SUPPLEMENTARY INFORMATION FOR

A tripartite carbohydrate-binding module to functionalize cellulose nanocrystal

Angeline Pelus[1‡], Gaëlle Bordes[2‡], Sophie Barbe[1], Younes Bouchiba[1], Callum Burnard[1], Juan Cortés[3], Brice Enjalbert[1], Jeremy Esque[1], Alejandro Estaña[3], Régis Fauré[1], Anthony K. Henras[2], Stéphanie Heux[1], Claude Le Men[1], Pierre Millard[1], Sébastien Nouaille[1], Julien Pérochon[1], Marion Toanen[1], Gilles Truan[1], Amandine Verdier[1], Camille Wagner[1], Yves Romeo[2]*, Cédric Y. Montanier[1]*

[‡]These authors contributed equally to this work

*Correspondence: yves.romeo@univ-tlse3.fr, Tel. +33 (0)5 61 33 58 08 - cedric.montanier@insa-toulouse.fr, Tel. +33 (0)5 61 55 97 13.

[1] TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

[2] Molecular, Cellular and Developmental biology department (MCD), Centre de Biologie Integrative (CBI), Université de Toulouse, CNRS, UPS, 31062 Toulouse, France.

[3] LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

This Supplementary Information contains 5 sections:

1. Cloning of the DNA sequences (Table S1-S2)

2. Dynamic of RAC sedimentation by magnetic beads grafted with fluorescent mSA-CBM3-AzF (Figures S2-S4). See Electronic Supplementary Information for Movie S1 and S2

3. Molecular modelling assesses the flexibility of mSA-CBM3 (Table S3 and Figure S7-S8)

4. Expression and purification of the chimeric CBM3 based proteins (Figures S8-S11, Table S5)

5. RAC is recovered using fluorescent mSA-CBM3-AzF grafted on magnetic beads (Figure S9-S10)

1. Cloning of the DNA sequences encoding the hybrid proteins

At position 153 of *CBM3a* (C to T exchange), HindIII restriction site at position 241 was removed (CTC → CTG). Undesired restriction sites present in the gene coding for mRFP1 were also removed by punctual mutation: HindIII at position 209 (GCT → GCG), NcoI at position 419 (ACC → ACG) and HindIII at position 431 (GCT → GCG).

The sequence coding for the monomeric hybrid streptavidin mSA2 from *Streptomyces avidinii* (Part:BBa_K1896000) was fused to the *N*-terminal of CBM3a with an additional sequence coding for a TEV protease site (GAA AAC CTG TAT TTT CAG GGC). The corresponding construction mSA2-Tev-linker1-CBM3a-linker2-His_6 (pET28-mSA2-CBM3 for short) is detailed in Table S2. In order to generate mSA2-Tev-linker1-CBM3a-linker2-AzF-His_6 (pET28-mSA2-CBM3a-AzF for short), codon TTT (Phe_{364}) was replaced by a stop amber codon TAG at position 1090 of *mSA2-Tev-linker1-CBM3a-linker2-His_6* (Part:BBa_K2668010 , Table S2). Apart from point modifications of the DNA sequences (see Table S2 for details), all the genes correspond to the wild type sequences, except for CBM3, whose DNA sequence has been optimized to more closely reflect the codon usage of the recipiendary host organism (http://parts.igem.org as Part:BBa_K1321014).

Table S1: Primers used in construction of fusion proteins

pET28-mRFP1	For	5'- AGGAGATATACCATGGCTTCCTCCGAAGACGTTATCAAAG-3'
	Rev	5'- GTGCGGCCGCAAGCTTAGCACCGGTGGAGTGACG-3'
CBM3-RFP1	For	5'-GGACGATCCGATGGCTTCCTCCGAAGACG-3'
	Rev	5'-AGGAAGCCATCGGATCGTCCAATGATGGAGGAATC-3'
pET28-CBM3-mRFP1	For	5'- TAAGAAGGAGATATACCATGAATGCTACGCCAACTAAGG GTGC-3'
	Rev	5'- CTCGAGTGCGGCCGCAAGCTTAGCACCGGTGGAGTGACG-3'
CBM3-Met_{127}→Gly_{127}	For	5'- CGTAAAGGGCAGCTCAAGCACAAATAACGC-3'
	Rev	5'- GAGCTGCCCTTTACGAATGTTCCTTTTACATTT-3'
pET28-mSA-CBM3	For	5'- TAAGAAGGAGATATACCATGGCGGAAGCGGGTATCACC-3'
	Rev	5'- CTCGAGTGCGGCCGCAAGCTTCGGATCGTCCTATGATGGAGG-3'
pET28-mSA-CBM3-AzF	For	5'- CCATCATAGGACGATCCGAAGCTTGCG-3'
	Rev	5'- ATCGTCCTATGATGGAGGAATCGTGGTAGCC-3'

Table S2: Details of the DNA sequences and corresponding primary sequences of the chimeric proteins used in this study. Linker1 sequence is in green, linker2 sequence is in blue, CBM3 sequence is in black, mRFP1 sequence is in purple, mSA sequence is in purple and TEV protease site sequence is underlined in black. Ambre codon in linker2 DNA sequence is in bold and corresponding emplacement in primary sequence is symbolized by *.

Construct	DNA sequence	Protein
linker1-*Ct*CBM3a-linker2-RFP1-His_6 Part:BBa_K2668020	atgaatgctacgccaactaag ggtgcaaccccgaccaacaca gcaacgcctacaaaaagcgct acagcaacacctacaagaccg tcagttcctacaaacacaccg actaacacaccggcaaataca cctgtttcaggcaacttgaag gtcgaattttataactcaaat ccgagtgatacaactaacagt attaatccgcagtttaaagta acaaatacaggatcaagtgca attgatctttcaaagctgaca ttgagatactattacaccgtt gatggccagaaggaccagact ttctggtgtgaccatgcagct atcataggtagcaacggctca tacaacggcatcacatcaaat gtaaaaggaacattcgtaaag ggcagctcaagcacaaataac gcagacacatacctcgaaata	MNATPTKGATPTNTAT PTKSATATPTRPSVPT NTPTNTPANTPVSGNL KVEFYNSNPSDTTNSI NPQFKVTNTGSSAIDL SKLTLRYYYTVDGQKD QTFWCDHAAIIGSNGS YNGITSNVKGTFVKGS SSTNNADTYLEISFTG GTLEPGAHVQIQGRFA KNDWSNYTQSNDYSFK SASQFVEWDQVTAYLN GVLVWGKEPGGSVVPS TQPVTTPPATTKPPAT TKPPATTIPPSLDDPM ASSEDVIKEFMRFKVR MEGSVNGHEFEIEGEG EGRPYEGTQTAKLKVT KGGPLPFAWDILSPQF QYGSKAYVKHPADIPD

	agtttcacaggtggcactttg gaacctggtgctcatgtacag atacagggtaggtttgcgaaa aatgactggagtaattataca cagtcaaatgattactcattt aagtcagcatcacagttcgta gaatgggatcaggttacagca tatttgaatggagtacttgta tggggtaaagaaccaggagga tcagtagttccgtcaacacag ccggtaacaaccccaccggca acaaccaagccgccagcaaca accaaaccaccggctaccacg attcctccatcagacgatccg atggcttcctccgaagacgtt atcaaagagttcatgcgtttc aaagttcgtatggaaggttcc gttaacggtcacgagttcgaa atcgaaggtgaaggtgaaggt cgtccgtacgaaggtacccag accgctaaactgaaagttacc aaaggtggtccgctgccgttc gcttgggacatcctgtccccg cagttccagtacggttccaaa gcgtacgttaaacacccggct gacatcccggactacctgaaa ctgtccttcccggaaggtttc aaatgggaacgtgttatgaac ttcgaagacggtggtgttgtt accgttacccaggactcctcc ctgcaagacggtgagttcatc tacaaagttaaactgcgtggt accaacttcccgtccgacggt ccggttatgcagaaaaaaacG atgggttgggaagcgtccacc gaacgtatgtacccggaagac ggtgctctgaaaggtgaaatc aaaatgcgtctgaaactgaaa gacggtggtcactacgacgct gaagttaaaaccacctacatg gctaaaaaaccggttcagctg ccgggtgcttacaaaaccgac atcaaactggacatcacctcc cacaacgaagactacaccatc gttgaacagtacgaacgtgct gaaggtcgtcactccaccggt gctaagcttgcggccgcactc gagcaccaccaccaccaccac	YLKLSFPEGFKWERVM NFEDGGVVTVTQDSSL QDGEFIYKVKLRGTNF PSDGPVMQKKTMGWEA STERMYPEDGALKGEI KMRLKLKDGGHYDAEV KTTYMAKKPVQLPGAY KTDIKLDITSHNEDYT IVEQYERAEGRHSTGA KLAAALEHHHHHH
mSA2-Tev-linker1-*Ct*CBM3a-linker2-His$_6$	atggcggaagcgggtatcacc ggcacgtggtacaaccagcat ggttctaccttcaccgttacc	MAEAGITGTWYNQHGS TFTVTAGADGNLTGQY ENRAQGTGCQNSPYTL

	gcgggtgcggacggtaacctg accggccagtacgaaaaccgt gcgcagggcactggttgccag aactctccgtacaccctgacc ggccgttacaacggtaccaaa ctggaatggcgtgttgaatgg aacaactctaccgaaaactgc cactctcgtaccgaatggcgt ggtcagtaccagggtggtgcg gaagcgcgtatcaacacccag tggaacctgacctacgaaggt ggttctggtccggcgaccgaa cagggtcaggacaccttcacc aaagttaaaccgtctgcggcg tccaccggtgaaaacctgtat tttcagggcaatgctacgcca actaagggtgcaaccccgacc aacacagcaacgcctacaaaa agcgctacagcaacacctaca agaccgtcagttcctacaaac acaccgactaacacaccggca aatacacctgtttcaggcaac ttgaaggtcgaattttataac tcaaatccgagtgatacaact aacagtattaatccgcagttt aaagtaacaaatacaggatca agtgcaattgatctttcaaag ctgacattgagatactattac accgttgatggccagaaggac cagactttctggtgtgaccat gcagctatcataggtagcaac ggctcatacaacggcatcaca tcaaatgtaaaaggaacattc gtaaagggcagctcaagcaca aataacgcagacacatacctc gaaataagtttcacaggtggc actttggaacctggtgctcat gtacagatacagggtaggttt gcgaaaaatgactggagtaat tatacacagtcaaatgattac tcatttaagtcagcatcacag ttcgtagaatgggatcaggtt acagcatatttgaatggagta cttgtatggggtaaagaacca ggaggatcagtagttccgtca acacagccggtaacaacccca ccggcaacaaccaagccgcca gcaacaaccaaaccaccggct accacgattcctccatcattt gacgatccgaagcttgcggcc	TGRYNGTKLEWRVEWN NSTENCHSRTEWRGQY QGGAEARINTQWNLTY EGGSGPATEQGQDTFT KVKPSAASTGENLYFQ GNATPTKGATPTNTAT PTKSATATPTRPSVPT NTPTNTPANTPVSGNL KVEFYNSNPSDTTNSI NPQFKVTNTGSSAIDL SKLTLRYYYTVDGQKD QTFWCDHAAIIGSNGS YNGITSNVKGTFVKGS SSTNNADTYLEISFTG GTLEPGAHVQIQGRFA KNDWSNYTQSNDYSFK SASQFVEWDQVTAYLN GVLVWGKEPGGSVVPS TQPVTTPPATTKPPAT TKPPATTIPPSFDDPK LAAALEHHHHHH

	gcactcgagcaccaccaccac caccac	
mSA2-Tev-linker1-*Ct*CBM3a-linker2-AzF-His_6 Part:BBa_K2668010	atggcggaagcgggtatcacc ggcacgtggtacaaccagcat ggttctaccttcaccgttacc gcgggtgcggacggtaacctg accggccagtacgaaaaccgt gcgcagggcactggttgccag aactctccgtacaccctgacc ggccgttacaacggtaccaaa ctggaatggcgtgttgaatgg aacaactctaccgaaaactgc cactctcgtaccgaatggcgt ggtcagtaccagggtggtgcg gaagcgcgtatcaacacccag tggaacctgacctacgaaggt ggttctggtccggcgaccgaa cagggtcaggacaccttcacc aaagttaaaccgtctgcggcg tccaccggtgaaaacctgtat tttcagggcaatgctacgcca actaagggtgcaaccccgacc aacacagcaacgcctacaaaa agcgctacagcaacacctaca agaccgtcagttcctacaaac acaccgactaacacaccggca aatacacctgtttcaggcaac ttgaaggtcgaattttataac tcaaatccgagtgatacaact aacagtattaatccgcagttt aaagtaacaaatacaggatca agtgcaattgatctttcaaag ctgacattgagatactattac accgttgatggccagaaggac cagactttctggtgtgaccat gcagctatcataggtagcaac ggctcatacaacggcatcaca tcaaatgtaaaaggaacattc gtaaagggcagctcaagcaca aataacgcagacacatacctc gaaataagtttcacaggtggc actttggaacctggtgctcat gtacagatacagggtaggttt gcgaaaaatgactggagtaat tatacacagtcaaatgattac tcatttaagtcagcatcacag ttcgtagaatgggatcaggtt acagcatatttgaatggagta cttgtatggggtaaagaacca ggaggatcagtagttccgtca acacagccggtaacaacccca	MAEAGITGTWYNQHGS TFTVTAGADGNLTGQY ENRAQGTGCQNSPYTL TGRYNGTKLEWRVEWN NSTENCHSRTEWRGQY QGGAEARINTQWNLTY EGGSGPATEQGQDTFT KVKPSAASTGENLYFQ GNATPTKGATPTNTAT PTKSATATPTRPSVPT NTPTNTPANTPVSGNL KVEFYNSNPSDTTNSI NPQFKVTNTGSSAIDL SKLTLRYYYTVDGQKD QTFWCDHAAIIGSNGS YNGITSNVKGTFVKGS SSTNNADTYLEISFTG GTLEPGAHVQIQGRFA KNDWSNYTQSNDYSFK SASQFVEWDQVTAYLN GVLVWGKEPGGSVVPS TQPVTTPPATTKPPAT TKPPATTIPPS*DDPK LAAALEHHHHHH

	ccggcaacaaccaagccgcca gcaacaaccaaaccaccggct accacgattcctccatca**TAG** gacgatccgaagcttgcggcc gcactcgagcaccaccaccac caccac	

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

The imaging system is based on a acA1920-150 µm CMOS camera (Basler, Deutschland), with a sensor size of 1920×1200 pixels, a binarization of 12 bits (4096 levels of grey), allowing to acquire and grab 25 frames per second (fps) at full-frame. Associated with a 25 mm focal length lens, the system is designed to observe in transmission a field of 41.7 mm width × 14.6 mm height in a ROI (Region Of Interest) of 1000×350 pixels with a spatial sampling of 41.66 µm/pixel.

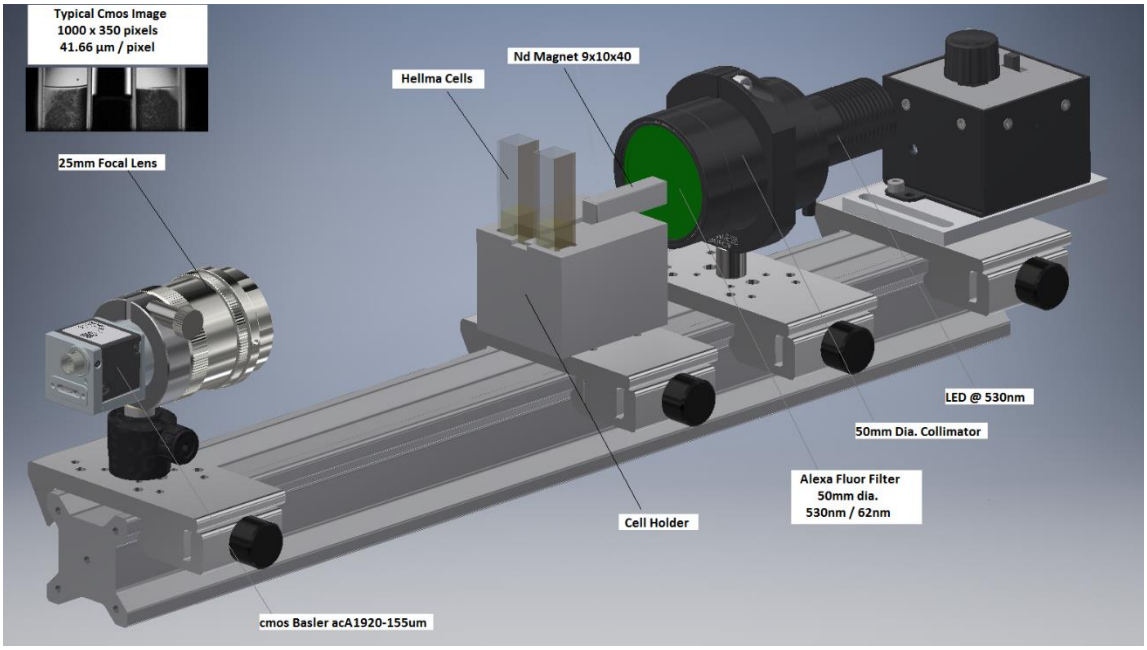

Figure S1: Experimental set-up in transmission

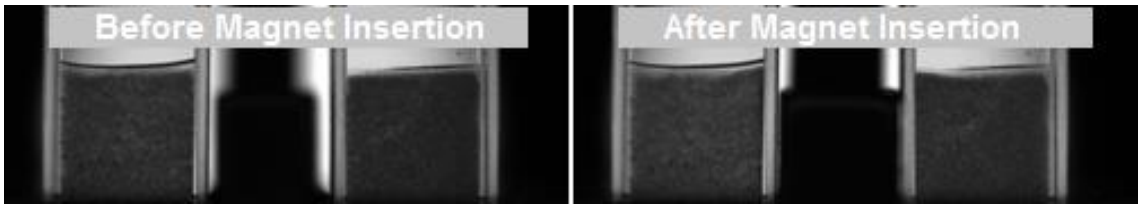

Figure S2: Two 1000 × 350 pixels acquired images. Left: before magnet insertion; Right: after magnet insertion.

Typically, samples to be compared are poured in individual sample cells, far from the magnet. Image acquisition is then started (two frequencies of 4 frames per second (fps) and 25 fps were used) and after about 10 s, beads recovery is initiated by insertion of the magnet. Acquisition is maintained for further 90 s (*i.e.* 360 images at 4 fps or 2250 images at 25 fps, Electronic Supplementary Information Movie S1). Treatment of the data is illustrated in Fig. S3. After background correction, two areas are isolated for each set of images (Fig. S3, in blue or in red, respectively). These areas are 220 pixels wide, pixel 0 being opposite to the position of the magnet, and 100 pixels high (delimited by the two white horizontal lines, Fig. S3A). The normalized intensity of each area is plotted, taking into account the intensity of the solution in each cuvette before and after complete sedimentation of the beads (Fig. S3B). The position of the sedimentation front in each sample corresponds to the value of the abscissa for 50% of the threshold level.

A B

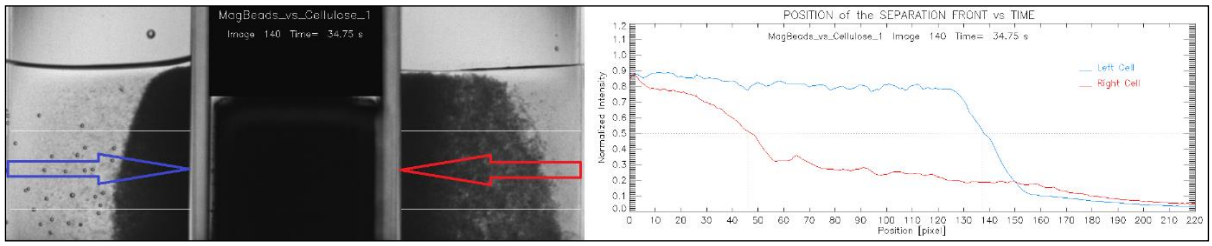

Figure S3: Example of the image data treatment (image 140 corresponding to 37.75 s). **A:** Position and direction of the selected area are represented by blue and red arrows. **B:** Screenshot

of the corresponding area plots intensity vs. pixel position. The 1 mL spectrophotometer semi-micro cuvettes were filled with 600 µL of RAC (2%) in the presence of 80 µl of magnetic beads grafted with TMR biocytin/mSA-CBM3-AzF (left cell, in blue) or bar beads (right cell, in red). Example of the evolution of the recovery front is provided here (Electronic Supplementary Information Movie S2)

The sedimentation front position of the different samples was plotted as a function of time (Fig. S4).

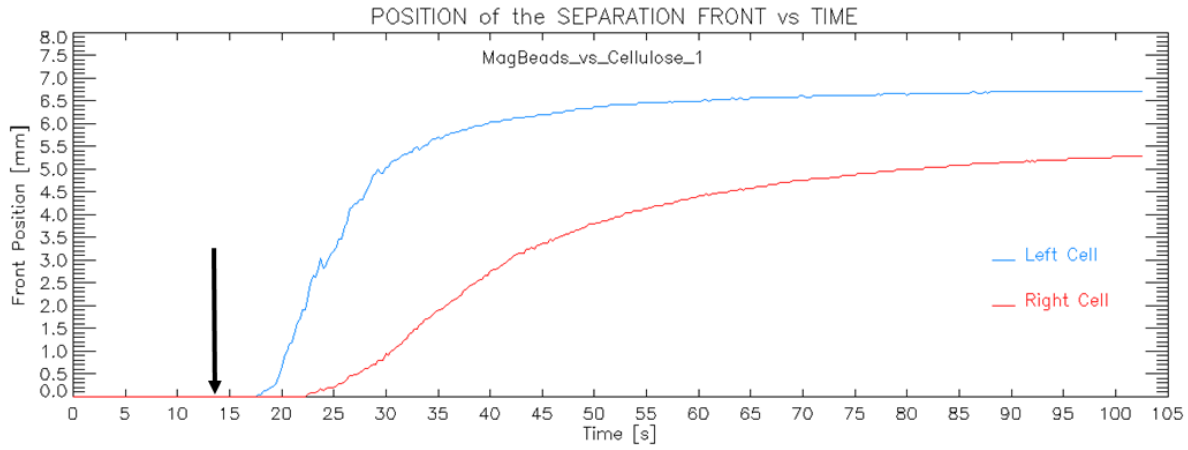

Figure S4: Evolution of the sedimentation front over the time, exemplified from run 1, at 4 fps. Black arrow corresponds to the insertion of the magnet. The cuvettes were filled with 600 µL of RAC (2%) in the presence of 80 µl of magnetic beads grafted with TMR biocytin/mSA-CBM3-AzF (left cell, in blue) or non-grafted beads (right cell, in red).

Curves were modeled as a classical first order step response time with the Equation 1:

$$Pos(t) = Pos(\infty)\left(1 - \exp - \frac{(t - t_0)}{\tau}\right) \qquad \text{Eq. 1}$$

where t_0 is the starting time when the sedimentation front starts to move, $Pos(\infty)$ is the maximum position at "infinite" time and τ is time constant. The time constant that best fits the data is further called τ_{fit}.

From this first order model, the theoretical rising time T_r was obtained by the Equation 2:

$$T_r = T_{90\%} - T_{10\%} \qquad \text{Eq. 2}$$

that can be linked to the time constant τ with regards to the Equation 3:

$$\tau = T_r / 2.2 \qquad \text{Eq. 3}$$

3. Molecular modelling assesses the flexibility of mSA-CBM3

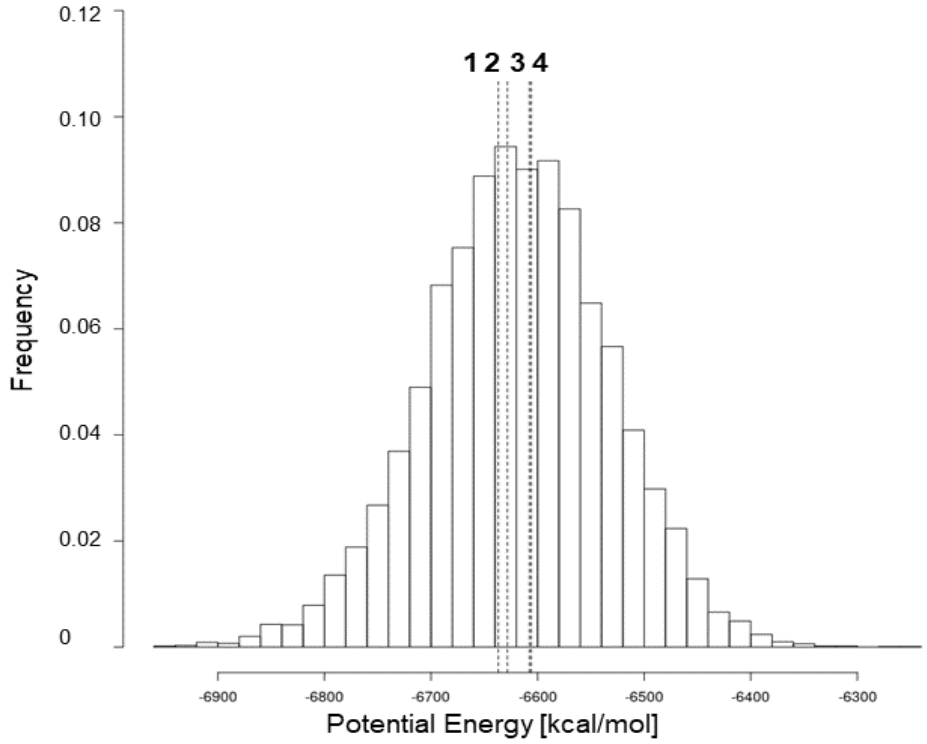

Figure S5: Potential energy distribution of the ensemble of conformations generated using the IDP conformational ensemble model generation method. Conformations (1, 2, 3, 4) selected for MD studies are indicated by vertical dashed lines.

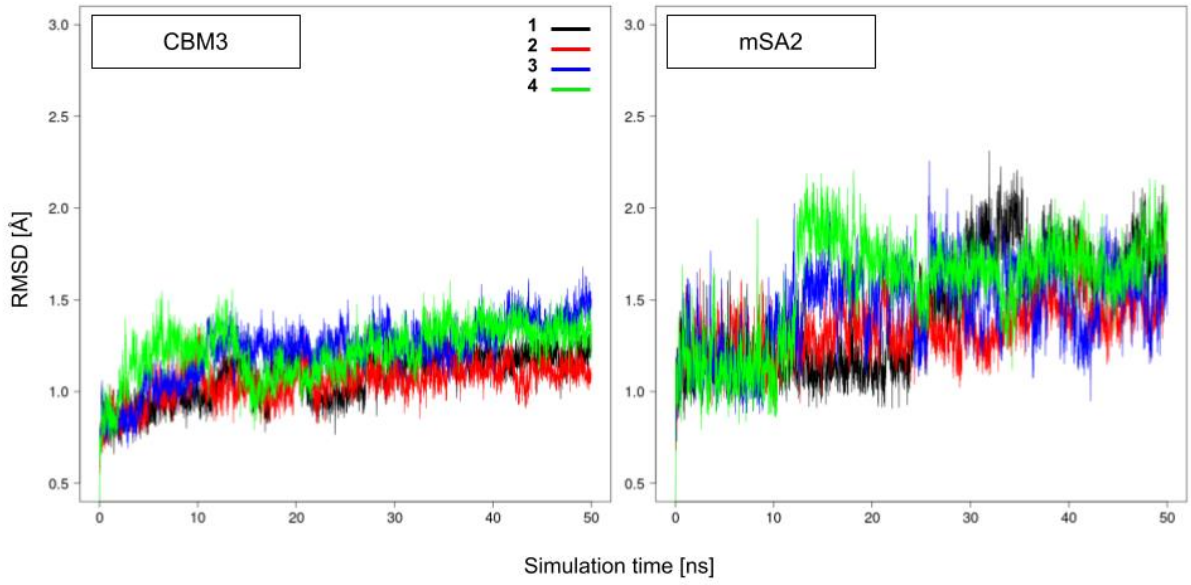

Figure S6: Backbone RMSD of globular domains (CBM3 on the left and mSA2 on the right) along 50 ns-MD simulations carried out using four starting conformational states of the protein platform extracted from the conformational ensemble model generated.

4. Expression and purification of the chimeric CBM3 based proteins

Table S3: Molar extinction coefficient and molecular weight of the proteins studied in this work. Theoretical molar extinction coefficients and molecular weight were calculated using ProtParam online software (https://web.expasy.org/protparam/)

Protein	Molar extinction coefficient ($M^{-1}.cm^{-1}$)	Molecular weight (g/mol)
mRFP1	32,890	26,943
CBM3-RFP	68,300	52,164
mSA-CBM3	73,340	40,885
mSA-CBM3-AzF	73,340	40,885

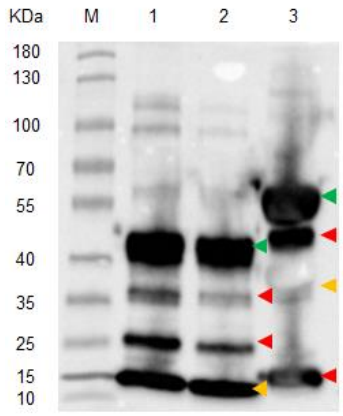

Figure S7: Western blot analysis of purified chimeric proteins, using anti-His-tag-HRP to reveal the presence of His_6-tagged protein. Lanes: M, molecular mass markers; 1: mSA-CBM3; 2: mSA-CBM3-AzF; 3: CBM3-RFP. Green arrows indicate the expected size of the proteins (40.8 and 52 kDa, respectively), orange arrows indicate protein fragment sizes from a common origin (14 and 38 kDa, respectively) and red arrows indicate undetermined fragments of proteolysed proteins.

Figure S8: Presence of a putative additional promotor in the sequence coding mSA-CBM3. CBM3 corresponds to the original sequence (GenBank HF912722.1) and CBM3 optimized corresponds to the sequence that has been codon optimized (iGEM registry, Part:BBa_K1321014). Alignment numbering is based on mSA-CBM3 coding sequence. Promoter region has been predicted using Bprom (Softberry, Mount Kisco, NY, USA).

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

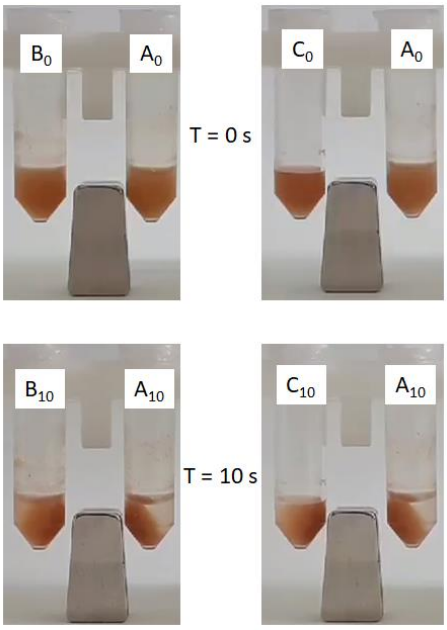

Figure S9: Recovery of magnetic beads previously incubated 1 h with 2% of RAC with a magnet after 10 s in presence of: **A**, magnetic bead grafted with mSA-CBM3-MB; **B**, magnetic beads in presence of mSA-CBM3; **C**, magnetic beads only. Brownish colour is due to the ferric cores of the magnetic beads. RAC in solution present a white cloudy aspect.

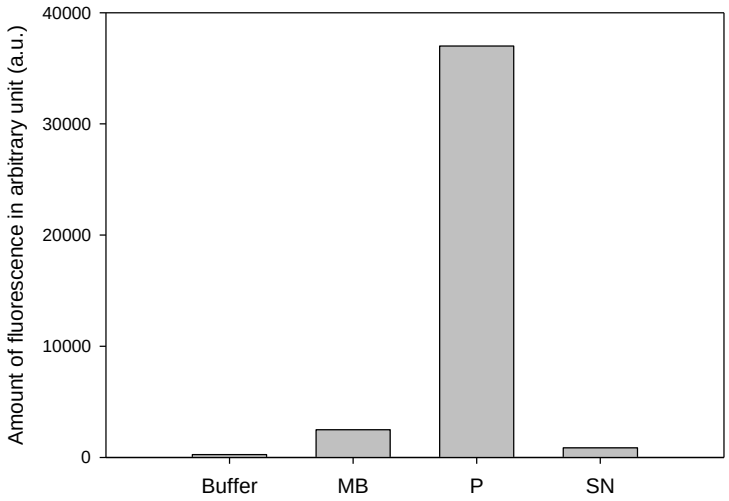

Figure S10: Amount of fluorescence measured after recovering of the beads with a magnet. TMR-mSA-CBM3-MB was incubated with 2% RAC for 1 h and the amount of fluorescence in the pellet (P) and the supernatant (SN) was measured after complete beads sedimentation. Fluorescence of the buffer and the magnetic beads (MB) was also measured.