JOURNAL OF BACTERIOLOGY, July 1992, p. 4677–4682
0021-9193/92/144677-06$02.00/0

Vol. 174, No. 14

The Catalytic Domain of Endoglucanase A from *Clostridium cellulolyticum*: Effects of Arginine 79 and Histidine 122 Mutations on Catalysis

A. BELAICH,[1]* H.-P. FIEROBE,[1] D. BATY,[2] B. BUSETTA,[3] C. BAGNARA-TARDIF,[1,4] C. GAUDIN,[1] AND J.-P. BELAICH[1,4]

Laboratoire de Chimie Bactérienne[1] and Laboratoire de Biochimie et Biologie Moléculaire,[2] Centre National de la Recherche Scientifique, and Université de Provence,[4] Marseille, and Laboratoire de Cristallographie et de Physique Cristalline, Bordeaux,[3] France

Received 25 February 1992/Accepted 18 May 1992

Sequence analysis of the endoglucanase EGCCA of *Clostridium cellulolyticum* indicates the existence of two domains: a catalytic domain extending from residue 1 to residue 376 and a reiterated domain running from residue 390 to 450. A small deletion in the C terminal end of the catalytic domain inactivated the protein. From the analysis of the sequences of 26 endoglucanases belonging to family A, we focused on seven amino acids which were totally conserved in all the catalytic domains compared. The roles of two of these, Arg-79 and His-122, were studied and defined on the basis of the mutants obtained by introducing various substitutions. Our findings suggest that Arg-79 is involved in the structural organization of the protein; the His-122 residue seems to be more essential for catalysis. The role of His-123, which is conserved only in subfamily A4, was also investigated.

Clostridium cellulolyticum is a mesophilic anaerobic bacterium which is able to degrade crystalline cellulose (13). Studies of the various components of the cellulolytic complex of this bacterium are in progress, and previous studies have dealt with the cloning (8) and sequencing (9) of the *celCCA* gene, which encodes the endoglucanase EGCCA. Numerous genes of cellulases have now been sequenced, and the corresponding enzymes have been divided into nine different families on the basis of sequence comparisons and hydrophobic cluster analysis (14, 17, 18). According to this classification, EGCCA is a member of family A. At least 26 genes of cellulases belonging to family A have been sequenced to date, but none of their three-dimensional (3-D) structures have been determined. At present, the only 3-D structure described in the literature is that of cellobiohydrolase II from *Trichoderma reesei* (25), which belongs to family B. A second structure description should be published soon, as *Clostridium thermocellum* endoglucanase CELD (family E) has been crystallized (19) and its 3-D structure has been determined (20). Since the structure of the cellulases of family A is unknown, information about the structure-function relationships must be deduced from the primary sequences. Using hydrophobic cluster analysis, Henrissat et al. (17) found that five segments were conserved, including several amino acids which were particularly well conserved. In the present study, we explored the roles of three of these amino acids: arginine 79 and histidine 122, which are conserved throughout the family, and histidine 123, which is conserved only in subfamily A4.

MATERIALS AND METHODS

Strains and media. *Escherichia coli* TG1 (30) was used as the host for recombinant plasmids. *E. coli* cells were grown aerobically with shaking at 37°C in Luria-Bertani (LB) medium. Ampicillin (100 μg/ml), isopropyl-β-D-thiogalactopyranoside (IPTG; 1 mM), or both were added to the medium when necessary.

Labeling experiments were carried out in M9 minimal medium supplemented with ampicillin and with 1% glycerol as the carbon source.

DNA techniques. Plasmid DNA preparations, single-stranded DNA preparations, and DNA fragment isolation procedures used were based on those described by Sambrook et al. (26). DNA sequencing was carried out by the dideoxy method with a United States Biochemical Sequenase kit.

Site-directed mutagenesis. Plasmid pA2 (10) was the parent of all the plasmids used throughout this study. This plasmid was first digested by *Nco*I at the unique *Nco*I site, and the end was filled in by the Klenow enzyme in the presence of deoxynucleoside triphosphates. Afterwards, it was digested by *Eco*RI, yielding a 700-bp *Eco*RI-*Nco*I-filled-in fragment of *celCCA* (containing the first 700 bp of the coding sequence of *celCCA*). This fragment was then inserted into the *Eco*RI-*Hin*cII sites of the M13 mp18 bacteriophage vector, which restores the *Nco*I site. Site-directed mutagenesis was performed with an oligonucleotide-directed in vitro mutagenesis kit from Amersham by the method of Taylor et al. (27). Mutagenic oligonucleotides (22- to 24-mers) were synthesized with a DNA synthesizer. The whole 700-bp *Eco*RI-*Nco*I *celCCA* fragment carrying a mutation was sequenced to ensure the absence of any secondary mutations. To construct the mutants, the wild-type *Eco*RI-*Nco*I fragment of pA2 was replaced by homologous fragments having a single mutation.

Assay of EGCCA protein. The concentrations of EGCCA present in cultures expressing the mutated genes and the half-lives of these proteins were estimated by radiolabeling with [^{35}S]methionine. The cells to be labeled were grown overnight in minimal medium and then diluted to 10 ml with fresh minimal medium to give an optical density of 0.8 at 600 nm. IPTG (2 mM final concentration) was added, and the

* Corresponding author.

induction was continued for 10 min. The culture was then divided into three parts: the first part (2.5 ml) for long-term labeling, the second part (2.5 ml) for pulse-chase labeling, and the third part (5 ml) to measure the biochemical activities and optical densities. For long-term labeling, 63 μCi of [^{35}S]methionine and 10 μl of a 1% methionine solution were added to the first portion. For pulse-chase labeling, 42 μCi of [^{35}S]methionine was added to the second portion, and after a pulse of 1 min, 50 μl of a 5% methionine solution was added to chase. Aliquots of 400 μl were removed at specific intervals and centrifuged. The pellets from portions 1 and 2 were resuspended in 80 μl of sodium dodecyl sulfate (SDS) sample buffer prior to electrophoresis. The pellets from portion 3 were resuspended in Tris buffer (pH 7.0), sonicated in a Branson sonicator, and centrifuged. The pellet was discarded, and the supernatant was used to measure the enzymatic activities.

Obtaining cell extracts. In some cases, when larger quantities of cell extract were necessary to perform the experiments, the experimental protocol used was as follows. Cells harboring mutations in the *celCCA* gene were grown in 500 ml of LB medium supplemented with ampicillin and IPTG until an optical density at 600 nm of 2 was reached. The cells were pelleted by centrifugation, resuspended in buffer I (50 mM Tris [pH 7], 1 mM EDTA) and broken up in a French press. Nucleic acids were precipitated with streptomycin sulfate (20 g/liter). Ammonium sulfate (60% saturation) was then added to the supernatant, and the extract was centrifuged again. The resulting pellet, resuspended in 20 ml of buffer I, was the cell extract. The amount of mutated protein in the extract was estimated by SDS-polyacrylamide gel electrophoresis (PAGE) with gels stained with Coomassie blue, by comparing the intensity of the protein band with that of wild-type EGCCA obtained under the same conditions. Since these proteins (wild-type or mutated EGCCA) have a high level of expression in *E. coli*, they can be unambiguously identified.

Enzyme assays. The carboxymethyl cellulase (CMCase) activity and the *p*-nitrophenyl cellobiosidase (pNPCase) activity were assayed as previously described (10). The protein concentration was determined by Lowry's method (21) with bovine serum albumin as the standard.

RESULTS

All of the experiments were carried out with plasmid pA2 (10), which corresponds to a genetic construction in which the *celCCA* gene has been placed under the control of the P_{tac} promoter in plasmid pJF118EH, leading to overexpression of the EGCCA in *E. coli*.

Analysis of truncated EGCCA derivatives. The primary sequence of EGCCA shows that this protein consists of two domains: a catalytic domain and a repeated stretch of 21 amino acid (aa) residues at the C terminus of the protein. This reiterated domain seems to be typical of the cellulases belonging to cellulolytic clostridia (2–4, 11). The catalytic domain has been compared with other catalytic domains of cellulases belonging to family A (17), and the five conserved sequence elements identified are shown in Fig. 1.

Deletions in the C terminus were performed with the specific restriction sites in *celCCA*, namely, *Asu*II and *Nco*I, which are located in the DNA region encoding the 376th and 258th amino acid residues, respectively. A universal translation terminator was inserted; the end of the gene was deleted so that the *rrnB* transcriptional terminator would be placed just after the translational terminator. The corresponding plasmids pA3 and pA4 encoded truncated versions of EGCCA with molecular masses of 43 kDa (EGCCA-A3) and 30 kDa (EGCCA-A4), respectively (Fig. 1).

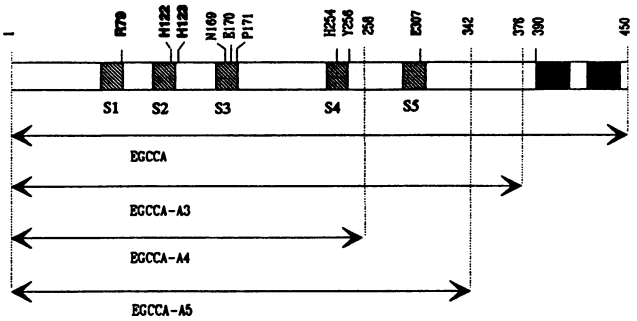

FIG. 1. Organization of domains in EGCCA. Boldface characters indicate the amino acids which were chosen for mutagenesis, and dotted lines indicate the limits of the deletion mutant proteins. Symbols: ▧, conserved clusters in family A cellulases; ■, clostridial reiterated sequences.

EGCCA-A3 does not contain the reiterated sequence. It is approximately the same size as the truncated form of the EGCCA obtained by proteolysis of the entire EGCCA (10) (the two proteins differ by only 5 aa). It is very stable; the half-life of the protein within recombinant *E. coli* cells is at least fivefold longer than that of EGCCA (about 8 h, compared with 1.6 h for EGCCA) (data not shown). The protein was purified as described by Fierobe et al. (10) to purify the small form of EGCCA. The K_m and V_{max} measured with carboxymethyl cellulose (CMC) as a substrate were found to be 4.3 g/liter and 130 IU/mg, respectively. These values are almost identical to those found for the truncated EGCCA. A slight difference in the pNPCase activity was observed; the level of this activity was 3 IU/M for EGCCA-A3 and 1.4 IU/M for EGCCA. EGCCA-A4 corresponds to a protein truncated just after the fourth conserved segment (S4 in Fig. 1). This protein (produced after induction by IPTG) is completely inactive on CMC.

To determine the minimum size of the protein displaying catalytic properties, *celCCA* was linearized with *Asu*II and digested with exonuclease BAL 31; the termini of the DNA were filled in with the Klenow fragment of *E. coli* polymerase I, and a translational terminator was inserted. The clones which harbored a plasmid containing the *Nco*I site were analyzed. The largest one corresponded to a truncation of 96 bp of the gene from the *Asu*II site. The corresponding protein (named EGCCA-A5), which contained all five segments identified by Henrissat et al. (17), was inactive on CMC, indicating that the region downstream from the conserved fragments is essential for activity.

Analysis of amino acid substitution in EGCCA. As mentioned above, five conserved segments (designated S1 to S5) were identified in the catalytic domain of the cellulases belonging to family A. Among these stretches, the following amino acid residues were found to be particularly well conserved: Arg-79, His-122, Asn-169, Glu-170, His-254, Tyr-256, and Glu-307. The catalysis probably involves an acid-base mechanism similar to that proposed in the case of lysozyme. In our case, His-122 was suspected by Henrissat et al. (17) of acting as a proton donor; the negatively charged side chain of Glu-170 may stabilize the reaction intermediate carbonium cation. We focused on the two amino acids, His-122 (segment 2) and Arg-79 (segment 1), which are

TABLE 1. Variations in the specific activities of the various EGCCA mutants relative to the activity of native EGCCA

Protein	Relative sp act (%) of:			
	CMCase		pNPCase	
	pH 7.00	pH 5.6	pH 7.00	pH 5.9
EGCCA	100.0	100.0	100.0	100.0
EGCCA-A3	178.0	ND[a]	225.0	ND
$EGCCA_{R79K}$	22.0	21.0	9.0	ND
$EGCCA_{R79S}$	3.0	ND	9.7	ND
$EGCCA_{R79V}$	3.7	3.7	9.0	ND
$EGCCA_{H122S}$	0.8	5.0	0.0	0.0
$EGCCA_{H122G}$[b]	1.5	7.6	0.0	0.0
$EGCCA_{H122E}$	<0.02	0.8	0.0	0.0
$EGCCA_{H122F}$[b]	1.4	12.4	0.0	0.0
$EGCCA_{H123V}$[b]	1.4	16.4	0.5	0.6
$EGCCA_{H122S,H123V}$[b]	<0.01	0.1	0.0	0.0

[a] ND, not determined.
[b] The specific activities of these mutants were estimated by using the apparent synthesis level observed on the polyacrylamide gel.

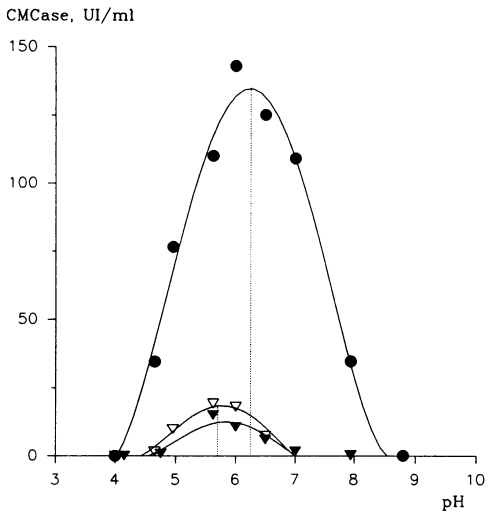

FIG. 2. Influence of pH on the CMCase activities of wild-type EGCCA (●) and mutated proteins $EGCCA_{H122F}$ (▼) and $EGCCA_{H123V}$ (▽). Dotted lines indicate the maximum activities. UI, international units.

conserved in all members of family A, and on His-123, which is conserved only in subfamily A4.

(i) Site-directed mutagenesis of the Arg-79 residue. By using synthetic oligonucleotides, Arg-79 of wild-type EGCCA was replaced by either lysine, serine, or valine. The corresponding mutant proteins are referred to as $EGCCA_{R79K}$, $EGCCA_{R79S}$, and $EGCCA_{R79V}$, respectively. The half-lives of the mutated proteins were measured by pulse-chase experiments (see Materials and Methods) and then compared with that of wild-type EGCCA. The half-lives were found to be practically identical. This result indicates that all these mutants have the same sensitivity to cellular proteases and suggests that their 3-D structures are very similar. The CMCase and pNPCase activities of these mutated proteins were measured by using cell extracts. In order to correct for variations in the level of expression between the various mutants, the EGCCA concentration in the extracts was estimated either by performing radiolabeling experiments or by estimating the intensity of the protein bands in SDS-PAGE (see Materials and Methods). The results are given in Table 1. Replacing Arg-79 with Ser or Val results in a dramatic loss of CMCase activity; as expected, the CMCase activity was less strikingly reduced in the case of replacement by Lys. This result can easily be explained by the fact that Arg and Lys are functionally quite similar. The pNPCase activity was also measured, and the results (Table 1) showed no significant difference between $EGCCA_{R79K}$, $EGCCA_{R79S}$, and $EGCCA_{R79V}$. The K_ms of $EGCCA_{R79V}$ and $EGCCA_{R79K}$ for CMC in the cell extracts were found to be six and three times higher, respectively, than the K_m of the wild-type EGCCA (2 g/liter).

(ii) Site-directed mutagenesis of His-122. As can be seen in Fig. 3, histidine 122 is conserved in all of the cellulases of family A. It was replaced by either serine, glutamic acid, glycine, or phenylalanine, and the corresponding mutant proteins are referred to as $EGCCA_{H122S}$, $EGCCA_{H122E}$, $EGCCA_{H122G}$, and $EGCCA_{H122F}$, respectively. The half-lives of the mutated proteins were measured (data not shown) and found to be identical to that of EGCCA. A strong decrease in all of the CMCase activities was observed, with $EGCCA_{H122E}$ showing the greatest decrease. With all the His-122 mutants, the maximum activity was found at a lower pH (Table 1 and Fig. 2 for $EGCCA_{H122F}$); this was not so in the case of the mutations involving Arg-79. The K_m for CMC was measured by using cell extracts containing $EGCCA_{H122F}$ or $EGCCA_{H122G}$; it was found to be similar to the K_m of wild-type EGCCA. In all cases and at every pH tested, the pNPCase activity was abolished.

These results suggest that His-122 is fundamental to catalysis. No pNPCase activity was measured with the mutated proteins, but a significant level of CMCase activity was found, for example, at pH 5.6, especially with mutant $EGCCA_{H122F}$. If His-122 acts as the proton donor, it is possible that, in the case of the mutations, another amino acid serves as a substitute for the proton donor in catalysis. It can be seen in Fig. 3 that in subfamily A4 another histidine occurs after the first one, and the second histidine might be involved in catalysis.

(iii) Site-directed mutagenesis of His-123. To verify the assumption that the second histidine is involved in catalysis, His-123 was replaced by valine. This amino acid was chosen because it occurs after the conserved histidine in some of the cellulases in family A. The mutated protein was called $EGCCA_{H123V}$. Surprisingly, this mutation entailed a considerable decrease in the CMCase activity, and, as in the case of the various His-122 mutants of EGCCA, the maximum activity was shifted to lower pH levels (Table 1 and Fig. 2). In this case, the pNPCase activity was very low but nevertheless significant compared with the nonexistent activity in all of the His-122 mutants. The K_m of CMC was measured with cell extracts containing $EGCCA_{H123V}$ and was found to be similar to the K_m of wild-type EGCCA.

(iv) Mutagenesis of His-122 and His-123. A double mutant in which His-122 and His-123 were replaced by serine and valine, respectively, was obtained. The mutated protein $EGCCA_{H122S,H123V}$ was tested to determine the CMCase and pNPCase activities. This protein is almost totally inactive, as the maximum CMCase activity found at pH 5.6 was only 0.1% of the CMCase activity of the wild-type protein.

DISCUSSION

The results of this study confirm that EGCCA consists of two separate domains: a catalytic domain and a reiterated

FIG. 3. Sequence alignment of a conserved segment in family A cellulases (Cel A1 to A5), a β-xylosidase (BX), and family A β-glucosidases (BGA). The subclassification of cellulases (A1 to A5) is that of Béguin (3). The designations used consist of the enzyme abbreviation (EG for endoglucanase, BX for β-xylosidase, BG for β-glucosidase, BGAL for β-galactosidase, PGLU for phosphoglucosidase, PGAL for phosphogalactosidase, and LP for lactase-phlorizin hydrolase) followed by the origin of the organism (CT for *C. thermocellum*, CC for *C. cellulolyticum*, CA for *C. acetobutylicum*, CCV for *C. cellulovorans*, CS for *Caldocellum saccharolyticum*, BN4 for *Bacillus* strain N4, BSP for *Bacillus* sp., BP for *Bacillus polymyxa*, BS for *Bacillus subtilis*, EC for *E. coli*, ECH for *Erwinia chrysanthemi*, BF for *Butyrivibrio fibrisolvens*, RA for *Ruminococcus albus*, BR for *Bacteroides ruminicola*, ASP for *Agrobacterium* sp., LC for *Lactobacillus casei*, LL for *Lactococcus lactis*, SA for *Staphylococcus aureus*, SL for *Streptococcus lactis*, SS for *Sulfolobus solfataricus*, TR for *T. reesei*, and H for human) followed by a number or a letter when the organism produces several of these enzymes. Further details of the sequences analyzed can be found in the reports by Béguin (3), Gilkes et al. (14), and Gräbnitz et al. (16). The β-xylosidase sequence (BXCSB) is from the report of Lüthi et al. (22). R (or K in one case), H, and E in large type correspond to R79, H122, and E170 from segments S1, S2, and S3 in EGCCA (see Fig. 1).

domain. This reiterated domain has been observed up to now only in clostridia, and its role is likely to be a polyvalent one. Fierobe et al. (10) have demonstrated that its deletion from EGCCA resulted in an enhancement of the CMCase activity, which was correlated with a decrease in the avicelase activity. On the other hand, Tokatlidis et al. (28) reported that its presence in CelD from *C. thermocellum* was responsible for protein-protein interactions with other cellulases and especially with the large noncellulase subunit from the cellulosome. The two sets of data may both be valid, and, as previously observed (15) with the loss of cellulose-binding domains in *Cellulomonas fimi* cellulases, the deletion of the protein-anchoring domain might induce considerable changes in the enzyme activities. The protein EGCCA-A3, which is produced by gene truncation at the *Asu*II site, has the same properties as those obtained by proteolysis of the entire EGCCA (10). When 33 aa were deleted from the C terminus of EGCCA-A3, no CMCase activity was observed, although the five stretches of amino acids conserved in family A were preserved. These 33 aa may serve to maintain the structure of the enzyme. Similar data were obtained in the case of the alkaline cellulase of a *Bacillus* species (12) and in that of the endoglucanase Z (EGZ) of *Erwinia chrysanthemi* (23), where truncations of 32 and 22 aa, respectively, from the C-terminal end of the catalytic domains led to inactive proteins.

To obtain some information about the 3-D structure of the catalytic domain of EGCCA, we compared the distribution of its predicted secondary structure with similar distributions in proteins with known X-ray structures as described

by Busetta (6). This comparison suggested a 3-D structure which was very similar to the complete so-called triose phosphate isomerase (TIM) eight-stranded α/β barrel often encountered in glycosidic enzymes (24) as well as in (TIM) (1) or the larger domain of xylose isomerase (7), rather than the truncated barrel recently described as present in the cellobiohydrolase of *T. reesei* (25). Furthermore, with this model, the conserved residues, which are presumably involved in the active site, are mainly located at the carboxyl-terminal end of the predicted β-sheet structures, as is the case in all of the parallel β-sheet enzymes (5).

A homologous region has recently been assumed to exist in cellulase family A and β-glucosidase family A (BGA) (16). This region, corresponding to the N-terminal side in the two enzyme families, involves Arg-79, His-122, and Glu-170 in EGCCA. Modeling sequences of family BGA confirmed that a common 3-D structure predictably exists in the N-terminal moieties of the two families. The cellulases of family A do not cleave disaccharides, so the only common action performed by both of these enzymes is the breakdown of the β(1-4) glycosidic bond, and it is tempting to attribute this precise role to this homologous region. In the case of β-glucosidase from an *Agrobacterium* species, Wither et al. (31) demonstrated that Glu-346 is involved at the active site of this enzyme, and the mechanism may therefore be more complex than that deduced simply on the basis of regional homologies.

Arg-79 clearly plays an important role in endoglucanase activity, but it does not seem to be absolutely necessary for catalytic activity, since the three mutated proteins were still active on CMC and *p*-nitrophenyl cellobioside (pNPC). Substituting Ser or Val for Arg induced a dramatic decrease in the CMCase activity, and this activity was not enhanced at pH 5.6. This decrease was less significant when lysine was substituted for arginine. Lysine, like arginine, has a positive charge and is a conservative replacement. This relatively high level of conservation was not observed with the pNPCase activity, however; in this case, replacing arginine with lysine had the same effect as replacing arginine with valine or serine. The affinity of the mutated proteins for CMC decreased significantly, but this decrease was smaller in the case of the $EGCCA_{R79K}$ protein. It is likely that Arg-79 is involved in the formation of a salt bridge in the protein or forms a hydrogen bond to the substrate. His-122, which is conserved throughout family A, appears to be essential for catalysis. A histidyl residue (His-516) has been identified in the catalytic center of endoglucanase D of *C. thermocellum* (29), which belongs to family E. Substituting Ser for His reduced the pNPCase activity of the protein by only 66%, and the authors concluded that this residue was not the proton donor (29). In the case of the EGZ of *E. chrysanthemi* (family A, subfamily A2), it has been shown that His-98 and Glu-133, which correspond to His-122 and Glu-179 of EGCCA from *C. cellulolyticum*, respectively, play an important role in catalysis (23). The recent studies by Wither et al. (31) indicate in fact that in family BGA, the corresponding Glu is not the nucleophile partner in catalysis. In the case of EGCCA, there are several arguments in favor of the hypothesis that His-122 is directly involved in catalysis. Substituting Ser, Gly, or Phe for His-122 resulted in a dramatic decrease in CMCase activity at pH 7.0, but this effect was less pronounced at pH 5.6, where the maximum activity was shifted. Substitutions involving a charged amino acid such as Glu had even more dramatic results. The pNPCase activity was equal to zero regardless of the pH. The presence of two histidines (His-122 and His-123) gives the protein a nearly neutral isoelectric point, and the maximum activity was found at pH 6.5. A mutation of His-122 or His-123 shifted the maximum CMCase activity to a lower pH (about 1.0 pH unit lower). The His-123 mutation also resulted in a considerable loss of CMCase activity; in this case, however, the pNPCase activity was very low, although significant in comparison with the His-122 mutations. His-123 probably plays a role in the local structure of the catalytic site or in the positioning of the substrate. If we refer to the xylose isomerase model, reactive His-54 is correctly oriented by Asp-57. In the case of family A, an aspartate is located 2 aa before the reactive histidine except in the case of subfamily A4, where an aspartate or a glutamate occurs after the His-His sequence. This aspartate may influence the orientation of His-122 and His-123 in the protein.

One noteworthy point which emerged here concerns the changes observed in the substrate specificity. The loss of activity was more dramatic when pNPC was used as the substrate instead of CMC. These changes may be attributable to the considerable conformational differences which exist between these substrates. pNPC is a short molecule, but the *p*-nitrophenyl group gives it steric bulkiness, whereas CMC is a long molecule presumably occupying all the active and binding sites of the enzyme. At first analysis, there are two hypotheses which might explain the observed changes. (i) A mutation induces such a major change in the active site that it becomes inaccessible to pNPC. (ii) CMC, by binding to a larger number of residues in a protein than pNPC does, stabilizes the protein structure so that the activity is partly conserved.

At the present stage, our assumptions can be confirmed only through X-ray diffraction analysis. As the production of EGCCA-A3 in large amounts is now possible in our laboratory, crystallogenesis trials are now being undertaken to complete this analysis; the preliminary results are promising. By analyzing EGCCA-A3 by X-ray diffraction along with other directed mutagenesis experiments, we hope to be able to elucidate the catalytic mechanism of this enzyme.

ACKNOWLEDGMENTS

We thank the Service de production de bactéries, LCB, for providing bacterial cultures. We are indebted to Claude Lazdunski and Eric Faure for their interest and helpful discussions. We are grateful to Jessica Blanc for corrections to the manuscript.

This work was supported by grants from the European Economic Community and from the Gaz de France Co.

REFERENCES

1. **Alber, T., D. W. Banner, A. C. Bloomer, G. A. Petsko, D. Phillips, P. S. Rivers, and I. A. Wilson.** 1981. On the three dimensional structure and catalytic mechanism of triose phosphate isomerase. Philos. Trans. R. Soc. Lond. B **293**:159–171.
2. **Aubert, J.-P., P. Beguin, and J. Millet.** Genes involved in cellulose and xylane degradation by *Clostridium thermocellum*. *In* Genetics and molecular biology of anaerobes, in press.
3. **Béguin, P.** 1990. Molecular biology of cellulose degradation. Annu. Rev. Microbiol. **44**:219–248.
4. **Belaich, J. P., A. Belaich, C. Gaudin, and C. Bagnara.** Genes and proteins involved in cellulose degradation by mesophilic *Clostridia*. *In* Genetics and molecular biology of anaerobes, in press.
5. **Branden, C. J.** 1980. Relation between structure and function of α/β proteins. Q. Rev. Biophys. **13**:317–338.
6. **Busetta, B.** 1988. The use of folding patterns in the search of protein structural similarities: a three dimensional model of phosphoribosyl transferases. Biochim. Biophys. Acta **957**:21–33.
7. **Carrel, H. L., J. P. Glusker, V. Burger, F. Monfre, D. Tritsch,**

and J. F. Biellmann. 1989. X-ray analysis of D-xylose isomerase at 1.9 Å: native enzyme in complex with substrate and with a mechanism-designed inactivator. Proc. Natl. Acad. Sci. USA **86:**4440–4444.

8. **Faure, E., C. Bagnara, A. Belaich, and J. P. Belaich.** 1988. Cloning and expression of two cellulase genes of *Clostridium cellulolyticum* in *Escherichia coli*. Gene **65:**51–58.
9. **Faure, E., A. Belaich, C. Bagnara, C. Gaudin, and J. P. Belaich.** 1989. Sequence analysis of the *Clostridium cellulolyticum* endoglucanase A encoding gene, *celCCA*. Gene **84:**39–46.
10. **Fierobe, H.-P., C. Gaudin, A. Belaich, M. Loutfi, E. Faure, C. Bagnara, D. Baty, and J.-P. Belaich.** 1991. Characterization of endoglucanase A from *Clostridium cellulolyticum*. J. Bacteriol. **173:**7956–7962.
11. **Foong, F., T. Hamamoto, O. Shoseyov, and R. H. Doi.** 1991. Nucleotide sequence and characteristics of endoglucanase gene *engB* from *Clostridium cellulovorans*. J. Gen. Microbiol. **137:** 1729–1736.
12. **Fukumori, F., T. Kudo, and K. Horikoshi.** 1987. Truncation analysis of an alkaline cellulase from an alkalophilic *Bacillus* species. FEMS Microbiol. Lett. **40:**311–314.
13. **Giallo, J., C. Gaudin, and J.-P. Belaich.** 1985. Metabolism and solubilization of cellulose by *Clostridium cellulolyticum* H10. Appl. Environ. Microbiol. **49:**1216–1221.
14. **Gilkes, N. R., B. Henrissat, D. G. Kilburn, R. C. Miller, Jr., and R. A. J. Warren.** 1991. Domains in microbial β-1,4-glycanases: sequence conservation, function, and enzyme families. Microbiol. Rev. **55:**303–315.
15. **Gilkes, N. R., R. A. J. Warren, R. C. Miller, and D. G. Kilburn.** 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis. J. Biol. Chem. **263:**10401–10407.
16. **Gräbnitz, F., M. Seiss, K. P. Rücknagel, and W. L. Staudenbauer.** 1991. Structure of the β-glucosidase gene *bglA* of *Clostridium thermocellum*. Eur. J. Biochem. **200:**301–309.
17. **Henrissat, B., M. Claeyssens, M. Tomme, P. Lemesle, and J.-P. Mornon.** 1989. Cellulase families revealed by hydrophobic cluster analysis. Gene **81:**83–95.
18. **Henrissat, B., and J.-P. Mornon.** 1990. Comparison of *Trichoderma* cellulases with other β-glycanases, p. 12–29. *In* C. P. Kubicek, T. E. Eveleigh, H. Esterbauer, W. Steiner, and E. M. Kubicek-Pranz (ed.), *Trichoderma* cellulases: biochemistry, genetics, physiology and applications. Royal Society of Chemistry, Cambridge.
19. **Joliff, G., P. Beguin, J. Millet, J.-P. Aubert, P. Alzari, M. Juy, and R. J. Poljak.** 1986. Crystallization and preliminary X-ray diffraction study of an endoglucanase from *Clostridium thermocellum*. J. Mol. Biol. **189:**249–250.
20. **Juy, M. (Institut Pasteur, Paris, France).** 1991. Personal communication.
21. **Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall.** 1951. Protein measurements with the Folin phenol reagent. J. Biol. Chem. **193:**265–275.
22. **Lüthi, E., D. R. Love, J. McAnulty, C. Wallace, P. A. Caughey, D. Saul, and P. L. Bergquist.** 1990. Cloning, sequence analysis, and expression of genes encoding xylan-degrading enzymes from the thermophile "*Caldocellum saccharolyticum*." Appl. Environ. Microbiol. **56:**1017–1024.
23. **Py, B., I. Bortoli-German, J. Haiech, M. Chippaux, and F. Barras.** 1990. Cellulase EGZ of *Erwinia chrysanthemi*: structural organization and importance of His 98 and Glu 133 residues for catalysis. Protein Eng. **4:**325–333.
24. **Quiocho, F. A.** 1986. Carbohydrate-binding proteins: tertiary structures and protein-sugar interactions. Annu. Rev. Biochem. **55:**287–315.
25. **Rouvinen, J., T. Bergfors, T. Teeri, J. K. C. Knowles, and T. A. Jones.** 1990. Three dimensional structure of cellobiohydrolase II from *Trichoderma reesei*. Science **249:**380–386.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. **Taylor, J. W., J. Ott, and F. Eckstein.** 1985. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. Nucleic Acids Res. **13:**8765–8785.
28. **Tokatlidis, K., S. Salamitou, P. Béguin, P. Dhurjati, and J.-P. Aubert.** 1991. Interaction of the duplicated segment carried by *Clostridium thermocellum* cellulases with cellulosome components. FEBS Lett. **291:**185–188.
29. **Tomme, P., S. Chauvaux, P. Beguin, J. Millet, J.-P. Aubert, and M. Claeyssens.** 1991. Identification of a histidyl residue in the active center of endoglucanase D from *Clostridium thermocellum*. J. Biol. Chem. **266:**10313–10318.
30. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40:**9–17.
31. **Withers, S. G., R. Antony, J. Warren, I. P. Street, K. Rupitz, J. P. Kempton, and R. Aebersold.** 1990. Unequivocal demonstration of the involvement of a glutamate residue as a nucleophile in the mechanism of a "retaining" glycosidase. J. Am. Chem. Soc. **112:**5887–5889.