

Homology swapping of intrinsic secondary structural elements between cellulosomal types I and II cohesins and their effect on dockerin binding*

Ilit Noach¹, Yoav Barak^{1,2}, Felix Frolow^{3,4}, Raphael Lamed³, and Edward A. Bayer^{1,‡}

¹Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100 Israel; ²Chemical and Biophysical NanoSciences, Chemical Research Support, The Weizmann Institute of Science, Rehovot 76100, Israel; ³Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel; ⁴The Daniella Rich Institute for Structural Biology, Tel Aviv University, Ramat Aviv 69978, Israel

Abstract: The high-affinity cohesin–dockerin interaction dictates the suprastructural assembly of the multienzyme cellulosome complex. The interaction between these two complementary families of protein modules was found to be species-specific and type-dependent. The structure of the type II cohesin module possesses additional intrinsic secondary structural elements absent in the type I cohesin, i.e., an α -helix and two singular “ β -flaps”. The role of these extra secondary structures in dockerin recognition was studied in this work using gene swapping, in which corresponding homologous stretches of types I and II cohesins were interchanged. The specificity of binding of the resultant chimaeric cohesins was determined by enzyme-linked affinity assay. Several chimaeric cohesins retained dockerin recognition properties. Hence, these cohesins may undergo manipulations (insertion/deletion of peptide segments) without altering their affinity toward their counterpart dockerin, although type-dependent binding specificity cannot be converted by swapping of the additional secondary structural elements between the two cohesin types. The results further emphasize the strong affinity and plasticity between the cohesin and dockerin pair and are consistent with the known findings on specificity of the types I and II interactions. These studies provide insight into the structural and functional resilience of the cohesins and thus have direct bearing on their potential use in biotechnological applications.

Keywords: chimaera; *Clostridium thermocellum*; enzyme-linked immunoabsorbent assay; β -flaps; α -helix; multienzyme complex; protein–protein interaction.

INTRODUCTION

The cellulosome is a multienzyme complex, an exocellular macromolecular machine, designed for efficient degradation of cellulose and associated plant cell wall polysaccharides [1]. The first cellulosome

*Paper based on a presentation at the 13th International Biotechnology Symposium (IBS 2008): “Biotechnology for the Sustainability of Human Society”, 12–17 October 2008, Dalian, China. Other presentations are published in this issue, pp. 1–347.

‡Corresponding author: E-mail: ed.bayer@weizmann.ac.il

was discovered in the anaerobic thermophilic cellulolytic bacterium, *Clostridium thermocellum* [2,3]. The cellulosomal enzyme subunits were found to be united into the complex by means of a unique class of nonenzymatic, multimodular polypeptide subunit, termed “scaffoldin” [4]. The scaffoldin usually contains a module termed cellulose-binding module (CBM) which is responsible for the binding of the complex to the substrate. The scaffoldin also contains multiple copies of a definitive type of module, termed “cohesin”. The cellulosomal enzyme subunits, on the other hand, contain a complementary type of module, termed “dockerin”. The tenacious interaction between the cohesin and dockerin modules provides the definitive molecular mechanism that integrates the enzyme subunits into the cellulosome complex [5–8]. The various modules of the cellulosome complex and the strong cohesin–dockerin interaction can also be utilized for biotechnology purposes, such as affinity tags and designer cellulosomes, i.e., native or artificially constructed cellulosomes that may serve to better degrade cellulosic materials for the production of bio-ethanol.

Historically, the cohesins have been divided into different types according to phylogenetic sequence analysis [6,7,9] (Fig. 1). For many species of scaffoldin, the cohesins have been classified as type I on the basis of sequence homology. The scaffoldin of *C. thermocellum* also contains a special type of dockerin module at its carboxy terminus. This dockerin failed to bind to the cohesins from the same scaffoldin subunit, but instead interacted with a different type of cohesin, termed type II [6]. These cohesins are somewhat different than those of type I, having additional segments within the polypep-

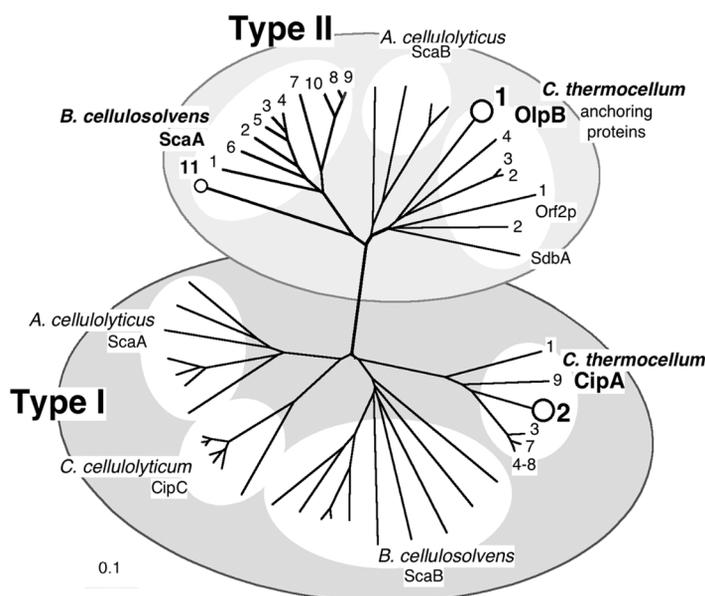


Fig. 1 Phylogenetic relationship of types I and II cohesins. Type II cohesins include the surface anchoring proteins of *C. thermocellum* (OlpB, SdbA, and Orf2p) those from the *Bacteroides cellululosolvens* ScaA scaffoldin and cohesins from *Acetivibrio cellulolyticus* ScaB scaffoldin. The cohesins are enumerated according to their position from the N-terminus. Denoted in the figure are the *C. thermocellum* cohesins participate in this work, the type I cohesin 2 from the primary CipA scaffoldin and the first type II cohesin from OlpB anchoring scaffoldin (large circles). Also marked is the 11th cohesin of ScaA from *B. cellululosolvens* whose crystal structure was previously determined by our lab and was used as a prototype of the type II cohesin model in this work (small circle). Scale bar indicates the percentage (0.1) of amino acid substitutions. The GenBank or Swiss-Prot accession numbers for scaffoldin sequences used to construct this tree are as follows: the primary scaffoldin, CipA (Q06851), and anchoring scaffoldins, SdbA (U49980), OlpB (Q06852), and Orf2p (Q06853) from *C. thermocellum*; ScaA and ScaB (AF224509) from *B. cellululosolvens*; ScaA (AF155197), ScaB (AY221112), and ScaC (AY221113) from *A. cellulolyticus* and CipC (U40345) from *C. cellulolyticum*.

tide and clear diversity in the latter half of the sequence. The type II cohesins were discovered as component parts of a group of noncatalytic cell-surface “anchoring” proteins on *C. thermocellum* [6,9–11]. Each of the anchoring proteins also contains an S-layer homology (SLH) module that implants the protein onto the cell envelope. The type II cohesins selectively bind to type II dockerins, and the cellulosome is thereby incorporated into the cell surface of *C. thermocellum*. The interaction between the cohesin and the dockerin is species-specific and type-dependent. This specificity ultimately determines the potential overall architecture of the cellulosome complex.

The 3D structures of three type I cohesins were determined by X-ray crystallography: Cohesins 2 and 7 from the CipA scaffoldin of *C. thermocellum* (PDB codes: 1ANU and 1AOH, respectively) [12,13] and cohesin 1 from the CipC scaffoldin of *Clostridium cellulolyticum* (PDB code: 1G1K) [14]. The type I cohesin modules comprise an elongated, somewhat conical molecule, with a jellyroll topology that folds into a nine-stranded β -sandwich (Fig. 2A). The two sheets of the sandwich are composed of strands 8-3-6-5 and 9-1-2-7-4, respectively, where β -strand 9 (C-terminus) and β -strand 1 (N-terminus) run parallel, and the remaining strands are antiparallel. Prior to this work, only two crystal structures of type II cohesins had been determined: the 11th cohesin of ScaA from *Bacteroides cellulosolvens* (PDB code: 1TYJ) [15] and the first ScaB cohesin of *Acetivibrio cellulolyticus* (PDB codes: 1QZN) [16]. Subsequently, the 3D structure of the lone SdbA cohesin of *C. thermocellum* was determined (PDB code: 2BM3) [17]. Despite the similarity in overall structure and topology of the type I and II cohesins, several distinctive differences were noted. In this context, a short α -helical element occurs between β -strands 6 and 7 in the type II cohesins, thus flanking the “crown” of the molecule. More intriguingly, β -strands 4 and 8 are each disrupted by a “ β -flap”, which interrupts and digresses from the original course of the β -strand, but then returns to complete the respective strand (Fig. 2B). In the present work, we studied the possible role of these extra secondary structural elements upon dockerin binding by employing a recombinant genetic approach, recently implemented in our laboratory, which involves progressive swapping of corresponding homologous stretches [18] of the cohesin modules from types I and II. This work provides valuable biochemical evidence, which supports the observed differences between the two types of cohesin in complex with their respective dockerin, as revealed in the crystal structures determined subsequently.

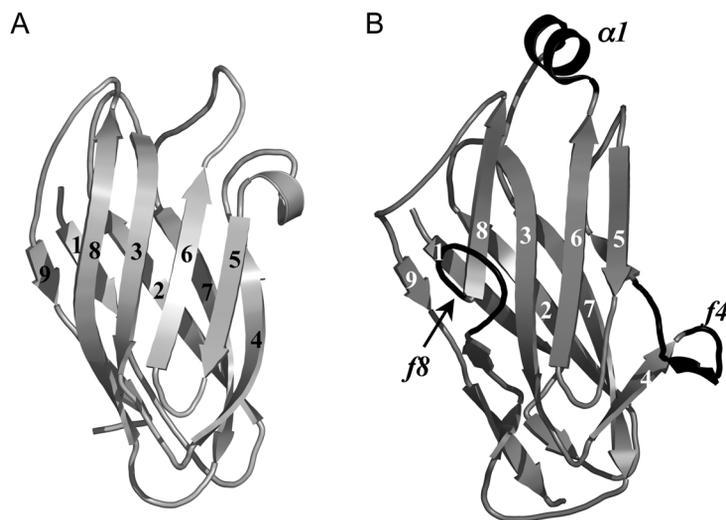


Fig. 2 Schematic ribbon diagram of representative types I and II cohesin 3D structures. (A) The type I cohesin 2 structure from *C. thermocellum* (Ct-CohA2) (PDB code: 1ANU). (B) The type II cohesin 11 structure from *B. cellulosolvens* (Bc-CohA11) (PDB code: 1TYJ). The additional secondary structures, the two “ β -flaps” (f4 and f8) and the α -helix are black. The β -strands are numbered 1–9.

RESULTS

To study the possible effect of the additional secondary structural elements revealed in the type II cohesins on dockerin recognition we combined homology swapping with an enzyme-linked immunosorbent assay (ELISA)-based assay. Chimaeric cohesins were constructed whereby corresponding homologous stretches of the types I and II cohesins were interchanged and fused to the C-terminus of a family-3 CBM from *C. thermocellum* (CBM-Cohs). In this context, the additional secondary structural elements (the α -helix and two β -flaps) were deleted from the type II cohesin and inserted into the type I cohesin, in a progressive manner. The resultant chimaeras were tested for specificity of binding by employing an ELISA-based assay [19] using dockerins of both rival types fused to a xylanase (Xyn-Doc). The corresponding wild-type (WT) cohesins were used as controls.

Construction of WT and chimaeric proteins

A prototype system, comprising the 11th type II cohesin module of the cellulosomal scaffoldin polypeptide ScaA from *B. cellulosolvens* (*Bc-CohA11*), whose 3D structure was recently determined by our lab (PDB code: 1TYJ) [15] and a type I cohesin from scaffoldin B (ScaB) of the same cellulosome system were initially selected together with their counterpart dockerins (i.e., the Cel48A enzyme-borne type I dockerin and the type II ScaA dockerin, respectively). Unfortunately, we encountered unexpected obstacles with the resultant type II cohesin–dockerin interaction, and we decided to examine an alternative cohesin–dockerin system from the well-studied cellulosome of *C. thermocellum*. Due to the lack of other type II cohesin structures at that time, sequence-based alignment between *Bc-CohA11* and the entire complement of the type II *C. thermocellum* cohesins was applied using ClustalW [20]. The analogous *C. thermocellum* cohesins we selected for this work were: (i) the first type II cohesin from the OlpB anchoring scaffoldin (*Ct-OlpB1*), which retained the highest sequence homology with the additional secondary element segments of *Bc-CohA11* and (ii) the type I cohesin from CipA (*Ct-CohA2*), whose 3D structure was determined previously (PDB code: 1ANU) [12]. The chosen counterpart dockerins were the type II CipA dockerin and the type I CelS (Cel48A) dockerin.

Structure-based alignment between the 3D *Bc-CohA11* structure (PDB code: 1TYJ) and the *C. thermocellum* cohesins—the type I *Ct-CohA2* structure (PDB code: 1ANU) and the type II *Ct-OlpB1*—was achieved using PROMALS3D [21], in order to determine the preferred positions of the homologous stretches which correspond to the additional secondary structural elements (Fig. 3).

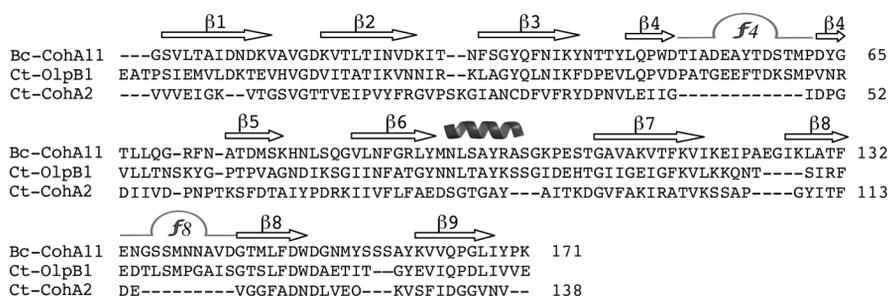


Fig. 3 Structure-based alignment of type II vs. I cohesin sequences. Structure-based alignment between the sequences of the type II *B. cellulosolvens* CohA11 (*Bc-CohA11*; PDB code 1TYJ), the type I *C. thermocellum* CohA2 (*Ct-CohA2*; PDB code 1ANU), and the type II *OlpB1* cohesin from *C. thermocellum* (*Ct-OlpB1*) using the PROMALS3D server [21]. The nine β -strands comprising the 1TYJ and 1ANU structures are designated by enumerated arrows. The additional β -flaps (*f4* and *f8*) and the α -helix are represented symbolically. The residues are enumerated according to the respective PDB code.

In the first round of experiments, three constructs of the type II *Ct-OlpB1*, which lack either the α -helix or one of the two β -flaps, were produced (Fig. 4). Likewise, three constructs of the type I *Ct-CohA2* were initially produced, each of which contained one of the type II secondary structural el-

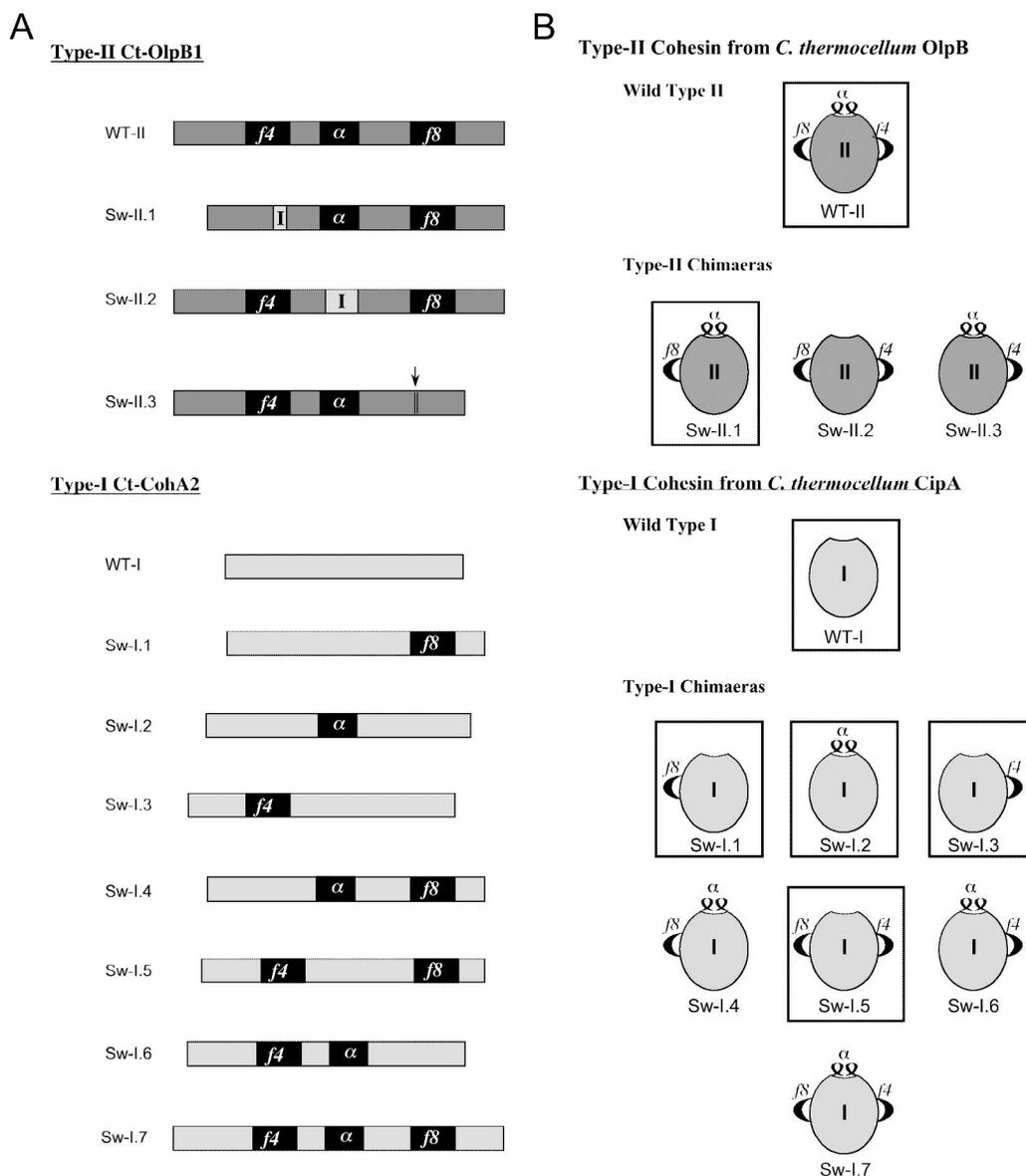


Fig. 4 Schematic representation of types II and I chimaeric cohesins. The figure shows symbolic views of the approximate positions of the designated secondary structures in the various WT and swaps (Sw) in the DNA (A) and protein (B), respectively. Each of the three type II chimaeric cohesins from *C. thermocellum* *OlpB* lacks one of the additional secondary elements: i.e., β -flap 4 (Sw-II.1), the α -helix (Sw-II.2) and β -flap 8 (Sw-II.3); the resultant spaces were filled in with corresponding DNA fragments derived from the type I cohesin gene (designated I). Single, double, and triple Sw of the seven type I chimaeric cohesins from *C. thermocellum* *CipA* resulted in the addition of one of the secondary elements (Sw-I.1, Sw-I.2, and Sw-I.3), a combination of two secondary elements (Sw-I.4, Sw-I.5, Sw-I.6) and all three secondary elements (Sw-I.7), respectively. The cohesins that successfully interact with their counterpart dockerin in the ELISA assay are framed in boxes.

ements. Upon obtaining the results of subsequent binding studies, a second round was performed in which four additional type I chimaeras were produced, each with an insertion of complementary multiple secondary structural elements (double and triple Sw). Three of them contained insertion of two secondary structural elements and the fourth included all three additional secondary structural elements (Fig. 4).

Expression and purification of fusion proteins

Desired cohesins (WT and chimaeras) were fused to the scaffoldin-based CBM3a of *C. thermocellum* [22], and the resultant hybrid proteins were affinity-purified on cellulose resin [18,19]. In each case, sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) of the eluted CBM-Coh fusion proteins showed a purified product consistent with the calculated molecular weight (Fig. 5). This purification procedure was performed for several different proteins in simultaneous fashion by batch adsorption to and desorption from cellulose matrices in separate samples. This approach resulted in ample protein yields (14–21 mg/l of cell culture).

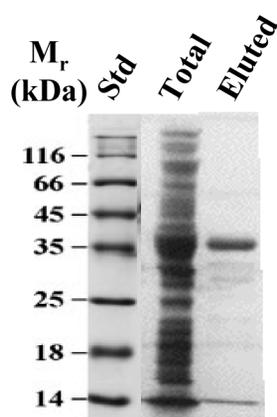


Fig. 5 Representative batchwise purification of an expressed CBM-Coh fusion protein. The type I chimaera of cohesin 2, derived from the *C. thermocellum* CipA scaffoldin that contains an insertion of an α -helix (Sw-I.2; representative of CBM-Coh fusion proteins) was expressed in the *E. coli* host-cell system. The cells were sonicated, the soluble proteins were applied batchwise to cellulose resin, and the adsorbed proteins were eluted with 1 % triethylamine. Total protein (the cell sonicate prior to application to the cellulose resin) and the eluted fractions were subjected to SDS-PAGE on 12 % gels. The eluted CBM-Coh proteins were remarkably similar in their purity, and their respective mobility on the gels was consistent with their calculated molecular size. Std, molecular weight standards.

Cohesin–dockerin binding specificity

The resultant chimaeras were tested for specificity of dockerin binding by employing an ELISA-based affinity assay developed by Barak et al. [19]. The *C. thermocellum* system provided good results for the WT controls in the ELISA-based assay: the WT type II Ct-OlpB1 cohesin interacted with its counterpart dockerin, the type II CipA dockerin (Doc-II), but not with the rival type I Cel48S dockerin (Doc-I) and vice versa, thus verifying the lack of cross-reactivity between these two types of cohesin.

For the type II cohesin–dockerin interaction, the results of the binding assay showed that only the OlpB1 chimaera Sw-II.1, which lacks β -flap 4, remained active. The binding to the counterpart dockerin for this chimaera was preserved with essentially the same affinity displayed by the WT cohesin. The other two chimaeras (deficient in either the α -helix or the β -flap 8) proved diminished in their re-

spective ability to bind the type II dockerin (Fig. 6A). In addition, in control experiments, all three chimaeras failed to interact with the rival type I dockerin. In contrast, strong binding to the counterpart type I dockerin was preserved in all of the type I chimaeras that contained one of the additional secondary structural elements (Sw-I.1–Sw-I.3; see Fig. 4), with essentially the same affinity as the WT cohesin (Fig. 6B).

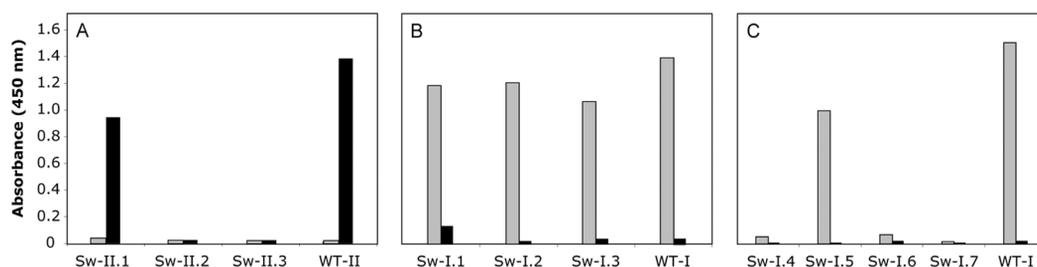


Fig. 6 Binding analysis of types I and II *C. thermocellum* chimaeras. The chimaeric cohesins were fused to the CBM3 module (CBM-Coh). The resultant fusion chimaeras were purified and analyzed on ELISA plates for binding to Xyn-Doc fusion proteins of both types. Similarly fused, WT cohesins were used as references. (A) Bar graph representing the three type II chimaeric cohesin interaction with the counterpart CipA dockerin (in black) and with the type I Cel48A dockerin (in gray). (B) Interaction of the type I chimaeric cohesins that possess only one of the additional secondary elements (single swap). Gray bars represent affinity to the type I Cel48A dockerin while black bars stand for the affinity to the type II CipA dockerin. (C) Interaction of type I chimaeric cohesins that undergo combinations of double and triple Sw. The names of the chimaeras correspond to those designated in Fig. 4.

In view of these results, we proceeded to extend the decoration of the type I chimaeras and produced four more chimaeras with insertions of multiple secondary structural elements (double and triple Sw). Three of them (Sw-I.4–Sw-I.6) contained combinations of two secondary structures and the fourth included all three additional secondary structural elements (Sw-I.7; see Fig. 4). From these four chimaeras, only the chimaera which incorporated the two β -flaps (Sw-I.5) recognized the counterpart dockerin (Fig. 6C). Similar to the results obtained for the type II chimaeras, all of the type I chimaeras failed to interact with the rival (type II) dockerin control. Nonfunctional chimaeras were shown to be in a folded state by circular dichroism analysis and Fourier transform infrared (FT-IR) spectroscopy (data not shown).

DISCUSSION

In efforts to determine the molecular elements that dictate binding and species specificity, crystallization studies were performed to solve the 3D structure of the cohesin–dockerin complex. The first structure of a cohesin–dockerin complex that was determined involved the type I cohesin 2 module and xylanase 10B dockerin module from *C. thermocellum* [23]. The cohesin module was found to undergo very little conformational change upon binding the dockerin module, with the latter interacting with the 8-3-6-5 face of the cohesin module via an extensive hydrogen-bonding network and supporting hydrophobic interactions. These contacts are made primarily through only one of the two helices present in the dockerin structure, which was at the time somewhat surprising given the internal two-fold symmetry of the dockerin sequence and corresponding structure. Follow-up mutational and X-ray structural studies by Carvalho, Pinheiro, and co-workers [24,25] elegantly illustrated that the two-fold symmetry of the type I dockerin sequences and structures from *C. thermocellum* and *C. cellulolyticum* provides a second mode of interaction, whereby the dockerin module is positioned 180° on the 8-3-6-5 face of its cohesin partner.

The homology-swapping approach described in this work was achieved during the period following publication of the 3D structure of the type I cohesin-dockerin heterodimer but before elucidation of the type II structure. In fact, at the time the initial experiments were commenced, only the in-house type II cohesin structures alone (from *A. cellulolyticus* and *B. cellulosolvans* [15,16]) were known. For this reason, the *C. thermocellum* OlpB cohesin and not the SdbA cohesin, the structure of which was determined subsequently [17,26], was selected for this work as described above.

Crystallography provides structural evidence alone, which is unidimensional in the absence of supportive biochemical evidence. The biochemical findings revealed from homology swapping studies performed in this work correlate well with the crystal structures of the *C. thermocellum* type I cohesin-dockerin complexes [23,24]. The structures revealed that several hydrophobic residues from β -strands 3, 5, and 6 of the cohesin module, participate in complex formation. In addition, the protein modules also interact via a series of hydrogen-bonding interactions, mainly through residues located on strand 3 and also by strand 5, the C-terminal part of strand 6, and the loops connecting strands 4–5, 5–6, and 8–9 of the cohesin module. Therefore, altering strands 4 or 8 (Sw-I.3 and Sw-I.1, respectively), the loop between strands 6 and 7 (Sw-I.2), and even the addition of both β -flaps to type I cohesin (Sw-I.5), failed to disrupt dockerin recognition, because the main interacting areas were essentially preserved. Moreover, the preservation of dockerin binding of the type I cohesin chimaera that incorporated both β -flaps may be due to the flexible nature and possible structural reciprocity of these inserted flaps. Conversely, the type I cohesin chimaeras with the addition of the α -helix together with one of the two β -flaps (Sw-I.4 and Sw-I.6), as well as the chimaera with all three additional secondary elements (Sw-I.7), are all severely affected in their dockerin recognition properties, presumably due to greater disruption of the interacting regions. The preservation of dockerin recognition among many of the type I chimaeras can be further explained by the dual-binding mode that was demonstrated in the type I cohesin-dockerin complex structures [24,25]. This mode of interaction provides plasticity in cohesin recognition and may overcome steric effects of the insertions in the functional type I cohesin chimaeras.

Recently, the crystal structure of a type II complex from *C. thermocellum* was determined, consisting of the SdbA cohesin module and the CipA scaffoldin dockerin module together with its adjacent X-module. The type II complex displayed striking differences from the type I complex, which, for the most part, appear to arise from the lack of obvious sequential symmetry of the dockerin module [26]. The dockerin module is rotated $\sim 20^\circ$ on the cohesin surface compared to its type I counterpart [23] and forms contacts with the 8-3-6-5 face of the type II cohesin module across the entire length of both dockerin helices. The contact face also includes β -flap 8 and the loop region leading into the α -helix between β -strands 6–7. This capacious interface has a similar extent of hydrogen-bonding network to that of the type I interaction but comprises a much more prominent hydrophobic component, which apparently contributes to the higher affinity observed for the type II interaction and the specificity between the two types.

Our results thus correlate well with these structural findings in which the type II chimaeras that lack the α -helix (Sw-II.2) or β -flap 8 (Sw-II.3) fail to bind the counterpart dockerin, and only the chimaera, which lacks β -flap 4, retained binding affinity to the dockerin. The apparent reason for this is that flap 4 is not located in the binding region and therefore does not interfere in the cohesin-dockerin interaction. Both types of chimaeras failed to recognize the rival dockerin, and it seems that the specificity of the types I and II cohesin-dockerin interactions cannot be converted by displacement of the additional secondary elements that appear in the type II cohesins. Nevertheless, this work emphasizes the strength and the plasticity of the cohesin-dockerin interaction by showing that even addition of large insertions in the *C. thermocellum* type I cohesin resulted in only a minor effect on its affinity to the counterpart dockerin. This study complements recent X-ray studies providing basic insight into the structural plasticity and functional resilience of the cohesins and thus has direct bearing on their potential as versatile binding elements in biotechnological applications.

METHODS

Construction of recombinant proteins

The CBM-Coh (pET28*acbm3*) and Xyn-Doc (pET9*dxyn*) cassettes for expression of fusion proteins were designed by Barak et al. [19]. Gene swapping was performed using overlap-extension polymerase chain reaction (PCR). The genes coding for the resultant chimaeric cohesins were prepared using appropriate primers as listed in Table 1. For each chimaera, two successive PCR reactions were required as described in Table 2. The PCR fragments of the various cohesin chimaeras and the WT cohesins were cleaved and ligated into a linearized pET28*acbm3* plasmid at sites between the *Bam*HI and *Xho*I of this vector. The PCR fragments of the dockerin modules were cleaved and ligated into a linearized pET9*dxyn* plasmid between its *Kpn*I and *Bam*HI sites, and the resulting expressed product constituted a His-tagged xylanase T-6 fusion protein bearing a dockerin at the C-terminus.

Table 1 Primers used in this study for homology swapping. The names of clones correspond to those designated in Fig. 4. Bold segments indicate restriction sites. F, forward primer; R, reverse primer.

Clone	Primer	Sequence
OlpB1, WT-II	F_II	GCGAGGATCCGACGGTGTGGTAGTAGAA
	R_II	GGTCTCGAGTTATGTTGCATTGCCAACGTT
Sw-II.1	F_II-f4	GCCGGTAGGGATAGTAAATAGGGTTTTGCTGAC
	R_II-f4	TATTTACTATCCCTACCGGCTGAAAACTTCAG
Sw-II.2	F_II-H	CAGCGGAACAGGAGCGTATGCAATAGAACATACAGGAATAATAGG
	R_II-H	CATACGCTCCTGTTCCGCTGTCTTCTCCCGTAGCAAATAATAA
Sw-II.3	F_II-f8	CAAATACGTCTATTAGGTTTGAAGGAACAAGTT
	R_II-f8	CCAGTCAAACAACTTGTTTCCTCAAACCTA
CohA2, WT-I	F_I	GCTGGGATCCGAAGCAACTCCAAGTATTG
	R_I	GGTCCCTCGAGTTATGCCTCTACAACATAAGATC
Sw-I.1	F_I+f8	TATCGATGCCCCGGGGCAATATCGGGAGTAGGTGGATTTGCAGATAA
	R_I+f8	TATTGCCCCGGGCATCGATAATGTATCGTCGAAAGTAATATAGCCCC
Sw-I.2	F_I+H	TAACAGCGTACAAATCCAGCGGAATAGACACTAAAGACGGAGTATTGTC
	R_I+H	GCTGGATTTGTACGCTGTTAAATTGTTATACGCAAACAGGAATACTATTA
Sw-I.3	F_I+f4	CAGGAGAGGAATTTACTGATAAGTCCATGCCGGATCCCGGAGACATAATAGT
	R_I+f4	CTTATCAGTAAATTCCTCTCCTGTTGCAGGGTCTATAATTTCCAATACATTCG
Doc-II	F	CAGCGGTACCAATAATGATGTGGGTAGG
	R	TCGCGGATCCTTACTGTGCGTCGTAATCAC
Doc-I	F	ATTTGGTACCTGGTACTCCTTCT
	R	CGGGATCCTTAGTTCCTGTACGG

Table 2 Components used for overlap-extension PCR to obtain the different chimaeras. The first PCR products (1A, 1B) were purified, diluted ten-fold and used for the final PCR reaction (PCR 2). For each reaction, the two designated primers (Pr. 1 and Pr. 2) and the appropriate template are listed. The PCR fragments were subsequently cloned into pET28acbm3. *olpB1* and *cohA2* refer to the WT, types II and I cohesin genes of *C. thermocellum*, respectively. The names of the chimaeras correspond to those designated in Fig. 4.

Clone	PCR 1A			PCR 1B			PCR 2		
	Pr. 1	Pr. 2	Template	Pr. 1	Pr. 2	Template	Pr. 1	Pr. 2	Template
Sw-II.1	F_II	R_II-f4	<i>olpB1</i>	F_II-f4	R_II	<i>olpB1</i>	F_II	R_II	1A+1B
Sw-II.2	F_II	R_II-H	<i>olpB1</i>	F_II-H	R_II	<i>olpB1</i>	F_II	R_II	1A+1B
Sw-II.3	F_II	R_II-f8	<i>olpB1</i>	F_II-f8	R_II	<i>olpB1</i>	F_II	R_II	1A+1B
Sw-I.1	F_I	R_I+f8	<i>cohA2</i>	F_I+f8	R_I	<i>cohA2</i>	F_I	R_I	1A+1B
Sw-I.2	F_I	R_I+H	<i>cohA2</i>	F_I+H	R_I	<i>cohA2</i>	F_I	R_I	1A+1B
Sw-I.3	F_I	R_I+f4	<i>cohA2</i>	F_I+f4	R_I	<i>cohA2</i>	F_I	R_I	1A+1B
Sw-I.4	F_I	R_I+f8	<i>cohA2+H</i>	F_I+f8	R_I	<i>cohA2+H</i>	F_I	R_I	1A+1B
Sw-I.5	F_I	R_I+f4	<i>cohA2+f8</i>	F_I+f4	R_I	<i>cohA2+f8</i>	F_I	R_I	1A+1B
Sw-I.6	F_I	R_I+f4	<i>cohA2+H</i>	F_I+f4	R_I	<i>cohA2+H</i>	F_I	R_I	1A+1B
Sw-I.7	F_I	R_I+f4	<i>cohA2+H+f8</i>	F_I+f4	R_I	<i>cohA2+H+f8</i>	F_I	R_I	1A+1B

Expression and purification of fusion proteins

The expression and purification of CBM-cohesin fusion proteins were accomplished as described previously [19] with minor modifications. Briefly, the host cells were grown in 100 ml LB (Luria-Bertani) medium, supplemented with 50 µg/ml kanamycin. Induction was initiated with 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside) following incubation of 3 h at 37 °C. Two milliliters of amorphous cellulose were then added to the supernatant and incubated for 1–2 h on a rotator at 4 °C. Centrifugation and washing steps were done accordingly using 12 ml *tert*-butyldimethylsilyl (TBS) instead of 45 ml TBS as previously reported [19]. The protein was eluted from the pellet with 2 vol (1 ml each) of 1 % (v/v) triethylamine, and the eluted fractions were neutralized with 100 µl of 1 M MES at pH 5.5. Purity was assessed by SDS-PAGE, and proteins were stored in 50 % (v/v) glycerol at –20 °C. The His-tagged Xyn-Doc proteins were expressed in *Escherichia coli* and purified by metal-ion affinity chromatography according to Barak et al. [19].

Protein concentration

The concentration of the purified proteins was estimated by absorbance (280 nm) based on the known amino acid composition of the desired protein using the ProtParam tool (<www.expasy.org/tools/protparam.html>) on the EXPASY Server [27]. The extinction coefficients at 280 nm for the different type I CBM-Coh constructs were typically between 40 500 and 42 000 M⁻¹ cm⁻¹ and higher values (between 45 000 and 46 400 M⁻¹ cm⁻¹) were calculated for the type II CBM-Coh constructs. Those of the Xyn-Doc constructs were generally around 85 000 M⁻¹ cm⁻¹.

Affinity-based ELISA

The detection of cohesin–dockerin interactions was carried out by the ELISA-based affinity assay as described in Barak et al. [19]. Briefly, the ELISA plate was coated with saturating amounts of CBM-Cohs chimaeras of both rival types (~30 nM) and the desired Xyn-Doc (~0.2 nM) was applied to the plate. The corresponding WT cohesins were used as references. The amount of interacting dockerin was determined immunochemically using anti-xylanase primary antibody and HRP-labeled secondary antibody. In this approach, we evaluated by the ELISA-based matching fusion protein system, the binding

affinity of the WT and the chimaeric cohesins to the counterpart dockerin in the presence of 1 mM CaCl₂.

ACKNOWLEDGMENTS

This research was supported in part by the Israel Science Foundation (Grant Nos. 966/09, 159/07, and 291/08) and by grants from the United States–Israel Binational Science Foundation (BSF), Jerusalem, Israel. E.A.B. holds The Maynard I. and Elaine Wishner Chair of Bio-organic Chemistry.

REFERENCES

1. E. A. Bayer, L. J. W. Shimon, R. Lamed, Y. Shoham. *J. Struct. Biol.* **124**, 221 (1998).
2. R. Lamed, E. Setter, R. Kenig, E. A. Bayer. *Biotechnol. Bioeng. Symp.* **13**, 163 (1983).
3. R. Lamed, E. Setter, E. A. Bayer. *J. Bacteriol.* **156**, 828 (1983).
4. E. A. Bayer, E. Morag, R. Lamed. *Trends Biotechnol.* **12**, 379 (1994).
5. B. Lytle, C. Myers, K. Kruus, J. H. D. Wu. *J. Bacteriol.* **178**, 1200 (1996).
6. S. Salamiou, O. Raynaud, M. Lemaire, M. Coughlan, P. Béguin, J.-P. Aubert. *J. Bacteriol.* **176**, 2822 (1994).
7. K. Tokatlidis, P. Dhurjati, P. Béguin. *Protein Eng.* **6**, 947 (1993).
8. K. Tokatlidis, S. Salamiou, P. Béguin, P. Dhurjati, J.-P. Aubert. *FEBS Lett.* **291**, 185 (1991).
9. E. Leibovitz, P. Béguin. *J. Bacteriol.* **178**, 3077 (1996).
10. E. Leibovitz, H. Ohayon, P. Gounon, P. Béguin. *J. Bacteriol.* **179**, 2519 (1997).
11. M. Lemaire, H. Ohayon, P. Gounon, T. Fujino, P. Béguin. *J. Bacteriol.* **177**, 2451 (1995).
12. L. J. W. Shimon, E. A. Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham, F. Frolow. *Structure* **5**, 381 (1997).
13. G. A. Tavares, P. Béguin, P. M. Alzari. *J. Mol. Biol.* **273**, 701 (1997).
14. S. Spinelli, H. P. Fierobe, A. Belaich, J. P. Belaich, B. Henrissat, C. Cambillau. *J. Mol. Biol.* **304**, 189 (2000).
15. I. Noach, F. Frolow, H. Jakoby, S. Rosenheck, L. J. W. Shimon, R. Lamed, E. A. Bayer. *J. Mol. Biol.* **348**, 1 (2005).
16. I. Noach, R. Lamed, Q. Xu, S. Rosenheck, L. J. W. Shimon, R. Kenig, E. A. Bayer, F. Frolow. *Acta Crystallogr., D* **59**, 1670 (2003).
17. A. L. Carvalho, V. M. Pires, T. M. Gloster, J. P. Turkenburg, J. A. Prates, L. M. Ferreira, M. J. Romão, G. J. Davies, C. M. Fontes, H. J. Gilbert. *J. Mol. Biol.* **349**, 909 (2005).
18. D. Nakar, T. Handelsman, Y. Shoham, H.-P. Fierobe, J. P. Belaich, E. Morag, R. Lamed, E. A. Bayer. *J. Biol. Chem.* **279**, 42881 (2004).
19. Y. Barak, T. Handelsman, D. Nakar, A. Mechaly, R. Lamed, Y. Shoham, E. A. Bayer. *J. Mol. Recognit.* **18**, 491 (2005).
20. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins. *Bioinformatics* **23**, 2947 (2007).
21. J. Pei, N. V. Grishin. *Bioinformatics* **23**, 802 (2007).
22. E. Morag, A. Lapidot, D. Govorko, R. Lamed, M. Wilchek, E. A. Bayer, Y. Shoham. *Appl. Environ. Microbiol.* **61**, 1980 (1995).
23. A. L. Carvalho, F. M. Dias, J. A. Prates, T. Nagy, H. J. Gilbert, G. J. Davies, L. M. Ferreira, M. J. Romão, C. M. Fontes. *Proc. Natl. Acad. Sci. USA* **100**, 13809 (2003).
24. A. L. Carvalho, F. M. V. Dias, T. Nagy, J. A. M. Prates, M. R. Proctor, N. Smith, E. A. Bayer, G. J. Davies, L. M. A. Ferreira, M. J. Romão, C. M. G. A. Fontes, H. J. Gilbert. *Proc. Natl. Acad. Sci. USA* **104**, 3089 (2007).

25. B. A. Pinheiro, M. R. Proctor, C. C. Martinez-Fleites, J. A. M. Prates, V. A. Money, G. J. Davies, E. A. Bayer, C. M. G. A. Fontes, H.-P. Fierobe, H. J. Gilbert. *J. Biol. Chem* **283**, 18422 (2008).
26. J. J. Adams, G. Pal, Z. Jia, S. P. Smith. *Proc. Natl. Acad. Sci. USA* **103**, 305 (2006).
27. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch. In *The Proteomics Protocols Handbook*, J. M. Walker (Ed.), pp. 571–607, Humana Press, Tolowa, NJ (2005).