# 13
# Structure of Cellulases and Their Applications

KUNIO OHMIYA[1], KAZUO SAKKA[1], SHUICHI KARITA[2] AND TETSUYA KIMURA[1]

[1]*Applied Microbiology, School of Bioresources, and* [2]*Center for Molecular Biology and Genetics, Mie University, Kamihama-cho, Tsu, 514 Japan*

## Introduction

It is well known that cellulosic biomass photosynthesized by solar energy with $CO_2$ and $H_2O$ is one of the most important energy reservoirs and renewable resources on the earth. Therefore, its effective utilization through biological process will be an important key to overcome the shortage of foods, feeds and fuels, which will hit us in the near future because of the explosive increase in human population. Since the main components of the biomass are polymers consisting of glucose and other related sugars, they will become the main carbon and energy sources for microbial fermentation when they are converted into monosaccharides. However, because biomass largely originates from rigid tissue consisting of cellulose, hemicellulose, lignin and protein, effective methods of solubilization must be devised. For enhancing solubilization of fibrous plant materials, cellulases and related enzymes with higher activity and their genes have been screened for a long time. So far, more than 200 genes encoding cellulases and related enzymes have been cloned and characterized genetically and biochemically. This work has been summarized elsewhere (Béguin and Aubert, 1994; Shimada *et al.*, 1994; Tomme *et al.*, 1995b). New findings about cellulolytic enzymes and their genes have appeared since these reviews were published, as the pace of progress is continuously accelerating. Based on these fundamental data, functions of the cellulolytic enzymes have been clarified and will be improved further by using recombinant DNA techniques. Construction of new host-vector systems for cellulolytic rumen anaerobes gives us a clue to breeding cellulolytic microorganisms, leading to efficient digestion of grass in cattle. Softening of the plant tissue has been already attempted by integrating genes encoding plant cell wall-degrading enzymes such as cellulases and xylanases into plants. In addition, industrial application of cellulases for biopulping and for improving laundry detergents have been realized. In this review, we attempt to summarize some important results on the

*Biotechnology and Genetic Engineering Reviews* – Vol. 14, April 1997
0264-8725/365-414 $20.00 + $0.00 © Intercept Ltd, P.O. Box 716, Andover, Hampshire SP10 1YG, UK.

characterization of cellulases and xylanases from the recent papers. We also describe industrial application of cellulases and xylanases, and molecular breeding of cellulolytic organisms.

## Molecular structures of cellulases

### CATALYTIC DOMAINS OF CELLULASES AND XYLANASES

#### *Modular structures of cellulases*

The fungal and bacterial cellulases usually comprise two or more structural and functional domains, i.e., a catalytic domain joined to a cellulose-binding domain (CBD) is a common arrangement in enzymes of non-complexed systems from aerobic organisms, and a catalytic domain joined to a dockerin domain which is responsible for the cellulosome assembly is a common arrangement in enzymes of complexed systems from some anaerobic organisms such as *Clostridium thermocellum* (Bayer *et al.*, 1994; Béguin and Lemaire, 1996). In addition to these domains, many cellulases and xylanases include S-layer homologous (SLH) domain, fibronectin-type III domain, NodB-like domain, and various regions of unknown function (Tomme *et al.*, 1995b). These domains are often connected by linker sequences enriched in Pro and hydroxyamino acids (Thr and Ser). In *Table 1*, data on domain organizations of cellulases and xylanases are collected, including the sizes of enzymes and the location of functional domains in each polypeptide. Among various domains, catalytic and cellulose-binding domains are generally recognized to be important for hydrolysis of cellulose. Therefore, this short review focuses on the structure and function of these two important domains.

#### *Classification of catalytic domains*

Catalytic domains of cellulases and xylanases were first classified into 6 families with alphabetical letters, i.e., families A to F, on the basis of amino acid sequence homology and hydrophobic cluster analysis by Henrissat *et al.* (1989). Six families were later added to the classification (Tomme *et al.*, 1995b). On the other hand, Henrissat and Bairoch (1993, 1996) compared all the available sequences of glycosyl hydrolases (E.C. 3.2.1.x), using hydrophobic cluster analysis, and classified these enzymes into families with Arabic numerals, currently families 1 to 58 (Henrissat and Bairoch, 1996). Recently, they have constructed a permanently updated version of the classification, which is available through the ExPASy WWW server at the URL: 'http://expasy.hcuge.ch/cgi-bin/lists?glycosid.txt'. The classification of glycosyl hydrolase related to degradation of cellulose and xylan is summarized in *Table 2*. Since this classification is based on the structural features of the glycosyl hydrolases, it is not necessarily consistent with the IUB enzyme nomenclature of this class of enzymes, which is based on substrate specificity. For example, family 1 includes various enzymes with different EC number such as β-glucosidase (3.2.1.21), 6-phospho-β-galactosidase (3.2.1.85), 6-phospho-β-glucosidase (3.2.1.86), lactase-phlorizin hydrolase (3.2.1.108), and myrosinase (3.2.3.1). Family 5 comprises endoglucanase (EC 3.2.1.4), β-mannanase (EC 3.2.1.78), exo-1,3-glycanase (EC 3.2.1.58), and

**Table 1.** Classification of cellulase and xylanase catalytic domains

| Family and Organism | Strain | Enzyme | Data Base Accession No. Protein | DNA | EC Number | Size | Mature | Proton donor & Nucleophile | Domain and comment |
|---|---|---|---|---|---|---|---|---|---|
| family 5 | | | | | | | | | |
| *Anaerocellum thermophilum* | | CELD | | Z77855 | EC 3.2.1.4 | 749Aa | | 189 286 | CD-390, SLH |
| *Bacillus lautus* | PL236 | CELB | P23550 | M33762 | EC 3.2.1.4 | 566Aa | 31-566 | 177 299 | |
| *Bacillus polymyxa* | | | P23548 | M33791 | EC 3.2.1.4 | 397Aa | | 194 317 | CD |
| *Bacillus* sp. | 1139 | | P06564 | M15743 | EC 3.2.1.4 | 800Aa | 31-800 | 190 305 | CD, Uk |
| *Bacillus* sp. | 186-1 | CELB1 | | Z33876 | EC 3.2.1.4 | 389Aa | 26-389 | 165 254 | CD |
| *Bacillus* sp. | D04 | CEL | | U27084 | EC 3.2.1.4 | 486Aa | 30-486 | 169 257 | |
| *Bacillus* sp. | KSM635 | | P19424 | M27420 | EC 3.2.1.4 | 941Aa | 30-941 | 373 485 | 40-SLH1-99, 100-SLH2-151, 152-SLH3 225, CD, Uk |
| *Bacillus* sp. | KSM64 | | | M84963 | EC 3.2.1.4 | 822Aa | | 190 305 | |
| *Bacillus* sp. | N-4 | CELA | P06566 | M14781 | EC 3.2.1.4 | 488Aa | | 163 252 | |
| *Bacillus* sp. | N-4 | CELB | P06565 | M14729 | EC 3.2.1.4 | 409Aa | | 165 254 | |
| *Bacillus* sp. | N-4 | CELC | P19570 | M25500 | EC 3.2.1.4 | 825Aa | | 219 335 | |
| *Bacillus subtilis* | | BGLC | P07983 | M16185 | EC 3.2.1.4 | 499Aa | 30-499 | 169 257 | CD, 350-CBD-499 |
| *Bacillus subtilis* | 168trpC2 | EGLS | P10475 | Z29076 | EC 3.2.1.4 | 499Aa | 30-499 | 169 257 | CD, 350-CBD-499 |
| *Bacillus subtilis* | | BGLC | P23549 | D01057 | EC 3.2.1.4 | 499Aa | 30-499 | 169 257 | CD, 350-CBD-499 |
| *Burkholderia solanacearum* | | EGL | P17974 | M84922 | EC 3.2.1.4 | 426Aa | 46-426 | 249 361 | CD |
| *Butyrivibrio fibrisolvens* | A46 | CELA | P22541 | M37031 | EC 3.2.1.4 | 429Aa | 35-429 | 249 334 | CD |
| *Butyrivibrio fibrisolvens* | H17c | END1 | P20847 | X17538 | EC 3.2.1.4 | 547Aa | | 189 321 | CD, 451-CBD-547 |
| *Butyrutyrrllum saccharolyticum* | | CELB | P10474 | X13602 | EC 3.2.1.4 | 1039Aa | 29-1039 | 792 934 | 29-CD(10)-375, 376-Lin-416, 417-CBD-570, 571-Lin-618, 619-CD(5)-1039 |
| *Caldocellum saccharolyticum* | | MANA | P22533 | L01257 | EC 3.2.1.78 | 1331Aa | 42-1331 | 162 257 | 42-CD(5)-325, 326-Lin-361, 362-CBD-518, 519-Lin-564, 565-CBD-720, 721-Lin-780, 781-CD(44)-1331 |
| *Candida albicans* | | XOG1 | P29717 | X56556 | EC 3.2.1.58 | 438Aa | 39-438 | 230 330 | |
| *Cellulomonas fimi* | | CEND | P50400 | L02544 | EC 3.2.1.4 | 747Aa | 40-747 | 208 349 | CD, 448-Fn3-542, 544-Fn3-639, 640-CBD-747 |
| *Clostridium aceto-butylicam* | P262 | EGLA | P15704 | M31311 | EC 3.2.1.4 | 448Aa | 35-448 | 175 263 | CD |

**Table 1 cont.**

| Family and Organism | Strain | Enzyme | Data Base Accession No. Protein | DNA | EC Number | Size | Mature | Proton donor | Domain and comment & Nucleophile |
|---|---|---|---|---|---|---|---|---|---|
| *Clostridium cellulolyticum* | ATCC35319 | CELCCA | P17901 | M93096 | EC 3.2.1.4 | 475Aa | 27-475 | 195 332 | CD, Doc |
| *Clostridium cellulolyticum* | ATCC35319 | CELCCD | P25472 | D90341 | EC 3.2.1.4 | 584Aa | 25-584 | 159 264 | 25-CD-328, 329-Lin-353, 354-CBD-529, 530-Doc-584 |
| *Clostridium cellulovorans* | ATCC35296 | ENGB | P28621 | M75706 | EC 3.2.1.4 | 440Aa | 34-440 | 179 305 | CD, 87-Doc-440 |
| *Clostridium cellulovorans* | ATCC35296 | ENGD | P28623 | M37434 | EC 3.2.1.4 | 515Aa | 32-515 | 180 303 | 32-CD-376, 377-Lin-407, 408-CBD-515 |
| *Clostridium josui* | | CELA | | D85526 | EC 3.2.1.4 | 930Aa | | 184 287 | CD, Uk, SLH |
| *Clostridium longisporum* | | CELA | | L02868 | EC 3.2.1.4 | 517Aa | | 185 309 | CD, 421-CBD-517 |
| *Clostridium thermocellum* | F1 | CEL307 | P23340 | D00945 | EC 3.2.1.4 | 343Aa | 22-343 | 140 280 | CD |
| *Clostridium thermocellum* | NCIB10682 | CELB | P04956 | X03592 | EC 3.2.1.4 | 563Aa | 28-563 | 204 363 | CD, Lin, 502-Doc-563 |
| *Clostridium themocellum* | | CELC | P07985 | M19422 | EC 3.2.1.4 | 343Aa | | 140 280 | CD |
| *Clostridium thermocellum* | | CELE | P10477 | M22759 | EC 3.2.1.4 | 814Aa | 35-814 | 193 316 | CD, 415-Doc-474, CBD |
| *Clostridium thermocellum* | NCIB10682 | CELG | Q05332 | X69390 | EC 3.2.1.4 | 566Aa | 31 566 | 226 381 | CD, 503-Doc-549 |
| *Clostridium thermocellum* | NCIB10682 | CELH | P16218 | M31903 | EC 3.2.1.4 | 900Aa | 45-900 | 460 565 | 44-CD-630, 631-Lin-654, 655-CBD-832, 833-Doc-895 |
| *Cryptococcus flavus* | | CMC1 | Q04469 | D13967 | EC 3.2.1.4 | 341Aa | 31-341 | 166 275 | CD |
| *Erwinia carotovora* subsp. | *atroseptica* | CELN | | L39788 | EC 3.2.1.4 | 444Aa | 32-444 | 168 256 | 32-CD-338, 339-Lin-357, 358-CBD-444 |
| *Erwinia carotovora* | SCC3193 | CELV1 | | X79241 | EC 3.2.1.4 | 504Aa | | 168-256 | CD, Lin, CBD |
| *Erwinia chrysanthemi* | 3937 | CELZ | P07103 | Y00540 | EC 3.2.1.4 | 426Aa | 44-385 | 176 263 | 44-CD-332, 333-Lin-366, 367-CBD-426 |
| *Fibrobacter succinogenes* | A3c | END3 | | L39839 | EC 3.2.1.4 | 657Aa | 26-657 | 448 597 | CD |
| *Fibrobacter succinogenes* | RH9-1 | END3 | | L39838 | EC 3.2.1.4 | 669Aa | 24-669 | 454 603 | CD |
| *Fibrobacter succinogenes* | S85 | CEDA | | U07419 | EC 3.2.1.91 | 357Aa | 17-357 | 151 284 | CD |
| *Fibrobacter succinogenes* | S85 | CEL-3 | P14250 | M29047 | EC 3.2.1.4 | 658Aa | 26-658 | 448 597 | |
| *Fibrobacter succinogenes* | S85 | CELG | | U33887 | EC 3.2.1.4 | 519Aa | | 166 324 | CD |
| *Humicola grisea* | IF09854 | CEL | | D84470 | EC 3.2.1.4 | 388Aa | | 215 | |
| *Macrophomina phaseolina* | | EGL1 | | U14948 | EC 3.2.1.4 | 332Aa | | | |
| *Macrophomina phaseolina* | | EGL2 | | U13914 | EC 3.2.1.4 | 368Aa | | 175 287 | CD |
| *Neocallimastix fontalis* | MCH3 | CELA | | U38843 | EC 3.2.1.4 | 433Aa | | 142 266 | CD, Doc |
| *Neocallimastix patriciarum* | | CELB | | Z31364 | EC 3.2.1.4 | 473Aa | 18-473 | 173 295 | CD, 366-Lin-387, 388-Doc-473 |
| *Penicillium janthinellum* | | EGII | | X89564 | EC 3.2.1.4 | 410Aa | | | CD, Lin, CBD |
| *Pseudomonas fluorescens* sp. | *cellulosa* | CELC | P27033 | X61299 | EC 3.2.1.4 | 748Aa | 39-748 | 503 653 | 39-CBD-136, 137-Lin-179, 228-Lin-280, 281-CD-748 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Pseudomonas fluorescens* sp. *cellulosa* | | CELE | | X86798 | EC 3.2.1.4 | 570Aa | | 171 259 | CD-340, 341-Lin-378, 379-Uk-424, 425-Lin-470, 471-CBD-570 |
| *Robillarda* sp. | Y-20 | CMCaseI | P23044 | | EC 3.2.1.4 | 375Aa | | | |
| *Prevotella ruminicola* | AR20 | END | | S22458 | EC 3.2.1.4 | 500Aa | | 293 425 | CD |
| *Prevotella ruminicola* | 23 | XYNA | | M83379 | EC 3.2.1.8 | 584Aa | | 189 322 | CD |
| *Prevotella ruminicola* | $B_14$ | CMCASE | | M38216 | EC 3.2.1.4 | 925Aa | | 734 852 | CD(26), CD(5) |
| *Ruminococcus albus* | AR67 | CELA | | | EC 3.2.1.4 | 414Aa | | 217 343 | CD |
| *Ruminococcus albus* | 8 | CELA | | L04563 | EC 3.2.1.4 | 411Aa | | 179 305 | CD |
| *Ruminococcus albus* | F-40 | EG I | P16216 | M30928 | EC 3.2.1.4 | 406Aa | 44-406 | 210 330 | CD |
| *Ruminococcus albus* | F-40 | EG IV | Q07940 | D16315 | EC 3.2.1.4 | 312Aa | | 135 222 | CD |
| *Ruminococcus albus* | SY3 | CELA | P23660 | X54931 | EC 3.2.1.4 | 364Aa | | 169 293 | CD |
| *Ruminococcus albus* | SY3 | CELB | P23661 | X54932 | EC 3.2.1.4 | 409Aa | | 212 332 | CD |
| *Ruminococcus flavefaciens* | 17 | ENDA | | S55178 | EC 3.2.1.4 | 455Aa | | 196 325 | CD |
| *Ruminococcus flavefaciens* | FD-1 | CELA | P16169 | X51944 | EC 3.2.1.4 | 336Aa | | 141 | |
| *Ruminococcus flavefaciens* | FD-1 | CELE | Q05622 | L03800 | EC 3.2.1.4 | 320Aa | 30-320 | 150 | |
| *Saccharomyces cerevisiae* | | EXG1 | P23776 | M34341 | EC 3.2.1.58 | 448Aa | 41-448 | 232 334 | |
| *Saccharomyces cerevisiae* | | EXG2 | P52911 | Z46870 | EC 3.2.1.58 | 562Aa | 23-562 | 254 334 | |
| *Saccharomyces cerevisiae* | GRF88 | SPR1 | P32603 | X59259 | EC 3.2.1.58 | 445Aa | 22-445 | 233 335 | |
| *Saccharomyces cerevisiae* | 5288C | | P38081 | Z35925 | | 501Aa | | 236 333 | |
| *Saccharomyces cerevisiae* | 5288C | | P40566 | Z47047 | | 764Aa | | 265 | |
| *Schizosaccharomyces pombe* | | | Q10444 | Z70721 | EC 3.2.1.58 | 570Aa | 48-570 | 338 439 | |
| *Streptomyces lividans* | 66 | CELA | P27035 | M82807 | EC 3.2.1.4 | 459Aa | 28-459 | 286 378 | 28-CBD-135, CD |
| *Streptomyces lividans* | 66 | MANA | P51529 | M92297 | EC 3.2.1.78 | 363Aa | 36-363 | | |
| *Thermomonospora fusca* | | CELE | Q01786 | L01577 | EC 3.2.1.4 | 466Aa | 37-466 | 299 391 | 37-CBD-142,143-Lin-163, 164-CD-466 |
| *Trichoderma reesei* | | EGL2 | P07982 | M19373 | EC 3.2.1.4 | 418Aa | 22-418 | 239 350 | 22-CBD-57, 58-Lin-91, 91-CD-418 |
| *Trichoderma reesei* | RutC30 | MAN | | L25310 | EC 3.2.1.78 | 437Aa | | 196 314 | Lin, CBD |
| *Xanthomonas abilineans* | | ENG | | L26543 | EC 3.2.1.4 | 361Aa | | 194 ? | |
| *Xanthomonas campestris* | | ENGXCA | P19487 | M32700 | EC 3.2.1.4 | 493Aa | 26-493 | 182 303 | CD, 375-Lin-399 |
| unidentified thermoanaerobe | | CEL | | U12011 | EC 3.2.1.4 | 332Aa | | 138 275 | |
| Family 6 | | | | | | | | | |
| *Agricus bisporus* | C54-CARB8 | CEL3A | | Z34007 | EC 3.2.1.91 | 438Aa | | | |
| *Agaricus bisporus* | D649 | CEL3 | P49075 | L24519 | EC 3.2.1.91 | 438Aa | 21-438 | 215 393 | 21-CBD-59, 60-Lin-59, CD |
| *Cellulomonas fimi* | | CBHA | | L25809 | EC 3.2.1.91 | 872Aa | | | |
| *Cellulomonas fimi* | | CENA | P07984 | M15823 | EC 3.2.1.4 | 449Aa | 32-449 | 283 423 | 32-CBD-138, 139-Lin-168, CD |
| *Fusarium oxysporum* | | | P46236 | L29377 | EC 3.2.1.4 | 462Aa | 17-462 | 236 416 | 17-CBD-65, 66-Lin-99, CD |
| *Microbispora bispora* | | CELA | P26414 | | EC 3.2.1.4 | 456Aa | 31-456 | 151 300 | CD, 323-Lin-355, 356-CBD-456 |

**Table 1 cont.**

| Family and Organism | Strain | Enzyme | Data Base Accession No. Protein | DNA | EC Number | Size | Mature | Proton donor | Domain and comment & Nucleophile |
|---|---|---|---|---|---|---|---|---|---|
| *Neocallimastix patriciarum* | | CELA | | U29872 | EC 3.2.1.91 | 428Aa | 20-228 | 222 383 | |
| *Streptomyces halstedii* | JM8 | CELA1 | P33682 | Z12157 | EC 3.2.1.4 | 321Aa | 28-321 | 149 295 | |
| *Streptomyces* sp. | KSM-9 | CASA | P13933 | L03218 | EC 3.2.1.4 | 359Aa | 71-359 | 192 339 | |
| *Thermomonospora fusca* | YX | CELB | P26222 | M73321 | EC 3.2.1.4 | 441Aa | 32-441 | 148 296 | 32-CD-320, 321-Lin-340, 341-CBD-441 |
| *Thermomonospora fusca* | YX | CELC | | U18978 | EC 3.2.1.91 | 596Aa | 39-596 | | 39-CBD-140, 141-Lin-175, 176-CD-596 |
| *Trichoderma reesei* | QM9414 | CBH2 | P07987 | M16190 | EC 3.2.1.91 | 471Aa | 25-471 | 245 425 | 25-CBD-65, 66-Lin-106, CD |
| Family 7 | | | | | | | | | |
| *Aspergillus oryzae* | KBN616 | CELB | | D83732 | EC 3.2.1.4 | 416Aa | | | CD |
| *Cochliobolus carbonum* | RACE1 | CEL1 | | U25129 | EC 3.2.1.91 | 456Aa | | | |
| *Cryphonectria parasitica* | | CBH-1 | | L43048 | EC 3.2.1.91 | 452Aa | | | |
| *Fusarium oxysporum* | | | P46237 | L29378 | EC 3.2.1.4 | 429Aa | 19-429 | 144 | |
| *Fusarium oxysporum* | | | P46238 | L29379 | EC 3.2.1.91 | 514Aa | 18-514 | | 18-CD-439, 440-Lin-482, 483-CBD-514 |
| *Humicola grisea* | IFO9854 | CBH-1 | | D63515 | EC 3.2.1.91 | 525Aa | 19-525 | 145 | 19-CD-467, 468-Lin-489, 490-CBD-525 |
| *Humicola grisea* | | CBH1 | P15828 | X17258 | EC 3.2.1.91 | 525Aa | 19-525 | 145 | 19-CD-467, 468-Lin-489, 490-CBD-525 |
| *Neurospora crassa* | 74-OR23-1A | CBH-1 | P38676 | X77788 | EC 3.2.1.91 | 516Aa | 18-516 | | 18-CD-445, 446-Lin-480, 481-CBD-516 |
| *Penicillium janthinellum* | C41 | CBH1 | Q06886 | X59054 | EC 3.2.1.91 | 537Aa | 19-537 | | 19-CD-453, 454-Lin-477, 478-CBD-537 |
| *Phanerochaete chrysosporium* | BKM-1767 | CBHI-1 | | X54411 | EC 3.2.1.91 | 451Aa | | | |
| *Phanerochaete chrysosporium* | ME446 | CBH1 | P13860 | M22220 | EC 3.2.1.91 | 516Aa | | | CD, 450-Lin-480, 481-CBD-516 |
| *Phanerochaete chrysosporium* | ME446 | CBHI-2 | | Z29653 | EC 3.2.1.91 | 540Aa | | | CD, Lin, CBD |
| *Phanerochaete chrysoporium* | | CBHI-4 | | | EC 3.2.1.91 | 510Aa | | | |
| *Trichoderma koningii* | G-39 | CBHI | | X69976 | EC 3.2.1.91 | 513Aa | 18-513 | | CD, Lin, CBD |
| *Trichoderma longibrachiatum* | | EGLI | | X60652 | EC 3.2.1.91 | 463Aa | 23-463 | | CD, Lin, CBD |
| *Trichoderma reesei* | L-27 | CBH1 | P00725 | X69976 | EC 3.2.1.91 | 513Aa | 18-513 | 234 229 | 18-CD-453, 454-Lin-477, 478-CBD-513 |
| *Trichoderma reesei* | VTT-D-80133 | EGL1 | P07981 | M15665 | EC 3.2.1.4 | 459Aa | 23-459 | | 23-CD-397, 398-Lin-423, 424-CBD-459 |
| *Trichoderma viride* | | CBH1 | P19355 | X53931 | EC 3.2.1.91 | 513Aa | 18-513 | | 18-CD-453, 454-Lin-477, 478-CBD-513 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Family 8 | | | | | | | | | |
| *Acetobacter xylinum* | ATCC23769 | CMCAX | P37696 | M69060 | EC 3.2.1.4 | 342Aa | 21-342 | 57 114 | |
| *Bacillus circulans* | WL-12 | BGC | P19254 | X52880 | EC 3.2.1.73 | 409Aa | | 95 156 | |
| *Bacillus* sp. | KSM330 | | P29019 | M68872 | EC 3.2.1.4 | 463Aa | 56-463 | 130 191 | |
| *Cellulomonasu uda* | CB4 | | P18336 | M36503 | EC 3.2.1.4 | 359Aa | 24-359 | 53 110 | CD |
| *Clostridium josui* | | CELB | P37701 | D16670 | EC 3.2.1.4 | 460Aa | | 99 155 | CD, Doc |
| *Clostridium thermocellum* | NCIB10682 | CELA | P04955 | K03088 | EC 3.2.1.4 | 477Aa | 33-477 | 95 152 | CD, 416-Doc-472 |
| *Clostridium cellulolyticum* | ATCC35319 | CELCCC | P37699 | M87018 | EC 3.2.1.4 | 460Aa | 33-460 | 99 155 | CD, 405-Doc-460 |
| *Erwinia chrysanthemi* | 3937 | CELY | P27032 | M74044 | EC 3.2.1.4 | 332Aa | 24-332 | 53 110 | CD |
| *Escherichia coli* | K12 | YHJM | P37651 | U00039 | | 368Aa | 22-368 | 55 116 | CD |
| Family 9 | | | | | | | | | |
| *Arabidopsis thaliana* | | OR16 | | U37702 | EC 3.2.1.4 | 621Aa | 35-621 | 570 165 | |
| *Arabidopsis thaliana* | | EGL | | U17888 | EC 3.2.1.4 | | | | Partial |
| *Butyrivibrio fibrisolvens* | H17c | CED1 | P23658 | X55732 | EC 3.2.1.- | 547Aa | | 529 148 | CD |
| *Caldocellum saccharolyticum* | | CELA | | L32742 | EC 3.2.1.4 | 1742Aa | | 443 80 | CD(9), Lin, CBD, Lin, CBD, CD(48) |
| *Cellulomonas fimi* | ATCC484 | CENB | P26225 | M64644 | EC 3.2.1.4 | 1045Aa | 34-1045 | 458 91 | 34-CD-643, 644-Fn3-944, 945-CBD-1045 |
| *Cellulomonas fimi* | ATCC484 | CENC | P14090 | X57858 | EC 3.2.1.4 | 1101Aa | 33-1101 | 891 506 | 64-CBD1-173, 212-CBD2-318, 329-CD-880, 918-Rp-1101 |
| *Clostridium cellulolyticum* | ATCC35319 | CELCCG | P37700 | M87018 | EC 3.2.1.4 | 725Aa | 36-725 | 455 93 | CD, 485-CBD-662, 664-Doc-719 |
| *Clostridium cellulovorans* | ATCC35296 | ENGC | P28622 | M95180 | EC 3.2.1.4 | 553AA | | 466 101 | CD, Doc |
| *Clostridium stercorarium* | NCIB11745 | CELZ | P23659 | X55299 | EC 3.2.1.4 | 986Aa | 26-986 | 447 84 | CD, 835-CBD-986 |
| *Clostridium thermocellum* | NCIB10682 | CELD | P04954 | X04584 | EC 3.2.1.4 | 649AA | 42-649 | 555 201 | CD, 584-Doc-644 |
| *Clostridium thermocellum* | NCIB10682 | CELF | P26224 | X60545 | EC 3.2.1.4 | 739Aa | 28-739 | 447 84 | 28-CD-470, 505-CBD-638, 670-Doc-732 |
| *Clostridium thermocellum* | NCIB10682 | CELI | Q02934 | L04735 | EC 3.2.1.4 | 879Aa | 56-879 | 495 131 | 56-CD-518, 729-CBD-879 |
| *Clostridium thermocellum* | F1 | CELJ | | D83704 | EC 3.2.1.4 | 1601Aa | | 735 377 | SLH, Uk, CD(9), CD(44), Doc, Uk |
| *Dictyostelium discoideum* | | CELA | P22699 | M33861 | EC 3.2.1.4 | 705Aa | 24-705 | 437 85 | CD, Lin, CBD |
| *Fibrobacter succinogenes* | AR1 | ENDA | P23665 | M58520 | EC 3.2.1.4 | 453Aa | 27-453 | | |
| *Fibrobacter succinogenes* | BL2 | EGC | | L48039 | EC 3.2.1.4 | 620Aa | | 539 177 | CD |
| *Fibrobacter succinogenes* | S85 | EGB | | L14436 | EC 3.2.1.4 | 555Aa | | 537 160 | CD |
| *Fibrobacter succinogenes* | S85 | CELD | | U05897 | EC 3.2.1.4 | 668Aa | | 538 177 | CD |
| *Fibrobacter succinogenes* | S85 | CELE | | U05897 | EC 3.2.1.4 | 467Aa | | 389 24 | CD |
| *Lycopersicon esculentum* | | CEL1 | | U20590 | EC 3.2.1.4 | 501Aa | | 462 79 | CD |
| *Lycopersicon esculentum* | | CEL2 | | U13055 | EC 3.2.1.4 | 489Aa | | 468 80 | CD |
| *Persea americana* cv. *Haas* | | CEL1 | P05522 | M17634 | EC 3.2.1.4 | 494Aa | 26-494 | 471 82 | CD |
| *Phaseolus vulgaris* | | | P22503 | M57400 | EC 3.2.1.4 | 496Aa | 24-496 | 475 93 | CD |
| *Pisum sativum* | | EGP1 | | L41046 | EC 3.2.1.4 | 486Aa | 25-489 | 466 80 | CD |

**Table 1 cont.**

| Family and Organism | Strain | Enzyme | Data Base Accession No. Protein | DNA | EC Number | Size | Mature | Proton donor | Domain and comment & Nucleophile |
|---|---|---|---|---|---|---|---|---|---|
| *Populus alba* | | | | D32166 | EC 3.2.1.4 | 494Aa | | 474 85 | CD |
| *Pseudomonas fluorescens* subsp. *cellulosa* | | CELA | P10476 | X12570 | EC 3.2.1.4 | 962Aa | | 582 203 | CD, 608-Lin-664, Uk, 823-Lin-859, 866-CBD-962 |
| *Prunus persica* | | | P38534 | ZI23119 | EC 3.2.1.4 | 251Aa | | | Partial |
| *Streptomyces reticuli* | TU45 | CEL1 | Q05156 | X65616 | EC 3.2.1.4 | 746Aa | 30-746 | 726 342 | 45-CBD-170, 171-Lin-211, 212-CD-746 |
| *Thermomonospora fusca* | | CELA | | L20094 | EC 3.2.1.4 | 974Aa | | 732 345 | CD, CBD |
| *Thermomonospora fusca* | | CELD | P26221 | M73322 | EC 3.2.1.4 | 707Aa | 47-507 | 469 104 | CD, CBD |
| Family 10 | | | | | | | | | |
| *Actinomadura* sp. FC7 | | XYNII | | U08894 | EC 3.2.1.8 | 419Aa | 38-419 | 166 274 | |
| *Anaerocellum thermophilum* | | XYNA | | Z69782 | EC 3.2.1.8 | 688Aa | | 492 600 | |
| *Aspergillus awamori* | IFO4308 | XYNA | P33559 | D14847 | EC 3.2.1.8 | 327Aa | 26-327 | 157 263 | CD |
| *Aspergillus nidulans* | | XYLC | | Z49894 | EC 3.2.1.8 | 309Aa | | 154 244 | CD |
| *Bacillus* sp. | 137 | XYL | | Z35497 | EC 3.2.1.8 | 331Aa | | 136 243 | CD |
| *Bacillus* sp. | C-125 | XYNA | P07528 | D00087 | EC 3.2.1.8 | 396Aa | 29-396 | 195 301 | CD |
| *Bacillus stearothermophilus* | No.21 | XYNA | P45703 | D28121 | EC 3.2.1.8 | 330Aa | 7-330 | 133 240 | CD |
| *Bacillus stearothermophilus* | T-6 | | P40943 | Z29080 | EC 3.2.1.8 | 407Aa | 29-407 | 187 293 | |
| *Bacteroides ovatus* | V975 | XYLI | P49942 | U04957 | EC 3.2.1.8 | 376Aa | 25-376 | 160 265 | CD |
| *Butyrivibrio fibrisolvens* | 49 | XYNA | P23551 | | EC 3.2.1.8 | 411Aa | 34-411 | 201 311 | |
| *Butyrivibrio fibrisolvens* | H17c | XYNB | P26223 | X61495 | EC 3.2.1.8 | 635Aa | | 150 255 | |
| *Caldocellum saccharolyticum* | | CELB | P10474 | X13602 | EC 3.2.1.8 | 1039Aa | 29-1039 | | 29-CD(10)-375, 376-Lin-416, 417-CBD-570, 571-Lin-618, 619-CD(5)-1039 |
| *Caldocellum saccharolyticum* | | XYNA | P23556 | M34459 | EC 3.2.1.8 | 342Aa | 34-342 | 144 252 | CD |
| *Caldocellum saccharolyticum* | | ORF4 | P23557 | M34459 | EC 3.2.1.8 | 312Aa | | 104 216 | |
| *Cellulomonas fimi* | | XYNB | P07986 | M15824 | EC 3.2.1.91 | 484Aa | 42-484 | 168 274 | 42-CD-356, 357-Lin-376, 377-CBD-484 |
| *Cellulomonas fimi* | | XYNC | | Z50866 | EC 3.2.1.8 | 1188Aa | | 547 659 | Uk, CD, CBD |
| *Cellvibrio mixtus* | | XYNB | | Z48926 | EC 3.2.1.8 | 621Aa | 20-621 | 403 516 | |
| *Clostridium stercorarium* | F-9 | XYNB | P40942 | D12504 | EC 3.2.1.8 | 387Aa | | 185 293 | CD |
| *Clostridium thermocellum* | | XYNX | P38535 | M67438 | EC 3.2.1.8 | 1087Aa | 31-1087 | 347 452 | Uk, CD, CBDI, CBDII, 907-SLH1-965, 966-SLH2-1029, 1030-SLH3-1087 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Clostridium thermocellum* | YS | XYNY | P51584 | X83269 | EC 3.2.1.8 | 1077Aa | 27-1077 | 337 460 | CD, 734-Doc-791 |
| *Clostridium thermocellum* | NCIB10682 | XYNZ | P10478 | M22624 | EC 3.2.1.8 | 837Aa | 29-837 | 645 754 | 328-CBD-416, 429-Doc-487 |
| *Cryptococcus albidus* | CCY17-4-4 | | P07529 | X12596 | EC 3.2.1.8 | 332Aa | 22-332 | 120 214 | CD |
| *Dictyoglomus thermophilum* | Rt46B.1 | XYNA | | L39866 | EC 3.2.1.8 | 352Aa | | 155 262 | CD |
| *Eubacterium ruminantium* | | XYNA | | D63938 | EC 3.2.1.8 | 827Aa | | 408 534 | Uk, CD, CBD |
| *Fusarium oxyrporum* | | | P46239 | L29380 | EC 3.2.1.4 | 385Aa | 20-385 | 209 321 | 25-CBD-52, 53-Lin-84, 85-CD-385 |
| *Magnaporthe grisea* | Ken60-19 | XYN33 | | L37530 | EC 3.2.1.8 | 331Aa | 29-331 | 159 266 | CD |
| *Neocallimastix patriciarum* | | XYLB | | X76919 | EC 3.2.1.8 | 860Aa | | 144 255 | |
| *Penicillium chrysogenum* | Q176 | XYLP | P29417 | M98458 | EC 3.2.1.8 | 353Aa | 24-353 | 161 267 | CD |
| *Prevotella ruminicola* | $B_1$4 | XYNA | P48789 | Z49241 | EC 3.2.1.8 | 369Aa | 21-369 | 156 261 | CD |
| *Prevotella ruminicola* | $B_1$4 | XYNC | | Z79595 | EC 3.2.1.8 | 560Aa | 20-560 | 337 468 | CD |
| *Prevotella ruminicola* | D31d | XYN | | U53926 | EC 3.2.1.8 | 707Aa | | 476 615 | |
| *Pseudomonas fluorescens* subsp. *cellulosa* | | XYNA | P14768 | X15429 | EC 3.2.1.8 | 611Aa | 27-611 | 391 510 | 27-CBD-130, 131-Lin-179, 227-Lin-259, CD |
| *Pseudomonas fluorescens* subsp. *cellulosa* | | XYNB | P23030 | X54523 | EC 3.2.1.8 | 592Aa | 39-592 | 431 530 | 39-CBD-134, 135-Lin-160, 300-Lin-320, CD |
| *Pseudomonas fluorescens* subsp. *cellulosa* | | XYNF | | Z48928 | EC 3.2.1.8 | 605Aa | 20-605 | 385 497 | CBD, CD |
| *Ruminococcus flavefacience* | 17 | XYNA | P29126 | Z11127 | EC 3.2.1.8 | 954Aa | 28-954 | 774 884 | 245-Lin-622, 623-CD-954 |
| *Streptomyces lividans* | 66 | XLNA | P26514 | M64551 | EC 3.2.1.8 | 477Aa | 42-477 | 169 277 | CD |
| *Thermoascus aurantiacus* | | | P23360 | | EC 3.2.1.8 | 269Aa | | | |
| *Thermoanaerobacter saccharolyticum* | B6A-RI | XYNA | P36917 | M97882 | EC 3.2.1.8 | 1157Aa | 34-1157 | 495 600 | Uk, CD, CBDI, CBDII, 1055-SLH1-1113, 1114-SLH2-1157 |
| *Thermoanaerobacterium thermosulfurigenes* | | XYNA | | U50952 | EC 3.2.1.8 | 1234Aa | 33-1234 | 494 599 | Uk, CD, CBDI, CBDII, SLH |
| *Thermotoga maritima* | MSB8 | XYNA | | Z46264 | EC 3.2.1.8 | 1059Aa | 31-1055 | 502 608 | Uk, CD, CBDI, CBDII |
| *Thermotoga neapolitana* | Z2706-MC24 | XYNA | | Z46945 | EC 3.2.1.8 | 1055Aa | | 498 604 | |
| *Thermotoga* sp. | FjSS3-B.1 | XYNA | | U33060 | EC 3.2.1.8 | 346Aa | | 153 259 | CD |
| *Thermophilic bacterium* | RT8.B4 | XYNA | P40944 | L18965 | EC 3.2.1.8 | 684Aa | 35-684 | 490 598 | Uk, CD |
| Family 11 | | | | | | | | | |
| *Ascochyta pisi* | | XYN | | Z68891 | EC 3.2.1.8 | 227Aa | 20-227 | 212 121 | CD |
| *Aspergillus awamori* | ATCC11358 | EXLA | | X78115 | EC 3.2.1.8 | 211Aa | | 197 106 | CD |
| *Aspergillus kawachii* | IFO4308 | XYNB | P48824 | D38070 | EC 3.2.1.8 | 225Aa | 19-225 | 212 121 | CD |
| *Aspergillus kawachii* | IFO4308 | XYNC | P33557 | D14848 | EC 3.2.1.8 | 211Aa | 28-211 | 197 106 | CD |

**Table 1 cont.**

| Family and Organism | Strain | Enzyme | Data Base Accession No. Protein | DNA | EC Number | Size | Mature | Proton donor | Domain and comment & Nucleophile |
|---|---|---|---|---|---|---|---|---|---|
| *Aspergillus nidulans* | | XYLA | | Z49892 | EC 3.2.1.8 | 225Aa | | 212 121 | CD |
| *Aspergillus nidulans* | | XYLB | | Z49893 | EC 3.2.1.8 | 221Aa | | 208 117 | CD |
| *Aspergillus niger* | | XYN | | A19535 | EC 3.2.1.8 | 211Aa | | 197 106 | CD |
| *Aspergillus niger* | ATCC90196 | XYN4 | | U39785 | EC 3.2.1.8 | 169Aa | | | |
| *Aspergillus niger* | ATCC90196 | XYN5 | | U39784 | EC 3.2.1.8 | 211Aa | | 197 106 | CD |
| *Aspergillus niger* | IFO4066 | XYNB | | D38071 | EC 3.2.1.8 | 225Aa | | 212 121 | CD |
| *Aspergillus tubigensis* | NW756 | XLNA | | L26988 | EC 3.2.1.8 | 210Aa | | 196 105 | CD |
| *Aureobasidium pullulans* | | XYNA | | U10298 | EC 3.2.1.8 | 221Aa | | 208 114 | CD |
| *Bacillus circulans* | | XLNA | P09850 | X07723 | EC 3.2.1.8 | 213Aa | 29-213 | 200 106 | CD |
| *Bacillus pumilus* | IPO | XYNA | P00694 | X00660 | EC 3.2.1.8 | 228Aa | 28-228 | 209 120 | CD |
| *Bacillus* sp. | YA-14 | XYNS | | X59058 | EC 3.2.1.8 | 213Aa | 29-213 | 200 106 | CD |
| *Bacillus subtilis* | | XYNA | P18429 | M36648 | EC 3.2.1.8 | 213Aa | 29-213 | 200 106 | CD |
| *Bacillus subtilis* | 168trpC2 | XYNA | | Z34519 | EC 3.2.1.8 | 214Aa | | 201 107 | CD |
| *Bacillus stearothermophilus* | No.236 | XYNA | P45705 | U15985 | EC 3.2.1.8 | 211Aa | 20-211 | 198 104 | CD |
| *Cellulomonas fimi* | | XYND | | X76729 | EC 3.2.1.8 | 717Aa | | 216 126 | CD, CBDI, CBDII |
| *Cellvibrio mixtus* | | XYNA | | Z48925 | EC 3.2.1.8 | 656Aa | 23-656 | 212 115 | Lin, CBD |
| *Chaetomium gracile* | | CGXA | | D49850 | EC 3.2.1.8 | 219Aa | | 206 115 | CD |
| *Chaetomium gracile* | | CGXB | | D49851 | EC 3.2.1.8 | 241Aa | | 207 115 | CD |
| *Clostridium acetobutylicum* | P262 | XYNB | P17137 | M31726 | EC 3.2.1.8 | 261Aa | 29-261 | 242 152 | CD |
| *Clostridium stercorarium* | F-9 | XYNA | P33558 | D13325 | EC 3.2.1.8 | 512Aa | 31-512 | 215 124 | CD, 236-CBDI-365, 416-CBDII-504 |
| *Cochliobolus carbonum* | RACE1 | XYL1 | Q06562 | L13596 | EC 3.2.1.8 | 221Aa | 31-221 | 206 115 | CD |
| *Cochliobolus carbonum* | SB111 | XYL2 | | U58915 | EC 3.2.1.8 | 231Aa | 41-231 | 216 125 | CD |
| *Cochliobolus carbonum* | SB111 | XYL3 | | U58916 | EC 3.2.1.8 | 222Aa | 40-222 | 206 113 | CD |
| *Cryptococcus* sp. | S-2 | XYN-CS2 | | D63381 | EC 3.2.1.8 | 209Aa | | 194 101 | |
| *Fibrobacter succinogenes* | S85 | XYNC | P35811 | U01037 | EC 3.2.1.8 | 608Aa | 26-608 | 237 142 | 26-CD-262, 263-Lin-285, 286-CD-529 |
| *Humicola insolens* | | XYL1 | X76047 | | EC 3.2.1.8 | 227Aa | | 212 121 | |
| *Magnaporthe grisea* | Ken60-19 | XYN22 | L37529 | | EC 3.2.1.8 | 233Aa | 40-233 | 217 126 | |
| *Neocallimastix frontalis* | MCH3 | XYN1 | | X82266 | EC 3.2.1.8 | 607Aa | | 229 141 | CD, CD 474 386 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Neocallimastix frontalis* | | XYN2 | | X82439 | EC 3.2.1.8 | 266Aa | | 223 150 | CD |
| *Neocallimastix patriciarum* | | XYNA | P29127 | X65526 | EC 3.2.1.8 | 607Aa | | 223 141 | 30-CD-255, 256-Lin-274, 275-CD-499, 500-Lin-523, 524-Doc-6 |
| *Penicillium purpurogenenum* | | XYNB | | Z50050 | EC 3.2.1.8 | 208Aa | 26-208 | 194 101 | CD |
| *Piromyces* sp. | | XYNA | | X91858 | EC 3.2.1.8 | 625Aa | | 603 510 (133 218) | CD, Lin, Doc, Lin, CD |
| *Pseudomonas fluorescens* subsp. *cellulosa* | | XYNE | | Z48927 | EC 3.2.1.8 | 661Aa | 26-661 | 213 116 | CD, Lin, CBD |
| *Ruminococcus albus* | 7 | XYNA | | U43089 | EC 3.2.1.8 | 680Aa | | 244 147 | CD, Lin, Uk |
| *Ruminococcus flavefaciens* | 17 | XYNA | | Z11127 | EC 3.2.1.8 | 954Aa | | 223 122 | CD(11), 245-Lin-622, 623-CD(10) |
| *Ruminococcus flavefaciens* | 17 | XYNB | | Z35226 | EC 3.2.1.8 | 781Aa | | 226 124 | CD, Uk |
| *Ruminococcus flavefaciens* | 17 | XYND | | S61204 | EC 3.2.1.8 | 802Aa | | 226 124 | CD, Uk, CD |
| *Ruminococcus* sp. | | XYN1 | | Z49970 | EC 3.2.1.8 | 789Aa | | 224 123 | CD, Lin, Uk |
| *Schzophyllum commune* | ATCC388548 | XYNA | P35809 | | EC 3.2.1.8 | 197Aa | | 184 87 | CD |
| *Streptomyces* sp. | S38 | XYL1 | | X98518 | EC 3.2.1.8 | 228Aa | 39-228 | 215 125 | CD |
| *Streptomyces* sp. | EC3 | XLN | | X81045 | EC 3.2.1.8 | 240Aa | 50-240 | 226 135 | CD |
| *Streptomyces lividans* | 66 | XLNB | P26515 | M64552 | EC 3.2.1.8 | 335Aa | 42-335 | 218 128 | CD, CBD |
| *Streptomyces lividans* | 66 | XLNC | P26220 | M64553 | EC 3.2.1.8 | 240Aa | 50-240 | 226 134 | CD |
| *Thermomonospora alba* | | XYLA | | Z81013 | EC 3.2.1.8 | 482Aa | | 169 274 | CD, Lin, CBD |
| *Thermomonospora fusca* | YX | TFXA | | U01242 | EC 3.2.1.8 | 338Aa | | 216 127 | CD, 253-XBD-338 |
| *Trichoderma harzianum* | | | P48793 | | EC 3.2.1.8 | 190Aa | | 177 86 | CD |
| *Trichoderma reesei* | C30 | XYN1 | P36218 | X69574 | EC 3.2.1.8 | 229Aa | 52-229 | 215 126 | CD |
| *Trichoderma reesei* | C30 | XYN2 | P36217 | X69573 | EC 3.2.1.8 | 222Aa | 33-197 | 209 118 | CD |
| Family 12 | | | | | | | | | |
| *Erwinia carotovora* | | CELS | P16630 | M32399 | EC 3.2.1.4 | 264Aa | 33-264 | | CD |
| *Aspergillus aculeatus* | F-50 | | P22669 | D00546 | EC 3.2.1.4 | 237Aa | 17-237 | | CD |
| *Aspergillus kawachii* | | CELK | | D12901 | EC 3.2.1.4 | 239Aa | | | CD |
| *Aspergillus oryzae* | KBN616 | CELA | | D83731 | EC 3.2.1.4 | 239Aa | | | CD |
| *Streptomyces lividans* | 66 | CELB | | U04629 | EC 3.2.1.4 | 408Aa | 41-408 | | |
| *Streptomyces rochei* | A2 | EGLS | | X73953 | EC 3.2.1.4 | 382Aa | | | CD, 277-CBD-382 |
| Family 26 | | | | | | | | | |
| *Bacillus* sp. AM-001 | | MANA | P16699 | M31797 | EC 3.2.1.78 | 513Aa | 27-513 | | |
| *Clostridium thermocellum* | NCIB10682 | CELH | P16218 | M31903 | EC 3.2.1.4 | 900Aa | 45-900 | | Partial similarity N-terminal region, CD is family 5 |
| *Piromyces* sp. | | MANA | | X91857 | EC 3.2.1.78 | 606Aa | | | CD, Lin, Doc |

**Table 1 cont.**

| Family and Organism | Strain | Enzyme | Data Base Accession No. Protein | DNA | EC Number | Size | Mature | Proton donor | Domain and comment & Nucleophile |
|---|---|---|---|---|---|---|---|---|---|
| *Piromyces* sp. | | MANB | | X97408 | EC 3.2.1.78 | 571Aa | | | CD, Lin, Doc |
| *Piromyces* sp. | | MANC | | X97520 | EC 3.2.1.78 | 569Aa | | | CD, Lin, Doc |
| *Prevotella ruminicola* | $B_14$ | CMCASE | | M38216 | EC 3.2.1.4 | 925Aa | | | CD(26), CD(5) |
| *Pseudomonas fluorescens* subsp. *cellulosa* | | MANA | P49424 | X82179 | EC 3.2.1.78 | 419Aa | 39-419 | | |
| *Rhodothermus marinus* | | MANA | P49425 | X09047 | EC 3.2.1.78 | 968Aa | | | |
| Family 44 | | | | | | | | | |
| *Bacillus latus* | | CELA | P29719 | M76588 | EC 3.2.1.4 | 700Aa | 34-700 | | 548-CBD-700 |
| *Caldocellum saccharolyticum* | | MANA | P22533 | L01257 | EC 3.2.1.78 | 1331Aa | 42-1331 | | 42-CD(5)-325, 326-Lin-361, 362-CBD-518, 519-Lin-564, 565-CBD-720, 721-Lin-780, 781-CD(44)-1331 |
| *Clostridium thermocellum* | F1 | CELJ | | D83704 | EC 3.2.1.4 | 1601Aa | | | SLH, Uk, CD(9), CD(44), Doc, Uk |
| *Ruminococcus flavefaciens* | | CELB | | U08621 | EC 3.2.1.4 | 553Aa | | | |
| Family 45 | | | | | | | | | |
| *Fusarium oxysporum* | | CEL | P45699 | L29381 | EC 3.2.1.4 | 376Aa | 19-376 | 140 29 | 19-CD-308, 309-Lin-338, 339-CBD-376 |
| *Humicola insolens* | | EGV | P43316 | | EC 3.2.1.4 | 213Aa | | 121 10 | |
| *Pseudomonas fluorescens* subsp. *cellulosa* | | CELB | P18126 | X52615 | EC 3.2.1.4 | 511Aa | 30-482 | 393 276 | 30-CBD-131, 132-Lin-173, 223-Lin-259 |
| *Trichoderma reesei* | | EGV | P43317 | Z33381 | EC 3.2.1.4 | 242Aa | 18-242 | 134 27 | 18-CD-182, 183-Lin-205, 206-CBD-242 |
| Family 48 | | | | | | | | | |
| *Caldocellum saccharolyticum* | | CELA | | L32742 | EC 3.2.1.4 | 1742Aa | | | CD(9), Lin, CBD, Lin, CBD, CD(48) |
| *Clostridium cellulolyticum* | | CELCCF | P37698 | U30321 | EC 3.2.1.4 | 722Aa | 30-722 | | CD, 667-Doc-722 |
| *Clostridium josui* | | ORF1 | | D16670 | | | | | Partial |
| *Clostridium thermocellum* | ATCC27405 | CELS | P38686 | S56455 | EC 3.2.1.4 | 741Aa | 28-741 | | CD, 679-Doc-743 |
| *Clostridium stercorarium* | NCIB11754 | CELY | P50900 | Z69359 | EC 3.2.1.91 | 914Aa | 34-914 | 34-914 | |
| *Cellulomonas fimi* | ATCC484 | CBHB | P50899 | L38827 | EC 3.2.1.91 | 1090Aa | 54-1090 | | 989-CBD-1090 |

Abbreviation: CD: Catalytic domain; Lin: Linker sequence; Doc: Dockerin domain; Fn3: Fibronectin typeIII domain; CBD: Cellulose-binding domain; XBD: Xylan-binding domain; SLH: Slime layer homologous domain; Uk: Unknown domain; Aa: Amino acids; Rp: Repeated sequence. Number is position of amino acid sequence.
Proton-donor and necleophile residues were deduced by multiple alignment with a 3-D structure known sequence in same family.

cellobiohydrolase (EC 3.2.1.91). The fact that there are differences in substrate specificities among the enzymes belonging to the same family suggests that substrate specificities vary, depending on slight differences in tertiary structures derived from amino acid replacements. On the other hand, the catalytic mechanism of enzymes in a certain family seems to be invariable. Fundamentally, there are two catalytic mechanisms, a retaining mechanism and an inverting mechanism. Hydrolysis of β-glucosidic linkage proceeds stereoselectively by the catalysis of glycosyl hydrolases. In the retaining mechanism, the reaction proceeds in two steps (double displacement reaction) leading to retention of configuration at the level of anomeric carbon. In the inverting mechanism, the reaction proceeds in a single step (single displacement

**Table 2.** Classification of enzymes involved in cellulose and xylan hydrolysis based on amino acid sequence similarities (Henrissat and Bairoch, 1993; Henrissat and Bairoch, 1996; Gebler *et al.*, 1992; Shen *et al.*, 1993; Kawaminami *et al.*, 1995)

| Family | Enzyme | EC 3.2.1.X | Mechanism |
|---|---|---|---|
| 1 | GBLU, PBGAL, PBGLU, LPH, MYRO | 21, 85, 86, 108, (3.2.3.1) | Retaining |
| 3 | BGLU | 21 | Retaining |
| 5 [A] | EG, BMAN, EX-1,3-G, CBH | 4,78,58,91 | Retaining |
| 6 [B] | EG, CBH | 4, 91 | Inverting |
| 7 [C] | EG, CBH | 4, 91 | Retaining |
| 8 [D] | EG, LIC | 4, 73 | Inverting |
| 9 [E] | EG | 4 | Inverting |
| 10 [F] | XYN, CBH | 8, 91 | Retaining |
| 11 [G] | XYN | 8 | Retaining |
| 12 [H] | EG | 4 | Retaining |
| 26 [I] | EG, BMAN | 4, 78 | |
| 39 | BXYL, AIDU | 37, 76 | Retaining |
| 43 | BXYL, ARAF | 37, 55 | Inverting |
| 44 [J] | EG | 4 | Inverting |
| 45 [K] | EG | 4 | Inverting |
| 48 [L] | EG, CBH | 4, 91 | Inverting |

BGLU: β-glucosidase, PBGAL: phospho-β-galactosidase, PBGLU: phospho-β-glucosidase, LPH: lactase-phlorizin hydrolase, MYRO: myrosinase, EG: endoglucanase, BMAN: β-mannanase, EX-1,3-G: exo-1,3-glycanase, CBH: cellobiohydrolase, LIC: lichenase, XYN: xylanase, BXYL: β-xylosidase, AIDU: α-iduronidase, ARAF: α-arabinofuranosidase.

reaction) leading to inversion of configuration at the level of anomeric carbon. Enzymes belonging to families 1, 3, 5, 7, 10, 11, 12 and 39 catalyze hydrolysis of β-glucosidic linkage by the retaining mechanism and enzymes 6, 8, 9, 43, 44, 45 and 48 catalyze hydrolysis by the inverting mechanism (*Table 2*). The former mechanism is represented by *Cellulomonas fimi* exoglucanase/xylanase Cex (recently renamed as XynB) (Tull and Withers, 1994; MacLeod *et al.*, 1994; Tomme *et al.*, 1995b). In the first step, cleavage of the glycosidic bond is promoted by an acidic group, Glu-127 (proton donor), which denotes a proton to the glycosidic oxygen. Cleavage of the bond results in the release of a fragment carrying a new non-reducing end group and the formation of a glycosyl-enzyme intermediate via oxocarbonium ion-like transition states. The nucleophile Glu-233, which was identified by using a mechanism-based inhibitor, participates in the formation of a glycosyl-enzyme intermediate. In the second step, this intermediate is hydrolyzed by Glu-127 via an oxocarbonium ion-like transition state. The intermediate formed during catalysis was trapped using a mecha-

nism-based inhibitor (Tull *et al.*, 1991). Thus, successive inversion leads to overall retention of configuration at the level of the anomeric carbon. The inverting mechanism is represented by *C. thermocellum* endoglucanase CelD (Béguin and Aubert, 1992; Béguin and Lemaire, 1996). Glycosidic oxygen is protonated by a proton donor (Glu-555). Simultaneously, a nucleophile, the negatively charged Asp-201 residue, promotes the ionization of a water molecule. The resulting OH ion attacks the anomeric center directly, thus leading to inversion of the configuration. In this case, no glycosyl-enzyme intermediate is formed while the reaction proceeds via oxocarbonium ion-like transition state.

*Three-dimensional structures of cellulases and xylanases*

The resolution of three-dimensional structures of the enzymes is indispensable for understanding the catalytic mechanism and the difference in substrate specificities among the enzymes belonging to the same family, and for improving further the function of enzymatic properties through protein engineering. In this context, the number of cellulases and xylanases whose three-dimensional structures have been solved is rapidly increasing. Currently three-dimensional structures have been solved for enzymes from families 5, 6, 7, 8, 9, 10, 11 and 45.

The three-dimensional structure of *C. thermocellum* endoglucanase CelC, a member of the largest cellulase family (family 5), is determined by X-ray crystallography at 2.15 Å solution using multiple isomorphous replacement and density averaging between two crystal forms (Dominguez *et al.*, 1995). The protein is a cylindrical $(\alpha/\beta)_8$ barrel, which was first observed in triose phosphate isomerase and hence called the TIM barrel, with two β-bulges at strands 3 and 7 and an acidic cleft containing the active site on the carboxyl-terminal side of the barrel. A segment of 54 amino acids adjacent to the active site cleft of CelC folds into a distinct subdomain consisting of four α-helixes and a short, two-stranded β-structure. The subdomain which is inserted between strand 6 and helix 6 extends the top of the barrel on one side, thus creating a deep substrate-binding cleft. Glu-280, identified as the nucleophile in catalysis, lies at the bottom of the active-site crevice. Cellotriose binds to the bottom of the crevice, adjacent to Glu-280 and in contact with Asn-139, His-90 and Glu-140. Glu-140 is the proton donor in the catalytic reaction. The crystal structure of *Clostridium cellulolyticum* endoglucanase CelCCA, a member of family 5, is also solved at a resolution of 1.6 Å (Ducros *et al.*, 1995). The overall structural fold of CelCCA, $(\alpha/\beta)_8$ barrel, resembles that of CelC. The proton donor, Glu-170, and the nucleophile, Glu-307, are located on β-strands 4 and 7, respectively, as in CelC. Since CelCCA carries a deletion of subdomain present in CelC and the specific specificities of them are slightly different from each other (Schwarz *et al.*, 1988; Fierobe *et al.*, 1991), comparison between the structures of them should disclose the structural determinants influencing catalytic specificity.

*Trichoderma reesei* cellobiohydrolase CbhII, a member of family 6, is the first enzyme whose catalytic domain structure was solved by X-ray crystallography at 2.0 Å resolution (Rouvinen *et al.*, 1990). The molecule is a α/β barrel made up of seven parallel strands of which the first six strands are connected by α-helixes but the connection between the sixth and seventh strands is irregular. This fold is similar to but different from $(\alpha/\beta)_8$ barrel. Two of the loops at the carboxyl-terminal end of the barrel

are extensive. Side chains from these loops and from the barrel form an almost perfectly enclosed 20 Å-long tunnel through which the non-reducing end of the cellulose threads. Asp-221 and Asp-175 residues located in the center of the tunnel are assigned to the proton donor and the nucleophile, respectively. Four clear binding sites (subsites) for glucosyl units referred to as A, B, C and D exist in the tunnel and the site of cellulose cleavage is likely to be between B and C. This structure, therefore, gives the protein cellobiohydrolase activity, i.e., exo-type activity. On the other hand, *Thermomonospora fusca* endoglucanase E2 is also classified in family 6. The fold of the catalytic domain of E2 is equivalent to that of *T. reesei* CbhII, however, there is a significance difference in the organization of the active site (Spezio *et al.*, 1993). In contrast with the finding that the active site of CbhII is almost enclosed, the active site cleft of E2, approximately 11 Å deep and running the entire length of the molecule, is completely free for ligand binding in the crystal. The difference in the organization of the active site between CbhII and E2 suggests that the main difference between endo- and exo-cellulases is the degree to which their active sites are accessible to substrate.

The structure of the family 7 catalytic domain from *T. reesei* cellobiohydrolase CbhI is determined at 1.8 Å resolution (Divine *et al.*, 1994). The protein comprises a large, single domain with overall dimensions of approximately 60 Å by 50 Å by 40 Å. About one-third of this 434-residue domain is arranged in two large antiparellel β-sheets that stack face-to-face to form a β-sandwich. Except for four short α-helixes, the rest of the protein consists almost entirely of loops connecting the β-strands. The two β-sheets are highly curved, forming concave and convex faces that contain seven and eight antiparallel β-strands, respectively. Most of the loops between strands in the convex sheet are short, whereas those in the concave sheet and the connections between sheets are longer. Together with local twists in some of the strands, this creates a ~40 Å-long tunnel that runs the length of the concave sheet. The tunnel is the binding site of the substrate and the active site. Glu-217 is the proton donor and Glu-212 is the nucleophile. This structure is very similar to those of bacterial β-glucanases (lichenases) of family 16 with the main-chain topology of the plant legume lectins. Although family 7 contains endoglucanases in addition to cellobiohydrolase, the three-dimensional structures of endoglucanases in this family have not been reported.

The crystal structure of *C. thermocellum* endoglucanase CelA, a family 8 catalytic domain, is determined at 1.65 Å resolution (Alzari *et al.*, 1996). The protein folds into a regular $(\alpha/\alpha)6$ barrel formed by six inner and six outer α-helixes. The globular core has an overall spherical form about 50 Å in diameter with a long acidic cleft running across the molecular surface at the N-terminal end of the central helixes. Cello-oligosaccharides bind to an acidic cleft which contains at least five glycosyl-binding subsites (A-E from the non-reducing end of the substrate). Although substrate binding promotes no significant conformational rearrangements in the active site, the cellulose chain is clearly bent within the enzyme cleft upon binding. The scissile glycosidic linkage occurs between subsites C and D. The Glu-95 residue is the proton donor in the catalytic mechanism and either Asp-152 or Asp-278 seem to serve as general base catalyst while the general base catalyst must be clearly identified by further structural evidence.

The three-dimensional structure of the family 9 catalytic domain of *C. thermocellum* CelD is determined by X-ray crystallography at 2.3 Å resolution (Juy *et al.*, 1992). CelD has a globular, slightly elongated, shape with rough dimensions of 50 × 50 ×

70 Å. It contains two distinct structural domains, a small N-terminal β-barrel and a larger α-helical domain, the overall topology of which is as in the immunoglobulin-like domain tightly packed against a larger catalytic domain. The latter shows a protein fold shaped like an $(\alpha/\alpha)_6$ barrel of 12 helices connected by loops that form the active site, in which a long, open groove runs across one face of the molecule. The long groove contains the active site. Glu-555 acts as a proton donor in the reaction and Asp-201 is the nucleophile (Chauvaux *et al.*, 1992). Since the structure of the active sites of CelD is similar to that of lysozyme, a scissile glycosidic bond lies between the D and E subsites when glucosyl-subsites are labeled A to F from the non-reducing end of the oligosaccharide, with the catalytic residues positioned on both sides of it. As described above, family 8 CelA folds into the $(\alpha/\alpha)_6$ barrel topology in addition to family 9 CelD, although they have unrelated amino acid sequences (Béguin *et al.*, 1985; Joliff *et al.*, 1986). Both enzymes have an acidic active site cleft on the N-terminal end of the inner α-helices and hydrolyze the glycosidic linkage via an inverting mechanism. Because of dissimilar packing of α-helices within the barrel, however, there are important differences in their structures, i.e., the cross-section is nearly circular in CelA but more elliptical in CelD and the architecture and orientation of the corresponding active site clefts differ considerably.

The three-dimensional structures of the family 10 catalytic domains are reported for *Streptomyces lividans* xylanase XlnA (Derewenda *et al.*, 1994), *C. fimi* Cex (White *et al.*, 1994), *Pseudomonas fluorescens* xylanase XynA (Harris *et al.*, 1994), and *C. thermocellum* xylanase XynZ (Domínguez *et al.*, 1995). The catalytic domain of *S. lividans* XlnA exhibits a tertiary fold of a typical $(\alpha/\beta)_8$ barrel, similar to the structure of family 5 catalytic domains. Seen from the side the molecule has a 'salad bowl' shape. The face of the molecule on the carboxyl-terminal side of the β-barrel (or the top face) has a larger radius, ~45 Å, due to a more elaborate architecture of the β-α loops. The bottom face, consisting of a simple α-β turns, has a radius of approximately 30 Å. The active site is formed by an acidic cleft on the carboxyl-terminal side of the β-barrel. Glu-128 as the proton donor and Glu-236 as the nucleophile are located at the carboxyl-terminal ends of strands 4 and 7, respectively. The three-dimensional structures of other catalytic domains in family 10 are very similar to the protein fold of XlnA described above. Two glutamates, 128 and 236 in XlnA, are conserved as Glu-127 and Glu-233 in *C. fimi* Cex, Glu-127 and Glu-246 in *P. fluorescens* XynA, and Glu-645 and Glu-754 in *C. thermocellum* XynZ. In *C. fimi* Cex, the two key catalytic residues are suitably disposed within the site, with their carboxyl groups facing together at a separation of 5.5 Å. This separation is similar to that in other retaining β-glycanases whose structures have been determined, i.e., 5.4 Å between Glu-78 and Glu-172 for the *Bacillus circulans* xylanase (Campbell *et al.*, 1993) and 5.5 Å between Asp-52 and Glu-35 of hen egg white lysozyme (Imoto *et al.*, 1972). Such a separation is presumably optimal for the efficient formation of a glycosyl-enzyme intermediate on Glu-233 of Cex, while at the same time allowing Glu-127 to protonate the departing aglycone in a concerted manner. Xylopentaose binds to five subsites A-E of *Pseudomonas fluorescens* xylanase XynA and is cleaved between subsites D and E.

Three-dimensional structures of family 11 catalytic domains are reported for *Bacillus pumilus* xylanase XynA (Katsube *et al.*, 1990), *Trichoderma harzianum* xylanase (Campbell *et al.*, 1993), and *T. reesei* xylanase XynI (Törrönen and Rouvinen,

1995) and XynII (Törrönen *et al.*, 1994; Törrönen and Rouvinen, 1995). XynI and XynII, major xylanases from *T. reesei*, are highly homologous (sequence identity approximately 50%). Overall structures of XynI and XynII resemble each other, i.e., both enzymes exist as a single domain that contains two mostly antiparallel β-sheets which are packed against each other. The β-sheet structure is twisted, forming a large cleft on one side where the active site is situated. The proton donor and the nucleophiles are Glu-164 and Glu-75 in XynI, and Glu-177 and Glu-86 in XynII, respectively. The width of the active site cleft and the number of subsites are different in these enzymes. The active site of XynI is narrower in XynI and appears to contain only three subsites (A-C from the non-reducing end of the substrate) and the scissile glycosidic linkage occurs between subsites B and C. On the other hand, XynII most probably contains five subsites (A-E from the non-reducing end of the substrate) and the scissile glycosidic linkage occurs between subsites B and C. In spite of overall resemblance of their tertiary structures, there is a difference in their pH optimum. XynI is an acidic xylanase, being active in the pH range of 3–6 with maximum activity at pH 3.5, whereas XynII is active over a larger pH range of 4–8, having a maximum at pH 5.3. The pH optimum depends mostly on the properties of the acid/base catalyst. The conformational change observed at pH 6.5 in XynII places the carboxylate group of Glu-177 in a totally different position to change its p*K*a value, leading to the shift of the active pH range of XynII. Alternatively, the pH optimum difference between these enzymes is explained by the variation in the hydrogen bonding of the acid/base catalyst. In XynI, Asp-33 makes a strong hydrogen bond to Glu-164. This lowers the p*K*a value of this glutamic acid. On the other hand, the Asp-33 residue of XynI is replaced with an asparagine residue (Asn-44) in XynII. Therefore, interaction between Glu-177 and Asn-44 in XynII is weaker than the corresponding one in XynI. This may raise the p*K*a value of Glu-177 and affects the pH optimum of the enzyme. In case of *B. pumilus* XynA (Katsube *et al.*, 1990), although the structure was predominantly characterized as three large β-sheets but not two β-sheets, it is actually similar to the structures of *T. reesei* xylanases.

An endoglucanase of family 12 from *Aspergillus aculeatus* shows no sequence homology on the level of primary structure, but has a three-dimensional structure similar to that of family 11 catalytic domains described above (Okada, 1991; Törrönen *et al.*, 1993).

Among family 45 cellulases, the catalytic domain of *Humicola insolens* endoglucanase EGV is determined by X-ray analysis at 1.6 Å (Davies *et al.*, 1995). The enzyme has a flattened spheroidal shape with rough dimensions of 42 × 42 × 22 Å. The major structural feature is a six-stranded β-barrel domain. The barrel core of the structure is similar to that found in 'barwin', a plant defence protein (Ludvigsen and Poulson, 1992). A large deep groove runs across the surface of the molecule, partitioning the β-barrel from the loop region. Two aspartates, Asp-10 and Asp-121, are identified as critical residues for activity by site-directed mutagenesis. They are located on either side of the groove with their Cα atoms some 11.5 Å apart. EGV has seven subsites for sugar binding, subsites A-G, aiding cleavage of the glycosidic bond between subsites D and E.

As described above, comparison of the crystalline structures shows that there is common protein fold and similar active site in two unrelated families, i.e., families 5 and 10 (Domínguez *et al.*, 1995), and families 11 and 12 (Törrönen and Rouvinen,

1995; Okada, 1991). Jenkins *et al.* (1995) compared the recently determined crystalline structures and sequences of β-glycanases, i.e., family 5 cellulase, family 10 xylanase, 1,3-β-glucanase, 1,3-,1,4-β-glucanase, and β-galactosidase and concluded that they belong to a superfamily of 8-fold α/β barrels with similar amino acid residues at their active sites. They termed this group the 4/7 superfamily on the basis of the findings that the nucleophile, a glutamate residue, is located close to the carboxyl-terminus of β-strand seven and these enzymes have the sequence Asn-Glu close to the carboxyl-terminus of β-strand four. On the other hand, Henrissat *et al.* (1995) compared the regions surrounding the catalytic amino acids previously identified in a few retaining glycosyl hydrolases, by hydrophobic cluster analysis, and they concluded that the enzymes of families 1, 2, 5, 10, 30, 35, 39 and 42 which share a $(\alpha/\beta)_8$ barrel structure evolved from the same ancestral $(\alpha/\beta)_8$ barrel structure and acquired a diversified substrate specificity through evolutionary events. More recently, Henrissat and Bairoch introduced a new concept 'clan' into classification of glycosyl hydrolase (Henrissat and Bairoch, 1996). Clan grouping of glycosyl hydrolase is shown in *Table 3*. A 'clan' is a group of families that are thought to have a common ancestry and are recognized by significant similarities in tertiary structure together with conservation of the catalytic residues and catalytic mechanism.

**Table 3.** Clan grouping of glycosyl hydrolase families (Henrissat and Bairoch, 1996)

| Clan | Families grouped |
|---|---|
| GH-A | 1, 2, 5, 10, 17, 30, 35, 39, 42, 53 |
| GH-B | 7, 16 |
| GH-C | 11, 12 |
| GH-D | 27, 36 |
| GH-E | 33, 34 |

Family 1: contains β-glucosidases, 6-phospho-β-galactosidases, 6-phospho-β-glucosidases, pactase-phlorizin hydrolases, myrosinases
Family 2: β-galactosidases, β-glucosidases
Family 5: endoglucanases, β-mannanases, exo-1,3-glycanases
Family 10: xylanases
Family 17: endo-1,3-β-glucosidases, lichenases
Family 30: mammalian glucosylceramidases
Family 35: β-galactosidases
Family 39: mammalian α-L-iduronidases, bacterial β-xylosidases
Family 42: β-galactosidases
Family 53: endo-1,4-β-galactanases
Family 7: endoglucanases, cellobiohydrolases
Family 16: lichenases
Family 11: xylanases
Family 12: endoglucanases
Family 27: α-galactosidases, α-N-acetylgalactosaminidase
Family 36: α-galactosidases
Family 33: sialidases
Family 34: sialidases (influenza neuraminidases)

### *Improvement of cellulase function by mutagenesis of catalytic domains*

It is generally believed that enzymes have become well-adapted to their physiological environment and are at their optimum state. If this is true, the improvement of enzymes has to be carried out by well-designed site-directed mutagenesis rather than by random mutagenesis. Accumulation of data on three-dimensional structures of cellulases and

xylanases enables us to identify amino acid residues of a certain enzyme which are important for expressing its characteristic properties and to introduce mutations in the target positions to change the properties by site-directed mutagenesis. One successful example of such mutagenesis is conversion of exocellobiohydrolase activity of *C. fimi* CbhA to endo-glucanase activity by deletion of a surface loop (Meinke *et al.*, 1995). As described above, the comparison between three-dimensional structures of *T. reesei* CbhII and *T. fusca* E2 belonging to family 6 suggested that the basic difference between exo- and endoglucanase is the accessibility of their active sites to internal β-1,4-glucosidic bonds in polymeric substrates, i.e., the exoglucanase contains the tunnel-shaped active site which restricts hydrolysis to β-1,4-glucosidic bonds at the ends of cellulose molecules, reversely, the endoglucanase has the active site which is open against the substrate. If this is the case, the removal of a surface loop of the exoglucanase covering its active site to form a tunnel should unhinder the active site against the substrate, leading to enhancement of endoglucanase activity. When this hypothesis was examined by deletion of a region in *C. fimi* cellobiohydrolase CbhA corresponding to part of the carboxyl-proximal loop of *T. reesei* CbhII, the mutation actually enhanced the endoglucanase activity of the enzyme on CMC and altered its activities on some small substrates, suggesting that academic researches on the structures and functions of many cellulases and xylanases lead to applicative studies such as molecular breeding of the enzymes.

Similarly, data on three-dimensional structures of cellulases and xylanases may enable us to change their enzymatic properties, such as optimum pH and substrate specificities. Since difference in optimum pHs between closely related xylanases XynI and XynII of *T. reesei* were ascribed to different hydrogen bondings of the catalytic residues (see above), displacement of the amino acid residues involved in the hydrogen bonding formation with the catalytic amino acids may change the optimum pH of the target enzyme. Both *C. cellulolyticum* CelCCA and *C. thermocellum* CelC are endoglucanases belonging to family 5. The former is active on xylan as well as on cellulosic materials (Fierobe *et al.*, 1991) but the latter is not (Schwarz *et al.*, 1988). *C. fimi* Cex (XynB) in family 10 hydrolyses CMC in exo-mode in addition to xylan (Gilkes *et al.*, 1991) while *C. thermocellum* XynZ (Grépinet *et al.*, 1988a, b), in the same family, is not active on CMC. These differences in the substrate specificities are thought to be due to slight differences in tertiary structures at active sites; detailed comparison between the structures of these enzymes should disclose the structural determinants influencing catalytic specificity, leading to an artificial change of substrate specificity of a target enzyme.

In general, enzymes from thermophilic organisms are thermophilic and those from mesophilic organisms are mesophilic even though they belong to the same family. This suggests that enzymes may not be at their optimal state with respect to stability, i.e., stability-increased mutants may be isolated by artificial mutagenesis. Since the mechanism of stabilization of enzymes is not clear, unfortunately, random mutagenesis should be a strong method for the stabilization of enzymes. The *xynA* gene of *B. punilus* was randomly mutagenized by chemical reagents (Arase *et al.*, 1993). As a result of screening for heat-resistant mutants, four heat-resistant mutants were selected from 60,000 mutant genes, and Gly-38 and Ser-12 were assigned as the main target points for stability-increasing mutations. By substituting other amino acid residues at these points, the mutations at these points can be optimized for the stabilization of this

enzyme. In conclusion, random mutagenesis provides us not only with some desired mutants but also with many promising strategies for further improvement of the mutants if the three-dimensional structure of the enzyme is known.

## CELLULOSE-BINDING DOMAINS (CBDS) OF CELLULASES AND XYLANASES

Efficient enzymatic degradation of insoluble polysaccharides often requires a tight interaction between the enzymes and their substrates. In the case of cellulose degradation, many cellulases are known to bind to crystalline and/or amorphous cellulose *via* cellulose-binding domains (CBDs) which are distinct from catalytic domains. In most cases, CBDs are separated from catalytic domains by linker sequences which are highly enriched in proline and hydroxyamino acids as described below. CBDs are also found in xylanases (Sakka *et al.*, 1993; Millward-Sadler *et al.*, 1994) and other plant cell wall hydrolases, such as α-L-arabinofuranosidase and acetylxylan esterase (Ferreria *et al.*, 1993) and β-mannanase (Stalbrand *et al.*, 1995) in addition to endoglucanases and exoglucanases. It seems illogical for an enzyme to have an affinity with a substrate which cannot be hydrolysed. But this is not surprising since various types of enzymes, including xylanase and other hemicellulases, are necessary to act cooperatively on plant cell walls which consist of cellulose, hemicellulose and lignin and the binding of hemicellulases to cellulose *via* CBDs should be advantageous in hydrolysis of hemicellulose in plant cell walls. Although xylan-binding domains have also been reported (Black *et al.*, 1995) in xylanase, the widely distributed substrate-binding domains in xylanases are probably CBD, due to the high frequency of adsorption to the homogeneous cellulose. If xylan-binding domains specifically evolved for each xylan, a variety of xylan-binding domains have to be examined because of the heterogeneity of xylan caused by variations in plant growth conditions such as weather or plant species. It might be ineffective to waste the gene capacity for only xylan degradation. Therefore, xylanases have chosen the simplest way to integrate CBD. CBDs are present in both bacterial and fungal enzymes. CBDs exhibit cellulose-binding activity even if they are separated from catalytic domains by proteolysis (Gilkes *et al.*, 1988; Changs and Wilson, 1988; Owolabi *et al.*, 1988; McGavin and Forsberg, 1989; Jauris *et al.*, 1990) or by gene manipulation (Din *et al.*, 1991; Coutinho *et al.*, 1992; Ong *et al.*, 1993; Goldstein *et al.*, 1993; Brun *et al.*, 1995; Morag *et al.*, 1995). Furthermore, CBDs retain their functions in hybrid proteins which consist of a CBD and a catalytic domain derived from different origins (Poole *et al.*, 1991; Maglione *et al.*, 1992; Tomme *et al.*, 1995a; Karita *et al.*, 1996).

Klyosov reported that for effective hydrolysis of crystalline cellulose, not only the quantity of cellulases but also the ability to absorb tightly on cellulose is essential. Some tightly bound enzymes induce expansion of the cellulose surface followed by a release of small particles (Klyosov, 1990). This suggests that binding of enzymes to cellulose is one of the important factors for degradation of cellulose. Removal of CBDs from some cellulases severely reduces their activities against insoluble cellulose (Tomme *et al.*, 1988; Gilkes *et al.*, 1988; Ghangas and Wilson, 1988; Stahlberg *et al.*, 1988; Irwin *et al.*, 1933; Hall *et al.*, 1995). A CBD from *C. fimi* CenA expressed in *E. coli* without any catalytic activity disrupts the structure of cellulose fibres and fragments it, but produces no detectable amount of reducing sugars (Din *et al.*, 1991). Conversely, artificial addition of a heterologous CBD on a catalytic domain increases

its specific activities on insoluble cellulose allomorphs (Maglione *et al.*, 1992). These results show that CBDs seem to have a role in disruption and dispersion of cellulose crystalline. On the other hand, there is a report stating that the function of CBD is restricted to enhancing cellulose-hydrolysing efficiency of the catalytic domain by increasing the adsorption partition coefficient on crystalline cellulose (Nidetzky *et al.*, 1994). The functions of CBDs may vary depending on their origins.

Strength of binding of CBD to the substrate differs according to the source of the CBD or CBD family (see below). The CBDs from cellulases and xylanases of *P. fluorescens* subsp. *cellulosa* exhibit high affinity to cellulose and the enzyme once adsorbed onto cellulose can be quantitatively eluted from cellulose-enzyme complex by boiling in 10% SDS (Poole *et al.*, 1991). The CBDs of CenA from *C. fimi* (Gilkes *et al.*, 1988) and an endoglucanase E2 from *T. fusca* (Ghangas *et al.*, 1989) bind tightly to cellulose and can be eluted by 8M and 6M guanidine hydrochloride, respectively. Interaction between cellulose of CelE from *C. thermocellum* (Durrant *et al.*, 1991) or CelZ from *C. stercorarium* (Jauris *et al.*, 1990) is affected by ionic strength, and these enzymes can be recovered by elution with distilled water. XynA from *C. stercorarium* can be desorped from the cellulose-XynA by cellobiose (Takada *et al.*, 1996).

By comparing the amino acid sequences of CBDs, they are grouped into ten families, i.e., Families I to X (Tomme *et al.*, 1995c) using Roman numerals, as proposed by Coutinho *et al.* (1992), to avoid confusion with the Arabic numbering scheme for catalytic domain classification (Henrissat and Bairoch, 1993).

Family I CBDs, including 36 members, are peculiar to fungal enzymes, including endoglucanases, cellobiohydrolases, xylanases and mannanases (Tomme *et al.*, 1995a). These are composed of a relatively small number of amino acids, about 36, and the amino acid sequences are highly conserved. The CBD of cellobiohydrolase I (CBHI) from *T. reesei* is a wedge-shaped β-sheet structure composed of three antiparallel strands with one hydrophobic and one hydrophilic face (Kraulis *et al.*, 1989). The hydrophilic surface contains three conserved tyrosine residues. One of them (Y492), located at the tip of the wedge-shaped domain, is essential for functionality (Reinikainen *et al.*, 1992, 1995). At least two tyrosine residues, including Y492, and a glutamine residue on the hydrophilic face are essential for tight binding of CBD to cellulose (Linder *et al.*, 1995a). The cellulose-binding affinities of CBDs of CBHI and endoglucanase I (EGI) have been compared. The CBD of EGI has significantly higher affinity than that of CBHI. This difference could be ascribed to a replacement of a tyrosine by a tryptophan on the hydrophilic face (Linder *et al.*, 1995b). The binding of CBHI *via* CBD to cellulose was significantly affected by ionic strength, suggesting that interaction between cellulose and CBD includes hydrophobic effect (Reinikainen *et al.*, 1995).

Family II CBDs are composed of about 100 amino acid residues and are found in bacterial hydrolases such as cellulases, xylanases, one α-L-arabinofuranosidase, two acetylxylan esterase and two chitinases. Presently, 36 CBDs are known to belong to this family (Tomme *et al.*, 1995b), in which several highly conserved aromatic residues are implicated in substrate binding (Poole *et al.*, 1993; Din *et al.*, 1994a). Structural analysis by NMR indicates that the family II CBD from the *C. fimi* exoglucanase Cex forms an extensive β-sheet structure with a β-barrel fold. Two tryptophan residues participate in cellulose binding (Xu *et al.*, 1995). The family II CBD isolated from *C. fimi* endoglucanase CenA disrupts the structure of cellulose

fibres under electron microscopic observations (Din *et al.*, 1991). When the CBDs from CenA and Cex, which have no detectable cellulolytic activity, are incubated with crystalline cellulose, they release small particles from the cellulose (Din *et al.*, 1991; Tomme *et al.*, 1995a), suggesting that these CBDs loosen hydrogen bonds between microfibrils of cellulose and degrade the crystal structure without any cleavage of covalent bonds. Furthermore, the synergistic effect is observed between the CBD of and the catalytic domain from CenA upon hydrolysis of crystalline cellulose (Din *et al.*, 1994b). These observations clearly indicate the significance of CBDs present in *C. fimi* cellulases in cellulose hydrolysis.

Family III CBD has 24 members of bacterial origin. About 170 amino acids residues can be aligned as a conserved region. In this family, CBDs are located not only in catalytic proteins but also in non-catalytic proteins such as a cellulosome integrating protein (CipA) from *C. thermocellum*, which is a scaffolding protein of a cellulosome, a complex of cellulases (see below). Also, there are several conserved aromatic amino acid residues in this family. The binding capacity of the CBD from *C. thermocellum* CipA with amorphous cellulose is about 20-fold higher than that with crystalline cellulose (Morag *et al.*, 1995). The CBD from CbpA, a scaffolding protein of a cellulosome of *Clostridium cellulovorans*, was expressed in *E. coli* and characterized. This CBD prefers crystalline cellulose to amorphous cellulose (Goldstein *et al.*, 1993). Mutation analysis of this CBD shows that the entire region of the CBD is required for maximal binding to crystalline cellulose (Goldstein and Doi, 1994). Both CBDs from these scaffolding proteins also can bind to chitin, a polymer of *N*-acetylglucosamine in β-1,4 linkage (Goldstein and Doi, 1994; Morag *et al.*, 1995).

Family IV currently consists of 6 CBDs which have weak sequence similarity with family II CBDs and are now classified in an independent family. The two tandem CBDs in CenC from *C. fimi* bind unusually to amorphous cellulose but not to crystalline cellulose (Coutinho *et al.*, 1992).

Family V includes only a single representative from *Erwinia chrysanthemi* EGZ (Guiseppi *et al.*, 1988). This CBD consists of 60 amino acid sequence and has a dispersion effect of cellulose crystal like family II CBDs (Brun *et al.*, 1995).

Family VI include CBDs, which consist of about 90 amino acids, from various xylanases. Two CBDs are located tandemly in C-terminal of *Clostridium stercorarium* XynA containing family 11 catalytic domain (Sakka *et al.*, 1996). These CBDs function independently as CBD and bind preferentially to amorphous cellulose such as ball-milled cellulose and acid-swollen cellulose. The binding to cellulose of XynA is inhibited by cellobiose and cellobiose can release the enzyme once adsorbed to cellulose from cellulose (Takada *et al.*, 1996). When the tandem CBDs were joined to the endoglucanase EGIV of *Ruminococcus albus*, hybrid enzyme showed an enhanced activity toward insoluble cellulose at low concentration, but no change of specific activities against the same substrate was observed at the saturated level of the substrate (Karita *et al.*, 1996), suggesting that the CBDs of XynA allow the enzyme to bind to the substrate and increase the substrate concentration around the enzyme.

Families VII and VIII also consist of only a single representative from *C. thermocellum* CelE and *Dictyostelium discoidum* CelA, respectively.

Family IX CBDs are present as tandem repeats at the C-termini of thermostable xylanases, such as *Thermotoga maritima* XynA, *C. thermocellum* XynX and *Thermoanaerobacterium saccharolyticum* XynA, with family 10 catalytic domains

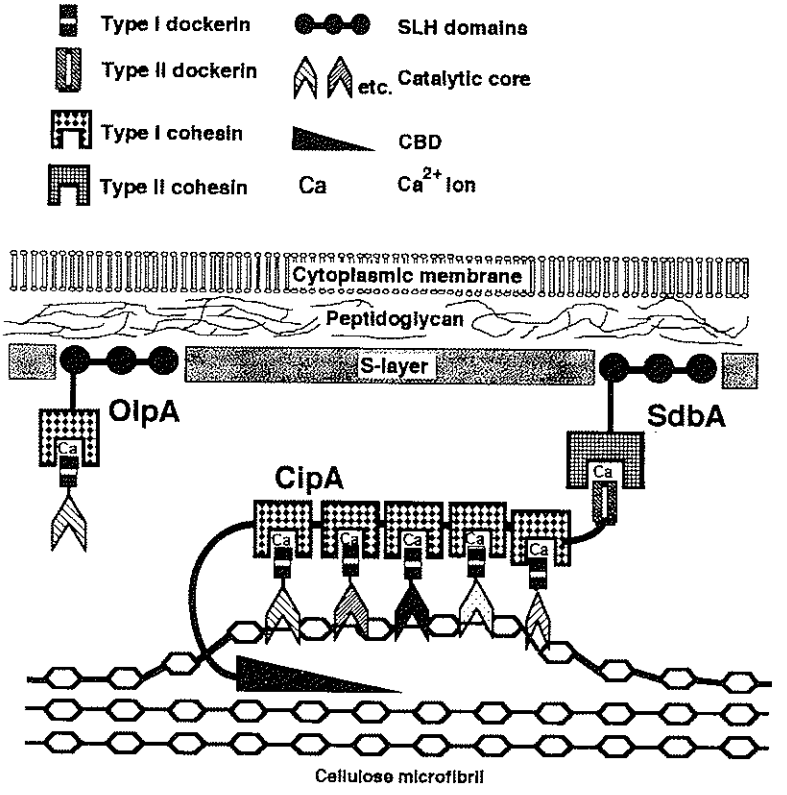


**Figure 1.** Hypothetical organization of the *C. thermocellum* cellulosome. Nine type I cohesins of a scaffolding protein CipA interact with type I dockerins on various catalytic subunits. A CBD of CipA adsorbs onto cellulose and disorders its crystalline structure to supply the catalytic subunits with easily hydrolyzable substrate, i.e., amorphous region in the cellulose. CipA is attached to the cell by SdbA containing SLH domains and a type II cohesin. Interaction between cohesins and dockerins is mediated by $Ca^{2+}$ ion. OlpA containing a type I cohesin and three SLHs anchor a protein containing a type I dockerin.

(Winterhalter *et al.*, 1995). Recently CBDs in this family have also been found in a mesophilic xylanase, xylanase C from *C. fimi* (Clarke *et al.*, 1996) and xylanase A from *Eubacterium ruminantium* (Kobayashi *et al.*, 1996).

Family X CBDs consist of 55 amino acid residues and are found in cellulases and xylanases from *P. fluorescens* subsp. *cellulosa* (Tomme *et al.*, 1995c).

An increasing awareness of the importance of the CBDs in cellulose degradation will stimulate our interest in CBDs and allow us to identify and analyze novel CBDs. It seems certain that new CBD families will be added to the present classification in the future.

Since CBDs specifically bind to cellulosic materials and are eluted from the cellulose-protein complex by procedures suitable for respective CBDs, they can be used as a cellulose-binding ligand which is available for purification and immobilization of a fusion protein with CBDs (Ong *et al.*, 1989, 1995; Greenwood *et al.*, 1992; Assouline *et al.*, 1993). Recently, a plasmid vector containing a CBD encoding gene was constructed, by which a fusion protein with a parent protein on an insert gene is

expressed in *E. coli*. As a result the fusion protein with CBD as a tag is readily purified with a single step by using a cellulose column as an affinity ligand (Graham *et al.*, 1995).

## THE CELLULOSOME OF *C. THERMOCELLUM*

Aerobic cellulolytic fungi and bacteria secrete a combination of endoglucanases and exoglucanases (cellobiohydrolases), as non-complexed systems, into the surrounding environment. In these non-complexed cellulolytic systems, synergistic action of individual enzymes is the focus of many studies on cellulase interactions and has been reviewed recently (Tomme *et al.*, 1995b).

On the other hand, multienzyme complexes having high activity against crystalline cellulose, known as a cellulosome, were identified and characterized in cellulolytic clostridia such as *C. cellulolyticum, C. cellulovorans, C. papyrosolvens* and *C. thermocellum* (for reviews, see Béguin and Lemaire, 1996; Bayer *et al.*, 1994; Doi *et al.*, 1994; Felix and Ljungdahl, 1993), and anaerobic cellulolytic fungi such as *Neocallimastix patriciarum* and *Piromyces* sp. (for a review, see Teunissen and Op den Camp, 1993). Among the cellulases of these organisms, genetic and biochemical analyses of the *C. thermocellum* cellulosome have progressed rapidly in recent years. The strong cellulolytic activity of the *C. thermocellum* cellulase system must be ascribed to the ordered structure of the cellulosome but not to simple synergism of individual cellulases, because each enzyme purely isolated can hardly hydrolyze crystalline cellulose by itself. Therefore, we describe here in brief the mechanism of cellulosome assembly deduced from the recent findings.

The *C. thermocellum* cellulosome is a high molecular mass, extracellular cellulase complex composed of at least 14 proteins in *C. thermocellum* strain YS (Lamed *et al.*, 1983) and 50 proteins in strain JM20 (Morag *et al.*, 1992), detectable on sodium dodecyl sulfate-acrylamide gel electrophoresis, ranging in molecular weight from 20,000 to 250,000. In the cellulosome, catalytic subunits such as endoglucanases, exoglucanase and xylanases are held together by a non-catalytic scaffolding protein (cellulosome integrating protein, CipA) to assemble into a huge complex. Hypothetical organization of the cellulosome and the structures of its catalytic or structural components are schematically depicted in *Figure 1* and *Figure 2*. CipA, which comprises a CBD, a type II dockerin domain, and nine cohesin domains (Gerngross *et al.*, 1993), is thought to have two main functions, i.e., first, it binds a series of catalytic subunits together to form a complex; second, it causes the complex to adsorb onto cellulose and it may disorder the crystalline structure of cellulose to produce easily hydrolyzable substrate. Cohesin domains are subunit-binding domains which are responsible for integrating catalytic subunits into the cellulosome complex and all amino acid sequences of cohesins (about 160 amino acids) are highly conserved, i.e., 57%–100% sequence identities are observed between respective cohesins. Dockerin is a docking domain present in catalytic subunits that interacts with a cohesin domain of CipA but one dockerin exits in CipA itself. Catalytic subunits, therefore, are composed of at least two domains, a catalytic domain and a dockerin domain, and some enzymes contain additional functional domains such as CBD and SLH. A dockerin domain comprises two highly conserved duplicated sequences each composed of about 24 amino acids (*Figure 3*). The catalytic subunits bind strongly to the

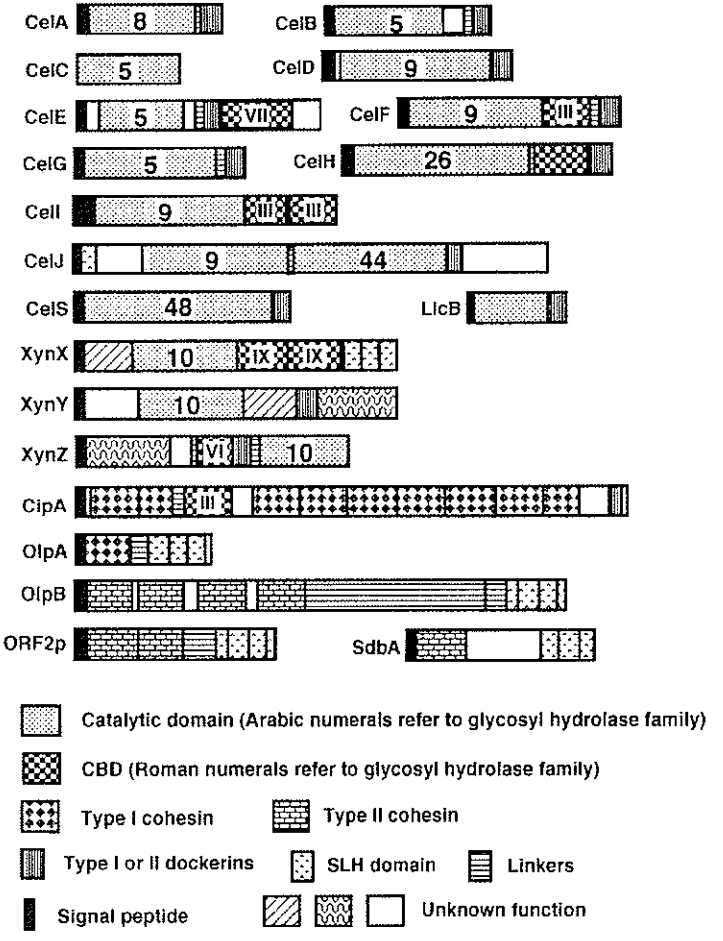


**Figure 2.** Schematic representation of catalytic subunits and structural proteins of the *C. thermocellum* cellulosome showing the diversity of structural and functional organization. The lengths of the domains are proportional to the number of amino acid residues. Catalytic and cellulose-binding domain families are indicated by numbers. CelS, CelS, and XynZ contain a type I dockerin and CipA, a type II dockerin. Regions of unknown function showing significant similarity are filled with the same pattern.

CipA protein to assemble into a cellulosome by association of a dockerin of the former and a cohesin of the latter. The dockerins of CelD (Tokatlidis *et al.*, 1993), XynZ (Tokatlidis *et al.*, 1991), and CelS (Kruus *et al.*, 1995) bind to CipA. It has recently been demonstrated that the interaction between cohesins and dockerins is mediated by $Ca^{2+}$ ions and the first stretch of the tandem repeats of a dockerin is important for $Ca^{2+}$-binding (Choi and Ljungdahl, 1996a, b), while a part of the repeats was found to be similar to EF hand-type $Ca^{2+}$-binding sites and the CelD was actually found to bind to $Ca^{2+}$ in the earlier experiment (Chauvaux *et al.*, 1990). Surprisingly, the dockerin of CipA itself cannot bind to the seventh cohesin of CipA (Salamitou *et al.*, 1994) although the dockerin of CipA is highly homologous with the other dockerins reported

from *C. thermocellum* (*Figure 3*). Therefore, dockerins found in catalytic subunits and CipA are classified into 'type I' and 'type II', respectively. As a receptor domain for the type II dockerin of CipA, a new type of cohesin domain has recently been identified in the protein encoded by the *sdbA* (scaffoldin dockerin binding) gene (Leibovits and Béguin, 1996). Since the new cohesin of SdbA is not homologous with the cohesins of CipA, the former is termed 'type II' cohesin and the latter, 'type I' cohesin. Type II cohesin domains are conserved in *C. thermocellum* OlpB and ORF2p (Fujino *et al.*, 1993). Since SdbA contains three repeats that are highly similar to the segments termed S-layer homologous (SLH) domain present in several proteins located on the cell surface of various bacteria (Lupas *et al.*, 1994), it appears to play a role for anchoring CipA, and therefore the cellulosome, to the cell surface (Leibovits and Béguin, 1996). Both OlpB and ORF2p which comprise four and two type II cohesins, respectively, and SLH domains will also participate in anchoring CipA to the cell surface. On the other hand, OlpA that consists of a type I cohesin and three SLHs may anchor a catalytic subunit on the cell surface (Salamitou *et al.*, 1994). The endoglucanase CelJ has a dockerin highly homologous to many other dockerin domains (Ahsan *et al.*, 1996). This domain may be classified into type II, for the second and third cohesins of CipA from *C. thermocellum* YS did not associate with a catalytic subunit S2 that was equivalent to CelJ (Yaron *et al.*, 1995). Up to the present, only CelD, CelS and XynZ as catalytic subunits were experimentally shown to associate with CipA although many cellulase and xylanase genes and their translated products were characterized. Therefore, the third type of dockerin and cohesin may be found in catalytic subunits and unknown structural protein(s). SLH domains are found in not only structural proteins such as SdbA, OlpA and OlpB but also in catalytic subunits such as CelJ (Ahsan *et al.*, 1996) and XynX (GenBank accession no M67438). It is possible that SLHs of the catalytic subunits associate directly with the peptidoglycan to anchor these proteins, which may lead to the anchorage of the cellulosome containing these proteins to the cell surface.

## Applications of cellulases

### CELLULASES IN DETERGENT

Synthetic heavy-duty detergents now commercially available have good detergency as a result of numerous technical improvements over the 40 years since the invention of detergents. At present, the degree of detergency seems to have peaked; all detergents contain similar ingredients and are based on the same mechanisms. In these mechanisms, soil adsorbed onto the surfaces of fibers or in their interstices is removed by surfactants and builders, which lower interfacial tension and enhance the repulsive force between the soil and the fabric. All of these mechanisms involve interaction between the ingredients of the detergent and the surfaces of the fabric and the soil. According to the electron microscopic observation by Murata *et al.* (1991), the soil and oily drops, however, still remain in the interior of the cotton fibers even though proteases and lipases are often used to hydrolyze protein and oil in soil. Therefore, another washing mechanism should be employed for releasing the soil in different kinds of fibers. For this purpose, cellulase treatment of the soiled cotton cloths was tried as a new washing method. A cellulase compatible with the alkaline ingredients

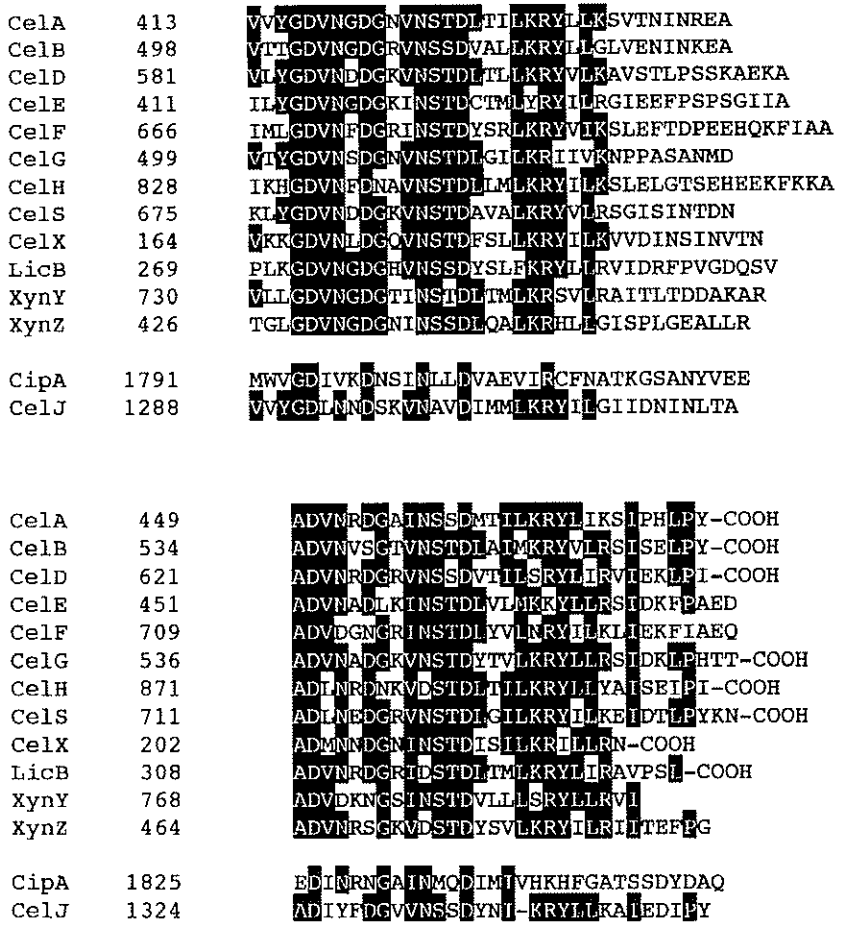


**Figure 3.** Alignment of type I and II dockerin domains from the *C. thermocellum* enzymes. Residues that are identical or similar in the majority of the displayed sequences are highlighted. Numbering of residues starts with putative initiation codon. Among these dockerins, those of CelD, CelS and XynZ were experimentally shown to belong to type I. The dockerin of CipA belongs to type II and that of CelJ seems to belong to type II or the third type.

of heavy-duty detergents and suitable for use in washing at low temperature is required. Therefore, a novel alkaline cellulase with these properties was screened for and found from *Bacillus* sp. KSM (Ozaki *et al.*, 1995), which catalyzes endohydrolysis under the conditions of an optimum pH at 9.5, an optimum temperature at 40°C and an activity of 1,500 units/g (Murata *et al.*, 1991, 1993).

Cotton cloth artificially soiled with oleic acid was incubated at pH 9 and 30°C for 120 min in the alkaline cellulase solution in the presence of 0.1 wt% heptaoxyethylene dodecyl ether, a detergent or in inactivated alkaline cellulase solution in the presence of the same surfactant. As the concentration of alkaline cellulase in the detergent solution increases, the amount of oleic acid remaining in the cotton cloth decreases. However, adding a coresponding amount of inactivated alkaline cellulase does not

affect the residual amount of soil. Thus, it is confirmed that the soils, which are not to be substrates for cellulase, are removed from the cotton cloths by active cellulase. This indicates that the indirect influence of the alkaline cellulase such as its fiber-degrading action releases the oily soil from the amorphous fiber spaces, to where the ordinary surfactant cannot penetrate to reach the soil but the alkaline cellulase might have penetrated into the amorphous space of fibers by adsorbing with CBD, as was mentioned before in this review. Depending on the increase in the amount of adsorption of alkaline cellulase, the relative crystallinity index was increased. This indicates that the adsorption of alkaline cellulase decreases in the amorphous part of insoluble cellulose powder, leading to the conclusion that the more cellulase adsorbed to the cellulose, the more cellulose is hydrolyzed. This hydration may allow for easier enzyme action. In the washing experiment with the alkaline cellulase, however, a negligible increase in reducing sugar was observed, suggesting that the cotton fiber was not solubilized but cleaved inside the fiber to release oily soil due to the endotype enzymic action. This result was well supported from electron microscopic observation. Therefore, attack on cotton fiber with the alkaline cellulase may relax the rigidity of the fiber to release soil soil contained in the interior of the fiber.

Enzymatic bio-polishing with cellulase preparations is familiar as a biological finishing process for textiles made from cellulosic fibers (Lange, 1993). Preparations with alkaline pH optima from *Humicola* or *Bacillus* spp. are suitable as detergent additives to assist soil removal and improve fabric appearance by reducing fuzz or piling or to enhance the softness, lustre and color brightening of cotton fabrics. Therefore, cellulase could replace the use of pumice in the manufacture of stone-washed denim (Lange, 1993).

## XYLANASES FOR BIOPULPING

The annual world production of wood pulp for paper-making is estimated at more than 160 million metric tons per year at present and most of it is produced via the processes of the alkaline kraft pulping and bleaching. Kraft pulping as the first processing of wood is for getting rid of the lignin which causes brownish colored paper, but possible damage on pulp constituents, mainly cellulose, should be reduced to get high quality of paper. In detail, the paper making process is initiated by debarking and chipping wod logs, followed by a strong alkaline cooking where the main part of the lignin is dissolved and then washed away. This process removes most of the lignin, but the residual lignin (4%–5%), covalently bound to carbohydrate moieties, imparts a dark brown color to the kraft pulp (Yamasaki *et al.*, 1981). The resultant kraft pulp should be bleached in order to get a white pulp for high quality paper production. The traditional bleaching consists of a chlorine multistep based on the chemical sequence. During the first stage, elemental chlorine is the main bleaching chemical, but part of it can be replaced with the milder chlorine dioxide. Then alkaline extraction follows in order to remove the dissolved lignin and is succeeded by two chloride dioxide treatments to remove traces of lignin. The lignin-containing fraction dissolved by chlorine, highly colored chlorinated organic compounds, causes serious environmental problems. In a softwood kraft mill using a conventional bleaching sequence, approximately 5 kg of total organically bound chlorine is discharged per ton of bleached pulp and more than 300 different organic compounds, along with a small

quantity of highly toxic dioxins, have been detected in waste bleach water (Eriksson, 1991). In addition, conventionally bleached pulps contain 10–15 μg/g of absorbable halogen halides (AOX) (Jamieson, 1991). There is strong pressure on the industry to diminish the waste of the organic chlorine compounds and to produce environmentally acceptable pulps. The basic idea of enzyme-aided bleaching (Viikari *et al.*, 1986) was developed to hydrolyze hemicelluloses in pulps. Since then, there have been many reports of methods to improve efficiency of hemicellulase pretreatment of biobleaching (Viikari *et al.*, 1994). The major component of hemicellulose, xylan, a polymer of xylose, is highly branched with side chains of acetyl, arabinosyl and glucuronosyl residues (Aspinall, 1988). Hardwoods contain acetylated xylan and softwoods contain arabinoxylan, the acetyl and arabinosyl substituents occurring on approximately 70% and 12% of the xylosyl residues, respectively (Timell, 1962). Xylan was reported to link to lignin, and phenolic compounds may also be involved in cross-linking between xylan molecules and other polysaccharides (Markwalder and Neukom, 1976). Xylan tends to adsorb to cellulose fiber via hydrogen bonds (McNeil *et al.*, 1975). Xylan plays a significant role, in association, among lignin, hemicellulose and cellulose fibers (Kato, 1981). Degradation and removal of xylan is, therefore, an important step in kraft pulping and bleaching to release lignin from cellulose fiber. During the kraft pulping process by heating in alkaline solution, a part of the xylan is dissolved into the pulping liquor. As the washing proceeds and the alkali concentration decreases, however, dissolved xylan reprecipitates onto the surface of cellulose microfibers. In this process, a part of the lignin also reprecipitates onto the cellulose fibers. These redepositions suggest that liginin and xylan are covalently linked to each other (Jansson and Palenius, 1972; Jansson *et al.*, 1975; Iverson and Wannstrom, 1986). In birch kraft pulps, the amount of reprecipitated xylan has been estimated to be up to 12% of the total xylan in the pulp (Axellsson *et al.*, 1962). This reprecipitated lignin should be removed to prepare high quality of paper with high brightness. Therefore, degradation of reprecipitated xylan by xylanase seems to be an effective way to release lignin from cellulose fiber in the process of the bleaching of the kraft pulp. The enzymatic bleaching process is of great interest to the pulp industry as a way of reducing environmental charges.

### *Action of various hemicellulases*

Because of the complex structure of hemicelluloses, several different enzymes are needed for their enzymatic degradation. Since xylan and mannan are two major components of the backbone of hemicellulose, endo-β-1,4-xylanase (EC 3.2.1.8) and endo-β-1,4-mannanase (EC 3.2.1.78) are supposed to be the main enzymes for hemicellulose degradation. In addition, α-arabinosidase, α-glucuronidase and α-galactosidase are important to cleave side groups of the xylan backbone. Acetylesterase for cutting off the esterified side groups also enhances the degradation of hemicellulose.

The effects of various hemicellulases on pulp bleaching have been studied by different methods (Bailey *et al.*, 1992), of which the brightness measurement is the most convenient. The main enzyme required for kraft pulp is known to be endo-β-1,4-xylanase from *T. reesei*. In xylanase-aided bleaching, side group cleaving enzymes, arabinosidase, glucuronidase and acetyl esterase can be expected to be positive effects on

pulp bleaching as described above. Partially purified acetyl esterase of *T. reesei*, arabinosidase from *Aspergillus niger* and glucuronidase of *Agricus bisporus* are employed, in which glucuronidase only increases brightness in chemical pulping (Kantelinen, 1992). Mannanase, which degrades main chain of hemicellulose, gives no positive effect, since mannan is already degraded during alkali treatment in kraft pulping (Ratto *et al.*, 1993). In the peroxide delignification method, however, mannanase is effective (Buchert *et al.*, 1993). Cellulase is harmful in pulp treatments, since a rapid depolymerization of cellulose fibers occurs and lowers pulp quality (Buchert *et al.*, 1994).

*Xylanase in enzyme-aided bleaching.* The mechanisms of xylanase action on pulp bleaching is thought to be as follows. Xylanase acts mainly on the relocated, reprecipitated xylan on the surface of kraft pulp fibers and removes them. The resulting pulp becomes more permeable because the reprecipitated xylan, which is a barrier to lignin extraction in the inner layers, is removed and easily bleached by subsequent chemical treatment. Also, the hydrolysis of inner xylan and side chains by diffused xylanase helps the delignification and bleaching by chemical compounds.

Since the first scientific work on enzyme-aided pulp bleaching was reported by Viikari *et al.* (1986), there have been many reports on the efficiency of xylanase pretreatment and a number of commercially available xylanases (*T. reesei*, *Thermomyces lanuginosus*, *Aureobasidium pullulans* and *S. lividans*) have been tested for their properties in bleaching experiments with various methods. A variety of microorganisms such as fungi and bacteria have been reported to produce xylanases and development of efficient enzyme production or optimized conditions have also been reported. In practical process conditions for enzyme-aided bleaching, cellulase-free xylanase preparations that are both active and stable at high temperature and high pH conditions are favored. Thermostable xylanases and/or alkaline xylanases are good candidates for enzyme-aided bleaching. Some alkaline-stable xylanases (Hogman *et al.*, 1992) and thermostable xylanases (Jager *et al.*, 1992; Senior *et al.*, 1992; Davis *et al.*, 1992) are used for enzyme-aided bleaching.

The practical purpose of the enzymatic pretreatment of kraft pulp is to reduce the chemical consumption and increase the final brightness of pulp. Unbleached kraft pulps have brown-colored lignin which should be removed in successive bleaching stages. Bleaching procedures with and without chlorine are chosen with respect to the final properties of pulps. Chlorine (C), ozone (Z) and peroxy acids react with all of the aromatic structure of lignin. Chlorine dioxide (D), oxygen (O), sodium hypochloride (H) and hydrogen peroxide (P) bind to the certain structure of lignin. Alkaline washing (E) extracts and degrades lignin. The multiple steps of these procedures are chosen for bleaching. In several basic studies, enzymes are used in the pretreatment of kraft pulp before chlorine or peroxide bleaching (*Table 4*). Reduction of total chlorine consumption (average 15%) in chemical bleaching and adsorbable organic halogens in bleached pulp are obtained. Also, a final brightness is higher than the control pulp. Recent work has reported the optimization of detail conditions in enzyme-aided bleaching is important to get the good results (Garg *et al.*, 1996).

In recent years, the production of totally chlorine-free pulps has increased dramatically due to the requirements for environmentally friendly production. Enzyme-aided bleaching without harmful chemical compounds is now a key to reduce environmental pollution in the paper industry.

**Table 4.** Enzyme-aided bleaching in different sequences

| Bleaching | Sequence | Aim |
|---|---|---|
| Conventional | X/(C/D)EDED | Reduction of chlorine consumption and AOX reduction |
| Elementary chlorine free | XDEDED | Reduction of chlorine consumption |
| | XD(EP)DED | AOX reduction |
| Totally chlorine free | OXQPP | Increase brightness |
| | OXQZP | Reduction in chemical consumption |

X, enzyme; C, chlorine; D, chlorine dioxide; P, hydrogen peroxide; O, oxygen delignification; Z, ozone; Q, chelation of metals; E, alkaline washing

## BREEDING OF CELLULOLYTIC MICROORGANISMS

Based on the characterization of cellulases, xylanases and their genes, fundamental enzymatic reaction systems for the degradation or solubilization of cellulosic materials have been revealed and clarified, which may enhance utilization of biomass resources. For further developing enzymatic attack on biomass, improvement of the ability of microorganisms to produce enzymes is one possibility. Especially in small closed ecosystems such as rumen of cattle, the ability to digest grass may directly affect growth of the ruminants. This is obviously when forages mainly containing grass are replaced by condensed ones containing high energy compounds, indicating that the growth of the ruminant may be limited by the intake rate of digested energy compounds. Therefore, improvement of the cellulolytic ability of microorganisms, especially rumen cellulolytic bacteria, is now one of the targets for genetic modification. We now discuss some recent information about new host-vector systems available to rumen bacteria.

### *Development of host-vector system*

Since rumen bacteria are thought to make an important contribution to digestion of plant material in the rumen, as described above, the ability to alter these bacteria genetically might help to improve the efficiency of rumen fermentation. However, previously there has been no way to introduce foreign DNA to rumen bacteria. Scientists in the field of cellulolytic rumen anaerobes have experienced much trouble with deletion of DNA fragments from the anaerobes in the well established *Escherichia coli* host-vector systems, so a new host-vector system for cloning such easily deletable genes is urgently required. For this purpose, a stably replicable plasmid DNA in the rumen anaerobes should be found.

A small cryptic plasmid was isolated from the rumen bacterium *Butyrivibrio fibrisolvens* and sequenced (Kobayashi *et al.*, 1995). When the region between the inverted repeats of the plasmid is deleted and replaced by the erythromycin resistance gene from pAMβ1 together with pUC18, to produce the 7.9 kb chimeric plasmid, the construct was successfully transformed into *E. coli* and *B. fibrisolvens* by electroporation, and stably maintained in both hosts. Using electroporation, two strains of *B. fibrisolvens* have been transformed with plasmid vectors (Ware *et al.*, 1992). It has also been demonstrated that the conjugative transposon Tn*916* can be transferred into *B. fibrisolvens* from *Enterococcus faecalis* (Hespell and Whitehead,

1991). These studies demonstrate that *B. fibrisolvens* can be transformed and provide methods which are available to introduce genes into this organism. A limitation of genetic studies of *B. fibrisolvens*, has been the availability of suitable vectors and transfer systems (Clarke *et al.*, 1994). Using the conjugative tetracycline resistant transposon, Tn*916*, the *Staphylococcus aureus* plasmid, pUB110, and the pUB110-based shuttle vector, a conjugative transfer system has been developed for *B. fibrisolvens*. Results indicated that Tn*916* was necessary for mobilization of pUB110 as transformants were not detected when the transposon is absent from the donor strains. The ability to mobilize pUB110 and its shuttle plasmid between *B. fibrisolvens* strains provides a conjugative transfer system that circumvents problems encountered with electroporation.

Strains of *E. coli* originally isolated from the rumen of sheep have been shown to be capable ot exchanging a 60 kb plasmid, conferring resistance to tetracycline and ampiciflin, at low frequencies under anaerobic conditions (Scott and Flint, 1995). The plasmid transfer between certain *E. coli* strains can occur under conditions that closely simulate an anaerobic gut environment.

Genetic tools have been designed in human colonic species belonging to the family Bacteroidaceae. These are mainly shuttle vectors derived from indigenous *Bacteroides* plasmids (Guiney *et al.*, 1984, 1988; Shoemaker *et al.*, 1985; Smith, 1985; Valentine *et al.*, 1988; Pheulpin *et al.*, 1988). It has been shown recently that one of them is able to mediate transfer and overexpression in *Bacteroides fragilis* and *Bacteroides uniformis* of a xylanase gene previously cloned from *Prevotella ruminicola* (Whitehead and Hespell, 1989, 1990). Thus it will be useful to set up new plasmid vehicles able to replicate in a broad range of ruminal genera or species, either by extending the host range of the previously described 'colonic' vectors, or constructing new ones. Flint *et al.* (1989b) found a conjugal plasmid in *P. ruminicola* which carried a gene for tetracycline resistance and showed that this plasmid could be transferred into another *P. ruminicola* strain by conjugation. Subsequently, Thomson and Flint (1989) demonstrated that plasmid DNA isolated from a *P. ruminicola* transconjugant could be introduced into wild-type strain by electroporation. Transfer of antibiotic resistance plasmids has also been demonstrated between rumen *P. ruminicola* strains under anaerobic conditions *in vitro* (Flint *et al.*, 1988). A new shuttle vector was constructed for use in *Prevotella/Bacteroides* host strains by combining a replicon from *P. ruminicola*, pBluescript sequences and a *tetQ* marker gene for selection in *Prevotella/Bacteroides* hosts (Daniel *et al.*, 1995). Following insertion of a fragment carrying an endoglucanase/xylanase gene from *P. ruminicola* 23 into the multiple cloning site, the resulting construct was introduced into *Bacteroides vulgatus*, *B. uniformis* and *P. ruminicola*. This resulted in an increase of between 4 and 50-fold in CM-cellulase and xylanase activities in cells grown with glucose. In contrast, activities are barely detectable for the same construct in *E. coli*. Most of the total xylanase activity produced is found within the cell in *P. ruminicola* and *B. vulgatus* transformed with the constructed plasmid and in *P. ruminicola* 23. An osmotic shock experiment indicated that a significant proportion of the xylanase activity in *B. vulgatus* transformant is secreted into periplasm. An *E. coli-Bacteroides* shuttle vector (14.7 kb) was constructed by combining the pBR322 derivative with a 4.6 kb cryptic plasmid from *Bacteroides fragilis* and shortened to a size of 10.5 kb derivative (Pheulpin *et al.*, 1988). They are mobilized by R751 into *Bacteroides distasonis* where they replicated stably.

A small cryptic plasmid from *P. ruminicola* is subcloned in *E. coli* and completely

sequenced (Ogata *et al.*, 1996). Two open reading frames, encoding potential polypeptides with limited sequence similarity to replication initiation and mobilization proteins, could be identified within the sequence. The plasmid hybridized at high stringency with plasmids from *Bacteroides/Prevotella* and *Butyrivibrio*, and with pBR322, suggesting that at least regions of the plasmid are widespread. A 2.6 kb plasmid from a rumen bacterium of the genus *Butyrivibrio* shows the high AT content typical of plasmids from gram-positive organisms (Hefford *et al.*, 1993). Computer analysis of sequence data reveals two major open reading frames encoded on the same strand but in different frames.

Although these results are important because they showed that DNA could be introduced into the microorganism by conjugation and electroporation, the plasmid could not be used as a shuttle vector because it was relatively large (19.5 kb) and did not replicate in *E. coli*. So, a shuttle vector from the colonic species *B. uniformis* was transferred to *P. ruminicola* by the mobilization method, via *E. coli* (Shoemaker *et al.*, 1991). The transfer frequency is $10^{-6}$ to $10^{-7}$ per recipient. The tetracycline resistant colonies obtained after mating are true transconjugants because they contained the plasmid (pRDB5). This plasmid replicates in many colonic *Bacteroides* species as well as in *P. ruminicola* $B_1$4, indicating that its host range is very wide. Although the plasmid is relatively large, there are several usable cloning sites inside on antibiotic resistance genes. Thus pRDB5 could be used without further modification as a vector for introducing cloned DNA into *P. ruminicola*. New shuttle vectors based on a *P. ruminicola* cryptic plasmid (pRR17, 9.5 kb) inserted within the *E. coli* vector carrying the clindamycin resistance and erythromycin resistance *Bacteroides* marker, are constructed (Bechet *et al.*, 1993). These constructs were transferred into IP. ruminicora, *B. distasonis* and other *Bacteroides* sp. by conjugal mobilization using helper plasmid and by electroporation.

A carboxymethyl cellulase gene from *P. ruminicola* B14 was reconstructed by adding a cellulose binding domain from a *T. fusca* cellulase and was conjugally transferred from *E. coli* to *B. uniformis* by using a chloramphenicol and tetracycline resistance shuttle vector (Gardner *et al.*, 1996). The vector was specifically constructed to facilitate conjugal transfer of vectors from *B. uniformis* donors to *P. ruminicola* recipients. The CMCase is produced, when a xylanase promoter from IP. ruminicola 23 was placed upstream of the CMCase gene, suggesting that not all of the promoters recognized by *B. uniformis* and *P. ruminicola* 23 are functional in *P. ruminicola* $B_1$4. Although the reconstructed CMCase alone does not allow *B. uniformis* to grow on acid-swollen cellulose, rapid growth was observed when two exocellulases were added to the culture supernatant. Under these conditions, the reconstructed CMCase permitted faster growth than the wild-type CMCase. A *Selenomonas ruminantium* (an gram-negative ruminal anaerobe) plasmid has been cloned in *E. coli* and completely sequenced (Attwood and Brooker, 1992). This plasmid was shown to replicate independently in *E. coli* by a DNA polymerase I-dependent mechanism. This plasmid may form the basis of a *Selenomonas/E. coli* shuttle vector. Another plasmid was characterized and sequenced from *S. ruminantium* (Zhang and Brooker, 1993). Gene expression and DNA structural features of the plasmid have been shown to replicate by a rolling circle-type mechanism which is found from many gram-positive organisms.

A plasmid DNA was isolated from *Ruminococcus flavefaciens* with 5.2 kb size by

utilizing mutanolysin and proteinase K for the rapid lysis of the cells. The plasmid has potential in the development of genetic vectors for rumen bacteria (Asmundson and Kelly, 1987).

Another important cellulolytic microorganism, Rumicococciae, remains to be precisely studied for establishing new host-vector systems to accelerate degradation of cellulosic biomass.

*Establishment of cellulolytic recombinants in the rumen ecosystem*

The cellulolytic action of bacteria in the rumen is the major way to digest forage that has been taken in by ruminants, since mammals cannot synthesize cellulose-degrading enzymes by themselves. As solubilization of these materials in the rumen is slow and imperfect, there is substantial interest in improving the rate and extent of cellulose and hemicellulose digestion in the rumen. One possible and promising attempt is to enhance the rate and extent of forage degradation by introducing some genes into rumen bacteria that encode enzymes effective in forage degradation to increase enzyme production depending on the fundamental data as mentioned above. The next essential step is that these improved recombinants must be able to survive in the rumen micro-ecosystem. The deliberate release of anaerobic microorganisms into the rumen is used as, and is considered to be, a means to improve digestion rate of cellulosic materials in rumen. One of the trials was carried out by Flint *et al.* (1989). Selective plating procedures were used to follow the fate of rifampicin-resistant mutant strains of the obligately naerobic species *Bacteroides multiacidus* and *S. ruminantium* after their introduction at numbers around $10^7$/ml into the rumen of sheep. *B. multiacidus* strain F100 shows an initially rapid rate of loss (49%/h) but subsequently numbers decline more gradually, approaching the limits of detection ($<10^3$/ml) after 100 h. Viable cell numbers also decrease *in vitro* upon addition of the cells to whole rumen contents, but remain stable upon addition to cell-free rumen fluid, suggesting protozoal predation. In the case of *S. ruminantium*, introduced rifampicin-resistant strains persist in the rumen at levels around $10^6$ ml for at least 30 days. Both rumen bacteria remain in the rumen at a cell number around $10^3$/ml or higher with the organisms mutated on, under the cultivation of gradual increase in those antibiotics. In another experiment (Miyagi *et al.*, 1995), a transconjugant of *R. albus* A3 harboring pAMβ1 was released into a goat rumen. The A3 strain was obtained by filter mating with *Bacillus thringiensis* that harbored pAMβ1 and the *R. albus* A2 strain mutated spontaneously to give rifampicin and streptomycin resistance. To investigate the possibility of distinguishing *R. albus* A3 cells from rumen microflora, indigenous rumen microorganisms were inoculated into medium without antibiotics in Hungate tubes. After this culture had been incubated at 37°C overnight, it was diluted appropriately with anaerobic dilution buffer and inoculated into M10 agar medium that contained rifampicin and erythromycin in roll tube. For specific selection of *R. albus* A3 cells with resistance to rifampicin and erythromycin among rumen microflora, the possible use of these antibiotics in combination to inhibit the growth of indigenous rumen microorganisms was examined. In the absence of both antibiotics, the viable cell number in intact rumen liquid was $10^9$ colony-forming units per millilitre, and with the combination of rifampin and erythromycin, only one type of colony, a flat, rough and lacerate colony, was formed at $10^3$ colony formation units per millilitre by

a filamentous anaerobic fungus. This fungus colony was easily distinguished from the *R. albus* A3 colony, a capitate, smooth and entire colony. Therefore, exact numbers of *R. albus* A3 cells were determined on the basis of colony morphology under selection pressure by antibiotics. The diplococcal bacterium which grows in the presence of these antibiotics was confirmed to be *R. albus* A3 by dot blot hybridization, which detects pAMβ1. The A3 cells remained roughly constant for 14 days in this goat rumen.

These basic data show that the transformed microorganisms can survive in rumen, although the cell numbers are not large, so far. The improvement of the growth rate of such transformants depends on the integration of a cellulase gene which can hydrolyze tough cellulose, having an affinity for crystalline cellulose much higher than for amorphous cellulose.

## BIOTECHNOLOGY OF CELLULASES AND XYLANASES IN PLANTS

In general, biomass is the mass of living cells, which propagate by themselves and die, and possess both biosynthetic and biodegradation systems. If the former system is much more active than the latter one, the increase of biomass is remarkable. This phase corresponds to the growth phase. When the biodegradation system becomes more active than the biosynthetic system, however, the amount of biomass synthesis decreases, corresponding to the death phase in the growth curve of cells. Depending on the variety of time scales for the growth curve, the diversity of plant growth rate seems to cause tissue 'toughness'. When a plant in the rapid growth phase expresses some genes encoding the biomass degrading enzymes, some changes may occur in the properties of the plant tissue. It is believed that this kind of integration of exogenous fiber-degrading enzymes into tobacco cells, for example, may enhance relaxation of plant tissue and enhance biodegradation.

### *Expression of xylanase genes in tobacco cells*

Xylanase Z gene (*xynZ*) from *C. thermocellum* is integrated into tobacco plant (*Nicotiana tabacum* Cv Samsun NN) by an integration system with *Agrobacterium tumefaciens* (Helbers *et al.*, 1995). The xylanase Z specifically hydrolyzes xylan, but not cellulose and carboxymethyl cellulose, which might have importance for potential applications in cellulase-free systems like the specific processing of plant fibers (Grépinet *et al.*, 1988a). This enzyme is a heat-stable protein having maximum activity at around 70°C, so that only 1/5 of maximum activity will be revealed at around 25°C, the optimum temperature for the growth of the tobacco plant. This suggests that the integrated gene product XynZ will not affect plant growth even if the expression of the enzyme is remarkably high. Depending on these available properties, the expression of XynZ was examined in the tobacco plant. For the effective expression of the gene in tobacco, 1,000 bp of C-terminal side of the full size gene (2,500 bp) (Grépinet *et al.*, 1988b) of the *xynZ* gene which encodes catalytic domain was amplified by a polymerase chain reaction (PCR) and inserted into the binary vector pBinAR between the constitutive cauliflower mosaic virus (CaMV) 35S promoter and the terminator of the octopine synthase gene from *A. tumefaciens*. The signal peptide of the proteinase inhibitor II protein was inserted in front of the coding

region to target the xylanase into the apoplastic space. The rationale behind this strategy was to stabilize the heterologous enzyme by removing it from protease-rich intracellular environments. The *xynZ* gene expression construct was integrated into the tobacco plant using *Agrobacterium*-mediated transfer (Deblaere *et al.*, 1985). For the isolation of XynZ from transgenic tobacco, fresh leaves were vacuum-infiltrated. The isolate hydrolyzed xylan well. Quantification of the XynZ protein in plant extracts was performed by comparing the densitometrically determined amount of the XynZ enzyme in Coomassie-stained gels to the total amount of protein. The relative amount of the enzyme was calculated to be around 4%. The sugar components of hemicellulose from transgenic tobacco stalks were very similar to those of wild tobacco, indicating that the expression of bacterial xylanase does cause defects in the cell wall constituents of tobacco plant, even though the exogenous xylanase expresses as much as 4%. This insignificant difference between the transgenic and wild tobacco might be caused by a lower content of xylan in the tobacco, as a dicot, than monocots such as rice plants (Rosahl *et al.*, 1987).

Xylanase B (Xyn B) from *C. stercorarium* is also expressed in tobacco suspension cells (*Nicotiana tabacum L.cv* BY2 cell) under the control of CaMV 35S promoter and noparin synthetase terminator (Sun *et al.*, 1996). The plasmid constructed using pUC118 was introduced into BY2 protoplasts by electroporation and transformed cells incubated in a medium containing kanamycin in agarose beads type culture. After a 4–6 week cultivation, the calli depicting clear halos on the xylan containing agar plates were selected. Depending on the quantitative measurement of xylanase activity of transformed suspension cells, a transformant having the highest activity in the tested calli was selected for further study. Xylanase activity roughly equal to that in the intracellular fractions was detected in the culture supernatant, indicating that the integrated gene product leaks out or was secreted out of the cells due to the presence of a specified signal peptide sequence. The amounts of expressed XynB protein was also around 4%–5% of total proteins in the soluble extracts of tobacco suspension cells. The soluble fraction prepared from the transformant by sonication was used as crude xylanase to examine solubilization of barley straw powder for 15 h at 60°C. When the insoluble fraction of tobacco suspension cells containing xylan much lower than barley straw was used as a substrate, it was negligibly degraded by the crude xylanase due to the low content of xylan in dicot tobacco.

In these experiments, the bacterial xylanases are stably expressed in tobacco plant or its suspension cells as high as around 4% of total proteins without any inhibitory effects on plant growth. In addition the thermostable enzymes are easily isolated from other tobacco proteins by heating, which may allow us to use tobacco for the production of enzymes of interest.

### *Expression of cellulase in tobacco cells*

A cellulase (endoglucanase I: EgI) gene (*egI*) from a rumen bacterium *R. albus*, was also expressed in tobacco cells BY2 using the gene construction and integration methods similar to those for XynB. The *egI* gene is truncated, the 5' moiety encoding signal peptide and further 15 amino acids for the effective expression of activity. Schematic methods of *egI* gene truncation and construction of plasmid harboring *egI*

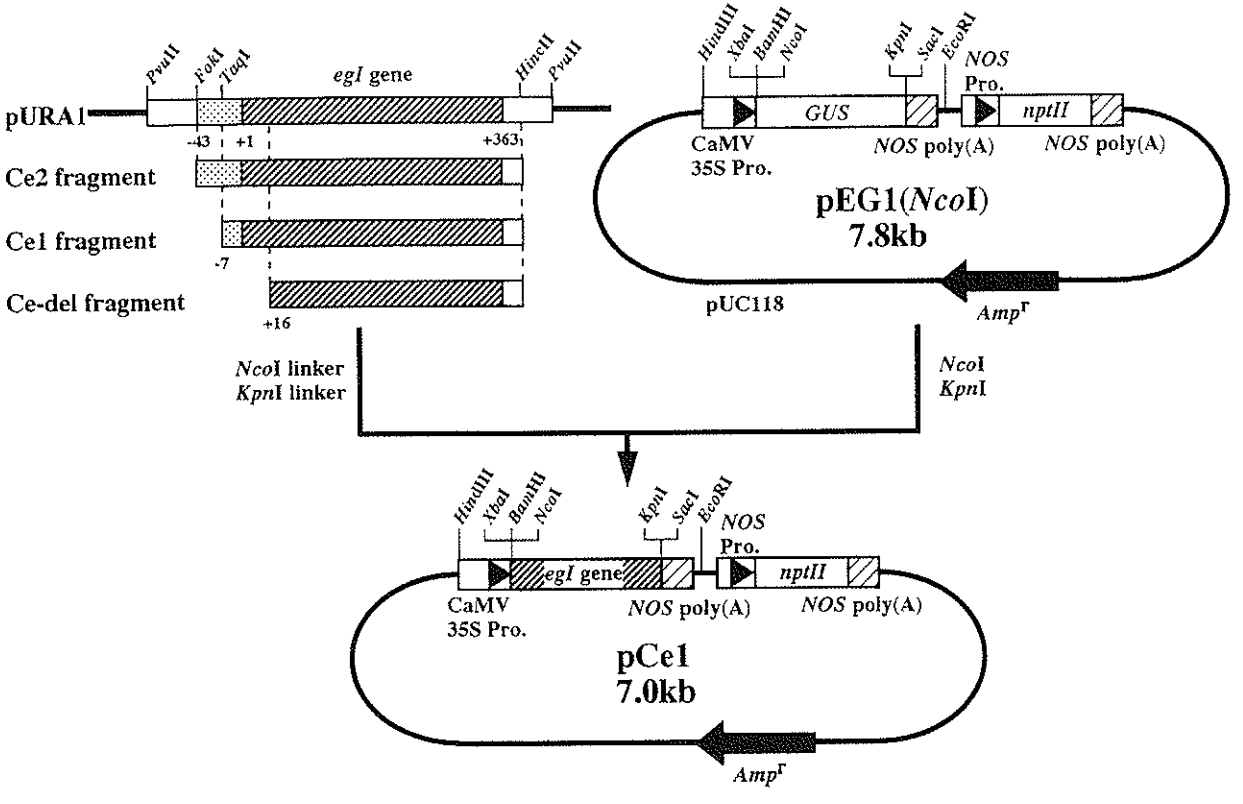


**Figure 4.** Structures of the introduced *egI* and its derivatives. The shaded box represents the region encoding the mature enzyme and the dotted box the region encoding the EgI signal sequence. Numbers refer to amino acid residues.

for tobacco cell are shown in *Figure 4* (Kawazu *et al.*, 1996). Tobacco cells BY2 are transformed with the constructed plasmids by electroporation. The modified EgI truncated 15 amino acids at N-terminal (Ce-del) revealed the highest activity in our test (Ohmiya *et al.*, 1991). In the supernatant of transformed tobacco cells negligible activity was detected, indicating that almost all the activity was retained in the cells. The molecular size of Ce-del protein prepared from the transformed cell was the same to that from *E. coli* transformant, showing that no proteolytic action occurred on the exogenous cellulase Ce-del in the host tobacco cells. The Ce-del protein recovered from sonicated cell suspension was evaluated to be around 0.1% of total soluble protein of the tobacco when estimated by Western blot analysis with anti-EgI antibody. This cellulase activity against CMC in transformed cells was about 30 times higher than wild type of cells at the late log phase of growth curve, as shown in *Figure 5*. The cell wall degrading activity in the cell extract was significantly present after reaction at 35°C overnight. In other words, Ce-del expressed in tobacco cells hydrolyzed host cell wall. Therefore, Ce-del expression in tobacco cells will somewhat inhibit synthesis of cell wall formation, which caused delay of cell propagation as shown in the growth curve (*Figure 5*). This degradative effect of Ce-del may cause some relaxation of tobacco cell wall. To confirm this, protoplasts were prepared from the tobacco cells transformed by Ce-del. The protoplast formation numbers from the transformed cells are higher than those of wild cells. This result clearly shows that the relaxation of tobacco cell wall may be caused by the expression of a cellulase Ce-del.

A transgenic tobacco plant was bred by integrating the truncated *egI* gene, using the *Agrobacterium*-mediated transformation method. The transgenic tobacco plants are morphologically indistinguishable from wild type grown in the greenhouse under the same conditions. Remarkable CMC degrading activity is depicted from all the cross-sections of transgenic tobacco leaves we tested, indicating that endoglucanase from *egI* gene leaks out from the destroyed tissue and cells. When this kind of transgenic plant is fed to cattle, the cellulase produced and stored in plants cells might be released by mastication and disruption and may attack cellulosic compounds to solubilize grass. It may enhance rates of ensilage and digestion of grass in rumen.

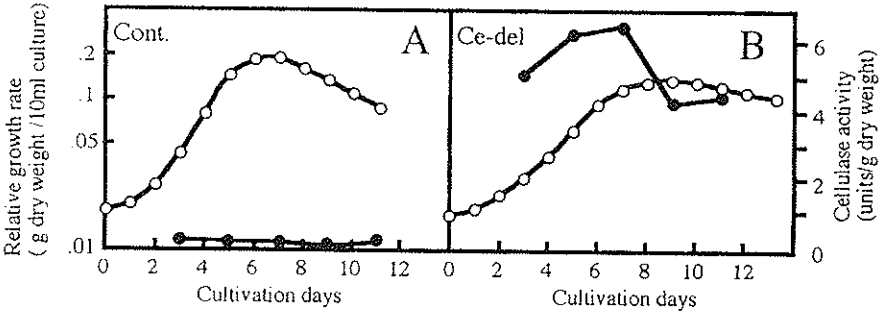


**Figure 5.** Time courses of cellulase activity (closed circles) and cell density (open circles). A: non-transformed cells. B: Ce-del.

*Expression of a thermostable (1,3-1,4)-β-glucanase in barley*

A thermostable (1,3-1,4)-β-glucanase obtained by intragenic recombination *in vitro* between the genes from *Bacillus amyloliquefaciens* and *Bacillus macerans*, was tested for expression with the barley (1,3-1,4)-β-glucanase isozyme EII gene promoter in aleurone protoplasts of barley (Jensen *et al.*, 1996). The glucanases from both *Bacillus* sp. are synthesized and secreted with the same specificity as the barley enzyme but are more thermotolerant than their barley counterparts. In addition, the codon usage for the barley glucanase exhibits strong preference for G or C in the third position, resulting in a G–C content of 66% in the coding region, while such codon bias is not observed in both *Bacillus* glucanase genes. Since modification of a bacterial gene toward plant gene codon usage increases its expression in plants, the codons for the hybrid gene were modified to match those of the gene encoding the barley glucanase. The modified *Bacillus* enzyme gene-containing plasmid is transfected to aleurone protoplasts by PEG-mediated DNA uptake. The G+C rich construct gives an average production of 40 ng enzyme/2 × $10^5$ protoplasts after cultivation for 110 h, while no (1,3-1,4)-β-glucanase activity is detectable in protoplasts transfected with the bacterial glucanase without any G+C adaptation. In the case of plant transformation, the plasmids containing the modified barley endoglucanase genes are introduced into immature embryos by particle bombardment. Two of the transformation experiments gave a total of 14 green plants. These plants were morphologically indistinguishable from control plants grown under identical conditions in the greenhouse. The transgenic plants were self crossed and the endosperms obtained from the segregants in which both chromosomes had endoglucanase genes. These endosperms were analyzed for expression of glucanase gene. In this plant breeding project, Jensen *et al.* endeavoured to produce barley plants that during steeping and germination express a (1,3-1,4)-β-glucanase that survives the high temperatures used for kiln-drying of green malt. This would allow the enzyme to act in the mash tun as do the thermostable α-amylases from barley, wheat and rice. Such a heat-stable (1,3-1,4)-β-glucanase synthesized during germination might eliminate the requirement of complete endosperm wall depolymerization in the malting schedule and thereby provide new opportunities for the application of malting and mashing in production of conventional and novel biotechnological commodities. The modified enzyme with high G+C content produced in *E. coli* has been successfully tested in pilot mashing and feed pellet production.

As described above, there are a few papers concerning successful expression, in plants, of bacterial genes encoding the plant fiber-degrading enzymes. The production of xylanases in transgenic plants is much higher than cellulases, suggesting that the enzymes degrading the main structural compounds of plants might be inhibited in their expression in plant cells, or the transgenic cells might be lethal when exogenous cellulolytic enzyme is highly expressed, due to the incomplete formation of cell wall. From the results of xylanase expression at as much as 4% (w/w) in plant cells, a new enzyme production system and a new immobilization system of enzymes in plant cells becomes of use. In addition, efficiency increase in protoplast formation indicates relaxation of the tough and rigid texture of plants. This may enhance digestion of these transgenic plants as feed in rumen or solubilization of cellulosic materials for the use of carbon and energy sources in fermentation.

## Concluding remarks

Tremendously rapid improvement of living conditions and shifting food intake from plant to animal sources in developing countries, as their economic situation improves are causing logarithmic increases in energy consumption in convenient forms for utilization, such as fossil energy and cereal grains. Some grains which are used as human foods are consumed now as feed for animal husbandry. This has triggered an amazing increase in the cost of maize in world trade markets and indicates an absolute bulk shortage of foods and feeds. Even under these circumstances, rapid increase in their production or an abundant harvest cannot be expected in the near future. To overcome this situation, we have been studying cellulases and related enzymes for exploiting effective utilization and solubilization of fibrous plant materials. This progress in understanding the fundamental aspects of degradation of cellulosic materials may promote unutilized cellulosic materials to foods, feeds and fuels. The major interests of scientists in the field are now focusing toward the following applications, based on the accumulated fundamental knowledge described above: (1) Construction of lignocellulose-degrading enzymes with higher activity by domain shuffling of catalytic and substrate binding domains from many kinds of enzymes; (2) Breeding microorganisms having strong degradation activity against lignocellulosic materials for preparing fermentable substances by using genetic engineering; (3) Breeding grass with integrated foreign cellulase and xylanase genes for accelerating its digestion in silage and/or ruminants.

## References

AHSAN, M.M., KIMURA, T., KARITA, S., SAKKA, K. AND OHMIYA, K. (1996). Cloning, DNA sequencing, and expression of the gene encoding *Clostridium thermocellum* cellulase CelJ, the largest catalytic component of the cellulosome. *Journal of Bacteriology* **178**, 5732–5740.

ALZARI, P.M., SOUCHON, H. AND DOMINGUEZ, R. (1996). The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from *Clostridium thermocellum*. *Structure* **4**, 265–275.

ARASE, A., YOMO, T., URABE, I., KATSUBE, Y. AND OKADA, H. (1993). Stabilization of xylanase by random mutagenesis. *FEBS Letters* **316**, 123–127.

ASMUNDSON, R.V. AND KELLY, W.J. (1987). Isolation and characterization of plasmid DNA from *Ruminococcus*. *Current Microbiology* **16**, 97–100.

ASPINALL, G.O. (1988). Chemistry of cell wall polysaccharides. In: *The Biochemistry of Plants (A Comprehensive Treatise), Vol. 3. Carbohydrates: Structure and Function* (J. Preiss, ed.), pp. 473–500. Academic Press, New York.

ASSOULINE, Z., SHEN, H., KILBURN, D.G. AND WARREN, R.A. (1993). Production and properties of a factor X-cellulose-binding domain fusion protein. *Protein Engineering* **6**, 787–792.

ATTWOOD, G.T. AND BROOKER, J.D. (1992). Complete nucleotide sequence of a *Selenomonas ruminantium* plasmid and definition of a region necessary for its replication in *Escherichia coli*. *Plasmid* 28, 123–129.

AXELLSSON, S., CROON, I. AND ENSTROM, B. (1962). Dissolution of hemicelluloses during surphate pulping. *Svensk Papperstid.* **65**, 693–697.

BAILEY, M.J., BIELY, P. AND POUTANEN, K. (1992). Interlaboratory testing of methods for assay of xylanase activity. *Journal of Biotechnology* **23**, 257–270.

BAYER, E.A., MORAG, E. AND LAMED, R. (1994). The cellulosome-a-treasure-trove for biotechnology. *Trends in Biotechnology* **12**, 379–386.

BEARD, C.E., HEFFORD, M.A., FOSTER, R.J., SONTAKKE, S., TEATHER, R.M. AND GREGG, K. (1995). A stable and efficient transformation system for *Butyrivibrio fibrisovens* OB156. *Current Microbiology* **30**, 105–109.

BÉCHET, M., PHEULPIN, P., FLINT, H.J., MARTIN, J. AND DUBOURGUIER, H.-C. (1993). Transfer of hybrid plasmids based on the replicon pRR17 from *Escherichia coli* to *Bacteroides* and *Prevotella* strains. *Journal of Applied Biotechnology* **74**, 542–548.

BÉGUIN, P. AND AUBERT, J.-P. (1992). La dégradation de la cellulose par les microorganisms. *Ann. Inst. Pasteur/Actualités* **3**, 91–115.

BÉGUIN, P. AND AUBERT, J.-P. (1994). The biological degradation of cellulose. *FEMS Microbiology Reviews* **13**, 25–58.

BÉGUIN, P. AND LEMAIRE, M. (1996). The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation. *Critical Reviews of Biochemistry and Molecular Biology* **31**, 201–236.

BÉGUIN, P., CORNET, P. AND AUBERT, J.-P. (1985). Sequence of a cellulase gene of the thermophilic bacterium *Clostridium thermocellum*. *Journal of Bacteriology* **162**, 102–105.

BLACK, G.W., HAZLEWOOD, G.P., MILLWARD-SADLER, S.J., LAURIE, J.I. AND GILBERT, H.J. (1995). A modular xylanase containing a novel non-catalytic xylan specific binding domain. *Biochemical Journal* **307**, 191–195.

BRUN, E., GANS, P., MARION, D. AND BARRAS, F. (1995). Overproduction, purification and characterization of the cellulose-binding domain of the *Erwinia chrysanthemi* secreted endoglucanase EGZ. *European Journal of Biochemistry* **231**, 142–148.

BUCHERT, J., RANUA, M., SIIKA-AHO, M., PERE, J. AND VIIKARI, L. (1994). *Trichoderma reesei* cellulase in the bleaching of kraft pulp. *Applied Microbiological Biotechnology* **40**, 941–945.

BUCHERT, J., SALMINEN, J., SIIKA-AHO, M., RANUA, M. AND VIIKARI, L. (1993). *Trichoderma reesei* xylanase and mannannase in enzyme-aided bleaching. *Holzforshung* **47**, 473–478.

CAMPBELL, R., ROSE, D., WAKARCHUK, W., TO, R., SUNG, W. AND YAGUCHI, M. (1993). A comparison of the structures of the 20 Kd xylanases from *Trichoderma hazianum* and *Bacillus circulans*. In: *Trichoderma reesei cellulases and other hydrolases* (P. Suominen and T. Reinikainen, eds), pp. 63–72. Foundation for biotechnological and industrial fermentation, Helsinki.

CHAUVAUX, S., BÉGUIN, P., AUBERT, J.-P., BHAT, K.M., GOW, L.A., WOOD, T.M. AND BAIROCH, A. (1990). Calcium-binding affinity and calcium-enhanced activity of *Clostridium thermocellum* endoglucanase D. *Biochemical Journal* **265**, 261–265.

CHAUVAUX, S., BÉGUIN, P. AND AUBERT, J.-P. (1992). Site-directed mutagenesis of essential carboxylic residues in *Clostridium thermocellum* endoglucanase CelD. *Journal of Biological Chemistry* **267**, 4472–4478.

CHOI, S.K. AND LJUNGDAHL, L.G. (1996a). Dissociation of the cellulosome of *Clostridium thermocellum* in the presence of ethylenediaminetetraacetic acid occurs with the formation of truncated polypeptides. *Biochemistry* **35**, 4897–4905.

CHOI, S.K. AND LJUNGDAHL, L.G. (1996b). Structural role of calcium for the organization of the cellulosome of *Clostridium thermocellum*. *Biochemistry* **35**, 4906–4910.

CLARKE, J.H., DAVIDSON, K., GILBERT, H.J., FONTES, C.M.G.A. AND HAZLEWOOD, G.P. (1996). A modular xylanase from mesophilic *Cellulomonas fimi* contains the same cellulose-binding and thermostabilizing domains as xylanases from thermophilic bacteria. *FEMS Microbiology Letters* **139**, 27–35.

CLARKE, R.G., CHENG, K.J., SELINGER, L.B. AND HYNES, M.F. (1994). A conjugative transfer system for the rumen bacterium *Butyrivibrio fibrosolvens*, based on Tn-916-mediated transfer of the *Staphylococcus aureus* plasmid pUB110. *Plasmid* **32**, 295–305.

COUTINHO, J.B., GILKES, N.R., WARREN, R.A.J., KILBURN, D.G. AND MILLER, JR, R.C. (1992). The binding of *Cellulomonas fimi* endoglucanase C (CenC) to cellulose and Sephadex is mediated by the N-terminal repeats. *Molecular Microbiology* **6**, 1243–1252.

DANIEL, A.S., MARTIN, J., VANAT, I., WHITEHEAD, T.R. AND FLINT, H.J. (1995). Expression of a cloned cellulase/xylanase gene from *Prevotella ruminicola* in *Bacteroides vulgatus*, *Bacteroides uniformis* and *Prevotella ruminicola*. *Journal of Applied Bacteriology* **79**, 417–424.

DAVIES, G.J., DODSON, G.G., HUBBARD, R.E., TOLLEY, S.P., DAUTER, Z., WILSON, K.S., HJORT, C., MIKKELSEN, J.M., RASMUSSEN, G. AND SCHÜLEIN, M. (1993). Structure and function of endoglucanase V. *Nature* **365**, 362–364.

DAVIES, G.J., TOLLEY, S.P., HENRISSAT, B., HJORT, C. AND SCHÜLEIN, M. (1995). Structures of oligosaccharide-bound forms of the endoglucanase V from *Humicola insolens* at 1.9 Å resolution. *Biochemistry* **34**, 16210–16220.

DAVIS, S., GYSIN, B., CASIMIR, J. AND ZIMMERMANN, W. (1992). Thermostable xylanases from actinomycetes for pulp bleaching. In: *Xylans and Xylanases Progress in Biotechnology, Vol. 7* (J. Visser, G. Beldman, M.A. Kustersvan Someren and A.G.J. Voragen, eds), pp. 551–554. Elsevier, Amsterdam.

DEBLAERE, R., BYTEBIER, B., DE GREVE, H., DEBROECK, F.M., SCHELL, J., VAN MONTAGU, M. AND LEEMANNS, J. (1985). Efficient octopine Ti-plasmid derived vectors for *Agrobacterium*-mediated gene transfer. *Nucleic Acids Research* **13**, 4777–4788.

DEREWENDA, U., SWENSON, L., GREEN, R., WEI, Y., MOROSOLI, R., SHARECK, F., KLUEPFEL, D. AND DEREWENDA, Z.S. (1994). Crystal structure, at 2.6-Å resolution, of the *Streptomyces lividans* xylanase A, a member of the family of β-1,4-D-glycanases. *Journal of Biological Chemistry* **269**, 20811–20814.

DIN, N., GILKES, N.R., TEKANT, B., MILLER, JR, R.C., WARREN, R.A.J. AND KILBURN, D.G. (1991). Non-hydrolytic disruption of cellulose fibers by the binding domain of a bacterial cellulase. *Bio/Technology* **9**, 1096–1099.

DIN, N., FORSYTHE, I.J., BURTNICK, L.D., GILKES, N.R., MILLER, JR, R.C., WARREN, R.A.J. AND KILBURN, D.G. (1994a). The cellulose-binding domain of endoglucanase A (CenA) from *Cellulomonas fimi*: evidence for the involvement of tryptophan residues in binding. *Molecular Microbiology* **11**, 747–755.

DIN, N., DAMUDE, H.G., GILKES, N.R., MILLER, JR, R.C., WARREN, R.A.J. AND KILBURN, D.G. (1994b). C1-Cx revisited: intramolecular synergism in a cellulase. *Proceedings of the National Academy of Sciences, USA* **91**, 11383–11387.

DIVINE, C., STAHLBERG, J., REINIKAINEN, T., ROUHONEN, L., PETTERSSON, G., KNOWLES, J.K.C., TEERI, T.T. AND JONES, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**, 524–528.

DOI, R.H., GOLDSTEIN, M., HASHIDA, S., PARK, J.-S. AND TAKAGI, M. (1994). The *Clostridium cellulovorans* cellulosome. *Critical Reviews in Microbiology* **20**, 87–93.

DOMÍNGUEZ, R., SOUCHON, H., SPINELLI, S., DAUTER, Z., WILSON, K.S., CHAUVAUX, S., BÉGUIN, P. AND ALZARI, P.M. (1995). A common protein fold and similar active site in two distinct families of β-glycanases. *Nature, Structural Biology* **2**, 569–576.

DUCROS, V., CZJZEK, M., BELAICH, A., GAUDIN, C., FIEROBE, H.-P., BELAICH, J.-P., DAVIES, G.J. AND HASER, R. (1995). Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* **3**, 939–949.

DURRANT, A., HALL, J., HAZLEWOOD, G.P. AND GILBERT, H.J. (1991). The non-catalytic carboxyl-terminal region of endoglucanase E from *Clostridium thermocellum* contains a cellulose-binding domain. *Biochemical Journal* **273**, 289–293.

ERIKSSON, K.E.L. (1991). Biotechnology: Three approaches to reduce the environmental impact of the pulp and paper industry. *Science Progress, Oxford* **75**, 175–189.

FELIX, C.R. AND LJUNGDAHL, L.G. (1993). The cellulosome: the exocellular organelle of *Clostridium*. *Annual Reviews of Microbiology* **47**, 791–819.

FERREIRA, L.M.A., WOOD, T.M., WILLIAMSON, G., FAULDS, C., HAZLEWOOD, G.P. AND GILBERT, H.J. (1993). A modular esterase from *Pseudomonas fluorescens* subsp. *cellulosa* contains a non-catalytic cellulose binding domain. *Biochemical Journal* **294**, 349–355.

FIEROBE, H.-P., GAUDIN, C., BELAICH, A., LOUTFI, M., FAURE, E., BAGNARA, C., BATY, D. AND BELAICH, J.-P. (1991). Characterization of endoglucanase A from *Clostridium cellulolyticum*. *Journal of Bacteriology* **173**, 7956–7962.

FLINT, H.J., THOMSON, A.M. AND BISSET, J. (1988). Plasmid-associated transfer of tetracycline resistance in *Bacteroides ruminicola*. *Applied and Environmental Microbiology* **54**, 855–859.

FLINT, H.J., BISSET, J. AND WEBB, J. (1989). Use of antibiotic resistance mutations to track strains of obligately anaerobic bacteria introduced into the rumen of sheep. *Journal of Applied Bacteriology* **67**, 177–183.

FUJINO, T., BÉGUIN, P. AND AUBERT, J.-P. (1993). Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. *Journal of Bacteriology* **175**, 1891–1899.

GARDNER, R.G., RUSSEL, J.B., WILSON, D.B., WANG, G.-R. AND SHOEMAKER, N.B. (1996). Use of a modified *Bacteroides-Prevotella* shuttle vector to transfer a reconstructed β-1,4-endoglucanase gene into *Bacteroides uniformis* and *Prevotella ruminicola* $B_1$4. *Applied Environmental Microbiology* **62**, 196–202.

GARG, A.P., MCCARTHY, A.J. AND ROBERTS, J.C. (1996). Biobleaching effect of *Streptomyces thermoviolaceus* xylanase preparations on birchwood kraft pulp. *Enzyme Microbial Technology* **18**, 261–267.

GEBLER, J., GILKES, N.R., CLAEYSSENS, M., WILSON, D.B., BÉGUIN, P., WAKARCHUK, W.W., KILBURN, D.G., MILLER, JR. R.C., WARREN, R.A.J. AND WITHERS, S.G. (1992). Stereoselective hydrolysis catalyzed by related β-1,4-glucanases and β-1,4-xylanases. *Journal of Biological Chemistry* **267**, 12559–12561.

GERNGROSS, U.T., ROMANIEC, M.P.M., KOBAYASHI, T., HUSKISSON, N.S. AND DEMAIN, A.L. (1993). Sequencing of a *Clostridium thermocellum* gene (*cipA*) encoding the cellulosomal SL protein reveals an unusual degree of internal homology. *Molecular Microbiology* **8**, 325–334.

GHANGAS, G.S. AND WILSON, D.B. (1988). Cloning of the *Thermomonospora fusca* endoglucanase E2 gene in *Streptomyces lividans*: affinity purification and functional domain of the cloned gene product. *Applied Environmental Microbiology* **54**, 2521–2526.

GHANGAS, G.S., HU, Y.-J. AND WILSON, D.B. (1989). Cloning of a *Thermomonospora fusca* xylanase gene and its expression in *Escherichia coli* and *Streptomyces lividans*. *Journal of Bacteriology* **171**, 2963–2969.

GILKES, N.R., WARREN, R.A.J., MILLER, JR. R.C. AND KILBURN, D.G. (1988). Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. *Journal of Biological Chemistry* **263**, 10401–10407.

GOLDSTEIN, M.A., TAKAGI, M., HASHIDA, S., SHOSEYOR, O., DOI, R.H. AND SEGEL, I.H. (1993). Characterization of the cellulose-binding domain of the *Clostridium cellulovorans* cellulose-binding protein. *Journal of Bacteriology* **175**, 5762–5768.

GOLDSTEIN, M.A. AND DOI, R.H. (1994). Mutation analysis of the cellulose-binding domain of the *Clostridium cellulovorans* cellulose-binding protein. *Journal of Bacteriology* **176**, 7328–7334.

GRAHAM, R.W., GREENWOOD, J.M., WARREN, R.A.J., KILBURN, D.G. AND TRIMBUR, D.E. (1995). The pTugA and pTugAS vectors for high-level expression of cloned genes in *Escherichia coli*. *Gene* **158**, 51–54.

GREENWOOD, J.M., ONG, E., GILKES, N.R., WARREN, R.A.J., MILLER, JR. R.C. AND KILBURN, D.G. (1992). Cellulose-binding domains: potential for purification of complex protein. *Protein Engineering* **5**, 361–365.

GRÉPINET, O., CHEBROU, M.-C. AND BÉGUIN, P. (1988a). Purification of *Clostridium thermocellum* xylanase Z expressed in *Escherichia coli* and identification of the corresponding product in the culture medium of *C. thermocellum*. *Journal of Bacteriology* **170**, 4576–4581.

GRÉPINET, O., CHEBROU, M.-C. AND BÉGUIN, P. (1988b). Nucleotide sequence and deletion analysis of the xylanase gene (*xynZ*) of *Clostridium thermocellum*. *Journal of Bacteriology* **170**, 4582–4588.

GUINEY, D.G., HASEGAWA, P. AND DAVIS, C.E. (1984). Plasmid transfer from *Escherichia coli* to *Bacteroides fragilis*: differential expression of antibiotic resistance phenotypes. *Proceedings of the National Academy of Sciences, USA* **83** 7203–7206.

GUINEY, D.G., BOUIC, K., HASEGAWA, P. AND MATTHEWS, B. (1988). Construction of shuttle cloning vectors for *Bacteroides fragilis* and use in assaying foreign tetracycline resistance gene expression. *Plasmid* **20**, 17–22.

GUISEPPI, A., CAMI, B., AYMERIC, J.L., BAIL, G. AND CREUZET, N. (1988). Homology between endoglucanase Z of *Erwinia chrysanthemi* and endoglucanases of *Bacillus subtilis* and alkalophilic *Bacillus*. *Molecular Microbiology* **2**, 159–164.

HALL, J., BLACK, G.W., FERREIRA, L.M.A., MILLWARD-SADLER, S.J., ALI, B.R.S., HAZLEWOOD, G.P. AND GILBERT, H.J. (1995). The non-catalytic cellulose-binding domain of a novel cellulase from *Pseudomonas fluorescens* subsp. *cellulosa* is important for the efficient hydrolysis of Avicel. *Biochemical Journal* **309**, 749–756.

HARRIS, G.W., JENKINS, J.A., CONNERTON, I., CUMMINGS, N., LEGGIO, L.L., SCOTT, M., HAZLEWOOD, J.P., LALURIE, J.I., GILBERT, H.J. AND PICKERSGILL, R.W. (1994). Structure of the catalytic core of the family F xylanase from *Pseudomonas fluorescens* and identification of the xylopentaose-binding sites. *Structure* **2**, 1107–1116.

HEFFORD, M.A., TEATHER, R.M. AND FORSTER, R.J. (1993). The complete nucleotide sequence of a small cryptic plasmid from a rumen bacterium of the genus *Butyrivibrio*. *Plasmid* **29**, 6369.

HELBERS, K., WILKE, I. AND SONNEWALD, U. (1995). A thermostable xylanase from *Clostridium thermocellum* expressed at high levels in the apoplast of transgenic tobacco has no detrimental effects and is easily purified. *Bio/Technology* **13**, 63–66.

HENRISSAT, B. AND BAIROCH, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* **293**, 781–788.

HENRISSAT, B. AND BAIROCH, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. *Biochemical Journal* **316**, 695–696.

HENRISSAT, B., CLAEYSSENS, M., TOMME, P., LEMESLE, L. AND MORNON, J.-P. (1989). Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81**, 83–95.

HENRISSAT, B., CALLEBAUT, I., FABREGA, S., LEHN, P., MORNON, J.-P. AND DAVIES, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proceedings of the National Academy of Sciences USA* **92**, 7090–7094.

HESPELL, R.B. AND WHITEHEAD, T.R. (1991). Conjugal transfer of Tn*916*, Tn*916*ΔE and pAMβ1 from *Enterococcus faecalis* to *Butyrivibrio fibrisolvens* strains. *Applied and Environmental Microbiology* **57**, 2703–2709.

HOGMAN, S., JOVES, H., ROSENBERG, E. AND SHOHAM, Y. (1992). Bleachability improvement of softwood kraft pulp through treatment with an alkali- and thermostable xylanase. In: *Biotechnology in the Pulp and Paper Industry* (M. Kuwahara and M. Shimada, eds), pp. 107–113. UNI Publishers Co, Tokyo.

IMOTO, T., JOHNSON, L.N., NORTH, A.C.T., PHILLOPS, D.C. AND RUPLEY, J.A. (1972). In: *The enzymes* (P.D. Boyer, ed.), pp. 666–668. Academic Press, New York.

IRWIN, D., SPEZIO, M., WALKER, L.P. AND WILSON, D.B. (1993). Activity studies of eight purified cellulases: specificity, synergism and binding domain effects. *Biotechnology and Bioengineering* **42**, 1002–1013.

IVERSON, T. AND WANNSTROM, S. (1986). Lignin-carbohydrate bonds in a residual lignin isolated from pine kraft pulp. *Holzforshung* **40**, 19–22.

JAGER, A., SINNER, M., PURKARTHOFER, H., ESTERBAUER, H. AND DITZELMULLER, G. (1992). Biobleaching with xylanase from a thermophilic fungus. In: *Biotechnology in the Pulp and Paper Industry* (M. Kuwahara and M. Shimada, eds), pp. 115–121. UNI Publishers Co, Tokyo.

JAMIESON, S. (1991). Pulp environmental concerns put another demand on pulp producers. *Pulp Pap. Can.* **92**, 27–28.

JANSSON, J. AND PALENIUS, I. (1972). Aspects of the colour of kraft pulp and kraft lignin. *Paper Timber* **54**, 343–352.

JANSSON, J., PALENIUS, I., STENKUD, B. AND SAGFORSS, P.E. (1975). Differences in colour and strength of kraft pulps from batch and flow cooking. *Paper Timber* **57**, 387–390.

JAURIS, S., RUCKNAGEL, K.P., SCHWARZ, W.H., KRATZSCH, P., BRONNENMEIER, K. AND STAUDENBAUER, W.L. (1990). Sequence analysis of the *Clostridium stercorarium celZ* gene encoding a thermoactive cellulase (Avicelase I): identification of catalytic and cellulose-binding domains. *Molecular General Genetics* **223**, 258–267.

JENKINS, J., LEGGIO, L.L., HARRIS, G. AND PICKERSGILL, R. (1995). β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven. *FEBS Letters* **362**, 281–285.

JENSEN, L.G., OLSEN, O., COPS, O., WOLF, N., THOMSEN, K.K. AND VAN WETTSTEIN, D. (1996). Transgenic barley expressing a protein-engineered, thermostable (1,3-1,4)-β-glucanase during germination. *Proceedings of the National Academy of Sciences USA* **93**, 3487–3491.

JOLIFF, G., BÉGUIN, P. AND AUBERT, J.-P. (1986). Nucleotide sequence of the cellulase gene *celD* encoding endoglucanase D of *Clostridium thermocellum*. *Nucleic Acids Research* **14**, 8605–8613.

JUY, M., AMIT, A.G., ALZARI, P.M., POLJAK, R.J., CLAEYSSENS, M., BÉGUIN, P. AND AUBERT, J.-P. (1992). Three-dimensional structure of a thermostable bacterial cellulose. *Nature* **357**, 89–91.

KANTELINEN, A. (1992). Enzyme in bleaching of kraft pulp. In: *VTT Publications 144*. pp. 86. Technical Research Centre of Finland, Espoo.

KARITA, S., SAKKA, K. AND OHMIYA, K. (1996). Cellulose-binding domains confer enhanced activity against insoluble cellulose to *Ruminococcus albus* endoglucanase IV. *Journal of Fermentation and Bioengineering* **81**, 553–556.

KATO, K. (1981). Ultrastructure of the plant cell wall: biochemical viewpoint. *Encyclopedia of Plant Physiology New Series* **13B**, 29–46.

KATSUBE, Y., HATA, Y. AND YAMAGUCHI, H. (1990). Estimation of xylanase active site from crystalline structure. In *Protein Engineering* (M. Ikehara, ed.), pp. 91–96. Japan Scientific Societies Press, Tokyo.

KAWAMINAMI, S., OZAKI, K. AND ITO, S. (1995). Stereoselective hydrolysis catalyzed by a *Bacillus* endoglucanase in family D. *Biochemical and Biophysical Research Communications* 212, 539–543.

KAWAZU, T., OHTA, T., ITO, K., SHIBATA, M., KIMURA, T., SAKKA, K. AND OHMIYA, K. (1996). Expression of a *Ruminococcus albus* cellulase gene in tobacco suspension cells. *Journal of Fermentation and Bioengineering* **82**, 205–209.

KLYOSOV, A.A. (1990). Trends in biochemistry and enzymology of cellulose degradation. *Biochemistry* **29**, 10578–10585.

KOBAYASHI, Y. Personal communication. Accession Number D63938.

KOBAYASHI, Y., FOSTER, R.J., HEFFORD, M.A., TEATHER, R.M., WAKITA, M., OHMIYA, K. AND HOSHINO, S. (1995). Analysis of the sequence of a new cryptic plasmid, pRJF2, from a rumen bacterium of the genus *Butyrivibrio*; Comparison with other *Butyrivibrio* plasmids and application in the development of a cloning vector. *FEMS Microbiology Letters* **130**, 137–144.

KRAULIS, P.J., CLORE, G.M., NILGES, M., JONES, T.A., PETTERSON, G., KNOWLES, J. AND GRONENBORN, A.M. (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.

KRUUS, K., LUA, A.C., DEMAIN, A.L. AND WU, J.H.D. (1995). The anchorage function of CipA (CelL), a scaffolding protein of the *Clostridium thermocellum* cellulosome. *Proceedings of the National Academy of Sciences USA* **92**, 9254–9258.

LAMED, R., SETTER, L.E. AND BAYER, E.A. (1983). Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology* **156**, 828–836.

LANGE, N.K. (1993). In *Trichoderma reesei cellulase and other hydrolases* (P. Souminen and T. Reinikainen, eds) pp. 263–272. Foundation for Biotechnical and Industrial Fermentation, Helsinki.

LEIBOVITZ, E. AND BÉGUIN, P. (1996). A new type of cohesin domain that specifically binds the dockerin domain of the *Clostridium thermocellum* cellulosome integrating protein CipA. *Journal of Bacteriology* **178**, 3077–3084.

LINDER, M., LINDEBERG, G., REINIKAINEN, T., TEERI, T.T. AND PETTERSSON, G. (1995a). The difference in affinity between two fungal cellulose-binding domains is dominated by a single amino acid substitution. *FEBS Letters* **372**, 96–98.

LINDER, M., MATTINEN, M.L., KONTTELI, M., LINDEBERG, G., STAHLBERG, J., DRAKENBERG, T., REINIKAINEN, T., PETTERSSON, G. AND ANNILA, A. (1995b). Identification of function-

ally important amino acids in the cellulose-binding domain of *Trichoderma reesei* cellobiohydrolase I. *Protein Science* **4**, 1056–1064.

LUDVIGSEN, S. AND POULSON, M.F. (1992). Three-dimensional structure in solution of Barwin, a protein from barley seed. *Biochemistry* **31**, 8783–8789.

LUPAS, A., EHGELHARDT, H., PETERS, J., SANTARIUS, U., VOLKER, S. AND BAUMEISTER, W. (1994). Domain structure of the *Acetogenium kuvui* surface layer revealed by electron crystallography and sequence analysis. *Journal of Bacteriology* **176**, 1224–1233.

MACLEOD, A.M., LINDHORST, T., WHITHERS, S.G. AND WARREN, R.A.J. (1994). The acid/base catalyst in the exoglucanase/xylanase from *Cellulomonas fimi* is glutamic acid 127: evidence from detailed kinetic study of mutants. *Biochemistry* **33**, 6371–6376.

MAGLIONE, G., MATSUSHITA, O., RUSSELL, J.B. AND WILSON, D.B. (1992). Properties of a genetically reconstructed *Prevotella ruminiola* endoglucanase. *Applied Environmental Microbiology* **58**, 3593–3597.

MARKWALKER, H.U. AND NEUKOM, H. (1976). Diferulic acid as a possible crosslink in hemicelluloses from wheat germ. *Phytochemistry* **15**, 836–837.

MCGAVIN, M. AND FORSBERG, C.W. (1989). Catalytic and substrate-binding domains of endoglucanase 2 from *Bacteroides succinogenes*. *Journal of Bacteriology* **171**, 3310–3315.

MCNEIL, M., ALBERSHEIN, P., TAIZ, L. AND JONES, R.L. (1975). The structure of plant cell walls. VII. Barley aleurone cells. *Plant Physiology* **55**, 64–68.

MEINKE, A., DAMUDE, H.G., TOMME, P., KWAN, E., KILBURN, D.G., MILLER, JR, R.C., WARREN, R.A.J. AND GILKES, N.R. (1995). Enhancement of the endo-β-1,4-glucanase activity of an exocellobiohydrolase by deletion of a surface loop. *Journal of Biological Chemistry* **270**, 4383–4386.

MILLWARD-SADLER, S.J., POOLE, D.M., HENRISSAT, B., HAZLEWOOD, G.P., CLARKE, J.H. AND GILBERT, H.J. (1994). Evidence for a general role for high-affinity non-catalytic cellulose binding domains in microbial plant cell wall hydrolases. *Molecular Microbiology* **11**, 375–382.

MILLWARD-SADLER, S.J., DAVIDSON, K., HAZLEWOOD, G.P., BLACK, G.W., GILBERT, H.J. AND CLARKE, J.H. (1995). Novel cellulose-binding domains, NodB homologues and conserved modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas fluorescens* subsp. *cellulosa* and *Cellovibrio mixtus*. *Biochemical Journal* **312**, 39–48.

MIYAGI, T., KANEICHI, K., AMINOV, R.I., KOBAYASHI, Y., SAKKA, K., HOSHINO, S. AND OHMIYA, K. (1995). Enumeration of transconjugated *Ruminococcus albus* and its survival in the goat rumen microcosm. *Applied and Environmental Microbiology* **61**, 2030–2032.

MORAG, E., BAYER, E.A. AND LAMED, R. (1992). Affinity digestion for the near-total recovery of purified cellulosome from *Clostridium thermocellum*. *Enzyme Microbial Technology* **14**, 289–292.

MORAG, E., LAPIDOT, A., GOVORKO, D., LAMED, R., WILCHEK, M., BAYER, E.A. AND SHOHAM, Y. (1995). Expression, purification, and characterization of the cellulose-binding domain of the scaffolding subunit from the cellulosome of *Clostridium thermocellum*. *Applied and Environmental Microbiology* **61**, 1980–1986.

MURATA, M., HOSHINO, E., YOKOSUKA, M. AND SUZUKI, A. (1991). New detergent mechanism with use of novel alkaline cellulase. *J. Am. Oil Chem. Soc.* **68**, 553–558.

MURATA, M., HOSHINO, E., YOKOSUKA, M. AND SUZUKI, A. (1993). New detergent mechanism using cellulase revealed by change in physicochemical properties of cellulose. *J. Am. Oil Chem. Soc.* **70**, 53–58.

NIDETZKY, B., STEINER, W. AND CLAEYSSENS, M. (1994). Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: adsorption of two cellobiohydrolases, two endoglucanases and their core proteins on filter paper and their relation to hydrolysis. *Biochemical Journal* **303**, 817–823.

OGATA, K., AMINOV, R.I., NAGAMINE, T., BENNO, Y., SEKIZAKI, T., MITSUMORI, M., MINATO, H. AND ITABASHI, H. (1996). Structural organisation of pRAM4, a cryptic plasmid from *Prevotella ruminocola*. *Plasmid* **35**, 91–97.

OHMIYA, K., DEGUCHI, H. AND SHIMIZU, S. (1991). Modification of the properties of a

*Ruminococcus albus* endo-1,4-β-glucanase by the gene truncation. *Journal of Bacteriology* **173**, 636–641.

OKADA, H. (1991). Comparison of primary, secondary and tertiary structures of xylanase of *Bacillus pumilus* and cellulase of *Aspergillus acleatus*. In: *Microbial utilization of renewable resources* (S. Kinoshita and A. Bhumiratana, eds), pp. 1–7. International Center of Cooperative Research in Biotechnology, Japan.

ONG, E., GREENWOOD, J.M., GILKES, N.R., KILBURN, D.G., MILLER, JR, R.C. AND WARREN, R.A.J. (1989). The cellulose-binding domains of cellulases: tools for biotechnology. *Trends in Biotechnology* **7**, 239–243.

ONG, E., GILKES, N.R., MILLER, JR, R.C., WARREN, R.A.J. AND KILBURN, D.G. (1993). The cellulose-binding domain (CBDCex) of an exoglucanase from *Cellulomonas fimi*: production in *Escherichia coli* and characterization of the polypeptide. *Biotechnology and Bioengineering* **42**, 401–409.

ONG, E., ALIMONTI, J.B., GREENWOOD, J.M., MILLER, JR, R.C., WARREN, R.A. AND KILBURN, D.G. (1995). Purification of human interleukin-2 using the cellulose-binding domain of a prokaryotic cellulase. *Bioseparation* **5**, 95–104.

OWOLABI, J.B., BÉGUIN, P., KILBURN, D.G., MILLER, JR, R.C. AND WARREN, R.A.J. (1988). Expression in *Escherichia coli* of the *Cellulomonas fimi* structural gene for endoglucanase B. *Applied and Environmental Microbiology* **64**, 518–523.

OZAKI, K., HAYASHI, Y., SUMITOMO, N., KAWAI, S. AND ITO, S. (1995). Construction, purification and properties of a truncated alkaline endoglucanase from *Bacillus* sp. KSM-635. *Bioscience, Biotechnology and Biochemistry* **59**, 1613–1618.

PHEULPIN, P., TIERNY, Y., BECHET, M. AND GUILLAUME, J.-B. (1988). Construction of new shuttle plasmid vectors for *Escherichia coli*=*Bacteroides* transgeneric cloning. *FEMS Microbiology Letters* **55**, 15–22.

POOLE, D.M., DURRANAT, A.J., HAZLEWOOD, G.P. AND GILBERT, H.J. (1991). Characterization of hybrid proteins of the catalytic domains of *Clostridium* and *Ruminococcus* endoglucanases fused to *Pseudomonas* non-catalytic cellulose binding domains. *Biochemical Journal* **279**, 787–792.

POOLE, D.M., HAZLEWOOD, G.P., HUSKISSON, N.S., VIRDEN, R. AND GILBERT, H.J. (1993). The role of conserved tryptophan residues in the interaction of a bacterial cellulose binding domain with its ligand. *FEMS Microbiology Letters* **106**, 7784.

RATTO, M., SIIKA-AHO, M., BUCHERT, J., VALKEJARVI, A. AND VIIKARI, L. (1993). Enzymatic hydrolysis of isolated and fibre-bound galactoglucomannan from pine wood and pine kraft pulp. *Applied Microbiology and Biotechnology* **40**, 449–454.

REINIKAINEN, T., RUOHONEN, L., NEVANEN, T., LAAKSONEN, L., KRAULIS, P., JONES, T.A., KNOWLES, J.K.C. AND TEERI, T.T. (1992). Investigation of the function of mutated cellulose-binding domains of *Trichoderma reesei* cellobiohydrolase I. *Proteins* **14**, 475–482.

REINIKAINEN, T., TELLEMAN, O. AND TEERI, T.T. (1995). Effects of pH and high ionic strength on the adsorption and activity of native and mutated cellobiohydrolase I from *Trichoderma reesei*. *Proteins* **22**, 392–403.

ROSAHL, S., SCHMIDT, R., SCHELL, J. AND WILLMITZER, L. (1987). Expression of a tuber-specific storage protein in transgenic tobacco plants: demonstration of an esterase activity. *EMBO Journal* **6**, 11551159.

ROUVINEN, T., BERGFORS, T., TEERI, T.T., KNOWLES, J.K.C. AND JONES, T.A. (1990). Three-dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. *Science* **249**, 380–385.

SAKKA, K., KOJIMA, Y., KONDO, T., KARITA, S., OHMIYA, K. AND SHIMADA, K. (1993). Nucleotide sequence of the *Clostridium stercorarium xynA* gene encoding xylanase A: identification of catalytic and cellulose binding domains. *Bioscience, Biotechnology and Biochemistry* **57**, 273–277.

SAKKA, K., TAKADA, G., KARITA, S. AND OHMIYA, K. (1996). Identification and characterization of cellulose-binding domains in xylanase A of *Clostridium stercorarium*. *Annals of the New York Academy of Science* **782**, 241–251.

SALAMITOU, S., LEMAIRE, M., FUJINO, T., OHAYON, H., GOUNON, P., BÉGUIN, P. AND AUBERT,

J.-P. (1994a). Subcellular localization of *Clostridium thermocellum* ORF3p, a protein carrying a receptor for the docking sequence borne by the catalytic components of the cellulosome. *Journal of Bacteriology* **176**, 2828–2834.

SALAMITOU, S., RAYNAUD, O., LEMAIRE, M., COUGHLAN, M., BÉGUIN, P. AND AUBERT, J.-P. (1994b). Recognition specificity of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA. *Journal of Bacteriology* **176**, 2822–2877.

SCHWARZ, W.H., SCHIMMING, S. AND STAUDENBAUER, W.L. (1988). Degradation of barley β-glucan by endoglucanase C of *Clostridium thermocellum*. *Applied Microbiology and Biotechnology* **29**, 25–31.

SCOTT, K.P. AND FLINT, H.J. (1995). Transfer of plasmids between strains of *Escherichia coli* under rumen conditions. *Journal of Applied Biotechnology* **76**, 189–193.

SENIOR, D.J., HAMILTON, J. AND BERNIER, JR, R.L. (1992). Use of Streptomyces lividans xylanase for biobleaching of kraft pulps. In: *Xylans and Xylanase Progress in Biotechnology, Vol. 7.* (J. Visser, G. Beldman, M.A. Kusters van Someren and A.G.J. Voragen, eds), pp. 555–558. Elsevier Science Publishers, Amsterdam.

SHEN, H., TOMME, P., MEINKE, A., GILKES, N.R., KILBURN, D.G., WARREN, R.A.J. AND MILLER, JR, R.C. (1993). Stereochemical course of hydrolysis catalyzed by *Cellulomonas fimi* CenE, a member of a new family of β-1,4-glucanases. *Biochemical and Biophysical Research Communications* **199**, 1223–1227.

SHIMADA, K., KARITA, S., SAKKA, K. AND OHMIYA, K. (1994). Cellulases, xylanases and their genes from bacteria. In *Recombinant Microbes for Industrial and Agricultural Applications* (Y. Murooka and T. Imanaka, eds) pp. 395–429. Marcel Dekker Inc, New York.

SHOEMAKER, N.B., ANDERSON, K.L., SMITHSON, S.L., WANG, G.-R. AND SALYERS, A.A. (1991). Conjugal transfer of a shuttle vector from the human colinic anaerobe *Bactteroides uniformis* to the ruminal anaerobe *Prevotella (Bacteroides) ruminicola* $B_1$4. *Applied and Environmental Microbiology* **57**, 2114–2120.

SHOEMAKER, N.B., GUTHRIE, E.P., SALYERS, A.A. AND GARDNER, J.F. (1985). Evidence that the clindamycin-erythromycin resistance gene of *Bacteroides* plasmid pBF4 is on a transposable element. *Journal of Bacteriology* **162**, 626–632.

SMITH, C.J. (1985). Development and use of cloning systems for *Bacteroides fragilis*: cloning of a plasmid-encoded clindamycin resistance determinant. *Journal of Bacteriology* **164**, 294–301.

SPEZIO, M., WILSON, D.B. AND KARPLUS, P.A. (1993). Crystal structure of the catalytic domain of a thermophilic endoglucanase. *Biochemistry* **32**, 9906–9916.

STÅHLBERG, J., JOHANSSON, G. AND PETTERSSON, G. (1988). A binding-site-deficient, catalytically active, core protein of endoglucanase III from the culture filtrate of *Trichoderma reesei*. *European Journal of Biochemistry* **173**, 179–183.

STALBRAND, H., SALOHEIMO, A., VEHMAANPERÄ, J., HENRISSAT, B. AND PENTTILÄ, M. (1995). Cloning and expression in *Saccharomyces cerevisiae* of a *Trichoderma reesei* β-mannanase gene containing a cellulose binding domain. *Applied Environmental Microbiology* **61**, 1090–1097.

SUN, J., KAWAZU, T., KIMURA, T., SAKKA, K. AND OHMIYA, K. (1996). Highly expression of xylanase B gene from *Clostridium stercorarium* in tobacco cells. *Journal of Fermentation and Bioengineering*, submitted.

TAKADA, G., KARITA, S., TAKEUCHI, A., AHSAN, M.M., KIMURA, T., SAKKA, K. AND OHMIYA, K. (1996). Specific adsorption of *Clostridium stercorarium* xylanase to amorphous cellulose and its desorption by cellobiose. *Biosci. Biotech. Biochem.* **60**, 1183–1185.

TEUNISSEN, M.J. AND OP DEN CAMP, H.J.M. (1993). Anaerobic fungi and their cellulolytic and xylanolytic enzymes. *Antoni van Leewenhoek International Journal of Genetics and Molecular Microbiology* **63**, 63–76.

THOMSON, A.M. AND FLINT, H.J. (1989). Electroporation induced transformation of *Bacteroides ruminicola* and *Bacteroides uniformis* by plasmid DNA. *FEMS Microbiology Letters* **61**, 101–104.

TIMELL, T.E. (1962) Enzymatic hydrolysis of a 4-*O*-methylglucuronoxylan from the wood of the white birch (Betula papyrifera Marsh.). *Sven. Papperstidn.* **65**, 435–447.

TOKATLIDIS, K., DHURJATI, P. AND BÉGUIN, P. (1993). Properties conferred on *Clostridium thermocellum* endoglucanase CelC by grafting the duplicated segment of endoglucanase CelD. *Protein Engineering* **6**, 947–952.

TOKATLIDIS, K., SALAMITOU, S., BÉGUIN, P., DHURJATI, P. AND AUBERT, J.-P. (1991). Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. *FEBS Letters* **291**, 185–188.

TOMME, P., VAN TILBEURGH, H., PETTERSSON, G., VAN DAMME, J., VAN DERKEREKHOVE, J., KNOWLES, J., TEERI, T. AND CLAEYSSENS, M. (1988). Studies of the cellulolytic system of *Trichoderma reesei* QM9414, analysis of domain function in two cellobiohydrolases by limited proteolysis. *European Journal of Biochemistry* **170**, 575–581.

TOMME, P., GILKES, N.R., MILLER, JR, R.C., WARREN, R.A.J. AND KILBURN, D.G. (1994). An internal cellulose-binding domain mediates adsorption of an engineered bifunctional xylanase/cellulase. *Protein Engineering* **7**, 117123.

TOMME, P., DRIVER, D.P., AMANDORON, E.A., MILLER, JR, R.C., WARREN, R.A.J. AND KILBURN, D.G. (1995a). Comparison of a fungal (family I) and bacterial (family II) cellulose-binding domain. *Journal of Bacteriology* **177**, 4356–4363.

TOMME, P., WARREN, R.A.J. AND GILKES, N.R. (1995b). Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology* **37**, 1–81.

TOMME, P., WARREN, R.A.J., MILLER, JR, R.C., KILBURN, D.G. AND GILKES, N.R. (1995c). Cellulose-binding domains – classification and properties. In: *The enzymatic degradation of insoluble polysaccharides* (W.M. Saddler and M. Penner, eds), pp. 142–161. American Chemical Society, Washington.

TÖRRÖNEN, A., HARKKI, A. AND ROUVINEN, J. (1994). Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: two conformational states in the active site. *EMBO Journal* **13**, 2493–2501.

TÖRRÖNEN, A. AND ROUVINEN, J. (1994). Structural comparison of two major endo-1,4-β-xylanase from *Trichoderma reesei*. *Biochemistry* **34**, 847–856.

TÖRRÖNEN, A., KUBICEK, C.P. AND HENRISSAT, B. (1993). Amino acid sequence similarities between low molecular weight endo-1,4-β-xylanase and family H cellulases revealed by clustering analysis. *FEBS Letters* **321**, 135–139.

TULL, D. AND WITHERS, S.G. (1994). Mechanism of cellulases and xylanases: a detailed kinetic study of the exo β-1,4-glycanase from *Cellulomonas fimi*. *Biochemistry* **33**, 6363–6370.

TULL, D., WITHERS, S.G., GILKES, N.R., KILBURN, D.G., WARREN, R.A.J. AND AEBERSOLD, R. (1991). Glutamic acid 274 is the nucleophile in the active site of a 'retaining' exoglucanase from *Cellulominas fimi*. *Journal of Biological Chemistry* **266**, 15621–15625.

VALENTINE, P.J., SHOEMAKER, N.B. AND SALYERS, A.A. (1988). Mobilization of *Bacteroides* plasmids by *Bacteroides* conjugal elements. *Journal of Bacteriology* **170**, 1319–1324.

WARE, C.E., BAUCHOP, T., HUDMAN, J.F. AND GREGG, K. (1992). Cryptic plasmid pBf1 from *Butyrivibrio fibrisolvens* AR10: Its use as a replicon for recombinant plasmids. *Current Microbiology* **24**, 193–197.

VIIKARI, L., RANUA, M., KANTELINEN, A., SUNDQUIST, J. AND LINKO, M. (1986). Bleaching with enzymes. In: *Proc 3rd Int Conf Biotechnol Pulp Pap Ind, Stockholm*, pp. 67–69.

VIIKARI, L., KANTELINEN, A., SUNDQUIST, J. AND LINKO, M. (1994). Xylanases in bleaching: from an idea to the industry. *FEMS Microbiology Reviews* **13**, 335–350.

WHITE, A., WITHERS, S.G., GILKES, N.R. AND ROSE, D.R. (1994). Crystal structure of the catalytic domain of the β-1,4-glycanase Cex from *Cellulomonas fimi*. *Biochemistry* **33**, 12546–12552.

WHITEHEAD, T.R. (1992). Genetic transformation of the ruminal bacteria *Butyrivibrio fibrisolvens* and *Streptococcus bovis* by electroporation. *Lett Appl Microbiol* **15**, 186–189.

WHITEHEAD, T.R. AND HESPELL, R.B. (1989). Cloning and expression in *Escherichia coli* of a xylanase gene from *Bacteroides ruminicola 23*. *Applied and Environmental Microbiology* **55**, 893–396.

WHITEHEAD, T.R. AND HESPELL, R.B. (1990). Heterologous expression of the *Bacteroides ruminicola* xylanase gene in *Bacteroides flagilis* and *Bacteroides uniformis*. *FEMS Microbiology Letters* **66**, 61–66.

WINTERHALTER, C., HEINRICH, P., CANDUSSIO, A., WICH, G. AND LIEBL, W. (1985). Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium *Thermotoga maritima*. *Molecular Microbiology* ***15***, 431–444.

XU, G.-Y., ONG, E., GILKES, N.R., KILBERN, D.G., MUHANDIRAM, D.R., HARRIS-BRANDTS, M., CARVER, J.P., KAY, L.E. AND HARVEY, T.S. (1995). Solution structure of a cellulose-binding domain from *Cellulomonas fimi* by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**, 6993–7009.

YAMASAKI, T., HOSOYA, S., CHEN, C.-L., GRATZL, J.S. AND CHANG, H.-M. (1981). Characterization of residual lignin in kraft pulp. In: *The Ekman-days international symposium on wood and pulping chemistry. Vol. 2*, pp. 34–42. Swedish Forest Products Laboratory. Stockholm.

YARON, S., MORAG, E., BAYER, E.A., LAMED, R. AND SHOHAM, Y. (1995). Expression, purification and subunit-binding properties of cohesin 2 and 3 of the *Clostridium thermocellum* cellulosome. *FEBS Letters* **360**, 121–124.

ZHANG, N. AND BROOKER, J.D. (1993). Characterization, sequence and replication of a small cryptic plasmid from *Selenomonas ruminantium* subspecies *lactilytica*. *Plasmid* **29**, 125–134.