A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins

Qi Xu[a], Mark Morrison[b,c,*], Karen E. Nelson[c,d], Edward A. Bayer[e], Nof Atamna[a], Raphael Lamed[a]

[a]*Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Ramat Aviv, Israel*
[b]*The MAPLE Research Initiative, Department of Animal Sciences, The Ohio State University, Columbus, OH, USA*
[c]*The North American Consortium for Genomics of Fibrolytic Ruminal Bacteria, USA*
[d]*The Institute for Genomic Research, Rockville, MD, USA*
[e]*Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot, Israel*

Received 26 February 2004; revised 24 March 2004; accepted 1 April 2004

Available online 20 April 2004

Edited by Ulf-Ingo Flügge

Abstract We recently showed that some of the enzymes underpinning cellulose solubilization by *Ruminococcus albus* 8 lack the conventional type of dockerin module characteristic of cellulosomal proteins and instead, bear an "X" domain of unknown function at their C-termini. We have now subcloned and expressed six X domains and showed that five of them bind to xylan, chitin, microcrystalline and phosphoric-acid swollen cellulose, as well as more heterogenous substrates such as alfalfa cell walls, banana stem and wheat straw. The X domain that did not bind to these substrates was derived from a family-5 glycoside hydrolase (Cel5G), which possesses two X domains in tandem. Whereas the internal X domain failed to bind to the substrates, the recombinant dyad exhibited markedly enhanced binding relative to that observed for the C-terminal X domain alone. The evidence supports a distinctive carbohydrate-binding role of broad specificity for this type of domain, and we propose a novel family (designated family 37) of carbohydrate-binding modules that appear to be peculiar to *R. albus*.

Keywords: Cellulose degradation; Glycoside hydrolase; Carbohydrate-binding module; Family 37 CBM

1. Introduction

Ruminococcus albus is a Gram-positive anaerobic bacterium that has been studied largely because of its ability to efficiently degrade and use cellulose as a carbohydrate source. Many such anaerobic bacteria are characterized by multi-protein complexes called cellulosomes that are considered to facilitate both substrate adhesion and efficient degradation of the intricate polysaccharide matrix of the plant cell wall. The cellulosome was initially discovered and described two decades ago from *Clostridium thermocellum* [1–3]. Since the original description, distinctive cellulosomes were established for other cellulolytic bacteria, including *Clostridium cellulovorans*, *Clostridium cellulolyticus*, *Clostridium josui*, *Acetivibrio cellulolyticus*, *Bacteroides cellulosolvens*, and *Ruminococcus flavefaciens* [4–16]. The presence of a dockerin domain in several *R. albus* enzymes implies the existence of a cellulosome in this bacterium as well [17–19].

We recently reported the isolation of a group of adhesion-defective mutants from *R. albus* that were deficient in their degradation of cellulosic substrates [20]. Combined analyses of the proteome and genome sequence data served to identify two glycoside hydrolases (Cel9B and Cel48A) that proved deficient in the mutant strains. Both Cel9B and Cel48A possess modular architectures and share characteristic features with processive endocellulases that have been characterized as key enzymes in the process of cellulose solubilization by other bacteria. Interestingly, both gene products lack dockerin sequences, but instead possess a novel type of X domain (of unknown function) at their C-termini.

In the present work, we considered three alternative hypotheses as to the possible function of the C-terminal X domain: (i) the domain may serve as an alternative type of dockerin-like domain to mediate assembly into a multi-protein cellulosome-like complex, (ii) the domain could mediate the attachment of the parent protein (enzyme in this case) to the cell surface or (iii) the domain could facilitate adhesion to cellulose and/or other insoluble polysaccharide substrates. In order to examine these hypotheses, we identified several additional X domains of this type that comprise components of the polypeptide chains of *R. albus* glycoside hydrolases. The individual X domains were subcloned and the expressed proteins were analyzed for their respective binding activities. In conflict with the first two hypotheses, none of the domains bound to any cellular components. On the other hand, most of the other expressed proteins bound to various insoluble polysaccharides, indicating a carbohydrate-binding role for this module.

2. Materials and methods

2.1. Genome sequence analysis

The peptide sequence of the Cel9B and Cel48A X domains was used as query sequences in tBLASTx searches of the *R. albus* strain 8 genome sequence data available via The Institute for Genomic Research's (TIGR) unfinished genomes website (http://www.tigr.org). Several contigs were identified which contained sequences with a high level of identity to the query sequences. The open reading frames

* Corresponding author. Fax: +1-614-292 7116.
E-mail address: morrison.234@osu.edu (M. Morrison).

0014-5793/$22.00 © 2004 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/j.febslet.2004.04.005

(ORF) within these contigs were identified via the ExPASy web site (http://www.expasy.ch).

2.2. Protein sequence analysis

Potential signal sequences were determined by the SignalP V2.0 program [21]. The parameters for molecular weight, theoretical p*I*, amino acid composition and extinction coefficient were computed using the ProtParam Tool (http://www.expasy.org/tools/protparam.html), available via the SWISS-PROT Website [22]. Multiple sequence alignment and phylogenetic trees were generated using the ClustalW program (http://www2.ebi.ac.uk/clustalw/). The X domains, Fn3-like domains, and enzyme sequences were obtained from either the GenBank Website (http://www.ncbi.nlm.nih.gov/), the SWISS-PROT Website or via the Carbohydrate-Active Enzymes server (CAZy Website, http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html), designed by Coutinho and Henrissat [23,24].

2.3. DNA-based methods

Ruminococcus albus 8 genomic DNA was isolated according to the protocol of Murray and Thompson [25]. PCRs were performed using a Master Personal device (Eppendorf, Hamburg, Germany) at various annealing temperatures, 55–60 °C. *Taq* polymerase used was TaKaRa Ex Taq™ (Takara Shuzo Co., Ltd., Kyoto, Japan). The resulting PCR fragments were cloned using the pGEM-T Vector System 1 (Promega Corporation, Madison, WI). *Escherichia coli* XL-1 strains were used as host cells for transformation. DNA samples were purified using either the QIAquick PCR Purification Kit (Qiagen, Inc., Valencia, CA), or Agarose Gel DNA Extraction Kit (Roche Diagnostics Corporation, Indianapolis, IN). Plasmids were purified using the High Puri Plasmid Isolation Kit (Boehringer Mannheim, Germany).

2.4. DNA sequencing

DNA sequencing was performed either directly on PCR products or on cloned fragments on an ABI PRISM® 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA) at the Sequencing Lab of Tel Aviv University, Israel. The resulting sequences were compared to known cellulosome-related proteins.

2.5. Cloning and overexpression of recombinant proteins

The appropriate genes were subcloned into expression vectors via PCR (see Fig. 1 for details). The PCR products were cloned into either pET28a or pMalC-NN-EGFP vectors, and their intact sequences were verified by DNA sequencing. The clones were expressed in *E. coli* BL21 (DE3) at 16 °C and grown in the presence of 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Following growth, the cultures were lysed by sonication according to Ding et al. [7]. The expressed proteins were identified by SDS–PAGE (using 10–12% resolving gels) and Coomassie brilliant blue staining.

2.6. Recombinant protein purification and protein assay

Recombinant proteins, fused with a His-tag, were purified by Ni–NTA affinity chromatography (Qiagen) following the manufacturer's specifications. Recombinant proteins fused with maltose-binding protein (MBP) were purified by Amylose Resin affinity chromatography (New England Biolabs) following manufacturer's supplier specifications. The protein concentrations were determined by Bradford assay (Bio-Rad) with bovine serum albumin fraction V as standard.

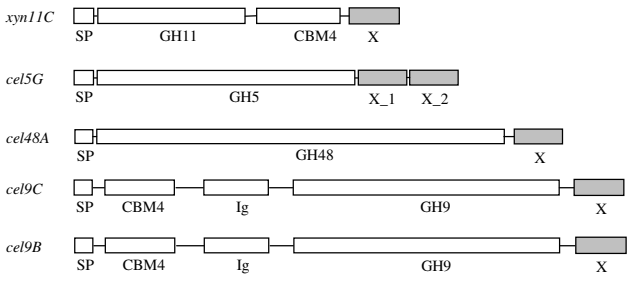

Fig. 1. Supramolecular architecture of the genes studied in this work. Each gene encodes for a protein that harbors a C-terminal X domain. *cel5G* encodes for a protein characterized by both an internal and a C-terminal X domain in tandem.

2.7. Binding assays

The recombinant proteins (20–40 μg) were mixed with suspended polysaccharides (0.3–2 mg) in a final volume of 0.2 ml with buffer (50 mM phosphate, 300 mM NaCl, pH 8.0). Avicel, xylan (from beechwood), lichenan and chitin were all purchased from Sigma Chem. Co. (St. Louis, MO). Sepharose was purchased from Amersham Biosciences (Glyfada, Greece). Neutral detergent fiber of alfalfa cell walls and banana fruit stem were prepared according to Van Soest et al. [26].

The mixtures were maintained at room temperature for 20 min with gentle rotation, then centrifuged at 12 000 × *g* for 6 min to sediment the polysaccharide and bound proteins. The supernatant (containing unbound proteins) was recovered and the protein concentration measured by Bradford assay (as described above), or alternatively, by running the recovered proteins on SDS–PAGE gels. The polysaccharide particles were also washed four times with 1-ml aliquots of the buffer solution described above. After centrifugation, the polysaccharides were resuspended with 80 μl of the same buffer and placed in a boiling water bath for 10 min. After centrifugation, the supernatant was recovered and the proteins subjected to SDS–PAGE.

3. Results

3.1. Identification and description of X domains from R. albus 8 glycoside hydrolases

The X domains from *cel9B* and *cel48A* were used as query sequences in tBLASTx searches of the *R. albus* strain 8 genome sequence data, and three full-length glycoside hydrolase sequences, containing X domains at their C-termini, were identified. These genes are hereafter referred to as *cel5G*, *cel9C* and *xyn11C*, according to the CAZy nomenclature, and their modular architecture is presented in Fig. 1.

The *cel9C* gene exhibits high homology and identical domain architecture with the previously described enzyme [20], *cel9B* (64% identity, 75% overall similarity). It encodes for a 999-residue protein that contains a signal peptide, a family-4 CBM, an Ig-like domain, a family-9 catalytic domain and a C-terminal X domain (homologous to those of *cel9B* and *cel48A*). The *xyn11C* gene encodes a 542-residue protein that includes a signal peptide, a family-11 catalytic domain, a family-4 CBM and an X domain. The protein encoded by the *cel5G* gene is composed of 682 amino acids, including a signal peptide, a family-5 catalytic domain and a tandem repeat of the X domain at the C-terminus. The same type of X domain has also been reported for two other previously sequenced *R. albus* enzymes – xylanases B (*xyn11B*) and C (*xyn5A*) [27]. To date, this type of domain has only been associated with proteins of *R. albus* and has not yet been observed in proteins from any other type of bacterium – including *R. flavefaciens*.

Phylogenetic analysis of the *R. albus* X domains is presented in Fig. 2, relative to other related X domains whose functions have not been fully characterized. The similarity to fibronectin-III domains (Fn3) and other X domains has been noted earlier [20]. Both X1 and X57 domains were erroneously designated Fn3 domains, but the distinction was later confirmed [28,29]. Likewise, the *R. albus* X domains map together on a branch of the tree separate from those of the Fn3 and other related domains and form a clearly distinct group. Most of the latter form a relatively close cluster, that of XynC is somewhat removed from the cluster, whereas the internally positioned Cel5G X_1 domain is clearly distant from the others. Although the internal Cel5G X_1 domain radiates from a focal point close to that of the Fn3 domains, it clearly belongs to the *R. albus* cluster by virtue of three conserved tryptophans (not

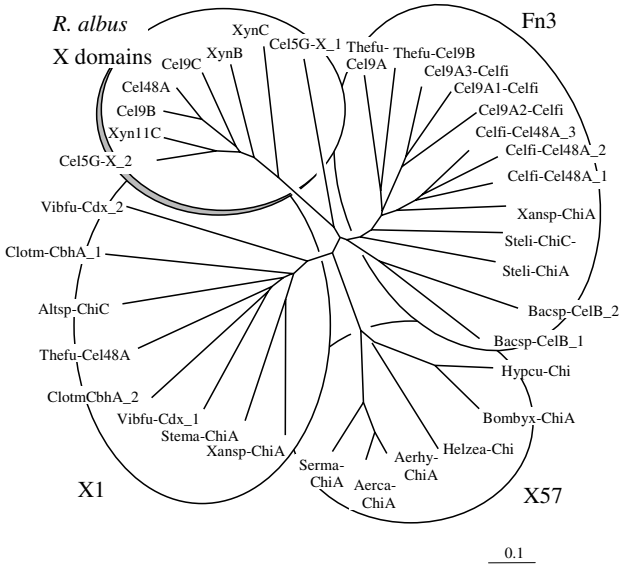

Fig. 2. Phylogenetic analysis of the X domains of *R. albus* enzymes, used in this study. The latter domains are mapped on a background of very similar domains, classified as Fn3, X1 and X57. Scale bars indicate percentage (0.1) of amino acid substitutions. See Fig. 6 of [20] and Table 2 of this article for sources of the sequences and abbreviations used in this figure.

shown) in the respective sequences, which are distinctive to their family.

3.2. Expression of X domains

The X domains of *cel5G*, *cel9B*, *cel9C*, *xyn11C*, and *cel48A* were cloned and expressed, using the primers listed in Table 1. The expressed recombinant proteins and their solubility properties are shown in Table 2. When expressed as His-tag fusion proteins, the recombinant proteins were soluble in crude *E. coli* lysates, but all the X domains were found to aggregate readily after Ni-column purification. Conversely, when fused to a MBP the X domains remained soluble or mostly soluble upon purification, and, unless otherwise stated, their properties were investigated in this work mainly as the respective recombinant MBP fusion protein. The X domain was positioned at the C-terminal end of the MBP carrier protein – similar to its native state.

3.3. X domains fail to function as a dockerin

We have previously hypothesized whether this type of domain would function as an alternative type of dockerin. Since conventional dockerins have been discovered in numerous *R. albus* proteins, a novel type of dockerin might imply two separate types of cellulosome in this bacterium. In order to investigate whether X domains react with other cell components, *R. albus* cell extracts were subjected to SDS–PAGE and Western blotting, using recombinant *E. coli* cell lysates containing either the His-tagged Cel5G-X_1 or Cel5G-X_2 as probes. The former probe reacted with a single polypeptide of ~18 kDa molecular mass. Combined MALDI-TOF and genome sequence analyses served to identify the reactive peptide as a ribosomal protein (L5). The Cel5G-X_2-containing probe failed to react with any *R. albus* 8 protein, and the interaction between the X_1 probe and the L5-like protein does not appear to be of any biological significance. It therefore seems unlikely that any of the X domains function in a manner analogous to that of dockerins and cellulosome formation. This is further

Table 1
Primers used in this study

Name[a]	Nucleotide sequence[b]	Location	Comments
F-X299-1	ATATCCATGGATCCCGAAACACCTGCTGTTGTA	X_1, *cel5G*	C-terminal His-tag, or N-terminal MBP
R-X299-1	CCCCGCGGCCGCACCCGCAACATCTTTCGCCAT	X_1, *cel5G*	C-terminal His-tag, or N-terminal MBP
F-X299-2	ATAT CCATGGATGTTGCGGGTACTGTTACTTAC	X_2, *cel5G*	C-terminal His-tag, or N-terminal MBP
R-X299-2	CCCCGCGGCCGCCTTTACAGTGATAGTCACAGC	X_2, *cel5G*	C-terminal His-tag, or N-terminal MBP
F-EX-X172	ATATCCATGGATCGTTTCGGCGGTTCGAATCCTG	X domain, *cel9C*	C-terminal His-tag, or N-terminal MBP
R-EX-X172	CCCCGCGGCCGCCTTTATAGTAACAGTACAAGCACG	X domain, *cel9C*	C-terminal His-tag, or N-terminal MBP
F-EX-X13	ATATCCATGGATGATAAGACTTATCCTACCAAC	X domain, *cel48A*	C-terminal His-tag, or N-terminal MBP
R-EX-X13	CCCCGCGGCCGCCTTAACTGTAACGTTAACTACAGA	X domain, *cel48A*	C-terminal His-tag, or N-terminal MBP
F-EX-X234	TTTCCATGGGCGGCACTGTATATTCTTCCAAC	X domain, *xyn11C*	C-terminal His-tag
R-EX-X234	GGGCTCGAGCCTTACAGTAACAACTATAGC	X domain, *xyn11C*	C-terminal His-tag

[a] Primers designated: F, forward, R, reverse.
[b] Residues shown in italics were added to the 5′ ends to generate restriction enzyme cleavage sites (underlined).

Table 2
Expressed proteins prepared in this study

Name	Plasmid used	Fused tag or protein	Solubility of recombinant protein
Cel5G-X_1	pET28a	C-terminal His-tag	Partially soluble, precipitated quickly after purification
	pMalC-NN-EGFP	N-terminal MBP	Soluble
Cel5G-X_2	pET28a	C-terminal His-tag	Partially soluble, precipitated quickly after purification
	pMalC-NN-EGFP	N-terminal MBP	Soluble
Cel5G-X_1–X_2	pET28a	C-terminal His-tag	Partially soluble, precipitated quickly after purification
	pMalC-NN-EGFP	N-terminal MBP	Soluble
Cel48A-X	pET28a	C-terminal His-tag	Insoluble
	pMalC-NN-EGFP	N-terminal MBP	Soluble
Cel9C-X	pET28a	C-terminal His-tag	Insoluble
	pMalC-NN-EGFP	N-terminal MBP	Soluble
Xyn11C-X	pET28a	C-terminal His-tag	Mainly insoluble

supported by the relatively small size of the interacting polypeptide, which could contain at most a single cohesin-like module.

3.4. The X domains bind to various polysaccharides

The binding of the respective recombinant MBP-fused X domain to different polysaccharides was investigated. The experimental results showed that all of the X domains, with the notable exception of the internal Cel5G-X_1 domain, can bind to amorphous cellulose, crystalline cellulose, insoluble xylan, chitin and lichenan (illustrated in Fig. 3 and Table 3 for the Cel9C-derived X domain, respectively). The internal Cel5G-X_1 domain failed to bind significantly to any of these polymers (Fig. 4 and Table 3). Two other *R. albus* X domains were selected for further study on natural substrates. The binding of MBP-fused C-terminal X domains from Cel9C and Cel5G-X_2 bound to cell wall preparations derived from wheat straw, alfalfa and banana fruit stem (shown for the X domain of Cel9C in Fig. 3).

3.5. Inhibition of binding by soluble hydroxyethylcellulose

The inhibition of binding of MBP-fused C-terminal Cel5G-X_2 domain to insoluble xylan was investigated using different soluble substrates. The binding to xylan was not inhibited by xylose, cellobiose, ribose, and sucrose (up to 9.0% w/v).

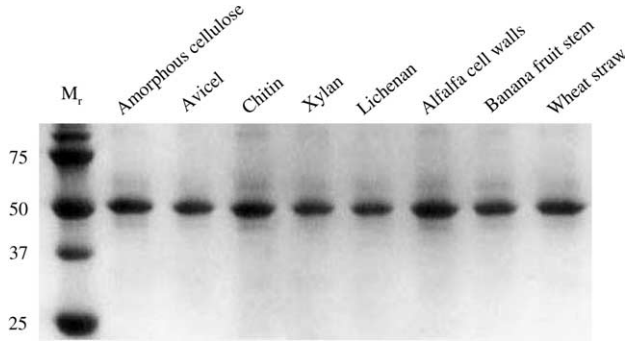

Fig. 3. Binding of X domain to different polysaccharides. The MBP-fused *cel9C*-derived X domain was over-expressed as an MBP fusion protein, and the affinity-purified recombinant protein was examined for its binding to a variety of insoluble polysaccharides as designated in the figure. In addition to purified polysaccharides, the construct also bound to natural polymers, e.g., alfalfa cell walls, banana fruit stem and wheat straw. In each case, the recombinant X domain was first introduced to a suspension of the insoluble substrate, the mixture was washed four times by centrifugation, then the attached protein was removed by introduction of SDS-containing sample buffer and subjected to SDS–PAGE.

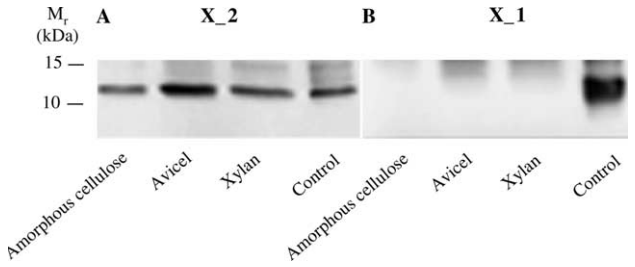

Fig. 4. Binding of C-terminal versus internal X domain to purified polysaccharides. The internal and C-terminal X domains (X_1 and X_2, respectively) of Cel5G were over-expressed separately as the corresponding His-tagged construct. The recombinant proteins were introduced to a suspension of the designated insoluble polysaccharide, washed by centrifugation, and the attached protein was subjected to SDS–PAGE. The C-terminal X_2 domain (panel A) binds to all three polysaccharides tested, whereas the internal X_1 domain (panel B) failed to bind significantly to any of the target polymers. The control lane shows the respective construct prior to addition of polymer.

However, the binding was inhibited clearly by as little as 0.5% hydroxyethylcellulose (Fig. 5).

3.6. Interaction between tandem X domains

Because the two X domains from *R. albus* Cel5G acted differently in polysaccharide-binding assays, the two domains were cloned and expressed together as a dyad, and the binding properties of this recombinant protein were compared to those for the individual domains (Table 3 and Fig. 6). The binding capacity of the Cel5G-X_1-X_2-MBP recombinant protein was 5- to 7-fold higher than that obtained for the Cel5G-X_2-MBP recombinant protein, suggesting that the X_1-X_2 domains interact in a synergistic fashion to promote adhesion of the protein to the polysaccharides examined.

4. Discussion

Many non-catalytic domains from different glycoside hydrolases show striking sequence similarities and can thus be grouped into a family of domains on the basis of homologous primary structures [23,30]. If a function has not been demonstrated for any of the members of the latter group, they are collectively referred to as X domains. More than 100 such X domains have now been described (B. Henrissat, personal communication). The transition of a recognized X domain to a designated classification usually requires biochemical characterization of its function. Typically, the biochemical characterization and clas-

Table 3
Binding capacity of expressed MBP-fused X domains to various polymers[a]

Name	Source	Amorphous cellulose	Avicel	Xylan	Chitin	Lichenan
X_2	Cel5G	2.7 ± 0.6	0.24 ± 0.12	1.2 ± 0.4	0.11 ± 0.08	2.2 ± 0.8
X_1	Cel5G	<0.27	<0.02	<0.12	<0.01	<0.2
[X_1–X_2]	Cel5G	16.2 ± 2.5	2.5 ± 0.5	4.9 ± 1.6	nd[b]	14.7 ± 2.2
X	Cel9C	0.80 ± 0.10	0.60 ± 0.08	0.65 ± 0.09	0.18 ± 0.02	0.70 ± 0.10
X	Cel48A	2.6 ± 0.7	1.5 ± 0.3	2.2 ± 1.2	1.6 ± 0.6	2.1 ± 1.5
Control[c]	MBP	<0.01	<0.01	<0.01	<0.01	<0.01

[a] Values (means ± S.D. from at least three separate experiments) indicate mg of bound protein/100 mg of the designated polymer.
[b] Not determined.
[c] Recombinant MBP without added domain.

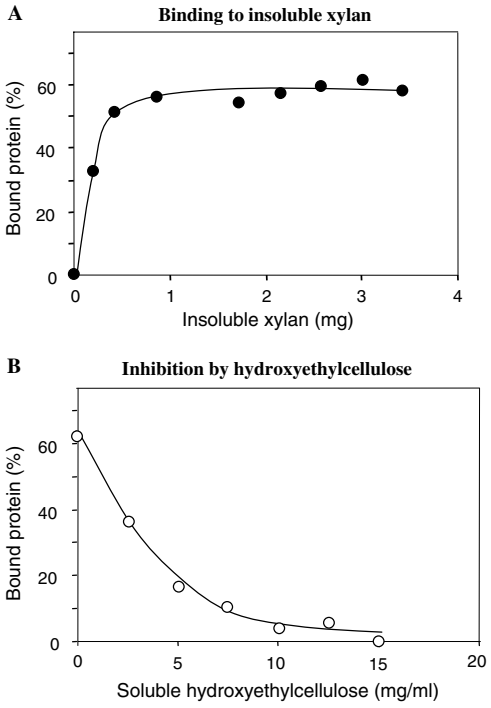

Fig. 5. Binding of X domain to insoluble xylan and selective inhibition by hydroxyethylcellulose. (A) The recombinant C-terminal X_2 domain (MBP-fused, Cel5G-derived) binds to insoluble xylan. (B) The binding of the X domain to insoluble xylan was inhibited by hydroxyethylcellulose.

sification of a new X domain family is based upon the results obtained with one member of the family. In this work, we have cloned and expressed 6 different X domains, from 5 different *R. albus* proteins, belonging to four different families of GHs. In one case, the enzyme contained a tandem repeat of the X-module, in which, upon ClustalW analysis, the internal module (X_1) was found to branch separately at a distance from the remainder. Further genome mining has now revealed that there are seven other instances in the *R. albus* 8 genome of coding sequences containing two, tandemly arranged X domains. All of the 8 internal (X) domains are more similar to each other than to the other 30 or so X domains identified so far (not shown), but a more detailed analysis of both types of domains will await closure and assembly of the genome. In any case, the results presented here suggest that these X domains are widespread throughout the *R. albus* 8 proteome, are unique to this bacterium, and appear to underscore the binding capacity of the protein to a variety of polysaccharides.

The X domains described in this work have been provisionally classified as X94 (B. Henrissat, personal communication). In previous work [20], both Cel48A and Cel9B were shown to be selectively recovered from *R. albus* 8 cell extracts by cellulose-affinity procedures. Accordingly, we hypothesized that the X domains in these proteins may serve in one of the following capacities: (i) they may represent cellulose-binding modules, or (ii) they may coordinate the interaction of Cel9B and Cel48A with a cellulosome-like complex in a manner analogous to a cohesin-dockerin interaction; or (iii) they may coordinate the anchoring of Cel9B and Cel48A to the bacterial cell surface by some other means. The lack of clear binding of the His-tagged Cel5G-X1 or Cel5G-X2 recombinant proteins with any cell-wall associated protein supports the contention that these X domains do not coordinate a dockerin-cohesin like interaction between any glycoside hydrolase of *R. albus* 8 and other surface associated proteins. Conventional types of dockerin sequences have already been observed in numerous *R. albus* proteins, so it is not particularly surprising that X94 would lack such activity. Although the binding of the internal (X_1) domain to the presumptive ribosomal L5 protein suggests that this domain might possibly mediate attachment to the cell surface, it does not explain how the other proteins that lack an internal X_1 domain – notably Cel48A and Cel9B – would be attached to the cell surface.

Although several types of CBM (e.g., from families 3, 4 and 22) have been identified previously in *R. albus* enzymes, we describe here a novel type of CBM that appears to be unique to this species. The *R. albus* X94 domain clearly binds strongly to cellulose and to several other polysaccharides derived from plant and fungal cell walls (and/or arthropodal exoskeleton). The domain thus shows an unusual versatility in its recognition of polysaccharides, although the overall specificity of interaction was demonstrated by inhibition in the presence of a soluble cellulosic derivative. Structurally, the domain appears

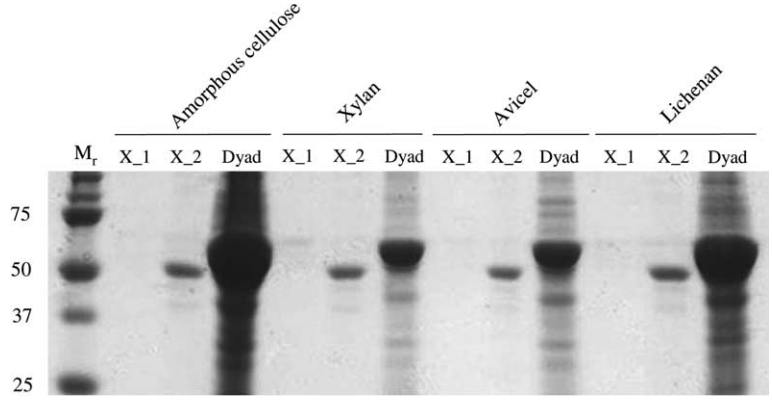

Fig. 6. Binding capacity of single X domains and the tandem dyad to different polymers. The expressed MBP-fusion protein, containing either a single internal X_1 domain, a C-terminal X_2 domain, or both domains in tandem (all derived from Cel5G), was subjected to interaction with the designated insoluble polysaccharide, and the amount of bound protein was assessed by SDS–PAGE. In all cases, the internal X_1 domain displayed little or no binding to the target polymer, whereas the C-terminal X_2 domain bound clearly to all polysaccharides. The tandem dyad (X_1–X_2) exhibited highly enhanced binding (5- to 7-fold) compared to that of the single domain.

to be similar to the Fn3-like domains, since initial BLAST searches revealed a distant similarity to the latter family of domains. Nevertheless, the presence of 3 conserved tryptophans in distinctive positions (not shown) indicated the distinction from the Fn3 domains, as further borne out by mapping on a phylogenetic tree (Fig. 2). Consequently, the uniqueness of this domain and its defined carbohydrate-binding properties allow its classification into a novel family of CBM – hereby designated as family 37 (B. Henrissat, personal communication).

The exact role of the CBM37 in the *R. albus* cellulase system is still unknown. Although its members recognize and bind strongly to cellulose and numerous other polysaccharides, its broad specificity and apparent uniqueness to this species is rather suspicious and could also imply a more innovative function in this particular bacterium. For example, the *R. albus* CBM37 could possibly be involved in recognition and binding to cell-surface polysaccharides, which would explain the original report [20] regarding the cell-associated status of Cel9B and Cel48A. The precise function of the CBM37 in *R. albus* will be the topic of future investigation.

Acknowledgements: This research has been supported by Research Grants 394/03 and 771/01 from the Israel Science Foundation (Jerusalem), US-3106-99C from BARD, the United States-Israel Binational Agricultural Research and Development Fund, and 99-35206-8688 from USDA's NRICGP. The preliminary sequence data used in this study were obtained from The Institute for Genomic Research through the website at http://www.tigr.org. Sequencing of *Ruminococcus albus* was accomplished with support from USDA's Initiative for Future Agriculture and Food Systems 00-52100-9618 made available to the North American Consortium for Genomics of Fibrolytic Ruminal Bacteria.

References

[1] Bayer, E.A., Kenig, R. and Lamed, R. (1983) J. Bacteriol. 156, 818–827.
[2] Lamed, R., Setter, E. and Bayer, E.A. (1983) J. Bacteriol. 156, 828–836.
[3] Lamed, R., Setter, E., Kenig, R. and Bayer, E.A. (1983) Biotechnol. Bioeng. Symp. 13, 163–181.
[4] Lamed, R., Naimark, J., Morgenstern, E. and Bayer, E.A. (1987) J. Bacteriol. 169, 3792–3800.
[5] Ding, S.-Y., Bayer, E.A., Steiner, D., Shoham, Y. and Lamed, R. (1999) J. Bacteriol. 181, 6720–6729.
[6] Ding, S.-Y., Bayer, E.A., Steiner, D., Shoham, Y. and Lamed, R. (2000) J. Bacteriol. 182, 4915–4925.
[7] Ding, S.-Y. et al. (2001) J. Bacteriol. 183, 1945–1953.
[8] Xu, Q., Kenig, R., Shoham, Y., Bayer, E.A. and Lamed, R. (2003) J. Bacteriol. 185, 4548–4557.
[9] Rincon, M.T. et al. (2003) J. Bacteriol. 185, 703–713.
[10] Shoseyov, O., Takagi, M., Goldstein, M.A. and Doi, R.H. (1992) Proc. Natl. Acad. Sci. USA 89, 3483–3487.
[11] Pagès, S., Belaich, A., Fierobe, H.-P., Tardif, C., Gaudin, C. and Belaich, J.-P. (1999) J. Bacteriol. 181, 1801–1810.
[12] Kakiuchi, M. et al. (1998) J. Bacteriol. 180, 4303–4308.
[13] Kirby, J., Martin, J.C., Daniel, A.S. and Flint, H.J. (1997) FEMS Microbiol.. Lett. 149, 213–219.
[14] Doi, R.H., Goldstein, M., Hashida, S., Park, J.S. and Takagi, M. (1994) Crit. Rev. Microbiol. 20, 87–93.
[15] Doi, R.H., Kosugi, A., Murashima, K., Tamaru, Y. and Han, S.O. (2003) J. Bacteriol. 185, 5907–5914.
[16] Belaich, J.-P., Tardif, C., Belaich, A. and Gaudin, C. (1997) J. Biotechnol. 57, 3–14.
[17] Karita, S., Sakka, K. and Ohmiya, K. (1996) J. Ferment. Bioeng. 81, 553–556.
[18] Ohara, H., Noguchi, J., Karita, S., Kimura, T., Sakka, K. and Ohmiya, K. (2000) Biosci. Biotechnol. Biochem. 64, 80–88.
[19] Morrison, M. and Miron, J. (2000) FEMS Microbiol. Lett. 185, 109–115.
[20] Devillard, E., Goodheart, D.E., Karnati, S.K., Bayer, E.A., Lamed, R., Miron, J., Nelson, K.E. and Morrison, M. (2004) J. Bacteriol. 186, 136–145.
[21] Nielsen, H., Brunak, S. and von Heijne, G. (1999) Protein Eng. 12, 3–9.
[22] Bairoch, A. and Apweiler, R. (2000) Nucl. Acids Res. 28, 45–48.
[23] Bourne, Y. and Henrissat, B. (2001) Curr. Opin. Struct. Biol. 11, 593–600.
[24] Coutinho, P.M. and Henrissat, B. (1999) in: Recent Advances in Carbohydrate Bioengineering (Gilbert, H.J., Davies, G.J., Henrissat, B. and Svensson, B., Eds.), pp. 3–12, The Royal Society of Chemistry, Cambridge.
[25] Murray, M.G. and Thompson, W.F. (1980) Nucleic Acids Res. 8, 4321–4325.
[26] Van Soest, P.J., Robertson, J.B. and Lewis, B.A. (1991) J. Dairy Sci. 74, 3583–3597.
[27] Nakamura, M. et al. (2002) Anim. Sci. J. 73, 347–352.
[28] Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A.B., Chet, I., Wilson, K.S. and Vorgias, C.E. (1994) Structure 2, 1169–1180.
[29] Kataeva, I.A., Seidel III, R.D., Shah, A., West, L.T., Li, X.-L. and Ljungdahl, L.G. (2002) Appl. Environ. Microbiol. 68, 4292–4300.
[30] Henrissat, B. and Davies, G.J. (2000) Plant Physiol. 124, 1515–1519.