Endoglucanase Sensitivity for Substituents in Methyl Cellulose Hydrolysis Studied Using MALDI-TOFMS for Oligosaccharide Analysis and Structural Analysis of Enzyme Active Sites

U. Schagerlöf,[†] H. Schagerlöf,*[,‡] D. Momcilovic,[§] G. Brinkmalm,[||] and F. Tjerneld[‡]

Departments of Molecular Biophysics, Biochemistry, and Technical Analytical Chemistry, Lund University, P. O. Box 124, SE-221 00 Lund, Sweden, and AstraZeneca R&D Mölndal, SE-431 83 Mölndal, Sweden

Received January 31, 2007; Revised Manuscript Received April 19, 2007

The properties of modified cellulose polymers, such as methylcellulose, are significantly influenced by the distribution of substituents along the polymer backbone. This distribution is difficult to determine due to the lack of suitable analytical methods. One approach is to use cellulose-degrading enzymes to gain information from the capability of the enzymes to cleave the bonds between glucose units. Endoglucanases are cellulase enzymes that can break internal glycosidic linkages and degrade low substituted regions of modified cellulose where the substituents do not interfere with the enzyme active site. In this work methyl cellulose was degraded using five endoglucanases from glycosyl hydrolase families 5 and 7 from three different species. The products were analyzed with reducing end analysis, chromatography (SEC-MALS-RI), and MALDI-TOFMS. The results were correlated with available determined enzyme structures and using structural alignment for unknown enzyme structures. This was performed in order to elucidate the relationship between active site structures and sensitivity for substituents on derivatized cellulose. The evaluation of endoglucanase hydrolysis of methyl cellulose showed that differences in sensitivity could be related to differences in steric hindrance of substituents in the active site, which could explain differences within family 5 and 7 enzymes, as well as the generally higher substituent tolerance for family 5 enzymes. This information is important for use of endoglucanases as tools for characterization of substituent distribution. The results are also valuable since soluble cellulose derivatives are generally used as substrates during enzyme characterization and in endoglucanase activity assays.

Introduction

Chemical modification of cellulose introduces new characteristics, such as improved solubility in various solvents. This opens up a broad range of applications for modified cellulose, for example in building, pharmaceutical, and food industries.[1] To achieve the desired properties, the hydroxyl groups along the polymer chain are substituted with various groups like carboxymethyl, methyl, hydroxypropyl, or a mixture of different groups.[2] The properties of the modified cellulose will depend on the polymer chain length, the nature of the substituent, the degree of substitution (DS), and the distribution of substituents along the backbone.[3,4] The distribution of substituents influences thickening, gelation, film formation, and flocculation, making a thorough characterization of substituent distribution important.[5] This can be analyzed on two levels, within the anhydroglucose unit (AGU) where OH groups at positions 2, 3, and 6 can be substituted or along the cellulose backbone where substitution can be more or less heterogeneous. The monomeric distribution can be analyzed using for example carbon nuclear magnetic resonance spectroscopy (^{13}C NMR) or gas chromatography in combination with flame ionization detector and mass spectrometry (GC-FID-MS), whereas characterization of distribution along the polymer is more complex.[6] The use of cellulose degrading endoglucanases (i.e., enzymes that selectively hydrolyze the (1→4)-β-D-glycosidic linkages within the cellulose chain) is one approach for investigation of the substituent distribution heterogeneity.[7−10] Due to steric interference from the substituents, the enzymes cannot degrade more heavily substituted regions but only regions with sufficiently low DS. By analysis of the produced fragments after hydrolysis, the substituent distribution along the polymer chain may be investigated. A polymer with even distribution is expected to be less degraded than a polymer with blockwise distribution if both have the same average DS.[11] If the polymer is highly substituted, the enzyme may not be able to degrade it at all, and in this case more detailed information concerning the distribution pattern cannot be collected.[12] The intact parts of the polymer may be further analyzed using fractionation and acid hydrolysis combined with for example NMR or MS analysis.[13,14]

The extent to which the enzyme is hindered is determined by the active site structure of the enzyme. More specific information on the substitution pattern can be achieved by comparing the degradation pattern from several different endoglucanases.

Further detailed structural information of the substituent pattern requires more information of how substituents affect the enzyme activity. This can be achieved either by comparing degradation patterns for various enzymes of an already characterized substrate or by using an enzyme with known degradation pattern. Partially modified celluloses with fully known substituent patterns are not commercially available; hence, it has become necessary to rely on structurally characterized enzymes. Since the number of structurally determined endo-

* Corresponding author: E-mail: Herje.Schagerlof@biochemistry.lu.se.
[†] Department of Molecular Biophysics, Lund University.
[‡] Department of Biochemistry, Lund University.
[§] Department of Technical Analytical Chemistry, Lund University.
[||] AstraZeneca R&D Mölndal.

10.1021/bm0701200 CCC: $37.00
Published on Web 07/07/2007

Table 1. Reducing Ends from Enzymatically Degraded MC Reported as Percent of the Total Number of Glycosidic Bonds[a]

	reducing ends (%)	DP (reducing ends)[b]	DP_w (SEC-MALS-RI)[c]
intact			1062
BaCel5A	4.6	22	62
HiCel5A	2.5	40	123
TrCel5A	3.6	28	105
HiCel7B	1.6	63	238
TrCel7B	2.1	48	136

[a] The average length of the produced fragments (DP) was calculated on the basis of the reducing end and SEC-MALS-RI analysis. [b] Calculated from reducing ends. [c] Calculated from SEC-MALS-RI (M_w/(average $M_{monomer}$) (187.4 g·mol^{-1})).

glucanases with substrate occupancy in the active site are few, this is highly limiting for the characterization of modified cellulose.

It is not known to what extent the enzymes are hindered by substituents or at which positions of the glucose residues substitution will interfere with hydrolysis. Earlier investigations indicate that the C-2 position on the glucose containing the new reducing end is the most important.[15−17] The sensitivity toward the position of the substituent also differs between endoglucanases. The minimal requirements for hydrolysis are undetermined, i.e., the number of glucose units in sequence, which fulfils the enzyme requirements regarding substitution that allow hydrolysis of the glycosidic bond. These characteristics will depend on the individual endoglucanase active site and the number of subsites present.

In this study the relationship between the structure of the enzyme active site and the capacity to accept substituents is investigated. A commercial batch of methyl cellulose (MC) with a DS of 1.80 was chosen as substrate for the study. MC is a nontoxic, water soluble cellulose derivative, produced by etherification of the hydroxyl groups along the cellulose backbone with methyl groups. Since the introduced methyl groups are uncharged and relatively small, it can be anticipated that the enzymes will be less hindered than by many other substituents, for example carboxymethyl or acetate groups.[8,11,17] The introduced substituent does not polymerize, and thus the interpretation of mass spectrum will be more straightforward.

To determine to which degree the enzymes accept substituents, MC was degraded using five endoglucanases from two different families and three different organisms. The degraded products were analyzed with size exclusion chromatography, multiangle light scattering, and refractive index detection (SEC-MALS-RI) as well as reducing end analysis to get an estimation of the enzyme capacity to degrade the substrate. Short-chained oligosaccharides (with a degree of polymerization (DP) up to 10) were analyzed using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOFMS). These products have been formed by enzymatic hydrolysis of the glycosidic bonds; thus, both ends of the oligosaccharide have been positioned in the active site. Hence, the measured DS will reflect the enzyme capacity to accommodate substituents in the active site. The results from these investigations are then correlated with differences in the active site for the structurally determined enzymes. The conclusions drawn could to some extent also be applied when evaluating the endoglucanases used without detailed information of the substrate interaction in the active site. Furthermore, modification of the cocrystallized substrate was simulated to evaluate the extracted information from the analysis of the enzymatically hydrolyzed MC.

Experimental Section

Chemicals. The MALDI matrix 2,5-dihydroxybenzoic acid (DHB) was from Aldrich (Steinheim, Germany) and methyl cellulose SM-1500 (viscosity type, 1500 cP; lot no., 103674) with the stated DS 1.80, was from Shin-Etsu Chemical Co. (Tokyo, Japan). Trifluoroacetic acid (TFA) was from Riedel-de Haën (Seelze, Germany), potassium hexacyanoferrate(III), NaCl, glucose, and 50% NaOH solution were from Merck (Darmstadt, Germany), and sodium carbonate decahydrate was from ICN (Aurora, OH). All other chemicals were of analytical grade. The water used in all experiments was purified in a Milli-Q system, (18 MΩ cm, Millipore Bedford, MA).

Enzymes. The following endoglucanases were used: *Bacillus agaradhaerens* Cel5A (BaCel5A), *Humicola insolens* Cel5A (HiCel5A), *H. insolens* Cel7B (HiCel7B), *Trichoderma reesei* Cel5A (TrCel5A), and *T. reesei* Cel7B (TrCel7B). BaCel5A, HiCel5A and HiCel7B were provided purified as monocomponents by Novozymes, Bagsværd, Denmark, and TrCel5A and TrCel7B were purified as described by Karlsson et al.[18]

Enzymatic Degradation of MC. For enzymatic degradation, MC was dissolved to a concentration of 10 g/L in water. To each sample enzyme was added to a final concentration of 1 μM. The degradation was allowed to continue for 72 h at room temperature. To verify that no further degradation occurs, a sample from each enzyme was analyzed also after 168 h, with no further degradation observed. The samples were stored at 4 °C until analysis. Prior to MALDI-TOFMS analysis the samples were filtered through 10 kDa cutoff Nanosep Omega filters (Pall, Ann Arbor, MI) after which the permeate was collected and dried in a Christ rotation vacuum concentrator model 2-18 (Martin Christ GmbH, Osterode am Harz, Germany).

Reducing End Analysis. The amount of reducing sugars formed during enzymatic degradation was measured using a ferricyanide method adapted from Kidby and Davidson.[19] A reagent was prepared by dissolving 150 mg of $K_3[Fe^{III}(CN)_6]$, 7 g of sodium carbonate decahydrate, and 200 μL of a 50% sodium hydroxide solution in 250 mL of water. Determination was carried out by pipetting 200 μL of reagent and 50 μL of sample into a well of a 96-well microtiter plate (Whatman, Haverhill, MA). After sample deposition, the microtiter plate was sealed and incubated at 37 °C for 90 min. Absorbance was measured at 405 nm using a SpectraCount microplate photometer (Packard, Maryland, CT). Calibration was performed using glucose solutions in the concentration range of 0−10 mM.

SEC-MALS-RI. The molar mass distributions of intact and enzymatically degraded MC were determined by SEC-MALS-RI. The analytes were separated on a TSK-GEL GMPWXL 7.8 × 300 mm, particle size 13 μm, linear mixed bed size-exclusion column (TosoHaas Bioseparation Specialists, Stuttgart, Germany) at a flow rate of 0.5 mL/min. The mobile phase was a 0.10 M NaCl solution filtered with a 0.22 mm mixed cellulose ester filter GSWP (Millipore Corp.). The pump was a Shimadzu LC-10AD liquid chromatography pump and the degasser a Shimadzu DGU-14A (Shimadzu Corp., Tokyo, Japan). Injection of the polymer solution was carried out by a Waters 717 plus Autosampler (Waters, Milford, MA), equipped with a 100 μL sample loop. The injected amount of sample was 100 μg as the polymer concentration was held at 1 g/L.

The light scattering photometer was a DAWN-DSP multiangle light scattering instrument (Wyatt Technology, Santa Barbara, CA). Simultaneous concentration detection was performed using an Optilab DSP interferometric refractometer (Wyatt Technology). Both detectors operated at a wavelength of 690 nm. The output signals from the detectors were analyzed

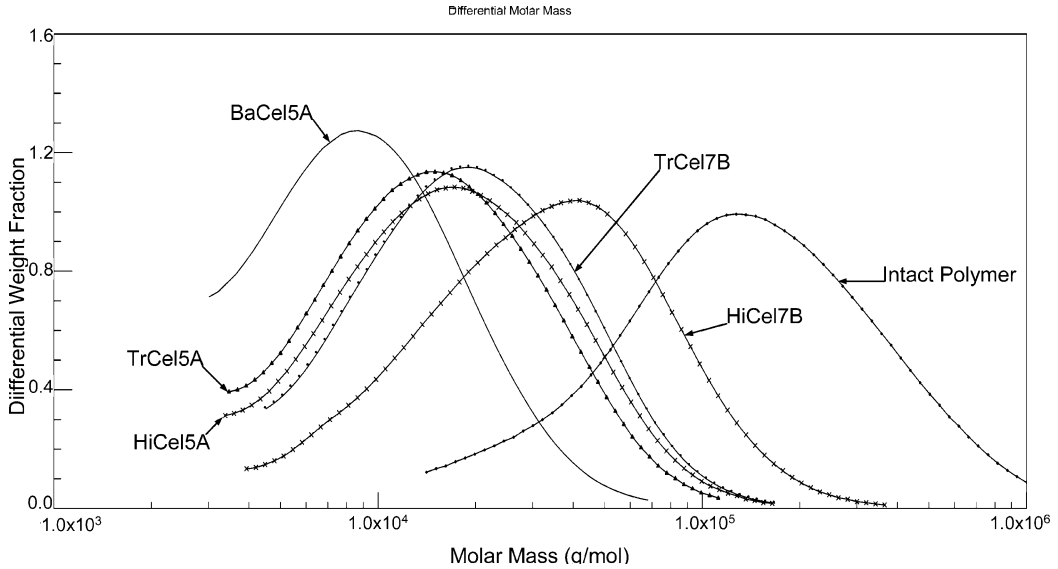

Figure 1. Differential molar mass distribution of intact and enzymatically degraded MC as measured with SEC-MALS-RI. Enzymes (endoglucanases) used to degrade the polymer are indicated in the figure.

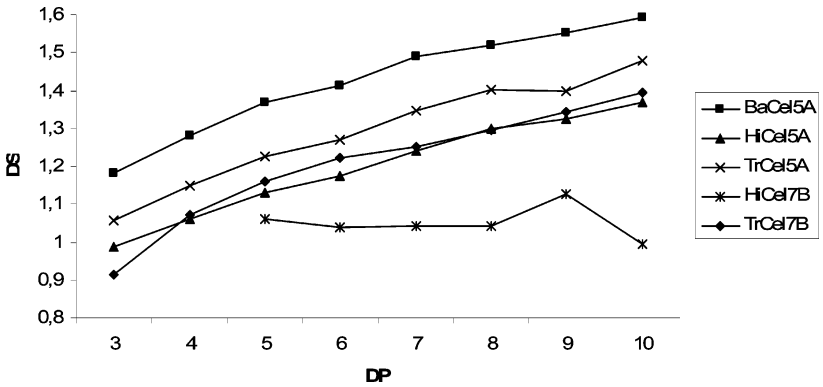

Figure 2. Average DS (degree of substitution) shown as a function of DP (degree of polymerization) for enzymatically degraded MC. A higher value of DS indicates a higher tolerance toward substituents present in the active site. Values for DP3 and 4 from HiCel7A were omitted due to impurities interfering with the signal for these products.

by the ASTRA 4.90.07 software (Wyatt Technology). The weight-average molar mass (M_w) was obtained by Debye plots using Berry's method.[20] A d*n*/d*c* value of 0.141 was used, and it was assumed to be constant for the molar mass range studied.[21]

MALDI-TOFMS. MALDI-TOFMS experiments were performed using a PerSeptive Voyager-DE STR (Applied Biosystems, Framingham, MA) equipped with an N_2-laser, time-lag focusing, reflector, and a tandem coupled microchannel plate detector in reflector mode. Mass spectra were acquired in positive ion mode with the time-lag focusing and the reflector activated at all times. The accelerating voltage was 20 kV and the reflector voltage was 13% higher. The laser intensity was held slightly above threshold, and the lag time was 150 ns plus instrument offset. To prevent detector saturation, the detector activation gate was set at *m*/*z* 300. For most acquisitions, the guide wire was set to 0.01% of the acceleration voltage. Spectra were accumulated for 200 laser shots. Several different positions within the sample spot were used for the acquisition.

DHB was used as MALDI matrix for the analysis of the degraded MC and was dissolved to 10 g/L in H_2O. The filtered samples were dissolved to approximately 1 g/L in H_2O. Thereafter the analyte solutions and the matrix solution where mixed 1:4 (v/v) and vortexed. A 1 μL aliquot of the mixture was applied on a MALDI sample plate and then allowed to dry at reduced pressure by insertion of the sample plate directly into the loading chamber of the mass spectrometer. When the pressure was lower than 2×10^{-2} Torr, the sample was moved into the source chamber.

Structural Analysis. All structures and sequences used refer only to the catalytic core of the respective endoglucanase. The definition of the catalytic core was according to the information available for the individual enzymes at the CAzy database.[22] The X-ray structures where downloaded from the RCSB Protein Data Bank.[23] The accession numbers of the main structures used are 1QI0, 1QI2, 1H11, 1H2J, 1HF6, and 1H5V for BaCel5A,[24−26] 1OJK for HiCel7B,[27] and 1EG1 for TrCel7B.[28]

Sequence alignment was performed using the softwares ClustalW[29] and T-Coffee;[30] all parameters were left at default values. Previously identified conserved amino acids of family 5 endoglucanases[31] were used to improve alignment accuracy together with information of the catalytic trio identity for each enzyme.[22] The structures were visualized and compared using Pymol.[32] By overlapping the structures of several cocrystallized substrates present in the active site in the atomic structures of BaCel5A, a model of a cellopentaose molecule in the active site could be constructed. Structural alignment of the α carbons (C_α's) in the peptide chains according to Kleywegt et al.[28] was performed by using the Magic Fit option of SPDB viewer,[33] thereby enabling the visualization of the substrate from HiCel7B in the active site of TrCel7B. Simulated modification of the substrates in enzyme active sites was performed by artificially inserting methyl group coordinates into the protein data bank

Table 2. Hydrolysis Products from Enzymatically Degraded MC, Detected with MALDI-TOF−MS, Shown as Chain Length (DP) and the Number of Methyl Groups[a]

no. of Me	DP 3	4	5	6	7	8
0						
1	a b c e					
2	a b c e	a b c d e	b d			
3	a b c e	a b c d e	a b c d e	b d		
4	a b c e	a b c d e	a b c d e	a b c d e	d	
5	a b c e	a b c d e	a b c d e	a b c d e	a b d e	d
6	a b c	a b c d e	a b c d e	a b c d e	a b c d e	a b d e
7	a c	a b c d e	a b c d e	a b c d e	a b c d e	a b c d e
8		a b c d e	a b c d e	a b c d e	a b c d e	a b c d e
9		a b c	a b c d e	a b c d e	a b c d e	a b c d e
10		a	a b c d e	a b c d e	a b c d e	a b c d e
11			a c	a b c d e	a b c d e	a b c d e
12			a	a b c e	a b c d e	a b c d e
13				a c	a b c d e	a b c d e
14					a b c e	a b c e
15					a c	a b c e
16					a	a b c e
17						a c
18						a

[a] a, BaCel5A; b, HiCel5A; c, TrCel5A; d, HiCel7B; e, TrCel7B.

Table 3. Summary of the Results from the Different Analysis Methods with the Most Effective Enzyme on Top and Then in Decreasing Efficiency[a]

reducing ends[b]	SEC-MALS-RI[c]	MS[d]
BaCel5A	BaCel5A	BaCel5A
TrCel5A	TrCel5A	TrCel5A
HiCel5A	HiCel5A	TrCel7B, HiCel5A
TrCel7B	TrCel7B	HiCel7B
HiCel7B	HiCel7B	

[a] The more effective enzyme gives a product which is degraded to a higher extent, i.e., has more reducing ends released, and gives shorter average chain lengths and a higher DS per DP (i.e., is capable of degrading higher substituted parts). [b] Formed reducing ends measures the total enzyme efficiency to hydrolyze the glycosidic bonds of the substrate (Table 1). [c] SEC-MALS-RI gives information of the length of the produced fragments (Figure 1). [d] MS data gives an estimation of the DS of the produced fragments (Figure 2).

(PDB) structure files. Coordinates for the methyl groups were downloaded from the Hetero-compound Information Centre-Uppsala (HIC-Up) database.[34] Atom distances and torsion angles were monitored in the SPDB viewer.[33]

Results and Discussion

In this study, MC was degraded with a total of five endoglucanases belonging to two glycosyl hydrolase families and the resulting products were analyzed. The results were evaluated and used to elucidate information concerning enzyme sensitivity for substituents. This information in turn was correlated with the available structural information of the enzymes.

Enzymatic Hydrolysis Sensitivity. To determine the molar mass distribution prior to and after degradation, the samples were analyzed with SEC-MALS-RI. The enzymes had significantly different degradation capability, even though all of them degrade MC (Figure 1). HiCel7B could not degrade the cellulose as efficiently as the other investigated enzymes, resulting in a very broad molar mass distribution. The most effective enzyme was BaCel5A, degrading the polymer extensively, while the remaining three enzymes had similar average degradation capacity.

As a complement to the SEC-MALS-RI data an analysis of the formed reducing ends was used to estimate to what degree the different endoglucanases were able to degrade MC. This method gives an estimation of the production of new reducing ends, i.e., how many β-1−4 linkages that have been hydrolyzed by the enzymes. BaCel5A produced most reducing ends of the studied enzymes followed by TrCel5A. The least effective of the investigated enzymes was HiCel7B (Table 1).

The average length of the produced fragments was calculated from the obtained M_w from SEC-MALS-RI measurements divided with the average monomer molar mass ($M_{monomer}$, 187.4 g·mol^{-1}). The average length was also calculated from the reducing end measurements (Table 1). The SEC-MALS-RI data and the reducing ends analysis data clearly agree on the relative order of substituent sensitivity between the enzymes although the lengths of the produced fragments that the individual enzymes produce do not correlate. The large differences in product length are caused by the fact that the MALS detectors lose sensitivity in the low-M_w range. Notwithstanding, both techniques confirmed the fact that the different enzymes degraded MC to various degrees, clearly demonstrating the enzymes varying sensitivity toward substituents.

MALDI-TOFMS. To enable a more effective ionization of the short fragments of the degraded MC, the longer fragments have to be removed. In previous studies,[9,35,36] this sample preparation was obtained through fractionation of the sample with SEC. Since this method is both time-consuming and elaborate, it was replaced by a filtration step using a Nanosep 10 kDa Omega filter instead. To ensure that filtration did not affect the result, the filtrated sample was compared with earlier obtained results from SEC fractionated samples in previous studies.[36] It could be concluded that filtration sample pretreatment does not affect the quality of the results but offers a faster and more efficient method compared with SEC fractionation.

When the different enzymes were compared, it was observed that shorter products exhibit a lower DS (Figure 2). A higher

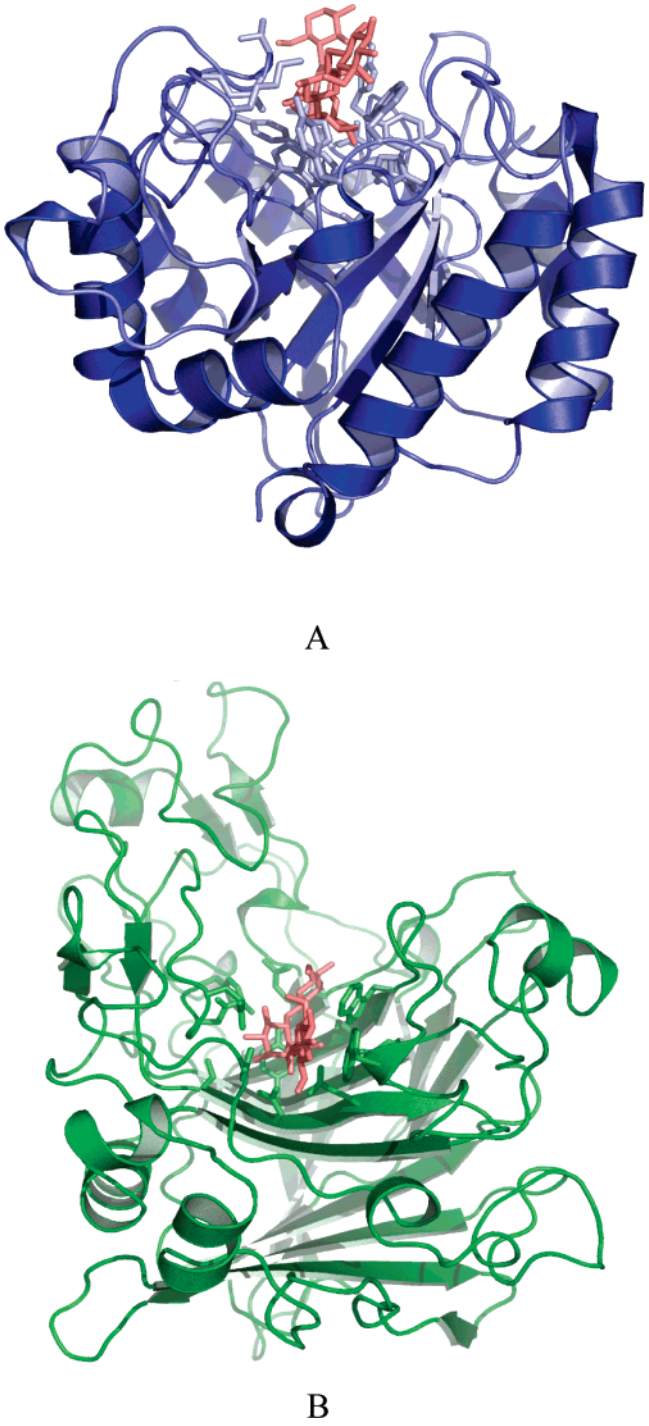

Figure 3. General active site structures of endoglucanase family 5 in cyan (A) and endoglucanase family 7 in green (B) in cartoon representations. The catalytic residue side chains are represented, and the axis of the catalytic site is perpendicular to the image.

value of DS for the lower DP indicates a higher tolerance toward substituents present in the active site. For example, an oligosaccharide of DP 3 has had 2/3 of the total number AGU in the position next to the cleavage point (−1 and +1). In comparison, an oligosaccharide of DP 7 has only had 2/7 of the total number of AGU in the corresponding position. For longer products DS will approach the value of the intact polymer (1.8) and not give any additional information about the sensitivity of the active site. The measured DS of short fragments are somewhat overestimated depending on the lower ionization efficiency for low substituted oligosaccharides.[16] This will however not affect the comparison of the relative enzyme efficiency, since

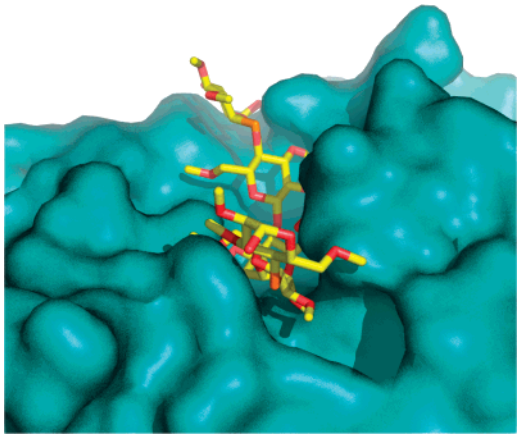

Figure 4. Active site of *Bacillus agaradhaerens* Cel5A (BaCel5A) with a simulation of modified methyl cellopentaose in the active site with the −3 subsite in the right bottom corner of the image. The simulated substitution confirms the conclusion that no hindrance for full substitution is present except in the −1 subsite.

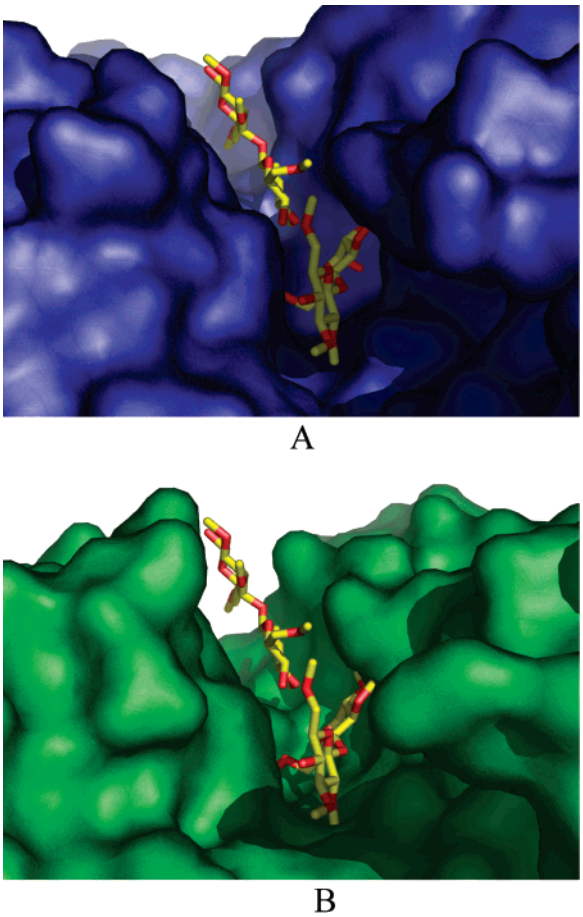

Figure 5. Active site of *Humicola insolens* Cel7B (HiCel7B) (A) and *Trichoderma reesei* Cel7B (TrCel7B) (B) with two methyl cellobiose molecules in the active site with the −2 subsite at the bottom of the image. The simulated modification with methyl groups of the cellobiose shows the difficulties to accommodate a fully substitutated substrate in the active site. Structural alignment of the C_α's from HiCel7B and TrCel7B enables the visualization of the modified substrate from the HiCel7B structure in the TrCel7B active site.[28]

the relative difference of the investigated endoglucanases is not influenced. Due to the fact that the DS for longer products approach the DS of the intact polymer, the shorter fragments are the most interesting for this investigation.

CDV

Table 4. Schematic Representation of the Substrate Binding Amino Acids in the Individual Subsites for the Different Endoglucanases Used[a]

subsite	BaCel5A	TrCel5A	HiCel5A	TrCel7B	HiCel7B
−3	Trp39 Tyr40 Lys267	Gly17 Ser18 Phe278	Trp269		
−2	His35 Tyr66 Ser69 Thr235 Trp262 Lys267 Glu269	Asn14 Gly45 Asn53 Trp273 Phe278 Ser280	His19 Ser44 Gln52 Trp264 Trp269 Gly271	Arg108 Tyr170 Ser318 Trp320	Arg108 Tyr171 Ser345 Trp347
−1	Arg62 Tyr66 His101 Asn138 Glu139 Tyr202 Glu228 Ala234 Thr235	Arg41 Gly45 His85 Asn128 Glu129 Tyr201 Glu240	Arg40 Ser44 His84 Asn123 Glu124 Tyr191 Glu231	Asn142 Tyr146 Asp172 Gln174 Glu196 Asp198 Glu201	Asn143 Tyr147 Asp173 Gln175 Glu197 Asp199 Glu202
+1	Trp178	Trp166	Trp161	Trp329	Trp356
+2	Gln180 His206	Ser168 Asp205	Gly163 Asp195	Ala222	Asn237

[a]This table is based on available structural information for BaCel5A, TrCel7B, and HiCel7B in combination with the results from the sequence alignment of family 5 (Figure 6) and the structural alignment of family 7.[28]

Qualitative analysis of the MS data, i.e., consideration of the individually detected fragments, gives an estimated overview of how substituted fragments an enzyme can produce (Table 2). From Table 2 it can be concluded that both *H. insolens* enzymes produce oligosaccharides with lower DS than the average. Hence it can be concluded that the enzymes from *H. insolens* have higher sensitivity for substituents present in the active site than the other enzymes included in this work. It is also observed that BaCel5A produces fragments with significantly higher DS, indicating a capability to degrade higher substituted regions of the polymer.

The samples from HiCel7B were contaminated with a substance that interfered with the MALDI-TOFMS measurements and gave significantly lower peak intensities for the detected oligosaccharides. The lower intensities introduced a higher uncertainty than for the other enzymes. However, the low signal obtained will not affect the DS values significantly, and it is therefore still a relevant conclusion that HiCel7B, which has significantly lower DS than the other enzymes, is more sensitive for substituents than the other enzymes.

Structural Aspects of Endoglucanase Specificity. There are significant variations in substitution sensitivity between endoglucanases, within and between families. In Table 3 the results from the various analysis methods used have been summarized with the most efficient enzyme on top and then in falling order. From this it is clear that the family 5 endoglucanases hydrolyze MC more efficiently than family 7. It can also be concluded that in comparison the enzymes used from *T. reesei* are more efficient than the *H. insolens* enzymes. The most efficient of the enzymes used for MC degradation however proved to be BaCel5A. These results can then be directly related to the substituent sensitivity and enzyme properties. The dimensions of the active site in combination with the enzyme affinity for both substrate and product influence the enzyme sensitivity. The sensitivity differences between families that can be seen in these results can be related to structural differences by comparing the general active site structures. Atomic structures for all of the used enzymes were not available but by combining the information of the formed products with structural alignment and bioinformatics more information of the unknown enzyme structures could be gained.

General Active Site Structure. The family 5 catalytic core has a general structure of an eight stranded TIM barrel, which is a $\beta-\alpha-\beta$ motif forming a parallel β-sheet that is twisted into a barrel with the α-helices located on the outside of the barrel. The active site is suspended across the top of the barrel with highly flexible sides and little vertical restrictions[24−26] (Figure 3A). This gives a high degree of freedom for substrate size in the vertical direction. There are five individual subsites, three at the nonreducing end and two at the reducing end.

The catalytic core of family 7 consists of two antiparallel β-sheets that pack face to face to form a slightly twisted β-sandwich. On the concave side of the sandwich long loops form a groove where the active site is located across the face of the β-sheet[27,28,37] (Figure 3B). These enzymes may accommodate up to seven individual subsites but only four of them contribute significantly[27] when hydrolyzing natural substrate. However, since the enzyme substrate interaction is not fully understood, it is possible that more of the subsites have to be taken into account when considering artificial substrates.

In comparison family 5 enzymes were more effective than family 7 at degrading MC (Figure 2, Table 3). Data indicates that family 7 enzymes prefer longer low substituted regions (Table 2). This may depend on a substituent hindrance due to less available space in the active site. Family 5 enzymes were able to accommodate a larger number of substituents in the active site (Figure 4) and thereby able to degrade shorter low substituted regions in comparison to family 7 enzymes. The options of family 7 endoglucanases to accommodate substituents are to some extent limited by the fact that the catalytic residues are located on strands in the β-sheet forming the floor of the active site[27,28] (Figure 3B). This implies that the substrate has to be positioned close to the β-sheet to bind in the −1 subsite. Hence, there is a natural limit for the possibility to vertically adjust the substrate position to compensate for the presence of substituents in the active site. This would account for the higher substituent sensitivity of family 7 endoglucanases since the substituents pointing into the enzyme will have higher probability of preventing a proper binding needed for hydrolysis to take place (Figure 5).

Subsite Interactions. There is not yet enough knowledge of endoglucanase interactions in the active site with natural substrates to predict enzyme sensitivity for any given derivatized cellulose. However, structurally determined endoglucanases have revealed information concerning the active sites. Table 4 shows the substrate binding amino acids for the structurally determined enzymes of all the individual subsites. In addition, the sequentially aligned amino acids for the unknown structures (Figure 6) are listed for each individual subsite. In general, within a family there is a high conservation of the substrate binding amino acids in subsite −1 and only small variations in adjacent subsites.[28,31] Hence, variations in sensitivity toward substituents

Figure 6. Sequence alignment of family five endoglucanases used for MC hydrolysis. α helices and β strands of the *Bacillus agaradhaerens* Cel5A (BaCel5A) structure are indicated above the sequence (a for α helices and b for β strands). Conserved amino acids are indicated with a star below sequences; bold stars indicate amino acids conserved within the whole family. In gray boxes substrate binding amino acids are indicated. The amino acid numbering is indicated for each individual sequence at the end of the rows.

within a family seem to originate mainly from minute differences in the surrounding subsites.

Earlier results show that BaCel5A can accept full substitution of MC at the +1 and +2 subsites and that −1 accepts C3 or C6 substitutions but not C2.[15] However, DP 3 with 7 substituents and DP 4 with 10 substituents can be detected in the MS data (Table 2), indicating that BaCel5A can degrade polymers, fully substituted either at the −3 and −2 or at the +1 and +2 subsites. Hence the present data show that BaCel5A also can accept full substitution at the −2 and −3 subsite, but if all four subsites indeed could be fully substituted at the same time, it is reasonable to expect a more complete degradation. Simulated substitution of the constructed cellopentaose in the active site of BaCel5A by inserting methyl group coordinates into the PDB file shows no visible interference of hydrolysis by any substituents except in the −1 subsite (Figure 4). Hence, it can be concluded that BaCel5A does not have any requirements for a specific substitution pattern to allow hydrolysis apart from a glucose unit in the −1 subsite with C2 and either C3 or C6 unsubstituted. BaCel5A's only other limitation for hydrolysis is that the DS should not exceed a certain total number for all the subsites together. Rather than a lack of space in the active site this limitation may originate from disturbances in the hydrogen-bonding network, leading to a decrease in gained energy upon substrate binding. Subsequently, while the loss of entropy is constant upon substrate binding, a lower number of side chains that strongly interacts with the substrate reduce the gain in enthalpy. The enzyme ability to lower the activation energy for hydrolysis, however, is proportional to the difference between entropy loss and enthalpy gain. At a certain threshold value of binding energy, the enzyme will no longer be able to lower the activation energy enough for hydrolysis to occur.

Sequence alignment of the three family 5 endoglucanases in this work reveals highly conserved subsites (Figure 6). However it is noticeable that BaCel5A has twice as many charged substrate-binding residues compared with the other two family 5 members but no glycines. TrCel5A and HiCel5A have three and two glycines respectively in the active site as well as residues with very short side chains in general compared to BaCel5A. Hence, the main difference within family 5 is the length of the side chains and the number of charges present in the active site. This apparently affects the enzyme ability to interact with the substrate since there is a clear difference between the substitution sensitivity within the family 5 enzymes. For BaCel5A, DP 4 with 10 substituents was found but for TrCel5A and HiCel5A, DP 4 with only 9 substituents was found (Table 2). This is an indication that these two enzymes require either a fully unsubstituted AGU in the −1 subsite or that the minimal requirement is a longer sequence than the one single AGU that BaCel5A needs. The conclusion concerning the BaCel5A enzyme was a confirmation that certain positions cannot be substituted in the −1 subsite for hydrolysis to occur.[16]

From Table 2 it is possible to discuss a "local threshold value" for DS (DS_{TV}), i.e. the number of substituents per glucose unit that the endoglucanase can tolerate in the active site. The threshold value will be equal to the least substituted oligomers, not degraded by the endoglucanase. From Table 2, in which the detected end products from enzymatic degradation are listed, the DS_{TV} can be estimated. Of special interest are the end products of DP 5 and 6 since a difference in the tolerance of substituent can be observed for the investigated enzymes. For HiCel5A and HiCel7B, the products DP 5 with 2 substituents as well as DP 6 with 3 substituents can be observed. However, for BaCel5A, TrCel5A, and TrCel7B these are substrates and can thus be hydrolyzed. Thus, the local threshold values for HiCel5A and HiCel7B are 0.4 for DP5 and 0.5 for DP6, in contrast to the higher DS_{TV} for BaCel5A, TrCel5A, and TrCel7B which are 0.6 for DP5 and 0.67 for DP6. It is possible that this

difference in sensitivity is due to lower tolerance in disturbances of the hydrogen-bonding network for HiCel5A and HiCel7B.

The substrate present in the HiCel7B structure was modified into methyl cellulose by inserting methyl group coordinates into the substrate PDB file of the determined HiCel7B structure[27] (Figure 5A). The modeled substrate active site interactions indicate that this enzyme does not have any specific requirements except for in the −1 subsite. Due to the positioning of the catalytic residues in the −1 subsite any other substituent position than C3 would sterically hinder hydrolysis of the polymer. As for the other subsites possibly only the −2 subsite would suffer any hindrance from substitution at any position on the AGU. Structural alignment of the C_α's from HiCel7B and TrCel7B enables the visualization of the substrate from the HiCel7B structure in the TrCel7B active site.[28] The natural substrate has approximately the same interactions with the catalytic residues in TrCel7B as it does in HiCel7B. It was therefore concluded that the actual substrate position in HiCel7B can be directly transferred to the positioning of the structurally aligned substrate in TrCel7B[38] (Figure 5B). Hence the steric hindrance of substituents for derivatized cellulose in the TrCel7B active site should be similar to that of HiCel7B.

When comparing the sensitivity of the two enzymes from family 7, TrCel7B is less sensitive to substituents than HiCel7B (Table 3). Since the substrate binding residues are 100% conserved in all the subsites except subsite +2, where Ala222 in TrCel7B corresponds to Asn237 in HiCel7B,[28] the difference in substituent sensitivity may depend more on the available space inside the active site rather than the environment of the individual subsites (Figure 5). In HiCel7B strand 192−194 together with strand 221−223 forms an antiparallel β-sheet and together with a small 3_{10} helix 225−228[27] forms an indent next to the subsites +1 and +2. This could make it more difficult to accept substituents in these subsites.

Conclusions

In this study formed products from enzymatic methyl cellulose degradation were analyzed using reducing end analysis, SEC-MALS-RI, and MALDI-TOFMS in combination with bioinformatics and structural analysis of endoglucanase active sites. These methods gave information of the general active site structure and subsite interactions of importance in hydrolysis of substituted cellulose. The evaluation of endoglucanase specificity for methyl cellulose hydrolysis resulted in a confirmation that certain positions cannot be substituted in the −1 subsite for hydrolysis to occur.[16] The analysis of the products can be discussed in terms of a threshold value of local DS for the glucose units in the active site rather than a specific substituent pattern. By combining hydrolysis data from enzymes with divergent threshold values, a more extensive characterization of the substituent pattern can be obtained. This will simplify and improve the enzymatic characterization of substituent distribution in cellulose derivatives. Since soluble cellulose derivatives are generally used as substrates during enzyme characterization and comparison as well as in endoglucanase activity assays, it is important to be aware that the enzymes sensitivity toward substituents differs.

Acknowledgment. This work was financed by the Centre for Amphiphilic Polymers from renewable resources (CAP), Lund, Sweden. The enzymes from *B. agaradhaerens* and *H. insolens* were graciously provided by the late Martin Schülein (Novozymes AS).

References and Notes

(1) Richardson, S.; Gorton, L. *Anal. Chim. Acta* **2003**, *497*, 27−65.
(2) Brandt, L. In *Ullmann's Encyclopedia of Industrial Chemistry*, 5th ed.; Campbell, F. T., Pfefferkorn, R., Rounsaville, J. F., Eds.; VCH Verlagsgesellschaft: Weinheim, Germany, 1986; Vol. A5, pp 461−488.
(3) Kobayashi, K.; Huang, C. I.; Lodge, T. P. *Macromolecules* **1999**, *32*, 7070−7077.
(4) Rinaudo, M. *Biomacromolecules* **2004**, *5*, 1155−1165.
(5) Mischnick, P. *Macromol. Symp.* **1997**, *120*, 281−290.
(6) Mischnick, P.; Heinrich, J.; Gohdes, M.; Wilke, O.; Rogmann, N. *Macromol. Chem. Phys.* **2000**, *201*, 1985−1995.
(7) Horner, S.; Puls, J.; Saake, B.; Klohr, E. A.; Thielking, H. *Carbohydr. Polym.* **1999**, *40*, 1−7.
(8) Saake, B.; Lebioda, S.; Puls, J. *Holzforschung* **2004**, *58*, 97−104.
(9) Karlsson, J.; Momcilovic, D.; Wittgren, B.; Schulein, M.; Tjerneld, F.; Brinkmalm, G. *Biopolymers* **2002**, *63*, 32−40.
(10) Schagerlof, H.; Johansson, M.; Richardson, S.; Brinkmalm, G.; Wittgren, B.; Tjerneld, F. *Biomacromolecules* **2006**, *7*, 3474−3481.
(11) Wirick, M. G. *J. Polym. Sci., Polym. Chem. Ed.* **1968**, *6*, 1965−1974.
(12) Schagerlof, H.; Richardson, S.; Momcilovic, D.; Brinkmalm, G.; Wittgren, B.; Tjerneld, F. *Biomacromolecules* **2006**, *7*, 80−85.
(13) Mischnick, P.; Niedner, W.; Adden, R. *Macromol. Symp.* **2005**, *223*, 67−77.
(14) Puls, J.; Altaner, C.; Saake, B. *Macromol. Symp.* **2004**, *208*, 239−253.
(15) Cohen, A. S.; Nilsson, C.; Schagerlof, H.; Tjerneld, F.; Gorton, L. *Anal. Chim. Acta* **2006**, *561*, 16−24.
(16) Momcilovic, D.; Schagerlof, H.; Rome, D.; Jornten-Karlsson, M.; Karlsson, K. E.; Wittgren, B.; Tjerneld, F.; Wahlund, K. G.; Brinkmalm, G. *Anal. Chem.* **2005**, *77*, 2948−2959.
(17) Nojiri, M.; Kondo, T. *Macromolecules* **1996**, *29*, 2392−2395.
(18) Karlsson, J.; Siika-aho, M.; Tenkanen, M.; Tjerneld, F. *J. Biotechnol.* **2002**, *99*, 63−78.
(19) Kidby, D. K.; Davidson, D. J. *Anal. Biochem.* **1973**, *55*, 321−325.
(20) Andersson, M.; Wittgren, B.; Wahlund, K. G. *Anal. Chem.* **2003**, *75*, 4279−4291.
(21) Poche, D. S.; Ribes, A. J.; Tipton, D. L. *J. Appl. Polym. Sci.* **1998**, *70*, 2197−2210.
(22) Coutinho, P. M.; Henrissat, B. Carbohydrate-Active Enzymes server, http://afmb.cnrs-mrs.fr/CAZY/, 1999.
(23) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(24) Varrot, A.; Schulein, M.; Davies, G. J. *J. Mol. Biol.* **2000**, *297*, 819−828.
(25) Varrot, A.; Schuelein, M.; Fruchard, S.; Driguez, H.; Davies, G. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *D57*, 1739−1742.
(26) Varrot, A.; Tarling, C. A.; MacDonald, J. M.; Stick, R. V.; Zechel, D. L.; Withers, S. G.; Davies, G. J. *J. Am. Chem. Soc.* **2003**, *125*, 7496−7497.
(27) Ducros, V. M. A.; Tarling, C. A.; Zechel, D. L.; Brzozowski, A. M.; Frandsen, T. P.; von Ossowski, I.; Schulein, M.; Withers, S. G.; Davies, G. J. *Chem. Biol.* **2003**, *10*, 619−628.
(28) Kleywegt, G.; Zou, J.-Y.; Divne, C.; Davies, G. J.; Sinning, I.; Ståhlberg, J.; Reinikainen, T.; Srisodsuk, M.; Teeri, T.; Jones, T. A. *J. Mol. Biol.* **1997**, *272*, 383−397.
(29) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673−4680.
(30) Notredame, C.; Higgins, D. G.; Heringa, J. *J. Mol. Biol.* **2000**, *302*, 205−217.
(31) Wang, Q. P.; Tull, D.; Meinke, A.; Gilkes, N. R.; Warren, R. A. J.; Aebersold, R.; Withers, S. G. *J. Biol. Chem.* **1993**, *268*, 14096−14102.
(32) DeLano, W. L. *Pymol*; DeLano Scientific: San Carlos, CA, 2002.
(33) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714−2723.
(34) Kleywegt, G. J.; Jones, T. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *D54*, 1119−1131.
(35) Momcilovic, D.; Wittgren, B.; Wahlund, K.-G.; Karlsson, J.; Brinkmalm, G. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1107−1115.
(36) Momcilovic, D.; Wittgren, B.; Wahlund, K. G.; Karlsson, J.; Brinkmalm, G. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 1116−1124.
(37) Divne, C.; Ståhlberg, J.; Teeri, T. T.; Jones, T. A. *J. Mol. Biol.* **1998**, *275*, 309−325.
(38) Carugo, O. *J. Appl. Crystallogr.* **2003**, *36*, 125−128.

BM0701200