Impact of scaffoldin mechanostability on cellulosomal activity

Albert Galera-Prat, Andrés M. Vera, Sarah Moraïs, Yael Vazana, Edward A. Bayer &
Mariano Carrión-Vázquez

Supporting Information

Protein engineering methods

The methods used to generate the desired constructions were based on standard procedures [1]. All oligonucleotides were synthesized by Sigma-Aldrich (St. Louis, MO).

The four monovalent scaffoldins used in this work were built from components of *C. thermocellum* CipA scaffoldin. In particular, X-module and type II dockerin (XDock), were amplified from pET28-XDock [2, 3]. Cohesins 1 and 7 were amplified from pAFM-CtA1 [4] and pAFM-c7A [5], respectively. pET28-CBM [6] was used as a template for the CBM. First XDock was amplified with primers incorporating NdeI, NheI, KpnI and SpeI at the 5' end and two STOP codons as well as a XhoI site at the 3' end. The resulting fragment was cloned into the pET28a vector using NdeI and XhoI sites, to generate pET28-rs-XDock. Then, CBM was amplified and cloned into the last vector using NdeI and NheI sites (pET28-CBM-rs-XDock) or in SpeI site (pET28-rs-CBM-XDock), which were used to generate the monovalent scaffoldins where the cohesin was positioned in the connecting and external region, respectively. Finally, to each plasmid, both CtA1 and CtA7 cohesins were cloned using KpnI and SpeI sites, thus generating the four monovalent scaffoldins: pET28-CBM-CtA1-XDock (C1), pET28-CBM-CtA7-XDock (C7), pET28-CtA1-CBM-XDock (E1) and pET28-CtA7-CBM-XDock (E7).

To generate GFP-XDock, the GFP coding sequence was amplified from the GFP-ssrA plasmid [7], a kind gift from Prof. A. Goldberg, and NdeI and EcoRI sites were introduced at its ends. The sequence was cloned into pET28 using the same restriction sites. Then the genes for the X-module and type II dockerin from *C. thermocellum* CipA were amplified, introducing an EcoRI site at the 5' end and two STOP codons and a XhoI site at the 3' end, and the resultant sequence was cloned using the same sites of the vector generated in the previous step.

Cel5A-t consists of the catalytic module from *Thermobifida fusca* Cel5A endoglucanase where its native CBM has been removed and a dockerin module from *C. thermocellum* Xyn10Z has

been fused at the C-terminus [8]. The pET28 vector containing the *C. thermocellum* SdbA cohesin (CtS1) was described previously [9].

For AFM experiments, two constructs were used: $(CtA1)_8$, containing 8 tandem repeats of CtA1, described previously [4] and pET-Cel8A generated for this study. To generate pET-Cel8A, *C. thermocellum* Cel8A dockerin was PCR amplified from plasmid Dockerin Basic (I27)3[10],and cloned between the NcoI XhoI sites of pET24 d expression vector. This construct includes a stop codon after the dockerin gene so the histag sequence in the plasmid backbone is not translated.

Protein expression and purification

Proteins were expressed in *Escherichia coli* BL21 (DE3), BL21 star (DE3) or C41 (DE3) [11]. Cells were grown in LB broth at 37°C and expression was induced by addition of Isopropyl-D-1-thiogalactopyranoside (IPTG). In particular, the monovalent scaffoldins were expressed for 16 h at 16°C with 0.1 mM IPTG and 1 mM $CaCl_2$, GFP-XDock was expressed for 16 h at 25°C with 0.1 mM IPTG and 1 mM $CaCl_2$.

For AFM experiments, $(CtA1)_8$ and pET-Cel8A vectors were co-transformed in Bl21Star *E. coli* strain. Protein expression was induced by adding 0.5 mM IPTG and cells were incubated overnight at 16 °C.

After expression, cells were harvested by centrifugation at 6000xg for 10 min and snap frozen in liquid N_2. Lysate samples were centrifuged at 39000xg for 30 min and, except for GFP-XDock, clarified samples were incubated first at 55°C for 30 min and then in an ice bath for 10 min before centrifugation for 10 min at 4000xg. Proteins were purified by Ni^{2+} affinity chromatography in a HisTrap HP column (GE Healthcare, Uppsala, Sweeden) using FPLC (ÄKTA Purifier, GE Healthcare), the fractions were analyzed by SDS-PAGE, pooled and further re-purified if necessary by ionic-exchange in HiTrap Q HP column (GE Healthcare) and using 50 mM Tris, 100 mM NaCl, 1mM $CaCl_2$ at pH 8.0 and eluting using a gradient to a similar buffer containing 1M NaCl) or size-exclusion chromatography using HiLoad 16/60 column (GE Healthcare). Purified samples were concentrated in 50 mM Tris, 300 mM NaCl, 1 mM $CaCl_2$ at pH 7.4 (activity buffer) and stored in 50% glycerol at -20°C. Protein concentration was determined spectrophotometrically by using a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and the predicted extinction coefficient as calculated by Protparam (http://web.expasy.org/protparam/).

Co-expressed $(CtA1)_8$ and Cel8A proteins were purified together in the same chromatographic experiment. Protein purification was conducted as already described, combining an Histag affinity purification followed by a size exclusion chromatography. As the dockerin protein has no Histag, co-purification of this protein simultaneously with the $(CtA1)_8$ sample reports for correct cohesin-dockerin interaction.

Atomic Force Microscopy – Single Molecule Force Spectroscopy experiments

Atomic Force Microscopy experiment were conducting in an AFS- Single molecule Atomic Force Spectroscope (Luigs & Newmann, GmbH) using Ni^{2+} functionalized coverslips as described elsewhere [10]. In short, $(CtA1)_8$ –Cel8A samples were incubated in 20 mM, 100mM NaCl, 1 mM $CaCl_2$ at a concentration of 0.1-0.2 mg/ml in the functionalized coverslips for 15 min. After the incubation step, the surface was washed out twice and the experiment conducted at 400nm/s.

Selection of microparticle size

The polystyrene microparticles (Spherotech Inc., Lake Forest, IL) used in this study mimic the presence of a bacterial cell. To this end, we selected microparticles with a size that provided a similar volume and surface comparable to that of *C. thermocellum* cells. Assuming that *C. thermocellum* cells have a cylindrical shape (0.6 to 0.7 µm in diameter and 2.5 to 3.5 µm long [12]) and the microparticles are spherical, we calculated that the diameter of the microparticles that would provide the same surface of a *C. thermocellum* cell would be:

$$D_{particle|Surface} = 2\sqrt{\frac{r(r+h)}{2}} = 1.47 \ \mu m$$

While the diameter that would result in the same volume would be

$$D_{particle|Volume} = 2\sqrt[3]{\frac{3}{4}r^2 h} = 1.24 \ \mu m$$

Where *r* corresponds to the average radius and *h* to the average length of the *C. thermocellum* cell.

The selected microparticles had an average diameter of 1.39 µm, which therefore show a similar surface and volume to those of the bacterial cell.

Substrate selection

Since the monovalent scaffoldins used here only include a single cohesin, the increase in activity observed for the Cel5A-t enzyme upon binding to it is likely derived from the targeting effect of the CBM. This effect is expected to play an important role when trying to degrade a crystalline and poorly accessible substrate [13, 14]. However, in amorphous substrates, where accessibility is increased, this effect usually has a lower contribution, and may even hinder the activity of the system [13]. Hence, we used microcrystalline cellulose in our experiments, since we expected this to increase the difference between the conditions studied. The difficulty to degrade a substrate usually depends on the accessibility of enzymes [13, 15]. Since Avicel has limited accessibility, comparable to that of natural substrates [15], we expect that our results could be readily extrapolated to the degradation of natural substrates of industrial interest.

Geometry of force application

Since force is a vectorial parameter and the response of proteins to force is highly anisotropic, it is important to consider the geometry in which force is being applied [16].

Some cellulosomes can simultaneously bind to a cell surface and to its substrate. In those cases, the modules located between the two anchoring points may be exposed to force. This force would stretch the cohesin modules in the direction of their N- and C-termini, similar to that used to characterize the mechanical stability of several cohesin modules by AFM [4, 5]. The constructions used in those studies ensured that force was applied in the same geometry thus making the results directly comparable to the real system.

In our study we use an experimental design that ensures that the monovalent cellulosomes designed were oriented such that force was applied in the same geometrical fashion as in the natural system (N- to C- terminal for the elements between the two anchoring points) so that our results would be physiologically meaningful.

References

1. Sambrook, J. and D.W. Russell, *Molecular cloning: a laboratory manual*. 2001: CSHL Press.
2. Adams, J.J., et al., *Expression, purification and structural characterization of the scaffoldin hydrophilic X-module from the cellulosome of Clostridium thermocellum.* Protein Expr Purif, 2004. **38**(2): p. 258-63.
3. Adams, J.J., et al., *Mechanism of bacterial cell-surface attachment revealed by the structure of cellulosomal type II cohesin-dockerin complex.* Proc Natl Acad Sci U S A, 2006. **103**(2): p. 305-10.
4. Galera-Prat, A., et al., *The cohesin module is a major determinant of cellulosome mechanical stability.* J Biol Chem, 2018. **293**(19): p. 7139-7147.
5. Valbuena, A., et al., *On the remarkable mechanostability of scaffoldins and the mechanical clamp motif.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 13791-6.
6. Morag, E., et al., *Expression, purification, and characterization of the cellulose-binding domain of the scaffoldin subunit from the cellulosome of Clostridium thermocellum.* Appl Environ Microbiol, 1995. **61**(5): p. 1980-6.
7. Benaroudj, N. and A.L. Goldberg, *PAN, the proteasome-activating nucleotidase from archaebacteria, is a protein-unfolding molecular chaperone.* Nat Cell Biol, 2000. **2**(11): p. 833-9.
8. Caspi, J., et al., *Effect of linker length and dockerin position on conversion of a Thermobifida fusca endoglucanase to the cellulosomal mode.* Appl Environ Microbiol, 2009. **75**(23): p. 7335-42.
9. Haimovitz, R., et al., *Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules.* Proteomics, 2008. **8**(5): p. 968-79.
10. Vera, A.M. and M. Carrion-Vazquez, *Direct Identification of Protein-Protein Interactions by Single-Molecule Force Spectroscopy.* Angew Chem Int Ed Engl, 2016. **55**(45): p. 13970-13973.
11. Miroux, B. and J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.* J Mol Biol, 1996. **260**(3): p. 289-98.
12. McBee, R., *The characteristics of Clostridium thermocellum.* J Bacteriol, 1954. **67**(4): p. 505-6.
13. Fierobe, H.P., et al., *Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components.* J Biol Chem, 2002. **277**(51): p. 49621-30.
14. Fierobe, H.P., et al., *Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin.* J Biol Chem, 2005. **280**(16): p. 16325-34.
15. Zhang, Y.H. and L.R. Lynd, *Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems.* Biotechnol Bioeng, 2004. **88**(7): p. 797-824.
16. Carrion-Vazquez, M., et al., *The mechanical stability of ubiquitin is linkage dependent.* Nat Struct Biol, 2003. **10**(9): p. 738-43.
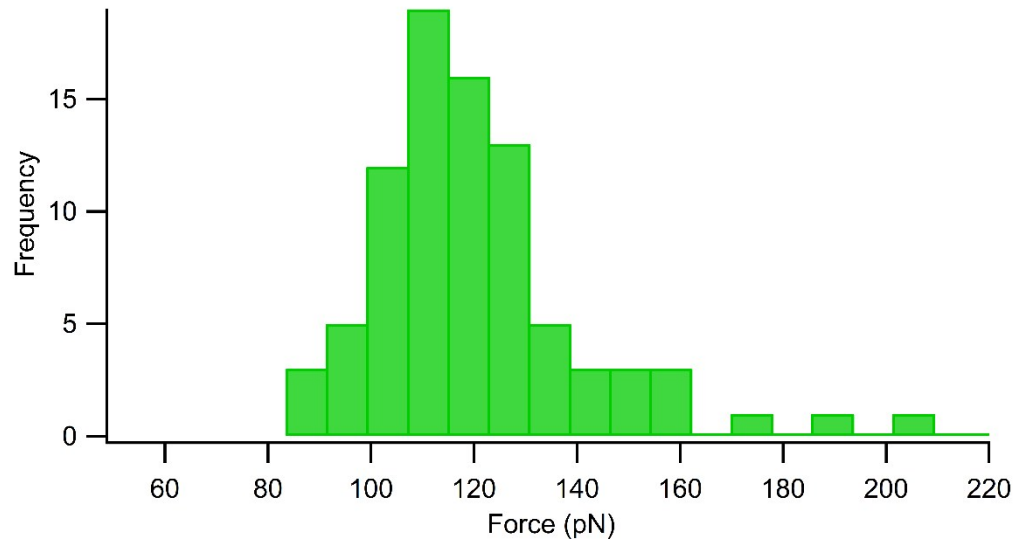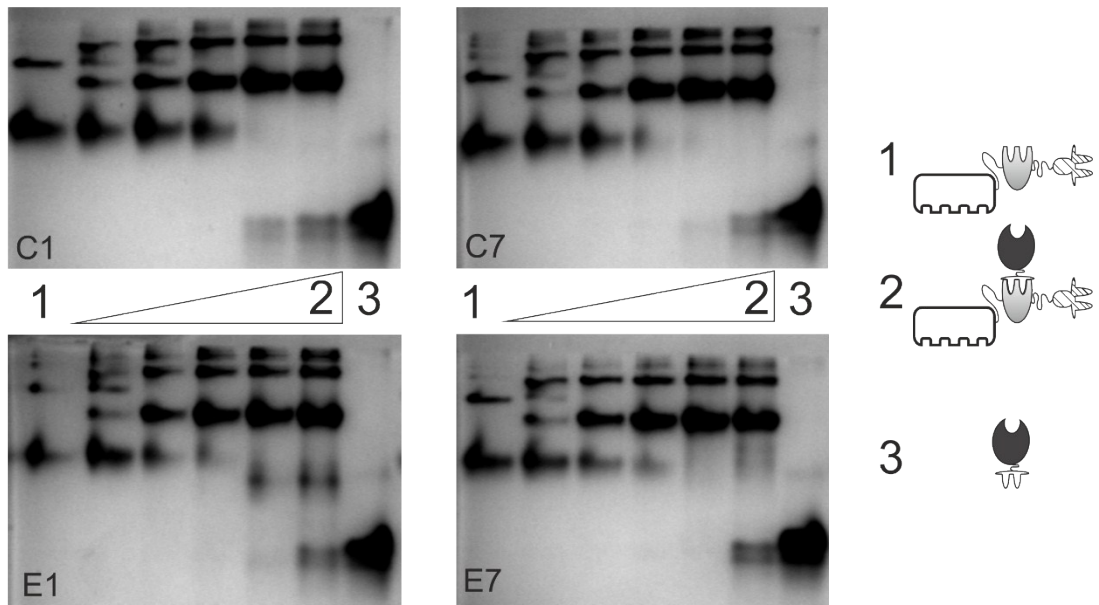
Supplementary Figure 1: The mechanical stability of cohesin CtA1 is not affected by dockerin binding. Unfolding force histogram of CtA1 in the presence of dockerin. The average unfolding force is 118 ± 27 pN (n=85, average ± s.d.) very close to that measured under same conditions for CtA1 without dockerin (124±25pN).
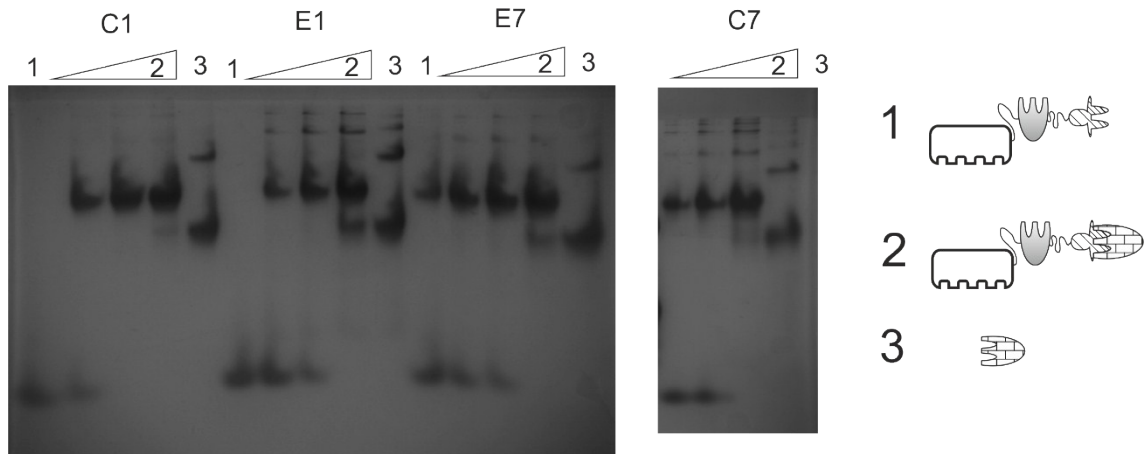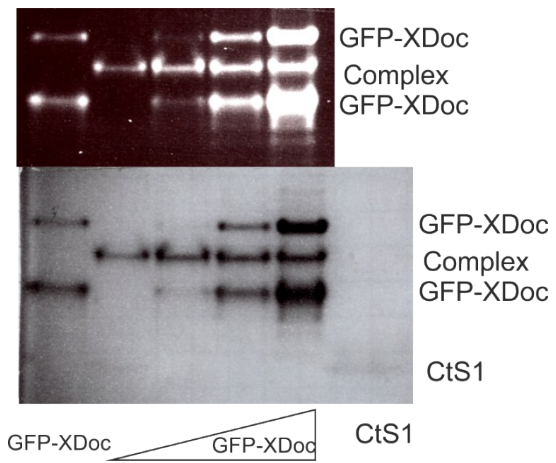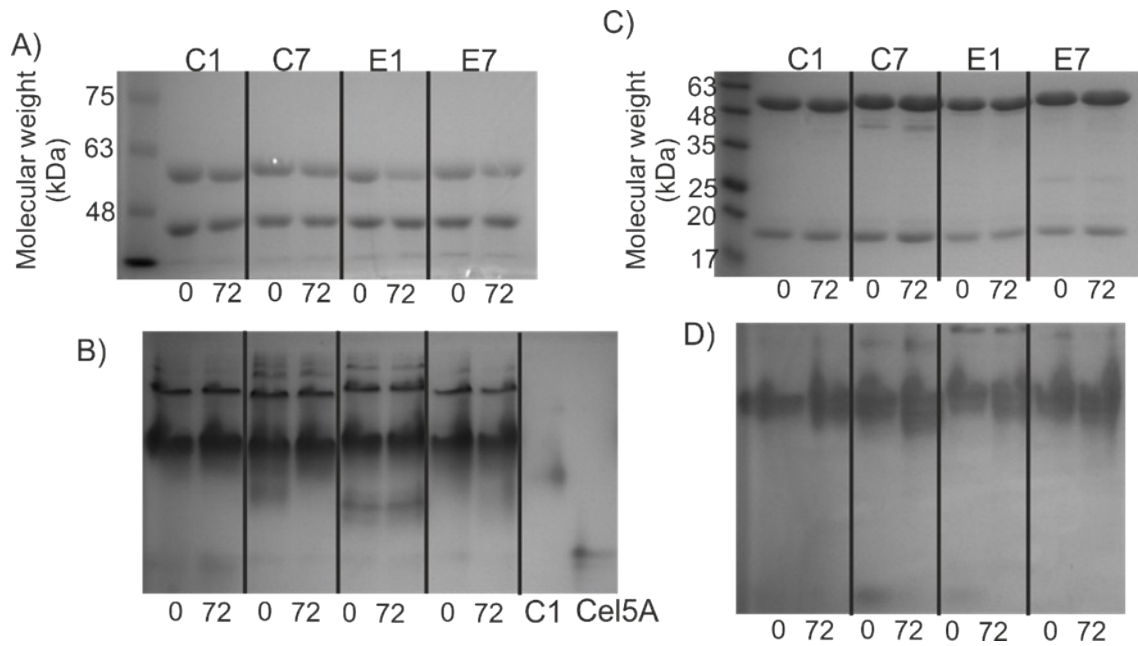
Supplementary Figure 2: Full gels for controls in Figure 2. Each mini-scaffoldin was titrated with increasing concentrations of CtS1 (a) and Cel5A (b).

Supplementary Figure 3: Full gels for GFP-XDoc and CtS1 interaction in Figure 3. CtS1 was titrated with increasing concentrations of GFP-XDoc. Before Coomasie staining of the gel (bottom), a UV image was acquired (top).

Supplementary Figure 4: Stability controls of the different mini-cellulosomes used in this study. Fresh samples (0) and samples incubated for 72 h at 50°C (72) of the different mini-scaffoldins, with either Cel5A-t enzyme (A and B) or with CtS1 (C and D), were compared both using SDS-PAGE (A and C) and native PAGE (B and D).
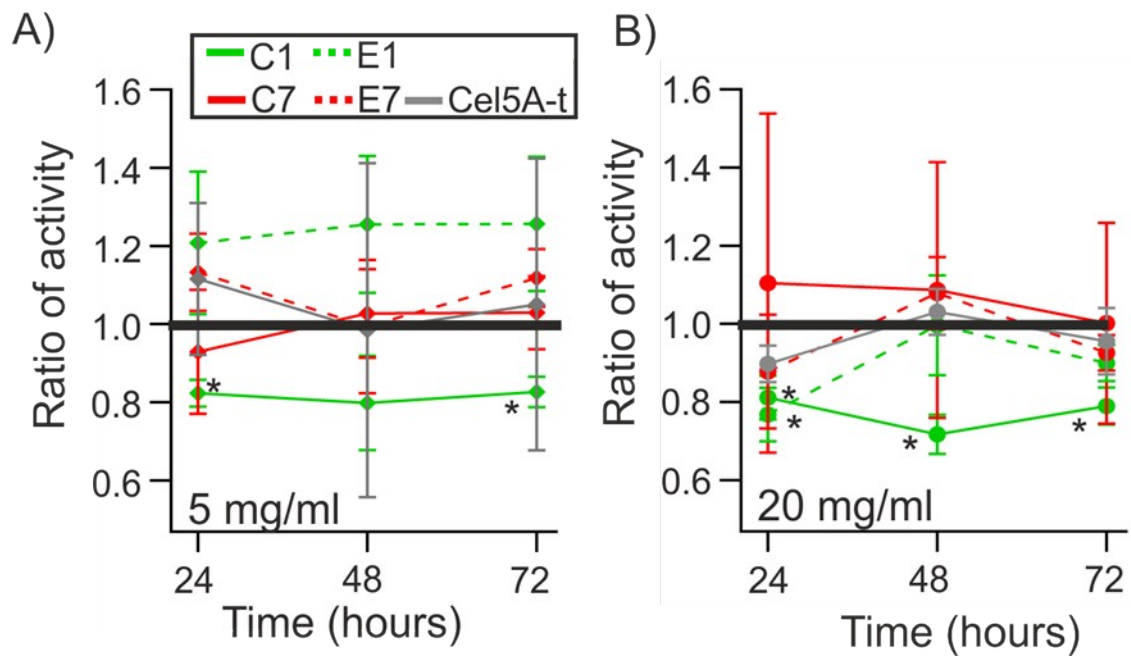
Supplementary Figure 5: Time course of the ratio of activity of bound over free cellulosomes.
Time dependence of the ratio of activity with 5 mg/ml (A) and 20 mg/ml (B). Values are given as mean ± sem to facilitate visualization.

Sequences of monovalent scaffoldins
C1:
mgsshhhhhhssglvprgshmMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVD
GQKDQTFWCDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKND
WSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGgsvvpasigtATMTVEIGKVTAAVGSKVEIPITL
KGVPSKGMANCDFVLGYDPNVLEVTEVKPGSIIKDPDPSKSFDSAIYPDRKMIVFLFAEDSGRGTYAITQDGV
FATIVATVKSAAAAPITLLEVGAFADNDLVEISTTFVAGGVNLGSSVPTTQPNVPSDtsnkpviegYKVSGYILPD
FSFDATVAPLVKAGFKVEIVGTELYAVTDANGYFEITGVPANASGYTLKISRATYLDRVIANVVVTGDTSVSTS
QDPIMMWVGDIVKDNSINLLDVAEVIRCFNATKGSANYVEELDINRNGAINMQDIMIVHKHFGATSSDYD
AQ

C7:
mgsshhhhhhssglvprgshmMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYYTVD
GQKDQTFWCDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFAKND
WSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGgsvvpasigtAVRIKVDTVNAKPGDTVRIPVRF
SGIPSKGIANCDFVYSYDPNVLEIIEIEPGELIVDPNPTKSFDTAVYPDRKMIVFLFAEDSGTGAYAITEDGVFA
TIVAKVKSGAPNGLSVIKFVEVGGFANNDLVEQKTQFFDGGVNVGtsnkpviegYKVSGYILPDFSFDATVAPL
VKAGFKVEIVGTELYAVTDANGYFEITGVPANASGYTLKISRATYLDRVIANVVVTGDTSVSTSQDPIMMWV
GDIVKDNSINLLDVAEVIRCFNATKGSANYVEELDINRNGAINMQDIMIVHKHFGATSSDYDAQ

E1:
mgsshhhhhhssglvprgshmgasigtATMTVEIGKVTAAVGSKVEIPITLKGVPSKGMANCDFVLGYDPNVLEV
TEVKPGSIIKDPDPSKSFDSAIYPDRKMIVFLFAEDSGRGTYAITQDGVFATIVATVKSAAAAPITLLEVGAFAD
NDLVEISTTFVAGGVNLGSSVPTTQPNVPSDtsMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAI
DLSKLTLRYYYTVDGQKDQTFWCDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPG
AHVQIQGRFAKNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGgsvvptsnkpviegYKVSG
YILPDFSFDATVAPLVKAGFKVEIVGTELYAVTDANGYFEITGVPANASGYTLKISRATYLDRVIANVVVTGDT
SVSTSQDPIMMWVGDIVKDNSINLLDVAEVIRCFNATKGSANYVEELDINRNGAINMQDIMIVHKHFGATS
SDYDAQ

E7:
mgsshhhhhhssglvprgshmgasigtAVRIKVDTVNAKPGDTVRIPVRFSGIPSKGIANCDFVYSYDPNVLEIIEIE
PGELIVDPNPTKSFDTAVYPDRKMIVFLFAEDSGTGAYAITEDGVFATIVAKVKSGAPNGLSVIKFVEVGGFA
NNDLVEQKTQFFDGGVNVGtsMANTPVSGNLKVEFYNSNPSDTTNSINPQFKVTNTGSSAIDLSKLTLRYYY
TVDGQKDQTFWCDHAAIIGSNGSYNGITSNVKGTFVKMSSSTNNADTYLEISFTGGTLEPGAHVQIQGRFA
KNDWSNYTQSNDYSFKSASQFVEWDQVTAYLNGVLVWGKEPGgsvvptsnkpviegYKVSGYILPDFSFDAT
VAPLVKAGFKVEIVGTELYAVTDANGYFEITGVPANASGYTLKISRATYLDRVIANVVVTGDTSVSTSQDPIM
MWVGDIVKDNSINLLDVAEVIRCFNATKGSANYVEELDINRNGAINMQDIMIVHKHFGATSSDYDAQ

CBM
CtA1
CtA7
XDock
Sequence from vector/restriction sites

Uppercase: Modules
Lowercase: linkers