
Article:
Roth, Christian orcid.org/0000-0001-5806-0987, Moroz, Olga V., Ariza, Antonio et al. (4 more authors) (2018) Structural insight into industrially relevant glucoamylases : flexible positions of starch-binding domains. Acta crystallographica. Section D, Structural biology. pp. 463-470. ISSN 2059-7983

https://doi.org/10.1107/S2059798318004989

---

manuscript rr5156 for review

Structural insight into industrially-relevant glucoamylases: flexible positions of starch-binding domains

Christian Roth, Olga V. Moroz, Antonio Ariza, Lars K. Skov, Keiichi Ayabe, Gideon J. Davies and Keith S. Wilson*

Category: *research papers*

Co-editor:

Professor R. Read

Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 0XY, UK

Telephone: 01223 336500

Fax: 01223 336827

Email: rjr27@cam.ac.uk

Contact author:

Keith Wilson

Department of Chemistry, University of York, Wentworth Way, York, YO10 5DD, United Kingdom

Telephone: 07904850375

Fax: 01904 328266

Email: keith.wilson@york.ac.uk

Structural insight into industrially-relevant glucoamylases: flexible positions of starch-binding domains

Authors

Christian Roth[ac#] , Olga V Moroz[a#] , Antonio Ariza[ab#], Lars K. Skov[b], Keiichi Ayabe[d], Gideon J Davies[a] and Keith S Wilson[a]*

[a] Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK

[b] Current address: Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, UK

[c] Novozymes A/S, Krogshøjvej 36, DK-2880 Bagsværd, Denmark

[d] Novozymes Japan Ltd., 261-8501 Chiba-shi, Japan

[e] Current address: Carbohydrates: Structure and Function, Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Berlin, Germany

[#]The first three authors contributed equally to this work.

* Correspondence email: keith.wilson@york.ac.uk

Synopsis Three industrially-relevant glucoamylase structures have been determined revealing how the starch-binding module can adopt different orientations relative to the catalytic domain.

Abstract Glucoamylases (GA's) are one of the most important classes of enzymes in the industrial degradation of starch biomass. They consist of a catalytic domain and a carbohydrate binding domain (CBM), with the latter being important for the interaction with polymeric substrate. Whereas the catalytic mechanism and the structure of the individual domains are well known, the spatial arrangement of the domains with each other and its influence on activity are not fully understood. We have crystallised and determined the structure of three industrially used fungal glucoamylases, two of which are full length. We show for the first time that the relative orientation between the CBM and the catalytic domain is flexible as they can adopt different orientations independently of ligand binding, suggesting a role as an anchor to increase the contact time and relative concentration of substrate near the active site. The flexibility in orientation of the two domains presented a considerable challenge for the crystallisation of the enzymes.

Keywords: Starch; glucoamylase; carbohydrate-binding module;

1. Introduction

Starch and glycogen are one of the major reserves of carbon and energy for all life. Furthermore, starch is one of the most important commodities for the food industry as well as for biofuel production (reviewed in (Torney *et al.*, 2007, Lovegrove *et al.*, 2017)). Due to the high stability of the glycosidic bond (Wolfenden *et al.*, 1998) and the internal structure of starch, current industrial processes to modify and break it down usually require harsh conditions. In nature there is a huge variety of specialised enzymes that modify and degrade glycogen and raw starch (reviewed in (Whelan, 1971)). In contrast, in industry the non-enzymatic modification and breakdown of raw starch mostly requires harsh conditions not desired in green environmentally friendly processes (Xu *et al.*, 2016). The use of enzymes proved to be a sustainable alternative to chemical processes and is now a multi-billion-dollar market (Chemier *et al.*, 2009). Among the most important classes of enzymes for the complete degradation of starch are glucoamylases, members of the glycoside hydrolase family 15 (GH15; reviewed in CAZypedia (Consortium, 2017) at https://www.cazypedia.org/index.php/Glycoside_Hydrolase_Family_15) of the CAZy database (http://www.cazy.org) (Lombard *et al.*, 2014), which catalyse the cleavage of the α-1,4- and α-1,6-glycosidic bonds. Glucoamylases use a classical acid/base Koshland type inverting mechanism, releasing β-glucose from the non-reducing end of an α-glucan chain (Koshland, 1953, Weil, 1954, Pazur & Ando, 1959;1960). Fungal enzymes in particular are widely used for the complete degradation of starch to glucose (reviewed in (Norouzian *et al.*, 2006)) and early on their heavy use sparked high interest in the structural and functional characterisation of microbial glucoamylases from a variety of sources (bacterial, fungal and eukaryotic). All glucoamylases possess a common catalytic domain, which can be followed or preceded by additional domains, usually carbohydrate binding modules (CBM's) (reviewed in (Boraston *et al.*, 2004, Marin-Navarro & Polaina, 2011)). The catalytic domain alone is able to degrade oligosaccharides, but for the interaction and degradation of raw starch, the carbohydrate binding domain proved to be essential (Stoffer *et al.*, 1993, Sauer *et al.*, 2000).

The first crystal structure of a glucoamylase was that of the catalytic domain of the *Aspergillus awamori* var. X100 glucoamylase (Aleshin *et al.*, 1992). The structure revealed a 13 α-helix fold with an $(\alpha/\alpha)_6$-helical bundle as core, with the active site in a deep funnel-like structure at the N-terminal side of the helical bundle. A subsequent structure of a complex with the well-known inhibitor acarbose allowed the identification of two conserved catalytic glutamic acid residues and important interactions in up to four well defined subsites (Aleshin *et al.*, 1994). The structure confirmed that glucoamylases employ a single displacement mechanism with inversion of the anomeric centre first deduced by Weil *et al.* (Weil, 1954). Insight into the interaction with raw starch was gained through various NMR and crystal structures of an isolated CBM alone and in complex with β-cyclodextrin or isomaltose (Sorimachi *et al.*, 1997, Chu *et al.*, 2014). The CBM forms a twisted β-sandwich domain

and at least two different starch binding sites have been identified in the *A. niger* CBM (Sorimachi *et al.*, 1997). The relative orientation of the catalytic domain and the CBM to one another and their interaction was studied with scanning tunnelling microscopy, suggesting an average distance of 90 Å and multiple different conformations (Kramer *et al.*, 1993, Sauer *et al.*, 2000). The first crystal structure of a full length glucoamylase was solved from *Hypocrea jecorina* in 2008 (Bott *et al.*, 2008), and showed the CBM in a single orientation consistent with feeding an amylose chain into the active site. However, a single structure cannot reflect the flexible orientation observed using scanning tunnelling microscopy and light scattering, thought to be important for the complete degradation of starch (Kramer *et al.*, 1993, Payre *et al.*, 1999).

Here, we have crystallised and determined the X-ray structures of three industrially-relevant glucoamylases. We describe the full-length structures for *Hormoconis resinae* GA (HrGA) and *Penicillium oxalicum* GA (PoGA) in their native states, as well as the catalytic domain of *Aspergillus niger* GA (AnGA). Two distinct relative conformations for the CBM were observed and reveal for the first time that, independent of ligand binding, multiple conformations are adopted. This advances our understanding of how glucoamylases interact with the substrate and subsequently degrade the starch polymer into its monomers.

2. Material and Methods

2.1. Cloning, expression, purification

The genes encoding the glucoamylases (UNIPROT accession numbers: AnGA P69328, HrGA Q03045 and PoGA S7ZIW0) were cloned, expressed in *A. niger* and the proteins purified as described for PoGA in Patent WO 2011127802 A1. Briefly, the culture broth from fermentation of *A. niger* MBin118 harbouring the glucoamylase gene was filtrated through a 0.22 μm PES filter, and applied on an α-cyclodextrin affinity gel column previously equilibrated in 50 mM NaOAc, 150 mM NaCI, pH 4.5 buffer. Unbound material was washed off the column with equilibration buffer and the glucoamylase was eluted using the same buffer containing 10 mM β-cyclodextrin over 3 column volumes. The glucoamylase activity of the eluent was checked to see if the glucoamylase had bound to the α-cyclodextrin affinity gel. The purified glucoamylase sample was then dialysed against 20 mM NaOAc, pH 5.0. The purity was finally checked by SDS-PAGE, and only a single band was observed. Purified proteins were provided from Novozymes to the University of York. PoGA was further treated with endo-H to minimize N-glycosylation.

2.2. Crystallisation

For all three proteins, initial crystallisation screening was carried out using sitting-drop vapour-diffusion with drops set up using a *Mosquito Crystal* liquid handling robot (TTP LabTech, UK) with 150 nl protein solution plus 150 nl reservoir solution in 96-well format plates (MRC 2-well

crystallisation microplate, Swissci, Switzerland) equilibrated against 54 µl reservoir solution. Experiments were carried out at room temperature with a number of commercial screens. Crystallisation of the intact proteins proved to be a challenge, presumably due to microheterogenity as a result of non-uniform glycosylation.

Penicillium oxalicum

Prior to crystallisation, the protein was concentrated to 48 mg/ml by ultrafiltration in an Amicon centrifugation filter unit (Millipore), aliquoted to 50 µl; aliquots that were not immediately set up for crystallisation were flash frozen in liquid nitrogen and stored at -80°C to use later in optimisations. The protein concentration was determined using a Coomassie (Bradford) assay (Bradford, 1976), with bovine serum albumin (BSA) as standard. The protein was diluted to several concentrations in the range 10-48 mg/ml for the initial crystallisation trials.

No hits appeared in the initial screens until an additional purification step by ion exchange was introduced. Anion exchange (pI is 6.0) was carried out in 20 mM Tris-HCl pH 7.5, with shallow gradient elution in 20 mM TrisHCl pH 7.5, 1 M NaCl. The asymmetrical peak started eluting at 50 mM NaCl, the shoulder was separated from the main peak and fractions corresponding to the main peak were pooled and concentrated to 10 mg/ml.

Initial hits were obtained in PACT premier™ HT-96 (Molecular Dimensions), with the best being very small clusters in condition E8 (0.2 M Na_2SO_4, 20% PEG3350). Seeding stock was prepared from condition E8 and MMS (microseed matrix screening, recent review in (D'Arcy *et al.*, 2014)) performed according to the published protocols (Shaw Stewart *et al.*, 2011). Briefly, crystals from the initial successful drop were transferred onto a glass slide, crushed, and collected in a Seed Bead™ (HR2-320, Hampton research) with 50 µl well solution added, vortexed for one minute, and used as an initial seeding stock: unused seeding stocks were stored at -20°C for later experiments. MMS was carried out with 150 nl protein solution plus 100 nl reservoir solution plus 50nl seeding stock, using a *Mosquito Crystal* liquid handling robot, in the PACT premier™ HT-96 screen. Following MMS, the best hit was obtained in condition C1 (PCB – sodium propionate + sodium cacodylate + Bis-Tris-propane buffer pH 4.0, 25% PEG1500). These again were clusters of inter-grown crystals, but bigger and better defined than the initial ones. Final optimisation was carried out using the Silver Bullets screen (HR2-096, Hampton Research), where different additives found successful for various proteins in the past are added to the same "hit" condition while seeding with the same seeding stock as before (PACT E8). This was successful and produced diffraction-quality crystals in condition G2 (0.2% thiodiglycolic acid, 0.2% adipic acid, 0.2% benzoic acid, 0.2% oxalic acid anhydrous, 0.2% terephthalic acid, 20 mM Hepes pH 6.8). To summarise, the condition that gave the final crystals was PACT C1 (100 nl) with added Silver Bullets G2 (50 nl) and seeding stock from PACT E8 (Table 1). The crystals were cryoprotected with PEG1500, added to the crystallisation condition to a final

concentration of 34%. Data were collected at the Diamond Light Source beamline I04-1 to 2.0Å resolution. The data were indexed and integrated with XDS (Kabsch, 2010) and subsequently scaled and merged with Aimless (Evans & Murshudov, 2013).

Aspergillus niger

A number of hits appeared in the initial screens, a single crystal suitable for diffraction being obtained in condition H8 of the Index screen (HR2-144, Hampton Research) - 0.1 M Hepes pH 7.5, 25% PEG 3350. Without further cryoprotection due to the 25% PEG content, the crystals were flash frozen in liquid nitrogen and data were collected at beamline ID14-1 of the ESRF to a resolution of 2.3 Å. However, the diffraction was of limited quality with streaking of the spots, some evidence of splitting and possible ice rings. In addition, after structure solution, the crystal, as described below, only contained the catalytic domain of the protein. X-ray data were processed with XDS (Kabsch, 2010), followed by AIMLESS (Evans & Murshudov, 2013) for scaling and merging.

Hormoconis resinae

The protein was co-crystallised with 7.5 mM acarbose. Large orthogonal crystals formed in condition C5 of the Index screen (60% tacsimate pH 7.0), the best of which diffracted to 6 Å at a home source (Rigaku MicroMax 007HF rotating anode), using 25% glycerol as cryoprotectant. Data were collected on beamline ID14-2 at the ESRF to a maximum resolution of 3.6 Å, processed with MOSFLM (Leslie, 2006) and scaled with AIMLESS (Evans & Murshudov, 2013).

Data processing statistics for all three enzymes are given in Table 2.

2.3. Structure solution and refinement

The structures were solved by molecular replacement using PHASER (McCoy *et al.*, 2007) with the catalytic domain of *A. awamori* GA as search model (PDB-ID: 1GLM). The solution of AnGA was problematical. The images were initially integrated in space group P1, while intensity statistics suggested the crystal was twinned. Merging the data and indeed refining the model was tried in space group P1 with four molecules per AU, in P 2_1 with two, and in P $2_12_12_1$ with a single molecular in the AU: for each of these the refined R and R_{free} were around 25 and 32% respectively. It was therefore decided to use the higher symmetry orthorhombic space group for the analysis, while accepting that the data and refinement statistics are poor for a structure at this resolution. Fortunately the AnGA structure is of least importance to the conclusions of this work, since it lacks the starch binding domain and another structure of the catalytic domain solved at higher resolution, is already available (Lee & Paetzel, 2011). For HrGA and PoGA additional unexplained density was attributed to the CBM, which was subsequently placed by an additional round of molecular replacement using the CBM from *H. jecorina* (PDB-ID:2VN4) as a model. The structures were subsequently rebuilt in real

space using Coot (Emsley *et al.*, 2010) followed by reciprocal space refinement with REFMAC (Murshudov *et al.*, 1997) within CCP4i2 (Potterton *et al.*, 2018). The respective glycosylation was added in Coot and refined in REFMAC using dictionaries created with PRIVATEER (Agirre *et al.*, 2015). The dictionaries allowed the use of additional monoperiodic torsion restraints to stabilise the conformation of the carbohydrates rings. All other linker parameters were taken from the CCP4 monomer library. The quality of the final models was evaluated using Molprobity (Chen *et al.*, 2010) as part of the PHENIX package (Adams *et al.*, 2011) and PRIVATEER. Figures of the structural models were prepared with CCP4mg (McNicholas *et al.*, 2011). Final refinement statistics are given in Table 3.

3. Results

3.1. Crystallisation and Structure determination.

Purified recombinantly-expressed glucoamylases from the three organisms were crystallised in either glycosylated form for AnGA and HrGA or de-glycosylated form in the case of PoGA. Data were collected for the partially deglycosylated PoGA to 2.0 Å, for AnGA to 2.3 Å, and for the glycosylated HrGA to 3.6 Å resolution. The structures were solved by molecular replacement using the catalytic domain of *A. awamori* glucoamylase. Initial refinement revealed additional density for the linker region in all three structures, as well as several N- and O-glycosylation sites. Surprisingly the CBM in *A. niger* could not be seen, and closer inspection of the crystal packing indicated that the CBM cannot be accommodated in the crystal lattice and was probably cleaved off during crystallisation. For the other two structures the full CBM could be built. The final model of AnGA includes one protein molecule in the asymmetric unit (AU) with residues built from 25 to 491, three N-acetyl glucosamine, two mannose residues, as part of the N- and O-glycosylation sites and was refined to a final R/R_{free} of 25.1/34.3 %. The PoGA model included one molecule in the AU comprising residues 30 to 616, three N-glycosylation sites, a Bis-Tris-propane (BTP), two PEG molecules and was refined to an R/R_{free} of 18.7/22.0 %. The HrGA model with two molecules in the AU included residues 29 to 616, up to seven N-glycosylation sites and had acarbose bound in the active site. The model was refined to an R/R_{free} of 23.8/25.4 %.

3.2. Overall structure

The catalytic domains of all three structures superimpose with an r.m.s.d. of 1.2Å over max. 452 residues, despite an overall low sequence identity of approximately 50 % (Fig. 1, Sup. Fig. 1). The domain follows the canonical fold with an $(\alpha/\alpha)_6$ barrel. In all three structures, the part of the linker domain interacting with the catalytic domain is structurally conserved.

The C-terminal CBM adopts the well-known β-sandwich motif, a hallmark of carbohydrate binding modules. A sequence comparison classifies both the PoCBM and the HrCBM into CAZy family CBM20 (Boraston *et al.*, 2004, Lombard *et al.*, 2014). In addition, the CBMs of HrGA and PoGA adopt two different relative conformations with respect to the catalytic domain with concomitant differences in the position of the C-terminal part of the linker (Fig. 1). Due to the different positions HrGA has an extended arrangement of both domains, whereas PoGA adopts an overall more compact structure.

3.3. Protein glycosylation

3.3.1. N-glycosylation

The catalytic domain of AnGA has two resolved N-glycosylation sites at N195 and N419, with two GlcNAc residues visible at N195 and one at N419.

The full-length structures of HrGA and PoGA show a greater degree of N-glycosylation, which is most extensive for HrGA, where the resolved sites are at N99, N200, N427, N500, N514, N528 and N587. N200 is the only site that shows branching of the glycosylation chain. In PoGA three glycosylation sites at N184, N410 and N514 are observed. The most extensive at N184 is structurally equivalent to N200 in HrGA and N195 in AnGA.

3.3.2. O-glycosylation

O-glycosylation was only observed in AnGA, in particular in the linker domain, which would connect to the CBM. Only two sites, at S483 and S484 could be modelled with confidence.

3.4. Inhibitor binding

HrGA was co-crystallised with the well-known inhibitor acarbose. Clear density corresponding to the inhibitor was found in the active site of both independent monomers. In both monomers an acarbose molecule was fitted and refined, assuming full occupancy. The resulting average B-value of 41.3 $Å^2$ is similar to the surrounding residues and supports the full occupancy. The inhibitor occupies the active site pocket with the cyclohexitol moiety of acarviosine populating the -1 subsite (subsite nomenclature in (Davies *et al.*, 1997) – briefly, subsites are labelled from -*n* to +*n*, with -*n* at the non-reducing end and +*n* the reducing end. Cleavage occurs between the -1 and +1 subsites.). There are further interactions with the sugars in subsite +1 (4-amino-4,6-Deoxy-glucose) and +2 (glucose), whereas the terminal sugar is not stabilised by direct interactions with the protein. The inhibitor interacts with HrGA by multiple hydrogen bonds and hydrophobic interactions with conserved residues in the active site (Fig. S2). The catalytic acid E208 forms a hydrogen bond with the bridging nitrogen in acarbose as expected for a productive complex with substrate having a glycosidic bond.

The catalytic base E432 is at a distance of 3.9 Å from the anomeric carbon, but is hydrogen bonded to Y76. The PoGA substrate binding site is occupied with a Bis-Tris-propane molecule (BTP), part of the crystallisation medium, which spans subsite -1 and +1 with multiple hydrogen bonds including to the catalytic acid. The BTP molecule refines with an average B of 38 $Å^2$ at full occupancy, which is similar to the surrounding residues. In AnGA, the active site is empty. There are no significant shifts for side chains lining the active site between the three structures indicating a rigid active site.

4. Discussion

The structures of the three glucoamylases show a high degree of conservation in the catalytic domain, as well as for the part of the linker which interacts with the catalytic domain. Whereas for AnGA the CBM-domain was missing, a full-length model could be built for HrGA and PoGA.

In the active site of HrGA one acarbose molecule was identified. The residues interacting with acarbose are identical to those described for *A. awamori* GA (Aleshin *et al.*, 1994) with the catalytic acid E208 interacting with the iminolinkage of acarbose and E432, the catalytic base, approximately 4 Å away. The arrangement and distance between the two catalytic residues support the proposed inverting mechanism for these enzymes (Weil, 1954, Pazur & Ando, 1959;1960, McCarter & Withers, 1994). In PoGA a BTP molecule was modelled into the active site forming key interactions with residues in subsite -1 and +1. A similar situation was observed for the structure of the catalytic domain of AnGA (Lee & Paetzel, 2011), where a Tris and a glycerol molecule were found in subsites -1 and +1 respectively. Tris and its derivatives are well known inhibitors of glycoside hydrolases (Roberts & Davies, 2012), which even led to the development of inhibitors sharing common characteristics (Taylor *et al.*, 2007).

All three structures reveal several glycosylation sites in the N-terminal domain, as well as in the CBM for PoGA and HrGA. The N-glycosylation sites are highly conserved, suggesting a functional role for example in protein secretion or stability. The N-glycosylation site in AnGA at N195 corresponds to N184 in PoGA and N200 in HrGA, which shows the highest degree of complexity of all sites in the latter two. Residual density suggests additional sugars in AnGA at this site, but the quality of the electron density did not allow a further extension of the glycosylation tree. This site is close to the linker connecting the CBM and might be of special importance for the linker stability. The importance of this site is further supported by the fact that the treatment of PoGA with endo-H still left this site intact. The N-glycosylation sites N528 and 587 in the CBM of HrGA are in close proximity to the proposed starch binding site 1 in AnGA CBM (Sorimachi *et al.*, 1997), suggesting that only binding site 2 is involved in the interaction with the substrate in HrGA.

Interestingly, O-glycosylation was only found in AnGA, specifically in the linker region. Though more sites are known (Lee & Paetzel, 2011), only two (S483 and S484) could be modelled with confidence. Another model, refined to higher resolution, showed up to seven glycosylation sites with an eighth suggested, but not modelled (Lee & Paetzel, 2011). The role of this specific linker glycosylation is not known, but a role in protein stability was proposed (Goto *et al.*, 1997). Additionally it was shown that the extent of the O-glycosylation influences the susceptibility of AnGA to proteolytic degradation, which decreases with higher glycosylation (Le Gal-Coeffet *et al.*, 1995).

The potential variability in the glycosylation pattern might also explain the positive effect of ion exchange purification fine tuning on crystallisation, which is in agreement with the suggestion that the protein naturally occurs in multiple species, due to possible glycosylation microheterogenity and/or conformational variability resulting in slight variations of the overall surface charge. Selecting a species with homogenous surface charge by separating a narrow peak fragment facilitated forming crystal contacts favourable for crystallisation.

Of particular interest in this study is the CBM domain (CAZy family CBM20) and its role in the degradation of raw starch. For both HrGA and PoGA the CBM linked to the catalytic domain could be clearly modelled, with both domains forming a β-sandwich, a well-known structural motif for CBM's. Glycoside hydrolase structures with intact appended CBMs are extremely rare, rarer still if connected by a flexible linker. It is therefore difficult to know if these rare observations reflect a unique orientation, or one favoured by a certain crystal packing environment. Therefore, the most important observation is that the CBM domains in the two structures, reported here, adopt very different relative orientations with respect to the catalytic domain (Fig. 2) and to the orientation observed in the crystal structure for *H. jecorina* GA solved independently (Bott *et al.*, 2008). Whereas HrGA adopts a more extended structure with the two domains separated, PoGA adopts a rather compact arrangement. A flexible arrangement with different relative conformations for the CBM was previously proposed based on single molecule scanning tunnelling microscopy and light scattering data collected on AnGA (Kramer *et al.*, 1993, Payre *et al.*, 1999), which suggested that the average distance between the two domains is about 90 Å (Kramer *et al.*, 1993), corresponding to a rather extended arrangement. However, it was proposed that for the proper function a closer distance is necessary (Payre *et al.*, 1999). Indeed dynamic light scattering experiments pointed to a hydrodynamic radius between 60 to 75 Å upon ligand binding (Payre *et al.*, 1999). The structure of *H. jecorina* GA was solved independently in two different crystal forms, with both domains in close proximity (Bott *et al.*, 2008), and these authors concluded that the compact conformation is the dominant active one. Further support for this model comes from the fact that this would bring the second CBM substrate binding site close to the active site. A model with amylose as substrate was developed, which takes advantage of the close proximity of the second substrate binding site in CBM

and the active site in the catalytic module (Bott *et al.*, 2008). Such a compact conformation is also supported by heterobifunctional inhibitors that can bridge the catalytic site and the CBM binding site (Payre *et al.*, 1999, Sauer *et al.*, 2000). Furthermore, multiple salt bridges and hydrophobic interactions have been proposed to stabilise this conformation. Nevertheless, in HrGa compared to HjGA the CBM is in a conformation where the proposed starch binding site is remote from the active site, despite co-crystallisation with acarbose. Comparing the compact conformation of PoGA with HjGA shows that the CBM in PoGA is in a more elevated position relative to the catalytic domain (Fig. 2), and in addition is stabilised by multiple hydrogen bonds and hydrophobic interactions as well. Furthermore, no GA specific inhibitor was present during the crystallisation. Though a BisTris-propane buffer molecule was identified in the active site, which might be a weak inhibitor for GA's (Roberts & Davies, 2012). Taken together all the results gathered on different GA's point to the conclusion that the precise orientation of the CBM with respect to the catalytic domain is neither important, nor stabilised by specific interactions between the domains, or favoured by binding of inhibitors in the active site and therefore does not influence the activity. Furthermore, GA variants with different linker lengths showed virtually no difference in activity provided the linker has a minimal length of 17 residues to prevent steric clashes between the two domains (Sauer *et al.*, 2001). Additionally, the issue still engenders controversy due to an alternative model, involving an extended conformation and subsequent dimerization, based on the results of small angle X-ray scattering (SAXS) of AnGA (Jorgensen *et al.*, 2008). Indeed, analysis of the potential stable oligomers in the crystal revealed a dimer for HrGA with the two CBM's part of the interface (Fig. 3) showing some resemblance to the proposed SAXS model.

In summary, we solved the crystal structures of three industrially relevant glucoamylases, with two in the intact two domain form. We show for the first time that the carbohydrate binding module can adopt multiple relative orientations with respect to the catalytic module, independent of a bound ligand in the active site. The results are in agreement with single molecule data and kinetic analysis of linker variants. Nevertheless, further research is needed to clarify the mode of action with respect to synergy of binding and oligomerisation on raw starch. Taken together the data strongly indicate that many relative orientations are accessible in solution and contribute to the enhanced activity towards polymeric substrate by increasing the relative local substrate concentration, the probability of contact with the substrate and may allow the catalytic domain to reach multiple structurally weak points in starch without dissociation from the substrate.

Figure 1 Stereo ribbon diagram of the tertiary structure of GA. A) Side view of AnGA (red), HrGA (yellow) and PoGA (blue) with the relative domain orientation for the CBM of HrGA and PoGA. The active site is indicated with acarbose in sphere representation, observed in HrGA. The corresponding glycosylation sites are shown as glycoblocks (McNicholas & Agirre, 2017), coloured according to the residue type. The superposition is based on secondary structure superposition of the catalytic domains. B) Top view along the $(\alpha/\alpha)_6$-barrel.

Figure 2 Stereo representation of the relative orientation of the CBM with respect to the catalytic domain from the side (a) and from the front (b). HrGA is coloured in yellow, PoGA in blue and HjGA (PDB-ID 2vn7) in green. The active site is indicated with acarbose in sphere representation.The glycosylation sites are shown as glycoblocks (McNicholas & Agirre, 2017) and coloured according to the residue type.

Figure 3 Potential dimer of HrGA determined from the crystal structure using PISA (Krissinel, 2015). The bound inhibitor acarbose is shown in sphere representation, whereas the glycosylation is shown as glycoblocks (McNicholas & Agirre, 2017).

Table 1 Crystallisation

	PoGA	AnGA	HrGA
Method	Vapour diffusion, sitting drop, MMS	Vapour diffusion, sitting drop	Vapour diffusion, sitting drop
Plate type	MRC 2-well crystallization microplate, Swissci, Switzerland		
Temperature (K)	293		
Protein concentration	10 mg/ml	18 mg/ml	37 mg/ml
Buffer composition of protein solution	20 mM Tris-HCl pH 7.5, 50mM NaCl	25 mM piperazine pH 5.0, 150mM NaCl	20 mM Na-acetate pH 5.0
Seeding stock composition	PACT E8: 0.2 M Na_2SO_4, 20%	N/A	N/A

Composition of reservoir solution	PEG3350 PACT condition C1: PCB – sodium propionate + sodium cacodylate + Bis-Tris-propane buffer pH 4.0, 25% PEG1500; + 1/3 Silver bullet condition G2: 0.2% thiodiglycolic acid, 0.2% adipic acid, 0.2% benzoic acid, 0.2% oxalic acid anhydrous, 0.2% terephthalic acid, 20 mM Hepes pH 6.8	Index H8: 0.1 M Hepes pH 7.5, 25% PEG 3350	Index C5 (60% tacsimate pH 7.0)
Volume and ratio of drop	200nl protein + 150nl reservoir (containing 100nl screen condition + 50nl Silver bullets)+50nl seeding stock	150nl protein + 150nl reservoir	150nl protein + 150nl reservoir
Volume of reservoir		54 µl	

Table 2 Data collection and processing

Values for the outer shell are given in parentheses.

	PoGA	AnGA	HrGA
Diffraction source	Diamond I04-1	ESRF ID14-1	ESRF ID14-2
Wavelength (Å)	0.9173	0.9334	0.933
Temperature	100	100	100

(K)			
Detector	Pilatus 2M	ADSC Q210 CCD	ADSC Q210 CCD
Crystal-detector distance (mm)	222.7	227.2	307.5
Rotation range per image (°)	0.2	0.5	1.0
Total rotation range (°)	180	180	121
Exposure time per image (s)	0.2	3	10
Space group	H32	$P2_12_12_1$	$P2_12_12_1$
a, *b*, *c* (Å)	189.3,189.3,115.4	56.4,73.1,102.9	138.0,149.8,192.4
α, β, γ (°)	90,90,120	90, 90, 90	90,90,90
Mosaicity (°)	0.24	1.13	1.53
Resolution range (Å)	47.31-2.00	44.64-2.3	59.98-3.60
Total No. of reflections	548252(35057)	138362(13756)	248567(25376)
No. of unique reflections	53301(3921)	19454(1870)	38033(3798)
Completeness (%)	100(100)	99.5(99.9)	81.8(68.0)
Redundancy	10.3	7.1(7.4)	6.5(5.5)
$\langle I/\sigma(I)\rangle$	6.2(1.0) #	6.3(1.3)	10.5(2.5)
$R_{r.i.m.}$	0.225(2.284)	0.071(0.556)	0.142(0.675)
CC(1/2)	0.994(0.394)	0.995(0.816)	0.994(0.889)
Overall *B* factor from	28.3	36.16	45.0

Wilson plot ($Å^2$)	

Resolution limit judged based on CC(1/2), which is 0.394 for PoGA and 0.816 for AnGA in highest resolution shell. The $\langle I/\sigma(I)\rangle$ drops below 2.0 at a resolution of 2.18 Å for PoGA and 2.58 Å for AnGA.

† Redundancy-independent merging *R* factor $R_{r.i.m}$ (Diederichs & Karplus, 1997)

Table 3 Structure solution and refinement

	PoGA	AnGA	HrGA
PDB code	6FHV	6FRV	6FHW
Resolution range (Å)	47.31-2.00	44.64-2.30	59.58-3.60
Completeness (%)	100	99.6	81.8
No. of reflections, working set	53296	19520	37085
No. of reflections, test set	2614	974	1887
Final R_{cryst}	18.7	25.1	23.6
Final R_{free}	22.0	34.2	25.4
Cruickshank DPI	0.1525	0.5677	
No. of non-H atoms			
Protein	4576	3542	8959
Ligand	120	64	429
Water	257	27	
Total	5403	3633	9388

R.m.s. deviations			
Bonds (Å)	0.0146	0.0089	0.0101
Angles (°)	1.632	1.267	1.684
Average *B* factors (Å^2)			
Protein	24.1	58.0	55
Ligand	44.7	60.0	83.6
Water	37.0	40.7	
Ramachandran plot			
Most favoured (%)	96.4	91.3	91.2
Allowed (%)	3.6	7.6	8.2

Acknowledgements

We thank ESRF for the access to beamlines ID14-1 and ID14-2 and Diamond Light Source for access to beamline I041 (proposal number mx-1221) that contributed to the results presented here. The authors also thank Dr. Johan Turkenburg and Sam Hart for assistance during data collection.

Supporting information

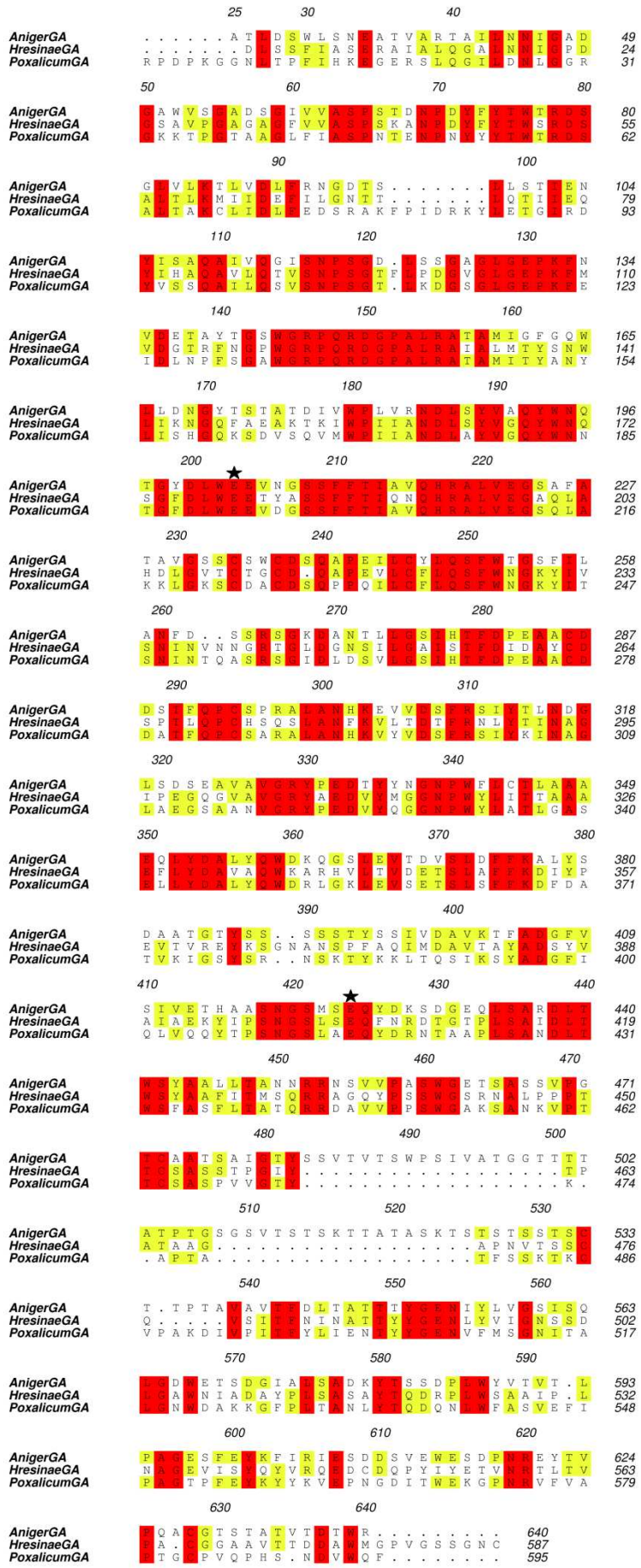
AnigerGA
HresinaeGA
PoxalicumGA

Figure S1 Sequence alignment of AnGA, HrGA and PoGA obtained using MUSCLE (Edgar, 2004) and visualized using ALINE (Bond & Schuttelkopf, 2009). Amino acids identical for all three proteins are outlined in red, for two – in yellow. The catalytic acid and base are marked with a star.

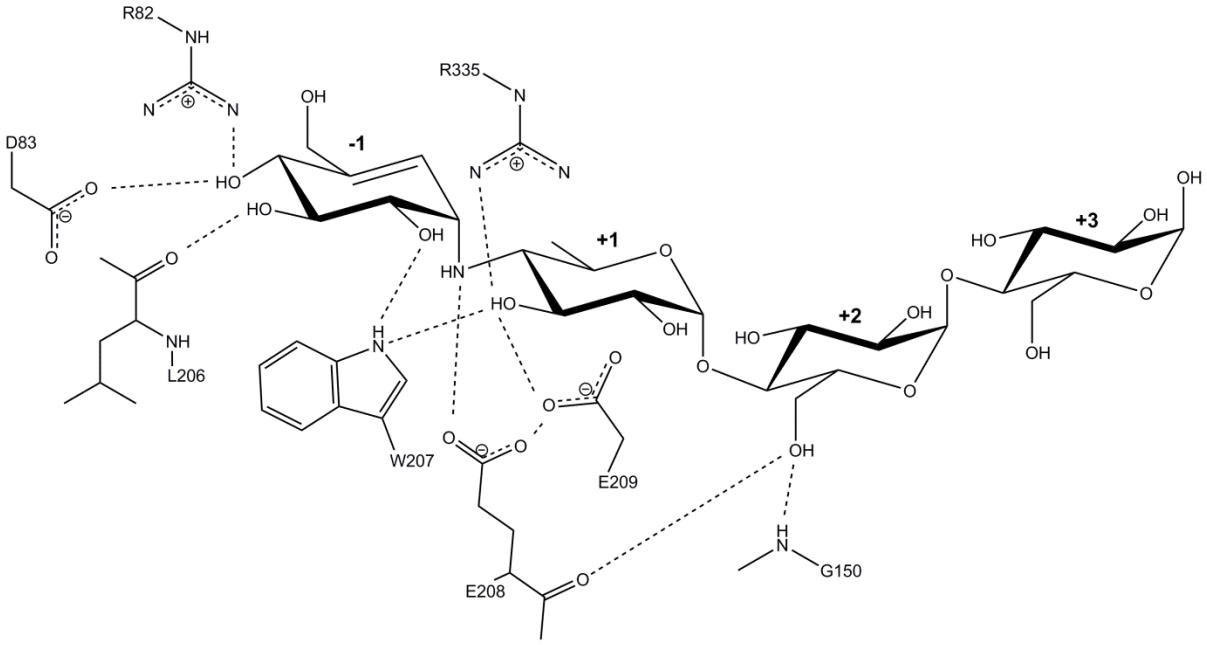

Figure S2 Schematic representation of the interaction of acarbose in the active site of HrGA. Hydrogen bonds are shown as dotted lines and the monomers of acarbose are numbered according their position in the subsites of the GA binding site Figure was prepared using ChemDraw (Perkin Elmer Informatics Inc.).

References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L. W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2011). *Methods* **55**, 94-106.

Agirre, J., Iglesias-Fernandez, J., Rovira, C., Davies, G. J., Wilson, K. S. & Cowtan, K. D. (2015). *Nat Struct Mol Biol* **22**, 833-834.

Aleshin, A., Golubev, A., Firsov, L. M. & Honzatko, R. B. (1992). *J Biol Chem* **267**, 19291-19298.

Aleshin, A. E., Firsov, L. M. & Honzatko, R. B. (1994). *J Biol Chem* **269**, 15631-15639.

Bond, C. S. & Schuttelkopf, A. W. (2009). *Acta Crystallogr D Biol Crystallogr* **65**, 510-512.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. (2004). *Biochemical Journal* **382**, 769-781.

Bott, R., Saldajeno, M., Cuevas, W., Ward, D., Scheffers, M., Aehle, W., Karkehabadi, S., Sandgren, M. & Hansson, H. (2008). *Biochemistry* **47**, 5746-5754.

Bradford, M. M. (1976). *Anal Biochem* **72**, 248-254.

Chemier, J. A., Fowler, Z. L., Koffas, M. A. & Leonard, E. (2009). *Adv Enzymol Relat Areas Mol Biol* **76**, 151-217.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 12-21.

Chu, C. H., Li, K. M., Lin, S. W., Chang, M. D., Jiang, T. Y. & Sun, Y. J. (2014). *Proteins* **82**, 1079-1085.

Consortium, C. (2017). *Glycobiology* **28**, 3-8.

D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W. & Marsh, M. (2014). *Acta Crystallogr F Struct Biol Commun* **70**, 1117-1126.

Davies, G. J., Wilson, K. S. & Henrissat, B. (1997). *Biochemical Journal* **321**, 557-559.

Diederichs, K. & Karplus, P. A. (1997). *Nat Struct Biol* **4**, 269-275.

Edgar, R. C. (2004). *Nucleic Acids Res* **32**, 1792-1797.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 486-501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214.

Goto, M., Ekino, K. & Furukawa, K. (1997). *Appl Environ Microbiol* **63**, 2940-2943.

Jorgensen, A. D., Nohr, J., Kastrup, J. S., Gajhede, M., Sigurskjold, B. W., Sauer, J., Svergun, D. I., Svensson, B. & Vestergaard, B. (2008). *J Biol Chem* **283**, 14772-14780.

Kabsch, W. (2010). *Acta Crystallogr D Biol Crystallogr* **66**, 125-132.
Koshland, D. E. (1953). *Biol Rev* **28**, 416-436.
Kramer, G. F. H., Gunning, A. P., Morris, V. J., Belshaw, N. J. & Williamson, G. (1993). *J Chem Soc Faraday T* **89**, 2595-2602.
Krissinel, E. (2015). *Nucleic Acids Res* **43**, W314-319.
Le Gal-Coeffet, M. F., Jacks, A. J., Sorimachi, K., Williamson, M. P., Williamson, G. & Archer, D. B. (1995). *Eur J Biochem* **233**, 561-567.
Lee, J. & Paetzel, M. (2011). *Acta Crystallogr Sect F Struct Biol Cryst Commun* **67**, 188-192.
Leslie, A. G. (2006). *Acta Crystallogr D Biol Crystallogr* **62**, 48-57.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res* **42**, D490-495.
Lovegrove, A., Edwards, C. H., De Noni, I., Patel, H., El, S. N., Grassby, T., Zielke, C., Ulmius, M., Nilsson, L., Butterworth, P. J., Ellis, P. R. & Shewry, P. R. (2017). *Crit Rev Food Sci Nutr* **57**, 237-253.
Marin-Navarro, J. & Polaina, J. (2011). *Appl Microbiol Biotechnol* **89**, 1267-1273.
McCarter, J. D. & Withers, S. G. (1994). *Curr Opin Struct Biol* **4**, 885-892.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J Appl Crystallogr* **40**, 658-674.
McNicholas, S. & Agirre, J. (2017). *Acta Crystallogr D Struct Biol* **73**, 187-194.
McNicholas, S., Potterton, E., Wilson, K. S. & Noble, M. E. (2011). *Acta Crystallogr D Biol Crystallogr* **67**, 386-394.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Crystallogr D Biol Crystallogr* **53**, 240-255.
Norouzian, D., Akbarzadeh, A., Scharer, J. M. & Moo Young, M. (2006). *Biotechnol Adv* **24**, 80-85.
Payre, N., Cottaz, S., Boisset, C., Borsali, R., Svensson, B., Henrissat, B. & Driguez, H. (1999). *Angew Chem Int Edit* **38**, 974-977.
Pazur, J. H. & Ando, T. (1959). *J Biol Chem* **234**, 1966-1970.
Pazur, J. H. & Ando, T. (1960). *J Biol Chem* **235**, 297-302.
Potterton, L., Agirre, J., Ballard, C., Cowtan, K., Dodson, E., Evans, P. R., Jenkins, H. T., Keegan, R., Krissinel, E., Stevenson, K., Lebedev, A., McNicholas, S. J., Nicholls, R. A., Noble, M., Pannu, N. S., Roth, C., Sheldrick, G., Skubak, P., Turkenburg, J., Uski, V., von Delft, F., Waterman, D., Wilson, K., Winn, M. & Wojdyr, M. (2018). *Acta Crystallogr D Struct Biol* **74**, 68-84.
Roberts, S. M. & Davies, G. J. (2012). *Methods Enzymol* **510**, 141-168.
Sauer, J., Christensen, T., Frandsen, T. P., Mirgorodskaya, E., McGuire, K. A., Driguez, H., Roepstorff, P., Sigurskjold, B. W. & Svensson, B. (2001). *Biochemistry* **40**, 9336-9346.
Sauer, J., Sigurskjold, B. W., Christensen, U., Frandsen, T. P., Mirgorodskaya, E., Harrison, M., Roepstorff, P. & Svensson, B. (2000). *Biochim Biophys Acta* **1543**, 275-293.
Shaw Stewart, P. D., Kolek, S. A., Briggs, A. R., Chayen, N. E. & Baldock, P. F. M. (2011). *Crystal Growth & Design* **11**, 3432-3441.
Sorimachi, K., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B. & Williamson, M. P. (1997). *Structure* **5**, 647-661.
Stoffer, B., Frandsen, T. P., Busk, P. K., Schneider, P., Svendsen, I. & Svensson, B. (1993). *Biochem J* **292 (Pt 1)**, 197-202.
Taylor, E. A., Clinch, K., Kelly, P. M., Li, L., Evans, G. B., Tyler, P. C. & Schramm, V. L. (2007). *J Am Chem Soc* **129**, 6984-+.
Torney, F., Moeller, L., Scarpa, A. & Wang, K. (2007). *Curr Opin Biotechnol* **18**, 193-199.
Weil, C. E. B., R.J.; Van Dyk, J.W. (1954). *Cereal Chemistry* **31**, 510-518.
Whelan, W. J. (1971). *Biochem J* **122**, 609-622.
Wolfenden, R., Lu, X. D. & Young, G. (1998). *J Am Chem Soc* **120**, 6814-6815.
Xu, Q. S., Yan, Y. S. & Feng, J. X. (2016). *Biotechnol Biofuels* **9**, 216.

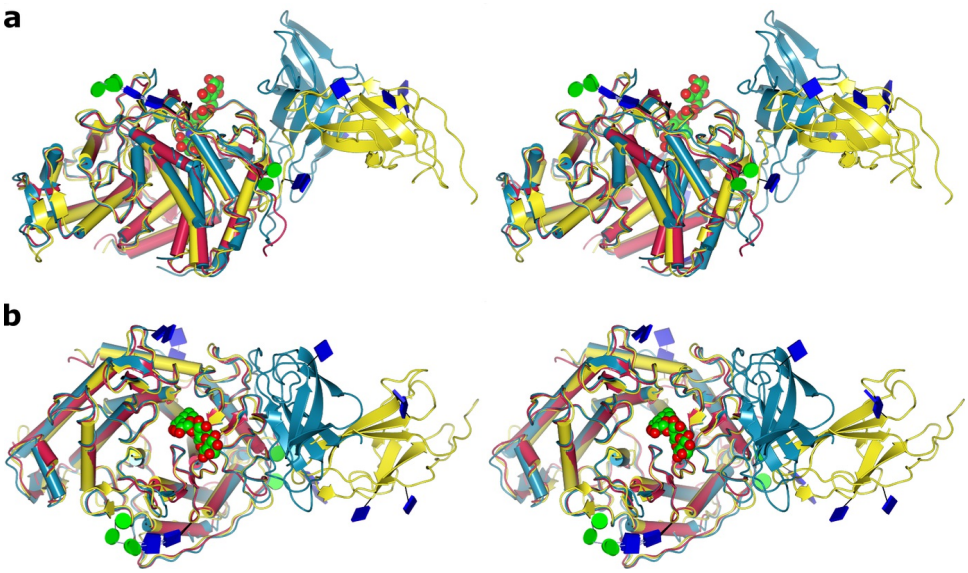

Figure 1

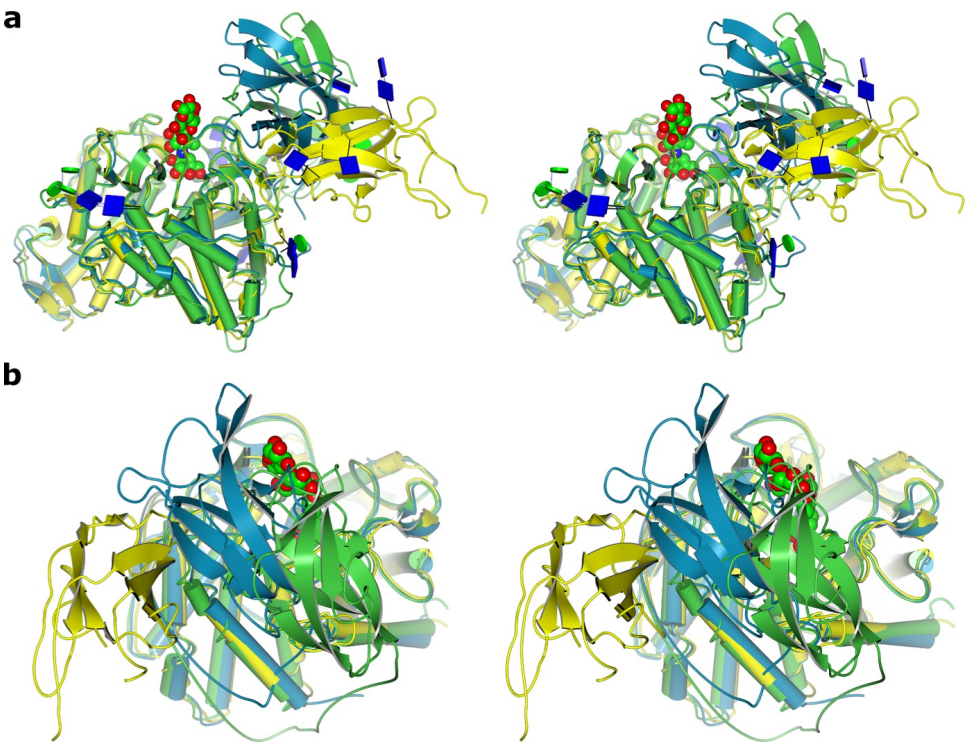

Figure 2

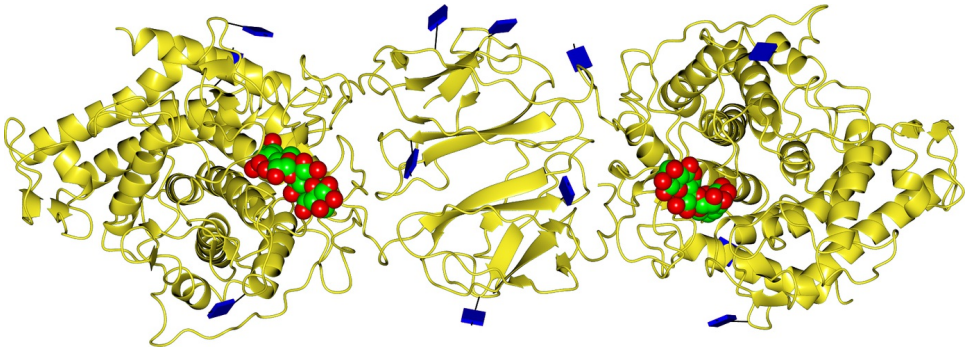

Figure 3