JOURNAL OF BACTERIOLOGY, Oct. 1992, p. 6707–6710
0021-9193/92/206707-04$02.00/0

Vol. 174, No. 20

Evidence for a Modular Structure of the Homologous Repetitive C-Terminal Carbohydrate-Binding Sites of *Clostridium difficile* Toxins and *Streptococcus mutans* Glucosyltransferases

CHRISTOPH VON EICHEL-STREIBER,[1]* MARKUS SAUERBORN,[1] AND HOWARD K. KURAMITSU[2]

Institut für Medizinische Mikrobiologie, Johannes-Gutenberg-Universität, Hochhaus am Augustusplatz, D-6500 Mainz, Federal Republic of Germany,[1] and Department of Pediatric Dentistry, University of Texas Health Science Center, San Antonio Texas 78284[2]

Received 17 March 1992/Accepted 31 July 1992

The homologous C-terminal repeats of *Clostridium difficile* toxins (ToxA and ToxB) and streptococcal glucosyltransferases appear to mediate protein-carbohydrate interactions at cellular binding sites with sugar moieties as substrates. A consensus sequence of 134 repeating units from gram-positive bacteria indicates that these repeats have a modular design with (i) a stretch of aromatic amino acids proposed to be involved in the primary carbohydrate-protein interaction, (ii) an amplification of this interaction by repetition of the respective sequences, and (iii) a second domain, not characterized, that is responsible for carbohydrate specificity.

Clostridium difficile, *Streptococcus mutans*, and *S. downei* are gram-positive bacteria of the mammalian alimentary canal. The virulence factors ToxA and ToxB of *C. difficile* and the streptococcal glucosyltransferases (GTF) are high-molecular-weight (150- to 310-kDa) extracellular proteins involved in the development of pseudomembranous colitis (10) and dental plaque (9), respectively.

C-terminal repeating units of gram-positive bacteria. In streptococci, the glucan-binding proteins (1) and two types of GTF that mediate the synthesis of water-soluble (GTF-S) and water-insoluble (GTF-I) glucans (12) show mutual homology and characteristically harbor repeating units at their carboxy termini (2, 5, 14, 17). *C. difficile* ToxA and ToxB have 63% amino acid sequence homology, are composed of two domains (18), and similar to GTF, have C-terminal repeating units (18).

(Parts of the data presented herein are contained in the doctoral thesis of M.S.)

Computer analysis of the C-terminal repeats. Computer analysis (with DNASIS and PROSIS [Pharmacia, Freiburg, Federal Republic of Germany] and PC/Gene, version 6.0, Amos Bairoch [Genofit, Heidelberg, Federal Republic of Germany]) predicted amino acid homology between the C termini of the *C. difficile* ToxA, *S. mutans* GTF-I and GTF-SI, and *S. downei* GTF-I and GTF-S (19) proteins. This homology is more extensive between GTF-I and ToxA than between GTF-I and ToxB (see Fig. 1). The homologous sequences map to the C-terminal repeats of the proteins (2, 19).

Involvement of the repeats in carbohydrate-protein interaction. GTF have a polysaccharide-binding domain with known specificity for α-1,6 glucans. Successive deletion of the C-terminal repeats of GTF suggested that the repeats are involved in the polysaccharide-acceptor binding necessary for glucan synthesis (2, 7). The *C. difficile* ToxA repeats are involved in interactions with the oligosaccharide components of receptor molecules on target cells (16). The primary sequences of some ToxA-binding oligosaccharides have been described recently (16). Although experimental evidence for the carbohydrate specificity of ToxB has yet to be presented, it is tempting to speculate that the ToxB C-terminal repeats have a similar function.

Initially, the homology detected between streptococcal and clostridial proteins (GTF and ToxA-ToxB) was surprising. Streptococci are facultative anaerobes, while *C. difficile* is a strictly anaerobic bacterium. Being gram-positive organisms, they have similar membrane structures, and both secrete virulence factors. However, despite the functional differences between these proteins, the homology detected may reflect their common carbohydrate-binding capacities.

A second remarkable homology between the C-terminal repeats of ToxA-ToxB and all GTF proteins was observed at a sequence near the N termini of the latter enzymes (1; Fig. 1). These N-terminal domains could reflect a second carbohydrate-accepting domain of glucosyltransferases or may be associated with the catalytic domain of these enzymes. Sucrose could be this second interacting carbohydrate, since its binding site has been identified at the amino-terminal region of GTF (11). The facts that the glucan-binding protein of *S. mutans* consists essentially of repetitive domains and that a catalytic domain is absent (1) may be taken as a hint for the congruity of the catalytic domain of GTF to the second region of homology.

Two monoclonal antibodies, PCG-4 and TTC8, block the interaction of ToxA with cells (18). These monoclonal antibodies may prove useful as a probe for detecting ToxA-binding sites (19), especially as monoclonal antibody TTC8 did not react either with ToxB or GTF (data not shown). This indicates that the carbohydrate-binding specificities of GTF and ToxA-ToxB are distinct and mediated by different protein sequences.

Consensus sequence of the repeating units. Comparison of as many as 134 distinct direct repeating sequences derived from the appropriate proteins of *C. difficile*, *S. downei*, *S. mutans*, and *S. pneumoniae* (*LytA*) (4), and CPL-1 of bacteriophage *Cp-1* (3) (Table 1) led us to define a consensus sequence for these C-terminal repeats (Fig. 2). Three consecutive aromatic amino acids occur in almost all of these sequences, and all of the sequences were aligned accordingly. As a consequence, the A repeats of GTF, ToxA ALICE-KV, and ToxB BALI-BV were divided into two

* Corresponding author.

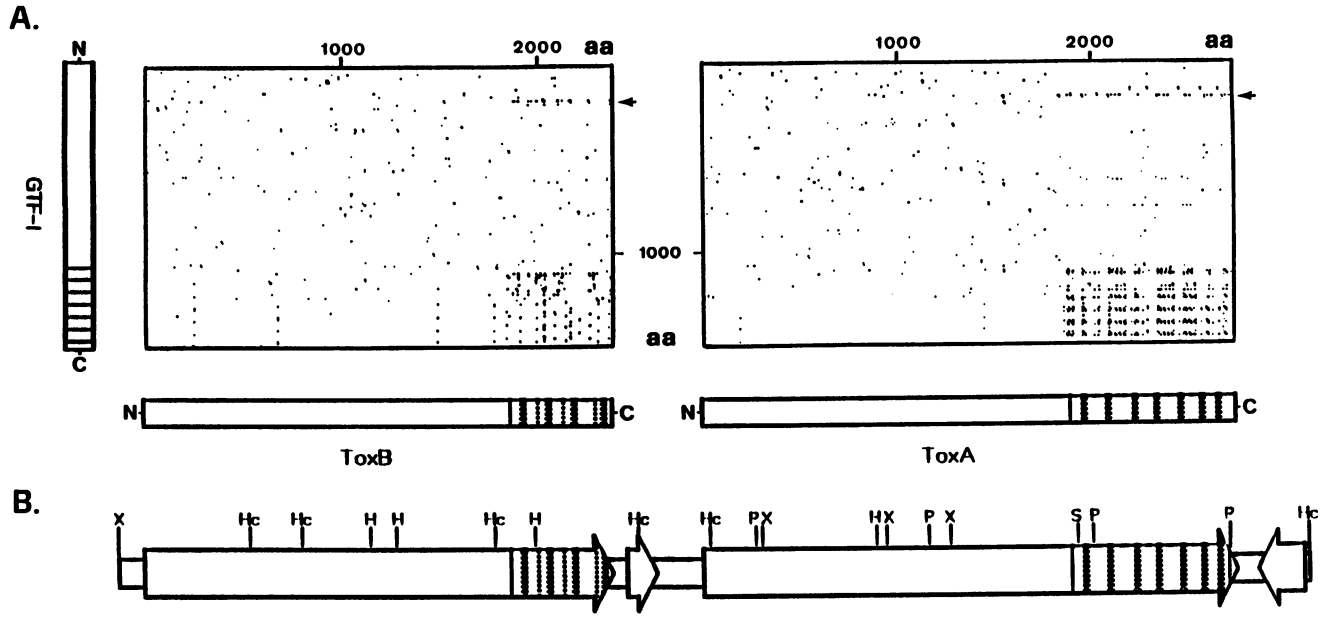

FIG. 1. (A) Dot matrix plot of GTF and ToxA-ToxB. The arrows indicate a second N-terminal homology of GTF-I to ToxA-ToxB. (B) Partial map of the *C. difficile tox* locus. X, *Xba*I; *Hin*cII; H, *Hin*dIII; P, *Pst*I; S, *Spe*I.

separate segments and the repeats were not always listed in accordance with their original identifications (compare references in Table 1). Within the consensus sequence [x(M/A)(V/A)TGw(q/r/k)tIBG(Q/K)(k/w)(Y/F)YFBxB*G], six positions alternate between two and three amino acids, two positions are variant (x), and at one position there may be a one-amino-acid shift (+) (Table 1 and Fig. 2). In addition, computer analysis predicted an antigenic epitope for the amino acid on the N-terminal side of the aromatic amino acid. Recently, the synthetic oligopeptide TIDGKKYYFN was successfully used for induction of an antiserum which reacted with ToxA following Western blot (immunoblot) analysis (20).

The repeating sequence as a general theme for protein-carbohydrate interaction. The distinctive structure of the repeating units, the extent of the repeats, and their position within the proteins make it likely that such C-terminal repeating units are carbohydrate-binding modules. All of the analyzed sequences share a stretch of aromatic amino acids (IDGkwYYFD+N+G; Fig. 2) flanked by additional conserved residues (underlined). As the functional epitopes of these proteins are distinct (i.e., carbohydrate specificity and TTC8 reactivity), this sequence alone cannot directly account for the specificity of carbohydrate-protein binding but may play a more general role in protein-carbohydrate interaction.

Several examples for the role of aromatic amino acids in carbohydrate binding have been previously reported (6, 13, 15, 21). The hydrophobic portions of the interacting carbohydrate species would initially interact with these aromatic residues. Additionally, three of the adjacent conserved amino acids (Asn or Asp, respectively) may each form two hydrogen bonds to different H bond donors or acceptors. Thereby, the required space for entry of the carbohydrates into their binding pockets may be minimized. Lemieux (8) postulated that for binding between protein and carbohydrates, flexible residues are the sites of primary contact which guide the interacting molecules into proper binding positions. In this regard, the flexibility of amino acid chains has been postulated as significant for the carbohydrate-binding domains of wheat germ hemagglutinin and lysozyme (6, 21). Prediction of the secondary structure of the repetitive segments listed in Table 1 suggests that the three aromatic amino acids are located between two potential movable segments (like hinges) bordered by the conserved glycine residues. Thus, the requirements for a potential carbohydrate-binding pocket appear to be possible. However, these computer-based predictions require confirmation by physical characterization of the actual binding domains.

In summary, we propose that the highly homologous sequences may act as the initial attachment sites for carbohydrates and mediate a nonspecific carbohydrate interaction with the proteins. This interpretation is in agreement with

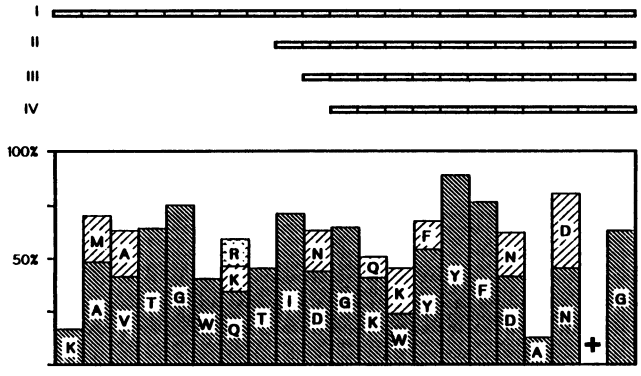

FIG. 2. Conservation of amino acids within the proposed carbohydrate-binding regions. Conservation of amino acids was calculated on the basis of 75 full-length protein sequences (I), 37 half-length sequences (IV), and 3 and 19 sequences of the lengths indicated in II and III, respectively. The bars indicate the relative frequencies of the individual amino acids. For some positions, up to three different residues are listed. +, shift of one amino acid present at high frequency in the aligned sequence (compare Table 1).

TABLE 1. Characterization of the repeating units

Source	Protein	Length (amino acids)	Repeat				Reference
			Designation	Length	Frequency	Amino acid sequence[a]	
S. downei	GTF-I	1,597	A	35	6	GNWYYFGKD-G KAVTGAQTIGGQKLYFDAN-G	2
			B	48	2	QxFNKSVxVNGKTYYFGSD-G	
			C	20	5	DGKVRYFDANSGBMVTNTFV	
	GTF-S	1,365	A	32	6	GNWYYFNxD-G QAATGLQTIDGQTVYFDDN-G	5
			C	21	4	NGKLYYFDADQGxxVxKGWx	
S. mutans	GTF-I	1,475	A	32–33	4.5	GQWFYFDNN-G YAVTGARTINGQHLYFRAN-G	14
			C	20	4	HGRISYYDGNSGDQIRNRFV	
	GTF-SI	1,375	A	32–33	4	GEWFLFDHN-G IAVTGAVTVNGARLYFKPN-G	17
			C	20	4	BGxLRYFDxBSGNQIxNRFV	
	GBP	563	A	32–34	5.5	xKWYYKGAD-G KRVTGWQTIDGKQYYFDQD-G	1
			C	20	2	DGKxYYFDPDSGQxWTNRFV	
C. difficile	ToxA	2,710	ALICE-KV	50 + 21	7	GKKYYFGNN-S KAVTGWQTIBGKKYYFNPN-T	19
			EA	21	6	EAATGWQTIBGKKYYFNTN-T	
			AS	21	7	IASTGYKTIBGKHFYFBTD-G	
			IA	21	3	IAAxGLQTIBNBKYYFBxD-T	
	ToxB	2,366	BALI-BK	49 + 21	4	EKIYYFDDx-Y TAAVGWKxLDGEKYYFDPDTA	18
			BA	21	6	IAxIGLTxIDDKNYYFNEB-G	
			BE	20–22	2	xxxxGxxxxxDxKYYFBxx-G	
			X	19–20	3	INDKxFYFxDG-G	
Phage *Cp-1*	CPL-1	339	P	21	6	AMVTGWVKDGGKWYYMBBN-G	3
S. pneumoniae	LytA	318	P	21	6	AMxTGW-KIDDTWYYFDSD-G	4

[a] The amino acid sequences are based upon, but not congruent with, the repeat sequences defined in the references. Use of capital letters does not imply that each position is necessarily highly conserved. Only if an amino acid is varied in the same position of each sequence of the repeat was it indicated by an x. A dash indicates an amino acid shift. The triplet of aromatic amino acids is in bold letters. Alternation of Asn and Asp at one position is indicated as B.

the ability of antisera against the peptide TIDGKKYYFN to inhibit erythrocyte agglutination and the cytotoxic effects of ToxA (20). Enhancement of attractive forces would be achieved by the existence of several repetitions of these modules. The specificity of carbohydrate binding would be mediated by still undefined sequential and structural features present in different carbohydrate-binding proteins.

This work was supported by grant 01KI8832/3 from the Bundesministerium für Forschung und Technologie (C.v.E.-S.) and by National Institutes of Health grant DE03258 (H.K.K.).

REFERENCES

1. **Banas, J. A., R. R. B. Russell, and J. J. Ferretti.** 1990. Sequence analysis of the gene for the glucan-binding protein of *Streptococcus mutans* Ingbritt. Infect. Immun. **58**:667–673.
2. **Ferretti, J. J., M. L. Glipin, and R. R. B. Russell.** 1987. Nucleotide sequence of a glucosyltransferase gene from *Streptococcus sobrinus* MFe28. J. Bacteriol. **169**:4271–4278.
3. **Garcia, E., J. L. Garcia, P. Garcia, A. Arraras, J. M. Sanchez-Puelles, and R. Lopez.** 1988. Molecular evolution of lytic enzymes of Streptococcus pneumoniae and its bacteriophages. Proc. Natl. Acad. Sci. USA **85**:914–918.
4. **Garcia, P., J. L. Garcia, E. Garcia, and R. Lopez.** 1986. Nucleotide sequence and expression of pneumococcal autolysin gene from its own promoter in *E. coli*. Gene **43**:265–272.
5. **Gilmore, K. S., R. R. B. Russell, and J. J. Ferretti.** 1990. Analysis of the *Streptococcus downei gtfS* gene, which specifies a glucosyltransferase that synthesizes soluble glucans. Infect. Immun. **58**:2452–2458.
6. **Johnson, L. N., J. Cheetham, P. J. McLaughlin, K. R. Acharya, D. Barford, and D. C. Phillips.** 1988. Protein-oligosaccharide interactions: lysozyme, phosphorylase, amylases. Curr. Top. Microbiol. Immunol. **139**:81–134.
7. **Kato, C., and H. K. Kuramitsu.** 1990. Carboxyl-terminal deletion analysis of the *Streptococcus mutans* glucosyltransferase-I enzyme. FEMS Microbiol. Lett. **72**:299–302.
8. **Lemieux, R. U.** 1989. The origin of the specificity in recognition of oligosaccharides by proteins. Chem. Soc. Rev. **18**:347–374.
9. **Loesche, W. J.** 1986. Role of *Streptococcus mutans* in dental decay. Microbiol. Rev. **50**:353–380.
10. **Lyerly, D. M., H. C. Krivan, and T. D. Wilkins.** 1988. *Clostridium difficile*: its disease and toxins. Clin. Microbiol. Rev. **1**:1–18.
11. **Mooser, G., S. A. Hefta, R. J. Paxton, J. E. Shively, and T. D. Lee.** 1991. Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus α-glucosyltransferases. J. Biol. Chem. **226**:8916–8922.
12. **Mukasa, H.** 1986. Properties of *Streptococcus mutans* glucosyltransferases, p. 121–132. *In* S. Hamada, S. M. Michalek, H. Kiyono, L. Menaker, and J. R. McGhee (ed.), Molecular biology and immunobiology of *Streptococcus mutans*. Elsevier

Science Publishers, New York.
13. **Quiocho, F. A.** 1988. Molecular features and basic understanding of protein-carbohydrate interactions: the arabinose-binding protein-sugar complex. Curr. Top. Microbiol. Immunol. **139:** 135–148.
14. **Shiroza, T., S. Ueda, and H. K. Kuramitsu.** 1987. Sequence analysis of the *gtfB* gene from *Streptococcus mutans*. J. Bacteriol. **169:**4263–4270.
15. **Sixma, T. K., S. E. Pronk, D. H. Kalk, B. A. M. van Zanten, A. M. Berghuis, and W. G. J. Hol.** 1992. Lactose binding to heat-labile enterotoxin revealed by X-ray crystallography. Nature (London) **355:**561–564.
16. **Tucker, K. D., and T. D. Wilkins.** 1991. Toxin A of *Clostridium difficile* binds to the human carbohydrate antigens I, X, and Y. Infect. Immun. **59:**73–78.
17. **Ueda, S., T. Shiroza, and H. K. Kuramitsu.** 1988. Sequence analysis of the *gtfC* gene from *Streptococcus mutans* GS-5. Gene **69:**101–109.
18. **von Eichel-Streiber, C., R. Laufenberg-Feldmann, S. Sartingen, J. Schulze, and M. Sauerborn.** 1992. Comparative sequence analysis of the *Clostridium difficile* toxins A and B. Mol. Gen. Genet. **233:**260–268.
19. **von Eichel-Streiber, C., and M. Sauerborn.** 1990. *Clostridium difficile* toxin A carries a C-terminal repetitive structure homologous to the carbohydrate binding region of streptococcal glucosyltransferases. Gene **96:**107–113.
20. **Wren, B. W., R. R. B. Russell, and S. Tabaqchali.** 1991. Antigenic cross-reactivity and functional inhibition by antibodies to *Clostridium difficile* toxin A, *Streptococcus mutans* glucan-binding protein, and a synthetic peptide. Infect. Immun. **59:**3151–3155.
21. **Wright, C. S.** 1990. 2.2Å resolution structure analysis of two refined *N*-acetylneuraminyl-lactose wheat germ agglutinin isolectin complexes. J. Mol. Biol. **215:**635–651.