A tripartite carbohydrate-binding module to functionalize cellulose nanocrystal

Angeline Pelus, Gaëlle Bordes, Sophie Barbe, Younes Bouchiba, Callum Burnard, Juan Cortés, Brice Enjalbert, Jérémy Esque, Alejandro Estaña, Régis Fauré, et al.

A tripartite carbohydrate-binding module to functionalize cellulose nanocrystal

Angeline Pelus[a†], Gaëlle Bordes[b†], Sophie Barbe[a], Younes Bouchiba[a], Callum Burnard[a], Juan Cortés[c], Brice Enjalbert[a], Jeremy Esque[a], Alejandro Estaña[c], Régis Fauré[a], Anthony K. Henras[b], Stéphanie Heux[a], Claude Le Men[a], Pierre Millard[a], Sébastien Nouaille[a], Julien Pérochon[a], Marion Toanen[a], Gilles Truan[a], Amandine Verdier[a], Camille Wagner[a], Yves Romeo[b]*, Cédric Y. Montanier[a]*

The development of protein and microorganism engineering have led to rising expectations of biotechnology in the design of emerging biomaterials, putatively of high interest to reduce our dependence on fossil carbon resources. In this way, cellulose, a renewable carbon based polysaccharide and derived products, display unique properties used in many industrial applications. Although the functionalization of cellulose is common, it is however limited in terms of number and type of functions. In this work, a Carbohydrate-Binding Module (CBM) was used as a central core to provide a versatile strategy to bring a large diversity of functions to cellulose surfaces. CBM3a from *Clostridium thermocellum*, which has a high affinity for crystalline cellulose, was flanked through linkers with a streptavidin domain and an azide group introduced through a non-canonical amino acid. Each of these two extra domains were effectively produced and functionalized with a variety of biological and chemical molecules. Structural properties of the resulting tripartite chimeric protein were investigated using molecular modelling approaches, and its potential for the multi-functionalization of cellulose was confirmed experimentally. As a proof of concept, we show that cellulose can be labelled with a fluorescent version of the tripartite protein grafted to magnetic beads and captured using a magnet.

Introduction

Decades of progress in protein and microorganism engineering have led to the development of emergent multifunctional biomaterials[1–3]. Beyond the indisputable technical advantages brought by these emerging biomaterials, their development must also accompany our modern societies towards a greater independence from fossil carbon resources. Cellulose is the most abundant naturally occurring organic substance in the biosphere. While some bacteria are able to produce cellulose[4], it is a basic structural material of most plants, synthetized by cellulose synthase complexes present in the Golgi apparatus[5]. Cellulose consists of unbranched water-insoluble homopolysaccharide of β-1,4 linked glucopyranose residues, which are between 100 and 15,000 sugar units in length, organized in strong microfibril meshes[6]. Cellulose offers the enormous advantage of being a safe and inert macromolecule with versatile physical properties[7]. It is also inexpensive and commercially available in various forms, not only for paperboard and paper but also as derivatives used for food packaging[8]. Cellulose is also used as a bio-based reinforcement in composites[9] or as insulator[10]. This plant polysaccharide has also been approved for pharmaceutical, cosmetic, analytical devices or tissue engineering[11–13]. Previous work has shown that the properties of cellulose derivatives such as cellulose nanocrystals (CNCs) can be improved and modulated *via* chemical grafting of new functional groups[14]. This cheap and abundant material presents a very attractive solution to create CNCs functionalized with custom chemical functions and/or labelling for an increasing number of biotechnological and medical applications[15–17]. An easy and flexible process for efficient and convenient functionalization of CNCs would broaden the scope of industrial applications.

Currently applied cellulose surface modification methods require tedious drying/solvent exchange of the cellulose and the use of organic solvents, which has a negative impact on the environment[18]. Besides, several studies reported the use of some carbohydrate-binding modules (CBMs) to reversibly interact with cellulose, providing innovating solid support for protein purification[19] or to functionalize cellulose[20–22]. CBMs are non-catalytic modules of multi-modular glycoside hydrolases (GHs) produced by lignocellulolytic microorganisms[23]. Many of the CBMs contain from 30 to about 200 amino acids and exist as single or repeated copies into proteins. They often bind to specific oligo- and polysaccharides derived from the plant cell wall. CBMs modulate the activity of the appended enzyme through i) a proximity effect, ii) a targeting function and iii) in a lesser extent a contribution to non-hydrolytic substrate disruption[23]. Three distinct topologies of CBM ligand-binding sites complement the conformation of the target polysaccharide. Type A displays a planar

[a.] *TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France.*
[b.] *Molecular, Cellular and Developmental biology department (MCD), Centre de Biologie Integrative (CBI), Université de Toulouse, CNRS, UPS, 31062 Toulouse, France.*
[c.] *LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France.*
† These authors contributed equally to this work.
* Correspondence: yves.romeo@univ-tlse3.fr, Tel. +33 (0)5 61 33 58 08 and cedric.montanier@insa-toulouse.fr, Tel. +33 (0)5 61 55 97 13.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

hydrophobic platform containing exposed aromatic amino acids which interact with a flat surface of cellulose through carbohydrate-aromatic stacking interactions[24]. Type B accommodates the ligand within extended clefts of varying depths[25] while type C binds to small or individual glycosyl unit[26]. CBMs are widely used in fusion with recombinant protein for technological purposes such as protein purification, protein immobilization, targeting activity to cellulose, labelling or biomaterial functionalization[27]. For the last case, the functionality is actually provided by the module fused to the CBM, while in a recent study, CBM was used to introduce an alkyne group to cellulose surfaces, being subsequently combined with polyethylene glycol[28]. Whatever the approach currently chosen to functionalize cellulose, it is restricted to a unique function[21,29,30]. Our motivation was to develop an original fusion protein composed of three distinct domains, one being designed to bind to CNCs while two additional domains provide a wide range of biological and chemical functionalization. Studies reporting CNCs functionalization for advanced applications have considerably increased in the last decade[14]. Enlarging the functionalization possibilities of CNCs would foster the development of new applications by combining electronic, magnetic, catalytic or optical properties for future nanomaterials[31].

Here we report the design and the production of a tripartite chimeric protein named "mSA-CBM3-AzF", combining three different types of protein/ligand interactions. The core structure consists of the CBM3a from *Clostridium thermocellum* (CBM3), which has a strong affinity for crystalline and amorphous cellulose[32]. As a second part, the N-terminus of the CBM3 is flanked by the monomeric engineered mSA2 form of the streptavidin of *Streptomyces avidinii* (mSA)[33], which displays one the strongest noncovalent affinities in nature for a wide range of biotinylated molecules[34–36]. The third part consists in incorporating in the linker flanking the C-terminus of CBM3 the non-canonical amino acid (ncAA) 4-azido-L-phenylalanine (AzF)[37,38]. AzF contains an azide group able to covalently bond cyclooctyne (i.e. strained cycloalkyne) derivate molecules in a copper-free Strain-Promoted azide-alkyne cycloaddition (SPAAC), under physiological condition[39]. In this work, we produced and purified mSA-CBM3-AzF and provide a 3D structural model in solution of this chimeric proteins. Furthermore, we demonstrate the functionality of each of its part individually and finally bring evidence of the versatility of this chimeric protein in functionalizing cellulose using a fluorescent mSA-CBM-AzF grafted to magnetic beads.

Materials and methods

Cloning of the DNA sequences encoding the hybrid proteins

The sequence corresponding to CBM3a module of CipA from *Clostridium thermocellum*, surrounded by its endogenous *N*- and *C*-terminal linker sequences and the gene coding for the monomeric mRFP1 fluorescent protein from *Discosoma striata* were recovered from the iGEM registry parts (http://parts.igem.org) as Part:BBa_K1321014 and Part:BBa_E1010, respectively. *CBM3a* and *mRFP1* nucleotide sequences were modified as specified in the Supplementary Information and synthesized (IDT, Coralville, IA, USA). DNA cloning and mutagenesis were performed by homologous recombination (In-Fusion® HD cloning kit, Clonetech, Mountain View, CA, USA), using primers listed in Table S1. All constructions in this study were introduced into NcoI/HindIII-restricted pET28a(+) to ensure that the recombinant proteins harboured a *C*-terminal His_6-tag. Met_{127} of CBM3 in plasmid pET28-CBM3-mRFP1 was changed to Gly. The resulting gene *linker1-CBM3a-linker2-RFP1-*His_6 is registered in the iGEM registry (http://parts.igem.org/Help:Parts) as Part:BBa_K2668020 (Table S2). This plasmid was used to generate descendent plasmids pET28-mSA-CBM3 (lacking RFP1 moiety) and pET28-mSA-CBM3-AzF (with TAG amber codon for AzF incorporation at Phe_{364} position) as detailed in the Supplementary Information.

Protein expression and purification

Homemade *Escherichia coli* BL21 (DE3) competent cells (Mix & Go! *E.coli* Transformation Kit, Zymo Research, Irvine, CA, USA) harboring the appropriate recombinant plasmid were cultured in Luria-Bertani (LB) broth containing 50 µg/mL kanamycin at 37 °C to mid-exponential phase (OD_{600nm} = 0.6). Recombinant protein expression was induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM for 4h at 37°C. To produce the ncAA containing protein mSA-CBM3-AzF, *E.coli* BL21 (DE3) competent cells were co-transformed with the plasmid coding for mSA-CBM3-AzF and pEVOL-Azf (Addgene, Watertown, MA, USA, plasmid #31186) coding for the aminoacyl-tRNA synthetase (aaRS)/tRNA pair for AzF incorporation at the amber codon[40]. Cells were cultured in LB broth containing 50 µg/mL kanamycin, 34 µg/mL chloramphenicol, 1 % of D-glucose and 2 mM AzF (resuspended at 2 M in 1 N NaOH; the pH was adjusted with 1 mM HCl). mSA-CBM-AzF and aaRS/tRNA expression were induced by addition of 1 mM final IPTG and 0.02 % L-arabinose, respectively. The culture was incubated at 37 °C for further 4 h. Cells were harvested (5,000 g, 10 min), re-suspended in 10 ml of 50 mM Tris/HCl buffer pH 8.0 containing 150 mM NaCl and a protease inhibitor cocktail (SigmaFast™ Protease Inhibitor Cocktail, Merck KGaA, Darmstadt, Germany) and frozen at -20 °C until purification. Cells were disrupted by sonication on ice for 1 min and the lysate was clarified by centrifugation (30 min at 60,000 g at 8 °C). Proteins were purified by immobilized metal ion affinity chromatography (IMAC) using Talon resin (TALON® Metal Affinity Resin, Clontech, Mountain View, CA, USA) and eluted in 50 mM Tris/HCl buffer, pH 8.0 containing 150 mM NaCl supplemented with 100 mM imidazole. The eluted proteins were buffer exchanged in 50 mM Tris/HCl buffer pH 8.0 using a PD-10 desalting column (GE Healthcare Life Sciences, Chicago, IL, USA). Purified proteins were adjudged homogenous by SDS-PAGE (Any kD™ Mini-PROTEAN® TGX Stain-Free™ Protein Gels, Bio-Rad, Hercules, CA, USA). Protein concentrations were determined by measuring absorbance at 280 nm with NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Theoretical molecular weight and molar extinction coefficients (Table S3) were calculated using ProtParam online software (https://web.expasy.org/protparam/).

Protein analyses

Proteins were resuspended in Laemmli Buffer (40 mM Trizma base, 2 % SDS v/v, 5 % glycerol w/v, 0.08 % bromophenol blue w/v, 25 mM DTT). Western blot experiments were performed as follow: protein samples were heated 5 min at 95 °C, loaded on SDS-polyacrylamide gels (10 %) and transferred to nitrocellulose membranes using Trans-blot turbo transfer system (Bio-Rad, Hercules, CA, USA). Membranes

were saturated for 1 h with TBST buffer (150 mM NaCl, 20 mM Tris pH 8.0, 0.001% Tween-20) containing 5 % powder milk, and incubated over night with the same buffer containing HRP Anti-6X His tag antibody (Abcam, UK) antibodies. After 3 washes with TBST buffer, ECL detection was performed. ImageLab software (Bio-Rad, Hercules, CA, USA) was used to detect and quantify protein signals.

Pull-down assay on RAC

Solution of regenerated amorphous cellulose (RAC) was prepared as previously described[41,42] at a final concentration of 20 mg/mL. Determination of the equilibrium binding constant of CBM3, CBM3-RFP and mSA-CBM3 was performed by adding 2 µM of studied protein on increasing concentration of RAC (from 0 to 8 mg/mL) in 50 mM Tris/HCl pH 8.0, in a final volume of 100 µL for 1 h at 21° C with 1,000 rpm shaking (ThermoMixer® C, Eppendorf, Germany). Unbound protein recovered in the supernatant after centrifuging for 5 min at 2,250 g was quantified at 280 nm (NanoDrop™ 2000, ThermoFischer, Waltham, MA, USA). Equilibrium binding constant was determined using Sigma plot (Systat Software, San Jose, CA, USA).

msA-CBM3 labelling with biotinylated fluorophore and RAC binding

Commercial 5-(and-6)-tetramethylrhodamine biocytin (Invitrogene™ biocytin TMR, ThermoFisher Scientific, Waltham, MA, USA) was used to label mSA-CBM3 through streptavidin-biotin association. A mSA-CBM3:TMR-Biocytin ratio of 1:3 was mixed in 400 µL of 50 mM Tris/HCl pH 8.0 (12 µM of msA-CBM3 and 37 µM of TMR-Biocytin) and incubated protected from light 1 h at 21 °C under constant agitation at 1,000 rpm (ThermoMixer® C, Eppendorf, Germany). Labelled mSA-CBM3 was purified from unreactive TMR-biocytin using 100 µL of Talon resin (TALON® Metal Affinity Resin, Clontech, Mountain View, CA, USA) in a 1.5 mL Eppendorf tube. Each purification step was performed by pelleting the resin at 1500 g in 50 mM Tris/HCl pH 8.0. Labelled mSA-CBM3 was recovered in 200 µL by addition of 50 mM Tris/HCl pH 8.0 containing 100 mM imidazole. Fractions of interest were pooled and buffer exchanged in 50 mM Tris/HCl pH 8.0, using a PD-10 desalting column with the spin protocol in 2.5 mL (GE Healthcare Life Sciences, Chicago, IL, USA). Binding to RAC was performed with 100 µL of labelled mSA-CBM3 mixed with 20 µL of RAC 20 mg/mL, filled up to 200 µL with 50 mM Tris/HCl pH 8.0 buffer and incubated 1 h under constant agitation at 1,000 rpm (ThermoMixer® C, Eppendorf, Germany). Unbound proteins were recovered by centrifugation at 14,600 g. The pellet was washed two times with 200 µL of 50 mM Tris/HCl pH 8.0 buffer before being resuspended with 200 µL of TEV protease buffer (10 mM Tris/HCl pH 8.0, 150 mM NaCl, 0.5 mM EDTA) supplemented with 2 µg of TEV protease (GST-tagged and histidine-tagged Tev protease, T4455, Merck KGaA, Germany). The reaction was incubated 2 h at room temperature under constant agitation at 1,000 rpm. The solution was centrifuged 10 min at 14,600 g and the supernatant was collected. Fractions of interest were analysed by both SDS-PAGE (Any kD™ Mini-PROTEAN® TGX Stain-Free™ Protein Gels, Bio-rad) and fluorescence measurement (ex/em = 560/590 nm) in microplate (Corning™ 96-Well Clear Bottom Black Polystyrene Microplates, Fisher Scientific) using BMG FLUOstar Optima microplate reader (BMG Labtech, France).

msA-CBM3-AzF labelling with cyclooctyne functionalized fluorophore

Recombinant mSA-CBM3-AzF was labelled with an alkyne functionalized fluorophore DBCO-Alexa Fluor 488 (DBCO-AF488, Jena Bioscience, Germany) by SPAAC. Stock solution of DBCO-AF488 (10 mg/mL) was prepared in DMSO. SPAAC was performed in 500 µL of 50 mM Tris/HCl pH 8.0 buffer with a ratio azide:cyclooctyne of 1:2, meaning that 250 µL of labelled protein (35 µM) were mixed with 250 µL of DBCO-AF488 (0.1 mg/mL). Reaction was incubated protected from light overnight at 21°C with constant shaking (1,000 rpm). Labelled mSA-CBM3-AzF was purified from unreactive DBCO-AF488 using IMAC purification as previously described. Fractions of interest were analysed by both SDS-PAGE (Any kD™ Mini-PROTEAN® TGX Stain-Free™ Protein Gels, Bio-rad) and fluorescence measurement (ex/em = 500/530 nm) in microplate (Corning™ 96-Well Clear Bottom Black Polystyrene Microplates, ThermoFisher Scientific) using BMG FLUOstar Optima microplate reader (BMG Labtech, France).

Functionalization of mSA-CBM3-AzF with paramagnetic beads

An aliquot of mSA-CBM3-AzF (100 µL, 14 µM) was grafted to 30 µL of DBCO-magnetic beads (mean bead diameter 0.97 µm, 1 mg/mL, Jena Bioscience, Germany) by SPAAC click chemistry in a final volume of 200 µL filled up with 50 mM Tris/HCl pH 8.0 buffer. Reaction was performed overnight at 21 °C, protected from light with constant agitation (1,000 rpm). Similarly, a sample of magnetic beads was mixed with mSA-CBM3. The magnetic beads were washed 3 times with 1 mL of 50 mM Tris/HCl pH 8.0 buffer using a magnetic stand (PureProteome™ Magnetic Stand, 8-well, Merckmillipore, Ma, USA). Beads were resuspended in 200 µL of 50 mM Tris/HCl pH 8.0 buffer containing 2 mg/mL of RAC and incubated for 1 h at 21 °C under constant agitation (1,000 rpm). Ability of the different grafted magnetic beads to sediment RAC was finally judged by eye after 10 s contact with a magnet.

An aliquot of mSA-CBM3-AzF (2 mL, 8.5 µM) was mixed with TMR-biocytin (25 µL, 25 µM) and filled up to 2.5 mL of 50 mM Tris/HCl pH 8.0 buffer. After purification, 2 mL of TMR-biocytin/mSA-CBM3-AzF were mixed to 80 µL of DBCO-magnetic beads (10 mg/mL) and incubated as described above. A control consisting in replacing TMR-biocytin/mSA-CBM3-AzF with a buffer was performed. Washed magnetic beads were resuspended with 1.14 mL of 50 mM Tris/HCl pH 8.0 buffer and filled up to 1.5 mL with 0.15 mL of RAC (20 mg/mL) and incubated for 1 h at 21 °C under constant agitation (1,000 rpm). Each reaction was diluted twice with 50 mM Tris/HCl pH 8.0 buffer and 600 µL were used to follow the ability of the grafted magnetic beads to sediment RAC with a magnet using an optical bench in 1 mL spectrophotometer semi-micro cuvette (Biosigma S.p.A., Italy).

Dynamic of RAC sedimentation by magnetic beads grafted with fluorescent mSA-CBM3-AzF

The experimental set-up is shown in Fig. S1. Samples to be compared are poured into1 mL spectrophotometer semi-micro cuvette (Biosigma S.p.A.). The cells are spaced by 9.1 mm, allowing a 9×10 mm section Neodyme magnet (Supermagnete, Deutschland) to slide quickly between them. The camera is operated by a free Basler Pylon software (v5.1) that allows friendly configuration, visualization and acquisition of the images. The camera and lens system is mounted on a simple *Micro-contrôle X48 bench and carrier*, allowing easy

rotation of the camera, and translation of the system with respect to the object to adjust the focus. The illumination is performed by a home-made collimated source giving a 50 mm diameter beam, perfectly parallel for a low cost led at a wavelength of 530 nm (Fig. S1 and Supplementary Information). Two typical images recorded before and just after the magnet insertion (between the cuvettes) are shown in Fig. S2. The suspension monitored in transmission mode allows to measure the evolution of the suspension over time once the magnet is in place (see Supplementary Information for details, Fig. S3 and Fig. S4).

3D structural model building and conformational exploration of the tripartite protein

A 3D structural model of the His-tagged tripartite protein constituted by mSA and CBM3a domains flanked with the endogenous CipA *N*- and *C*-terminal linkers (involving residues 11-169 and 331-380, respectively) was built from the crystallographic structures of CBM3 (PDB: 4JO5)[43] and mSA (PDB: 4JNJ)[44] using a recently-developed method to generate 3D models of intrinsically disordered proteins (IDP)[45]. An ensemble of 10,000 conformations was generated considering CBM3 and mSA domains as rigid bodies, and sampling the flexible linker regions using the IDP conformational ensemble modelling method. This method exploits an extensive database of three-residue fragments extracted from coil regions in high-resolution experimentally-determined protein structures. It builds conformations, residue-by-residue, sampling dihedral angles values from the database, while avoiding steric clashes. In this case, we applied a simple strategy that disregards the local sequence context. For each sampled conformation of the backbone, statistically-relevant collision-free side-chain conformations were sampled using continuous rotamers[46]. The resulting models were energy minimized (steepest descent and conjugate gradient algorithms) using the sander module of the AMBER16 suite of programs[47] with the AMBER ff14SB force field[48] and the generalized born implicit solvation model[49]. For each minimized conformation, distance between the center of mass of mSA and the *C*-terminal Phe_{364} amino acid residue as well as the angle between the centers of mass of mSA, CBM3 and the *C*-terminal Phe_{364} of mSA-CBM3 were computed using the *cpptraj* AMBER module[50]. Distinct conformational states of the mSA-CBM3 molecular system were selected from the conformational ensemble model and studied through Molecular Dynamics (MD) simulations. Each conformational state model was first neutralized with three counter-ions and solvated with TIP3P water molecules, using an octahedral box with a minimum distance of 0.15 nm between the solute and the simulation box edges. The molecular systems were then subjected to an energy minimization schedule (steepest descent and conjugate gradient methods) with harmonic positional restraints of 25.0 kcal/mol/Å^2 that were gradually removed along the energy minimization schedule. Following the energy minimization steps, the systems were heated incrementally over 100 ps to 300 K, under constant volume conditions and with harmonic positional restraints of 25.0 kcal/mol/Å^2 on the solute atoms. At the final required temperature (300 K), the harmonic potential restraints were gradually removed along 150 ps system equilibration under constant pressure conditions (1 bar). The production phase of the MD simulations were then carried out over 50 ns at constant temperature and pressure. The temperature (300 K) and the pressure (1 bar) were controlled using the Berendsen algorithms[51]. Long-range electrostatic forces were handled by using the Particle-Mesh Ewald method[52]. A 9Å cut-off for non-bonded interactions was used. The integration time-step of each simulation was 2.0 fs and the SHAKE algorithm[53] was used to constrain the lengths of all chemical bonds involving hydrogen atoms to their equilibrium values. Atomic coordinates for each simulation were saved every 10 ps. The *sander* program of the AMBER16 suite of programs was used for MD preparations (minimization, heating and equilibration) while the *GPU pmemd.CUDA* program[54] as used for the production phase of the MD simulations. The root mean square deviation (RMSD) of backbone atoms relative to the starting conformational state was calculated for mSA and CBM3 domains along each MD simulation using the *cpptraj* module of the AMBER16 package. All plots were generated using R (v 3.4.4) (R Core Team (2020)).

Results

Rational protein design for custom cellulose functionalisation

The aim of our work was to develop a versatile protein providing cellulose with original combinations of biological and/or chemical functions for a wide range of biotechnological applications. The protein should also ensure sufficient independence for the different functions to avoid unproductive interactions or to promote possible cascade events. Hence, we designed a tripartite chimeric protein (Fig.1 and Table S2) whose cellulose-binding core corresponds to the well characterized CBM3a from the *Clostridium thermocellum* cellulosome scaffolding protein CipA flanked with its *N*- and *C*-terminal linker1 and 2, respectively[43]. This first component of our hybrid protein presents an affinity for crystalline cellulose estimated to 10^5 M^{-1} [55].

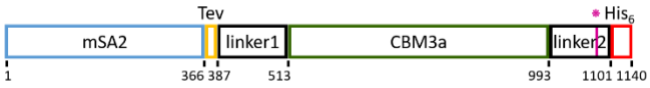

Fig. 1. Schematic representation of the tripartite chimeric protein. CBM3 (in green) is able to bind to cellulose, while Streptavidin (mSA2, in blue) is functionalized with a biotinylated molecule and the ncAA (AzF, in pink) through SPAAC click-chemistry. The CBM3 is connected to the other two parts with two flexible linkers1 and 2 (in black). A Tev protease site (in orange) between mSA2 and CBM3 allows the release of mSA2 part. A His6-tag (in red) eases the purification of the protein. Codon numbering is represented.

To functionalize cellulose with a molecule of biological origin, the second component of the hybrid protein relies on the streptavidin/biotin interaction, which is commonly used in biotechnology[35]. The monomeric engineered mSA form of the Streptavidin (SA) of *Streptomyces avidinii*[35] was fused to the *N*-terminal part of CBM3. Streptavidin is a homotetrameric protein that binds to up to four biotin molecules with an affinity of 10^{15} M^{-1} [36], one of the strongest noncovalent interactions in nature. The monomeric mSA version was used to avoid any unproductive tetramerization in our system. Nowadays a wide range of commercial biotinylated

molecules are available for various applications[35]. It is also possible to generate custom-made molecules chemically, chemoenzymatically[34] or alternatively *in vivo*[56]. A Tev protease site has been introduced between mSA and linker1 to allow splitting the chimeric protein if required. The third component of our tripartite protein is a reactive chemical function inserted at the *C*-terminal of CBM3 through the introduction of ncAA within the amino acid sequence[57]. Such targeted incorporation of ncAA into proteins involves the use of orthogonal tRNA/aaRS pairs that uniquely recognize a nonsense codon to reassign it to a ncAA. The amino acid specificity of the aaRS is modified such that it aminoacylates its cognate tRNA with only the desired ncAA into recombinant proteins, exploiting thus the translational machinery of the host cell *in vivo*[38]. A large diversity of ncAA harbouring various chemical group is available[58], enlarging the possibility of grafting numerous chemical molecules at a dedicated position. In our study, AzF was introduced in place of Phe_{364}, displaying an azide group able to form under physiological condition a covalent bond with cyclooctyne derivatives through SPAAC[39]. The resulting tripartite protein is referred to as mSA-CMB3-AzF.

Molecular modelling assesses the flexibility of mSA-CBM3

The tripartite protein uses linker1 and 2 to associate the CBM3 with the other two parts, mSA and AzF. These linkers, originated from *Ct*CBM3a, contain 49 and 36 amino acids, respectively. As all CipA linkers, they are supposed to be highly flexible[59]. This flexibility, though to be important for the high catalytic efficiency of the cellulosome[60], could impair the design of our chimeric protein because of steric hindrance due to a too short distance or an unfavourable geometry between CBM3, mSA and AzF, incompatible with the recognition or the functionalization of cellulose. To investigate structural properties of our tripartite protein, a conformational ensemble model was generated using a recent method dedicated to highly-flexible proteins (see Materials and Methods). As shown in Fig.2, mSA-CBM3 can adopt a large ensemble of low-energy conformations characterized by mSA - Phe_{364} (position of AzF) distance values ranging from 20 Å to 200 Å, and mSA - CBM3 - Phe_{364} angle values ranging from 5° to 180°.

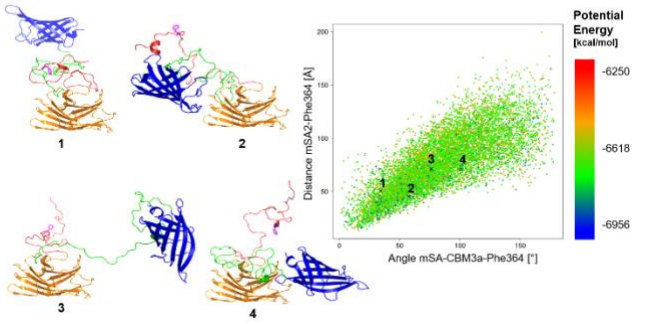

Fig. 2. Potential energy landscape of mSA-CBM3 projected on two geometric descriptors. The X axis corresponds to the angle between the centers of mass of mSA, CBM3a and the Phe364 at the extremity of the C-terminal linker2 while the Y axis corresponds to the distance between mSA and Phe364 (position of AzF) centers of mass. The color scale depicts the variation of the potential energy of the conformations. The average-energy conformations of four regions were extracted and represented in cartoon visualization with in blue the mSA domain, in green the linker1 between mSA and CBM3a domains, in orange the CBM3a domain and in red the C-terminal linker2. The Phe364 is shown in stick representation and is colored in magenta.

The potential energies are homogeneously distributed for the generated conformations, as expected for proteins with large intrinsically disordered regions (IDRs)[61]. Furthermore, the energies lower than -6,100 kcal/mol (which accounts for 99.9% of the sampled data) are normally distributed (Kolmogorov-Smirnov test p-value = 0.4249) around an energy value of -6,620 kcal/mol ± 86 (Fig. S5). In order to investigate the stability of the globular domains within different regions of the conformational space for the whole system, we then selected four distinct conformations with close-to-average energy values. The conformational states are included in the most populated regions of the landscape (Fig. 2 and Fig. S5). For each selected conformational state, we performed 50 ns- MD simulations of the whole mSA-CBM3 system in explicit solvent. The RMSD of backbone atoms (Fig. S6) is on average 1.0 Å for the CBM3 domain and 1.15 Å for mSA2, which indicates a stable behaviour of each globular domain of the mSA-CBM3 model. It appears that RMSD fluctuations are larger in mSA (σ=0.27Å) than in CBM3 (σ=0.14Å), which is expected given the presence of flexible loops in the structure of mSA[62,63] compared to the more packed structure of the CBM3. These four conformations represent thus different states that mSA-CBM3 may adopt in solution. The regions on both sides of CBM3 can adopt diverse orientations: they can be relatively close or remote. None of the modeled conformations present any interaction between mSA or linker2 (position of AzF) with the cellulose binding site of CBM3a. Such conformational variability makes this molecular system a tripartite platform compatible with the diverse functionalization envisaged.

Expression and purification of the chimeric CBM3 based proteins.

To investigate the biological activity of each individual part of mSA-CBM3-AzF we produced intermediate protein versions: the central domain CBM3 flanked with its linkers 1 and 2; CBM3 only in fusion with the fluorescent protein RFP1[64] to easily label CBM3: and CBM3 in fusion with mSA and with the incorporation of AzF (Table 1 and Table S3).

Table 1 Chimeric proteins produced in this study.

Chimeric Protein	Protein composition	Additional part	
CBM3-RFP	linker1-*Ct*CBM3a-linker2-RFP1-His_6	-	RFP1
mSA-CBM3	mSA2-Tev-linker1-*Ct*CBM3a-linker2-His_6	Tev site	mSA2
mSA-CBM3-AzF	mSA2-Tev-linker1-*Ct*CBM3a-linker2-AzF-His_6	Tev site	azido-phenylalanine

Each protein has a *C*-terminal His_6-tag in anticipation of purifying mSA-CBM3-AzF from truncated form of the protein having the translation aborted (no incorporation of AzF). Unexpectedly, expression of the three chimeric proteins in *E.coli* resulted in additional band of about 38 kDa for CBM3-RFP and of about 15 kDa for mSA-CBM3 and mSA-CBM3-AzF, corresponding to a putative proteolysis site in CBM3. Presence of extra bands reflects specific site proteolysis, potentially in linker regions (Fig. S7). A deeper analysis of the DNA sequence coding CBM3 revealed significant divergences between the sequence from iGEM registry (Part:BBa_K1321014) used and the sequence from GenBank (HF912722.1), probably emerging from codon optimization usage. As a consequence, a putative promoter region has been created at position 680 and 700 in the gene coding for mSA-CBM3, with a putative +1 transcription start site at position 713 and an A/G-rich Shine Dalgarno-like region at position 745, making ATG_{763} a good candidate as an internal start codon (Fig. S8).

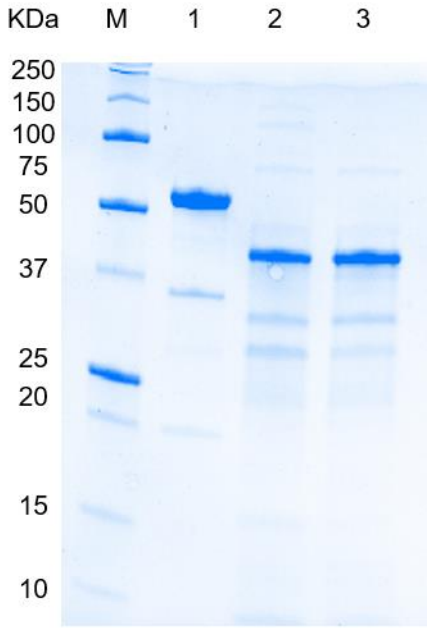

Fig. 3. SDS-PAGE of the proteins used in this study. Lanes: M, molecular mass markers; 1, CBM3-RFP; 2, mSA-CBM3; 3, mSA-CBM3-AzF.

Transcription initiated at this putative promoter and translation initiated at ATG_{763} produce a protein with theoretical molecular mass corresponding to the extra band observed on gel (Fig. S7). To confirm this hypothesis, ATG_{763} coding for methionine Met_{255} has been mutated into valine or glycine (GTG and GGC, respectively). The point mutation of Met_{255} to valine, an alternative start codon[65], reduced by 70 % the amount of truncated form of mSA-CBM3 (data not shown). On the other hand, mutation to glycine fully abrogated expression of truncated form. In addition, the deletion of the 6 nucleotides coding for Lys_{249} and Gly_{250} in the putative Shine Dalgarno area (745 to 750, Fig. S8) led also to the absence of truncated form of mSA-CBM3 (data not shown). Taken together, our data revealed the presence of an alternative translational start site, preceded by an active ribosome binding site. As a consequence, Met_{255} has been also replaced by a glycine in CBM3-RFP and mSA-CBM3-AzF coding sequence. Thereby, all three proteins were expressed in *E.coli* and obtained with high purity (> 90 %, Fig. 3). Typical amounts of pure proteins were about 73, 5 and 2 mg/L of culture for CBM3-RFP, mSA-CBM3 and mSA-CBM3-AzF, respectively.

Each individual part of mSA-CBM3-AzF displays active function

The biological activity of each part was assessed individually using CBM3-RFP, mSA-CBM3 and mSA-CBM3-AzF, in order to evaluate any potential deleterious cross effect. Firstly, the ability of CBM3 to bind to cellulose was confirmed using CBM3-RFP with a fast fluorescence test. Both purified CBM3-RFP and mRFP1 were incubated with insoluble regenerated amorphous cellulose (RAC) and the amount of fluorescence in the RAC pellet after centrifugation was measured (Fig. 4). We focused on RAC, the non-crystalline domains of cellulose, because it provides a large binding surface to CBM[66]. We observed a clear capture of the fluorescence in the RAC pellet incubated with CBM3-RFP (Fig. 4, A3), but not with the mRFP1 control (Fig. 4, A1), indicating that the former interacts efficiently with cellulose. Almost 88 % of the fluorescence is measured in the pellet when RAC was

incubated with CBM3-RFP (Fig. 4, B3), while on the other hand nearly 99 % of the fluorescence from mRF1 is measured in the supernatant (Fig. 4, B2).

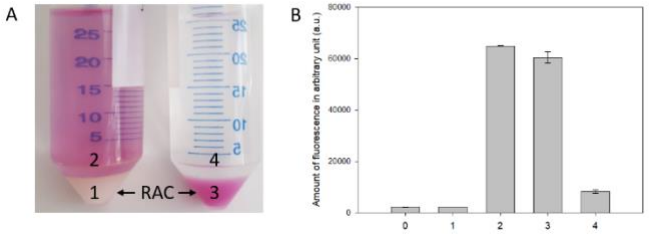

Fig. 4. Binding of CBM3-RFP to regenerated amorphous cellulose (RAC). **A:** Pellet of 2 mL of RAC 2% after having been mixed and incubated with mRFP1 (79.5 µM, 1) and CBM3-RFP (68.1 µM, 3). Fractions 2 and 4 correspond to the supernatant. **B:** Fluorescence intensity measured on RAC previously mixed with mRFP1 or CBM3-RFP. 0, Buffer; 1, bound fraction of mRFP1; 2, unbound fraction of mRFP1; 3, bound fraction of CBM3-RFP; 4, unbound fraction of CBM3-RFP. The values are shown as mean ± standard deviation of replicates (n = 3).

The ability of the isolated CBM3 domain to bind to RAC was then compared to CBM3-RFP and mSA-CBM3 to assess if the presence of RFP or mSA part at the *C*-or *N*-terminus of CBM3 alter its affinity to RAC (Fig. 5).

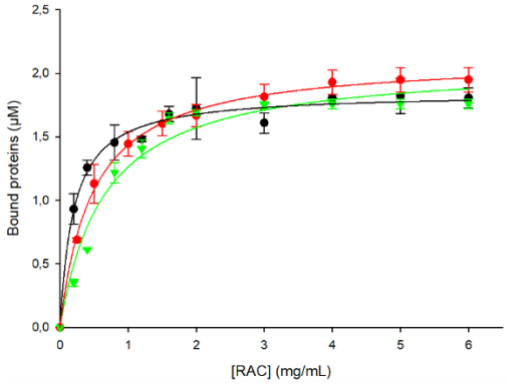

Fig. 5. Quantitative pull-down assay with RAC. Binding isotherm of CBM3 (black line), CBM3-RFP (red line) and mSA-CBM3 (green line) for an increasing concentration of regenerated amorphous cellulose (RAC). Data are represented as the amount of bound protein from the remaining measured fraction of unbound protein in the supernatant. The values are shown as mean ± standard deviation of replicates (n = 3).

The equilibrium binding constant K_d of CBM3, CBM3-RFP and mSA-CBM3 was determined as 0.20 ± 0.02, 0.47 ± 0.02 and 0.64 ± 0.12 mg/mL, respectively. The affinity of CBM3 to RAC is reduced by the presence of either mRFP or mSA, indicating a slight effect of the presence of an additional domain linked to both *N*- or *C*-terminus of CBM3. We cannot exclude that the differences observed between CBM3-RFP and mSA-CBM3 are linked to the nature of the fused domain, independently of their position relative to CBM3. However, the three proteins displayed a similar binding capacity of 1.84 ± 0.03, 2.12 ± 0.02 and 2.08 ± 0.09 µM, respectively, meaning that the size of the proteins shouldn't impact significantly the amount of CBM3 bound to RAC. This result suggests that flexible linker1 and 2 do not reduce the accessibility of CBM3 to the surface of RAC.

Beside cellulose binding, the second biological function presents in mSA-CBM3 is to specifically bind to biotinylated biomolecules, as the streptavidin part should display the ability to bind to biotin with a very high affinity. This function was assessed using binding to the fluorescent TMR-biocytin, that was first incubated with mSA-CBM3 to obtain TMR-mSA-CBM3 followed by performing and pull-down assays in presence of RAC to capture bound proteins. Fluorescence of both pellet and supernatant was measured. As shown in Fig. 6A, an excess of TMR-mSA-CBM3 accounting for 57,392 ± 983 a.u. of fluorescence (Fig. 6A, bar plot 1) was mixed to 2 % of RAC. After RAC pellet recovery, the remaining fluorescence in the supernatant was quantified as 9828 ± 415 a.u. (Fig. 6A, bar plot 2). Pellet washed two times released in total 4522 a.u. in the supernatant (Fig. 6A, bar plot 3 and 4). These measures indicate that about 75% of the initial fluorescence was captured in the RAC pellet, reflecting an efficient binding of TMR-mSA-CBM3 to cellulose. After incubation of the pellet of TMR-mSA-CBM3 associated to RAC with TEV protease, 42,505 ± 599 a.u. of fluorescence were recovered in the supernatant (Fig. 6A, bar plot 5), which corresponds to approximately 99 % of the fluorescence retained in the pellet. These results show that the TMR-mSA part can be efficiently released from cellulose-bound TMR-mSA-CBM3 by TEV protease. In parallel, SDS-PAGE analysis of fractions 1 to 5 was performed (Fig. 6B). Proteolysis of TMR-mSA-CBM3 by TEV released two protein fragments displaying apparent molecular weights of 27 and 13 kDa, corresponding to linker1-CBM3-linker2 and mSA, respectively (Fig. 6B lane 0). When applied to the RAC pellet incubated with TMR-mSA-CBM3, TEV cleavage released in the supernatant a fragment of 13 kDa (Fig. 6B lane 5) while from the insoluble RAC, a fragment of 27 kDa was generated (Fig. 6B lane 6). These results are in a good agreement with fluorescence intensity measurements of the different fractions from the pulldown assay.

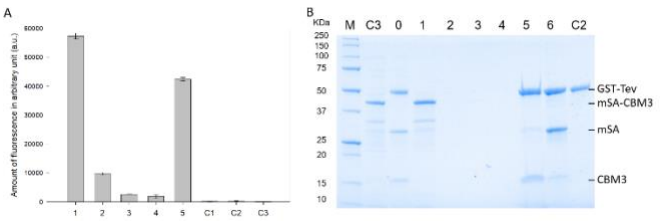

Fig. 6. Pull down assay of TMR-mSA-CBM3 in presence of 2 % RAC. **A:** Fluorescence intensity for 1, TMR-mSA-CBM3 before incubation with RAC; 2, unbound fraction of TMR-mSA-CBM3; 3 and 4, successive washes of the RAC; 5, supernatant after TEV proteolysis; C1, incubation buffer; C2, TEV buffer; C3, mSA-CBM3. The values are shown as mean ± standard deviation of replicates (n = 3). **B:** SDS-PAGE of the sample 1 to 5, C3 and C2. Lanes: M, molecular mass markers; C3, mSA-CBM3; 0, mSA-CBM3+TEV; 1, TMR-mSA-CBM3; 2, unbound fraction of TMR-mSA-CBM3; 3 and 4, successive washes of

the RAC; 5, supernatant after TEV proteolysis of RAC; 6, RAC after TEV proteolysis; C2, TEV buffer.

The third functional group carried by our tripartite protein corresponds to the introduction of a unique non-natural azide group brought by the introduction of AzF in the amino acid sequence of mSA-CBM3, hereafter referred to as mSA-CBM-AzF. The reactive azide group is able to react with a cyclooctyne group *via* a copper-free click-chemistry[38]. Alexa Fluor 488-DBCO (DBCO-AF488) contains such alkyne group and was used to fluorescently label mSA-CBM3-AzF. Data in Fig. 7 present the different steps of the purification of the labelled mSA-CBM3-AzF. About half of the fluorescence measured in the loaded fraction (Fig. 7A, bar plot 1, 65,000 a.u.) is recovered in the unbound fraction (Fig. 7A, bar plot 2, 35,621 a.u.) mainly in the form of unreacted fluorophore containing no trace of labelled mSA-CBM3-AzF (in agreement with the SDS-PAGE, Fig. 7B, lane 2). Successive washes released decreasing amounts of fluorescence (Fig. 7A, bar plot 3-5 with respectively 10,230, 2822 and 468 a.u.) corresponding to unreacted fluorophore (Fig. 7B, lanes 3-5). Bound fluorescence was recovered after two consecutive elutions with imidazole (Fig. 7A, bar plot 6 and 7, with 21,560 a.u. and 7895 a.u., respectively). Eluted fractions were pooled and extensively dialyzed as fraction 8. The fluorescence measured in this fraction is in the range of the sum of each eluted fraction (Fig. 7A, bar plot 8 with 25,640 a.u.) while SDS-PAGE analysis of the different fractions demonstrated that the fluorescence is associated to mSA-CBM3-AzF (Fig. 7B, lanes 6-8).

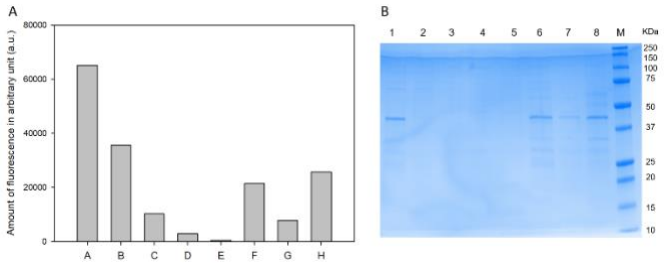

Fig. 7. Purification of mSA-CBM3-AzF by cobalt affinity chromatography after being labelled with Alexa Fluor 488-DBCO. A: Fluorescence measured for: 1, free labelled mSA-CBM3-AzF; 2, unbound labelled mSA-CBM3-AzF; 3 to 5, successive washes; 6 and 7, eluted labelled mSA-CBM3-AzF; 8, fraction 6 and 7 pooled and dialysed. B: SDS-PAGE of the samples 1 to 8.

Taken together, our data clearly demonstrated the availability of each single domain of mSA-CBM3-AzF to act individually on its dedicated ligand or target molecule.

RAC is recovered using fluorescent mSA-CBM3-AzF grafted on magnetic beads

The proof of concept of the versatility of our tripartite protein in functionalizing cellulose was demonstrated in two steps. First, mSA-CBM3-AzF was grafted to magnetic beads in order to specifically recover RAC using a magnet. Hence, mSA-CBM3-AzF was primary immobilized on magnetic beads activated with DBCO groups by click-chemistry. The mSA-CBM3-AzF coated magnetic beads (hereafter referred to as mSA-CBM3-MB) were incubated with RAC, to assess their binding capabilities to bind to RAC through the CBM3 moiety. Effect of a magnet on RAC recovery with mSA-CBM3-MB was compared to mSA-CBM3 or bare beads (Fig. S9). After 10 s in contact with a magnet, mSA-CBM3-MB displayed a more compact form (Fig. S9, A_{10}) than bare beads with (Fig. S9, B_{10}) or without mSA-CBM3 (Fig. S9, C_{10}), the last two presenting a similar aspect. Furthermore, the supernatant was clearer with mSA-CBM3-MB than with mSA-CBM3 or bare beads. We hypothesize that, as mSA-CBM3-MB is bound to RAC, both magnetic beads and cellulose are in the mobile phase, whereas with no interaction between beads and RAC, the latter prevents the beads from moving. We then analysed the fully functionalized version TMR-mSA-CBM3-MB. The mSA part was labelled with TMR prior being grafted *via* AzF on magnetic bead. Similar experiments as described above were performed using a home-made optical bench (Fig. S1) to quantify the dynamic of RAC movement under magnetic field. Results were compared to bare beads. Experiments followed with the digital camera, made in triplicate with the same couple of samples (Fig. 8 and Supplementary Information) provide an example of the evolution of the magnetic recovery front from the third run. Curves from Fig.8 represent the evolution over time of a selected area of the sedimentation front position of the different samples as described in Figs. S3 and S4.

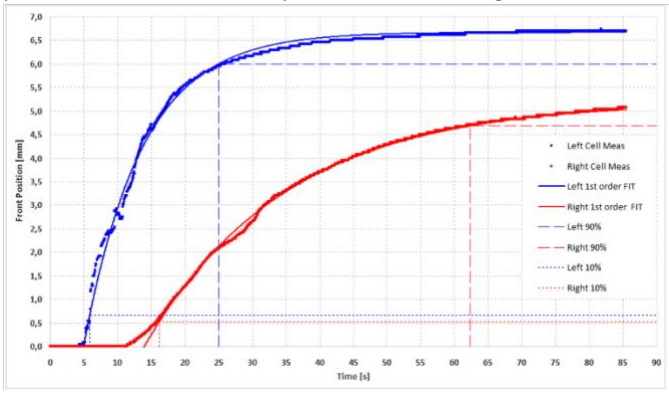

Fig. 8. Dynamic evolution of the sedimentation front followed by a digital camera. The spectrophotometer cuvettes were filled with 600 µL of RAC (2 %) in the presence of 80 µl of magnetic beads grafted with TMR-mSA-CBM3-MB (in blue) or bare beads (in red). See Fig. S2 and Fig. S3 for correspondence with left cell and right cell. Samples were submitted to 4 cycles of mixing and magnetic sedimentation. Data from the third run, collected at 25 frames per second, are shown. Raw data were plotted as diamond dots and the model as solid line. The value of the rising time T_r was determined from Eq.2 (Supporting Information) between 10 and 90 % of the maximum value of the sedimentation front position. The period of time before the insertion of the magnet was not taken into account (between 10 and 20 s).

The modelized first order curves present a very good fit with our data, which clearly indicates that the RAC recovery front displacement is faster in the presence of TMR-mSA-CBM3-MB compared to bare beads. From the model, we determined the starting time t_0 from which the recovery front begins to move, the τ_{fit} reflecting the time constant that best fits the data, the rising time T_r and the recovery time constant τ (Table 2).

Table 2 Temporal parameters of the magnetic recovery front displacement. Magnetic beads grafted or not with TMR-mSA-CBM3-MB were both incubated with 2 % RAC. First order curve models were used to determine: t_0, the starting time of the front displacement; τ_{fit}, reflecting the time constant that best fits the data to the model; T_r the rising time; τ the sedimentation time constant. The values are shown as ± mean standard deviations out of 4 replicates.

Magnetic beads	t_0 (s)	τ_{fit} (s)	T_r (s)	τ (s)
TMR-mSA-CBM3-MB	5.5 ± 0.5	8.7 ± 0.1	19.2 ± 0.2	8.7 ± 0.1
Bare beads	13.7 ± 0.5	20.2 ± 0.7	45.6 ± 0.9	20.7 ± 0.4

Firstly, our data indicate that the onset of the movement of the recovery front t_0 begins 5.5 s after insertion of the magnet in presence of TMR-mSA-CBM3-MB, while it begins after 13.7 s for the bare beads. It is also evident that the value of the rising time T_r is 2.4 fold lower for the grafted beads than for the bare beads, reflecting a faster movement of the recovery front in the first case. As hypothesised above, unbound RAC probably limit the displacement of bare beads. However, even if the rising time T_r for the bare beads is higher than for the grafted beads (Table 2), this number is unexpectedly high, reflecting nevertheless a movement of the RAC, even if no interaction between the substrate and the beads is evident. We assume that 2 % of RAC is a concentration sufficiently high enough to allow the beads (mean bead diameter 0.97 µm compared to ≈100 nm for RAC) to induce a RAC displacement on their own movement under magnet. Anyway, the existence of a strong vortex in the presence of grafted beads (Electronic Supplementary Information Movie S1) and the value of the t_0 (Table 2), clearly indicate a different phenomenon. Furthermore, the experiments were replicated four times with the same samples, reflecting the robustness of the data. In addition, the similarity between the time constant τ_{fit} and τ demonstrates the very good correlation between the chosen model and our data. Unfortunately, the fluorescence carried by TMR-mSA-CBM3-MB was not intense enough to be captured by our digital camera. It has been measured *ex temporarily* on the fourth run of the dynamic experiment with a fluorimeter (Fig. S10). Data indicate that the fluorescence measured in the fully recovered pellet beads is significantly higher than in the supernatant (37,001 and 864 a.u., respectively). Altogether, the optical bench results and fluorescence measurement indicates that we were able to graft our fluorescent tripartite protein mSA-CBM3-AzF on magnetic beads and specifically recover RAC from the solution simply by using magnet.

Discussion

In this work, we created and characterized a tripartite protein whose originality lies in its ability to be functionalized on a custom basis, either with biological or chemical molecules and to provide a protein platform to functionalize cellulose. The protein was designed around the CBM family 3a from *Clostridium thermocellum*, known to provide a strong affinity for crystalline and non-crystalline cellulose, a renewable carbon-based polymer widely used in biotechnology. Flanked to CBM3 by its native flexible linkers, the chimeric protein displayed a streptavidin module (mSA) at the *N*-terminus and an AzF at the *C*-terminus, resulting in mSA-CBM3-AzF. The ability to retain strong affinity for cellulose was assessed against RAC, a soluble derivative of cellulose. Neither the presence of the additional mSA at the *N*-terminus of CBM3 nor the presence of the fluorescent protein RFP1 at the *C*-terminus is deleterious for RAC interaction. These results are in agreement with previous studies exemplifying the use of CBM3 in fusion with a large variety of proteins such as lytic polysaccharide monooxygenase[67], antimicrobial peptide[68], or eGFP[69]. Similarly, in our study the additional activity was evidenced by labelling mSA-CBM3 with a fluorophore TMR-biocytin at the *N*-terminus, a fluorophore DBCO-AF488 or a cyclooctyne-coated magnetic beads using AzF introduced at the *C*-terminus of our tripartite protein. We went yet another step further by introducing two additional functions on CBM-based hybrid protein, grafting TMR-biocytin/mSA-CBM3-AzF to cyclooctyne-coated magnetic beads. The reason for this was to propose a proof of concept for multi-functionalization of cellulose, possibly bringing two functions to interact to each other in a geometric arrangement in order to facilitate for example cascade reaction, as previously published[70]. In nature, the family 3a CBM allows the cellulosome to target lignocellulose and direct the action of the enzymatic complex towards its substrate[71]. The reason for the catalytic efficiency of the cellulosome is provided by the complementary between the GHs anchored to the scaffoldin but also by the spatial proximity between the enzymes[72]. A major concern is the high degree of flexibility of the cellulosome due to the presence of linker regions, preventing crystallographic structure of the ones flanking CBM3[43]. We therefore proposed to model mSA-CBM3. Our results confirmed the high degree of flexibility of the linker regions. However, the distance values between mSA and Phe_{364} in linker2 (ranging from 20 to 120 Å) are compatible with channelling events between two contiguous enzymes[37,73,74]. Interestingly, the fact that none of the modelled conformations present any steric hindrance with the cellulose binding site of CBM3 is in agreement with the similar affinity of our different constructs to bind to cellulose. It is also interesting to note that the relative position of the flanking regions mSA and linker2 to CBM3 is compatible with the X-ray structure analysis of the cellulose bind domain expressed with its linker[43]. Thus, the multi-functionalized cellulose proof of concept was experimented by collecting RAC using magnetic beads grafted with fluorescent TMR-mSA-CBM3-AzF. Our design provides a single point association with each part, allowing to specifically control the molarity of the functionalization or control the number of contact with the beads. Finally, we clearly developed a platform CBM based protein interacting with cellulose and displaying also a highly versatile functionalization. The two additional modules, a streptavidin and an

azide group, allow to graft biotinylated protein and chemical molecule, respectively. This tripartite chimeric protein is thus able to spatially localized two different molecules on cellulose. We expect many applications of this platform in biotechnology, from innovating protein scaffolding, increasing possibilities in enzymatic cascade reaction, to functionalized material such as cellulose fibres displaying biological or physical properties.

Author Contributions

A. P., G. B. Y. B., C. B., B. E., R. F., A. K. H., S. H., P. M., J. P., M. T., A. V., C. W., Y. R., C. Y. M. were involved in the conceptualization, data curation and formal analysis of the data. A. P., G. B., , Y. B., C. B., , J. P., M. T., , A. V., C. W., run the experiments. S.B., Y. B., J. C., J. E., A. E. performed molecular modeling and analysis of the data. A. P., G. B., C L-M., C. Y. M. performed the experiments on the optical bench and C L-M. analyzed the results. B. E., R. F., A. K. H., S. H., P. M., Y. R., C. Y. M., S. N., G. T. provided methodology development and supervision. C. Y. M. wrote the initial draft. B E., R. F., A. K. H., S. H., P. M., Y. R., C. Y. M., S. N. commented and reviewed the draft.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was initiated by the Toulouse INSA-UPS team which participated in the 2018 iGEM competition (http://2018.igem.org/Team:Toulouse-INSA-UPS). The team was composed of G.B, A. P., Y. B., C. B., A. V., J. P., M. T., C. W., S. B., R. F., A. H., P. M., C.Y. M., Y. R., S. H., and B. E. This work was supported by academic and private sponsors whose list is available here: http://2018.igem.org/Team:Toulouse-INSA-UPS/Sponsors. This work was granted access to the HPC resources on the Computing mesocenter of Région Midi-Pyrénées (CALMIP, Toulouse, France).

References

1 C. Gilbert and T. Ellis, *ACS Synth. Biol.*, 2019, **8**, 1–15.

2 J. Xue, X. Wang, E. Wang, T. Li, J. Chang and C. Wu, *Acta Biomater.*, 2019, **100**, 270–279.

3 C. Wang, M. Wang, T. Xu, X. Zhang, C. Lin, W. Gao, H. Xu, B. Lei and C. Mao, *Theranostics*, 2019, **9**, 65–76.

4 P. Ross, R. Mayer and M. Benziman, *Microbiol. Rev.*, 1991, **55**, 35–58.

5 S. Kimura, W. Laosinchai, T. Itoh, X. Cui, C. R. Linder and R. M. Brown, *Plant Cell*, 1999, **11**, 2075–2085.

6 L. C. Duchesne and D. W. Larson, *Bioscience*, 1989, **39**, 238–241.

7 Y. Nishiyama, 2014, pp. 5–17.

8 L. Motelica, D. Ficai, A. Ficai, O. C. Oprea, D. A. Kaya and E. Andronescu, *Foods*, 2020, **9**, 1438.

9 S. J. Eichhorn, C. A. Baillie, N. Zafeiropoulos, L. Y. Mwaikambo, M. P. Ansell, A. Dufresne, K. M. Entwistle, P. J. Herrera-Franco, G. C. Escamilla, L. Groom, M. Hughes, C. Hill, T. G. Rials and P. M. Wild, *J. Mater. Sci.*, 2001, **36**, 2107–2131.

10 P. Lopez Hurtado, A. Rouilly, V. Vandenbossche and C. Raynaud, *Build. Environ.*, 2016, **96**, 170–177.

11 R. J. Hickey and A. E. Pelling, *Front. Bioeng. Biotechnol.*, 2019, 7, 45.

12 Y. Zhu, X. Xu, N. D. Brault, A. J. Keefe, X. Han, Y. Deng, J. Xu, Q. Yu and S. Jiang, *Anal. Chem.*, 2014, **86**, 2871–2875.

13 J. Shokri and K. Adibki, in *Cellulose - Medical, Pharmaceutical and Electronic Applications*, InTech, 2013.

14 J. Tang, J. Sisler, N. Grishkewich and K. C. Tam, *J. Colloid Interface Sci.*, 2017, **494**, 397–409.

15 L. Bacakova, J. Pajorova, M. Tomkova, R. Matejka, A. Broz, J. Stepanovska, S. Prazak, A. Skogberg, S. Siljander and P. Kallio, *Nanomaterials*, 2020, **10**, 196.

16 W. Y. Hamad, C. Miao and S. Beck, *Ind. Biotechnol.*, 2019, **15**, 133–137.

17 C. Calvino, N. Macke, R. Kato and S. J. Rowan, *Prog. Polym. Sci.*, 2020, **103**, 101221.

18 Y. Yoo and J. P. Youngblood, *ACS Sustain. Chem. Eng.*, 2016, **4**, 3927–3938.

19 W. Wan, D. Wang, X. Gao and J. Hong, *Appl. Microbiol. Biotechnol.*, 2011, **91**, 789–798.

20 A. Almeida, A. M. M. Rosa, A. M. Azevedo and D. M. F. Prazeres, *J. Mol. Recognit.*, 2017, **30**, e2634.

21 F. K. Andrade, R. Costa, L. Domingues, R. Soares and M. Gama, *Acta Biomater.*, 2010, **6**, 4034–4041.

22 C. I. P. D. Guerreiro, C. M. G. A. Fontes, M. Gama and L. Domingues, *Protein Expr. Purif.*, 2008, **59**, 161–168.

23 A. B. Boraston, D. N. Bolam, H. J. Gilbert and G. J. Davies, *Biochem. J.*, 2004, **382**, 769–781.

24 J. Tormo, R. Lamed, a J. Chirino, E. Morag, E. a Bayer, Y. Shoham and T. a Steitz, *EMBO J.*, 1996, **15**, 5739–51.

25 C. Montanier, J. E. Flint, D. N. Bolam, H. Xie, Z. Liu, A. Rogowski, D. P. Weiner, S. Ratnaparkhe, D. Nurizzo, S. M. Roberts, J. P. Turkenburg, G. J. Davies and H. J. Gilbert, *J. Biol. Chem.*, 2010, **285**, 31742–31754.

26 F. Cuskin, J. E. Flint, T. M. Gloster, C. Morland, A. Basle, B. Henrissat, P. M. Coutinho, A. Strazzulli, A. S. Solovyova, G. J. Davies and H. J. Gilbert, *Proc. Natl. Acad. Sci.*, 2012, **109**, 20889–20894.

27 C. Oliveira, V. Carvalho, L. Domingues and F. M. Gama, *Biotechnol. Adv.*, 2015, **33**, 358–369.

28 K. Aïssa, M. A. Karaaslan, S. Renneckar and J. N. Saddler, *Biomacromolecules*, 2019, **20**, 3087–3093.

29 T. Kitaoka and H. Tanaka, *J. Wood Sci.*, 2001, **47**, 322–324.

30 A. Elter, T. Bock, D. Spiehl, G. Russo, S. C. Hinz, S. Bitsch, E. Baum, M. Langhans, T. Meckel, E. Dörsam, H. Kolmar and G. Schwall, *Sci. Rep.*, 2021, **11**, 7880.

31 C. Moreau, A. Villares, I. Capron and B. Cathala, Ind. Crops Prod., 2016, 93, 96–107.

32 D. Poole, FEMS Microbiol. Lett., 1992, 99, 181–186.

33 J. K. Mann, D. Demonte, C. M. Dundas and S. Park, Technology, 2016, 04, 152–158.

34 C. M. Dundas, D. Demonte and S. Park, Appl. Microbiol. Biotechnol., 2013, 97, 9343–9353.

35 Q. Le, V. Nguyen and S. Park, Appl. Microbiol. Biotechnol., 2019, 103, 7355–7365.

36 L. Chaiet and F. J. Wolf, Arch. Biochem. Biophys., 1964, 106, 1–5.

37 C. C. Liu and P. G. Schultz, Annu. Rev. Biochem., 2010, 79, 413–444.

38 L. Wang, J. Xie and P. G. Schultz, Annu. Rev. Biophys. Biomol. Struct., 2006, 35, 225–249.

39 N. J. Agard, J. A. Prescher and C. R. Bertozzi, J. Am. Chem. Soc., 2004, 126, 15046–15047.

40 J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang and P. G. Schultz, J. Am. Chem. Soc., 2002, 124, 9026–9027.

41 J. Hong, X. Ye, Y. Wang and Y.-H. P. Zhang, Anal. Chim. Acta, 2008, 621, 193–199.

42 X. Jia, Y. Chen, C. Shi, Y. Ye, P. Wang, X. Zeng and T. Wu, J. Agric. Food Chem., 2013, 61, 12405–12414.

43 O. Yaniv, E. Morag, I. Borovok, E. A. Bayer, R. Lamed, F. Frolow and L. J. W. Shimon, Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun., 2013, 69, 733–737.

44 D. DeMonte, E. J. Drake, K. H. Lim, A. M. Gulick and S. Park, Proteins Struct. Funct. Bioinforma., 2013, 81, 1621–1633.

45 A. Estaña, N. Sibille, E. Delaforge, M. Vaisset, J. Cortés and P. Bernadó, Structure, 2019, 27, 381-391.e2.

46 T. Harder, W. Boomsma, M. Paluszewski, J. Frellsen, K. E. Johansson and T. Hamelryck, BMC Bioinformatics, 2010, 11, 306.

47 H. G. D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, C. A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, I. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, J. S. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith and L. X. and P. A. K. R.C. Walker, J. Wang, R.M. Wolf, X. Wu, 2016.

48 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, J. Chem. Theory Comput., 2015, 11, 3696–3713.

49 A. Onufriev, D. Bashford and D. A. Case, Proteins Struct. Funct. Bioinforma., 2004, 55, 383–394.

50 D. R. Roe and T. E. Cheatham, J. Chem. Theory Comput., 2013, 9, 3084–3095.

51 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, J. Chem. Phys., 1984, 81, 3684–3690.

52 T. Darden, D. York and L. Pedersen, J. Chem. Phys., 1993, 98, 10089–10092.

53 W. F. van Gunsteren and H. J. C. Berendsen, Mol. Phys., 1977, 34, 1311–1327.

54 R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand and R. C. Walker, J. Chem. Theory Comput., 2013, 9, 3878–3888.

55 M. C. Hernandez-Gomez, M. G. Rydahl, A. Rogowski, C. Morland, A. Cartmell, L. Crouch, A. Labourel, C. M. G. A. Fontes, W. G. T. Willats, H. J. Gilbert and J. P. Knox, FEBS Lett., 2015, 589, 2297–2303.

56 Y. Li and R. Sousa, Biotechnol. Lett., 2012, 34, 1629–1635.

57 C. Noren, S. Anthony-Cahill, M. Griffith and P. Schultz, Science (80-.)., 1989, 244, 182–188.

58 V. Rigolot, C. Biot and C. Lion, Angew. Chemie Int. Ed., 2021, anie.202101502.

59 Y. J. Bomble, G. T. Beckham, J. F. Matthews, M. R. Nimlos, M. E. Himmel and M. F. Crowley, J. Biol. Chem., 2011, 286, 5614–5623.

60 Y. Vazana, Y. Barak, T. Unger, Y. Peleg, M. Shamshoum, T. Ben-Yehezkel, Y. Mazor, E. Shapiro, R. Lamed and E. A. Bayer, Biotechnol. Biofuels, 2013, 6, 182.

61 V. Burger, T. Gurry and C. Stultz, Polymers (Basel)., 2014, 6, 2684–2719.

62 I. Le Trong, Z. Wang, D. E. Hyre, T. P. Lybrand, P. S. Stayton and R. E. Stenkamp, Acta Crystallogr. Sect. D Biol. Crystallogr., 2011, 67, 813–821.

63 A. Barozet, M. Bianciotto, M. Vaisset, T. Siméon, H. Minoux and J. Cortés, Proteins Struct. Funct. Bioinforma., 2021, 89, 218–231.

64 D. E. Johnson, H. Ai, P. Wong, J. D. Young, R. E. Campbell and J. R. Casey, J. Biol. Chem., 2009, 284, 20499–20511.

65 F. Belinky, I. B. Rogozin and E. V. Koonin, Sci. Rep., 2017, 7, 12422.

66 J. Hong, Y. Wang, X. Ye and Y.-H. P. Zhang, J. Chromatogr. A, 2008, 1194, 150–154.

67 L. I. Crouch, A. Labourel, P. H. Walton, G. J. Davies and H. J. Gilbert, J. Biol. Chem., 2016, 291, 7439–7449.

68 R. Ramos, S. Moreira, A. Rodrigues, M. Gama and L. Domingues, Biotechnol. Prog., 2013, 29, 17–22.

69 Y. Hébert-Ouellet, F. Meddeb-Mouelhi, V. Khatri, L. Cui,

B. Janse, K. MacDonald and M. Beauregard, Green Chem., 2017, 19, 2603–2611.

70 M. Liu, J. Fu, X. Qi, S. Wootten, N. W. Woodbury, Y. Liu and H. Yan, ChemBioChem, 2016, 17, 1097–1101.

71 E. A. Bayer, J.-P. Belaich, Y. Shoham and R. Lamed, Annu. Rev. Microbiol., 2004, 58, 521–554.

72 C. M. G. A. Fontes and H. J. Gilbert, Annu. Rev. Biochem., 2010, 79, 655–681.

73 T. A. Ngo, E. Nakata, M. Saimura and T. Morii, J. Am. Chem. Soc., 2016, 138, 3012–3021.

74 A. Küchler, M. Yoshimoto, S. Luginbühl, F. Mavelli and P. Walde, Nat. Nanotechnol., 2016, 11, 409–420.