PCCP

Accepted Manuscript

This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the author guidelines.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the ethical guidelines, outlined in our author and reviewer resource centre, still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

PCCP

ARTICLE

Steered molecular dynamics simulations reveal the role of Ca^{2+} in regulating mechanostability of cellulose-binding proteins

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Melissabye Gunnoo,[a] Pierre-André Cazade,[a] Adam Orlowski,[a] Mateusz Chwastyk,[b,c] Haipei Liu,[d,e] Duy Tien Ta,[d,e] Marek Cieplak,[c] Michael Nash,[d,e] Damien Thompson[a*]

The conversion of cellulosic biomass into biofuels requires degradation of the biomass into fermentable sugars. The most efficient natural cellulase system for carrying out this conversion is an extracellular multi-enzymatic complex named the cellulosome. In addition to temperature and pH stability, mechanical stability is important for functioning of cellulosome domains, and experimental techniques such as Single Molecule Force Spectroscopy (SMFS) have been used to measure the mechanical strength of several cellulosomal proteins. Molecular dynamics computer simulations provide complementary atomic-resolution quantitative maps of domain mechanical stability for identification of experimental leads for protein stabilization. In this study, we used multi-scale steered molecular dynamics computer simulations, benchmarked against new SMFS measurements, to measure the intermolecular contacts that confer high mechanical stability to a family 3 Carbohydrate Binding Module protein (CBM3) derived from the archetypal *Clostridium thermocellum* cellulosome. Our data predicts that electrostatic interactions in the calcium binding pocket modulate the mechanostability of the cellulose-binding module, which provides an additional design rule for the rational re-engineering of designer cellulosomes for biotechnology. Our data offers new molecular insights into the origins of mechanostability in cellulose binding domains and gives leads for synthesis of more robust cellulose-binding protein modules. On the other hand, simulations predict that insertion of a flexible strand can promote alternative unfolding pathways and dramatically reduce the mechanostability of the carbohydrate binding module, which gives routes to rational design of tailormade fingerprint complexes for force spectroscopy experiments.

Introduction

Lignocellulosic biomass is a renewable and abundant natural source of organic carbon. However, the component polysaccharides are extremely difficult to degrade and their recalcitrance severely limits their application as a carbon source for commercial production of biofuels and other value-added chemicals.[1] Many anaerobic bacteria have evolved a diverse class of multi-protein complexes, collectively referred to as cellulosomes, that efficiently degrade crystalline cellulose found in plant cell walls.[2] The cellulosome complex links together a diverse set of enzymes necessary for cellulose degradation through a "plug and socket" modular interaction.[3] Complementary protein domains termed dockerin and cohesin are key components for these interactions, which work in tandem with the carbohydrate binding modules (CBM) [4, 5] that target the complex to its cellulose substrate.

The sub-micron sized cellulosomes are constantly moving and rearranging due to random Brownian motion, and also experience hydrodynamic shear forces under flow. As a result, the CBM domain which is attached simultaneously to the enzyme and the substrate (Fig. 1) is subjected to high mechanical stress. Earlier studies showed that cohesins placed in connecting regions (located between two anchoring points on the scaffoldin; Fig. 1) exhibit very high mechanical stability with rupture forces <400 pN, larger than any known folded protein at the time.[6] By contrast, cohesins located in the non-connecting or 'hanging' region showed much lower mechanical stability.[6] Previous studies have also focused on the mechanostability of cohesin and dockerin complexes,[3] but comparatively little has been reported on mechanical stability of CBM domains.[7, 8] Similar to cohesins,[6] re-engineering of CBM mechanical stability is a potentially powerful means of amplifying the catalytic activity of bacterial cellulosomes. Due to their importance in biomass deconstruction, studies have

[a] Department of Physics, Bernal Institute, University of Limerick, V94 T9PX, Ireland.
[b] Department of Physics, Arizona State University Tempe, Arizona 85287, USA.
[c] Laboratory of Biological Physics, Institute of Physics, Polish Academy of Sciences, Warsaw, Poland.
[d] Department of Chemistry, University of Basel, 4056 Basel, Switzerland.
[e] Department of Biosystems Science and Engineering, Swiss Federal Institute of Technology (ETH Zurich), 4058 Basel, Switzerland.
* Correspondence to: damien.thompson@ul.ie

focussed particularly on optimizing properties of the cellulose enzyme units such as increased stability of cellulases at elevated temperatures and at non-physiological pH, higher tolerance to end-product inhibition, and higher catalytic efficiency on poorly soluble cellulosic substrates.[9, 10] However, the ability of cellulases to depolymerize cellulose and release fermentable sugars depends not only on their intrinsic activity but also, critically, on their ability to access individual glucan chains on the cellulose. Stable anchoring *via* CBM domains helps increase the effective enzyme concentration on the cellulose and ensures that the cellulase enzymes can target the glucan chains.[11, 12]

The mechanical stability of a protein is now a familiar biological and biophysical property, and mechanical forces are known to be important for numerous biological systems and processes including protein folding, organelle transport, and muscle elasticity.[13] The mechanical strength of a biomolecule or biomolecular complex is measured using atomic force microscopy (AFM),[14] tuned in its single molecule force spectroscopy (SMFS) mode to quantify the force required to unfold a protein domain.[15] Molecular dynamics computer simulations can provide complementary maps of the forced unfolding energy landscape, providing experimentally-inaccessible atomic detail on dynamics. Steered molecular dynamics (SMD) techniques provide a computationally efficient means of modelling AFM experiments[16] and can guide the design and interpretation of SMFS experiments.[17] For example, SMD simulations revealed that high mechanical stability in cohesins is commonly associated with a patch of highly localized hydrogen bonds between long beta strands.[6]

The purpose of the present work is to identify the protein residues and regions (*e.g.*, hydrogen bond patches and ion-binding pockets) that provide the strongest resistance to unfolding of CBM. We model the carbohydrate binding module family 3 (CBM3) protein from the cellulosomal scaffoldin subunit CipA of the bacteria *Clostridium thermocellum*. CBM3 is 155 residues long[18] and folds into the common beta sandwich structure containing two antiparallel beta sheets (Fig. 1), nine beta strands (henceforth termed simply "strands") which are interconnected by hydrogen bonds, and a calcium binding pocket. We identify the protein sites that regulate the mechanical properties of CBM, by comparing the stretching and unfolding of wild type CBM3 together with ~40 mutants (both single and multi-site substitutions, see Table S1 and Fig. S1). This strategy allows us to identify the most significant mechanostable motifs in native CBM, and predict mutations that could further improve the stability of re-engineered CBM modules in so-called "designer cellulosomes"[19]. Furthermore, we were interested in studying the effects of introducing a peptide tag (ybbR-tag) as a means of modulating the mechanical stability of CBM3. The ybbR-tag has been used previously to site specifically immobilize proteins to surfaces for single-molecule mechanical experiments[20], and has also been used for site-specific post-translational modification of proteins[21], therefore introduction of ybbR-tags into CBMs could provide a convenient additional functional group for assembly of synthetic cellulosomes.

Complex charge balancing and conformational plasticity make it difficult to modulate the mechanical properties of proteins.[22, 23] Yet significant advances have been made – some through trial-and-error methods and others in a systematic and rational manner. For example, Manteca *et al.* showed that disulfide bonds behave as mechanical lockers to block one unfolding pathway and force unfolding *via* the pathway with the higher energy barrier, thereby increasing the mechanical stability.[24, 25] Cao *et al.* increased the mechanical stability of elastomeric proteins by combining improved protein-protein interaction[26] with engineered metal chelation.[27] More generally, improved understanding of how to modulate protein stability would benefit drug design in, *e.g.*, targeting disease-causing protein variants,[28] and would also speed up biophysical characterization of membrane proteins.[29]

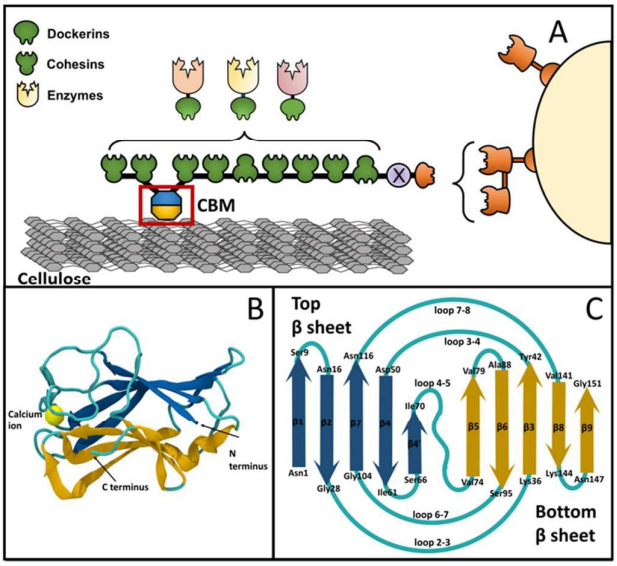

Figure 1. A – Schematic representation of scaffoldin subunit CipA of bacteria Clostridium thermocellum. Dockerin-containing enzymes are bound to the scaffoldin via complexation with cohesins. One single CBM in the scaffoldin anchors the cellulosome complex (and the parent cell) to the cellulose substrate. **B** – The overall three-dimensional structure of CBM3 domain. Cartoon representation of CBM3 β-sandwich structure drawn with VMD software[51]. The top beta sheet is colored navy, the bottom sheet is colored gold, and the linker regions are colored light blue. **C** – Strands and loops numbering in the beta structure of CBM3.

Results and Discussion

Quantifying the mechanical stability of carbohydrate binding module

(a) Identifying the force regime of CBM3 unfolding.

Extensive steered molecular dynamics simulations of CBM3 forced unfolding were performed as described in the Methods section (in Supporting Information). The N-to-C distance between opposite ends of the protein was monitored as forces of 350pN, 400pN, 450pN, and 500pN were applied. Three distinct unfolding events corresponding to high structural resistance were observed (Fig. 2A). The first part of the trajectories confirms the high mechanostability of the CBM3 domain in its native state as it retains its native fold for at least five nanoseconds (for forces of 400pN and below). Shortly after, the protein undergoes a rapid extension of 126 Å from

the 22 Å native N-C distance to form a long-lived intermediate state (i1) after which the domain extends by a further 120 Å to form state i2 that precedes the complete unfolding of CBM3.

Structural snapshots of these trajectories show that the first intermediate corresponds to the loss of contacts in the N-terminus region, along strands 8 and 9, spanning residues 120 to 155 (Fig. 2B). In the i1 and i2 states, disruptions were observed in the C-terminus region of the protein with residues 1-20 separating from the main beta sandwich structure.

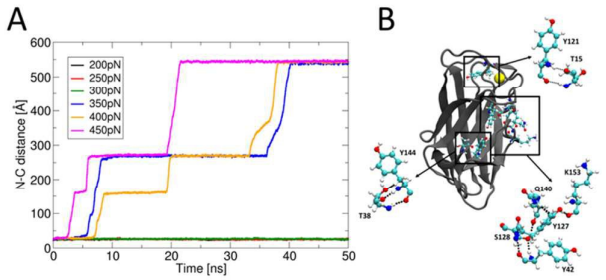

Figure 2. **A** – Molecular dynamics forced unfolding trajectories performed at constant forces ranging from 200 pN to 450 pN (unravelling occurs during 50 ns for applied forces of 350pN, 400pN, and 450pN). **B** – Hydrogen bond networks that are broken in constant force (400 pN) SMD simulations as the protein unravels from its native fold (zoom-in panels) to intermediate state i1. Residues 120-155 span beta strands 8 and 9 and loop 7-8. The contacts include the hydrogen bond pairs Tyr121-The15, Tyr127-Lys153, Ser128-Tyr42, Ser128-Gln140, and Tyr144-Thr38.

(b) Mapping the unfolding landscape of CBM3.

Constant-velocity steered MD simulations were carried out as described in the Methods section (in Supporting Information) and the data presented here is for simulations in which the protein was pulled from its N-terminus, to compare with SMFS experiments.[30, 31] Control simulations were performed by pulling from the C-terminus (Fig. S3). The most commonly occurring pathway (which we label Pathway A, found in 7 out of 10 simulations) for N-to-C pulling consisted of three well-defined peaks (Fig. 3 and Fig. S4-5). These correspond to stable states: native, i1 and i2 discussed in section (a) above. Peaks I and III showed similar trends in all trajectories (Fig. S4) and their respective average maximum peak heights were calculated to be F_1 =1023 ± 43 pN and F_3 = 1392 ± 73 pN.

Depending on which unravelling pathway was followed, out of the three pathways observed in 70, 20, and 10% of the CBM3 stretching simulations respectively, the position and height of peak II varied significantly (Fig. S4). Paci *et al.* reported similar-sized shifts in peak position and height for unfolding a smaller beta sandwich structure protein (fibronectin type III domain from the human form of tenascin, TNfn3).[32] The authors suggested that the unfolding pathways for TNfn3 were broad, compared to other proteins such as the octameric TI 127 poly-protein studied using equivalent techniques.[16, 33-36]

In CBM3, beta strands that are not involved in the calcium coordination sphere separate during the first 250 Å extension. As they do not always separate in the same order, numerous pathways become available. Hence we find variation in the shape and height of peak II corresponding to regions where strands 1, 2, 8 and 9 (Fig. 1C) separate, in random order. In the majority of trajectories, beta strands separate in the following order, β9, β1, β8 and finally β2 (Movie S1). The computed force-extension curves for CBM unfolding by two different pathways are shown in Figure 3, with three clear peaks found for the major pathway A and three less distinguishable peaks in minor pathway B.

The mechanical stability of a protein has been described as a property governed by the specific non-covalent interactions in the main regions of the protein.[15, 37, 38] In the most common pathway (A) of CBM3 stretching, a combination of longitudinal shearing of two anti-parallel beta strands (breaking 12 localized hydrogen bonds) and core hydrophobic disruptions causes the first (peak I) force barrier (Figures. S5-S6 and Table. S2).

The molecular structures sampled around peak II reveal that a number of backbone hydrogen patches (Fig. S5C) form barriers termed mechanical clamps,[33] of varying resistance strength. Most hydrogen bonds involved in those barriers are perpendicular to the direction of the force vector, which gives rise to longitudinal shearing instead of an unzipping (lateral shearing) mechanism for hydrogen bond breaking (Fig. S7-8). Moreover, in the majority of trajectories, hydrogen bond rupture between a pair of beta strands occurs in stages; at 100 Å extension two hydrogen bond pairs separate between strand pair β2-β7 while the remaining three hydrogen bonds between β2 and β7 break later at 150 Å extension. This multiple stage bond breaking results in a broad peak II compared to the sharp peak III at 250 Å where the calcium ion separates from one of the aspartic acid residues (either D46 or D126) in its interaction sphere (Fig. S5).

To further characterize the unfolding of native CBM, experimental SMFS measurements were carried out using automated AFM-based SMFS and coarse-grained (CG) simulations of CBM unfolding were performed using a Go-like model[39-43]. These control experiments and simulations are described in the supporting information (Fig. S9-10).

Rational design of cellulose binding proteins with tailormade mechanostabilities

(a) Targeting sites for re-engineering

The computed unfolding pathway of native CBM highlights the importance of the β-strand regions, the hydrophobic core, and the charged calcium-binding sites to the overall mechanostability of CBM. A set of *in silico* mutations were tested for each of these regions. In the first set, at least one mutation was made in each β-strand (Fig. S11A). Residues forming two or more hydrogen bonds with an adjacent strand were targeted in order to maximize destabilization in the protein β-structure. Secondly, hydrophobic sidechains (Fig. S11B) were replaced in an attempt to disrupt the packing interactions in the protein core. Finally, one aspartate residue Asp46 of the highly-charged calcium binding pocket (Fig. S12A) was mutated to Gly to decrease both electrostatic and van der Waals interactions with the calcium ion, and was also mutated to the bulkier Pro (Fig. S12B) (formally an imino acid, with a non-polar sidechain pyrrolidine).

As well as pulling the full length of the protein, we also stretched CBM from a central region to test the effect that an

altered pulling geometry has on its stability. Brockwell *et al.* have shown that the mechanical resistance depends on not only the amino acid sequence, topology, and the unfolding rate, but is also critically dependent on the direction of the applied force.[44] Hence we created protein structure models in which an 11-residue peptide DSLEFIASKLA is added to the CBM3 domain. This ybbR-tag is commonly used in single molecule experiments such as SMFS[20, 45] either to fuse one end of a protein chain to a surface or to label proteins.[46] We inserted the ybbR-tag into flexible loop 5-6 (Fig. 1C) between residues Ser84 and Thr85. This site was selected due to its high RMSF (root mean square fluctuations of the backbone Cα atoms) values (1.9 Å *vs.* max of 0.7 Å in the more ordered regions) and because this point is near the center of the 155-residue protein sequence. Two single mutant variants were also generated along with the native CBM-ybbR. Mutants D46P-ybbR and D126A-ybbR displayed the combined effect of pulling from the middle of the protein and weakening the calcium binding site.

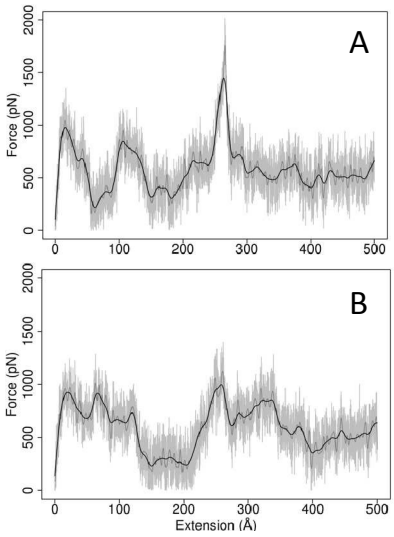

Figure 3. Representative computed force-extension profiles for CBM3 unfolding from the N-terminus *via* two different pathways. The gray data points show the spread of force values recorded every 0.5 picoseconds throughout the trajectory, and the black trace marks a cubic spline interpolation of the raw data using the R package[52].

(b) The calcium binding site imposes high mechanostability

To verify the effect of calcium binding on the mechanostability of the carbohydrate binding module, we performed control simulations with the calcium ion deleted. The force-extension profiles were similar to that of the native except for a large reduction in peak III (Fig. S20). It should be noted that the (artificial) removal of the calcium ion gives rise to additional unfolding pathways as beta strands can separate more freely. In any case, we estimate using Poisson-Boltzmann electrostatic binding free energy calculations a barrier in excess of 200 kcal/mol for Ca^{2+} removal, which corresponds to a vanishingly small probability of CBM losing its calcium ion.[47] Zhang *et al.* tested the force-induced unfolding pathway of another protein, Calmodulin, in both the presence and the absence of Ca^{2+} and found that the unfolding order of the N- and C-terminal regions was related to the calcium binding states as well as the interactions between the two EF-hand motifs in Calmodulin.[48]

(c) Comparing computed F_1 values across all protein variants

To compare the effect of mutation on the mechanostability of the CBM3 domain, the first force peak value (F_1) was measured from the force-extension curves (Fig. S21-S26) of each protein variant. The results are illustrated in Figure 4. Taking into account the occurrence of multiple pathways in the simulations, it is important that the force peak we choose as comparator does not vary from one pathway to another. Visual inspection of the CBM3 steered molecular dynamics repeat curves (Fig. S4) confirms that Peak I does not vary according to the pathway taken. For this reason, all mutant force measurements were taken from peak I (Fig. 4).

1. Beta-strand disruption

The removal of two or three hydrogen bonds (Fig. S11A) only slightly destabilizes the highly mechanostable CBM3 protein (Fig. 4 and Fig. S21). This is because beta strands are held together by approximately 50 hydrogen bonds and when a few of those are disrupted, the remaining 45+ hydrogen bonds are able to preserve the structure or form a slightly altered structure by reconfiguring the hydrogen bonding network. The F_1 for most beta mutants is reduced by less than 10% except in the case of E93A which shows a 17% decrease (Fig. 4) due to loosening of loop 5-6.

2. Destabilizing the hydrophobic core

Single sites mutations in the hydrophobic interior of the protein have minimal effect on the mechanostability of CBM3, yielding F_1 reductions of less than 10% (Fig. 4 and Fig. S22). By contrast, multiple combined mutations in the hydrophobic core and of charged residues (Table S1 and Fig. S1) gave decreases of up to 19% in F_1, with quadruple mutant E3A/E5A/D46G/V136A showing the strongest effect. This reduction in force stems from a combination of disruptions that allow easier separation of β1 from β2 (a main contributor to Peak I (Fig. S5 and Table S2)) and disruptions in the calcium pocket that increase the number of possible unfolding pathways.

3. Breaking the calcium coordination sphere

The Ca^{2+} binding loop of CBM3 plays a prominent role in the intramolecular interactions within the protein and therefore its mechanical stability. The calcium sphere disruption is marked by the highest peak (III) in native CBM3 force-extension curves (Fig. S20A), and all mutants generated to target the calcium pocket resulted in ~45% reduction in peak III. Results for the corresponding peak I were different for each mutant (Fig. S27).

Amongst all the substitutions made in the calcium binding pocket, the largest decrease in F_1 of 23% compared to native, was observed on substituting Asp46 with a Pro residue (Fig. 4 and S23). D46P gives the lowest force peak I among all three sets of single as well as the multiple mutants tested (Fig. 4). Conformational changes due to single residue mutation from Asp to Pro cause major loss in hydrogen interactions (Fig. S12B/D) not only at the mutated site; in particular, Tyr7 and Gln21 do not form hydrogen bonds anymore, a long range effect[22] that extends 20 Å from the mutation site. Our results suggest that Ca^{2+} in CBM3 has the ability to modulate the conformation of loop 7-8 (Fig. S12), a "hook-like" structure that

contributes to CBM3 mechanical stability by forming approx. 11 hydrogen bonds.

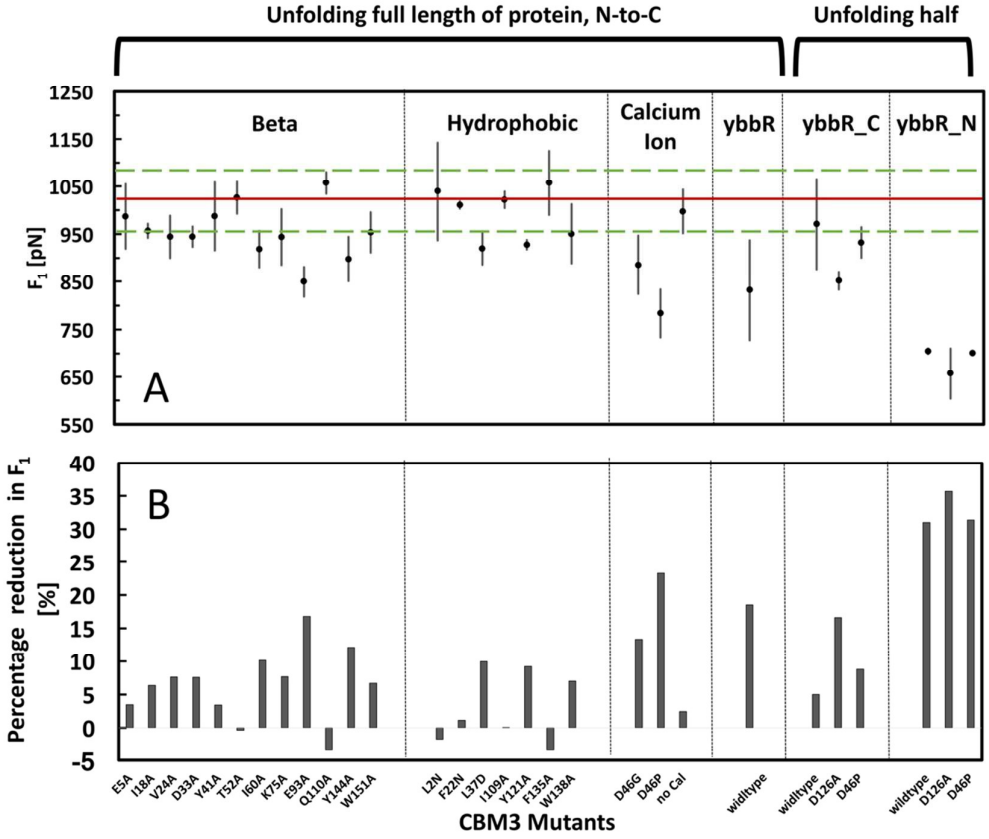

Figure 4. **A** – F_1 distribution for all point mutants, calculated from at least three repeats. The red line represents the average F_1 (1023 pN) of native protein and green lines represent the error calculated as 95% confidence interval of the mean. **B** – Percentage decrease in average F_1 of mutants compared to that of native CBM3. The free energies of the unfolding intermediates are in principle obtainable using Jarzynski's equality[53, 54] (also known as the nonequilibrium work relation), and could be used in the future to very precisely calculate and compare potentials of mean force (PMF) for the native protein and promising mutants.

Regulation of protein stiffness by cation binding has been shown in previous studies. For example Hocky *et al.*[49] recently discussed how the presence of Mg^{2+} affects not only the structure of actin monomer but also the polymeric filament structure, assembly, and mechanical stability. Mg^{2+} binds to a loop on an actin protein that adheres to an adjacent actin subunit, which increases torsional stiffness in the actin filaments.[49] In CBM3, mutation-induced shifts in the Ca^{2+} sphere combined with disruptions in loop 7-8 cause strong destabilization. Similar weakening effects were reported in a recent simulation study of an extracellular depolymerase, PhaZ7 from *Paucimonas lemoignei*, in which the (forced) open conformation of PhaZ7 mutant Y104E had large RMSD > 2 Å. Even after 200 ns of free dynamics, the structure did not stabilize due to non-ordered motion of flexible loop 281-295.[50]

4. Grafting a ybbR-tag into the protein module

The ybbR-CBM structure did not show an increase in RMSD (as described in Supporting Information) but its F_1, when stretched from N-to-C, was lowered by 19% (Fig. 4 and S23) due to the loosening of loop 5-6 when the ybbR-tag is inserted. On the other hand, when pulling directly from the tag (which means only half of CBM3 was extended, see Fig. S24-25), F_1 was further reduced by 35% and 16% in the case of ybbR_N and ybbR_C (Fig. 4) respectively. This variation in F_1 can be explained by the number of lateral (unzipping) and longitudinal shearing events between beta strands pairs. The unfolding pathways of both ybbr_N and ybbR_C are shown in Movies S1 and S2 respectively and are described in Supporting Information.

Methods

The description of CBM3 models and all simulations protocols are presented in the supporting information.

Conclusions

Our simulation dataset provides atomic resolution details of the forced unfolding pathway of a carbohydrate binding module (CBM) protein in its native state as well as approx. 40 mutated states. The simulations revealed three defined force peaks in constant velocity pulling and three plateaus in constant force pulling simulations, corresponding to barriers between three stable folded intermediates and their

corresponding unfolded states. The first mechanical resistance to the external force application occurs within the initial 11 Å extension and arises predominantly due to a combination of core hydrophobic disruptions and inter-strand hydrogen bonds between β8-β9 and β1-β2. This is followed by a barrier due to lateral shearing of hydrogen bonds between strand pairs β3-β8 and β2-β7 while the final barrier involves the separation of aspartic acid residue 46 or 126 from the calcium ion.

In the second part of this study, we explored the possibility of altering the mechanical stability of the CBM domain by testing over 20 single mutations as well as 19 multi-site mutants. Three sets of mutants were investigated in the β-structure, hydrophobic core and calcium ion sphere. Single-site mutations on the beta strands made little impact as removal of only two or three hydrogen bonds did not cause enough structural perturbation to significantly lower the unfolding force of the protein. Compared to the native protein, unfolding forces of all beta mutants were reduced by less than 10% except in the case of E93A. Similarly, when a single residue in the hydrophobic core was replaced by a polar one, unfolding forces were only reduced by 10%, while multiple hydrophobic mutations resulted in a decrease of up to 19% in F_1. The most significant decrease in unfolding force of 23% was predicted for mutating a negative carboxylate in the calcium binding sphere (D46P). This mutation resulted in bending of loop 7-8 to form new interactions between the calcium ion and aspartic acid 139.

The third aim of the study was to investigate the effect of an alternative pulling geometry on CBM3 in which only half the protein is unraveled. The data suggest that unfolding forces required to extend CBM3 depend strongly upon which part of protein was pulled. The C-terminal half of the protein required higher unfolding forces compared to the N-terminal half, mainly because the calcium coordination sphere must be broken in C-terminus unfolding. We hope that future experiments will test our hypotheses concerning the predicted effects of mutants on the force response of the CBM. In general, judicious choice of mutants coupled with loop insertion appears to be a promising means of modulating the mechanostability of protein modules, which has immediate applications in synthesis of designer cellulosomes for biofuel production.

Funding sources

This research has been supported by the European Framework Programme VII NMP grant 604530-2 (CellulosomePlus). DT acknowledges also support through Science Foundation Ireland (SFI) (grant no. 15/CDA/3491).

Abbreviations

CBM Carbohydrate-binding module
CBM3 Family 3 carbohydrate binding module
MD Molecular dynamics
SMD Steered molecular dynamics
SMFS Single molecule force spectroscopy

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank Ed Bayer for discussions and Science Foundation Ireland (SFI) for provision of computing time at the SFI/Higher Education Authority Irish Center for High-End Computing (ICHEC).

Author contributions

MG and DT designed the study and prepared the manuscript. MG and MCh performed the computer simulations, and MG, MCh, MCi and DT analyzed the results. HL and DTT performed the experiments and analyzed the SMFS data with MN. All authors contributed to writing and editing the manuscript.

Notes and references

1. E. A. Bayer, R. Lamed and M. E. Himmel, *Curr. Opin. Biotechnol.*, 2007, **18**, 237-245.
2. Y. Shoham, R. Lamed and E. A. Bayer, *Trends Microbiol*, 1999, **7**, 275-281.
3. M. Gunnoo, P. A. Cazade, A. Galera-Prat, M. A. Nash, M. Czjzek, M. Cieplak, B. Alvarez, M. Aguilar, A. Karpol, H. Gaub, M. Carrion-Vazquez, E. A. Bayer and D. Thompson, *Adv Mater*, 2016, **28**, 5619-5647.
4. 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**.
5. J. Caspi, Y. Barak, R. Haimovitz, D. Irwin, R. Lamed, D. B. Wilson and E. A. Bayer, *Appl Environ Microbiol*, 2009, **75**, 7335-7342.
6. A. Valbuena, J. Oroz, R. Hervas, A. Manuel Vera, D. Rodriguez, M. Menendez, J. I. Sulkowska, M. Cieplak and M. Carrion-Vazquez, *P Natl Acad Sci USA*, 2009, **106**, 13791-13796.
7. M. Abou-Hachem, E. N. Karlsson, P. J. Simpson, S. Linse, P. Sellers, M. P. Williamson, S. J. Jamieson, H. J. Gilbert, D. N. Bolam and O. Holst, *Biochemistry*, 2002, **41**, 5720-5729.
8. M. I. M. Khan, M. Sajjad, S. Sadaf, R. Zafar, U. H. K. Niazi and M. W. Akhtar, *J. Biotechnol.*, 2013, **168**, 403-408.
9. K. H. Malinowska, T. Rind, T. Verdorfer, H. E. Gaub and M. A. Nash, *Anal Chem*, 2015, **87**, 7133-7140.
10. Y. H. P. Zhang, M. E. Himmel and J. R. Mielenz, *Biotechnol Adv*, 2006, **24**, 452-481.
11. C. Hervé, A. Rogowski, A. W. Blake, S. E. Marcus, H. J. Gilbert and J. P. Knox, *Proceedings of the National Academy of Sciences*, 2010, **107**, 15293-15298.
12. A. Várnai, M. Siika-aho and L. Viikari, *Biotechnol Biofuels*, 2013, **6**, 30.

13. P. Y. Chen, A. Y. M. Lin, Y. S. Lin, Y. Seki, A. G. Stokes, J. Peyras, E. A. Olevsky, M. A. Meyers and J. McKittrick, *J Mech Behav Biomed*, 2008, **1**, 208-226.
14. A. Vinckier and G. Semenza, *Febs Lett*, 1998, **430**, 12-16.
15. M. Carrion-Vazquez, A. F. Oberhauser, T. E. Fisher, P. E. Marszalek, H. B. Li and J. M. Fernandez, *Prog Biophys Mol Bio*, 2000, **74**, 63-91.
16. M. Sotomayor and K. Schulten, *Science*, 2007, **316**, 1144-1148.
17. R. B. Best, S. B. Fowler, J. L. T. Herrera, A. Steward, E. Paci and J. Clarke, *J Mol Biol*, 2003, **330**, 867-877.
18. J. Tormo, R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham and T. A. Steitz, *Embo J.*, 1996, **15**, 5739-5751.
19. L. Artzi, E. A. Bayer and S. Moraïs, *Nature Reviews Microbiology*, 2016, **15**, 83.
20. C. Schoeler, R. C. Bernardi, K. H. Malinowska, E. Durner, W. Ott, E. A. Bayer, K. Schulten, M. A. Nash and H. E. Gaub, *Nano letters*, 2015, **15**, 7370-7376.
21. W. Ott, T. Nicolaus, H. E. Gaub and M. A. Nash, *Biomacromolecules*, 2016, **17**, 1330-1338.
22. M. Chwastyk, A. M. Vera, A. Galera-Prat, M. Gunnoo, D. Thompson, M. Carrión-Vázquez and M. Cieplak, *The Journal of Chemical Physics*, 2017, **147**, 105101.
23. A. Aleksandrov, D. Thompson and T. Simonson, *Journal of molecular recognition : JMR*, 2010, **23**, 117-127.
24. H. Li, *Organic & Biomolecular Chemistry*, 2007, **5**, 3399-3406.
25. A. Manteca, Á. Alonso-Caballero, M. Fertin, S. Poly, D. De Sancho and R. Perez-Jimenez, *Journal of Biological Chemistry*, 2017, **292**, 13374-13380.
26. Y. Cao, T. Yoo, S. Zhuang and H. Li, *Journal of molecular biology*, 2008, **378**, 1132-1141.
27. Y. Cao, Yongnan D. Li and H. Li, *Biophysical Journal*, 2011, **100**, 1794-1799.
28. S. M. A. Hermans, C. Pfleger, C. Nutschel, C. A. Hanke and H. Gohlke, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2017, **7**, e1311-n/a.
29. F. Magnani, Y. Shibata, M. J. Serrano-Vega and C. G. Tate, *Proceedings of the National Academy of Sciences*, 2008, **105**, 10744-10749.
30. M. A. Jobst, C. Schoeler, K. Malinowska and M. A. Nash, *J Vis Exp*, 2013, DOI: 10.3791/50950, e50950.
31. M. A. Jobst, L. F. Milles, C. Schoeler, W. Ott, D. B. Fried, E. A. Bayer, H. E. Gaub and M. A. Nash, *Elife*, 2015, **4**.
32. S. P. Ng, R. W. S. Rounsevell, A. Steward, C. D. Geierhaas, P. M. Williams, E. Paci and J. Clarke, *J Mol Biol*, 2005, **350**, 776-789.
33. H. Lu, B. Isralewitz, A. Krammer, V. Vogel and K. Schulten, *Biophysical Journal*, 1998, **75**, 662-671.
34. M. Gao, M. Wilmanns and K. Schulten, *Biophysical Journal*, 2002, **83**, 3435-3445.
35. H. Lu and K. Schulten, *Chem Phys*, 1999, **247**, 141-153.
36. H. Lu, B. Isralewitz and K. Schulten, *Biophysical Journal*, 1999, **76**, A176-A176.
37. B. Isralewitz, M. Gao, H. Lu and K. Schulten, *Biophysical Journal*, 2000, **78**, 28a-28a.
38. E. Paci and M. Karplus, *P Natl Acad Sci USA*, 2000, **97**, 6521-6526.
39. N. Go, *Annual review of biophysics and bioengineering*, 1983, **12**, 183-210.
40. M. Cieplak and T. X. Hoang, *Biophysical Journal*, 2003, **84**, 475-488.
41. M. Cieplak, T. X. Hoang and M. O. Robbins, *Proteins: Structure, Function, and Bioinformatics*, 2004, **56**, 285-297.
42. J. I. Sułkowska and C. Marek, *Journal of Physics: Condensed Matter*, 2007, **19**, 283201.
43. J. I. Sułkowska and M. Cieplak, *Biophysical Journal*, 2008, **95**, 3174-3191.
44. D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham and S. E. Radford, *Nat Struct Biol*, 2003, **10**, 731-737.
45. L. F. Milles, E. A. Bayer, M. A. Nash and H. E. Gaub, *J Phys Chem B*, 2017, **121**, 3620-3625.
46. J. Yin, P. D. Straight, S. M. McLoughlin, Z. Zhou, A. J. Lin, D. E. Golan, N. L. Kelleher, R. Kolter and C. T. Walsh, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 15815-15820.
47. D. Thompson and T. Simonson, *Journal of Biological Chemistry*, 2006, **281**, 23792-23803.
48. Y. Zhang and J. Z. Lou, *Plos One*, 2012, **7**.
49. G. M. Hocky, J. L. Baker, M. J. Bradley, A. V. Sinitskiy, E. M. De La Cruz and G. A. Voth, *The Journal of Physical Chemistry B*, 2016, **120**, 4558-4567.
50. T. F. Kellici, T. Mavromoustakos, D. Jendrossek and A. C. Papageorgiou, *Proteins: Structure, Function, and Bioinformatics*, 2017, **85**, 1351-1361.