Prediction and Rationalization of the pH Dependence of the Activity and Stability of Family 11 Xylanases†

Jacob Kongsted,‡ Ulf Ryde,‡ James Wydra,§ and Jan H. Jensen*,‖

Department of Theoretical Chemistry, Chemical Center, University of Lund, S221 00 Lund, Sweden, Department of Chemical and Biochemical Engineering, University of Iowa, Iowa City, Iowa 52242, and Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

Received August 14, 2007; Revised Manuscript Received September 14, 2007

ABSTRACT: This paper presents a study of the pH dependence of the activity and stability of a set of family 11 xylanases for which X-ray structures are available, using the PROPKA approach. The xylanases are traditionally divided into basic and acidic xylanases, depending on whether the catalytic acid is hydrogen bonded to an Asn or Asp residue. Using X-ray structures, the predicted pH values of optimal activity of the basic xylanases are in the range of 5.2−6.9, which is in reasonable agreement with the available experimental values of 5−6.5. In the case of acidic xylanases, there are only four X-ray structures available, and using these structures, the predicted pHs of optimal activity are in the range of 4.2−5.0, compared to an observed range of 2−4.6. The influence of dynamical fluctuations of the protein structure is investigated for *Bacillus agaradhaerens* and *Aspergillus kawachii* xylanase using molecular dynamics (MD) simulations to provide snapshots from which average values can be computed. This decreases the respective predicted pH optima from 6.2−6.7 and 4.8 to 5.3 ± 0.3 and 4.0 ± 0.2, respectively, which are in better agreement with the observed values of 5.6 and 2, respectively. The change is primarily due to structural fluctuations of an Arg residue near the catalytic nucleophile, which lowers its pK_a value compared to using the X-ray structure. The MD simulations and some X-ray structures indicate that this Arg residue can form a hydrogen bond to the catalytic base, and it is hypothesized that this hydrogen bond is stabilized by an additional hydrogen bond to another Glu residue present only in acidic xylanases. Formation of such a hydrogen bond is predicted to lower the pH optimum of *A. kawachii* xylanase to 2.9 ± 0.3, which is in reasonable agreement with the observed value of 2. The predicted pH of optimal stability is in excellent agreement with the pH value at which the melting temperature (T_m) is greatest. Some correlation is observed between the pH-dependent free energy of unfolding and T_m, suggesting that the thermostability of the xylanases is partly due to a difference in residues with shifted pK_a values. Thus, the thermostability of xylanases (and proteins in general) can perhaps be increased by mutations that introduce ionizable residues with pK_a values significantly lower than standard values.

Family 11 xylanases catalyze the random endohydrolysis of 1,4-β-D-xylosidic linkages in xylan, the second most abundant polysaccharide in nature. Family 11 xylanases are widely used in baking and industrial processes, such as paper bleaching, and there is great deal of interest in engineering xylanases that work at extreme pHs or temperatures (*1*). A significant amount of experimental data has been generated over the past two decades, including X-ray geometries of 16 different family 11 xylanases and pK_a determinations of all Asp, Glu, and His residues for two different xylanases. The latter is made possible by the fact that family 11 xylanases are small enough (∼200 residues) to be studied by NMR, which is relatively rare for industrially important enzymes, and the family 11 xylanase from *Bacillus circulans* has one of the best characterized enzymatic mechanisms due to the pioneering work of Withers and McIntosh (*2*−*4*). This makes family 11 xylanases an ideal case for validation of new theoretical methods for the prediction of pH effects on enzymatic activity and stability, such as the work by Nielsen, Withers, and McIntosh on *B. circulans* xylanase (*2*).

Here we present the prediction, validation, and rationalization of the pH dependence of the activity and stability of most of the family 11 xylanases for which X-ray structures are available. The paper is organized as follows. First, we present the computational methodology used to predict the pH dependence of activity and stability, including a brief review of the PROPKA method used to predict the pK_a values. Second, we validate the predicted pK_a values and pH−activity profiles for the two most well-studied xylanses, from *B. circulans* (including mutants) and *Bacillus agaradhaerens*, and present predictions or rationalizations for the remaining xylanases. Third, we discuss the effect of including molecular dynamics on a few of the proteins. Fourth, we

† This work was supported by the U.S. National Science Foundation (J.H.J., MCB 209941), a Skou Fellowship (J.H.J.) from the Danish Research Agency (Forskningsrådet for Natur og Univers), a postdoctoral fellowship from the Villum Kann Rasmussen Foundation (J.K.), and the Swedish Research Council (U.R.).

* To whom correspondence should be addressed. E-mail: jhjensen@kemi.ku.dk. Phone: +45 3532 0239. Fax: +45 3532 2014.

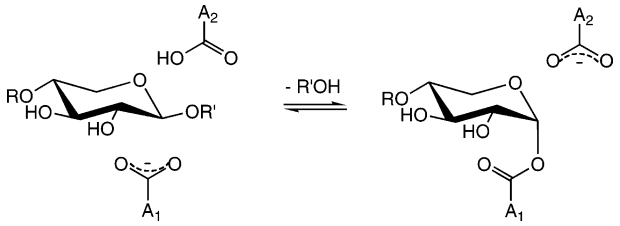
‡ University of Lund.

§ University of Iowa.

‖ University of Copenhagen.

10.1021/bi7016365 CCC: $37.00
Published on Web 10/26/2007

Scheme 1



discuss some family xylanases that have or may have usual pK_a determinants. Fifth, we present predictions of the pH dependence of protein stability, including some relatively limited validation against available experimental data. Finally, we summarize our results and present some suggestions for future experimental work on family 11 xylanases.

COMPUTATIONAL METHODOLOGY

Predictions of pH−Activity Optima. The pH−activity profile of family 11 xylanases is typically bell shaped, which is consistent with a double-displacement hydrolysis mechanism involving a catalytic nucleophile (A_1^-) and a general acid (HA_2), which are usually both Glu residues (Scheme 1).

The assumption is therefore that only one protonation state of the catalytic residues of the enzyme ($^-A_1A_2H$) contributes to catalysis and that the activity is related to the pH-dependent fraction of this protomer

$$\frac{[^-A_1A_2H]}{[A]_0} = \frac{1}{1 + 10^{pK_{a1}-pH} + 10^{pH-pK_{a2}}} \quad (1a)$$

$$\approx (1 - \theta_1)\theta_2 \quad (1b)$$

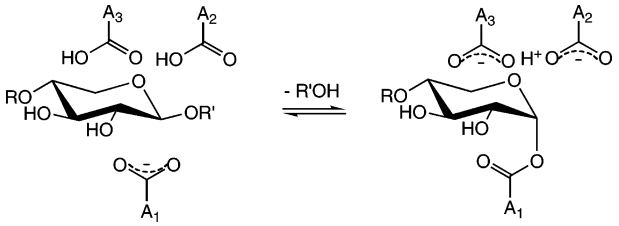
where $[A]_0$ is the total protein concentration and θ_i is the protonation probability of site i $[(1 + 10^{pH-pK_{a,i}})^{-1}]$. The approximation is valid when pK_{a1} is more than 1 pH unit higher than pK_{a2}. The apparent pK_a values of the two sites can be extracted from the pH−activity profile by fitting to the following functional form (*3*):

$$\left(\frac{k_{cat}}{K_M}\right)_{obs} = \left(\frac{k_{cat}}{K_M}\right)_{max}\left(\frac{1}{1 + 10^{pK_{a1}-pH} + 10^{pH-pK_{a2}}}\right) \quad (2)$$

From this, it is apparent that the maximum activity occurs at pH = (pK_{a1} + pK_{a2})/2. Furthermore, in cases where eq 1b is valid, the activity is ≥50% of the maximum value in the pH range from pK_{a1} to pK_{a2}. Equations 1 and 2 are derived assuming a low substrate concentration compared to K_M, and the pK_a values are those of the enzyme free of substrate. In some of the pH−activity studies cited below, the substrate is structurally very diverse (xylan from wood) and the exact substrate concentration is not always well defined. This can contribute to the discrepancy between prediction and experimental observation. Thus, we use X-ray structures obtained without ligands to predict pK_a values whenever possible.

Withers, McIntosh, and co-workers (*3*) have convincingly argued that in enzymes where A_2H forms a hydrogen bond to another acidic group [$A_2H\cdots A_3H$ (Scheme 2)] pK_{a2} (as defined in eq 2) actually is the pK_a value of A_3H, which is lower than pK_{a1}.

Scheme 2



Predictions of the pH Dependence of Protein Stability. The free energy of unfolding, in kilocalories per mole at 298 K and a given pH, can be expressed as (*5*, *6*)

$$\Delta G_U(pH) = \Delta G_U(pH_{ref}) + 1.36\int_{pH_{ref}}^{pH}(Q_U - Q_F)\,dpH$$

$$= \Delta G_U(pH_{ref}) + \Delta\Delta G_U(pH,pH_{ref}) \quad (3)$$

where $\Delta G_U(pH_{ref})$ is the unfolding free energy at a reference pH, Q_F and Q_U are the total charge of the protein of the folded and unfolded state, respectively, and $\Delta\Delta G_U(pH,pH_{ref})$ is the pH-dependent change in free energy of unfolding due to the pH change. In this study, we compute only the latter term in the pH range of 0−13 (with a pH increment of 0.01 pH unit) to determine the pH at which $\Delta G_U(pH)$ is a maximum, i.e., the pH of optimal stability. The pK_a values used to compute Q_U are the pK_{model} values used by PROPKA. Several studies have shown that the pK_a values of unfolded proteins can still be shifted relative to the model values, but it is believed that this has a relatively small effect on the pH value at which $\Delta G_U(pH)$ is a maximum (*7*−*10*). Thus, the predicted pH of optimal stability for two proteins can be compared with some degree of confidence.

Becktel and Schellman's (*11*) "law of small perturbations on protein stability" states that the change in T_m[1] due to a small perturbation such as mutations or changes in pH is given by

$$\Delta T_m = \frac{\Delta\Delta G_U}{\Delta S_{U,T_m}} = \frac{\Delta\Delta G_U T_m}{\Delta H_{U,T_m}} \quad (4)$$

where $\Delta\Delta G_U$ is the change in the free energy of unfolding due to the perturbation and $\Delta S_{U,T_m}$ and $\Delta H_{U,T_m}$ are the enthalpy and entropy of unfolding of the unperturbed protein at its T_m, respectively. The underlying assumption here is that the perturbation is sufficiently small that $\Delta S_{U,T_m}$ values for the unperturbed and perturbed state are similar. Thus, a change in T_m due to an increase in pH is expected to depend linearly on the pH-dependent free energy defined in eq 3.

$$\Delta T_m = \frac{\Delta\Delta G_U(pH,pH_{ref})}{\Delta S_{U,T_m}(pH_{ref})} \quad (5)$$

Furthermore, if $\Delta\Delta G_U$ values for two different proteins are entirely pH-dependent, and their $\Delta S_{U,T_m}$ values are similar at a given pH, then the difference in T_m is proportional to the difference in $\Delta\Delta G_U(pH,pH_{ref})$. However, $\Delta S_{U,T_m}$ could be very different for different proteins due to effects such

[1] Abbreviations: AMBER, Assisted Model Building with Energy Refinement; MD, molecular dynamics; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; pI, isoelectric point; T_m, melting temperature.

as configurational entropy, hydrophobic interactions, and disulfide bridges (*6*).

An alternative rationalization of protein stability, due to Linderstrom-Lang (*12*), is that the pH of maximum stability occurs at the isoelectric point (pI, where $Q_F = 0$), where the repulsive and attractive interactions between charge groups are optimized. However, this correlation does not seem to born out in practice (*6*, *8*, *13*). For example, in a recent computational study of 43 proteins, Alexov (*8*) found no correlation been the pH of optimal stability and pI.

pK_a Prediction Using PROPKA. PROPKA 1.0 (*14*) was used to predict the pK_a values used in this study. PROPKA utilizes a very fast empirical method to predict pK_a values and is successful at predicting unusual pK_a values. Recently, a comparative study of several protein pK_a prediction methods showed that PROPKA was the most accurate method overall (*15*). The program computes the pK_a perturbations due to desolvation, hydrogen bonding, and charge−charge interactions. In the current version of PROPKA, contributions from heteroatoms such as bound ions or transition metals to the pK_a values are not included. While the program is accessible via a web interface (http://propka.ki.ku.dk), we used an off-line version that automatically generates different structures from Protein Data Bank (PDB) files with fractional occupancy. This version (together with the scripts used to compute the pH-dependent free energy) is distributed freely under the GNU GPL and can be obtained by contacting the corresponding author.

Protein Structures. The PDB entries for the crystal structures used in this study can be found in Tables 2−4. Any heteroatoms such as the ligand in 1QH7 or the saccharide covalently attached to Asn7 in 1MW4 are removed prior to the pK_a calculations. Occasionally, the X-ray structure contains more than one protein chain in the asymmetric unit cell and/or one or more residues with fractional occupancy. In these cases, we construct individual protein structures for each chain and/or occupancy and compute pK_a values for each structure. For example, the 1QH7 crystal structure contains two chains (A and B) in the asymmetric unit cell, and both chains have several side chain conformers with fractional occupancy (1 and 2). In principle, separate structures can be constructed for an individual rotamer in each chain. However, to reduce the number of structures and simplify the analysis, we consider only the subset in which *all* residues are either in the first or in the second conformation listed in the PDB file for each chain. Thus, the 1QH7 structure gives rise to four separate structures (A1, A2, B1, and B2) and four sets of predicted pK_a values.

Molecular Dynamics Simulations. Alexov (*16*) has shown that structural fluctuations of charged side chain orientations can have a significant effect on pK_a predictions. To probe dynamical effects on the predicted pK_a values, we have performed classical molecular dynamics (MD) simulations for *B. agaradhaerens* and *Aspergillus kawachii* xylanase. From these MD simulations, a number of snapshots have been extracted and the pK_a values have been calculated for each snapshot using PROPKA. The final estimate of the residue pK_a value was obtained by averaging over the individual pK_a values calculated from different snapshots. The MD simulations have been performed using the sander module in Amber 8.0 (*17*) employing the ff03 (*18*) force field. This force field includes polarization implicitly through enhanced values of the partial point charges (as compared to the corresponding charges obtained for the isolated molecule). The initial geometry of the protein was taken as the X-ray structure from the Protein Data Bank. The protein was solvated in an octahedral water box extending 10 Å outside the protein. Electrostatic interactions were treated using the particle-mesh Ewald method (*19*). If not stated otherwise, all Asp and Glu residues were negatively charged and all Lys and Arg residues were positively charged. All other residues were assumed to be neutral, with His residues protonated at the δ site. Note that Amber automatically neutralizes the system by adding a uniform neutralizing plasma. All simulations made use of the SHAKE (*20*) algorithm to constrain bonds involving hydrogens. The temperature was kept constant at 300 K using the Berendsen weak-coupling algorithm (*21*) with a time constant of 1 ps. The time step in the MD simulaiton was 2 fs, and a nonbonded cutoff of 8 Å was employed. All simulations made use of periodic boundary conditions. The MD simulations were conducted using the following protocol: (i) energy minimization by 1000 steps keeping all atoms except hydrogen atoms and water molecules restrained to their position in the crystal structure using a force constant of 418 kJ mol^{-1} Å^{-2}, (ii) a 20 ps MD simulation with constant pressure and a reduced restraining force constant, (iii) a 50 ps equilibration with constant pressure and no restraints, (iv) a 200 ps MD equilibration with a constant volume, and (v) a production MD with either 400 ps or 1.2 ns in which snapshots were saved every 10 ps.

Miscellaneous. Sequence alignments and computations of percent similarity were performed with the Clustal W web interface (*22*). The display and overlay of structures were performed with SwissPDB (*23*).

RESULTS AND DISCUSSION

B. circulans Xylanase. Xylanase from *B. circulans* is the most well-studied of the xylanases considered here. The pK_a values of all Asp, Glu, and His residues have been measured by NMR (*24*, *25*) and are listed in Table 1 together with the PROPKA predictions based on the 1XNB structure. The largest error is 0.8 pH unit (His156), and the pK_a values of the catalytic residues (Glu78 and Glu172) are 5.1 and 7.3, respectively, in good agreement with the experimental values of 4.6 and 6.8, respectively. The latter values are identical (with an experimental error of 0.1 pH unit) with the apparent pK_a values extracted from the pH−activity profile (Table 2). Since the NMR measurements are conducted in the absence of substrate, this indicates that the substrate concentrations used in determining the pH optimum are sufficiently low compared to the K_M for this enzyme and that the apparent pK_a values extracted from the activity profile reflect that of the free enzyme (cf. eq 2). The predicted pH optimum of 6.2 is therefore also in good agreement with the experimental value of 5.7.

The main pK_a determinants identified by PROPKA are summarized in Figure 1. The pK_a values of both residues are elevated by 2.9−3.3 pH units due to desolvation; however, the pK_a of Glu78 is lowered by three hydrogen bonding interactions and an attractive charge−charge interaction with Arg112, while the pK_a value of Glu172 is lowered by only two hydrogen bonding interactions. Thus, the pK_a

Table 1: Comparison of Experimental and Predicted pK_a Values[a]

B. circulans			*B. agaradhaerens*					
					PROPKA			
residue	exp pK_a	PROPKA	residue	exp pK_a	A1	A2	B1	B2
Asp4	3.0	2.4	Asp5	3.8	3.1	3.1	3.9	3.9
Asp11	2.5	2.0	Asp12	3.9	2.8	2.8	3.1	3.1
Asp83	<2.0	1.4	Asp15	3.4	2.2	2.2	2.1	2.1
Asp101	<2.0	1.5	Asp21	3.5	3.0	2.2	3.3	3.3
Asp106	2.7	3.0	Asp90	3.9	3.8	3.8	3.9	3.9
Asp119	3.2	3.5	Asp99	<2.7	0.9	0.9	−0.1	−0.1
Asp121	3.6	3.9	Asp118	<2.7	2.2	2.2	2.0	2.0
Glu78	4.6	5.1	Asp123	<2.7	1.2	1.2	1.1	1.1
Glu172	6.7	7.3	Glu17	4.3	1.7	1.6	1.6	1.7
His149	<2.3	1.9	Glu56	4.6	4.1	4.1	4.1	4.1
His156	~6.5	7.3	Glu94	3.9	4.9	4.9	5.4	5.8
			Glu126	4.5	3.8	3.8	4.0	4.0
			Glu167	3.6	3.8	3.9	2.6	2.7
			Glu178	4.1	−0.5	−0.5	−0.5	−0.5
			Glu184	6.5	7.4	7.6	7.5	7.2
			His11	6.1−6.8	6.7	6.7	6.8	6.8
			His32	6.7	6.5	6.5	6.5	6.5
			His60	4.1	4.1	4.1	4.1	4.2
			His162	<2.7	1.6	1.6	1.5	1.5

[a] The X-ray structure of *B. agaradhaerens* contains two crystallographically independent protein chains (A and B) and two alternative conformations (1 and 2) for several residues, and we calculate the pK_a values for each separately.

Table 2: pH Optima (pH opt) and Associated pK_a Values Obtained Experimentally and Predicted by the PROPKA Method [(pK_{a1} + pK_{a2})/2] for *B. circulans* Xylanase Wild Type and Mutants[a]

	experimental			PROPKA		
catalytic residues	pH opt	pK_{a1}	pK_{a2}	pH opt	pK_{a1}	pK_{a2}
wild type E78 and E172 (1XNB)	5.7	4.6 (4.6)	6.8 (6.7)	6.2	5.1	7.3
Y80F E78 and E172 (1HV0)	6.3	4.8 (5.0)	7.7 (7.9)	7.0	5.4	8.5
Q127A E78 and E172 (1HV1)	5.7	4.1 (4.2)	7.3 (7.3)	6.4	5.7	7.2
Y69F E78 and E172 (2BVV)		(4.9)	(8.3)	6.6	5.9	7.2
N35D E78 and D35 (1C5H)	4.6	5.8 (5.7)	3.5 (3.7)	4.2	4.8	3.5

[a] pK_a values in parentheses were obtained from NMR titration curves for the active site residues indicated in the first column.

value of Glu78 is intrinsically lower than that of Glu172, and the deprotonation of Glu172 is further disfavored by a repulsive charge−charge interaction with Glu78.

The residues displayed in Figure 1 have all been subjected to site-directed mutagenesis (*2*). Here, we confine ourselves to analyses of mutants for which X-ray structures have been obtained (*2*) to avoid errors in the predicted mutant structures as possible sources of error. The results are summarized in Table 2. According to PROPKA, Tyr80 lowers the pK_a of Glu78 and Glu172 by 0.3 and 0.8 pH unit, respectively, in the wild-type (WT) structure. On the basis of the structure of the Tyr80Phe mutant, PROPKA predicts an increase in the corresponding pK_a values of 0.3 and 1.2 pH units, respectively. The reason for the greater increase in the pK_a value of Glu172 is that the Tyr80Phe mutant structure displays a longer Gly172−Asn35 hydrogen bond, which therefore lowers the pK_a value less than in the WT structure. The predicted increases in pK_a values agree well with the values determined by NMR (0.2 and 0.9, respectively) and extracted from the pH−activity profiles (0.4 and 1.1, respectively).

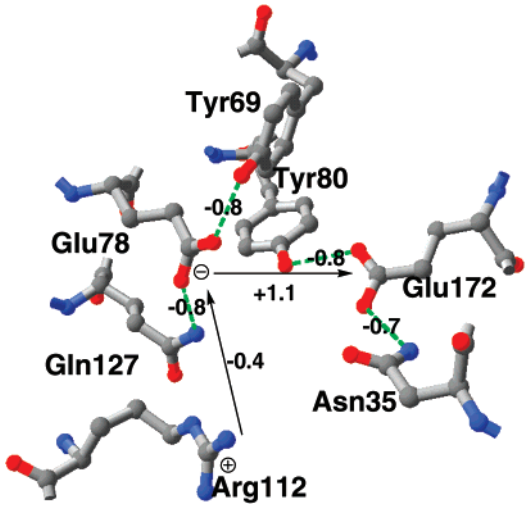


FIGURE 1: Key determinants of the pK_a values of the catalytic nucleophile (Glu78) and acid (Glu172) in *B. circulans* xylanase computed using the 1XNB X-ray structure.

On the basis of the WT structure, the mutations of Gln127 to Ala and Tyr69 to Phe are predicted to affect only the pK_a value of Glu78, both by +0.8 pH unit, and this compares well to the predictions based on the structure of the mutants of +0.6 and +0.8, respectively. However, in the case of the Gln127Ala mutant, the pK_a value of Glu78 is experimentally observed to decrease (−0.5) while the pK_a value of Glu172 increases by a corresponding amount. In the case of the Tyr69Phe mutant, the pK_a of Glu78 is seen to increase slightly (+0.3) while the pK_a value of Glu172 is increased significantly, to 8.3. The source of this discrepancy is not known, but it is interesting to note that the Tyr89Phe mutation leaves the positions of the crystallographic water molecules in the active site intact, while their number and position are significantly altered in the two other mutants. Mikami and co-workers (*26*) have previously implicated such a change in the microhydration to explain similarly puzzling results for *Bacillus cereus* β-amylase, and we plan to address this potentially important issue further in future studies. The effect, if any, of such water molecules on the pK_a values is currently not included in PROPKA.

Finally, the effect of mutating Asn35 to Asp on the pH optimum is predicted reasonably well: a decrease from 6.2 to 4.2, compared to the experimental values of 5.7 and 4.6. Here pK_{a2} refers to the pK_a value of Asp35, in accordance with the reverse protonation mechanism of Joshi et al. (*3*).

B. agaradhaerens Xylanase. Xylanase from *B. agaradhaerens* is the only other family 11 xylanase for which the pK_a values of all Asp, Glu, and His residues have been measured by NMR (*27*). These values are listed in Table 1 together with the PROPKA predictions based on the 1QH7 structure of *B. agaradhaerens* xylanase. This structure contains a β-D-xylanopyranoside molecule, but it was used because no crystal structure of the apo form is available. The asymmetric unit cell contains two protein chains (A and B), and both chains have several side chain conformers with fractional occupancy (1 and 2). Thus, the 1QH7 structure was used to construct four separate structures (A1, A2, B1, and B2) and four sets of predicted pK_a values, as described in Computational Methodology. The predicted values are all within 1.3 pH units of experiment, with the exception of those of Glu17 and Glu178, for which the deviation is

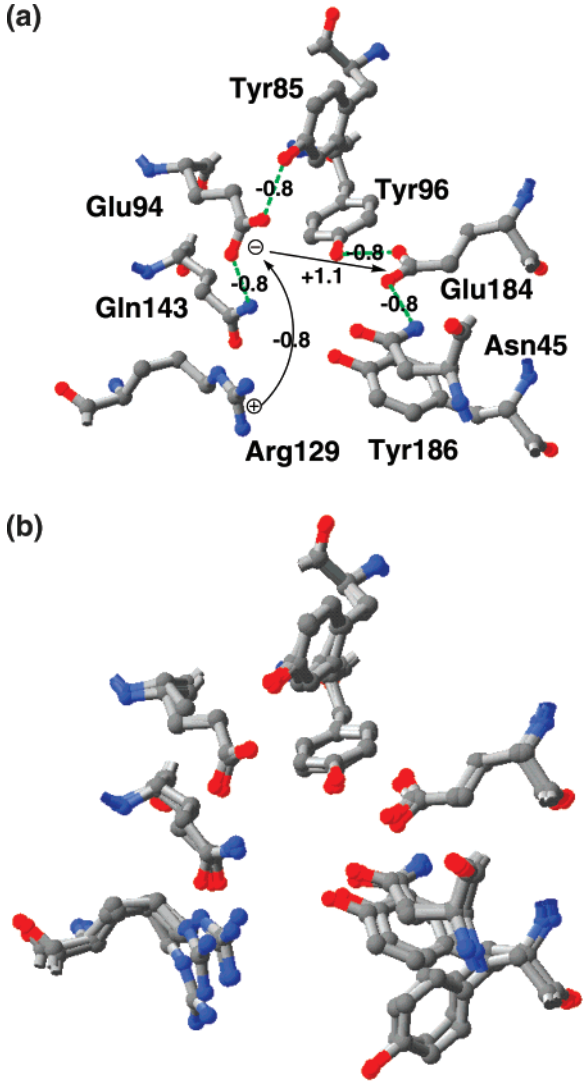


FIGURE 2: (a) Key determinants of the pK_a values of the catalytic nucleophile (Glu94) and acid (Glu184) in *B. agaradhaerens* xylanase computed using the 1QH7_A1 X-ray structure. (b) Overlay of the four structures (A1, A2, B1, and B2) generated from the 1QH7 PDB file (see the text for further discussion).

between 2.6 and 4.6 pH units. One possible explanation for this discrepancy is the structural influence of the ligand: Glu17 makes a direct hydrogen bond to the ligand, while the pK_a value of Glu178 is significantly lowered by Arg49, which forms two hydrogen bonds to the ligand. However, a ligand-free X-ray structure of a very similar xylanse from *Bacillus subtilitis* B230 (1IGO, 72% sequence identity) is available, and this structure gives rise to near-identical predicted pK_a values for Glu17 and Glu148, which lead us to consider alternate explanations for this discrepancy. It is interesting to note that the titration of Glu178 could not be monitored directly from its C^{δ} chemical shift but was inferred from chemical shift changes of a nearby amide group (*27*).

It is possible that this chemical shift change refers to titration of Glu17 (which is also nearby) and that Glu178 does not titrate in the pH range of 3−9 used in the experimental study, in agreement with the PROPA predictions. Glu17 exhibits an unusual π-stacking interaction with Arg49, which is also observed between Asp75 and Arg87 in the 1A2P X-ray structure of barnase. Using this structure, PROPKA predicts a pK_a value of −1.2 for Asp75 while the

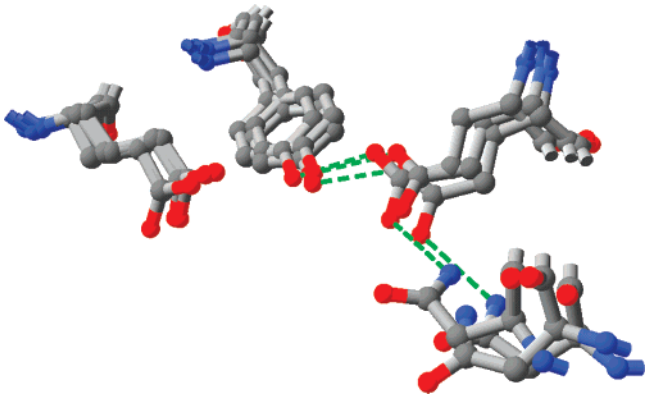


FIGURE 3: Overlay of the active site residues in *B. circulans* xylanse (1XNB), *Tr. reesei* xylanase II (1XYP), and *C. thermophilium* xylanase (1H1A_A1).

experimental value is 3.1, i.e., an underestimation of 4.3 (*14*), which is similar to the 2.6 pH unit error observed for Glu17. It is therefore possible that Asp/Glu-Arg π-stacking interactions require a special treatment in PROPKA or are especially sensitive to pH-induced structural changes. We will investigate this issue in future studies.

The pK_a predictions for the catalytic acid Glu184 are in the range of 7.2−7.6, which are in reasonable agreement with the value determined by NMR (6.5) and in good agreement with the value extracted from the pH−activity profile (*28*) [7.1 (Table 2)]. Interestingly, the pK_a predictions for the nucleophile Glu94 span a larger range (4.9−5.8). This variation is due to a difference in interaction with Arg129 in three of the four structures: −0.8, −0.8, −0.4, and 0.0 in structures A1, A2, B1, and B2, respectively, which is due to a difference in the position of the Arg129 residue (Figure 2b). The lower value, 4.9, is in better agreement with the value determined by NMR (3.9) and the value extracted from the pH−activity profile [4.2 (Table 2)]. This issue is considered further below. Overall, the residues that make key contributions to the pK_a values of the two catalytic residues in *B. agaradhaerens* xylanase are the same as for *B. circulans* xylanase (Figure 2a).

Other Basic Xylanases. The pH dependence of the activity has been reported for four other basic family 11 xylanses for which structures are available [Table 3; the pH optimum for *B. subtilitis* B230 was quoted by Oakley et al. (*29*), with reference to an Australian patent, and will not be considered further]: *Thermomyces langinosus* (*30*), *Streptomyces* S38 (*31*), *Trichoderma reesei* II (*32*), and *Nonomuraea flexuosa* (*33*). The pH optima for these enzymes are in the range of 5−6.5, which is in good agreement with the predicted range of 5.6−6.7. The apparent pK_a values are also in reasonable agreement with the experimental estimates, with the largest deviation being observed for pK_{a2} in *T. langinosus*: ∼7.5 versus a predicted value of 6.4. The underlying predicted pK_a values span a relatively narrow range for both pK_{a1} (4.7−5.8) and pK_{a2} (6.3−7.6). The spread in the former is caused primarily by variation in the pK_a shift induced by an Arg residue as discussed for *B. agaradhaerens*, while the spread in the latter is caused by the lack of a positive pK_a shift due to the nucleophile [e.g., Glu78 in *B. circulans* xylanse (cf. Figure 1)] in *B. agaradhaerens* and *Tr. reesei*. In the case of *T. langinosus* xylanase, this interaction is missing because Glu184 adopts a different conformation compared to, for

Table 3: pH Optima (pH opt) and Associated pK_a Values Obtained Experimentally and Predicted by the PROPKA Method [(pK_{a1} + pK_{a2})/2] for Basic Xylanases[a]

	experimental			PROPKA		
catalytic residues	pH opt	pK_{a1}	pK_{a2}	pH opt	pK_{a1}	pK_{a2}
B. circulans E78 and E172 (1XNB)	5.7	4.6 (4.6)	6.8 (6.7)	6.2	5.1	7.3
B. agaradhaerens E94 and E184 (1QH7)	5.6	4.2 (3.9)	7.1 (6.5)	6.2−6.7	4.8−5.8	7.2−7.6
T. langinosus E86 and E178 (1YNA)	6.5	∼5	∼7.5	5.7	5.0	6.4
Streptomyces S38 E87 and E177 (1HIX)	6.0	4.3	7.7	6.0	4.7	7.4
Tr. reesei II E86 and E177 (1XYP)	5−5.5	∼4	∼7	5.6	5.0	6.3
N. flexuosa E87 and E176 (1M4W)	5−6		∼7.5	6.3	5.2	7.4
B. subtilitis B230 E94 and E183 (1IGO)	8			6.2−6.3	5.3−5.5	7.0−7.3
B. subtilitis strain 168 E78 and E172 (1XXN)				6.5	5.2	7.8
D. thermophilium Rt E90 and E180 (1F5J)				5.2−6.0	3.3−4.9	7.1−7.2
P. varioti E86 and E178 (1PVX)				6.1	5.4	6.8
C. thermophilum E87 and E178 (1H1A)				5.5−6.9	2.9−5.8	7.0−8.0
T. harzanum E58 E86 and E177 (1XND)				6.0	4.7	7.3

[a] pK_a values in parentheses were obtained from NMR titration curves for the active site residues indicated in the first column.

example, *B. circulans* xylanse, which increases the Glu178−Glu86 distance. This conformation is very similar to that observed in the structure of *Tr. reesei* xylanase II obtained at pH 4.5 (1XYO) (*34*), consistent with the fact that the structure of *B. agaradhaerens* xylanase was obtained at pH 4.0. In the case of *Tr. reesei* xylanase II (1XYP), the large Glu177−Glu86 distance is due to a more "open" conformation compared to, say, that of *B. circulans* xylanase as discussed by Törrönen and Rouvinen (*34*).

The pH optima have not been reported for the remaining six basic family 11 xylanases in Table 3, so our computed pH optima may serve as predictions (*B. subtilitis* strain 168 xylanase differs from *B. circulans* xylanase by only one residue and is included only for completeness). The pH optima of these xylanases are in the range of 5.5−7.0 and do not differ significantly from the ones already discussed. However, the pK_a value of the nucleophilic residue is seen to vary considerably for *Dictyoglomus thermophilium* and especially for *Chaetomium thermophilium*. In the latter case, four separate structures (A1, A2, B1, and B2, as discussed for *B. agaradhaerens*) can be constructed leading to the following respective predictions for pK_{a1} and pK_{a2}: 5.8 and 8.0, 2.9 and 8.0, 5.1 and 7.0, and 5.1 and 7.0, respectively. As seen before, the spread in pK_{a1} is caused primarily by different conformations of a nearby Arg residue (Arg123), which in structure A2 forms a hydrogen bond with the catalytic nucleophile. The difference in pK_{a2} is caused by the fact that E178 has a different conformation in the A and B structures, where the A conformation resembles that observed in 1XNB (but without a hydrogen bond to an Asn group), while the B conformation resembles that of the observed in the structure of *T. langinosus* xylanase (1YNA).

In summary, the experimentally measured pH optima of the basic family 11 xylanses are in the range of 5−6.5, which is in reasonable agreement with the computed range of 5.2−6.9. The observed pK_a values of the nucleophile and catalytic acid are ∼4−5 and 6.5−7.7, respectively, which is in reasonable agreement with the computed values of 2.9−5.8 and 6.3−8.0, respectively. The spread in pK_a values are caused exclusively by differences in key distances rather than differences in the sequence, because the residues identified as key determinants for *B. circulans* and *B. agaradhaerens* xylanase (Figures 1 and 2) are conserved in all 12 basic xylanases discussed here, while the average degree of sequence similarity is 55% (excluding *B. subtilitis* strain 168 xylanse which differs from *B. circulans* xylanase by one residue). The differences in distance are caused either by different side chain conformations or by differences in the compactness of the protein. For example, the (sometimes slight) differences in the side chain conformation of an Arg residue (Figure 2b) modulate the pK_a of the nucleophile. The pK_a of the catalytic acid is modulated either by adoption of different a side chain conformation (e.g., observed in 1YNA and 1H1A_B1) or by changes in the compactness of the protein structure. For example, the pK_a values of Glu177, Glu172, and Glu178 computed with the 1XYP, 1XNB, and 1H1A_A1 structures are 6.3, 7.3, and 8.0, respectively. Compared to 1XNB, 1XYP lacks the repulsive interaction with the catalytic nucleophile, while 1H1A_A1 lacks the attractive interaction with Asn35 [*B. circulans* numbering (cf. Figure 1)]. However, the side chain conformations are all similar (Figure 3), so the differences in inter-residue distances are primarily due to different backbone positions, reflecting the compactness of the protein structure.

Acidic Xylanases. X-ray structures have been determined for four acidic family 11 xylanases, if one includes the N35D mutant of *B. circulans* xylanase (Table 4) (*3*, *32*, *35*−*37*). The observed pH optima vary from 2 to 4.6, while the predicted pH optima vary much less (4.2−5.0) since the key pK_a determinants and the active site environment are very

Table 4: pH Optima (pH opt) and Associated pK_a Values Obtained Experimentally and Predicted by the PROPKA Method [(pK_{a1} + pK_{a2})/2] for Acidic Xylanases[a]

	experimental			PROPKA		
catalytic residues	pH opt	pK_{a1}	pK_{a2}	pH opt	pK_{a1}	pK_{a2}
B. circulans N35D E78 and D35 (1C5H)	4.6	5.8 (5.7)	3.5 (3.7)	4.2	4.8	3.5
Tr. reesei I E75 and D37 (1XYN)	4.0−4.5			4.7	5.4	4.0
A. niger E79 and D37 (1UKR)	3−3.5			5.0	5.8	4.2
A. kawachii E79 and D37 (1BK1)	2			4.8	5.4	4.3

[a] pK_a values in parentheses were obtained from NMR titration curves for the active site residues indicated in the first column.

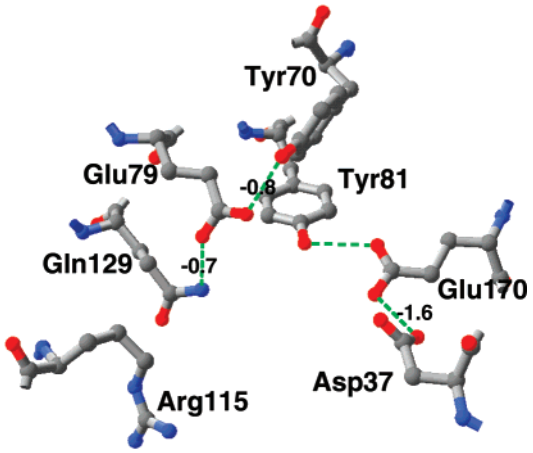


FIGURE 4: Key determinants of the pK_a values of the catalytic nucleophile (Glu79) and acid (Glu170) in *A. kawachii* xylanase computed using the 1BK1 X-ray structure.

similar for all four enzymes and also similar to those of the basic xylanases (see Figure 4 for a representative example). The predicted pH optima are thus in considerable error (2.0−2.8) for *Aspergillus niger* and *A. kawachii* xylanase. Part of the discrepancy may result from uncertainty in the experimental data. For example, *A. niger* and *A. kawachii* xylanase differ only in three amino acids located far from the active site. It is therefore not immediately clear why the observed pH optima differ by 1−1.5 pH units. In addition, the reported pH activity curves for these two proteins deviate significantly from the expected Gaussian shape (*35*−*37*). Honda et al. (*38*) have shown that the use of complex sugars as substrates can lead to a pH-dependent product inhibition that complicated the determination of the pH optimum in *Cellulomas fimi* family 10 xylanase. It would therefore be very interesting to determine the pH−activity profiles for the acidic xylanases with a synthetic substrate such as $ONPX_2$, which was done for the N35D mutant of *B. circulans* xylanase (*3*). On the other hand, it is known from the basic xylanase that the pK_a value of the catalytic portion can be affected significantly by the position of a nearby Arg residue. This issue is considered further in the next section.

Predictions Using Molecular Dynamics. As detailed above, many of the X-ray structures used in this study exhibit several possible conformations for some residues, notably Arg129 (*B. agaradhaerens* numbering) which results in different pK_a values for the catalytic nucleophile. To investigate this issue further, we performed 400 ps MD simulations starting from each of the four structures from 1QH7 (A1, A2, B1, and B2, described above). Here Glu184 is protonated, while all other residues are in their standard state. This resulted in the respective pK_a values for the catalytic nucleophile (4.4 ± 0.3, 2.9 ± 1.8, 4.4 ± 0.6, and 4.1 ± 0.6) and catalytic acid (7.6 ± 0.6, 6.2 ± 0.5, 6.6 ± 0.3, and 6.5 ± 0.7). The lowering of pK_{a1} is due to fluctuations in the conformation of Arg129, which occasionally shortens the Glu94−Arg129 distance considerably (Figure 5). In the MD simulation started using the A2 structure, a Glu94−Arg129 hydrogen bond is formed after 200 ps, which persists until the end of the simulation [which was continued for an additional 800 ps (data not shown)]. This lowers the average pK_a of the catalytic nucleophile significantly compared to those of the other MD simulations. Similar hydrogen bonds have been seen in the 1HA1 and 1F5J X-ray structures. The average over all four trajectories yields a pK_{a1} of 3.9 ± 0.6, identical to the experimental value measured by NMR and very similar to the apparent pK_a value extracted from the activity profile (4.2). This suggests that Arg129 has significant mobility in solution and can form a hydrogen bond with Glu94 in *B. agaradhaerens* xylanse, and presumably all family 11 xylanases. The Glu94−Arg129 hydrogen bond is probably transient on the basis of comparison with experimental pK_a values. Inclusion of the dynamical motion increases the accuracy of the PROPKA prediction for this residue and, therefore, of the pH of optimal activity: 5.3 ± 0.3 compared to 6.2−6.7 using X-ray structures (the observed value is 5.6).

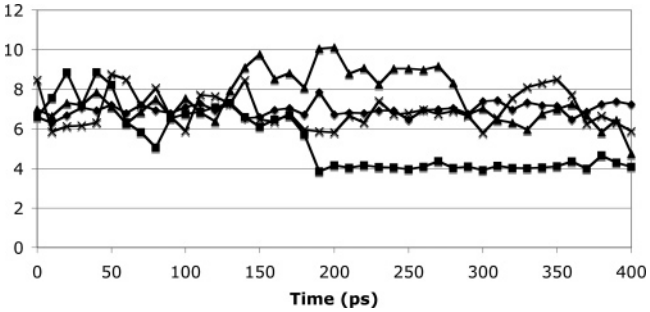


FIGURE 5: Plot of the distance between the CD atom in Glu94 and the CZ atom in Arg129 along MD simulations starting from the A1 (◆), A2 (■), B1 (▲), and B2 (×) structures generated from the 1QH7 PDB file (*B. agaradhaerens* xylanase).

The effect of dynamical motion was also studied for *A. kawacii* xylanase since the largest discrepancy between predicted and observed pH−activity optima is for this enzyme. Using 120 structures extracted from a 1.2 ns MD simulation starting from the 1BK1 X-ray structure resulted in predicted pK_a values of 4.2 ± 0.4 and 3.9 ± 0.2 for Glu79 and Asp37, respectively. The resulting pH of optimal activity (4.0 ± 0.3) is in somewhat better agreement with the observed value of 2 than when using the PDB structure (4.8). However, the error is still substantial. Notably, a Glu79−Arg115 hydrogen bond, which could lower the pK_a value of Glu79 considerably, was not observed during the 1.2 ns MD simulation (average Glu79−Arg115 distance of 8.2 ± 0.9 Å), but it is not clear whether this is due to insufficient sampling.[2] To investigate this issue further, a new starting

[2] In the first MD simulation, Glu170 (the catalytic acid) and Glu118 were protonated. Glu118 was protonated to weaken its interaction with Arg115 and, hence, increase the probability of an Arg115−Glu79 hydrogen bond. In the second MD simulation, Glu118 was deprotonated to further stabilize the hydrogen bond network.

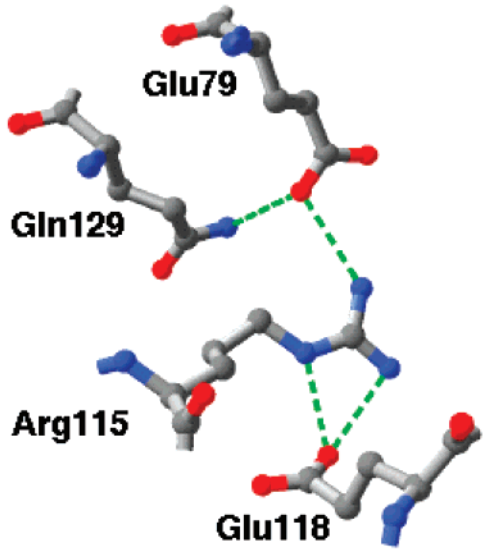


FIGURE 6: Local structure around Arg115 at $t = 400$ ps from the MD simulation initiated from a structure constructed from 1BK1 (see the text for further details).

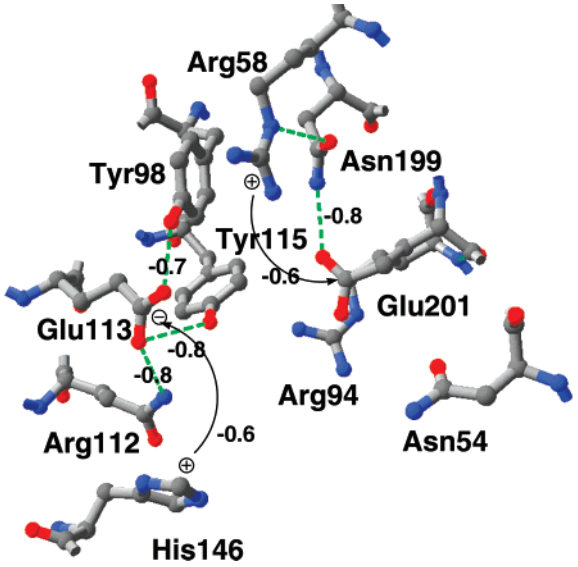


FIGURE 7: Key determinants of the pK_a values of the catalytic nucleophile (Glu113) and acid (Glu201) in *N. patriciarum* xylanase computed using the 2C1F X-ray structure.

geometry with a Glu79−Arg115 hydrogen bond was constructed from the X-ray structure by changing select dihedral angles in the Arg115 side chain (starting Glu79−Arg115 distance of 4.1 Å). Interestingly, this hydrogen bond persists during a 400 ps MD[2] simulation [average Glu79−Arg115 distance of 4.8 ± 0.5 Å (Figure 5)], resulting in pK_a values of 2.0 ± 0.6 and 3.7 ± 0.2 for Glu79 and Asp37, respectively. In this simulation, Glu118 was deprotonated. The resulting pH of optimal activity (2.9 ± 0.3) is in better agreement with the observed value of 2.

It is tempting to propose that an enzyme with a Glu79−Arg115 hydrogen bond is the prime contributor to the activity of acidic xylanases, especially since it does not involve reverse protonation (note that pK_{a1} < pK_{a2}). However, such a Glu−Arg hydrogen bond is presumably also possible in the N35D mutant of *B. circulans* xylanase, yet pK_{a1} is firmly established to be 5.8 for this enzyme. One possible explanation for why the Glu−Arg hydrogen bond is favored in acidic xylanases over the N35D mutant of *B. circulans* is the identity of residue $i + 3$, where i is the residue number of the Arg in question (Figure 6). This residue is Glu is the three acidic xylanases considered here, while it is either Ala or Gln in the basic xylanases. The presence of this Glu allows for a Glu−Arg$_i$−Glu$_{i+3}$ hydrogen bonding network (Figure 6), which might favor the Glu−Arg hydrogen bond further in acidic xylanases. Ultimately, NMR measurement of the pK_a values of the catalytic residues in *A. kawachii* xylanase is needed to resolve this issue. Alternatively, a measured pH optimum of an N35D/A115E mutant of *B. circulans* xylanase would also be interesting.

New and Needed X-ray Structures. During the course of this study, X-ray structures from two new basic xylanases were released in the Protein Data Bank: one from *Bacillus* sp. 41M-1 (2DCJ) and the other from *Neocallimastix patriciarum* (2C1F). *Bacillus* sp. 41M-1 xylanase shares a relatively high degree of sequence homology with *B. agarahaerends* (83%), and the key determinants of the pH optimum are similar in nature to those of the basic xylanases discussed so far (Figures 1 and 2). The predicted pH optimum is in the range of 4.8−6.3, where the variation stems mainly from different conformations of Arg128 observed in the X-ray structure.

N. patriciarum xylanase on the other hand has a low degree of sequence homology with the other basic xylanases; the highest degree of sequence homology (43%) is with *B. subtilis* B230 xylanase, and most interestingly, the Arg112 (*B. circulans* numbering) residue is replaced with a His residue at position 146. Furthermore, PROPKA reveals several new pK_a determinants for the catalytic acid (Glu201): a hydrogen bond to Asn199 rather than Asn54 (which corresponds to Asn35 in *B. circulans* xylanase) and charge−charge interactions with Arg94 and Arg58. Asn199, Arg94, and Arg58 correspond to Ala170, Tyr65, and Gly39 in *B. circulans* xylanase, respectively. Interestingly, Tyr95 and Gly39 are conserved in all other xylanase structures considered here, while Ala170 is conserved in all but three of the xylanase structures considered here: *B. agaradharens*, *B. subtilis* B230, and *D. thermophilium* xylanase which have Thr, Thr, and Cys residues, respectively, instead.

The predicted pK_a value of the catalytic nucleophile (Glu113) is 4.6, which is similar to that predicted for other basic xylanases, because His146 has a higher pK_a value than Glu113 and therefore performs essentially the same role as the usual Arg residue (Figure 7). However, the pK_a of the catalytic acid Glu201 is predicted to be 3.7, significantly lower than the normal values of basic xylanases, and lower than the pK_a of the catalytic nucleophile! Furthermore, the resulting pH optimum is 4.2, significantly lower than the observed value of 7.0 (*39*). Closer inspection of the structure reveals that that Arg94 exhibits a π-stacking interaction with Glu113, similar to that observed for the Glu17−Arg69 pair in *B. agaradhaerens* xylanase and the Asp75−Arg87 pair in barnase as discussed above. The errors in the pK_a values observed for Glu17 and Asp75 are 2.6 and 4.3 pH units, respectively, so it is reasonable to assume that the pK_a value of Glu201 is closer to 6.3−8.0, which results in a predicted pH optimum of 5.2−6.3, which is in better agreement with the experimental value.

More generally, with the exception of *N. patriciarum* xylanase, all family 11 xylanase structures exhibit the same

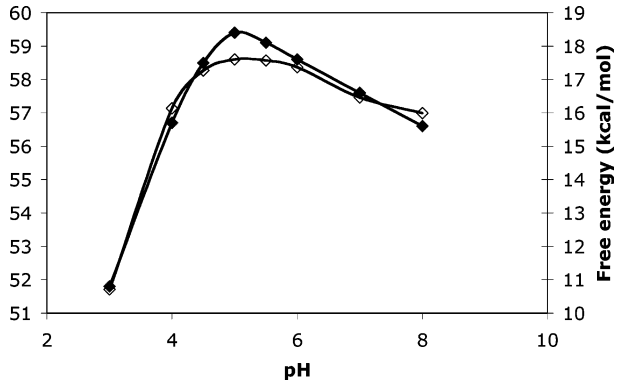


FIGURE 8: Plot of observed T_m values (◆, left *y*-axis) and predicted pH-dependent free energies of unfolding (◇, right *y*-axis) for *B. circulans* xylanase.

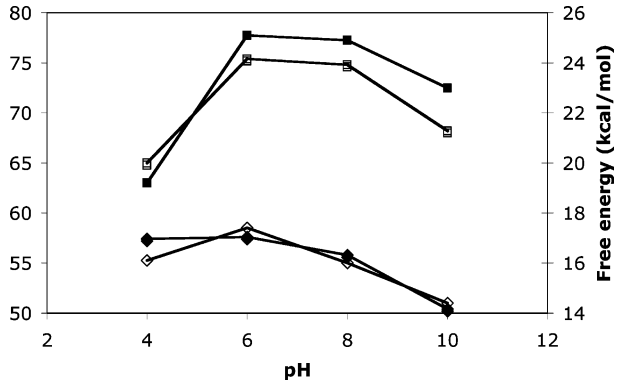


FIGURE 9: Plot of observed T_m values (◆ and ■, left *y*-axis) and predicted pH-dependent free energies of unfolding (◇ and □, right *y*-axis) for *B. circulans* xylanase (◆ and ◇) and *B. agaradhaerens* xylanase (■ and □).

key determinants of the pH optimum [except that Asn35 is Asp35 in acidic xylanases (cf. Figure 1)] and thus provide no new design ideas for engineering pH optima. However, an analysis of 254 sequences of family 11 xylanases on the Cazy web site (http://www.cazy.org) reveals a few interesting exceptions: for example xylanase B from *C. thermophilium* (sequence Q8J1V4), xylanase B from *Penicillum purpurogenum* (Q96W72), and xylanase from *Penicillum* sp. 40 (Q9UUQ2) have Gly, His, and Gln residues in place of Asn35, Tyr80, and Arg112 (*B. circulans* numbering), respectively. Clearly, X-ray structures of these proteins would be very useful.

pH-Dependent Stabilities. Table 4 lists the pH value at which the pH-dependent unfolding free energy (cf. eq 3) attains a maximum (pH_{max}), and the corresponding free energy value relative to pH 0 [$\Delta\Delta G_U(pH_{max},0)$]. For most xylanases considered here, the free energy varies relatively little in certain pH ranges. Therefore, the pH range for which the free energy is >80% of the maximum value is also listed. As we discuss below, this pH range can be a reasonably good approximation of the pH range at which the enzyme shows >80% residual activity. Next we compare these results to the experimental data on stability available in the literature for these systems and offer possible rationalization of some of the experimental observations.

Davoodi et al. (*40*) have reported midpoint denaturation temperatures (T_m) for *B. circulans* xylanase in the pH range of 3−8 measured by isothermal calorimetry, and there data are plotted together with the corresponding pH-dependent folding free energies in Figure 8. There is excellent correlation between the observed T_m values and computed $\Delta\Delta G_U$(pH,0) for *B. circulans* xylanase ($r = 0.98$), and the predicted pH of optimal stability (5.2) is in good agreement with the experimental value of 5. Similar T_m data measured by CD spectroscopy have been reported for *B. circulans* and *B. agaradhaerens* xylanase for pH 4.0, 6.0, 8.0, and 10.0 by Poon et al. (*28*). These data are shown in Figure 9 (right *y*-axis) along with the computed free energies[3] at the corresponding pH values (left *y*-axis). Here there is good ($r = 0.94$) and reasonable ($r = 0.91$) correlation between the pH-dependent unfolding free energy and the T_m data. The predicted optimal stability pH values are 5.2 and 6.5, which is good agreement with the experimentally observed T_m maxima of 5 and 6 for *B. circulans* and *B. agaradhaerens* xylanse, respectively. However, we note that it is not possible to determine whether the pH−stability optimum is closer to pH 5 since experimental data were not obtained for this pH value. Both studies note that the denaturation is irreversible, so thermodynamic parameters cannot be extracted from the experimental data.

The pH-dependent folding free energies for two proteins are not rigorously comparable because the reference free energy may be different. However, it is interesting to note that the higher T_m (at a given pH) observed for *B. agaradhaerens* xylanse compared to that of *B. circulans* xylanase correlates with a larger $\Delta\Delta G_U$(pH,0), suggesting that the thermostability of *B. agaradhaerens* xylanase is in part due to a difference in residues with shifted pK_a values and that the entropies of unfolding at the respective T_m values are similar (cf. eq 4). This prediction is also consistent with the fact that the relative thermostability of *B. agaradhaerens* compared to that of *B. circulans* xylanase is shown to be pH-dependent (Figure 9). Analysis of the data (not shown) shows that the larger $\Delta\Delta G_U$(pH,0) predicted for *B. agaradhaerens* xylanase relative to *B. circulans* xylanase is primarily due to seven residues with shifted pK_a values in *B. agaradhaerens* xylanase: two residues (Asp12 and Asp15) which have no counterparts in the *B. circulans* xylanase sequence, five residues (Glu56, His60, Glu126, Glu167, and Glu178) which are Thr, Phe, Tyr, Lys, and Tyr residues, respectively, in *B. circulans* xylanase, and Asp123 whose pK_a (1.2) is significantly lower than that of its counterpart in *B. circulans* xylanase (Asp106, pK_a = 3.0), due to two hydrogen bonds with Arg149 (Gln133 in *B. circulans*).

The correlation between T_m and $\Delta\Delta G_U$(pH,0) is further probed in Figure 10, which shows a good correlation of measured T_m values at pH 6.0 (by CD spectroscopy) of *B. circulans* (*28*), *B. agaradhaerens* (*28*), and *Streptomyces* S38 (*31*) xylanase with the predicted pH-dependent unfolding free energy. However, a similar comparison of CD-measured T_m and $\Delta\Delta G_U$(pH,0) at pH 8.0 for *B. agaradhaerens* [75 °C and 25 kcal/mol (*28*)] and *B. subtillis* B230 [67 °C and 28−29 kcal/mol (*29*)] shows that this correlation does not always hold. Furthermore, the half-lives of *Tr. reesei* II, *C. thermophilium*, and *N. flexuosa* xylanase at pH 6.0 and 75 °C are 0, 7, and 1500 min, respectively (*41*), which does not

[3] As discussed previously, there are large discrepancies with experiment for the predicted pK_a values of Glu17 and Glu178. Therefore, we have used experimental pK_a values for these two residues.

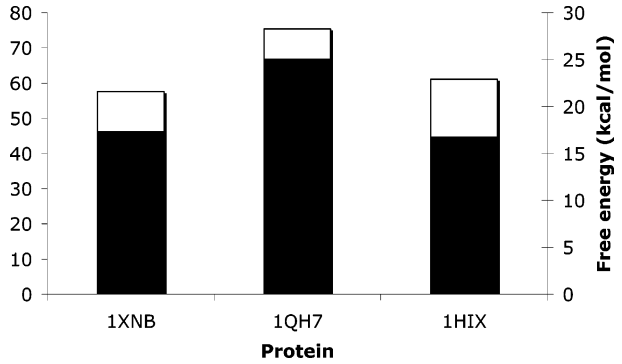


FIGURE 10: Plot of observed T_m values (white bars, left *y*-axis) and predicted pH-dependent free energies of unfolding (black bars, right *y*-axis) for *B. circulans* xylanase (1XNB), *B. agaradhaerens* xylanase (1QH7), and *Streptomyces* S38 (1HIX) at pH 6.

correlate well with the corresponding predicted $\Delta\Delta G_U$(6.0,0) values of 23, 21−27, and 22 kcal/mol, respectively.

The pH-dependent unfolding free energy of *B. agaradhaerens* xylanase is within 80% of the maximum value in the pH range of 4.3−10.3, while the residual activity is >80% at pH 6, 8, and 10, respectively, after incubation at 55 °C for up to 6 min (no data reported for pH 4) (*28*).

The pH-dependent stability of xylanase from *T. lanuginosus* is predicted to be largest for pH 7.8, which is consistent with the observation that the residual activity after incubation for 3 h at room temperature is largest around roughly pH 6.0−8.0 (*30*). The free energy is within 80% of the maximum value in the pH range of 4.3−9.9, which is in reasonable agreement with the pH range of 5−12 within which the residual activity is >80%.

Xylanase I and II from *Tr. reesei* have 100% residual activity after 24 h at room temperature in the pH ranges of 2.5−8.5 (i.e., at all values tested) and 3.0−8.5, respectively (*32*), while the corresponding free energies are within 80% of the maximum values in the pH ranges of 3.5−8.3 and 4.1−10.7. It is worth noting that the residual activity is very temperature-dependent: at 40 °C the respective ranges are 2.5−4.5 and 4.0−7.5 (*32*).

Finally, after 1 h at 30 °C, the xylanase from *A. kawachii* has 80% residual activity in the pH range of 1.0−9.0 (*37*), while the corresponding free energies are within 80% of the maximum values in the pH range of 3.8−7.8. Furthermore, the experimentally determined pH of maximum stability is 2, much lower than the theoretical prediction of 5.1.

In summary, the predicted pH of optimal stability is in excellent agreement with the pH value at which the T_m is largest, though the pH dependence of T_m has only been measured for *B. circulans* and *B. agaradhaerens*. The predicted $\Delta\Delta G_U$(pH,0) value tends to vary relatively little (within 80% of the maximum pH-dependent free energy of unfolding) with pH in the range of ∼4−10, which corresponds roughly to the pH range at which the xylanases exhibit >80% residual activity, though such data are available for xylanases from only *B. agaradhaerens*, *T. lanuginosus*, *Tr. reesei*, and *A. kawachii*. Most notably, the predicted pH optima for the three xylanases from acidophilic organisms (*Tr. reesei*, *A. niger*, and *A. kawachii*) are not lower than, say, that of *B. circulans*, nor are their pH-dependent free energies of unfolding predicted to be significant larger at low pH.

Some correlation is observed between $\Delta\Delta G_U$(pH,0) and T_m, suggesting that the thermophilicity for at least some of the xylanases is in part due to a difference in residue with shifted pK_a values. In this regard, it is interesting to note that the average predicted maximum $\Delta\Delta G_U$(pH,0) value for xylanases from thermophilic organisms (*T. langinuosus*, *N. flexuosa*, *C. thermophilum*, *D. thermophilum*, and *Paecilomyces varioti*) is 24 kcal/mol compared to 19 kcal/mol for xylanases from mesophilic organisms. In general, $\Delta\Delta G_U$(pH,0) values are not directly comparable (as mentioned previously), and a recent study by Greaves and Warwicker (*42*) found no correlation with their predicted maximum $\Delta\Delta G_U$(pH,0) values and T_m in a study of 140 proteins. However, the more similar the proteins are, the better the correlation is expected to be (cf. eq 4), and the proteins here are much more similar compared to the proteins studied by Greaves and Warwicker (*42*). With that in mind, our results suggest that the thermostability of xylanases (and protein in general) can be obtained by mutations that introduce ionizable residues such as Asp, Glu, and His with pK_a values significantly lower than standard values or resdues that significantly lower pK_a values of such residues already present in the WT.

Finally, it is clear from the data in Table 5 that there is no correlation between the observed or predicted isoelectric points (pI) and the pH of optimal stability.

SUMMARY AND OUTLOOK

This paper presents a study of the pH dependence of the activity and stability of family 11 xylanases for which X-ray structures are available. Following previous work, the xylanases are separated into basic and acidic xylanases, depending on whether the catalytic acid is hydrogen bonded to an Asn or Asp residue. Using X-ray structures, the predicted pH values of optimal activity of the basic xylanases are in the range of 5.2−6.9, which is in reasonable agreement with the available experimental values of 5−6.5. Several of the X-ray structures exhibit several conformations for an Arg residue, which results in differences in pK_a values of the catalytic nucleophile that are as large as 2.9 pH units. We investigated this issue further by performing MD simulations of *B. agaradhaerens* xylanase and computing average pK_a values at snapshots along the MD simulation. The MD simulations indicate that the position of the Arg residue fluctuates significantly and that a (transient) hydrogen bond to the catalytic nucleophile is possible. Such a hydrogen bond has been observed in X-ray structures of family 11 xylanases from *D. thermophilium* and *C. termophilium*. Inclusion of this dynamical effect decreases the predicted optimum from 6.2−6.7 (obtained using the structures for the different conformation observed in X-ray structure 1QH7) to 5.3 ± 0.3, which is in good agreement with the observed value of 5.6.

In the case of acidic xylanases, there are only four X-ray structures available (counting the N35D mutant of *B. circulans* xylanase), and using these structures, the predicted pHs of optimal activity are in the range of 4.2−5.0, compared to an observed range of 2−4.6. The largest discrepancy is observed for *A. kawachii* xylanase where the observed pH optimum is 2 while the predicted value if 4.8. *A. niger* xylanase differs from *A. kawachii* xylanse by only three

Table 5: Predicted pHs of Optimal Stability (pH_{max}), pH-Dependent Free Energies at pH_{max} ($\Delta\Delta G_{max}$), pH Ranges at Which the pH-Dependent Free Energy Is within 80% of $\Delta\Delta G_{max}$, Numbers of Disulfide Bridges in the Enzyme (S−S), Experimental pIs of the Folded Enzyme, and Corresponding PROPKA Predictions for the Folded (pI F) and Unfolded (pI U) State of the Protein

species	pH_{max}	$\Delta\Delta G_{max}$	80% of ΔG_{max}	S−S	pI exp	pI F	pI U
			Basic Xylanases				
B. circulans	5.2	17.6	3.7−10.0	0	9.0	9.2	9.2
B. agaradhaerens	6.5	26.2	4.3−10.4	0		8.7	8.7
	6.4	24.6	4.1−10.4			8.7	
T. langinosus	7.8	25.7	4.3−9.9	1	3.8	4.1	4.6
Streptomyces S38	9.3	19.8	4.8−10.3	0	9.8	9.5	9.0
	9.3	17.4	4.2−10.2			8.7	
Tr. reesei II	9.3	24.3	4.3−10.7	0	9.0	8.7	8.8
	9.3	22.2	4.1−10.5			8.7	
N. flexuosa	9.4	22.8	4.4−10.5	0	8.2	7.6	7.5
B. subtilitis B230	6.4	29.1	4.1−10.6	0	8.8	9.0	9.0
	8.8	27.7	4.0−10.6			9.0	9.0
B. subtilitis 168	5.3	18.6	3.7−9.9	0		9.3	9.2
D. thermophilium	5.9	25.7	4.0−9.7	0		9.4	8.8
	5.7	23.6	3.9−9.6			9.3	
P. varioti	7.9	25.5	4.3−9.9	1		3.4	4.7
C. thermophilum	6.2	20.8	4.2−9.8	0		9.0	8.8
	6.4	24.7	4.4−10.0			8.8	
	6.8	27.0	4.2−10.1			9.0	
T. harzanum E58	6.1	19.0	4.0−10.4	0		8.8	8.8
			Acidic Xylanases				
B. circulans N35D	5.4	17.9	3.8−9.0	0		9.1	8.9
Tr. reesei I	4.8	11.6	3.5−8.3	0	5.5	4.5	4.6
A. niger	5.5	16.8	4.1−6.8	1		3.3	3.9
A. kawachii	5.1	14.3	3.8−7.8	1	3.1	3.2	3.8

residues and has an observed pH optimum of ~3, so it is possible that the pH optimum of *A. kawachii* is closer to 3. Averaging pK_a values over a 1.2 ns MD simulation, to include the dynamical fluctuation of the Arg residue, lowers the predicted pH optimum to 4.0 ± 0.2. Contrary to MD simulations of *B. agaradhaerens* xylanase, a hydrogen bond between the catalytic nucleophile and the Arg residue was not observed during this simulation. However, a MD simulation starting from a structure with such a hydrogen bond indicates that it can persist for at least 1 ns, and that it may be stabilized by an additional hydrogen bond to a nearby Glu residue, not present in the basic xylanases. The pH optimum based on this MD simulation is 2.9 ± 0.3, which is in reasonable agreement with the observed value of 2.

The predicted pH of optimal stability is in excellent agreement with the pH value at which T_m is largest, although the pH dependence of T_m has been measured for only *B. circulans* and *B. agaradhaerens*. The predicted $\Delta\Delta G_U$(pH,0) values tend to vary relatively little (within 80% of the maximum pH-dependent free energy of unfolding) with pH in the range of ca. 4−10, which corresponds roughly to the pH range at which the xylanases exhibit >80% residual activity. Some correlation is observed between $\Delta\Delta G_U$(pH,0) and T_m for the three xylanases for which T_m values have been published, suggesting that the thermostability of at least some of the xylanases is in part due to a difference in residue with shifted pK_a values. Thus, the thermostability of xylanases (and protein in general) can perhaps be obtained by mutations that introduce ionizable residues such as Asp, Glu, and His with pK_a values significantly lower than standard values or significantly lower pK_a values of such residues already present in the WT.

This study indicates that PROPKA is a useful tool for identifying key residues that determine the pH dependence of the activity and stability of family 11 xylanases, and presumably proteins in general. However, significantly more data are needed to properly validate the computational approach, such as microscopic pK_a values measured by NMR, pH−activity profiles (preferably measured using synthetic substrates) that allow the apparent pK_a values to be extracted, and pH−T_m profiles that span pH values at which the protein unfolds at room temperature (which allows the reference free energy to be extracted and compared for different enzymes). We hope that this study will inspire further work.

ACKNOWLEDGMENT

We thank David Rogers for the use of his propka.pl script and Jamshid Davoodi, Lawrence McIntosh, and David Poon for sharing experimental T_m data. J.H.J. thanks Leonardo de Maria, Esben Friis (who alerted us to the 2C1F structure), Marco Betz, Lawrence McIntosh, and Jens Erik Nielsen for very helpful discussions.

REFERENCES

1. Collins, T., Gerday, C., and Feller, G. (2005) Xylanases, xylanase families and extremophilic xylanases, *FEMS Microbiol. Rev. 29*, 3−23.
2. Joshi, M. D., Sidhu, G., Nielsen, J. E., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2001) Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase, *Biochemistry 40*, 10115−10139.
3. Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2000) Hydrogen bonding and catalysis: A novel explanation for how a single amino acid substitution can change the pH optimum of a glycosidase, *J. Mol. Biol. 299*, 255−279.
4. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) The pK_a of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A C-13-NMR study of *Bacillus circulans* xylanase, *Biochemistry 35*, 9958−9966.
5. Tanford, C. (1970) Protein denaturation. C. Theoretical models for the mechanism of denaturation, *Adv. Protein Chem. 24*, 1−95.
6. Yang, A. S., and Honig, B. (1993) On the Ph-Dependence of Protein Stability, *J. Mol. Biol. 231*, 459−474.

7. Elcock, A. H. (1999) Realistic modeling of the denatured states of proteins allows accurate calculations of the pH dependence of protein stability, *J. Mol. Biol. 294*, 1051−1062.
8. Alexov, E. (2004) Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and three-dimensional structure, *FEBS J. 271*, 173−185.
9. Tollinger, M., Crowhurst, K. A., Kay, L. E., and Forman-Kay, J. D. (2003) Site-specific contributions to the pH dependence of protein stability, *Proc. Natl. Acad. Sci. U.S.A. 100*, 4545−4550.
10. Zhou, H. X. (2002) A Gaussian-chain model for treating residual charge-charge interactions in the unfolded state of proteins, *Proc. Natl. Acad. Sci. U.S.A. 99*, 3569−3574.
11. Becktel, W. J., and Schellman, J. A. (1987) Protein stability curves, *Biopolymers 26*, 1859−1877.
12. Linderstrom-Lang, K. U. (1924) On the ionisation of proteins, *C. R. Trav. Lab. Carlsberg 15*, 1−29.
13. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A., and Pace, C. N. (2001) The effect of net charge on the solubility, activity, and stability of ribonuclease Sa, *Protein Sci. 10*, 1206−1215.
14. Li, H., Robertson, A. D., and Jensen, J. H. (2005) Very fast empirical prediction and rationalization of protein pK_a values, *Proteins: Struct., Funct., Bioinf. 61*, 704−721.
15. Davies, M. N., Toseland, C. P., Moss, D. S., and Flower, D. R. (2006) Benchmarking pKa Prediction, *BMC Biochem. 7*, 18.
16. Alexov, E. (2003) Role of the protein side-chain fluctuations on the strength of pair-wise electrostatic interactions: Comparing experimental with computed pK_as, *Proteins 50*, 94−103.
17. Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. III (2005) The Amber biomolecular simulation programs, *J. Comput. Chem. 26*, 1668−1688.
18. Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, *J. Comput. Chem. 24*, 1999−2012.
19. Darden, T., York, D., and Pedersen, L. (1993) Particle Mesh Ewald: An N.Log(N) Method for Ewald Sums in Large Systems, *J. Chem. Phys. 98*, 10089−10092.
20. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical-Integration of Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes, *J. Comput. Phys. 23*, 327−341.
21. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular Dynamics with Coupling to an External Bath, *J. Chem. Phys. 81*, 3684−3690.
22. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Clustal-W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice, *Nucleic Acids Res. 22*, 4673−4680.
23. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis 18*, 2714−2723.
24. Joshi, M. D., Hedberg, A., and McIntosh, L. P. (1997) Complete measurement of the pK_a values of the carboxyl and imidazole groups in *Bacillus circulans* xylanase, *Protein Sci. 6*, 2667−2670.
25. Plesniak, L. A., Connelly, G. P., Wakarchuk, W. W., and McIntosh, L. P. (1996) Characterization of a buried neutral histidine residue in *Bacillus circulans* xylanase: NMR assignments, pH titration, and hydrogen exchange, *Protein Sci. 5*, 2319−2328.
26. Hirata, A., Adachi, M., Utsumi, S., and Mikami, B. (2004) Engineering of the pH optimum of *Bacillus cereus* β-amylase: Conversion of the pH optimum from a bacterial type to a higher-plant type, *Biochemistry 43*, 12523−12531.
27. Betz, M., Lohr, F., Wienk, H., and Ruterjans, H. (2004) Long-range nature of the interactions between titratable groups in *Bacillus agaradhaerens* family 11 xylanase: pH titration of *B. agaradhaerens* xylanase, *Biochemistry 43*, 5820−5831.
28. Poon, D. K. Y., Webster, P., Withers, S. G., and McIntosh, L. P. (2003) Characterizing the pH-dependent stability and catalytic mechanism of the family 11 xylanase from the alkalophilic *Bacillus agaradhaerens*, *Carbohydr. Res. 338*, 415−421.
29. Oakley, A. J., Heinrich, T., Thompson, C. A., and Wilce, M. C. (2003) Characterization of a family 11 xylanase from *Bacillus subtillis* B230 used for paper bleaching, *Acta Crystallogr. 59*, 627−636.
30. Lin, J., Ndlovu, L. M., Singh, S., and Pillay, B. (1999) Purification and biochemical characteristics of β-D-xylanase from a thermophilic fungus, *Thermomyces lanuginosus* SSBP, *Biotechnol. Appl. Biochem. 30*, 73−79.
31. Esteves, F. D., Ruelle, V., Lamotte-Brasseur, J., Quinting, B., and Frere, J. M. (2004) Acidophilic adaptation of family 11 endo-β-1,4-xylanases: Modeling and mutational analysis, *Protein Sci. 13*, 1209−1218.
32. Tenkanen, M., Puls, J., and Poutanen, K. (1992) Two major xylanases from *Trichoderma reesei*, *Enzyme Microb. Technol. 14*, 566−574.
33. Leskinen, S., Mantyla, A., Fagerstrom, R., Vehmaanpera, J., Lantto, R., Paloheimo, M., and Suominen, P. (2005) Thermostable xylanases, Xyn10A and Xyn11A, from the actinomycete *Nonomuraea flexuosa*: Isolation of the genes and characterization of recombinant Xyn11A polypeptides produced in *Trichoderma reesei*, *Appl. Microbiol. Biotechnol. 67*, 495−505.
34. Torronen, A., and Rouvinen, J. (1995) Structural Comparison of Two Major endo-1,4-Xylanases from *Trichoderma reesei*, *Biochemistry 34*, 847−856.
35. Krengel, U., and Dijkstra, B. W. (1996) Three-dimensional structure of endo-1,4-β-xylanase I from *Aspergillus niger*: Molecular basis for its low pH optimum, *J. Mol. Biol. 263*, 70−78.
36. Fushinobu, S., Ito, K., Konno, M., Wakagi, T., and Matsuzawa, H. (1998) Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: Biased distribution of acidic residues and importance of Asp37 for catalysis at low pH, *Protein Eng. 11*, 1121−1128.
37. Ito, K., Ikemasu, T., and Ishikawa, T. (1992) Purification and properties of acid stable xylanases from *Aspergillus kawachii*, *Biosci., Biotechnol., Biochem. 56*, 547−550.
38. Honda, Y., Kitaoka, M., Sakka, K., Ohmiya, K., and Hayashi, K. (2002) An investigation of the pH-activity relationships of Cex, a family 10 xylanase from *Cellulomonas fimi*: Xylan inhibition and the influence of nitro-substituted aryl-β-D-xylobiosides on xylanase activity, *J. Biosci. Bioeng. 93*, 313−317.
39. Huang, Y. H., Huang, C. T., and Hseu, R. S. (2005) Effects of dockerin domains on *Neocallimastix frontalis* xylanases, *FEMS Microbiol. Lett. 243*, 455−460.
40. Davoodi, J., Wakarchuk, W. W., Campbell, R. L., Carey, P. R., and Surewicz, W. K. (1995) Abnormally high pK_a of an active-site glutamic acid residue in *Bacillus circulans* xylanase. The role of electrostatic interactions, *FEBS J. 232*, 839−843.
41. Hakulinen, N., Turunen, O., Janis, J., Leisola, M., and Rouvinen, J. (2003) Three-dimensional structures of thermophilic β-1,4-xylanases from *Chaetomium thermophilum* and *Nonomuraea flexuosa*. Comparison of twelve xylanases in relation to their thermal stability, *FEBS J. 270*, 1399−1412.
42. Greaves, R. B., and Warwicker, J. (2007) Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles, *BMC Struct. Biol. 7*, 18.

BI7016365