Revisiting the NMR solution structure of the Ce148S type-I dockerin module from Clostridium thermocellum reveals a cohesin-primed conformation

ARTICLE *in* JOURNAL OF STRUCTURAL BIOLOGY · NOVEMBER 2014

Impact Factor: 3.23 · DOI: 10.1016/j.jsb.2014.09.006

Journal of Structural Biology 188 (2014) 188–193

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Structure Report

Revisiting the NMR solution structure of the Cel48S type-I dockerin module from *Clostridium thermocellum* reveals a cohesin-primed conformation

Chao Chen [a,c], Zhenling Cui [a,c], Yan Xiao [a], Qiu Cui [a,b], Steven P. Smith [d], Raphael Lamed [e], Edward A. Bayer [f], Yingang Feng [a,*]

[a] *Shandong Provincial Key Laboratory of Energy Genetics, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, 189 Songling Road, Qingdao 266101, China*
[b] *Key Laboratory of Biofuels, Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, 189 Songling Road, Qingdao, Shandong 266101, China*
[c] *University of Chinese Academy of Sciences, Beijing 100049, China*
[d] *Department of Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario K7L 3N6, Canada*
[e] *Department of Molecular Microbiology and Biotechnology, Tel Aviv University, 69978 Tel Aviv, Israel*
[f] *Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel*

ARTICLE INFO

Article history:
Received 21 August 2014
Received in revised form 20 September 2014
Accepted 22 September 2014
Available online 28 September 2014

Keywords:
Protein structure
Protein–protein interaction
Cellulosome
Dockerin
Cohesin
Cis-proline

ABSTRACT

Dockerin modules of the cellulosomal enzyme subunits play an important role in the assembly of the cellulosome by binding tenaciously to cohesin modules of the scaffoldin subunit. A previously reported NMR-derived solution structure of the type-I dockerin module from Cel48S of *Clostridium thermocellum*, which utilized two-dimensional homonuclear ^{1}H-^{1}H NOESY and three-dimensional ^{15}N-edited NOESY distance restraints, displayed substantial conformational differences from subsequent structures of dockerin modules in complex with their cognate cohesin modules, raising the question whether the source of the observed differences resulted from cohesin-induced structural rearrangements. Here, we determined the solution structure of the Cel48S type-I dockerin based on ^{15}N- and ^{13}C-edited NOESY-derived distance restraints. The structure adopted a fold similar to X-ray crystal structures of dockerin modules in complex with their cohesin partners. A unique cis-peptide bond between Leu-65 and Pro-66 in the Cel48S type-I dockerin module was also identified in the present structure. Our structural analysis of the Cel48S type-I dockerin module indicates that it does not undergo appreciable cohesin-induced structural alterations but rather assumes an inherent calcium-dependent cohesin-primed conformation.

1. Introduction

Plant cell wall polysaccharides comprise the most abundant reservoir of organic carbon in the biosphere. To facilitate the efficient degradation of valuable carbon, a subset of anaerobic cellulolytic organisms have integrated complementary catalytic enzymes, such as cellulases and hemicellulases, into non-catalytic scaffoldin protein subunits to form a multi-enzyme complex termed the cellulosome (Bayer et al., 2004). Indeed, the cellulosome affords synergistic and conjugative interactions among the enzyme constituents allowing for efficient degradation of cellulosic biomass (Demain et al., 2005; Fierobe et al., 2002). The cellulosomal enzyme subunits are integrated into the cellulosome through a high-affinity interaction between their associated type-I dockerin modules and the type-I cohesin modules of the scaffoldin subunit (Bayer et al., 2008).

A single structure of an isolated cellulosomal dockerin (Doc48S, PDB: 1DAV and 1DAQ) from the most abundance cellulosomal enzyme Cel48S of *Clostridium thermocellum* has been reported (Lytle et al., 2001), while several X-ray crystal structures of type-I dockerins in complex with their cognate type-I cohesin binding partners have been determined (Bras et al., 2012; Carvalho et al., 2003; Currie et al., 2012; Pinheiro et al., 2008). These latter structures revealed two nearly anti-parallel α-helices corresponding to a tandem duplicated amino acid sequence in the type-I dockerin modules, which forms two symmetric cohesin-binding sites resulting in a dual binding mode (Carvalho et al., 2003, 2007; Pinheiro et al., 2008). However, the NMR structure of the isolated Cel48S

* Corresponding author. Fax: +86 532 80662707.
E-mail address: fengyg@qibebt.ac.cn (Y. Feng).

http://dx.doi.org/10.1016/j.jsb.2014.09.006
1047-8477/

dockerin of *C. thermocellum* showed significant deviation from the cohesin-bound type-I dockerin structures. It displayed a more open conformation with a distance of 9–11 Å between the Cα atoms of the two α-helices, compared with 6–7 Å in the cohesin-bound dockerin crystal structures (Carvalho et al., 2003). The symmetry of the two helices and cohesin-binding sites in the Cel48S structure were poorly defined in the NMR-derived solution structure of the Cel48S type-I dockerin module.

In this study, we revisited the NMR structure of the Cel48S dockerin module (Doc48S) using a suite of ^{15}N- and ^{13}C-edited experiments. The resultant Doc48S structure revealed the protein module adopts a cohesin-primed conformation that undergoes very little appreciable structure rearrangement upon binding cohesin.

2. Methods and materials

2.1. Protein expression and purification

The gene encoding the dockerin module of *C. thermocellum* Cel48S (Doc48S; UniProtKB accession code A3DH67; residues 673–741) was cloned into a modified pET28a vector containing a hexahistidine-SMT3-tag (Mossessova and Lima, 2000) using the upstream primer (5′-CGCGGATCCACTAAATTATACGGCGAC-3′) contained the *Bam*HI site and the downstream primer (5′-CCGCTCGAGTTAGTTCTTGTACGGCAATGT-3′) contained the *Xho*I site. The resultant plasmid was transformed into *Escherichia coli* strain BL21 (DE3) for protein expression. Uniformly ^{15}N- and ^{15}N/^{13}C-labeled Doc48S was obtained by growth of transformed bacteria on M9 minimal media, supplemented with $^{15}NH_4Cl$ and ^{13}C-glucose as sole nitrogen and carbon sources, respectively. Upon reaching an optical density at 600 nm of 0.9–1.0, recombinant protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside to a final concentration of 0.5 mM, followed by growth for an additional 5 h at 37 °C. Bacterial cells were harvested by centrifugation, resuspended in 20 ml of binding buffer (20 mM Na_3PO_4, 500 mM NaCl, 30 mM imidazole, pH 8.0), and lysed by freeze–thaw followed by sonication. After centrifugation at 10,000*g* for 30 min, the supernatant was applied onto a Histrap™ HP (GE Healthcare) column equilibrated with binding buffer. Bound proteins were eluted using binding buffer increasing in imidazole concentration (30–500 mM) of imidazole. The eluent was pooled, treated with ULP1 protease, dialyzed overnight against binding buffer, and reapplied to the Histrap™ HP (GE Healthcare) column. SDS–PAGE analysis indicated that the flowthrough contained Doc48S, which, after concentrating by ultrafiltration, was applied to a Superdex 75 column (GE Healthcare), equilibrated with 50 mM Bis-Tris pH 6.6, 100 mM KCl, 20 mM $CaCl_2$. Fractions containing purified Doc48S, identified by SDS–PAGE analysis, were pooled and concentrated. NMR samples consisted of 0.3–0.5 mM Doc48S in 50 mM Bis-Tris, pH 6.6, 100 mM KCl, 20 mM $CaCl_2$, 0.02% (w/v) sodium 2,2-dimethylsilapentane-5-sulfonate (DSS), 1× protease inhibitor cocktail (Roche), 90% H_2O/10% D_2O.

2.2. NMR spectroscopy and structure calculations

All NMR data were collected at 298 K on a Bruker Avance III 600 MHz NMR spectrometer equipped with a *z*-gradient triple resonance cryoprobe. Backbone and side chain ^{1}H, ^{13}C, and ^{15}N resonance assignments were derived from 2D ^{1}H–^{15}N HSQC, 2D ^{1}H–^{13}C HSQC, 3D ^{1}H–^{13}C–^{15}N HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HBHA(CBCA)(CO)NH, HBHA(CBCA)NH, H(C)CH–TOCSY and (H)CCH–TOCSY datasets. Distance restraints for structure calculations were generated by using the 3D ^{15}N and ^{13}C-edited NOESY-HSQC spectra with mixing times of 200 ms. ^{1}H, ^{13}C, and ^{15}N chemical shifts were referenced according to IUPAC recommendations using the internal DSS (Markley et al., 1998). All NMR spectra were processed using NMRPipe (Delaglio et al., 1995) and analyzed with NMRViewJ (Johnson and Blevins, 1994).

Initial Doc48S structures were calculated using the CANDID module of the CYANA software (Herrmann et al., 2002) with the NOE peak lists from the ^{15}N-edited and aliphatic ^{13}C-edited NOESY-HSQC datasets and dihedral angle restraints derived from the chemical shifts using the program TALOS+ (Shen et al., 2009). The Doc48S structures were refined using the distance restraints derived by the semi-automatic assignment program SANE (Duggan et al., 2001), as well as hydrogen bond restraints introduced according to the secondary structure elements and Ca^{2+} restraints according to the conserved Ca^{2+}-coordinating residues (Carvalho et al., 2003; Strynadka and James, 1989). Using these restraints, 100 structures were calculated by CNS (Brunger et al., 1998), and the 20 lowest energy conformers were selected to represent the final ensemble of structures of Doc48S. The final structures were analyzed using PROCHECK_NMR (Laskowski et al., 1996), MOLMOL (Koradi et al., 1996) and WHAT_CHECK (Hooft et al., 1996). The Doc48S structures have been deposited into Protein Data Bank (PDB ID: 2MTE), and the chemical shift assignments have been deposited into the BioMagResBank under accession number 25158.

3. Results and discussion

3.1. Analysis of previous type-I dockerin structures

Owing to the observed discrepancies between the isolated Doc48S structure (Lytle et al., 2001) and the Xyn10B dockerin structure from the original type-I cohesin-dockerin complex

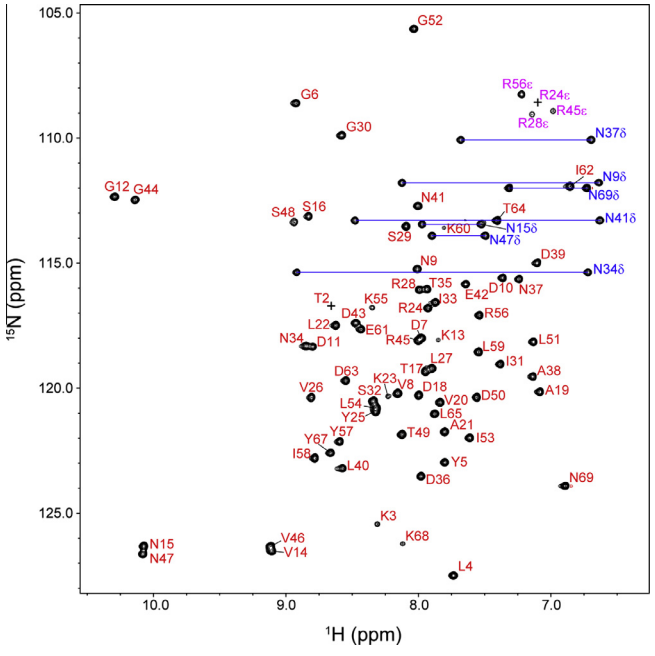

Fig.1. ^{1}H-^{15}N HSQC spectrum of Doc48S at pH 6.6 and 298 K. Resonances for backbone and side chain amide groups that have been assigned are labeled with one-letter code for amino acid residue type followed by the position in the sequence. Labels for Asn side chain NH_2 resonances are identified in blue with an accompanying δ, and are connected by a horizontal line. Resonances corresponding to the side chain NH^{ε} of Arg are labeled with an accompanying ε in magenta, and are folded into the spectrum from their original ^{15}N chemical shifts by adding a spectrum width of 24 ppm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Carvalho et al., 2003), it had been proposed that the type-I dockerin module adopts a flexible conformation in solution, which undergoes a substantial conformational change upon binding to its cognate cohesin module. Nevertheless, Lytle et al. also reported the NMR backbone dynamics parameters, which showed that the average order parameter S^2 value for residues 5 to 29 and 32 to 66 corresponding to the structured regions is 0.81 ± 0.05, indicating that the structure is well ordered in solution (Lytle et al., 2001). Inspection of these data associated with this structure revealed that ^{1}H and ^{15}N chemical shift assignments were incomplete (~88% and ~85% of backbone and side chain, respectively) and that no ^{13}C chemical shift assignments were made. As such, the number of NOE-derived distance restraints used to determine the Doc48S structure was limited to approximately 10 NOEs per residue on average (Lytle et al., 2001). The resultant structure exhibited an unusually low percentage of residues in favored regions in the Ramachandran plot (55.5% and 63.5% for the structure ensemble and average structure, respectively) and large structural derivations between conformers (root mean square derivation (RMSD) for backbone heavy atoms in secondary structure elements of 0.39 ± 0.08 Å). These features thus need to be taken into consideration with comparing the Doc48S structure to type-I dockerin structures in complex with type-I cohesin modules.

To address these discrepancies, in the current study we have revisited the NMR-derived solution structure of Doc48S by incorporating ^{13}C chemical shift and NOESY-based distance restraints with ^{15}N-based NMR information.

3.2. The revisited Doc48S structure

The ^{1}H-^{15}N–HSQC spectrum of Doc48S displays a single set of well-resolved resonances, indicative of a homogeneous and structured protein in solution and displayed similar spectral features to Doc48S spectra previously reported (Lytle et al., 2000), suggesting that solution properties of Doc48S were comparable (Fig. 1). Backbone amide resonances for 67 of the 69 total residues of Doc48S were unambiguously assigned; the exceptions being the backbone amide resonance of Ser1 and the backbone nitrogen of Pro66, which are generally not observed in triple resonance experiments. In total, 98% of the expected ^{1}H, ^{13}C, and ^{15}N chemical backbone and side chain chemical shifts for Doc48S were assigned.

The revisited solution structure of Doc48S was solved using 2264 distance-derived restraints, 36 hydrogen bond restraints, and 120 TALOS+-derived dihedral restraints. An ensemble of the 20-lowest energy Doc48S conformers, which displayed good structural convergence (Fig. 2a), was selected for further analysis. The

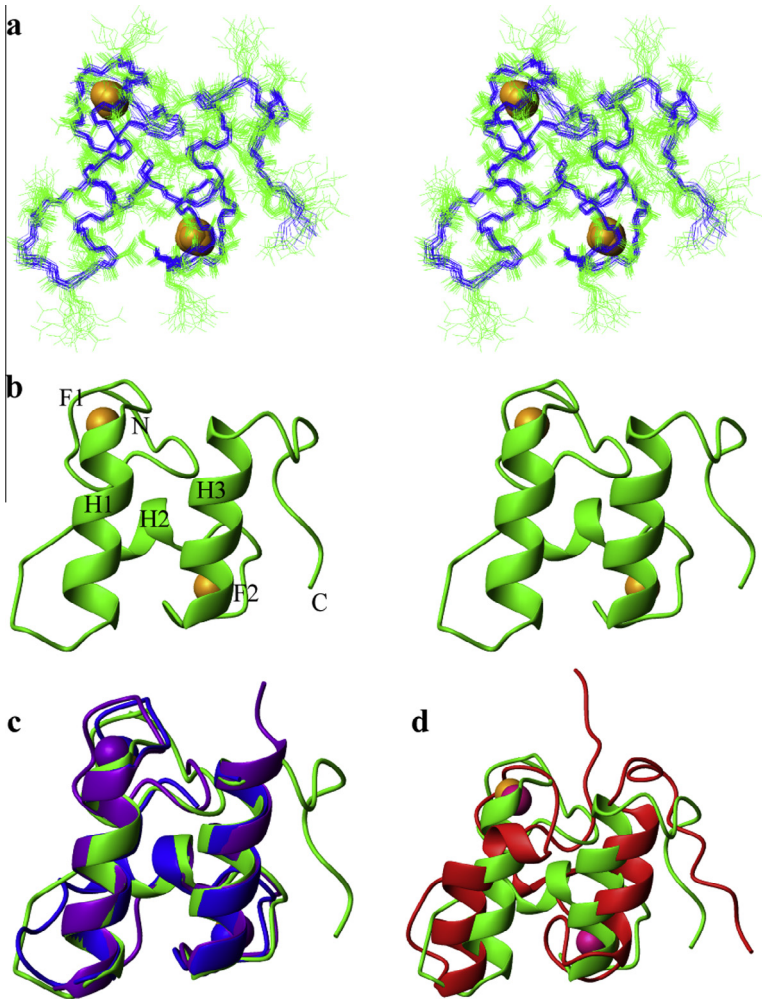

Fig.2. The NMR solution structure of Doc48S. (a) Stereo view of 20 structure ensemble of Doc48S. The backbone and side chain heavy atoms are in blue and green, respectively, and the calcium ions are depicted as yellow spheres. (b) Stereo view of backbone ribbon representation of Doc48S. The calcium irons are shown as yellow balls. The three helices (H1, H2 and H3) and two F-hand-like structures (F1 and F2) are labeled accordingly. (c) Structural comparison of Doc48S structure presented here (green) and crystal structures of cellulosomal dockerin modules (PDB 1OHZ, blue; PDB 2VN5, purple) (d). Structural comparison of the Doc48S structure presented here (green) and that previously determined (PDB 1DAQ, red). The structures were superimposed by the backbone heavy atoms of residues 16–27, 35–38, 48–59.

lowest-energy structure from the ensemble was selected for the ribbon representation (Fig. 2b). A summary of the structural statistics is shown in Table 1.

The 69-residue Doc48S structure comprises three α-helices (H1, residues 16–27; H2, residues 35–38; H3, residues 48–59). Helices H1 and H3, which are antiparallel to one another, and the two calcium-binding loops (Ca1 and Ca2) correspond to the tandem duplicated sequences that form the two F-hand motifs (Pages et al., 1997) (Fig. 2). A short loop region (residues 28–34) and helix H2 connect the F-hand motifs.

3.3. The revised solution structure of Doc48S is similar to the crystal structures of type-I dockerins in the cohesin–dockerin complexes

A comparison of the NMR-derived solution structure of Doc48S presented here to previously reported type-I dockerin structures bound to their cognate cohesin binding partners (Bras et al., 2012; Carvalho et al., 2003; Currie et al., 2012; Pinheiro et al., 2008) indicated that the Doc48S structure is more similar to the cohesin-bound dockerin modules than to the previously determined isolated Doc48S solution structure (Lytle et al., 2001). The backbone RMSD between the revised Doc48S structure and the Xyn10B type-I dockerin module from the type-I cohesin-dockerin complex (Carvalho et al., 2003) was 0.48 Å (Fig. 2c), while the RMSD between our Doc48S structure and the previously reported Doc48S solution structure (Lytle et al., 2001) was 3.53 Å (Fig. 2d) (Supplementary Table S1). The interhelical angle of 165.3° between helices H1 and H3 in the revised Doc48S structure is similar to analogous angles in cohesin-bound dockerin structures, which range from 164.2° to 168.9°. H2 linking the two F-hand motifs in the current Doc48S structure superimposes on the corresponding helices of the *C. thermocellum* Xyn10B (Carvalho et al., 2003) and *Clostridium cellulolyticum* Cel5A (Pinheiro et al., 2008) dockerin modules despite the presence of only one turn as opposed to the two turns observed in the dockerin crystal structures. The two F-hand-like structures in the revised Doc48S are also very similar to those in dockerin crystal structures. The previous solution structure of Doc48S (Lytle et al., 2001) showed large derivations in all these regions (Fig. 2d) with the H1-H3 interhelical angle measuring 152.1° (Table S1). The Doc48S structure presented here thus appears to be better defined than that previously reported, when evaluated by a variety of measures, including Ramachandran statistics, RMSD of backbone and side chain atom coordinates, and WHAT_CHECK *Z*-scores (Table 1); an improvement correlated with a more complete set of ^{1}H, ^{13}C, and ^{15}N chemical shift assignments and an increased number of distance restraints used in the structure calculations (current 32.8 vs. previous 10.3 NOEs per residue).

Table 1
Statistics of revisited and previous NMR structures of Doc48S.

Parameters	Revisited structure (PDB 2MTE)	Previous structure (PDB 1DAQ/1DAV)
NOE restraints		
Intra-residue	515	127
Sequential	276	144
Medium-range	174	209
Long-range	275	248
Ambiguous	1024	
Total	2264	728
Calcium restraints	12	12
Hydrogen bond restraints	36	
Backbone torsion angle restraints	120	79
Phi (Φ) angle restraints	60	
Psi (Ψ) angle restraints	60	
Violations		
Max. NOE violation (Å)	0.045	0.3
Max. torsion angle violation (°)	0.39	0.12
R.M.S.D from mean structure (Å)		
Backbone in structured region[a]	0.28 ± 0.06	0.39 ± 0.08
Heavy atoms in structured region[a]	0.68 ± 0.07	0.82 ± 0.13
Backbone for residues 1–69	0.43 ± 0.06	1.03 ± 0.42
Heavy atoms for residues 1–69	0.91 ± 0.09	1.55 ± 0.38
Ramachandran statistics		
Most favored region (%)	80.7	63.5/55.5[b]
Additionally allowed (%)	19.3	30.2/35.7
Generously allowed (%)	0	6.3/8.3
Disallowed (%)	0	0.0/0.6
WHAT_CHECK *Z*-scores[c]		
1st generation packing quality	−2.251	−4.900/−4.686[d]
2nd generation packing quality	−0.918	−4.392/−3.944
Ramachandran plot appearance	−3.519	−8.170/−8.163
Chi-1/chi-2 rotamer normality	−5.617	−8.298/−8.305
Backbone conformation	0.112	−7.059/−6.206
Inside/outside distribution	1.037	1.104/1.103

[a] The structured regions include residues 16–27, 35–38, 48–59.
[b and d] Before and after "/" stand for ensemble and average of the previous Cel48S NMR structures, respectively.
[c] For the *Z*-scores, more positive value is better.

3.4. The cis-conformation of the Leu65-Pro66 peptide bond

Doc48S has a single proline residue (Pro66) located at the C-terminus, which was observed to form a peptide bond with the preceding leucine (Leu65) in the cis-conformation (Fig. 3a). This conformation was confirmed by both the chemical shifts of Pro66 and the NOE connectivities between Leu65 and Pro66. Specifically, the difference of proline C_β and C_γ chemical shifts ($\Delta C_{\beta\gamma}$) can be used to predict the cis–trans conformation of the peptide bond (Schubert et al., 2002). The C_β and C_γ chemical shifts of Doc48S Pro66 are 35.23 ppm and 24.26 ppm, respectively, and the $\Delta C_{\beta\gamma}$ value of 10.97 ppm was suggestive of the cis-conformation. Further, NOE connectivities between the H_α atoms of Leu65 and Pro66 were observed in ^{13}C NOESY-HSQC dataset, whereas no connectivities between Leu65 H_α and Pro66 H_δ were present (Fig. 3b), indicating a cis-conformation of the Leu65-Pro66 peptide bond (Wuthrich, 1986).

In the previously reported NMR-derived solution structure of Doc48S, chemical shifts for Pro66 were not assigned (Lytle et al., 2001), and the cis-conformation of the Leu65-Pro66 peptide bond was therefore not identified. Pro66 at the C-terminus of Doc48S is highly conserved among several *C. thermocellum* type-I dockerin modules, while the residue preceding the proline possesses a relatively bulky hydrophobic side chain moiety (i.e., Leu, Ile, Phe; Fig. 3f). Of those dockerin modules for which structures exist, only Cthe0435 dockerin (Bras et al., 2012) and the Cel9D dockerin (Currie et al., 2012) contain the conserved proline (Fig. 3d and e), and both are in the trans-conformation with the preceding residue. Of note, the C-terminal proline residues in the *C. thermocellum* Xyn10B and *C. cellulolyticum* Cel5A dockerin modules (Carvalho et al., 2003; Pinheiro et al., 2008) were deleted during the construction of the expression plasmids. The Cel48S dockerin is the first discovered dockerin containing a cis-proline.

Detailed analysis of the structure around Pro66 in Doc48S revealed that Pro66 and Tyr67 are well ordered, which was further evidenced by ^{1}H-^{15}N heteronuclear steady-state NOE experiments (Supplementary Fig. S1). Pro66 makes a hydrophobic, ring-stacking interaction with Tyr5, which forms a conserved clasp in dockerin modules proposed in a previous study (Slutzki et al., 2013) (Fig. 3c). However, a cis-conformation is not an absolute requirement for the formation of this clasp, since the Cel9D dockerin module contains the conserved Tyr-Pro interaction, where the proline is in the trans-conformation (Fig. 3e). Tyr67 is involved in hydrophobic interactions with Ile53 and Ile62, which causes a sharp kink in

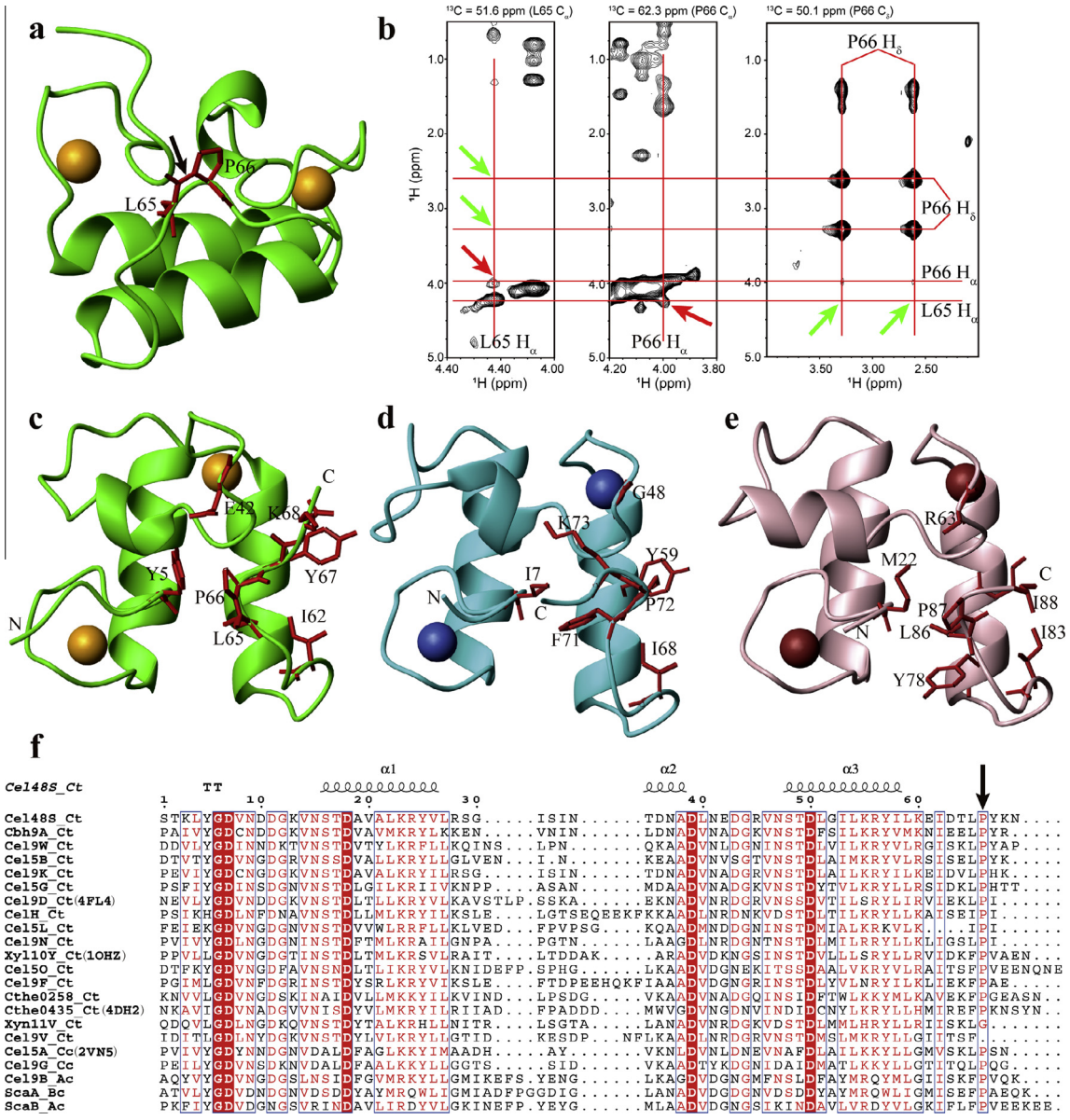

Fig.3. The conformation of the C-terminal X-Proline peptide bond in dockerin modules. (a) The cis-conformation of Leu65 and Pro66 peptide in the NMR structure of Doc48S. The bonds of Leu65 and Pro66 heavy atoms are showed as red sticks and the cis-peptide bond is indicated by an arrow. (b) The ^{13}C-edited planes of the 3D ^{13}C NOESY-HSQC spectrum. The peaks evidenced for the cis-conformation are indicated by red arrows, while the positions of absent peaks for trans-conformation are indicated by green arrows. (c) The structure details around Pro66 in Doc48S. (d) The structure details around P72 of Cthe_0435 dockerin in the crystal structure 4DH2. (e) The structure details around Pro87 of Cel9D dockerin in the crystal structure 4FL4. (f) Sequential alignment of *C. thermocellum* dockerins. The secondary structures of Doc48S are shown at the top. The C-terminal prolines of these dockerins are indicated by an arrow. The species are appended to the enzyme names in abbreviated form with Ct for *Clostridium thermocellum*, Cc for *Clostridium cellulolyticum*, Ac for *Acetivibrio cellulolyticus*, and Bc for *Bacteroides cellulosolvens*. PDB codes are indicated in parentheses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the backbone at Pro66 (the angle formed by nitrogen atoms of Leu65, Pro66 and Tyr67 is 87.8 ± 1.8°, likely favoring the cis-conformation of Pro66). A hydrogen bond between the backbone amide of Tyr67 and carbonyl group of Leu40, and a salt bridge between Glu42 and Lys68 were observed in the Doc48S structure, which appear to contribute to the stability of the cis-conformation. The Cel9D dockerin has only one residue, Ile88, following the conserved proline (Pro87), and the kink angle (105.2°) is slightly larger than that of Doc48S. The HN of Ile88 and CO of Pro87 form a very short hydrogen bond (1.3 Å), and a smaller angle would make the two atoms clash if proline is in a cis-conformation.

From the sequence alignments, type-I dockerin modules from Cel9K, Cbh9A, Cel9W, and Cel5B would be predicted to have cis-prolines similar to that of the Cel48S dockerin. Because of the poor sequence similarity around the prolines, however, more structural studies are needed to identify the conformation of the X-proline peptide bond in other dockerins.

4. Conclusions

The revisited solution structure of the Cel48S type-I dockerin module presented here revealed a compact conformation unlike the previously determined Cel48S dockerin structure, which displayed a more open conformation. These structural differences can be accounted for by a more comprehensive set of ^{15}N- and ^{13}C-edited NOESY-derived distance restraints used in the current structure determination. It thus appears that the Cel48S type-I

dockerin module, and more generally all type-I dockerins, adopt a cohesin-primed conformation in solution. These results refute previous suggestions that the dockerin module undergoes marked conformational alterations upon binding to its cohesin partner.

Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program, Grant No. 2011CB707404) and the National Natural Science Foundation of China (Grant Nos. 31270784 and 31100568). E.A.B. is the incumbent of The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry at the Weizmann Institute of Science.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jsb.2014.09.006.

References

Bayer, E.A., Belaich, J.P., Shoham, Y., Lamed, R., 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. Annu. Rev. Microbiol. 58, 521–554.

Bayer, E.A., Lamed, R., White, B.A., Flint, H.J., 2008. From cellulosomes to cellulosomics. Chem. Rec. 8, 364–377.

Bras, J.L., Alves, V.D., Carvalho, A.L., Najmudin, S., Prates, J.A., Ferreira, L.M., Bolam, D.N., Romao, M.J., Gilbert, H.J., Fontes, C.M., 2012. Novel clostridium thermocellum type I cohesin–dockerin complexes reveal a single binding mode. J. Biol. Chem. 287, 44394–44405.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., Warren, G.L., 1998. Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D Biol. Crystallogr. 54, 905–921.

Carvalho, A.L., Dias, F.M., Prates, J.A., Nagy, T., Gilbert, H.J., Davies, G.J., Ferreira, L.M., Romao, M.J., Fontes, C.M., 2003. Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex. Proc. Natl. Acad. Sci. U.S.A. 100, 13809–13814.

Carvalho, A.L., Dias, F.M., Nagy, T., Prates, J.A., Proctor, M.R., Smith, N., Bayer, E.A., Davies, G.J., Ferreira, L.M., Romao, M.J., Fontes, C.M., Gilbert, H.J., 2007. Evidence for a dual binding mode of dockerin modules to cohesins. Proc. Natl. Acad. Sci. U.S.A. 104, 3089–3094.

Currie, M.A., Adams, J.J., Faucher, F., Bayer, E.A., Jia, Z., Smith, S.P., 2012. Scaffolding conformation and dynamics revealed by a ternary complex from the *Clostridium thermocellum* cellulosome. J. Biol. Chem. 287, 26953–26961.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A., 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277–293.

Demain, A.L., Newcomb, M., Wu, J.H., 2005. Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. 69, 124–154.

Duggan, B.M., Legge, G.B., Dyson, H.J., Wright, P.E., 2001. SANE (structure assisted NOE evaluation): an automated model-based approach for NOE assignment. J. Biomol. NMR 19, 321–329.

Fierobe, H.P., Bayer, E.A., Tardif, C., Czjzek, M., Mechaly, A., Belaich, A., Lamed, R., Shoham, Y., Belaich, J.P., 2002. Degradation of cellulose substrates by cellulosome chimeras. Substrate targeting versus proximity of enzyme components. J. Biol. Chem. 277, 49621–49630.

Herrmann, T., Guntert, P., Wuthrich, K., 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. 319, 209–227.

Hooft, R.W., Vriend, G., Sander, C., Abola, E.E., 1996. Errors in protein structures. Nature 381, 272.

Johnson, B.A., Blevins, R.A., 1994. NMR view: a computer program for the visualization and analysis of NMR data. J. Biomol. NMR 4, 603–614.

Koradi, R., Billeter, M., Wuthrich, K., 1996. MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. 14, 51–55.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486.

Lytle, B.L., Volkman, B.F., Westler, W.M., Heckman, M.P., Wu, J.H., 2001. Solution structure of a type I dockerin domain, a novel prokaryotic, extracellular calcium-binding domain. J. Mol. Biol. 307, 745–753.

Lytle, B.L., Volkman, B.F., Westler, W.M., Wu, J.H., 2000. Secondary structure and calcium-induced folding of the *Clostridium thermocellum* dockerin domain determined by NMR spectroscopy. Arch. Biochem. Biophys. 379, 237–244.

Markley, J.L., Bax, A., Arata, Y., Hilbers, C.W., Kaptein, R., Sykes, B.D., Wright, P.E., Wuthrich, K., 1998. Recommendations for the presentation of NMR structures of proteins and nucleic acids. J. Biomol. NMR 12, 1–23.

Mossessova, E., Lima, C.D., 2000. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. Mol. Cell 5, 865–876.

Pages, S., Belaich, A., Belaich, J.P., Morag, E., Lamed, R., Shoham, Y., Bayer, E.A., 1997. Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: prediction of specificity determinants of the dockerin domain. Proteins 29, 517–527.

Pinheiro, B.A., Proctor, M.R., Martinez-Fleites, C., Prates, J.A., Money, V.A., Davies, G.J., Bayer, E.A., Fontesm, C.M., Fierobe, H.P., Gilbert, H.J., 2008. The *Clostridium cellulolyticum* dockerin displays a dual binding mode for its cohesin partner. J. Biol. Chem. 283, 18422–18430.

Schubert, M., Labudde, D., Oschkinat, H., Schmieder, P., 2002. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. J. Biomol. NMR 24, 149–154.

Shen, Y., Delaglio, F., Cornilescu, G., Bax, A., 2009. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J. Biomol. NMR 44, 213–223.

Slutzki, M., Jobby, M.K., Chitayat, S., Karpol, A., Dassa, B., Barak, Y., Lamed, R., Smith, S.P., Bayer, E.A., 2013. Intramolecular clasp of the cellulosomal *Ruminococcus flavefaciens* ScaA dockerin module confers structural stability. FEBS Open Biol. 3, 398–405.

Strynadka, N.C.J., James, M.N.G., 1989. Crystal-structures of the helix-loop-helix calcium-binding proteins. Annu. Rev. Biochem. 58, 951–998.

Wuthrich, K., 1986. NMR of Proteins and Nucleic Acids. John Wiley & Sons Inc., New York (pp. 123–125).