Mechanism of glycoside hydrolysis: A comparative QM/MM molecular dynamics analysis for wild type and Y69F mutant retaining xylanases†

Mahmoud E. S. Soliman,[a,b] J. Javier Ruiz Pernía,[a] Ian R. Greig[a] and Ian H. Williams*[a]

Computational simulations have been performed using hybrid quantum-mechanical/molecular-mechanical potentials to investigate the catalytic mechanism of the retaining *endo*-β-1, 4-xylanase (BCX) from *B. circulans*. Two-dimensional potential-of-mean-force calculations based upon molecular dynamics with the AM1/OPLS method for wild-type BCX with a *p*-nitrophenyl xylobioside substrate in water clearly indicates a stepwise mechanism for glycosylation: the rate-determining step is nucleophilic substitution by Glu78 to form the covalently bonded enzyme-substrate intermediate without protonation of the leaving group by Glu172. The geometrical configuration of the transition state for the enzymic reaction is essentially the same as found for a gas-phase model involving only the substrate and a propionate/propionic acid pair to represent the catalytic glutamate/glutamic acid groups. In addition to stabilizing the $^{2,5}B$ boat conformation of the proximal xylose in the non-covalent reactant complex of the substrate with BCX, Tyr69 lowers the free-energy barrier for glycosylation by 42 kJ mol^{-1} relative to that calculated for the Y69F mutant, which lacks the oxygen atom O_Y. B3LYP/6-31+G* energy corrections reduce the absolute height of the barrier to reaction. In the oxacarbenium ion-like transition state O_Y approaches closer to the endocyclic oxygen O_{ring} of the sugar ring but donates its hydrogen bond not to O_{ring} but rather to the nucleophilic oxygen of Glu78. Comparison of the average atomic charge distributions for the wild-type and mutant indicates that charge separation along the bond between the anomeric carbon and O_{ring} is matched in the former by a complementary separation of charge along the O_Y–H_Y bond, corresponding to a pair of roughly antiparallel bond dipoles, which is not present in the latter.

Introduction

The *endo*-1,4-β-xylanase (BCX) from *Bacillus circulans* catalyses the hydrolysis of xylan and β-xylobiosides with net retention of anomeric configuration. As a retaining β-glycosidase of family GH11,[1] the proposed double displacement mechanism involves a covalent glycosyl-enzyme intermediate.[2] Formation and hydrolysis of this covalent intermediate occur *via* oxacarbenium ion-like transition states, with the assistance of two key active site glutamic acid residues. One (Glu78) acts as a nucleophile: it is deprotonated in the enzyme-substrate complex[3] and it attacks the anomeric carbon of the substrate and displacing the leaving group. The other (Glu172) is protonated in the enzyme-substrate complex and plays a dual role of an acid/base catalyst:[3] in the glycosylation step it assists formation of the glycosyl-enzyme intermediate by donating a proton to the leaving group, and in the subsequent deglycosylation step it serves as a base, deprotonating the attacking water molecule. Despite extensive experimental studies, including mutational and crystallographic analyses of active-site residues in mutant BCXs,[4–6] there remain details of the catalytic mechanism that are still not well understood. Scheme 1 shows alternative
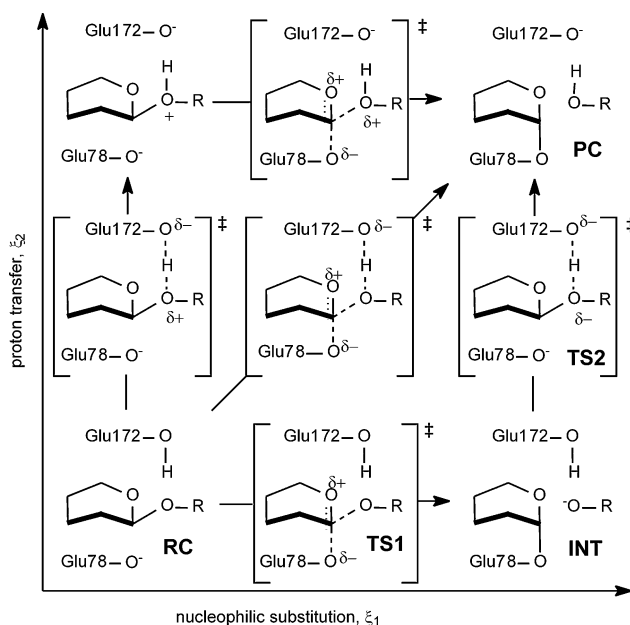
Scheme 1 Alternative mechanisms for formation of a covalent glycosyl-enzyme intermediate PC from the enzyme-substrate complex RC.

mechanisms for glycosylation: starting from the non-covalent enzyme-substrate complex RC, the covalent glycosyl-enzyme intermediate PC may be formed by a wholly concerted general acid-catalyzed mechanism (bottom left to top right) or by means

[a]Department of Chemistry, University of Bath, Bath, UK BA2 7AY. E-mail: i.h.williams@bath.ac.uk; Fax: +44 1225 386231; Tel: +44 1225 386625
[b]Department of Pharmaceutical Organic Chemistry, Faculty of Pharmacy, Zagazig University, Zagazig, Egypt
† Electronic supplementary information (ESI) available: Cartesian coordinates of QM atoms for optimised structures of key stationary structures. See DOI: 10.1039/b911644c

of either of two stepwise mechanisms which differ in the order of discrete nucleophilic substitution and proton transfer steps.

In the non-covalent complex, Tyr69 donates a strong hydrogen bond to the nucleophilic oxygen atom ($O^{\epsilon 2}$) of Glu78 (hereinafter denoted O_{nuc}) and accepts a hydrogen bond from the hydroxyl group at the 2-position of the distal xylose moiety.[4] In contrast, in the covalent intermediate, the hydrogen bond donated to the O_{nuc} is weaker, consistent with the ether character of its partner, and a stronger interaction is formed between Tyr69 and the ring oxygen (O_{ring}) of the proximal xylose moiety. The phenolic oxygen (O_Y) of Tyr69 is very important for catalysis, as evidenced by the observation that the Y69F mutant exhibits no detectable enzyme activity.[6] The nature of this $O_Y H_Y \cdots O_{ring}$ interaction is therefore an intriguing question that we seek to investigate here. Our previous work shed some light upon the role of Tyr69 in stabilizing the $^{2,5}B$ conformation of the distorted proximal sugar ring of a xylobioside substrate,[7] but now we employ computational modelling using quantum mechanical/molecular mechanical (QM/MM) potentials[8] to simulate free-energy changes occurring in the glycosylation step of the mechanism. Our aims are two-fold. First, to determine which of the pathways shown in Scheme 1 is preferred for the substrate ($R = C_6H_4NO_2$) employed in the experimental kinetics studies; we achieve this by considering a 2D free energy surface allowing both nucleophilic substitution and proton transfer to occur independently. Second, to obtain deeper understanding of the catalytic role of Tyr69; we achieve this by comparative analysis of the transition states involved in the reaction pathway for the wild-type BCX enzyme and the Y69F mutant. As yet very few computational modelling studies of glycosidase mechanisms include consideration of chemical bond making and breaking within the fully solvated protein environment,[9,10] and the only other example of a 2D free energy surface of which we are aware is for an inverting glycosyl hydrolase.[11]

Computational methods

Model system

As before,[7] we initiate the modelling from the PDB structure for the covalent enzyme-inhibitor complex of wild-type BCX (accession code 1BVV)[4] as this fulfills several criteria: it is a high-resolution (1.8 Å) crystal structure; the sugar substrate is covalently bound to the enzyme, thus allowing us to avoid errors involved with docking the substrate into the active site; the stereochemical configuration of the substrate is clearly defined; there are no mutations within the protein. A non-covalent reactant complex (RC, Fig. 1) with a xylobioside with a *p*-nitrophenyl (PNP) leaving group attached to anomeric carbon (C1, but hereinafter denoted C_{ano}) of the proximal sugar was constructed as previously described.[7] Note that the 2-F substituent of the covalent complex in the crystal structure is replaced by 2-OH in the RC model, and also that the unnatural aglycone PNP is used in the modelling in order to facilitate comparison with experimental kinetics results.

QM/MM simulations

All simulations were done using the Dynamo library[12] of programs for molecular dynamics (MD). Except for transition-structure
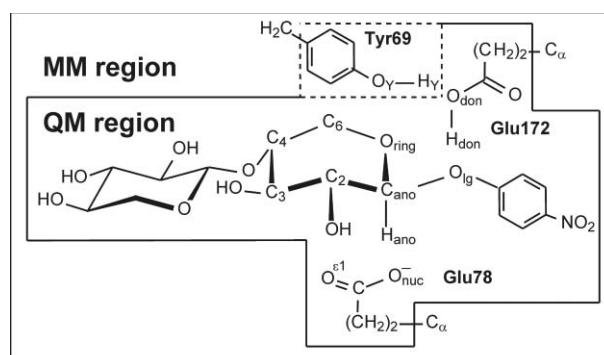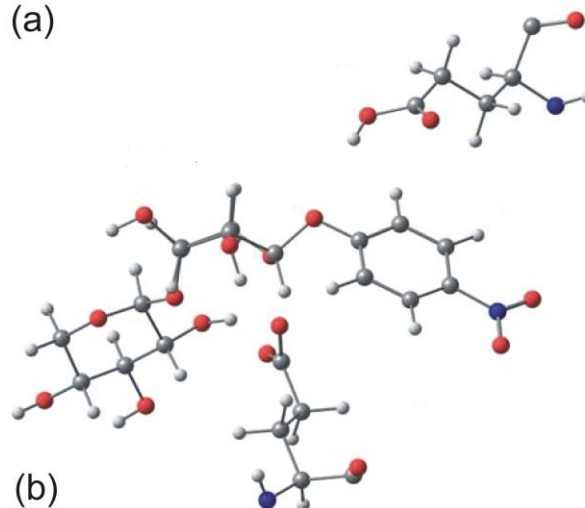
Fig. 1 BCX active site as modified for the non-covalent reactant complex showing (a) QM/MM partitioning and (b) actual structure.

(TS) localization, hessian calculation, and normal mode analysis (see below), the enzyme system was divided into two regions (Fig. 1a) in all simulations: the QM region contained active-site atoms described by the AM1 semi-empirical Hamiltonian,[13] and the MM region contained atoms in the rest of the enzyme and the surrounding solvent molecules described by the OPLS-AA potential.[14] For most simulations, the QM region contained 70 atoms consisting of the substrate, a xylobioside (XYL) with PNP aglycone, and the two catalytic residues Glu172 and Glu78. In some simulations we included the Tyr69 residue in the QM region; this larger model had QM 86 atoms. All covalent bonds between atoms of the QM and MM regions were treated by the link-atom approach in the Dynamo program: QM link atoms were placed along the $C_\beta(QM)–C_\alpha(MM)$ bonds of Glu78 and Glu172 and, where appropriate, Tyr69. The whole system was enveloped in a cubic box of TIP3P water[15] of side-length 55.5 Å. The total simulation system had 16476 atoms. QM/MM calculations were performed to obtain 2D free energy surfaces, to locate saddle-points, to determine reaction paths, and to evaluate Hessians.

2D-PMF calculations

To construct a two-dimensional free energy surface for the glycosylation step, potentials of mean force (PMFs) were computed using the method of umbrella sampling[16] as follows. Two

composite reaction coordinates, ξ_1 and ξ_2, which respectively describe nucleophilic substitution and proton transfer, were found (after many trials with other geometrical parameters) to give a good description of the mechanistic process; the oxygen of a glycosidic bond to the leaving group is denoted as O_{lg} whereas $O^{\epsilon 2}$ of the Glu172 proton donor is denoted as O_{don}. It was important to consider both variables in order to allow for possible coupling of proton transfer with nucleophilic attack. An umbrella potential U_i of harmonic form (eqn (3)) was defined for each coordinate ξ_i, where k_i is a force constant and ξ_i° is a constant reference value.

$$\xi_1 = d(O_{lg} \cdots C_{ano}) - d(O_{nuc} \cdots C_{ano}) \qquad (1)$$

$$\xi_2 = d(O_{don} \cdots H_{don}) - d(H_{don} \cdots O_{lg}) \qquad (2)$$

$$U_i(\xi_i) = 1/2 k_i(\xi_i - \xi_i^\circ)^2 \qquad (3)$$

All MD simulations employed the QM/MM potential described in the previous section and were performed in a series of windows using $k_1 = k_2 = 2500$ kJ mol^{-1} Å^{-2}, and with ξ_1° and ξ_2° being varied in increments of 0.1 Å from −1.8 to +3.0 Å and from −1.6 to +0.5 Å, respectively. The simulation in each window was ~10 ps length, with 2 ps equilibration: these values were found adequate to ensure sufficient overlapping of the simulations in neighbouring windows along each reaction coordinate in order to obtain satisfactory convergence of the PMF. A total of 846 individual MD trajectories requiring >10 ns of simulation time overall were performed. Longer simulations were performed for a few windows but did not lead to any significant changes; NMR studies[17] have shown BCX to have "high inherent conformational rigidity" both in its native state and as the covalent glycosyl-enzyme intermediate. All simulations were performed in the *NVT* ensemble at a temperature of 300 K; a velocity-Verlet–Langevin algorithm was used with a friction coefficient of 10 ps^{-1} for each atom and a time step of 1 fs.[18] The nonbonding interactions were calculated using an atom-based force-switching truncation function with inner and outer cutoffs of 10.5 and 11.5 Å, respectively. At each step of the data collection, the values of ξ_1 and ξ_2 were saved on an external file. The reaction-coordinate distribution functions were then determined for each window and pieced together using the weighted histogram analysis method.[19] The resulting distribution function $\rho(\xi_1, \xi_2)$ was then used to calculate the PMF, W (eqn (4)), where k_B is the Boltzmann constant, T is the temperature, and c is an arbitrary additive constant.[20] Differences in W provide estimates for changes in the Helmholtz free energy which, for a condensed phase system, are essentially equivalent to Gibbs energy changes.

$$W(\xi_1, \xi_2) = c - k_B T \ln \langle \rho(\xi_1, \xi_2) \rangle \qquad (4)$$

It is important to note that these 2D-PMF-derived free energy changes are slightly deficient because they are obtained by integrating over all but the two degrees of freedom ξ_1 and ξ_2, whereas true activation free energies are obtained by integrating over all degrees of freedom in the RC and all but one in the TS. Exact determination of free energy differences would necessarily involve consideration of the missing degrees of freedom, but the error from this approximation is unlikely to be large.

TS and reaction-path calculations

It is not generally feasible to perform calculations of two-dimensional free-energy surfaces directly with density-functional theory (DFT) techniques for a system of the size considered here—although a recent study of an inverting endoglucanase from family GH8 is a noteworthy exception.[11] Instead we estimated the probable errors in the QM part of the QM/MM calculations by means of the following procedure.

A 1D AM1/OPLS energy profile was generated in turn for each step of the glycosylation in the full system of enzyme, substrate and solvent. Up to 15 structures in the vicinity of the relevant transition state on the AM1/OPLS 2D-free-energy surface were selected, each of which was refined to a first-order saddle point (a TS) using standard second-derivative saddle-point location techniques. The QM region together with 17 active-site residues of the immediately surrounding MM region were optimized using second-derivative techniques, and the remainder of the system was optimized using a standard gradient minimizer, as implemented in Dynamo. Normal-mode analyses were done on all stationary points, and visual inspection of the normal mode displacements was carried out with the Molden program.[21] All starting points were found to converge to one and the same TS. From this refined TS a sequence of structures along the chosen reaction coordinate (in turn either ξ_1 or ξ_2) was generated by incrementing and decrementing the reference value ξ_i° by 0.05 Å and, at each point, reminimizing the whole system subject to a very strong harmonic constraint ($k_i = 20\,000$ kJ mol^{-1} Å^{-2}) applied to that reaction coordinate variables close to its reference value while allowing the rest of the system to minimize freely. This process allowed a continuous reaction path (similar to, but not identical with, an intrinsic reaction coordinate) to be generated between initial and final structures for each step of the reaction. Owing to the nature of this procedure note, however, that the final structure for the nucleophilic substitution was not exactly the same as the initial structure for the proton transfer.

To obtain higher-level corrected values for the reaction barriers, single-point energies were calculated for (only) the QM atoms of the reactant, TS, and product structures in vacuum with AM1 and with a DFT method; a total of 11 structures was adequate to span the complete range from initial to final for each step. The DFT calculations were performed by means of the Gaussian03 program[22] with B3LYP[23] and the 6-31+G* basis set. Fully optimized TS structures and reaction paths were computed in vacuum for a truncated model of 53 QM atoms comprising the proximal sugar ring and leaving group of the substrate, together with propionate and propionic acid moieties to represent the Glu78 and Glu172 residues of the active site. A similar procedure has been successfully employed by other workers in the field.[10]

Y69F mutant calculations

AM1/OPLS MD trajectories, optimizations of energy minima and first-order saddle points, and PMFs were computed as described above. Atomic charges were obtained by natural population analysis,[24] as implemented in Gaussian03; single-point calculations were performed using an enlarged QM region that included Tyr69 (or Phe in the Y69F mutant) and averages were

taken over ten structures randomly selected from the AM1/OPLS MD trajectories.

Results and discussion

Reaction mechanism for wild-type BCX

Fig. 2 shows the 2D surface of free-energy, with respect to the composite reaction coordinates ξ_1 for nucleophilic substitution at C_{ano} and ξ_2 for proton transfer to O_{lg}, for glycosylation of wild-type BCX by a PNP xylobioside substrate; ranges of $-1.5 \leq \xi_1 \leq 1.4$ and $-1.6 \leq \xi_2 \leq 0.5$ are displayed here. It was important to consider both variables in order to allow for possible coupling of proton transfer with nucleophilic attack in a concerted, general-acid catalyzed mechanism. However, the preferred pathway, shown by the dashed line on the figure, proceeds in a stepwise fashion from RC with nucleophilic substitution occurring first, with a barrier of 176 kJ mol^{-1}, to form the glycosyl-enzyme intermediate INT *via* TS1. The second step, from INT to PC *via* TS2, involves the proton transfer from Glu172 to the O_{lg} of the PNP anion with a barrier of 79 kJ mol^{-1}. However, the AM1/OPLS method predicts that this protonation of the PNP anion is unfavourable: the first-formed species INT is the glycosyl-enzyme with protonated Glu172 and anionic PNP, and PC would not be formed to any significant extent during the overall course of glycosylation and deglycosylation.
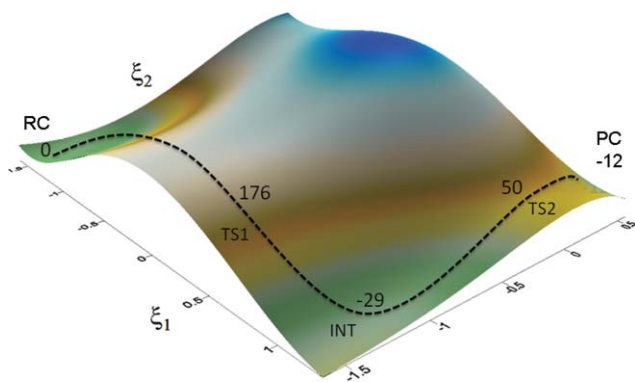

Fig. 2 AM1/OPLS calculated 2D-free-energy profile for glycosylation as a function of the coordinates for nucleophilic substitution ξ_1 and proton transfer ξ_2 (relative energies in kJ mol^{-1}).

Withers and co-workers[25] reported linear Brønsted correlations of log k_{cat}/K_m with aglycone pK_a for wild-type BCX and for Glu78Cys carboxymethylated at Cys78, Glu172Asp and Glu172Cys carboxymethylated at Cys172 mutants with slopes of −0.7 or −0.8 for hydrolysis of aryl xylobiosides with good leaving groups. Their interpretation was that there was very substantial glycosidic bond cleavage at the glycosylation transition state with very little proton donation; the Glu172Cys and Glu172Gln mutants lacking a proton-donor group were still capable of hydrolyzing substrates with good leaving groups (pK_a < 5.5) although they had no measurable activity with phenyl xylobioside or the natural substrate xylan. In the light of these experimental kinetic studies, our computational result for PNP xylobioside seems entirely reasonable. For the INT species however, the calculated AM1/OPLS free-energy barrier is much too high: the experimental value for k_{cat} = 24 s^{-1} for reaction of wild-type BCX with this substrate at 25 °C corresponds[25] to $\Delta G^{\ddagger} \approx 65$ kJ mol^{-1}.

TS and reaction-path calculations

Fig. 3a shows the AM1/OPLS energy profile (solid line) along the reaction coordinate ξ_1 for nucleophilic substitution in the full model including the MM environment of the protein and solvent. The enthalpic barrier height $\Delta H^{\ddagger}$ = 150 kJ mol^{-1} from this 1D calculation is somewhat lower than the free-energy barrier height $\Delta G^{\ddagger}$ = 176 kJ mol^{-1} from the 2D calculation; the difference suggests a small decrease in entropy in the transition state. The dashed line shows the corresponding gas-phase AM1 energy profile for the QM atoms only with geometries frozen as in the full enzymic model; the curve shown is a fit to a series of 11 single-point energy calculations for structures along the reaction coordinate on either side of TS1. The barrier height is a little lower than for the full model, but the overall enthalpy change from RC to INT underestimates the exothermicity; this highlights the importance in QM/MM treatments not only of E_{QM} (the energy of the QM region) but also of E_{MM} and $E_{QM/MM}$ (the energy of the MM region and the interaction energy between the two regions). The dotted line in Fig. 3a shows the gas-phase B3LYP/6-31+G* energy profile for the QM atoms only, fit to single-point energies for the same 11 frozen geometries as above. Although the DFT potential energy barrier is appreciably lower than the AM1 barrier, the overall reaction energy change is predicted to be considerably endothermic; this demonstrates the inadequacy of using a "high-level" DFT method to estimate a correction to the "low-level" semi-empirical QM/MM result if only E_{QM} is considered and if geometrical relaxation is ignored. Note that a comparison of enzyme and vacuum environments has no meaning from the point of view of catalysis, since the gas phase is not a reasonable reference state for an uncatalyzed reaction.
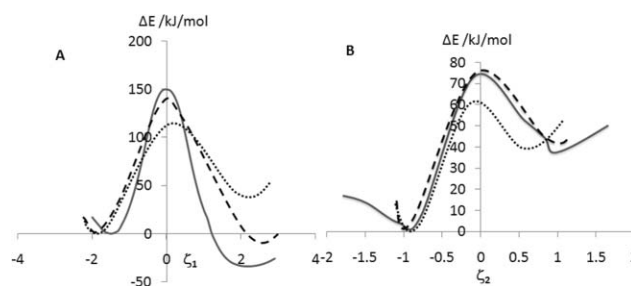

Fig. 3 Relative energies for nucleophilic substitution (a) with respect to RC, and proton transfer (b) with respect to INT calculated using AM1/OPLS in enzyme (solid line), AM1 single points (dashed line), and B3LYP/6-31+G* single points (dotted line).

Fig. 3b shows energy profiles along the reaction coordinate ξ_2 for proton transfer from Glu172 to O_{lg} in the full AM1/OPLS model (solid line) and for the gas-phase single-point energies calculated for the QM atoms only with AM1 (dashed line) and B3LYP/6-31+G* (dotted line). In this case all three methods give similar barrier heights and overall reaction energies. The profiles for nucleophilic substitution (Fig. 3a) cannot be combined simply with those for proton transfer (Fig. 3b) since the product geometry INT for the former is not quite the same as the reactant geometry INT for the latter, owing to the way that the profiles were computed with applied constraints.

It is of interest to consider the AM1 or B3LYP/6-31+G* geometries and relative energies of RC, TS1 and INT optimized in
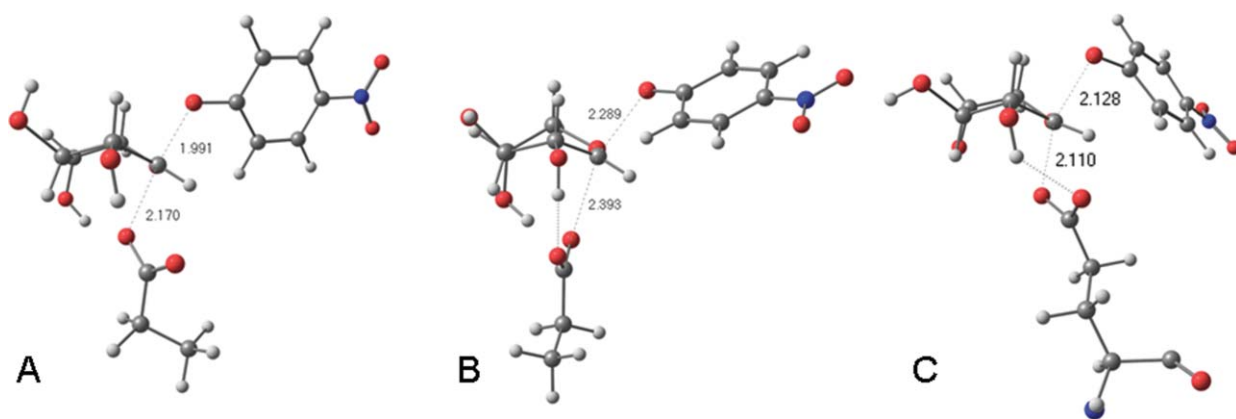
Fig. 4 Transition structures for the nucleophilic substitution optimized using three different methods: (a) gas-phase AM1, (b) gas-phase B3LYP/6-31+G* and (c) enzymic AM1/OPLS.

vacuum without any constraints for the nucleophilic substitution and proton transfer steps. To this end we performed unconstrained TS optimizations with both methods for a model system of 53 QM atoms comprising the proximal sugar ring and leaving group of the substrate, together with propionate and propionic acid moieties to represent the Glu78 and Glu172 residues of the active site. Fig. 4 and Table 1 show the resulting geometries of TS1 for the nucleophilic substitution step for both methods alongside that for the full QM/MM model. It is remarkable that they are all very similar. The gas-phase AM1 TS geometry (Fig. 4a) has shorter distances for the breaking $O_{lg} \cdots C_{ano}$ and making $O_{nuc} \cdots C_{ano}$ bonds than does the AM1/OPLS TS geometry (Fig. 4c), whereas the B3LYP/6-31+G* TS geometry (Fig. 4b) has longer distances. However, the nucleophilic carboxylate and acidic carboxyl moieties are located in essentially the same positions in all three TSs: this suggests that in the enzyme the catalytic residues are positioned more or less exactly where they prefer to be in vacuum. Furthermore, there is a hydrogen bonding interaction between $O^{\epsilon 1}$ of the nucleophilic carboxylate and the hydroxyl group at C2 of the proximal xylose ring: kinetic analysis indicates that this hydrogen bond contributes at least 18 kJ mol⁻¹ to stabilization of the transition state for glycosylation in the retaining β-glucosidase from *Agrobacterium faecalis*.[26]

Inspection of the interatomic distances presented in Table 1 suggests that the key geometrical features of the substrate and its interactions with catalytic groups within the enzyme active site are well described by the AM1/OPLS method, as compared with freely optimized structures obtained in vacuum using both the AM1 and B3LYP/6-31+G* methods.

Another point of interest is the distance $O_{nuc} \cdots O_{don}$ between the nucleophilic carboxylate and acidic carboxyl groups, corresponding to Glu78 and Glu172 in BCX. As Table 1 shows, in the AM1/OPLS optimized structures for RC, TS1 and INT of the nucleophilic substitution step this separation is found to vary between about 6 and 7 Å. The AM1 and DFT structures optimized in vacuum yield separations between 6.4 and 6.8 Å and 6.1 and 6.6 Å, respectively. It should be noted that, while all of these distances are greater than the often-quoted value of 5.5 Å for retaining β-glycosidases,[27] the structures considered here all contain the PNP aglycone group whereas this was not present in any of the X-ray crystallographic structures used to evaluate the experimental average separation between the catalytic groups.

These calculations indicate that proton transfer from Glu172 to the good leaving group PNP is thermodynamically unfavourable. Nevertheless it is of interest to note that the dihedral angle $O_{nuc} \cdots C_{ano} \cdots O_{lg} \cdots O_{don}$ in the optimised enzymic TS1 (Fig. 4c) has a value of 109° consistent with the designation of Glu172 as a *syn* proton donor in BCX.[28] This feature emerges directly from the calculations and is not the result of any constraint, let alone the artificiality of cartoon representations such as Scheme 1 and Fig. 1a. (It is not meaningful to consider this dihedral in TS2 for the proton transfer step because the PNP is displaced away from the covalent intermediate.)

Computational studies of protected glucosyl oxacarbenium ions by Whitfield and co-workers[29] have suggested a preference for the $^5S_1$ conformer. However, conformational analysis of TS1 structures from the AM1/OPLS MD trajectory shows no evidence for this conformer but only for $^{2,5}B$ and $^2S_0$, as assigned by these

Table 1 Selected interatomic distances (Å) and transition frequencies for the QM region in different simulation methods. RC, TS1, INT, TS2, and PC refer to the reactant complex, transition structure for the nucleophilic attack, xylosyl-intermediate, transition structure for proton transfer, and product complex, respectively

	AM1/OPLS (enzyme)					AM1 (gas-phase)					B3LYP/6-31+G* (gas-phase)				
	RC	TS1	INT	TS2	PC	RC	TS1	INT	TS2	PC	RC	TS1	INT	TS2	PC
$C_{ano} \cdots O_{lg}$	1.44	2.13	3.71	4.61	4.30	1.44	1.99	2.96	3.01	4.90	1.47	2.28	4.11	4.22	4.31
$O_{nuc} \cdots C_{ano}$	2.97	2.11	1.44	1.44	1.44	2.98	2.17	1.44	1.44	1.44	3.16	2.39	1.46	1.44	1.44
$H_{don} \cdots O_{lg}$	3.21	2.95	3.95	1.15	0.98	2.08	1.98	1.91	1.13	0.99	1.85	1.58	1.57	1.20	0.98
$O_{nuc} \cdots O_{don}$	6.96	7.63	7.93	8.31	8.13	6.76	6.5	6.44	6.71	7.20	6.64	6.11	6.34	6.45	6.57
$\nu^{\ddagger}$/cm⁻¹		460i		953i			413i		744i			119i		452i	

5240 | *Org. Biomol. Chem.*, 2009, **7**, 5236–5244

This journal is © The Royal Society of Chemistry 2009

authors' canonical scheme.[30] As noted above, both RC and TS1 maintain a hydrogen bond between the hydroxyl group at C2 of the proximal xylose ring and $O^{\epsilon 1}$ of the nucleophilic carboxylate, thereby enforcing an approximately *anti* conformation (162–163°) upon the dihedral angle H–C2–O2–H, which is only released in PC (66°) where this hydrogen bond is not present. Again, this behaviour of the oxacarbenium-like TS1 in the active-site of BCX is apparently different from that of pyranosyl oxacarbenium ions in continuum solvation calculations.[31]

Fig. 5 shows the energy profiles in vacuum corresponding to Fig. 4a and 4b. In these cases the intermediate is common to both parts of the profile, since the no constraints were applied on these calculations. The overall energy change for formation of the covalent intermediate in this simple model is essentially the same for both methods, and very similar to the QM/MM energy change for the full enzyme. However, the DFT potential energy barrier $\Delta E^{\ddagger} = 52$ kJ mol^{-1} is markedly lower than the AM1 barrier.
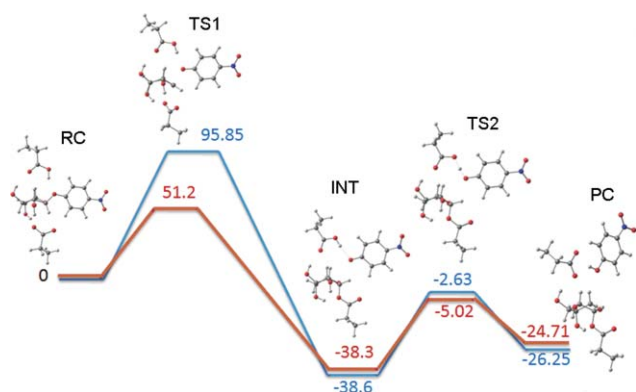

Fig. 5 Energy profile for a simplified model of glycosylation using B3LYP/6-31+G* (red) and AM1 (blue) methods in vacuum.

Overall it appears that the AM1/OPLS treatment of the solvated enzyme-substrate complex provides reasonably good geometries and predicts the important mechanistic features correctly, whereas calculated energies may contain errors. However, it is also evident that attempts to correct the deficiencies of the AM1 semiempirical method by means of a higher-level QM method (*e.g.* DFT) may not be successful unless they involve geometrical relaxation and consider not only E_{QM} but also E_{MM} and $E_{QM/MM}$; this is not feasible for free-energy calculations on systems of the size considered here. We note that, while AM1 does not predict structures and relative energies of carbohydrate conformations very accurately in the gas phase or solution,[32] this is unlikely to be as problematic within an enzyme active site in which multiple hydrogen-bonding contacts between the protein and the substrate (among other effects) are the principal determinant of conformation. Although subtle differences might exist between the true conformations and those found in this study (*e.g.* $^{2,5}B$ and adjacent conformations) dramatic differences are not possible, especially within the rather rigid active site of BCX.

Role of Tyr69: hydrogen bond analysis

We showed previously[7] that the phenolic OH group of Tyr69 donates a hydrogen bond to either O_{nuc} of Glu78 or O_{ring} of the proximal xylose, fluctuating between the two during the course of

an AM1/OPLS MD trajectory for the non-covalent RC of wild-type BCX. PMF calculations revealed[7] that these interactions stabilize the $^{2,5}B$ boat conformation of the sugar ring by about 20 kJ mol^{-1} relative to the $^{4}C_{1}$ chair conformation, thereby favouring an approximately coplanar alignment of the atoms about the C_{ano}–O_{ring} bond in the proximal ring of the substrate, as required for the oxacarbenium-like TS.

Fig. 6 extends this same analysis to the TS (TS1) and product (INT) of the nucleophilic substitution step of glycosylation; the previously published[7] trajectory for RC is also shown for comparison; over a longer (93 ps) trajectory the mean $H_Y \cdots O_{nuc}$ distance is consistently longer (3.293 ± 0.481 Å) than the $H_Y \cdots O_{ring}$ distance (2.466 ± 0.488 Å). However, during the course of a 30 ps AM1/OPLS MD trajectory for the TS, it is apparent that the $H_Y \cdots O_{nuc}$ distance is consistently shorter than the $H_Y \cdots O_{ring}$ distance, indicating that the hydrogen bond between Tyr69 and Glu78 is favored, although both distances are shorter than the averages found in RC. However, in the covalently-bonded INT the $H_Y \cdots O_{nuc}$ distance is consistently longer than the $H_Y \cdots O_{ring}$ distance, indicating that Tyr69 now donates its hydrogen bond exclusively to the xylose ring rather than to Glu78, although the average distance to the latter is similar to that in RC.
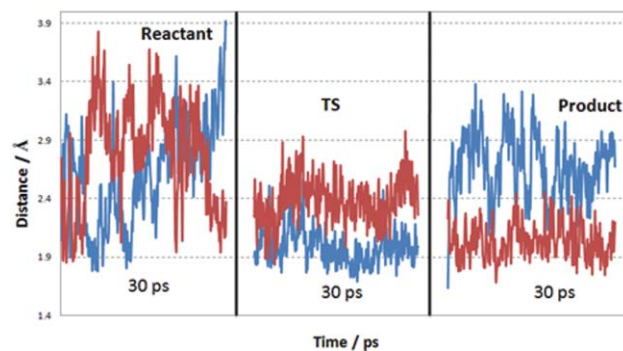

Fig. 6 Plot of the hydrogen bonding distances $H_Y \cdots O_{nuc}$ (blue) between Tyr69 and Glu78 and $H_Y \cdots O_{ring}$ (red) between Tyr69 and xylose during AM1/OPLS MD trajectories for reactant, TS and product of the nucleophilic substitution step for wild-type BCX.

Fig. 7 shows the $O_Y \cdots O_{ring}$ distance over 30 ps AM1/OPLS MD trajectories for the RC (as before),[7] TS (TS1) and product (INT) of nucleophilic substitution. It is evident that the average distance in both the TS and the covalently-bonded INT is similar and
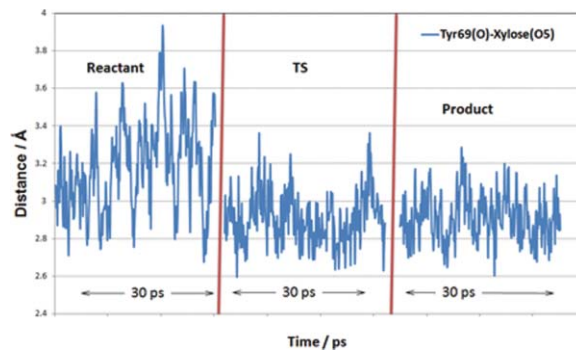

Fig. 7 A plot of the $H_Y \cdots O_{ring}$ distance between Tyr69 and the proximal xylose residue during AM1/OPLS MD trajectories for reactant, TS and product of the nucleophilic substitution step for wild-type BCX.

This journal is © The Royal Society of Chemistry 2009

Org. Biomol. Chem., 2009, **7**, 5236–5244 | 5241

significantly shorter than in the non-covalent reactant complex. The catalytic implications of this close approach between Tyr69 and the xylose ring in the TS are considered below.

Free-energy profiles for wild-type and Y69F mutant

To further understand the catalytic function of Tyr69, it is instructive to compare AM1/OPLS PMFs calculated for reactions of both wild-type BCX and the Y69F mutant. Since the free-energy profiles for the proton transfer step did not show any noticeable difference, only those for the nucleophilic substitution step are shown in Fig. 8. The shapes of the profiles are similar, but the free-energy barrier for reaction of the Y69F mutant is higher by about 42 kJ mol^{-1}. Although the absolute magnitudes of each barrier are undoubtedly overestimated by the AM1 description of the QM region, most of the systematic errors should cancel in the comparison. As commented above, it is not a straightforward task to estimate the magnitude of the error as compared to a large-basis DFT, or other high-level theoretical method. Our concern is not with the energy barrier itself, but rather with the influence of enzyme environment upon it. The source of the difference must lie in the interaction between Tyr69 and the substrate. We anticipate that the change from tyrosine to phenylalanine (both in the MM region) will polarize higher-level QM wavefunctions in a qualitatively similar (though undoubtedly quantitatively different) fashion to AM1. We therefore consider our AM1/OPLS estimate of the difference in the free-energy barrier for glycosylation to be at least qualitatively reliable. A quantitative assessment of error, as compared to an appropriate high-level QM method, would be meaningless unless it were evaluated for completely converged populations of conformations accessible under condensed-phase conditions, since there are large fluctuations between energy differences taken at arbitrary "snapshot" structures taken from the MD trajectories. It is not the intrinsic accuracy of the QM method for gas-phase calculations that it is important but rather the quality of the van der Waals parameters used in the QM/MM interface: these are well established for AM1/OPLS but are not for other QM methods.
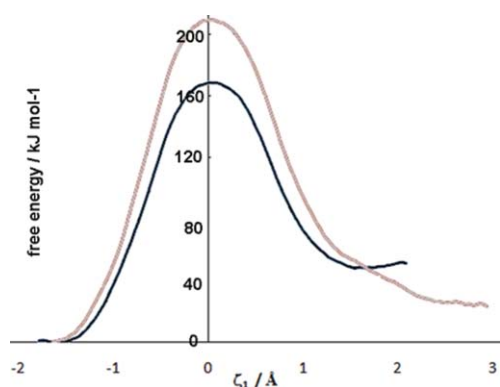

Fig. 8 Free-energy profiles for formation of the glycosyl-enzyme intermediate in wild-type BCX (black) and Y69F mutant (grey).

The 42 kJ mol^{-1} reduction in activation free energy corresponds to a factor of $\sim 2 \times 10^7$ in k_{cat} at 25 °C, in accord with the observation that the Y69F mutant has less than 0.01% of the activity of wild-type BCX.[6] Note that the two systems differ only by a single atom, O_Y.

Table 2 AM1/OPLS optimized interatomic distances (Å) for first-order saddle points for the transition states of wild-type BCX and the Y69F mutant

	Wild-type	Y69F
$C_{ano} \cdots O_{lg}$	2.127	2.150
$O_{nuc} \cdots C_{ano}$	2.110	2.129
$H_Y \cdots O_{nuc}$	1.909	
$O_Y \cdots O_{ring}$	2.829	
$\nu^{\ddagger}/cm^{-1}$	460i	422i

Transition state analysis for wild-type and Y69F mutant

Table 2 contains transition frequencies and selected interatomic distances and for optimised transition structures corresponding to first-order saddle points on the AM1/OPLS potential energy surfaces for the wild-type and mutant BCX. The mutation does not cause any significant alteration to the active-site geometry.

In order to obtain insight into the changes occurring in the electronic charge distribution as between the RC and the TS1 in the wild-type and mutant enzymes, detailed analysis was performed on sets of ten "snapshot" structures selected randomly from along the 30 ps QM/MM MD trajectories. Average atomic charges (Table 3) were computed with the B3LYP/6-31+G* method for a QM region which was extended to include the active-site tyrosine or phenylalanine. Again, it is important to note that the mutation causes no significant geometrical perturbation in the active site region (Fig. 9): the position of Glu78 remains unchanged, implying that Tyr69 plays no role in placing the Glu78 in a position appropriate for nucleophilic attack, in agreement with earlier X-ray crystallographic analysis.[4]
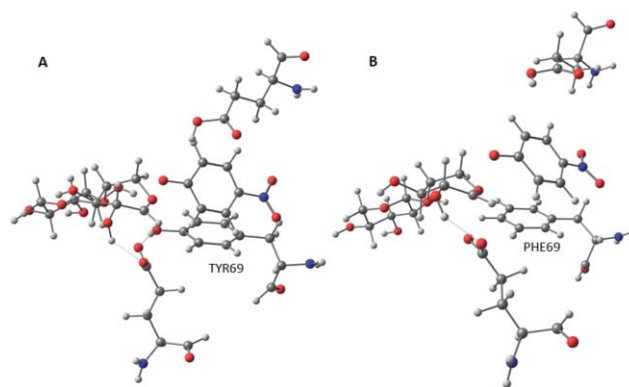

Fig. 9 Representative B3LYP/6-31+G* optimized transition structures selected for atomic charge calculations for (a) wild-type BCX and (b) the Y69F mutant, showing the full extent of the QM region.

Zechel and Withers suggested that in the TS there might be a 'stabilizing electrostatic or dipolar interaction' between O_Y and the partial positive charge on O_{ring}, which they depicted graphically as an arrow from an electronic lone pair on the former towards the latter.[33] Their description may now be re-interpreted in the light of the atomic charges now computed for the RC and the TS, which have been averaged over a selection of structures from MD trajectories for these species. Table 3 contains average charges obtained from natural population analysis of the calculated electron density distributions. Regardless of which method is

5242 | *Org. Biomol. Chem.*, 2009, **7**, 5236–5244

This journal is © The Royal Society of Chemistry 2009

Table 3 Average B3LYP/6-31+G* atomic charges ($|e|$) on selected atoms of first-order saddle points optimized from structures selected randomly along AM1/OPLS MD trajectories for the transition states of wild-type BCX and the Y69F mutant

| | AM1/OPLS | | | | B3LYP/6-31+G* | | | |
| | Wild-type | | Y69F | | Wild-type | | Y69F | |
	R	TS	R	TS	R	TS	R	TS
H_{ano}	0.196	0.246	0.210	0.258	0.138 ± 0.001	0.190 ± 0.005	0.152 ± 0.005	0.190 ± 0.005
C_{ano}	0.111	0.318	0.117	0.312	0.350 ± 0.003	0.353 ± 0.002	0.339 ± 0.003	0.347 ± 0.002
O_{ring}	−0.273	−0.128	−0.258	−0.120	-0.489 ± 0.004	-0.431 ± 0.001	-0.502 ± 0.000	-0.411 ± 0.001
O_{lg}	−0.272	−0.593	−0.263	−0.580	-0.551 ± 0.001	-0.672 ± 0.001	-0.470 ± 0.004	-0.658 ± 0.001
O_Y	−0.585	−0.619			-0.604 ± 0.002	-0.619 ± 0.004		
H_Y	0.435	0.366			0.350 ± 0.003	0.366 ± 0.003		

used to define atomic charges (alternatives include Mulliken and electrostatic potential charges), oxygen atoms invariably carry partial negative charges. Thus, as the glycosidic bond undergoes cleavage, and the glycon acquires oxacarbenium-ion character, so the charge on the O_{ring} becomes less negative rather than actually positive. Considering C_{ano} and H_{ano} as a unit, the charge separation along the O_{ring}–(CH)$_{ano}$ bond differs little between the RC (0.977 $|e|$) and the TS (0.974 $|e|$) for wild-type BCX, but is matched by a complementary separation of charge along the O_Y–H_Y bond as between the RC (0.954) and TS (0.985), corresponding to a pair of roughly antiparallel bond dipoles. In the Y69F there is no O_Y–H_Y bond dipole present to stabilize the O_{ring}–(CH)$_{ano}$ bond dipole.

Human purine nucleoside phosphorylase (hPNP) catalyses glycosyl transfer by means of a transition state with oxacarbenium-ion character;[34,35] its ribofuranoside substrate possesses a hydroxyl group C5. It has been suggested that the neighbouring His257 provides a mechanical push upon O5 towards the endocyclic O4 in a compressive motion with the phosphate nucleophile such that the build-up of electron density stabilizes the oxacarbenium-like transition state and facilitates the reaction.[34,35] Of course, the xylose ring in the −1 subsite of BCX has no hydroxymethyl substituent at C5; indeed, it has been noted that there is no space around C5 to accommodate any substituent.[4] Instead, however, the active site of wild-type BCX presents Tyr69 in close proximity to O5. We suggest that the OH groups of Tyr69 in BCX and of the hydroxylmethyl substituent of the ribofuranoside substrate of hPNP may have similar roles, which may be better described in terms of electrostatics than of dynamics.

Conclusions

2D-PMF calculations based upon MD with the hybrid QM/MM AM1/OPLS method for wild-type BCX with a PNP xylobiose substrate in water clearly indicate a stepwise mechanism for glycosylation: the rate-determining step is nucleophilic substitution by Glu78 to form the covalently bonded enzyme–substrate intermediate without protonation of the leaving group by Glu172. The geometrical configuration of the TS for the enzymatic reaction is essentially the same as found for a gas-phase model involving only the substrate and a propionate/propionic acid pair to represent the catalytic glutamate/glutamic acid groups. In addition to stabilizing the $^{2,5}B$ boat conformation of the proximal xylose in the non-covalent reactant complex of the substrate with BCX, Tyr69 lowers the free-energy barrier for glycosylation by 42 kJ mol^{-1} relative to that calculated for the Y69F mutant, which lacks the

oxygen atom O_Y. B3LYP/6-31+G* energy corrections reduce the absolute height of the barrier to reaction. In the oxacarbenium ion-like TS, O_Y approaches closer to the endocyclic oxygen O_{ring} of the sugar ring but donates its hydrogen bond not to O_{ring} but rather to the nucleophilic oxygen of Glu78. Comparison of the average atomic charge distributions for the wild-type and mutant indicates that charge separation along the bond between the anomeric carbon and O_{ring} is matched in the former by a complementary separation of charge along the O_Y–H_Y bond, corresponding to a pair of roughly antiparallel bond dipoles, which is not present in the latter.

References

1 B. Henrissat and G. Davies, *Curr. Opin. Struct. Biol.*, 1997, **7**, 637–644.
2 (a) J. D. McCarter and S. G. Withers, *Curr. Opin. Struct. Biol.*, 1994, **4**, 885–892; (b) D. J. Vocadlo and G. J. Davies, *Curr. Opin. Chem. Biol.*, 2008, **12**, 539–555.
3 L. P. McIntosh, G. Hand, P. E. Johnson, M. D. Joshi, M. Korner, L. A. Plesniak, L. Ziser, W. W. Wakarchuk and S. G. Withers, *Biochemistry*, 1996, **35**, 9958–9966.
4 G. Sidhu, S. G. Withers, N. T. Nguyen, L. P. McIntosh, L. Ziser and G. D. Brayer, *Biochemistry*, 1999, **38**, 5346–5354.
5 S. C. Miao, L. Ziser, R. Aebersold and S. G. Withers, *Biochemistry*, 1994, **33**, 7027–7032.
6 W. W. Wakarchuk, R. L. Campbell, W. L. Sung, J. Davoodi and M. Yaguchi, *Protein Sci.*, 1994, **3**, 467–475.
7 M. E. S. Soliman, G. D. Ruggiero, J. J. Ruiz Pernía, I. R. Greig and I. H. Williams, *Org. Biomol. Chem.*, 2009, **7**, 460.
8 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
9 S. Kozmon and I. Tvaroška, *J. Am. Chem. Soc.*, 2006, **128**, 16921–16927.
10 A. L. Bowman, I. M. Grant and A. J. Mulholland, *Chem. Commun.*, 2008, 4425–4427.
11 L. Petersen, A. Ardèvol, C. Rovira and P. J. Reilly, *J. Phys. Chem. B*, 2009, **113**, 7331–7339.
12 M. J. Field, M. Albe, C. Bret, F. Proust-De and Martin A. Thomas, *J. Comput. Chem.*, 2000, **21**, 1088–1100.
13 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.
14 G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, *J. Phys. Chem. B*, 2001, **105**, 6474–6487.
15 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

16 J. P. Valleau, G. M. Torrie, in *Statistical Mechanics*, B. J. Berne, Ed.; Plenum Press, New York, 1977, Part A, pp 169-194; G. M. Torrie and J. P. Valleau, *Chem. Phys. Lett.*, 1974, **28**, 578–581.

17 G. P. Connelly, S. G. Withers and L. P. McIntosh, *Protein Sci.*, 2000, **9**, 512–524.

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

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

20 M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford, UK, 1987.

21 G. Schaftenaar and J. H. Noordik, *J. Comput.-Aided Mol. Des.*, 2000, **14**, 123–134.

22 R. C. Gaussian03, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaroni, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, Gaussian, Inc., Wallingford CT, 2004.

23 A. D. Becke, *J. Chem. Phys.*, 1992, **96**, 2155–2160.

24 A. E. Reed, R. B. Weinstock and F. Weinhold, *J. Chem. Phys.*, 1985, **83**, 735–746.

25 (*a*) S. L. Lawson, W. W. Wakarchuk and S. G. Withers, *Biochemistry*, 1996, **35**, 10110–10118; (*b*) S. L. Lawson, W. W. Wakarchuk and S. G. Withers, *Biochemistry*, 1997, **36**, 2257–2265.

26 M. N. Namchuk and S. G. Withers, *Biochemistry*, 1995, **34**, 16194–16202.

27 Q. P. Wang, R. W. Graham, D. Trimbur, R. A. J. Warren and S. G. Withers, *J. Am. Chem. Soc.*, 1994, **116**, 11594–11595.

28 W. Nerinckx, T. Desmet, K. Piens and M. Claeyssens, *FEBS Lett.*, 2005, **579**, 302–312.

29 A. R. Ionescu, D. M. Whitfield, M. Z. Zgierski and T. Nukada, *Carbohydr. Res.*, 2006, **341**, 2912–2920.

30 A. Bérces, D. M. Whitfield and T. Nukada, *Tetrahedron*, 2001, **57**, 477–491.

31 A. R. Ionescu, D. M. Whitfield and M. Z. Zgierski, *Carbohydr. Res.*, 2007, **342**, 2793–2800.

32 J. P. McNamara, A.-M. Muslim, H. Abdel-Aal, H. Wang, M. Mohr, Ian H. Hillier and R. A. Bryce, *Chem. Phys. Lett.*, 2004, **394**, 429–436.

33 D. L. Zechel and S. G. Withers, *Acc. Chem. Res*, 2000, **33**, 11–18.

34 A. S. Murkin, M. R. Brinck, A. Rinaldo-Matthis, W. Shi, E. A. Taylor, S. C. Almo and V. L. Schramm, *Biochemistry*, 2007, **46**, 5038–5049.

35 S. Núñez, D. Antoniou, V. L. Schramm and S. D. Schwartz, *J. Am. Chem. Soc.*, 2004, **126**, 15720–15729.