



**HAL**  
open science

## Continually emerging mechanistic complexity of the multi-enzyme cellulosome complex

Steven P Smith, Edward A Bayer, Mirjam Czjzek

► **To cite this version:**

Steven P Smith, Edward A Bayer, Mirjam Czjzek. Continually emerging mechanistic complexity of the multi-enzyme cellulosome complex. *Current Opinion in Structural Biology*, 2017, 44, pp.151 - 160. 10.1016/j.sbi.2017.03.009 . hal-03758476

**HAL Id: hal-03758476**

**<https://hal.science/hal-03758476>**

Submitted on 29 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Continually emerging mechanistic complexity of the multi-enzyme cellulosome complex

*Steven P Smith<sup>1</sup>, Edward A Bayer<sup>2,\*</sup>, and Mirjam Czjzek<sup>3,\*</sup>*

<sup>1</sup> Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, K7L 3N6, Canada

<sup>2</sup> Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel

<sup>3</sup> Sorbonne Universités, UPMC, Université Paris 06, and Centre National de la Recherche Scientifique, UMR 8227, Integrative Biology of Marine Models, Station Biologique Roscoff, CS 90074, F-29688 Roscoff cedex, Bretagne, France

\*Correspondence to: [ed.bayer@weizmann.ac.il](mailto:ed.bayer@weizmann.ac.il) and [czjzek@sb-roscoff.fr](mailto:czjzek@sb-roscoff.fr)

Short title: Structural complexity of the cellulosome

## Highlights

- Cellulosome complexes mediate efficient hydrolysis of plant-derived polysaccharides
- Computational biology reveals structure-function relationships of cellulosomes
- Subtle structural variations allow for fine-tuning of cellulolytic GH9 enzymes
- Intermodular linkers guide non-random assembly of the enzymes onto the scaffoldin
- Designer cellulosomes provide insight into complementary action of novel enzymes

## **Abstract**

The robust plant cell wall polysaccharide-degrading properties of anaerobic bacteria are harnessed within elegant, macromolecular assemblages called cellulosomes, in which proteins of complementary activities amass on scaffold protein networks. Research efforts have focused and continue to focus on providing detailed mechanistic insights into cellulosomal complex assembly, topology, and function. The accumulated information is expanding our fundamental understanding of the lignocellulosic biomass decomposition process and enhancing the potential of engineered cellulosomal systems for biotechnological purposes. Ongoing biochemical studies continue to reveal unexpected functional diversity within traditional cellulase families. Genomic, proteomic, and functional analyses have uncovered unanticipated cellulosomal proteins that augment the function of the native and designer cellulosomes. In addition, complementary structural and computational methods are continuing to provide much needed insights on the influence of cellulosomal interdomain linker regions on cellulosomal assembly and activity.

## Introduction

Bacteria and fungi have evolved various strategies to cope with the complexity and recalcitrance of polysaccharide networks within the plant cell wall. Heterotrophic bacteria that are specialized in plant polysaccharide degradation harbor sophisticated gene structures coding for complete co-regulated catabolic systems, which include proteins for sensing, uptake and degradation [1]. In contrast, certain aerobic bacteria and fungi secrete large amounts of highly efficient free (uncomplexed) enzymes that can deconstruct plant-derived cellulosic polysaccharides [2]. A particularly remarkable cellulolytic system, however, is found in anaerobic bacteria, which have evolved an alternative strategy involving the production of organized, large, extracellular, multi-protein assemblages called cellulosomes (for review see [3], [4] and [5]).

The typical cellulosomal complex is made up of a large 'backbone' scaffoldin protein that contains several tandem modules, termed cohesins, to which the active protein components adhere through their so-called dockerin modules (**Figure 1**). These are small, roughly 80-residue calcium-binding modules that non-covalently associate with the scaffoldin-borne cohesins at astonishingly high affinities (in the sub-nM range) [6]. The megadalton cellulosomes not only assemble the key enzymatic players of lignocellulosic degradation but also many other auxiliary modules, such as carbohydrate binding modules (CBMs) as well as other types of enzymes and proteins [7]. All cellulosomal components can be identified by the 'signature' dockerin sequence, and the presence of a scaffoldin protein in a bacterial genome indicates the cellulolytic nature of the organism. Since the first observations described in 1983 [8], where cellulosomes were considered to consist exclusively of various cellulases having different modes of action and CBMs, many other dockerin-

containing protein actors, such as the hemicellulases (e.g., xylanases, mannanases) [9], pectin lyases [10], expansins [11,12], and proteases and protease inhibitors [13,14], have since been identified in cellulosome-bearing organisms, adding further structural and functional complexity to these polysaccharide-degrading nanomachines.

The cellulosome structure-function relationship is not yet completely understood. However, studies using purified and synthetically assembled cellulosomes show that their catalytic efficiency towards natural complex (i.e., lignin-containing) cellulosic substrates is significantly enhanced compared to an enzyme cocktail of the same composition [15-17]. This synergistic effect is particularly linked to the higher order structure of the cellulosome [18,19]. However, these large and sophisticated protein-protein complexes are also difficult to assess by classical structural methods, as they are by definition heterogeneous in composition and contain oligosaccharides, flexible linkers and regions of intrinsic disorder [20,21]. A recent review on higher order cellulosome structure [5] describes the 'dissect and reconstruct' strategies that have been employed in concert with structural biology approaches in order to understand, at least partially, the three-dimensional arrangements of these nanomachines (**Figure 1**).

However, comprehensive appreciation of the structure-function relationship of these large assemblages goes beyond providing a glimpse of what such a cellulosome might look like in three-dimensional space, but also necessitates 1) understanding the fine-tuning and regulation of selecting auxiliary modules and enzymes with different specificities and activities, 2) quantitative description of the forces and affinities that hold the interacting components together, and 3) how

intrinsic flexibility and structural heterogeneity are linked to the synergistic power of these macromolecular complexes.

Recent advancements in our understanding of cellulosome structure and function include biochemical studies illustrating the diversity of functions that are integrated into the cellulosome, such as divergence of both enzymatic and newly discovered non-enzymatic activities. Recent studies have also focused on the fine-tuning of contacts and forces that govern the interactions associated with cellulosomal assembly and function. Transcriptomic and proteomic studies are further dissecting the regulation of which components are integrated into the cellulosomal complex in response to specific substrates and/or environment. The cohesin-dockerin interaction is well documented as essential for cellulosome assembly and function, and its unique binding properties and mechanostability have been recently summarized [22]. Here, we will focus on recent observations obtained by complementary biochemical, structural, and biophysical methods that shed light on the cellulosome structure-function relationship.

### **Environmental influences on cellulosome composition and function**

Towards providing a comprehensive understanding of the structural and functional complexity of cellulosomes one needs to consider the regulation of the expression of cellulosomal components at the transcriptional level. Defined RNA structures (such as stem-loops) of intergenic processed sites serve as a general mechanism for regulating the transcript abundance of encoded genes. This type of regulatory apparatus has been associated with quantitatively defining the stoichiometry of the various protein components found in cellulosomal complexes [23\*]. Moreover, these types of regulatory RNA secondary structures are conserved among different

cellulolytic clostridia. The authors interpret their results by a hierarchical regulation: mobilization of a particular cellulolytic subunit is sensitive to specific carbon substrates at the transcriptional level, yet composition and stoichiometry within an individual complex are precisely determined at the post-transcriptional level. The resultant recipes of cellulosomes are likely organism-specific. Given that cellulose and associated plant-derived polysaccharides are frequently the only available carbon source, these recipes are crucial to the strategy of adaptive competition in individual cellulolytic microbes [23\*].

### **Expanding the functional diversity among the ‘classical’ cellulosomal enzymes**

Cellulolytic organisms generally produce a similar cadre of enzymes belonging to specific glycoside hydrolase (GH) families, several members being cellulases [24,25]. Indeed, the cellulolytic properties have largely been associated with three complementary types of cellulases adapted to the different natural structures of cellulose - endo-acting cellulases, (exo-)processive cellulases, and cellobiose-degrading  $\beta$ -glucosidases. Prominent among these different types of enzymes are members of the GH48, GH9, GH8 and GH5 families. The era of intensive ‘omics’ has provided many examples of multigenic GH families within a single bacterial species. A long held assumption was that individual members within a single family had a common activity and mechanism of action. One such example is family GH9, which contains almost exclusively EC numbers of  $\beta$ -1,4-glucosidases (EC 3.2.1.4, EC 3.2.1.91 and EC 3.2.1.176). However, recent studies on GH5 and GH9 family members illustrate the necessity to comprehensively analyze the structure-function relationships of multiple family members since bioinformatic annotations are not yet



at the level of identifying subtle differences that play crucial roles in catalytic efficiency.

Family 9 glycoside hydrolases (GH9) were historically among the first discovered cellulase families (i.e., cellulase family E) that are highly expressed and present in numerous paralogous gene copies in anaerobic cellulolytic bacteria. The first structures of family GH9 cellulases proposed the co-existence of endo-acting and processive enzymes, depending on the presence of a tightly attached and positioned CBM3 module that was proposed to confer a processive mode of action by extending the active site groove beyond the catalytic module [26-29]. A recent, systematic biochemical study of all thirteen GH9 representatives in *Clostridium (Ruminiclostridium) cellulolyticum* revealed that the story is much more complex and highlights that the cellulolytic activities are fine-tuned by the subtle but significant variability of structural composition and enzymatic activities of these GH9 enzymes [30\*\*] (**Figure 2**). For example, one of the family members, Cel9V, was found to be inactive despite possessing the conserved catalytic residues and being detected in native purified cellulosomes. Such an observation raises the question of the functional role that such an “inactive” protein could play in substrate degradation. Another especially notable finding from this study was the discovery of a *bona fide* endoxyloglucanase member of the GH9 family, Cel9X, which would not have been predictable despite prior knowledge of the structure-function relationship of this family of enzymes. The authors propose that GH9s were selected as a type of ‘scaffold’ from which diverse modes of activity could be created [30\*\*].

GH5 family members are active on both cellulosic and various hemicellulosic  $\beta$ -glucan substrates and the panel of substrate specificity can vary from one member of the GH5 family to another. Moreover, since this enzyme family is one among the

very large and multi-specific GH families, the members were recently sub-divided into 43 subfamilies using phylogenetic analyses based on sequence similarity [31]. None of the sub-families coincides exclusively with enzymes containing a dockerin module, again showing that synergistic implications of this family of enzyme in cellulosomal activities appears to be fine-tuned by their respective specificity.

Growing attention is also being given to cellulolytic GH5 and GH9 enzymes that do not bear scaffoldin-binding dockerin modules. Rather, some of these extra-cellulosomal enzymes contain S-layer homology (SLH) domains, diverse CBMs, and have been found attached to the cell surface [32]. In rare cases, modified GHs have been proposed to be sensors for induction of glycoside hydrolase-encoding genes [33], where they simultaneously induce and/or complement the action of cellulosomes. These ancillary modules have also been suggested to provide an alternative mechanism of cellulosome attachment to the bacterial cell surface [34].

### **Computational insights into unique features of cellulosomal enzyme topology and mechanism**

Computational biology is emerging as a unique methodology, complementary to traditional biochemical and structural approaches, to model the synergistic activity of cellulosomes and their components. Such strategies include the modeling of chemical reactions at the quantum level to models of the macromolecular architectures using less detailed “coarse-grained” approaches. These techniques can provide qualitative structural and dynamic descriptions of the cellulosome and its isolated components and can be used to study chemical reactivity (e.g., catalytic turnover rates) [35].

Two recent representative examples that use computational biology to expand

our understanding about the role of enzyme topology and dynamics in cellulosome assembly and function focus on two multimodular cellulosomal enzymes. The combination of coarse-grained and molecular-dynamic (MD) simulations recently showed that the rigidity of the linker tethering the *Clostridium (Ruminiclostridium) thermocellum* Cel8A catalytic domain to its type-I dockerin module can influence the topology of the catalytic domain, when the enzyme is bound to its substrate [36]. The authors suggest that such structural changes may have an effect on the activity of the enzyme, prior to and/or upon integration into a cellulosomal complex. Molecular modeling of multimodular cellulosomal components and their interdomain linkers have also suggested that linker elasticity may enhance the activity of the resident catalytic domains. When assessing the full-length bifunctional xylanase, Xyn10Z, from *C. thermocellum*, which comprises a family 1 carbohydrate esterase (CE1) catalytic domain, a family 6 CBM, a type-I dockerin, and a C-terminal GH10 domain, the interdomain linker properties allowed more spatial liberty for the connected catalytic domains (i.e., CE1 and GH10) at opposite ends of the polypeptide [37]. The simulations also predicted that the spatial range of the two Xyn10Z catalytic domains was even larger when the resident type-I dockerin module was bound to a cognate scaffoldin-bourne cohesin module. These observations support the proposition that the catalytic efficiency of the enzyme is enhanced when integrated into the cellulosome [37].

Computational biology has also provided greater insight into the mechanism of processivity of cellulolytic enzymes including GH48, which is one of the principal cellulosomal components responsible for its extraordinary catalytic efficiency [38]. The processivity of an enzyme involves initial binding of the glycan whereupon it repeatedly cleaves the same type of glycosyl unit until it dissociates from the

substrate. Based on X-ray crystal structures of enzyme-substrate complexes, the modeling study by Qian et al. [38] sheds light on the importance of solvent effects and the amino-acid residues that define the trajectory of the product upon release from the catalytic active site. Another recent computational study modeled the mode of cellulose degradation by fungal cellobiohydrolases [39]. Computational simulations revealed that slow complexation of glucan chains and excessive enzyme jamming are rate-limiting mechanisms, which need to be fine-tuned to enhance the reaction rate. In contrast to fungal degradation systems that can be described in a simplified manner as consisting of large quantities of free enzyme cocktails, bacterial cellulosomes, and scaffoldin subunit in particular, represent a mechanism for minimizing these 'enzymatic traffic jams', thereby circumventing these rate-limiting factors during polysaccharide degradation.

### **Cellulosome-based enhancers of cellulose hydrolysis – novel dockerin-containing catalytic and non-catalytic components**

Utilizing a dockerin signature sequence-based mining approach, recent genomic and proteomic studies of cellulosome-producing bacteria have vastly expanded the catalog of cellulosomal components beyond the traditional enzymatic, scaffoldin and cell-surface subunits to include enzymes and proteins whose functions apparently complement and augment the degradative properties of the cellulosome. Pectinolytic activity of the *C. thermocellum* cellulosome was recently associated with three multimodular endo-pectate lyases from polysaccharide lyase families 1 and 9 (PL1A, PL1B, and PL9), each of which comprised a catalytic domain, a CBM, and a dockerin module. These enzymes displayed unique activities on polygalacturonic acid, methyl-esterified pectin, and rhamnogalacturonan [10\*].

Using a structure-guided approach to identifying novel cellulosomal components, Levy-Assaraf and coworkers discovered a dockerin-containing cysteine protease encoded in the genome of the cellulosome-producing bacterium *Ruminococcus flavefaciens* [13]. The presence of such non-cellulolytic enzymes leads to speculation as to whether cellulosomes possess secondary activities or if the combination of several, apparently unrelated, activities are necessary to match specific environmental conditions, in which these polymers may be encountered side by side.

A particularly striking and novel discovery was the recent identification and characterization of two dockerin-containing expansins encoded in the genome of the cellulosome-producing bacterium *Clostridium clariflavum* [11<sup>\*\*</sup>,12<sup>\*</sup>,40<sup>\*</sup>]. The expansin family of proteins, which have been identified in various plants and bacteria, are essentially non-catalytic proteins that destabilize the plant cell wall by disrupting the hydrogen-bonding networks of the associated polysaccharides; a “wedge-like” mechanism that has the potential to augment the hydrolytic properties of the cellulosome. A representative expansin was shown to preferentially adhere to microcrystalline cellulose and decreased the tensile strength of filter paper via a loosening effect [40<sup>\*</sup>], and its addition to native and designer cellulosomes significantly enhanced cellulose hydrolysis [11<sup>\*\*</sup>,12<sup>\*</sup>], supporting the notion of assisting in cellulose creeping [41]. Despite the apparent functional advantage that expansins impart to cellulosomes, this intriguing type of non-catalytic protein has only been shown to reside in those from *C. clariflavum*. Thus, an outstanding question is whether, in the absence of expansins, other yet-to-be-identified cellulosomal components assist in the relaxing of the cellulose structure. Regardless, such non-

catalytic cellulosomal components will need to be considered in future application of engineered enzyme complexes in biomass decomposition.

### **Extension of designer cellulosomes and integration of oxidative enzymes**

The concept of designer cellulosomes was proposed over two decades ago [42]. These artificial cellulosomes are engineered to comprise chimeric scaffoldins, which bear a multiplicity of cohesins of divergent dockerin-binding specificities. Desired chimeric enzymes are fitted with matching dockerins for their controlled integration at predetermined positions into the complex (**Figure 3**). Following the original proposal, it took seven years for the initial experimental feasibility study to be published in the scientific literature [43], wherein rudimentary designer cellulosomes were described. Since these early investigations, there has been systematic progress in the number and types of lignocellulosic biomass-degrading enzymes that have been incorporated into designer cellulosomes, and the performance of these artificial enzyme complexes has, in consequence, also increased. Chimeric scaffoldins of up to six cohesins have been successfully prepared, but hexavalent scaffoldins seem to be a limit for cloning and expression in *Escherichia coli* [44]. In order to resolve this barrier, a novel approach involving ‘adaptor’ scaffoldins was introduced, whereby two interacting scaffoldins enabled integration of increasing numbers of enzymes, thereby further enhancing the performance of the resultant designer cellulosome on lignocellulose substrates [18].

Designer cellulosome technology has recently been used to incorporate types of enzymes that are inevitably foreign to native cellulosome systems. In this context, oxidative enzymes have been integrated into designer cellulosomes. Dockerin modules were thus appended onto both lytic polysaccharide monooxygenases

(LPMOs) [15'] and a lignin-modifying laccase [45''] (**Figure 3**). In both cases, the resultant enzymes were shown to enhance significantly the hydrolysis of cellulosic substrates by the respective designer cellulosomes that contain these oxidative enzymes in addition to conventional glycoside hydrolases. This is surprising, in view of the fact that native cellulosome systems have been described exclusively in anaerobic bacteria, which would presumably preclude the inclusion of oxidative enzymes.

### **Cellulosomal carbohydrate-binding modules – the continually expanding 'sticky' repertoire**

The proximity effect imparted by resident CBMs to the efficiency with which the cellulosome degrades cellulosic substrates is well established [46], with the scaffoldin family-3a CBM being the archetypal example. Genomic sequencing and in-depth structural and biochemical studies continue to extend the repertoire of CBMs, or the CBMome, being employed by cellulolytic bacteria and their mechanisms of action. A recent study of the *R. flavefaciens* cellulosome characterized six previously unidentified CBM families that target  $\beta$ -glucans,  $\beta$ -mannans, and pectin homogalacturonan by employing unique ligand-binding aromatic conformations [47'']. A notable seminal finding in this study is that binding of homogalacturonan by the newly identified CBM77 family involves basic residues of the former and carboxylate groups of the latter [47'']. These binding specificity determinants were not predicted by bioinformatics alone and emphasize the need for comprehensive biochemical, biophysical and structural studies of cellulosomal components.

## **Functional influence of cellulosomal interdomain linker conformations**

The precise role and influence of linkers on catalytic efficiency of a native cellulosome continues to be a topic of debate. Their presumed intrinsic disorder, based on sequence analysis, has no doubt influenced our perspective that they simply serve as physical tethers on multiple and complementary functional components. The engineering of designer cellulosomes that function optimally and as expected is also critically reliant on a precise knowledge of the influence of interdomain linker amino-acid composition, length and dynamics. Molinier et al. utilized a designed divalent (two-cohesin) scaffoldin chimera to test the impact of intercohesin linker length on the synergistic hydrolysis of cellulose by resident cellulases [21]. Despite observing significant differences in the length and conformational plasticity of the various intercohesin linkers by small-angle X-ray scattering, all of the engineered bi-enzyme complexes had similar activities on crystalline cellulose, suggesting that cellulosomal interdomain linkers had little functional impact [21]. In contrast, a methodical analysis of the influence of scaffoldin modular organization on cellulose hydrolysis illustrated that longer intercohesin linkers are favored in the design of efficient engineered cellulosomes [48].

Another study involving engineered cohesin-containing miniscaffoldins revealed an unexpected structural and functional influence on cellulosome assembly [49\*\*]. In this work, the scaffoldin-borne cohesins possessed similar or identical specificities for the same type of dockerin on the various enzymes examined. Designed to assess the heterogeneity and random assembly of cellulosomal enzymes onto the scaffoldin subunit, these divalent and trivalent miniscaffoldins comprised interdomain linkers between 11 and 50 amino acid residues in length. Rather than observing random association of the enzymatic subunits into the



scaffoldins there appeared to be favored integration, where the presence of one type of enzyme on the scaffoldin influenced the type of enzyme that was subsequently integrated into the complex [49\*\*]. Notably, the length of the scaffoldin linker regions connecting adjacent cohesin modules directly influenced this preferential assimilation. This relationship, at least in part, ensures the assembly of enzymatic subunits of varying sizes and structural complexity, and with complementary enzymatic activities, into a functional cellulosome.

## **Conclusions**

Collectively, these recent findings highlight the importance of the diversity and heterogeneity of cellulosomes, which are achieved through 'variations of a theme' at the protein level, but also regulated at the gene transcription level. The structural heterogeneity also forms the basis of efficient synergy amongst the cellulosomal components within the macromolecular nanomachine. The more traditional view of cellulosomes as optimized machineries that evolved solely for dedicated breakdown of cellulose and associated polysaccharides is undergoing transformation. In this context, the complexes can be considered in a broader sense as plant cell wall polysaccharide-degrading factories in which each component is recruited to fulfill, a given task that augments the overall synergy of action [35]. Ongoing complementary structural, biochemical, biophysical, and computational information of novel cellulosomal components has thus contributed to a greater understanding of their combined influence on cellulosome assembly, dynamics, and function. Advanced knowledge of the involvement of these novel components will help perfect the design and generation of non-natural, engineered designer cellulosomes with maximal cellulolytic efficiency.

## Acknowledgements

This work was supported by a Natural Science and Engineering Research Council of Canada discovery grant to SPS (RGPIN 2015-06667), by an Israel Science Foundation grant to EAB (No. 1349), and by an ADEME ERA-NET IB grant to MC (No. 1201C0104). Additional support to EAB and MC was provided by the European Union, Area NMP.2013.1.1-2: Self-assembly of naturally occurring nanosystems: CellulosomePlus Project number: 604530. EAB is the incumbent of the Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.

## References and recommended reading

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

- of special interest
- of outstanding interest

1. Bjursell MK, Martens EC, Gordon JI: **Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period.** *J Biol Chem* 2006, **281**:36269-36279.
2. Wilson DB: **Microbial diversity of cellulose hydrolysis.** *Curr Opin Microbiol* 2011, **14**:259-263.
3. Bayer EA, Belaich JP, Shoham Y, Lamed R: **The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides.** *Annu Rev Microbiol* 2004, **58**:521-554.
4. Fontes CM, Gilbert HJ: **Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates.** *Annu Rev Biochem* 2010, **79**:655-681.
5. Smith SP, Bayer EA: **Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex.** *Curr Opin Struct Biol* 2013, **23**:686-694.
6. Carvalho AL, Dias FM, Prates JA, Nagy T, Gilbert HJ, Davies GJ, Ferreira LM, Romao MJ, Fontes CM: **Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex.** *Proc Natl Acad Sci U S A* 2003, **100**:13809-13814.

7. Blouzard JC, Coutinho PM, Fierobe HP, Henrissat B, Lignon S, Tardif C, Pages S, de Philip P: **Modulation of cellulosome composition in *Clostridium cellulolyticum*: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses.** *Proteomics* 2010, **10**:541-554.
8. Lamed R, Setter E, Bayer EA: **Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*.** *J Bacteriol* 1983, **156**:828-836.
9. Morag E, Bayer EA, Lamed R: **Relationship of cellulosomal and noncellulosomal xylanases of *Clostridium thermocellum* to cellulose-degrading enzymes.** *J Bacteriol* 1990, **172**:6098-6105.
10. Chakraborty S, Fernandes VO, Dias FM, Prates JA, Ferreira LM, Fontes CM,
  - Goyal A, Centeno MS: **Role of pectinolytic enzymes identified in *Clostridium thermocellum* cellulosome.** *PLoS One* 2015, **10**:e0116787.
 The biochemical characterization of three dockerin-bearing pectate lyases from *C. thermocellum* is reported, in the context of their action within the cellulosomal complex. These proteins not only contain catalytic domains and dockerins, but two of them also comprise CBMs, which is unusual for pectate lyases.
11. Artzi L, Morag E, Shamshoum M, Bayer EA: **Cellulosomal expansin:
 
  - functionality and incorporation into the complex.** *Biotechnol Biofuels* 2016, **9**:61.

 These experimental analyses explore the functional characteristics of a cellulosomal expansin from *C. clariflavum*. In particular, the cellulose-disruption activity of the expansin, its substrate-binding characteristics, the binding specificity of its associated dockerin module, and its impact on the degradation of cellulosic substrates by partner enzymes or cellulosomes are described. The data suggest that the mechanism of action of this type of protein is highly complicated, and its contribution to biomass degradation is largely multifactorial.
 
12. Chen C, Cui Z, Song X, Liu YJ, Cui Q, Feng Y: **Integration of bacterial
 
  - expansin-like proteins into cellulosome promotes the cellulose degradation.** *Appl Microbiol Biotechnol* 2016, **100**:2203-2212.

 This work describes the ability of two expansin-like proteins from *C. clariflavum*, Clocl\_1298 and Clocl\_1862, to synergistically enhance cellulose degradation by native and designer cellulosomes.
 
13. Levy-Assaraf M, Voronov-Goldman M, Rozman Grinberg I, Weiserman G, Shimon LJ, Jindou S, Borovok I, White BA, Bayer EA, Lamed R, et al.: **Crystal structure of an uncommon cellulosome-related protein module from *Ruminococcus flavefaciens* that resembles papain-like cysteine peptidases.** *PLoS One* 2013, **8**:e56138.
14. O Cuív C, Gupta R, Goswami HP, Morrison M: **Extending the cellulosome paradigm: the modular *Clostridium thermocellum* cellulosomal serpin PinA**

**is a broad-spectrum inhibitor of subtilisin-like proteases.** *Appl Environ Microbiol* 2013, **79**:6173-6175.

15. Arfi Y, Shamshoum M, Rogachev I, Peleg Y, Bayer EA: **Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation.** *Proc Natl Acad Sci U S A* 2014, **111**:9109-9114.

In this study, the self-assembly and integration into designer cellulosomes of dockerin-fused lytic polysaccharide monooxygenases (LPMOs) is investigated. The chimeric proteins were demonstrated to be functional, and enhanced cellulolytic activity was achieved when these enzymes were incorporated alongside with endo- and exo-cellulases, showing that the benefit of a proximity effect can be extended to oxidative enzymes within a designer cellulosome complex.

16. Fendri I, Tardif C, Fierobe HP, Lignon S, Valette O, Pages S, Perret S: **The cellulosomes from *Clostridium cellulolyticum*: identification of new components and synergies between complexes.** *FEBS J* 2009, **276**:3076-3086.
17. Fierobe HP, Mingardon F, Mechaly A, Belaich A, Rincon MT, Pages S, Lamed R, Tardif C, Belaich JP, Bayer EA: **Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin.** *J Biol Chem* 2005, **280**:16325-16334.
18. Stern J, Morais S, Lamed R, Bayer EA: **Adaptor scaffoldins: An original strategy for extended designer cellulosomes, inspired from Nature.** *MBio* 2016, **7**:e00083.
19. Vazana Y, Morais S, Barak Y, Lamed R, Bayer EA: **Designer cellulosomes for enhanced hydrolysis of cellulosic substrates.** *Methods Enzymol* 2012, **510**:429-452.
20. Bayer EA, Smith SP, Noach I, Alber O, Adams JJ, Lamed R, Shimon LJ, Frolow F: **Can we crystallize a cellulosome?** In *Biotechnology of lignocellulose degradation and biomass utilization*. Edited by Saaka K, Karita S, Kimura T, Saaka M, Matsui H, Miyake H, Tanaka A: Ito Print Publishing; 2009:183-205.
21. Molinier AL, Nouailler M, Valette O, Tardif C, Receveur-Brechot V, Fierobe HP: **Synergy, structure and conformational flexibility of hybrid cellulosomes displaying various inter-cohesins linkers.** *J Mol Biol* 2011, **405**:143-157.
22. Nash MA, Smith SP, Fontes CM, Bayer EA: **Single versus dual-binding conformations in cellulosomal cohesin-dockerin complexes.** *Curr Opin Struct Biol* 2016, **40**:89-96.
23. Xu C, Huang R, Teng L, Jing X, Hu J, Cui G, Wang Y, Cui Q, Xu J: **Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA processing and stabilization.** *Nat Commun* 2015, **6**:6900.

The study unravels the role of RNA structures on the regulation of cellulosomal function and composition, suggesting that the evolution of cellulosomes is also influenced by the number, position, and nature of such genomic elements. These findings have general implications in designing and engineering protein complex assemblies.

24. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B: **The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics**. *Nucleic Acids Res* 2009, **37**:D233-238.
25. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active enzymes database (CAZy) in 2013**. *Nucleic Acids Res* 2014, **42**:D490-495.
26. Irwin D, Shin DH, Zhang S, Barr BK, Sakon J, Karplus PA, Wilson DB: **Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis**. *J Bacteriol* 1998, **180**:1709-1714.
27. Jindou S, Xu Q, Kenig R, Shulman M, Shoham Y, Bayer EA, Lamed R: **Novel architecture of family-9 glycoside hydrolases identified in cellulosomal enzymes of *Acetivibrio cellulolyticus* and *Clostridium thermocellum***. *FEMS Microbiol Lett* 2006, **254**:308-316.
28. Petkun S, Rozman Grinberg I, Lamed R, Jindou S, Burstein T, Yaniv O, Shoham Y, Shimon LJ, Bayer EA, Frolow F: **Reassembly and co-crystallization of a family 9 processive endoglucanase from its component parts: structural and functional significance of the intermodular linker**. *PeerJ* 2015, **3**:e1126.
29. Sakon J, Irwin D, Wilson DB, Karplus PA: **Structure and mechanism of endo/exocellulase E4 from *Thermomonospora fusca***. *Nat Struct Biol* 1997, **4**:810-818.
30. Ravachol J, Borne R, Tardif C, de Philip P, Fierobe HP: **Characterization of all family-9 glycoside hydrolases synthesized by the cellulosome-producing bacterium *Clostridium cellulolyticum***. *J Biol Chem* 2014, **289**:7335-7348.  
This study provides detailed structure-based, biochemical analyses of the substrate specificities of all thirteen family GH9 enzyme-encoding genes of *C. cellulolyticum*, a cellulosome-producing anaerobe. Systematic characterization of these enzymes not only showed that the novel cellulases display different activity patterns, but also identified Cel9X as a genuine endoxyloglucanase, an activity that previously was only considered as a side activity of some rare members of this family.
31. Aspeborg H, Coutinho PM, Wang Y, Brumer H, 3rd, Henrissat B: **Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5)**. *BMC Evol Biol* 2012, **12**:186.
32. Franche N, Tardif C, Ravachol J, Harchouni S, Ferdinand PH, Borne R, Fierobe HP, Perret S: **Cel5I, a SLH-containing glycoside hydrolase: Characterization**

**and investigation on its role in *Ruminiclostridium cellulolyticum*. PLoS One** 2016, **11**:e0160812.

33. Bahari L, Gilad Y, Borovok I, Kahel-Raifer H, Dassa B, Nataf Y, Shoham Y, Lamed R, Bayer EA: **Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. J Ind Microbiol Biotechnol** 2011, **38**:825-832.
34. Kosugi A, Murashima K, Tamaru Y, Doi RH: **Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. J Bacteriol** 2002, **184**:884-888.
35. Gunnoo M, Cazade PA, Galera-Prat A, Nash MA, Czjzek M, Cieplak M, Alvarez B, Aguilar M, Karpol A, Gaub H, et al.: **Nanoscale engineering of designer cellulosomes. Adv Mater** 2016, **28**:5619-5647.
36. Rozycki B, Cieplak M: **Stiffness of the C-terminal disordered linker affects the geometry of the active site in endoglucanase Cel8A. Mol Biosyst** 2016, **12**:3589-3599.

Coarse-grained and all-atom modeling was used in this study to investigate the influence of the disordered C-terminal linker on the structural conformations of the connected catalytic domain. The study elegantly shows that a change in rigidity of the linker region, for example through binding to a cognate cohesin, does impact topology of the active site, thus suggesting an effect on enzymatic activity.
37. Rozycki B, Cieplak M, Czjzek M: **Large conformational fluctuations of the multi-domain xylanase Z of *Clostridium thermocellum*. J Struct Biol** 2015, **191**:68-75.
38. Qian M, Guan S, Shan Y, Zhang H, Wang S: **Structural and molecular basis of cellulase Cel48F by computational modeling: Insight into catalytic and product release mechanism. J Struct Biol** 2016, **194**:347-356.
39. Shang BZ, Chang R, Chu JW: **Systems-level modeling with molecular resolution elucidates the rate-limiting mechanisms of cellulose decomposition by cellobiohydrolases. J Biol Chem** 2013, **288**:29081-29089.
40. Artzi L, Morag E, Barak Y, Lamed R, Bayer EA: ***Clostridium clariflavum*: Key cellulosome players are revealed by proteomic analysis. MBio** 2015, **6**:e00411-00415.

This work describes the comprehensive mass spectrometry-based proteomic and biochemical analysis of native cellulosomes produced by *C. clariflavum*. Notable findings included eleven scaffoldins identified, substrate-specific expression of cellulosomal proteins, and that cellulosomes derived from *C. clariflavum* grown on microcrystalline cellulose medium were the most active of those tested.
41. Cosgrove DJ: **Plant expansins: diversity and interactions with plant cell walls. Curr Opin Plant Biol** 2015, **25**:162-172.

42. Bayer EA, Morag E, Lamed R: **The cellulosome--a treasure-trove for biotechnology.** *Trends Biotechnol* 1994, **12**:379-386.
43. Fierobe HP, Mechaly A, Tardif C, Belaich A, Lamed R, Shoham Y, Belaich JP, Bayer EA: **Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes.** *J Biol Chem* 2001, **276**:21257-21261.
44. Morais S, Morag E, Barak Y, Goldman D, Hadar Y, Lamed R, Shoham Y, Wilson DB, Bayer EA: **Deconstruction of lignocellulose into soluble sugars by native and designer cellulosomes.** *MBio* 2012, **3**.
45. Davidi L, Morais S, Artzi L, Knop D, Hadar Y, Arfi Y, Bayer EA: **Toward**  
 •• **combined delignification and saccharification of wheat straw by a laccase-containing designer cellulosome.** *Proc Natl Acad Sci U S A* 2016, **113**:10854-10859.  
 This study describes the first reported incorporation of a lignin-active enzyme into a designer cellulosome system. Addition of a dockerin-containing laccase variant originating from the aerobic bacterium *Thermobifida fusca* into a scaffoldin chimera comprising complementary cellulases enhanced wheat straw polysaccharide degradation two-fold over the same designer cellulosomes lacking the laccase.
46. Fierobe HP, Bayer EA, Tardif C, Czjzek M, Mechaly A, Belaich A, Lamed R, Shoham Y, Belaich JP: **Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components.** *J Biol Chem* 2002, **277**:49621-49630.
47. Venditto I, Luis AS, Rydahl M, Schuckel J, Fernandes VO, Vidal-Melgosa S, Bule  
 •• P, Goyal A, Pires VM, Dourado CG, et al.: **Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition.** *Proc Natl Acad Sci U S A* 2016, **113**:7136-7141.  
 This study uncovers six previously unidentified carbohydrate-binding module (CBM) families that bind  $\beta$ -glucans,  $\beta$ -mannans, and homogalacturonan, four of which were structurally characterized.
48. Vazana Y, Barak Y, Unger T, Peleg Y, Shamshoum M, Ben-Yehezkel T, Mazor Y, Shapiro E, Lamed R, Bayer EA: **A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates.** *Biotechnol Biofuels* 2013, **6**:182.
49. Borne R, Bayer EA, Pages S, Perret S, Fierobe HP: **Unraveling enzyme**  
 •• **discrimination during cellulosome assembly independent of cohesin-dockerin affinity.** *FEBS J* 2013, **280**:5764-5779.  
 This work employed designer cellulosomes to assess the extent of heterogeneity and randomness in the assembly of enzymatic subunits onto a cellulosomal scaffoldin. Resident enzymatic subunits caused the preferential integration of additional cellulosomal components onto a scaffoldin, a mechanism also influenced by the size of the inter-cohesin linker regions.

50. Alzari PM, Souchon H, Dominguez R: **The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*.** *Structure* 1996, **4**:265-275.
51. Garcia-Alvarez B, Melero R, Dias FM, Prates JA, Fontes CM, Smith SP, Romao MJ, Carvalho AL, Llorca O: **Molecular architecture and structural transitions of a *Clostridium thermocellum* mini-cellulosome.** *J Mol Biol* 2011, **407**:571-580.
52. Adams JJ, Pal G, Jia Z, Smith SP: **Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex.** *Proc Natl Acad Sci U S A* 2006, **103**:305-310.
53. Tormo J, Lamed R, Chirino AJ, Morag E, Bayer EA, Shoham Y, Steitz TA: **Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose.** *EMBO J* 1996, **15**:5739-5751.
54. Yaniv O, Shimon LJ, Bayer EA, Lamed R, Frolow F: **Scaffoldin-borne family 3b carbohydrate-binding module from the cellulosome of *Bacteroides cellulosolvens*: structural diversity and significance of calcium for carbohydrate binding.** *Acta Crystallogr D Biol Crystallogr* 2011, **67**:506-515.
55. Czjzek M, Fierobe HP, Receveur-Brechot V: **Small-angle X-ray scattering and crystallography: a winning combination for exploring the multimodular organization of cellulolytic macromolecular complexes.** *Methods Enzymol* 2012, **510**:183-210.
56. Chen C, Cui Z, Xiao Y, Cui Q, Smith SP, Lamed R, Bayer EA, Feng Y: **Revisiting the NMR solution structure of the Cel48S type-I dockerin module from *Clostridium thermocellum* reveals a cohesin-primed conformation.** *J Struct Biol* 2014, **188**:188-193.
57. Mandelman D, Belaich A, Belaich JP, Aghajari N, Driguez H, Haser R: **X-ray crystal structure of the multidomain endoglucanase Cel9G from *Clostridium cellulolyticum* complexed with natural and synthetic cello-oligosaccharides.** *J Bacteriol* 2003, **185**:4127-4135.
58. Parsieglia G, Belaich A, Belaich JP, Haser R: **Crystal structure of the cellulase Cel9M enlightens structure/function relationships of the variable catalytic modules in glycoside hydrolases.** *Biochemistry* 2002, **41**:11134-11142.
59. Schubot FD, Kataeva IA, Blum DL, Shah AK, Ljungdahl LG, Rose JP, Wang BC: **Structural basis for the exocellulase activity of the cellobiohydrolase CbhA from *Clostridium thermocellum*.** *Biochemistry* 2004, **43**:1163-1170.
60. Chitarra V, Souchon H, Spinelli S, Juy M, Beguin P, Alzari PM: **Multiple crystal forms of endoglucanase CelD: signal peptide residues modulate lattice formation.** *J Mol Biol* 1995, **248**:225-232.



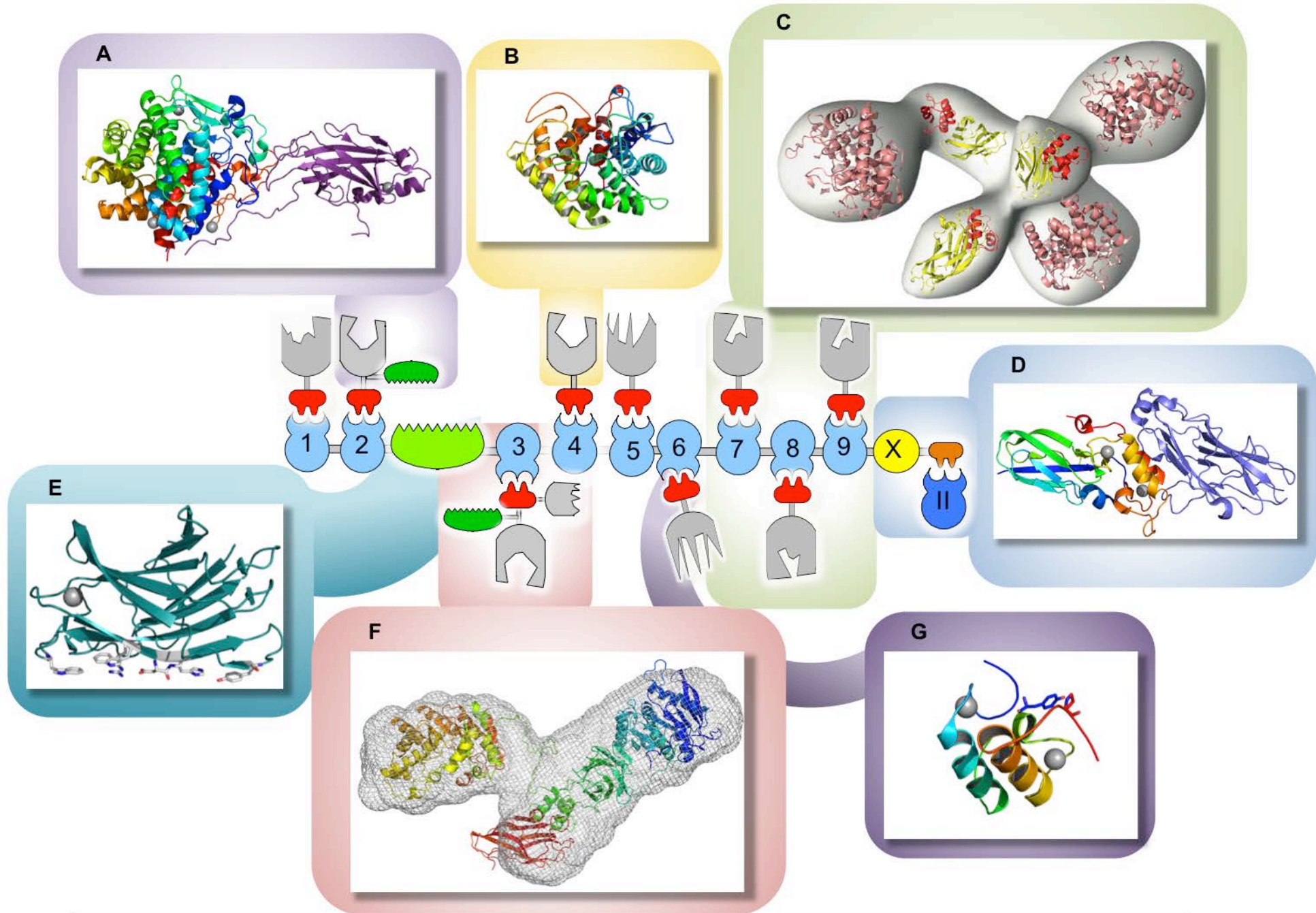
61. Georgelis N, Yennawar HN, Cosgrove DJ: **Structural basis for entropy driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin.** *Proc Natl Acad Sci U S A* 2012, **109**:14380-14385.
62. Pinheiro BA, Proctor MR, Martinez-Fietes C, Prates JAM, Money VA, Davies GJ, Bayer EA, Fontes CMGA, Fierobe HP, Gilbert HJ: **The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner.** *J Biol Chem* **283**: 18422 -18430.
63. Carvalho AL, Dias FMV, Nagy T, Prates JAM, Proctor MR, Smith SP, Bayer EA, Davies GJ, Ferreira LMA, Romao MJ, Fontes CMGA, Gilbert HJ: **Evidence for a dual binding mode of dockerin modules to cohesins.** *Proc Natl Acad Sci U S A* **104**: 3089-3094.
64. Salama-Alber O, Jobby MK, Chitayat S, Smith SP, White BA, Shimon LJ, Lamed R, Frolow F, Bayer EA: **Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: Binding fidelity, promiscuity, and structural buttresses.** *J Biol Chem* **288**: 16827-16838.

## Figure Legends

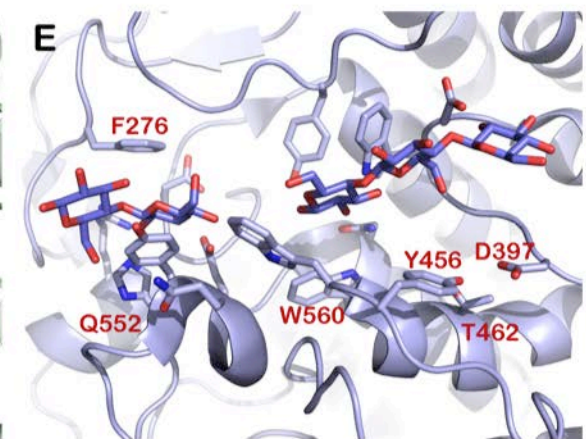
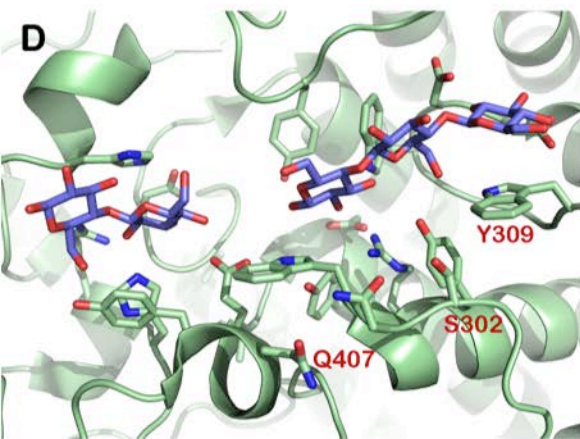
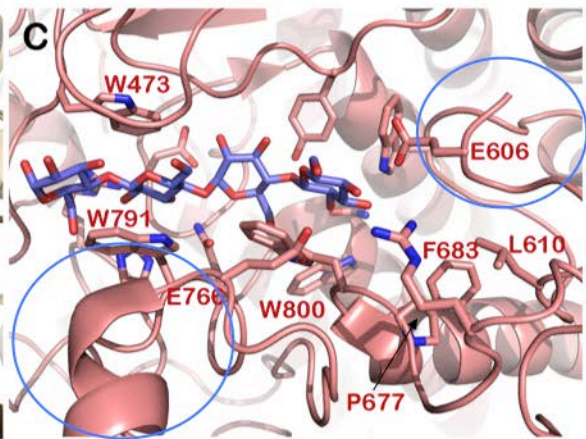
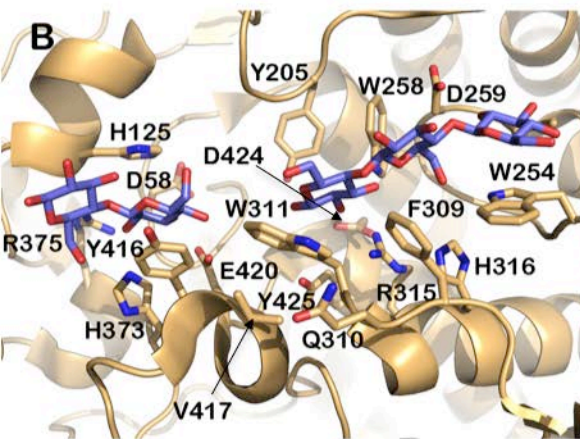
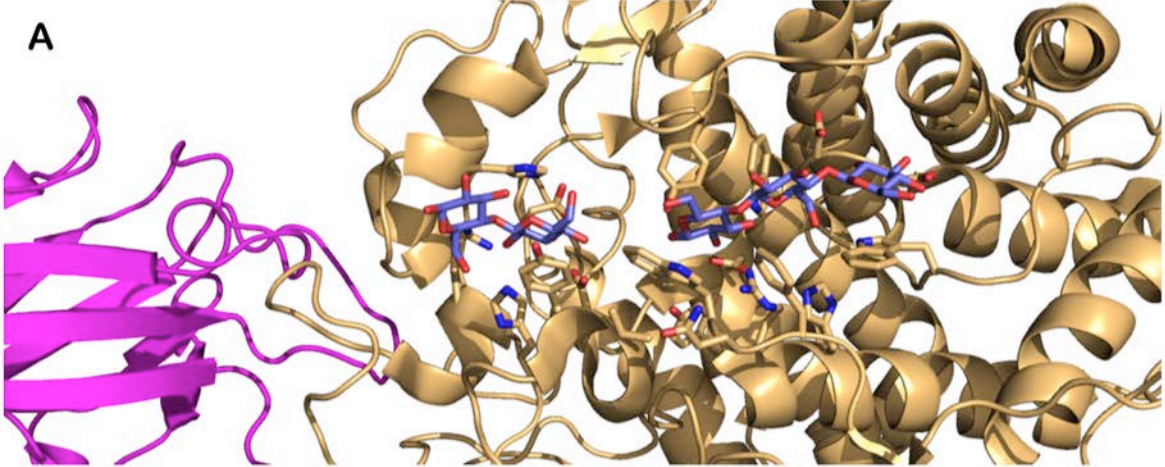
**Figure 1.** Architecture of a cellulosome from *Clostridium (Ruminiclostridium) thermocellum*. Three-dimensional structures of various modular components of the cellulosome, determined by NMR spectroscopy, small-angle X-ray scattering, or X-ray crystallography, are shown and shaded according their position in the central schematic cellulosome representation. Specifically, (A) The X-ray crystal structure of the processive endoglucanase Cel9I, where the N-terminal catalytic module is rainbow coloured and the C-terminal CBM3c is illustrated in purple. (PDB-code 2XFG; [28]). (B) The first X-ray crystal structure of a  $\beta$ -endoglucanase, the cellulosomal cellulase Cel8A (PDB-code 1CEM; [50]). (C) A cellulosomal fragment comprising the three C-terminal scaffoldin-based cohesin modules (yellow) in complex with Cel8A, where the type-I dockerin and catalytic modules are coloured red and pink, respectively (cryo-electron microscopy and X-ray crystal structure, EMD-code EMD-1823, PDB-codes 2CCL, 1CEM; [51]). (D) The X-ray crystal structure of scaffoldin X module/type-II dockerin dyad (rainbow-coloured) in complex with a cell surface subunit type-II cohesin module in blue (PDB-code 2B59; [52]). (E) X-ray crystal structure of the scaffoldin-bourne CBM3a module (PDB-codes 1NBC, 4B9F; [53,54]). (F). Small angle X-ray scattering (SAXS) envelope superimposed with a structural model of the full length enzyme Xyn10Z (N-terminal CE1 module (blue), a CBM6 module (green), a type-I dockerin module (light green) and a C-terminal GH10 xylanase (orange/yellow) in complex with a scaffoldin-bourne type-I cohesin module (red) (PDB-codes 1JJF, 1GMM, 1XYZ, 1OHZ; [55]). (G) NMR-derived structure of an isolated type-I dockerin module (PDB-code 2MTE; [56]).

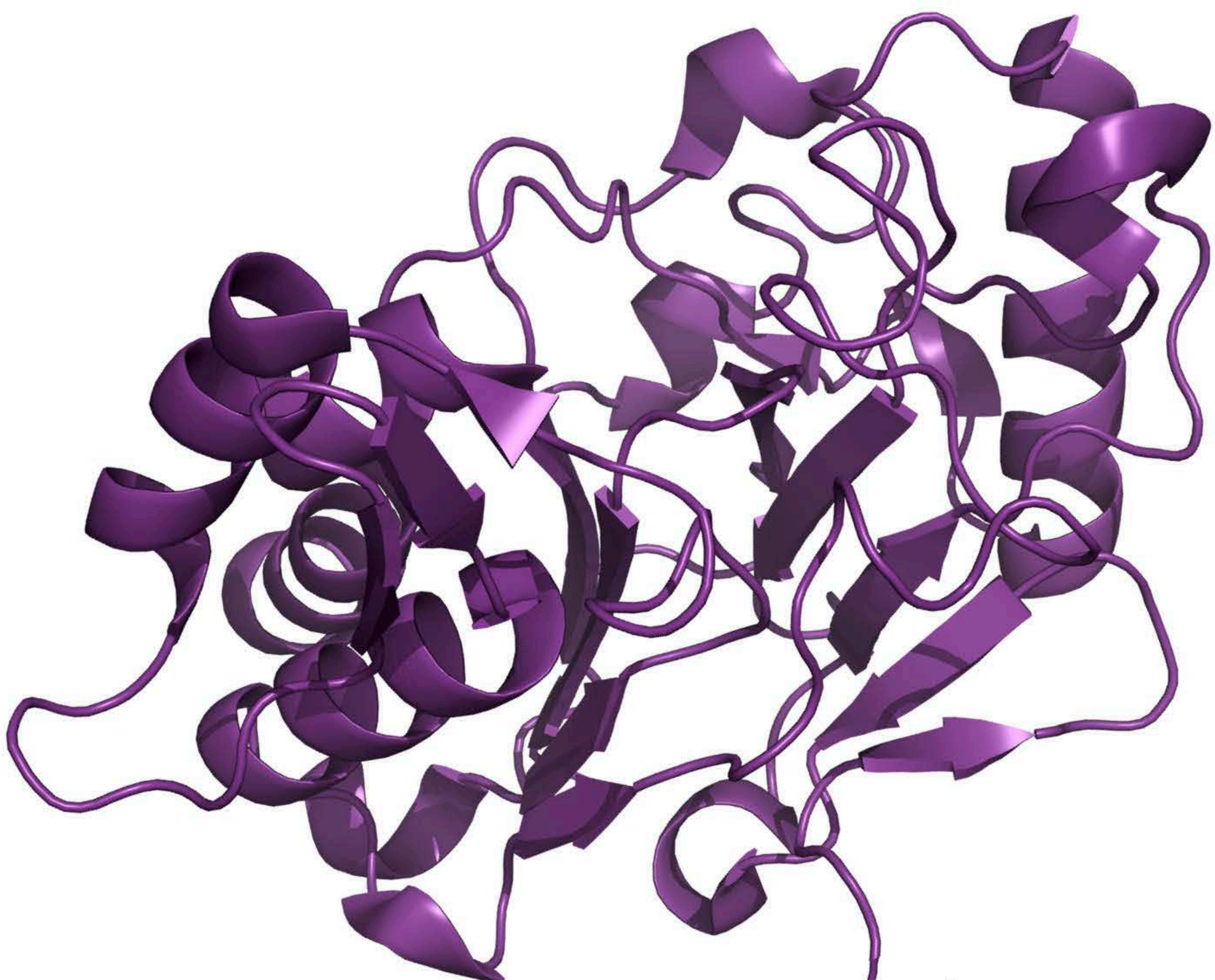
**Figure 2.** Structures of GH9 enzymes that reflect the observed biochemical variability. (A) Ribbon representation of full length *C. cellulolyticum* Cel9G in complex with  $\beta$ -1,4-linked cellodextrin substrate molecules (carbon atoms in purple, oxygen atoms in red). The N-terminal catalytic module is colored in light orange and the C-terminal CBM3c module in magenta, highlighting the continuation of the catalytic binding site through the appended CBM (PDB-code 1GA2; [57]). (B) Expanded view of the catalytic active site of Cel9G (endoglucanase or cellulase), where residues interacting with the bound substrate molecules are labeled by single-letter amino acid code and position in sequence. (C) Expanded view of the catalytic active site of the *C. thermocellum* GH9 exocellulase CbhA in salmon, in complex with a cellotetraose substrate molecule (PDB-code 1RQ5; [59]). Only those amino-acid residues that are different from *C. cellulolyticum* Cel9G are labeled in red. Two secondary-structural elements (blue circles) are highlighted that seal one end of the active-site cleft, changing the mode of action of this GH9 enzyme. (D) Expanded view of the catalytic active site of *C. cellulolyticum* Cel9M endoglucanase in light green (PDB-code 1IA6; [58]). Only those amino-acid residues that are different from Cel9G are labeled in red. (E) Close-up view of the catalytic active site of *C. thermocellum* Cel9D endoglucanase in light blue (PDB-code 4CJ0; [60]). Only those amino-acid residues that are different from *C. cellulolyticum* Cel9G are labeled in red. The substrate molecules were positioned in the active sites of Cel9M (D) and Cel9D (E) by superimposition. The differences in substrate specificity of these very subtle structural differences observed in the active sites of these enzymes are difficult to predict without precise biochemical characterizations. GH9 enzymes very often occur as multigenic families within cellulosome containing bacterial genomes, such as *Clostridium thermocellum* and *Clostridium cellulolyticum*, for example.

**Figure 3.** Schematic representation of the incorporation of novel enzymatic and non-enzymatic functionalities into designer cellulosomes. Backbone ribbon representations of a structural homolog of the Tfu1114 laccase from *Thermobifida fusca* (purple; PDB-code 1T8H), the E7 lytic polysaccharide monooxygenase (LPMO) from *Thermobifida fusca* (dark red; PDB-code 5UIZ), and a structural homolog of the CclEXL1 expansin from *C. clariflavum* (blue; PDB-code 4FG4; [61]) are tethered to type-I dockerin modules from *C. cellulolyticum* (Type I doc\_Cc; purple), *C. thermocellum* (Type I doc\_Ct; dark red), and the type-III dockerin module from *R. flavefaciens* (Type III doc\_Rf; blue), respectively. The dockerin modules are bound to their cognate cohesin modules (light pink, yellow, and green surface representations; PDB-codes 2VN5, 2CCL, 4IU3; [62-64]), which comprise a chimeric scaffoldin.

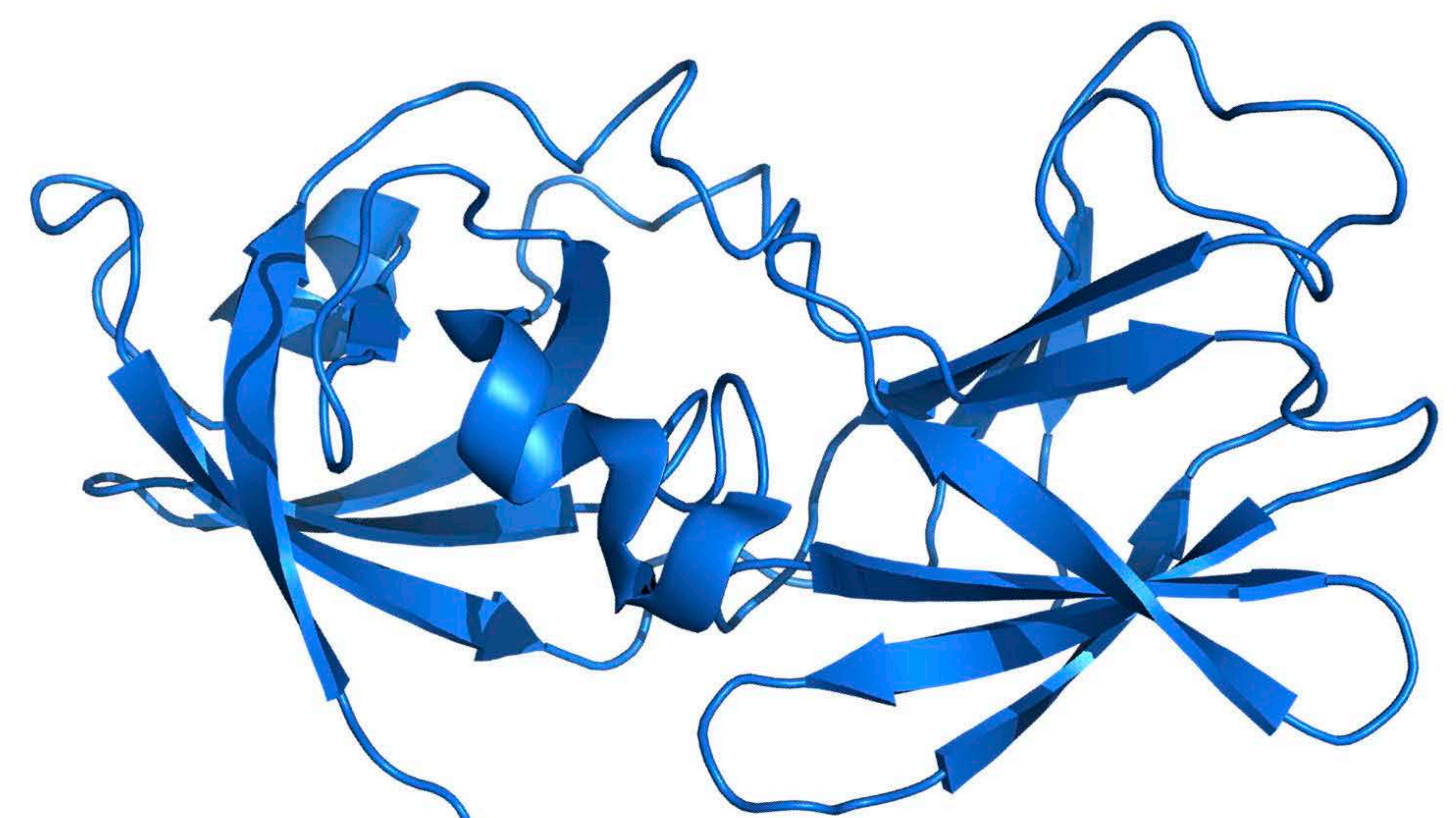


 Carbohydrate Binding Modules CBMs
  X Domain
  Cohesins
  Dockerins
  Catalytic Modules

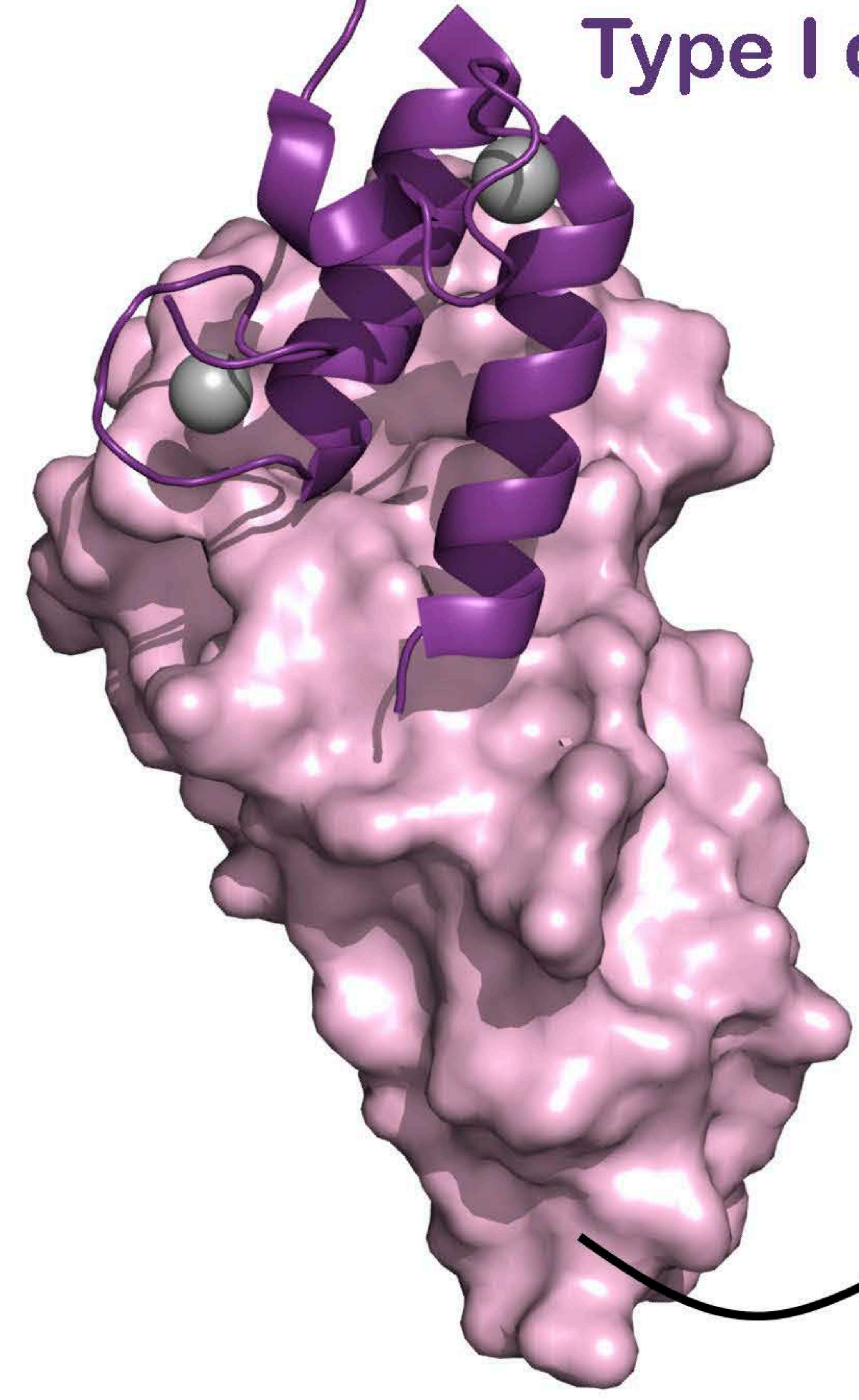




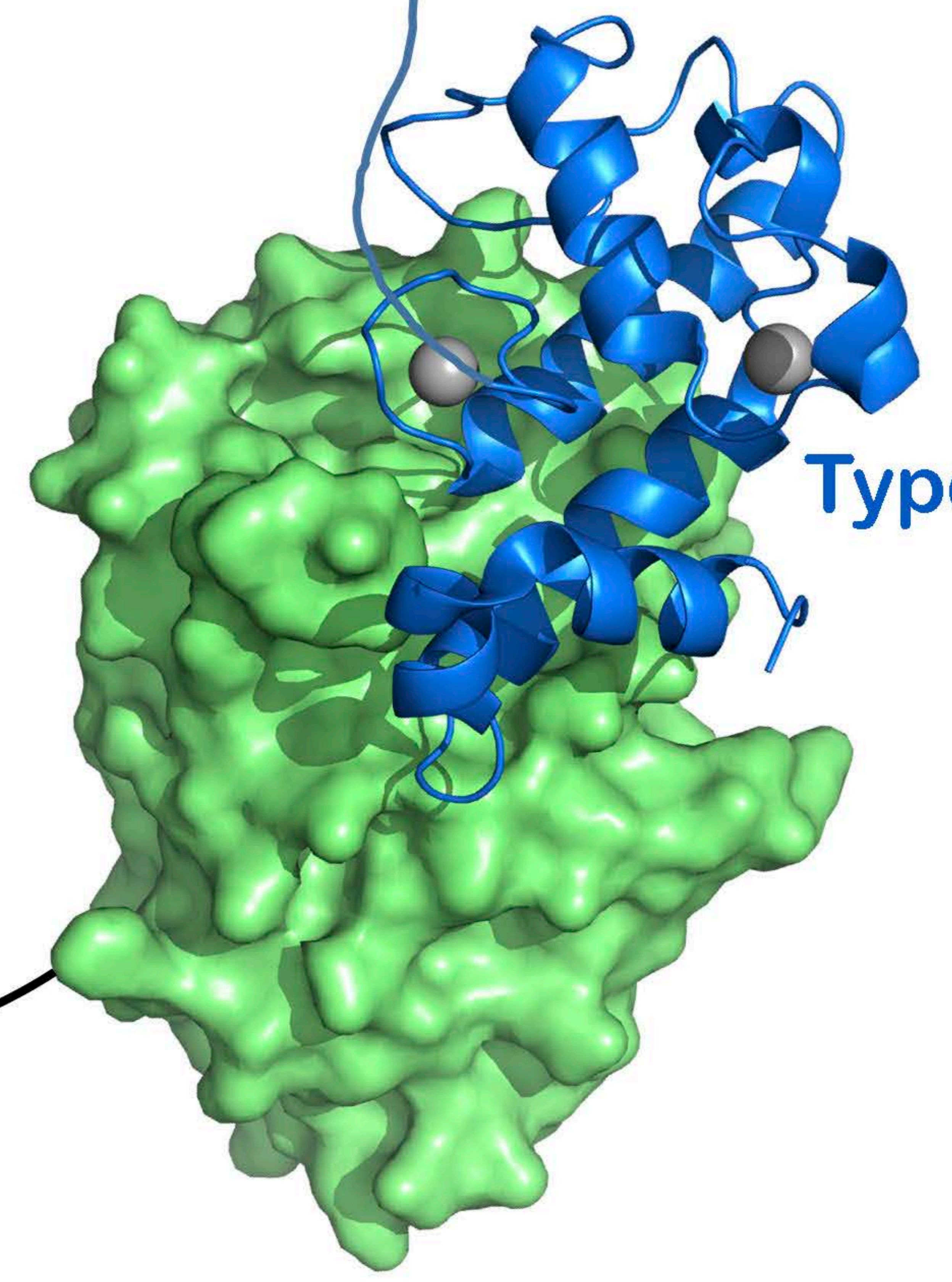
Laccase



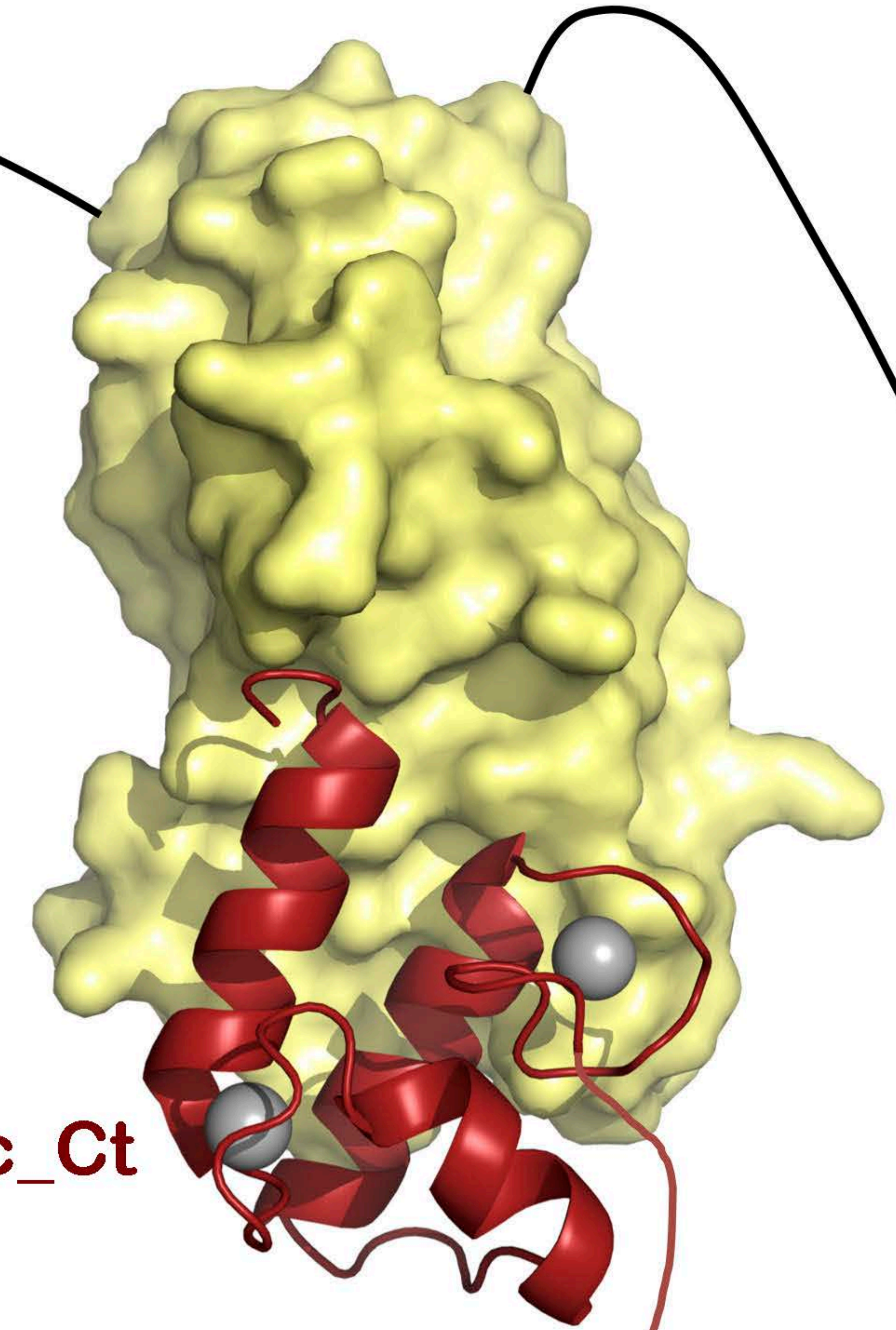
Expansin



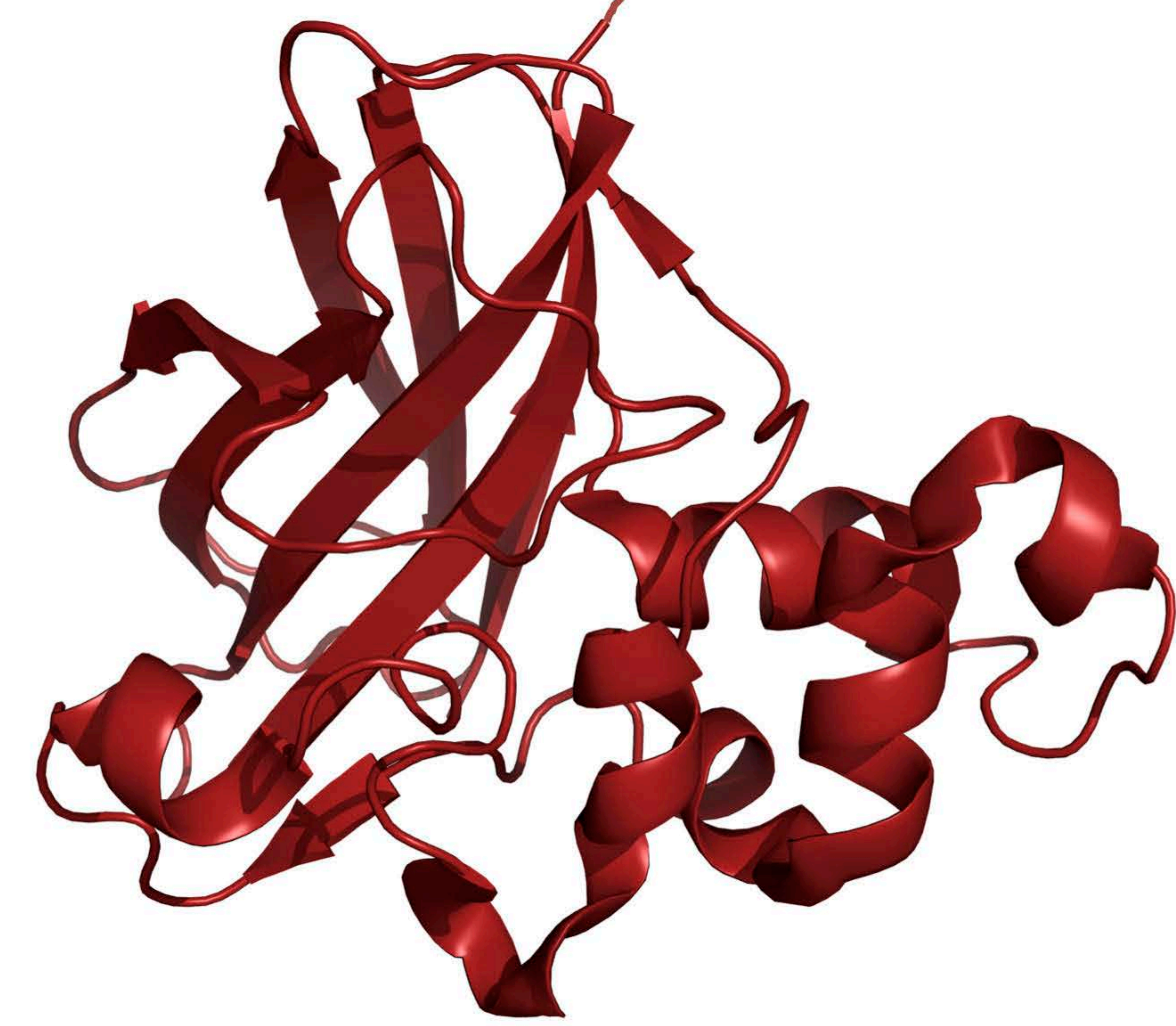
Type I doc\_Cc



Type III doc\_Rf



Type I doc\_Ct



LPMO

