



Accepted Article

Title: Release of soybean isoflavones using a β -Glucosidase from *Alicyclobacillus herbarius*.

Authors: Lidia Delgado, Christian M. Heckmann, Flavio Di Pisa, Louise Gourlay, and Francesca Paradisi

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *ChemBioChem* 10.1002/cbic.202000688

Link to VoR: <https://doi.org/10.1002/cbic.202000688>

Release of soybean isoflavones using a β -Glucosidase from *Alicyclobacillus herbarius*.

Lidia Delgado^{a,#}, Christian M. Heckmann^{a,#}, Flavio Di Pisa^b, Louise Gourlay^b,

Francesca Paradisi^{a,c,*}

^aUniversity of Nottingham, School of Chemistry, Department of Chemical Biology, University Park, Nottingham NG7 2RD, United Kingdom.

^bDipartimento di Bioscienze, Università di Milano, Via Celoria 26, 20133, Milan, Italy.

^cUniversity of Bern, Department of Chemistry and Biochemistry, Freiestrasse 3, CH-3012 Bern Switzerland

These authors share equal contribution to the manuscript

*Corresponding author, E-mail: Francesca.paradisi@dcb.unibe.ch (F. Paradisi); Telephone number +41(0)31 631 4002

ABSTRACT

β -Glucosidases are used in the food industry to hydrolyse glycosidic bonds in complex sugars, with enzymes sourced from extremophiles better able to tolerate the process conditions. In this work, a novel β -glycosidase from the acidophilic organism *Alicyclobacillus herbarius* was cloned and heterologously expressed in *E. coli* BL21 (DE3). *AheGH1* was stable over a broad range of pH values (5-11) and temperatures (4 °C-55 °C). The enzyme exhibited excellent tolerance to fructose and good tolerance to glucose, retaining 65% activity in the presence of 10% (w/v) glucose. It also tolerated organic solvents, some of which appeared to have a stimulating effect, in particular ethanol with a 1.7-fold increase in activity at 10% (v/v). The enzyme was then applied for the cleavage of isoflavone from isoflavone glucosides in an ethanolic extract of soy flour, to produce soy isoflavones which constitute a valuable food supplement, achieving full conversion within 15 min at 30 °C.

Accepted Manuscript

Introduction

β -glucosidases (EC 3.2.1.21) constitute a group of enzymes that hydrolyse terminal, non-reducing glycosyl residues from glycosides.^[1] These enzymes have been successfully applied in a broad range of industrial applications,^[2] and they have become key tools to hydrolyse very stable glycosidic bonds in a clean and efficient way.

However, many food industrial processes involve harsh conditions (high concentrations of solvents and sugars, low pH and high temperatures) that can lead to enzyme inactivation.^[3,4] Extremophiles are organisms well adapted to extreme environmental conditions^[5] and they constitute a novel and alternative source of enzymes (extremozymes) for industrial applications. Extremozymes are generally more resistant in demanding industrial processes when compared to mesophilic enzymes.^[6]

Soybean (*Glycine max*) originating in China, constitutes one of the largest sources of vegetable oil in the world and has the highest protein content among all others food crops.^[7] Its consumption has become increasingly popular in recent years as it is an excellent protein source for the human diet. In addition, soybean contains several compounds considered important food supplements due to their health properties, especially isoflavones. Soybeans contain mainly three types of isoflavones (daidzein, genistein, and glycitein), which can be found in four different forms (Fig. 1): as aglycons, 7-*O*- β -D-glucosides, 7-*O*-(6''-*O*-acetyl)glucosides, or 7-*O*-(6''-*O*-malonyl) glucosides.^[8] Recently, commercial preparations of isoflavones have come to the public attention following studies on their reported positive effects on cognitive function.^[9] However, when the biological activities of these compounds are considered, the bioavailability of the aglycone has been suggested to be higher than that of the glycoside; but it represents only a minor constituent of unfermented soy products.^[10]

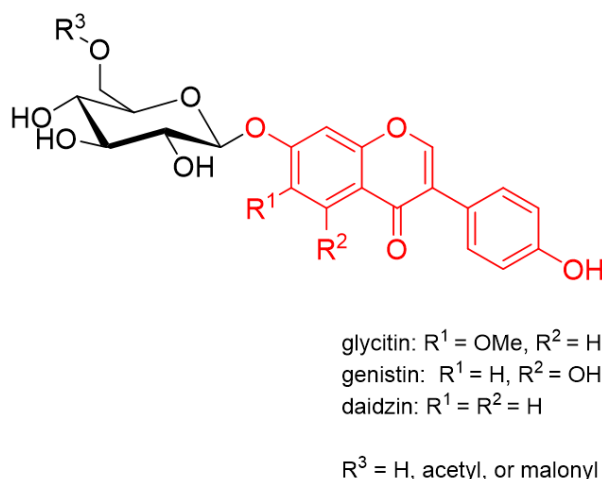


Figure 1: Structure of the main isoflavone glucosides (glycitin, genistin, and daidzin) found in soybean. The isoflavone moiety (glycitein, genistein, and daidzein) is highlighted in red. The sugar may be further acetylated or malonylated.

β -glucosidases can be used to hydrolyse isoflavone glucosides to their aglycons. The industrial processing for extracting isoflavones from soybeans includes the use of organic solvents (mostly ethanol) which is necessary to solubilise the isoflavones.^[11] Soy isoflavones are released as a side-product during the industrial fermentation process for the production of erythromycin, using the microorganism *Saccharopolyspora erythraea* (by endogenous β -glucosidases). However, the isoflavone aglycones themselves are metabolized by that organism, requiring either metabolic engineering or subsequent enzymatic treatment (or acid hydrolysis) of the spent broth to obtain the aglycon.^[12–14] The use of purified β -glucosidases for isoflavone aglycon production from soy flour has also been investigated, in particular from the thermophilic bacteria *Thermotoga maritima* and *Thermoanaerobacter ethanolicus JW200*. However, these enzymes required defatting of the soy flour using hexane (an undesirable neurotoxic solvent^[15]) and high reaction temperatures (65–80 °C).^[16,17] The use of a GH1 β -glucosidase from *Alicyclobacillus sp. A4* as a supplement in soy-based animal feed to aid in the release of the aglycon in monogastric animals has also been investigated under milder conditions (37 °C), but in the absence of the co-solvents required for extracting the aglycons.^[18] Again, many β -glucosidases reported to date from mesophilic organisms are inhibited

by both organic solvents.^[19] and glucose,^[20,21] which would be present at increasing concentrations as the hydrolytic process progresses.

In this work, a novel β -glucosidase (WP_026963033.1) of the glycosyl hydrolase family 1 (GH1) from the extremophilic organism *Alicyclobacillus herbarius* (*Ahe*) has been identified and investigated. *Ahe* was first isolated from a herbal tea made from dried flowers of hibiscus, it has been described as a thermo-acidophilic Gram-positive bacterium that grows at a range of temperatures between 35-65° C and a pH between 3.5-6, features which are very appealing for different applications in the food industry market.^[22] The enzyme has been cloned and expressed in *E. coli*, its crystal structure has been solved and the enzyme has been characterized to assess its performance under different operational conditions (glucose, fructose, organic co-solvents and range of pHs and temperatures) usually found in food industrial processes. Following the initial characterization, *AheGH1* has been applied to hydrolyse the main isoflavone glucosides present in soybean flour.

Results and discussion

Protein expression and purification

AheGH1 was expressed and purified with excellent yields between 75-100 mg/L of culture. The purification, by metal affinity chromatography, was analysed by SDS-PAGE (Fig. S1). *AheGH1* has a theoretical monomeric molecular weight of 52133.66 Da (\approx 52 kDa) estimated by the online tool ProtParam^[23] and sufficient degree of purification was confirmed by SDS-PAGE. The molecular weight was determined to be 99.7 KDa by gel filtration, consistent with it being a dimer in solution (Fig. S2 & Table S1).

The 3D structure of AheGH1

Crystals of *AheGH1* enzyme belonging to the monoclinic space group P21 diffracted at 1.98 Å resolution. The asymmetric unit of *AheGH1* contains four independent molecules, with a calculated Matthews coefficient of 2.42 Å³ Da⁻¹ (estimated solvent content of 49.31%) (Fig. 2). The overall

structure of the four *AheGH1* subunits is identical (RMSD values of 0.05 to 0.08 Å over 356 backbone Ca atoms).

The final model was refined to R values of R_{work} 24.7% and R_{free} 29.8%. The presence of strong translational noncrystallographic symmetry (tNCS), as confirmed by a peak in the Patterson map, may justify the higher than expected final R_{work} value, despite the high resolution of the structure (Table S2). Electron density map was overall of good quality, with electron density coverage across residues 2-450 (chain A), 2-451 (chain B), 3-447, (chain C) and 3-446 (chain D), except for a short stretch (P304-D322). Several ethylene glycol molecules derived from the crystallization buffer were modelled into the electron density. In addition, two Nickel cations were identified in the model, present during the affinity chromatography purification step. The metal cations are located at the dimer interfaces between chain A and the symmetry-related monomer C and between the B and D subunits and are coordinated in a tetrahedral arrangement by the side chains contributed from two histidine residues (H61) and two glutamate (E29) residues.

AheGHI is arranged into a single $(\beta/\alpha)_8$ barrel fold, common to this family of glycosidases, with negligible main-chain displacements in peripheral loops and α -helices. Regions which show more significant changes in the secondary structure correspond to residues 272 to 281, folded into a gamma and two β -turns, and residues 409 to 416 that form two β -turns instead of the more common α -helices (Fig. 2).

A feature of the GH1 family is that despite low sequence identity (as low as 17%) they share high structure conservation. 3D structure-based comparisons performed with the DALI server (<http://ekhidna2.biocenter.helsinki.fi/dali/>)^[24] revealed that the *AheGH1* protein has the highest structural similarity with a family 1 glucoside hydrolase from *Paenibacillus polymyxa* (*BglB*; PDB: 2O9P).^[25] Despite average sequence identity (52%), superposition between the two proteins revealed a high degree of structural similarity (RMSD of 0.6 Å) over 441 aligned residues.

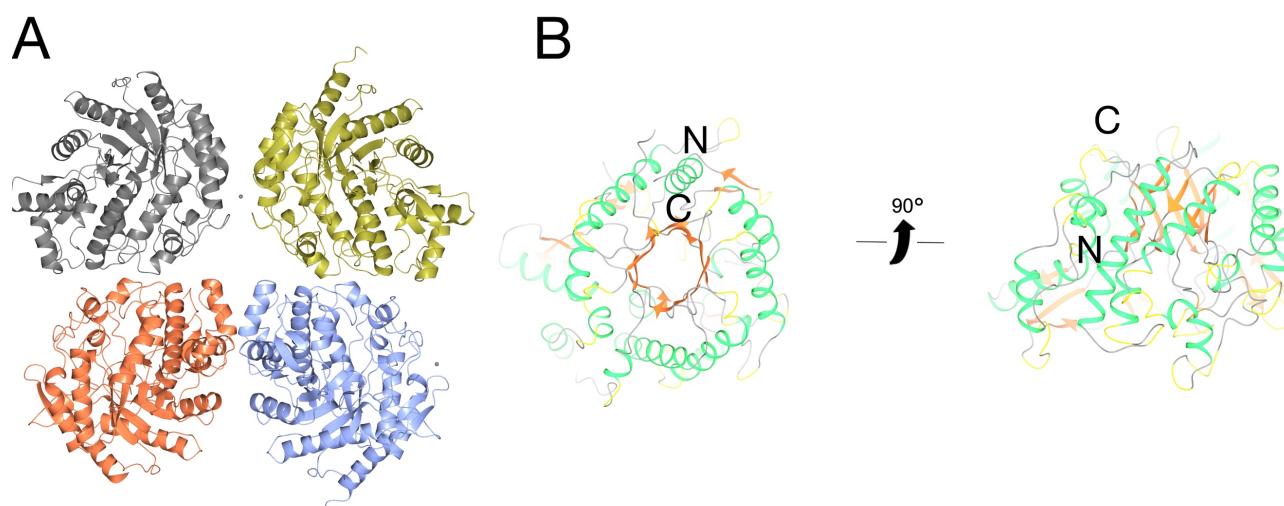


Figure 2. 3D structure of *AheGH1*. **A)** The *AheGH1* tetramer present in the asymmetric unit; **B)** Top and side views of the *AheGH1* monomer, with ribbons coloured according to secondary structure (β -strands in orange, α -helices in green, 3-, 4- and 5- turns in yellow, unstructured regions in grey). N- and C-terminal regions are labelled.

1.1 The *AheGH1* active site

An enzyme template search of catalytic site templates made with the ProFunc server (<http://www.ebi.ac.uk/thornton-srv/databases/ProFunc>),^[26] identified the cyanogenic family 1 glycosyl hydrolase (CBG) from white clover (PDB:1CBG)^[27] as the top hit. *AheGH1* and CBG share low (36.94%) overall sequence identity, yet high (73.47%) local sequence identity in their active sites over 49 equivalenced residues. At a structural level, they share high structural similarity (99.5% over 493 matched residues) and as for *AheGH1*, CBG is a homodimer in solution.^[27] In agreement with CBG and other GH1 members in general, the active site pocket contains several conserved polar and aromatic residues that are typically present in carbohydrate recognition sites that binds the nonreducing end of the substrate.^[21] Based on comparisons with GH1 members, *AheGH1* the main active site residues present are: R79, H122, E167, N166, N296, Y298, E356 and W402 (Fig. 3, & Fig. S4).

The mechanism of catalysis generally described for this class of enzyme involves a double displacement reaction, requiring a proton donor and a nucleophile.^[27] Previous results are consistent with an ionized carboxylic acid group acting as a catalytic nucleophile and a histidine residue or carboxylic acid group (with a significantly elevated pK_a) behaving as a general acid catalyst. Based

on comparisons made with GH1 members in general, E167 and E356 are likely to be the acid-base catalyst and nucleophile, respectively in the reaction (Fig. 3 & Fig. S4).

With regards to the active site architecture of GH1 enzymes in general, structural and thus functional differences are attributed to the loop regions that are present between the β/α motifs that shape the active site cavity. This may be appreciated by comparison of the overall sequence and structure conservation of *AheGH1* with all other enzymes of similar structure, using the ENDscript 2 server (<http://endscript.ibcp.fr/ESPript/ENDscript/>).^[28] As expected, structure and sequence divergence occurs mainly in the loop regions connecting the β/α motifs (Fig. S3).

In *BlgB*, nine active site residues are reported to delineate the entrance to the active site: Y169, T178, E180, R243, E225, Q316, H318, W328 and W412 (*BlgB* numeration)^[25]. These residues render the active site cavity narrower in comparison with its homolog *BglA* (PDB: 1E4Y).^[29] Except for W328, all remaining residues in *AheGH1* are not conserved (Fig. S4). Higher similarity, in terms of sequence identity, was on the other hand detected between *AheGH1* and the *BglA* active site, defined by residues W168, L177, V179, S224, T242, E314, N316, W326 and E408. Three out of the nine residues (i.e. W168, V179 and W326; Fig. S4) are conserved between the two proteins. In addition, according to the nature and steric hindrance of the side-chains of the residues defining the active site, the *AheGH1* cavity is more similar to that of *BglA*.

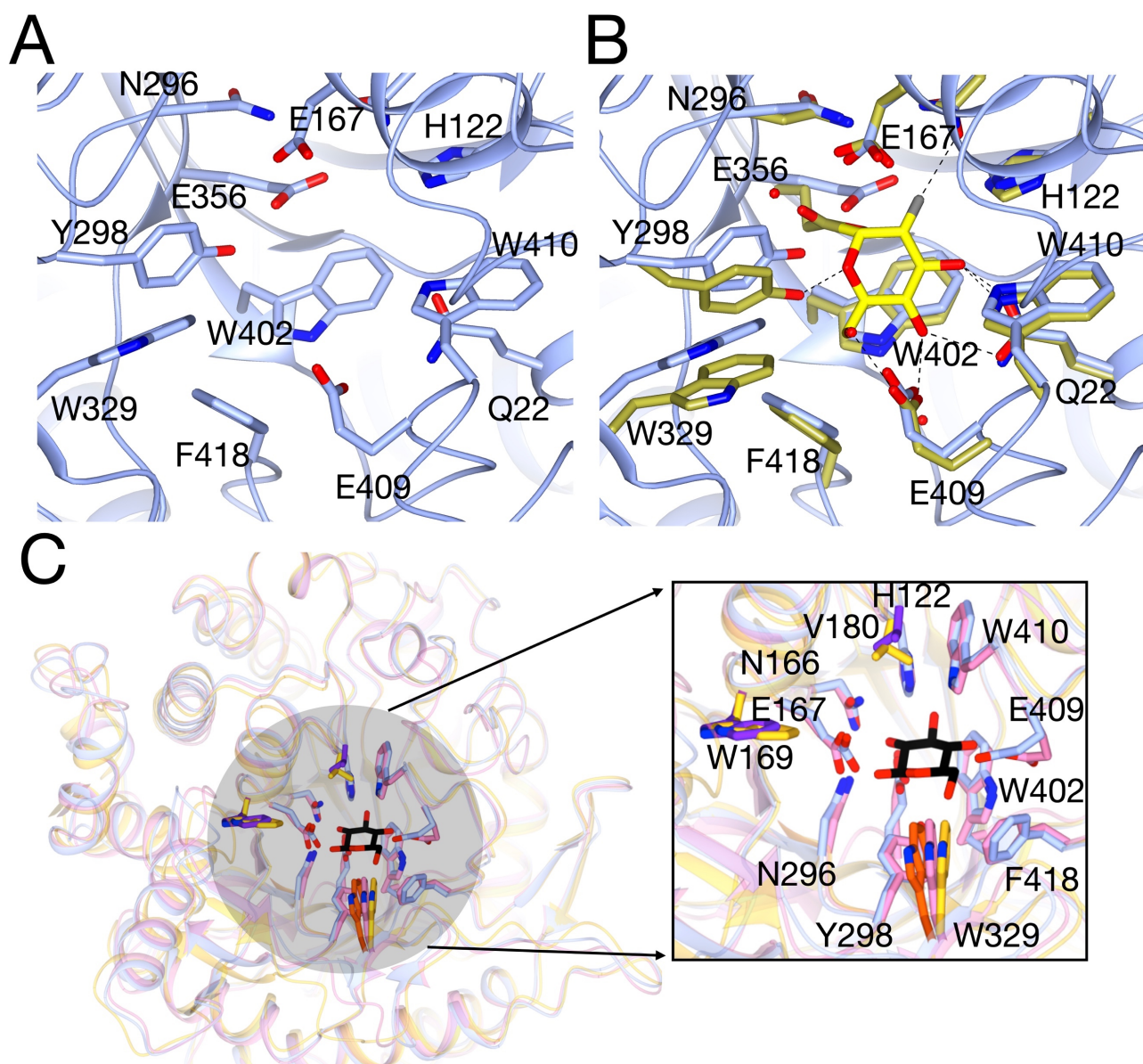


Figure 3. The *AheGH1* active site. **A)** Stereo-view of the *AheGH1* putative active site showing active site residues delineating the pocket (sticks). Proposed catalytic residues E356 and E167 are indicated; **B)** structural superposition of the active sites of *AheGH1* (blue) and the homologous protein *BglA* (gold, PDB: 1E4I) complexed with 2-deoxy-2-fluoro- α -D-glucopyranose catalytic intermediate (yellow sticks) and **C)** Superposition of *AheGH1* (ice blue) with *BglA* (yellow) and *BglB* (pink) in complex with glucose (PDB: 2O9T) and a detailed view of the glucose binding site. Glucose is shown as black sticks. *BglB* residues interacting with the glucose molecule, and corresponding *AheGH1* residues, are depicted as sticks and colored accordingly. *AheGH1* residues conserved in *BglA* are colored in purple. *AheGH1* W329, which is conserved in all the three homologues is colored in orange. Residue numbering in all panels is for *AheGH1*. This figure was made with CCP4mg.^[30]

Activity assay and kinetic parameters

The activity assay was performed spectrophotometrically using *p*-nitrophenyl- β -D-glucopyranoside (*p*NPG) at 25 °C. 10 μ L of the suitable dilution of the enzyme was added to a 96-well plate in triplicate together with 0.29 mL of the reaction solution *p*NPG (10 mM), HEPES (50 mM), pH 7.4. The *p*-nitrophenol formation was followed for 10 minutes at 420 nm. The enzyme showed a specific activity

of 20 U/mg. *AheGH1* also showed activity with *p*-nitrophenyl- β -D-fucopyranoside, -galactopyranoside, and -xylopyranoside. While the activity with the fucoside was unchanged, it was reduced by approx. 30% for the galactoside and 98% for the xyloside (Fig. 4A).

The kinetic parameters of *AheGH1* for pNPG were determined under the same conditions of the activity assay but using different substrate concentrations.^[31] K_m (1.4 ± 0.1 mM), K_i (37 ± 6 mM) and k_{cat} (2.7 ± 0.1 s⁻¹) were calculated in triplicate by nonlinear regression using GraphPad Prism 8 (Fig. 4B; see method section for details).

Enzymatic characterization

A thorough characterization of this novel enzyme was performed, to probe the activity and stability in the presence of sugars, solvents, different pH values and different temperatures.

Activity assays

The effect of glucose, fructose and a different range of co-solvents is reported in Fig. 4.

Glucose inhibition (Fig. 4C) is a common problem among β -glucosidases,^[21,32] as the accumulation of the product of the hydrolytic process naturally reduces the catalytic efficiency of the enzyme. However, it has been reported that some β -glucosidases belonging to the family GH1 can be exceptionally tolerant or even stimulated by glucose, but the mechanism of that tolerance/stimulation is still not clear.^[3,18,33] The structure of *BglB* in complex with glucose (PDB: 2O9T)^[25] has been solved; the residues involved in ligand binding (Q22, H122, N166, E157, H181, E365, W410, E409, W412:*BglB* numbering) are conserved in *AheGH1* (Fig. 3C). A mechanism of product inhibition, whereby the glucose molecule remains tightly bound to the active site until it is displaced by the arrival of a new substrate molecule, has been suggested for a plant β -D-glucan glucohydrolase GH3 enzyme.^[34] Further studies to assess the behaviour of *AheGH1* in the presence of glucose were carried out. A broad range of glucose concentrations, from 0.1 to 25% (*w/v*), were evaluated, showing that the *AheGH1* activity slowly decreases as the concentration of glucose increases. With 25% (*w/v*) of glucose in the reaction, *AheGH1* retains 30% of its activity. In contrast, the β -glucosidase from

Aspergillus niger, a glycosidase commonly used in food industrial processes, only retains 2 % activity in the presence of 10 % (w/v) glucose in the mixture.^[35] Fructose (Fig. 4C) has a milder effect on the enzyme activity than glucose. The retained activity varies between 80%, with 5% (w/v) of fructose, and 67%, with 25% of fructose in the reaction mixture. It is unclear whether the effect of fructose is due to inhibition or caused by the 2-3-fold increase in viscosity as concentration increases.^[36]

Commonly, many enzymes from mesophilic organisms are significantly destabilised by organic solvents. This could be attributed to the loss of crucial water molecules that maintain the protein conformation (desolvation), affecting the K_m and v_{max} values, and, in the most dramatic cases the overall protein folding. Retained activity in the presence of organic solvents is possible only when the protein surface and the active site remain well hydrated. *AheGH1* presents an impressive co-solvent tolerance, retaining over 50% activity in all cases (Fig. 4D). Particularly relevant for the intended application of this enzyme, is its performance in the presence of ethanol which is one of the most common solvents in food processing; with 10 and 20% of ethanol in the reaction mix, *AheGH1* is 50% more active than without ethanol at all. This behaviour has been previously reported for other glucosidases in literature.^[37]

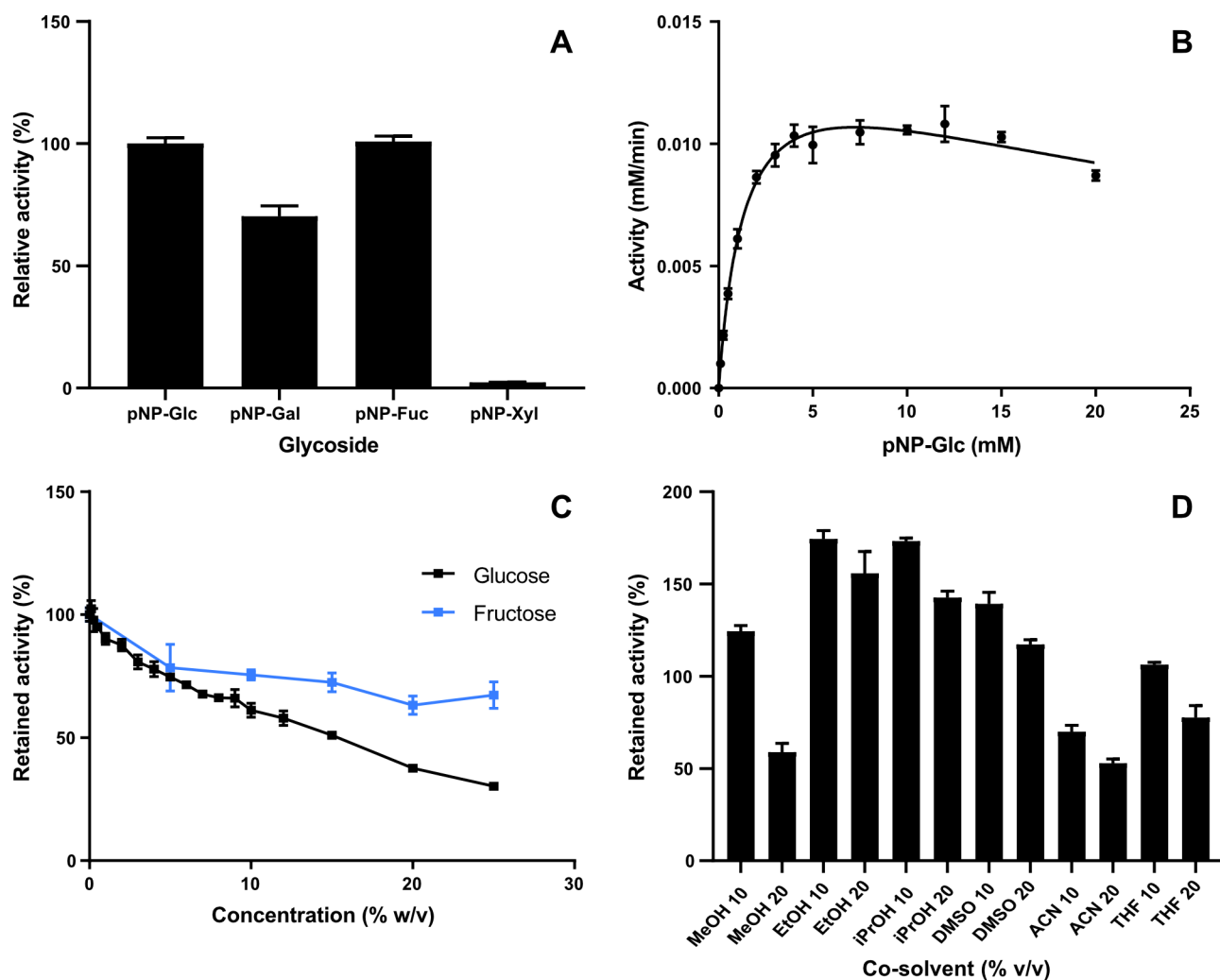


Figure 4. (A) Specific activity with various pNP-glucosides. (B) Substrate inhibition kinetics with pNPG. (C) Effect of Glucose and Fructose on activity and (D) Effect of co-solvents on activity. Represented in the figures are the mean and the standard deviation. The activity was tested using the standard activity assay at 25 °C. Error bars represent standard deviations (n = 3).

Stability assays

Enzyme long term stability is also a key parameter to assess the potential implementation of the catalyst in biotechnological processes. Results for the *AheGH1* stability assay in the presence of co-solvents and a different range of temperatures and pH values have been evaluated.

When the enzyme is incubated with co-solvents (Fig. 5), a very stable behaviour over significant period of time is observed. *AheGH1* retains over 55 % activity after 48 h incubation in all 12 conditions.

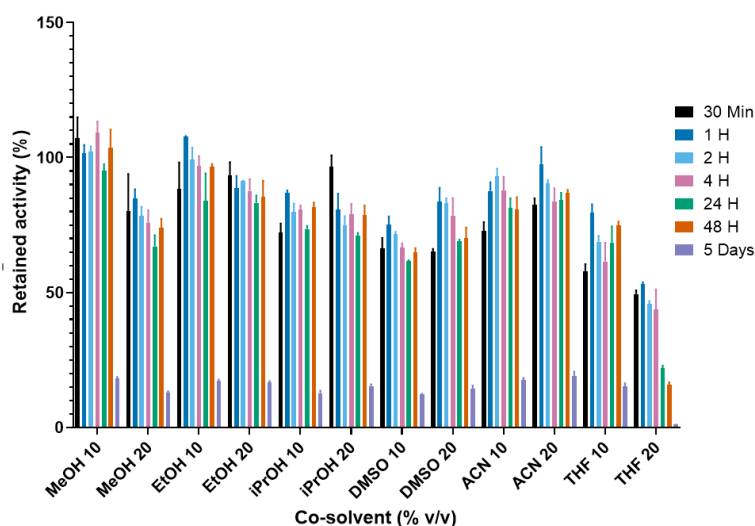


Figure 5. Effect of the incubation in different co-solvents on *AheGH1* stability. The activity was tested using the standard activity assay at 25 °C. Activity expressed relative to the activity at $t = 0$. Error bars represent standard deviations ($n = 3$).

The results obtained for the enzymatic stability at a different pH values (Fig. S5) showed that *AheGH1* is stable between pH 5 and 11, less active at pH 4 and 12 and completely unstable at pH 3. The enzyme shows no activity at pH 3 despite being selected from an acidophilic organism. Indeed, the ability of the source microorganism in dealing with acidic pH does not always translate to the isolated catalyst. Most adaptation mechanisms developed by acidophiles to survive at low pH involve very efficient homeostasis which prevents the ingress of protons to the cytoplasm.^[38] Consequently, the cytoplasmatic enzymes of those organism do not necessarily deal with acidic conditions and hence are not adapted to it.

Regarding stability at different temperatures (Fig. S6), *AheGH1* retains above 40% activity from 4 °C up to 55 °C, covering the usual range of temperatures of food industrial processes. This is lower than for the GH1 from thermoacidophile *Alicyclobacillus acidocaldarius* (WP_008336965.1; 52% identity),^[39] which was stable at up to 65 °C. On the other hand, the aforementioned *BglB* of *Paenibacillus polymyxa* has a half-life of 2.7 min at 55 °C. Arrizubieta and Polaina^[40] reported four mutations (three residues) that enhance its thermostability, H62R, M319I/V and M361I, with H62M resulting in the largest increase in thermostability. The equivalent residues in *AheGH1* are R62, A319 and F361 (see Supporting information). Thus, the most significant mutation in *BglB* is mirrored and

the other two residues are altered, which might contribute to the enhanced thermostability of *AheGH1*. Yet, while the authors propose a salt-bridge between R62 and E429 in *BglB*, the equivalent residue in *AheGH1* is Q429.

The thermal stability of *AheGH1* was measured via thermofluorimetry studies, following the increase in fluorescence intensity (excitation wavelength = 470-505 nm; emission = 540 nm-700 nm) that arises due to the binding of the fluorophore SYPRO™Orange to internally-located hydrophobic residues of the protein, that become exposed during thermal denaturation (see Materials and Methods). *AheGH1* unfolded in a single step, with a melting temperature (T_m) of 67°C (Fig. S7). This is approx. 10 °C higher than what has been reported for wild-type *BglB*,^[41,42] further confirming the increased thermostability of *AheGH1*.

Activity of AheGH1 with isoflavone glucosides

The hydrolytic activity of *AheGH1* towards three isoflavones glucosides (daidzin, glycitin, genistin) was assessed over 15 min incubation at 30 °C (Fig. 6). At the control reaction, where no enzyme was added, no hydrolysis occurred, and the 3 isoflavone glucosides remained intact. When *AheGH1* was added to the reaction, 100% conversion to the correspondent aglycons, daidzein, glycitein and genistein was consistently achieved. It is important to highlight that this hydrolysis occurs in the presence of 30% DMSO in addition to 15% ethanol present in the reaction, as both solvents have been used to solubilise the isoflavone glucosides in their stock solutions. These results are in line with the findings from the activity and stability assays in the presence of different co-solvents explained above.

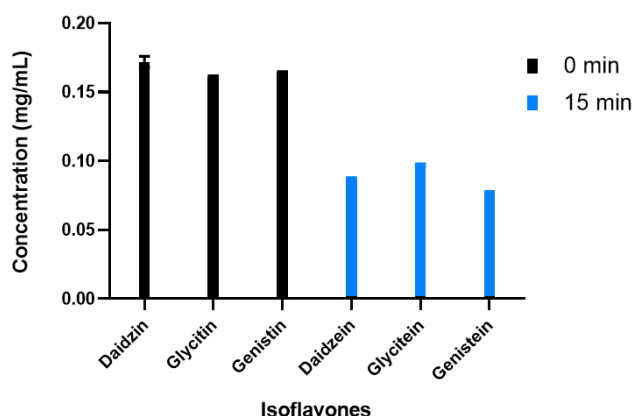


Figure 6. Enzymatic hydrolysis of 3 glucosides: daidzin, glycitin and genistin at time 0 and after 15 min reaction at 30 °C. Error bars represent standard deviations (n = 3).

Performance of *AheGH1* over soybean flour

Following the initial testing of *AheGH1* with the synthetic isoflavone glucosides, the performance of the enzyme was tested with a real isoflavone mixture extracted from soybean flour. In this case, the reaction was followed over time to monitor the hydrolysis of the isoflavone glucosides and aliquots were taken from the biotransformation at time 0 and after 15 min, 30 min, 1 h, 3 h, 24 h and 48 h. The evolution in the hydrolysis rate achieved by *AheGH1* is represented in Fig. 7.

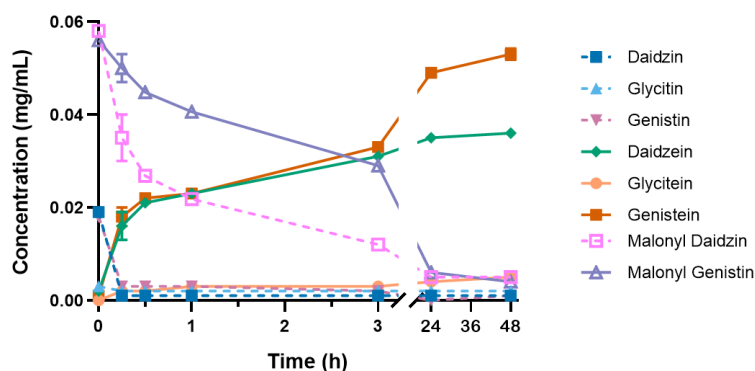


Figure 7. Hydrolysis reaction over soybean isoflavones glucosides at time 0, after 15 m, 30 m, 1 h, 3 h, 24 h and 48 h at 30 °C when *AheGH1* is present in the reaction. Error bars represent standard deviations (n = 3).

Within 15 min of incubation, almost no isoflavone glucosides remained in the mixture, matching what had been observed with the pure standards. However, the amount of aglycons in the sample treated with *AheGH1* continued to increase as the incubation time progresses. Concurrently, a decrease in two additional peaks was observed, which matched the standards malonyl daidzin and

malonyl genistin (Figures S8-13). The hydrolytic capacity of *AheGH1* had initially only been evaluated towards the 7-*O*- β -D-glucosides daidzin, glycitin and genistin. After quantification, the increase in aglycons closely matched the decrease observed for the malonyl-glucosides (Fig. 7), confirming that the enzyme was also capable of hydrolysing this form, albeit at a slower rate.

Conclusions

The extremozyme *AheGH1* was successfully cloned, purified and characterized. Its performance was tested under different conditions usually found in food industrial processes such as glucose, fructose, different co-solvents, broad range of temperatures and pH values showing promising results. The enzyme showed good tolerance to sugar (glucose and fructose) along with excellent performance in the presence of organic solvents, in particular ethanol. *AheGH1* also exhibited wide pH-stability and broad thermostability, making of it an outstanding candidate for its application in the food industry.

As a particular example, the evaluation of the hydrolytic capacity of *AheGH1* towards the most common isoflavone glucosides found in soybeans, concluded that the enzyme can efficiently convert isoflavone glucosides into aglycones, and hence, it constitutes a promising candidate for the enzymatic production of soybean isoflavones.

Experimental Section

Chemicals and materials

Commercially available reagents, organic solvents, and media were purchased from ACROS Organics, Thermo Fisher Scientific, Merck, or Sigma-Aldrich. The synthetic gene was purchased from Thermo and plasmid DNA purification kit from Macherey-Nagel. DNA ladder, protein marker and restriction were purchased from New England Biolabs.

Soybean flour and isoflavone standards were purchased from Merk and Carbosynth.

Discovery of Ahe sequence

A protein BLAST search was performed using the sequence of the halophilic β -glucosidase *BglA* from *Halothermothrix orenii*^[43,44] against the Alicyclobacillus genera. *BglA* was used as model protein for its excellent activity and stability on a broad variety of conditions. Alicyclobacillus genera was selected because of its acidophilic properties. Candidate sequences from different Alicyclobacillus species were obtained, and *Ahe* was selected for showing better tolerance to acidic pH.

Microbial strains and plasmids

The synthetic gene coding for *Alicyclobacillus herbarius* β -GH1 (*AheGH1*), with restriction enzymes BamHI and HindIII flanking the sequences as restriction sites, was codon optimised for *E. coli* and then ordered from GeneArt (ThermoFisher) in pMA, a commercial cloning vector. The gene was digested with BamHI and HindIII and ligated into pCH93b, an expression vector produced in-house,^[45] which includes a C-terminal poly-His tag for purification. *E. coli* strain XL10-Gold harbouring the plasmid was grown at 37 °C in LB medium supplemented with ampicillin (0.1 mg/mL). The gene was sequenced to confirm that the cloning was successful.

Expression and purification

Cells of *E. coli* BL21(DE3) harbouring the recombinant plasmid were grown at 37 °C in LB medium supplemented with ampicillin (0.1 mg/mL). When the OD₆₀₀ was between 0.6-0.8, isopropyl β -D-1-thiogalactopyranoside (1 mM) was added to induce protein expression and the culture was left at 30 °C overnight. Cells were harvested at 4500 g, 4 °C, 20 min and the pellet stored at -20 °C until purification.

The cell pellet was resuspended in buffer (HEPES (50 mM), sodium chloride (150 mM), imidazole (10 mM), pH 7.5) and cells were broken by sonication (6 min cycle, 5s on, 5s off, 50% amplitude, ¼ inch probe, Fisherbrand™ Model 120 Sonic Dismembrator). The lysate was collected by centrifugation at 14500 G, 1 h, 4 °C, and the pellet was discarded.

The supernatant was then filtered through PES 0.45 μm filters before loading it onto a IMAC column previously loaded with NiSO_4 0.1 M and washed with loading buffer (HEPES (50 mM), sodium chloride (150 mM), imidazole (10 mM), pH 7.5). The column was washed with loading buffer until a plateau in the UV_{280} absorbance was reached. Low affinity binding proteins were eluted using a step gradient 10% elution buffer and the protein of interest was eluted using 100% elution buffer (HEPES (50 mM), sodium chloride (150 mM), imidazole (300 mM), pH 7.5). The purified enzyme was dialysed overnight. Protein quantification was performed by measuring absorbance at 220, 250 and 280 nm using a BioTek Take3 Microplate reader using predicted extinction coefficients (54854.15 Da, $108415 \text{ M}^{-1} \text{ cm}^{-1}$).^[23]

Gel filtration

Gel filtration chromatography was performed on a Superdex 200 10/300 GL column (GE Healthcare), using a mobile phase consisting of Tris-HCl (50 mM), KCl (100 mM), pH 7.5. Injection volume: 750 μL , flow rate 0.75 mL/min. Samples were prepared to a final protein concentration of approx. 1 mg/mL. A calibration curve was generated using the Sigma Aldrich Gel Filtration Markers Kit for Protein Molecular Weights 12,000–200,000 Da (MWGF200).

Thermal shift Assay

6 μL of *AheGH1* (8.8 mg/ml) was mixed with 9 μL SYPROTMOrange Protein Stain (Sigma-Aldrich) solution (prepared by diluting 0.7 μL in 250 μL water). Thermal shift assays were carried out using the MiniOpticon Real-time thermocycler (Bio-rad), over a temperature gradient of 15-99°C, increasing at a rate of 2°C/min. The final melting temperature (T_m) was calculated from the derivative of each sigmoidal, melting curve, as an average of quadruple values.

Crystallization of AheGH1

AheGH1 crystals were grown using the sitting drop vapor diffusion technique, using an Orxy4 crystallization robot (Douglas Instruments). Briefly, 400 nl drops comprising 10 mg/ml *AheGH1*, in 10 mM HEPES pH 7.5, 100 mM NaCl. Drops were set up in 3-drop, round CrystalQuick 96-well plates (Greiner Bio-One). Each reservoir contained 100 μL of 96 different crystallization conditions

from the PACT Premier screen (Molecular Dimensions). *AheGH1* crystals grew after 5 days in condition F2 (0.2 M NaBr, 20% (w/v) PEG 3350, 0.1 M Bis Tris Propane pH 6.5, at room temperature. Crystals were cryoprotected 0.1 M Bis Tris Propane pH 6.5, 30% (w/v) PEG 3350 and 30% ethylene glycol cryocooled in liquid nitrogen.

Data collection and 3D structure determination of AheGH1

X-ray diffraction data were collected on the I04 beamline at the Diamond Light Source (DLS Didcot, UK). Diffraction data were reduced using XDS^[46] and anisotropically truncated and scaled with STARANISO.^[47] Data collection statistics are reported in Table S1.

The structure was solved by molecular replacement using MOLREP^[48] from the CCP4 suite using the crystal structure of a thermostable β -glucosidase from *Halothermothrix orenii* as a search model (PDB: 4PTV.^[49] The model was further built using Coot^[50] and refined using phenix.refine^[51] and BUSTER.^[52] Water molecules were added using ARP/wARP suite^[53] and manually inspected in Coot. The final model was inspected and validated with MolProbity.^[54] Coordinates and structure factors of the *AheGH1* have been deposited in the Protein Data Bank (www.rscb.org) with accession code 6YN7.

Activity assay and kinetics

Standard β -glucosidase activity was determined spectrophotometrically using 10 mM *p*-nitrophenyl- β -D-glucopyranoside (*p*NPG) at 25 °C. 10 μ L of a suitable enzyme dilution was added to a 96-well plate per triplicate. Immediately before the assay, 0.29 mL of the reaction solution HEPES (50 mM), *p*NPG (10 mM), pH 7.4 were added and the *p*-nitrophenol formation was followed at 420 nm for 10 minutes. The specific activity (U/mg) was expressed as μ mol of product formed per minute per milligram of protein. The extinction coefficient used for the *p*-nitrophenol was calculated to be 8.64 $\text{mM}^{-1}\text{cm}^{-1}$.

To measure the kinetic properties of the studied enzyme, different concentrations of the substrate were used, and the enzymatic activity was measured using the same method as in the standard activity

assay. Data were then plotted and fitted to a substrate inhibition curve^[31] $\left(v = \frac{k_{cat}}{1 + \frac{K_m + [S]}{K_i}} \right)$ using

GraphPad Prism 8.

Synthetic isoflavones standards

Stock solutions

Stock solutions for the β -glucoside isoflavones (daidzin, glycitin and genistin) and for the aglycon isoflavones (daidzein, glycitein and genistein) were prepared from authentic samples (1 mg/mL in ethanol for daidzin, daidzein and genistein and 1 mg/mL in DMSO for genistin, glycitin and glycitein). Calibration solutions were prepared by dilution of the main stock solutions.

Enzymatic reaction

Prior to the testing of the enzyme in the real matrix, a solution containing the 3 main isoflavone glucosides present in soybeans (daidzin, genistin and glycitin) was prepared in triplicate. 100 μ L sample containing the 3 isoflavones glucosides (150 μ g/mL final concentration) and the enzyme (0.05 mg/mL final concentration) were left in agitation at 30 °C for 15 minutes. After that time, 450 μ L of ACN were added to stop the enzymatic reaction and 450 μ L of distilled water were added to top up until 1 mL total volume. The samples were then analysed by HPLC.

Isoflavone extraction from soybean flour

1 g of soybean flour was weighted out and suspended in 4 mL of ethanol and 16 mL of distilled water. The sample was then sonicated for 20 min (5 s on, 5 s off, 60% amplitude, ¼ inch probe, Fisherbrand™ Model 120 Sonic Dismembrator) centrifuged for 30 min at 4500 g and filtered using a 0.45 μ m filter.

Enzymatic biotransformation

After the extraction, 0.5 mg/mL of *AheGH1* were added to a tube containing 3 mL of soybean extraction, in triplicate, and the samples were left in shaking at 30 °C. To follow the progress of the

biotransformation, 100 μ L aliquots were taken from the sample after 15 min, 30 min, 1 h, 3 h, 24 h and 48 h and transferred to HPLC vials. 450 μ L of ACN and 450 μ L of distilled water were added to stop the enzymatic reaction. The samples were then filtered (0.45 μ m) and analysed by HPLC.

Reverse-phase HPLC analysis of conversions

Samples were analysed using a ThermoFisher Ultimate 3000 Reverse-phase HPLC (diode array detector) on a Waters XBridge C18 column (3.5 μ m, 2.1 x 150 mm) with the following method: A: 0.1% TFA in water, B: 0.1% TFA in acetonitrile. Gradient: 0 min 95% A 5% B; 1 min 95% A 5% B; 5 min 5% A 95% B; 5.10 min 0% A 100% B; 6.60 min 0% A 100% B; 7 min 95% A 5% B; 10 min 95% A 5% B. Injection volume 2 μ L, at 45 °C with a flow rate of 0.8 mL/min. Retention times in minutes: daidzin (3.17), glycitin (3.22), genistin (3.40), malonyl-daidzin (3.44), malonyl-genistin (3.62), daidzein (3.76), glycitein (3.82), genistein (4.04).

Conversions were calculated from a calibration curve of authentic standards. Peak areas were manually integrated, and the correlation coefficients were obtained using the Microsoft Excel® linear regression model application.

Funding

LD was supported by the UK Engineering and Physical Sciences Research Council [grant number EP/L015633/1]. CH was supported by the Biotechnology and Biological Sciences Research Council through the iCASE scheme in collaboration with Johnson Matthey [grant number BB/M008770/1]. LJG was gratefully funded by Università degli Studi di Milano ‘Linea 2’ funding.

Conflict of interest

The authors declare no conflict of interest in publishing this work.

Acknowledgements

Authors would like to thank the Centre for Doctoral Training in Sustainable Chemistry of the University of Nottingham for its support and the School of Chemistry for access to analytical infrastructure. We thank Diamond staff from the I04 beamline and iNEXT (PID: 5912) for financial support.

Bibliography

- [1] A. Vasella, G. J. Davies, M. Böhm, *Curr. Opin. Chem. Biol.* **2002**, *6*, 619–629.
- [2] G. Singh, A. K. Verma, V. Kumar, *3 Biotech* **2016**, *6*, 1–14.
- [3] Y. Yang, X. Zhang, Q. Yin, W. Fang, Z. Fang, X. Wang, X. Zhang, Y. Xiao, *Sci. Reports* **2015**, *5*, 17296.
- [4] A. Sørensen, M. Lübeck, P. S. Lübeck, B. K. Ahring, *Biomolecules* **2013**, *3*, 612–631.
- [5] P. H. Rampelotto, *Life (Basel, Switzerland)* **2013**, *3*, 482–5.
- [6] S. Elleuche, C. Schröder, K. Sahm, G. Antranikian, *Curr. Opin. Biotechnol.* **2014**, *29*, 116–123.
- [7] M. C. Pagano, M. Miransari, *Production Worldwide*, Elsevier Inc., **2016**.
- [8] S. Kudou, Y. Fleury, D. Welti, D. Magnolato, T. Uchida, K. Kitamura, K. Okubo, *Agric. Biol. Chem.* **1991**, *55*, 2227–2233.
- [9] C. Cui, R. L. Birru, B. E. Snitz, M. Ihara, C. Kakuta, B. J. Lopresti, H. J. Aizenstein, O. L. Lopez, C. A. Mathis, Y. Miyamoto, L. H. Kuller, A. Sekikawa, *Nutr. Rev.* **2020**, *78*, 134–144.
- [10] D. C. Vitale, C. Piazza, B. Melilli, F. Drago, S. Salomone, *Eur. J. drug Metab. Pharmacokinet.* **2013**, *38*, 15–25.

- [11] P. A. Murphy, K. Barua, C. C. Hauck, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2002**, 777, 129–138.
- [12] P. E. Hessler, P. E. Larsen, A. I. Constantinou, K. H. Schram, J. M. Weber, *Appl. Microbiol. Biotechnol.* **1997**, 47, 398–404.
- [13] A. R. Reeves, R. Seshadri, I. A. Brikun, W. H. Cernota, M. C. Gonzalez, J. M. Weber, *Appl. Environ. Microbiol.* **2008**, 74, 7383–7390.
- [14] J. M. Weber, A. R. Reeves, R. Seshadri, W. H. Cernota, M. C. Gonzalez, D. L. Gray, R. K. Wesley, *Appl. Microbiol. Biotechnol.* **2013**, 97, 6427–6437.
- [15] D. Prat, J. Hayler, A. Wells, *Green Chem.* **2014**, 16, 4546–4551.
- [16] Y. Xue, X. Song, J. Yu, *World J. Microbiol. Biotechnol.* **2009**, 25, 2165–2172.
- [17] X. Song, Y. Xue, Q. Wang, X. Wu, *J. Agric. Food. Chem.* **2011**, 59, 1954–1961.
- [18] H. Cao, Y. Zhang, P. Shi, R. Ma, H. Yang, W. Xia, Y. Cui, H. Luo, Y. Bai, B. Yao, *J. Ind. Microbiol. Biotechnol.* **2018**, 45, 369–378.
- [19] J. Batra, S. Mishra, *J. Mol. Catal. B: Enzym.* **2013**, 96, 61–66.
- [20] P. O. De Giuseppe, T. D. A. C. B. Souza, F. H. M. Souza, L. M. Zanphorlin, C. B. Machado, R. J. Ward, J. A. Jorge, R. D. P. M. Furriel, M. T. Murakami, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2014**, 70, 1631–1639.
- [21] C. S. Chan, L. L. Sin, K.-G. Chan, M. S. Shamsir, F. A. Manan, R. K. Sani, K. M. Goh, *Biotechnol. Biofuels* **2016**, 9, 174.
- [22] K. Goto, H. Matsubara, K. Mochida, T. Matsumura, Y. Hara, M. Niwa, K. Yamasato, *Int. J. Syst. Evol. Microbiol.* **2002**, 52, 109–113.
- [23] B. A. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., *The Proteomics Protocols Handbook*, Humana Press, **2005**.

- [24] L. Holm, *Bioinformatics* **2019**, 1–2.
- [25] P. Isorna, J. Polaina, L. Latorre-García, F. J. Cañada, B. González, J. Sanz-Aparicio, *J. Mol. Biol.* **2007**, *371*, 1204–1218.
- [26] R. A. Laskowski, J. D. Watson, J. M. Thornton, *Nucleic Acids Res.* **2005**, *33*, W89–93.
- [27] T. Barrett, C. G. Suresh, S. P. Tolley, E. J. Dodson, M. A. Hughes, *Structure* **1995**, *3*, 951–960.
- [28] X. Robert, P. Gouet, *Nucleic Acids Res.* **2014**, *42*, W320–W324.
- [29] J. Sanz-Aparicio, J. A. Hermoso, M. Martínez-Ripoll, B. González, C. López-Camacho, J. Polaina, *Proteins Struct. Funct. Genet.* **1998**, *33*, 567–576.
- [30] S. McNicholas, E. Potterton, K. S. Wilson, M. E. M. Noble, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2011**, *67*, 386–394.
- [31] R. A. Copeland, *Enzymes. A Practical Introduction to Structure, Mechanism, and Data Analysis*, Wiley-VCH, Inc., New York, **2000**.
- [32] P. Pang, L. chuang Cao, Y. huan Liu, W. Xie, Z. Wang, *J. Struct. Biol.* **2017**, *198*, 154–162.
- [33] L. S. C. Costa, D. C. B. Mariano, R. E. O. Rocha, J. Kraml, C. H. da Silveira, K. R. Liedl, R. C. De Melo-Minardi, L. H. F. De Lima, *Molecules* **2019**, *24*, DOI 10.3390/molecules24183215.
- [34] M. Hrmova, G. B. Fincher, *Carbohydr. Res.* **2007**, *342*, 1613–1623.
- [35] A. Martino, C. Schiraldi, A. Di Lazzaro, I. Fiume, G. Spagna, P. . Pifferi, M. De Rosa, *Process Biochem.* **2000**, *36*, 93–102.
- [36] V. R. N. Telis, J. Telis-Romero, H. B. Mazzotti, A. L. Gabas, *Int. J. Food Prop.* **2007**, *10*, 185–195.

- [37] H. Michlmayr, C. Schümann, N. M. Barreira Braz Da Silva, K. D. Kulbe, A. M. Del Hierro, *J. Appl. Microbiol.* **2010**, *108*, 550–559.
- [38] H. Stan-Lotter, S. Fendrihan, *Adapt. Microb. Life to Environ. Extrem. Nov. Res. Results Appl. Second Ed.* **2017**, 1–342.
- [39] B. Di Lauro, M. Rossi, M. Moracci, *Extremophiles* **2006**, *10*, 301–310.
- [40] M. J. Arrizubieta, J. Polaina, *J. Biol. Chem.* **2000**, *275*, 28843–28848.
- [41] M. Camarillo-Cadena, G. Garza-Ramos, M. Peimbert, G. Pérez-Hernández, R. A. Zubillaga, *Protein J.* **2011**, *30*, 318–323.
- [42] M. Camarillo-Cadena, G. Garza-Ramos, M. Peimbert, J. Polaina, G. Pérez-Hernández, R. A. Zubillaga, *Protein J.* **2012**, *31*, 615–622.
- [43] A. Bhattacharya, B. I. Pletschke, *Enzyme Microb. Technol.* **2014**, *55*, 159–169.
- [44] L. Delgado, M. Parker, I. Fisk, F. Paradisi, *Food Chem.* **2020**, *323*, 126825.
- [45] C. M. Heckmann, L. J. Gourlay, B. Dominguez, F. Paradisi, *Front. Bioeng. Biotechnol.* **2020**, *8*, 707.
- [46] W. Kabsch, *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2010**, *66*, 125–132.
- [47] A. Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, G. Vornrhein, C., Bricogne, **n.d.**
- [48] A. Vagin, A. Teplyakov, *J. Appl. Crystallogr.* **1997**, *30*, 1022–1025.
- [49] N. Hassan, T.-H. Nguyen, M. Intanon, L. D. Kori, B. K. C. Patel, D. Haltrich, C. Divne, T. C. Tan, *Appl. Microbiol. Biotechnol.* **2015**, *99*, 1731–1744.
- [50] P. Emsley, K. Cowtan, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2004**, *60*, 2126–2132.
- [51] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd,

L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010**, *66*, 213–221.

[52] T. O. Bricogne, G; Blanc, E; Brandl, M; Flensburg, C; Keller, P; Paciorek, W; Roversi, P; Sharff, A; Smart, O.S; Vornrhein, C; Womack, **2017**.

[53] G. Langer, S. X. Cohen, V. S. Lamzin, A. Perrakis, *Nat. Protoc.* **2008**, *3*, 1171–1179.

[54] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **2010**, *66*, 12–21.