. Bayer, H. J. Flint, *J. Bacteriol.* **2001**, *1*, 1945.
- [92] A. Mechaly, H. P. Fierobe, A. Belaich, J. P. Belaich, R. Lamed, Y. Shoham, E. A. Bayer, *J. Biol. Chem.* **2001**, *276*, 19678.
- [93] J. J. Adams, G. Pal, Z. Jia, S. P. Smith, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 305.
- [94] F. Schaeffer, M. Matuschek, G. Guglielmi, I. Miras, P. M. Alzari, P. Beguin, *Biochemistry* **2002**, *41*, 2106.
- [95] C. Schoeler, K. H. Malinowska, R. C. Bernardi, L. F. Milles, M. A. Jobst, E. Durner, W. Ott, D. B. Fried, E. A. Bayer, K. Schulten, H. E. Gaub, M. A. Nash, *Nat. Commun.* **2014**, *5*, 1.
- [96] I. Noach, F. Frolow, H. Jakoby, S. Rosenheck, L. W. Shimon, R. Lamed, E. A. Bayer, *J. Mol. Biol.* **2005**, *348*, 1.
- [97] S. P. Smith, E. A. Bayer, *Curr. Opin. Struct. Biol.* **2013**, *23*, 686.
- [98] O. Salama-Alber, M. K. Jobby, S. Chitayat, S. P. Smith, B. A. White, L. J. W. Shimon, R. Lamed, F. Frolow, E. A. Bayer, *J. Biol. Chem.* **2013**, *288*, 16827.
- [99] A. Karpol, M. K. Jobby, M. Slutzki, I. Noach, S. Chitayat, S. P. Smith, E. A. Bayer, *FEBS Lett.* **2013**, *587*, 30.
- [100] O. Alber, I. Noach, M. T. Rincon, H. J. Flint, L. J. W. Shimon, R. Lamed, F. Frolow, E. A. Bayer, *Proteins: Struct., Funct., Bioinf.* **2009**, *77*, 699.
- [101] a) B. L. Lytle, B. F. Volkman, W. M. Westler, J. H. D. Wu, *Arch. Biochem. Biophys.* **2000**, *379*, 237; b) B. L. Lytle, B. F. Volkman, W. M. Westler, M. P. Heckman, J. H. D. Wu, *J. Mol. Biol.* **2001**, *307*, 745.
- [102] M. Sara, U. B. Sleytr, *J. Bacteriol.* **2000**, *182*, 859.
- [103] H. Engelhardt, J. Peters, *J. Struct. Biol.* **1998**, *124*, 276.
- [104] S. Chauvaux, M. Matuschek, P. Beguin, *J. Bacteriol.* **1999**, *181*, 2455.
- [105] T. Fujino, P. Beguin, J. P. Aubert, *J. Bacteriol.* **1993**, *175*, 1891.
- [106] M. Lemaire, H. Ohayon, P. Gounon, T. Fujino, P. Beguin, *J. Bacteriol.* **1995**, *177*, 2451.
- [107] Q. Xu, W. C. Gao, S. Y. Ding, R. Kenig, Y. Shoham, E. A. Bayer, R. Lamed, *J. Bacteriol.* **2003**, *185*, 4548.
- [108] Q. Xu, Y. Barak, R. Kenig, Y. Shoham, E. A. Bayer, R. Lamed, *J. Bacteriol.* **2004**, *186*, 5782.
- [109] Q. Xu, E. A. Bayer, M. Goldman, R. Kenig, Y. Shoham, R. Lamed, *J. Bacteriol.* **2004**, *186*, 968.
- [110] M. T. Rincon, S. Y. Ding, S. I. McCrae, J. C. Martin, V. Aurilia, R. Lamed, Y. Shoham, E. A. Bayer, H. J. Flint, *J. Bacteriol.* **2003**, *185*, 703.
- [111] M. T. Rincon, T. Cepeljnik, J. C. Martin, Y. Barak, R. Lamed, E. A. Bayer, H. J. Flint, *J. Bacteriol.* **2007**, *189*, 4774.
- [112] C. Sanchez, H. Arribart, M. M. Giraud Guille, *Nat. Mater.* **2005**, *4*, 277.
- [113] M. T. Rincon, T. Cepeljnik, J. C. Martin, R. Lamed, Y. Barak, E. A. Bayer, H. J. Flint, *J. Bacteriol.* **2005**, *187*, 7569.
- [114] M. T. Rincon, S. Y. Ding, S. I. McCrae, J. C. Martin, V. Aurilia, R. Lamed, Y. Shoham, E. A. Bayer, H. J. Flint, in *Genetics, Biotechnology of Lignocellulose Degradation and Biomass Utilization*, (Ed: K. Ohmiya, K. Sakka, S. Karita, T. Kimura, M. Sakka), Uni Publishers Co., Ltd., Tokyo, Japan **2004**, 241.
- [115] M. T. Rincon, J. C. Martin, V. Aurilia, S. I. McCrae, G. J. Rucklidge, M. D. Reid, E. A. Bayer, R. Lamed, H. J. Flint, *J. Bacteriol.* **2004**, *186*, 2576.
- [116] G. Guglielmi, P. Beguin, *FEMS Microbiol. Lett.* **1998**, *161*, 209.
- [117] R. H. Doi, in *Incredible Anaerobes: From Physiology to Genomics to Fuels, Vol. 1125* (Eds: J. Wiegand, R. J. Maier, M. W. W. Adams), **2008**, 267.
- [118] B. Garcia-Alvarez, R. Melero, F. M. V. Dias, J. A. M. Prates, C. Fontes, S. P. Smith, M. J. Romao, A. L. Carvalho, O. Llorca, *J. Mol. Biol.* **2011**, *407*, 571.
- [119] A. Valbuena, J. Oroz, R. Hervas, A. Manuel Vera, D. Rodriguez, M. Menendez, J. I. Sulkowska, M. Cieplak, M. Carrion-Vazquez, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13791.
- [120] M. Budayova, J. P. Astier, S. Veessler, M. Czjzek, A. Belaich, R. Boistelle, *J. Cryst. Growth* **1999**, *196*, 297.
- [121] a) O. Shoseyov, M. Takagi, M. A. Goldstein, R. H. Doi, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 3483; b) U. T. Gerngross, M. P. M. Romaniec, T. Kobayashi, N. S. Huskisson, A. L. Demain, *Mol. Microbiol.* **1993**, *8*, 325.
- [122] R. H. Doi, A. Kosugi, *Nat. Rev. Microbiol.* **2004**, *2*, 541.
- [123] W. H. Schwarz, *Appl. Microbiol. Biotechnol.* **2001**, *56*, 634.
- [124] a) G. J. Gerwig, P. Dewaard, J. P. Kamerling, J. F. G. Vliegthart, E. Morgenstern, R. Lamed, E. A. Bayer, *J. Biol. Chem.* **1989**, *264*, 1027; b) G. J. Gerwig, J. P. Kamerling, J. F. G. Vliegthart, E. Morag, R. Lamed, E. A. Bayer, *Eur. J. Biochem.* **1991**, *196*, 115.
- [125] a) G. J. Gerwig, J. P. Kamerling, J. F. G. Vliegthart, E. Morag, R. Lamed, E. A. Bayer, *J. Biol. Chem.* **1993**, *268*, 26956; b) G. J. Gerwig, J. P. Kamerling, J. F. G. Vliegthart, E. Morag, R. Lamed, E. A. Bayer, *Eur. J. Biochem.* **1992**, *205*, 799.
- [126] F. Mingardon, A. Chanal, C. Tardif, E. A. Bayer, H.-P. Fierobe, *Appl. Environ. Microbiol.* **2007**, *73*, 7138.
- [127] M. Voronov-Goldman, O. Yaniv, O. Gul, H. Yoffe, O. Salama-Alber, M. Slutzki, M. Levy-Assaraf, S. Jindou, L. J. W. Shimon, I. Borovok, E. A. Bayer, R. Lamed, F. Frolow, *FEBS Lett.* **2015**, *589*, 1569.
- [128] M. Slutzki, D. Reshef, Y. Barak, R. Haimovitz, S. Rotem-Bamberger, R. Lamed, E. A. Bayer, O. Schueler-Furman, *J. Biol. Chem.* **2015**, *290*, 13654.
- [129] a) J. Stern, M. Anbar, S. Morais, R. Lamed, E. A. Bayer, *Carbohydr. Res.* **2014**, *389*, 78; b) R. Borne, E. A. Bayer, S. Pages, S. Perret, H.-P. Fierobe, *FEBS J.* **2013**, *280*, 5764.

- [130] Y. Vazana, S. Morais, Y. Barak, R. Lamed, E. A. Bayer, *Appl. Environ. Microbiol.* **2010**, *76*, 3236.
- [131] A. Koivula, T. Reinikainen, L. Ruohonen, A. Valkeajarvi, M. Claeysens, O. Teleman, G. J. Kleywegt, M. Szardenings, J. Rouvinen, T. A. Jones, T. T. Teeri, *Protein Eng.* **1996**, *9*, 691.
- [132] L. Poidevin, J. Feliu, A. Doan, J. G. Berrin, M. Bey, P. M. Coutinho, B. Henrissat, E. Record, S. Heiss-Blanquet, *Appl. Environ. Microbiol.* **2013**, *79*, 4220.
- [133] M. Rief, M. Gautel, F. Oesterhelt, J. M. Fernandez, H. E. Gaub, *Science* **1997**, *276*, 1109.
- [134] M. Otten, W. Ott, M. A. Jobst, L. F. Milles, T. Verdorfer, D. A. Pippig, M. A. Nash, H. E. Gaub, *Nat. Methods* **2014**, *11*, 1127.
- [135] a) S. W. Stahl, M. A. Nash, D. B. Fried, M. Slutzki, Y. Barak, E. A. Bayer, H. E. Gaub, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 20431; b) M. A. Jobst, C. Schoeler, K. Malinowska, M. A. Nash, *J. Visualized Exp.* **2013**, *82*, e50950.
- [136] J. R. King, C. M. Bowers, E. J. Toone, *Langmuir* **2015**, *31*, 3431.
- [137] F. J. Kolpak, J. Blackwell, *Polymer* **1978**, *19*, 132.
- [138] S. Aravindanath, S. Sreenivasan, P. B. Iyer, *J. Polym. Sci., Part C: Polym. Lett.* **1986**, *24*, 207.
- [139] a) A. D. French, G. P. Johnson, *Cellulose* **2004**, *11*, 449; b) S. Mendonca, G. P. Johnson, A. D. French, R. A. Laine, *J. Phys. Chem. A* **2002**, *106*, 4115.
- [140] Y. J. Bomble, G. T. Beckham, J. F. Matthews, M. R. Nimlos, M. E. Himmel, M. F. Crowley, *J. Biol. Chem.* **2011**, *286*, 5614.
- [141] M. Wojciechowski, D. Thompson, M. Cieplak, *J. Chem. Phys.* **2014**, *141*, 245103.
- [142] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
- [143] N. Go, *Annu. Rev. Biophys. Bioeng.* **1983**, *12*, 183.
- [144] a) S. J. Marrink, A. H. de Vries, A. E. Mark, *J. Phys. Chem. B* **2004**, *108*, 750; b) S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J. Phys. Chem. B* **2007**, *111*, 7812.
- [145] M. Balsara, S. Stepaniants, S. Izrailev, Y. Oono, K. Schulten, *Bioophys. J.* **1997**, *73*, 1281.
- [146] J. I. Sulkowska, M. Cieplak, *J. Phys.: Condens. Matter* **2007**, *19*, 283201.
- [147] M. Sikora, J. I. Sulkowska, M. Cieplak, *PLoS Comput. Biol.* **2009**, *5*, e1000547.
- [148] a) L. Peplowski, M. Sikora, W. Nowak, M. Cieplak, *J. Chem. Phys.* **2011**, *134*, 085102; b) M. Sikora, M. Cieplak, *Phys. Rev. Lett.* **2012**, *109*, 208101.
- [149] M. Chwastyk, A. Galera-Prat, M. Sikora, A. Gomez-Sicilia, M. Carrion-Vazquez, M. Cieplak, *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 717.
- [150] M. Wojciechowski, P. Szymczak, M. Carrión-Vázquez, M. Cieplak, *Bioophys. J.* **2014**, *107*, 1661.
- [151] a) M. Sikora, M. Cieplak, *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1786; b) B. Rozycki, L. Mioduszewski, M. Cieplak, *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 3144.
- [152] B. A. Hall, M. S. P. Sansom, *J. Chem. Theory Comput.* **2009**, *5*, 2465.
- [153] F. Mayer, M. P. Coughlan, Y. Mori, L. G. Ljungdahl, *Appl. Environ. Microbiol.* **1987**, *53*, 2785.
- [154] a) K. T. Simons, R. Bonneau, I. Ruczinski, D. Baker, *Proteins: Struct., Funct., Bioinf.* **1999**, *37*, 171; b) C. A. Rohl, C. E. M. Strauss, K. M. S. Misura, D. Baker, in *Methods in Enzymology*, Vol. 383 (Eds: B. Ludwig, L. J. Michael), Academic Press, **2004**, *66*; c) J. Meiler, D. Baker, *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 538.
- [155] T. Reinikainen, L. Ruohonen, T. Nevanen, L. Laaksonen, P. Kraulis, T. A. Jones, J. K. C. Knowles, T. T. Teeri, *Proteins: Struct., Funct., Genet.* **1992**, *14*, 475.
- [156] a) A. L. Creagh, E. Ong, E. Jervis, D. G. Kilburn, C. A. Haynes, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12229; b) J. L. Henshaw, D. N. Bolam, V. M. R. Pires, M. Czjzek, B. Henrissat, L. M. A. Ferreira, C. Fontes, H. J. Gilbert, *J. Biol. Chem.* **2004**, *279*, 21552; c) R. Lamed, J. Tormo, A. J. Chirino, E. Morag, E. A. Bayer, *J. Mol. Biol.* **1994**, *244*, 236; d) P. Tomme, D. P. Driver, E. A. Amandoron, R. C. Miller, R. Antony, J. Warren, D. G. Kilburn, *J. Bacteriol.* **1995**, *177*, 4356; e) J. Tormo, R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham, T. A. Steitz, *EMBO J.* **1996**, *15*, 5739.
- [157] M. Linder, T. T. Teeri, *J. Biotechnol.* **1997**, *57*, 15.
- [158] N. Din, I. J. Forsythe, L. D. Burtnick, N. R. Gilkes, R. C. Miller, R. A. J. Warren, D. G. Kilburn, *Mol. Microbiol.* **1994**, *11*, 747.
- [159] B. Fan, J. Maranas, *Cellulose* **2015**, *22*, 31.
- [160] V. H. Rusu, R. Baron, R. D. Lins, *J. Chem. Theory Comput.* **2014**, *10*, 5068.
- [161] S. Markutsya, A. Devarajan, J. Y. Baluyut, T. L. Windus, M. S. Gordon, M. H. Lamm, *J. Chem. Phys.* **2013**, *138*, 214108.
- [162] a) L. Bu, G. T. Beckham, M. F. Crowley, C. H. Chang, J. F. Matthews, Y. J. Bomble, W. S. Adney, M. E. Himmel, M. R. Nimlos, *J. Phys. Chem. B* **2009**, *113*, 10994; b) L. T. Bu, M. E. Himmel, M. R. Nimlos, in *Computational Modeling in Lignocellulosic Biofuel Production*, Vol. 1052 (Eds: M. R. Nimlos, M. F. Crowley), **2010**, 99.
- [163] J. L. Sarver, J. E. Townsend, G. Rajapakse, L. Jen-Jacobson, S. Saxena, *J. Phys. Chem. B* **2012**, *116*, 4024.
- [164] F. P. Reis, A. Barbas, A. A. Klauer-King, B. Tsanova, D. Schaeffer, E. Lopez-Vinas, P. Gomez-Puertas, A. van Hoof, C. M. Arraiano, *PLoS One* **2013**, *8*, e76504.
- [165] a) F. Beck, P. Unverdorben, S. Bohn, A. Schweitzer, G. Pfeifer, E. Sakata, S. Nickell, J. M. Plitzko, E. Villa, W. Baumeister, F. Forster, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 14870; b) H. Ishida, *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 1985.
- [166] a) J. Jitonnom, M. A. L. Limb, A. J. Mulholland, *J. Phys. Chem. B* **2014**, *118*, 4771; b) J. Jitonnom, V. S. Lee, P. Nimmanpipug, H. A. Rowlands, A. J. Mulholland, *Biochemistry* **2011**, *50*, 4697.
- [167] R. Bernardi, I. Cann, K. Schulten, *Biotechnol. Biofuels* **2014**, *7*, 1.
- [168] R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, *55*, 2471.
- [169] a) G. M. Torrie, J. P. Valleau, *Chem. Phys. Lett.* **1974**, *28*, 578; b) S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, *J. Comput. Chem.* **1992**, *13*, 1011.
- [170] C. Micheletti, A. Laio, M. Parrinello, *Phys. Rev. Lett.* **2004**, *92*, 170601.
- [171] H. T. Dong, X. H. Qian, in *Computational Modeling in Lignocellulosic Biofuel Production*, Vol. 1052 (Eds: M. R. Nimlos, M. F. Crowley), **2010**, 1.
- [172] CPMD, <http://www.cpmd.org/>, Copyright IBM Corp 1990–2015, Copyright MPI fur Festkorperforschung Stuttgart 1997–2001. (accessed: November 2015).
- [173] M. Saharay, H. B. Guo, J. C. Smith, H. Guo, in *Computational Modeling in Lignocellulosic Biofuel Production*, Vol. 1052 (Eds: M. R. Nimlos, M. F. Crowley), **2010**, 135.
- [174] P. Lian, H. B. Guo, J. C. Smith, D. Q. Wei, H. Guo, *Cellulose* **2014**, *21*, 937.
- [175] X. L. Pan, W. Liu, J. Y. Liu, *J. Phys. Chem. B* **2013**, *117*, 484.
- [176] J. Liu, X. Wang, D. Xu, *J. Phys. Chem. B* **2010**, *114*, 1462.
- [177] M. Saharay, H. Guo, J. C. Smith, *PLoS One* **2010**, *5*, e12947.
- [178] a) O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor, A. D. Mackerell Jr., *J. Comput. Chem.*

- 2008, 29, 2543; b) O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor, A. D. MacKerell Jr., *J. Chem. Theory Comput.* **2009**, 5, 2353; b) O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison, A. D. MacKerell, *J. Chem. Theory Comput.* **2011**, 7, 3162; c) E. Hatcher, O. Guvench, A. D. MacKerell, *J. Phys. Chem. B* **2009**, 113, 12466; d) E. R. Hatcher, O. Guvench, A. D. MacKerell, *J. Chem. Theory Comput.* **2009**, 5, 1315; e) S. S. Mallajosyula, A. D. MacKerell, *J. Phys. Chem. B* **2011**, 115, 11215; f) E. P. Raman, O. Guvench, A. D. MacKerell, *J. Phys. Chem. B* **2010**, 114, 12981.
- [179] R. D. Lins, P. H. Hunenberger, *J. Comput. Chem.* **2005**, 26, 1400.
- [180] K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. Gonzalez-Outeirino, C. R. Daniels, B. L. Foley, R. J. Woods, *J. Comput. Chem.* **2008**, 29, 622.
- [181] a) J. F. Matthews, M. E. Himmel, J. W. Brady, in *Computational Modeling in Lignocellulosic Biofuel Production*, Vol. 1052 (Eds: M. R. Nimlos, M. F. Crowley), **2010**, 17; b) J. F. Matthews, G. T. Beckham, M. Bergenstrahle-Wohlert, J. W. Brady, M. E. Himmel, M. F. Crowley, *J. Chem. Theory Comput.* **2012**, 8, 735.
- [182] C. W. Fu, Y. P. Wang, T. Y. Fang, T. H. Lin, *PLoS One* **2013**, 8, e68565.
- [183] a) D. Sitkoff, K. A. Sharp, B. Honig, *J. Phys. Chem.* **1994**, 98, 1978; b) G. D. Hawkins, C. J. Cramer, D. G. Truhlar, *Chem. Phys. Lett.* **1995**, 246, 122; c) G. D. Hawkins, C. J. Cramer, D. G. Truhlar, *J. Phys. Chem.* **1996**, 100, 19824; d) J. Weiser, P. S. Shenkin, W. C. Still, *J. Comput. Chem.* **1999**, 20, 217; e) P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. H. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, T. E. Cheatham, *Acc. Chem. Res.* **2000**, 33, 889; f) A. Onufriev, D. Bashford, D. A. Case, *J. Phys. Chem. B* **2000**, 104, 3712.
- [184] M. R. Nimlos, G. T. Beckham, J. F. Matthews, L. T. Bu, M. E. Himmel, M. F. Crowley, *J. Biol. Chem.* **2012**, 287, 20603.
- [185] T. Li, S. H. Yan, L. S. Yao, *J. Mol. Model.* **2012**, 18, 1355.
- [186] G. T. Beckham, J. F. Matthews, Y. J. Bomble, L. T. Bu, W. S. Adney, M. E. Himmel, M. R. Nimlos, M. F. Crowley, *J. Phys. Chem. B* **2010**, 114, 1447.
- [187] N. F. Bras, N. M. F. S. A. Cerqueira, P. A. Fernandes, M. J. Ramos, *Int. J. Quantum Chem.* **2008**, 108, 2030.
- [188] a) H. Shiiba, S. Hayashi, T. Yui, *Cellulose* **2012**, 19, 635; b) T. Yui, H. Shiiba, Y. Tsutsumi, S. Hayashi, T. Miyata, F. Hirata, *J. Phys. Chem. B* **2010**, 114, 49.
- [189] M. Alahuhta, Q. Xu, Y. J. Bomble, R. Brunecky, W. S. Adney, S.-Y. Ding, M. E. Himmel, V. V. Lunin, *J. Mol. Biol.* **2010**, 402, 374.
- [190] D. D. Boehr, R. Nussinov, P. E. Wright, *Nat. Chem. Biol.* **2009**, 5, 789.
- [191] P. Andric, A. S. Meyer, P. A. Jensen, K. Dam-Johansen, *Biotechnol. Adv.* **2010**, 28, 308.
- [192] L. Bu, G. T. Beckham, M. R. Shirts, M. R. Nimlos, W. S. Adney, M. E. Himmel, M. F. Crowley, *J. Biol. Chem.* **2011**, 286, 18161.
- [193] J. Xu, M. F. Crowley, J. C. Smith, *Protein Sci.* **2009**, 18, 949.
- [194] J. Xu, J. C. Smith, *Protein, Eng., Des. Sel.* **2010**, 23, 759.
- [195] a) M. J. Harrison, A. S. Nouwens, D. R. Jardine, N. E. Zachara, A. A. Gooley, H. Nevalainen, N. H. Packer, *Eur. J. Biochem.* **1998**, 256, 119; b) J. P. Hu, P. Lanthier, T. C. White, S. G. McHugh, M. Yaguchi, R. Roy, P. Thibault, *J. Chromatogr. B: Biomed. Sci. Appl.* **2001**, 752, 349; c) I. Stals, K. Sandra, S. Geysens, R. Contreras, J. Van Beeumen, M. Claeysens, *Glycobiology* **2004**, 14, 713.
- [196] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, 26, 1781.
- [197] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, 102, 3586.
- [198] C. M. Payne, M. G. Resch, L. Q. Chen, M. F. Crowley, M. E. Himmel, L. E. Taylor, M. Sandgren, J. Stahlberg, I. Stals, Z. P. Tan, G. T. Beckham, *Proc. Natl. Acad. Sci. USA* **2013**, 110, 14646.
- [199] a) B. C. Knott, M. H. Momeni, M. F. Crowley, L. F. Mackenzie, A. W. Gotz, M. Sandgren, S. G. Withers, J. Stahlberg, G. T. Beckham, *J. Am. Chem. Soc.* **2014**, 136, 321; b) G. T. Beckham, J. Stahlberg, B. C. Knott, M. E. Himmel, M. F. Crowley, M. Sandgren, M. Sorlie, C. M. Payne, *Curr. Opin. Biotechnol.* **2014**, 27, 96.
- [200] B. C. Knott, M. F. Crowley, M. E. Himmel, J. Stahlberg, G. T. Beckham, *J. Am. Chem. Soc.* **2014**, 136, 8810.
- [201] J. Stahlberg, C. Divne, A. Koivula, K. Piens, M. Claeysens, T. T. Teeri, T. A. Jones, *J. Mol. Biol.* **1996**, 264, 337.
- [202] W. Materi, D. S. Wishart, *Drug Discovery Today* **2007**, 12, 295.
- [203] U. Heiz, U. Landman, *Nanocatalysis*, Springer-Verlag, Berlin, Heidelberg **2007**.
- [204] E. A. Bayer, E. Morag, R. Lamed, *Trends Biotechnol.* **1994**, 12, 379.
- [205] a) E. A. Bayer, R. Lamed, *Biodegradation* **1992**, 3, 171; b) K. Ohmiya, K. Sakka, T. Kimura, K. Morimoto, *J. Biosci. Bioeng.* **2003**, 95, 549.
- [206] a) F. Tao, L. Nguyen, S. Zhang, in *Metal Nanoparticles for Catalysis: Advances and Applications*, The Royal Society of Chemistry London, **2014**, 1; b) B. R. Cuenya, *Thin Solid Films* **2010**, 518, 3127; c) D. Wang, W. Q. Niu, M. H. Tan, M. B. Wu, X. J. Zheng, Y. P. Li, N. Tsubaki, *ChemSusChem* **2014**, 7, 1398.
- [207] A. L. Demain, *J. Ind. Microbiol. Biotechnol.* **2009**, 36, 319.
- [208] S. Morais, E. Morag, Y. Barak, D. Goldman, Y. Hadar, R. Lamed, Y. Shoham, D. B. Wilson, E. A. Bayer, *mBio* **2012**, 3, e00508.
- [209] H. P. Fierobe, E. A. Bayer, C. Tardif, M. Czjzek, A. Mechaly, A. Balaich, R. Lamed, Y. Shoham, J. P. Belaich, *J. Biol. Chem.* **2002**, 277, 49621.
- [210] H. P. Fierobe, E. A. Bayer, C. Tardif, M. Czjzek, A. Mechaly, A. Balaich, R. Lamed, Y. Shoham, J. P. Belaich, *J. Biol. Chem.* **2002**, 277, 49621.
- [211] H. P. Fierobe, F. Mingardon, A. Mechaly, A. Balaich, M. T. Rincon, R. Lamed, C. Tardif, J. P. Belaich, E. A. Bayer, *J. Biol. Chem.* **2005**, 280, 16325.
- [212] K. Murashima, C. L. Chen, A. Kosugi, Y. Tamaru, R. H. Doi, S. L. Wong, *J. Bacteriol.* **2002**, 184, 76.
- [213] J. Cha, S. Matsuoka, H. Chan, H. Yukawa, M. Inui, R. H. Doi, *J. Microbiol. Biotechnol.* **2007**, 17, 1782.
- [214] H. P. Fierobe, S. Pages, A. Balaich, S. Champ, D. Lexa, J. P. Belaich, *Biochemistry* **1999**, 38, 12822.
- [215] R. Koukiekolo, H. Y. Cho, A. Kosugi, M. Inui, H. Yukawa, R. H. Doi, *Appl. Environ. Microbiol.* **2005**, 71, 3504.
- [216] F. Mingardon, A. Chanal, C. Tardif, E. A. Bayer, H. P. Fierobe, *Appl. Environ. Microbiol.* **2007**, 73, 7138.
- [217] A. B. Boraston, D. N. Bolam, H. J. Gilbert, G. J. Davies, *Biochem. J.* **2004**, 382, 769.
- [218] J. Caspi, D. Irwin, R. Lamed, Y. Shoham, H. P. Fierobe, D. B. Wilson, E. A. Bayer, *Biocatal. Biotransform.* **2006**, 24, 3.
- [219] J. Caspi, D. Irwin, R. Lamed, Y. C. Li, H. P. Fierobe, D. B. Wilson, E. A. Bayer, *J. Biotechnol.* **2008**, 135, 351.
- [220] J. Caspi, Y. Barak, R. Haimovitz, D. Irwin, R. Lamed, D. B. Wilson, E. A. Bayer, *Appl. Environ. Microbiol.* **2009**, 75, 7335.
- [221] J. Caspi, Y. Barak, R. Haimovitz, H. Gilyar, D. C. Irwin, R. Lamed, D. B. Wilson, E. A. Bayer, *Syst. Biol. Synth. Biol.* **2010**, 4, 193.
- [222] M. Hammel, H. P. Fierobe, M. Czjzek, S. Finet, V. Receveur-Brechot, *J. Biol. Chem.* **2004**, 279, 55985.

- [223] Y. Vazana, Y. Barak, T. Unger, Y. Peleg, M. Shamshoum, T. Ben-Yehezkel, Y. Mazor, E. Shapiro, R. Lamed, E. A. Bayer, *Biotechnol. Biofuels* **2013**, *6*, 18.
- [224] Q. Xu, S. Y. Ding, R. Brunecky, Y. J. Bomble, M. E. Himmel, J. O. Baker, *Biotechnol. Biofuels* **2013**, *6*, 9.
- [225] J. Krauss, V. V. Zverlov, W. H. Schwarz, *Appl. Environ. Microbiol.* **2012**, *78*, 4301.
- [226] R. Schulz, B. Lindner, L. Petridis, J. C. Smith, *J. Chem. Theory Comput.* **2009**, *5*, 2798.
- [227] L. Petridis, S. V. Pingali, V. Urban, W. T. Heller, H. M. O'Neil, M. Foston, A. Ragauskas, J. C. Smith, *Phys. Rev. E* **2011**, *83*, 061911.
- [228] L. Petridis, R. Schulz, J. C. Smith, *J. Am. Chem. Soc.* **2011**, *133*, 20277.
- [229] B. Lindner, L. Petridis, R. Schulz, J. C. Smith, *Biomacromolecules* **2013**, *14*, 3390.
- [230] W. Roche, *The Kaleidoscope: or, Literary and Scientific Mirror*. **1825**, VI (275), 105.
- [231] Y. Tamaru, H. Miyake, K. Kuroda, A. Nakanishi, Y. Kawade, K. Yamamoto, M. Uemura, Y. Fujita, R. H. Doi, M. Ueda, *J. Bacteriol.* **2010**, *192*, 901.
- [232] R. Waeonukul, A. Kosugi, C. Tachaapaikoon, P. Pason, K. Ratanakhanokchai, P. Prawitwong, L. Deng, M. Saito, Y. Mori, *Bioresour. Technol.* **2012**, *107*, 352.
- [233] L.-H. Fan, Z.-J. Zhang, X.-Y. Yu, Y.-X. Xue, T.-W. Tan, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 13260.
- [234] T. D. Anderson, J. I. Miller, H.-P. Fierobe, R. T. Clubb, *Appl. Environ. Microbiol.* **2013**, *79*, 867.
- [235] S. Mitsuzawa, H. Kagawa, Y. Li, S. L. Chan, C. D. Paavola, J. D. Trent, *J. Biotechnol.* **2009**, *143*, 139.
- [236] A. Heyman, Y. Barak, J. Caspi, D. B. Wilson, A. Altmana, E. A. Bayer, O. Shoseyov, *J. Biotechnol.* **2007**, *131*, 433.
- [237] E. S. Cunha, C. L. Hatem, D. Barrick, *Appl. Environ. Microbiol.* **2013**, *79*, 6684.
- [238] S. Morais, A. Heyman, Y. Barak, J. Caspi, D. B. Wilson, R. Lamed, O. Shoseyov, E. A. Bayer, *J. Biotechnol.* **2010**, *147*, 205.
- [239] a) S. Y. Ding, Y. S. Liu, Y. N. Zeng, M. E. Himmel, J. O. Baker, E. A. Bayer, *Science* **2012**, *338*, 1055; b) a) H. Y. Cho, H. Yukawa, M. Inui, R. H. Doi, S. L. Wong, *Appl. Environ. Microbiol.* **2004**, *70*, 5704; b) J. E. Hyeon, W. J. Jeon, S. Y. Whang, S. O. Han, *Enzyme Microb. Technol.* **2011**, *48*, 371.
- [240] D.-M. Kim, M. Umetsu, K. Takai, T. Matsuyama, N. Ishida, H. Takahashi, R. Asano, I. Kumagai, *Small* **2011**, *7*, 656.
- [241] S.-L. Tsai, M. Park, W. Chen, *Biotechnol. J.* **2013**, *8*, 257.
- [242] C. Blanchette, C. I. Lacayo, N. O. Fischer, M. Hwang, M. P. Thelen, *PLoS One* **2012**, *7*, e42116.
- [243] Y. Mori, S. Ozasa, M. Kitaoka, S. Noda, T. Tanaka, H. Ichinose, N. Kamiya, *Chem. Commun.* **2013**, *49*, 6971.
- [244] Q. Sun, B. Madan, S.-L. Tsai, M. P. DeLisa, W. Chen, *Chem. Commun.* **2014**, *50*, 1423.
- [245] European Renewable Energy Council. Renewable Energy Technology Roadmap 20% by 2020. http://www.erec.org/fileadmin/erec_docs/Documents/Publications/Renewable_Energy_Technology_Roadmap.pdf (accessed: September 2015).