MICROBIOLOGICAL REVIEWS, June 1991, p. 303–315
0146-0749/91/020303-13$02.00/0

Vol. 55, No. 2

Domains in Microbial β-1,4-Glycanases: Sequence Conservation, Function, and Enzyme Families

N. R. GILKES,[1] B. HENRISSAT,[2] D. G. KILBURN,[1] R. C. MILLER, JR.,[1] AND R. A. J. WARREN[1]*

Department of Microbiology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5,[1] and Centre de Recherches sur les Macromolécules Végétales, Centre National de la Recherche Scientifique, B.P. 53X, 38041 Grenoble, France[2]

INTRODUCTION 303
STRUCTURAL ELEMENTS IN CELLULASES AND XYLANASES 303
Linkers 303
Repeated Sequences 305
Cellulose-Binding Domains 306
Catalytic Domains 308
EVOLUTION OF CELLULASES AND XYLANASES 310
CONCLUSIONS 312
ACKNOWLEDGMENTS 312
REFERENCES 312

INTRODUCTION

Cellulose is a polysaccharide composed of β-D-glucopyranosyl units joined by 1,4-glycosidic bonds. In xylan, the repeating unit is the β-1,4-D-xylopyranosyl residue. Although chemically similar, these two polysaccharides adopt different conformations. Cellulose molecules have a fully extended, flat conformation (37, 64) and are tightly packed into microfibrils to form a fibrous, naturally crystalline, insoluble material. Xylan molecules are twisted and are more flexible than cellulose chains (102), and the backbone is substituted with arabinose, glucuronic acid, or methylglucuronic acid. Chitin resembles cellulose since it is composed of *N*-acetyl-2-amino-2-deoxy-β-D-glucopyranosyl residues joined by 1,4-glycosidic bonds. Like cellulose, chitin chains have an extended conformation and form insoluble and crystalline microfibrils (12). The microbial conversion of cellulose and xylan to soluble products requires several types of enzyme: endoglucanases (1,4-β-D-glucan glucanohydrolase; EC 3.2.1.4), cellobiohydrolases (1,4-β-D-glucan cellobiohydrolase; EC 3.2.1.91), xylanases (1,4-β-D-xylan xylanohydrolase; EC 3.2.1.8), and β-xylosidases (1,4-β-D-xylan xylohydrolase; EC 3.2.1.37). Microorganisms capable of degrading lignocellulose usually produce complex, extracellular cellulase systems comprising combinations of these enzymes (22, 23). Microorganisms that degrade chitin also produce a variety of enzymes. Cellulases, xylanases, and chitinases all hydrolyze β-1,4-glycosidic bonds between pyranose units, but they show subtle differences in specificity. Structural features common to such enzymes may be related to their general catalytic activities as glycosidases. Those peculiar to each may be related to their specificities, i.e., exoglycanase or endoglycanase, cellulase or xylanase. Such features should be reflected in the amino acid sequences of the enzymes.

Gene cloning and DNA sequencing have allowed rapid determination of the amino acid sequences of cellulases and xylanases (1a, 7). Analysis and comparison of the sequences have revealed conserved stretches which are common to both cellulases and xylanases. The conserved sequences occur in discrete domains connected by linkers which allow the domains to function independently (30, 39, 61, 111, 113). The conserved sequences can be used to group the enzymes into families (5, 8, 54, 55, 62). It seems that the various β-1,4-glycanases arose from a few progenitor sequences by mutation and domain shuffling. In this context, it should be noted that some enzymes show a mixed specificity: enzymes that hydrolyze β-1,4 bonds in cellulosic substrates may also hydrolyze xylan, chitin, and related substrates at significant rates (41).

This review summarizes the domain organizations of cellulases and xylanases analyzed to date. It also compares the enzymes with other proteins which interact with various polysaccharides.

STRUCTURAL ELEMENTS IN CELLULASES AND XYLANASES

Linkers

Proteolytic cleavage of cellulases into separate catalytic and cellulose-binding fragments first demonstrated the presence of true domains within these enzymes (42, 111, 113). The primary sites of cleavage in an exoglucanase and an endoglucanase from *Cellulomonas fimi* (42) and in two cellobiohydrolases from *Trichoderma reesei* (111) are within or adjacent to short sequences of amino acids rich in proline or hydroxyamino acids or both. The short sequences appear to be linkers joining discrete catalytic domains and cellulose-binding domains (CBDs) in these enzymes. Similar sequences occur in other cellulases and xylanases (Table 1). The sequences vary considerably in length (6 to 59 amino acids for those reported to date) and in their proline and hydroxyamino acid contents. Some of them contain runs of hydroxyamino acids or consecutive repeats of shorter sequences of amino acids or both. Some are relatively rich in aspartate or glutamate or both. There is some sequence identity between linkers in enzymes from the same organism, but, except for CenA and Cex from *C. fimi* and EngXCA from *Xanthomonas campestris* (44), there is little if

* Corresponding author.

TABLE 1. Amino acid sequences of linker regions in cellulases and other proteins

Organism[a]	Protein	Designation	Linker sequence[b]	No. of: Amino acids	Prolines	Threonines and serines	Reference
A							
Bacillus sp. strain N4	Endoglucanase	CelA	$T_2P_2SDPTP_2SDPDGEPGPDPGEPDPTP_2SDP$	33	15	7	35
Bacillus sp. strain N4	Endoglucanase	CelB	(i) $P_2SDPTP_2SDPDPGEPDPTP_2SDPGEYP$	28	13	5	35
			(ii) $P_2SEPSDP_4SEPE(PDPGE)_3PDPTP_2SDPEYP$	42	20	5	
Bacillus sp. strain 1139	Endoglucanase	Egl	$T_2EPVEPEPVDPGE_2TP_2$	18	6	3	33
Butyrivibrio fibrisolvens	Endoglucanase	EndI	$(PDPTPVD)_4PDPQPVDPTP$	38	17	5	9
Butyrivibrio succinogenes	Xylanase	Xyn	$PGSFTPQPTITPQ(PT)_2PSGQT$	22	6	8	73
Caldocellum saccharolyticum	Bifunctional exo-endoglucanase	CelB	(i) $T_2S_2(PT)_4(VT)_2(PT)_5VTAT(PT)_3PVSTPAT$	43	14	23	100
			(ii) $PAPTMTVAPTAT(PT)_2LSPTV(TP)_2APTQTAI(PT)_2LTPN(PT)_2$	44	14	17	
Cellulomonas fimi	Endoglucanase	CenA	$PT_2S(PT)_4T(PT)_7VTPQPT$	33	14	17	117
Cellulomonas fimi	Endoglucanase	CenB	(i) $PTGT_3DT_2P_2T_2PGTP$	17	5	9	77
			(ii) $T_2DT_2GETEP_2T_2PGTP$	17	4	8	
			(iii) $T_2A_2PVDTVAPTVPGTP$	17	4	5	
			(iv) $S_2PVTFT_2LPVTSTPS$	16	3	8	
Cellulomonas fimi	Endoglucanase	CenC	(i) SLT_2SATP_3	10	3	5	24
			(ii) PVPTAP	6	3	1	
Cellulomonas fimi	Exoglucanase	Cex	$(PT)_3T(PT)_3T(PT)_3S$	21	9	12	83
Cellulomonas flavigena		ORF A[c]	$PDPTD_2PTQDPTD_2PT$	16	5	4	
Cellulomonas flavigena		ORF X	$PDPTDEPTEDPT(DDPT)_5EDPT$	36	10	9	1
Clostridium acetobutylicum	Endoglucanase	Egl	$T_2PTS_3PVYTSPITISKT_3$	21	3	13	123
Clostridium thermocellum	Endoglucanase	CelA	$PLSDLSGQPTP_2SNPTPSLP_2$	21	8	6	6
Clostridium thermocellum	Endoglucanase	CelB	$TPSVT(PS)_2ATPSPT_2ITAP_2T$	22	7	11	45
Clostridium thermocellum	Endoglucanase	CelE	$PLVS(PT)_3LMPTPSPTVT$	20	7	8	49
Clostridium thermocellum	Endoglucanase	CelH	(i) $(PT)_3WTSTP_2S_3P$	16	6	9	120
			(ii) $PGTYPSYSPKPSPTPRPTKP_2VTP$	24	10	7	
Clostridium thermocellum	Xylanase	XynZ	(i) TPVPTPSPKP	10	5	3	46
			(ii) TPNPSVTPTQTPIPT	15	5	6	
Dictyostelium discoideum	SGSP[d]	270-6	$PS_2TSVPT_3PTVTET(PTET)_{17}VT(PT)_2VTPTETPS_2$	99	24	52	43
Dictyostelium discoideum	SGSP	270-11	(i) $PIT_3S_2T_3DGS_2TPSTPTST_2SAST_3SG_2SAT_4GEPITDGSNG_2$ $AS_2T_2GNSGT_2GSAT_4S_4DNSDGSVGTST_3SPAIT_2S_2GSI_2$ $DPTSP_2T_2DS_3NSG_2YGS_4$	118	8	71	43
			(ii) $SDS_3(PT)_2(PTET)_8ET(PT)_2PS_5DVDSGS_3EIET(PT)_2$ $ETDT(PT)_2PS_5E(GS)_2S_2ETQP_2ITP_3T_2GTS$	105	23	56	
Erwinia chrysanthemi	Endoglucanase	CelZ	$T_2DPSTDT_2MTP_2LTNRPQPT$	21	5	9	47
Fibrobacter succinogenes	Endoglucanase	CelC	$PVS_3DMSPTS_2DAVIDPTS_3A_2V_2DPST$	30	4	13	76
Fibrobacter succinogenes	β-Glucanase		$PQS_4APAS_4(VPAS_4)_2AFVP_2S_4$	36	6	20	109
Microbispora bisopora	Endoglucanase	CelA	$P_2TYSPSPTPST(PS)_3QSDPGS(PS)_3$	30	12	14	119
Pseudomonas fluorescens subsp. *cellulosa*	Endoglucanase	EndA	(i) $S_{11}VPVS_7I_2PS_6IQPS_6MPS_8V_2AS_5VS$	59	4	44	48
			(ii) $S_4ASNINS_{12}AIVS_5V_2S_6$	37	0	28	
Pseudomonas fluorescens subsp. *cellulosa*	Endoglucanase	EndB	(i) $S_2APS_2VAS_7V_2S_2TPRS_5VS_3VPGTS_7$	42	3	30	39
			(ii) $STS_3TPLS_6RS_2VAS_4LS_2ATS_3AS_2VS2$	37	1	28	
Pseudomonas fluorescens subsp. *cellulosa*	Xylanase	XynA	(i) $S_3APAS_2VPS_2IAS_3PS_2VAS_2VIS_2MAS_3PVS_4VAS_2TPGS_3$	49	5	28	50
			(ii) $S_6LS_4V_2S_2IRS_9$	26	0	21	
Pseudomonas fluorescens subsp. *cellulosa*	Xylanases	XynB and XynC	(i) $SAT_2S_2VAS_4TPT_2S_4AS_2VAS$	26	1	19	61
			(ii) $SVS_5VQS_9A_2S$	21	0	16	
Streptomyces sp. strain KSM-9	Endoglucanase	CasA	(i) $PRT_2(PT)_2P$	9	4	4	79
			(ii) $PA_2TGA(SP)_2AP_2ASPAPSADS$	22	7	6	
Trichoderma reesei	Endoglucanase	EglI	$P_5AS_2T_2FSTTR_2S_2T_2S_3PSCTQT$	29	5	16	88
Trichoderma reesei	Endoglucanase	EglIII	$PGAT_2IT_2STRP_2SGPT_4RA(TS)_2S_2TP_2TS_2$	34	6	21	98

Trichoderma reesei	Exoglucanase	CbhI	$PG_2NRGT_4R_2PAT_3GS_2PGPTQS$	26	4	11	28
Trichoderma reesei	Exoglucanase	CbhII	$PGA_2S_2TRA_2S_2T_2SRVSPT_2SRS_3ATP_3GST_3RVP_2VG$	44	7	22	110
Xanthomonas campestris pv. *campestris*	Endoglucanase	EngXCA	$T_2(PT)_{11}$	24	11	13	44
B							
Escherichia coli	Outer membrane protein	OmpA	$APV_2(AP)_4$	12	5	0	20
Escherichia coli	Mannose permease	III^{Man}	$KA_2PAPA_4PKAAPTPAKP$	21	6	1	26
Escherichia coli	Pyruvate dehydrogenase	E2	$APA_3PAKQEA_2(AP)_2A_2KAEAPA_3PA_2K$	30	6	0	89
Escherichia coli	Outer membrane protein	TonB	$(EP)_4IPEP_2KEAPV_2IEK(PK)_5P$	33	14	0	27
Human	Immunoglobulin	IgA_1	$PVPSTP_2TPSPSTP_2T$	16	8	7	68
Klebsiella pneumoniae	Pullulanase	PulA	$S_5TS(GSP)_2(GNP)_2(GTP)_2$	18	6	4	60
Rabbit	Myosin light chain		$KPA_5PAPK(AP)_6K$	25	9	0	11
Streptomyces griseus	Ribosomal protein	L12	$(AV)_2AGPA_2G_2APA$	14	2	0	17
Trypanosoma brucei	Procyclin		$(DP)_2(EP)_{22-29}$	48–62	24–31	0	94
C							
Bacillus subtilis		Spo0A	$SGNAS_2VTHRAPS_2QS_2I$	18	1	8	118
Escherichia coli		CheB	$PT_2LKAGPL_2S_2EKL$	15	2	4	118
Klebsiella pneumoniae		NifA	$QAPQ_2SPRIERPRA$	14	3	1	118

[a] A, Cellulases; B, representative proteins with proline-rich linkers; C, representative proteins with Q linkers.
[b] (i) and (ii), etc., are multiple linkers in a single polypeptide.
[c] ORF, Open reading frames in a fragment of DNA encoding CMCase activity.
[d] SGSP, Spore germination-specific polypeptide.

any sequence identity between linkers from different organisms (Table 1).

Spore germination-specific polypeptide 270-11 from the slime mold *Dictyostelium discoideum* contains two sequences, each about 100 amino acids long, which are rich in proline and hydroxyamino acids. The sequences are characterized by contiguous repeats of the tetrapeptide TETP (Table 1) (43). The walls of *D. discoideum* spores contain cellulose, and it is possible that polypeptide 270-11 is involved in cellulose hydrolysis during germination (see below) (43).

It must be emphasized, however, that some cellulases do not contain obvious linker sequences like those given in Table 1. Such enzymes may not lack discrete domains, because CenA from *C. fimi* (102a) and XynA from *Pseudomonas fluorescens* subsp. *cellulosa* (30), from which the linkers have been deleted, have catalytic and cellulose-binding properties similar to those of the wild-type enzymes.

The linker sequences, especially those rich in proline, are similar to proline-rich linkers connecting different functional domains in other proteins (Table 1). Such sequences in immunoglobulin A1 (16), ribosomal protein L12 (17), and the pyruvate dehydrogenase complex (92) form flexible, extended hinge regions between different domains, and those in outer membrane protein TonB (27) and procyclin (94) form extended structures. It seems likely, therefore, that the linker sequences in cellulases form extended, flexible hinges between domains.

Some of the linkers in cellulases are also relatively rich in arginine, glutamine, glutamate, and hydrophobic amino acids. In this they are similar to Q-linkers, a recently proposed class of interdomain linkers found in a number of bacterial regulatory and sensory transduction proteins (118). Q-linkers are rich in glutamine, arginine, glutamate, serine, and proline and contain some hydrophobic amino acids (Table 1).

Repeated Sequences

Repeated sequences ranging in length from 20 to 150 amino acids occur in a number of cellulases (Table 2). The most striking example is a highly conserved sequence of ca. 24 amino acids which occurs twice, in close proximity, in five endoglucanases, a xylanase, and a partial open reading frame from *Clostridium thermocellum* and in an endoglucanase from *Clostridium cellulolyticum*. The identity is >70%. In CelA, CelB, CelCCA, CelD, CelH, and the open reading frame, the repeats are very close to the C termini of the polypeptides; in CelE and XynZ, the repeats are in the middle of the polypeptides (5, 29, 45, 46, 49, 58, 120). In CelA, CelB, CelE, and the open reading frame, the repeats are preceded by linker sequences; in XynZ, the repeat is followed by a linker sequence. The function(s) of the repeated sequences is not known, but they are not required for the catalytic activity of CelE (49) or CelH (120).

CelB of alkalophilic *Bacillus* strain N4 has a repeat of 61 amino acids at its C terminus. The repeats are 28 amino acids apart and are 90% identical. The polypeptide also has two linker sequences, each of which overlaps the N-terminal sequence of a repeat (35). CelB and CelA of strain N4 have very similar sequences throughout, but CelA has only a single copy, again at its C terminus, of the repeated sequence in CelB (35).

Endoglucanase CenB from *C. fimi* contains three contiguous repeats of a sequence of 98 amino acids which are >60% identical and separated by linkers. The repeats join an

TABLE 2. Repeated sequences in cellulases and xylanases

Organism	Enzyme	Size (amino acids)	No. of amino acids in repeat	No. of amino acids separating repeats	Location of repeat (residues)	Adjacent to linker	Required for enzymatic activity	Reference
Bacillus strain N4	CelB	465	61	28	318–464	+[a]	–	35
Cellulomonas fimi	CenB	1,012	98	0	610–905	+[b]	–	77
Cellulomonas fimi	CenC	1,069	(i) 50	0	1–300	+		24
			(ii) 100	0	886–1069	–		
Clostridium cellulolyticum	CelCCA	449	24	7	389–443	–		29
Clostridium thermocellum	CelA	445	24	8	385–440	+		6
Clostridium thermocellum	CelB	531	24	8	470–525	+		45
Clostridium thermocellum	CelD	608	24	12	544–603	–		58
Clostridium thermocellum	CelE	780	24	12	381–440	+	–	49
Clostridium thermocellum	CelH	856	24	16	789–851	–	–	120
Clostridium thermocellum	CelX[c]		22	12		+		49
Clostridium thermocellum	XynZ	809	24	10	402–459	+	–	46
Clostridium stercorarium	CelZ	961	(i) 88	4	651–832	–	–	56
			(ii) 140	234	495–616	–	–	
					850–961	–	–	
Trichoderma reesei	CbhII	447	20	0	43–83[d]		–	110

[a] Linker sequences overlap the N termini of both repeats.
[b] The N termini of all three repeats and the C terminus of the third repeat are linker sequences.
[c] CelX is a partial open reading frame (49).
[d] The repeats are linker sequences.

N-terminal catalytic domain to a C terminal CBD (77). The sequences of the repeats are 50% identical to the sequences of two tandem repeats in chitinase A1 from *Bacillus circulans* which are related to fibronectin type III repeats (115).

The C terminus of Avicelase I from *Clostridium stercorarium* comprises two contiguous repeats of 88 amino acids, each flanked by repeats of 140 amino acids. The two types of repeat are unrelated (56). The sequence of the longer repeat is related to a sequence forming the C termini of several endoglucanases from *Bacillus subtilis* (56, 69).

Endoglucanase CenC from *C. fimi* has two contiguous repeats of a sequence of 150 amino acids at its N terminus and two contiguous repeats of an unrelated sequence of 100 amino acids at its C terminus (24).

The repeated sequences in CelZ form or contain a CBD (56), and the fibronectin type III-like repeats in chitinase A1 may also form part of a chitin-binding site (115). The repeats in CenB and CenC could have a similar function. Alternatively, the fibronectin type III-like repeats in CenB could be involved in protein-protein interactions within the cellulase complex of *C. fimi* because the fibronectin type III repeats themselves are probably involved in protein-protein interactions (96).

EglI, EglIII, CbhI, and CbhII from *T. reesei* and CbhI from *Phanerochaete chrysosporium* contain two short, contiguous conserved sequences, termed A and B. Sequence A is a CBD, and sequence B, which is heavily glycosylated, is a linker (111, 113). They are at the N termini of EglIII and CbhII in the order A-B, but at the C termini of EglI and CbhI in the order B-A (28, 88, 98, 103, 110). In CbhII, the B sequence is repeated, so that the N terminus is ABB′ (Table 1) (110). The A and B sequences are not required for some catalytic activities of CbhI and CbhII (111) or EglIII (106).

Cellulose-Binding Domains

Many cellulases bind to cellulose, but the mechanism and significance of this interaction are unclear. At least some of them comprise discrete catalytic domains and CBDs which retain their functions when separated by proteolysis (38, 42, 52, 56, 75, 78, 85, 106, 111, 113). *N*-Bromosuccinimide-inactivated CbhI of *Aspergillus ficum* still binds to cellulose, indicating that this enzyme also comprises discrete catalytic and binding domains (52). Although these CBDs are not essential for catalytic activity, they do modulate the specific activities of the enzymes on soluble and insoluble cellulosic substrates.

To date only the CBDs of exoglucanase Cex and endoglucanase CenA of *C. fimi* (42) and those of cellobiohydrolases CBHI and CBHII of *T. reesei* (111, 113) have been characterized in any detail.

The CBDs of Cex and CenA of *C. fimi* are 108 and 111 amino acids long, respectively, and their sequences are more than 50% identical (83, 117). Very similar sequences are found in other cellulases and xylanases (Fig. 1). As in Cex and CenA, the sequences are N or C terminal, about 100 amino acids long, connected to the remainders of the polypeptides by linkers, and, except in EndI of *Butyrivibrio fibrisolvens* (9), not required for enzymatic activity (see Table 4). All of them are probably CBDs, although that of EndI of *Butyrivibrio fibrisolvens* may be an exception. Interestingly, it contains more than twice as many charged amino acids as the *C. fimi* CBDs and lacks the conserved cysteines near the N and C termini of the other bacterial CBDs (Fig. 1; see also Table 4).

Striking features of these bacterial CBD sequences are (i) low contents of charged amino acids; (ii) high contents of hydroxyamino acids; and (iii) conserved tryptophan, asparagine, and glycine residues (Fig. 1). Furthermore, there are two cysteines in identical positions close to the N and C termini in all but one of the sequences and hydrophobic residues at 10 conserved sites in all of the sequences (Fig. 1). Tryptophans are conserved in other types of polypeptide which interact with polysaccharides: the discrete starch-binding domains of microbial enzymes which degrade starch (107), the pilus-associated adhesion proteins of various *E. coli* strains (70), and the carbohydrate recognition domains of one class of animal lectins (25). Tryptophans in some of these polypeptides do interact with saccharides (57, 91, 107).

The CBDs of Cex and CenA of *C. fimi* bind the enzymes to cellulose (42), as do similar sequences in an endoglucanase, xylanases XynA and XynB, and an arabinofuranosidase from *P. fluorescens* subsp. *cellulosa* (30, 39, 50, 61). Cex hydrolyzes both cellulose and xylan (41), and the amino acid sequence of the catalytic domain of Cex is similar to sequences in several xylanases (see Table 6). Cellulose is usually associated with hemicelluloses such as xylan when found in nature. It is not surprising that some enzymes hydrolyze both cellulose and xylan and that some xylanases bind to cellulose.

The CBDs of CbhI, CbhII, EglI, and EglIII of *T. reesei* correspond to the conserved A sequences of the enzymes (Table 3) (28, 88, 98, 106, 110, 111, 113). They contain 33 amino acids, in contrast to the approximately 100-amino-acid bacterial CBDs (Fig. 1). They can be N or C terminal (Table 4). The identical residues include four cysteines, which form two disulfide bridges, two glutamines, and four aromatic residues. The CBD of CbhII was prepared by chemical synthesis and shown by two-dimensional nuclear magnetic resonance to be wedge shaped and to contain two disulfide bridges (65). Surprisingly, the catalytic domain of EglIII contains a sequence of ca. 100 amino acids which is very similar to the bacterial CBDs (76a).

Polypeptide 270-11 from *D. discoideum* contains two sequences of about 100 amino acids each, which are similar in sequence to the bacterial CBDs (43, 76a). It is not surprising that such sequences occur in both procaryotic and eucaryotic polypeptides, given the sequence similarities between the catalytic domains of some bacterial and fungal cellulases and between some bacterial cellulases and spore germination-specific polypeptide 270-6 of *D. discoideum* (Table 5).

Proteolytic removal of the N-terminal half of the CBD of CenA of *C. fimi*, which leaves only a single cysteine in the CBD, does not prevent the truncated enzyme from binding to cellulose (40). This suggests that a disulfide bridge between the two cysteines of this CBD is not essential for binding. The sequences of the C-terminal segments of the bacterial CBDs and the *T. reesei* CBDs are not related (Fig. 1; Table 3). The interactions between CBDs and cellulose have not been elucidated. It will be interesting to see whether the bacterial and *T. reesei* CBDs adsorb to cellulose in a similar manner and at the same sites.

Since CBDs in both groups are N or C terminal and are attached to catalytic domains of different specificities, they appear not to be determinants of specificity. An interesting example is provided by the endoglucanases from *C. fimi* and *Microbispora bispora*, which display significant homology both in their catalytic cores and in their binding domains. In the *M. bispora* enzyme the binding domain is at the C terminus; in the *C. fimi* enzyme it is at the N terminus. Since both enzymes are endoglucanases, the location of the binding domain at the N or C terminus of the catalytic domain clearly does not determine the endo versus exo specificity of these enzymes. There are differences in the sequences of the CBDs within each group, however, and further analysis is required to determine their exact contributions to enzyme function. The cellulases of *Clostridium thermocellum* characterized to date apparently lack CBDs, but they form a multienzyme complex, the cellulosome, which binds in toto to cellulose. Cell-associated cellulosomes can bind *Clostridium thermocellum* itself to cellulose (66). Such binding may be mediated by a noncatalytic component of the cellulosome (66). Cellulosomelike enzyme aggregates have not been observed in *T. reesei* and the bacteria which produce en-

FIG. 1. Amino acid sequences of bacterial CBDs. The sequences listed are *Cellulomonas fimi* Cex (83), *C. fimi* CenA (117), *C. fimi* CenB (77), *C. flavigena* CflX (1), *Microbispora bispora* CelA (119), *Pseudomonas fluorescens* subsp. *cellulosa* EndA (48), *P. fluorescens* subsp. *cellulosa* EndB (39), *P. fluorescens* subsp. *cellulosa* XynA (50), *P. fluorescens* subsp. *cellulosa* XynB and XynC (61), and *Butyrivibrio fibrisolvens* End1 (9). Amino acids which are conserved in at least 6 of the 10 sequences are boxed. Numbers refer to amino acid residues at the start of respective lines; all sequences are numbered from Met1 of the leader peptide. Symbols: *, N or C terminus of mature polypeptide; -, gap left to improve alignment.

TABLE 3. Amino acid sequences of fungal CBD

Enzyme[a]	Sequence[b]	Reference
PcCbhI	T v p q W G Q C G G I G Y t G s T T C A S p y T C h v L N P Y Y S Q C y	103
TrCbhI	T Q S H y G Q C G G I G Y S G P T v C A S G T T C Q Y L N P Y Y S Q C L ***	28
TrCbhII	* c s S V W G Q C G G q n W S G P T c C A S G S T C v Y S N D Y Y S Q C L	110
TrEgII	T Q T H W G Q C G G I G Y S G c k T C T S G T T C Q Y S N D Y Y S Q C L ***	88
TrEglIII	* q Q T V W G Q C G G I G W S G P T n C A p G S A C s t L N P Y Y a Q C i	98

[a] Pc, *P. chrysosporium* enzyme; Tr, *T. reesei* enzyme.
[b] Amino acid residues are indicated in the single-letter code. Boldface capital letters indicate conservation; lightface capital letters indicate partial conservation; lowercase letters indicate nonconservation. Symbols: *, N terminus of the mature enzyme; ***, C terminus.

zymes with CBDs. Perhaps these organisms are relatively static and are better served by diffusible enzymes which bind to the substrate than by multienzyme aggregates or adhesion of cells to the substrate or both.

Catalytic Domains

As with CBDs, catalytic domains have been delineated in some cellulases by proteolysis (38, 40, 42, 52, 56, 75, 78, 106, 111, 113). Other cellulases for which discrete domains have not been identified can be truncated by proteolysis without loss of catalytic activity (14, 18, 76, 93, 108). Still others, including two xylanases, can be truncated without loss of catalytic activity by deleting ends of the genes encoding them (32, 46, 48, 50, 69, 72, 120, 123). These enzymes may also have discrete catalytic domains.

More than 60 cellulase and xylanase genes have been sequenced. Other than the amino acid sequences deduced from the nucleotide sequences of the genes, little, if anything, is known about the enzymes encoded by many of them. However, sequence identity between characterized and uncharacterized enzymes is a strong indication of functional domains in the uncharacterized enzymes. Linker sequences join the domains of a number of the characterized enzymes (28, 30, 50, 77, 83, 110, 117, 119). The amino acid sequences between or next to putative linkers in newly sequenced enzymes can be analyzed for similarities to the sequences of known domains in other cellulases and xylanases. However, enzymes such as endoglucanase CelC of *Clostridium thermocellum* (101) and an endoglucanase of *Cellulomonas uda* (81), which contain neither repeated sequences nor putative linkers, may comprise catalytic domains only. Their amino acid sequences are similar to those of the catalytic domains of other cellulases (Table 5).

Cellulases and xylanases can be grouped into families of related enzymes on the basis of amino acid sequence identities in their putative catalytic domains (5, 8, 54, 55, 62). This grouping is confirmed and extended by hydrophobic cluster analysis, which reveals similarities in apparent secondary structures with proteins of very low sequence identity, even when domains are separated by variable segments of widely differing sizes (54, 55). Hydrophobic cluster analysis is especially useful for cellulases and xylanases, with their discrete domains and repeated sequences and linkers of various lengths.

Sequence identity in the catalytic domains of cellulases and xylanases has been reviewed recently (5, 8, 54, 55). The known sequences can be grouped into nine families (Table 5), which are quite distinct (5, 54, 55, 84, 97). Families A, B, F, and H contain fungal and bacterial enzymes. Family E

TABLE 4. Characteristics of CBDs

Organism[a]	Enzyme	Location (terminus)	Required for catalytic activity	No. of:						Reference
				Amino acids	Charged amino acids		Threonines and serines	Tryptophans	Cysteines	
					+	−				
A										
Butyrivibrio fibrisolvens	EglI	C	+	97	6	9	21	4	1	9
Cellulomonas fimi	CenA	N	−	111	3	3	27	6	2	117
Cellulomonas fimi	CenB	C	−	101	2	1	33	5	2	77
Cellulomonas fimi	Cex	C	−	108	3	1	28	5	2	83
Cellulomonas flavigena	ORF X		ND[b]	106	5	3	27	5	2	1
Microbispora bispora	CelA	C	−	104	3	3	26	5	2	119
Pseudomonas fluorescens subsp. *cellulosa*	CelA	C	−	103	6	2	18	5	2	48
Pseudomonas fluorescens subsp. *cellulosa*	CelB	N	−	100	4	4	24	5	2	39
Pseudomonas fluorescens subsp. *cellulosa*	XynA	N	ND	104	4	3	25	5	2	50
Pseudomonas fluorescens subsp. *cellulosa*	XynB/XynC[c]	N	−	97	4	5	23	5	2	61
B										
Phanerochaete chrysosporium	CbhI	C	ND	33	1	0	8	1	4	103
Trichoderma reesei	CbhI	C	−	33	1	0	6	0	4	28
Trichoderma reesei	CbhII	N	−	33	0	1	9	2	4	110
Trichoderma reesei	EglI	C	ND	33	2	1	8	1	4	88
Trichoderma reesei	EglIII	N	−	33	0	0	5	2	4	98

[a] A, Bacterial enzymes; B, fungal enzymes.
[b] ND, Not determined.
[c] Identical sequences in two enzymes.

TABLE 5. Families of cellulase and xylanase catalytic domains

Family	Organism	Enzyme	No. of amino acids	Catalytic domain[a]		Reference
				Terminus	No. of amino acids	
A						
1	*Bacillus* sp. strain 1139	Egl	770	N	385	33
2	*Bacillus* sp. strain KSM-635	Egl	912		≤555	86
3	*Bacillus* sp. strain N-4 gene pNK1	CelB	465	N	307	35
4	*Bacillus* sp. strain N-4 gene pNK2	CelA	384	N	305	35
5	*Bacillus* sp. strain N-4 gene pNK3	CelC	>800	N	~350	34
6	*Bacillus lautus*	CelB	536			59
7	*Bacillus polymyxa*	Egl	365			4
8	*Bacillus subtilis* N-24	Egl	463			80
9	*Bacteroides ruminicola*	Egl	>363	N		74
10	*Butyrivibrio fibrisolvens* A46	CelA	396			53
11	*Butyrivibrio fibrisolvens* H17c	End1	521	N	~385	9
12	*Caldocellum saccharolyticum*	CelB[b]	1,011	C	388	100
13	*Clostridium acetobutylicum*	Egl	409	N	~300	123
14	*Clostridium cellulolyticum*	CelA	449	N	~380	29
15	*Clostridium thermocellum*	CelB	528	N	>469	45
16	*Clostridium thermocellum*	CelC	322			101
17	*Clostridium thermocellum*	CelE	780	N	340	49
18	*Clostridium thermocellum*	CelH	~860	C	~305	120
19	*Erwinia chrysanthemi*	CelZ	385	N	305	47
20	*Fibrobacter succinogenes*	Egl3	635	C	~416	76
21	*Robillarda* sp. strain Y-20	Egl	375			122
22	*Ruminococcus albus* F-40	Egl1	363			82
23	*Ruminococcus albus* SY3	CelA	~365			90
24	*Ruminococcus albus* SY3	CelB	~385			90
25	*Trichoderma reesei*	EglIII	397	C	327	98
26	*Xanthomonas campestris*	EngXCA	468	N	~350	44
B						
1	*Cellulomonas fimi*	CenA	418	C	284	117
2	*Microbispora bispora*	CelA	426	N	~290	119
3	*Streptomyces* sp. strain KSM-9	CasA	318			79
4	*Trichoderma reesei*	CbhII	447	C	385	110
C						
1	*Humicola grisea*	CbhI	506			2
2	*Phanerochaete chrysoporium*	CbhI	~495	N	~425	103
3	*Trichoderma reesei*	CbhI	~495	N	~425	28
4	*Trichoderma reesei*	EglI	431	N	363	88
5	*Trichoderma viride*	Cbh	496	N	~435	21
D						
1	*Bacillus circulans*	Bgc	~378			15
2	*Cellulomonas uda*	Egl	336			81
3	*Clostridium thermocellum*	CelA	445	N	>384	6
4	*Erwinia chrysanthemi*	CelY				46a
E						
1	*Butyrivibrio fibrisolvens*	CedI	547			10
2	*Cellulomonas fimi*	CenB	1,012	N	607	77
3	*Cellulomonas fimi*	CenC	1,069	Internal	589	24
4	*Clostridium thermocellum*	CelD	608	N	>543	58
5	*Clostridium stercorarium*	CelZ	961	N	474	56
6	*Dictyostelium discoideum*	SGSP270-6	705	N	~450	43
7	*Persea americana*	Egl	469			112
8	*Persea americana*	Cell	484			19
9	*Persea americana*	Cel2				19
10	*Pseudomonas fluorescens* subsp. *cellulosa*	Egl	930			48
F						
1	*Bacillus* sp. strain C-125	XynA	396			51
2	*Butyrivibrio fibrisolvens*	XynA	378	N	~350	73
3	*Caldocellum saccharolyticum*	CelB[b]	1,011	N	347	100
4	*Caldocellum saccharolyticum*	XynA	312			71
5	*Caldocellum saccharolyticum*	ORF 4	312			71
6	*Cellulomonas fimi*	Cex	443	N	315	83
7	*Clostridium thermocellum*	XynZ	809	C	~350	46

Continued on following page

TABLE 5—*Continued*

Family	Organism	Enzyme	No. of amino acids	Catalytic domain[a]		Reference
				Terminus	No. of amino acids	
8	*Cryptococcus albidus*	Xyn	311			13
9	*Pseudomonas fluorescens* subsp. *cellulosa*	XynA	585	C	345	50
10	*Pseudomonas fluorescens* subsp. *cellulosa*	XynB	555	C	272	61
11	*Thermoascus aurantiacus*	Xyn	269			105
G						
1	*Bacillus circulans*	Xyn	185			121
2	*Bacillus pumilis*	XynA	201			36
3	*Bacillus subtilis*	Xyn	182			87
4	*Clostridium acetobutylicum*	XynB	234			124
H						
1	*Aspergillus aculeatus*	Egl	237			84
2	*Erwinia carotovora*	CelS	232			97
I						
1	*Ruminococcus flavefaciens*	CelA	352			114

[a] Deduced from positions of putative linkers, sequence comparison, and truncation experiments.
[b] Catalytic domain 1 (family F, no. 3) catalytic domain 2 (family A, no. 12) of the bifunctional cellulase of *Caldocellum saccharolyticum*.

contains bacterial enzymes and plant enzymes, thereby raising the possibility of a lateral transfer. It includes spore germination-specific polypeptide 270-6 from *D. discoideum*, whose spore coats contain cellulose. Polypeptides 270-6 and 270-11 could be involved in cellulose hydrolysis during spore germination (43). The avocado cellulases in family E appear to be involved in fruit ripening (19). At present, family C contains only fungal enzymes and families D and G contain only bacterial enzymes. Cellulases and xylanases vary widely in the numbers of amino acids they contain, but their catalytic domains tend to be more uniform in size (Table 5). It should be noted, however, that relatively few catalytic domains have been identified other than by sequence relatedness to known domains and the presence of adjacent linkers.

All enzymes reported to have exoglycosidase activity fall into families which have members with only endoglycosidase activity (Table 5). In other words, enzymes with similar sequences have different specificities. This suggests that exoglycosidase versus endoglycosidase activity may be a consequence of fine details of three-dimensional structure rather than of overall conformation. The only catalytic domain for which the three-dimensional structure is known is that of cellobiohydrolase II from *T. reesei* (95). The active site is in an enclosed tunnel through which a cellulose molecule threads. Two aspartyl residues located in the middle of the tunnel may be catalytic residues (95). CbhII is in family B of cellulases and xylanases, which contains exoglucanases and endoglucanases (5, 54). Modeling of other enzymes in the family, with the structure of CbhII as a guide, should give useful insights into their possible structures. Realistic comparisons, however, will require determination of the three-dimensional structures of other enzymes.

Hydrolysis of the β-1,4-glycosidic bond with retention or inversion of anomeric configuration may be a better indicator of similarity than enzyme specificity (116). Hydrolysis with retention of configuration requires a quite different mechanism than does hydrolysis with inversion (104). CenA of *C. fimi* and CbhII of *T. reesei* are in family B, and both cause inversion of configuration (63, 116). CenB of *C. fimi*, which is in family E, also hydrolyzes with inversion of configuration (77). The other enzymes which have been characterized, Cex of *C. fimi* and CbhI of *T. reesei*, hydrolyze with retention of configuration (63, 116) but are in families F and C, respectively.

At least one cellulase has two catalytic domains. CelB of *Caldocellum saccharolyticum* has an N-terminal exoglucanase domain and a C-terminal endoglucanase domain which belong to different families. A linker connects each catalytic domain to a central amino acid sequence of unknown function (99, 100) which is related to a sequence found at the C termini of several endoglucanases from *B. subtilis* (72, 80) and within avicelase I from *Clostridium stercorarium* (56) and endoglucanase CenB from *C. fimi* (76a). XynZ of *Clostridium thermocellum* contains a centrally located repeated sequence flanked by linkers, with a xylanase catalytic domain at the C terminus of the polypeptide and a sequence of 401 amino acids of unknown function at the N terminus (46).

All CBDs described to date are N or C terminal. Catalytic domains from each family are found in various combinations with other conserved sequences, such as CBDs and repeated sequences. This gives rise to a number of different types of primary structures in cellulases and xylanases. It is possible that further families remain to be identified.

Hydrolysis by glycosyl hydrolases often involves general acid catalysis, usually promoted by aspartate or glutamate residues or both. Active-site residues are usually highly conserved during evolution. The catalytic domains of cellulases and xylanases have been analyzed for conserved aspartates and glutamates in an attempt to target catalytic residues in the families (54, 55). In family A, two particular residues emerged as candidates. Site-directed mutations in two of the enzymes in family A support the involvement of the targeted residues in the active site (3, 90a).

EVOLUTION OF CELLULASES AND XYLANASES

It is obvious that cellulases and xylanases evolved by domain shuffling, with subsequent modifications of the domains. This is well illustrated by the fact that catalytic domains from different families are associated with the same

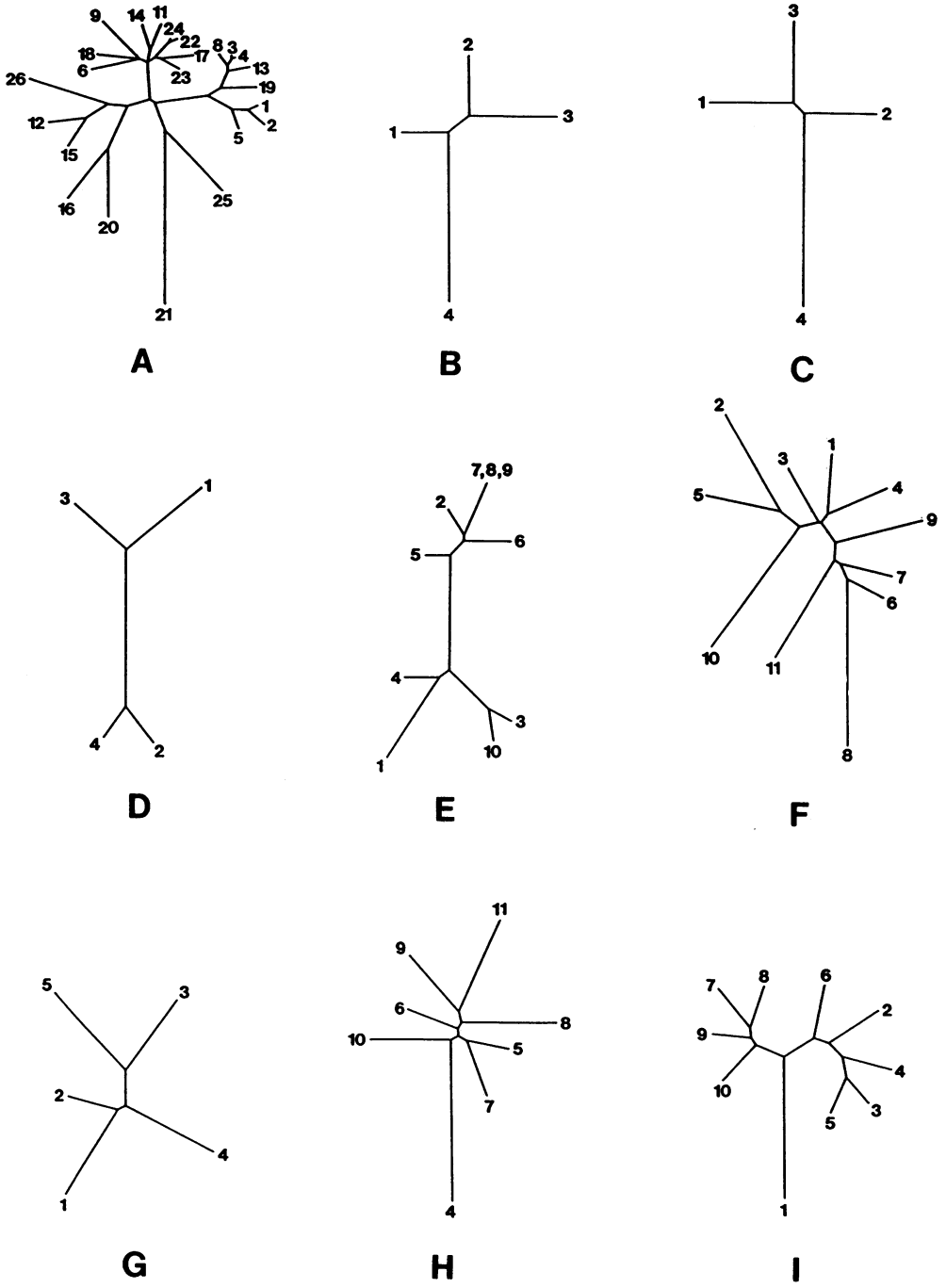

FIG. 2. Unrooted phylogenetic trees for the various related sequences and families of domains. Trees A through F, respectively, are for the families A through F of catalytic domains in Table 5; the enzymes are designated by the numbers in the table. Tree G is for the CBDs of the *T. reesei* enzymes in Table 3: 1, PcCbhI; 2, TrCbhI; 3, TrCbhII; 4, TrEgII; and 5, TrEgIII. Tree H is for the terminal domains with repeated sequences in the *Clostridium* enzymes in Table 2: 4, CelCCA; 5, CelA; 6, CelB; 7, CelD; 8, CelF; 9, CelH; 10, CelX; and 11, XynZ. Tree I is for the bacterial CBDs in Fig. 1: 1, End1; 2, CenA; 3, CenB; 4, Cex; 5, CelA; 6,CflX; 7, EndA; 8, EndB; 9, XynA; and 10, XynB/C.

type of CBD (Fig. 1; Table 5). Although at first sight the catalytic domain families B, C, and F appear to contain enzymes of different types, it must be emphasized again that the specificities of many cellulases and xylanases are not absolute.

It is striking that a given organism can possess enzymes from several families. *Clostridium thermocellum*, for example, has enzymes with catalytic domains from four of the six families (Table 5). In contrast, only one type of CBD has been found in a given organism to date, and there appear to be fewer CBD families than catalytic domain families. However, relatively few CBDs have been identified as such. Amino acid sequences allowing binding may be more constrained than those with catalytic activity because the latter

confer the subtle differences in specificity and mechanism within this group of enzymes.

Unrooted phylogenetic trees were computed (67) for the various families of domains (Fig. 2). Given the complexity of some of these trees, it will be interesting to see whether all members of a given family of catalytic domains do indeed hydrolyze with inversion or retention of anomeric configuration.

The diversity of the linker sequences stands in contrast to the families of related domains. Presumably, the linkers serve to optimize the activities or roles of the domains they join. Some of them may participate in enzyme-enzyme interaction, which could explain their varied compositions and lengths. Repeated copies of sequences such as $PX_2\alpha X_2LX_2LX_2LXLX_2NX\alpha X\alpha$ (where α is M, I, L, or V) are thought to participate in protein-protein interactions (31).

CONCLUSIONS

Microbial cellulases and xylanases comprise various combinations of discrete functional elements: catalytic domains, CBDs, linkers connecting such domains, and repeated sequences of amino acids. The enzymes can be grouped into families on the basis of conserved amino acid sequences in the catalytic domains and by hydrophobic cluster analysis. There are conserved sequences in some of the other elements, especially the CBDs, which are present in enzymes from different catalytic domain families. The enzymes appear to have arisen from a limited number of progenitor sequences by fusion or shuffling, or both, of domains. The binding domains have some features in common with other proteins that interact with polysaccharides, such as lectins, chitinases, and amylases.

Knowledge of the mechanisms of action and of the three-dimensional structures of microbial β-1,4-glycanases is needed to corroborate and extend the conclusions drawn from analysis of the amino acid sequences of these enzymes. Catalytic domains within a family would be expected to have similar conformations and mechanisms of action. For example, they should all hydrolyze the glycosidic bond with retention or inversion of configuration. The use of site-directed mutagenesis to change the conserved amino acids of the CBDs, especially the aromatic residues, could give critical insights into the ways in which the enzymes interact with cellulose.

ACKNOWLEDGMENTS

N.R.G., D.G.K., R.C.M., and R.A.J.W. thank the Natural Sciences and Engineering Research Council of Canada for support. B.H. thanks S. Altschul for providing the MSA program.

We are greatly indebted to Carrie Hirsch for typing the manuscript with care and precision.

REFERENCES

1. **Al-Tawheed, A. R.** 1988. M.Sc. thesis. Trinity College, Dublin, Ireland.
1a. **Aubert, J.-P., P. Béguin, and J. Millet (ed.).** 1988. Biochemistry and genetics of cellulose degradation. FEMS Symp. **43**:1–428.
2. **Azevedo, M. D., M. S. S. Felipe, S. Astolfi-Filho, and A. Radford.** 1990. Cloning, sequencing and homologies of the *cbh*-1 (exoglucanase) gene of *Humicola grisea* var. *thermoidea*. J. Gen. Microbiol. **136**:2569–2576.
3. **Baird, S. D., M. A. Hefford, D. A. Johnson, W. L. Sung, M. Yaguchi, and V. Seligy.** 1990. The glu residue in the conserved asn-glu-pro sequence of two highly divergent endo-β-1,4-glucanases is essential for enzymatic activity. Biochem. Biophys. Res. Commun. **169**:1035–1039.
4. **Baird, S. D., D. A. Johnson, and V. Seligy.** 1990. Molecular cloning, expression, and characterization of endo-β-1,4-glucanase genes from *Bacillus polymyxa* and *Bacillus circulans*. J. Bacteriol. **172**:1576–1586.
5. **Béguin, P.** 1990. Molecular biology of cellulose degradation. Annu. Rev. Microbiol. **44**:219–248.
6. **Béguin, P., P. Cornet, and J.-P. Aubert.** 1985. Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. J. Bacteriol. **162**:102–105.
7. **Béguin, P., N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., G. P. O'Neill, and R. A. J. Warren.** 1987. Cloning of cellulase genes. Crit. Rev. Biotechnol. **6**:129–162.
8. **Béguin, P., J. Millet, S. Chauvaux, E. Yagüe, P. Tomme, and J.-P. Aubert.** 1989. Genetics of bacterial cellulases, p. 57–72. *In* M. P. Coughlan (ed.), Enzyme systems for lignocellulose degradation. Elsevier Applied Science, London.
9. **Berger, E., W. A. Jones, D. T. Jones, and D. R. Woods.** 1989. Cloning and sequencing of an endoglucanase (*end*1) gene from *Butyrivibrio fibrisolvens* H17c. Mol. Gen. Genet. **219**:193–198.
10. **Berger, E., W. A. Jones, D. T. Jones, and D. R. Woods.** 1990. Sequencing and expression of a cellodextrinase (*ced*1) gene from *Butyrivibrio fibrosolvens* H17c cloned in *Escherichia coli*. Mol. Gen. Genet. **223**:310–318.
11. **Bhandari, D. G., B. A. Levine, I. P. Trayer, and M. E. Yeadon.** 1986. ^{1}H-NMR study of mobility and conformational constraints within the proline-rich N-terminal of the LC1 alkali light chain of skeletal myosin. Correlation with similar segments in other protein systems. Eur. J. Biochem. **160**:349–356.
12. **Blackwell, J.** 1982. The macromolecular organization of cellulose and chitin, p. 403–428. *In* R. M. Brown, Jr. (ed.), Cellulose and other natural polymer systems. Plenum Press, New York.
13. **Boucher, F., R. Morosoli, and S. Durand.** 1988. Complete nucleotide sequence of the xylanase gene from the yeast *Cryptococcus albidus*. Nucleic Acids Res. **16**:9874.
14. **Boyer, M. H., J. P. Chambost, M. Magnan, and J. Caltaneo.** 1984. Carboxymethylcellulase from *Erwinia chrysanthemi*. II. Purification and partial characterization of an endo-β-1,4-glucanase. J. Biotechnol. **1**:241–252.
15. **Bueno, A., C. R. Vazquez de Aldana, J. Correa, and F. del Rey.** 1990. Nucleotide sequence of 1,3-1,4-β-glucanase-encoding gene in *Bacillus circulans* WL-12. Nucleic Acids Res. **18**:4248.
16. **Burton, J., S. G. Wood, A. Pedyczak, and I. Z. Siemion.** 1989. Conformational preferences of sequential fragments of the hinge region of human IgA_1 immunoglobulin molecule. II. Biophys. Chem. **33**:39–45.
17. **Bushuev, V. N., A. T. Gudkov, A. Liljas, and N. F. Sepetov.** 1989. The flexible region of protein L12 from bacterial ribosomes studied by nuclear magnetic resonance. J. Biol. Chem. **264**:4498–4505.
18. **Calza, R. E., D. C. Irwin, and D. B. Wilson.** 1985. Purification and characterization of two β-1,4-endoglucanases from *Thermomonospora fusca*. Biochemistry **24**:7797–7804.
19. **Cass, L. G., K. A. Kirven, and R. E. Christoffersen.** 1990. Isolation and characterization of a cellulase gene family member expressed during avocado fruit ripening. Mol. Gen. Genet. **223**:76–86.
20. **Chen, R., W. Schmidmayr, C. Kramer, U. Chen-Schmeisser, and U. Henning.** 1980. Primary structure of a major outer membrane protein II* (*ompA* protein) of *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **77**:4592–4596.
21. **Cheng, C., N. Tsukagoshi, and S. Udaka.** 1990. Nucleotide sequence of the cellobiohydrolase gene from *Trichoderma viride*. Nucleic Acids Res. **18**:5559.
22. **Coughlan, M. P.** 1985. The properties of fungal and bacterial cellulases with comment on their production and application. Biotechnol. Genet. Eng. Rev. **3**:39–109.
23. **Coughlan, M. P., and L. G. Ljungdahl.** 1988. Comparative biochemistry of fungal and bacterial cellulolytic systems. FEMS Symp. **43**:11–30.
24. **Coutinho, J. B., B. Moser, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1991. Mol. Microbiol., in press.

25. **Drickamer, K.** 1988. Two distinct classes of carbohydrate-recognition domains in animal lectins. J. Biol. Chem. **263:** 9557–9560.
26. **Erni, B., B. Zanolari, P. Graff, and H. P. Kocher.** 1989. Mannose permease of *Escherichia coli*. Domain structure and function of the phosphorylating subunit. J. Biol. Chem. **264:** 18733–18741.
27. **Evans, J. S., B. A. Levine, I. P. Trayer, C. J. Dorman, and C. F. Higgins.** 1986. Sequence-imposed structural constraints in the TonB protein of *E. coli*. FEBS Lett. **208:**211–216.
28. **Fagerstam, L. G., G. Pettersson, and J. A. Engstrom.** 1984. The primary structure of a 1,4-β-glucan cellobiohydrolase from the fungus *Trichoderma reesei* QM9414. FEBS Lett. **167:**309–315.
29. **Faure, E., A. Belaich, C. Bagnara, C. Gaudin, and J.-P. Belaich.** 1990. Sequence analysis of the *Clostridium cellulolyticum cel*CCA endoglucanase gene. Gene **65:**51–58.
30. **Ferreira, L. M. A., A. J. Durrant, J. Hall, G. P. Hazlewood, and H. J. Gilbert.** 1990. Spatial separation of protein domains is not necessary for catalytic activity or substrate binding in a xylanase. Biochem. J. **269:**261–264.
31. **Field, J., H.-P. Xu, T. Michaeli, R. Ballester, P. Sass, M. Wigler, and J. Colicelli.** 1990. Mutations of the adenyl cyclase gene that block RAS function in *Saccharomyces cerevisiae*. Science **247:**464–467.
32. **Fukumori, F., T. Kudo, and K. Horikoshi.** 1987. Truncation analysis of an alkaline cellulase from an alkalophilic *Bacillus* species. FEMS Microbiol. Lett. **40:**311–314.
33. **Fukumori, F., T. Kudo, Y. Narahashi, and K. Horikoshi.** 1986. Molecular cloning and nucleotide sequence of the alkaline cellulase gene from the alkalophilic *Bacillus* sp. strain 1139. J. Gen. Microbiol. **132:**2329–2335.
34. **Fukumori, F., T. Kudo, N. Sashihara, Y. Nagata, K. Ito, and K. Horikoshi.** 1989. The third cellulase of alkalophilic *Bacillus* sp. strain N-4: evolutionary relationships within the *cel* gene family. Gene **76:**289–298.
35. **Fukumori, F., N. Sashihara, T. Kudo, and K. Horikoshi.** 1986. Nucleotide sequences of two cellulase genes from alkalophilic *Bacillus* sp. strain N-4 and their strong homology. J. Bacteriol. **168:**479–485.
36. **Fukusaki, E., W. Panbangred, W. Shinmyo, and H. Okada.** 1984. The complete nucleotide sequence of the xylanase gene (*xyn*A) of *Bacillus pumilis*. FEBS Lett. **171:**197–201.
37. **Gardner, K. H., and J. Blackwell.** 1974. The structure of native cellulose. Biopolymers **13:**1975–2001.
38. **Ghangas, G. S., and D. B. Wilson.** 1988. Cloning of the *Thermomonospora fusca* endoglucanase E2 gene in *Streptomyces lividans*: affinity purification and functional domains of the cloned gene product. Appl. Environ. Microbiol. **54:**2521–2526.
39. **Gilbert, H. J., J. Hall, G. P. Hazlewood, and L. M. A. Ferreira.** 1990. The N-terminal region of an endoglucanase from *Pseudomonas fluorescens* subspecies *cellulosa* constitutes a cellulose-binding domain that is distinct from the catalytic centre. Mol. Microbiol. **4:**759–767.
40. **Gilkes, N. R., D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1989. Structural and functional analysis of a bacterial cellulase by proteolysis. J. Biol. Chem. **264:**17802–17808.
41. **Gilkes, N. R., M. L. Langsford, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1984. Mode of action and substrate specificities of cellulases from cloned bacterial genes. J. Biol. Chem. **259:**10455–10459.
42. **Gilkes, N. R., R. A. J. Warren, R. C. Miller, Jr., and D. G. Kilburn.** 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. **263:**10401–10407.
43. **Giorda, R., T. Ohmachi, D. R. Shaw, and H. L. Ennis.** 1990. A shared internal threonine-glutamic acid-threonine-proline repeat defines a family of *Dictyostelium discoideum* spore germination specific proteins. Biochemistry **29:**7264–7269.
44. **Gough, C. L., J. M. Dow, J. Keen, B. Henrissat, and M. J. Daniels.** 1990. Nucleotide sequence of the gene encoding the major endoglucanase of *Xanthomonas campestris* pv. *campestris*. Gene **89:**53–59.
45. **Grépinet, O., and P. Béguin.** 1986. Sequence of the cellulase gene of *Clostridium thermocellum* coding for endoglucanase B. Nucleic Acids Res. **14:**1791–1799.
46. **Grépinet, O., M.-C. Chebrou, and P. Béguin.** 1988. Nucleotide sequence and deletion analysis of the xylanase gene (*xynZ*) of *Clostridium thermocellum*. J. Bacteriol. **170:**4582–4588.
46a. **Guiseppi, A.** 1988. Ph.D. thesis. Université d'Aix-Marseille I, Aix-Marseille, France.
47. **Guiseppi, A., B. Cami, J.-L. Aymeric, G. Ball, and N. Cruezet.** 1988. Homology between endoglucanase Z of *Erwinia chrysanthemi* and endoglucanases of *Bacillus subtilis* and alkalophilic *Bacillus*. Mol. Microbiol. **2:**159–164.
48. **Hall, J., and H. J. Gilbert.** 1988. The nucleotide sequence of a carboxymethylcellulase gene from *Pseudomonas fluorescens* subsp. *cellulosa*. Mol. Gen. Genet. **213:**112–117.
49. **Hall, J., G. P. Hazlewood, P. J. Barker, and H. J. Gilbert.** 1988. Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. Gene **69:**29–38.
50. **Hall, J., G. P. Hazlewood, N. S. Huskisson, A. J. Durrant, and H. J. Gilbert.** 1989. Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subspecies *cellulosa*: internal signal sequence and unusual protein processing. Mol. Microbiol. **3:**1211–1219.
51. **Hammamoto, T., H. Honda, T. Kudo, and K. Horikoshi.** 1987. Nucleotide sequence of the xylanase A gene of alkalophilic *Bacillus* sp. C-125. Agric. Biol. Chem. **51:**953–955.
52. **Hayashida, S., K. Mo, and A. Hosada.** 1988. Production and characteristics of Avicel-digesting and non-Avicel-digesting cellobiohydrolases from *Aspergillus ficum*. Appl. Environ. Microbiol. **54:**1523–1529.
53. **Hazlewood, G. P., K. Davidson, J. I. Laurie, M. P. M. Romaniec, and H. J. Gilbert.** 1990. Cloning and sequencing of the *cel*A gene encoding endoglucanase A of *Butyrivibrio fibrisolvens* strain A46. J. Gen. Microbiol. **136:**2089–2097.
54. **Henrissat, B., M. Claeyssens, P. Tomme, L. Lemesle, and J.-P. Mornon.** 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene **81:**83–95.
55. **Henrissat, B., and J. P. Mornon.** *In Trichoderma* cellulases: biochemistry, genetics, physiology and applications, in press. Springer-Verlag, New York.
56. **Jauris, S., K. P. Rücknagel, W. H. Schwarz, P. Kratzsch, K. Bronnenmeir, and W. L. Staudenbauer.** 1990. Sequence analysis of the *Clostridium stercorarium celZ* gene encoding a thermoactive cellulase (Avicelase I): identification of catalytic and cellulose-binding domains. Mol. Gen. Genet. **223:**258–267.
57. **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.
58. **Joliff, G., P. Béguin, and J.-P. Aubert.** 1986. Nucleotide sequence of the cellulase gene *cel*D encoding endoglucanase D of *Clostridium thermocellum*. Nucleic Acids Res. **14:**8605–8613.
59. **Jørgensen, P. L., and C. K. Hansen.** 1990. Multiple endo-β-1,4-glucanase-encoding genes from *Bacillus lautus* PL236 and characterization of the *cel*B gene. Gene **93:**55–60.
60. **Katsuragi, N., N. Takizawa, and Y. Murooka.** 1987. Entire nucleotide sequence of the pullulanase gene of *Klebsiella aerogenes* W70. J. Bacteriol. **169:**2301–2306.
61. **Kellett, L. E., D. M. Poole, L. M. A. Ferreira, A. J. Durrant, G. P. Hazlewood, and H. J. Gilbert.** 1990. Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens* subsp. *cellulosa* contain identical cellulose-binding domains and are encoded by adjacent genes. Biochem. J. **272:**369–376.
62. **Knowles, J., P. Lehtovaara, and T. Teeri.** 1987. Cellulase families and their genes. Trends Biotechnol. **5:**255–261.
63. **Knowles, J. K. C., P. Lehtovaara, M. Murray, and M. L. Sinnott.** 1988. Sterochemical course of the action of the cellobioside hydrolases I and II of *Trichoderma reesei*. J. Chem. Soc. Chem. Commun. **1988:**1401–1402.
64. **Kolpak, F. J., and J. Blackwell.** 1976. Determination of the

structure of cellulose II. Macromolecules **9**:273–278.
65. **Kraulis, P. M., M. G. Clore, M. Nilges, T. A. Jones, G. Pettersson, J. Knowles, and A. M. Gronenborn.** 1989. Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei*. A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing. Biochemistry **28**: 7241–7257.
66. **Lamed, R., and E. A. Bayer.** 1988. The cellulosome of *Clostridium thermocellum*. Adv. Appl. Microbiol. **33**:1–46.
67. **Lipman, D. J., S. F. Altschul, and J. D. Kececioglu.** 1989. A tool for multiple sequence alignment. Proc. Natl. Acad. Sci. USA **86**:4412–4415.
68. **Liu, Y. S. V., T. L. K. Low, A. Infante, and F. W. Putnam.** 1976. Complete covalent structure of a human IgA_1 immunoglobulin. Science **193**:1017–1020.
69. **Lo, A. M., R. M. MacKay, V. M. Seligy, and G. E. Willick.** 1988. *Bacillus subtilis* β-1,4-endoglucanase products from intact and truncated genes are secreted into the extracellular medium by *Escherichia coli*. Appl. Environ. Microbiol. **54**: 2287–2292.
70. **Lund, B., F. Lindberg, and S. Normark.** 1988. Structure and antigenic properties of the tip-located P pilus proteins of uropathogenic *Escherichia coli*. J. Bacteriol. **170**:1887–1894.
71. **Lüthi, E., D. R. Love, J. McAnulty, C. Wallace, P. A. Caughey, D. Saul, and P. L. Bergquist.** 1990. Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophile "*Caldocellum saccharolyticum*." Appl. Environ. Microbiol. **56**:1017–1024.
72. **MacKay, R. M., A. Lo, G. Willick, M. Zuker, S. Baird, M. Dove, F. Moranelli, and V. Seligy.** 1986. Structure of a *Bacillus subtilis* endo-β-1,4-glucanase gene. Nucleic Acids Res. **14**: 9159–9170.
73. **Mannarelli, B. M., S. Evans, and D. Lee.** 1990. Cloning, sequencing, and expression of a xylanase gene from the anaerobic ruminal bacterium *Butyrivibrio succinogenes*. J. Bacteriol. **172**:4247–4254.
74. **Matsushita, O., J. B. Russell, and D. B. Wilson.** 1990. Cloning and sequencing of a *Bacteroides ruminicola* B_14 endoglucanase gene. J. Bacteriol. **172**:3620–3630.
75. **McGavin, M., and C. W. Forsberg.** 1989. Catalytic and substrate-binding domains of endoglucanase 2 from *Bacteroides succinogenes*. J. Bacteriol. **171**:3310–3315.
76. **McGavin, M. J., C. W. Forsberg, B. Bell, A. W. Crosby, D. Dignard, and D. Y. Thomas.** 1989. Structure of the *cel-3* gene from *Fibrobacter succinogenes* S85 and characteristics of the encoded gene product, endoglucanase 3. J. Bacteriol. **171**: 5587–5595.
76a. **Meinke, A.** Personal communication.
77. **Meinke, A., C. Braun, N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1991. Unusual sequence organization in CenB, an inverting endoglucanase from *Cellulomonas fimi*. J. Bacteriol. **171**:308–314.
78. **Mo, K., and S. Hayashida.** 1988. Conversion of *Geotrichum candidum* endocellulase I to endocellulase II by limited proteolysis. Agric. Biol. Chem. **52**:1683–1688.
79. **Nakai, R., S. Horinouchi, and T. Beppu.** 1988. Cloning and nucleotide sequence of a cellulase gene *cas*A, from an alkalophilic *Streptomyces* strain. Gene **65**:229–238.
80. **Nakamura, A., T. Uozumi, and T. Beppu.** 1987. Nucleotide sequence of a cellulase gene of *Bacillus subtilis*. Eur. J. Biochem. **164**:317–320.
81. **Nakamura, K., N. Misawa, and K. Kitamura.** 1986. Sequence of a cellulase gene of *Cellulomonas uda* CB4. J. Biotechnol. **4**:247–254.
82. **Ohmiya, K., T. Kajino, A. Kato, and S. Shimizu.** 1989. Structure of a *Ruminococcus albus* endo-1,4-β-glucanase gene. J. Bacteriol. **171**:6771–6775.
83. **O'Neill, G. P., S. H. Goh, R. A. J. Warren, D. G. Kilburn, and R. C. Miller, Jr.** 1986. Structure of the gene encoding the exoglucanase of *Cellulomonas fimi*. Gene **44**:325–330.
84. **Ooi, T., A. Shinmyo, H. Okada, S. Murao, T. Kawaguchi, and M. Arai.** 1990. Complete nucleotide sequence of a gene coding for *Aspergillus aculeatus* cellulase (F1-CMCase). Nucleic Acids Res. **18**:5884.
85. **Owolabi, J. B., P. Béguin, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1988. Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. Appl. Environ. Microbiol. **54**:518–523.
86. **Ozaki, K., S. Shikata, S. Kawai, S. Ito, and K. Okamoto.** 1990. Molecular cloning and nucleotide sequence of a gene for alkaline cellulase from *Bacillus* sp. KSM-635. J. Gen. Microbiol. **136**:1327–1334.
87. **Paice, M. G., R. Bourbonnais, M. Desrochers, L. Jurasek, and M. Yaguchi.** 1986. A xylanase gene from *Bacillus subtilis*: nucleotide sequence and comparison with *B. pumilis* gene. Arch. Microbiol. **144**:201–206.
88. **Penttila, M., P. Lehtovaara, H. Nevalainen, R. Bhikhabhai, and J. Knowles.** 1986. Homology between cellulase genes of *Trichoderma reesei*: complete nucleotide sequence of the endoglucanase I gene. Gene **45**:253–263.
89. **Perham, R. N., and L. J. Packman.** 1989. 2-Oxo acid dehydrogenase multienzyme complexes: domains, dynamics, and design. Ann. N.Y. Acad. Sci. **573**:1–20.
90. **Poole, D. M., G. P. Hazlewood, J. I. Laurie, P. J. Barker, and H. J. Gilbert.** 1990. Nucleotide sequence of the *Ruminococcus albus* SY3 endoglucanase genes *cel*A and *cel*B. Mol. Gen. Genet. **223**:217–223.
90a. **Py, B., I. Bortoli-German, J. Haiech, M. Chippaux, and F. Barras.** 1991. Cellulase EGZ of *Erwinia chrysanthemi*: structural organization and importance of His-98 and Glu-133 residues for catalysis. Protein Eng. **4**:325–333.
91. **Quiocho, F. A.** 1986. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. Annu. Rev. Biochem. **55**:287–315.
92. **Radford, S. E., E. D. Laue, R. N. Perham, S. R. Martin, and E. Appella.** 1989. Conformational flexibility and folding of synthetic peptides representing an interdomain segment of polypeptide chain in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. J. Biol. Chem. **264**:767–775.
93. **Robson, L. M., and G. H. Chambliss.** 1986. Cloning of the *Bacillus subtilis* DLG β-1,4-glucanase gene and its expression in *Escherichia coli* and *Bacillus subtilis*. J. Bacteriol. **165**:612–619.
94. **Roditi, I., H. Schwarz, T. W. Pearson, R. P. Beecroft, M. K. Liu, J. P. Richardson, H.-J. Bühring, J. Pleiss, R. Bülow, R. O. Williams, and P. Overath.** 1989. Procyclin gene expression and loss of the variant surface glycoprotein during differentiation of *Trypanosoma brucei*. J. Cell Biol. **108**:737–746.
95. **Rouvinen, J., T. Bergfors, T. Teeri, J. K. C. Knowles, and T. A. Jones.** 1990. Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science **249**:380–386.
96. **Ruoslahti, E.** 1988. Fibronectin and its receptors. Annu. Rev. Biochem. **57**:375–413.
97. **Saarilahti, H. T., B. Henrissat, and E. T. Palva.** 1990. CelS: a novel endoglucanase identified from *Erwinia carotovora* subsp. *carotovora*. Gene **90**:9–14.
98. **Saloheimo, M., P. Lehtovaara, M. Penttila, T. T. Teeri, J. Stahlberg, G. Johansson, G. Pettersson, M. Claeyssens, P. Tomme, and J. C. Knowles.** 1988. EGIII, a new endoglucanase from *Trichoderma reesei*: the characterization of both gene and enzyme. Gene **63**:11–21.
99. **Saul, D. J., L. C. Williams, R. W. Grayling, L. W. Chamley, D. R. Love, and P. L. Bergquist.** 1990. *celB*, a gene coding for a bifunctional cellulase from the extreme thermophile "*Caldocellum saccharolyticum*." Appl. Environ. Microbiol. **56**:3117–3124.
100. **Saul, D. J., L. C. Williams, D. R. Love, L. W. Chamley, and P. L. Bergquist.** 1989. Nucleotide sequence of a gene from *Caldocellum saccharolyticum* encoding for exocellulase and endocellulase activity. Nucleic Acids Res. **17**:439.
101. **Schwarz, W. H., S. Schimming, K. P. Rücknagel, S. Burgschwaiger, G. Kreil, and W. L. Staudenbauer.** 1988. Nucleotide sequence of the *cel*C gene encoding endoglucanase C of *Clostridium thermocellum*. Gene **63**:23–30.
102. **Settineri, W. J., and R. H. Marchessault.** 1965. Derivation of

possible chain conformations for poly(β-1,4-anhydroxylose). J. Polym. Sci. Part C **11**:253–264.

102a. **Shen, H., M. Schmuck, I. Pilz, N. R. Gilkes, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** J. Biol. Chem., in press.

103. **Sims, P. F. G., C. James, and P. Broda.** 1988. The identification, molecular cloning and characterization of a gene from *Phanerochaete chrysosporium* that shows strong homology to the exo-cellobiohydrolase I gene from *Trichoderma reesei*. Gene **74**:411–422.

104. **Sinnott, M. L.** 1987. Glycosyl group transfer, p. 259–296. *In* M. I. Page and A. Williams (ed.), Enzyme mechanisms. Royal Society of Chemistry, London.

105. **Srinivasa, B. R., P. J. Vithayathil, R. P. Roy, and K. R. Swaminathan.** 1990. Significance of structural homology of *Thermoascus aurantiacus* xylanase with the exoglucanase of *Cellulomonas fimi*. J. Protein Chem. **9**:337–338.

106. **Stahlberg, J., G. Johansson, and G. Pettersson.** 1988. A binding-site-deficient, catalytically active, core protein of endoglucanase III from the culture filtrate of *Trichoderma reesei*. Eur. J. Biochem. **173**:179–183.

107. **Svensson, B., H. Jespersen, M. R. Sierks, and E. A. Macgregor.** 1989. Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem. J. **264**:309–311.

108. **Taylor, K. A., B. Crosby, M. McGavin, C. W. Forsberg, and D. Y. Thomas.** 1987. Characteristics of the endoglucanase encoded by a *cel* gene from *Bacteroides succinogenes* expressed in *Escherichia coli*. Appl. Environ. Microbiol. **53**:41–46.

109. **Teather, R. M., and J. D. Erfle.** 1990. DNA sequence of a *Fibrobacter succinogenes* mixed-linkage β-glucanase (1,3-1,4-β-D-glucan 4-glucanohydrolase) gene. J. Bacteriol. **172**:3837–3841.

110. **Teeri, T. T., P. Lehtovaara, S. Kauppinen, I. Salovuori, and J. Knowles.** 1987. Homologous domains in *Trichoderma reesei* cellulolytic enzymes: gene sequence and expression of cellobiohydrolase II. Gene **51**:43–52.

111. **Tomme, P., H. Van Tilbeurgh, G. Pettersson, J. Van Damme, J. Vandekerckhove, J. Knowles, T. Teeri, and M. Claeyssens.** 1988. Studies of the cellulolytic system of *Trichoderma reesei* QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. Eur. J. Biochem. **170**:575–581.

112. **Tucker, M. L., M. L. Durbin, M. T. Clegg, and L. N. Lewis.** 1987. Avocado cellulase: nucleotide sequence of a putative full-length cDNA clone and evidence for a small gene family. Plant Mol. Biol. **9**:197–203.

113. **Van Tilbeurgh, H., P. Tomme, M. Claeyssens, R. Bhikhabhai, and G. Pettersson.** 1986. Limited proteolysis of the cellobiohydrolase I from *Trichoderma reesei*. Separation of functional domains. FEBS Lett. **204**:223–227.

114. **Wang, W., and J. A. Thomson.** 1990. Nucleotide sequence of the *cel*A gene encoding a cellodextrinase of *Ruminococcus flavefaciens* FD-1. Mol. Gen. Genet. **222**:265–269.

115. **Watanabe, T., K. Suzuki, W. Oyanagi, K. Ohnishi, and H. Tanaka.** 1990. Gene cloning of chitinase A1 from *Bacillus circulans* WL-12 revealed its evolutionary relationship to *Serratia* chitinase and to the type III homology units of fibronectin. J. Biol. Chem. **265**:15659–15665.

116. **Withers, S. G., D. Dombroski, L. A. Berven, D. G. Kilburn, R. C. Miller, Jr., R. A. J. Warren, and N. R. Gilkes.** 1986. Direct ^{1}H N.M.R. determination of the stereochemical course of hydrolyses catalyzed by glucanase components of the cellulase complex. Biochem. Biophys. Res. Commun. **139**:487–494.

117. **Wong, W. K. R., B. Gerhard, Z. M. Guo, D. G. Kilburn, R. A. J. Warren, and R. C. Miller, Jr.** 1986. Characterization and structure of an endoglucanase gene *cen*A of *Cellulomonas fimi*. Gene **44**:315–324.

118. **Wootton, J. C., and M. H. Drummond.** 1989. The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. Protein Eng. **2**:535–543.

119. **Yablonsky, M. D., T. Bartley, K. O. Elliston, S. K. Kahrs, Z. P. Shalita, and D. E. Eveleigh.** 1988. Characterization and cloning of the cellulase complex of *Microbispora bispora*. FEMS Symp. **43**:249–266.

120. **Yagüe, E., P. Béguin, and J.-P. Aubert.** 1990. Nucleotide sequence and deletion analysis of the cellulase-encoding gene *cel*H of *Clostridium thermocellum*. Gene **89**:61–67.

121. **Yang, R. C. A., C. R. MacKenzie, and S. A. Narang.** 1988. Nucleotide sequence of a *Bacillus circulans* xylanase gene. Nucleic Acids Res. **16**:7187.

122. **Yoshigi, N., H. Taniguchi, and T. Sasaki.** 1990. Cloning and sequencing of the *endo*-cellulase cDNA from *Robillarda* sp. Y-20. J. Biochem. **108**:388–392.

123. **Zappe, H., W. A. Jones, D. T. Jones, and D. R. Woods.** 1988. Structure of an endo-β-1,4-glucanase gene from *Clostridium acetobutylicum* P262 showing homology with endoglucanase genes from *Bacillus* spp. Appl. Environ. Microbiol. **54**:1289–1292.

124. **Zappe, H., W. A. Jones, and D. R. Woods.** 1990. Nucleotide sequence of a *Clostridium acetobutylicum* P262 xylanase gene (*xyn*B). Nucleic Acids Res. **18**:2179.