Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2003) 151–159
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Can thioglycosides imitate the oxonium intermediate in glycosyl hydrolases?

Brian J. Smith*

Department of Structural Biology, The Walter and Eliza Hall Institute of Medical Research, 1G Royal Parade, Parkville, Victoria 3050, Australia

Received 30 January 2003; received in revised form 6 June 2003; accepted 6 June 2003

Abstract

Glycosyl hydrolases catalyse the acid hydrolysis of the glycosidic bond of glycans. The structure of barley β-D-glucan glucohydrolase in complex with a thiol substrate analogue presents very short contacts between the carboxyl oxygen atoms of the catalytic acid and the sulphur atom of the inhibitor. The geometries of acetic acid and dimethylsulfide in various ionisation states from ab initio molecular orbital calculations predict similar short contacts when an acetate anion forms a complex with a sulphonium cation. The energy of this complex is, however, significantly greater than the energy of the complex where both acetic acid and dimethylsulfide are neutral. Calculations on an active site model of barley β-D-glucan glucohydrolase indicate that the protein environment is able to significantly reduce this energy. The energy required for mechanical constraint of the short S···O separations, however, is identical to that required for the transfer of the proton from the acid to the sulphur. The identity of the species participating in the short contacts remains unanswered.
© 2003 Elsevier Inc. All rights reserved.

Keywords: Glycosidic hydrolysis; Molecular orbital calculations; Oxonium ions; Sulfonium ions; Thioglycosides

1. Introduction

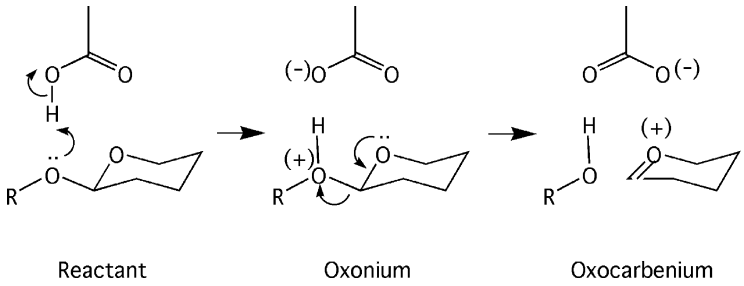
There are several distinctly identifiable species that might exist along the generic reaction pathway of glycosidic hydrolysis (Scheme 1). Abstraction of a proton by the glycosidic oxygen from the catalytic acid produces an oxonium ion intermediate. Cleavage of the glycosidic bond in this intermediate yields an oxocarbenium ion. Inverting enzymes use a single-displacement mechanism through an oxocarbenium ion transition state, while in retaining enzymes a covalent glycosyl-enzyme intermediate is formed; in either case the oxocarbenium ion itself does not exist as a distinct intermediate. While these features are generic to most glycosidases, there are many exceptions. For example, there are enzymes capable of hydrolyzing glucopyranosyl pyridimium salts involving C–N bond cleavage, in which no proton donation occurs [1–3].

Calculations [4] on the simplest glycoside, 2-oxanol, show that ring-distortion of the carbohydrate substrate prevents formation of a stable oxonium ion intermediate. For the thioglycosides, however, distortion of the ring is not predicted to prevent formation of a sulfonium ion [5]. The chair and twist-boat conformations of the sulfonium ions are predicted to have stable geometries. Dissociation of the glycoside bond in 2-methoxyoxane in the chair conformation is slightly exothermic. In contrast, in the sulfanyl derivative it is highly exothermic. Therefore, the greater stability of the sulfonium ions toward cleavage by acid hydrolysis [6], compared with the oxonium ions, accounts for their ability to avoid cleavage by glycosyl hydrolases. Myrosinase is able to hydrolyze thio-glucosides [7]. Although the catalytic machinery is similar to that of related *O*-glycosidases, the reaction is not acid catalysed; catalysis is due to the excellent leaving group properties of the substrate aglycon. While in the gas phase the proton affinity of alkyl sulfides is generally significantly greater than that of alkyl ethers, for example, the proton affinity of methanethiol [8] (CH_3SCH_3) is 830.9 kJ mol^{-1}, whereas for dimethylether [8] it is 792.0 kJ mol^{-1}; in solution alkyl thio-glucosides react more slowly than alkyl glucosides, presumably because the sulfur analogue has a reduced proton affinity [9]. One cannot predict a priori whether protonation of the sulfur of thioglycosides is more or less favoured than protonation of the oxygen of the natural substrate in the active site of the enzyme.

Since the sulfonium ions are predicted to be such stable species, it might be possible to detect their presence when bound to an enzyme. X-ray crystallography, however, does not immediately lend itself to determining the position of protons in enzymes. There have been, however, a small

* Tel.: +61-3-9345-2687; fax: +61-3-9345-2686.
E-mail address: bsmith@wehi.edu.au (B.J. Smith).

1093-3263/$ – see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S1093-3263(03)00156-6



Scheme 1.

number of reports of anomalous contacts between the sulphur atom of inhibitors and the carboxylate oxygen atoms of glycosyl hydrolase proteins. The structure of a nonhydrolyzable thioglycoside inhibitor bound to endoglucanase I from *Fusarium oxysporum* [10] exhibits an unusually short separation between one of the carboxylate oxygen atoms of the proposed proton donor (Glu202) and the sulphur atom of the inhibitor (2.71–2.86 Å, average separation 2.80 Å). This is considerably shorter than the sum of van der Waals radii, 3.25 Å (van der Waals' radii for oxygen and sulfur are 1.40 and 1.85 Å, respectively). The distance between the sulphur atom and the second carboxylate oxygen of the donor is much longer (3.41–3.52 Å, average separation 3.47 Å) and more in keeping with a noncovalent interaction. In another example, the structure of cellobiohydrolase CelC6 from *Trichoderma reesei* [11] reveals that the closest approach of the acid (Asp221) to the sulphur atom of a thioglycoside inhibitor is 3.6 Å. The active site residues of this enzyme and endoglucanase E2 from *Thermomonosporum fusca* [12] have several residues occupying the same relative positions in the two enzymes with the notable exception of the proposed catalytic acid residue (Asp221 in *T. reesei* and Asp117 in *T. fusca*). In the superposition of the two structures the separation of the carboxylate oxygen atoms of the acid of E2 and the sulphur atom of the inhibitor of CelC6 is just 2.6 Å. However, conformational changes are believed to occur upon ligand binding in *T. fusca* that may displace the acid and increase the separation between the acid and substrate [13].

The X-ray crystal structure of barley β-D-glucan glucohydrolase in complex with a thio-linked (1–3)-β-glucan has been recently reported [14]. In this structure the separation between the two carboxylate-oxygen atoms of the proposed proton donor (Glu491) and the sulphur atom is 2.72 and 2.97 Å, both significantly shorter than the sum of van der Waals radii. Quantum mechanical (QM) studies of small molecule model systems at the HF/3-21G* level indicated that these distances were suggestive of a sulfonium ion in complex with a carboxylate [14]. The X-ray crystal structure was solved to 2.20 Å resolution. Crudely, this corresponds to atom positions determined with precision within the range 0.2–0.4 Å. The difference between the sum of van der Waals' radii and the short contact distances is only slightly larger than the uncertainty expected from the X-ray experiment. Thus, the identification, particularly the ionization state, of the bound species remains in question. Reported here are further computational studies that consider what molecular species are likely to produce these short separations. The calculations of the small model systems reported earlier are extended to higher levels of theory, and the influence of the protein environment is investigated.

2. Methods and results

2.1. Complexes between acetic acid and dimethyl sulphide

Standard ab initio molecular orbital calculations were performed using the GAUSSIAN 98 [15] program. Geometries of the complexes between dimethyl sulfide and acetic acid in various ionisation states are presented in Fig. 1. Unless otherwise stated, all geometries were optimised at the MP2(fu)/6-31G(d) level.

In the complex between acetic acid and dimethylsulfide (**A**) the distances between the carboxylate oxygen atoms and the sulphur atom, 3.336 and 3.514 Å, are more than 0.5 Å larger than those observed in the experimental X-ray crystal structure of barley β-D-glucan glucohydrolase in complex with the thio-glycan, 2.72 and 2.97 Å.

The S···H distance in complex **A** is 2.354 Å. In Fig. 2 we show the effect on the geometry of this complex as the S···H distance is made shorter. Initially, the effect of the constraint is only to bring the two molecules closer together. The distances between the sulphur and two carboxylate-oxygen atoms reduces by roughly the same amount as the applied constraint, while the $H\cdots O^{(1)}$ distance remains unchanged. When the S···H separation is 1.6 and 0.75 Å shorter than in the complex **A**, the $S\cdots O^{(1)}$ distance has reduced to 2.70 Å, roughly 0.6 Å shorter than in **A**, while the $S\cdots O^{(2)}$ distance is 3.25 Å, less than 0.3 Å shorter than in **A**. The $H\cdots O^{(1)}$ distance increases by only 0.12 Å. At this point, the constraint has increased the energy of the complex by 66 kJ mol^{-1}. Further reduction of the S···H distance sees a continued reduction in the $S\cdots O^{(2)}$ distance, while the $S\cdots O^{(1)}$ distance undergoes only a small change. The $H\cdots O^{(1)}$ distance exhibits a larger change, increasing by 0.3 Å as the S···H separation is reduced by 0.3 Å from 1.6 to 1.3 Å.

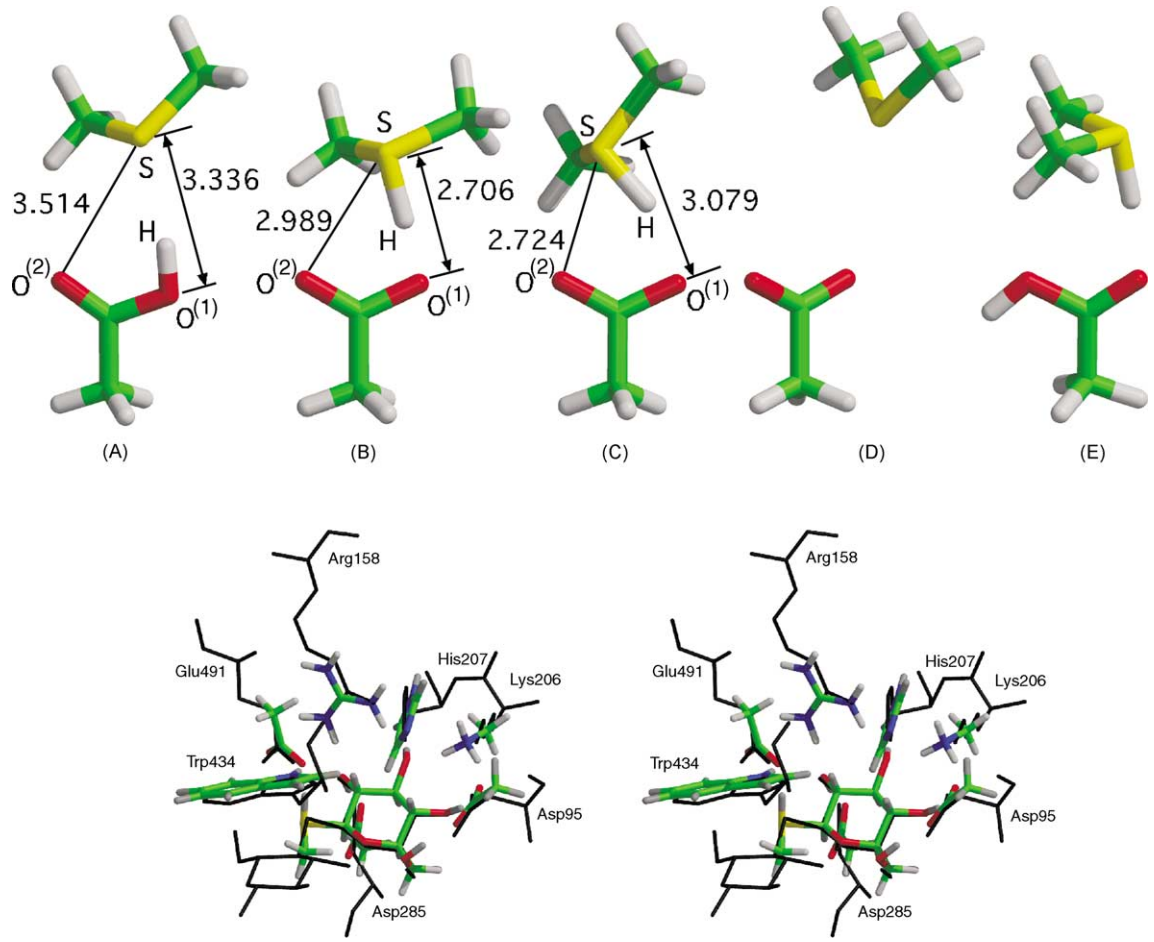


Fig. 1. Optimized geometries. Top panel: complexes between (**A**) acetic acid and dimethylsulfide, (**B**, **C**) the acetate anion and dimethyl-hydrogen-sulfonium, (**D**) the actate anion and dimethylsulfide, and (**E**) acetic acid and dimethyl-hydrogen-sulfonium optimised at the MP2/6-31G(d) level. In complex **B** the S–H distance is fixed at 1.345 Å, while in complex **C** the O···H separation is fixed at 1.75 Å. Bottom panel—stereo view of the optimised active site model (ball-and-stick representation coloured carbon green, nitrogen blue, oxygen red, sulphur yellow) with restraints to the heavy-atom positions of the X-ray structure (black lines). The S–H distance has been constrained to 1.327 Å. This diagram was generated using MOLSCRIPT [32] and Raster3D [33].

The S–H distance in the dimethyl-hydrogen-sulfonium cation at this level of theory is 1.345 Å. In the complex (**B**) between the acetate anion and dimethyl-hydrogen-sulfonium cation in which the distance between the sulphur and hydrogen atom was fixed at 1.345 Å, the distances between the carboxylate oxygen atoms and sulphur were found to be 2.706 and 2.989 Å, differing from the experimental values by less than 0.02 Å. Without the constraint on the S–H distance the geometry of **B** collapses to that of complex **A**. In complex **B** the H···O separation is 1.370 Å, only marginally longer than the S–H distance (fixed 1.345 Å). The proton is, therefore, situated at roughly the midpoint between the $O^{(1)}$ and S atoms.

The geometries of **A** and **B** were determined at several different levels to examine the influence of basis set size and role of electron correlation beyond the MP2 level on the S···O distances. Presented in Table 1 are the S···O distances calculated at six different levels of theory, including the MP2(fu)/6-31G(d) results discussed earlier. The effect of basis set size was examined at the MP2 level; five basis sets were examined, the 6-31G(d) and cc-pVDZ [16,17] bases with and without diffuse functions, and the cc-pVTZ basis. The root mean square deviation (rmsd) from the 6-31G(d) basis level distances is 0.08 Å, with a largest deviation of 0.15 Å (in **A**). The effect of incorporating diffuse functions was to change the distances by less than 0.1 Å. The difference in S···O distances between MP2 and QCISD [18]

Table 1
Comparison of calculated geometries at different levels of theory

	A		B[a]	
	S···$O^{(1)}$	S···$O^{(2)}$	S···$O^{(1)}$	S···$O^{(2)}$
MP2(fu)/6-31G(d)	3.336	3.514	2.706	2.989
MP2(fc)/6-31+G(d)	3.288	3.601	2.750	3.073
MP2(fc)/cc-pVDZ	3.277	3.589	2.658	2.991
MP2(fc)/aug-cc-pVDZ	3.205	3.598	2.700	3.059
MP2(fc)/cc-pVTZ	3.213	3.668	2.668	3.037
QCISD/6-31G(d)	3.396	3.569	2.733	2.992

[a] S–H distance constrained to 1.327 Å.

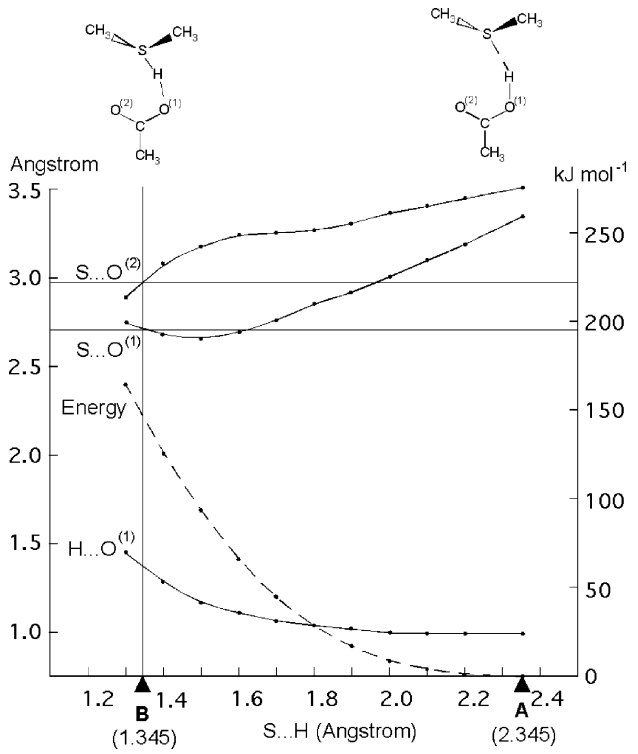


Fig. 2. Variation in structural parameters with restrained S · · · H separation in the complex between acetic acid and dimethylsulfide.

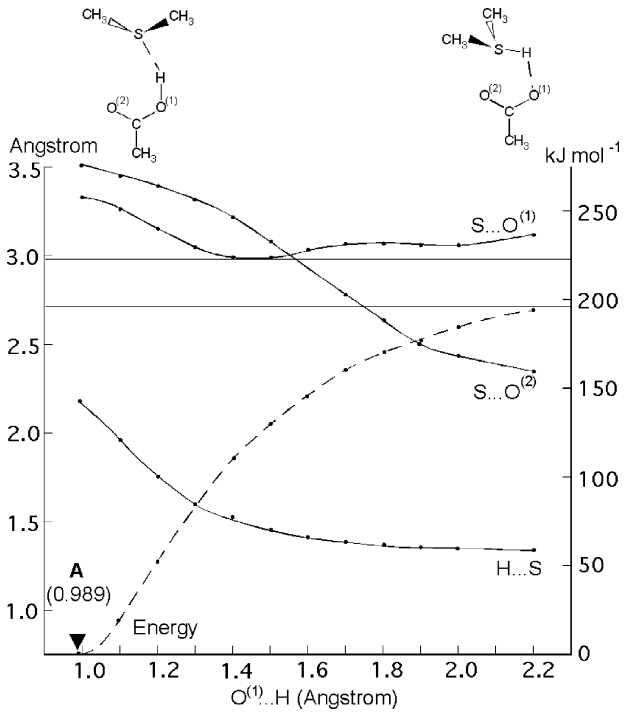


Fig. 3. Variation in structural parameters with restrained H · · · O separation in the complex between acetic acid and dimethylsulfide.

methods was less than 0.06 Å. The S · · · O distances in **A** are more than 0.5 Å longer than in **B** at all levels considered here. The rmsd from the MP2(fu)/6-31G(d) results for all other levels of theory was 0.06 Å. Thus, the effect of increasing the basis set size and level of theory from the MP2(fu)/6-31G(d) level on the S · · · O distances is only (relatively) small, and the MP2(fu)/6-31G(d) results are representative of the higher level calculations.

In Fig. 3 is presented the change in geometry of complex **A** as the O–H distance is increased. Initially the O–H distance is 0.989 Å. When the O–H distance is increased to 1.5 Å the S–H separation is reduced by more than 0.8 Å, from 2.354 to 1.524 Å. At this H · · · O separation, the S · · · $O^{(1)}$ and S · · · $O^{(2)}$ distances are roughly equal (3.005 and 3.082 Å, respectively). This S · · · O separation is similar to the longer of the two distances observed experimentally. The constraint increases the energy by 130 kJ mol^{-1} from **A**. The changes in S · · · $O^{(1)}$ and S · · · H separations are relatively small as the O–H distance is further increased beyond 1.5 Å. The S · · · $O^{(2)}$ separation, however, continues to decrease. At an O · · · H separation of 1.75 Å the S · · · $O^{(2)}$ distance is the same as that observed experimentally, while the S · · · $O^{(1)}$ distance is slightly longer than experiment, 3.07 Å.

The geometry of the complex between the acetate anion and dimethyl-hydrogen-sulfonium cation in which the distance between the thiol-hydrogen atom and carboxylate-oxygen atom was fixed at 1.75 Å (**C**), closely resembles that observed experimentally. Again, without the restraint on the H · · · $O^{(1)}$ distance the geometry of **C** collapses to that of complex **A**.

In both complexes **B** and **C**, obtained by restraining either a short S · · · H separation or a long H · · · O separation, both S · · · O separations are significantly shorter than in the complex **A**. The nature of the interaction between acetate anion and sulphonium cation that results from the two constraints are, however, very different. Complex **B** is characterised by having the proton attached to the sulphur mediating the shortest S · · · O separation, whereas in complex **C** the longest S · · · O separation is mediated by the proton. In both cases, the restraining energy is very large; in excess of 150 kJ mol^{-1} at the MP2(fu)/6-31G(d) level.

In addition to the complexes **A**–**C** we have considered the complex between the acetate anion and dimethyl sulfide (**D**), and the complex between acetic acid and the dimethyl-hydrogen-sulfonium cation (**E**). The distances between the sulphur and carboxylate oxygen atoms in complex **D** are much larger than observed in the experimental structure; one distance (4.501 Å) is more than 1.5 Å longer than the longest separation observed in the X-ray structure. In complex **E** the distance between the sulphur and oxygen atoms involved in forming a hydrogen bond (3.073 Å) differs from the longer of the two X-ray distances by just 0.1 Å. However, the other distance (3.601 Å) is considerably larger than either experimental separation.

The short separation observed between the inhibitor sulphur atom and catalytic acid carboxylate atoms observed in the X-ray crystal structure are compatible only with

complexes **B** and **C**. Complexes **B** and **C** are, however, significantly higher in energy than **A**, and in both cases the geometries are unstable to minimization without constraints. While in complexes **C** and **E** it is clear that the thiol is ionised, in complex **B** the position of the proton midway between the sulphur and oxygen atoms makes the distinction somewhat ambiguous.

Relative energies of complexes **A**, **B** and **C** were calculated at the QCISD(T) level with a large basis set using the G3(MP2) [19] additivity scheme. Since complexes **B** and **C** are not stationary points on the potential energy surface no zero-point vibrational contributions to the energies were considered. Complex **B** lies 130.6 kJ mol^{-1} higher in energy than **A**, while complex **C** lies 146.2 kJ mol^{-1} higher than **A** at this level. When the two O···S distances are fixed to 2.72 and 2.97 Å (**F**), the experimental distances, the minimised H–O distance is 0.994 Å and the S···H separation 1.797 Å. The QCISD(T) energy of this complex lies 36.0 kJ mol^{-1} above **A**. Thus, the energy required to form the two short S···O contacts between the neutral acetic acid and dimethylsulfide is significantly less than that required to move the proton from the acid oxygen atom to the sulphur.

In the systems analogues to **A** and **B** in which the sulphur is replaced with an oxygen atom (i.e. the complex between acetic acid and dimethylether (**A**$_0$) or the complex between the acetate anion and the dimethyloxonium cation (**B**$_0$), the H–O distance was fixed at 0.985 Å, the distance found in the minimised structure of the dimethyl-hydrogen-oxonium cation at the MP2(fu)/6-31G(d) level) the relative energy difference at the QCISD(T) level is 114.4 kJ mol^{-1}. Thus, there is a large energy requirement for transfer of the proton from the acid to both the sulphur and oxygen atom. In **A**$_0$ the distances between the carboxyl oxygen atoms and the ether oxygen atom are 2.760 and 3.353. In **B**$_0$ these distances are 2.349 and 2.987 Å. Notably, the decrease in S···O distances (as the proton is moved from the carboxylate oxygen to the sulphur, i.e. from **A** to **B**) are roughly 0.3 Å larger than those observed in the transition from **A**$_0$ to **B**$_0$.

At the HF/3-21G* level the S–H distance in the dimethyl-hydrogen-sulfonium molecule is 1.327 Å. At this level of theory and S–H distance, the S···O separations in complex **B** are 2.723 and 2.969 Å, thereby differing from the distances observed in the complex between the thiol inhibitor and barley β-D-glucan glucohydrolase by no more then 0.003 Å. Although the remarkable level of agreement is obviously fortuitous, it does support this level of calculation as a good choice for reproducing geometries [20]. These same distances in complex **A** are 3.403 and 3.570 Å. At this level, complex **B** lies 183.2 kJ mol^{-1} higher in energy than complex **A**. In complex **C** the S···O distances are 3.024 and 2.604 Å when the H···O distance is fixed at 1.75 Å. Complex **C** lies 196.6 kJ mol^{-1} higher than **A** at this level. Thus, the HF/3-21G* relative energies differ significantly from those obtained at the QCISD(T) level with the large basis set using MP2 optimised geometries. Density functional theory calculations, at the B3LYP/6-31+G(d) level with the HF/3-21G* optimised geometries, however, predict energies of **B** and **C** relative to **A** of 132.9 and 141.8 kJ mol^{-1}, in remarkably good agreement with the estimates using the G3(MP2) additivity scheme. Thus, the B3LYP/6-31+G(d)//HF/3-21G* calculations provide good quality relative-energies.

2.2. Active site model of barley β-D-glucan glucohydrolase

The large size of proteins precludes the use of QM methods on the whole system. An alternative approach uses a model system that represents only the essential features of the enzyme active site and substrate [21]. Active site residues are replaced with small molecule fragments in the same relative orientation as that found experimentally. Minimisation of the geometry in the gas-phase, however, can lead to large movements of the fragments. In addition, the calculated energy neglects the effects of surrounding protein and solvent.

A flat-bottomed harmonic potential (Eq. (1)) can be applied to any atom to restrain its position to a reference point. The restraint energy $E_{\text{restraint}}$ depends on the displacement $r = (r_x, r_y, r_z)$ of the atom from a predefined reference point by a distance r, the force constant of the harmonic restraint k, and the switch distance S. Thus, a restoring force is applied to the individual components of the gradient (Eq. (2)) for those centres that are displaced beyond the switch distance.

$$E_{\text{restraint}} = \begin{cases} 0 & \text{if } r \leq S \\ k(r-S)^2 & \text{if } r > S \end{cases} \quad (1)$$

$$\frac{\delta E_{\text{restraint}}(r)}{\delta r_i} = \begin{cases} 0 & \text{if } r \leq S \\ 2kr_i \dfrac{(r-S)}{r} & \text{if } r > S \end{cases} \quad (2)$$

Restraints provide atoms with limited mobility that enables them to be tethered to positions without being held fixed during minimisation. The switch distance can be chosen to account for uncertainties associated with the experimental structure, and to permit minimisation of functional groups within the macromolecular assemblies without the spatial arrangement of these groups being compromised. The bias in the final QM minimised geometry, then, is toward the experimental geometry only, atoms beyond the active site have no electronic influence on the geometry. This method differs from methods that apply a molecular mechanical (MM) treatment to the region omitted in the QM calculation, QM/MM calculations [22], where the bias is toward the MM model.

The GAMESS program [23] was modified to include restraints during optimisation. A model system of the active site of barley β-D-glucan glucohydrolase was assembled including the residues Asp95, Arg158, Lys206, His207, Asp285, Trp434, Glu491 and the thio-glycoside inhibitor. The histidine residue was replaced with an imidazole ring, glutamate and aspartate residues with an acetate anion,

tryptophane with an indole, and lysine with a methylammonium cation. This represents all side-chain functional groups that from hydrogen bonds with the substrate. Trp434 does not interact directly with the substrate, but forms a hydrogen bond between the $N_{\varepsilon 1}$ of the indole sidechain and a carboxylate oxygen atom of the catalytic acid. The protein binds the substrate with the terminal glycan in the −1 subsite, and the penultimate glycan in the +1 subsite. The glycan in the +1 subsite was replaced by a methyl group. The final system comprised a total of 92 atoms (45 nonhydrogen atoms). All nonhydrogen atoms (including the sulphur atom of the thiol glycosidic link) were restrained to the positions found in the refined X-ray structure (PDB accession code 1IEX). A switch distance of 0.5 Å and force constant of 0.5 mdyn/Å was applied. The switch distance is sufficiently large to accommodate the range of separations between the sulphur atom of the thioglycosidic link and the carboxylate oxygen atoms of the acid expected from the calculations above without imposing large restraints. The force constant used is comparable to the stretch energy of a covalent bond. The geometry of four minima were obtained at the HF/3-21G* level. Two of these minima had the inhibitor and acetic acid group corresponding to Glu491 neutral, with the carboxyl proton on either oxygen (I and II). The other two minima had the acetate corresponding to Glu491 ionised (negatively charged) and the inhibitor sulphur atom protonated (positively charged), with the proton on the inhibitor sulphur atom in positions corresponding to complexes **B** or **C** (III and IV, respectively). System III collapses to I upon minimisation; to prevent this, the S–H distance was fixed at 1.327 Å. Minimisation of system IV leaves the proton bonded to the inhibitor sulphur atom (at a distance of 1.37 Å). The indole is necessary to stabilise the proton on the inhibitor sulphur atom; without the indole, system IV collapses to II. An overlay of the inhibitor and active site residues from the X-ray analysis with system III is presented in Fig. 1. The relative energies of complexes I–IV are presented in Table 2. The separation between the carboxylate oxygens and sulphur atom, along with the rmsd of the distances from the X-ray positions of the nonhydrogen atoms in each system, are presented in Table 3.

Table 2
Calculated relative energies (kJ mol^{-1})

	I	II	III[a]	IV	V[b]
HF/3-21G*	0.0	8.0	96.2	138.0	45.5
B3LYP/6-31+G(d)[c]	0.0	7.0	62.5	100.6	33.6
B3LYP/6-31+G(d)/$\mathcal{F}$[d]	0.0	−18.1	37.4	79.4	6.5
B3LYP/6-31+G(d)/$\mathcal{F}$+$\nabla\phi$[e]	0.0	1.4	5.7	56.3	6.5

[a] S···H distance constrained to 1.327 Å.
[b] Both S···O distances constrained.
[c] Calculated on the geometries obtained at the HF/3-21G* level using harmonic restraints.
[d] Point charge field included in energy.
[e] Solvation energies obtained from DelPhi program added to B3LYP/6-31+G(d)/$\mathcal{F}$ energies.

Table 3
Comparison of S···O distances (Å)

	Experimental	I	II	III[a]	IV
S···O$^{(1)}$	2.72	3.18	3.52	2.88	2.78
S···O$^{(2)}$	2.97	3.58	3.34	3.26	3.02
RMSD[b]		0.411	0.411	0.413	0.421

[a] S–H distance constrained to 1.327 Å.
[b] Root-mean-square deviation of distances of restrained atoms.

At the HF/3-21G* level, system I has the lowest energy. Systems III and IV lie 96.2 and 138.0 kJ mol^{-1}, respectively, higher. The distances between the carboxylate oxygen atoms and the sulphur atom of the inhibitor for systems I and II were significantly larger than the experimental values, whereas for III and IV the distances are in reasonable agreement; in particular, for IV the difference in distances is less than 0.1 Å. At the B3LYP/6-31+G(d) level, the energy of II relative to I is roughly the same as the HF/3-21G* level predictions, whereas the energies of III and IV (relative to I) are more than 30 kJ mol^{-1} lower. However, the energies of III and IV remain very high relative to I.

The effect of the protein environment on the proton-transfer energy can be estimated by including the atomic partial charges from the remainder of the protein in the QM calculation as point charges. For protein residues within 8.0 Å of the ligand, charges were taken from the SIP data set for amino acids [24]. Ionizable amino acids, ASP, GLU, LYS and ARG, beyond 8.0 Å of the ligand were assigned unit charge on the C_γ, C_δ, N_ξ and C_ζ atoms, respectively. Energies calculated at the B3LYP/6-31+G(d) level including partial atomic charges (B3LYP/6-31+G(d)/$\mathcal{F}$) are presented in Table 2.

The effect on the B3LYP energies of including the electrostatic field of the point charges from the remainder of the protein was to reduced the energies of II, III and IV relative to I by roughly 20 kJ mol^{-1}. Transfer of the proton from the acetic acid to the sulphur atom, however, continues to require a significant amount of energy, ∼55 kJ mol^{-1}.

The interaction energy of the solvent with the protein can also be determined. The electrostatic contribution to the free energy of solvation of the system was calculated using the DelPhi program [25] (Molecular Simulations, San Diego, CA, 1997). Hydrogen atoms were added to all protein and inhibitor atoms of the X-ray structure to fill unsatisfied valencies using the InsightII package (Molecular Simulations, San Diego, CA, 1997). The position of all hydrogen atoms was subjected to energy minimisation (keeping protein, inhibitor and water atoms fixed at the positions determined from X-ray crystallography) using the Discover program (Molecular Simulations, San Diego, CA, 1997). Force field parameters for protein and inhibitor atoms were taken from the CVFF force field. SIP atomic radii and charges [26] were used for all atoms in the calculation of solvation energies. CHELPG charges [27] were calculated for the inhibitor (disaccharide) at the HF/6-31+G(d) level. Charges

and radii were taken from the SIP data set for amino acids [24]. All water molecules were removed. The ionic strength was set to zero and dielectric constant values of 1 and 78.54 were assumed for solutes and solvent, respectively. A grid of 197 × 197 × 197 points was used, yielding a grid resolution of 0.5 Å. This maintained a solvent boundary of at least 10.0 Å around the protein. By combining these energies with the QM energies in the point-charge field an estimate of the relative aqueous free energies can be obtained. The relative energies in the aqueous protein environment of each of the systems I–IV (B3LYP/6-31+G(d)/F+$\nabla\phi$) are presented in Table 2. The effect of the solvent on the system is to dramatically reduce the energy of III and IV relative to I.

The long S···O separations in I indicate that the environment of the active site is not capable of supporting both neutral acid and inhibitor with the short S···O separations observed experimentally. Constraining the S···O separations to those distances observed experimentally (like **F**) leads to a structure (V) that lies 45.5 kJ mol^{-1} higher than I at the HF/3-21G* level. The proton remains bound the acid in this structure, 1.78 Å from the sulphur of the inhibitor. At the B3LYP/6-31+G(d) level the energy difference between I and V is 33.6 kJ mol^{-1}. Inclusion of the electrostatic field reduces this difference to 6.5 kJ mol^{-1}, the solvation energy contributions are identical to those calculated for I. Thus the energy of III and this system (V) in which the S···O separations are fixed are indistinguishable at this level.

Stabilization of negative charge on the carboxylate (corresponding to Glu491) by the indole (Trp434) appears to be important in the description of this system. Replacing the atoms of the indole group with partial atomic charges in the B3LYP calculations increases the energy between I and III to 79.1 kJ mol^{-1} (i.e. the energy difference increases by 41.7 kJ mol^{-1}). Clearly, a quantum mechanical treatment is required to correctly describe the effect of the interaction of the indole with the acid. The residues Trp286 and Trp434 serve as a molecular 'clamp' around the glucosyl residue bound at the +1 subsite [14]. Trp434 is conserved in all known sequences of higher plant β-D-glucan glucohydrolases [28].

Indole is a weak base, in aqueous solution the acidity constant, $pK_{BH+} = -4$, is significantly lower than that of methylamine [29], $pK_{BH+} = 10.62$. In the gas phase, however, the proton affinity of indole [8] 933.4 kJ mol^{-1}, is significantly greater than that of methylamine, [8] 899.0 kJ mol^{-1}. In the protein environment, the protonation state of the indole side chain of tryptophane will be determined by the charge state of its neighbours. Protonation of Trp434 is likely to stabilize the negative charge on Glu491. We have considered the effect on the energy difference between I and IV of having the indole of Trp434 in the +1 charge state. At the B3LYP/6-31+G(d) level with point charge field and solvation energies included, the energy of III is 9.6 kJ mol^{-1} lower than that of I. The energy of IV relative to I is also reduced, but IV remains higher in energy than I by 47.0 kJ mol^{-1}. When the atoms of the positively charged indole are replaced with point charges, I lies 35.3 kJ mol^{-1} higher than III, again highlighting the importance of ensuring an adequate quantum mechanical description in these systems.

3. Discussion

In the paradigmatic mechanism [30] of glycoside hydrolysis proton transfer from the catalytic acid to the glycosidic oxygen of the substrate leads to an oxocarbenium intermediate. A stepwise path involving an oxonium intermediate is predicted for substrates in which the aglycon confers an increased proton affinity on the glycosidic oxygen [4]. In general, the oxonium ion does not present itself as a stable intermediate.

The enzymatic mechanism of barley β-D-glycan glucohydrolase has been studied extensively. The identity of several intermediate analogues, including the complexes with product [31] and the glycosyl-enzyme covalent intermediate [14], has been determined and their structures are known.

The identity of the *S*-cellobioside moiety of the nonhydrolysable $4^{I},4^{III},4^{V}$-*S*-trithiocellohexaose has been investigated in the current study. The X-ray crystal structure [14] revealed two very short contacts between the thio-glycosidic sulphur atom and the carboxylate oxygen atoms of the catalytic acid. The remarkable agreement in these distances between experiment and the MP2 optimised complex of the acetate anion and sulfonium cation is compelling evidence that the catalytic acid and inhibitor sulphur atom are both charged in the protein. In contrast, however, the extremely large energy required for displacement of the proton from the acid oxygen to the sulphur is equally compelling evidence that such a complex is unlikely. With such conflicting evidence, the identity of the bound species cannot be satisfactorily resolved from the results of these calculations.

We have shown here that the immediate environment of the inhibitor significantly reduces the energy difference. The energy difference of **B** and **C** relative to **A** is reduced from 132.9 and 141.8 kJ mol^{-1}, respectively, to 62.5 and 100.6 kJ mol^{-1} (energy of III and IV, respectively, relative to I at the B3LYP/6-31+G(d)//HF/3-21G* level). The electrostatic environment of the remainder of the protein reduces these differences by 25.1 and 21.2 kJ mol^{-1}, respectively, while the solvent reduces the differences by 31.7 and 23.1 kJ mol^{-1}. Thus, the proton is still predicted to be bound to the carboxyl oxygen, but one of the systems with the proton bound to the inhibitor sulphur (III) lies just 7 kJ mol^{-1} higher in energy. For III the S···O distances differ from experiment by 0.16 and 0.29 Å. For IV, however, the difference in S···O distances are 0.06 and 0.05 Å, yet the energy of this system lies 56.3 kJ mol^{-1} above I. Thus, where the calculated geometry is in good agreement with experiment the calculated energy remains high, and where the energy is low the calculated geometry is in significant disagreement with experiment.

Given the greater C–S bond lengths in the thiol-inhibitor than the C–O lengths of the natural substrate, it might be expected that, in order for the thio-glycoside to be accommodated within the active site, the sulphur must necessarily have shorter contacts with atoms of the neighbouring residues. The distance between carboxylate carbon atoms of the catalytic acid (Glu491) and nucleophile (Asp285) in the X-ray structure with the thio-glycoside is 7.21 Å, indistinguishable from the distance with glucose in the active site (7.20 Å). In structures I and II this distance is 7.33 and 7.41 Å, respectively, while in structures II and IV the distance is 7.10 and 7.30 Å. Thus, by this measure, there is a small contraction in the active when the sulphur of the inhibitor is protonated. The calculations on the systems I–IV were conducted with harmonic restraints on the heavy atoms. The atoms are, nonetheless, afforded limited mobility, which manifests in an rmsd of restrained atoms of roughly 0.4 Å in all systems. This level of mobility may not, however, be available to the individual atoms within the complete protein environment, and the protein environment may mechanically restrict the S···O distances. The energy required to constrain the S···O distances to those observed experimentally has been calculated at 6.5 kJ mol^{-1}. It is notable that the electrostatic environment of the protein beyond the active site can stabilise the energy of **V** relative to **I** by 27.1 kJ mol^{-1}.

In all the calculations presented here there has been no attempt to consider vibrational or thermal contributions to the energy differences. Also, only the electrostatic component of the solvation energy has been taken into account. The effects of these contributions to the energy of proton transfer energy, however, are unlikely to be large. The results presented here, especially the relative energies, should therefore be interpreted as providing a qualitative insight into the potential for the various systems. Thus, the energies of I, III and V are indistinguishable from one another.

4. Conclusions

In the X-ray structure of barley β-D-glycan glucohydrolase in complex with a nonhydrolyzable thioglycoside inhibitor the distances between the catalytic acid carboxyl oxygen atoms and the inhibitor sulphur atom are 2.72 and 2.97 Å. At the MP2/6-31G(d) level there is remarkable agreement in the complex between the acetate anion and the dimethyl-hydrogen-sulfonium cation, where the distances are 2.706 and 2.989 Å. High level estimates for the energy of this complex suggest it lies 130.6 kJ mol^{-1} higher in energy than the complex between the neutral acetic acid and dimethylsulfide, but collapses to the complex of neutral species without constraining the S–H distance. Incorporation of components of the active site into the QM calculations significantly reduces the difference in energy, and can maintain the proton on the sulphur atom of the inhibitor without imposing constraints. Incorporation of the effects of the electrostatic field of the remainder of the protein and of the solvent environment reduces the energy difference further. At the level of the calculations presented here the energy of the two states with the proton bound to the acid or the inhibitor sulphur atom are indistinguishable. The energy required for mechanical constraint of the S···O distances, however, is negligible. Thus, it is not possible to resolve the origins of the short contacts, although we can confirm either mechanical constraint or sulphonium ion formation is possible.

Acknowledgements

A generous allocation of time from the Victorian and Australian Partnerships for Advanced Computing is gratefully acknowledged. Insightful discussion with Drs. Maria Hrmova, Jose Varghese and Peter Colman is gratefully appreciated.

References

[1] L. Hoise, P.J. Marshall, M.L. Sinnott, Failure of the antiperiplanar lone pair hypothesis in glycoside hydrolysis. Synthesis, conformation, and hydrolysis of α-D-xylopyranosyl- and α-D-glucopyranosyl-pyridinium salts, J. Chem. Soc. Perkin Trans. II (1984) 1121–1131.
[2] X. Huang, K.S.E. Tanaka, A.J. Bennet, Glucosidase-catalysed hydrolysis of α-D-glucopyranosyl pyridinium salts: kinetic evidence for nucleophilic involvement at the glucosidation transition state, J. Am. Chem. Soc. 119 (1997) 11147–11154.
[3] K.S.E. Tanaka, J. Zhu, X. Huang, F. Lipari, A.J. Bennet, Glycosidase-catalyzed hydrolysis of 2-deoxyglucopyranosyl pyridinium salts: effect of the 2-OH group on binding and catalysis, Can. J. Chem. 78 (2000) 577–582.
[4] B.J. Smith, A conformational study of 2-oxanol: insight into the role of ring distortion on enzyme-catalysed glycosidic bond cleavage, J. Am. Chem. Soc. 119 (1997) 2699–2706.
[5] B.J. Smith, Effect of ring distorion on the acid hydrolysis of 2-methylsulfanyloxane, J. Phys. Chem. A 102 (1998) 4728–4733.
[6] H. Driguez, Thiooligosaccharides in glycobiology, Top. Curr. Chem. 187 (1997) 85–116.
[7] W.P. Burmeister, S. Cottaz, H. Driguez, R. Iori, S. Palmieri, B. Henrissat, The crystal structure of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an *S*-glycosidase, Structure 5 (1997) 663–675.
[8] W.G. Mallard, P.J. Linstrom, NIST Chemistry Webbook, NIST Standard Reference Database Number 69, July 2001.
[9] M.L. Sinnot, M.I. Page, The Chemistry of Enzyme Action, Elsevier, Amsterdam, 1984, pp. 389–431.
[10] G. Sulzenbacher, H. Driguez, B. Henrissat, M. Schülein, G.J. Davies, Structure of the fusarium oxysporum endoglucanase I with a nonhydrolysable substrate analogue—substrate distortion gives rise to the preferred axial orientation for the leaving group, Biochemistry 35 (1996) 15280–15287.
[11] J. Zou, G.J. Kleywegt, J. Ståhlberg, H. Driguez, W. Nerinckx, M. Claeyssens, A. Koivula, T.T. Teeri, T.A. Jones, Crystallographic evidence for substrate ring distortion and protein conformational changes during catalysis in cellobiohydrolase Cel6A from *Trichoderma reesei*, Structure 7 (1999) 1035–1045.
[12] M. Spezio, D.B. Wilson, P.A. Karplus, Crystal structure of the catalytic domain of a thermophilic endocellulase, Biochemistry 32 (1993) 9906–9916.

[13] B.K. Barr, D.E. Wolfgang, K. Piens, M. Claeyssens, D.B. Wilson, Active-site binding of glycosides by *Thermomonospora fusca* endocellulase E2, Biochemistry 37 (1998) 9220–9229.
[14] M. Hrmova, J.N. Varghese, D.R., B.J. Smith, H. Driguez, G.B. Fincher, Catalytic mechanisms and reaction intermediates along the hydrolytic pathway of a plant β-D-glucan glucosidase, Structure 9 (2001) 1005–1016.
[15] H.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O.T. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomeiil, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, J.A. Pople, Gaussian 98, Revision A.7., 1998.
[16] T.H. Dunning Jr., Gaussian basis sets for use in correlated molecular orbital calculations. Part I. The atoms boron through neon and hydrogen, J. Chem. Phys. 90 (1989) 1007–1023.
[17] R.A. Kendall, T.H. Dunning Jr., R.J. Harrison, Electron affinities of the first row atoms revisited. Sytematic basis set and wavefunctions, J. Chem. Phys. 96 (1992) 6796–6806.
[18] J.A. Pople, M. Head-Gordon, K. Raghavachari, Quadratic configuration interaction: a general technique for determining electron correlation energies, J. Chem. Phys. 87 (1987) 5968–5975.
[19] L.A. Curtiss, P.C. Redfern, K. Raghavachari, V. Rassolov, J.A. Pople, Gaussian-3 theory using reduced Moller–Plesset order, J. Chem. Phys. 110 (1999) 4703–4709.
[20] W.J. Hehre, Practical Strategies for Electronic Structure Calculations, Wavefunction, California, 1995.
[21] A.J. Mulholland, G.H. Grant, W.G. Richards, Computer modelling of enzyme catalysed reaction mechanisms, Protein Eng. 6 (1993) 133–147.
[22] J. Gao, in: K.B. Lipkowitz, D.B. Boyd, Methods and applications of combined quantum mechanical and molecular mechanical potentials, Reviews in Computational Chemistry, vol. 7, 1995, pp. 119–185.
[23] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.J. Jensen, S. Koseki, N. Matsunaga, K.A. Nguyen, S. Su, T.L. Windus, M. Dupuis, J.A. Montgomery Jr., General atomic and molecular electronic structure system, J. Comput. Chem. 14 (1993) 1347–1363.
[24] B.J. Smith, Solvation parameters for amino acids, J. Comput. Chem. 20 (1999) 428–442.
[25] A. Nicholls, B. Honig, A rapid finite difference algorithm, utilizing successive over-relaxation to solve Poisson–Boltzmann equations, J. Comput. Chem. 12 (1991) 435–445.
[26] B.J. Smith, N.E. Hall, Atomic radii: incorporation of solvation effects, J. Comput. Chem. 19 (1998) 1482–1493.
[27] C.M. Breneman, K.B. Wiberg, Determining atom-centered monoploes from electrostatic potentials. The need for high sampling density in formaldehyde conformational analysis, J. Comput. Chem. 11 (1990) 361–373.
[28] M. Hrmova, R. DeGori, B.J. Smith, J.K. Fairweather, H. Driguez, J.N. Varghese, G.B. Fincher, Structural basis for broad substrate specificity in higher plant β-D-glucan glucohydrolases, The Plant Cell 14 (2002) 1003–1052.
[29] A. Albert, E.P. Sergeant, Ionization Constants of Acids and Bases, Methuen, London, 1962.
[30] D.L. Zechel, S.G. Withers, Glycosidase mechanisms: anatomy of a finely tuned catalyst, Acc. Chem. Res. 33 (2000) 11–18.
[31] J.N. Varghese, M. Hrmova, G.B. Fincher, Three-dimensional structure of a barley β-D-glucan exohydrolase, a family 3 glycosyl hydrolase, Structure 7 (1999) 179–190.
[32] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, Appl. Cryst. 24 (1991) 946–950.
[33] E.A. Merritt, D.J. Bacon, Raster3D: photorealistic molecular graphics, Methods Enzymol. 277 (1997) 505–524.