View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Bath Research Portal

Citation for published version:
Mhlongo, NN, Skelton, AA, Kruger, G, Soliman, MES & Williams, IH 2014, 'A critical survey of average distances between catalytic carboxyl groups in glycoside hydrolases', Proteins: Structure, Function, and Bioinformatics, vol. 82, no. 9, pp. 1747-1755. https://doi.org/10.1002/prot.24528

DOI:
10.1002/prot.24528

Publication date:
2014

Document Version
Peer reviewed version

Link to publication

This is the accepted version of the following article: Mhlongo, N. N., Skelton, A. A., Kruger, G., Soliman, M. E.S. and Williams, I. H. (2014), A critical survey of average distances between catalytic carboxyl groups in glycoside hydrolases. Proteins, 82: 1747–1755, which has been published in final form at http://dx.doi.org/10.1002/prot.24528

University of Bath

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 13. May. 2019

A critical survey of average distances between catalytic carboxyl groups in glycoside hydrolases

Ndumiso N. Mhlongo,[a] Adam A. Skelton,[a] Gert Kruger,[b] Mahmoud E. S. Soliman[a*] and Ian H. Williams[c*]

[a] School of Health Sciences, University of KwaZulu-Natal, Durban 4001, South Africa

[b] Catalysis and Peptide Research Unit, School of Health Sciences, University of KwaZulu-Natal, Durban 4001, South Africa

[c] Department of Chemistry, University of Bath, Bath BA2 7AY, United Kingdom

* *Corresponding authors:* soliman@ukzn.ac.za, i.h.williams@bath.ac.uk

Abstract

Published X-ray crystallographic structures for glycoside hydrolases (GHs) from 39 different families are surveyed according to some rigorous selection criteria and the distances separating 208 pairs of catalytic carboxyl groups (20 α-retaining, 87 β-retaining, 38 α-inverting and 63 β-inverting) are analysed. First, the average of all four inter-carboxyl O···O distances for each pair is determined; second, the mean of all the pair-averages within each GH family is determined; third, means are determined for groups of GH families. No significant differences are found for free structures as compared with those complexed with a ligand in the active site of the enzyme, nor for α-GHs as compared with β-GHs. The mean and standard deviation (1σ) of the unimodal distribution of average O···O distances for all families of inverting GHs is 8 ± 2 Å, with a very wide range from 5 Å (GH82) to nearly 13 Å (GH46). The distribution of average O···O distances for all families of retaining GHs appears to be bimodal: the means and standard deviations of the two groups are 4.8 ± 0.3 Å and 6.4 ± 0.6 Å. These average values are more representative, and more likely to be meaningful, than the often-quoted literature values, which are based on a very small sample of structures. The newly-updated average values proposed here may alter perceptions about what separations between catalytic residues are "normal" or "abnormal" for GHs.

Key words: Glycoside hydrolases (GHs), structural analysis, inter-carboxyl separation

Introduction

Glycoside hydrolases (GHs, or glycosidases) play a variety of vital roles in biological processes and are enzymes that catalyse the hydrolysis of glycosidic linkages between carbohydrate molecules. Their structures, functions and mechanisms have been the subject of several recent reviews.[1-7] Most GHs use one or the other of two distinct mechanisms.[8] Retaining glycosidases employ a double-displacement mechanism, involving a covalent glycosyl-enzyme intermediate (Fig. 1), leading to net retention of the stereochemical configuration at the anomeric centre. Inverting glycosidases function through a direct displacement mechanism (Fig. 2) leading to net inversion of the stereochemical configuration at the anomeric centre. Both mechanisms operate via oxacarbenium-ion-like transition states and both involve a pair of carboxylic acid functional groups (either aspartic acid or glutamic acid) as the catalytic residues. In retaining GHs one carboxyl group functions as an acid/base catalyst and the other as a nucleophile, whereas in inverting GHs one functions as a general acid and the other as a general base (Figs. 1 and 2). Some GHs employ completely different mechanisms including, for example, neighbouring-group participation,[9] but these are not the primary subject of this survey.

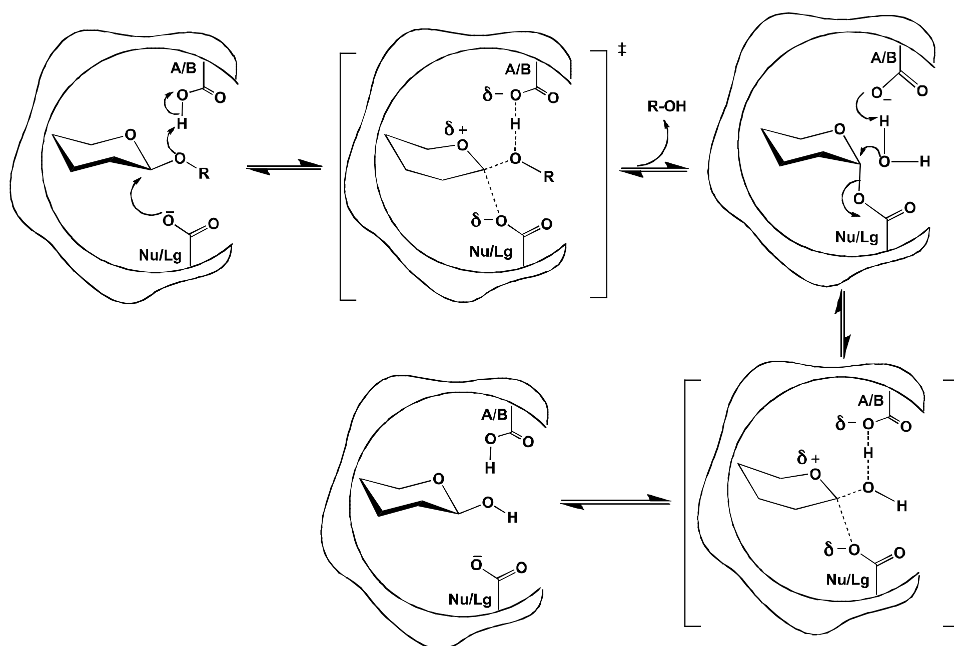

Figure 1.
Proposed mechanism of retaining glycosidases: A/B refers to acid/base catalytic residue and Nu/Lg refers to nucleophile/leaving group residue.

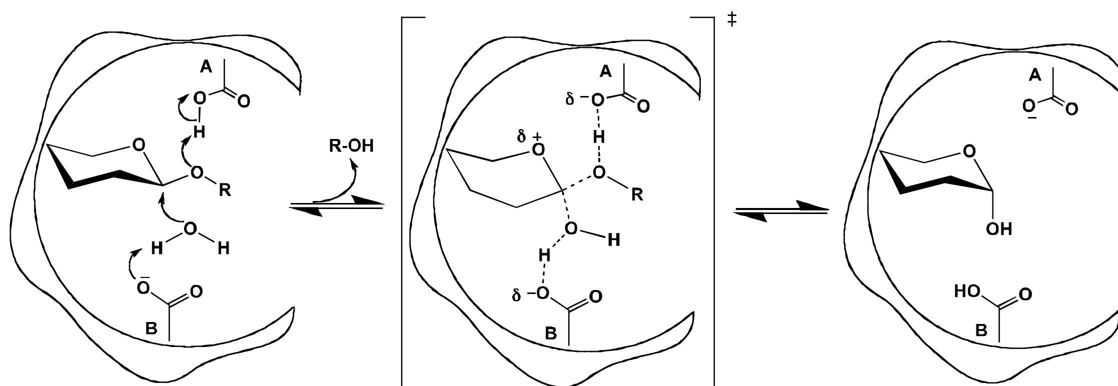

Figure 2.
Proposed mechanism of inverting glycosidases: A refers to catalytic acid residue and B refers to catalytic base residue.

In their influential 1994 review, McCarter and Withers wrote:[10]

> 'Inverting enzymes have an active-site architecture superficially similar to that of retaining enzymes, with two essential carboxylic acids on opposite faces of the substrate binding cavity. Closer inspection, however, reveals that the distance between the catalytic residues (a general acid and a general base) is significantly greater than that between the catalytic carboxylates of retaining enzymes. Indeed, measurement of the separations between the carboxyl oxygens (the average of the four possible distances between the four oxygens) of ten structurally defined glycosidases yielded average distances of 4.8 and 5.3 Å for retaining α- and β-glycosidases, respectively, whereas the average for inverting α- and β-glycosidases was 9.0 and 9.5 Å, respectively. The greater separation for the inverting enzymes is significant and is presumably required for an inverting mechanism in which the nucleophilic water, as well as the substrate, must be positioned between the carboxyl groups.'[10]

The primary source is a contemporaneous paper by Withers and co-workers which cites the actual enzymes considered: three each of α-retaining, β-retaining and α-inverting GHs together with a single β-inverting GH.[11] In view of the very limited extent of their sample, the four average distances reported should have been regarded as indicative and preliminary rather than fully representative and definitive in nature. However, since then it has been quite common to read in the GH literature statements along the lines of 'the distance between the two catalytic residues is longer than the 5 Å usually observed in GHs with a retaining mechanism' or 'the catalytic residues could not be identified unequivocally on the basis of the distance criteria suggested for the inverting mechanism' which appear to imply normative roles for these average distances.

The purpose of this paper is to provide more reliable evaluations of the average distances between the catalytic carboxyl groups in retaining α- and β-GHs and inverting α- and β-GHs, based upon

critical consideration of high-resolution X-ray structural data for a wider range of GH families (as classified by Henrissat[12]) now available and including both free and ligand-bound enzymes. This survey should help to determine what ranges are "normal" for the several classes of GHs and thus how normative the previously published and frequently cited averages actually are. In turn this provides insight into whether or not these inter-carboxyl separations are indeed mechanistically determinative.

METHODS

The CAZY database[13] and the CAZypedia website[14] are valuable and (almost) comprehensive sources of curated information on all aspects of GH structure, function and mechanism, and they provided the starting points for this survey. The CAZY database points to 775 (or more) GH X-ray crystallographic structures in the Protein Data Bank[15] now covering 132 GH families. For our present purposes, a total of 136 structures were selected (69 retaining, 67 inverting) subject to several criteria:

- only GHs thought to function by means of the double-displacement (retaining) and direct displacement (inverting) mechanisms (*cf*. Figs 1 and 2) were considered;
- only GHs containing a pair of catalytically functional carboxyl residues (general acid/base and nucleophile for retaining and general acid and general base for inverting) were considered;
- only GHs in which the specific pair of catalytically functional carboxyl groups has been identified with at least a degree of certainty were considered;
- only wild-type GHs were considered;
- only X-ray structures of the highest resolution (≤ 2.0 Å for retaining and ≤ 2.5 Å for inverting GHs) were considered.

The 136 selected structures include at least one example from each of 39 different GH families and, because some X-ray structures contain more than one sub-unit in the unit cell, the total number of carboxyl-group pairs in this survey is 211. In every case the original published paper has been consulted as the authoritative source of structural and mechanistic information, such as whether the enzyme is retaining or inverting, which carboxylic acid residues are the catalytically functional pair and which role is played by each member of the pair. In a few cases with three catalytic residues in the active site, the two most aligned carboxyl groups were selected.

Additionally, a further 91 structures of retaining GHs employing neighbouring-group participation (NGP) mechanisms were considered for comparison.

We have followed Withers[10,11] by determining all interatomic distances involving either oxygen atom ($O_{\delta1}$ and $O_{\delta2}$ or $O_{\varepsilon1}$ and $O_{\varepsilon2}$ of the side chains of aspartic acid or glutamic acid, respectively)

of one carboxyl group with either oxygen of the other carboxyl group (Fig.3); distances between oxygen atoms in the same carboxyl group ($O_{\delta1}\cdots O_{\delta2}$ or $O_{\varepsilon1}\cdots O_{\varepsilon2}$) are not of interest here. These four distances (d_{11}, d_{12}, d_{21} and d_{22}) were measured (using the selection tools in the VMD visualization program[16]) and the arithmetic mean $D_{OO} = ¼(d_{11} + d_{12} + d_{21} + d_{22})$ was determined.

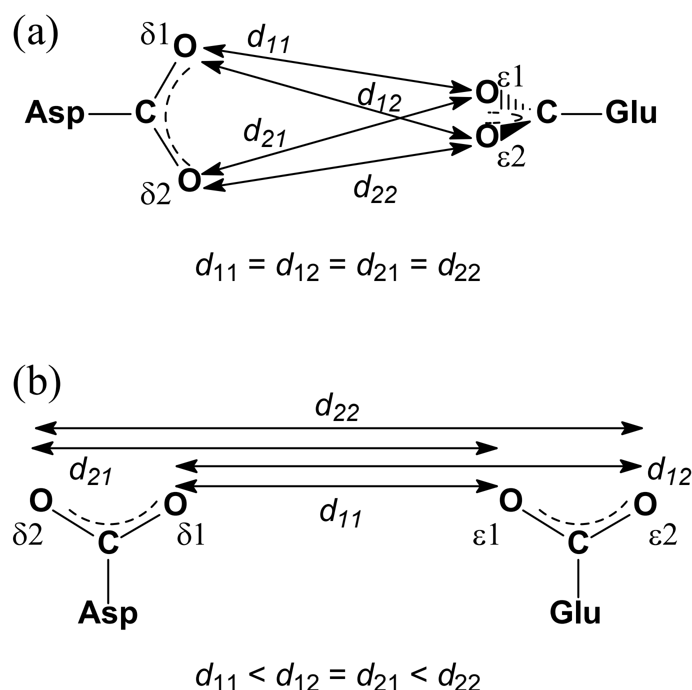

Figure 3.
Definition of inter-carboxyl group $O\cdots O$ distances and their inter-relationships for extreme cases of O atoms (a) at the vertices of an elongated tetrahedron (local D_{2d} symmetry and (b) aligned in a collinear manner.

RESULTS

A complete compilation of the relevant data for the 136 selected structures and 208 carboxyl-group pairs included in this survey is presented as Table S1 of the Supplementary Information. This includes the GH family and its clan (where assigned), the PDB code, the resolution and R-factor of the X-ray crystal structure, the identity of the catalytic residues, whether the structure is free or complexed, together with the individual inter-carboxyl-group $O\cdots O$ distances (d_{11}, d_{12}, d_{21} and d_{22}) and their average value D_{OO} for each structure. Table I contains mean values $\langle D_{OO}\rangle$ of these average distances and standard deviations (1σ) over the total number of structures included in the survey for each GH family. Table II combines these mean values $\langle D_{OO}\rangle$ across the categories of α- and β-retaining and α- and β-inverting GHs.

Fig. 4 shows the mean value $\langle D_{OO}\rangle$ determined for each of the structures within each of the GH families included in this survey: these are arranged in ascending order from left to right, and for

each GH family structures for free enzymes (lighter shading) and complexed (ligand-bound) enzymes (darker shading) are shown. For most GH families there is no significant difference between the mean values of D_{OO} for free or complexed structures. In some families (*e.g.* GH35) it appears that D_{OO} distances are longer for complexed structures than for free structures, but in others (*e.g.* GH6) it seems that the opposite is true. The magnitudes of the standard deviations (1σ, vertical black bars) of the averages $\langle D_{OO} \rangle$ for structures within each GH family depend, of course, upon the structural variation within each family and the number of structures selected according to the stipulated criteria. Generally there are larger uncertainties associated with inverting GHs with longer inter-carboxyl O···O distances.

Figure 4.
Mean values $\langle D_{OO} \rangle$ of the average inter-carboxyl O···O distances D_{OO} determined for structures within each GH family included in this survey. Free enzymes are indicated by lighter shading; ligand-bound (complexed) enzymes are indicated by darker shading. Vertical black bars denote the standard deviation (1σ) of the averages for structures within each GH family. The histogram is arranged from left to right in ascending order of magnitude of $\langle D_{OO} \rangle$.

Fig. 5 shows the mean value of D_{OO} determined for each of the structures within each of the GH families, with both free and ligand-bound (complexed) enzymes combined. The colour-coding makes it very clear that the separation between the catalytic carboxyl groups is usually smaller for retaining GHs than for inverting GHs and that there is no systematic difference between α- and β-retaining or α- and β-inverting GHs. The mean values $\langle D_{OO} \rangle$ for retaining GHs vary over a wide range from as little as 4.2 Å (GH101) to as much as 7.5 Å (GH89); likewise, the $\langle D_{OO} \rangle$ values for inverting GHs range between 5.0 Å (GH82) and 12.6 Å (GH46). Clearly, these ranges for retaining and inverting enzymes overlap each other and there are some significant outliers, notably GHs 28, 82 and 97.

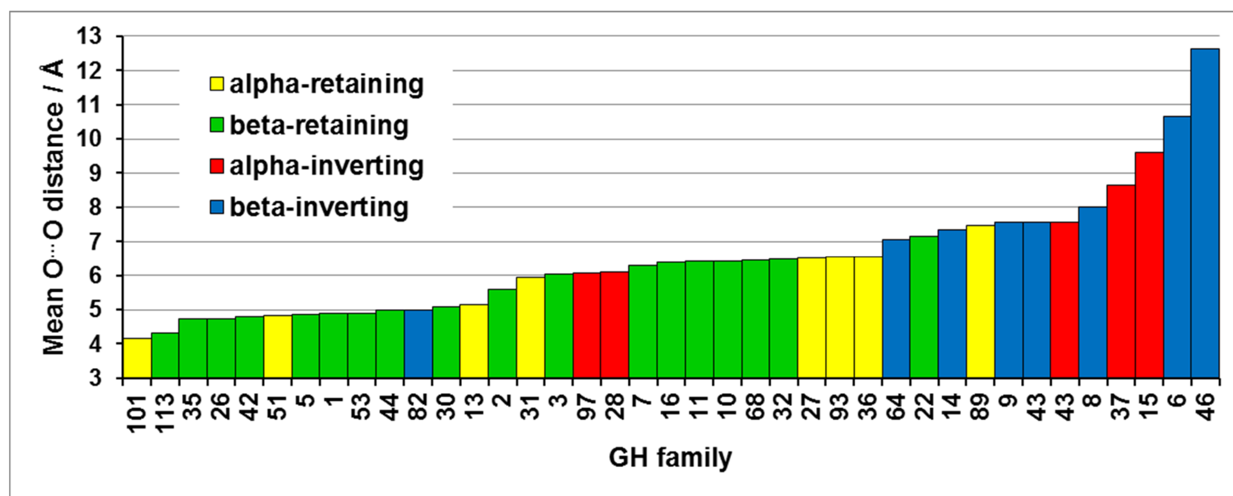

Figure 5.
Mean values $\langle D_{OO} \rangle$ of the average inter-carboxyl O···O distances D_{OO} determined for structures within each class of GH family: α-retaining (yellow), β-retaining (green), α-inverting (red) and β-inverting (blue). The histogram is arranged from left to right in ascending order of magnitude of $\langle D_{OO} \rangle$.

Fig. 6 shows the distributions of mean values $\langle D_{OO} \rangle$ for both retaining and inverting enzymes. The distribution for all inverting GHs is unimodal with a median very close to the overall mean value (8.0 Å) and also to the means of the α-inverting and β-inverting GHs treated separately. However, the distribution for retaining GHs is bimodal. There is no significant difference between the means of the α-retaining and β-retaining GHs treated separately, but the overall mean value (5.6 Å) falls between the two modes. Closer inspection of Fig. 4 suggests one cluster of retaining GH families (1, 5, 13, 26, 30, 35, 42, 44, 51 and 53) with values of $\langle D_{OO} \rangle$ close to 5 Å and another cluster (7, 10, 11, 16, 27, 32, 36, 68 and 93) with means close to 6.5 Å.

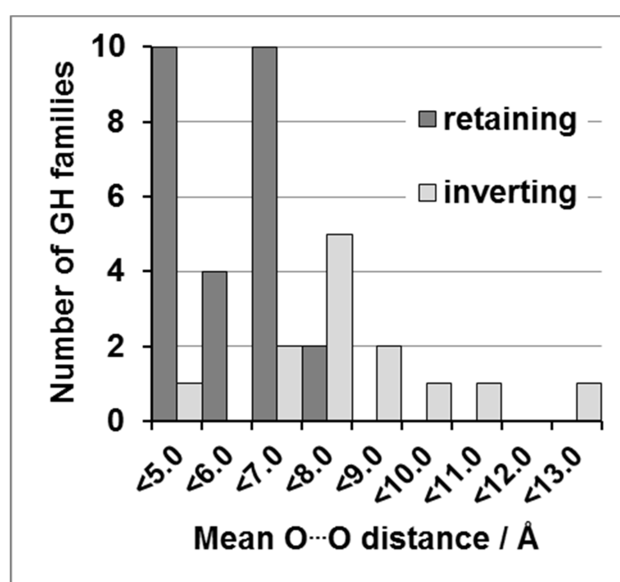

Figure 6.
Frequency distribution of mean values $\langle D_{OO} \rangle$ of O···O distances across GH families.

The value (5.6 Å, Table 1) of $\langle D_{OO} \rangle$ for the retaining NGP GHs in families 18, 20, 25, 56 and 84 is similar to that of the retaining GHs considered above, but the distribution of D_{OO} means has a larger standard deviation and a greater range between minimum and maximum values. (See Table S1 of the Supplementary Information for more details.)

DISCUSSION

It is instructive to compare the means and standard deviations of $\langle D_{OO} \rangle$ (Table II) considered in this survey with the averages proposed by Withers.[10,11] Our mean values for retaining GHs are higher than theirs (5.8 *vs.* 4.8 Å for α and 5.6 *vs.* 5.3 Å for β), but we have already noted the existence of a bimodal distribution in our survey. The retaining GHs may be split into two groups on the basis of the left-to-right ranking shown in Fig. 4:

- families 1, 5, 13, 26, 30, 35, 42, 44, 51, 53, 101 and 113 together have an overall mean value of $\langle D_{OO} \rangle$ equal to 4.8 ± 0.3 Å;
- families 2, 3, 7, 10, 11, 16, 22, 27, 31, 32, 36, 68, 89 and 93 together have an overall mean value of $\langle D_{OO} \rangle$ equal to 6.5 ± 0.5 Å.

These two groups each comprise both α- and β-GHs, so the similarity of our mean value for the first group with that of Withers and coworkers for α-GHs is merely coincidental. More noteworthy is that the standard deviation (1σ) for each group is much smaller than for all the retaining GHs taken together or grouped as α- and β-GHs. The question naturally arises as to what is responsible for the difference between these two groups of retaining GHs? One possibility is that there might be some relationship with the *syn-anti* proton donor concept[17] elaborated by Nerinckx *et al.*;[18] however, no correspondence is apparent between the GH families in the two groups and their *syn-* or *anti*-donor character.

However, there is another way by which to separate the retaining GHs into two groups: this is on the basis of their clans.[19]

- Clan A (or none): families 1, 2, 5, 10, 26, 30, 35, 42, 44, 51, 53, 101 and 113, which together have a mean value of $\langle D_{OO} \rangle$ equal to 5.0 ± 0.6 Å;
- Clans B, C, D, E, J (or none): 3, 7, 11, 13, 16, 22, 27, 31, 32, 36, 68, 89 and 93, which together have a mean value of $\langle D_{OO} \rangle$ equal to 6.4 ± 0.6 Å.

Although the standard deviations of the means for these groups are a little greater than for the ranking-based groups above, there is perhaps a rational basis for this separation: all of the GH families in the first group belong to clan A. The $(\beta/\alpha)_8$ fold of GHs in clan A is also shared by other clans, so it is not the protein fold itself that is the determining factor, but presumably some other structural characteristic that gives rise to generally shorter inter-carboxyl O···O distances D_{OO}. We

will not speculate upon this here, but we do note that the notion that clan A is associated with generally shorter D_{OO} distances is a testable hypothesis: structures for other members of this clan (GHs 17, 39 and 50) not included in this survey should also show shorter distances between their catalytic carboxyl groups.

One clear result from our survey is that the retaining GHs taken altogether tend to have significantly longer D_{OO} distances than was originally suggested by Withers[10,11] on the basis of a much smaller sample of structures.

Within the quite large single standard deviations for the D_{OO} distances found in our survey for retaining GHs, there is no significant difference between α- and β-inverting GHs. The overall mean we find for all inverting GHs taken together (8.0 Å) is markedly lower than either of the values (9.0 or 9.5 Å) suggested by Withers and co-workers for α- and β-GHs, respectively. This certainly arises out of the much larger numbers of GH families and of individual enzyme structures considered in the present study: from Fig. 4 it is evident that there is a much wider range of D_{OO} distances manifested in inverting GHs than has perhaps been commonly recognized.

It has been specifically noted for inverting GHs of family 6 that, while the identity of the catalytic general acid seems to be well established, the identity of the catalytic base is currently far less clear.[20] When there is no obvious carboxylate group to serve as a base within hydrogen-bonding distance of a water molecule that could act as the nucleophile in the inverting mechanism, it is possible that proton transfer might occur through a chain of water molecules.

However, it has been noted that nucleophilic attack by water and protonation of the leaving group oxygen by a general acid may occur from the same side of the glycosidic bond in α-linked carbohydrates, rather than from opposite sides of the active site.[21] Consequently, Benen and co-workers observed[22] that GH28 polygalacturonases diverge with respect to their active site configuration from the generally observed active site architecture found in inverting enzymes, by virtue of an unusually short distance between the catalytic acid group and the putative catalytic base. Similar exceptions to the general rule of longer D_{OO} distances between the catalytic acid and base groups for inverting GHs have also been noted for GH49 dextranases,[23] GH91 endorhamnosidases[24] as well as the GH82 ι-carrageenases[25] included in this survey.

Retaining GHs that employ NGP mechanisms have similar mean D_{OO} distances to those that use the double-displacement mechanism shown in Fig.1 but for a different reason: although one carboxylate residue does serve as a general acid/base, the other does not act as a nucleophile. A few of these NGP GHs from families 18, 20, 25, 56 and 84 have longer-than-usual D_{OO} distances but this should not be taken as evidence for an inverting mechanism by comparison to that shown in

Fig.2. Despite some early suggestions to the contrary, GHs from family 25 are now thought to adopt a retaining mechanism similar to that of other NGP GHs.[26]

The motivation for conducting this critical survey arose during the course of hybrid QM/MM molecular dynamics investigations of substrate conformation,[27] mechanism and reactivity[28] in wild-type and mutant GH11 β-retaining xylanases, for which the distance between the two catalytic residues was shown to play a crucial role. Withers and co-workers reported appropriate mutations and chemical modifications in order to shorten (Glu → Asp) or lengthen (Glu → carboxymethylated Cys) the side-chains of either the nucleophilic residue Glu78[29] of the endo-β-1, 4-xylanase from *Bacillus circulans* or the general acid residue Glu172:[30] from the concomitant changes in k_{cat}/K_m values for different substrates it was shown that the positional requirements for proton transfer from the general acid to the glycosidic oxygen were less demanding than those for nucleophilic attack at the anomeric centre. Furthermore, the Withers group showed that a Glu → Ala mutation of the nucleophilic residue in a β-glucosidase from *Agrobacterium faecalis* changed the reaction mechanism from retaining to inverting by virtue of shortening the side-chain and creating space to be occupied a different nucleophile (azide in this case rather than water).[11] In our opinion, elegant experimental studies such as these should be complemented by careful computational simulations which have the capability to shed light upon mechanistic details at the atomic level,[31] for example, by investigating the nature of thermal fluctuations in the positions of catalytic residues and the free-energy changes associated with conformational interconversions and reactive events. Thus, as part of a wider programme of investigation of unconventional catalytic mechanisms (*e.g.* ref. 32), it would in our view be of considerable interest to study the reaction mechanisms of inverting GHs with anomalously short separations between catalytic residues by means of appropriate QM/MM molecular dynamics simulations; detailed analysis of water structure within the active site could provide insight unobtainable by experiment. Furthermore, in order to obtain deeper insight into the underlying structural similarities and differences between groupings of GHs identifiable within the distribution of average inter-carboxyl separations between catalytic residues, it would be advisable to employ the methods of computational modeling to complement and to extend experimental tools. By these means it is possible to investigate the dynamics of enzyme structure in the presence of actual substrates rather than of inhibitors or other ligands which, in observed X-ray crystal structures, are often seen not to bind in the vicinity of the catalytic residues in orientations relevant for the reaction mechanism.

It is appropriate to comment upon some statistical aspects of the present survey. We have performed our analysis upon a finite and relatively small number of structures (even though this number is much larger than the ten considered originally by Withers and co-workers[10,11]); therefore care must be taken to treat the data appropriately. If the D_{OO} values for all the individual structures

(Table S1) were divided into bins according to increments of distance, and a histogram were plotted of the number of structures in each bin, then each of the four distributions (α- and β- retaining and inverting) would apparently be bimodal. However, these histograms would be biased by the varying numbers of structures included in this survey for each of the GH families. Certainly the "double-hump" that would be seen in each of these histograms for inverting enzymes would reflect more on the availability of structural data satisfying our criteria than it would on structural or mechanistic detail. It is better instead to obtain the arithmetic mean $\langle D_{OO} \rangle$ for each GH family, as reported in Table I, the resulting histogram (Fig. 5) is no longer bimodal for inverting GHs: the biasing effect of having relatively large numbers of structures for GHs 9 and 14 (with relatively shorter D_{OO} distances) and for GHs 6 and 15 (with relatively longer D_{OO} distances) but only a small number of structures for GHs 8 and 37 (with intermediate D_{OO} distances) is eliminated. However, as discussed above, the distribution for retaining GHs is still bimodal even after this biasing effect is eliminated: the presence of two groups, as proposed above, is not an obvious artefact.

We do not have a distribution of distances from an infinite population, which might be represented by a smooth and continuous mathematical function; therefore the histogram shown in Fig. 5 is necessarily a bar chart with discrete values. In a sense, because we are applying our selection criteria to the PDB files for all GHs containing a pair of carboxyl groups as the catalytic residues, we are sampling a larger population of GHs (many of which have crystal structures that do not satisfy our criteria, and many of which have not yet had their 3D structure determined). We would like to know the mean and variance of the distribution of distances for the whole population, but we are necessarily restricted to estimating these quantities from the properties of our limited sample. It is not even certain that the average distances considered here represent truly random variables over the whole population, so that the assumption of a normal distribution would be correct. It is all the more important, therefore, to ensure that we do not include any obvious biases within our sample of structures.

Finally, we note that for every value of the separation between the carbon atoms of the two carboxyl groups, there is an infinite number of relative orientations of the two groups. Each of these relative orientations could be described by six Euler angles, and for each one there are trigonometrical relationships between the four distances d_{11}, d_{12}, d_{21} and d_{22}. Clearly, all four distances would be equal if the two carboxyl groups were orientated such that the oxygen atoms were located at the vertices of an elongated tetrahedron (local D_{2d} symmetry, Fig.3a). At the other extreme, if the four oxygen atoms were aligned in a collinear manner (Fig. 3b), two of the distances ($d_{12} = d_{21}$) would be equal essentially to the inter-group C···C distance and the other two inter-group O···O distances would be larger and smaller by the value of the intra-group distance ($O_{\delta1} \cdots O_{\delta2}$ or $O_{\varepsilon1} \cdots O_{\varepsilon2}$) ~2.2 Å. Therefore the range between the lowest and the highest of the four distances d_{11},

d_{12}, d_{21} and d_{22} can vary between 0 and ~4.4 Å: these distances are not independent and it would be completely meaningless and inappropriate to determine and report standard deviations along with the mean values.

CONCLUSIONS

The present survey includes a total of 136 structures and 208 pairs of catalytic carboxyl groups (20 α-retaining, 87 β-retaining, 38 α-inverting and 63 β-inverting) satisfying the stated selection criteria and providing samples from 39 of the GH families. First, the average of all four inter-carboxyl O…O distances for each pair is determined; second, the mean of all D_{OO} of all four inter-carboxyl O···O distances for each pair is determined; second, the mean $\langle D_{OO} \rangle$ of all the pair-averages D_{OO} within each GH family is determined; third, means are determined for groups of GH families. No sensible differences are found for free structures as compared with those complexed with a ligand in the active site of the enzyme. Also, no significant difference is found for α-GHs as compared with β-GHs. The mean and standard deviation (1σ) of the unimodal distribution of $\langle D_{OO} \rangle$ distances for all families of inverting GHs is 8 ± 2 Å. The distribution of $\langle D_{OO} \rangle$ distances for all families of retaining GHs appears to be bimodal: the means and standard deviations of the two groups are 4.8 ± 0.3 Å and 6.4 ± 0.6 Å. We suggest that these average values are more representative, and possibly meaningful, than the often-quoted values originally proposed by Withers and co-workers.[10,11] The newly-updated average values proposed here may alter perceptions about what separations between catalytic residues are "normal" or "abnormal" for glycoside hydrolases.

ACKNOWLEDGEMENTS

IHW thanks the University of KwaZulu-Natal for hospitality during a research visit, and Professor Stephen Withers for helpful discussions.

REFERENCES

1. Henrissat B, Davies G. Structural and sequence-based classification of glycoside hydrolases. Curr Opinion Struct Biol 1997;7:637-644.
2. Zechel DL, Withers SG. Glycosidase mechanisms: anatomy of a finely tuned catalyst. Acc Chem Res 2000; 33:11-18.
3. Rye CS, Withers SG. Glycosidase mechanisms. Curr Opinion Chem Biol 2000;4:573–580.
4. Bourne Y, Henrissat B. Glycoside hydrolases and glycosyltransferases: families and functional modules. Curr Opinion Struct Biol 2001;11:593–600.
5. Vasella A, Davies GJ, Böhm M. Glycosidase mechanisms. Curr Opinion Chem Biol 2002;6:619–629.
6. Vocadlo DJ, Davies GJ. Mechanistic insights into glycosidase chemistry. Curr Opinion Chem Biol 2008;12:539–555.
7. Naumoff DG. Hierarchical classification of glycoside hydrolases. Biochemistry (Moscow) 2011;76:622-635.
8. Sinnott ML. Catalytic mechanisms of enzymatic glycosyl transfer. Chem Rev 1990;90:1171-1202.
9. Terwisscha van Scheltinga AC, Armand S, Kalk KH, Isogai A, Henrissat B, Dijkstra BW. Stereochemistry of chitin hydrolysis by a plant chitinase lysozyme and X-ray structure of a complex with allosamidin - evidence for substrate assisted catalysis. Biochemistry 1995;34:15619-15623.
10. McCarter JD, Withers SG. Mechanisms of enzymatic glycoside hydrolysis, Curr Opinion Struct Biol 1994;4:885-892.
11. Wang QP, Graham RW, Trimbur D, Warren RAJ, Withers SG. Changing enzymatic-reaction mechanisms by mutagenesis - conversion of a retaining glucosidase to an inverting enzyme. J Amer Chem Soc 1994;116:11594-11595.
12. Henrissat B. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 1991;280:309-316.
13. http://www.cazy.org/Glycoside-Hydrolases.html
14. http://www.cazypedia.org/index.php/Glycoside_Hydrolases
15. http://www.rcsb.org/pdb/home/home.do
16. Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. J Molec Graphics 1996;14:33-38. (http://www.ks.uiuc.edu/Research/)
17. Heightman TD, Vasella A. Recent insights into inhibition, structure and mechanism of configuration-retaining glycosidases. Angew Chem Intl Ed 1999;38:751-770.
18. Nerinckx W, Desmet T, Piens K, Claeyssens M. An elaboration on the syn-anti proton donor concept of glycoside hydrolases: electrostatic stabilization of the transition state as a general strategy. FEBS Letters 2005;579:302-312.
19. Henrissat B, Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. Biochem J 1996;316:695-6.
20. Piens K, Davies G. "Glycoside Hydrolase Family 8" in CAZypedia, available at URL http://www.cazypedia.org/, accessed 21 August 2013.

21. Shimizu, T Nakatsu T, Miyairi K, Okuno T, Kato H. Active-site architecture of endopolygalacturonase I from Stereum purpureum revealed by crystal structures in native and ligand-bound forms at atomic resolution. Biochemistry 2002;41:6651–6659.

22. Armand S, Wagemaker MJ, Sanchez-Tones P, Kester HC, van Santen Y, Dijkstra BW, Visser J, Benen JAE. The active site topology of Aspergillus niger endopolygalacturonase II as studied by site-directed mutagenesis. J Biol Chem 2000;275:691–696.

23. Larsson AM, Andersson R, Ståhlberg J, Kenne L, Jones TA. Dextranase from Penicillium minioluteum: reaction course, crystal structure, and product complex. Structure 2003;11:1111–1121.

24. Steinbacher S, Miller S, Baxa U, Budisa N, Weintraub A, Seckler R, Huber R. Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 Å, fully refined structure of the endorhamnosidase at 1.56 Å resolution, and the molecular basis of O-antigen recognition and cleavage. J Mol Biol 1997;267:865–880.

25. Michel G, Chantalat L, Fanchon E, Henrissat B, Kloareg B, Dideberg O. The iota-carrageenase of Alteromonas fortis - A beta-helix fold-containing enzyme for the degradation of a highly polyanionic polysaccharide. J Biol Chem 2001;276:40202-40209.

26. Martinez-Fleites C, Korczynska JE, Davies, GJ, Cope M, Turkenburg JP, Taylor EJ. Article The crystal structure of a family GH25 lysozyme from *Bacillus anthracis* implies a neighboring-group catalytic mechanism with retention of anomeric configuration. Carbohydr Res 2009;344:1753-1757.

27. Soliman MES, Ruggiero GD, Ruiz-Pernía, JJ, Greig IR, Williams IH. Computational mutagenesis reveals the role of active-site tyrosine in stabilising a boat conformation for the substrate: QM/MM molecular dynamics studies of wild-type and mutant xylanases. Org Biomol Chem 2009;7:460-468.

28. Soliman MES, Ruiz-Pernía, JJ, Greig IR, Williams IH. Mechanism of glycoside hydrolysis: A comparative QM/MM molecular dynamics analysis for wild type and Y69F mutant retaining xylanases. Org Biomol Chem 2009;7:5236-5244.

29. Lawson SL, Wakarchuk WW, Withers SG. Effects of both shortening and lengthening the active site nucleophile of Bacillus circulans xylanase on catalytic activity. Biochemistry 1996;35:10110-10118.

30. Lawson SL, Wakarchuk WW, Withers SG. Positioning the acid/base catalyst in a glycosidase: Studies with Bacillus circulans xylanase. Biochemistry 1997;36:2257-2265.

31. Ruiz-Pernía JJ, Tuñón I, Williams IH. Does glycosyl transfer involve an oxacarbenium intermediate? Computational simulation of the lifetime of the methoxymethyl cation in water. Pure Appl Chem 2011;83:1507-1514.

32. Chan J, Tang A, Bennet AJ. A stepwise solvent-promoted S_Ni reaction of α-D-glucopyranosyl fluoride: mechanistic implications for retaining glycosyltransferases. J Amer Chem Soc 2012;134:1212-1220.

Table I.

Mean value $\langle D_{OO} \rangle$ of average inter-carboxyl-group O···O distances D_{OO} (in Ångström), standard deviation (1σ) and number of structures N for each GH family.

	complexed enzymes				free enzymes		
GH family	$\langle D_{OO} \rangle$	1σ	N		$\langle D_{OO} \rangle$	1σ	N
α-retaining							
13	5.26	0.21	3		5.02	0.06	3
27	6.72	0.00	1		6.44	0.37	3
31	6.03	0.00	1		5.85	0.00	1
36	6.59	0.00	1		6.54	0.00	1
51					4.82	0.15	2
89					7.47	0.00	1
93	6.54	0.00	1		6.55	0.00	1
101					4.17	0.00	1
β-retaining							
1	4.83	0.14	13		5.02	0.03	4
2	5.74	0.04	4		5.46	0.04	4
3	6.04	0.00	1				
5	4.81	0.24	3		4.94	0.18	3
7	6.25	0.00	1		6.34	0.00	1
10	6.39	0.15	4		6.52	0.01	2
11	6.56	0.57	4		6.25	0.35	3
16	6.38	0.25	5		6.45	0.15	3
22	8.11	1.38	3		6.76	0.00	1
26	4.64	0.00	1		4.80	0.16	2
30	5.61	0.00	1		4.98	0.06	5
32	6.67	0.18	3		6.29	0.05	3
35	5.04	0.07	2		4.41	0.31	2
42					4.81	0.00	1
44	4.97	0.00	1		4.99	0.00	1
53					4.89	0.04	5
68					6.44	0.00	1
113					4.32	0.00	1
α-inverting							
15	9.39	0.16	8		9.75	0.32	7
28	6.51	0.00	1		6.02	0.74	5
37	8.64	0.02	2				
43	7.46	0.00	1		7.57	0.65	10
97	6.04	0.01	2		6.10	0.01	2
β-inverting							
6	10.07	0.30	5		11.22	0.83	5
8					8.00	0.02	4
9	7.33	0.52	7		7.74	1.16	9
14	7.17	0.39	17		7.80	0.23	6
43					7.56	0.18	4
46					12.62	0.00	1
64	7.08	0.00	1		7.03	0.00	1
82	5.21	0.00	1		4.87	0.14	2
NGP (free and complexed)							
18	5.37	1.44	50				
20	5.35	0.18	21	and	9.28	0.34	2
25	3.97	0.00	1	and	9.13	0.47	5
56	3.86	0.07	3				
84	5.53	0.21	8	and	8.58	0.00	1

Table II

Means and standard deviations (1σ) of $\langle D_{OO} \rangle$, minimum and maximum values across GH families.

	mean / Å	min / Å	max / Å	ref. 10
α-retaining	5.8 ± 0.9	4.2	7.5	4.8 ± 0.5
β-retaining	5.6 ± 0.8	4.2	7.7	5.3 ± 0.2
α-inverting	8.0 ± 1.5	6.0	10.2	9.0 ± 1.0
β-inverting	7.9 ± 1.6	4.8	12.6	9.5
all retaining	5.6 ± 0.8	4.2	7.7	
all inverting	8.0 ± 2.0	4.8	12.6	
NGP GHs[a]	5.6 ± 1.6	3.7	9.6	

a Neighbouring-group participation mechanism