Modular protein architectures for pH-dependent interactions and switchable assembly of nanocellulose

Sanni Voutilainen\textsuperscript{1,2}, Arja Paananen\textsuperscript{2}, Martina Lille\textsuperscript{2}, and Markus B. Linder\textsuperscript{1}

1. Department of Bioproducts and Biosystems, School of Chemical Engineering, Aalto University, Box 16100, 00076-Aalto, Espoo, Finland.
2. VTT Technical Research Centre of Finland Ltd., P.O. Box 1000, 02044 VTT Espoo, Finland

Corresponding author: Markus Linder, Department of Bioproducts and Biosystems, School of Chemical Engineering, Aalto University, Box 16100, 00076-Aalto, Espoo, Finland. markus.linder@aalto.fi.

Abstract
Protein engineering shows a wide range of possibilities for designing properties in novel materials. Following inspiration from natural systems we have studied how combinations or duplications of protein modules can be used to engineer their interactions and achieve functional properties. Here we used cellulose binding modules (CBM) coupled to spider silk N-terminal domains that dimerize in a pH-sensitive manner. We showed how the pH-sensitive switching into dimers affected cellulose binding affinity in relation to covalent coupling between CBMs. Finally, we showed how the pH-sensitive coupling could be used to assemble cellulose nanofibers in a dynamic pH-dependent way. The work shows how novel proteins can be designed by linking functional domains from widely different sources and thereby achieve new functions in the self-assembly of nanoscale materials.

Introduction
Biological materials are gaining attention among material scientists due to the wide range of possibilities for tuning their properties and integrating functionality [1–3]. The three main groups of biopolymers, i.e. proteins, polynucleotides, and carbohydrates, all have their own characteristics which are advantageous in different contexts. Carbohydrates, as for example cellulose and chitin are widely available and have excellent mechanical properties [4,5]. Polynucleotides (RNA and DNA) are superior in the storage and retrieval of information, and can be used for programming intricate structures in what is known as DNA-origami [6]. Proteins show a high versatility in binding interactions, molecular recognition, and switchable functions. One of the characteristics of proteins is how they often have evolved in natural systems by duplicating parts and recombining functionalities in a modular way [7]. Structural or specific binding elements are combined with
molecular switches to create overall architectures that can perform complicated functions. Evolution has finetuned the interplay of the functional elements into astoundingly well-performing systems [8].

In this work we created proteins with novel functional architectures by combining the high affinity cellulose binding of proteins called cellulose binding modules (CBMs) and the reversibly pH-dependent dimerization function of the N-terminal parts of spidroin proteins. The new fusion-proteins were then used to make reversibly pH-switching assemblies with cellulose. The type of cellulose used in this study was a highly dispersed form, called cellulose nanofibrils (CNF).

CBMs are non-catalytic protein modules that are found as parts of enzymes acting on lignocellulosic substrates, such as glycoside hydrolases. The CBMs have a biological function to bind specifically to cellulose. Often the CBMs are connected to the catalytically active parts through extended linkers. The binding specificity of a binding module is determined by the binding site topology. The majority of CBMs that bind crystalline cellulose have flat hydrophobic binding faces containing aromatic residues [9–11]. CBMs can be produced by recombinant DNA technology as independent domains fused to other proteins and have found uses in applications such as immobilization of antibodies [12] or enzymes [13] to cellulose, and even as components of protein-based adhesives for cellulose [14].

Spider dragline silk that is produced by orb-weaving spiders is a unique fibrous material composed of proteins called spidroins. Spidroins have complex structures, consisting of three functionally and structurally different sections. The core of the spidroin contains highly repetitive sequences of polyAla stretches that are interrupted by Gly, Pro, Gln-rich regions. This part comprises the bulk of the spidroin and forms extensive β-sheet structures in the final assembled material [15]. At both ends of the repetitive core the spidroin has two conserved globular modules, one called the N-terminal module and the other the C-terminal module. The terminal modules are not structurally related and they have different tasks [16]. Both N- and C-terminal modules have a role when spidroin protein molecules are being assembled to form silk fibers. The C-terminal module forms covalent dimers and has been proposed to trigger the linkage of spidroins by unfolding into a β-sheet amyloid form [17].

The N-terminal module forms a homodimer and firmly interconnects the spidroins in a pH-dependent way when the pH in the spider silk gland is lowered from pH 7 to pH 6 – 5.5. The 3D structure of the N-terminal module of major ampullate spidroin from the nursery web spider *Euprosthenops australis* has been solved [18]. The function of this NT module has been extensively studied through mutagenesis and the amino acids involved in the dimerization have been identified [19].
CNF is a form of cellulose that is typically obtained by disintegrating wood pulp by grinding and homogenizing, sometimes in the presence of enzymes [20]. This results in well dispersed fibrils with a crystalline packing of cellulose chains. They have a high aspect ratio with a diameter of around 5 nm and lengths of several micrometers. The high aspect ratio results in physical properties such as gelling by entanglement and shear thinning behavior. They can be assembled into for example fibers, films, or porous foams. These properties together with the possibility to functionalize NFC with other nanomaterials, chemically or through biotechnological means makes NFC highly useful for functional materials [5].

For our modular proteins we used a CBM from the Clostridium thermocellum cellulosome scaffoldin protein CipA and the spider silk N-terminal (NT) module from the E. australis major ampullate spidroin. This resulted in the fusion protein NT+CBM_CipA that we studied for functionalization of CNF. The idea was to utilize the pH-driven switching of the N-terminal module when the other end of the fusion protein is attached to cellulose through the CBM_CipA for pH dependent assembly of NFC. The binding properties of the CBM_CipA on nanocellulose were first determined for recombinantly produced single CBM_CipA and double-CBM_CipA proteins. This allowed an extensive understanding of the properties of spider silk N-terminal module containing CBM-fusion proteins. Finally, the effect of dimerization of the spider silk module was verified by rheological characterization of a CNF and NT+CBM_CipA mixture while slowly changing the pH.

**Materials and Methods**

*Materials.* Restriction enzymes were purchased from New England Biolabs and ligase from Promega. The plasmid pET28a(+) was purchased from Novagen. Site directed mutagenesis was performed with a Phusion site-directed mutagenesis kit. DNA plasmid purification was done with Nucleospin plasmid purification kits. E. coli strains XL1-blue (Stratagene) and BL21 (DE3) (Novagen) were used as cloning and as expression hosts, respectively. EnPresso B (BioSilta) growth medium was used for protein production. Other chemicals were from Sigma Aldrich unless otherwise stated. Nanofibrillated cellulose (NFC) was prepared as described earlier [20].

*Design of fusion proteins.* The fusion proteins were designed to contain the N-terminal (NT) module of E. australis major ampullate spidroin 1 (MaSp1) [19] and the cellulose-binding module (CBM) from C. thermocellum cellulosome scaffoldin subunit CipA, CBM_CipA (Uniprot Q06851)[9] connected by a linker region PTPTPTPTPTPTPTPTPTPTPS obtained from the C. fimi xylanase. A schematic illustration of the NT+CBM_CipA construct is shown in Fig. 1. The synthetized genes contained BsaI restriction sites in their 5’ and 3’ ends. The multiple cloning site of pET28a(+)
The expression vector was modified to include two BsaI restriction sites by inserting an adapter constructed from oligonucleotides 5’-CATGGGGAGACCGCGGATCCGAATTCGGGTCTC-3’ and 5’-TCGAGAGACCCGAATTCCGGATCCGCGGTCTCCC-3’ into NcoI and XhoI restriction sites. The fusion proteins were constructed from pieces coding for NT, CBM CipA and linker regions by utilizing the Golden gate method [21]. The genes were codon optimized for E. coli and synthetized by GenScript. The plasmids were transformed to E. coli XL1-blue strain and plated on kanamycin (50µg/mL) containing LB-plates. The single CBM CipA was cloned into the BsaI sites alone. All the expressed proteins contained a 6-His-tag in the C-terminal end.

The same linker peptide was also used to connect two CBM CipA proteins to form a double-CBM. An inactive variant, NT inact+CBM CipA, was made by site directed mutagenesis by replacing the Ala72 residue in NT by Arg to prevent the formation of a stable NT dimer [19]. The Ala72 is located in the dimer interface and is in contact with the same residue of the other partner of the homodimer across the interface. When Ala72 was replaced by a bulky Arg residue, dimerization was prevented [19]. This mutant protein NT inact+CBM CipA was utilized as a negative control protein for the experiments. In all, four different proteins were made, the single CBM CipA, the double-CBM CipA, the NT+CBM CipA, and the NT inact+CBM CipA. The full protein sequences are found in the supporting information.

Production and purification of the CBM and fusion proteins. The recombinant proteins were expressed in BL21(DE3) E. coli strain by cultivating the strains carrying the expression vectors in EnPresso B media (200 mL cultivations in 2 L Erlenmeyer flasks) according to the manufacturer’s instructions. The protein production was induced by 0.5mM isopropyl β-D-1-thiogalactopyranoside (IPTG, Roche Diagnostics). After 24h induction the cells were collected by centrifugation at 3900 g for 20 minutes. The cells were lysed by first freezing and thawing them once and then re-suspending them in 10 ml of 50 mM Tris-HCl pH 7.5 containing 1g/L of lysozyme from chicken egg white, 2.5 mM MgCl2 and DNAse 1 (Roche Diagnostics) in the presence of protease inhibitor cocktail (Complete, EDTA-free, Roche Diagnostics). The lysis reaction was carried out in room temperature for 2 to 3 hours after which the lysis was completed by sonication with 0.5 sec cycles for three minutes. The cell lysate was centrifuged for 30 min, 3900 g and the supernatant was applied onto 5ml His-Trap crude column (GE Healthcare) equilibrated with 20 mM sodium phosphate buffer pH 7.4 containing 0.5M NaCl and 20 mM imidazole. After washing with the equilibration buffer, the bound proteins were eluted with a linear gradient of imidazole from 20 mM to 500 mM in the equilibration buffer. Fractions were analyzed on SDS–PAGE and those containing pure protein were pooled and the buffer was exchanged to 50 mM sodium phosphate buffer pH 7.0 using Econo-Pac columns (Bio-Rad). The production levels of all variants were excellent; the protein yield after purification varied
from 400-500mg from 1 L of culture. Purified proteins are shown on SDS-PAGE (18%) stained with Coomassie blue (Fig.2).

**Protein concentration determination.** The protein concentrations of the purified proteins were determined by absorbance at 280 nm and calculated from the theoretical molar extinction coefficients calculated from the primary amino acid sequence [22]. CBM<sub>CipA</sub> \( \varepsilon = 35140 \, \text{M}^{-1} \, \text{cm}^{-1} \), NT+CBM<sub>CipA</sub> and NT<sub>inact</sub>+CBM<sub>CipA</sub> \( \varepsilon = 40910 \, \text{M}^{-1} \, \text{cm}^{-1} \), and double-CBM<sub>CipA</sub> \( \varepsilon = 70820 \, \text{M}^{-1} \, \text{cm}^{-1} \).

**Size exclusion chromatography.** Size exclusion chromatography (SEC) was performed with Äktapure system (Ge Healthcare) using Superdex 200 GL 10/300 column. The column was equilibrated with one of the following buffers: 50 mM citrate pH 3 or pH 3.5, 50 mM sodium acetate buffer pH 4, 4.5, 5 or 5.5, 50 mM sodium phosphate buffer pH 6, 6.5 or 7. All buffers contained 150 mM NaCl. Each sample was diluted in the equilibration buffer of the current run to protein concentration of about 0.5 to 0.7 g L<sup>-1</sup> and 250 \( \mu \text{L} \) was injected into the column. The elution (0.5 mLmin<sup>-1</sup>) was followed with a UV detector at 280 nm. SEC elution data were calibrated with the Gel Filtration Calibration Kit LMW (Ge Healthcare).

**Protein labelling with \( ^3 \text{H} \).** The single CBM<sub>CipA</sub> and fusion proteins were labelled with \( ^3 \text{H} \) by reductive methylation essentially as described earlier [23]. The buffer containing the proteins was changed to 200 mM borate (pH 8.5) using Econo-Pac gel-filtration columns. The concentration of amines (Lys and N-terminus) of each protein was calculated and 2.5 mg of each protein were mixed with formaldehyde in 5 times molar excess to the amine concentration. Tritium-enriched NaBH<sub>4</sub> (10.3 Ci mmol<sup>-1</sup>; 100 mCi; Perkin-Elmer) was dissolved in 0.1 M NaOH and added to the protein sample. The reactions were incubated 30 min on ice and terminated by applying the sample onto an Econo-Pac gel filtration column. The protein was eluted with 50 mM sodium phosphate pH 7.

**Binding to NFC.** Binding experiments were done in three different pH values using 50 mM sodium citrate pH 3.5, 50 mM sodium acetate pH 5.5, and 50 mM sodium phosphate pH 7 buffers. All buffers and protein dilutions contained 0.5% bovine serum albumin (BSA) to prevent non-specific binding. The initial protein concentrations ranged from 0.6 \( \mu \text{M} \) to 50 \( \mu \text{M} \). Equal volumes (100 \( \mu \text{L} \)) of protein and NFC suspension (2 g L<sup>-1</sup>) were dispensed in 24-well plates (Life Technologies) and mixed on orbital shaker in room temperature. After 60 min incubation, samples were filtered through Millex 0.22 \( \mu \text{m} \) GV13 filters (Millipore). The amount of \( ^3 \text{H} \) in the filtrate was quantified by liquid scintillation counting using Ultima Gold XR (PerkinElmer) scintillation counting liquid and a Tri-Carb 2810TR (PerkinElmer) scintillation counter. The amount of bound enzyme was calculated by
subtracting the measured free protein from the initial protein concentration. The reversibility of binding was tested by dilution experiments. A series of identical reaction mixtures were incubated for 60 min in order to obtain equilibrium (in pH 5.5). Then a 5-time excess of acetate buffer (0.5% BSA) without CBM_{CipA} or NT+CBM_{CipA} was added. After incubation for different times (15s to 20h), the mixtures were filtered, and the amounts of released protein were measured. Data were analyzed using the software Origin (Microcal).

**Rheology.** The pH dependent self-assembly of CNF and the protein variants NT+CBM_{CipA} and NT_{inact}+CBM_{CipA} were followed by rheological measurements. Samples were prepared by gently mixing 2 gL\(^{-1}\) CNF and 40 µM protein in 5 mM sodium phosphate buffer pH 7 in plastic tubes. The sample volume was 5 mL. Glucono-δ-lactone (GDL, Sigma Life Science) was added as a solid powder to the CNF-protein mixture (time point = 0, amount corresponds 0.4% final concentration) to change the sample pH gradually with time. Each sample was divided in two parts, one for pH measurement and the other for rheology. Rheological measurements were carried out at room temperature (22 °C) with a rheometer (AR-G2, TA instruments, UK) equipped with cross-hatched plate-plate geometry (diameter 40 mm). The viscoelastic properties were determined in small deformation oscillatory mode using 2.05 mL sample and a 1 mm gap. Time sweeps (frequency 0.1 Hz, strain 1 %) were run for 2 h. Evaporation was prevented by using a cover. The pH profiles were measured simultaneously as the rheological measurements using a pH meter with a data logger (pH 3310, WTW). The rheological and pH measurement were synchronized. Results were plotted according to corresponding time points. As a reference, 2 gL\(^{-1}\) CNF in 5 mM sodium phosphate buffer pH 7 without protein was used.

![Fig. 1. Schematic representation of NT+CBM_{CipA} (PDB id. for N-terminal module and CBM_{CipA} are 2LPJ and 1NBC, respectively). The aromatic and charged residues of CBM_{CipA} forming the cellulose-binding face are marked in red. The bars are showing the approximated sizes of the molecules.](image_url)
Fig. 2. SDS-PAGE analysis (18 % gel) of the purified proteins. Lane 1: single CBMCipA, lane 2: NT+CBMCipA, lane 3: NTinact+CBMCipA, lane 4: double-CBM CipA, leftmost lane: molecular weight marker

Results and Discussion

All fusion proteins, the single CBMCipA, the double-CBM CipA, the wild type spidroin NT-domain fused to the CBM (NT+CBM CipA), and the reference with a mutated and inactive spidroin NT variant (NTinact+CBM CipA), were produced with the E.coli expression system and were readily purified (Figure 2).

Initially we studied the pH dependent functionality of the two NT fusion proteins by SEC. At pH 7, the NT+CBM CipA protein eluted with a mass corresponding to its monomeric form (Fig 3a and b). Decreasing the pH to 6 led to dimerization and at pH 5.5 the protein was fully dimerized. This result supports previous reports of dimerization of the spidroin NT [18], and shows that the linkage to the CBM CipA, does not interfere with the NT dimer formation. As expected, the inactive mutant Ala72Arg (NTinact+CBM CipA) did not show dimerization at the same pH-range.
Fig. 3. Monomer/dimer analysis of NT+CBM fusion proteins. a) Molecular weights at different pH values calculated from size-exclusion chromatography; NT+CBMCipA (□), NT\textsubscript{inact}+CBMCipA (+) and double-CBM\textsubscript{CipA} (●) are shown. b) Size-exclusion chromatograms of NT+CBMCipA at two different pH values; NT+CBMCipA pH 5.5 (red), NT\textsubscript{inact}+CBMCipA pH 5.5 (black), NT+CBMCipA pH 7 (dashed line, red), NT\textsubscript{inact}+CBMCipA pH 7 (dashed line, black), including the calculated molecular weights. The theoretical molecular weights calculated from the amino acid sequences are: NT+CBMCipA = 34.6 kDa, NT\textsubscript{inact}+CBMCipA = 34.7 kDa, double-CBM\textsubscript{CipA} = 37.9 kDa.

Fig. 4. Comparison of the binding isotherms of CBM\textsubscript{CipA} (Δ) and double-CBM\textsubscript{CipA} (●) in different pH values. a) pH 7, b) pH 5.5 and c) pH 3.5. The same isotherms with closer zoom to the low protein concentrations are shown in the inserts.
We next determined the binding affinity and capacity of the isolated CBM\textsubscript{CipA} and the double-CBM\textsubscript{CipA} at pH 7 (Fig 4A). The parameters of binding were obtained by fitting a one-site binding model (Equation 1, where \(B\) is the amount of adsorbed protein, \(B_{\text{max}}\) is the maximum adsorbed protein, \(K_D\) is dissociation constant, and \(c\) is the concentration of free protein) to the data and are summarized in Table 1.

\[
B = \frac{B_{\text{max}} \cdot c}{K_D + c}
\]

We noted that the linkage of two CBM\textsubscript{CipA}-proteins to form the double-CBM\textsubscript{CipA} affected the binding by increasing affinity \(K_D\) from 0.57 \(\mu\text{M}\) to 0.2 \(\mu\text{M}\) and lowering capacity (from 25 to 16.7 \(\mu\text{mol g}^{-1}\)). The measured value \(K_D\) for the isolated CBM\textsubscript{CipA} fits very well with the previously reported \(K_D\) of 0.5 ± 0.2 \(\mu\text{M}\) [24] and 0.4 \(\mu\text{M}\) [25]. In Fig 4a, the insert shows the details of binding at low concentration, where the higher affinity of the double-CBM\textsubscript{CipA} is clearly visible. Another way of comparing affinities is through the partitioning coefficient \(K_r\), i.e., initial slopes of the curves (i.e. \(B_{\text{max}}/K_D\)). This value is the slope of the binding curve as the protein concentration approaches zero [26]. The partitioning coefficient is analogous to the specificity constant commonly used in Michaelis-Menten analysis (\(k_{\text{cat}}/K_M\), where \(k_{\text{cat}}\) is the turnover number and \(K_M\) is the Michaelis constant) [27]. Measured by \(K_r\), the double-linkage resulted in an increased initial affinity from 43.9 to 83.5 \(\text{L g}^{-1}\). Double-linkage thus results in a clearly higher affinity and also a lower capacity due to its larger size. This increase is expected due to a colocalization effect [7,28]. The double-CBM\textsubscript{CipA} binds first by one of its CBMs, and due to the colocalization of the second one, its local concentration increases, and it binds readily as also previously described [29].

The pH had a small effect on the CBM\textsubscript{CipA} binding going from pH 7 to pH 5.5, but more notable when going as low as pH 3.5 (Fig 4b and c). The ratio of \(K_r\) values (i.e. the \(K_r\) for the double-CBM\textsubscript{CipA} divided by the \(K_r\) for the single CBM\textsubscript{CipA}) remained very similar at around 2 regardless of the pH (Fig 4b and c, inserts, Table 1) That the ratio of \(K_r\) values does not change indicates that colocalization effects are not affected by pH [26].

Adding the N-terminal spidroin domain resulted in an overall decreased affinity (Fig 5, Table 1), with the \(K_r\) dropping to 2.7 \(\text{L g}^{-1}\) for the WT N-terminal domain and to 3.4 \(\text{L g}^{-1}\) for the inactive N-terminal domain mutant. \(B_{\text{max}}\) values show that the drop was due to a reduction in affinity, not on a change in capacity. The reason for the decrease in affinity is not obvious, although it is not unexpected that addition of a domain to the CBM\textsubscript{CipA} may result in a steric hindrance during binding. The two variants with wild-type NT and the inactive mutant behave very differently when the pH decreased. At pH 3.5, the inactive mutant had a low \(K_r\) of 2.95 \(\text{L g}^{-1}\) while the wild type had a \(K_r\) of 10.4 \(\text{L g}^{-1}\). This large
increase of the NT+CBM_{CipA}, which followed a different trend comparing to all other CBM variants is expected due to the dimerization of the NT domain at low pH, resulting in a colocalization effect and thereby increased affinity. The non-dimerizing mutant, on the other hand, would not be expected to show colocalization at any pH.

Fig. 5 Comparison of the binding isotherms of NT+CBM_{CipA} (□) and NT_{inact}+CBM_{CipA} (+) in different pH values. a) pH 7, b) pH 5.5 and c) pH 3.5. The same isotherms with closer zoom to the low protein concentrations are shown in the inserts.

Table 1. Affinity and capacity parameters obtained from binding isotherm fitting. Errors in B_{max} and K_{D} are standard errors non-linear curve-fit. B_{max} is the amount of adsorbed protein in saturation, K_{D} is dissociation constant. K_{r} is the partitioning coefficient which is defined as the ratio of B_{max} and K_{D}.

<table>
<thead>
<tr>
<th></th>
<th>pH 7</th>
<th></th>
<th></th>
<th>pH 5.5</th>
<th></th>
<th></th>
<th>pH 3.5</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>K_{D} (µM)</td>
<td>B_{max} (µmol g^{-1})</td>
<td>K_{r} (Lg^{-1})</td>
<td>K_{D} (µM)</td>
<td>B_{max} (µmol g^{-1})</td>
<td>K_{r} (Lg^{-1})</td>
<td>K_{D} (µM)</td>
<td>B_{max} (µmol g^{-1})</td>
<td>K_{r} (Lg^{-1})</td>
</tr>
<tr>
<td>CBM_{CipA}</td>
<td>0.57 ±0.03</td>
<td>25.0 ±0.4</td>
<td>43.9</td>
<td>0.54 ±0.02</td>
<td>21.9 ±0.2</td>
<td>40.6</td>
<td>5.0 ±0.4</td>
<td>21.2 ±0.4</td>
<td>4.2</td>
</tr>
<tr>
<td>double-CBM_{CipA}</td>
<td>0.20 ±0.01</td>
<td>16.7 ±0.2</td>
<td>83.5</td>
<td>0.14 ±0.01</td>
<td>13.0 ±0.2</td>
<td>92.9</td>
<td>0.79 ±0.05</td>
<td>9.1 ±0.2</td>
<td>11.5</td>
</tr>
<tr>
<td>NT+CBM_{CipA}</td>
<td>2.7 ±0.3</td>
<td>24.7 ±0.9</td>
<td>9.1</td>
<td>1.9 ±0.1</td>
<td>31.8 ±0.5</td>
<td>16.7</td>
<td>1.3 ±0.1</td>
<td>13.5 ±0.2</td>
<td>10.4</td>
</tr>
<tr>
<td>NT_{inact}+CBM_{CipA}</td>
<td>3.4 ±0.2</td>
<td>25.5 ±0.5</td>
<td>7.8</td>
<td>3.3 ±0.3</td>
<td>31.8 ±1</td>
<td>10.0</td>
<td>4.7 ±0.4</td>
<td>15.7 ±0.6</td>
<td>2.95</td>
</tr>
</tbody>
</table>
For exploring the magnitude of the colocalization effects on binding we calculated the change in free energy of binding for the different variants using equation 2.

\[
\Delta G = RT \ln K_D
\]

We note that the cooperativity in case of the double construct is significantly less that adding the free energies of the single CBMs, which is expected and fits well with previously observed systems of cooperativity [28]. The cooperativity effect is only a few percent of the sum of the free energies. The free energy change during binding of double-CBM\text{CipA} is slightly higher that for NT+CBM\text{CipA} at pH 3 indicating a greater steric hindrance due to the presence of the NT. The relatively low degree of cooperativity in the binding of the second CBM in the NT+CBM\text{CipA} suggest that a significant portion of the CBMs connected by dimerization are not cooperatively bound to the cellulose surface.

Table 2. Free energies of binding at different pH for all protein variants

<table>
<thead>
<tr>
<th></th>
<th>pH 7</th>
<th>pH 5.5</th>
<th>pH 3.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBM\text{CipA}</td>
<td>ΔG</td>
<td>ΔG</td>
<td>ΔG</td>
</tr>
<tr>
<td></td>
<td>kJmol(^{-1})</td>
<td>kJmol(^{-1})</td>
<td>kJmol(^{-1})</td>
</tr>
<tr>
<td>double-CBM\text{CipA}</td>
<td>-35.03</td>
<td>-35.16</td>
<td>-29.74</td>
</tr>
<tr>
<td>NT+CBM\text{CipA}</td>
<td>-37.58</td>
<td>-38.45</td>
<td>-34.24</td>
</tr>
<tr>
<td>NT\text{inact+CBM}\text{CipA}</td>
<td>-31.24</td>
<td>-32.10</td>
<td>-33.02</td>
</tr>
</tbody>
</table>

Another interesting property of CBM\text{CipA} binding is if the protein shows only a very slow desorption from the cellulose surface one it has bound. This property was studied using experiments were a set of identical tubes where CBM\text{CipA} and cellulose were mixed. Then buffer was added to some tubes and the bound amount in all tubes were measured. If the binding is reversible, the dilution would result in a lowering of bound amount according to Equation (1) [30]. The results showed that desorption happened only very slowly, around 10% in 20 hours (Fig 6). The NT domain linkages did not affect desorption rates.
Fig. 6 Reversibility of CBM$_{CipA}$ by dilution with buffer at pH 5.5, 22°C. Reversible protein would return to a point on the isotherm, whereas CBM$_{CipA}$ did not completely return to the isotherm during 20 h incubation. A) single CBM$_{CipA}$, B) NT+CBM$_{CipA}$, C) NT$_{inact}$+CBM$_{CipA}$

Fig 7. The elasticity of NFC with addition of the hybrid spidroin dimerizing protein and cellulose binding modules (NT+CBM$_{CipA}$). The pH-dependent dimerization of the NT domain results in a macroscopic response in materials properties. Filled squares (■) represent the NT-CBM$_{CipA}$, open squares (□) the inactive variant NT$_{inact}$+CBM$_{CipA}$, and filled triangles (▲) the buffer control. The points are averages of three separate runs and error bars show the standard deviation.

Having demonstrated that the silk N-terminal domain functioned to dimerize the CBM-fusion proteins in a pH dependent manner, we continued to investigate how this pH sensitive switch could be used in the molecular design for novel materials functions. We studied how a gradual pH change
would affect the viscoelastic properties of a network of NFC in the presence of the NT+CBM\textsubscript{CipA}. As a control we used the $NT_{inact}+CBM_{CipA}$ containing the inactive mutant. The pH change was accomplished by adding a lactone that slowly decomposes through hydrolysis when diluted into water. Through the carboxylic acid formed by the hydrolysis, the pH slowly decreased in the sample. This approach has the advantage that no additions of substances were needed during the experiment. In the sample containing CNF and the functional NT+CBM\textsubscript{CipA} an immediate increase in storage modulus was noted while the pH decreased, reaching an optimum at around pH 5.5 (Fig 7). The control samples with either no added protein or with the inactive $NT_{inact}+CBM_{CipA}$ mutant did not change the storage modulus appreciably. The pH-sensitive cross-linking of the NT in combination with the CBM binding to CNF could thus be used to modulate viscoelastic properties in a pH sensitive manner. The formation of a maximum peak at around pH 5.5 and the decrease of $G'$ at lower pH, where NT dimer formation is still strong, is probably due to the overall lower binding of the CBM\textsubscript{CipA} at low pH (Fig 5c).

Considering the opposed effects of cross-linking and increased affinity by colocalization we note that the degree of cooperativity in the binding of dimeric NT+CBM\textsubscript{CipA} ($\Delta G$ of binding was much smaller than the sum of the $\Delta G$ of the non-dimerized parts). We can therefore expect that much of the linked CBMs are free to interact with neighboring fibrils of cellulose resulting in a cross-linking. Also the slow desorption rate of the CBMs would contribute to making the cross-linking efficient by making exchange rates slow on the timescale of the rheology experiments.

**Conclusions**

The modularity of proteins such as spidroins and cellulolytic enzymes encourage us to explore modular architectures for building new functional biopolymers with novel combinations [31]. For material science this may be one of the most promising and versatile features of using proteins for molecular design. Functional properties from widely different sources can be combined, and resulting in proteins that can be produced as new multi-module polymers, allowing completely novel applications. Here we showed in detail how dimeric modules of CBMs interact with cellulose. The detection by tritium scintillation counting allowed highly precise binding data. Adding the pH sensitive spidroin NT-module resulted in a pH responsive mechanism for modulating protein interactions, which together with CNF gave a way to modulate viscoelastic properties of the combined protein-cellulose material. A detailed understanding of binding events and thermodynamics allowed us to understand the relation between cross-linking and cooperative binding.
Acknowledgements

The work was performed within the Academy of Finland Center of Excellence Programme (2014-2019) and Academy of Finland projects 307474, 317395 and 317019. We are grateful for the support by the FinnCERES Materials Bioeconomy Ecosystem

References


SUPPORTING INFORMATION

Modular protein architectures for pH-dependent interactions and switchable assembly of nanocellulose

Sanni Voutilainen\textsuperscript{1,2}, Arja Paananen\textsuperscript{2}, Martina Lille\textsuperscript{2}, and Markus Linder\textsuperscript{1}

1. Department of Bioproducts and Biosystems, School of Chemical Engineering, Aalto University, Espoo, Finland.
2. VTT Technical Research Centre of Finland Ltd., P.O. Box 1000, 02044 VTT Espoo, Finland

Amino acid sequences of the recombinant proteins used in the study.

\text{CBM}_{\text{CipA}}
MGNLKVEFYNNSNDTTSINPQFKVTNTGSSAIDLSKLTLRYYTVDGQKDQTFWCDHA AIIGSNGSYNGITSNVKGTTFVKMSSSTNNADTYLEISFTGTLTEPAHVIQGRIQGRFAKNDWSN YTQSNDYSFKSASQFVEWDVQTAYLNGVLVWGKEPLEHHHHHH

double-\text{CBM}_{\text{CipA}}
MGNLKVEFYNNSNDTTSINPQFKVTNTGSSAIDLSKLTLRYYTVDGQKDQTFWCDHA AIIGSNGSYNGITSNVKGTTFVKMSSSTNNADTYLEISFTGTLTEPAHVIQGRIQGRFAKNDWSN YTQSNDYSFKSASQFVEWDVQTAYLNGVLVWGKEPSAPTPTPTPTPTPTPTPTPSNLKVEFYNNSNDTTSINPQFKVTNTGSSAIDLSKLTLRYYTVDGQKDQTFWCDHAIIIGSNGSYNGITSNVKGTTFVKMSSSTNNADTYLEISFTGTLTEPAHVIQGRIQGRFAKNDWSN YTQSNDYSFKSASQFVEWDVQTAYLNGVLVWGKEPLEHHHHHH

\text{NT+CBM}_{\text{CipA}}
MGSGNSHTTPWTNPLAEFNFMNSFMQGLSSLMPGFTASQLDDMSTIAQSMVQSIQSLAAGQ RTSPNKQLALNMFASSMAEIAASEEIGGSLSTSTSSASAMSFLQTTTGVVQNFINE ITQVLSMAFAQAGMNDSAPTPTPTPTPTPTPTPTPSNLKVEFYNNSNDTTSINPQ FKVTNTGSSAIDLSKLTLRYYTVDGQKDQTFWCDHAIIIGSNGSYNGITSNVKGTTFVKM SSSTNNADTYLEISFTGTLTEPAHVIQGRIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDVQ VTAYLNGVLVWGKEPLEHHHHHH

\text{NT}_{\text{inact}+\text{CBM}_{\text{CipA}}}
MGSGNSHTTPWTNPLAEFNFMNSFMQGLSSLMPGFTASQLDDMSTIAQSMVQSIQSLAAGQ RTSPNKQLALNMFASSMAEIAASEEIGGSLSTSTSSASAMSFLQTTTGVVQNFINE ITQVLSMAFAQAGMNDSAPTPTPTPTPTPTPTPTSNLKVEFYNNSNDTTSINPQ FKVTNTGSSAIDLSKLTLRYYTVDGQKDQTFWCDHAIIIGSNGSYNGITSNVKGTTFVKM SSSTNNADTYLEISFTGTLTEPAHVIQGRIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDVQ VTAYLNGVLVWGKEPLEHHHHHH