Who's on base? Revealing the catalytic mechanism of inverting family 6 glycoside hydrolases†

Heather B. Mayes,[ab] Brandon C. Knott,[b] Michael F. Crowley,[c] Linda J. Broadbelt,[a] Jerry Ståhlberg[*d] and Gregg T. Beckham[*b]

In several important classes of inverting carbohydrate-active enzymes, the identity of the catalytic base remains elusive, including in family 6 Glycoside Hydrolase (GH6) enzymes, which are key components of cellulase cocktails for cellulose depolymerization. Despite many structural and kinetic studies with both wild-type and mutant enzymes, especially on the *Trichoderma reesei* (*Hypocrea jecorina*) GH6 cellulase (*Tr*Cel6A), the catalytic base in the single displacement inverting mechanism has not been definitively identified in the GH6 family. Here, we employ transition path sampling to gain insight into the catalytic mechanism, which provides unbiased atomic-level understanding of key order parameters involved in cleaving the strong glycosidic bond. Our hybrid quantum mechanics and molecular mechanics (QM/MM) simulations reveal a network of hydrogen bonding that aligns two active site water molecules that play key roles in hydrolysis: one water molecule drives the reaction by nucleophilic attack on the substrate and a second shuttles a proton to the putative base (D175) *via* a short water wire. We also investigated the case where the putative base is mutated to an alanine, an enzyme that is experimentally still partially active. The simulations predict that proton hopping along a water wire *via* a Grotthuss mechanism provides a mechanism of catalytic rescue. Further simulations reveal that substrate processive motion is 'driven' by strong electrostatic interactions with the protein at the product sites and that the −1 sugar adopts a ^2S_O ring configuration as it reaches its binding site. This work thus elucidates previously elusive steps in the processive catalytic mechanism of this important class of enzymes.

Introduction

Glycoside hydrolase (GH) enzymes are produced by all kingdoms of life.[1,2] They play vital roles in the turnover of carbon in nature,[3] fundamental mechanisms of cell biology, industrial biotechnology, and human health.[4–7] These ubiquitous enzymes are currently categorized into 135 families, catalogued on the Carbohydrate Active Enzymes database (CAZy, http://www.cazy.org).[1,2] One particularly important role of GH enzymes involves the turnover of carbohydrates, such as

cellulose, hemicellulose, and chitin in the biosphere, where enzymes from multiple GH families are secreted by microbes to produce sugars for food. Industrially, these natural systems are often starting points for designer enzyme cocktails to conduct cellulose depolymerization, with the ultimate aim of producing sugars for the production of renewable chemicals and fuels.[8] Given their fundamental and industrial importance, studying the elementary mechanisms of cellulose depolymerization could suggest methods for improving efficiency, which may profoundly impact the availability and affordability of cellulosic biofuels.[3,9] While significant advances have been made in characterizing cellulolytic GHs, many questions regarding their structure and function remain unanswered.[3]

Based on experimental observations of stereochemistry, Koshland postulated two general hydrolytic mechanisms that GHs employ for enzymatic action on anomeric carbon atoms: a single-displacement inverting mechanism or a double-displacement retaining mechanism.[10] In the first step of the double-displacement mechanism, a nucleophilic enzyme residue attacks the anomeric carbon while an acidic residue donates a proton to the glycosidic oxygen. In the second step, a water molecule attacks the anomeric carbon such that its hydroxide group bonds to the anomeric carbon, restoring the original stereochemistry and causing deglycosylation, while its

[a]*Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA*

[b]*National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO 80401, USA. E-mail: gregg.beckham@nrel.gov*

[c]*Biosciences Center, National Renewable Energy Laboratory, Golden, CO 80401, USA*

[d]*Department of Chemistry and Biotechnology, Swedish University of Agricultural Sciences, SE-75007, Uppsala, Sweden. E-mail: jerry.stahlberg@slu.se*
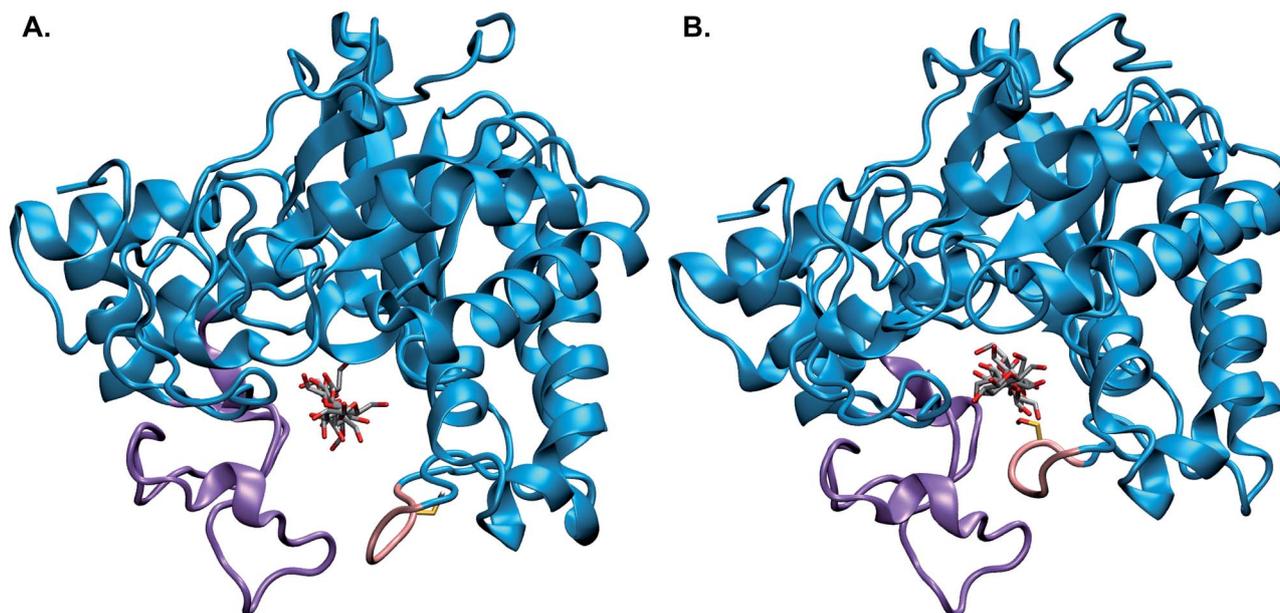
Fig. 1 The catalytic domain of *Tr*Cel6A showing the substrate binding-site tunnel enclosed by the N-terminal loop (purple) and the active-center (C-terminal) loop (pink) in the (A) "open" and (B) "closed" positions, with the entrance to the active site tunnel in the foreground. These images come from the processivity and hydrolysis simulations, respectively, using models based on crystal structures as described in the Computational methods.

proton transfers to the acidic residue, resetting the active site charge distribution for subsequent catalysis. Conversely, in the single-displacement mechanism, the water molecule attacks the anomeric carbon simultaneously with the acidic residue, donating a proton to the glycosidic oxygen and inverting the stereochemistry at the anomeric carbon. The required positioning of the attacking water and acidic residues prevents the proton of the attacking water from resetting the charge of the acidic residue in this single step. An acceptor must absorb the proton, and the active site charge distribution must be reset in a separate step.

Almost all known GH enzymes follow one of these two proposed mechanisms involving protein residues serving as catalytic acids and bases,[3,11] and identifying which parts of the enzymes serve as proton acceptors and proton donors is an active area of investigation.[12–14] Simulation can complement these experimental advances by providing dynamic, spatiotemporal models of enzyme mechanisms.[15–17] These models can test hypotheses and provide atomic-level insights toward the development of structure–function relationships.

Polysaccharide-active GHs are also characterized according to whether they perform non-processive chain hydrolysis to create new chain ends or processively cleave glycosidic bonds along a polymer chain.[3] These activities are synergistic, with the bulk of hydrolysis accomplished *via* processive action due to the efficiency of hydrolyzing multiple bonds in between enzyme-substrate binding and unbinding.[18,19] Recently, our group identified the molecular details of the catalytic[20] and processive[21] action of the retaining GH *T. reesei* Cel7A (*Tr*Cel7A), the main component of industrial enzyme cocktails for cellulose decomposition. The catalytic mechanism study identified key, non-intuitive contributions to the reaction coordinate (RC)

including protein side-chain conformational changes and a product-assisted step. The processivity study revealed that electrostatic interactions with the leading substrate glycosyl ring provide the driving force for chain translocation.

Additional computational studies validated Koshland's postulated mechanisms for other GH[22,23] and glycosyltransferase (GT) enzymes.[24–26] Yet, detailed mechanisms of other key GH enzymes have remained elusive, including the inverting *Tr*Cel6A (formerly designated CBHII[27]) shown in Fig. 1.[3,28,29] This enzyme is industrially significant due to its synergistic action with *Tr*Cel7A.[8,30–32] *Tr*Cel6A was the first reported fungal
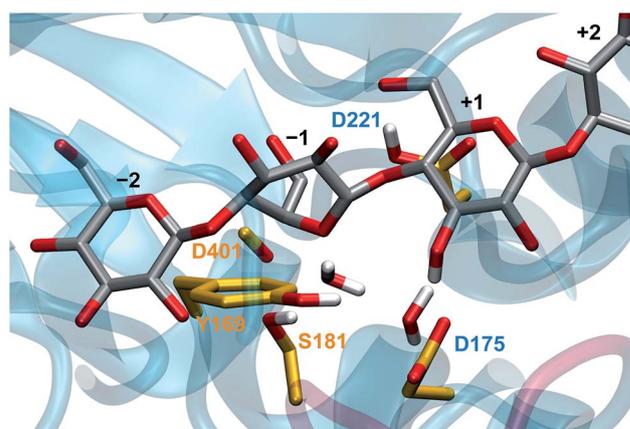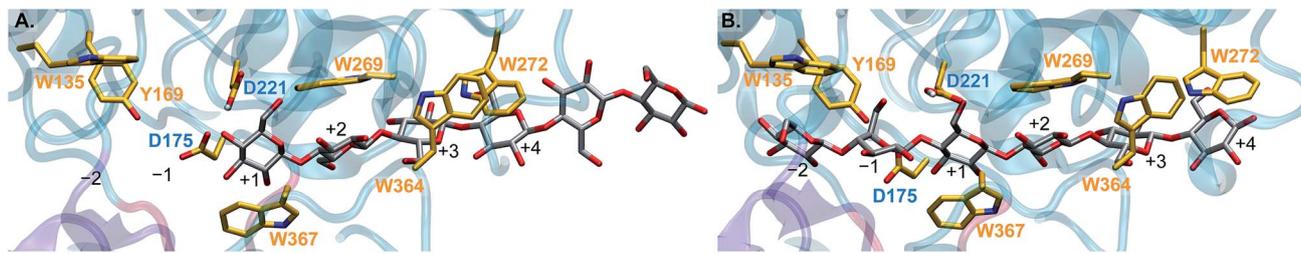

Fig. 2 The *Tr*Cel6A active site from the hydrolysis simulations, based on crystal structures as described in the Computational methods, several key residues and substrate binding sites labeled according to the conventional scheme.[41,42] The two water molecules shown are present in the crystal structure.

Fig. 3 Cross-sectional view of the *Tr*Cel6A tunnel in the (A) "pre-slide" and (B) "slide" modes from processivity simulations performed in this work as described below. Aromatic residues in the tunnel that interact with the substrate are shown and labeled in orange, and the substrate binding sites are labeled in black. The residues proposed to donate (D221) and accept (D175) a proton in the hydrolysis reaction are labeled in blue. Hydrolysis cleaves the glycosidic bond joining the glycosyl units at the −1 and +1 substrate binding sites.

cellulase crystal structure,[33] and since that time, GH6 cellulases have been the subject of multiple structure–function studies to elucidate the catalytic mechanism.[3] These studies led to the hypothesis of a four-step processive catalytic cycle involving (1) procession of the substrate into the active site, which occurs with the active site loops in the "open" position (Fig. 1A), (2) loop closure/activation, (3) hydrolysis with the active site loops in the "closed" position (Fig. 1B), and (4) product expulsion/loop opening.[3,29]

Experimental studies confirmed that D221 (shown in Fig. 2) acts as the catalytic acid,[34] yet the identification of its catalytic base has remained elusive, even leading to speculation that *Tr*Cel6A does not employ a proton-accepting residue (catalytic base).[35] Koivula *et al.*[34] proposed that the D175 residue can accept the excess proton through a short water wire (one bridge water molecule between the nucleophilic water and D175), rather than directly from the nucleophilic water, *via* the Grotthuss mechanism.[36] Water molecules have been resolves in the active-site of the substrate-bound *Tr*Cel6A[37] and *Humicola insolens* Cel6A (*Hi*Cel6A),[38] in accordance with this hypothesis. Studies of *Thermobifida fusca* Cel6B (*Tf*Cel6B) are also consistent with involvement of a proton-shuttling network,[39] including identification of two water molecules in the crystal structure of substrate-bound WT enzyme, leading Sandgren *et al.* to propose that a Grotthuss mechanism is also responsible for proton transfer in this bacterial member of the GH6 family.[40] However, Koivula *et al.*[34] also demonstrated that the enzyme retained 2–3% residual activity when the aspartic acid is mutated to the non-proton accepting alanine, rather than the null activity expected if D175 acted as a traditional GH catalytic base. A proposed mechanism to explain the residual activity is catalytic rescue by a longer water wire shuttling the excess proton to another proton acceptor or even to bulk water.[3,29]

To investigate this intriguing mechanism, we used a QM/MM molecular dynamics (MD) to simulate the catalytic domain of *Tr*Cel6A, allowing us to model bond rearrangement at the catalytic center of the enzyme. To elucidate the RC, we employed aimless shooting (AS) with likelihood maximization.[43,44] This transition path sampling (TPS) method[45] explores the transition state (TS) ensemble without modifying the simulation by bias potential along a preconceived RC.[46] The potential of mean force (PMF), which describes how the free energy changes along the RC, for the reaction was then

determined by equilibrium path sampling (EPS).[47] We also performed simulations of the *Tr*Cel6A D175A mutant to investigate mechanisms of catalytic rescue to account for the observed residual activity after mutation of the putative base.

Furthermore, we simulated the processive action wherein, with the active site loops in the "open" position (Fig. 1A), the substrate advances from its position immediately after hydrolysis and product expulsion (the "pre-slide" mode, Fig. 3A) into position for the next catalytic event (the "slide" mode, Fig. 3B).[3,29] Carbohydrate–aromatic interactions in the tunnel have been reported to play key roles in processivity,[18,29,48–51] and are shown and labeled in orange. The aspartic acid residues 175 and 221 are labeled in blue to indicate where the hydrolysis reaction will take place. The black labels identify the substrate binding sites +1 through +4.[3] The substrate enters from the right side of the figure (binding site +4). Binding sites −2 and −1 are occupied in the "slide" mode but not in the "pre-slide" mode. Hydrolysis breaks the glycosidic bond between the −1 and +1 binding sites.

These simulations allow us to calculate energy barriers and identify the interactions for these two important parts of the catalytic cycle. In combination with previous studies of product expulsion[52] and active site loop opening and closing,[37,53] this work greatly expands our understanding of the catalytic cycle of this industrially important enzyme.[3,29] Additionally, the findings presented here are expected to aid advances in the understanding of cellulase function and to apply to other inverting carbohydrate-active enzymes. As previously noted, while the mechanisms for some inverting GH enzymes have been reported,[22,23] they have been elusive for others beyond those in GH6, such as for the *Clostridium thermocellum* Cel124 inverting cellulase endo-β-1,4-glucanase, and it has been suggested that the water wire hypothesized in GH6 hydrolysis could also enable hydrolysis in the GH124 family enzymes.[54] Furthermore, identifying catalytic residues for some inverting glycosyltransferases[55–57] and DNA glycosylases[58] have also proven elusive, sparking debate and prompting analogies with *Tr*Cel6A.[58,59]

Computational methods

Detailed procedures are documented in the ESI† with an overview provided here. The CHARMM[60] package was used to build and equilibrate all systems and for some trajectory analyses.

The CHARMM36 force field was used for all simulations and all parts of the system: the enzyme,[61–63] carbohydrate ligands,[64,65] and sodium ions[66] with the TIP3P model for water.[67] All simulations were conducted at 300 K. After initial minimization and heating steps in the NVT ensemble, the remaining simulations employed the NPT ensemble with the pressure set to 1.0 bar.

For the hydrolysis simulations, the TrCel6A Michaelis complex was based on Protein Data Bank (PDB)[68] ID 1QJW[37] (TrCel6A Y169F in complex with Glc_2–S–Glc_2; with closed active-center loop) with the following modifications: the addition of the Tyr85 residue from PDB ID 1QK2[37] (TrCel6A WT in complex with Glc_2–S–Glc_2; with open active-center loop); replacing the mutant Phe169 with the Tyr169 from PDB ID 1QK2; using the Asp221 conformation from PDB ID 1HGW[34] (apo TrCel6A D175A mutant; with open active-center loop); and using the cellohexaose ligand from PDB ID 4AVO[69] (TfCel6B) D274A; with open active-center loop), with the leading four glycosyl units aligned to the tetramer in PDB ID 1QJW. This model was used to generate the mutant D175A by replacing Asp175 with Ala175 from crystal structure 1HGW of TrCel6A D175A.[34] For the processivity study, the TrCel6A "slide" and "pre-slide" conformations were based on PDB ID 1QK2 and the cellohexaose ligand from PDB ID 4AVO. For the "pre-slide" mode, the −1 and −2 ligand monomers were removed, and monomers were added in the +5 and +6 positions, in the same starting geometry as the +3 and +4 monomers. After the initial equilibration in CHARMM, the remaining simulations were primarily performed using the Amber[70] software package version 12. Some unrestrained MM-only simulations were run in NAMD[71] starting from the minimized and equilibrated structures from CHARMM.

In the hydrolysis simulations, the QM region was modeled with the self-consistent charge density functional tight-binding (SCC-DFTB) method with second-order terms in the charge density fluctuations.[72,73] The choice of this method is discussed in the ESI.† The ESI† also details the procedures employed for AS, likelihood maximization, committor analysis, and EPS. The processivity simulations employed umbrella sampling (US)[74] and the weighted histogram analysis method (WHAM)[75–77] to produce the PMF, with further details included in the ESI.† Images of the proteins were created in VMD.[78] Additional tools utilized in this work include CPPTRAJ[79] and CHAMBER13.[80] Plots were created in Igor Pro version 6.36 (Wavemetrics, Lake Oswego, OR, USA), with data for electrostatic and van der Waals interactions smoothed using the binomial algorithm with 250 passes over the 12 535 data points created by averaging the energy values within the collective variable (CV) bins of width 0.001 Å.

Results and discussion

WT hydrolysis

For TrCel6 to follow the canonical inverting mechanism, eight bond rearrangements (four cleaved and four formed, see Fig. 4) must occur in one step. As described in detail in the ESI,† simulation of this reaction began with biased simulations to create 48 potential TS structures, based on a range of inter-atomic distances for the eight bond rearrangements involving the WT protein, the substrate, and two active site water molecules. 37 of these hypothesized structures yielded "reactive" trajectories linking the reaction and product structures shown
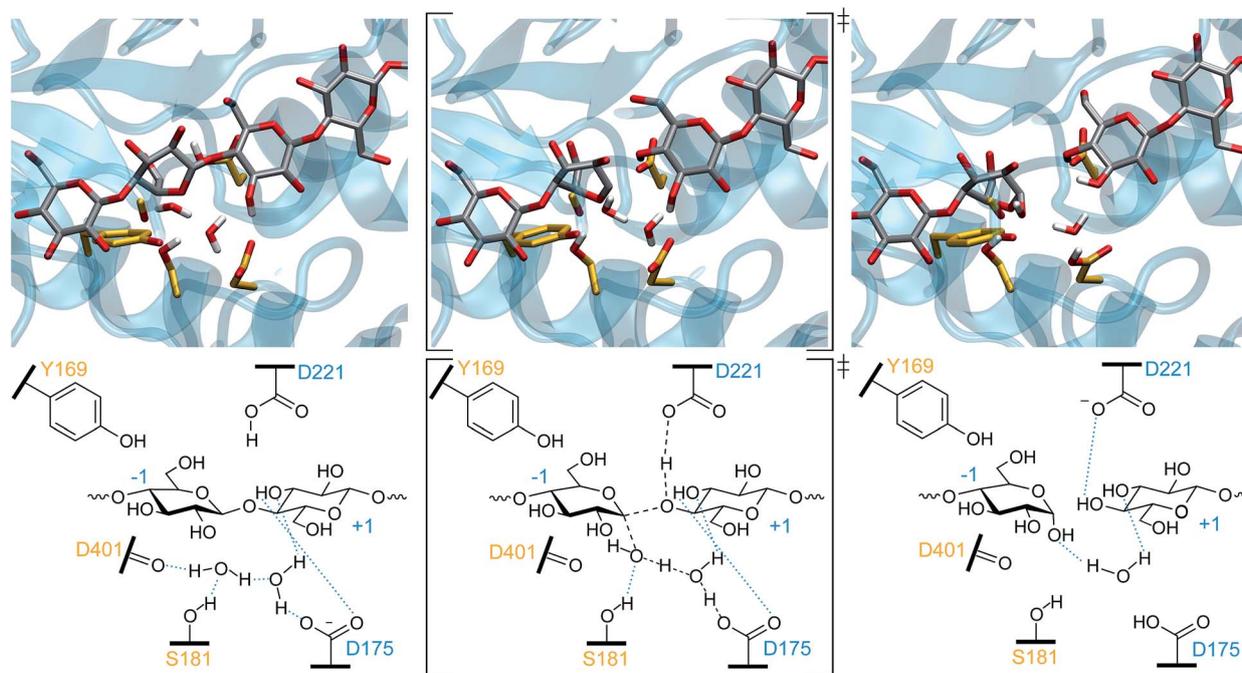

Fig. 4 The one-step inverting hydrolysis reaction by TrCel6A. In the reactant conformation (left), the −1 glucopyranoside ring is in a ^2S_O conformation. It is in a $^{2,5}B$ conformation in both the TS (middle) and product state (right). Intermediate bond lengths are indicated in the TS by dashed black lines. Blue dotted lines indicate hydrogen bonding.

in Fig. 4, supporting the hypothesis that D175 can accept the proton from the nucleophilic water *via* a second bridge water (Fig. 4). The biased MD simulations were followed with unbiased AS simulations allowing collection of thousands of conformations of the TS ensemble. As noted in the Computational methods, these TPS simulations use a Hamiltonian that is not biased along a preconceived RC, as employed by methods such as US[74] or metadynamics.[81]

We used likelihood maximization to test 87 order parameters (OPs), such as atomic distances or dihedral angles, to determine which OP or combination of OPs best represent the RC (see the ESI† for the full list of OPs tested). We excluded data from the first 500 AS points of each run to allow for decorrelation from the initial TS guesses. Another 2000 points were obtained in each of 32 independent runs. Likelihood maximization was used to identify the best RC and determine its parameters using half or all of the 64 000 resulting points to test for converged AS simulations and RC identification. A list of the best scoring one-, two-, and three-parameter RCs is included in the ESI.†

Of the set of tested OPs, the only ones for which inclusion in the RC significantly improved the likelihood score (as measured by the Bayesian information criterion[43,82]) were distances between atoms with bonds forming and breaking which included at least one atom of the nucleophilic water. The key distances identified are captured in the best identified two-parameter RC, OP6 and OP12 (Fig. 5). As shown, OP12 is a difference of distances involving the nucleophilic water oxygen: the distance between the oxygen atom and one hydrogen nucleus of the nucleophilic water (bond breaking), minus the distance between the nucleophilic water oxygen and the −1 glucopyranoside anomeric carbon (bond forming). OP6 is the distance between the hydrogen nucleus and the bridge water oxygen atom (bond forming). Committor analysis[83] was performed with this RC (a linear combination of OP6 and OP12), resulting in a relatively flat histogram. Sampling errors and the complexity of the RC likely contribute to deviations from an ideal committor.[20,84] To correct for the resulting overestimation of the rate coefficient,[85] we calculated the transmission coefficient for this RC, as discussed below.

While it was not surprising that the key distances involve the nucleophilic oxygen, it is surprising that the likelihood of the RC did not improve with the inclusion of OPs describing hydrogen bonding between water molecules and the D401 carbonyl or S181 hydroxyl group. Our simulations show that hydrogen bonding between the nucleophilic water and the S181 side chain and D401 backbone position the nucleophilic water for attack in the reactant conformation, consistent with the postulated roles for these residues.[3] As the distance between the nucleophilic water oxygen and the anomeric carbon decreases and the system moves to the TS conformation, the nucleophilic water oxygen is no longer close enough to the D401 carbonyl oxygen to allow for hydrogen bonding, but it remains hydrogen-bonded to the S181 hydroxyl group. The reactant conformation of the bridge water molecule is aligned by hydrogen bonding with a D175 carboxyl hydrogen, the bridge water, and with OH3 on the +1 glycosyl unit. While this hydrogen bonding appears to be crucial for positioning the water molecules for attack, the putative RC indicates that the nucleophilic water oxygen bond-forming and bond-cleaving distances better describe the hydrolysis reaction.

Fig. 6 shows the PMF and the values of key OPs (distances between atoms which have bonds that form or break) along the putative RC calculated using the data from five EPS runs. As discussed above, the RC is a linear combination of the OP6 and OP12, which describe nucleophilic attack. Note that the RC is dimensionless. The zero value of the RC aligns to the TS. In this case, the reactant basin corresponds to an RC value of −5.8 and the product basin to a value of 4.7. The scale is a function of a range of OP values for the coordinates used in likelihood maximization and has no inherent significance. The larger absolute value of the RC at the reactant well relative to the product well corresponds to a late TS.

Fig. 6A shows a free energy barrier of 12.3 ± 0.4 kcal mol^{-1} is consistent with that found for the hydrolysis (deglycosylation) of *Tr*Cel7A,[20] which was found to be 11.6 kcal mol^{-1} using the same Hamiltonian (CHARMM36 with SCC-DFTB for the QM region). The ΔG_{rxn} for *Tr*Cel6A is −12 ± 1 kcal mol^{-1}.

Based on simulations using the FF99SB Amber32 and GLY-CAM06 force fields with PBE functional, Petersen *et al.*[22] estimated a free energy barrier of 36 kcal mol^{-1} and ΔG_{rxn} of approximately −12 kcal mol^{-1} for the *C. thermocellum* inverting GH8 *endo*-1,4-glucanase A, following the classic inverting mechanism in which the proton from the attacking, nucleophilic water directly transfers to the catalytic base, rather than through a bridge water as shown here for *Tr*Cel6A. While some difference in the energy barriers is expected due to the different Hamiltonians used and differences between the enzymes, their reported barrier is likely a significant overestimate; transition state theory, assuming no recrossing and thus a transmission coefficient of 1.0, yields an upper-bound rate coefficient of 3.6 × 10^{-14} s^{-1} at 300 K for a $\Delta G^{\ddagger}$ of 36 kcal mol^{-1}, for a timescale much longer than the experimentally observed rate. For a $\Delta G^{\ddagger}$ of 12.3 kcal mol^{-1}, transition state theory predicts an upper-bound rate coefficient 17 orders of magnitude larger than for a $\Delta G^{\ddagger}$ of 36 kcal mol^{-1}, with a timescale on the order of 100 ms. Our calculated rate coefficient including recrossing is reported
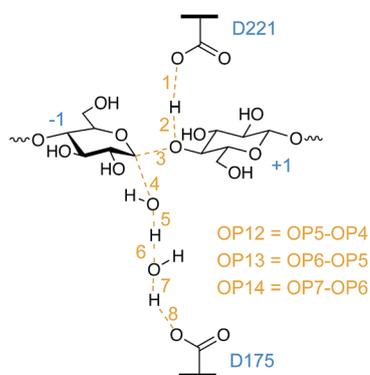

Fig. 5 Of the 87 OPs tested and listed in the ESI,† only OPs 4, 5, 6, 12, and 13 were identified by likelihood maximization as components of the best one-, two-, or three-parameter RCs.
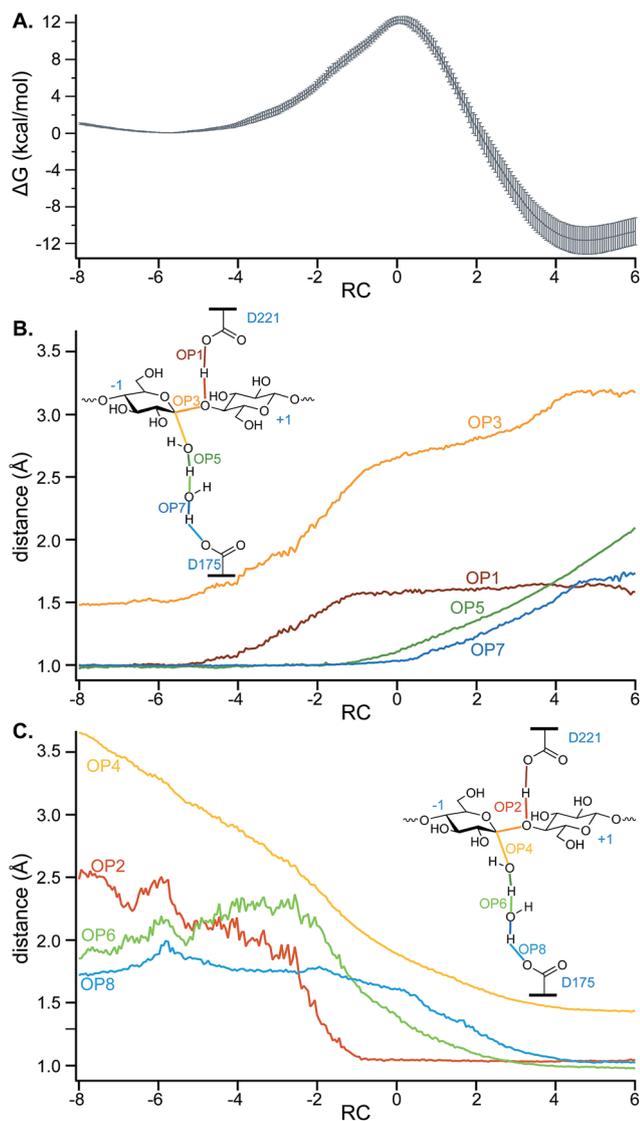
Fig. 6 (A) The average PMF (after setting the reactant energy to zero) and standard error generated from five independent EPS simulations using the two-parameter RC (unitless; a linear combination of OP6 and OP12, which is the difference between OP5 and OP4). A subset of the conformations generated in the EPS simulations were used to calculate, for the range of RC values aligning with the PMF, average values of OPs corresponding to distances between atoms of (B) breaking bonds and of (C) forming bonds.

below, along with the results of evaluating our TS ensemble by the histogram test.

To gain further insights into the structure of the TS, the values of OPs 1–8 (Fig. 5) were calculated for a subset of conformations generated during the EPS simulations (from 18 000 configurations) to allow calculation of the average values of these OPs in each of the bins of RC values created to calculate the PMF. For greater clarity, the OP values as a function of the RC are shown in two panels: Fig. 6B shows those that correspond to bond breaking and Fig. 6C for bond formation. The PMF indicates an overall single-barrier reaction, yet the eight bond cleavage and formation events do not occur simultaneously, as indicated both by the

projection of distances onto the RC (Fig. 6B and C) and by examining the simulations (a movie for one reactive trajectory is included in the ESI†). Instead, bond rearrangement occurs in a coordinated, step-wise fashion. The leading events are the elongation of the D221 carboxylic oxygen to acidic hydrogen bond, the shortening of the distance between the acidic hydrogen and glycosidic oxygen, the elongation of the glycosidic bond, and the approach of the nucleophilic water oxygen toward the −1 gluco-pyranoside anomeric carbon. The beginnings of the changes align with the beginnings of the rise in free energy along the RC from the reactant well located at RC = −5.8. Table 1 reports the progress of the key order parameters OP1–8 at the peak energy on the PMF, coincident with RC = 0. The D221 carboxylic oxygen to acidic hydrogen bond (OP1) is 91% along a change in distance from 1.00 to 1.63 Å, with the distance between the acidic hydrogen and accepting glycosidic oxygen (OP2) essentially already at its final distance and the glycosidic bond mostly broken (OP3), indicating a late TS. In contrast, the O–H bonds of the nucleophilic and bridge waters are only slightly elongated (OP5 at 15% and OP7 at 7%). This phenomenon highlights the crucial role of transferring the protons in the water wire for the reaction to proceed. It also underscores the importance of the alignment of the two active site water molecules for nucleophilic attack and proton transfer, as they must be in position on the reaction conformation for the reaction to proceed.

The low-energy product well is only reached when the transferring proton is accepted by a D175 carboxylic acid. The non-simultaneous multiple bond-cleavage and bond-formation events that depend on the participation of highly motile water molecules may all contribute to the difficulty of cleanly defining a one-dimensional RC that would yield a histogram sharply peaked at $p_B = 0.5$. As discussed below, we found that the simulations show that while water molecules occupy the active site as shown by the crystal structures, they are not static; the molecules can exchange positions with nearby water molecules. The unbiased simulations performed in this study confirm that D175 can indeed serve as the catalytic base. When it accepts the proton, the resulting product well is very stable, with a ΔG_{rxn} almost three times greater in magnitude than for TrCel7A, providing a driving force for the reaction.

The calculated transmission coefficient[86] is 0.43. This low value likely compensates for an imperfect RC, indicated by non-ideal results of the p_B histogram test.[46,84] The resulting rate coefficient of 2.9×10^3 s^{-1} at 300 K calculated by transition state theory is slightly lower than the second (deglycosylation) step of the TrCel7A hydrolysis mechanism, 5.3×10^3 s^{-1}.[20]

Fig. 7 shows the Cremer–Pople puckering coordinates[87] of the −1 glucopyranoside in the reactant, TS, and product conformations for the simulated hydrolysis reaction shown in Fig. 4, harvested as described in the ESI† and displaying results from every 500th 1 fs step. Fig. 7 also shows the puckering coordinates of the glycosyl moiety in the active site −1 position for crystal structures of GH6 cellulases co-crystallized with the non-hydrolyzable substrate mimic methyl 4-S-β-D-cellobiosyl-4-thio-β-D-cellobioside: TrCel6A wild-type (WT) (PDB ID 1QK2);[37] TrCel6A Y169F mutant (PDB ID 1QJW);[37] and HiCel6A D416A mutant (PDB ID 1GZ1).[88]

Table 1 Values of OP1–8 with their standard deviations in Å, at the reactant basin, putative TS, and product basin, and the percent change in distance from the reactant to TS compared to the change in distance from the reactant to product. For reference, the values of the RC and ΔG in kcal mol^{-1} (compared to the reactant basin and reported with their standard error) are included

	RC	ΔG	OP1	OP2	OP3	OP4	OP5	OP6	OP7	OP8
Reactant basin	−5.8	0.0	1.00 ± 0.03	2.48 ± 0.62	1.50 ± 0.05	3.24 ± 0.13	0.99 ± 0.03	2.16 ± 0.54	0.99 ± 0.03	1.99 ± 0.48
Putative TS	0.0	12.3 ± 0.4	1.58 ± 0.12	1.04 ± 0.04	2.66 ± 0.11	1.89 ± 0.04	1.11 ± 0.06	1.39 ± 0.12	1.04 ± 0.05	1.61 ± 0.12
Product basin	4.7	−11.7 ± 1.5	1.63 ± 0.13	1.04 ± 0.04	3.18 ± 0.19	1.45 ± 0.03	1.81 ± 0.03	0.99 ± 0.03	1.67 ± 0.15	1.03 ± 0.04
% change at TS			91%	99%	69%	75%	15%	66%	7%	40%
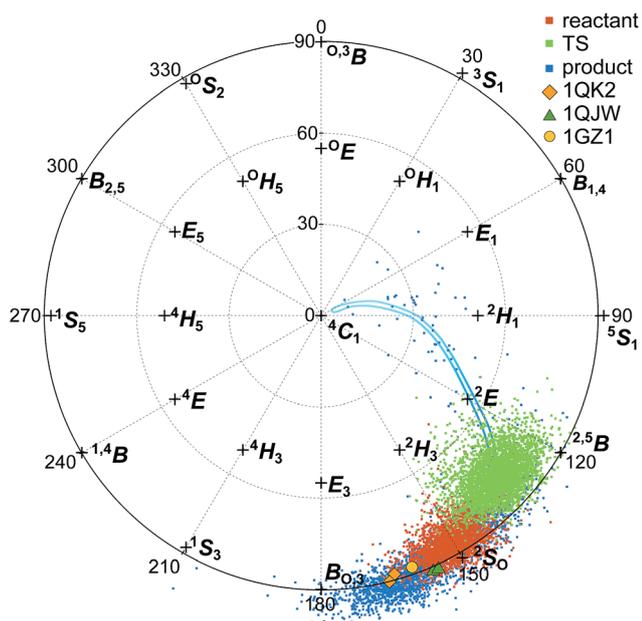
Fig. 7 Top-down view of the northern-hemisphere of the Cremer–Pople sphere[87] designating the −1 glucopyranoside puckering coordinates for the reactant (red), TS (green), and product (blue) for the elementary reaction shown in Fig. 4. Markers designate the puckering coordinates for the −1 glucopyranoside of the substrate mimic methyl 4-S-β-D-cellobiosyl-4-thio-β-D-cellobioside co-crystallized with TrCel6A wild-type (PDB ID 1QK2; orange diamond);[37] TrCel6A Y169F mutant (PDB ID 1QJW; dark green triangle);[37] and HiCel6A D416A mutant (PDB ID 1GZ1; yellow circle).[88] The two markers for PDB IDs 1QK2 and 1QJW represent the two chains present in the deposited structures. The blue line shows an approximate, smoothed ring-puckering path sampled during product ring fluctuations.

Interestingly, the puckering conformation changes along the reaction coordinate do not trace out straight path on the Cremer–Pople sphere, but a boomerang-like catalytic itinerary. The reactant structures reflected in Fig. 7 are in the Michaelis complex configuration with the −1 glucopyranoside adopting the ^2S_O pucker configuration. The TS ring conformations also remain close to the Cremer–Pople sphere equator, centered between the ^2S_O and the $^{2,5}B$ configurations. The product puckering configuration then shifts west, centered between the $B_{O,3}$ and the $^{2,5}B$ configurations. The cleaved cellobiose product remains in the −1 binding site for the five independent, 200 ps of unbiased, unconstrained simulations performed to collect the data shown in Fig. 7. However, the −1 glucopyranoside ring

shows more flexibility after cleavage, as evidenced by the wider spread of product ring conformations shown in Fig. 7. The ring readily samples the $B_{O,3}$, ^2S_O, and $^{2,5}B$ conformations, the later conformations mostly obscured in Fig. 7 by the reactant and TS conformations; this figure is reproduced as separate panels for each state in the ESI.† In one of the five simulations, the −1 glucopyranoside ring moved from the $B_{O,3}$ conformation to the ^2S_O to the $^{2,5}B$ conformation, and then continued to the ^2E conformation and on to $^4C_1$, which is the most stable orientation for a glucopyranoside ring in solution.[89] Following the trajectory for another 150 ps of QM/MM MD revealed that the ring returned to the $B_{O,3}/^2S_O/^{2,5}B$ conformations which are primarily sampled, rather than expelling from the enzyme product site in this short time period. This smoothed, approximate path is indicated by the blue line shown in Fig. 7. The simulations of the product conformation in the active site indicate the range of conformations accessible due to normal fluctuations and provide some information about the puckering itinerary between $B_{O,3}$ and $^4C_1$.

The ^2S_O and the $^{2,5}B$ conformations adopted by the reactant and product −1 glycosyl moieties are among the lowest free-energy ring conformations for the β-D-glucose monomer on the equator of the Cremer–Pople sphere.[89] The ring configurations for the −1 glycosyl moieties of the experimentally determined Michaelis structures also lie on the equator, between the canonical $B_{O,3}$ and ^2S_O coordinates. They overlap more closely with the product −1 puckering conformations than reactant conformations, although still close to the −1 puckering reactant conformation. Note that the substrates in the crystal structures contain a sulfur atom in place of the glycosidic oxygen atom connecting the −1 and +1 subunits. The larger atom may slightly perturb the position of the −1 subunit, which may in turn slightly perturb how the enzyme puckers that subunit. Still, the similarity in puckering structure demonstrates consistency between the simulations presented in this work and the experimentally reported crystal structures.

D175A mutant hydrolysis

As previously noted, when residue 175 is no longer able to accept a proton due to its mutation into an alanine residue, 2–3% residual activity remains, prompting our investigation of whether the hydrolysis event could be simulated with the TrCel6A D175A mutant. To infer whether a proton (or proton–water complex) would be able to escape the active site, we investigated the ease with which a water molecule could escape

the active site. Unrestrained MM-only simulations showed that the water molecules in the active site easily exchange positions with neighboring water molecules on a timescale of tenths of nanoseconds and diffuse into the bulk water. Protons are far more mobile than water molecules, and thus it is probable the excess proton liberated by the deprotonation of D221 could easily migrate beyond the active site, perhaps diffusing into the bulk water. Testing this hypothesis by expanding the QM region to include paths to bulk water would be prohibitively expensive. Instead, to determine whether the excess proton could hop beyond the second water in the active site, we modestly expanded the QM region to include a third water molecule initially 5.4 Å from the bridge water molecule that was also identified in the PDB ID 1QJW crystal structure. We excluded the mutated 175 alanine from the QM region. As noted in the detailed Computational procedures in the ESI,† we harvested sets of distances for OP1–6 from several well-equilibrated (more than 2000 trajectories completed), accepted WT AS points for use in targeted MD to create TS guesses for the D175A mutant cases. From these guesses, we were able to obtain multiple reactive trajectories for hydrolysis of the glycosidic bond.

During system energy minimization and equilibration, the third water molecule spontaneously moved toward the active site from its initial position determined from the crystal structure with an aspartic acid at residue 175 (PDB ID 1QJW). The mutation from aspartic acid to alanine created a larger cavity that was easily able to accommodate the third water molecule, as shown in the upper-left panel of Fig. 8. The PDB ID 1HGW structure of the *Tr*Cel6A D175A also resolved a water molecule closer to the bridge water (3.4 Å). From the initial TS guesses, we performed AS and continued to obtain reactive

trajectories as before. However, after proton transfer to the second ("bridge") water, the proton then hopped to the third water, and then to D221. This longer water wire performed catalytic rescue and reprotonated the catalytic acid. Had our simulation set-up included additional water molecules in the QM region, we might have observed proton transfer to the bulk water, as our simulations and experimental structural studies[34,37] indicate that bulk solvent is accessible from the active site even with the active site loop closed. Importantly, this simulation of the D175A mutant provides evidence for the potential of catalytic rescue by water. When residue 175 is part of the proton transfer mechanism, only two water molecules need to be aligned to allow proton hopping. Their alignment is aided by the S181 side chain and D401 backbone. In the case of the mutant, there must at least be a third water molecule properly aligned to form a water wire, which is a more rare event. Furthermore, previous studies have indicated that the presence of a negatively charged base, even separated by several water molecules, often accelerates proton dissociation from an acid.[90] Previous studies have also demonstrated that longer water wires transfer protons at slower rates than shorter water wires.[90–92] A recent QM/MM study of the inverting *O*-GlcNAc glycosyltransferase (OGT) calculated that the proton transfer to the catalytic base *via* a water molecule requires an activation energy 74% greater than direct transfer to α-phosphate, increasing the barrier from 23.5 kcal mol^{-1} to 41 kcal mol^{-1}.[93] Thus, our simulations provide the first results to support the hypothesis that a water wire can likely rescue catalytic activity at a lower rate than the WT due to the longer water wires involved, consistent with the lower residual activity observed for the D175A mutant.[34]
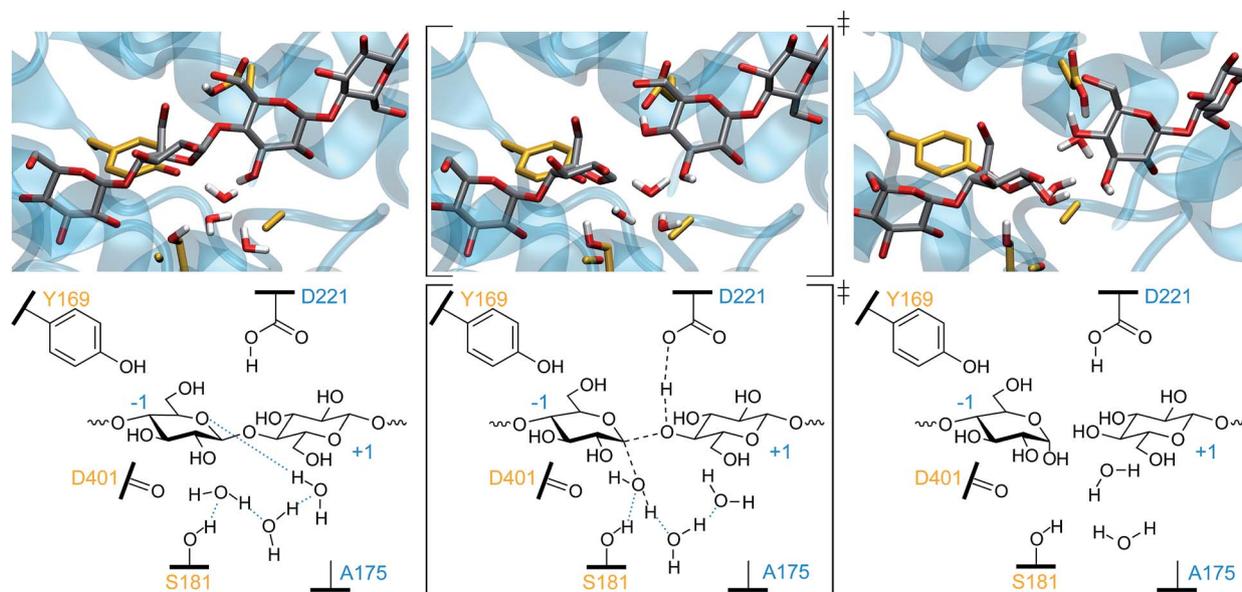

Fig. 8 The one-step inverting hydrolysis for the *Tr*Cel6A D175A mutant. The smaller residue 175 provides additional room for the third water molecule in the reactant (left) conformation. Only the nucleophilic water has an elongated O–H bond in the TS conformation (middle). In the product conformation (right), the catalytic acid (D221) has been reprotonated by proton transfer from nucleophilic water, to the bridge water, to the third water, and then to the carboxylic acid. Intermediate bond lengths are indicated in the TS by dashed black lines. Blue dotted lines indicate hydrogen bonding.

Processivity

Beyond the hydrolytic mechanism, we are also keenly interested in understanding cellulose chain processivity. As described in the Computational methods, US was used to simulate processivity from the "pre-slide" to "slide" conformations, *i.e.* the advancement of a cellohexaose molecule into subsites −1 and −2 in the tunnel of *Tr*Cel6A with the active site loops in open position.

Completely unbiased simulations run for over 50 ns provided well-equilibrated structures for the processivity simulations. During simulations of the pre-slide conformation, all substrate glycosyl rings were observed only in the solution-stable $^4C_1$ conformation. The same was true for the slide conformation simulations (both during the initial 50 ns of simulation and during an additional 250 ns of simulation) for all the substrate glycosyl rings except for the second glycosyl group (numbering from the leading monomer) which exclusively occupied the −1 binding site (Fig. 3). This subunit only adopted the $^2S_O/B_{O,3}$ conformation, which was also the reactant conformation observed in our hydrolysis studies and in crystal structures of cellulases trapped in the Michaelis complex (Fig. 7), indicating the importance of this subunit's ring distortion for reactivity.[3,89] The active site loop remained in the open conformation throughout these unbiased simulations and during the biased simulations that followed.

The CV used for US was the RMSD of the leading cellobiose ring atom coordinates compared to a reference equilibrated conformation in the pre-slide conformation. "Pulling" (or "pushing") the substrate from an equilibrated slide conformation toward (or further from) the pre-slide conformation, by biasing simulations along the CV, created initial configurations for each window.

An analysis of the puckering conformations revealed that all glycosyl units maintained the $^4C_1$ conformation throughout the processivity simulations, except for the second glycosyl unit. This monomer displayed a distinct conformational change near the CV value of 9.62 Å. In this region, the second glycosyl unit perturbs from the $^4C_1$ conformation to a $^2S_O/B_{O,3}$ conformation as it moves into the −1 position (Fig. 9). The ring did not freely move between the conformations in any of the windows, indicating that this conformational change was likely not well sampled. It is not surprising that a one-dimensional biasing coordinate was not sufficient to sample all important degrees of freedom for processivity. At least two more coordinates would be needed to correspond to the puckering coordinates shown in Fig. 9. Bias along three dimensions would be extremely costly, and the accuracy would be limited by level of accuracy of the Hamiltonian used. However, we can estimate the barrier for the puckering transition based on a previous study with a more accurate method: a QM study[89] with CCSD(T)//B3LYP, *versus* the pure MM CHARMM force field used for the processivity simulations. This QM study estimated a barrier of approximately 6 kcal mol^{-1} for free glucose, without any stabilization by an associated enzyme.

Therefore, a PMF generated based on biasing only the progress through the tunnel should be treated as an approximate pathway that does not include the full barrier corresponding to
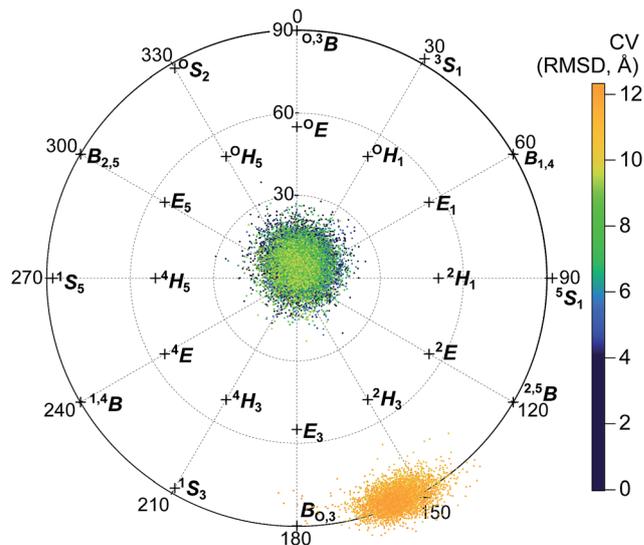

Fig. 9 Projection of the Cremer−Pople puckering coordinates of the second glycosyl unit onto the northern hemisphere of the Cremer−Pople sphere. Each point represents a conformation collected during umbrella sampling (separated by 100 ps each), colored according to the corresponding value of the CV (RMSD) tracking procession into the active-site tunnel, described in the text.

the puckering transition of the second glycosyl unit as it enters the −1 binding site. Still, analysis along this CV can be used to provide information about protein–sugar interactions in the tunnel. Thus, Fig. 10A shows the reconstructed PMF from WHAM constructed from separate analysis on (a) configurations in only the $^4C_1$ conformation (to the left of the red line) and (b) configurations only along the Cremer–Pople equator in the $^2S_O/B_{O,3}$ conformation (to the right of the red line). The low-energy well at a CV value of approximately 10.4 Å corresponds to the "slide" position and the lowest values of the *x*-axis correspond to the pre-slide conformation. The values of the CV higher than 10.4 Å are not expected to be visited during the catalytic cycle, except due to energy fluctuations, but were explored to verify that the substrate would not normally process beyond the slide conformation. We found that if we biased the substrate to advance beyond a CV value of approximately 12.3 Å, the second glycosyl unit again adopted the $^4C_1$ conformation as it was forced to leave the −1 position. The "pre-slide" snapshot in Fig. 3A was taken from near the end of the US simulations from a conformation with a CV value of 0.2 Å from the window centered at 0.0 Å. The "slide" snapshot in Fig. 3B was taken from near the end of the US simulation of the window centered at 10.5 Å (the CV value of that frame is 10.4 Å). The ESI† includes overlays of these two structures with their reference crystal structures.

The approximate PMF in Fig. 10A suggests that procession from the slide to pre-slide conformations occurs spontaneously. We used conformations generated during umbrella sampling to calculate the protein–sugar interactions shown in Fig. 10B and C. These unbiased interaction energies indicate that the low-energy well at the "slide" conformation is primarily due to strong electrostatic interactions between the protein and the leading cellobiose unit as it enters the product (−1 and −2) sites
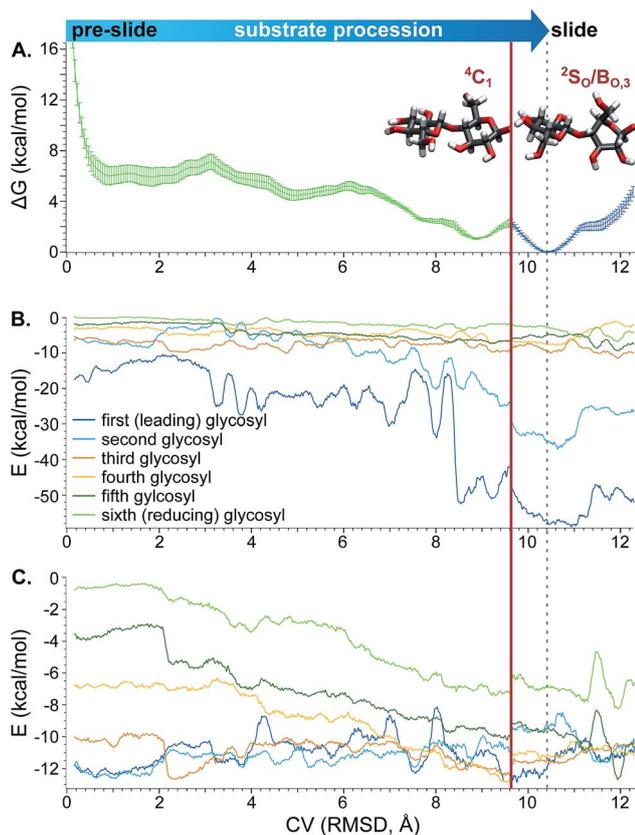
Fig. 10   (A) PMF for procession from the "pre-slide" to "slide" conformation constructed from US and WHAM. (B) Electrostatic and (C) van der Waals interactions between each of the substrate glycosyl units and the enzyme along the processive CV, smoothed as described in the Computational methods. For all three plots, the x-axis represents the CV for procession through the active-site containing tunnel, as described in the text. The solid red line designates a puckering dividing line; to the left of the line, the second glycosyl unit is exclusively in the $^4C_1$ conformation and to the right it is only in the $^2S_O/B_{O,3}$ conformation. The dotted gray line indicates the point along the CV corresponding to the slide conformation, at which point the substrate is in position in the active site for hydrolysis.

(Fig. 10B). As shown in the ESI,† the strongest of these electrostatic interactions involve residues that can hydrogen bond with hydroxyl groups of the substrate in the product sites (−1 and −2), occupied in the slide position: D137, at the end of the tunnel interacting with O3 at subsite −2; the side chain of D401 with O3 at −1; E399, also at the end of the tunnel, with O6 at −1; and K395 which hydrogen bonds with O3 at −1 and O6 at −2. These findings of hydrogen bonding between substrate and protein in the product site are consistent with previous studies suggesting that such interactions drive procession in CBH tunnels.[21,53]

Lower magnitude interactions with hydrophobic residues W269, W272, and W367 smoothly transfer between different glycosyl units along the processive path (see ESI†), increasing van der Waals interactions for the fourth, fifth, and sixth substrate glycosyl units as the substrate shifts into position for the next catalytic event (Fig. 10C). Such carbohydrate–aromatic residues have been shown to be important in processivity.[94]

Previous experimental studies indicated that Y169 is vital to puckering the glycosyl unit in the −1 position.[37,95] We found van der Waals interaction and electrostatic interactions of only a few kcal mol⁻¹ between Y169 and the first two glycosyl units, starting before the leading cellobiose unit occupies the product site (see ESI†). Previous computational studies have shown that W135, W269, and W367 are crucial for maintaining the ring conformation.[94] All three residues show even stronger interactions with the substrate in the slide position compared to Y169.

Conclusions

Here, we present the first atomic-level study of the dynamics of cellulose hydrolysis by the industrially important TrCel6A cellulase. The unbiased simulations provide evidence supporting the hypothesis that D175 can serve as the catalytic base via a bridging water molecule and highlight the importance of these water molecules in the reaction. Following the reaction path from the ensemble of collected TSs revealed that the glycosidic bond cleavage and transfer of the acidic hydrogen are almost complete before reaching the TS. Likelihood maximization identified that, of the 87 OPs tested, the key collective variables that best represent the RC are interatomic distances between the nucleophilic water oxygen and the anomeric carbon of the −1 glucopyranoside and between the nucleophilic water oxygen and the hydrogen that transfers to an adjacent water molecule, indicating that the nucleophilic attack drives the reaction. Determination of the PMF and transmission coefficient allowed calculation of the reaction barrier of 12.3 ± 0.4 kcal mol⁻¹ and rate coefficient of 2.9 × 10³ s⁻¹ at 300 K for this step. Hydrolysis thus has a larger barrier than our estimated processivity barrier (approximately 6 kcal mol⁻¹). Previous computational[96] and experimental[37] studies have shown the inherent flexibility of the active site loop, with both the open and closed positions accessible at ambient conditions, indicating that loop opening and closing also have smaller transition barriers than hydrolysis. Thus, it is likely that hydrolysis is the rate-limiting step in the TrCel6A processive catalytic cycle. The calculated rate coefficient for this step is larger than the experimentally determined k_{cat} of 14 ± 2 s⁻¹ for TrCel6A hydrolysis of cellohexaose,[34] which may be due in part to inaccuracy of the semi-empirical method used to calculate the barrier.

Our simulations provide evidence consistent with the hypothesis that in the absence of a catalytic base, as in the D175A mutant, additional water molecules can perform catalytic rescue, shuttling the excess proton back to the catalytic acid (D221), as simulated, or potentially to the bulk water through a longer water wire. Ability for migration of water between the active site and bulk has been observed in our simulations. The role of TrCel6A D175 analogs in homologous enzymes has been investigated by activity studies of WT and mutant TfCel6A (formerly E2)[97] and Cellulomonas fimi Cel6A (CfCel6A, formerly CenA).[98] Alanine mutants of the analogous residues to TrCel6A D175, TfCel6A D79A and CfCel6A D216A, also show decreased activity, with the extent of decrease dependent on the substrate. The substrate-dependence on the rate of decrease suggests that the stereochemistry of residue is

important. For example, we found that D175 in the *Tr*Cel6A WT hydrogen bonds with the second, "bridge" water, aiding in alignment of that molecule in a water wire. Additionally, the smaller side chain in *Tr*Cel6A D175A provided room for a third water molecule that could form a water wire with the two water molecules. The role of a second water molecule in WT hydrolysis raises the question of why the enzymes did not evolve a glutamic acid to serve as the catalytic base. It remains to be investigated whether the additional length would allow hydrolysis without the aid of a second water molecule. Even if that is possible, there is likely a reason why an aspartic acid is found at this position across homologous enzymes. The larger glutamic acid might interfere with substrate binding or movement into proper alignment in the active site. While no studies to our knowledge have reported a *Tr*Cel6A D175E mutant, activities of both *Tf*Cel6A D79E and *Cf*Cel6A D216E revealed that activity of such enzymes is decreased compared to WT. The residual activity of *Tf*Cel6A D79E is comparable to *Tf*Cel6A D79A for multiple substrates tested (phosphoric acid-swollen cellulose (SC), carboxymethyl-cellulose (SC), and filter paper).[97] For *Cf*Cel6A, the D216E mutant activity is comparable to D216A for phosphoric acid-swollen cellulose, yet more than an order of magnitude lower on the non-native substrates 2′,4′-dinitrophenyl β-D-cellobioside and carboxymethyl-cellulose.[98] Furthermore, the isosteric mutant *Tf*Cel6A D79N shows comparable activity to the D79A and D79E mutants, all reduced 2–3 orders of magnitude compared to WT, similar to the level of reduction for *Tr*Cel6A D175A compared to WT. Although D79N would not be able to accept a proton, its side chain can still participate in hydrogen bonding with water molecules to align them into a water wire. A mutation to a leucine would be interesting, as it is also isosteric but cannot participate in hydrogen bonding. Additionally, Vuong and Wilson studied the activity of the non-processive *endo*-glucanase *Tf*Cel6B WT and D226A mutant (homologous to *Tr*Cel6A D175A), which showed approximately one order of magnitude decrease in activity on bacterial microcrystalline cellulose, phosphoric acid-swollen cellulose, and phosphoric acid-treated cotton, yet a 10% increase in activity on carboxymethyl-cellulose.[39] These experimental studies and the computational work reported here indicate that both reactivity and stereochemistry of the residue at this position is important for hydrolysis; hydrolysis can occur even without the residue accepting a proton, as long as a water wire is able to fill this role.

The key role of water molecules in the WT reaction and in the D175A mutant may explain why experiment alone was unable to definitively map the *Tr*Cel6A mechanism. Catalytic rescue by a water wire has been indicated in other enzymes as well, such as human carbonic anhydrase II.[91] Like *Tr*Cel6A, carbonic anhydrase II shows residual activity when the primary proton acceptor is removed. Riccardi *et al.*'s investigation of mechanisms for catalytic rescue found that longer water wires required to reach an alternate stable state increased the barrier height for the reaction from 6.8 kcal mol^{-1} to 12.6 kcal mol^{-1} to 17.4 kcal mol^{-1} as the length of the water bridge was increased from two to three to four molecules, respectively.[91] The finding that longer water wires allow proton transfer at slower rates

than short water wires is supported by experimental and theoretical studies.[90,92,93]

We further investigated substrate processivity in *Tr*Cel6A. The simulations indicate that substrate movement into the product sites to be in position for the next catalytic event is a spontaneous reaction driven by multiple interactions with residues in the active site, primarily hydrogen bonding between charged side chains positioned to interact with the leading cellobiose unit when it occupies the −1 and −2 binding. The favorable interactions are consistent with previous studies of processive substrate motion and strong product inhibition.[3,21,53] These simulations further reveal the all substrate glycosyl rings maintain the $^4C_1$ conformation, except for the second-from-leading monomer. As this monomer moves into the −1 position, it distorts from the solution-stable $^4C_1$ conformation to a $^2S_O/B_{O,3}$ pucker.

During hydrolysis, the −1 glycosyl ring changes more subtly, remaining on the equator of the Cremer–Pople sphere as it shifts from the ^2S_O conformation in the reactant low-energy well toward the $^{2,5}B$ conformation at the TS, and then relaxes toward the $B_{O,3}$ conformation in the product low-energy well, with greater flexibility as the cleaved glycosidic bond no longer anchors the sugars as strongly in the −1 and −2 subsites. The functional importance of puckering for catalytic susceptibility activity has been previously studied,[89] and puckering has been observed in the lowest-energy transition-state conformation even in non-enzymatic, thermochemical cleavage of cellulose's glycosidic bond.[99] The energy barrier for non-enzymatic cleavage was calculated to be 54 kcal mol^{-1}, more than four times greater than the barrier calculated here, indicating the significant effect achieved by multiple interactions between the protein and substrate.

This study delivers on the promise of computational studies to provide a basic understanding of the underlying enzymatic mechanisms.[100–103] The studies presented here suggest two key features for inverting GH catalysis: (1) nucleophilic water molecule alignment to drive the reaction forward, and (2) water wires to shuttle the proton to the catalytic base. The important role of water in *Tr*Cel6A shown here may extend to inverting glycosyltranferases[59] and other enzymes for which a catalytic base has been difficult to experimentally determine.

Acknowledgements

References

1 B. L. Cantarel, P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat, *Nucleic Acids Res.*, 2009, **37**, D233–D238.

2 V. Lombard, H. Golaconda Ramulu, E. Drula, P. M. Coutinho and B. Henrissat, *Nucleic Acids Res.*, 2014, **42**, 490–495.

3 C. M. Payne, B. C. Knott, H. B. Mayes, H. Hansson, M. E. Himmel, M. Sandgren, J. Ståhlberg and G. T. Beckham, *Chem. Rev.*, 2015, **115**, 1308–1448.

4 N. Ramasubbu, L. M. Thomas, C. Ragunath and J. B. Kaplan, *J. Mol. Biol.*, 2005, **349**, 475–486.

5 A. Kato, E. Hayashi, S. Miyauchi, I. Adachi, T. Imahori, Y. Natori, Y. Yoshimura, R. J. Nash, H. Shimaoka, I. Nakagome, J. Koseki, S. Hirono and H. Takahata, *J. Med. Chem.*, 2012, **55**, 10347–10362.

6 A. El Kaoutari, F. Armougom, J. I. Gordon, D. Raoult and B. Henrissat, *Nat. Rev. Microbiol.*, 2013, **11**, 497–504.

7 T. A. Mousseau, G. Milinevsky, J. Kenney-Hunt and A. P. Møller, *Oecologia*, 2014, **175**, 429–437.

8 S. P. Chundawat, G. T. Beckham, M. E. Himmel and B. E. Dale, *Annu. Rev. Chem. Biomol. Eng.*, 2011, **2**, 121–145.

9 M. E. Himmel, S.-Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady and T. D. Foust, *Science*, 2007, **315**, 804–807.

10 D. E. Koshland Jr, *Biol. Rev.*, 1953, **28**, 413–436.

11 G. Speciale, A. J. Thompson, G. J. Davies and S. J. Williams, *Curr. Opin. Struct. Biol.*, 2014, **28**, 1–13.

12 Y. He, M. S. Macauley, K. A. Stubbs, D. J. Vocadlo and G. J. Davies, *J. Am. Chem. Soc.*, 2010, **132**, 1807–1809.

13 T. V. Vuong and D. B. Wilson, *Biotechnol. Bioeng.*, 2010, **107**, 195–205.

14 V. L. Y. Yip and S. G. Withers, *Biochemistry*, 2012, **51**, 8464–8474.

15 G. T. Beckham, Y. J. Bomble, E. A. Bayer, M. E. Himmel and M. F. Crowley, *Curr. Opin. Biotechnol.*, 2011, **22**, 231–238.

16 A. T. Bell and M. Head-Gordon, *Annu. Rev. Chem. Biomol. Eng.*, 2011, **2**, 453–477.

17 J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu and K. Schulten, *Curr. Opin. Struct. Biol.*, 2015, **31**, 64–74.

18 W. A. Breyer and B. W. Matthews, *Protein Sci.*, 2001, **10**, 1699–1711.

19 R. Stern and M. J. Jedrzejas, *Chem. Rev.*, 2008, **108**, 5061–5085.

20 B. C. Knott, M. Haddad Momeni, M. F. Crowley, L. F. Mackenzie, A. W. Götz, M. Sandgren, S. G. Withers, J. Ståhlberg and G. T. Beckham, *J. Am. Chem. Soc.*, 2014, **136**, 321–329.

21 B. C. Knott, M. F. Crowley, M. E. Himmel, J. Ståhlberg and G. T. Beckham, *J. Am. Chem. Soc.*, 2014, **136**, 8810–8819.

22 L. Petersen, A. Ardèvol, C. Rovira and P. J. Reilly, *J. Phys. Chem. B*, 2009, **113**, 7331–7339.

23 I. J. Barker, L. Petersen and P. J. Reilly, *J. Phys. Chem. B*, 2010, **114**, 15389–15393.

24 I. Tvaroska, I. André and J. P. Carver, *J. Am. Chem. Soc.*, 2000, **122**, 8762–8776.

25 S. Kozmon and I. Tvaroška, *J. Am. Chem. Soc.*, 2006, **128**, 16921–16927.

26 M. Krupička and I. Tvaroška, *J. Phys. Chem. B*, 2009, **113**, 11314–11319.

27 B. Henrissat, T. T. Teeri and R. A. J. Warren, *FEBS Lett.*, 1998, **425**, 352–354.

28 A. Vasella, G. J. Davies and M. Böhm, *Curr. Opin. Chem. Biol.*, 2002, **6**, 619–629.

29 G. T. Beckham, J. Ståhlberg, B. C. Knott, M. E. Himmel, M. F. Crowley, M. Sandgren, M. Sørlie and C. M. Payne, *Curr. Opin. Biotechnol.*, 2014, **27**, 96–106.

30 K. Igarashi, T. Uchihashi, A. Koivula, M. Wada, S. Kimura, T. Okamoto, M. Penttilä, T. Ando and M. Samejima, *Science*, 2011, **333**, 1279–1282.

31 K. Igarashi, *Nat. Chem. Biol.*, 2013, 1–2.

32 D. Martinez, R. M. Berka, B. Henrissat, M. Saloheimo, M. Arvas, S. E. Baker, J. Chapman, O. Chertkov, P. M. Coutinho, D. Cullen, E. G. J. Danchin, I. V. Grigoriev, P. Harris, M. Jackson, C. P. Kubicek, C. S. Han, I. Ho, L. F. Larrondo, A. L. de Leon, J. K. Magnuson, S. Merino, M. Misra, B. Nelson, N. Putnam, B. Robbertse, A. A. Salamov, M. Schmoll, A. Terry, N. Thayer, A. Westerholm-Parvinen, C. L. Schoch, J. Yao, R. Barabote, R. Barbote, M. A. Nelson, C. Detter, D. Bruce, C. R. Kuske, G. Xie, P. Richardson, D. S. Rokhsar, S. M. Lucas, E. M. Rubin, N. Dunn-Coleman, M. Ward and T. S. Brettin, *Nat. Biotechnol.*, 2008, **26**, 553–560.

33 J. Rouvinen, T. Bergfors, T. Teeri, J. K. C. Knowles and T. A. Jones, *Science*, 1990, **249**, 380–386.

34 A. Koivula, L. Ruohonen, G. Wohlfahrt, T. Reinikainen, T. T. Teeri, K. Piens, M. Claeyssens, M. Weber, A. Vasella, D. Becker, M. L. Sinnott, J.-y. Zou, G. J. Kleywegt, M. Szardenings, J. Ståhlberg and T. A. Jones, *J. Am. Chem. Soc.*, 2002, **124**, 10015–10024.

35 M. Schülein, *Biochim. Biophys. Acta*, 2000, **1543**, 239–252.

36 C. J. T. de Grotthuss, *Ann. Chim.*, 1806, **58**, 54–74.

37 J.-y. Zou, G. J. Kleywegt, J. Ståhlberg, H. Driguez, W. Nerinckx, M. Claeyssens, A. Koivula, T. T. Teeri and T. A. Jones, *Structure*, 1999, **7**, 1035–1045.

38 A. Varrot, T. P. Frandsen, I. von Ossowski, V. Boyer, S. Cottaz, H. Driguez, M. Schülein and G. J. Davies, *Structure*, 2003, **11**, 855–864.

39 T. V. Vuong and D. B. Wilson, *FEBS J.*, 2009, **276**, 3837–3845.

40 M. Sandgren, M. Wu, S. Karkehabadi, C. Mitchinson, B. R. Kelemen, E. A. Larenas, J. Ståhlberg and H. Hansson, *J. Mol. Biol.*, 2013, **425**, 622–635.

41 P. Biely, Z. Krátký and M. Vršanská, *Eur. J. Biochem.*, 1981, **119**, 559–564.

42 G. Davies and B. Henrissat, *Structure*, 1995, **3**, 853–859.

43 B. Peters and B. L. Trout, *J. Chem. Phys.*, 2006, **125**, 054108.

44 B. Peters, G. T. Beckham and B. L. Trout, *J. Chem. Phys.*, 2007, **127**, 034109.

45 P. G. Bolhuis, D. Chandler, C. Dellago and P. L. Geissler, *Annu. Rev. Phys. Chem.*, 2002, **53**, 291–318.

46 B. Peters, *J. Chem. Theory Comput.*, 2010, **6**, 1447–1454.

47 B. Peters, N. E. R. Zimmermann, G. T. Beckham, J. W. Tester and B. L. Trout, *J. Am. Chem. Soc.*, 2008, **130**, 17342–17350.

48 A. Koivula, T. Kinnari, V. Harjunpää, L. Ruohonen, A. Teleman, T. Drakenberg, J. Rouvinen, T. A. Jones and T. T. Teeri, *FEBS Lett.*, 1998, **429**, 341–346.

49 C. M. Payne, J. Baban, S. J. Horn, P. H. Backe, A. S. Arvai, B. Dalhus, M. Bjoras, V. G. H. Eijsink, M. Sørlie, G. T. Beckham and G. Vaaje-Kolstad, *J. Biol. Chem.*, 2012, **287**, 36322–36330.

50 C. M. Payne, W. Jiang, M. R. Shirts, M. E. Himmel, M. F. Crowley and G. T. Beckham, *J. Am. Chem. Soc.*, 2013, **135**, 18831–18839.

51 C. B. Taylor, C. M. Payne, M. E. Himmel, M. F. Crowley, C. McCabe and G. T. Beckham, *J. Phys. Chem. B*, 2013, **117**, 4924–4933.

52 L. Bu, G. T. Beckham, M. R. Shirts, M. R. Nimlos, W. S. Adney, M. E. Himmel and M. F. Crowley, *J. Biol. Chem.*, 2011, **286**, 18161–18169.

53 L. Bu, M. R. Nimlos, M. R. Shirts, J. Ståhlberg, M. E. Himmel, M. F. Crowley and G. T. Beckham, *J. Biol. Chem.*, 2012, **287**, 24807–24813.

54 J. L. A. Brás, A. Cartmell, A. L. M. Carvalho, G. Verzé, E. A. Bayer, Y. Vazana, M. A. S. Correia, J. A. M. Prates, S. Ratnaparkhe, A. B. Boraston, M. J. Romão, C. M. G. A. Fontes and H. J. Gilbert, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 5237–5242.

55 A. Fedorov, W. Shi, G. Kicska, E. Fedorov, P. C. Tyler, R. H. Furneaux, J. C. Hanson, G. J. Gainsford, J. Z. Larese, V. L. Schramm and S. C. Almo, *Biochemistry*, 2001, **40**, 853–860.

56 M. B. Lazarus, J. Jiang, T. M. Gloster, W. F. Zandberg, G. E. Whitworth, D. J. Vocadlo and S. Walker, *Nat. Chem. Biol.*, 2012, **8**, 966–968.

57 M. Schimpl, X. Zheng, V. S. Borodkin, D. E. Blair, A. T. Ferenbach, A. W. Schüttelkopf, I. Navratilova, T. Aristotelous, O. Albarbarawi, D. A. Robinson, M. A. Macnaughtan and D. M. F. van Aalten, *Nat. Chem. Biol.*, 2012, **8**, 969–974.

58 J. C. Fromme and G. L. Verdine, *EMBO J.*, 2003, **22**, 3461–3471.

59 S. G. Withers and G. J. Davies, *Nat. Chem. Biol.*, 2012, **8**, 952–953.

60 B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.

61 R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. Mackerell Jr, *J. Chem. Theory Comput.*, 2012, **8**, 3257–3273.

62 A. D. MacKerell Jr, M. Feig and C. L. Brooks III, *J. Am. Chem. Soc.*, 2004, **126**, 698–699.

63 A. D. Mackerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.

64 O. Guvench, S. N. Greene, G. Kamath, J. W. Brady, R. M. Venable, R. W. Pastor and A. D. Mackerell Jr, *J. Comput. Chem.*, 2008, **29**, 2543–2564.

65 O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. Mackerell Jr, *J. Chem. Theory Comput.*, 2009, **5**, 2353–2370.

66 D. Beglov and B. Roux, *J. Chem. Phys.*, 1994, **100**, 9050–9063.

67 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

68 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–242.

69 M. Wu, L. Bu, T. V. Vuong, D. B. Wilson, M. F. Crowley, M. Sandgren, J. Ståhlberg, G. T. Beckham and H. Hansson, *J. Biol. Chem.*, 2013, **288**, 33107–33117.

70 D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, *AMBER 12*, 2012.

71 J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *J. Comput. Chem.*, 2005, **26**, 1781–1802.

72 M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai and G. Seifert, *Phys. Rev. B*, 1998, **58**, 7260–7268.

73 G. d. M. Seabra, R. C. Walker, M. Elstner, D. A. Case and A. E. Roitberg, *J. Phys. Chem. A*, 2007, **111**, 5655–5664.

74 G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.

75 S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

76 B. Roux, *Comput. Phys. Commun.*, 1995, **91**, 275–282.

77 A. Grossfield, *WHAM: the weighted histogram analysis method, version 2.0.6*, 2012.

78 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.

79 D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084–3095.

80 M. F. Crowley, M. J. Williamson and R. C. Walker, *Int. J. Quantum Chem.*, 2009, **109**, 3767–3772.

81 A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–12566.

82 G. Schwarz, *Ann. Stat.*, 1978, **6**, 461–464.

83 B. Peters, *J. Chem. Phys.*, 2006, **125**, 241101.

84 G. T. Beckham and B. Peters, *Computational Modeling in Lignocellulosic Biofuel Production*, American Chemical Society, 2010, ch. 13, pp. 299–332.

85 D. G. Truhlar and B. C. Garrett, *Annu. Rev. Phys. Chem.*, 1984, **35**, 159–189.

86 A. C. Pan and D. Chandler, *J. Phys. Chem. B*, 2004, **108**, 19681–19686.

87 D. Cremer and J. A. Pople, *J. Am. Chem. Soc.*, 1975, **97**, 1354–1358.

88 A. Varrot, T. P. Frandsen, H. Driguez and G. J. Davies, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **58**, 2201–2204.

89 H. B. Mayes, L. J. Broadbelt and G. T. Beckham, *J. Am. Chem. Soc.*, 2014, **136**, 1008–1022.

90 O. F. Mohammed, D. Pines, E. T. J. Nibbering and E. Pines, *Angew. Chem., Int. Ed.*, 2007, **46**, 1458–1461.

91 D. Riccardi, P. König, H. Guo and Q. Cui, *Biochemistry*, 2008, **47**, 2369–2378.

92 B. J. Siwick, M. J. Cox and H. J. Bakker, *J. Phys. Chem. B*, 2008, **112**, 378–389.

93 M. Kumari, S. Kozmon, P. Kulhánek, J. Štepán, I. Tvaroška and J. Koča, *J. Phys. Chem. B*, 2015, **119**, 4371–4381.

94 C. M. Payne, Y. J. Bomble, C. B. Taylor, C. McCabe, M. E. Himmel, M. F. Crowley and G. T. Beckham, *J. Biol. Chem.*, 2011, **286**, 41028–41035.

95 A. Koivula, T. Reinikainen, L. Ruohonen, A. Valkeajärvi, M. Claeyssens, O. Teleman, G. J. Kleywegt, M. Szardenings, J. Rouvinen, T. A. Jones and T. T. Teeri, *Protein Eng.*, 1996, **9**, 691–699.

96 L. Bu, M. F. Crowley, M. E. Himmel and G. T. Beckham, *J. Biol. Chem.*, 2013, **288**, 12175–12186.

97 G. André, P. Kanchanawong, R. Palma, H. Cho, X. Deng, D. Irwin, M. E. Himmel, D. B. Wilson and J. W. Brady, *Protein Eng.*, 2003, **16**, 125–134.

98 H. G. Damude, S. G. Withers, D. G. Kilburn, R. C. Miller Jr and R. A. J. Warren, *Biochemistry*, 1995, **34**, 2220–2224.

99 H. B. Mayes and L. J. Broadbelt, *J. Phys. Chem. A*, 2012, **116**, 7098–7106.

100 S. Lutz, *Curr. Opin. Biotechnol.*, 2010, **21**, 734–743.

101 U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, *Nature*, 2012, **485**, 185–194.

102 M. T. Reetz, *J. Am. Chem. Soc.*, 2013, **135**, 12480–12496.

103 J. Damborsky and J. Brezovsky, *Curr. Opin. Chem. Biol.*, 2014, **19**, 8–16.

104 J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott and N. Wilkens-Diehr, *Comput. Sci. Eng.*, 2014, **16**, 62–74.