Improving the Efficiency of the Assembly of Cellulosomes Derived from Clostridium Thermocellum by in Silico Design of Docking Protein

Zirui Wang
Qilu University of Technology
Cuiping Yang
Qilu University of Technology
Le Xue
Huazhong University of Science and Technology
Jie Lu
Qilu University of Technology
Peng Du
Qilu University of Technology
Nan Li
Tianjin University of Science and Technology
Piwu Li
Qilu University of Technology
Junqing Wang (wjqtt.6082@163.com)
Qilu University of Technology https://orcid.org/0000-0001-9862-2366
Ruiming Wang
Qilu University of Technology

Research Article

Keywords: cellulosome, cohesin, dockerin, Clostridium thermocellum

Posted Date: February 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1349924/v1

Abstract

Objective

The docking and adhesion proteins present in cellulosomes are critical for the efficiency of their assembly. In this study, a *in silico* design method was used to directionally modify the calcium ion-binding site of a key component, namely, the type I docking protein DocA, of the cellulosomes of *Clostridium thermocellum*.

Results

The results indicated that a mutated DocA-D41 exhibited a highest binding capacity for the type I adhesion protein Coh at a calcium ion concentration of 5×10^{-4} mol/L, which was 4.11 times the capacity of the original DocA. A molecular dynamics simulation showed that the high-frequency RMSD values for DocA-D40 and DocA-D41 were 0.232 and 0.228, lower than that of the original DocA, which implies that the mutants DocA-D40 and DocA-D41 were more stable than that of DocA.

Conclusion

The results of this study provide an efficient method for constructing efficient *C. thermocellum*-derived cellulosomes, and will lay the foundation for the design of other types of cellulosomes.

Introduction

Lignocellulose, which the main component of plant cell walls, is currently the most abundant renewable resource worldwide and has a high exploitation value. However, lignocellulose is extremely difficult to degrade because it comprises a mixture of cellulose and lignin, which results in great waste of cellulose resources (Ragauskas et al. 2006). Cellulosomes, which are generally produced by anaerobic microorganisms, are macromolecular complexes assembled from scaffolding proteins and various enzymes that are able to degrade cellulose efficiently and have attracted much attention (Bayer et al. 2004; Doi et al.2004; Champreda et al. 2019). Cellulosomes have two functions, namely, the degradation of plant cell walls and the involvement of fiber vesicles in cellular metabolism, which has been shown to efficiently degrade plant cell walls in various environments (Bayer et al. 2008). Bensoussan found that proteins containing anchoring structural domains are not only involved in the degradation of cellulose but are also extensively involved in cellular metabolic processes (Bensoussan et al. 2017; Chowdhury et al. 2014; Bobik et al. 2015; Fletcher et al.2013).

Cellulosomes are highly efficient self-assembled multienzyme systems that are generally composed of two subunits, namely, a multienzyme subunit containing an anchoring structural domain (dockerin), which has a catalytic role, and a scaffolding protein subunit containing one or more cohesion structural domains, which assembles the cellulosome complex.

The concept of artificial cellulosomes was first proposed by Bayer erin terms of the artificial design and genetic engineering of cellulosomes for efficiently degrading lignocellulose (Bayer et al. 1994). Several laboratories used recombinant DNA techniques to construct genes encoding scaffolding proteins carrying adhesion structural domains and genes encoding cellulase carrying anchoring structural domains, which were expressed, purified, and assembled *in vitro* into the predicted multienzyme complexes. Fierobe designed a series of scaffolding proteins containing two adhesion structural domains and assembled *in vitro* a dual-enzyme complex containing two cellulase domains, which had a specific activity that was seven times that of the free enzyme (Fierobe et al. 2005). Moreover, fibrillar microsomes have been proposed for the exploitation of biomass resources. Artificial cellulosomes can efficiently degrade cellulose-like substances that are difficult to degrade and are present in plant cell wall polysaccharides, and they thereby play an important role in fermentation and the production of renewable energy and provide ideas for solving problems associated with the utilization of cellulose resources (Zverlov et al. 2008).

Cellulosome fractions can be functionally assembled in engineered organisms for the efficient production of biofuels from organic waste. However, there have been few studies of improving the interactions between key components of cellulosomes via protein engineering. In order to improve the efficiency of binding between docking proteins and adhesion proteins in cellulosomes, a type I docking protein (hereinafter referred to as DocA) and a type I adhesion protein (hereinafter referred to as Coh) from *Clostridium thermophilum* were selected as the main objects for *in silico* design (Shang et al. 2018; Sagong et al. 2017; Cameron et al. 2015; Artzi et al. 2017; Kosugi et al. 2002; Barth et al. 2018; Nash et al. 2016; Fontes et al. 2010).Mutation sites were selected with the aid of Rosetta and PyMOL software, and amino acids within 4 Å of the calcium ion-binding site of DocA were regarded as key residues involved in calcium binding. Using a Biacore T200 molecular interaction analyzer, we identified the two mutants that had the highest binding capacity for Coh, and then a molecular dynamics (MD) simulation was carried out to analyze the dynamic binding between the DocA mutants and Coh.

Materials And Methods

Strains and media

Escherichia coli BL21(DE3) was used as an expression host and was cultured in Luria broth (LB) medium at 37°C. The pET-28a(+) plasmid vectors (Sangon, Shanghai, China) were used for gene cloning. The enzymes used for DNA amplification and restriction and the plasmid preparation kit were obtained from Vazyme (Nanjing, China). The primers were synthesized by Qingke (Beijing, China). All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA). The strains and plasmids used are listed in Table 1.

Table 1
Strains and plasmids used in this chapter

Name	Effect	Source
E. *coli* DH5α	Preserve plasmids and use them for proliferation and extraction	Vazyme
E. *coli* BL21(DE3)	For plasmid preservation, protein expression and extraction	Vazyme
pET-28a(+)-DocA-G	Load and express fusion protein DocA-G	GEnscript
pET-28a(+)-DocA-D40	Load and express fusion protein DocA- D40	GEnscript
pET-28a(+)-DocA-D41	Load and express fusion protein DocA- D41	GEnscript
pET-28a(+)-Coh	Load and express fusion protein Coh	GEnscript

Please insert Table 1 here.

Selection Of Key Components Of Cellulosomes

The research model of cellulosomes was derived from *Clostridium thermocellum*, which has been fully studied. The 3D structures of a docking protein (referred to as DocA; Protein Data Bank [PDB] ID code: 2CCL) and an adhesion protein (referred to as Coh; PDB ID code: 1OHZ) from the cellulosomes of *C. thermocellum* were used for preliminary structural analysis using PyMOL 2.3.2 software. The homology sequence of DocA was searched on the NCBI website using the BLAST server, and homology alignment among a family of 10 xylanase primary structures was performed using the ClustalW2 program (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The 3D structures of the DocA mutants were predicted by multiple template-based homology modeling using the SALIGN program (http://salilab.org/salign) and the MODELLER 9.9 program (http://salilab.org/modeller/).

Heterologous Expression Of Doca-egfp And Coh

The protein DocA was fused with enhanced green fluorescent protein (EGFP) via a connecting peptide with the sequence SGGGSGGGSGGS to determine the expression status of DocA in terms of fluorescence intensity. The genes corresponding to DocA-G (DocA fused with EGFP) and Coh were codon-optimized according to the genome of *E. coli* BL21(DE3) and were synthesized by GenScript (Nanjing, China). The pET-28a(+) plasmid was used with an inducible T7 promoter for heterologous expression of DocA-G and Coh with a His-Tag. The pET-28a(+)-Coh and pET-28a(+)-DocA-G vectors that were obtained were separately transfected into *E. coli* BL21(DE3) by electroporation. *E. coli* BL21(DE3) transformants were selected on the basis of their ability to grow on an LB plate containing kanamycin and were then screened by colony polymerase chain reaction (PCR) with the primer pairs Coh-F, Coh-R and DocA-F, DocA-R. The primer sequences used are shown in Table 2. The expression of DocA-G and Coh in *E. coli*

BL21(DE3) was performed according to a previously reported method. The final concentration of the inducer isopropyl thiogalactoside (IPTG) was 0.2 mmol/L.

Table 2
The list of primers

Protein name	Primer name	Base sequence
DocA-G	DocA-F	GTACTATTAGGGGATGTTGAC
	DocA-R	CTTGTACAGCTCGTCCATGCCG
Coh	Coh-F	TCAGACGGTGTGGTAGTAG
	Coh-R	TGTTGCTCCCTTGGTCGGTGTTGC
DocA-D41	41-F	TGCCGAT accagcaat AATGGC acc ATTAATGCC
	41-R	CATTAAT ggt GCCATT attgctggt ATCGGCACGG
DocA-D40	40-F	GTGCCGAT accagcaatgat GGC tat ATTAATGCC
	40-R	GGCATTAAT ata GCC atcatgctggt ATCGGCACGGGCTTT

Please insert Table 2 here.

Purification Of Recombinant Doca-egfp And Coh

The induced bacterial cells were resuspended in 2 × phosphate-buffered saline (PBS) buffer and were then disrupted using an ultrasonic disintegrator (Scientz-650E, NingBo, China). After centrifugation at 12000 × g, a sample of 50 mL culture supernatant was brought to 75% saturation by the addition of solid $(NH_4)_2SO_4$. The precipitate was harvested, dissolved in 5 mL of 20 mmol/L Na_2HPO_4–NaH_2PO_4 buffer (pH 6.0), and dialyzed against the same buffer overnight. The dialysate was concentrated to 1 mL by ultrafiltration using a membrane with a 3 kDa cut-off (Millipore, Billerica, MA, USA) and was loaded onto a HisTrap HP affinity chromatography column (GE, PalosAlto, USA), followed by elution with a linear gradient of 0–400 mmol/L imidazole in the abovementioned buffer at a flow rate of 0.4 mL/min. Aliquots of 2 mL eluate that only contained the target xylanase were pooled, dialyzed against deionized water, and concentrated. The purified protein was detected by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), and its purity was ascertained. Protein samples that met the experimental requirements were collected, and desalting and concentration were separately performed. DocA-G was dialyzed to remove salts using PBS-EP + buffer, and Coh was dialyzed to remove salts using acetic acid–sodium acetate buffers with different pH values. All purification procedures were performed at 4°C unless stated otherwise.

Design And Identification Of Mutant Docking Proteins

The protein sequence data for DocA-G were imported into the Rosetta website (http://rosettadesign.med.unc.edu) for *in silico* design. One key sequence involved in calcium ion binding, namely, $DV^{40}D^{41}K^{42}N^{43}GS^{45}$, was selected as a potential mutation site after protein–protein docking and analysis. As shown in Fig. 1. Two single-site DocA-G mutants, which were predicted to have the most stable structures, were identified. We used rapid site-directed mutagenesis (Rapid Site-Directed Mutation Kit Tiangen, Beijing) to obtain these mutants.

Please insert Fig. 1 here.

Interaction between DocA-G and Coh and determination of binding ability

The DocA-G mutants were heterologously expressed and purified. A Biacore T200 molecular interaction analyzer was employed to investigate the binding mechanism of DocA-G and Coh. Coh was anchored on a CM5 sensor chip, and DocA-G mixed with different concentrations of $CaCl_2$ was allowed to flow through.

In accordance with the standard protocol, purified Coh was immobilized on the entire surface of a CM5 sensor chip, and the chip was then loaded into the analyzer. The channel of the DocA-G mutant was used as the detection channel, and the channel of the original DocA-G was used as the reference channel. 4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid-buffered saline was used as the mobile phase. The flow rate of the flow pool was 10 μL/min, and the temperature was set to 20 ℃. DocA-G was diluted to approximately the same concentration as the immobilized Coh, and $CaCl_2$ was serially diluted to concentrations of 1.00×10^{-2}, 1.00×10^{-3}, 1.00×10^{-4}, 1.00×10^{-5}, 1.00×10^{-6}, and 1.00×10^{-7} mol/L and stored at 4°C. The reaction time was strictly controlled at 30 min. We repeated the experiment three times using the same concentrations to confirm the repeatability of the results. After centrifugation, the ligand was injected into the detection and reference channels at a rate of 10 μL/min, and the binding status was determined according to the value of the absorption response.

Molecular Dynamics Simulation

We used RosettaDock 3.4 molecular docking software to construct the DocA–Coh complex. An MD simulation was performed using the GROMACS 4.5.4 package with the GROMOS 96 force field and the SPC/E explicit water model. Each system was minimized and equilibrated until the maximum force reached 10 kJ/(mol nm), as previously described. We gradually equilibrated the equilibration systems at 300 K for 100 ps with the restrained protein and ligand. After periodic boundary conditions were applied, electrostatic interactions were treated using the particle mesh Ewald method. The integration step was set to 0.002 ps, and bonds were constrained using the LINCS algorithm. After the first equilibration step, full equilibration was carried out for 5 ns without restraints, and then the g_rms tool was used to calculate the root mean square deviation (RMSD) values for the interacting enzymes.

Results

Selection and expression of the key components of cellulosomes

The docking protein DocA (PDB ID code: 2CCL) and the adhesion protein Coh (PDB ID code: 1OHZ), whose crystal structures had been investigated, were selected as the targets. The selected genes were initially optimized and synthesized in accordance with codons of *E. coli* and were then transfected into *E. coli* BL21(DE3) with the help of the pET-28a(+) plasmid for gene expression. When the optical density at 600 nm reached 1.0, IPTG at a final concentration of 0.2 mmol/L was added to induce protein expression, and induction was carried out at 22 °C for 10 h. After ultrasonic fragmentation and centrifugation, Coh and DocA-G were purified by salting out, ultrafiltration, and affinity chromatography using a HisTrap HP column and were then analyzed by SDS-PAGE. As shown in Fig. 2, the molecular weight of Coh was about 16.7 kDa (lanes 2 and 4), whereas the molecular weight of DocA-G was about 36.7 kDa (lanes 1 and 3). The molecular weights of the purified proteins were close to the theoretical values, which indicated that Coh and DocA-G, i.e., the key components of cellulosomes, were successfully purified.

Please insert Fig. 2 here.

Analysis of binding ability of DocA-G and Coh

After the purification process, the docking mechanisms of Coh and DocA-G at different calcium ion concentrations were investigated using a Biacore T200 molecular interaction analyzer (GE Healthcare, Chicago, IL, USA). It was found that the higher was the concentration of calcium ions in the range from 1.00×10^{-7} to 1.00×10^{-4} mol/L, the higher was the binding ability of Coh and DocA-G. However, when the calcium ion concentration was lower than 1.00×10^{-4} mol/L, the binding ability of Coh and DocA-G was similar to that when the calcium ion concentration was 1.00×10^{-7} mol/L. It can be tentatively concluded that the interaction between Coh and DocA-G to form a stable structure requires the participation of calcium ions at a concentration of about 1.00×10^{-4} mol/L.

Design and selection of DocA-G mutants

Mutation sites in DocA-G were selected with the help of PyMOL software. The amino acids V^{40}, D^{41}, K^{42}, N^{43}, and S^{45}, which were within 0.4 nm of the calcium ion-binding site of DocA-G, were selected as the key residues involved in calcium binding. Then, the key residues were altered and simulated using the Rosetta website. The two highest-scoring mutants, namely, DocA-D40 (containing T^{40}, S^{41}, N^{42}, D^{43}, and Y^{45}) and DocA-D41 (containing T^{40}, S^{41}, N^{42}, and T^{45}), were selected for protein–protein docking and analysis.

The genes encoding DocA-D40 and DocA-D41 were constructed using a rapid site-specific mutagenesis kit (Tiangen, Beijing, China). Using the pET-28a(+)-DocA-G vector as a template, the genes *docA-D40* and *docA-D41* were amplified with the 40-F, 40-R primers and the 41-F, 41-R primers, respectively. The primer sequences used are shown in Table 2. The target PCR products were gel-purified, inserted into the pET-

28a(+) plasmid, and transfected into *E. coli* BL21(DE3). The resulting recombinant vectors, namely, pET-28a(+)-DocA-D40 and pET-28a(+)-DocA-D41, were identified by DNA sequencing. The specific primer sequences used for mutagenesis are listed in Fig. 3.

Please insert Fig. 3 here.

In vitro confirmation of binding of DocA-G mutants to Coh.

As shown in Fig. 4, when the calcium ion concentration was in the range from 1.00×10^{-7} to 1.00×10^{-4} mol/L the binding capacities of DocA-D41 and DocA-G were almost identical, although the binding capacity of DocA-D41 was slightly higher at a calcium ion concentration of 1.00×10^{-4} mol/L and was about 1.2 times that of DocA-G. When the calcium ion concentration was in the range from 1.00×10^{-5} to 1.00×10^{-2} mol/L the binding capacities of DocA-D40 and DocA-D41 were about 3.68 times and 4.11 times that of the original protein DocA-G, respectively. Moreover, DocA-D41 exhibited the highest binding capacity for Coh at a calcium ion concentration of 5×10^{-4} mol/L.

Please insert Fig. 4 here.

Molecular dynamics simulation and structural analysis of DocA-G mutants

We constructed different DocA-D40-Coh and DocA-D41-Coh complexes using RosettaDock 3.4 and then performed an MD simulation for 5 ns using GROMACS 4.5 software. Using the g_rms tool in GROMACS 4.5, the difference parameters (i.e., RMSD) for the structures of the mutants and that of the original docking protein DocA (with/without Ca^{2+}) were calculated. The RMSD values for the mutants DocA-D40 and DocA-D41 (0.232 and 0.228, respectively) were lower than that for DocA (0.378), which implies that the structures of DocA-D40 and DocA-D41 are more stable than that of DocA.

Please insert Fig. 5 here.

Discussion

The value of cellulosomes in the conversion of cellulose has been recognized with advances in research on cellulosomes, which have led to new ideas for artificially designing and modifying natural cellulosomes to act more efficiently in the degradation of cellulose. The concept of artificial cellulosomes was first proposed by Bayer et al. in terms of the artificial design and genetic engineering of cellulosomes for efficiently degrading lignoc ellulose. Several researchers have used techniques such as DNA recombination to construct genes encoding scaffolding proteins carrying adhesion domains and genes encoding cellulase carrying anchoring domains, which were expressed and purified to assemble the desired multienzyme complex *in vitro* (Biswas et al.2015). Fierobe designed a series of scaffolding proteins containing two adhesion structural domains and assembled *in vitro* a dual-enzyme complex containing two cellulase domains, which had a specific activity that was seven times that of the free enzyme (Fierobe et al. 2005). Morais et al. constructed heat-stable exoglucanase Cel48S, endoglucanase

Cel8A, and heat-stable β-glucosidase from *C. thermophilum* by error-prone PCR and introduced them into artificial cellulosomes (Moraïs et al. 2016). The results showed that the degradation rate of the "heat-stable" artificial cellulosomes increased by a factor of 1.7 in comparison with conventionally designed artificial cellulosomes. Carvalho et al. found that DocA had two calcium ion-binding sites, of which one was used to stabilize the protein structure and the other was used for stable binding to Coh (Lytle et al. 2000). Research on cellulosomes currently mostly focuses on their structural analysis and applications, but reports on how to improve the binding efficiency of key components of cellulosomes by rational design have been rare (Igarashi et al. 2009; Jeon et al. 2012; Haimovitz et al. 2008).

In this study, we initially used a Biacore T200 molecular interaction analyzer to determine the binding affinities of Coh and DocA and found that binding between Coh and DocA occurred when the calcium ion concentration was in the range from 1.00×10^{-4} to 1.00×10^{-2} mol/L, whereas a high calcium ion concentration may inhibit the binding of Coh and DocA. In order to improve the binding efficiency of Coh and DocA, structure data for DocA-G were imported into the Rosetta website for *in silico* design. The two highest-scoring mutants, namely, DocA-D40 and DocA-D41, were selected for protein–protein docking and analysis. The results showed that the binding capacities of DocA-D40 and DocA-D41 were about 3.68 times and 4.11 times that of the original protein DocA-G, respectively. These results make it possible to improve the binding activity between components of cellulosomes via *in silico* design based on 3D structures. The mutant DocA-D40 exhibited a higher binding capacity when the calcium ion concentration was 1.00×10^{-4} mol/L, which implies that, via *in silico* design of the calcium ion-binding site, cellulosomes can be assembled in the presence of lower concentrations of calcium ions.

We also performed MD simulations to study the interactions of the DocA mutants with Coh. The results showed that the DocA mutants differed from the original protein to a greater or lesser extent and that the differences were mainly concentrated in the loop region. We used the g_rms tool in GROMACS 4.5.4 to calculate the RMSD values for the structures of the DocA mutants and that of the original protein DocA. As shown in Fig. 5. The mutants DocA-D40 and DocA-D41 had smaller RMSD values than DocA (without Ca^{2+}). In contrast to DocA (0.378), the high-frequency RMSD values for DocA-D40 and DocA-D41 were 0.232 and 0.228, respectively, which implies that the structures of the mutants DocA-D40 and DocA-D41 are more stable than that of DocA. Hence, the mutant docking proteins and adhesion protein are easier to assemble with calcium ions.

In conclusion, we developed a method based on the use of a Biacore T200 molecular interaction analyzer to measure activity involved in the assembly of cellulosomes. Moreover, via *in silico* design of the calcium ion-binding site based on the structure of DocA, two DocA mutants with higher binding capacities for Coh were obtained. As shown in Fig. 4 The binding capacities of the mutants DocA-D40 and DocA-D41 were about 3.68 times and 4.11 times that of the original protein DocA-G, respectively. DocA-D41 exhibited the highest binding capacity for Coh at a calcium ion concentration of 5×10^{-4} mol/L. By an MD simulation and structural analysis, we found that the RMSD values for the mutants DocA-D40 and DocA-D41 were lower than those for the original protein DocA, which implies that the structures of DocA-D40 and DocA-D41 are more stable as a result of mutation. Our findings provide an

effective method for constructing efficient cellulosomes derived from those in *C. thermocellum* and will lay a foundation for the design of other types of cellulosome.

Declarations

Acknowledgments

This work was supported by the National Science Foundation of China (31801527), Focus on Research and Development Plan in Shandong Province (2021ZDSYS10, 2020CXGC010603, 2019JZZY011003), Taishan industry leading talent (tscy20180103), National Key Research and Development Project (2019YFC1905900) and Cultivation Project of Shandong Synthetic Biotechnology Innovation Center (sdsynbio-2018-PY-02).

References

1. Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006) The path forward for biofuels and biomaterials. Science 311:484–489. doi: 10.1126/science.1114736
2. Bayer EA, Belaich JP, Shoham Y, Lamed R (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu Rev Microbiol 58:521–554. 10.1146/annurev.micro.57.030502.091022
3. Doi RH, Kosugi A (2004) Cellulosomes: plant-cell-wall-degrading enzyme complexes. Nat Rev Microbiol 2:541–551. doi:10.1038/nrmicro925
4. Champreda V, Mhuantong W, Lekakarn H, Bunterngsook B, Kanokratana P, Zhao XQ, Zhang F, Inoue H, Fujii T, Eurwilaichitr L (2019) Designing cellulolytic enzyme systems for biorefinery: From nature to application. J Biosci Bioeng 128:637–654. doi:10.1016/j.jbiosc.2019.05.007
5. Bayer EA, Lamed R, White BA, Flint HJ (2008) From cellulosomes to cellulosomics. Chem Rec 8:364–377. doi: 10.1002/tcr.20160
6. Bensoussan L, Moraïs S, Dassa B, Friedman N, Henrissat B, Lombard V, Bayer EA, Mizrahi I (2017) Broad phylogeny and functionality of cellulosomal components in the bovine rumen microbiome. Environ Microbiol 19:185–197. doi: 10.1111/1462-2920.13561
7. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA (2014) Diverse bacterial microcompartment organelles. Microbiol Mol Biol Rev 78:438–468. doi: 10.1128/MMBR.00009-14
8. Bobik TA, Lehman BP, Yeates TO (2015) Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. Mol Microbiol 98:193–207. doi: 10.1111/mmi.13117
9. Fletcher JM, Harniman RL, Barnes FR, Boyle AL, Collins A, Mantell J, Sharp TH, Antognozzi M, Booth PJ, Linden N, Miles MJ, Sessions RB, Verkade P, Woolfson DN (2013) Self-assembling cages from coiled-coil peptide modules. Science 340:595–599. doi: 10.1126/science.1233936

10. Bayer EA, Morag E, Lamed R (1994) The cellulosome–a treasure-trove for biotechnology. Trends Biotechnol 12:379–386. doi: 10.1016/0167-7799(94)90039-6
11. Fierobe HP, Mingardon F, Mechaly A, Bélaïch A, Rincon MT, Pagès S, Lamed R, Tardif C, Bélaïch JP, Bayer EA (2005) Action of designer cellulosomes on homogeneous versus complex substrates: controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. J Biol Chem 22:280:16325–16334. doi: 10.1074/jbc.M414449200
12. Zverlov VV, Klupp M, Krauss J, Schwarz WH (2008) Mutations in the scaffoldin gene, cipA, of *Clostridium thermocellum* with impaired cellulosome formation and cellulose hydrolysis: insertions of a new transposable element, IS1447, and implications for cellulase synergism on crystalline cellulose. J Bacteriol 190:4321–4327. doi: 10.1128/JB.00097-08
13. Shang X, Chai X, Lu X, Li Y, Zhang Y, Wang G, Zhang C, Liu S, Zhang Y, Ma J, Wen T (2018) Native promoters of *Corynebacterium glutamicum* and its application in L-lysine production. Biotechnol Lett 40:383–391. doi: 10.1007/s10529-017-2479-y
14. Sagong HY, Kim KJ (2017) Structural basis for redox sensitivity in *Corynebacterium glutamicum* diaminopimelate epimerase: an enzyme involved in L-lysine biosynthesis. Sci Rep 7:42318. doi: 10.1038/srep42318
15. Cameron K, Najmudin S, Alves VD, Bayer EA, Smith SP, Bule P, Waller H, Ferreira LM, Gilbert HJ, Fontes CM (2015) Cell-surface attachment of bacterial multienzyme complexes involves highly dynamic protein-protein anchors. J Biol Chem 290:13578–13590. doi: 10.1074/jbc.M114.633339
16. Artzi L, Bayer EA, Moraïs S (2017) Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. Nat Rev Microbiol 15:83–95. doi: 10.1038/nrmicro.2016.164
17. Kosugi A, Murashima K, Tamaru Y, Doi RH (2002) Cell-surface-anchoring role of N-terminal surface layer homology domains of *Clostridium cellulovorans* EngE. J Bacteriol 184:884–888. doi: 10.1128/jb.184.4.884-888.2002
18. Barth A, Hendrix J, Fried D, Barak Y, Bayer EA, Lamb DC (2018) Dynamic interactions of type I cohesin modules fine-tune the structure of the cellulosome of *Clostridium thermocellum*. Proc Natl Acad Sci U S A 115:11274–11283. doi: 10.1073/pnas.1809283115
19. Nash MA, Smith SP, Fontes CM, Bayer EA (2016) Single versus dual-binding conformations in cellulosomal cohesin-dockerin complexes. Curr Opin Struct Biol 40:89–96. doi: 10.1016/j.sbi.2016.08.002
20. Fontes CM, Gilbert HJ (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. Annu Rev Biochem 79:655–681. doi: 10.1146/annurev-biochem-091208-085603
21. Biswas R, Zheng T, Olson DG, Lynd LR, Guss AM (2015) Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in *Clostridium thermocellum*.Biotechnol Biofuels. 8:20. doi: 10.1186/s13068-015-0204-4.
22. Fierobe HP, Mingardon F, Mechaly A, Bélaïch A, Rincon MT, Pagès S, Lamed R, Tardif C, Bélaïch JP, Bayer EA (2005) Action of designer cellulosomes on homogeneous versus complex substrates:

controlled incorporation of three distinct enzymes into a defined trifunctional scaffoldin. J Biol Chem 280:16325–16334. doi: 10.1074/jbc.M414449200

23. Moraïs S, Stern J, Kahn A, Galanopoulou AP, Yoav S, Shamshoum M, Smith MA, Hatzinikolaou DG, Arnold FH, Bayer EA (2016) Enhancement of cellulosome-mediated deconstruction of cellulose by improving enzyme thermostability. Biotechnol Biofuels 9:164. doi: 10.1186/s13068-016-0577-z
24. Lytle BL, Volkman BF, Westler WM, Wu JH (2000) Secondary structure and calcium-induced folding of the *Clostridium thermocellum* dockerin domain determined by NMR spectroscopy. Arch Biochem Biophys 379:237–244. doi: 10.1006/abbi.2000.1882
25. Igarashi K, Koivula A, Wada M, Kimura S, Penttilä M, Samejima M (2009) High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. J Biol Chem 284:36186–36190. doi: 10.1074/jbc.M109.034611
26. Jeon SD, Lee JE, Kim SJ, Kim SW, Han SO (2012) Analysis of selective, high protein-protein binding interaction of cohesin-dockerin complex using biosensing methods. Biosens Bioelectron 35:382–389. doi: 10.1016/j.bios.2012.03.023
27. Haimovitz R, Barak Y, Morag E, Voronov-Goldman M, Shoham Y, Lamed R, Bayer EA (2008) Cohesin-dockerin microarray: Diverse specificities between two complementary families of interacting protein modules. Proteomics 8:968–979. doi: 10.1002/pmic.200700486

Figures

Figure 1

Ca^{2+} binding key amino acid selection and saturation mutation design.

PyMol 1.7 was used to analyze the structural characteristics of the protein DocA, and the amino acids in the protein DocA that were close to the calcium ion were selected as the key amino acids for calcium ion binding.

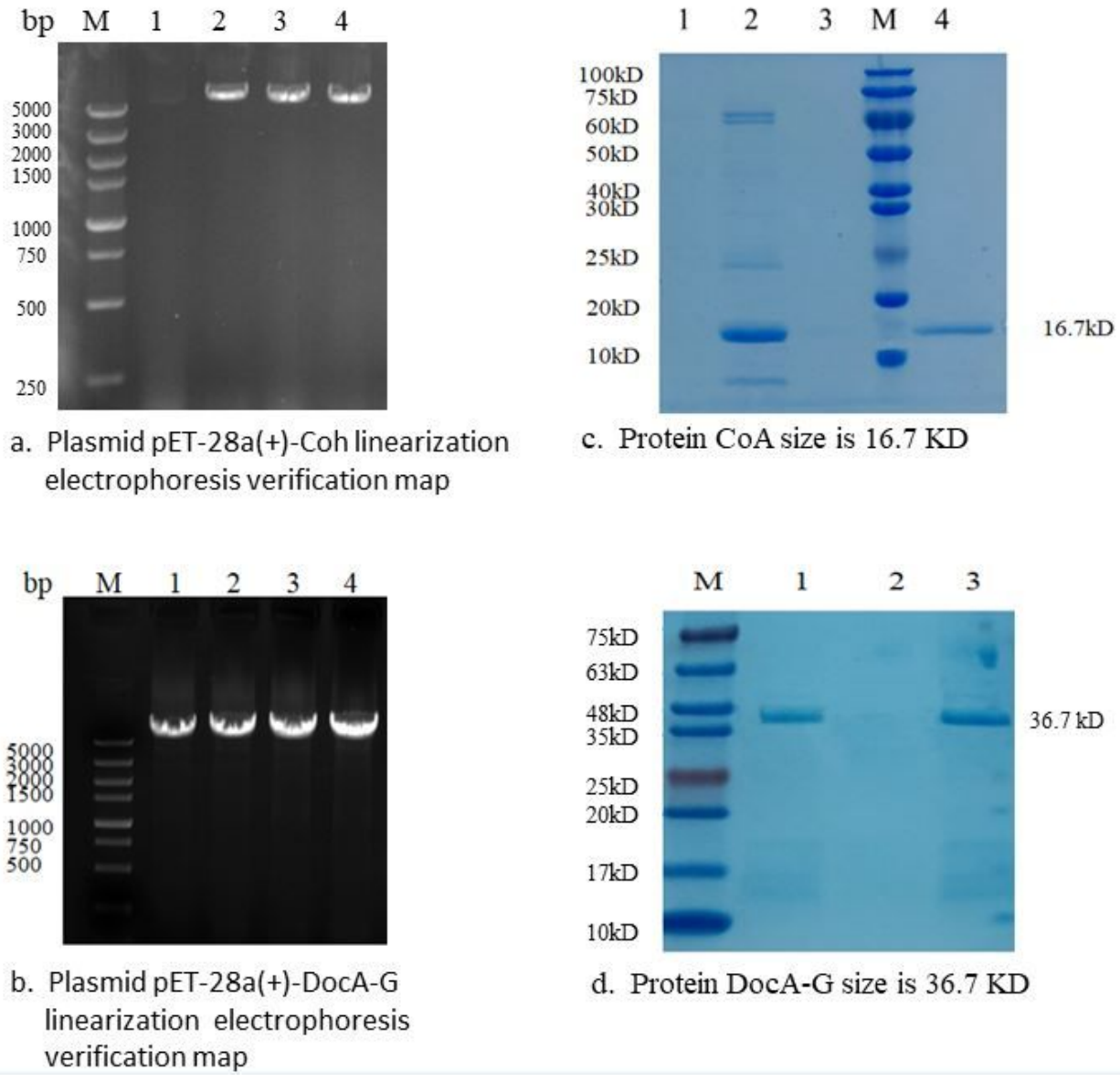

Figure 2

Electrophoretogram of purified Coh protein and DocA-G protein.

The docking protein DocA-G and the adhesion protein Coh, whose crystal structures had been investigated, were selected as the targets. The selected genes were initially optimized and synthesized in accordance with codons of *E. coli* and were then transfected into *E. coli* BL21(DE3) with the help of the pET-28a(+) plasmid for gene expression. After ultrasonic fragmentation and centrifugation, Coh and DocA-G were purified by salting out, ultrafiltration, and affinity chromatography using a HisTrap HP column and were then analyzed by SDS-PAGE. a. Lanes 2-4 are the linearized pET-28a-Coh plasmid (5758bp). b. Lanes 1-4 are the linearized pET-28a-DocA-G plasmid (6217bp). c. The molecular weight of Coh was about 16.7 kDa (lanes 2 and 4). d. The molecular weight of DocA-G was about 36.7 kDa (lanes 1 and 3).

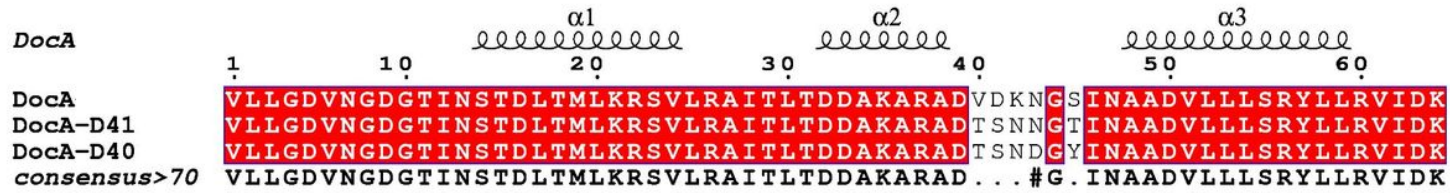

Figure 3

Multiple sequence alignment of DocA and its mutants.

Two DocA-G mutants, which were predicted to have the most stable structures, were obtained by Rosettadesign. The picture is a list of two DocA-G mutants amino acids and the original amino acid sequence.

Figure 4

Binding capacity map of single-site mutants and Coh at different Ca^{2+} concentrations.

When the calcium ion concentration was in the range from 1.00×10^{-7} to 1.00×10^{-4} mol/L the binding capacities of DocA-D41 and DocA-G were almost identical, although the binding capacity of DocA-D41 was slightly higher at a calcium ion concentration of 1.00×10^{-4} mol/L and was about 1.2 times that of DocA-G. When the calcium ion concentration was in the range from 1.00×10^{-5} to 1.00×10^{-2} mol/L the binding capacities of DocA-D40 and DocA-D41 were about 3.68 times and 4.11 times that of the original protein DocA-G, respectively. Moreover, DocA-D41 exhibited the highest binding capacity for Coh at a calcium ion concentration of 5×10^{-4} mol/L.

Figure 5

Scatter plots of RMSD for 4 kinds single-point mutants.

We constructed different DocA-D40-Coh and DocA-D41-Coh complexes using RosettaDock 3.4 and then performed an MD simulation for 5 ns using GROMACS 4.5 software. Using the g_rms tool in GROMACS 4.5, the difference parameters (i.e., RMSD) for the structures of the mutants and that of the original docking protein DocA (with/without Ca^{2+}) were calculated. The RMSD values for the mutants DocA-D40 and DocA-D41 (0.232 and 0.228, respectively) were lower than that for DocA (0.378), which implies that the structures of DocA-D40 and DocA-D41 are more stable than that of DocA.