# Producing natural vanilla extract from green vanilla beans using a β-glucosidase from *Alicyclobacillus acidiphilus*

Lidia Delgado[a], Christian M. Heckmann[a], Stefano De Benedetti[b], Marco Nardini[b], Louise J. Gourlay[b], Francesca Paradisi[a,c,*]

[a] University of Nottingham, School of Chemistry, Department of Chemical Biology, University Park, Nottingham NG7 2RD, United Kingdom.

[b] Dipartimento di Bioscienze, Università di Milano, Via Celoria 26, 20133, Milan, Italy.

[c] University of Bern, Department of Chemistry and Biochemistry, Freiestrasse 3, CH-3012 Bern Switzerland

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



## Abstract

Current methods for the production of natural vanilla extract are long and tedious, and the efficiency of the vanillin extraction is usually conditioned by different factors during the traditional curing process (temperatures and weather conditions). As an important fraction of vanillin is present in the form of glucovanillin in green beans, endogenous β-glucosidases contribute to its hydrolysis; however, these enzymes lose efficiency during the curing process. The use of extremophilic organisms as a source of an appropriate exogenous enzyme can offer a valid alternative when producing natural vanillin. Here, a β-glucosidase from the thermo-acidophilic organism *Alicyclobacillus acidiphilus* (*Aac*GH1) was cloned, expressed in *E. coli* BL21, and fully characterized in respect to both function and crystal structure. Notably, *Aac*GH1 was stable at a temperature up to 50 ºC and exhibited good tolerance to glucose, fructose and organic solvents, in particular it maintained full activity in the presence of up to 20% (*v/v*) ethanol. The enzyme was then successfully

applied to an ethanol-water (20% (*v/v*)) extract of green vanilla beans and the complete hydrolysis of glucovanillin (1.7 mM) to vanillin, and other flavour compounds commonly found in vanilla, was achieved using 0.5 mg/mL of enzyme in just 15 minutes at 30 ºC.

**Keywords:** Extremophile, β glucosidase, Glucovanillin, Vanilla, Vanillin.

## 1 Introduction

β-glucosidases (EC 3.2.1.21) are a diverse group of enzymes that hydrolyse terminal, non-reducing glycosyl residues from glycosides (Vasella et al., 2002). These enzymes are gaining importance due to the broad range of biotechnological applications where they can be used: surfactant, biofuel, agricultural, and food industries, among others (Singh et al., 2016). Industrial processes using enzymes are often ecologically friendlier, as they usually require less energy and are less hazardous than processes using chemical substances (Ventura-Sobrevilla et al., 2015).

In the specific case of the food industry, many processes involve the use of harsh conditions (low pH, high concentrations of solvents and sugars, and high temperatures) that can lead to enzyme inactivation (Sørensen et al., 2013; Yang et al., 2015). Several strategies exist to improve enzyme stability, such as directed evolution (Arnold, 2009), site-directed mutagenesis (Guo et al., 2016) or immobilization (Mohamad et al., 2015). The search for novel and improved β-glucosidases that meet the requirements for industrially suitable enzymes is ongoing. In that regard, extremophilic organisms, which are very well adapted to extreme environmental conditions (Rampelotto, 2013), constitute an alternative source of enzymes for industrial application as they are generally more able to withstand industrial processes in comparison with their mesophilic counterparts (Elleuche et al., 2014).

A very profitable industry where β-glucosidases can be applied is the vanilla industry. Vanilla aroma is widely used in the food, medicine, pharmaceutical, cosmetic, and perfume industries, among others.

Green vanilla pods do not have the characteristic scent of vanilla, because vanillin is principally present in the glycosylated non-volatile form of glucovanillin, which gradually increases in the pod as it matures and is harvested. The composition of the pod substantially changes after being subjected to the curing process, which traditionally comprises four major steps including killing, sweating, drying, and conditioning, with an overall duration of 150-180 days (Anuradha et al., 2013). During such a process, a series of biochemical and enzymatic changes generate about 200 compounds which give the characteristic flavour and aroma of vanilla (Baqueiro-Peña and Guerrero-Beltrán, 2017). The main contribution to the aroma comes from native glucosidases-mediated hydrolysis which releases several aroma compounds, the most concentrated being vanillin. By the end of the curing process, almost all of the glucoside has been converted into vanillin (Dignum et al., 2001), contributing to 1–2% (*w/w*) of the cured pod (Sinha et al., 2008); however, other additional compounds also contribute to the complexity and roundedness that distinguishes vanilla from synthetic vanillin.

The length, labour-intensity, and high sensitivity to the conditions during growth and processing cause high and fluctuant prices that have triggered a shift towards artificial vanilla flavour or pure synthetic vanillin (Pardío et al., 2018); however, this does not replace natural vanilla extract in high-quality products (Khoyratty, Shahnoo, Kodja and Verpoorte, 2018). Ways to accelerate the curing process have also been investigated (Waliszewski et al., 2007) but have so far not replaced the traditional process. As natural vanilla extract is commonly sold in the form of ethanolic extracts, the treatment of green vanilla extracts, avoiding the curing of whole beans altogether, is also of industrial interest.

Ruíz-Terán *et al* reported vanillin yields two to three times higher compared to Soxhlet extracts of traditionally cured vanilla, when simultaneously adding Viscozyme and Celluclast to aqueous suspensions of chopped green vanilla pods, incubating for 8 h at 50 °C, followed by addition of 47.5% ethanol (Ruiz-Terán et al., 2001). Similarly, Zhang *et al* reported the use of a pectinase from *Aspergillus niger* to hydrolyse the pectin that is compartmentalizing glucovanillin and endogenous β-glucosidases (50 °C, 7 h), resulting in an approx. 2-fold increase of vanillin production (Zhang et

al., 2014). Pardío *et al* found that additional vanillin could be released from homogenized green vanilla beans after release using the endogenous β-glucosidase had stalled by incubating with Crystalzyme PML-MX at pH 5.0 and 40_ºC for 72 h (Pardío et al., 2018). Perera and Owen reported the combined use of an exogenous cellulase, pectinase, and β-glucosidase on green vanilla homogenate (1 h at 37 °C) as well as whole green vanilla that had been freeze-thawed (4 days at 37 °C) to increase vanillin formation by up to 3.5-fold compared to traditionally cured vanilla (Perera and Owen, 2010).

Herein, a β-glucosidase of the glycosyl hydrolase family 1 from the thermo-acidophilic organism *Alicyclobacillus acidiphilus* (*Aac*) (Matsubara et al., 2002) was cloned and expressed in *E. coli*, its crystal structure was solved and the enzyme was functionally characterized with regards to different operational conditions (high glucose and fructose concentrations, organic co-solvents, and range of pH values and temperatures) often encountered in food industrial processes. Following the initial characterization, the hydrolytic capacity of *Aac*GH1 in ethanol-water (20% *v/v*) extracts of green vanilla pods was evaluated.

## 2 Materials and methods

### *2.1 Chemicals and materials*

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

Chemical standards were purchased from Merck. Madagascan cured vanilla pods (*Vanilla planifolia*) were purchased from Amazon UK. Green vanilla pods (*Vanilla planifolia*) was kindly donated by the Botanical Gardens of Birmingham (UK).

### *2.2 Microbial strains and plasmids*

The synthetic gene coding for *Alicyclobacillus acidiphilus* β-GH1 (*Aac*GH1), with restriction enzymes BamHI and HindIII flanking the sequences as restriction sites, were optimised for *E. coli* and then ordered from GeneArt (Thermo Fisher) already into a commercial cloning vector. The gene was digested with BamHI and HindIII and ligated into the expression vector produced in-house pCH93b, (Heckmann et al., 2020) which includes a C-terminal poly-His tag for purification. *E. coli* strain XL10-Gold harbouring the plasmid was grown at 37º C in Luria Bertani medium supplemented with ampicillin (0.1 mg/mL). The gene was sequenced to confirm that the cloning was successful. The sequencing was performed by the DNA Sequencing Facility based in the School of Life Sciences (University of Nottingham) using a 3130xl ABI PRISM Genetic Analyzer (Life Technologies).

### *2.3 Expression and purification*

Cells of *E. coli* BL21(DE3) harbouring the recombinant plasmid were grown at 37 ºC in LB medium supplemented with ampicillin (0.1 mg/mL). When an $OD_{600}$ between 0.6-0.8 was reached, isopropyl β-D-1-thiogalactopyranoside (1 mM) was added as an inductor for the overexpression of the enzyme,

and the culture left at 30 ºC 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 disrupted by sonication (6 min cycle, 5s on, 5s off, 50% amplitude, ¼ inch probe, Fisherbrand™ Model 120 Sonic Dismembrator). The supernatant was collected by centrifugation at 14500 g, 1 h, 4º C, and the pellet was discarded.

The supernatant was then filtered through Millex® PES 0.45 µm filter before loading it onto a HisTrap 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 into storage buffer (HEPES (50 mM), sodium chloride (150 mM), pH 7.5) overnight. Protein quantification was performed by measuring absorbance at 220, 250, and 280 nm using a BioTek Take3 Microplate reader using predicted extinction coefficients (54629.11 Da, 115865 $M^{-1}$ $cm^{-1}$; https://web.expasy.org/protparam/).

### *2.4 Gel filtration*

Gel filtration chromatography was performed on a GE Healthcare Superdex 200 10/300 GL column. Mobile phase consisted of Tris-HCl 50 mM, KCl 100 mM and pH 7.5. Injection volume was 750 µL and flow rate 0.8 mL/min. Samples were prepared to a final protein concentration of approx. 1 mg/mL. Sigma Aldrich Gel Filtration Markers Kit for Protein Molecular Weights 12,000–200,000 Da (MWGF200) was used to generate the calibration curve.

### 2.5 *Crystallization of AacGH1*

*Aac*GH1 (10 mg/ml; 10 mM HEPES pH 7.5, 100 mM NaCl) was crystallized in 0.8 µl microseeded sitting drops prepared using the Orxy 4 crystallization robot (Douglas Instruments) and flat-bottomed, Greiner CrystalQuick 96 well sitting drop plates (Greiner Bio-one), incubated at 20 °C. Microseeds were prepared by crushing *Aac*GH1 microcrystals that grew over 1 week in a 0.5 µl drops containing 50% protein mixed with condition H3 (0.1 M Bis-Tris pH 5.5 and 25% (w/v) PEG3350) from the JCSG screen (Molecular Dimensions), using the Seed Bead Kit (Hampton Research). Seeded drops contained 0.10 µl seed stock, 0.4 µl protein (10 mg/ml), and 0.30 µl reservoir solution and were suspended over 100 µl reservoir solution: H3 condition (0.2 M sodium iodide, 0.1 M Bis-Tris Propane pH 8.5 and 20% (w/v) PEG3350) from the PACT screen (Molecular Dimensions). Crystals were cryoprotected in 0.1 M Bis-Tris Propane pH 8.5 and 40% (w/v) PEG3350. X-ray diffraction data were collected on crystals that grew over two weeks.

### 2.6 *Data collection and 3D structure determination of AacGH1.*

X-ray diffraction data were collected on a single *Aac*GH1 crystal at 1.55 Å resolution on the XDR2 beamline at the ELETTRA synchrotron facility (Trieste, Italy). The space group was initially determined as a *P*2, however, data analysis using XTRIAGE available under the Phenix suite (Adams et al., 2010) revealed evidence of twinning. It was not possible to solve the structure in this space group, therefore the structure was solved in the triclinic (*P*1) space group with the following unit cell parameters: a = 62.7 Å, b= 91.6 Å, c=159.3 Å, $\alpha$ = 88.4°, $\beta$= 89.6°, $\gamma$ = 90.0° (Table S1). Eight *Aac*GH1 chains (Chains A to H) were present in the *P*1 asymmetric unit, with an estimated Matthew's coefficient of 2.86 Å3/Da (57.0 % solvent content). Data reduction was carried out using XDS (Kabsch, 2010) and scaled using STARANISO (Tickle et al., 2018). The search model identification and the molecular replacement were carried out by BALBES using the amino acid sequence as the input (Long et al., 2007). Structure completion was carried out manually and refined at 1.95 Å to convergence using COOT and REFMAC5, and structure geometry was validated by Molprobity in the PHENIX platform (Table S1) (Adams et al., 2010; Davis et al., 2007; Emsley et al., 2010; Vagin

et al., 2004). Atomic coordinates and structure factors were deposited in the RCSB Protein Data Bank (www.rcsb.org) under the accession code 6ZIV.

### *2.7 Activity assay and kinetics*

The β-glucosidase activity was determined spectrophotometrically using 10 mM *p*-nitrophenyl-β-D-glucopyranoside (*p*NPG) at 25 ºC. 10 µL of the 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 extinction coefficient used for the *p*-nitrophenol was calculated to be 8.64 $mM^{-1}cm^{-1}$. The specific activity (U/mg) was expressed as µmol of product formed per minute per milligram of protein.

To measure the kinetic properties of the studied enzyme, different concentrations of the substrate were used, the enzymatic activity was measured using the same method as in the standard activity assay. Data were then plotted and fitted to the standard Michaelis-Menten curve using GraphPad Prism 8.

### *2.8 Synthetic vanillin*

#### *2.8.1 Stock solution*

Vanillin 4-*O*-β-D-glucoside standard stock solution was prepared in water with a concentration of 10 mM. Calibration solutions were prepared by diluting the main stock solution.

#### *2.8.2 Enzymatic hydrolysis*

Prior to the testing of the enzyme in the real matrix, a solution containing 5 mM synthetic vanillin 4-*O*-β-D-glucoside was prepared. 9 replicates of 100 µL volume containing the synthetic vanillin glucoside (4.5 mM final concentration) and the enzyme (0.1 mg/mL final concentration) were left in agitation at 30 °C. After 10, 20, and 30 minutes, 3 replicates were taken to analyse the progression of the hydrolysis. 450 µL of ACN and 450 µL of 0.2% HCl were added to stop the enzymatic reaction and top up the sample until 1 mL total volume. The samples were then analysed by HPLC.

### *2.9 Vanillin extraction from vanilla pods*

#### *2.9.1 Cured vanilla pods*

1 g of the cured vanilla pod was cut into very small pieces. 8 mL of pure ethanol were added, and the solution was left macerating with shacking for 30 min at 30 ºC. After this time, 12 mL of distilled water were added to achieve a final ethanol concentration in the sample of 40% (*v/v*). The sample was then sonicated for 20 min (5 s on, 5 s off, 60 % amplitude) at 60ºC, centrifugated for 30 min at 10000 g, and filtered using a 0.45 µm filter.

#### *2.9.2 Green vanilla pods*

For the green vanilla pod, the same extraction protocol as with the cured vanilla was followed, but in order to be consistent, 2.7 g of the fresh vanilla pod were used instead of 1 g as the fresh pod was 2.7 times heavier than a cured one of similar length. 4 mL of ethanol were used for the maceration step and topped up with 16 mL of distilled water, giving a final ethanol concentration of 20% (*v/v*).

#### *2.9.3 Green vanilla pod treatment*

After the extraction, 0.5 mg/mL of *Aac*GH1 was added to a tube containing 2 mL of the cured and green vanilla extraction in each triplicate, and the samples were left under shaking at 30 ºC. To follow the progress of the biotransformation, 100 µL aliquots were taken from the 2 mL sample after 15 min, 30 min, 1 h, 3 h, 24 h, and 48 h and transferred to sample vials. 450 µL of ACN and 450 µL of 0.2 % HCL were added to the mix, the samples were filtered (0.45 µm) and analysed by HPLC.

### *2.10 Reverse-phase HPLC analysis of biotransformations*

Samples were analysed using a ThermoFisher Ultimate 3000 Reverse-phase HPLC (diode array detector) on a Waters XBridge C18 column (3.5 µm, 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 µL, at 45 °C with a flow rate of 0.8 mL/min. Retention times

in minutes: vanillin 4-*O*-β-D-Glucoside (1.53), 4-hydroxybenzoic acid (1.65), vanillic acid (2.46), 4-hydroxybenzaldehyde (2.47), vanillin (3.05).

Conversions were calculated from a calibration curve of authentic standards. Peak areas were integrated manually (220, 250, and 280 nm), and the correlation coefficients were obtained using the Microsoft Excel® linear regression model application. The coeluting vanillic acid and 4-hydroxybenzaldehyde were quantified by disentangling the peaks based on the different absorption characteristics of the compounds (see supporting information for details, Table S2).

## 3 Results and discussion

### *3.1 Protein expression and purification*

*Aac*GH1 was expressed in excellent yields (between 160-195 mg/L of culture), purified by metal affinity chromatography (Hochuli et al., 1988), and analysed by SDS-PAGE (Fig. S1). *Aac*GH1 has a theoretical molecular weight of 54629 Da (≈55 kDa) estimated by the online tool ProtParam (Gasteiger et al., 2005) which was confirmed by SDS-PAGE. The molecular weight was determined to be 120 kDa by gel filtration, consistent with it being a dimer in solution.

### *3.2 The 3D structure of AacGH1*

*Aac*GH1 was crystallized in the *P*1 spacegroup (eight molecules in the asymmetric unit) and the structure solved and refined to 1.95 Å resolution (Table S1). Electron density was well-defined from N-terminal residues 3-6 to C-terminal residues 448-449, depending on the *Aac*GH1 chain present in the crystal asymmetric unit. Electron density was absent for the initial N-terminal and last C-terminal residues, some loop regions, and some solvent-accessible side chains, which were flexible and lost to the solvent. As calculated using secondary structure matching of 443 main chain residues using the SUPERPOSE program, all eight *Aac*GH1 chains were highly similar (RMSD = 0.26 Å). Chains C and F were more complete, with regards to modelled residues, therefore all subsequent analyses were carried out on chain C.

*Aac*GH1 presents the canonical glycoside hydrolase family 1-fold, consisting of a central TIM barrel motif, comprised of alternating α-helices and β-strands $(\alpha/\beta)_8$ (Fig. 1). As observed for other members of this family, *Aac*GH1 contains several additional secondary structure elements, including a few short α-helices and a two-stranded and a three-stranded anti-parallel β-sheet that are peripheral to the TIM barrel. While SEC determined the molecular weight to be 120 kDa, suggesting a dimer in solution, no such dimer was observed in the crystal structure. This could be explained by transient dimers that may only form under certain buffer conditions (e.g. presence of certain metal ions or depending on the ionic strength of the buffer).

It has been reported that the stability of acidophilic proteins at low pH is aided by a reduced number of positive and negative surface charge patches that prevent undesirable, destabilising electrostatic repulsions between charged side chains at low pH (Schwermann et al., 1994). Coulombic electrostatic potential surface calculations made with Chimera on *Aac*GH1 clearly illustrate a prevalence of uncharged and negatively charged surface residues, with very few positive ones (Fig. 1B) (Pettersen et al., 2004). Increased negative surface charges has been reported as a typical feature for acidophilic proteins that also exhibit thermal adaptation, in line with our functional data (Reed et al., 2013).

The sequence- and structure-based conservation of *Aac*GH1 was assessed using the ENDscript 2.0 server (http://endscript.ibcp.fr/ESPript/cgi-bin/ENDscript.cgi) revealing, as expected, that the TIM barrel is the most highly conserved region with respect to both sequence and structure (Robert and Gouet, 2014) (Fig. S2). The highest structural homology (50% sequence identity) is shared with a metagenomic β-glycosidase (PDB entry 5XGZ) (RMSD of 1.4 Å over 434/445 aligned residues), as deduced using the DALI 3D structure comparison server (http://ekhidna2.biocenter.helsinki.fi/dali/) (Holm, 2019). Only flexible loops displayed structural divergence.

A 3D functional template search carried out using Profunc (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc/) (Laskowski et al., 2005) revealed that *Aac*GH1 has an active site architecture that is most similar (E-value 1.36E-18) to a cyanogenic β-glucosidase from white clover (*Trifolium*

*repens*; PDB entry 1CBG) (Barrett et al., 1995). The two proteins share a 40.0% sequence identity and 99.5% structural identity, and at the active site, they have 31 and 10 identical and similar residues, respectively.

### *3.2.1 The AacGH1 active site*

The active site represents the principal surface cavity of the enzyme, and electron density corresponding to Bis-Tris Propane, present as the buffer in the crystallization condition, was observed here for each *Aac*GH1 chain (Figs. 1 & S3). Based on comparisons made with the white clover β-glucosidase, *Aac*GH1 contains the active site residue E355 (E397 in β-glucosidase) housed in the conserved I(V)TENG motif, typical of glycosyl hydrolase family 1 members. In other members, E355 is responsible for the nucleophilic attack on the substrate (Wang et al., 2003). A salt bridge formed with R78 and hydrogen bonds with Y297 and a conserved active site water molecule, ensure it is deprotonated (Fig. S3). A second glutamate residue, E166 (E183 in β-glucosidase), is known to be the proton donor in the reaction. E166 is also housed in a conserved motif, although instead of the canonical LNEP motif, in *Aac*GH1 we find HNEP.

## *3.3 Kinetic parameters and enzymatic characterization*

The activity assay was performed under standard conditions showing a specific activity of 20 U/mg. The kinetic parameters of the β -glucosidase activity, determined spectrophotometrically using a *p*-nitrophenyl-β-D-glucopyranoside (*p*NPG) concentration range from 0 to 20 mM under optimal conditions (Copeland, 2000), $K_m$ (0.36 mM), $K_i$ (31 mM) and $V_{max}$ (0.0135 mM/min) were calculated in triplicate by nonlinear regression (Fig. S4).

A thorough characterization of *Aac*GH1was performed, to probe a range of different activity and stability assays in the presence of sugars, solvents, different pHs, and different temperatures.

### *3.3.1 Activity assays*

Glucose inhibition is a common problem among β-GH1 enzymes (Chan et al., 2016; Pang et al., 2017; Delgado et al., 2020), as the accumulation of the product of the hydrolytic process naturally reduces

1 the catalytic efficiency of the enzyme. Consequently, a broad range of glucose concentrations (0.1-
2 25% *w/v*) was tested (Fig. 2A), showing that the enzyme activity decreases as the concentration of
3 glucose increases. However, this decrease is not as pronounced as in other cases reported (Sabel et
4 al., 2014; Martino et al., 2000) and *Aac*GH1 maintains 50% of its activity in the presence of 25%
5 (*w/v*) glucose in the reaction. The majority of β-glucosidases present inhibition constants to glucose
6 between 0.5 mM and 100 mM, corresponding to 0.009% and 1.8% (*w/v*), respectively (Liu et al.,
7 2011). However, some β-glucosidases belonging to the family GH1 can be exceptionally tolerant or
8 even stimulated by glucose, but the mechanism of that tolerance/stimulation is still being investigated
9 (Cao et al., 2018a; Costa et al., 2019; Yang et al., 2015).

10 Interestingly, fructose does not exhibit an inhibitory effect like glucose but, on the contrary, it
11 enhances the activity of *Aac*GH1 towards the glucoconjugate. This phenomenon has been reported
12 previously with some GH1 β-glucosidases (Cao et al., 2018b), where enzymatic activity increases
13 with increasing concentrations of fructose (5-25% *w/v*), up to 1.5-fold of the standard activity at 25%
14 (*w/v*). The sugar content in cured vanilla beans is reported to range from 70 to 200 g/Kg (7-20%
15 *(w/w)*), with glucose and fructose being the main constituents (Ramachandra Rao and Ravishankar,
16 2000), well within the tolerated range of *Aac*GH1.

17 Enzymes from mesophilic organisms often show a significant loss of activity in the presence of
18 organic solvents. An explanation which has been suggested focusses on the loss of critical water
19 molecules which are essential to maintain the correct protein conformation (desolvation), affecting
20 the kinetic values, and, in some cases, the overall protein folding. Retained activity in the presence
21 of organic solvents is possible only when the surface and the active site remain well hydrated
22 (Alsafadi and Paradisi, 2013). In general, *Aac*GH1 presents good tolerance to the organic solvents
23 tested, apart from acetonitrile (Fig. 2B), with a marked loss of activity when increasing the
24 concentration of some solvents (methanol, DMSO, THF) from 10 to 20% (*v/v*). Particularly important
25 is its tolerance of ethanol, which is one of the most common solvents used in the food industry. In

this case, the enzyme retained activity is not affected by the presence of both 10 and 20% (v/v) ethanol in the mixture, indicating a high tolerance to ethanol by *Aac*GH1, making it attractive for the treatment of ethanolic vanilla extracts.

### *3.3.2 Stability assays*

Enzyme long term stability is also a key parameter to evaluate the possible implementation of the enzyme in biotechnological processes.

*Aac*GH1 shows a very stable behaviour when incubated for 48 h in the presence of methanol, ethanol, isopropanol, and DMSO (Fig. 3). Not surprisingly, following the activity tests mentioned above, the enzyme is less stable in the presence of 10% (*v/v*) acetonitrile and quickly loses activity with THF or 20% (*v/v*) acetonitrile.

Regarding pH stability (Fig. S5), *Aac*GH1 maintains >50% activity between pH 6 and 10 for 48 h at 25 ºC. At pH 5 and pH 11 the enzyme is less stable but retains around 20% and 35% activity respectively after 2h incubation, at pH 4 and 12 the activity falls to 14% and 8% after 2h and at pH 3 the enzyme completely loses activity within 30 min despite being selected from an acidophilic organism. Certainly, the ability of the source microorganism to deal with acidic pH does not always translate to acidic pH requirements of the isolated catalyst. Most adaptation mechanisms developed by acidophiles to survive at low pHs involve very efficient homeostasis which avoids the ingress of protons to the cytoplasm (Stan-Lotter and Fendrihan, 2017). Consequently, the cytoplasmatic enzymes of those organisms do not necessarily deal with acidic conditions and hence are not adapted to it.

The results obtained for enzymatic stability at a different temperature (Fig. S6) show that *Aac*GH1 retains above 50% activity during incubations at up to 45 °C for 48 h, but shows a complete loss of activity after 48 h at 50 °C. At 55 °C, *Aac*GH1 maintains 20% activity after 1 h and is completely inactive after 2 h. While β-glucosidases with much higher stability (retaining activity at >90 °C) have been reported (Dion et al., 1999; Gabelsberger et al., 1993), *Aac*GH1 can be used at moderate

temperatures (30 °C), which is often insufficient to activate the most thermophilic catalysts (Cao et al., 2018a). Additionally, while the β-glucosidase from *Thermotoga maritima* (Gabelsberger et al., 1993) retains 60% activity for 6 h at 95 °C, it loses 50% of activity in the presence of only 4% glucose. Several β-glucosidases from the genus *Alicyclobacillus* with similar stability, co-solvent tolerance, and lack of glucose inhibition have previously been reported (Cao et al., 2018a; Delgado et al., 2020a; Di Lauro et al., 2006).

### *3.4 Activity of AacGH1 with synthetic vanillin*

The hydrolytic activity of *Aac*GH1 (0.01mg/mL) towards synthetic vanillin 4-*O*-β-D-Glucoside (4.5 mM) was assessed (Fig. 4). The progression of the hydrolysis was checked after 10, 20 and 30 min, showing a complete conversion of the glucoside to the aglycon after 30 min.

### *3.5 Performance of Aac*GH1 *with real vanilla extract*

Following the process in Fig. 5, 1.7 mM vanillin glucoside and virtually no free vanillin were extracted from a green vanilla pod sample. In contrast, no glucovanillin was detected in the extract of cured vanilla beans, and vanillin, in its aglycon form, had an average concentration of 2.9 mM. The green vanilla pod extract was treated with *Aac*GH1 (0.5 mg/mL) at 30 °C and the reaction was monitored for 24 h (15 min, 30 min, 1 h, 3 h, and 24 h) by HPLC. The results show that within 15 min incubation time the enzyme hydrolyses the whole amount of glucoside to the aglycon (Fig. 6), while no spontaneous conversion takes place in the absence of the biocatalyst in the period of time. Besides, several additional peaks have been identified (and quantified, see supporting information for details) in the treated extract which are absent in the untreated one, therefore attributable to enzymatic hydrolysis (Fig. S7-9): vanillic acid (0.07 mM), 4-hydroxybenzoic acid (0.08 mM), and 4-hydroxybenzaldehyde (0.17 mM). These structural analogues of vanillin are known to be present in glycosylated form in green vanilla and contribute significantly to the complex aroma of cured vanilla (Klimes, I Lamparsky, 1976). The formation of additional compounds in low concentration was also observed, but they could not be identified.

## 4 Conclusions

The extremozyme *Aac*GH1 was successfully cloned, purified, and crystallised for the first time. A thorough characterization was done to assess its performance under different harsh conditions commonly found in several food industrial processes such as high concentrations of glucose and fructose and a different range of co-solvents, temperatures, and pHs, showing promising results. The enzyme showed good tolerance to glucose and excellent tolerance to fructose, which even stimulates enzymatic activity. Of particular note is the excellent performance of the enzyme in the presence of ethanol, a common solvent of the food industry. *Aac*GH1 also exhibited good stability at neutral pH and fair thermostability, making it a promising candidate to be used in the food industry.

As a particular example, the evaluation of the hydrolytic capacity of *Aac*GH1 towards vanillin glucoside, in its synthetic and natural form, concluded that the enzyme can efficiently hydrolyse the glucoside to its aglycon, vanillin, and hence, it constitutes a promising candidate to be used in the profitable vanilla industry, allowing a natural vanilla flavour to be obtained without the need for the lengthy curing process (2 hours in comparison to half a year, Fig. 6). Additionally, elevated temperatures of 40-50 °C could were not necessary and reaction times between 1 and 72 h (see introduction) were reduced to 15 min. This might be of particular interest for the processing of substandard harvests (such as damaged crops) where the traditional curing process would be economically unfeasible. Future work on the immobilization of *Aac*GH1 may further improve this process, facilitating enzyme recovery and reuse.

## 5 Funding

This work 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]. The structural work was funded by 'Linea 2' from the Università degli Studi di Milano.

## 6 Conflict of interest

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

## 7 Acknowledgements

The authors of this publication would like to thank Mr. Wayne Williams, senior horticulturalist at Birmingham Botanical Gardens, who kindly donated the sample of fresh vanilla used in this work. We also 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 the free access to HPLC.

## 8 Bibliography

Adams, P.D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: A comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D. Biol. Crystallogr. 66, 213–221. https://doi.org/10.1107/S0907444909052925

Alsafadi, D., Paradisi, F., 2013. Effect of organic solvents on the activity and stability of halophilic

alcohol dehydrogenase (ADH2) from Haloferax volcanii. Extremophiles 17, 115–122. https://doi.org/10.1007/s00792-012-0498-0

Anuradha, K., Shyamala, B.N., Naidu, M.M., 2013. Vanilla- Its Science of Cultivation, Curing, Chemistry, and Nutraceutical Properties. Crit. Rev. Food Sci. Nutr. 53, 1250–1276. https://doi.org/10.1080/10408398.2011.563879

Arnold, F.H., 2009. How proteins adapt: Lessons from directed evolution. Cold Spring Harb. Symp. Quant. Biol. 74, 41–46. https://doi.org/10.1101/sqb.2009.74.046

Baqueiro-Peña, I., Guerrero-Beltrán, J.Á., 2017. Vanilla (Vanilla planifolia Andr.), its residues and other industrial by-products for recovering high value flavor molecules: A review. J. Appl. Res. Med. Aromat. Plants 6, 1–9. https://doi.org/10.1016/j.jarmap.2016.10.003

Cao, H., Zhang, Y., Shi, P., Ma, R., Yang, H., Xia, W., Cui, Y., Luo, H., Bai, Y., Yao, B., 2018a. A highly glucose-tolerant GH1 β-glucosidase with greater conversion rate of soybean isoflavones in monogastric animals. J. Ind. Microbiol. Biotechnol. 45, 369–378. https://doi.org/10.1007/s10295-018-2040-6

Cao, H., Zhang, Y., Shi, P., Ma, R., Yang, H., Xia, W., Cui, Y., Luo, H., Bai, Y., Yao, B., 2018b. A highly glucose-tolerant GH1 β- glucosidase with greater conversion rate of soybean isofavones in monogastric animals. J. Ind. Microbiol. Biotechnol. 45, 369–378.

Chan, C.S., Sin, L.L., Chan, K.-G., Shamsir, M.S., Manan, F.A., Sani, R.K., Goh, K.M., 2016. Characterization of a glucose-tolerant β-glucosidase from Anoxybacillus sp. DT3-1. Biotechnol. Biofuels 9, 174. https://doi.org/10.1186/s13068-016-0587-x

Copeland, R.A., 2000. Enzymes. A Practical Introduction to Structure, Mechanism, and Data Analysis, 2nd ed. Wiley-VCH, Inc., New York. https://doi.org/10.1002/0471220639

Costa, L.S.C., Mariano, D.C.B., Rocha, R.E.O., Kraml, J., da Silveira, C.H., Liedl, K.R., De Melo-Minardi, R.C., De Lima, L.H.F., 2019. Molecular dynamics gives new insights into the glucose

tolerance and inhibition mechanisms on β-glucosidases. Molecules 24. https://doi.org/10.3390/molecules24183215

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., Snoeyink, J., Richardson, J.S., Richardson, D.C., 2007. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 35, 375–383. https://doi.org/10.1093/nar/gkm216

Delgado, L., Heckmann, C.M., Di Pisa, F., Gourlay, L., Paradisi, F., 2020a. Release of soybean isoflavones using a β-Glucosidase from Alicyclobacillus herbarius. ChemBioChem. https://doi.org/10.1002/cbic.202000688

Delgado, L., Parker, M., Fisk, I., Paradisi, F., 2020b. Performance of the extremophilic enzyme BglA in the hydrolysis of two aroma glucosides in a range of model and real wines and juices. Food Chem. 323, 126825. https://doi.org/10.1016/j.foodchem.2020.126825

Di Lauro, B., Rossi, M., Moracci, M., 2006. Characterization of a β-glycosidase from the thermoacidophilic bacterium Alicyclobacillus acidocaldarius. Extremophiles. https://doi.org/10.1007/s00792-005-0500-1

Dignum, M.J.W., Kerler, J., Verpoorte, R., 2001. Vanilla production: Technological, chemical, and biosynthetic aspects. Food Rev. Int. 17, 119–120. https://doi.org/10.1081/fri-100000269

Dion, M., Fourage, L., Hallet, J.N., Colas, B., 1999. Cloning and expression of a β-glycosidase gene from Thermus thermophilus. Sequence and biochemical characterization of the encoded enzyme. Glycoconjugate J. https://doi.org/10.1023/A:1006997602727

Elleuche, S., Schröder, C., Sahm, K., Antranikian, G., 2014. Extremozymes-biocatalysts with unique properties from extremophilic microorganisms. Curr. Opin. Biotechnol. 29, 116–123. https://doi.org/10.1016/j.copbio.2014.04.003

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. Acta

Crystallogr. Sect. D. Biol. Crystallogr. 66, 486–501. https://doi.org/10.1107/S0907444910007493

Gabelsberger, J., Liebl, W., Schleifer, K.H., 1993. Purification and properties of recombinant β-glucosidase of the hyperthermophilic bacterium Thermotoga maritima. Appl. Microbiol. Biotechnol. 40, 44–52. https://doi.org/10.1007/BF00170427

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., B.A., 2005. The Proteomics Protocols Handbook. Humana Press.

Guo, B., Amano, Y., Nozaki, K., 2016. Improvements in glucose sensitivity and stability of trichoderma reesei β-glucosidase using site-directed mutagenesis. PLoS ONE. https://doi.org/10.1371/journal.pone.0147301

Heckmann, C.M., Gourlay, L.J., Dominguez, B., Paradisi, F., 2020. An (R)-Selective Transaminase From Thermomyces stellatus: Stabilizing the Tetrameric Form. Front. Bioeng. Biotechnol. 8, 707. https://doi.org/10.3389/fbioe.2020.00707

Hochuli, E., Bannwarth, W., Döbeli, H., Gentz, R., Stüber, D., 1988. Genetic approach to facilitate purification of recombinant proteins with a novel metal chelate adsorbent. Bio/Technology 6, 1321–1325. https://doi.org/10.1038/nbt1188-1321

Holm, L., 2019. Benchmarking fold detection by DaliLite v.5. Bioinformatics 35, 5326'5327.

Kabsch, W., 2010. XDS. Acta Crystallogr. Sect. D Biol. Crystallogr. 66, 125–132. https://doi.org/10.1107/S0907444909047337

Khoyratty, Shahnoo, Kodja, H., Verpoorte, R., 2018. Vanilla flavor production methods: A review. Ind. Crop. Prod. 125, 433–442. https://doi.org/10.1016/j.indcrop.2018.09.028

Klimes, I Lamparsky, D., 1976. Vanilla volatiles: a comprehensive analysis. Int Flavours Food Addit. 272–291.

Laskowski, R.A., Watson, J.D., Thornton, J.M., 2005. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res. 33, W89-93. https://doi.org/10.1093/nar/gki414

Liu, J., Zhang, X., Fang, Z., Fang, W., Peng, H., Xiao, Y., 2011. The 184th residue of β-glucosidase Bgl1B plays an important role in glucose tolerance. J. Biosci. Bioeng. 112, 447–450. https://doi.org/10.1016/j.jbiosc.2011.07.017

Long, F., Vagin, A.A., Young, P., Murshudov, G.N., 2007. BALBES: A molecular-replacement pipeline. Acta Crystallogr. Sect. D. Biol. Crystallogr. 64, 125–132. https://doi.org/10.1107/S0907444907050172

Martino, A., Schiraldi, C., Di Lazzaro, A., Fiume, I., Spagna, G., Pifferi, P.., De Rosa, M., 2000. Improvement of the flavour of Falanghina white wine using a purified glycosidase preparation from Aspergillus niger. Process Biochem. 36, 93–102. https://doi.org/10.1016/S0032-9592(00)00181-3

Matsubara, H., Goto, K., Matsumura, T., Mochida, K., Iwaki, M., Niwa, M., Yamasato, K., 2002. Alicyclobacillus acidiphilus sp. nov., a novel thermo-acidophilic, ω-alicyclic fatty acid-containing bacterium isolated from acidic beverages. Int. J. Syst. Evol. Microbiol. 52, 1681–1685. https://doi.org/10.1099/ijs.0.02169-0

Mohamad, N.R., Marzuki, N.H.C., Buang, N.A., Huyop, F., Wahab, R.A., 2015. An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. Biotechnol. Biotechnol. Equip. 29, 205–220. https://doi.org/10.1080/13102818.2015.1008192

Pang, P., Cao, L. chuang, Liu, Y. huan, Xie, W., Wang, Z., 2017. Structures of a glucose-tolerant β-glucosidase provide insights into its mechanism. J. Struct. Biol. 198, 154–162. https://doi.org/10.1016/j.jsb.2017.02.001

Pardío, V.T., Flores, A., López, K.M., Martínez, D.I., Márquez, O., Waliszewski, K.N., 2018. Effect of endogenous and exogenous enzymatic treatment of green vanilla beans on extraction of vanillin and main aromatic compounds. J. Food Sci. Technol. 55, 2059–2067. https://doi.org/10.1007/s13197-018-3120-3

Perera, C.O., Owen, E., 2010. Effect of tissue disruption by different methods followed by incubation with hydrolyzing enzymes on the production of vanillin from tongan vanilla beans. Food Bioprocess Technol. 3, 49–54. https://doi.org/10.1007/s11947-007-0048-4

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera?A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Ramachandra Rao, S., Ravishankar, G., 2000. Vanilla flavour: production by conventional and biotechnological routes. J. Sci. Food Agric. 80, 289–304. https://doi.org/10.1002/1097-0010(200002)80:3<289::aid-jsfa543>3.0.co;2-2

Rampelotto, P.H., 2013. Extremophiles and extreme environments. Life (Basel, Switzerland) 3, 482–5. https://doi.org/10.3390/life3030482

Reed, C.J., Lewis, H., Trejo, E., Winston, V., Evilia, C., 2013. Protein adaptations in archaeal extremophiles. Archaea 2013. https://doi.org/10.1155/2013/373275

Robert, X., Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. 42, W320–W324. https://doi.org/10.1093/nar/gku316

Ruiz-Terán, F., Perez-Amador, I., López-Munguia, A., 2001. Enzymatic extraction and transformation of glucovanillin to vanillin from vanilla green pods. J. Agric. Food. Chem. 49, 5207–5209. https://doi.org/10.1021/jf010723h

Sabel, A., Martens, S., Petri, A., König, H., Claus, H., 2014. Wickerhamomyces anomalus AS1: A new strain with potential to improve wine aroma. Ann. Microbiol. 64, 483–491.

https://doi.org/10.1007/s13213-013-0678-x

Schwermann, B., Pfau, K., Liliensiek, B., Schleyer, M., Fischer, T., Bakker, E.P., 1994. Purification, Properties and Structural Aspects of a Thermoacidophilic α-Amylase from Alicyclobacillus Acidocaldarius Atcc 27009: Insight into Acidostability of Proteins. Eur. J. Biochem. 226, 981–991. https://doi.org/10.1111/j.1432-1033.1994.00981.x

Singh, G., Verma, A.K., Kumar, V., 2016. Catalytic properties, functional attributes and industrial applications of β-glucosidases. 3 Biotech. https://doi.org/10.1007/s13205-015-0328-z

Sinha, A.K., Sharma, U.K., Sharma, N., 2008. A comprehensive review on vanilla flavor: Extraction, isolation and quantification of vanillin and others constituents. Int. J. Food Sci. Nutr. 59, 299–326. https://doi.org/10.1080/09687630701539350

Sørensen, A., Lübeck, M., Lübeck, P.S., Ahring, B.K., 2013. Fungal beta-glucosidases: A bottleneck in industrial use of lignocellulosic materials. Biomolecules 3, 612–631. https://doi.org/10.3390/biom3030612

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., Murshudov, G.N., 2004. REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. Sect. D. Biol. Crystallogr. 60, 2184–2195. https://doi.org/10.1107/S0907444904023510

Vasella, A., Davies, G.J., Böhm, M., 2002. Glycosidase mechanisms. Curr. Opin. Chem. Biol. 6, 619–629. https://doi.org/10.1016/S1367-5931(02)00380-0

Ventura-Sobrevilla, J., Boone-Villa, D., Rodriguez, R., Martinez-Hernandez, J.L., Aguilar, C.N., 2015. Microbial biosynthesis of enzymes for food applications, Improving and Tailoring Enzymes for Food Quality and Functionality. Elsevier Ltd. https://doi.org/10.1016/B978-1-78242-285-3.00004-1

Waliszewski, K.N., Ovando, S.L., Pardio, V.T., 2007. Effect of hydration and enzymatic

pretreatment of vanilla beans on the kinetics of vanillin extraction. J. Food Eng. 78, 1267–1273. https://doi.org/10.1016/j.jfoodeng.2006.01.029

Wang, X., He, X., Yang, S., An, X., Chang, W., Liang, D., 2003. Structural basis for thermostability of β-glycosidase from the thermophilic eubacterium Thermus nonproteolyticus HG102. J. Bacteriol. 185, 4248–4255. https://doi.org/10.1128/JB.185.14.4248-4255.2003

Yang, Y., Zhang, Xinxin, Yin, Q., Fang, W., Fang, Z., Wang, X., Zhang, Xuecheng, Xiao, Y., 2015. A mechanism of glucose tolerance and stimulation of GH1 β-glucosidases. Sci. Rep. 5, 17296. https://doi.org/10.1038/srep17296

Zhang, Y., Mo, L., Chen, F., Lu, M., Dong, W., Wang, Q., Xu, F., Gu, F., 2014. Optimized production of vanillin from green vanilla pods by enzyme-assisted extraction combined with pre-freezing and thawing. Molecules 19, 2181–2198. https://doi.org/10.3390/molecules19022181

Figure 1

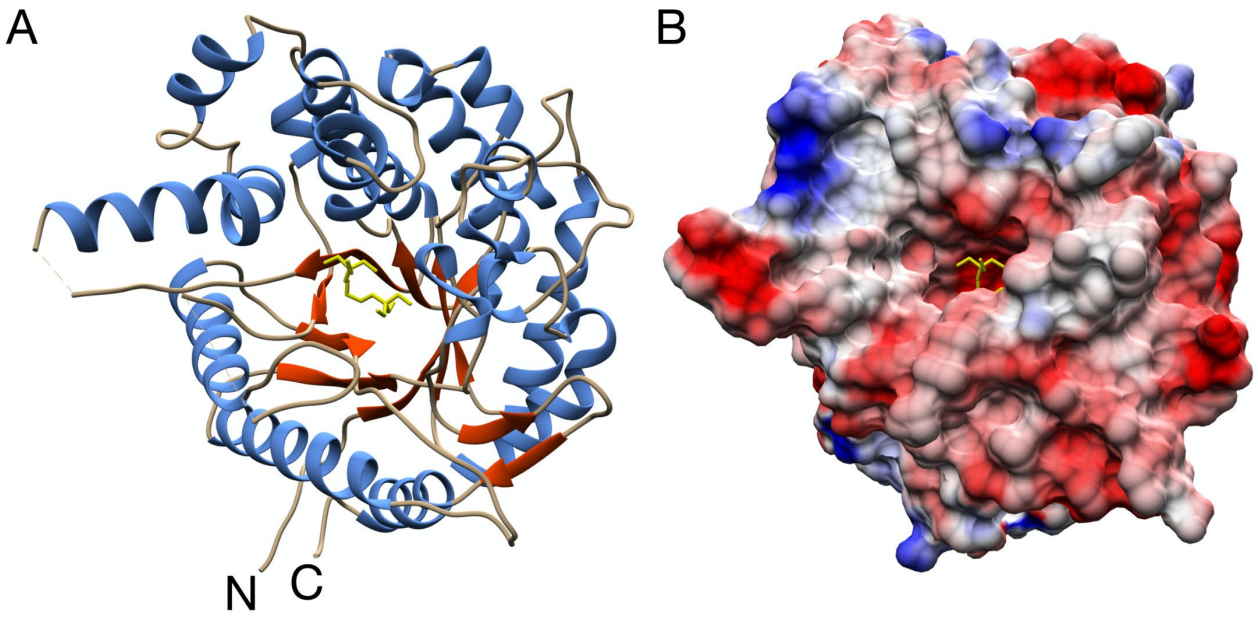



Figure 2

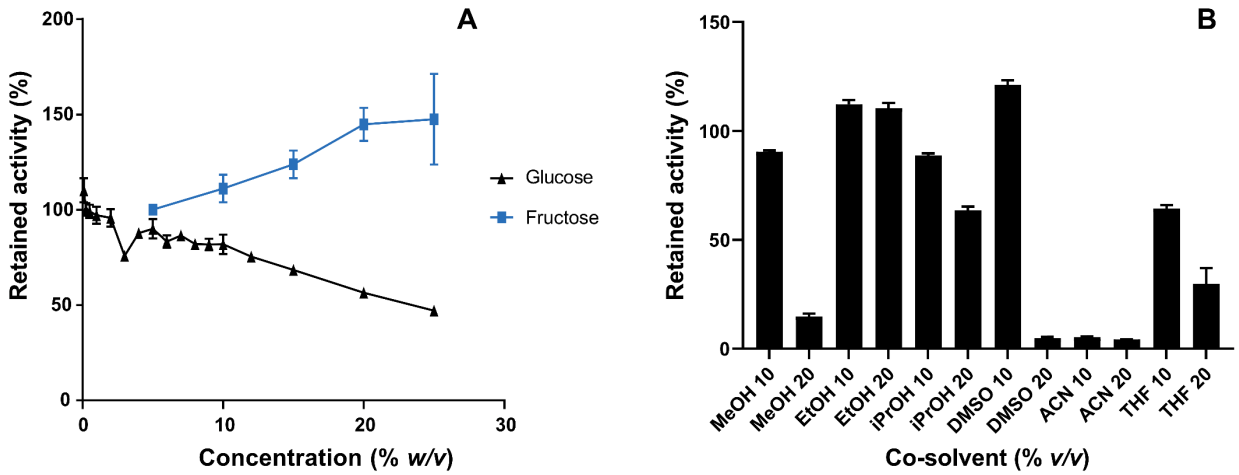


Figure 3

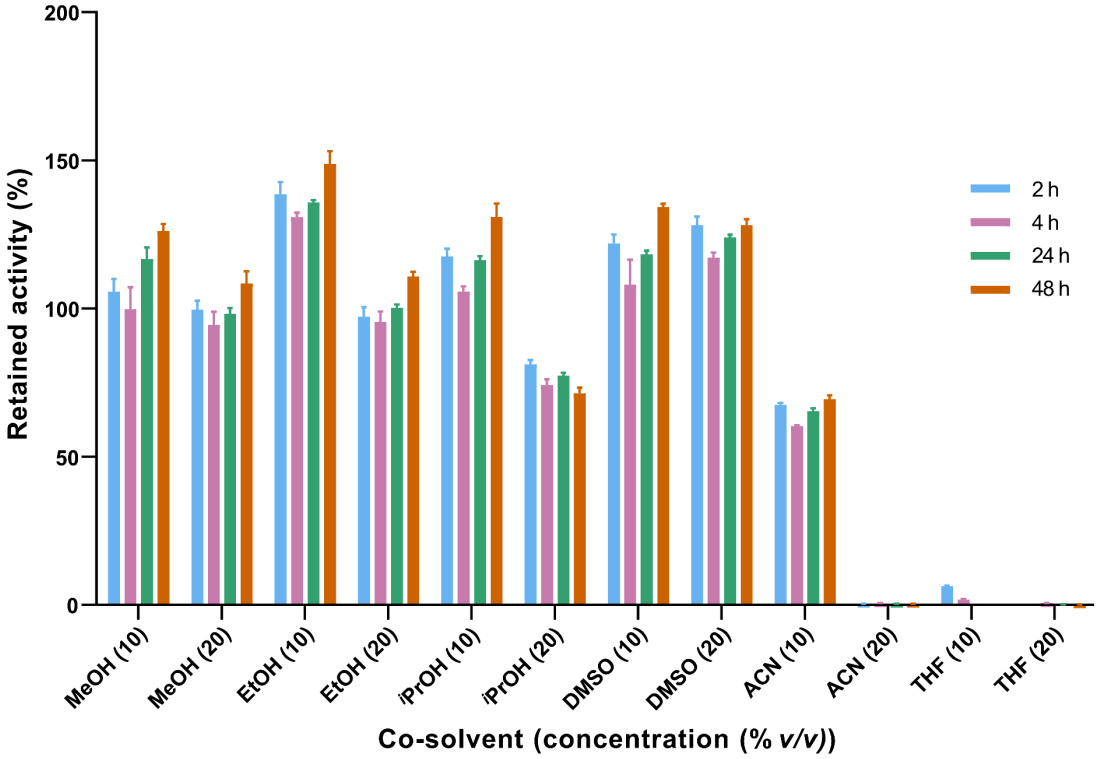



Figure 4

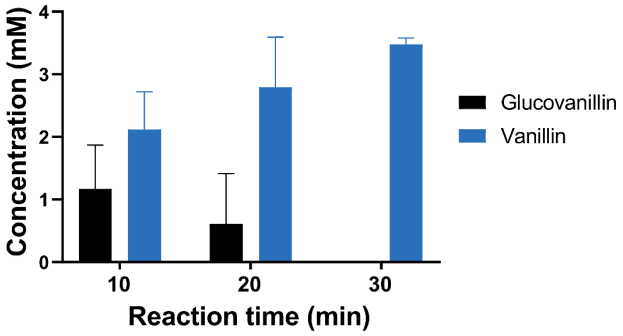



Figure 5

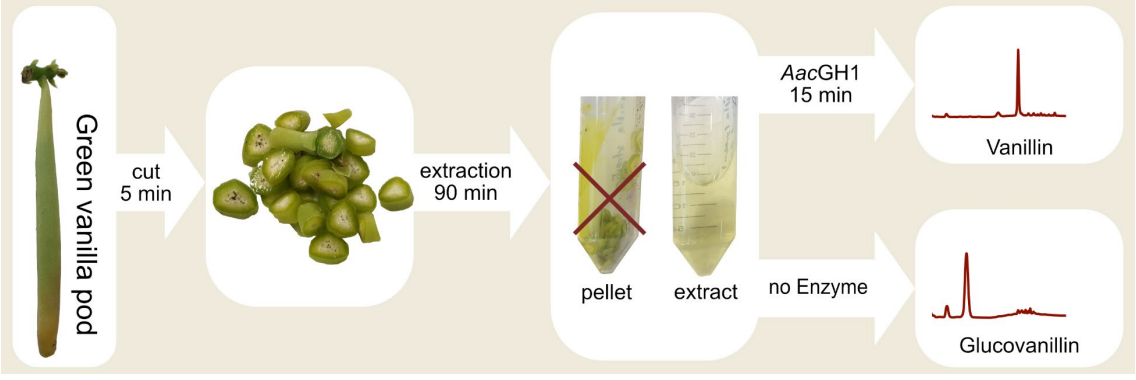
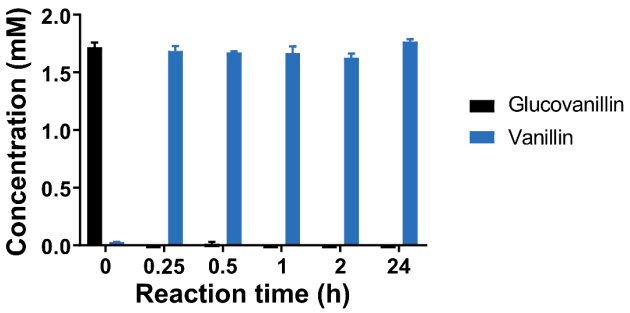


Figure 6


Concentration (mM)
2.0
1.5
1.0
0.5
0.0
0
0.25
0.5
1
2
24
Reaction time (h)
Glucovanillin
Vanillin