Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of an endoglucanase from termite, *Nasutitermes takasagoensis*

Shahram Khademi,[a] Linda A. Guarino,[a] Hirofumi Watanabe,[b] Gaku Tokuda[b] and Edgar F. Meyer[a]*

[a]Biographics Laboratory, Texas A&M University, Department of Biochemistry and Biophysics, College Station, TX 77843-2128, USA, and [b]Institute of Insect and Animal Sciences, National Institute of Agrobiological Sciences, Tsukuba, Ibaraki 305-8634, Japan

Correspondence e-mail: e-meyer@tamu.edu

Received 8 October 2001
Accepted 5 February 2002

PDB References: termite endoglucanase, pH 2.5, 1ks8, r1ks8sf; pH 5.6, 1ksc, r1kscsf; pH 6.5, 1ksd, r1ksdsf.

Contrary to conventional wisdom, it has been shown recently that termites do not necessarily depend on symbiotic bacteria to process cellulose. They secrete their own cellulases, mainly endo-β-1,4-glucanase and β-1,4-glucosidase. Here, the first structure of an endogenous endoglucanase from the higher termite *Nasutitermes takasagoensis* (NtEgl) is reported at 1.40 Å resolution. NtEgl has the general folding of an $(\alpha/\alpha)_6$ barrel, which is a common folding pattern for glycosyl hydrolase family 9. Three-dimensional structural analysis shows that the conserved Glu412 is the catalytic acid/base residue and the conserved Asp54 or Asp57 is the base. The enzyme has a Ca^{2+}-binding site near its substrate-binding cleft. Comparison between the structure of the Ca^{2+}-free enzyme produced by reducing the pH of the soaked crystal from 5.6 (the pH of optimum enzyme activity) to 2.5 with that of the Ca^{2+}-bound enzyme did not show significant differences in the locations of the C^{α} atoms. The main differences are in the conformation of the residue side chains ligating the Ca^{2+} ion. The overall structure of NtEgl at pH 6.5 is similar to that at pH 5.6. The major change observed was in the conformation of the side chain of the catalytic acid/base Glu412, which rotates from a hydrophobic cavity to a relatively hydrophilic environment. This side-chain displacement may decrease the enzyme activity at higher pH.

1. Introduction

One of the main components of endogenous cellulase from termites and cockroaches is a group of enzymes called the endo-β-1,4-glucanases, which are mainly active against carboxymethyl cellulose (CMC). In 1997, Tokuda and coworkers purified, characterized and cloned the endo-β-1,4-glucanase from the wood-eating higher termite *Nasutitermes takasagoensis* (NtEG; Tokuda *et al.*, 1997; Watanabe *et al.*, 1997, 1998).

On the basis of sequence comparisons, the endoglucanase NtEgl from the termite *N. takasagoensis* belongs to glycosyl hydrolase family 9. This is the most taxonomically diverse family, containing cellulases from plants, bacteria, slime molds and termites, but not from fungi (Fig. 1). The majority of cellulases so far classified, including those from nematodes, have a common structure composed of a catalytic domain linked to a cellulose-binding domain (CBD). All known plant and termite cellulases do not have linkers and CBDs and with one exception (Brummell *et al.*, 1997) consist only of a single catalytic domain. The sequence of NtEgl shows a relatively high similarity to the catalytic domain sequence of bacterial, protozoan and plant members of family 9 (Fig. 1).

Here, we report the overexpression, purification and structure determination of NtEgl, presenting the first structure of an insect cellulase.

Among cellulose-digesting invertebrates, termites are the most significant because of their economic damage and the spectacular mounds many of them construct. In the USA alone, some species of termites cause an estimated $1 billion a year in damage (Brown, 2000). Some colonies devour an average of 450 kg of wood a year. Knowing the structure of cellulase from termites is important for designing potent and environmentally safe inhibitors to control the damage they cause.

2. Materials and methods

2.1. Expression and purification of termite endoglucanase

A cDNA clone obtained from Noda (Tokuda *et al.*, 1999) was inserted into a pET3a vector. This clone was named pET-Nts. BL21(DE3) cells containing pXKEJ, which expresses chaperone proteins (dnaK, dnaJ and GroEL) under arabinose control, were transformed with pET-Nts. The expression of the chaperone proteins was induced by the addition of arabinose. Expression of NtEgl was then induced by the addition of 0.1 m*M* IPTG. Cells were harvested using centrifugation and lysed by sonicaton. The resulting supernatant was precipitated with 50% saturated ammonium sulfate. The precipitated protein was collected by centrifugation and resuspended in 50 m*M* Tris pH 8.0, 1 *M* ammonium sulfate. The protein solution was loaded onto a 10 ml phenyl Sepharose column previously equilibrated in the same buffer. The column was eluted with a linear ammonium sulfate gradient from 1.0 to 0.0 *M*. Peak fractions containing NtEgl were dialyzed against Tris pH 8.0 plus 50 m*M* NaCl and loaded onto a MonoQ column equilibrated in the same buffer. The column was eluted with a linear NaCl gradient from 50 to 550 m*M*. Peak fractions of NtEgl were pooled and concentrated for crystallization.

2.2. Crystallization and data collection

Crystallization trials were carried out using the hanging-drop vapor-diffusion technique. Initial screening of crystallization conditions using sparse-matrix sampling (Jancarik & Kim, 1991) yielded plate-shaped initial crystals, which were very thin. After optimization of crystallization conditions, single crystals of NtEgl appeared in a drop containing 8 mg ml^{-1} NtEgl, 0.1 *M* sodium citrate pH 4.5 and 0.8 *M* ammonium sulfate. After a few weeks, the crystal size was 0.2 × 0.2 × 0.5 mm (Table 1). At this point, the crystal was transferred into a harvesting solution (1.5 *M* ammonium sulfate and 0.1 *M* citrate buffer). By varying the pH of the harvesting solution (pH 2.5, 5.6 or 6.5), we could determine the structure of the protein at various pH values. In order to collect diffraction data at low temperature, the crystals were transferred into a cryoprotectant solution (1.5 *M* ammonium sulfate, 10–25% glycerol plus citrate buffer).

Data collection was performed at 120 K (Oxford Cryosystems) using a Nonius/MacScience DIP2030-k image-plate detector and a Rigaku rotating-anode X-ray source with Cu $K\alpha$ radiation focused by an MSC/Osmic mirror system. The data were evaluated with the program *DENZO* and scaled using *SCALEPACK* (Otwinowski & Minor, 1997) (Table 1).

2.3. Structure determination and refinement

Molecular-replacement calculations for the native NtEgl structure were performed with the program *CNS* (Brünger *et al.*, 1998) using the structure

```
T. fusca          EPAFNYAEALQKSMFFYEAQRSGKLPENNRVSWRGDSGLNDGA-----DVGLD--LTGGWYDAG  57
N. takasagoensis  -MAYDYKQVLRDSLLFYEAQRSGRLPADQKVTWRKDSALNDQG-----DQGQD--LTGGYFDAG  56
D. discoideum     ---TDYCSLLENALMFYKMNRAGRLP-DNDIPWRGNSALNDASPNSAKDANGDGNLSGGYFDAG  60
P. persica        ---QDYSDALEKSILFFEGQRSGKLPANQRATWRANSGLSDGS-----SYHVD--LVGGYYDAG  54
                     *   *     *    * * **      **  * * *          *  * **  ***

T. fusca          DHVKFGFPMAFTATMLAWGAIESPEGYIRSGQMPYLKDNLRWVNDYFIKAHPS-PNVLYVQVGD 120
N. takasagoensis  DFVKFGFPMAYTATVLAWGLIDFEAGYSSAGALDDGRKAVKWATDYFIKAHTS-QNEFYGQVGQ 119
D. discoideum     DGVKFGLPMAYSMTMLGWSFIEYESNIAQCGLTSLYLDTIKYGTDWLIAAHTA-DNEFAGQVGD 123
P. persica        DNVKFGLPMAFTTTLLAWSVIEFGDSMHN--QIENAKDAIRWSTDYLLKAATSTPGALYVQVAD 116
                  * **** ***   * * *  *                        *    *          **

T. fusca          GDADHKWWGPAEVMPMERPSFKVDPSCPGSDVAAETAAAMAASSIVFADDDPAYAATLVQHAKQ 184
N. takasagoensis  GDADHAFWGRPEDMTMARPAYKIDTSRPGSDLAGETAAALAAASIVFRNVDGTYSNNLLTHARQ 183
D. discoideum     GNVDHSWWGPPEDMTMARPTYMLTTEAPGTEIAMEAASALAAASIAFKSSNPTYAATCLAHAKT 187
P. persica        PNADHQCWERPEDMDTPRNVYKVSTQNPGSDVAAETAAALAAASIVFKDSDPSYSGKLLHTAMK 180
                     **  *   * *   *          **   * * * * ** ** *      *       *

T. fusca          LYTFADTYRGVYSDCV-P-AGAFYNSWSGYQDELVWGAYWLYKATGDDSYLAKAEYEYDFLSTE 246
N. takasagoensis  LFDFANNYRGKYSDSI-T-DARNFYASADYRDELVWAAAWLYRATNDNTYLNTAESLYDEFG-- 243
D. discoideum     LHNFGYTYRGVYSDSI-TNAQAFYNSWSGYKDDLVWGSIWLYKATQDSDYLTKAVADYASGG-- 248
P. persica        VFDFADRYRGSYSDSIGSVVCPFYCSYSGYHDELLWGASWIHRASQNSSYLAYIKSNGHILG-- 242
                     *   *** ***              * * * *   *   *     **

T. fusca          QQTDLRSYRWTIAWDDKSYGTYVLLAKETGKQ------KYIDDANRWLDYWTVGVNGQRVPYSP 304
N. takasagoensis  ----LQNWGGGLNWDSKVSGVQVLLAKLTNKQ------AYKDTVQSYVNYL---INNQQK--TP 292
D. discoideum     --VGGMAQGNSHDWDNKAPGCCLLLSKLVPTTS-----TYKTDFEGWLNYW---LPGGGVTYTP 302
P. persica        ----ADDDGFSFSWDDKRPGTKVLLSKNFLEKNNEEFQLYKAHSDNYICSLLPGTSNFQAQYTP 302
                            ** *  *   ** *          *                        *

T. fusca          GGMAVLDTWGALRYAANTAFVALVYAKVIDDP----------VRKQRYHDFAVRQINYALGDNP 358
N. takasagoensis  KGLLYIDMWGTLRHAANAAFIMLEAAELG-------------LSASSYRQFAQTQIDYALGDGG 343
D. discoideum     GGLAWIRQWGPARYAATAAFLGSLAG-----------------TEKGTDFTQKQVDYLIGNNP 348
P. persica        GGLLYKASESNLQYVTSTTLLLLTYAKYLRTNGGVATCGSSKVTAETLISEAKKQVDYILGNNP 366
                  *                                                 *  *  *

T. fusca          RNSSYVVGFGNNPPRNPHHRT-------AHGSWTDS------IASPAENRHVLYGALVGGPGSP 409
N. takasagoensis  R--SFVCGFGSNPPTRPHHRSSSCPPAPATCDWN-T------FNSPDPNYHVLSGALVGGP-DQ 397
D. discoideum     NQQSFVVGMGPNYPINPHHRA-------AHHSTTND------INNPVNNLYLLKGALVGGP-GS 398
P. persica        AKISYMVGFGKKYPLHIHHRGSSLPSVHEHPERISCNNGFQYLNSGSPNPNVLVGAIVGGP-DS 429
                     *   * *   *   ***                          *   * ** ****

T. fusca          NDAYTDDRQDYVANEVATDYNAGFSSALAMLVEE------                        443
N. takasagoensis  NDNYVDDRSDYVHNEVATDYNAGFQSALAALVALGY----                        433
D. discoideum     NDEYTDDRTDYISNEVATDYNAGFVGALASLVNPSSTSVP                        438
P. persica        KDSFSDDRNNYQQSEPATYINAPIVGALAFFSANTNPN--                        467
                   *   ***  *   * **  **    ***
```

Figure 1
Multiple sequence alignment among catalytic domains of four members of family 9. The conserved residues are marked by a star. The putative catalytic residues are in bold. The sequences are as follows: *T. fusca*, *Thermomonospora fusca* (P26221); *N. takasagoensis*, *Nasutitermes takasagoensis* (M64644); *D. discoideum*, *Dictyostelium discoideum* (P22699); *P. persica*, *Prunus persica* (peach tree) (P94114).

Table 1
Structure-determination summary of NtEgl at three pH values: 2.5, 5.6 and 6.5.

Values in parentheses are for the highest resolution shell.

	pH 2.5	pH 5.6	pH 6.5
Space group	*C*2	*C*2	*C*2
Unit-cell parameters			
a (Å)	89.71	89.11	89.25
b (Å)	83.83	83.94	83.92
c (Å)	60.68	60.84	60.88
α (°)	90.00	90.00	90.00
β (°)	95.29	95.60	95.47
γ (°)	90.00	90.00	90.00
Resolution (Å)	30.00–1.40 (1.45–1.40)	30.00–1.55 (1.61–1.55)	30.00–1.60 (1.66–1.60)
Completeness (%)	96.9 (89.4)	93.3 (83.8)	97.0 (92.4)
Total No. of reflections	775989	1124141	654717
Unique reflections	85117	60004	57195
$I/\sigma(I)$	22.0 (2.7)	24.7 (3.6)	19.6 (3.6)
$R_{\rm merge}$† (%)	6.2 (54.3)	7.9 (56.2)	8.9 (50.9)
Protein atoms	3382	3382	3382
Calcium ions	0	1	1
Water molecules	729	567	705
R value‡ (%)	17.0	17.8	17.8
$R_{\rm free}$ (%)	18.5	20.4	20.1
R.m.s.d. bonds (Å)	0.004	0.005	0.005
R.m.s.d. bond angles (°)	1.24	1.21	1.21
R.m.s.d. *B* factors (Å^2)			
Main chain	0.87	0.95	0.92
Side chain	1.81	1.72	1.78

† $R_{\rm merge}$ (%) = $100\sum_h \sum_i |I(h)_i - \langle I(h)\rangle|/\sum_h \sum_i \langle I(H)_i\rangle$, where I is the observed intensity and $\langle I\rangle$ is the average intensity of multiple observations of symmetry-related reflections. ‡ R value = $\sum \big||F_o| - |F_c|\big|/\sum |F_o|$.

of the catalytic domain of the endoglucanase from *Thermomonospora fusca* (E4; Sakon *et al.*, 1997), a member of family 9, as the search model. Native NtEgl diffraction data in the resolution range 15–4.0 Å were used in a direct rotation-search method (DeLano & Brünger, 1995; Brünger, 1997). The direct rotation function used reflections with $F_{\rm obs} > 2\sigma$. After rotation and fast translation-search steps (each step accompanied by Patterson correlation refinement), the ten highest translation-search peaks were determined.

The refinement of atomic positions was performed with the maximum-likelihood method based on amplitudes (Pannu & Read, 1996). Only 90% of the data in the resolution range 30–1.55 Å was used for the refinement; the remaining 10% was reserved as the test set for structure validation. During all the refinements the same test set was used for the calculation of $R_{\rm free}$. The best solution after rigid-body refinement had an R value of 47.4% ($R_{\rm free}$ = 47.0%). Simulated annealing using torsion-angle molecular dynamics employed a protocol used by Adams *et al.* (1997), with initial minimization followed by torsion-angle molecular dynamics. This first round of simulated-annealing refinement reduced the R value to 36.9% (and $R_{\rm free}$ to 40.6%). At this point, we started to replace amino acids of the model (E4) with corresponding residues from NtEgl based on the result of multiple sequence alignment (Fig. 1) and to rebuild the model manually. The refinement proceeded with simulated annealing (Brünger *et al.*, 1987), correcting errors introduced by the manual rebuilding and optimizing the fit of the retraced region. After many steps of simulated-annealing refinement using *CNS* and manual rebuilding using *PRONTO* (Laczkowski *et al.*, 1996) followed by refinement for individual B factors, the R value was reduced to 23.7% ($R_{\rm free}$ to 26.1%). The addition of calcium ion and water molecules to the model and final refinement reduced the R factor to 18.7% ($R_{\rm free}$ to 21.3%) (Table 1).

The structures of NtEgl at pH 2.5 and 6.5 were determined using the data collected from NtEgl soaked at pH 2.5 and 6.5, respectively. The final model of the native NtEgl (pH 5.6) was used as an initial model for the refinement. After some corrections in the side-chain and water positions, the final models were likewise obtained for pH 2.5 and 6.5 (Table 1).

3. Results and discussion

3.1. Crystallography

A crystal soaked in 0.1 *M* citrate buffer pH 5.6 was used for data collection. The crystal structure of NtEgl was determined at 1.55 Å resolution by the molecular-replacement method using the three-dimensional structure of the endoglucanase from *T. fusca* (E4) as a search model. The crystal structure consists of one molecule per asymmetric unit and four asymmetric units per cell in the space group *C*2, with unit-cell parameters a = 89.1, b = 83.9, c = 60.8 Å, $\alpha = \gamma = 90.0$, $\beta = 95.6^\circ$ (Table 1). The refined molecular structure contains 433 residues, 567 water molecules and one calcium ion. The final structure has a crystallographic R factor of 17.8% and an $R_{\rm free}$ of 20.4% against all reflections in the resolution range 30–1.55 Å (93.3% completeness; 83.8% completeness in the resolution range 1.61–1.55 Å).

In the Ramachandran plot (Ramakrishnan & Ramachandran, 1965) generated using the program *PROCHECK* (Laskowski *et al.*, 1993), 89.4% of the residues were found to be in the most favorable regions and 10% in favorable regions; only two residues, Ala55 and Ser323, were found in generously allowed regions. These two residues, located within good electron density and having overall B factors of 10.4 and 24.4 Å^2, respectively, are part of the active site and therefore their main-chain angles may quite possibly change upon ligand binding.

The second structure of NtEgl was determined using the data set from a crystal soaked at pH 6.5 (Table 1). The final model contains all protein residues (433 amino acids), 705 water molecules and one calcium ion. The R factor is 17.8% ($R_{\rm free}$ = 20.1%) for all reflections in the resolution range 30.0–1.60 Å (97.0% completeness; 92.4% completeness in the resolution range 1.66–1.60 Å). The root-mean-square (r.m.s.) difference between C^α positions of the structures of NtEgl at pH 5.6 and 6.5 is 0.11 Å.

A data set from a crystal soaked at pH 2.5 was used to determine a third structure of NtEgl (Table 1). The final model consists of 433 residues, 729 water molecules and no calcium ion. The structure was refined to a final R factor of 17.0% ($R_{\rm free}$ = 18.5%), including all reflections in the resolution range 30.0–1.4 Å (96.9% completeness; 89.4% completeness

in the resolution range 1.45–1.40 Å). The r.m.s. fit between the C^{α} atoms of NtEgl at pH 5.6 and 2.5 is 0.17 Å.

3.2. Overall structural description

Termite cellulase NtEgl is a monomeric protein with approximate dimensions of 65 × 60 × 45 Å. NtEgl has the general folding motif of a left-handed $(\alpha/\alpha)_6$ barrel, which is a common folding pattern of glycosyl hydrolase enzymes in families 8, 9 and 15 (Juy *et al.*, 1992; Sakon *et al.*, 1997; Parseigla *et al.*, 1998; Alzari *et al.*, 1996). The 12 helices of the $(\alpha/\alpha)_6$ barrel (*A*1–*A*12) have various lengths from 12 to 20 amino acids and show an alternating connecting pattern between outer and inner helices (Fig. 2). The inner helices *A*2, *A*4, *A*6, *A*8, *A*10 and *A*12 run parallel, forming a central barrel, with its central volume largely occupied by aliphatic (Val71, Ala157, Ala161, Val260, Val262, Leu263, Ile313 and Ala426) and aromatic (Tyr67, Trp74, Phe63, Trp217, Trp221 and Phe312) side chains. No water molecules are found in the central volume of the inner helices.

On the C-terminal side of the inner helices, the barrel helices are connected by relatively short loops of 3–11 residues in length. Six connections on the N-terminal side of the inner helices vary from 14 to 72 residues in length and form the cleft of the substrate-binding site. The loops, comprising about half of the total amino acids of the protein, have relatively higher temperature factors, showing that they are flexible; on ligand binding, some of them may change their location. In addition to the structural elements of the $(\alpha/\alpha)_6$ barrel, there are eight short α-helices and three short β-strands.

The enzyme contains three cysteine residues, Cys362, Cys365 and Cys372, which are not conserved in family 9. The two cysteines (Cys365 and Cys372) are involved in a disulfide bridge to stabilize the folding of the long loop 341–414. The only *cis*-proline of the enzyme (Pro369) is in this loop. Most ionic interactions between charged groups are partially or totally hydrated at the molecular surface, with the single exception of a salt bridge between the internal residues Glu154 and Lys256. This salt bridge occurs within the central hydrophobic core, adjacent to the catalytic center, and may be important for stability of the active-site architecture.

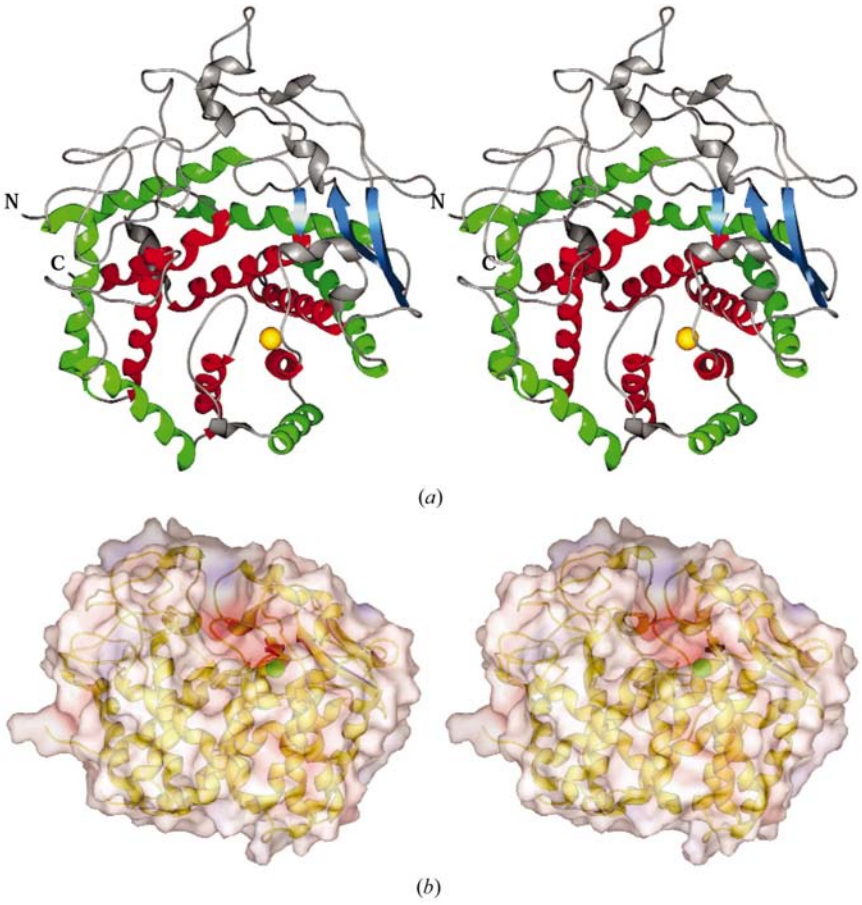

Figure 2
The structure of NtEgl. (*a*) Ribbon representation of NtEgl showing top view of 12 helices. Outer helices are in green, inner helices are red, β-strands are blue and the Ca^{2+} ion is brown. (*b*) Surface potential of NtEgl showing the binding cleft. The ribbon representation of NtEgl is visible through the transparent molecular surface, with positive charges in blue and negative charges in red; the Ca^{2+} ion is green.

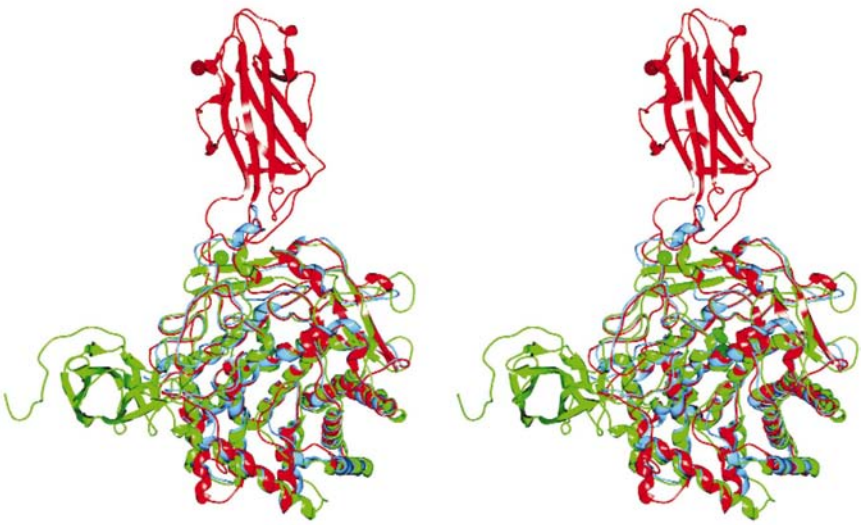

Figure 3
The superposition of the three-dimensional structure of NtEgl (in blue) on the three-dimensional structure of CelD (in green) and E4 (in red). CelD and E4 have a cellulose-binding domain at their N-terminal and C-terminal, respectively.

NtEgl has a calcium ion near the non-reducing end of the binding cleft. Based on the sequence and structural analysis (Fig. 2), this calcium site is likely to be present in the catalytic domain of all enzymes in family 9. The calcium ion is ligated by Asp210 O, Asp213 $O^{\delta1}$, Asp213 $O^{\delta2}$, Glu214 $O^{\varepsilon1}$, Glu214 $O^{\varepsilon2}$, Asp254 O and two water molecules.

3.3. The ligand-binding site cleft

Despite numerous attempts to form a stable enzyme–inhibitor complex, preceded by inhibition assays, neither we nor our collaborator Professor A. Vasella (personal communication) have been able to identify any effective inhibitors. This negative result actually emphasizes the difficulty of such research and the need to press ahead in our search for inhibitors, whether existing or novel. However, based on sequence analysis and comparison of the three-dimensional structure of NtEgl with bacterial endoglucanases from *T. fusca* (Sakon *et al.*, 1997) and *Clostridium thermocellum* (Juy *et al.*, 1992), we can speculate with confidence about the residues at the binding site. In order to compare the structure of NtEgl with the structure of bacterial cellulase from *T. fusca* (E4) and cellulase from *C. thermocellum* (CelD), the three-dimensional structure of NtEgl was superimposed on the three-dimensional structures of CelD and E4 by matching the C^{α} positions of the core region of the $(\alpha/\alpha)_6$ barrel (Fig. 3). The r.m.s. fit between the C^{α} atoms of the core region of NtEgl and CelD (214 atoms) is 2.3 Å and between that of NtEgl and E4 (215 atoms) is 1.6 Å. The core region of the $(\alpha/\alpha)_6$ barrel is similar, whereas significant differences are observed in the loop regions.

The binding cleft of NtEgl was identified based on sequence homology and three-dimensional structural comparison with CelD and E4. On the N-terminal side of the inner barrel, a cleft runs roughly from α-helix 1 to α-helix 8 and forms the site for substrate binding. One of the rims of the active-site cleft consists of two loops 342–414 and 287–302. The long loop 342–414 contains two short α-helices (Fig. 2). The loops 107–148 and 191–211 form a part of the other rim of the active-site cleft. The dimension of the cleft was estimated by calculating the solvent-accessible surface, giving a length of ∼30 Å and a depth of ∼9.0 Å. The average width of the cleft is ∼4 Å. A 30 Å long cleft is likely to bind around five glucopyranose units, which can be numbered from the subsite at the non-reducing end of the cellulose chain (subsite −3) to the subsite at the reducing end of the cellulose (subsite +2); cleavage takes place between subsites −1 and +1 (Fig. 4).

At the non-reducing end of the cleft, the conserved Trp253 side chain interacts with the glucopyranose ring in subsite −3. Invariant residue Asp254 may interact with a hydroxyl group in the glucopyranose rings at either the −3 or −2 subsite. The mostly conserved Trp301 would flank subsite −2. In subsite −1, the conserved residues Phe205 and Tyr417 may stabilize the hydrophobic glucopyranose ring. The stabilization of the substrate at subsite −1 occurs through hydrogen bonding to conserved Asp54. The conserved Tyr408 side chain is probably involved in a stacking interaction with a glycosyl moiety in the subsite +1. The invariant residues Asp57, His124, His359, Arg361 and acid/base Glu412 may help in stabilization of the substrate through hydrogen bonding in the subsite +1. The conserved Trp127 probably interacts with the glucopyranose ring at subsite +2.

Glycosyl hydrolases of family 9 cleave the sugar chain with inversion of configuration at the anomeric C atom. The mechanism requires an acid to protonate the glycosidic O

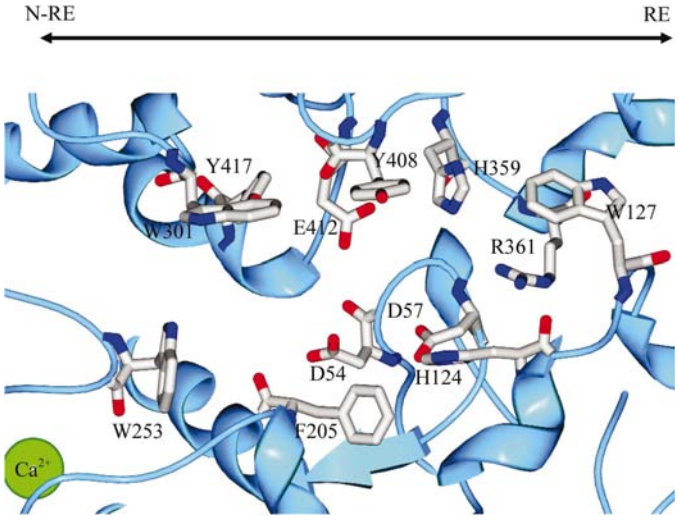

Figure 4
The binding cleft of NtEgl. The arrow shows the reducing end (RE) and non-reducing end (NRE) of the cleft.

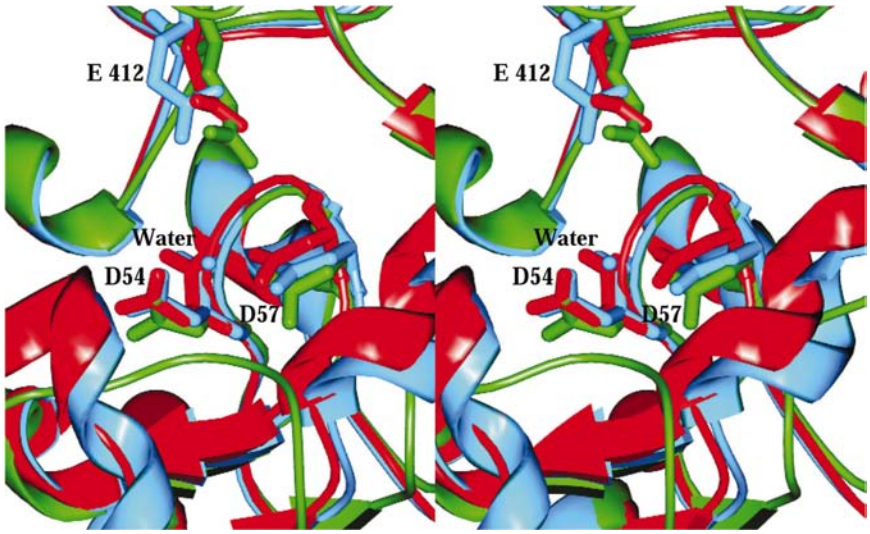

Figure 5
The superposition of residues Asp54, Asp57 and Glu412 from NtEgl (in blue) on 55, 58 and 424 from CelD (in green) and 198, 201 and 555 from E4 (in red), respectively.

atom and a base to extract a proton from a nucleophilic water molecule that subsequently attacks the anomeric C atom (Koshland, 1953). Sequence analysis and comparison of the three-dimensional structure of NtEgl with the structure of E4 and CelD (Fig. 5) show that the carboxyl group of the conserved Glu412 (corresponding to 555 and 424 in E4 and CelD) is the acid that protonates the leaving group and that Asp54 or Asp57 (corresponding to 198 and 201 in E4, and to 55 and 58 in CelD) may be the base that extracts a proton from a nucleophilic water. The distances between the O atom of nucleophilic water and Asp54 $O^{\delta2}$ and Asp57 $O^{\delta2}$ are 2.68 and 2.73 Å, respectively. This proposed mechanism agrees well with the mutagenesis studies of CelD (Chauvaux *et al.*, 1992). Mutation of either the Asp54, Asp57 or Glu412 equivalents in CelD reduces the activity to less than 1%.

3.4. The effect of pH changes on conformation of NtEgl

NtEgl was crystallized at pH 4.5. In order to obtain the structure of a ligand–enzyme complex, crystals were soaked in a solution containing cellopentose at pH 2.5. At this pH, the activity is significantly lower and therefore the substrate was expected to be stable at the active site. Unfortunately, no clear electron density was found for the ligand. Later, the effect of pH on the structure of the enzyme was studied by collecting three diffraction data sets at pH 2.5, 5.6 (the pH optimum) and 6.5. Structures corresponding to these three data sets were appropriately refined and compared.

The conformations of the backbone at three different pH values are quite similar. However, at pH 2.5 the enzyme had lost its calcium ion. The largest change was observed for the side chains of the residues involved in Ca^{2+} binding. At pH 5.6, the calcium ion is ligated by Asp210 O, Asp213 $O^{\delta1}$, Asp213 $O^{\delta2}$, Glu214 $O^{\varepsilon1}$, Glu214 $O^{\varepsilon2}$, Asp254 O and two water molecules (Fig. 6*a*). The coordination of the Ca^{2+} ion seems to be a distorted octahedral arrangement with two bidentate ligands, Asp213 and Glu214 (Fig. 6*a*). This coordination is similar to the Ca^{2+}-binding site in E4 (Sakon *et al.*, 1997) and the Ca^{2+}-binding site *B* in CelD (Chauvaux *et al.*, 1995). The main-chain carbonyl groups of Asp210 and Asp254 form the two vertices of the octa-

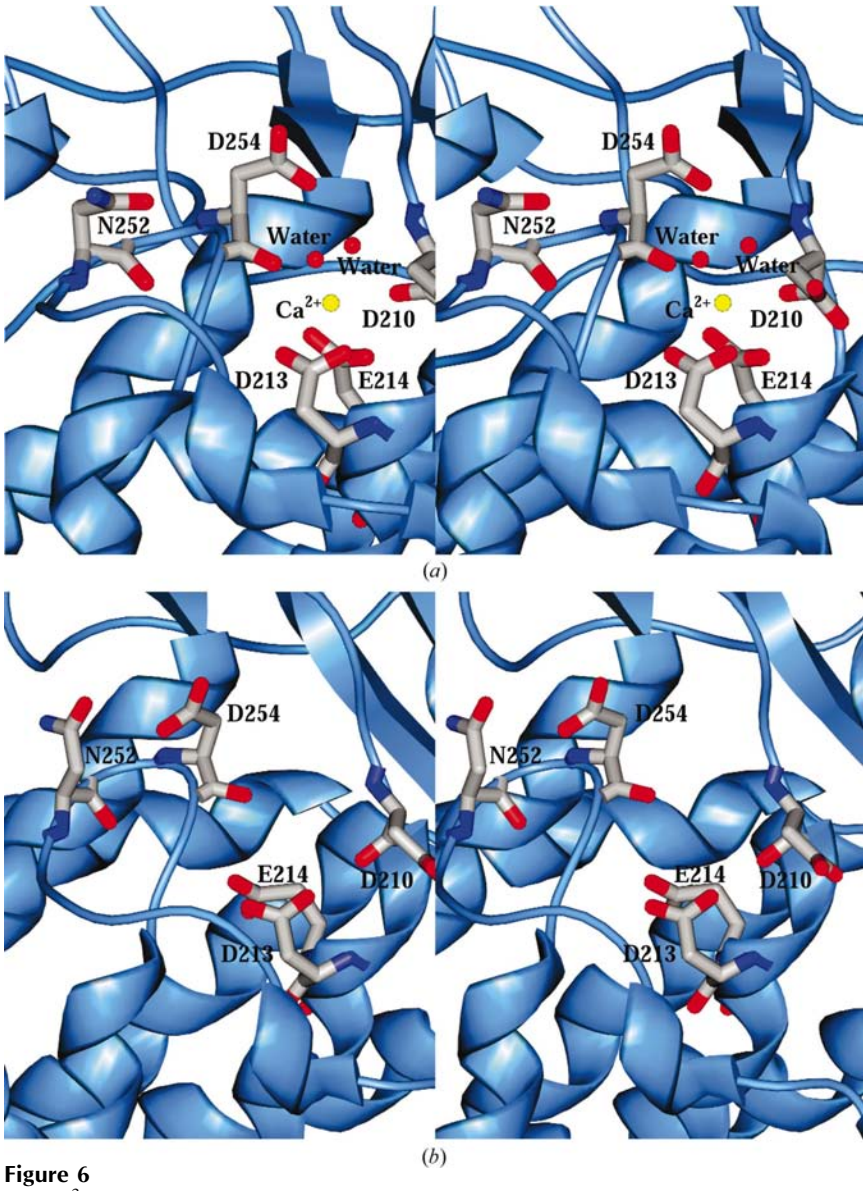

Figure 6
The Ca^{2+}-binding site of NtEgl at (*a*) pH 5.6 and (*b*) pH 2.5.

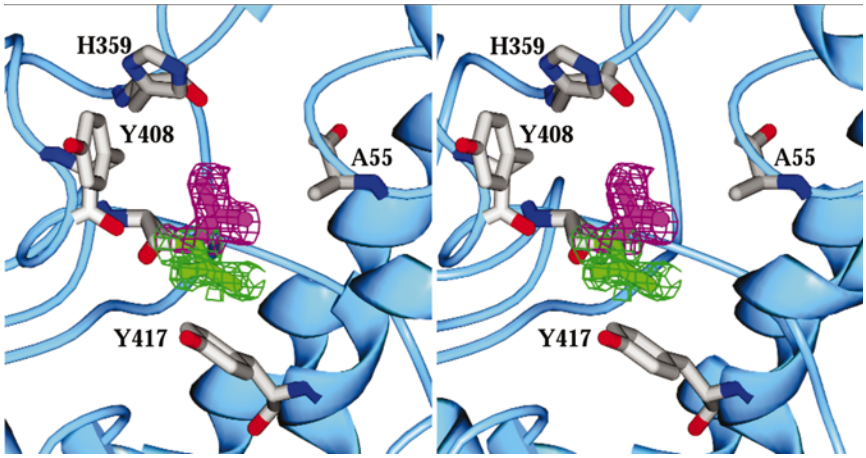

Figure 7
Stereo representation of environment of the catalytic acid/base residue Glu412 at pH 5.6 and 6.5. The carboxylate group of Glu412 at pH 5.6 is in purple and at pH 6.5 is in green. The simulated-annealing omit map density for Glu412 (at pH 5.6 is in purple and at pH 6.5 is in green) is contoured at 0.75σ.

hedron. At pH 2.5, the torsion angle between atoms C^{α} and C^{β} of the conserved Asp54 changed by $\sim$135° (from $-$164° at pH 5.6 to 61° at pH 2.5). This 1.7 Å displacement of Asp254 C^{γ} pushes the side chain of Asn252, causing a 141° rotation in the torsion angle between atoms C^{β} and C^{γ} of Asp252 (Fig. 6*b*). These changes in torsion angles place the side chains of Asp254 and Asn252 in new positions and result in a hydrogen bond between Asp254 $O^{\delta 2}$ and Asn252 $O^{\delta 1}$. The torsion angle between atoms C^{β} and C^{γ} of Glu214 also changed by $\sim$132° (from 74° at pH 5.6 to $-$58° at pH 2.5).

In CelD (an endoglucanase from *C. thermocellum*), there are three Ca^{2+}-binding sites (sites *A*, *B* and *C*) and one Zn^{2+}-binding site. The Ca^{2+}-binding site *B* in CelD corresponds to the Ca^{2+}-binding site in NtEgl. Chauvaux and coworkers performed extensive kinetic studies on the effect of each Ca^{2+}-binding site on the stability and kinetic properties of the enzyme (Chauvaux *et al.*, 1995). They showed that Ca^{2+}-binding site *C* is important for stability of the enzyme and that Ca^{2+}-binding sites *A* and *B* do not have any effect on either the structural stability or kinetic properties of CelD. Here, the results of structural studies on NtEgl show that Ca^{2+} binding does not have any significant effect on the backbone three-dimensional structure of NtEgl. The role of this calcium-binding site on the kinetics or structures of cellulases from family 9 is not yet clear.

At pH 6.5, the main conformational change was observed for the conserved acid/base Glu412, where the torsion angle between atoms C^{β} and C^{γ} changed by $\sim$118° (from 67 to $-$175°) (Fig. 7). This rotation places the carboxyl part of the Glu412 in a different position. The displacement is 2.4 Å, calculated from the position of the C^{δ} atom. The temperature factor for the carboxylate group of Glu412 is relatively high (32.9 $Å^2$ for pH 5.6 and 31.9 $Å^2$ for pH 6.5 relative to the average temperature factor of $\sim$18.5 $Å^2$ for all residues in NtEgl); therefore, the carboxylate group is relatively flexible.

At pH 5.6, the carboxyl group of Glu412 makes no hydrogen bonds with other residues; Glu412 $O^{\varepsilon 2}$ is located 3.8 Å from His359 $C^{\delta 2}$ and Glu412 $O^{\varepsilon 1}$ is 3.3 Å from Ala55 C^{β}. This somewhat hydrophobic environment could raise the pK_a value of Glu412. At pH 6.5, owing to the displacement of the carboxyl group of Glu412, the Glu412 $O^{\varepsilon 2}$ atom is located 2.8 Å from Tyr417 OH and Glu412 $O^{\varepsilon 1}$ is 3.1 Å from Tyr408 O. It seems that raising the pH from 5.6 to 6.5 deprotonates the carboxyl group of Asp412. Therefore, the resulting negatively charged carboxyl group cannot stay in an hydrophobic environment and the carboxyl group rotates toward a relatively hydrophilic environment. This deprotonation and side-chain shift may result in a reduction in enzyme activity at higher pH.

We thank Professor Michael Slaytor and Dr Andrew Scrivener for their assistance with related cockroach cellulase studies. We thank Dr Stanley M. Swanson for technical assistance. Financial support was provided by the Robert A. Welch Foundation (A-328) and the Texas Advanced Technology Program (LG and EM).

References

Adams, P. D., Pannu, N. S., Read, R. J. & Brünger, A. T. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 5018–5023.
Alzari, P. M., Souchon, H. & Dominguez, R. (1996). *Structure*, **4**, 265–275.
Brown, P. L. (2000). *The New York Times*, **51388**, 1.
Brummell, D. A., Catala, C., Lashbrook, C. C. & Bennett, A. B. (1997). *Proc. Natl Acad. Sci. USA*, **94**, 4794–4799.
Brünger, A. T. (1997). *Methods Enzymol.* **277**, 558–580.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–21.
Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
Chauvaux, S., Beguin, P. & Aubert, J. P. (1992). *J. Biol. Chem.* **267**, 4472–4478.
Chauvaux, S., Souchon, H., Alzari, P. M., Chariot, P. & Beguin, P. (1995). *J. Biol. Chem.* **270**, 9757–9762.
DeLano, W. L. & Brünger, A. T. (1995). *Acta Cryst.* D**51**, 740–748.
Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssena, M., Beguin, P. & Aubert, J. P. (1992). *Nature (London)*, **357**, 89–91.
Koshland, D. E. J. (1953). *Biol. Rev.* **28**, 416–436.
Laczkowski, A., Meyer, E. F., Swanson, S. M. & Worick, R. K. (1996). *J. Iran Univ. Med. Sci.* **3**, 31–41.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.
Parseigla, G., Juy, M., Reverbel-Leroy, C., Tardif, C., Balaich, J. P., Driguez, H. & Haser, R. (1998). *EMBO J.* **17**, 5551–5562.
Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
Sakon, J., Irwin, D., Wilson, D. B. & Karplus, A. (1997). *Nature Struct. Biol.* **4**, 810–818.
Tokuda, G., Lo, N., Watanabe, H., Slaytor, M., Matsumoto, T. & Noda, H. (1999). *Biochem. Biophys. Acta*, **1447**, 146–159.
Tokuda, G., Watanabe, H., Matsumoto, T. & Noda, H. (1997). *Zool. Sci.* **14**, 83–93.
Watanabe, H., Nakamura, M., Tokuda, G., Yamaoka, I., Scrivener, A. M. & Noda, H. (1997). *Insect. Biochem. Mol. Biol.* **27**, 305–313.
Watanabe, H. Noda, H., Tokuda, G. & Lo, N. (1998). *Nature (London)*, **394**, 330–331.