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

Sanni Voutilainen a,b, Arja Paananen b, Martina Lille b, Markus B. Linder a,*

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

A R T I C L E   I N F O
Article history:
Received 17 April 2019
Received in revised form 17 June 2019
Accepted 28 June 2019
Available online 29 June 2019

A B S T R A C T

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.

© 2019 Published by Elsevier B.V.

1. 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 hydro-lases. 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 poly-Ala stretches that are interrupted by Gly, Pro, Gln-rich regions. This part comprises the bulk of the spidroin and forms extensive j- 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

* Corresponding author.
E-mail address: markus.linder@aalto.fi (M.B. Linder).

https://doi.org/10.1016/j.ijbiomac.2019.06.227
0141-8130/© 2019 Published by Elsevier B.V.
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 *Euproctis chrysorrhoea* 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 make NFC highly useful for functional materials [5]. The surface of CNF is relatively inert as cellulose is composed only of repeating glucose linked by β-1,4-glycosidic bonds. The surface shows only hydroxyl groups that are not very reactive. For functionalization, the hydroxyl groups are often oxidized, typically to carboxyl groups. These can then be further reacted and modified with polymers, proteins, or other chemical functionalities [21,22]. CNF gels and other forms of nanocellulose have found multiple uses as for example carriers of drug molecules or as three dimensional scaffolds for growing stem cells [23,24]. The use of CBMs for functionalizing cellulose has the advantage that no covalent modifications need to be made since the binding of CBM is spontaneous. The ability to produce CBMs as fusion proteins has the advantage that no covalent modifications need to be made.

For our modular proteins we used a CBM from the *Clostridium thermocellum* cellulose 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<sub>Cipa</sub> 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<sub>Cipa</sub> for pH dependent assembly of NFC. The binding properties of the CBM<sub>Cipa</sub> on nanocellulose were first determined for recombinantly produced single CBM<sub>Cipa</sub> and double-CBM<sub>Cipa</sub> 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<sub>Cipa</sub> mixture while slowly changing the pH.

### 2. Materials and methods

#### 2.1. 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 Nucleosil 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].

#### 2.2. 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* cellulose scaffoldin subunit CipA, CBM<sub>Cipa</sub> (Uniprot Q06851) [9] connected by a linker region PTPITPTPTITPTPTPT (from the *C. fimii* xylanase. A schematic illustration of the NT + CBM<sub>Cipa</sub> construct is shown in Fig. 1. The synthetized genes contained Bsal restriction sites in their 5' and 3' ends. The multiple cloning site of pET28a(+) expression vector was modified to include two Bsal restriction sites by inserting an adapter constructed from oligonucleotides 5'-CATGGG CAGACCGGGATCCGAACTCCGGTCCTC-3' and 5'-TGGAGAGACCGAATT CCGATCCCGGTCTCCC-3' into NcoI and XhoI restriction sites. The fusion proteins were constructed from pieces coding for NT, CBM<sub>Cipa</sub> and linker regions by utilizing the Golden gate method [26]. 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 (50μgml<sup>−1</sup>) containing LB-plates. The single CBM<sub>Cipa</sub> was cloned into the Bsal 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<sub>Cipa</sub> proteins to form a double-CBM. An inactive variant, NT<sub>inact</sub> + CBM<sub>Cipa</sub>, was made by site directed mutagenesis by replacing the Ala72 residue in NT with Arg to prevent the formation of a stable NT dimer [19]. In all, four different proteins were made, the single CBM<sub>Cipa</sub>, the double-CBM<sub>Cipa</sub>, the 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 *Euproctis chrysorrhoea* 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].

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* cellulose scaffoldin subunit CipA, CBM<sub>Cipa</sub> (Uniprot Q06851) [9] connected by a linker region PTPITPTPTITPTPTPT (from the *C. fimii* xylanase. A schematic illustration of the NT + CBM<sub>Cipa</sub> construct is shown in Fig. 1. The synthetized genes contained Bsal restriction sites in their 5' and 3' ends. The multiple cloning site of pET28a(+) expression vector was modified to include two Bsal restriction sites by inserting an adapter constructed from oligonucleotides 5'-CATGGG CAGACCGGGATCCGAACTCCGGTCCTC-3' and 5'-TGGAGAGACCGAATT CCGATCCCGGTCTCCC-3' into NcoI and XhoI restriction sites. The fusion proteins were constructed from pieces coding for NT, CBM<sub>Cipa</sub> and linker regions by utilizing the Golden gate method [26]. 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 (50μgml<sup>−1</sup>) containing LB-plates. The single CBM<sub>Cipa</sub> was cloned into the Bsal 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<sub>Cipa</sub> proteins to form a double-CBM. An inactive variant, NT<sub>inact</sub> + CBM<sub>Cipa</sub>, 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]. In all, four different proteins were made, the single CBM<sub>Cipa</sub>, the double-CBM<sub>Cipa</sub>, the

---

**Fig. 1.** Schematic representation of NT + CBM<sub>Cipa</sub>. (PDB id. for N-terminal module and CBM<sub>Cipa</sub> are 2LPJ and 1NBC, respectively). The aromatic and charged residues of CBM<sub>Cipa</sub> forming the cellulose-binding face are marked in red. The bars are showing the approximated sizes of the molecules. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
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 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG, Roche Diagnostics). After 24 h, induction the cells were collected by centrifugation at 3900g for 20 min. 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 1 g L⁻¹ of lysozyme from chicken egg white, 2.5 mM MgCl₂ and DNase I (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 h after which the lysis was completed by sonication with 0.5 s cycles for 3 min. The cell lysate was centrifuged for 30 min, 3900g and the supernatant was applied onto 5 mL His-Trap crude column (GE Healthcare) equilibrated with 20 mM sodium phosphate buffer pH 7.4 containing 0.5 M 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 to 500 mg from 1 L of culture. Purified proteins are shown on SDS–PAGE (18%) stained with Coomassie blue (Fig. 2).

2.4. 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 [27]. $\varepsilon_{280} = 35140 \, \text{M}^{-1} \cdot \text{cm}^{-1}$, NT + CBMCipA and $\varepsilon_{280} = 40910 \, \text{M}^{-1} \cdot \text{cm}^{-1}$, and double-CBMCipA $\varepsilon = 70820 \, \text{M}^{-1} \cdot \text{cm}^{-1}$.

2.5. Size exclusion chromatography

Size exclusion chromatography (SEC) was performed with ÄktaPrime 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⁻¹ and 250 μL was injected into the column. The elution (0.5 mL min⁻¹) was followed with a UV detector at 280 nm. SEC elution data were calibrated with the Gel Filtration Calibration Kit LMW (GE Healthcare).

2.6. Protein labelling with $^3$H

The single CBMCipA and fusion proteins were labelled with $^3$H by reductive methylation essentially as described earlier [28]. 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 NaBH₄ (10.3 Ci mmol⁻¹; 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.
CBMC_{CipA} or NT + CBMC_{CipA} was added. After incubation for different times (15 s to 20 h), the mixtures were filtered, and the amounts of released protein were measured. Data were analyzed using the software Origin (Microcal).

2.8. Rheology

The pH dependent self-assembly of CNF and the protein variants NT + CBMC_{CipA} and NT_{inact} + CBMC_{CipA} were followed by rheological measurements. Samples were prepared by gently mixing 2 mL^{-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 Sci- ence) 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 of free protein) to the data and are summarized in Table 1.

\[ B = \frac{B_{max} \times C}{K_D - C} \]  

We noted that the linkage of two CBMC_{CipA}-proteins to form the double-CBMC_{CipA} affected the binding by increasing affinity \(K_D\) from 0.57 μM to 0.2 μM and lowering capacity (from 25 to 16.7 μmol g^{-1}). The measured value \(K_D\) for the isolated CBMC_{CipA} fits very well with the previously reported \(K_D\) of 0.5 ± 0.2 μM [29] and 0.4 μM [30]. One-site Langmuir isotherms describe the binding data very well suggesting a single type of binding interaction. This becomes especially clear at high concentrations with the high-affinity double-CBMC_{CipA}. In Fig. 4a, the insert shows the details of binding at low concentration, where the higher affinity of the double-CBMC_{CipA} is clearly visible. Another way of comparing affinities is through the partitioning coefficient \(K_c\), i.e., initial slopes of the curves (i.e. \(B_{max}/K_D\)). This value is the slope of the binding curve as the protein

Initially we studied the pH dependent functionality of the two NT fusion proteins by SEC. At pH 7, the NT + CBMC_{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 CBMC_{CipA} does not interfere with the NT dimer formation. As expected, the inactive mutant Ala72Arg (NT_{inact} + CBMC_{CipA}) did not show dimerization at the same pH-range. 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} + CBMC_{CipA} was utilized as a negative control protein for the experiments.

We next determined the binding affinity and capacity of the isolated CBMC_{CipA} and the double-CBMC_{CipA} at pH 7 (Fig. 4A). The parameters of binding were obtained by fitting a one-site binding model (Eq. (1), where \(B\) is the amount of adsorbed protein, \(B_{max}\) 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.

\[ B = \frac{B_{max} \times C}{K_D - C} \]  

We noted that the linkage of two CBMC_{CipA}-proteins to form the double-CBMC_{CipA} affected the binding by increasing affinity \(K_D\) from 0.57 μM to 0.2 μM and lowering capacity (from 25 to 16.7 μmol g^{-1}). The measured value \(K_D\) for the isolated CBMC_{CipA} fits very well with the previously reported \(K_D\) of 0.5 ± 0.2 μM [29] and 0.4 μM [30]. One-site Langmuir isotherms describe the binding data very well suggesting a single type of binding interaction. This becomes especially clear at high concentrations with the high-affinity double-CBMC_{CipA}. In Fig. 4a, the insert shows the details of binding at low concentration, where the higher affinity of the double-CBMC_{CipA} is clearly visible. Another way of comparing affinities is through the partitioning coefficient \(K_c\), i.e., initial slopes of the curves (i.e. \(B_{max}/K_D\)). This value is the slope of the binding curve as the protein

3. Results and discussion

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

Initially we studied the pH dependent functionality of the two NT fusion proteins by SEC. At pH 7, the NT + CBMC_{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 CBMC_{CipA} does not interfere with the NT dimer formation. As expected, the inactive mutant Ala72Arg (NT_{inact} + CBMC_{CipA}) did not show dimerization at the same pH-range. 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} + CBMC_{CipA} was utilized as a negative control protein for the experiments.

We next determined the binding affinity and capacity of the isolated CBMC_{CipA} and the double-CBMC_{CipA} at pH 7 (Fig. 4A). The parameters of binding were obtained by fitting a one-site binding model (Eq. (1), where \(B\) is the amount of adsorbed protein, \(B_{max}\) 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.

\[ B = \frac{B_{max} \times C}{K_D - C} \]  

Table 1

<table>
<thead>
<tr>
<th>pH 7</th>
<th>(K_D) (μM)</th>
<th>(B_{max}) (μmol g^{-1})</th>
<th>(K_c) (Lg^{-1})</th>
<th>(B_{max}) (μmol g^{-1})</th>
<th>(K_c) (Lg^{-1})</th>
<th>(B_{max}) (μmol g^{-1})</th>
<th>(K_c) (Lg^{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBMC_{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>
</tr>
<tr>
<td>Double-CBMC_{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>
</tr>
<tr>
<td>NT_{inact} + CBMC_{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>
</tr>
<tr>
<td>NT_{inact} + CBMC_{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>
</tr>
</tbody>
</table>

Bold numbers signify standard deviation.
The pH had a small effect on the CBMCipA binding going from pH 7 to pH 5.5, but more notable when going as low as pH 3.5 (Fig. 4b and c, Table 1). The exact reason for the decrease of binding is not clear, but we note that the binding face of CBMCipA does contain a His-residue, and that in engineered CBMs it has been shown that His residues lead to lower affinities at low pH due to their protonation [9,35]. The ratio of $K_r$ values (i.e. the $K_r$ for the double-CBMCipA divided by the $K_r$ for the single CBMCipA) 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 [31].

Adding the N-terminal spidroin domain resulted in an overall decreased affinity (Fig. 5, Table 1), with the $K_r$ dropping to 2.7 Lg$^{-1}$ for the WT N-terminal domain and to 3.4 Lg$^{-1}$ for the inactive N-terminal domain mutant. $K_d$ 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 in the fusion construct may be due to a steric hindrance. The fact that the inactive mutant and the wild type have exactly the same binding affinity compared to each other for simultaneous binding of both domains, although the geometry of the binding sites on cellulose may require a precise positioning of the two CBMs in relation to each other. A precise positioning of CBMs implies higher structural order, and hence a higher entropy cost of binding. This would reduce the free energy available for the cooperative binding, which is in line with our observations. The free energy change during binding of double-CBMCipA is slightly higher than for NT + CBMCipA at pH 3 indicating a greater steric hindrance due to the presence of the NT. Looking at binding capacities (Table 1), we note that at pH 3, the binding capacity of the other protein variant divided by that of the NT + CBMCipA is slightly higher than for NT + CBMCipA, suggesting that the dimerization interaction leads to a need for more space on the cellulose surface when the NT + CBMCipA binds compared to the inactive mutant.

Another interesting property of CBMCipA 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 CBMCipA and cellulose were mixed. Then buffer was added to some tubes and the bound amount in all tubes was measured. If the binding is reversible, the dilution would result in a lowering of bound amount according to Eq. (1) [36]. The results showed that concentration approaches zero [31]. The partitioning coefficient is analogous to the specificity constant commonly used in Michaelis-Menten analysis ([$k_{cat}/K_M$], where $k_{cat}$ is the turnover number and $K_M$ is the Michaelis constant) [32]. Measured by $K_r$, the double-linkage resulted in an increased initial affinity from 43.9 to 83.5 Lg$^{-1}$. 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 [7,33].

The free energies of binding at different pH for all protein variants are shown in Table 2. We note that the cooperativity in case of the double construct (double-CBMCipA) does not fully correspond to the sum of free energies of the single CBMs. This means that although there is a cooperativity in binding, it corresponds to less than the maximum possible. This observation is perfectly in line with studies on other cooperative systems, and as discussed in previously [33]. The likely reason for this commonly observed effect is that there is an energetic cost for orientation in binding and that the rearrangements of the geometries of the proteins is required. On the other hand, the observed cooperativity shows that the binding sites of CBMs on the cellulose surface are sufficiently close to each other for simultaneous binding of both domains, although the geometry of the binding sites on cellulose may require a precise positioning of the two CBMs in relation to each other. A precise positioning of CBMs implies higher structural order, and hence a higher entropy cost of binding. This would reduce the free energy available for the cooperative binding, which is in line with our observations. The free energy change during binding of double-CBMCipA is slightly higher than for NT + CBMCipA at pH 3 indicating a greater steric hindrance due to the presence of the NT. Looking at binding capacities (Table 1), we note that at pH 3, the binding capacity of double-CBMCipA is close to half of that of CBMCipA. Because this is on a molar basis, it suggests that both CBMs of double-CBMCipA are bound to the cellulose. The binding capacity of NT + CBMCipA is slightly higher than for NT + CBMCipA, suggesting that the dimersization interaction leads to a need for more space on the cellulose surface when the NT + CBMCipA binds compared to the inactive mutant.

Another interesting property of CBMCipA 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 CBMCipA and cellulose were mixed. Then buffer was added to some tubes and the bound amount in all tubes was measured. If the binding is reversible, the dilution would result in a lowering of bound amount according to Eq. (1) [36]. The results showed that

$$\Delta G = RT \ln K_D$$  \hspace{1cm} (2)

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>$\Delta G$</td>
<td>$\Delta G$</td>
<td>$\Delta G$</td>
<td></td>
</tr>
<tr>
<td>kJmol$^{-1}$</td>
<td>kJmol$^{-1}$</td>
<td>kJmol$^{-1}$</td>
<td></td>
</tr>
<tr>
<td>CBMCipA</td>
<td>-35.03</td>
<td>-35.16</td>
<td>-29.74</td>
</tr>
<tr>
<td>Double-CBMCipA</td>
<td>-37.58</td>
<td>-38.45</td>
<td>-34.24</td>
</tr>
<tr>
<td>NT + CBMCipA</td>
<td>-31.24</td>
<td>-32.10</td>
<td>-33.02</td>
</tr>
<tr>
<td>NT + CBMCipA</td>
<td>-30.68</td>
<td>-30.75</td>
<td>-29.89</td>
</tr>
</tbody>
</table>
desorption happened only very slowly, around 10% in 20 h (Fig. 6). The NT domain linkages did not affect desorption rates.

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 + CBMCipA. As a control we used the NTinact + CBMCipA 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. The decrease in pH resulted in a dimerization of NT + CBMCipA, but not NTinact + CBMCipA (Fig. 3). Accordingly, in the sample containing CNF and the functional NT + CBMCipA 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 NTinact + CBMCipA mutant did not change the storage modulus appreciably. We therefore find that the dimerization of the N-terminal domain leads to crosslinking of the NFC fibrils. 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 CBMCipA at low pH (Fig. 5c). We could therefore demonstrate a silk-like assembly of NFC in an analogous way as in which silk proteins assemble due to a pH trigger [18]. We propose that such pH-dependent gel stiffening could be useful for applications such as switchable hydrogels for cell culture or release of drugs from NFC colloids [23,24]. On another level the work shows the great versatility for proteins as components in functional materials. Functionalities such as cellulose binding and pH-dependent binding can be combined in a straightforward way to achieve combined properties that would be very difficult to achieve by use of for example synthetic polymers or molecules. The simple dimerization of CBMs in the double-CBMCipA shows that parameters such as affinity and binding capacity also can be engineered in a straightforward way and could find applications in areas such as immobilization, nanomaterials, or modification of cellulose degrading enzymes.

4. 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 [37]. 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...