PROTEINS: Structure, Function, and Genetics 39:170–177 (2000)

Cohesin-Dockerin Recognition in Cellulosome Assembly: Experiment Versus Hypothesis

Adva Mechaly,[1] Sima Yaron,[2] Raphael Lamed,[3] Henri-Pierre Fierobe,[4] Anne Belaich,[4] Jean-Pierre Belaich,[4,5] Yuval Shoham,[2] and Edward A. Bayer[1]*

[1]*Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel*
[2]*Department of Food Engineering and Biotechnology, and Institute of Catalysis Science and Technology, Technion—Israel Institute of Technology, Haifa, Israel*
[3]*Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Ramat Aviv, Israel*
[4]*Bioénergétique et Ingéniérie des Protéines, Centre National de la Recherche Scientifique, IBSM-IFR1, Marseille, France*
[5]*Université de Provence, Marseille, France*

ABSTRACT **The cohesin-dockerin interaction provides the basis for incorporation of the individual enzymatic subunits into the cellulosome complex. In a previous article (Pagés et al., Proteins 1997;29:517–527) we predicted that four amino acid residues of the ~70-residue dockerin domain would serve as recognition codes for binding to the cohesin domain. The validity of the prediction was examined by site-directed mutagenesis of the suspected residues, whereby the species-specificity of the cohesin-dockerin interaction was altered. The results support the premise that the four residues indeed play a role in biorecognition, while additional residues may also contribute to the specificity of the interaction. Proteins 2000;39:170–177.**

Key words: cellulosome; cellulases; multi-enzyme complex; scaffoldin subunit; protein-protein interaction; *Clostridium thermocellum; Clostridium cellulolyticum*

INTRODUCTION

Cellulosomes are exocellular macromolecular machines, designed for efficient degradation of cellulose and associated plant wall polysaccharides.[1] In a given microbe, the cellulosome complex is composed of a collection of subunits, each of which comprises a set of interacting functional modules.[2,3] The various enzymatic subunits are integrated into the complex by means of a major nonhydrolytic polypeptide, called scaffoldin. A complementary pair of modules—the cohesins on scaffoldin and the dockerins on the enzymatic subunits—provides the high-affinity-based interaction that defines the cellulosome complex.

Crystal structures of recombinant cohesin domains have recently been determined,[4,5] but structures for dockerin domains alone or for the cohesin-dockerin complex are still wanting. Over 50 different dockerin sequences from cellulosomal enzymes are currently known from about ten different organisms,[6] and the wealth of sequences provides a potential alternative resource for obtaining insight into the cohesin-dockerin interaction. In this context, we demonstrated previously that the interaction among cohesins and dockerins between the two cellulosome-producing clostridia, *C. thermocellum* and *C. cellulolyticum*, is species specific.[7] On the basis of these findings, we compared the different dockerin sequences of the two bacteria to determine which amino acids could account for the observed species specificity. Only four suspected residues were unambiguously indicated. The four residues are located within a 22-residue duplicated sequence, characteristic of the dockerin domain. In *C. thermocellum*, the two pairs of suspected recognition determinants include a serine residue in the first position and either a serine or a threonine in the second. In *C. cellulolyticum*, the two pairs include an alanine in the first position and a hydrophobic residue (usually leucine or isoleucine) in the second.

In order to test the validity of our prediction, we used site-directed mutagenesis to modify the suspected residues of two cellulosomal enzymes—one from each bacterium—in an attempt to convert the respective dockerin specificities to match the cohesin of the other species. The results support the original prediction that these four residues serve as recognition codes in the interaction with the cohesin. The data also suggest that additional dockerin residues apparently contribute to the specificity characteristics of the interaction.

MATERIALS AND METHODS

General

Agar, tryptone, and yeast extracts were from Difco Laboratories (Detroit, MI). Avicel Type PH-101 (microcrystalline cellulose) was from FMC BioProducts (Vallensbaek, Denmark). Ampicillin, kanamycin and N,N,N′,N′-

Abbreviations: CBD, cellulose-binding domain; CelS-ct, recombinant cellulosomal Family-48 cellulase from *C. thermocellum*; mCelS-ct, mutated CelS-ct; CelA-cc, recombinant cellulosomal Family-5 cellulase from *C. cellulolyticum*; mCelA-cc, mutated CelA-cc; Coh2-ct, recombinant probe from the scaffoldin subunit of *C. thermocellum*, which contains the cohesin-2 domain and the CBD; Coh1-cc, recombinant probe from the scaffoldin subunit of *C. cellulolyticum*, which contains the CBD, a domain of unknown function and the cohesin-1 domain PCR, polymerase chain reaction.

*Correspondence to: Edward A. Bayer, Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100 Israel. E-mail: bfbayer@wicc.weizmann.ac.il

Received 12 October 1999; Accepted 14 December 1999

TABLE I. Details of Primers Used for OE-PCR Mutagenesis[†]

Primer	Sequence	Restriction site
N-ter	5′-GAATTCATTAAAGAGGAGAAA	
C-ter	5′-CATTACTGGATCTATCAACAGG	
S688AT689L	5′-GGAAAAGTTAAC**GC**G**TTA**GACGCTGTAGCATTG	*Mlu*I
S720AT721L	5′-CAAAATTCCTAAGTC**TAA**TG**C**ATTAACTCTGCC	*Nsi*I
N-ter-NcoI	5′-CATATTGTCCATGGAATTTTGCAGG	
C-ter-ScaI	5′-GTGACTGGTGAGTACTCAACC	
A424SL425T	5′-GGAAAT**GTTGAC**T**C**AA**C**TGATTTTGCAG	*Hinc*II
A455SF456T	5′-GCAAGGTCA**GT**TG**A**ATTCACTTC	*Eco*RI

[†]Mutated bases shown in bold and restriction sites underlined.

tetra methyl ethylene diamine (TEMED) were from Sigma Chemical Co. (St. Louis, MO). Restriction and modification enzymes and buffers were from New England BioLabs (Beverly, MA). Pfu DNA polymerase (Promega, Madison, WI) was used for high-fidelity PCR.

Bacterial Strains and Vectors

Escherichia coli strain XL1-Blue was obtained from Stratagene (La Jolla, CA). *E. coli* strain MutS was obtained from CLONTECH Laboratories (Palo Alto, CA). *E. coli* strain BL21 (λDE3) and the T7 RNA polymerase expression vector pET9d (Novagen, Madison, WI) were described elsewhere.[8] Plasmids pQE-30 and pREP4 were from QIAGEN (Valencia, CA). *Clostridium thermocellum* strain YS, was originally isolated from soil samples obtained at the hot springs of Yellowstone National Park.[9,10]

DNA Manipulation

DNA was manipulated by standard procedures.[11,12] DNA transformation was performed using either the calcium chloride method[12] for strain BL21(DE3)pLysS or by electroporation using GeneZapper (IBI, New Haven, CT) for strains XL-1 Blue and MutS. DNA sequencing was performed by the DNA Sequencing Unit at the Weizmann Institute (Rehovot, Israel).

Cloning

The dockerin domain of the *C. thermocellum* Family-48 CelS was cloned into the His-tag-containing plasmid, pQE30. The primers were based on the gene sequence reported by Wang et al.[13] Since the *N*-terminal region of the dockerin domain of CelS is almost identical to that of CelE, the primer was designed to commence within the unique linker region. The primer (5′-GGAATTCGGATC-C**ATCGAAGGTCGTT**CTACTAAATTATACGGCGACG-3′) also contained a Factor-Xa protease cleavage site (shown in bold font), located downstream of *Eco*RI- and *Bam*HI-restriction sites (single and double underline, respectively), to enable future removal of the His-tag. The *Bam*HI restriction site was in frame with both the start codon and the His-tag. The *C*-terminal primer (5′-GACCTTAGATCTAAGCTTAGTTCTTGTACGGCAATG-3′) contained a *Hind*III site (underlined).

The entire *cel*S without its leader sequence was cloned into the pET9d vector. The *N*-terminal primer (5′-GGGG-TACCGCGGCCATGGGTCCTACAAAGGCACCTAC- 3′) contained an ATG translational start codon inside an *Nco*I restriction site (underlined). The *C*-terminal primer (5′-GACCTTAGATCTAAGCTTAGTTCTTGTACGGCAATG-3′) contained a stop codon and a *Bgl*II restriction site (underlined).

The DNA segments were amplified via PCR from total chromosomal DNA of *C. thermocellum* strainYS. Pfu DNA polymerase was used under conditions recommended by the manufacturer (Promega). Positive clones were verified by DNA sequencing.

Details of the cloning of cohesin constructs (Coh2-ct from *C. thermothelum* and Coh1-cc from *C. cellulolyticum*), as well as the Family-5 endoglucanase CelA-cc from *C. cellulolyticum*, were described elsewhere.[14–16]

Site-Directed Mutagenesis

The overlap-extension (OE) PCR method

Mutations were introduced into the dockerin domain of the intact *C. cellulolyticum* CelA-cc and into the detached, enzyme-free dockerin construct derived from CelS of *C. thermocellum* by OE-PCR site-directed mutagenesis method.[17,18] This strategy employs four oligonucleotides in two sequential PCR reactions, using two flanking and two different internal mutagenic primers, in an attempt to generate simultaneously the two pairs of desired mutations in a single step. The two resultant overlapping fragments were then annealed and extended to produce the desired full-length DNA. The flanking primers were then added to the reaction mixture, and the DNA was amplified by conventional PCR. The flanking and mutagenic primers used for the mutagenesis of the dockerin-containing constructs are shown in Table I.

For the mutagenesis of the detached enzyme-free dockerin, the flanking primer N-ter was used with the mutagenic primer S720AT721L and C-ter with S688AT689L, in order to amplify the two overlapping PCR fragments. After annealing and elongation, the two flanking primers were added to the mixture to produce the entire mutated dockerin domain. In a similar manner, the desired mutations were introduced into *celA* by using flanking primers N-ter-NcoI with A455S-F456T and C-ter-ScaI with A424S-L425T.

The unique-site eliminating method

The mutations in the detached enzyme-free dockerin construct were transferred to the intact parent enzyme

(CelS-ct), using the unique-site elimination procedure.[19] The selection primer (5′-CCAGCGCGTC**CGCGG**CCATGC-CGG-3′) was designed to replace a unique *Eag*I site in the partial tetracycline resistance gene of the pET vectors with a *Sac*II site (underlined, with the mutations shown in bold). Instead of a mutagenic primer, a 198-bp AatII-BglII fragment, containing the mutated dockerin, was used. Since this fragment was about 7-fold longer than a typical primer, approximately 700 ng were mixed with 100 ng of the selection primer for the annealing procedure.

Expression of Recombinant Proteins

The pET-derivatized plasmids were transformed into competent BL21(DE3)pLysS cells, containing the inducible T7 RNA polymerase. The pQE-30 clone, together with the pREP4 plasmid, was transformed into competent *E. coli* XL-1 Blue cells. All transformed cells were grown at 37°C (250 rpm) on Luria Broth medium, supplemented with glycerol (12 g/liter) and the appropriate antibiotics, to an OD_{600} of 2. The cells were then stored on ice for 3.5 h and at 15°C for another 30 min without shaking. Isopropyl-β-D-thiogalactoside (IPTG) was then added to a final concentration of 0.1 mM, and the cells were incubated for 17 h at 15°C. In the case of the dockerin-containing constructs, the medium also contained 2 mM $CaCl_2$. The OD_{600} of the overnight cultures varied from 3 to 6. Cultures were harvested by centrifugation (6000 rpm, 10 min). The cells were resuspended in a minimum amount of either ice-cold lysis buffer (50 mM NaH_2PO_4, pH 8, containing 300 mM NaCl and 10 mM imidazole, for the purification of the His-tag containing construct) or sonication buffer (50 mM sodium phosphate buffer, pH 7.8, containing 0.3 M NaCl). The cells were broken by sonication (4 × 20 sec, level 7, using an Ultrasonic processor XL2020, MISONIX Inc., Farmingdale, NY) and centrifuged (14,000 rpm, 20 min). Crude extracts originating from *C. thermocellum* were subjected to heat treatment (60°C, 10 min) before centrifugation.

Purification of Expressed Proteins

The CBD-containing constructs (Coh2-ct and Coh1-cc) were purified by affinity chromatography on microcrystalline cellulose (Avicel Type PH-101 FMC) as described previously.[14] The bound proteins were eluted from the cellulose matrix with 1% triethylamine (4 ml per gram Avicel). The eluents were neutralized with phosphate buffer (pH 7) and maintained at either 4°C (*C. cellulolyticum*) or −20°C (*C. thermocellum*).

The detached, enzyme-free dockerin constructs, containing the His-tag, were isolated by metal-chelate affinity chromatography using a nickel-containing resin,[20,21] nitrilo-tri-acetic acid (Ni-NTA), according to the manufacturer's instructions (QIAGEN).

For the purposes of the present study, the crude extracts of the native and mutant dockerin-containing, intact enzyme constructs were used for affinity-blotting studies without additional purification.

Complex Formation in Solution

In preliminary experiments, complex formation between detached, enzyme-free dockerins and recombinant cohesin constructs was carried out by combining the proteins in 50 mM Tris-HCl buffer (pH 7.5), containing 300 mM NaCl and 10 mM $CaCl_2$. The mixture was incubated for 10 min at room temperature, and the resulting complexes were analyzed by nondenaturing polyacrylamide gel electrophoresis as described earlier.[22]

Protein Biotinylation

When desired, cohesin-containing constructs and crude extracts of the dockerin-containing enzymes were biotinylated using biotinyl *N*-hydroxysuccinimide ester at a 15-fold molar ratio of reagent to protein as described previously.[23] The cohesin-containing constructs from both species and the native and mutant crude enzyme constructs from *C. thermocellum* all retained binding activity and were recognized by avidin probes. Biotinylation of crude enzyme constructs from *C. cellulolyticum* was ineffective in the binding assay, and conventional Western blotting was employed instead as described below.

Affinity Blot Analysis

Protein samples were separated by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) on 10% gels,[24] and the separated proteins were transferred electrophoretically onto 0.45-μm nitrocellulose membranes (Scheicher & Schuell, Dassel, Germany). The membranes were blocked using a 0.5% (w/v) solution of bovine serum albumin (BSA, Fraction V, Sigma) in PBS buffer (0.8% NaCl, 0.02% KCl, 0.144% $NaHPO_4$ and 0.024% KH_2PO_4, pH 7.4). Additional incubation and washing steps were performed in 50 mM Tris-HCl, pH 7.75, containing 0.05% (v/v) Tween 20, 0.5% BSA, 2 mM $CaCl_2$ and 0.9% NaCl.

The interaction between cohesin- and dockerin-containing constructs was analyzed using two complementary strategies: (1) Blotted, SDS-PAGE-separated, crude extracts of dockerin-containing enzyme constructs (~2 to 4 μg protein per lane) were reacted with biotinylated cohesin-containing constructs (~3 to 8 μg protein per ml per cm^2 nitrocellulose membrane). The blots were developed using ExtrAvidin-horseradish peroxidase complexes (1:2000 dilution, Sigma) as described by Morag et al.[25] (2) Blotted, SDS-PAGE-separated, cohesin-containing constructs (~ 2 μg protein per lane) were treated with the desired enzyme preparation (~20 to 50 μg protein per ml per cm^2 nitrocellulose membrane). In the case of CelS-ct and mCelS-ct, the membrane was reacted with biotinylated preparations of the respective crude extract and developed as above using ExtrAvidin-horseradish peroxidase complexes. For CelA-cc and mCelA-cc, the membrane was reacted successively with the desired crude extract, followed by anti-CelA antibodies (1:2000 dilution)[15,16] and developed with alkaline phosphatase goat anti-rabbit IgG (H+L) (1:2000 dilution, Jackson, ImmunoResearch Laboratories, Inc., West Grove, PA).

Miscellaneous Methods

Protein concentration was determined either according to Bradford,[26] using ovalbumin as a standard, or by using the absorbance at 280 nm, based on the known amino acid composition of the desired protein.[27] Purity was estimated by SDS-PAGE.[24]

RESULTS

In preliminary experiments, we tried to examine the interaction between enzyme-free dockerin constructs and the two different cohesin domains, both in solution and by affinity blotting. For this purpose, the suspected recognition residues in a detached recombinant dockerin derived from the *C. thermocellum* CelS subunit were mutated to resemble those of the dockerin of a typical cellulosomal subunit from *C. cellulolyticum*. However, using this approach, ambiguous results were obtained that were traced to the properties of the relatively small, free dockerin polypeptides. Indeed, in the previously described experimental system[7,14] the cohesin-dockerin interaction was examined using intact dockerin-containing enzymes. We therefore decided to introduce the mutations within the dockerin portion of the gene for the intact parent cellulosomal enzyme. In this manner, it was hoped that proper folding of the dockerin domain would be maintained.

Mutagenesis of the Intact Dockerin-Borne Cellulosomal Enzymes

Transformation of dockerin mutations into the CelS-ct gene

The gene encoding the Family-48 CelS subunit of the *C. thermocellum* cellulosome was cloned and expressed in *E. coli* as described previously.[28] In the present work, CelS-ct was amplified without its leader sequence, via PCR using chromosomal DNA from *C. thermocellum* strain YS, and cloned into the pET9d vector. Since the desired mutations had already been inserted into the detached recombinant dockerin from the same enzyme, we decided to transfer them en bloc to the parent enzyme subunit. Transformation of the mutations was accomplished by site-directed mutagenesis using the unique-site elimination procedure.[19] The entire mutant dockerin domain was used as the mutagenic primer, and all four mutations were successfully inserted in one step. The resultant construct contained the four desired mutations (S688A,[†] T689L, S720A, and T721L), and the expressed protein was termed mCelS-ct. The mutations were designed to change the interspecies specificity of the dockerin, based on the previously described homology-based prediction.[7] Thus, the putative recognition dyads, S688–T689 and S720–T721 of the CelS dockerin (representing positions 10 and 11 of the calcium-binding loop)[7] from *C. thermocellum*, were each modified to A–L. On the basis of the described model, the interconversion of these residues was expected to reverse the inherent specificities, such that the dockerin would no longer recognize the *C. thermocellum* cohesins, but would recognize those of *C. cellulolyticum*.

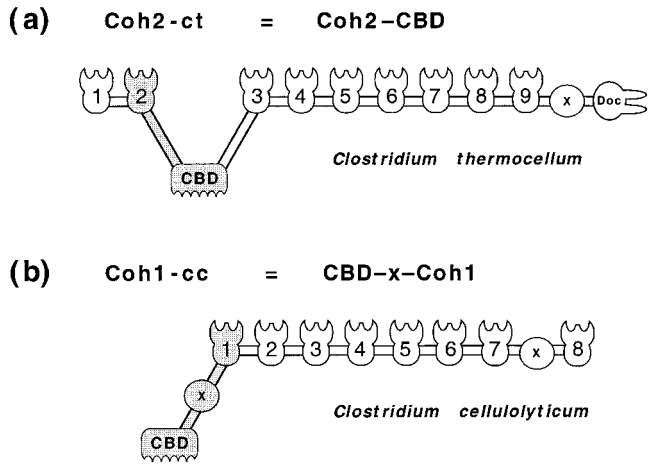

Fig. 1. Schematic representation of the recombinant cohesin-containing probes, used for interaction with dockerin-borne enzymes. The modular organization of the complete scaffoldin subunit, derived from the two species, is presented. The shaded portion of each indicates the contents of the respective cohesin-containing probe. Shown are the positions of the cohesin domains (numbered) relative to the CBD in the scaffoldin of the designated species. In addition to the CBD and cohesins, the *C. cellulolyticum* scaffoldin includes two domains of unknown function (designated x), and the *C. thermocellum* scaffoldin bears one such domain and a C-terminal type-II dockerin (Doc). The *C. thermocellum*-derived probe, Coh2-ct, contains the second cohesin domain, connected to the CBD by a linker segment. The *C. cellulolyticum*-derived probe, Coh1-cc, contains the first cohesin domain, connected to the N-terminal CBD via linkers and a domain (x) of unknown function.

Mutagenesis of the dockerin domain of the CelA-cc gene

The dockerin domain of the Family-5, cellulosomal glycosyl hydrolase CelA from *C. cellulolyticum*[16] was mutated using OE-PCR site-directed mutagenesis.[17] The enzyme was mutated in the following positions: A424S, L425T, A455S, and F456T, such that the new residues emulated the putative recognition residues of the *C. thermocellum* dockerins.[7]

Interaction Between Cohesins and Intact Dockerin-Mutated Enzymes

Two reciprocal experiments, based on affinity blotting, were used to examine the specificity of interaction between the cohesin probes and the cellulosomal enzymes mutated in the desired positions of their resident dockerin domains. In the first set of experiments, samples of the wild-type and mutated enzymes (CelS-ct, mCelS-ct, CelA-cc and mCelA-cc) were subjected to SDS-PAGE and blotted onto nitrocellulose. The separated bands were then challenged with the cohesin probes, Coh2-ct and Coh1-cc (Fig. 1). In the reciprocal set, the cohesin probes were first separated on SDS-PAGE and blotted onto nitrocellulose, followed by interaction with the enzymes.

The results of the enzyme blots (Fig. 2) clearly indicated that mutation of the four putative recognition residues of the enzymes from both species generated new and distinct

[†]Numbering of suspected recognition residues in all of the recombinant proteins includes the signal peptide.

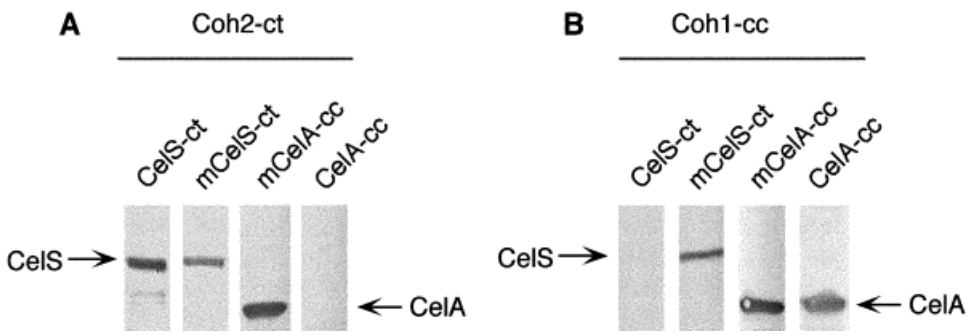

Fig. 2. Affinity-blotting of native and mutant dockerin-bearing enzymes using recombinant cohesin-containing probes. The native and mutant enzyme samples, CelS-ct and mCelS-ct from *C. thermocellum* and CelA-cc and mCelA-cc from *C. cellulolyticum*, were subjected to SDS-PAGE and blotted onto nitrocellulose membranes. The separated proteins were probed using the desired biotinylated cohesin construct: **Blot A**: Coh2-ct from *C. thermocellum* and **Blot B**: Coh1-cc from *C. cellulolyticum* (see Fig. 1 for composition of the respective probe). The blots were developed using an avidin-peroxidase conjugate. The positions for the native CelS and CelA are indicated.

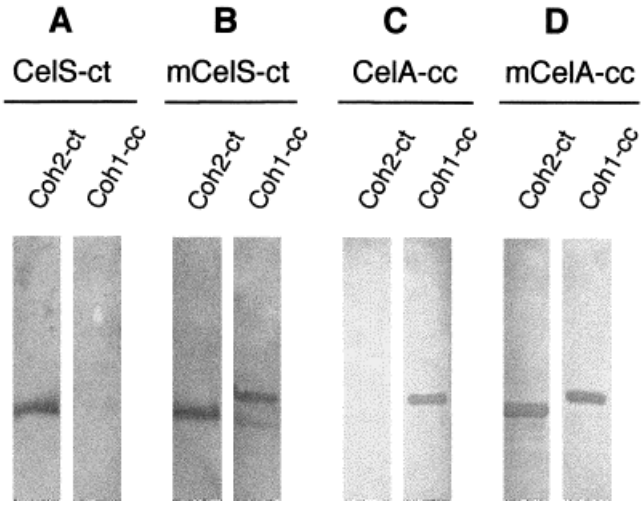

Fig. 3. Affinity-blotting of the recombinant cohesin constructs using native and mutant dockerin-bearing enzymes as probes. The cohesin-containing constructs, Coh2-ct and Coh1-cc, were subjected to SDS-PAGE and blotted onto nitrocellulose membranes. The blots were challenged by the respective native or mutated enzyme and developed as described in the Methods Section. Note, the native dockerin-borne enzymes (**Blots A and C**) recognize only the cohesin construct from the same species, whereas the mutated dockerins (**Blots B and D**) from either species recognize the cohesins from both.

interactions. The *C. thermocellum* cohesin probe (Coh2-ct) was now capable of recognizing the mutated enzyme from *C. cellulolyticum*. Likewise, the *C. cellulolyticum* cohesin probe (Coh1-cc) selectively bound to the mutated enzyme from *C. thermocellum*. Interestingly, the conversion was not absolute in that the cohesin probes from both species retained the capacity to recognize the mutated enzyme from the same species—i.e., in both cases, the mutations of the putative recognition residues of the dockerin domains failed to nullify the intraspecies interaction with the cohesin.

These results were confirmed in the reciprocal experiment, wherein the two species of cohesin probes were first blotted and then subjected to interaction using the mutated enzymes (Fig. 3). Once again, the mutated enzymes were now capable of recognizing the cohesin from the alternate species while retaining the original specificity for its own cohesin.

DISCUSSION

It has long been known that cellulase and cellulosomal components are composed of a collection of functional domains or modules,[29–31] the principle ones being catalytic modules that perform the enzymatic cleavage of the substrate.[32,33] In addition, accessory modules appear to regulate its activity in some way.[6] For example, the cellulose-binding domain—the CBD—serves to target the catalytic module to its substrate, thereby facilitating its hydrolysis.[34] Unique to the cellulosome are the cohesin and dockerin domains, the interaction of which forms the basis for incorporation of the enzyme subunits into the complex.[2,3,35]

The sequence of the dockerin domain is particularly intriguing, as it contains a duplicated segment that bears striking sequence resemblance to the EF-hand motif, characteristic of various calcium-binding proteins.[36,37] The cohesin-dockerin interaction was thus studied further, and a series of reports suggested the following: (1) the dockerin domain of cellulosomal enzymes interacts selectively with the cohesin domains of the scaffoldin subunit,[38,39] (2) the interaction is calcium dependent and the calcium-binding property resides within the dockerin component,[4,14,40] (3) the interaction is relatively unselective among the different cohesins and dockerins[14,22] and (4) both duplicated segments of the dockerin are required for the high-affinity interaction with the cohesins.[41,42]

The present study focuses on the biorecognition phenomenon between the cohesin and dockerin domains. In this work, we have succeeded in creating a new recognition specificity between the cohesin and dockerin domains from two cellulosome-producing species, *C. thermocellum* and *C. cellulolyticum*, based on our previously described prediction of putative recognition residues in the dockerin domain.[7] The theoretical rationale for this prediction was achieved through comparative bioinformatics of an aligned set of dockerin domains from the two species. Four amino

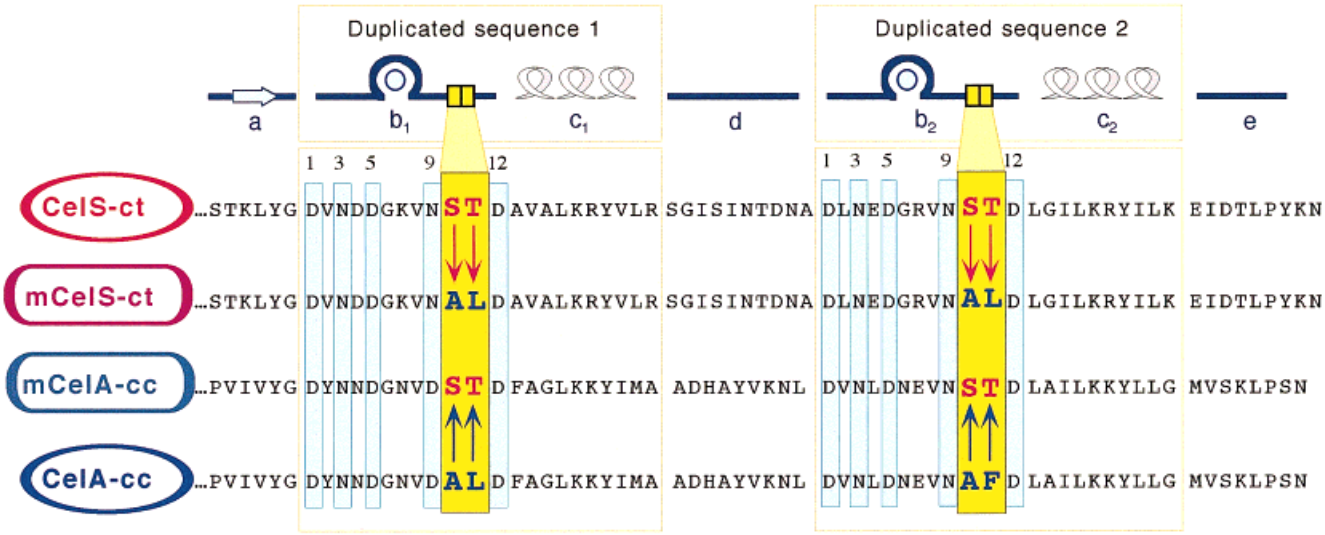

Fig. 4. Exchange of putative recognition determinants in dockerin-bearing enzymes from *C. thermocellum* and *C. cellulolyticum*. The dockerin sequences from *C. thermocellum* CelS and *C. cellulolyticum* CelA are shown, partitioned into seven parts (a–e) and color-coded according to Pagés et al.[7] The first 12 amino acids (segments b_1 and b_2) of the 22-residue duplicated sequences are believed to form calcium-binding loops, and the last 10 residues (segments c_1 and c_2) appear to form α helices, thus assuming a proposed F-helix variant of the EF-hand motif. The side-chains of blue-shaded residues presumably interact with a calcium ion. The yellow-shaded residues denote the suspected recognition residues—positions 10 and 11 of the duplicated sequence in the respective mutant dockerins (mCelS-ct and mCelA-cc)—which were mutated as shown to emulate the dockerin domain of the rival species.

acid residues were indicated, which were located in distinct positions in the dockerin sequences.

In order to test the validity of our prediction, the four designated positions in the dockerin domains of two representative enzymes (one from each species) were converted by site-directed mutagenesis to emulate the respective residues of the counterpart species. The relevant mutations from the two species are illustrated schematically in Fig. 4. Thus, mutation of the relevant positions from ST to AL in the *C. thermocellum* dockerin and conversely from AL(F)[‡] to ST in the *C. cellulolyticum* dockerin resulted in a conversion of specificity, such that the mutated dockerins recognized the cohesins of the rival species.

While a bona fide structural model is still lacking for the dockerin domain, an interim model has been suggested, based on sequence alignment and homology modeling between parts of the dockerin sequence and the EF-hand motif. The resultant model was called the F-hand motif, since, according to secondary structure analysis, the E-helix appeared to be lacking.[7,43] The four designated recognition residues comprise a repeated tandem pair or dyad, located within the presumed calcium-binding loop of the two duplicated stretches that characterize the dockerin sequence. In the proposed model, we postulated that the four recognition residues would not be directly involved in calcium binding, but that the binding of calcium by neighboring residues would be required for the exposure or correct display of the recognition residues. In contradiction to earlier work,[40] we also claimed that both duplicated stretches would be critical to binding,[7,43] and this view has since been corroborated experimentally in recent work.[41,42,44]

The establishment of a viable experimental system to challenge the proposed recognition determinants hindered our initial attempts to investigate the validity of our original prediction. It should be noted that the detached enzyme-free dockerin domains were found experimentally to unfold in solution and to bind nonspecifically to various proteins. This erratic behavior of the enzyme-free dockerins was also reflected by changes in electrophoretic mobility in nondenaturing PAGE and in gel filtration studies (R. Kenig and R. Lamed, unpublished results). Increasing the concentration of calcium appeared to induce proper folding of the free domain, and in the presence of the recombinant cohesin, the calcium requirement was diminished. In this respect, the dockerin component in the intact cellulosomal enzyme appeared to be more stable, and for this reason we eventually shifted our mutagenesis efforts to concentrate on the dockerin domain within the intact parent enzyme.

By creating the new specificity, we have in essence confirmed our initial model[7] that these four residues serve as recognition codes in the interaction with the cohesin. It is interesting that only four residues were implicated in the original prediction. Originally, we had hoped to interconvert the specificities between the two species by chang-

[‡]Of the eight known *C. cellulolyticum* enzyme sequences, the proposed "recognition dyads" are usually AL or AI, the two exceptions being the Family-5 CelA enzyme (used in the present study), wherein the second duplicated pair is AF, and the Family-9 CelE, wherein the second pair is AV.

ing only these four residues. We considered that such an interconversion would provide stronger evidence as to the involvement of the implicated residues than would be demonstrated by simple mutagenesis-induced interference in binding. The results indeed support this contention. However, we were unable to completely *destroy* the *intra*species interaction of the mutated dockerins with the respective cohesin from the same species. This implies that additional residues, perhaps contributing more subtle secondary interactions, would also be involved in cohesin-dockerin recognition.

Ultimately, we hope to determine the precise elements that control the specificity of interaction between cohesins and dockerins. Such an accomplishment would potentially have biotechnological value,[45] since a graded range of affinities among affinity partners would serve various types of application. The cohesin and dockerin domains are particularly appropriate for the production of complementary hybrid proteins, since both operate naturally as functional modules in parent protein pairs.

In some respects, the results of this study were somewhat different than that anticipated. It was thus surprising that exchange of the four recognition residues led to the establishment of a new specificity rather than the repression of the native interaction. In the final analysis, however, it is far more constructive and meaningful to create a specific interaction than to destroy one. Further investigation into the fine structure and biorecognition properties of the cohesin-dockerin interaction is now under way.

ACKNOWLEDGMENTS

Parts of this work were documented in the Ph.D thesis of S.Y. The authors appreciate the contribution of Dr. Orit Gat (Biological Institute, Nes Ziona, Israel) for her initial preparations of the free dockerin constructs. A contract from the European Commission (Biotechnology Programme, Proposal PL962303) and grants from the Israel Science Foundation (administered by the Israel Academy of Sciences and Humanities, Jerusalem) are gratefully acknowledged. Additional support was provided by the Otto Meyerhof Center for Biotechnology, established by the Minerva Foundation, (Munich, Germany).

REFERENCES

1. Shoham Y, Lamed R, Bayer EA. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. Trends Microbiol 1999;7:275–281.
2. Bayer EA, Chanzy H, Lamed R, Shoham Y. Cellulose, cellulases and cellulosomes. Curr Opin Struct Biol 1998;8:548–557.
3. Bayer EA, Shimon LJW, Lamed R, Shoham Y. Cellulosomes: structure and ultrastructure. J Struct Biol 1998;124:221–234.
4. Shimon LJW, Bayer EA, Morag E, et al. The crystal structure at 2.15 Å resolution of a cohesin domain of the cellulosome from *Clostridium thermocellum*. Structure 1997;5:381–390.
5. Tavares GA, Béguin P, Alzari PM. The crystal structure of a type I cohesin domain at 1.7 Å resolution. J Mol Biol 1997;273:701–713.
6. Coutinho PM, Henrissat B. Carbohydrate-Active enZYmes and associated MODular Organization server (CAZyModO Website). [Online]. http://afmb.cnrs-mrs.fr/~pedro/DB/db.html. [10 November 1999, last date accessed]. 1999.
7. Pagés S, Belaich A, Belaich J-P, et al. Species-specificity of the cohesin-dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: Prediction of specificity determinants of the dockerin domain. Proteins 1997;29:517–527.
8. Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol 1990;185:60–89.
9. Lamed R, Setter E, Bayer EA. Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. J Bacteriol 1983;156:828–836.
10. Lamed R, Bayer EA. The cellulosome of *Clostridium thermocellum*. Advan Appl Microbiol 1988;33:1–46.
11. Ausubel FM, Brent R, Kingston RE, et al., editors. Short protocols in molecular biology. New York: Greene Publishing Associates/ Wiley Interscience; 1992.
12. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratories; 1989.
13. Wang WK, Kruus K, Wu JHD. Cloning and DNA sequence of the gene coding for *Clostridium thermocellum* cellulase S_S (CelS), a major cellulosome component. J Bacteriol 1993;175:1293–1302.
14. Yaron S, Morag E, Bayer EA, Lamed R, Shoham Y. Expression, purification and subunit-binding properties of cohesins 2 and 3 of the *Clostridium thermocellum* cellulosome. FEBS Lett 1995;360: 121–124.
15. Pagés S, Belaich A, Tardif C, Reverbel-Leroy C, Gaudin C, Belaich J-P. Interaction between the endoglucanase CelA and the scaffolding protein CipC of the *Clostridium cellulolyticum* cellulosome. J Bacteriol 1996;178:2279–2286.
16. Fierobe H-P, Gaudin C, Belaich A, et al. Characterization of endoglucanase A from *Clostridium cellulolyticum*. J Bacteriol 1991;173:7956–7962.
17. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 1989;77:51–56.
18. Meza R, Nufiez-Valdez M-E, Sanchez J, Bravo A. Isolation of Cry1Ab protein mutants of *Bacillus thuringiensis* by a highly efficient PCR site-directed mutagenesis system. FEMS Microbiol Lett 1996;145:333–339.
19. Deng WP, Nickoloff JA. Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. Anal Biochem 1992;200:81–88.
20. Porath J, Carlsson J, Olsson I, Belfrage G. Metal chelate affinity chromatography, a new approach to protein fractionation. Nature 1975;258:598–599.
21. Janknecht R, de Martynoff G, Lou J, Hipskind RA, Nordheim A, Stunnenberg HG. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proc Natl Acad Sci USA 1991;88:8972–8976.
22. Lytle B, Myers C, Kruus K, Wu JHD. Interactions of the CelS binding ligand with various receptor domains of the *Clostridium thermocellum* cellulosomal scaffolding protein, CipA. J Bacteriol 1996;178:1200–1203.
23. Bayer EA, Wilchek M. Protein biotinylation. Methods Enzymol 1990;184:138–160.
24. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 1970;227:680–685.
25. Morag E, Halevy I, Bayer EA, Lamed R. Isolation and properties of a major cellobiohydrolase from the cellulosome of *Clostridium thermocellum*. J Bacteriol 1991;173:4155–4162.
26. Bradford M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976;72:248–254.
27. Edelhoch H. Spectroscopic determination of tryptophan and tyrosine in proteins. Biochemistry 1967;6:1948–1954.
28. Wang WK, Kruus K, Wu JHD. Cloning and expression of the *Clostridium thermocellum* cellulase celS gene in *Escherichia coli*. Appl Microbiol Biotechnol 1994;42:346–352.
29. Hall J, Hazlewood GP, Barker PJ, Gilbert HJ. Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for activity. Gene 1988;69:29–38.
30. Ferreira LMA, Durrant AJ, Hall J, Hazlewood GP, Gilbert HJ. Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. Biochem J 1990;269:261–264.
31. Gilkes NR, Henrissat B, Kilburn DG, Miller RCJ, Warren RAJ. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol Rev 1991;55:303–315.
32. Henrissat B, Claeyssens M, Tomme P, Lemesle L, Mornon J-P.

Cellulase families revealed by hydrophobic cluster analysis. Gene 1989;81:83–95.
33. Coutinho PM, Henrissat B. Carbohydrate-Active enZYmes server (CAZy Website). [Online]. http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html. [10 November 1999, last date accessed]. 1999.
34. Tomme P, Warren RAJ, Miller RC, Kilburn DG, Gilkes NR. Cellulose-binding domains—Classification and properties. In: Saddler JM, Penner MH, editors. Enzymatic degradation of insoluble polysaccharides. Washington, D.C.: American Chemical Society; 1995. p 142–161.
35. Béguin P, Lemaire M. The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. Crit Rev Biochem Molec Biol 1996;31:201–236.
36. Chauvaux S, Béguin P, Aubert J-P, Bhat KM, Gow LA, Wood TM, Bairoch A. Calcium-binding affinity and calcium-enhanced activity of *Clostridium thermocellum* endoglucanase D. Biochem J 1990;265:261–265.
37. Herzberg O, James MN. Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 Å resolution. J Mol Biol 1988;203:761–779.
38. Tokatlidis K, Salamitou S, Béguin P, Dhurjati P, Aubert J-P. Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. FEBS Lett 1991;291:185–188.
39. Salamitou S, Raynaud O, Lemaire M, Coughlan M, Béguin P, Aubert J-P. Recognition specificity of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA. J Bacteriol 1994;176:2822–2827.
40. Choi SK, Ljungdahl LG. Structural role of calcium for the organization of the cellulosome of *Clostridium thermocellum*. Biochemistry 1996;35:4906–4910.
41. Lytle B, Wu JHD. Involvement of both dockerin subdomains in assembly of the *Clostridium thermocellum* cellulosome. J Bacteriol 1998;180:6581–6585.
42. Fierobe H-P, Pagés S, Belaich A, Champ S, Lexa D, Belaich J-P. Cellulosome from *Clostridium cellulolyticum:* Molecular study of the dockerin/cohesin interaction. Biochemistry 1999;38:12822–12832.
43. Bayer EA, Morag E, Lamed R, Yaron S, Shoham Y. Cellulosome structure: four-pronged attack using biochemistry, molecular biology, crystallography and bioinformatics. In: Claeyssens M, Nerinckx W, Piens K. editors. Carbohydrases from *Trichoderma reesei* and other microorganisms. London: The Royal Society of Chemistry; 1998. p 39–67.
44. Lytle B, Westler WM, Wu JHD. Molecular assembly of the *Clostridium thermocellum* cellulosome. In: Ohmiya K, Hayashi K, Sakka K, Kobayashi Y, Karita S, Kimura T. editors. Genetics, biochemistry and ecology of cellulose degradation. Tokyo: Uni Publishers Co.; 1999. p 444–449.
45. Bayer EA, Morag E, Lamed R. The cellulosome—A treasure-trove for biotechnology. Trends Biotechnol 1994;12:378–386.