Probing Electrostatic Interactions along the Reaction Pathway of a Glycoside Hydrolase: Histidine Characterization by NMR Spectroscopy

Mario Schubert,1,8 David K. Y. Poon,1,11 Jacqueline Wicki,1,11 Chris A. Tarling,11 Emily M. Kwan,11 Jens E. Nielsen,1
Stephen G. Withers,1,11 and Lawrence P. McIntosh*,1,11

Department of Biochemistry and Molecular Biology, Department of Chemistry, The Michael Smith Laboratory, and The Protein Engineering Network of Centres of Excellence, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3, and Department of Biochemistry, Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland

Received February 5, 2007; Revised Manuscript Received April 17, 2007

ABSTRACT: We have characterized by NMR spectroscopy the three active site (His80, His85, and His205) and two non-active site (His107 and His114) histidines in the 34 kDa catalytic domain of Cellulomonas fimic xylanase Cex in its apo, noncovalently aza-sugar-inhibited, and trapped glycosyl−enzyme intermediate states. Due to protection from hydrogen exchange, the level of which increased upon inhibition, the labile 1H and 1H atoms of four histidines (t1H ∼ 0.1−300 s at 30 °C and pH ∼7), as well as the nitrogen-bound protons in the xylobio-imidazole and -isofagomine inhibitors, could be observed with chemical shifts between 10.2 and 17.6 ppm. The histidine pK_a values and neutral tautomeric forms were determined from their pH-dependent 13C-1−13H chemical shifts, combined with multiple-bond 1H=2J=15N scalar coupling patterns. Remarkably, these pK_a values span more than 8 log units such that at the pH optimum of ∼6 for Cex activity, His107 and His205 are positively charged (pK_a > 10.4), His85 is neutral (pK_a < 2.8), and both His80 (pK_a = 7.9) and His114 (pK_a = 8.1) are titrating between charged and neutral states. Furthermore, upon formation of the glycosyl−enzyme intermediate, the pK_a value of His80 drops from 7.9 to <2.8, becoming neutral and accepting a hydrogen bond from an exocyclic oxygen of the bound sugar moiety. Changes in the pH-dependent activity of Cex due to mutation of His80 to an alanine confirm the importance of this interaction. The diverse ionization behaviors of the histidine residues are discussed in terms of their structural and functional roles in this model glycoside hydrolase.

Dissecting the complex network of electrostatic interactions within the folded structure of a protein remains an important experimental and theoretical challenge. In particular, understanding these interactions and their changes along the reaction pathway of an enzyme is critical for understanding the ability of active site residues and cofactors to facilitate catalysis by acting in a coordinated manner as general acids/bases, nucleophiles, and electrophiles. One method for directly interrogating the electrostatic interactions within a protein involves the measurement of the site-specific protonation states and pK_a values of its constituent ionizable groups by NMR spectroscopy. This approach is particularly convenient for histidine residues, given the wealth of information that can be obtained for the 1H, 13C, and 15N nuclei within the imidazole ring (1). The pK_a of an unperturbed histidine side chain also falls near the physiological pH, and thus, the ionization state of this residue can be modulated readily by its protein environment. Furthermore, histidines often play key structural roles due to the planar aromatic nature of the heterocyclic imidazole side chain, along with its ability to act not only as a hydrogen bond donor and acceptor but also as the positively charged component of a salt bridge. As such, histidine residues serve critical enzymatic roles, functioning commonly in nucleophilic and general acid/base catalysis, as well as in the chelation of metal ions.

In this study, we have characterized by NMR spectroscopy the electrostatic interactions involving the histidine residues...
within the xylanase Cex (or ClXyn10A) from *Cellulomonas fimi*. This secreted enzyme, composed of separable N-terminal catalytic (34 kDa) and C-terminal cellulose-binding (11 kDa) domains, joined by a flexible glycosylated proline-threonine rich linker (2, 3), catalyzes the cleavage of β-1,4-glycosidic bonds in xylan, as well as a range of soluble ary glycosides based upon xylose and glucose (4). The catalytic domain of Cex adopts an (α/β)8-barrel structure (5) and is a family 10 member of the large GH-A clan of retaining glycoside hydrolases (6). In the retaining mechanism, a double-displacement reaction proceeds with direct nucleophile attack upon the anomeric center of the glycone by a covalent glycosyl—enzyme intermediate. A partner active site carboxyl serves a dual role of general acid and then base during the glycosylation and deglycosylation steps, respectively (7). Through detailed kinetic, crystallographic, and mutational studies, we and others have identified the nucleophile (Glu233) and the general acid/base catalyst (Glu127) in Cex and defined key structural and energetic determinants for substrate specificity and glycosyl—enzyme intermediate formation (5, 8, 9). These studies were facilitated by the availability of noncovalent inhibitors with micromolar to nanomolar affinities for this enzyme (10, 11), as well as slow substrates, such as 2-deoxy-2-fluorocellobiosides and -xylobiosides, that form long-lived glycosyl—enzyme intermediates (4, 5).

The catalytic domain of Cex (denoted as CexCD) contains five histidines, each in a distinct environment. Two of these, His107 and His114, are distant from the active site of the enzyme yet play important structural roles involving interactions with charged groups and bound water molecules. The remaining three are located in the active site of Cex, interacting with a constellation of polar groups, as well as with the bound substrate in the Michaelis complex and the glycosyl—enzyme intermediate. His85 is adjacent to the general acid/base Glu127, His205 directly contacts the nucleophile Glu233, and His80 forms a hydrogen bond to an exocyclic oxygen of the bound substrate. Using NMR spectroscopy, we demonstrate that the pK_a values of these histidines differ by at least 8 log units due to the distinct electrostatic interactions experienced by each within Cex. Furthermore, the pK_a value of His80 drops by more than 5 units upon formation of the trapped glycosyl—enzyme intermediate, illustrating the significant electrostatic changes that occur within the active site of this model glycoside hydrolase along its reaction pathway.

**Experimental Procedures.**

**Protein Preparation.** The expression and purification of full-length wild-type and mutant (H80A or H205A) Cex in *Escherichia coli*, and the subsequent isolation of the catalytic domain after proteolytic cleavage of the linker, were performed as described previously (12). The mutations were introduced using the QuikChange method (Stratagene). Unlabeled CexCD was obtained from bacteria grown in LB medium, whereas 15N-labeled wild-type and mutant CexCD were prepared using M9 medium containing 1.0 g/L [15 N]Celtone (Stable Spectral Isotopes, Inc.). CexCD, selectively labeled with [13 C]-[15 N]His, was expressed in a synthetic medium enriched with 75 mg/L (13 C, 99%) L-histidine·HCl·H2O (Cambridge Isotope Laboratories, Inc.) and 0.02% NaCl for subsequent NMR spectroscopic analyses. Covalently inhibited 2FCb-CexCD was generated by addition of an ~3-fold molar excess of 2′,4′-dinitrophenyl 2-deoxy-2-fluoro-β-celllobiose to the purified protein. This complex is stable with a half-life of many months (8, 12). Noncovalent complexes of CexCD were formed in the presence of an ~2-fold molar excess of xylobio-isofagomine [1,5-imino-1,4,5-trideoxy-3-O-β-D-xlylopyranosyl]-D-threo-pentitol] or xylobio-imidazolide [(6S,7S,8S)-7,8-dihydroxy-6-(β-D-xlylopyranosyl)-5,6,7,8-tetrahydroimidazole[1,2-α]pyridine] (11).

**NMR Spectroscopy.** Unless stated otherwise, all NMR spectra were recorded at 30 °C using a 600 MHz Varian INOVA spectrometer equipped with a gradient triple-resonance probe. NMR data were processed using NMRpipe (14) and analyzed using SPARKY (15). 1H and 13C chemical shifts were referenced to an external sample of DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) and 15N shifts referenced indirectly via chemical shift ratios (16). Experiments used for this study included gradient versions of a one-dimensional (1D) jump-and-return echo (17), two-dimensional (2D) jump-and-return echo NOESY (17), one-bond sensitivity-enhanced 15N HSQC (18), non-sensitivity-enhanced 15N HSQC with a total INEPT delay of 22 ms to select for multiple-bond correlations while suppressing direct couplings (19), one-bond constant-time 13C HSQC (20), one-bond sensitivity-enhanced 13C HSQC with a CPMG pulse train (21), and a CβHδ (22, 23). J_13H couplings were measured from 15N IPAP-HSQC spectra (24), as well as 1D jump-and-return echo and 2D 15N HSQC spectra recorded without 15N decoupling during the observation period.

**Hydrogen Exchange.** Proton–deuterium exchange (HX) measurements were carried out by passing unlabeled CexCD and 2FCb-CexCD through a 2.5 mL Sephadex G-25 medium spin column in a table top centrifuge to rapidly transfer the protein into 99% D_2O buffer containing 20 mM sodium phosphate and 0.02% NaN_3 (pH 6.5) (25). A time series of 1H NMR spectra were recorded at 20 °C starting ~5 min after transfer. Magnetization transfer-based proton–proton exchange measurements were recorded with 2FCb-CexCD at 10 °C using the CLEANEX-PM pulse sequence (26). Rate constants were determined using a MatLab macro provided by W.-Y. Choy (University of Toronto, Toronto, ON).

**pK_a Measurements.** Histidine pK_a values were determined from the pH dependence of the imidazole 1H and 13C chemical shifts of CexCD and 2FCb-CexCD selectively labeled with [13 C]-[15 N]His. The chemical shifts were measured using 13C-CPMG-HSQC spectra (21), each recorded in 10–30 min. The pH of the protein samples was first increased by 0.3 unit increments from a starting value of 6.5 by the addition of microliter aliquots of 0.1–0.5 M NaOH and HCl, respectively. The pK_a values were obtained by fitting the measured chemical shifts as a function of pH to the standard equations describing one or two macroscopic protonation equilibria (27). In the case of biphasic titration curves, the histidine pK_a was taken from the major, co-incident 1H and 13C chemical shift changes, with any minor changes attributed to the ionization of neighboring residues in the protein. Errors are estimated to be ±0.1 unit for all pH measurements and fit pK_a values.
Enzyme Kinetics. The pH dependence of $k_{\text{cat}}/K_m$ for full-length H80A Cex was determined at 37 °C in 150 mM NaCl and 0.1% (w/v) bovine serum albumin (BSA), according to published methods (8, 28). The synthetic substrate $p$-nitrophenyl $\beta$-xylobioside ($p$NP-Xb) was described previously (4).

RESULTS

Histidine NMR Assignments. A striking feature of the $^1$H-NMR spectrum of CexCD is the presence of peaks with resonances downfield of $\sim 11$ ppm (Figure 1). Upon formation of a long-lived covalent glycosyl intermediate with 2FCb, the chemical shifts of several of these peaks are perturbed and three additional downfield signals are observed. With $^{15}$N-labeled protein samples, each of these peaks exhibits a splitting of $\sim 85-100$ Hz due to a one-bond $^1$H-$^{15}$N coupling and thus must arise from nitrogen-bonded imidazole, indole, or amide protons. In $^{15}$N-HSQC spectra, three of these $^1$H signals in apo-CexCD and six in 2FCb-CexCD show one-bond correlations to $^{15}$N nuclei with chemical shifts in the range of 160–190 ppm (Figure 2), indicative of a protonated imidazole nitrogen (29). In general, the labile histidine $^1$H are observed only if protected from HX due to hydrogen bonding and/or burial within the interior of a protein.

The downfield region of the $^1$H-NMR spectrum of CexCD also changes significantly with the noncovalent binding of xylobio-isofagomine ($K_i = 0.13 \mu M$ at pH 7 (11)) and xylobio-imidazole ($K_i = 0.15 \mu M$) (Figure 1). In addition to six histidine signals, confirmed from $^{15}$N-HSQC spectra of the $^{15}$N-labeled protein (not shown), one extra peak is observed with each inhibitor arising from a $^1$H not directly bonded to $^{15}$N. These are tentatively assigned to the protected $^{14}$N-bonded protons from the unlabeled xylobio-imidazole in its cationic form and the xylobio-isofagomine in either its charged or neutral form (30). In principle, such signals could arise from a buried phenolic or carboxylic proton in the protein or, much less likely on the basis of their chemical shifts, from a hydroxyl proton in either the protein or bound inhibitor. However, this seems unlikely since similar signals...
The multiple-bond experiment also defined the charge states and tautomeric forms of the imidazole rings (29). These conclusions were corroborated by considering the corresponding $^{13}$C chemical shifts of each residue (32). The $^1$H$_{\alpha}$ were then connected to the directly bonded $^{13}$C$_{\alpha}$ in a constant-time $^{13}$C-HSQC spectrum of the uniformly $^{13}$C-labeled protein, as well as in a $^{13}$C-CPMG-HSQC spectrum of CexCD selectively labeled with $^{13}$C-His (Figure 4). The CPMG pulse train in this latter experiment led to a small increase in the signal-to-noise ratio, particularly with His80, likely due to refocusing of conformational exchange broadening. For protected N$_{\alpha}$H and N$_{\beta}$H, assignments of protonated nitrogens were determined from a one-bond $^{15}$N-HSQC spectrum (Figure 2). A second parallel approach was undertaken using $^1$H--$^1$H NOESY spectra to assign signals from the protected $^1$H$_{\alpha}$ and $^1$H$_{\beta}$ on the basis of intra- and interresidue interactions with neighboring protons in the two forms of the protein. Finally, to confirm these assignments, spectra were also recorded with apo and inhibited forms of CexCD containing alanine substitutions at His80 or His205 (not shown). These assignments, summarized in Table S1 of the Supporting Information, have been deposited in the BioMagResBank as accession numbers 7264 and 7265.

**Histidine HX.** The observation of $^{15}$N-HSQC signals from the labile nitrogen-bonded protons of His85 and His107 in CexCD, plus those of His80 and His205 in the inhibited species (Figures 1 and 2), indicates that these protons are protected from rapid HX. However, none of these signals were observed in 1D spectra of apo- and 2FCb-CexCD recorded immediately after either form of the protein was transferred into D$_2$O buffer (pH 6.5 and 20 °C), demonstrating that their exchange lifetimes are all less than the ~5 min dead time of this approach. Accordingly, the CLEANEX-PM experiment (26) was used to measure rapid proton--proton HX. Detectable exchange in 2FCb-CexCD was observed for only H85 $^1$H$_{\alpha}$ and H205 $^1$H$_{\alpha}$, with pseudo-first-order rate constants of 16.5 ± 1 and 12.5 ± 0.7 s$^{-1}$, respectively (pH 6.5 and 10 °C). Thus, the lifetimes for the remaining labile imidazole protons, observed in the $^{15}$N-HSQC spectra of apo- and 2FCb-CexCD, must fall between the limits of these two experimental approaches (i.e., from 0.1 to 300 s). These limits correspond to HX ~10$^{-2}$--10$^{-5}$ fold slower than that expected for a nitrogen-bonded proton in a free imidazole under similar conditions (~1 ms (25)). By way of comparison, the highly protected $^1$H$_{\alpha}$ of a buried neutral histidine in Bacillus circulans xylanase exchanges with a lifetime of ~7 h at 30 °C and pH 7 (3,8), whereas the detectable $^1$H$_{\alpha}$ of His57 in a rat trypsin mutant exchanges with lifetimes of 2–12.5 ms in its apo and two inhibited forms at 5 °C and pH 5.2 or 10 (34).

**Histidine pK$_{\alpha}$ Values.** The pK$_{\alpha}$ values (or limits thereto) of the five histidines in apo- and 2FCb-CexCD were determined from the pH dependence of their $^{13}$C$_{\alpha}$ and $^1$H$_{\alpha}$ shifts (Figure 5 and Table 1). These measurements were carried out using selectively $^{13}$C$_{\alpha}$-His-labeled protein, for which well-resolved $^{13}$C-HSQC spectra could be obtained within minutes (Figure 4). This approach (35, 36) is significantly more sensitive than direct $^{15}$N detection (33, 37) or indirect two- or three-bond $^1$H$_{\alpha}$-$^{15}$N-H$^2$ correlation experiments (19). Although not needed with the 34 kDa CexCD, background deuteration should facilitate the applicability of this approach to higher molecular mass systems.
Over a pH range of 4.3–10.5 for apo-CexCD, only His114 ($pK_a = 8.1$) and His80 ($pK_a = 7.9$) exhibited measurable titrations. Similarly for 2FCb-CexCD, over a pH range of 2.8–10.1, only His114 ($pK_a = 8.1$) titrated. Thus, the remaining histidines, with pH-independent chemical shifts, have $pK_a$ values that are at least 1 unit outside these ranges (i.e., given that clear spectral changes should be seen at a pH of $pK_a \pm 1$). However, on the basis of $^{13}$C- and $^1$H-
shifts alone, one cannot unambiguously establish if a histidine is neutral or charged within these pH limits. For example, although the $^{13}$C and $^1$H shifts of a neutral histidine in an unstructured polypeptide do not change with pH, relative to its charged form, CexCD, neutral (see below) His85 has a $^1$H shift diagnostic of a charged group whereas protonated His205 has a $^{13}$C shift indicative of a neutral imidazole. These discrepancies reflect the additional local environmental effects on the NMR signals of each particular histidine.

Due to the extreme difference between the chemical shifts of a protonated and unprotonated imidazole nitrogen ($^1$N, 37, 40), the charge states of the histidines in apo- and 2FCb-CexCD can be determined unambiguously by consideration of their $^{15}$N assignments obtained at pH 6.5 (Figure 3). Thus, in apo- and 2FCb-CexCD, the non-titrating His85 is neutral and, on the basis of the observed pattern of $^2$JNH and $^3$JNH couplings, in the common N$^2$ tautomeric state. By contrast, His107 and His205 are both positively charged. Upon formation of 2FCb-CexCD, His80 adopts the N$^2$ tautomeric form at neutral pH, with its $^1$K value dropping from 7.9 to a value below the measured limit of 2.8.

The $^1$K values of the histidines in the complexes of CexCD with xylobio-imidazole and -isofagomine were not measured. However, at pH 6.5, their ionization states in both complexes are likely similar to that of 2FCb-CexCD, which is evident from multiple-bond $^{15}$N-HSQC spectra (not shown). Thus, the $^1$K of His80 appears to drop from 7.9 to values significantly below 6.5 upon noncovalent inhibition with the two xylobiose derivatives.

**DISCUSSION**

Using a combination of $^1$H, $^{13}$C, and $^{15}$N chemical shift information and qualitative two- and three-bond $^1$H-$^1$N correlation measurements, we obtained a comprehensive description of the pH dependence of the ionization and tautomerization behavior of the five histidines in CexCD. In summary, for the apoenzyme at its pH optimum of ~6, His107 and His205 are positively charged, His85 exists as the neutral N$^2$ tautomer, and His114 and His80 are titrating between predominantly positively charged imidazolium ion and neutral imidazole states. Upon formation of two noncovalent enzyme-inhibitor complexes or a random-coil polypeptide. As noted previously, the least conserved electrostatic interactions

<table>
<thead>
<tr>
<th>residue</th>
<th>form</th>
<th>sequence conservation (%)</th>
<th>accessible surface area (Å²)</th>
<th>structural features</th>
<th>HX protected</th>
<th>charge and tautomer at pH 6.5</th>
<th>$^1$K</th>
</tr>
</thead>
<tbody>
<tr>
<td>His80</td>
<td>apo</td>
<td>100</td>
<td>2.7/0/2.1</td>
<td>N$^{1}$ H-bond donor to buried D123 (2.9 Å)</td>
<td>H$^1$</td>
<td>neutral N$^{1}$</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>2FCb</td>
<td>0.8/0/0.0</td>
<td></td>
<td>N$^{2}$ H-bond acceptor from inhibitor 3-OH group (2.7 Å)</td>
<td>H$^1$</td>
<td>neutral N$^{2}$</td>
<td>&lt;2.8</td>
</tr>
<tr>
<td>His85</td>
<td>apo</td>
<td>71</td>
<td>22.9/0/2/11.8</td>
<td>3.4 Å to E127; 3.5 Å to R136; N$^{1}$ H-bond acceptor from buried water (3.0 Å)</td>
<td>H$^2$</td>
<td>neutral N$^{2}$</td>
<td>&lt;4.0</td>
</tr>
<tr>
<td></td>
<td>2FCb</td>
<td>26.5/0/3/14.6</td>
<td></td>
<td>N$^{1}$ H-bond donor to buried E52 (2.7 Å); N$^{2}$ H-bond donor to buried water (2.8 Å)</td>
<td>H$^1$, H$^2$</td>
<td>positive</td>
<td>&gt;10.4</td>
</tr>
<tr>
<td>His107</td>
<td>apo</td>
<td>60</td>
<td>0.0/0.0/0.0</td>
<td>N$^{1}$ H-bond donor to surface water (2.8 Å); N$^{2}$ H-bond donor to D64 (2.7 Å)</td>
<td>H$^1$, H$^2$</td>
<td>neutral</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>2FCb</td>
<td>28</td>
<td>47.6/19.7/0.0</td>
<td>N$^{1}$ H-bond donor to E52 (2.7 Å); N$^{2}$ H-bond donor to E235 (2.9 Å)</td>
<td>H$^1$, H$^2$</td>
<td>neutral</td>
<td>&gt;10.4</td>
</tr>
<tr>
<td>His114</td>
<td>apo</td>
<td>99</td>
<td>48.2/20.7/0.0</td>
<td>N$^{1}$ H-bond donor to D235 (2.9 Å); N$^{2}$ H-bond donor to E233 (2.9 Å)</td>
<td>H$^1$, H$^2$</td>
<td>neutral</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>2FCb</td>
<td>6.4/0/3.5</td>
<td></td>
<td>N$^{1}$ H-bond donor to inhibitor O5 (3.0 Å)</td>
<td>H$^1$, H$^2$</td>
<td>positive</td>
<td>&gt;10.4</td>
</tr>
<tr>
<td>His205</td>
<td>apo</td>
<td>99</td>
<td>6.2/0/0.0</td>
<td>N$^{2}$ H-bond donor to inhibitor O5 (3.0 Å)</td>
<td>H$^1$, H$^2$</td>
<td>neutral</td>
<td>&gt;10.1</td>
</tr>
<tr>
<td></td>
<td>2FCb</td>
<td>99</td>
<td></td>
<td>N$^{2}$ H-bond donor to inhibitor O5 (3.0 Å)</td>
<td>H$^1$, H$^2$</td>
<td>neutral</td>
<td>&gt;10.1</td>
</tr>
</tbody>
</table>

*Sequence conservation (identity) based on the alignment of 92 family 10 glycoside hydrolases from the SMART database (http://smart.embl-heidelberg.de). *Calculated with MolMol (70) for each imidazole ring (ring/H$^2$/H$^2$) using a probe radius of 1.4 Å. *From PDB entry 2EXO for apo-CexCD and PDB entry 1EXP for 2FCb-CexCD. Structural features are given for only 2FCb-CexCD if they differ from the apoprotein. *Observed in a 1D jump-and-return or 2D $^{15}$N-HSQC spectrum. *$^1$K limits are set by the range of pH conditions utilized for a given titration. Actual $^1$K values are likely at least 1 unit beyond these limits. Errors are ±0.1. *See Discussion for further clarification.
in a family of related proteins generally occur with surface residues and are associated with relatively unperturbed \( pK_a \) values (41, 42). Interestingly, the imidazole of His114 stacks against a corner formed by the orthogonal rings of Phe60 and Phe115. Shielding due to the net ring currents from these aromatic groups alters the \(^1\text{H}\) chemical shift of this histidine to a highly anomalous value of 3.79 ppm.

**Active Site Histidines**

His80. In apo-CexCD, the completely conserved and slightly exposed His80 lies within the active site cleft of the protein, donating its \(^3\text{H}\) in a hydrogen bond to buried Asp123. Along with electrostatic effects (\( \sim 4 \) Å) from the nucleophile Glu233 and Lys47, this interaction helps to increase the \( pK_a \) value of His80 to 7.9 but does not significantly protect its labile imidazole protons from HX. Remarkably, upon formation of the covalent glycosyl–enzyme intermediate with 2Fcb, the \( pK_a \) of this residue drops by more than 5 units to 2.8 and its \(^1\text{H}\) can now be observed in a \(^1\text{H}\)-NMR spectrum. Although not determined quantitatively, the \( pK_a \) of His80 also decreases in the presence of the xylobio-isofagomine and -imidazole inhibitors. In each of these three cases, the neutral His80 adopts the less common N\(^{31}\)H tautomer and thereby continues donating a hydrogen bond to Asp123, while accepting one via its unprotonated N\(^{\alpha}\)2 from the exocyclic 3-OH of the bound sugar. Previously, either the ionized form or the opposite tautomer of this histidine has been assumed in X-ray crystallographic analyses of inhibited Cex and related family 10 glycosyl hydrolases such that both His80 and Lys47 would be donating hydrogen bonds to the 3-O of the substrate (5, 10, 30, 43). Burial of His80 by the inhibitor and/or possible changes in the dynamic properties of the surrounding active site region of Cex (12) likely lead to an increased level of protection of the labile \(^{31}\text{H}\) from rapid solvent exchange.

The reduced \( pK_a \) of His80 in 2Fcb-CexCD may be a necessary consequence of the hydrogen bonding arrangement observed at that location. That is, upon covalent linkage to CexCD, the 3-OH group of the bound substrate becomes completely buried within the active site of the enzyme. Since the oxygen is accepting a hydrogen bond from Lys47, which is unambiguously in its -NH\(^3\) form with a \( pK_a \) of 9 (44), it must be donating a hydrogen bond to His80. For the lone pair of electrons on the imidazole N\(^{\alpha}\)2 of His80 to accept this hydrogen bond, the imidazole cannot also remain protonated at that position. The neutralization of nearby Glu233 due to its covalent linkage to the bound inhibitor and/or possible changes in local dielectric due to solvent exclusion may also disfavor His80 remaining positively charged in 2Fcb-CexCD. In the cases of the xylobiose-derived inhibitors, repulsion from the potentially ionized isofagomine or imidazole groups (30, 43) may contribute further to the reduced \( pK_a \) value of His80. Interestingly, in apo-CexCD, the \(^1\text{H}\) and \(^1\text{C}\) of His80 also exhibit small chemical shift changes corresponding to a titration with an apparent \( pK_a \) of \( \approx 3.7 \) (Figure 5). This may reflect the deprotonation of Asp123 to which His80 is hydrogen bonded, thereby forming a buried ion pair. Furthermore, such chemical shift changes are not seen with 2Fcb-CexCD, suggesting that the \( pK_a \) of Asp123 also changes along with that of His80. If so, this might imply a maintenance of net charge neutrality in this region of Cex by the formal transfer of the imidazolium proton of His80 to the carboxylate of Asp123 upon formation of the glycosyl–enzyme intermediate. We are currently attempting to address this hypothesis by directly measuring the \( pK_a \) values of the Asp and Glu residues in CexCD. However, detecting and assigning the pH-dependent \(^1\text{C}\) and \(^1\text{H}\) signals, respectively, from the large number of carboxyl side chains in this 34 kDa protein is a challenging task.

By thermodynamic linkage, the decrease in the \( pK_a \) value of His80 upon inhibitor binding necessitates that the contribution by His80 toward substrate binding will be favored with an increase in pH. Thus, solely on the basis of its \( pK_a \) values, at the pH optimum of CexCD (\( \approx 6 \)), the deprotonation of His80 upon formation of the glycosyl–enzyme intermed-
more, as shown in Figure 7, the pH dependence of $k_{\text{cat}}/K_m$. The activity profile of full-length H80A Cex toward pNP-Xb (●) is fit to a bell-shaped curve with apparent $K_a$ values of 4.6 and 6.6. The activity of WT Cex toward pNP-Cb (○) follows $pK_a$ values of 4.1 and 7.7 (reproduced from ref 6). Note that, with WT Cex, $k_{\text{cat}}/K_m$ is ~100-fold higher toward pNP-Xb vs pNP-Cb (4). Also, the pH dependence of $k_{\text{cat}}/K_m$ reflects ionization events within the free enzyme and thus should not differ between the two closely related neutral substrates.

ate leads to a free energy penalty of ~2.8 kcal/mol (45). This linkage should be manifest in a pH dependence of the kinetic parameters of Cex for its substrates or of the $K_a$ values for its inhibitors. However, interpreting such effects is difficult due to the complex network of additional ionizable groups (also with changing $pK_a$ values upon substrate binding) in the active site of this enzyme, including Asp123 and Glu233. Also, the $K_m$ value of Cex is dependent upon the kinetic parameters governing both the glycosylation and deglycosylation steps in its double-displacement mechanism.

The catalytic importance of His80 is confirmed by an ~2000-fold decrease in $k_{\text{cat}}/K_m$ from 1500 mM$^{-1}$ s$^{-1}$ for wild-type Cex to 0.72 mM$^{-1}$ s$^{-1}$ for the H80A mutant toward the synthetic substrate pNP-Xb (pH 7 and 37 °C) (46). Furthermore, as shown in Figure 7, the pH dependence of $k_{\text{cat}}/K_m$ differs significantly between the two species. In its simplest interpretation, the bell-shaped activity profile of wild-type Cex indicates that a group with an apparent acidic limb $pK_a$ value of ~4.1 in the free enzyme must be deprotonated for catalysis to occur, whereas another with an apparent basic limb $pK_a$ value of ~7.7 must remain protonated (8, 9). These groups can be assigned tentatively to the nucleophile, Glu233, and the general acid, Glu127, respectively. However, site-specific $pK_a$ measurements for the two catalytic side chains, as well as for all other ionizable active site residues (which may exhibit coupled microscopic protonation equilibria with Glu127 and Glu233), are required to confirm this conclusion (47). In particular, the observation that the NMR-determined $pK_a$ of 7.9 for His80 is similar to the basic limb value of ~7.7 raises the interesting possibility that the ionization of this histidine contributes to a kinetic $pK_a$ governing the pH-dependent activity profile of CexCD. In contrast to the WT enzyme, the $k_{\text{cat}}/K_m$ profile of the H80A mutant follows apparent $pK_a$ values of 4.6 and 6.2. The acid limb shift toward a higher pH implies that the apparent $pK_a$ value of Glu233 increased upon removal of a favorable electrostatic interaction with the nearby His80 (N$^2$=O' separation of ~4 Å). However, the concomitant decrease in the basic limb $pK_a$ value is the opposite of that expected for loss of a favorable interaction with Glu127 (N$^2$=O' separation of ~7 Å). Thus, additional structural and electrostatic changes within the tightly coupled active site of Cex must accompany the mutation of His80. Although difficult to interpret, these kinetic measurements nevertheless demonstrate that His80 plays a key role in establishing the pH-dependent activity of Cex.

His85. The moderately conserved His85 lies partly buried near the surface of the active site of the protein and in the proximity to the guanidinium group of Arg136, as well as the carboxyl of the general acid/base catalyst Glu127. His85 has a $pK_a$ value of ~2.8 and remains in neutral N=H tautomeric form under all pH conditions that were examined. It is somewhat surprising that the labile H$^2$ is sufficiently protected from HX to be detected by $^1$H-NMR, as within the available crystal structures of CexCD, it is exposed to the solvent. Flipping the imidazole ring of His85 about its C$^\alpha$=C$^\gamma$ bond would decrease the accessible surface area of the N=H group, while also allowing a very weak hydrogen bond (~3.3 Å) to the side chain of Glu127. However, this would leave a buried water with an unsatisfied hydrogen bond to a site currently identified in the crystallographically determined structures as N$^3$ of His85. The 2D jump-and-return echo NOESY spectrum does not help resolve this discrepancy since no inter-residue NOEs were observed to the H$^2$ of His85. The low $pK_a$ value of this histidine, which likely results from Arg136 charge repulsion, is also not fully expected given that Glu127, as well as Asp138, are adjacent to its imidazole ring. However, by analogy to the well-studied example of B. circulans xylanase (47), it is anticipated that in apo-CexCD, Glu127 has a $pK_a$ value [tentatively 7.7 by kinetic measurements as discussed above (9)] that is greater than the pH optimum of the enzyme so that it can serve as a general acid yet lower than this value in the glycosyl–enzyme intermediate where it functions as a general base. Thus, at least for apo-CexCD, Glu127 would be negatively charged only at elevated pH values, and any stabilization of His85 in a protonated state by this catalytic glutamate would be offset by the intrinsic $pK_a$ of the histidine, which favors the neutral form under basic conditions. In the end, the net effect of these complex pH-dependent interactions is to dramatically decrease the $pK_a$ of His85 in the apo and inhibited forms of CexCD.

Unfortunately, for unknown reasons, we were unable to produce Cex with the H85A substitution despite testing a variety of bacterial expression conditions. Also, we were unsuccessful at generating mutants of Cex encoding a glutamine or asparagine at this position. Thus, the contributions of His85 to the pH-dependent activity of Cex could not be investigated further through site-directed mutagenesis.

His205. In apo-CexCD, the highly conserved His205 bridges Asp235 and the nucleophile, Glu233. Surprisingly, although donating hydrogen bonds to both carboxylates, neither of its labile imidazole protons is protected significantly from HX. pH-dependent kinetic measurements suggest that Glu233 has an apparent $pK_a$ value of ~4.1 in the free enzyme (9), whereas the charge state of buried Asp235 has not yet been determined. These two hydrogen bonding interactions must help increase the $pK_a$ of His205 to > 10.4 such that the histidine is positively charged under all conditions that have been examined. Upon formation of the glycosyl–enzyme intermediate, the $pK_a$ value of His205 remains > 10.1, despite the loss of a favorable electrostatic interaction with Glu233 (now neutral in a glycosidic linkage).
This may be compensated by formation of a bifurcated hydrogen bond from its N-H to both Glu233 (2.9 Å) and O5 of the proximal sugar ring (3.0 Å). In the cases of the xylobiose-derived inhibitors, His205 also remains positively charged, with its N-H strongly hydrogen bonded (∼2.7 Å) to Glu233, while also in van der Waals contact with carbons in the isofagomine or imidazole rings. As discussed below, the ∼3 ppm downfield shift of His205's H2 in Xbiso-CexCD and XbIm-CexCD relative to 2FCb-CexCD (Figure 1) likely results from differences in the ionization state of the hydrogen bond acceptor, Glu233. It is highly probable that the nucleophilic glutamate remains negatively charged in Xbiso-CexCD and XbIm-CexCD, whereas it becomes neutral due to covalent modification in 2FCb-CexCD. As with His80, upon formation of either the noncovalent inhibitor complexes or the covalent glycosyl—enzyme intermediate, the labile imidazole protons of His205 become protected from exchange with water, due to solvent exclusion and/or a reduction in local active site dynamics (12).

Consistent with the intimate contact between His205 and the nucleophile Glu233, mutation of this histidine residue to an alanine or asparagine severely (∼10^4-fold) decreases the k_d/K_m value of Cex toward substrates such as pNP-Cb (28). As also discussed in a previous structural study of these mutants, substitution of His205 with a neutral side chain dramatically alters the pH-dependent activity profile of Cex such that no essential ionization is detected between pH ∼4 and 9 (28). While not readily interpretable, this verifies the catalytic importance of His205.

**Xylobiose-Derived Aza-Sugar Inhibitors**

Signals tentatively assigned to nitrogen-bonded protons in xylobio-isofagomine and -imidazole were also detected when these inhibitors were complexed with CexCD at pH 6.5. These protons are likely protected from rapid HX due to hydrogen bonding with active site residues, including Glu127, Gln203, and Glu233. In a detailed crystallographic analysis of four aza-sugar inhibitors bound to CexCD at pH 4.6, each was inferred to have a positively charged nitrogen (10). This appears to hold true for the xylobio-imidazole since a proton assigned to its ring nitrogen is observed in the 1H-NMR spectrum of XbIm-CexCD. Furthermore, this indicates that the pK_a of the bound xylobio-imidazole is likely elevated from its solution value of 5.9 (J. Wicki and S. G. Withers, unpublished observations) due to favorable active site electrostatic interactions. This conclusion is also consistent with the atomic-resolution X-ray crystallographic structure of cationic glucoimidazole bound to a family 5 endoglucanase at pH 5.5 (48), and as inferred from the pH-dependent inhibition of a family 1 β-glucosidase by glucoimidazole derivatives (49).

In contrast, the charge state of the xylobiose-isofagomine cannot be determined solely from the 1H NMR spectrum of Xbiso-CexCD. However, on the basis of a measured pK_a value of 8.8 in solution (J. Wicki and S. G. Withers, unpublished observations), this inhibitor should also be positively charged when bound to CexCD. If so, then both nitrogen-bonded protons of the isofagomine moiety likely have degenerate chemical shifts in the spectrum in Figure 1C. In a 1.05 Å resolution crystal structure of inhibited *Bacillus agaradhaerens* Cel5A at pH 5, cellobio-isofagomine was clearly protonated with electron density observed for two hydrogens bonded to its ring nitrogen (43). In an ∼1 Å resolution structure of *Streptomyces lividans* Xyn10a, which is highly similar to Cex, the charge state of bound xylobiose-isofagomine remained ambiguous at both pH 5.8 and 7.5. However, the ligand nitrogen was inferred to be positively charged on the basis of the pH dependence of its inhibition constant, K_i (30). Kinetic inhibition and isothermal titration calorimetry studies also indicated that isofagomine binds preferentially to a family 1 β-glucosidase in its conjugate acid form (50). In principle, if synthesized in a _15_N-labeled form, the protonation state of the isofagomine group in XbIso-CexCD could be determined unambiguously from its pH-dependent _15_N chemical shifts and _1H_MAS coupling patterns (44), thereby providing further insights into the mechanism of CexCD inhibition.

**Histidine pK_a Values and pK_a Calculations**

The pK_a value of a histidine in a random coil polypeptide is typically ∼6. In stark contrast, the pK_a values of the five histidine residues in CexCD vary from <2.8 to >10.4 and that of His80 drops from 7.9 to <2.8 upon formation of the glycosyl—enzyme intermediate. Abnormally low pK_a values (<3) have also been reported for histidine residues in proteins such as *B. circulans* xylanase (33), *B. agaradhaerens* xylanase (51), and subtilisin BPN' (52), whereas elevated pK_a values of 9.1 and 10.4 have been measured for salt-bridged histidine side chains in T4 lysozyme (45) and superoxide dismutase (53), respectively (see also ref 42 and http://www.jenner.ac.uk/PPD/). Thus, the acid dissociation constant of a histidine can vary by at least 10^8-fold depending upon its structural and electrostatic environment within a protein and can dramatically change along the reaction pathway of a single enzyme.

The accurate calculation of pK_a values of ionizable groups in proteins remains a theoretical and practical challenge. A summary of predictions using WHAT IF (54) and PropKa (55) for the five histidines in CexCD is presented in Table S2 of the Supporting Information. After optimization of hydrogen bonding networks, WHAT IF calculates pK_a values by solving the linearized Poisson—Boltzmann equation to describe the electrostatic interactions within a protein, followed by a Monte Carlo analysis to define the lowest-energy ionization states as a function of pH (54). In contrast, PropKa rapidly predicts pK_a values for ligand-free protein residues using empirical rules for desolvation, hydrogen bonding, and Coulombic interactions (55). Relative to a random coil polypeptide reference state, PropKa correctly predicted the signs of the pK_a changes (i.e., positive or negative) for His85, His114, and His205 in apo-CexCD yet tended to underestimate their magnitudes. In the case of His80 and His107, the signs of the changes were also incorrect. These errors may result from an overestimation of the effects of desolvation to favor the neutral imidazole ring. In the case of apo-CexCD, WHAT IF correctly predicted the signs of the pK_a changes for all histidines yet both over- and underestimated their magnitudes by up to 2 log units. For 2FCb-CexCD, WHAT IF correctly predicted a decrease in the pK_a value of His80, attributable to an increased level of desolvation and neutralization of Glu233 upon formation of a glycosyl—enzyme intermediate yet
significant underestimated the magnitude of the change. Unexpectedly, the pKₐ value of His205 was also predicted to drop dramatically due to a loss of stabilizing electrostatic interactions. These errors likely reflect the challenges in calculating the pKₐ values of a system of highly coupled ionizable groups, as found in the active site of Cex.

Nevertheless, both approaches provided valuable insights, discussed above, into the factors dictating the titration behaviors of the histidines in this enzyme. Conversely, the data from this study should facilitate the development of improved algorithms for protein electrostatic calculations.

**Histidine Chemical Shifts**

The chemical shifts of a histidine can also provide important insights into its structural and electrostatic environment within a protein. For example, using data measured herein for CexCD (Table S1 of the Supporting Information) and reported in the BioMagResBank, a striking linear relationship is found to exist between the ¹H and ¹⁵N chemical shifts of directly bonded ¹H-¹⁵N and ¹H-¹⁵N pairs in proteins due to a complex interplay of sequence and structural effects. This simple relationship may help provide qualitative insights into the environments and charge states of histidine residues in proteins, for example, by allowing approximate chemical shift predictions for ¹H¹⁰¹² nuclei, which are not detected due to rapid HX, based on ¹⁵N‐¹⁰¹² shifts measured from HMBC spectra. Further inspection of the data in Table S1 reveals that, within each fully protonated histidine in CexCD, the ¹H- and ¹⁵N- nuclei have chemical shifts downfield from those of the corresponding ¹H- and ¹⁵N- nuclei, respectively. This pattern, seen also for free histidine (56), however, does not hold for proteins in general (Figure 8A), thus precluding assignments based on shifts alone. Also, as noted previously (1), most positively charged histidines in CexCD have ¹H- and ¹⁵N- chemical shifts downfield of those from neutral histidines. However, His205 in 2FCb-CexCD stands as a noteworthy exception, with its ¹H- and ¹⁵N- shifts rather upfield at 10.22 and 164.3 ppm, respectively. An explanation of this anomaly is provided by a comparison to the well-studied serine proteases.

His205 is hydrogen bonded to the nucleophile Glu233 and Asp235, forming a Asp235-His205-Glu233 constellation that is strikingly similar to the catalytic (Asp/Glu)-His-nucleophile triad found in serine proteases (1), phosphatidylinositol-specific phospholipase C (57), cholinesterases (58, 59), and hydroxynitrile lyase (60). In each case, the bridging His N- is close to the activated nucleophile, which can be Ser, Glu, Asp, or part of a substrate, while the N- forms a strong hydrogen bond to the carboxyl group of a second Glu or Asp. This latter interaction, best described as a short ionic hydrogen bond (61), rather than a low-barrier hydrogen bond (62), typically leads to a HX-protected His H- with a diagnostic downfield resonance. Recent quantum mechanical calculations (63, 64) demonstrated that these downfield proton chemical shifts are a consequence of lengthening of the N-–H- bond, combined with its polarization due to hydrogen bonding to a negatively charged carboxyl in the catalytic triad. Consistent with these calculations, the most downfield-shifted NMR signals in CexCD appear to arise from ¹H- and ¹⁵N- or ¹H- and ¹⁵N- pairs involved in hydrogen bonds with negatively charged carboxylate acceptors (i.e., His107-Glu53; Figure 8A and Table S1 of the Supporting Information). Furthermore, the ¹H- and ¹⁵N- atoms of His205 are unusually upfield shifted in 2FCb-CexCD, where its hydrogen bond acceptor Glu233 is neutralized in a glycosidic linkage, yet more typically downfield in apo-CexCD, XbIm-CexCD, and XbIso-CexCD where Glu233 most likely remains negatively charged. Interestingly, the ¹H- and ¹⁵N- chemical shifts of His80 in inhibited 2FCb-CexCD are also relatively upfield, which by this argument is indicative of hydrogen bonding to Asp123 in its neutral state, as hypothesized above.

**Histidine ¹J_HN Coupling Constants**

To characterize further the histidines in CexCD, we attempted to measure ¹H-¹⁵N- coupling constants between the imidazole ¹⁵N- and the side chain carboxyl ¹⁵C- of their Asp/Glu hydrogen bond partners. These ¹J_HN values are related to the covalent character of the hydrogen bond and the N=O distance (65). Unfortunately, the short ¹⁵N transverse relaxation times of this 34 kDa ²H- ¹³C- and ¹⁵N-
labeled protein precluded these measurements. However, as a function of both N–H (66) and donor-acceptor distances (67), the more easily determined \(^{1}J_{NH}\) coupling constants also provide insights into the nature of a hydrogen bond.

The observed \(^{1}J_{NH}\) coupling constants of the histidines in CexCD vary from \(\sim 85\) to \(100\) Hz and show an inverse correlation with their \(^{15}N\) and \(^{1}H\) (Figure 8b) and \(^{1}H\) (not shown) chemical shifts. Similar correlations were found for the OHN hydrogen bonds of pyridine acid complexes (68) and the NHN hydrogen bonds of DNA (69). This correlation is consistent with the above-mentioned quantum mechanical calculations (64), attributing a downfield \(^{1}H\) chemical shift to the lengthening of the N–H bond, which in turn leads to a smaller \(^{1}J_{NH}\) coupling constant (66). This is most readily seen with His107, where the downfield-shifted \(^{1}H\) (Table S1), which hydrogen bonds to Glu52, has a \(^{1}J_{NH}\) significantly smaller than that of the upfield-shifted \(^{1}H\) (Table S1), which hydrogen bonds to a buried water \(\sim 91\) Hz vs \(99\) Hz (Table S1 of the Supporting Information).

Concluding Remarks

By providing residue-specific \(pK_a\) values, NMR spectroscopy can yield detailed insights into the electrostatic bases for enzymatic catalysis. Through steady advances in NMR instrumentation and methodology, it is now possible to investigate in detail higher-molecular weight systems, such as enzymatic catalysis. Through steady advances in NMR instrumentation and methodology, it is now possible to investigate in detail higher-molecular weight systems, such as enzymatic catalysis. Through steady advances in NMR instrumentation and methodology, it is now possible to investigate in detail higher-molecular weight systems, such as enzymatic catalysis. Through steady advances in NMR instrumentation and methodology, it is now possible to investigate in detail higher-molecular weight systems, such as enzymatic catalysis.

SUPPORTING INFORMATION AVAILABLE

Chemical shifts and \(^{1}J_{NH}\) couplings (Table S1) and predicted \(pK_a\) values (Table S2) for the histidine residues in apo-CexCD and inhibited CexCD. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

22. Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1993) NMR experiments for correlating \(^{13}C\) and \(^{1}H\) chemical shifts of
aromatic residues in \(^{13}\)C-labelled proteins via scalar couplings, J. Am. Chem. Soc. 115, 11054–11055.


bond in the active site of hydroxynitrile lyase from *Hevea brasiliensis* explains a large pKa shift of the catalytic base induced by the reaction intermediate, *J. Biol. Chem.* 279, 3699–3707.


BI700249M