BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Conserved Water Molecules in Family 1 Glycosidases: A DXMS and Molecular Dynamics Study

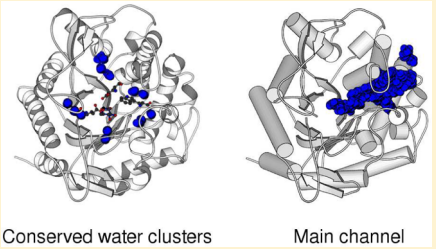
David Teze,[†] Johann Hendrickx,[†] Michel Dion,[†] Charles Tellier,[†] Virgil L. Woods, Jr.,[‡,⊥] Vinh Tran,[†] and Yves-Henri Sanejouand*[,†,§]

[†]UFIP, Université de Nantes, 2 rue de la Houssinière, 44322 Nantes, France

[‡]Department of Medicine, University of California−San Diego, 9500 Gilman Drive, La Jolla, California 92093-0652, United States

Supporting Information

ABSTRACT: By taking advantage of the wealth of structural data available for family 1 glycoside hydrolases, a study of the conservation of internal water molecules found in this ubiquitous family of enzymes was undertaken. Strikingly, seven water molecules are observed in more than 90% of the known structures. To gain insight into their possible function, the water dynamics inside *Thermus thermophilus* β-glycosidase was probed using deuterium exchange mass spectroscopy, allowing the pinpointing of peptide L117-A125, which exchanges most of its amide hydrogens quickly in spite of the fact that it is for the most part buried in the crystal structure. To help interpret this result, a molecular dynamics simulation was performed whose analysis suggests that two water channels are involved in the process. The longest one (∼16 Å) extends between the protein surface and W120, whose side chain interacts with E164 (the acid−base residue involved in the catalytic mechanism), whereas the other channel allows for the exchange with the bulk of the highly conserved water molecules belonging to the hydration shell of D121, a deeply buried residue. Our simulation also shows that another chain of highly conserved water molecules, going from the protein surface to the bottom of the active site cleft close to the nucleophile residue involved in the catalytic mechanism, is able to exchange with the bulk on the nanosecond time scale. It is tempting to speculate that at least one of these three water channels could be involved in the function of family 1 glycoside hydrolases.



Glycoside hydrolases (GH) are found in all kingdoms of life. They are highly efficient catalysts able to hydrolyze glycosidic bonds with rate constants up to 1000 s^{-1}, that is, to accelerate their spontaneous rate of hydrolysis by factors approaching 10^{17}-fold.[1] Such an extreme efficiency is achieved through (at least) two kinds of mechanisms, which are responsible for either the net inversion or retention of the anomeric configuration of the linkage to be cleaved. In the latter case, the reaction proceeds in two steps, with the outcome of the first one being the formation of a glycosyl−enzyme covalent intermediate, which is the species that reacts with water, whereas the hydrolysis of the glycosyl−enzyme bond itself is assumed to proceed through a rather straightforward mechanism: a key acid−base residue grasps a proton of the water molecule, which then becomes able to perform a nucleophilic attack on the glycosyl−enzyme linkage.[2]

There are lines of evidence suggesting that the full story may prove to be more complex. For instance, it has been shown that a single mutation, far from the glycosyl−enzyme bond to be cleaved, can prove enough for favoring transglycosylation over hydrolysis during the second step.[3,4] This means that as a consequence of a single mutation the glycosyl−enzyme lifetime can become long enough so that an acceptor sugar can access the active site, compete with water, and react with the covalent intermediate in spite of the fact that the active site is, in the meantime, filled with water molecules. On the other hand, while the sialidase of *Trypanosoma rangeli* and the trans-sialidase of *Trypanosoma cruzi* are 70% identical at the amino acid level, their catalytic sites are highly conserved, with only subtle structural differences left for rationalizing their spectacular difference of function.[5] Moreover, when mutations are performed with the aim of turning the sialidase activity of *Trypanosoma rangeli* into the trans-sialidase activity of *Trypanosoma cruzi*, six mutations are required before obtaining a significant switch. However, the trans-sialidase activity obtained is equivalent to 10% of that of *Trypanosoma cruzi*,[6] probably because the hydrolytic activity of the mutants remains as high as that of the wild type.

From such facts, it is tempting to speculate that water molecules involved in the hydrolytic activity of retaining glycoside hydrolases need more than the correct positioning of an acid−base residue in their neighborhood to be involved in the reaction mechanism. Indeed, when the structures of the glycosyl-enzymes of *Trypanosoma rangeli* sialidase[7] and *Trypanosoma cruzi* trans-sialidase[8] are superimposed, their aspartate acid−bases[9] are found to overlap.

For instance, in the case of a newly discovered glycoside hydrolase family (GH117), it has been hypothesized that a

Received: March 1, 2013
Revised: July 25, 2013
Published: July 29, 2013

chain of seven water molecules located in a channel going from the enzyme active site to the enzyme surface may play an important role for the provision of the catalytically added water molecule.[10] Although the hypothesis of a functional role for the water chain found in GH117 remains to be demonstrated, crystallographic and molecular dynamics studies have already suggested that water channels permitting efficient proton transfer through chains of hydrogen bonds or simply easier active-site access, hydration, and/or dehydration, may prove essential for the function of numerous enzymes, like serine proteases,[11] protein kinases,[12] cytochrome P450,[13,14] the *Rhizomucor miehei* lipase,[15] aldose reductase,[16] histone lysine methlytransferase SET79,[17] guanylate kinase,[18] and ATP-synthase.[19] In the case of haloalkane dehalogenase, redesigning the water-access channels leading to the buried active site to occlude them was even shown to have a dramatic effect on the enzyme function.[20] Also, in the case of catalase, the mutation of a single aspartate located halfway in the long channel joining the active site to the protein surface, reduces its activity by up to 90%.[21]

To provide clues indicating that this may also be the case for retaining glycoside hydrolases, a study of water structure and dynamics nearby the active site of *Thermus thermophilus* β-glycosidase was undertaken on the basis of the wealth of structural data available for glycoside hydrolases belonging to family 1 (GH1[22]) while taking advantage of the complementarity of deuterium-exchange mass spectroscopy and molecular dynamics techniques.

METHODS

Multiple Sequence Alignment. GH1 sequences were retrieved from the Uniprot database,[23] and known GH1 structures, namely, those gathered in the CAZy database,[22] were retrieved from the Protein Data Bank.[24,25] Multiple sequence alignments (MSA) were performed with CLUSTALW.[26] For assessing the degree of amino acid residue conservation, 548 bacterial sequences that are more than 30% identical to the sequence of *Thermus thermophilus* β-glycosidase (Ttβgly) and less than 90% identical to each other after the MSA were considered. Hereafter, amino acid residues conserved in more than 95% of these sequences are coined "highly conserved". This represents a subset of 37 residues (9% of Ttβgly residues).

Water Clustering. The clustering of crystallographic water molecules found in GH1 structures with a sufficiently high resolution (better than 2.0 Å) was performed using the highest resolution one (R = 0.99 Å) as a reference, that is, the structure of *Thermus thermophilus* β-glycosidase (Ttβgly; PDB ID 1UG6).

Because accurate water clustering requires high-quality structural superimposition and because the quality of a structural superimposition drops as the sequence identity decreases, GH1 with sequences less than 40% identical to the Ttβgly sequence after multiple sequence alignment were disregarded. The remaining 54 protein chains (there are up to four chains in a given PDB structure of our data set) were superimposed onto the structure of Ttβgly using only, as a reference frame, the backbone atoms of the 93 amino acids (22% of them) perfectly conserved in this data set. Such a strategy was chosen because this subset of amino acids has proved particularly rigid across evolution. Indeed, their C_α root-mean-square difference (C_α rmsd) with Ttβgly is less than 0.7 Å for all structures considered, except for the structure of the fungus *Trichoderma reesei* GH1 (PDB 3AHY), for which a C_α rmsd of 1.2 Å is observed. In practice, protein chain superimposition was performed with the Profit software (Martin, A.C.R., http://www.bioinf.org.uk/software/profit).

Crystallographic water molecules were then clustered using a cutoff criterion applied to distances between all pairs of water oxygens coming from the 54 protein chains of our data set. According to this criterion, a water molecule belongs to a given cluster if its oxygen atom is less than a given distance (R_c) away from the oxygen of a water molecule already belonging to that cluster. Therefore, when R_c has a value smaller than the length of an hydrogen bond, the size of the cluster gives the level of conservation of the position of a single water molecule in the structures of the data set considered. Note that such an approach relies on the high structural conservation of this family of enzymes across evolution.

Deuterium-Exchange Mass Spectroscopy. Initial isotopic labeling was performed for various timespans (10, 30, 100, 300, 1000, 3000, 10 000, and 30 000 s) at a low temperature (0 °C) so that most deuterium−hydrogen exchanges could occur during the deuterium-exchange mass spectroscopy (DXMS) analysis. In practice, 20 μL of purified protein (18 μg per μL) was incubated with 60 μL of buffered D_2O (Tris 8.3 mM, pH 7, and NaCl 50 mM), with 4 μL being picked for each individual analysis. The deuterated protein samples were then quenched, that is, unfolded (0.5 M guanidine hydrochloride) under low pH conditions (0.8% formic acid) to minimize the back-exchange rate. Cleavage into fragments was performed on a pepsin column, the peptides obtained were analyzed by standard mass spectrometry (MS), using a Thermo Finnigan LCQ Classic electrospray ion-trap mass spectrometer and the DXMS (Sierra Analytics) and SEQUEST (Thermo Finnigan) softwares. Deuterium loss by individual fragments during DXMS analysis was calibrated using protein samples that were initially fully labeled under unfolding conditions and then moved into H_2O solution before proteolysis, as was done for the other samples. To do so, deuterium incorporation in a given peptide was calculated by comparing the mass spectra of the partially deuterated, nondeuterated, and fully deuterated peptides. Next, theoretical mass spectra corresponding to the percentages of deuteration thus obtained were compared to the experimental ones. This last step allows the removal from further analysis of peptides for which experimental data happen to be ambiguous. In practice, all measurements described in the current study were performed for a *Thermus thermophilus* β-glycosidase bearing the F401S mutation.[3]

Molecular Dynamics Simulations. Crystal structures were first embedded into a shell of water molecules to have roughly 0.7 g of water per gram of protein, well above the 0.3−0.4 g/g of structurally bound water molecules expected around a protein,[27] with the proper water density being enforced using the Beglov−Roux confining potential, which also prevents evaporation.[28,29] For the purpose of the present study, using such a boundary potential has an additional advantage: it limits the number of water molecules to a few thousand, making the detailed analysis of water dynamics easier to manage.

Next, the system was relaxed step by step through 5000 steps of energy minimization followed by 200 ps of molecular dynamics (MD) simulation under harmonic constraints on protein heavy atoms as well as on crystallographic water oxygens, with the corresponding force constant being divided by 10 before each new step, from 1000 to 0.01 kcal/mol/Å^2.

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900−5910

Further unconstrained MD simulation was performed with the CHARMM27 force field,[30] the TIP3P water model,[31] shifted electrostatics, a cutoff at 9 Å, the SHAKE algorithm,[32] and a time step of 2 fs. Note that a similar level of description of the potential-energy surface has recently proved accurate enough to enable the ab initio folding of several fast folding proteins.[33] All calculations were done with version 35b3 of CHARMM[34] available at CINES (Montpellier, France) on the Jade SGI supercomputer.

RESULTS AND DISCUSSION

Conserved Clusters of Water Molecules in GH1 Structures. *Crystallographic Water Clustering.* Conserved water molecules were sought in the crystallographic structures of glycoside hydrolases belonging to family 1 among those whose sequence identity with *Thermus thermophilus* β-glycosidase (Ttβgly) is high enough (40% or more) to ensure high-quality structure superimposition and reliable water clustering. Herein, water conservation is probed when water molecules coming from most of the 25 PDB structures considered (54 protein chains) are found in a given cluster.

When R_c, the water−water distance threshold for being included in a given cluster, is larger than 2.0 Å, the corresponding (bulk) cluster contains ∼20 000 water molecules (370 per protein chain, on average). It covers the entire surface of the protein including the cleft leading to the catalytic site. As a matter of fact, dozens of water molecules belonging to this cluster are hydrogen bonded to E164, the acid−base involved in the catalytic mechanism. Moreover, after superimposing the structure of a glycosyl-enzyme-reaction intermediate analogue, obtained from the structure of the β-glucosidase of *Thermotoga maritima* (PDB ID 1OIN), seven of them are observed less than 4 Å away from the anomeric C1 atom involved in the glycosyl-protein linkage, close to the location expected for a nucleophilic attack. The fact that these water molecules belong to the bulk cluster illustrates how simple it is for a water molecule to access the active site and to adopt a near-attack configuration. Note that hereafter, because we are seeking water channels, only buried water molecules are considered.

Isolated Conserved Clusters. Near W385, which belongs to the highly conserved transition-state stabilizing hydrophobic platform,[35] there is a conserved cluster of water molecules whose oxygen atoms are more than 6.4 Å away from any other crystallographic water oxygen in our structure data set (Figure S1). The average distance from the cluster center of the 50 molecules belonging to this cluster is 0.2 Å, meaning that it corresponds to the site of a unique water molecule that is found in 92% of the 54 structures in our data set, that is, whose position is highly conserved. Indeed, in the structure of Ttβgly (PDB ID 1UG6), a single water molecule is observed there. Therefore, this water molecule is likely to be a characteristic feature of all GH1 structures. Looking at the few structures in which it has not been reported supports this point. For instance, although it is missing in three structures from an uncultured bacterium (PDB 3FIY, 3FIZ, and 3FJ0), it has been found in the fourth one from the same species available in our data set (PDB 3CMJ). The role of this isolated water molecule is likely to be a structural one because it is hydrogen bonded to the backbone carbonyls of W385 and F401 as well as to the backbone amide of A341 (Table 1). Thus, it may help stabilizing the peptidic bond between W385 and S386, which is in a rare and energetically unfavorable cis-configuration.[36]

Table 1. Conserved Crystal Water Molecules[a]

cluster	water index[b]	degree of conservation (%)	in all kingdoms	hydrogen bonds[c]
W385	×14	**94**	yes	A341n, **W385o**, F401o
T14	×1	**97**	yes	**E20**, **S77o**, ×7, ×16
	×7	88	yes	T14, S15n, A16n, **S77n**, ×1
T116	×2	**97**	yes	T14o, T116, **H119**[d]
D121	×3	87	no[e]	**S77**, V78o, **H119n**, **D121**
	×8	**93**	yes	**W80n**, L117o, **D121**
	×11	**95**	yes	S15, Y17n, **H119o**
G21	×13	**90**	yes	Y17o, **G21n**, G49o, ×18
	×18	80	yes	**G21o**, ×13, ×21
	×21	74	yes	**E20**, R28, ×18, ×23
	×23	73	yes	R28, **D121**, ×21
R82	×47	84	yes	D26o, Y56, **R82**
E392	×17	**93**	yes	**W385**, N390, E392, ×22
	×22	89	yes	N390n, R399o, F401n, ×17
D34	×60	76	no[e]	Y17, **D34**
N219	×19	76	no[e]	W166n, N219, ×92
	×92	73	no[e]	N219o, ×19, ×106
R399	×46	85	yes	N390o, R399o
W120	×4	67	no[e]	H178, ×10
	×10	73	yes	**W33**, A179o, ×4

[a]For each conserved cluster, the water molecules found in the *T. thermophilus* β-glycosidase structure (PDB ID 1UG6) whose position is conserved in more than 50% of all GH1 structures with a resolution better than 2.0 Å are given as well as the list of their hydrogen bonds in the crystal structure (highly conserved residues are shown in bold). Our data set includes 7, 27, and 60 structures from Archaea, Eubacteria, and Eukarya, respectively. [b]With respect to PDB, the crystallographic water numbering is shifted by 904. In PDB 1UG6, ×1 has number 905, ×2, number 906, and so forth. [c]In the Ttβgly structure; "o" and "n" indicate that the hydrogen bond is with the backbone carbonyl or amide, respectively. Hydrogen bonding is assumed when the donor and acceptor atoms are less than 3.2 Å from each other. [d]Observed in 90% of GH1 structures but not in Ttβgly. [e]Not found in Archaea.

The second most isolated cluster, near T14, is not a typical one because, when R_c = 5.1 Å, it appears as a loose group of two subclusters. However, when R_c is in the 2.7−4.8 Å range (Figure S1), it becomes a well-defined cluster of 166 water molecules (three per protein chain), with one of them being 11.5 Å away from the C1 anomeric carbon.

Hereafter, each cluster is named after one of the residues of Ttβgly hydrogen bonded with water molecules belonging to this cluster. With the exception of cluster D121, all other clusters are within hydrogen-bonding distance ($R_c \leq 3.2$ Å) of at least one water molecule belonging to the bulk cluster (Table S1). Several of them correspond to highly conserved groups of at least two (clusters E392 and W120), three (T308), or more (G21 and N219) hydrogen-bonded water molecules. Some of these clusters are tips of narrow water channels (see below). Note that in each of them, except for cluster G21, there is at least one water molecule less than 10 Å away from the C1 anomeric carbon (Table S1). As shown in Figure1, these water clusters are sitting all around the enzyme active site.

Highly Conserved Water Molecules of Ttβgly. Up to this point, to allow for optimal structure superimposition, water

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900−5910

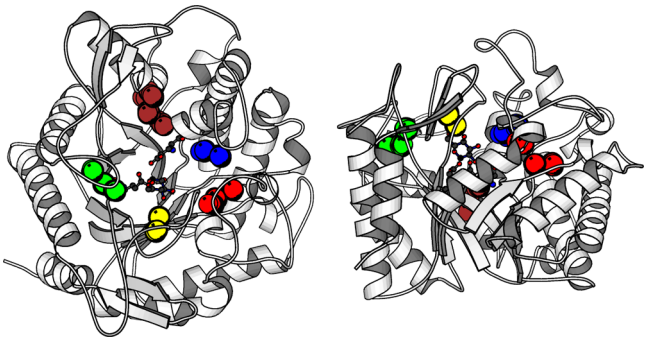


Figure 1. Five buried clusters with several conserved water molecules found in the vicinity of catalytic residues. They are shown on top of the *T. thermophilus* β-glycosidase structure (left, top view; right, side view). Water molecules belonging to clusters W120, D121, N219, T308, and E392 are depicted in blue, red, brown, green, and yellow, respectively. The two catalytic residues (including the glycosyl moiety) are shown as balls and sticks. These clusters were obtained by superimposing 54 protein chains found in 25 different GH1 structures with a resolution better than 2.0 Å. Drawn with Molscript.[37]

conservation in the GH1 family has been studied using a set of enzymes close to Ttβgly, with most of the structures considered coming from bacterial species. To confirm that the striking conservation of dozens of water molecules observed in this structure data set holds for most, if not all, GH1 structures, the 30 conserved water molecules of Ttβgly identified above (Table S1) were searched in all GH1 structures gathered in the CAZy database[22] among those with a resolution better than 2.0 Å, that is, 153 protein chains coming from 94 different PDB structures (as retrieved on November 8th, 2012). In spite of the fact that this later data set includes enzymes from Eubacteria, Archaea, and Eukarya and that, as a consequence, the sequence identity with Ttβgly can be as low as 27% after multiple sequence alignment, 20 amino acid residues are still found to be perfectly conserved. Moreover, their relative position in space is also highly conserved, with their C_α rmsd with Ttβgly being of 0.7 Å at most. As such, they can also be used as a reference frame, allowing the conservation of the position of individual water molecules to be accurately studied. Hereafter, the position of a water molecule is assumed to be conserved in a given structure if it does not change by more than 1.6 Å with respect to its position in Ttβgly. According to this criterion, 20 water molecules of Ttβgly are conserved in at least 65% of the cases (Table 1), the 10 remaining ones being found in less than 44% of them. Note that the former have all have been found both in Eubacteria and Eukarya, with most of them being found also in Archaea. Note also that seven among these 20 remarkable water molecules, including the noteworthy x14, the water molecule of Ttβgly belonging to cluster W385, are found in more than 90% of the structures. Such water molecules, whose position in space has been so well preserved all along evolution, are likely to play a major role in the folding and/or the function of the enzymes belonging to the GH1 family.

Moreover, as detailed in Table 1, 12 of these highly conserved water molecules are organized in chains of two (clusters T14, W120, N219, and E392) or even four (G21) hydrogen-bonded water molecules, suggesting that they may be part of water channels. The longest one, from cluster G21, is far from the enzyme active site (Table S1). However, two water molecules belonging to this cluster are hydrogen bonding the side chains of E20 and D121, respectively, which are two highly conserved residues also hydrogen bonded to several other remarkable water molecules (Table 1), bridging together clusters T14, G21, and D121.

Deuterium-Exchange Mass Spectroscopy. The conserved clusters of water molecules described above that are the closest to the active site (like clusters T116, W120, E392, ...) and/or involved in networks of highly conserved water molecules (like clusters G21, D121, ...) may be directly involved in the enzyme function. However, like cluster W385, they also could have a structural role. In the latter case, the water molecules belonging to these clusters are expected to be tightly bound to the protein scaffold. As a matter of fact, among the 15 water oxygens of Ttβgly with B-factors lower than 6 $Å^2$, 11 (73%) are among the 20 remarkable molecules described above (noteworthy, all those from clusters T116, W120, D121, and W385).

As a consequence, their lifetime within the cluster is expected to be a relatively long one. Accordingly, their rate of exchange with bulk water is expected to be slow. To help pinpoint which among these conserved water molecules have significant intrinsic dynamics, DXMS was used. Indeed, deuterium–hydrogen exchange is a method of choice for determining the accessibility of backbone amide hydrogens to bulk water molecules[38] because hydrogen bonding with water is required for allowing such exchange.[39] Coupled to mass spectrometry, it has the potential to reveal which stretches of buried backbone amide hydrogens are in contact with water molecules that are being exchanged relatively quickly with the bulk.[40,41] Specifically, this technique uses mass spectrometry to measure time-dependent deuterium incorporation into the backbone amides of peptides obtained by limited proteolysis.

In the case of Ttβgly, after 10 s (the time lag of the fastest measurements), 29% of all backbone amide hydrogens are exchanged (Figure 2). Such easy-to-exchange backbone amide hydrogens are usually on the protein surface itself among those

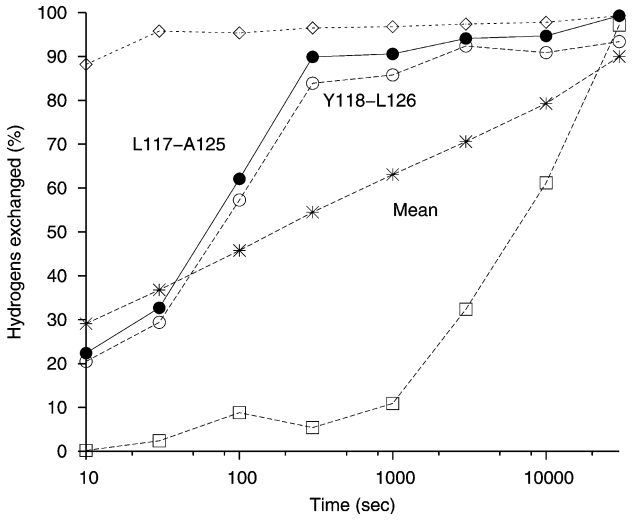


Figure 2. Kinetics of deuterium–hydrogen exchange for *T. thermophilus* β-glycosidase. Data are shown for the two overlapping peptides (open and black circles) with the highest exchange rate between 10 (the experiment time lag) and 300 s (when 54% of all backbone amide hydrogens are exchanged). For comparison, the overall percentage of the exchanged amide hydrogens is also given (stars) as well as data obtained for the peptides that have exchanged the least (F391-G396; open squares) and the most (W347-A351; open diamonds) after 300 s. Note the logarithmic scale for time.

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900–5910

that are not involved in secondary-structure elements. For instance, W347-A351, the peptide that exchanges its backbone amide hydrogens the most (88% of them) during the experimental time lag, belongs to an unstructured loop lying on the surface. Note that with the conditions of our experimental protocol it is necessary to wait roughly 300 s to have half of all backbone amide hydrogens exchanged (Figure 2). Because, in the crystal structure, 49% of backbone amides are involved in hydrogen bonds with backbone carbonyls, most exchanges observed after 300 s are expected to be those of such hydrogens. For instance, F391-G396, the peptide that exchanged its backbone amide hydrogens the least (5% of them) after 300 s, belongs to a γ-turn-α-helix motif.[42]

However, backbone amide hydrogens that are being exchanged during the 10 to 300 s time span have relatively high deuterium−hydrogen exchange rates but are not often in contact with bulk water molecules (as compared to backbone amides exposed on the protein surface). Thanks to mass spectrometry, it is possible to identify in which stretches of amino acids most of these amides stand. As shown in Table 2,

Table 2. Hydrogen-Deuterium Exchange in *T. thermophilus* β-Glycosidase Measured Using DXMS[a]

peptide	10 s (%)	300 s (%)	deuterium increase (%)
L117-A125	22	90	68
Y118-L126	21	84	63
V103-L106	0	56	56
A351-L362[b]	0	55	55
A205-L218	23	68	45
F242-L265	47	92	45
P297−W312	45	89	44
V352-A364	36	80	44
V233−F241	15	55	40
S266-L271	34	74	40

[a]The increase in the amount of deuterium incorporated between 10 and 300 s is given for the peptides that exchanged the most. On average, the 94 peptides for which good quality data could be obtained exchanged only 25% of their backbone amide hydrogens during this time span. [b]Mass spectrum quality is poor at 10 s, but the next points are fair and consistent with the first one.

in the case of Ttβgly, among the 94 peptides for which good quality mass spectra could be obtained all along the exchange kinetics, which represents a 93% coverage of the protein sequence, four are found to exchange more than 50% of their backbone amide hydrogens during this time span. For two of them, which are overlapping ones, namely L117-A125 and Y118-L126, nearly two-thirds of their backbone amide hydrogens (63−68% of them) are being exchanged. Note that the exchange kinetics of these two peptides are highly similar (Figure 2). This means that the backbone amide hydrogens of Y118 and L126 are being exchanged at a similar rate. Moreover, their exchange kinetics show that in both peptides one or two backbone amide hydrogens (20% of them) are being exchanged quickly, whereas another one is being exchanged very slowly (Table 2 and Figure 2).

Interestingly, although A125 is on the surface of Ttβgly and, as such, is accessible to bulk water, W120 is at the bottom of the active site cleft, with its aromatic ring being close to the carboxylate group of the acid−base residue (the distance between the C_β of E164 and the $N_{\varepsilon 1}$ of W120 is nearly 4 Å). As a matter of fact, between W120 and A125 the protein backbone is in an almost extended conformation, going straight through the protein, from the active site to the protein surface (Figure 3).

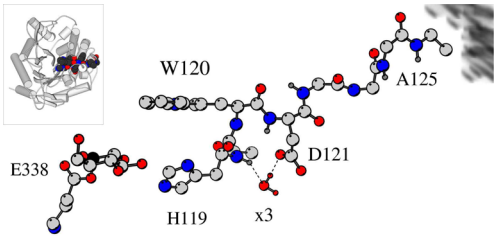


Figure 3. Extended part of L117-A125 (the buried peptide of *T. thermophilus* β-glycosidase with the highest deuterium exchange rate). x3, the only conserved crystallographic water molecule hydrogen bonded to a backbone amide hydrogen belonging to peptide L117-A125, is shown as well as E338, the nucleophile residue bearing the glycosyl moiety, in the glycosyl-enzyme intermediate. Atoms exposed on the protein surface are shadowed. The location in the structure of this peptide is specified in the inset (top view).

V103-L106, the next peptide highlighted by our DXMS data, belongs to a long helix sitting on the protein surface far away from the enzyme active site. As a matter of fact, all other peptides mentioned in Table 2 are also lying mostly on the protein surface, except for peptide A205-L218, which is half-buried, with R213-N219 being one of the β-strands of Ttβgly β-barrel. However, because after 300 s this peptide exchanged only 68% of its backbone amide hydrogens (Table 2), it is likely that most of the nonexchanged ones belong to the buried β-strand.

Molecular Dynamics Simulation. *Simulation Set Up.* Taken together, the previous results suggest that a water channel may run along the H119-A125 peptide stretch (whose backbone amides are common to the L117-A125 and Y118-L126 peptides), allowing for the deuterium−hydrogen exchange of almost all of its backbone amide hydrogens. However, in the crystal structure, this peptide is buried, whereas, among highly conserved ones, only a single water molecule is found hydrogen bonded to the backbone amide of an amino acid residue belonging to this stretch, namely, H119 (Table 1 and Figure 3).

Therefore, to help interpreting the DXMS results, a 50 ns long molecular dynamics simulation was performed starting from the structure of Ttβgly (PDB ID 1UG6). Because the experimental and simulation time scales are widely different, from seconds to hours for the former and dozens of nanoseconds for the latter, to speed up the studied process, the simulation setup was performed at 350 K, instead of 273 K, for the DXMS data acquisition. Note that this is the optimal temperature for the function of Ttβgly.[43]

Moreover, keeping in mind that the GH1 structures may have been optimized during the course of evolution for performing their natural function, namely, hydrolysis, a glycosyl moiety borrowed from a glycosyl-enzyme structure of *Thermotoga maritima* β-glucosidase (PDB 1OIN[44]) was plugged onto the structure of Ttβgly to mimic the pretransition state expected before the nucleophilic attack of a water molecule, with the fluoride in position 2 allowing for the trapping of the covalent intermediate being replaced in the simulation by an oxygen. Note that in the case of this family of enzymes it is well known that the structures of the native and

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900−5910

glycosylated forms are highly similar, except for the conformation of the nucleophile glutamate.[45,46]

However, during this 50 ns MD simulation, the conformation of the side chain of E338, the nucleophile residue of Ttβgly bearing the glycosyl moiety, is both remarkably stable and consistent with conformations observed in most GH1 glycosyl-enzyme structures (Figure S2). Even if a low-populated state ($\chi_1 \sim -90°$, $\chi_2 \sim -160°$, and $\chi_3 \sim 150°$) is shortly explored several times during the simulation (1% of total time), it happens to correspond to the conformation found in the two structures of the glycosyl-enzyme of *T. maritima* (there are two protein chains in the asymmetric unit of PDB 1OIN). Therefore, in spite of having been transferred from another crystal structure, the dynamical behavior of the glycosyl moiety of Ttβgly observed during our simulation looks fairly realistic.

Dynamics of Conserved Water Clusters. Our analysis of GH1 crystal structures has shown that 30 water molecules of Ttβgly belong to conserved clusters (Table S1). Hereafter, the position of each of these water molecules is coined a "cluster site". During our MD simulation, a water molecule (its water oxygen) is found less than 1.6 Å away from a given cluster site more than 75% of the time for all sites of clusters W385, H365, T308, G21, and E392 as well as for several sites of clusters T14 and N219. Less-populated sites (less than 25% of the time) are observed only in the case of clusters T14 (1 site out of 7), W120, and R399, with the less populated one being the later (with a water molecule being observed there only 9% of the time).

Reciprocally, a site is occupied by the same single water molecule all along the simulation in the case of cluster W385 only, where the most isolated crystal water molecule stands (Table S1), further confirming its stability and, as a corollary, its likely structural role. Overall, nine conserved cluster sites (out of 30) are visited by no more than five different water molecules during our simulation. This is particularly the case for all three sites of cluster D121, which are visited by the three water molecules belonging to this cluster in the crystal structure but also by x23, a water molecule of cluster G21 already found hydrogen bonded to D121 in the crystal structure (Table 1), and by w5030, a noncrystallographic water molecule.

However, during our simulation, no single water molecule goes through more than five different cluster sites (out of 30), suggesting that a much longer simulation would be necessary to fully describe the internal water dynamics of Ttβgly. Indeed, with Beglov–Roux boundary conditions,[28] water molecules can not leave the first hydration shell of the protein. Therefore, in a simulation that is long enough, each water molecule should be able to visit all 30 conserved sites.

Water Dynamics above the Anomeric Carbon. In GH1 enzymes, although not buried, a site of special interest for a water molecule is over the C1 anomeric carbon, where the nucleophilic attack has to take place. During our simulation, there is a water oxygen less than 3.5 Å away from the C1 carbon 30% of the time (Figure S3), with 605 different water molecules coming that close, illustrating again how easy it is for water molecules to access the enzyme active site and to adopt a near-attack configuration.

Comparison of MD with DXMS Data. Focusing on the L117-A125 peptidic stretch highlighted by DXMS, the three water molecules belonging to cluster D121 in the crystal structure (x3, x8, and x11) are found hydrogen bonded with five of the seven backbone amides of peptide Y118-A125 (Table 3). Because, as mentioned above, they are among the few ones to visit cluster D121, this means that, starting from the hydration shell of the buried side chain of D121 (Table 1), it is possible to access all backbone amide hydrogens of peptide Y118-L122. Reciprocally, it seems less likely for a water molecule to access the backbone amides of Y118, H119, and W120 without going first through cluster D121 (Table 3).

The fact that several water molecules are able to run along the Y118-L122 peptidic stretch during the 50 ns time span of our simulation, visiting all of its backbone amides (namely, x3, x8, and x11), and that other water molecules are also found hydrogen bonded with some of these amides is consistent with the DXMS data, which have shown that bulk water molecules can access these backbone amides rather easily.

In addition, our MD results suggest that the Y118 amide hydrogen is the most difficult one to access (Table 3). Taking into account the DXMS data discussed above, this means that the L126 amide hydrogen has also to be difficult to exchange. This is expected because in the crystal structure the L126 amide is involved in a hydrogen bond with the P123 backbone oxygen. Actually, the DXMS data confirm that in peptide L124-L126 there is indeed a backbone amide hydrogen that is difficult to exchange because after 300 s only 47% of the backbone amide hydrogens of this peptide are exchanged (one out of two). Moreover, because after 10 s 21% of the backbone amide hydrogens of peptide L124-L126 are exchanged, this means that the A125 amide hydrogen is being exchanged quicker than most of the other backbone amide hydrogens of peptide L117-A125. This is indeed what our simulation suggests also (Table 3). Therefore, at least in the case of the two peptides of Ttβgly highlighted by DXMS, the water dynamics on the nanosecond time scale, as obtained in a MD simulation at 350 K, is able to provide a fairly consistent interpretation of the experimental results.

Table 3. Hydrogen Bonds between Backbone Amides of Peptide L117-A125 and Water Molecules, During a 50 ns Simulation of *T. thermophilus* β-Glycosidase at 350 K[a]

amino acid residue[b]	number of molecules	time (%)	water molecule index
Y118	3	0.1	x3, x8, x11
H119	6	86	x3, x5, x8, x11, x23, w4001
W120	6	15	x3, x5, x8, x11, x23, w4001
D121	8	25	x3, x5, x8, x11, x23, w1020, w1950, w4157
L122	10	0.5	x3, x8, x11, x23, x454, **w1713**, w3973, w4001, w4157, w7232
L124	39	41	x32, x74, x98, x101, x164, x285, x316, x347, x437, ...
A125	918	65	x18, x32, x54, x56, x60, x64, x68, x74, x78, x84, ...

[a]For each amino acid residue, the number of different water molecules found hydrogen bonded is given as well as the percentage of the simulation spent hydrogen bonded. When there are no more than 10 of them, their index is given, with crystallographic water molecules being labeled "x" (others are labeled "w"). Otherwise, indices are given for 10 crystallographic water molecules. Indices of water molecules that are also visiting one of the carboxylate oxygens of acid–base residue (E164) are underlined, with the single one coming close (less than 3.5 Å) to the C1 anomeric carbon, being shown in bold. Hydrogen bonding is assumed when the water oxygen–amide nitrogen distance is less than 3.2 Å. [b]L117 and P123 have no amide hydrogen to exchange. For the former, this is due to cleavage by pepsin.

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900–5910

E127-L171 Water Channel. Several water molecules found hydrogen bonded with the backbone amides of peptide L117-A125 were also found during our MD simulation hydrogen bonded to E164, the acid−base residue, and/or close to the C1 anomeric carbon. In three cases (specified in Table 3), the time lag between both kinds of events is short (less than 1 nanosecond). This is because, as illustrated in Figure 4 for the case of w1713, these water molecules are entering (or exiting) a long channel that starts over the W120 aromatic moiety, on the enzyme active-site side, and ends on the protein surface, near E127, ~16 Å away from W120.

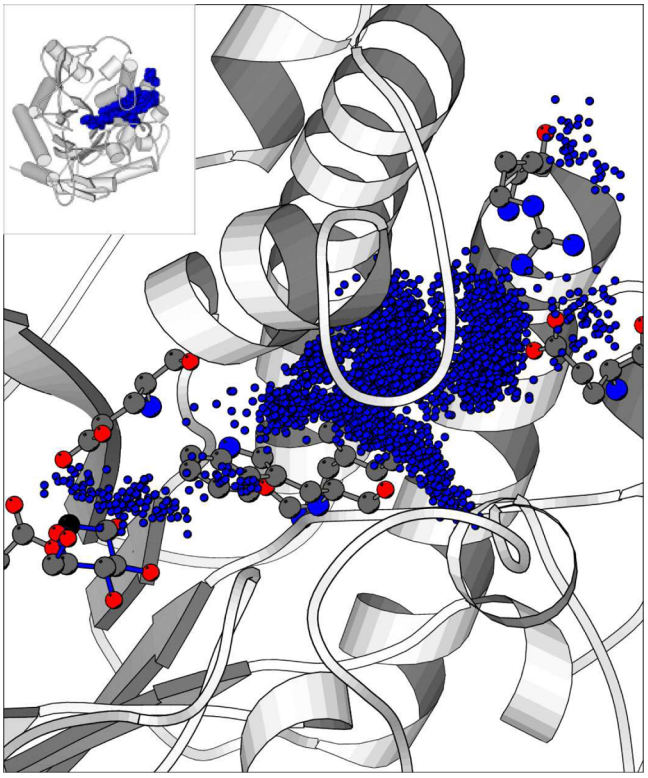


Figure 4. Channel explored by water molecule w1713 is shown on top of the *T. thermophilus* β-glycosidase structure. All blue dots correspond to water oxygen positions observed between t = 10.8 ns, when the molecule is close to the C1 anomeric carbon (depicted in black), and t = 30.5 ns, when it leaves the channel. The two catalytic residues (including the glycosyl moiety) are also shown (on the left) as well as W120 (middle), E127, and R133 (right) (whose salt bridge seems to control the channel entry on the protein surface). The location of the channel in the structure is specified in the inset (top view).

However, hydrogen-bonding with the backbone amides of peptide L117-A125 is not mandatory when visiting the water channel that starts over the W120 aromatic moiety. On the contrary, the backbone amides of peptide Y118-D121, which is in a β-turn conformation,[42] are shielded from the molecules crossing the channel by the aromatic ring of Y118 (not shown). As a matter of fact, a total of 15 different water molecules are hydrogen bonded to the Y118 hydroxyl during our MD simulation, and seven of them, like x12 (Figure 5), are also found hydrogen bonded as well, either with the acid−base residue and/or with the glycosyl moiety, shortly before entering or after exiting the channel over the W120 aromatic ring.

On the active-site side, the channel entry is narrow, being occluded most of the time by residues C167, L171, and H178

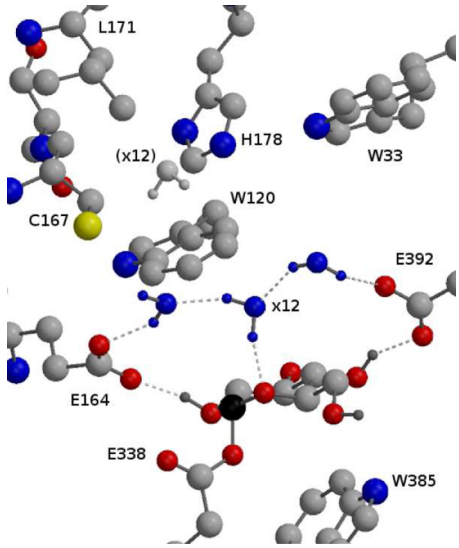


Figure 5. Jump of a water molecule on the C1 anomeric carbon after its exit from the water channel ending over the W120 aromatic ring. Prior to the jump (8 ps before), water molecule x12 (blue one) is hydrogen bonded to H178, whereas 79 ps before the jump it is on top of the W120 aromatic ring, close to the indole nitrogen (light-gray). The dashes pinpoint key pairs of oxygen atoms that are within hydrogen-bonding distance. In this snapshot of the simulation at t = 2.2 ns, the distance between the water x12 oxygen and the C1 anomeric carbon (black sphere) is 3.4 Å. Drawn with Molscript[37] and Raster3d.[47]

(Figure 5), whereas on the protein surface side, the E127-R133 salt bridge seems to control water entry/exit (Figure 4). These features allow us to take water molecules visiting the channel between W120 and E127 as probes for exploring its limits. As a result, the 10 buried residues that are able to establish, through their side chains, hydrogen bonds with them can be identified. On the enzyme active-site side, W33, C167, S168, and H178 are standing above and on both sides of the W120 aromatic moiety, whereas within the channel, W132 and H173 seem to act as relay sites between Y118 and the E127-R133 door on the protein-surface side. Note that W33 and W132 are both highly conserved and far away (~13 Å) from the catalytic carboxylates.

Two Other Internal Water Channels. To seek for other such channels in Ttβgly, time spans between the contacts of a given water molecule with the surface atoms were analyzed, focusing on water molecules remaining buried for 100 ps or more, and retaining those whose distance between their entry point into the protein (previous contact with a surface atom) and their exit point (new contact with a surface atom) is larger than 9.5 Å (three water−water hydrogen bonds) to avoid cases where a water molecule just hides in a pocket below the protein surface. Hereafter, the so-defined internal water channels of Ttβgly are named after their most frequent entry−exit amino acid residues. Note that among the 22 water molecules that remain under the protein surface for more than 1 ns after their entry inside the protein, 18 (82% of them) are visiting one of the three internal channels described below.

As shown in Table 4, although the E127-L171 internal water channel revealed by our analysis of the DXMS data is the longest one (W120 itself being rarely the last residue visited in the channel by a water molecule), there are two other channels in Ttβgly crossed by a similar number of water molecules during the simulation. E20-T23, although crossed by more

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900−5910

Table 4. Internal Water Channels of *T. thermophilus* β-Glycosidase Observed During a 50 ns MD Simulation at 350 K[a]

channel	main entry−exit	distance (Å)[b]	other entries−exits	total[c]
1	E20-T23 (13)	10.4	G46(4), E25(1), D34(1), S47(1)	20
2	E127-L171 (8)	16.5	W393(3), W120(2), L182(2)[d]	15
3	E392-Y397 (7)	10.0	F395(1), T398(1), R400(1)	10

[a]For each channel, the entry−exit points (amino acid residues) of the largest number of water molecules is given as well as the alternative entry or exit points observed, with the number of crossing events observed in each case given in parentheses. Only channels whose main entry−exit path is crossed by at least five water molecules are shown. [b]Mean distance between entry and exit points in the crystal structure. [c]Total number of crossing events considered. [d]L182 is close to the E127-R133 salt bridge.

water molecules, may look less interesting at first sight because E20 and T23 are both on the protein surface, far from the active site. However, in the crystal structure, E20 is hydrogen bonded to a water molecule belonging to cluster G21 (Table 1 and Figure 6a) that includes, as mentioned above, a hydrogen bonded chain of four highly conserved water molecules. Moreover, one of the molecules belonging to this chain is also hydrogen bonded to D121 and, like the three water molecules belonging to cluster D121 (Figure 6a), is among the few molecules found hydrogen bonded to four different amide hydrogens of peptide L117-A125 during our MD simulation (Table 2). Therefore, our MD simulation shows that water molecules can use channel E20-T23 to go from the protein surface to the hydration shell of D121. From there, they are also able to travel within the E127-L171 channel. Thus, through this long route inside the protein, they could also have access to the enzyme active site.

However, our MD simulation shows that all of these molecules are able to exchange with the bulk on the nanosecond time scale. Indeed, each site of cluster G21 is visited by several different water molecules and, at the end of the simulation, all four crystallographic water molecules have left the cluster. Note that x21 and x23 could have been expected to be tightly bound ones because they are both hydrogen bonding a pair of oppositely charged residues (Figure 6a).

The E392-Y397 water channel may also have a functional role. First, because E392 is one of the key residues lying at the bottom of the active site cleft.[48] Indeed, in the glycosyl-enzyme structure, it is hydrogen bonded to both the O4 and O6 atoms of the glycosyl moiety (Figure 6b). Therefore, its functional role could be to help recognize and/or position precisely the glycosyl moiety. Second, in the crystal structure, the two water molecules of cluster E392 together with the single molecule of cluster R399 form an almost linear chain going from the active site toward the protein surface, with x46 being only 3.5 Å away from x22 and in contact with backbone atoms of Y397, one of the end points of the channel (Figure 6b). Because the precise positioning of x17 is secured by hydrogen bonds with the side chains of N390, E392, and W385, this water molecule could help rigidify E392 during the catalytic process as well as, through W385 and E392, the glycosyl moiety. Indeed, it is well known that substrate distortion is a key step of the reaction catalyzed by glycoside hydrolases.[49] Therefore, it is tempting to

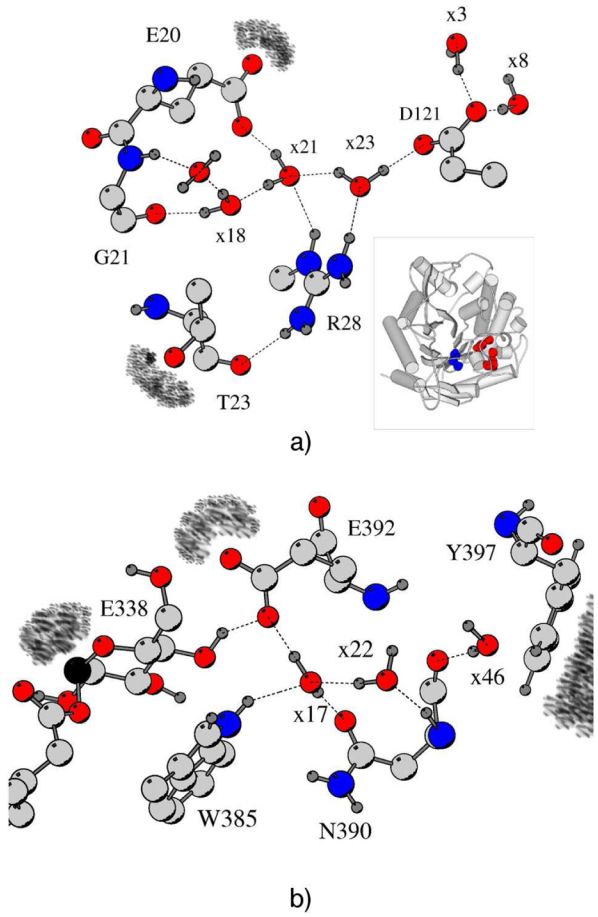


Figure 6. Two other internal water channels observed in the *T. thermophilus* β-glycosidase structure. (a) E20-T23 channel. Both entries of this channel are on the protein surface. They provide access to a chain of four highly conserved water molecules belonging to cluster G21 but also, through hydrogen bonding with D121, to the conserved cluster named after this residue. (b) E392-Y397 channel. This channel runs straight from the bottom of the enzyme active site cleft (E338 and E392) to the protein surface (Y397). It corresponds to a chain of three highly conserved water molecules belonging to clusters E392 and R399. The C1 anomeric carbon is depicted as a black sphere. Atoms exposed to bulk water, that is, either on the protein surface or at the bottom of the active site cleft, are shadowed. Dashes pinpoint key hydrogen bonds. The locations in the structure of clusters G21, D121 (both in red), E392, and R399 (both in blue) are specified in the inset.

speculate that networks of water molecules may be required for allowing such a distortion, for instance, by freezing the environment of the substrate to favor the distorted state. In turn, this would explain why cluster E392 is so well conserved (Table 1). In this respect, note the key role of N390, which is hydrogen bonded to all three highly conserved water molecules observed in this channel. Interestingly, the N390I mutation has been found to affect hydrolysis much more than the spontaneous transglycosylation activity of Tt*β*gly.[3] This could mean that the hydrolytic activity of Tt*β*gly requires a more precise positioning of the glycosyl moiety during the reaction process than the transglycosylation one.

However, the fact that, as observed in our MD simulation, the water molecules in channel E392-Y397 can exchange quickly with the bulk (after 22.3 ns all three crystallographic water molecules have left the channel) suggests that its

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900−5910

functional role could also have a more dynamical aspect. For instance, it may help to remove water molecules sitting at the bottom of the active site, allowing the incoming substrate to have easier access to the deepest part of the cleft, in order to become hydrogen bonded with E392.

Water-Channel Conservation Across GH1. To check whether the internal water channels of Ttβgly are as well conserved as many of its crystallographic water molecules, a second 50 ns MD simulation at 350 K was performed, starting from the glycosyl-enzyme structure of *Thermotoga maritima* (Tmβgly), another thermophilic organism. Note that the sequences of both enzymes are 48% identical. Indeed, *T. maritima* is the organism the closest to *T. thermophilus* for which a structure of a glycosyl-enzyme intermediate has been determined (PDB ID 1OIN[44]).

As shown in Table 5, in this case, only two main internal water channels are identified. Although significantly different in their details from those of Ttβgly, they share several key residues. Indeed, although channel Y19-W406 is shorter than its counterpart, its second main exit is the same as Ttβgly channel E127-L171, namely, W122 (used six times), which corresponds to W120 in Ttβgly. Reciprocally, W406 corresponds to Ttβgly W393, the alternative exit of Ttβgly channel E127-L171 (used three times; Table 4). Therefore, on the enzyme active-site side, this channel is conserved. However, on the other side, Y19 is in contact with P25, which corresponds to Ttβgly T23, as though Ttβgly channels E20-T23 and E127-L171 were fused in Tmβgly.

The third Ttβgly channel, E392-Y397, is conserved on both ends. Indeed, in Tmβgly, its main entry on the protein surface side is the same, namely, Y410, which corresponds to Ttβgly Y397. Although M322 is also on the protein surface, E405, which corresponds to Ttβgly E392 (note that it is on the active-site side), is used eight times, that is, as many times as E392 during our simulation of Ttβgly (Table 4).

Table 5. Internal Water Channels of *T. maritima* β-Glucosidase Observed During a 50 ns MD Simulation at 350 K[a]

channel	main entry−exit	distance (Å)[b]	other entries−exits	total[c]
1	Y19-W406 (17)	10.3	W122(6), N44(1), H180(1)	25
2	M322-Y410 (9)	11.0	E405(8), K46(1), N47(1), F312(1), I326(1), W398(1), F414(1)	23

[a]For each channel, the entry−exit points (amino acid residues) of the largest number of water molecules is given as well as the alternative exit−entry points observed, with the number of crossing events observed in each case given in parentheses. Only channels whose main entry−exit path is crossed by at least five water molecules are shown. The residues corresponding to those involved in the three main channels found in *T. thermophilus* β-glycosidase are underlined. [b]Mean distance between entry and exit points in the crystal structure. [c]Total number of crossing events.

CONCLUSIONS

Twenty crystallographic water molecules from the structure of Ttβgly, the β-glycosidase of *T. thermophilus*, are found in 67−97% of GH1 structures (Table 1), with seven of them being found in more than 90% of GH1 structures. Such a high level of conservation across evolution means that all these water molecules have important roles. In the case of the isolated water molecule that is hydrogen bonding the backbones of A341, W385, and F401 (Table 1), it is likely to be a structural one. However, because many of these conserved molecules are organized in chains that are lying near the enzyme active site (Figure 1), they may prove to be involved more directly in the enzyme function.

To gain insights into the possible role of these conserved water molecules, deuterium-exchange mass spectroscopy was used. Deuterium−hydrogen exchange is indeed a method of choice for probing internal water dynamics because hydrogen bonding is required for allowing the exchange to take place. In the case of Ttβgly, it revealed that hydrogen−deuterium exchange happens to be particularly quick in peptide L117-A125, although, in the crystal structure, the only backbone atoms of this peptide that are accessible to solvent are those of A125. Such a result may prove to have a functional significance because the aromatic moiety of W120 is in contact with the side chain of E164, the acid−base residue involved in the catalytic mechanism (Figure 5).

To help interpret the DXMS data, a molecular dynamics simulation was performed. On the nanosecond time scale, at T = 350 K, the backbone amides of peptide Y118-D121 were found to be accessible by two routes. First, through a long water channel ending over the W120 aromatic moiety, with its main entry on the protein surface being the E127-R133 salt bridge (Figure 4). Strikingly, water molecules leaving this channel on the side of the enzyme active site are able to jump directly on the C1 anomeric carbon (Figure 5) whose linkage with E338, the nucleophile residue, is to be hydrolyzed in the next step of the catalytic mechanism.

However, in our MD simulation of Ttβgly, only a small subset of water molecules were found to be able to become hydrogen bonded with all backbone amides of Y118-L122 (Table 3), but it is through another route because these water molecules belong, in the crystal structure, to the hydration shell of D121. Interestingly, D121 is both buried and highly conserved, with its hydration shell being also highly conserved, as it is observed in more than 90% of GH1 structures (Table 1). Moreover, in the crystal structure, D121 is also hydrogen bonded to a water molecule belonging to a chain of four conserved water molecules (Figure 6a) that are able to exchange with the bulk on the nanosecond time scale, as also shown during our MD simulation. Therefore, although during our simulation, probably as a consequence of the relatively long residence time of water molecules in the hydration shell of D121, no bulk water molecule was found able to reach the backbone amides of peptide Y118−W120 through this route, it also seems to be a likely one for explaining the DXMS data.

However, during our MD simulation, another chain of conserved water molecules was found able to exchange with the bulk on the nanosecond time scale. Overall, three rather long and often narrow water channels have been found in Ttβgly. One of them ends close to the acid−base residue, the second joins the former at the level of a buried, highly conserved aspartate, and the last one ends close to the nucleophile residue. It is tempting to speculate that at least one of these channels could be involved in providing the catalytically added water molecule. However, the fact that high water-exchange rates, namely, on the nanosecond time scale, are observed within these channels calls for another possible role that is associated with the need to remove water from the bottom of the enzyme active site when the incoming substrate approaches the catalytic residues. Within the frame of this later hypothesis, the

dx.doi.org/10.1021/bi400260b | *Biochemistry* 2013, 52, 5900−5910

narrowness of the channels would be a consequence of the necessity for the environment of the active site to also remain rigid, for the enzyme to achieve its function, during which substrate distortion needs to be performed.[49]

■ ASSOCIATED CONTENT

Supporting Information

Method details for the crystallographic water clustering. Behavior of the glycosyl moiety and the accessibility of the anomeric carbon during the molecular dynamics simulation of Ttβgly. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: yves-henri.sanejouand@univ-nantes.fr; Tel: +33251125776.

Funding

[§]A GENCI-CINES starting grant of computer time du directeur is acknowledged as well as financial support from Région Pays de la Loire through the GLYCONET network.

Notes

The authors declare no competing financial interest.

[⊥]Deceased on September 9th, 2012.

■ ACKNOWLEDGMENTS

D.T. thanks Henry Guan, Sheng Li, and Jun Lee for their help during his stay in the laboratory of the late professor Virgil L. Woods. Preliminary mass spectrometry studies were performed at the BIBS facility of the INRA Angers-Nantes Center.

■ REFERENCES

(1) Zechel, D. L., and Withers, S. G. (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res. 33*, 11−18.

(2) McCarter, J. D., and Withers, S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol. 4*, 885−892.

(3) Feng, H. Y., Drone, J., Hoffmann, L., Tran, V., Tellier, C., Rabiller, C., and Dion, M. (2005) Converting a β-glycosidase into a β-transglycosidase by directed evolution. *J. Biol. Chem. 280*, 37088−37097.

(4) Koné, F. M. T., Le Béchec, M., Sine, J. P., Dion, M., and Tellier, C. (2009) Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng., Des. Sel. 22*, 37−44.

(5) Amaya, M. F., Buschiazzo, A., Nguyen, T., and Alzari, P. M. (2003) The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T. cruzi* trans-sialidase. *J. Mol. Biol. 325*, 773−784.

(6) Paris, G., Ratier, L., Amaya, M. F., Nguyen, T., Alzari, P. M., and Frasch, A. C. C. (2005) A sialidase mutant displaying trans-sialidase activity. *J. Mol. Biol. 345*, 923−934.

(7) Watts, A. G., Oppezzo, P., Withers, S. G., Alzari, P. M., and Buschiazzo, A. (2006) Structural and kinetic analysis of two covalent sialosyl-enzyme intermediates on *Trypanosoma rangeli* sialidase. *J. Biol. Chem. 281*, 4149−4155.

(8) Amaya, M. F., Watts, A. G., Damager, I., Wehenkel, A., Nguyen, T., Buschiazzo, A., Paris, G., Frasch, A. C., Withers, S. G., and Alzari, P. M. (2004) Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. *Structure 12*, 775−784.

(9) Damager, I., Buchini, S., Amaya, M. F., Buschiazzo, A., Alzari, P., Frasch, A. C., Watts, A., and Withers, S. G. (2008) Kinetic and mechanistic analysis of *Trypanosoma cruzi* trans-sialidase reveals a classical ping-pong mechanism with acid/base catalysis. *Biochemistry 47*, 3507−3512.

(10) Rebuffet, E., Groisillier, A., Thompson, A., Jeudy, A., Barbeyron, T., Czjzek, M., and Michel, G. (2011) Discovery and structural characterization of a novel glycosidase family of marine origin. *Environ. Microbiol. 13*, 1253−1270.

(11) Meyer, E. (1992) Internal water molecules and H-bonding in biological macromolecules: A review of structural features with functional implications. *Protein Sci. 1*, 1543−1562.

(12) Knight, J. D. R., Hamelberg, D., McCammon, J. A., and Kothary, R. (2009) The role of conserved water molecules in the catalytic domain of protein kinases. *Proteins 76*, 527−535.

(13) Oprea, T. I., Hummer, G., and Garca, A. E. (1997) Identification of a functional water channel in cytochrome P450 enzymes. *Proc. Natl Acad. Sci. U.S.A. 94*, 2133−2138.

(14) Fishelovitch, D., Shaik, S., Wolfson, H. J., and Nussinov, R. (2010) How does the reductase help to regulate the catalytic cycle of cytochrome P450 3A4 using the conserved water channel? *J. Phys. Chem. B 114*, 5964−5970.

(15) Haeffner, F., and Norin, T. (1999) Molecular modelling of lipase catalysed reactions. Prediction of enantioselectivities. *Chem. Pharm. Bull. 47*, 591−600.

(16) Cachau, R., Howard, E., Barth, P., Mitschler, A., Chevrier, B., Lamour, V., Joachimiak, A., Sanishvili, R., Van Zandt, M., Sibley, E., Moras, D., and Podjarny, A. (2000) Model of the catalytic mechanism of human aldose reductase based on quantum chemical calculations. *J. Phys. IV 10*, 3−13.

(17) Zhang, X., and Bruice, T. C. (2007) Histone lysine methyltransferase SET79: Formation of a water channel precedes each methyl transfer. *Biochemistry 46*, 14838−14844.

(18) Delalande, O., Sacquin-Mora, S., and Baaden, M. (2011) Enzyme closure and nucleotide binding structurally lock guanylate kinase. *Biophys. J. 101*, 1440−1449.

(19) Gohlke, H., Schlieper, D., and Groth, G. (2012) Resolving the negative potential side (n-side) water-accessible proton pathway of F-type ATP synthase by molecular dynamics simulations. *J. Biol. Chem. 287*, 36536−36543.

(20) Pavlova, M., Klvana, M., Prokop, Z., Chaloupkova, R., Banas, P., Otyepka, M., Wade, R. C., Tsuda, M., Nagata, Y., and Damborsky, J. (2009) Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol. 5*, 727−733.

(21) Chelikani, P., Carpena, X., Fita, I., and Loewen, P. C. (2003) An electrical potential in the access channel of catalases enhances catalysis. *J. Biol. Chem. 278*, 31290−31296.

(22) Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The carbohydrate-active enzymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res. 37*, D233−D238.

(23) Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M. J., Natale, D. A., O' Donovan, C., Redaschi, N., and Yeh, L. S. (2004) UniProt: The universal protein knowledgebase. *Nucleic Acids Res. 32*, D115−D119.

(24) Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol. 112*, 535−542.

(25) Kouranov, A., Xie, L., De La Cruz, J., Chen, L., Westbrook, J., Bourne, P. E., and Berman, H. M. (2006) The RCSB PDB information portal for structural genomics. *Nucleic Acids Res. 34*, D302−D305.

(26) Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res. 31*, 3497−3500.

(27) Rupley, J. A., Gratton, E., and Careri, G. (1983) Water and globular proteins. *Trends Biochem. Sci. 8*, 18−22.

(28) Beglov, D., and Roux, B. (1995) Dominant solvation effects from the primary shell of hydration: Approximation for molecular dynamics simulations. *Biopolymers 35*, 171−178.

(29) Hamaneh, M. B., and Buck, M. (2007) Acceptable protein and solvent behavior in primary hydration shell simulations of hen lysozyme. *Biophys. J. 92*, L49−L51.
(30) Brooks, B. R., et al. (2009) CHARMM: The biomolecular simulation program. *J. Comput. Chem. 30*, 1545−1614.
(31) Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926−935.
(32) Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys. 23*, 327−341.
(33) Lindorff-Larsen, K., Piana, S., Dror, R. O., and Shaw, D. E. (2011) How fast-folding proteins fold. *Science 334*, 517−520.
(34) Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187−217.
(35) Nerinckx, W., Desmet, T., and Claeyssens, M. (2003) A hydrophobic platform as a mechanistically relevant transition state stabilising factor appears to be present in the active centre of all glycoside hydrolases. *FEBS Lett. 538*, 1−7.
(36) Weiss, M. S., Jabs, A., and Hilgenfeld, R. (1998) Peptide bonds revisited. *Nat. Struct. Biol. 5*, 676−676.
(37) Kraulis, P. (1991) Molscript: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr. 24*, 946−950.
(38) Hvidt, A., and Wallevik, K. (1972) Conformational changes in human serum albumin as revealed by hydrogen-deuterium exchange studies. *J. Biol. Chem. 247*, 1530−1535.
(39) Wagner, G., and Wüthrich, K. (1979) Structural interpretation of the amide proton exchange in the basic pancreatic trypsin inhibitor and related proteins. *J. Mol. Biol. 134*, 75−94.
(40) Englander, J. J., Del Mar, C., Li, W., Englander, S. W., Kim, J. S., Stranz, D. D., Hamuro, Y., and Woods, V. L. (2003) Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. U.S.A. 100*, 7057−7062.
(41) Chung, K. Y., Rasmussen, S. G. F., Liu, T., Li, S., DeVree, B. T., Chae, P. S., Calinski, D., Kobilka, B. K., Woods, V. L., Jr, and Sunahara, R. K. (2011) Conformational changes in the G protein Gs induced by the β_2 adrenergic receptor. *Nature 477*, 611−615.
(42) Laskowski, R. A. (2001) PDBsum: Summaries and analyses of PDB structures. *Nucleic Acids Res. 29*, 221−222.
(43) Dion, M., Fourage, L., Hallet, J. N., and Colas, B. (1999) Cloning and expression of a β-glycosidase gene from *Thermus thermophilus*. Sequence and biochemical characterization of the encoded enzyme. *Glycoconjugate J. 16*, 27−37.
(44) Zechel, D., Boraston, A., Gloster, T., Boraston, C., Macdonald, J., Tilbrook, D., Robert, V., and Davies, G. (2003) Iminosugar glycosidase inhibitors: Structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to β-glucosidases. *J. Am. Chem. Soc. 125*, 14313−14323.
(45) Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) The crystal structures of *Sinapis alba* myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure 5*, 663−676.
(46) Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Oonanant, W., Bevan, D. R., Esen, A., Chen, C. J., Opassiri, R., Svasti, J., and Ketudat Cairns, J. R. (2008) Structural insights into rice BGlu1 β-glucosidase oligosaccharide hydrolysis and transglycosylation. *J. Mol. Biol. 377*, 1200−1215.
(47) Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic molecular graphics. *Methods Enzymol. 277*, 505−524.
(48) Marana, S. R., Terra, W. R., and Ferreira, C. (2002) The role of amino-acid residues Q39 and E451 in the determination of substrate specificity of the *Spodoptera frugiperda* β-glycosidase. *Eur. J. Biochem. 269*, 3705−3714.
(49) Vocadlo, D. J., and Davies, G. J. (2008) Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol. 12*, 539−555.